Titre: Title:	Advanced Surgical Planning of Adolescent Idiopathic Scoliosis Instrumentation Combining Artificial Intelligence and Physics-Based Model Simulation
Auteur: Author:	Caroline Constant
Date:	2025
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Constant, C. (2025). Advanced Surgical Planning of Adolescent Idiopathic Scoliosis Instrumentation Combining Artificial Intelligence and Physics-Based Model Simulation [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/70987/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/70987/
Directeurs de recherche: Advisors:	Carl-Éric Aubin, & Noelle Larson
Programme: Program:	Génie biomédical

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Advanced surgical planning of adolescent idiopathic scoliosis instrumentation combining artificial intelligence and physics-based model simulation

CAROLINE CONSTANT

Institut de génie biomédical

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*

Génie biomédical

Novembre 2025

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

Advanced surgical planning of adolescent idiopathic scoliosis instrumentation combining artificial intelligence and physics-based model simulation

présentée par **Caroline CONSTANT**

en vue de l'obtention du diplôme de *Philosophiæ Doctor*

a été dûment acceptée par le jury d'examen constitué de :

Farida CHERIET, présidente

Carl-Éric AUBIN, membre et directeur de recherche

Noelle LARSON, membre et codirectrice de recherche

Jeremy RAWLINSON, membre

Brice ILHARREBORDE, membre

Paul SPONSELLER, membre externe

DEDICATION

“I'm sure I can pass this test

*Cause I'm a real tough kid, I can handle my sh***

They said, "Babe, you gotta fake it 'til you make it" and I did”

Honorary Doctor Swift

ACKNOWLEDGEMENTS

This is the hardest part of my thesis to write, but also the one that matters the most. I often thought the day I would write these words would never come. Then it did, together with the hardest moment of my life: when my mom passed away. Looking back at the mountains climbed and valleys crossed, I know my courage came from those around me. That is why I begin with her.

Maman, tu m'as toujours dit que je pouvais faire n'importe quoi si je le voulais vraiment. Je t'ai répondu que toutes les mères disent cela, puis tu m'as dit: « Oui, mais pour toi, c'est vrai.» Well, here is the proof and all the work that came with it! Thank you for showing me courage and perseverance, but most of all, for teaching me to care for others. As you devoted your life to helping those in need through your work as a nurse, I hope that the work in this thesis will, in its own way, continue that legacy by helping children who need it.

To **Emma** and **Matteo André**, you are what truly matters. Emma, your patience, encouragement, and endless support have carried me through every challenge of this journey. Matteo André, your laughter and light made every effort worthwhile. Everything I do, I do for you both.

To **Carl-Eric Aubin**, thank you for everything (and I really mean everything). I still don't know how you manage to do it all. Your combination of intelligence, ambition, compassion, and patience made you the best director I could have hoped for. I couldn't have done it without you (obviously)!

To **Dr. Noelle Larson**, thank you for being such a strong role model: an inspiring surgeon, scientist, and mother. Thank you for welcoming me as your student despite your many responsibilities. It has been an honor to learn from you and work together.

To **Dr. André Desrochers**, my gratitude to you is bigger than what words can express. You shaped me as a scientist, clinician and person. "Thank you" feels far too small, but I hope my work makes you proud.

To **Nora Goudsouzian**, thank you for pushing me when I thought I couldn't do it. On long nights, seeing the light in your office reminded me I wasn't alone. That mattered more than you know.

To **Stephan Zeiter**, thank you for always believing in me and supporting our sometimes crazy projects. People say it takes a village to raise someone; if so, you were certainly one of the chiefs.

A big thank you to my fellow students and the assistants de recherche at TransMedTech, CHU Sainte-Justine, and Polytechnique for the learning experience, invaluable help, and teamwork. A special mention goes to Christiane Caouette, Nathalie Bourassa, Christian Bellefleur, Xiaoyu Wang, Carolina Solorzano Barrera, Mathieu Chayer, and Maeva Lopez, for their precious help with lab tools, 3D reconstructions of patients, ADAMS simulations, and everything in between that needed debugging when I couldn't do it myself.

I would also like to thank the Orthopedic Surgery Artificial Intelligence Laboratory (OSAIL) at the Mayo Clinic for introducing me to the world of artificial intelligence and its application in cutting-edge orthopedic surgery. Special thanks to Hilal Maradit Kremers, Cody C. Wyles, Pouria Rouzrokh, and Michael J. Taunton for their help in shaping the project and guidance in the initiation of our own AI model.

This work was carried out with the generous support of several funding agencies and partners. It was realized within the framework of an NSERC-Alliance grant with Medtronic, initially established as the NSERC/Medtronic Industrial Research Chair in Spine Biomechanics, with special thanks to Jeremy Rawlinson for his continued support and guidance from afar. Additional support was provided by the Quattrone-Foderaro Grant for AI Innovation in Orthopedic Surgery, the TransMedTech Institute and its main financial partner (Canada First Research Excellence Fund and Fonds de Recherche Québec), as well as the Mayo Foundation, the Mayo Clinic, and the AO Foundation. I am also deeply grateful to the TransMedTech Institute for believing in me and my project and for awarding me the TransMedTech Institute Excellence Scholarship. I would also like to thank Polytechnique Montréal for its additional financial support, which has been instrumental in advancing my research.

I'm done!

RÉSUMÉ

La scoliose idiopathique de l'adolescent (SIA) est une déformation tridimensionnelle (3D) de la colonne vertébrale d'origine inconnue, touchant 2 à 4 % des enfants âgés de 10 à 18 ans. Bien que définie par une déviation latérale de la colonne vertébrale (angle de Cobb >10°), la SIA implique également des rotations vertébrales et un désalignement sagittal. Pour les courbures sévères (angle de Cobb >45–50°), l'arthrodèse rachidienne postérieure (ARP) instrumentée par vis pédiculaires constitue le traitement chirurgical de référence. Bien que la planification chirurgicale initiale se concentrait sur l'alignement coronal, l'obtention d'une correction 3D équilibrée est désormais reconnue comme essentielle pour garantir des résultats positifs à long terme et la satisfaction des patients.

La planification chirurgicale préopératoire est cruciale pour déterminer les niveaux de fusion, la répartition des vis et la courbure des tiges. Bien que les technologies d'imagerie 3D soient de plus en plus accessibles, la pratique repose encore principalement sur des radiographies 2D, des règles issues de systèmes de classification et l'expérience du chirurgien. Ces approches n'intègrent pas pleinement l'ensemble des paramètres d'instrumentation et des variables chirurgicales, particulièrement au-delà de la densité des vis et des caractéristiques des tiges, qui influencent pourtant une planification optimale. Cela contribue à une variabilité importante des stratégies chirurgicales utilisées en pratique et à des taux de chirurgie de révision rapportés entre 7% et 13% dans les cinq ans, souvent liés à des complications mécaniques ou à un alignement postopératoire sous-optimal.

Des outils d'assistance informatisés ont été développés pour soutenir la planification chirurgicale de la SIA, notamment pour l'ARP, mais la plupart restent limités à l'analyse géométrique à partir de radiographies et n'intègrent pas le comportement biomécanique de la colonne. Des plateformes commerciales telles que Surgimap, mediCAD et UNiD Hub peuvent aider à l'alignement et à la planification en 2D ou fournir des implants personnalisés, mais elles sont principalement conçues pour les pathologies dégénératives et ne prennent que partiellement en compte la complexité 3D de la SIA. Des outils comme SpineEOS permettent une visualisation 3D du rachis, mais leurs capacités de planification demeurent essentiellement basées sur la géométrie. En revanche, les simulations basées sur la physique, telles que la modélisation multicorps, permettent de prédire de manière réaliste les résultats chirurgicaux et les forces implantaires à partir de l'anatomie

spécifique du patient. Malgré ce potentiel, leur adoption clinique reste limitée en raison de la complexité technique, des besoins computationnels élevés et d'un accès restreint. Parallèlement, l'intelligence artificielle (IA), en particulier l'apprentissage profond (« *deep learning* »), émerge comme un outil précieux pour l'aide à la décision chirurgicale, mais son intégration à la modélisation biomécanique et son application à la planification instrumentée complète de la SIA demeurent rares.

Cette thèse aborde ces défis en posant la question de recherche suivante : *dans quelle mesure une approche hybride, combinant l'IA, un modèle biomécanique déterministe 3D spécifique au patient et un algorithme d'optimisation, peut-elle intégrer avec précision et exactitude les principaux paramètres de l'ARP (vertèbres instrumentées proximale et distale, densité des vis et courbure des tiges) afin d'optimiser la planification et la prédiction de la correction de la SIA en fonction des caractéristiques individuelles des patients ?* Pour y répondre, l'objectif général était de développer un outil intégrant l'IA à la simulation déterministe afin d'assister la planification chirurgicale préopératoire de l'ARP dans la SIA. Les objectifs spécifiques étaient: (1) développer un réseau neuronal multitâche (RNMT) pour prédire les principaux paramètres d'instrumentation; (2) intégrer ce modèle à une simulation biomécanique afin d'évaluer la correction spinale et les forces implantaires; (3) coupler le système avec un algorithme d'optimisation pour affiner les choix des instrumentations; et (4) évaluer la crédibilité et l'incertitude du modèle développé selon les lignes directrices internationales établies (ASME V&V40:2018).

La première étape a consisté en une revue de la littérature portant sur l'utilisation de l'apprentissage profond en imagerie rachidienne. En suivant la méthodologie PRISMA, 365 études ont été classées en quatre catégories d'application clinique: diagnostic, aide à la décision, évaluation de l'instrumentation et prédiction des résultats cliniques. Bien que les applications de l'apprentissage profond aient connu une croissance rapide dans les dernières années, en particulier pour le diagnostic, seulement 15 % des modèles ont été validés de manière externe, et peu concernaient la planification chirurgicale ou la biomécanique 3D. Cela a confirmé une lacune critique dans la littérature et ouvert la voie au développement d'outils de planification plus complets pour la chirurgie de la SIA, basés sur l'IA.

La deuxième partie du travail a consisté à développer un RNMT entraîné sur les données de 179 patients atteints de SIA inclus dans l'essai clinique MIMO (Lenke 1 et 2). Le modèle a été conçu pour prédire les vertèbres instrumentées proximale (UIV) et distale (LIV), la courbure des tiges et la densité des vis à partir des données cliniques et radiographiques préopératoires. Pour ces prédictions, le modèle a atteint une exactitude UIV/LIV de 82–95 % (pour les 2 meilleures prédictions) et des erreurs quadratiques moyennes de 0.2–0.3 pour la densité des vis et de 3.7–5.6° pour la courbure des tiges. Une validation externe, utilisant des données de 10 patients issues de deux hôpitaux indépendants, a confirmé la généralisabilité du modèle. Ces résultats soulignent la faisabilité de l'utilisation de l'apprentissage profond pour reproduire les décisions de planification de chirurgiens experts.

Dans la troisième étude, les constructions instrumentées prédites par le RNMT ont été simulées biomécaniquement et comparées aux plans réalisés par les chirurgiens chez 35 patients atteints de SIA, à l'aide de modèles multicorps rachidiens validés, spécifiques aux patients. Alors que les chirurgiens obtenaient une meilleure correction dans le plan coronal (différence moyenne de Cobb: 5.6°, $p < 0,001$), les constructions dérivées de l'IA restauraient plus efficacement la cyphose thoracique (32.7° vs 29.3°, $p = 0.008$) et, dans 77 % des cas, égalaient ou surpassaient la correction 3D globale. Fait notable, ces constructions issues de l'IA nécessitaient moins de vis et des longueurs de fusion plus courtes, sans augmentation des forces d'arrachement des vis. Ces résultats ont confirmé la faisabilité biomécanique d'une planification informée par l'IA. Néanmoins, les tentatives d'affiner ces prédictions par l'IA en ajustant les cibles de correction chirurgicale restaient limitées à une sélection statique parmi des options pré-générées, sans mise en œuvre d'une véritable optimisation ni intégration complète des considérations biomécaniques.

La quatrième étude a perfectionné le modèle hybride en intégrant une optimisation multi-objectifs. Pour 20 patients, les plans prédits par l'IA ont été affinés à l'aide d'un algorithme d'optimisation équilibrant correction 3D, forces d'arrachement des vis et préservation de la mobilité. Chaque patient a fait l'objet de 972 simulations, soit un total de 19 440 pour la cohorte. En moyenne, les plans optimisés utilisaient 4 vis de moins et fusionnaient deux niveaux de moins ($p < 0.001$) que les constructions réalisées par les chirurgiens, tout en égalant ou surpassant celles-ci dans tous les paramètres de correction 3D. Plus précisément, les stratégies optimisées amélioraient la rotation axiale vertébrale moyenne de 2.8° ($p = 0.05$) et l'angle de Cobb thoracique principal de 3° ($p = 0.019$), tout en restaurant la cyphose thoracique dans des plages cliniquement acceptables, similaires aux plans chirurgicaux (29.2° ± 6.4° vs 27.4° ± 8.9°, $p = 0.372$). Toutes les configurations maintenaient des charges biomécaniques acceptables, démontrant le potentiel du modèle hybride à fournir des stratégies d'instrumentation plus sûres et plus efficientes.

Afin de garantir que l'outil numérique développé puisse être considéré comme fiable et adapté à une future translation clinique, un processus structuré de vérification, validation et quantification de l'incertitude (VVUQ) a été mené. Le contexte d'utilisation du modèle a été défini comme un outil d'aide à la décision pour la planification de l'ARP dans la SIA, fournissant des recommandations d'instrumentation tout en laissant le jugement final au chirurgien. Selon la norme ASME V&V 40, ce rôle correspond à une influence moyenne du modèle et à une faible conséquence décisionnelle, aboutissant à une classification de risque moyen–faible qui a guidé le choix des activités de VVUQ. Les études de vérification ont confirmé la stabilité du solveur numérique, les perturbations de paramètres entraînant des variations négligeables de l'alignement de la colonne et des forces exercées sur les implants. La validation, réalisée sur 35 chirurgies réelles de SIA, a montré une concordance étroite entre les résultats simulés et les résultats postopératoires observés, avec des écarts moyens systématiquement inférieurs à ±5°. Des analyses de sensibilité menées chez 20 patients supplémentaires ont confirmé que le choix du niveau de fusion avait l'impact le plus important sur la correction et les forces exercées sur les implants, que la courbure des tiges influençait principalement l'alignement sagittal, et que la distribution des vis modifiait les forces sans compromettre la correction. Ces résultats démontrent que l'outil numérique développé satisfait aux attentes de crédibilité associées à sa classe de risque et qu'il est aligné sur les standards internationaux, renforçant à la fois sa rigueur méthodologique et sa préparation à une éventuelle translation clinique.

Pris ensemble, ces quatre travaux et les activités de VVUQ soulignent la faisabilité et les bénéfices potentiels de l'intégration de l'IA, de la modélisation biomécanique déterministe et de l'optimisation au sein d'un modèle numérique unifié de planification chirurgicale pour la SIA. Sur le plan méthodologique, cette thèse fait progresser le domaine en combinant la prédiction par réseau neuronal de type RNMT, la biomécanique multicorps et l'optimisation dans un cadre crédible et orienté vers un usage clinique futur. Elle intègre également des processus de validation et de quantification de l'incertitude alignés sur la norme ASME V&V 40, afin de renforcer la crédibilité du modèle et de son cadre d'utilisation, avec des implications potentielles pour une préparation réglementaire future. Les limites de ce travail incluent la nature rétrospective de l'ensemble de données (principalement des patientes Lenke 1/2, de sexe féminin, de type caucasien), ainsi que le fait que toutes les simulations demeurent *in silico.* Une validation prospective supplémentaire, accompagnée du développement d'interfaces et de leur intégration dans le flux clinique, reste essentielle.

En conclusion, cette thèse jette les bases d'un outil de planification préopératoire de nouvelle génération, capable de soutenir une prise de décision chirurgicale plus transparente, individualisée et informée par la biomécanique. Avec un développement continu, une validation rigoureuse et une collaboration interdisciplinaire soutenue, de tels outils pourraient améliorer à la fois la qualité et la constance des soins offerts aux adolescents subissant une arthrodèse rachidienne postérieure.

ABSTRACT

Adolescent idiopathic scoliosis (AIS) is a three-dimensional (3D) spinal deformity of unknown cause, affecting 2–4% of children aged 10 to 18. Although defined by a lateral deviation of the spine (Cobb angle >10°), AIS also involves vertebral rotation and sagittal misalignment. For severe curves (Cobb angle >45–50°), posterior spinal fusion (PSF) with pedicle screw instrumentation is the surgical standard. Although early surgical planning prioritized coronal alignment, achieving balanced 3D correction is now recognized as critical to long-term outcomes and patient satisfaction.

Preoperative surgical planning is essential to determining fusion levels, screw distribution, and rod contouring. Although 3D imaging technologies are becoming increasingly available, current practice still relies predominantly on 2D radiographs, rule-based classification systems, and surgeon experience. Such approaches may not fully integrate the range of instrumentation parameters and surgical variables, beyond screw density and rod characteristics, that influence optimal planning. This can contribute to considerable variability in surgical strategies and to revision surgery rates, which are reported to be between 7% and 13% within five years, often related to mechanical complications or suboptimal postoperative alignment.

Computer-assisted tools have been developed to support surgical planning in AIS, including PSF, but most remain limited to geometric radiographic-based analysis and do not incorporate the biomechanical behavior of the spine. Commercial platforms such as Surgimap, mediCAD, and UNiD Hub can assist with 2D alignment and planning or provide patient-specific implants, yet they are primarily designed for degenerative pathologies and only partially address the 3D complexity of AIS. Tools like SpineEOS enable 3D visualization, yet their planning capabilities remain predominantly geometry-based. In contrast, physics-based simulations such as multibody (MB) modeling can realistically predict surgical outcomes and implant forces using patient-specific anatomy. Despite their potential, clinical adoption has been hindered by technical complexity, high computational demands, and limited accessibility. Meanwhile, artificial intelligence (AI), particularly deep learning, is emerging as a valuable aid for surgical decision-making, but its integration with biomechanical modeling and its application to comprehensive instrumentation planning in AIS remain uncommon.

This thesis addresses these challenges by posing the following research question: *To what extent can a hybrid approach, combining AI, a deterministic 3D biomechanical patient-specific model,*

and an optimization algorithm, accurately and precisely integrate key PSF parameters (upper and lower instrumented vertebrae, screw density, and rod curvature) to optimize the planning and prediction of AIS correction based on individual patient characteristics? To address this question, the general objective was to develop a tool that integrates AI with deterministic model simulation to assist PSF preoperative surgical planning in AIS. Specific aims were: (1) to develop a neural network model for predicting key instrumentation parameters; (2) to integrate this model with biomechanical simulation for assessing spinal correction and implant forces; (3) to couple the system with an optimization algorithm to refine construct choices; and (4) to evaluate the framework's credibility and uncertainty following established international guidelines (ASME V&V40:2018).

The first investigation was a scoping review of deep learning and spine imaging. Using PRISMA methodology, 365 studies were categorized into four clinical applications: diagnostics, decision support, instrumentation assessment, and outcome prediction. While deep learning applications have grown rapidly, especially for diagnostics, only 15% of models have been externally validated, and few have addressed surgical planning or 3D biomechanics. This confirmed a critical gap in the literature and set the stage for the development of more comprehensive planning tools for AIS surgery using AI.

The second part of this work involved developing a multitask neural network (NNML) trained on data from 179 AIS patients enrolled in the MIMO clinical trial (Lenke 1 and 2). The NNML was designed to predict upper (UIV) and lower instrumented vertebrae (LIV), rod curvature, and screw density using preoperative clinical and radiographic data. For these predictions, the model achieved UIV/LIV accuracy of 82–95% (top-2 predictions) and root mean square errors of 0.2–0.3 for screw density and 3.7–5.6° for rod curvature. External validation using 10 patients from 2 independent hospitals confirmed the model's generalizability. These findings highlight the feasibility of using deep learning to emulate expert planning decisions.

In the third study, instrumentation constructs predicted by the NNML were biomechanically simulated and compared with surgeon-performed plans in 35 AIS patients, using validated patient-specific multibody spine models. While surgeons achieved greater coronal plane correction (mean Cobb difference: 5.6°, $p < 0.001$), AI-derived constructs restored thoracic kyphosis more effectively (32.7° vs. 29.3°, $p = .008$) and, in 77% of cases, matched or exceeded the overall 3D correction. Notably, these AI-derived constructs required fewer screws and shorter fusion lengths, without increasing implant loads. These findings confirmed the biomechanical feasibility of AI-informed planning. Nevertheless, attempts to refine these AI predictions through modified surgical correction targets remained limited to static selection among pre-generated options, without implementing true optimization or fully integrating biomechanical considerations.

The fourth study advanced the hybrid framework by incorporating multi-objective optimization. For 20 patients, the AI-predicted plans were refined using an optimization algorithm that balanced 3D correction, screw pullout force, and motion preservation. Each patient underwent 972 simulations, totaling 19,440 across the cohort. On average, optimized plans used 4 fewer screws and fused up to two fewer levels ($p < 0.001$) compared to surgeon-performed constructs, with scenarios equaling or surpassing them in all key correction metrics. Specifically, optimized strategies improved mean axial vertebral rotation by 2.8° ($p = .05$), main thoracic Cobb angle by 3° ($p = .019$), while also restoring thoracic kyphosis within acceptable clinical ranges, similar to surgeon-performed plans (29.2° ± 6.4° vs. 27.4° ± 8.9°, $p = .372$). All configurations maintained acceptable biomechanical loads, demonstrating the full pipeline's potential to yield safer and more efficient instrumentation strategies.

To ensure that the developed framework could be considered reliable and suitable for future clinical translation, a structured verification, validation, and uncertainty quantification (VVUQ) process was conducted. The framework's context of use was defined as a decision-support tool for PSF planning in AIS, providing recommendations on instrumentation while leaving final judgment to the surgeon. According to ASME V&V 40, this role corresponds to medium model influence and low decision consequence, resulting in a medium–low overall risk classification that informed the selection of VVUQ activities. Performed verification studies confirmed solver stability, with parameter perturbations producing negligible changes in alignment and implant forces. Validation against 35 real AIS surgeries showed close agreement between simulated and actual postoperative outcomes, with mean differences consistently within ±5°. Sensitivity analyses in 20 additional patients confirmed that fusion level selection had the most significant effect on spinal correction and implant loading, rod curvature primarily influenced sagittal alignment, and screw distribution altered forces without compromising correction. These findings demonstrate that the framework meets the credibility expectations for its risk class and is aligned with international standards, supporting both methodological rigor and translational readiness.

Collectively, these four studies and the performed VVUQ activities underscore the feasibility and potential benefits of integrating AI, deterministic biomechanical modeling, optimization, and simulation-based reasoning into a unified framework for surgical planning in AIS. Methodologically, this thesis advances the field by integrating neural network prediction, multibody biomechanics, and optimization within a credible, clinically oriented framework. It also integrates validation and uncertainty quantification processes aligned with ASME V&V 40 to comprehensively support model and framework credibility, with potential implications for regulatory readiness. Limitations include the retrospective nature of the dataset (primarily Lenke 1/2, female, white patients), and the fact that all simulations remain *in silico*. Further prospective validation, along with interface development and workflow integration, will be essential.

In conclusion, this thesis lays the groundwork for a next-generation planning tool that could support more transparent, individualized, and biomechanically informed surgical decision-making. With continued development, rigorous validation, and sustained interdisciplinary collaboration, such tools have the potential to improve both the quality and consistency of care for adolescents undergoing spinal fusion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
ADAM	Adaptive Moment Estimation
AI	Artificial intelligence
AIS	Adolescent idiopathic scoliosis
ALARA	“As low as reasonably achievable” principle
ANOVA	Analysis of variance
ASME	American Society of Mechanical Engineers
ATR	Angle of trunk rotation
AVR	Apical vertebral rotation
C7PL	C7 plumb line
CE	Categorical cross-entropy
CI	Confidence interval
CLAIM	Checklist for Artificial Intelligence in Medical Imaging
CM&S	Computational modeling and simulation
CNN	Convolutional neural network
CoCr	Cobalt-chrome
COU	Specific Context of Use
CSVL	Center sacral vertical line
CT	Computed tomography
DEXA	Dual-energy x-ray absorptiometry
DICOM	Digital Imaging and Communications in Medicine

DL	Deep learning
EHRs	Electronic health records
FDA	U.S. Food and Drug Administration
FE	Finite element
HRQOL	Health-related quality of life scores
ISO	International Organization for Standardization
LIV	Lower instrumented vertebra
LSD	Least significant difference
LSTM	Long Short-Term Memory
MANOVA	Multivariate analysis of variance
MB	Multibody modeling
MDD	Medical Devices Directorate
MIMO	Minimize Implants Maximize Outcomes
ML	Machine learning
MLP	Multilayer Perceptron
MRI	Magnetic resonance imaging
MSE	Mean Squared Error
MT	Main thoracic
NCV	Numerical code verification
NN	Neural network
NNML	Neural network multitask learning
NSE	Numerical solver error
ORPD	Orientation of the Regional Plane of Deformation
PA	Posteroanterior

PACS	Picture Archiving and Communication Systems
PMC	Plane of maximum curvature
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSF	Posterior spinal instrumentation and fusion
PT	Proximal thoracic
ReLU	Rectified Linear Unit
RNN	Recurrent neural network
SaMD	Software as a Medical Device
SD	Standard deviation
SMS	Sanders maturity scale
SQA	Software quality assurance
SRS	Scoliosis Research Society
SS	Stainless steel
ST-NN	Single-task neural networks
SVA	Sagittal vertical axis
Ti	Titanium alloys
TL/L	Thoracolumbar/lumbar
TPD	Therapeutic Products Directorate
UHSS	Ultrahigh-strength stainless steel
UIV	Upper instrumented vertebra
US	Ultrasound
VV	Verification and validation
VVUQ	Verification, validation, and uncertainty quantification
η^2	Partial eta squared

β	Weighting coefficients for the classification loss functions
α	Weighting coefficients for the regression loss functions

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

Idiopathic scoliosis is a complex, three-dimensional (3D) spinal deformity of unknown cause. It is defined by a lateral deviation in the coronal plane (Cobb angle >10°), but also involves segmental deviation and vertebral rotation in the transverse plane, as well as alterations in sagittal alignment [1]. Adolescent idiopathic scoliosis (AIS), the most common form of scoliosis, affects approximately 2–4% of individuals between the ages of 10 and 18 [2-4]. For many years, clinical evaluation and treatment of AIS focused mainly on the coronal curvature [5, 6]. Nevertheless, it is now widely recognized that AIS is not merely a lateral deviation, but a multidimensional condition. Additionally, coronal curve severity is often associated with significant axial and sagittal plane abnormalities [7-11].

Severe AIS can lead to long-term physical impairment, pain, and pulmonary dysfunction [12], necessitating surgical intervention when the Cobb angle exceeds 45–50° [13, 14]. Approximately 1% of AIS patients undergo surgery, totaling 15,000 cases annually in North America [15]. The current standard surgical treatment for AIS cases is posterior spinal instrumentation and fusion (PSF) with pedicle screws [16, 17]. The primary goal is realigning the vertebrae and achieving a balanced and stable fusion [18]. Typically, this involves multisegmental pedicle screws connected to dual rods over the length of the intended fusion for stability, along with bone graft placement to promote bony fusion. Although surgical correction historically prioritized coronal alignment, it is now clear that restoring overall 3D spinal balance leads to better functional outcomes and patient satisfaction [19, 20]. While modern pedicle screw constructs have improved the capacity for 3D correction [8, 21, 22], no single surgical technique has emerged as universally optimal. As a result, surgical approaches remain diverse, reflecting the complexity of the condition and the ongoing search for the most effective strategy.

Preoperative planning plays a critical role in scoliosis surgery, guiding decisions on fusion levels, implant placement, and rod selection to correct the patient's deformity effectively. Given the many factors influencing AIS correction, a well-structured plan is essential for achieving optimal postoperative outcomes. Classification systems, particularly the Lenke classification, support standardization of surgical decisions and enable more consistent outcome comparisons [6, 18]. One key step in planning is determining the scoliotic curve(s) to address and the appropriate fusion levels. The goal is to limit the number of fused vertebrae to preserve mobility, while ensuring the fusion is long enough to prevent malalignment or curve progression of the uninstrumented curves [23, 24]. The next consideration is the number and distribution of pedicle screws; sufficient anchor points are needed to perform effective correction maneuvers, but excessive screw malalignment increases implant stress [25]. Rod characteristics (diameter, shape, and material) are equally important and should be carefully chosen, as an inappropriate rod can lead to mechanical complications [26-28]. Stiff rods effectively maintain spinal alignment, but overly rigid constructs may contribute to adjacent segment disease [29].

Traditionally, AIS planning relies on two-dimensional (2D) radiographs, which are mainly descriptive assessments of sagittal alignment [6, 18]. While this approach simplifies evaluation, it fails to capture the 3D nature of AIS fully [30-32]. Furthermore, despite the knowledge that postoperative 3D spinal alignment significantly impacts patient outcomes [19, 33, 34], most planning methods still rely heavily on 2D parameters, limiting the ability to predict complex behavior of the spine in 3D. Currently available software, such as Surgimap (Nemaris, Inc., New York, USA) and mediCAD (Hectec GmbH, Landshut, Germany), enables preoperative simulations [35, 36] but remains primarily 2D-based. As such, they are better suited for degenerative spinal conditions than complex 3D deformities like AIS [33, 37]. Moreover, these tools focus exclusively on radiographic-based geometric correction, without incorporating biomechanical considerations. UNiD Hub (Medicrea, Lyon, France) introduces patient-specific rods, enabling customized implants, but also relies on geometric factors without explicitly accounting for biomechanical properties [37]. SpineEOS (EOS Imaging, Paris, France) offers 3D planning capabilities and deformity correction assessment of AIS [37], yet its decision-making process remains mostly confined to geometric considerations.

In contrast, physics-based computational modeling, particularly finite element and multibody (MB) methods, has emerged as a promising tool for evaluating instrumentation strategies using 3D patient-specific spine models [25, 38-41]. MB modeling, in particular, allows realistic simulation of surgical maneuvers and provides a non-invasive way to assess the biomechanical impact of spinal instrumentation prior to surgery [42]. It has shown promising accuracy, predicting postoperative alignment with an average Cobb angle deviation of only 1.2° compared to actual outcomes [43]. Despite these advantages, clinical implementation remains limited due to the complexity of patient-specific modeling, regulatory issues, and the high computational demands, with optimization sometimes requiring up to 36 hours per patient [42, 44].

To overcome these limitations, machine learning (ML), a branch of artificial intelligence (AI), is increasingly explored to enhance preoperative decision-making by leveraging large clinical datasets [45]. Notably, Lafage et al. developed a deep learning (DL) model that accurately identified the upper instrumented vertebra (UIV) in 87.5% of adult scoliosis cases, highlighting the potential of AI-assisted surgical planning. Despite these technological advances, revision surgery remains a significant concern, with 7–13% of AIS patients requiring reoperation within five years, primarily due to suboptimal patient-specific deformity correction or postoperative malalignment [46, 47]. Mechanical complications, such as rod breakage, are also common causes of revision, and they are influenced by instrumentation factors like rod alloy and diameter [48-50]. "Surgeon-modifiable factors" play a central role in PSF outcomes and postoperative spinal alignment, including upper and lower instrumented vertebra (LIV) selection [51-57], screw density and pattern [40, 58, 59], and patient-specific rod contouring [60, 61]. Postoperative malalignment is strongly associated with decreased health-related quality of life (HRQOL), pain, spinal decompensation, and degeneration of adjacent segments [62-66]. Restoring sagittal alignment, in particular, has been linked to reduced rod fracture risk and improved pulmonary function [67-69].

These observations underscore the importance of thorough preoperative surgical planning and thoughtful instrumentation selection to achieve a well-balanced spine and optimal outcomes. While classification systems provide a valuable framework for decision-making [6, 18], postoperative curve progression and compensatory curve development can still occur, even in surgeries that follow guideline-based classifications [16, 47, 53, 70, 71]. Moreover, many key instrumentation parameters, such as screw density, are not considered by these systems and remain largely reliant on the surgeon's experience [72-74]. These problems may partly explain the high variability in instrumentation strategies and the prevalence of so-called "rule breakers," as recently reported in a multicenter study [75].

Altogether, these findings highlight the need for a more comprehensive, patient-specific approach to PSF instrumentation strategies to improve surgical outcomes. Given the high inter-individual variability in AIS presentation, tailoring surgical constructs to each patient’s anatomy and biomechanics may reduce the risk of poor outcomes. Although not yet widely implemented, self-learning algorithms and AI-based planning tools hold promise in supporting surgeons with more precise, patient-specific preoperative planning, potentially improving correction prediction and long-term results. Building on these considerations, the overarching objective of this doctoral project was to develop and evaluate a hybrid planning framework that combines artificial intelligence, patient-specific biomechanical simulation, and optimization to guide key instrumentation parameters in AIS surgery (UIV, LIV, screw density, rod curvature). This framework was further subjected to verification, validation, and uncertainty quantification to ensure its credibility for potential clinical translation.

CHAPTER 2 REVIEW OF KNOWLEDGE

2.1 Descriptive and Functional Anatomy of the Asymptomatic Spine

The spine (or vertebral column) is a central structure of the human body, composed of vertebrae, muscles, intervertebral discs, and ligaments. It provides trunk mobility, transmits loads from the trunk and head to the pelvis, and protects the spinal cord. This section explores the descriptive and functional anatomy of the spine, focusing on its structural and biomechanical characteristics.

2.1.1 Descriptive Anatomy of the Spine

To understand the spine's function, it is essential to describe its major components. The spine consists of rigid structures (vertebrae) that support the upper body and flexible elements (intervertebral discs, muscles, and ligaments) that enable movement and provide stability. In tetrapods, the spine serves four key functions: (1) protecting the spinal cord, (2) supporting body weight, (3) allowing trunk motion, and (4) serving as an attachment site for muscles and limbs [76].

An anatomical study of the spine uses three orthogonal planes: frontal (coronal), sagittal (lateral), and transverse (axial), as illustrated in Figure 2.1A. In the coronal plane, a healthy spine appears straight, while in the sagittal plane, it exhibits an S-shaped curvature, with cervical lordosis, thoracic kyphosis, lumbar lordosis, and sacral kyphosis (Figure 2.1B). These natural curves are biomechanically optimized to efficiently distribute mechanical loads and maintain an upright posture with minimal energy expenditure [77, 78].

The spine is divided into five regions, each with unique structural and functional characteristics. The cervical spine comprises seven vertebrae (C1–C7) and connects the head to the trunk, allowing significant motion. The thoracic spine includes twelve vertebrae (T1–T12), which provide stability and support the rib cage. The lumbar spine, composed of five vertebrae (L1–L5), connects the trunk to the pelvis and supports the most significant load. Finally, the sacrum and coccyx form the base of the spine, with the sacrum comprising five fused vertebrae and the coccyx including three to five atrophied vertebrae [77].

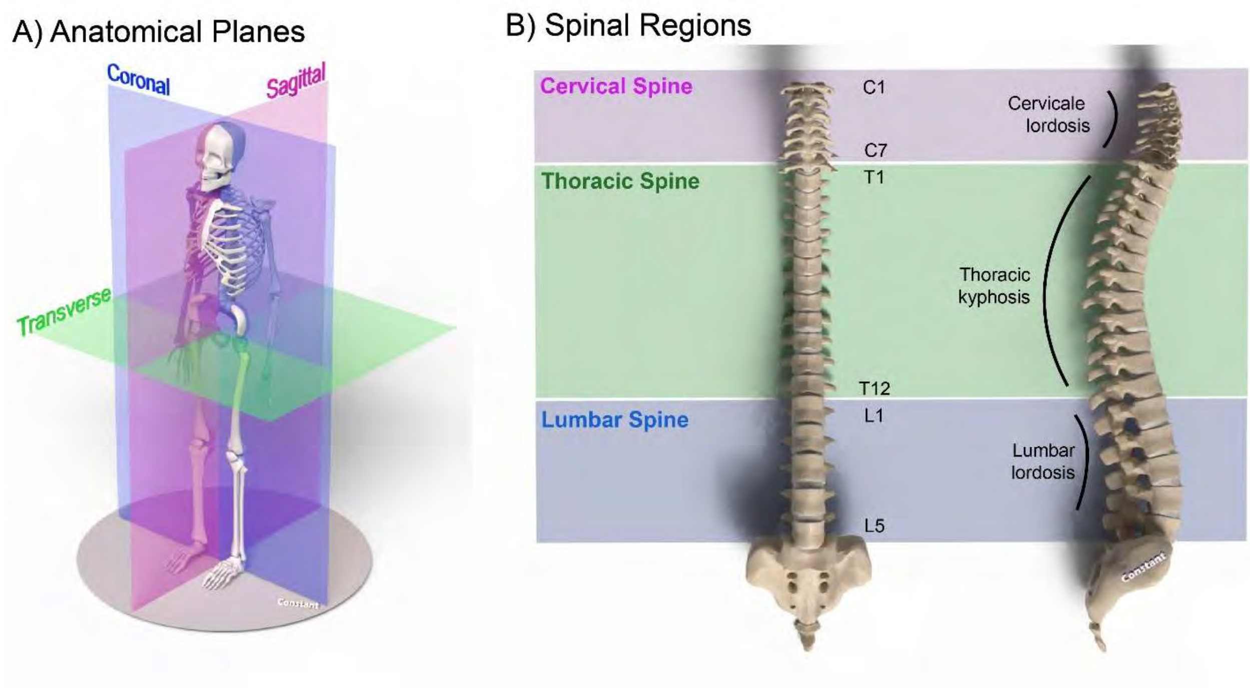


Figure 2.1 Anatomical Drawings of the Spine

A) the principal anatomical planes, and B) the spinal segments and their natural curvatures.

2.1.1.1 Vertebrae

Understanding the unique anatomy of vertebrae is crucial for surgical planning. Vertebrae, composed of cortical bone (outer layer) and trabecular bone (interior), vary in size and function across regions. Due to greater load demands, vertebral dimensions increase from the cervical to the lumbar spine.

Each vertebra follows a consistent structural pattern (Figure 2.2). The anterior vertebral body serves as the primary weight-bearing portion, while the posterior vertebral arch encloses the vertebral foramen, forming a protective passage for the spinal cord. Various processes extend from the vertebral arch, which play structural and functional roles. The spinous process projects posteriorly, providing an attachment site for muscles and ligaments, while the transverse processes extend laterally to serve as additional muscle attachment points. Superior and inferior articular processes form facet joints that connect adjacent vertebrae, influencing motion and stability [77].

Distinct anatomical adaptations are observed in different spinal regions. Cervical vertebrae exhibit unique features that allow a wide range of motion in the neck. The atlas (C1) and axis (C2) are specialized for head movement, with the atlas supporting the skull and the axis providing a pivot point for rotation. Thoracic vertebrae, in contrast, are less mobile due to their articulations with the ribs, which enhance stability and contribute to the characteristic kyphotic curvature of this region (Figure 2.2A). Lumbar vertebrae, the largest and strongest, lack transverse foramina and costal facets, reflecting their role in supporting significant axial loads while permitting greater flexion-extension movements (Figure 2.2B) [77, 79, 80].

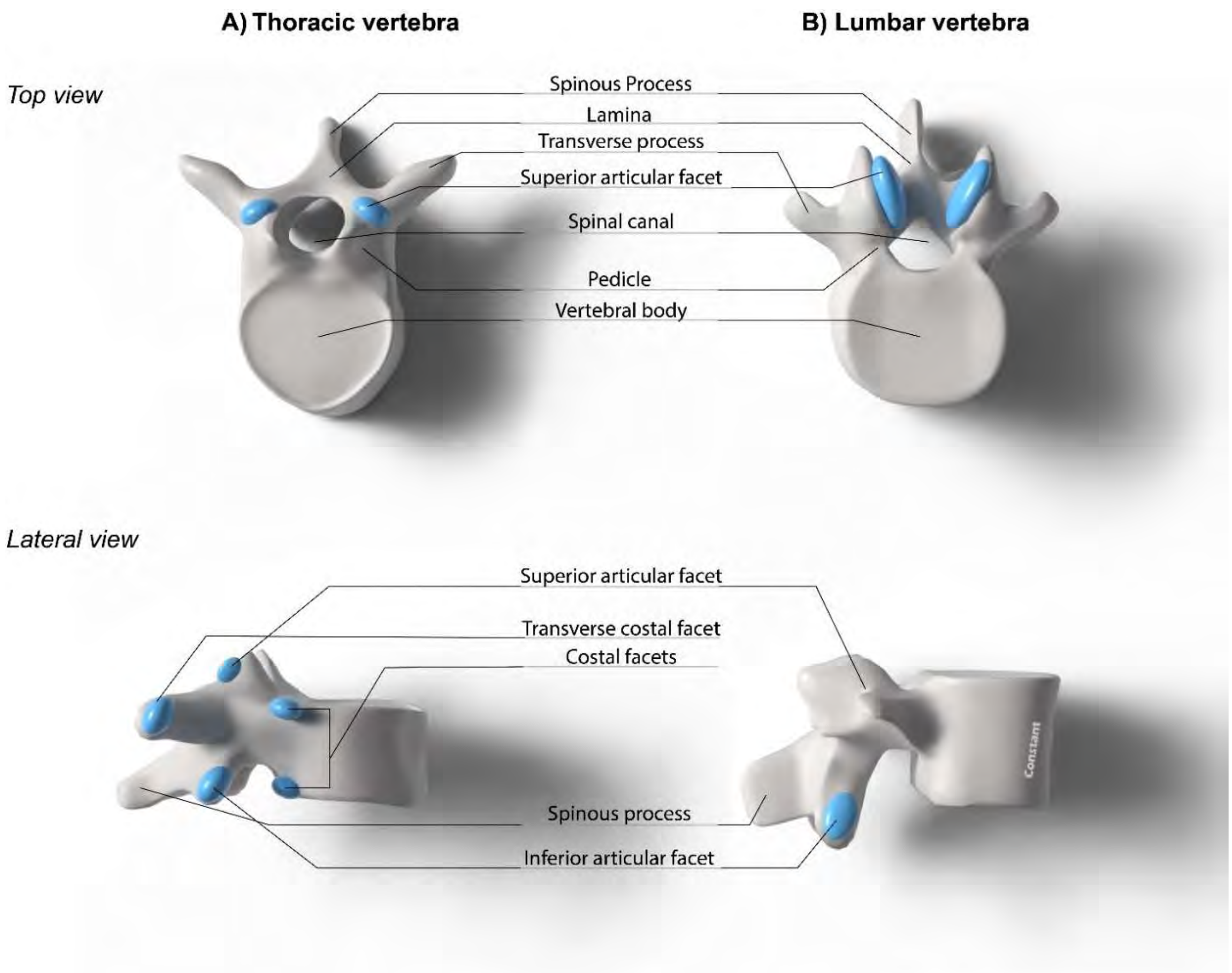


Figure 2.2 Anatomical Drawings of Vertebrae

A) Thoracic vertebra, and B) lumbar vertebra, from top and lateral views.

2.1.1.2 Intervertebral Discs and Ligaments

The structural integrity of the vertebral column depends on both vertebrae and supporting structures like intervertebral discs and ligaments that connect adjacent vertebrae. Intervertebral discs are fibrocartilaginous structures located between vertebral bodies. Each disc consists of an outer annulus fibrosus, a tough fibrous ring that anchors the disc to adjacent vertebrae, and an inner nucleus pulposus, a gel-like core that absorbs compressive forces. Discs account for approximately 25% of the spine's height, with thickness varying by region to accommodate different load demands, with the lumbar discs being the thickest due to higher load demands. Aging or spinal degeneration can lead to a loss of water content in the nucleus pulposus, reducing disc height and flexibility [77].

Ligaments further reinforce the spine and ensure structural stability. The anterior and posterior longitudinal ligaments run along the vertebral column, stabilizing it during movement. The ligamentum flavum connects adjacent laminae, while the interspinous and supraspinous ligaments link spinous processes. These structures work together to limit excessive motion and maintain alignment, ensuring that the spine remains stable under mechanical stress [81].

2.1.1.3 Pelvis

The pelvis serves as the structural connection between the spine and the lower limbs. It consists of the sacrum, coccyx, and two hip bones. The sacrum, composed of five fused vertebrae, articulates with the lumbar spine via an intervertebral disc. The coccyx, located at the tip of the sacrum, comprises three to five atrophied vertebrae. Each hip bone is formed by the fusion of the ilium, ischium, and pubis by the age of 12. These bones are joined at the pubic symphysis and sacroiliac joints, which provide rigidity but allow for limited movement during activities such as childbirth [77].

2.1.2 Biomechanics of the Spine

The biomechanics of the spine examines the relationship between its anatomical structures and the mechanical principles that govern its function. Advances in engineering have enabled the development of numerical models for *in silico* studies of load distribution, which are challenging to measure *in vivo*. Spinal mechanics are influenced by the interactions between vertebrae, intervertebral discs, ligaments, and muscles, which together ensure both stability and mobility under various mechanical loads. Entire books have been written on spine biomechanics and surgical instrumentation [82, 83]. Therefore, this section will focus on general principles of the thoracic spine that address anatomically and surgically relevant issues in the treatment of AIS.

2.1.2.1 Range of Motion of the Spine

The amplitude of spinal movement varies across regions and individual vertebral segments due to differences in the anatomical characteristics of bony structures and surrounding soft tissues, as discussed earlier. The thoracic region is the stiffest section of the spine, with motion constrained by the rib cage and facet joint orientations. Flexion and extension amplitudes increase progressively from the thoracic to the cervical spine, while rotational motion decreases caudally along the thoracic spine and becomes minimal in the lumbar region [84].

Thoracic spinal stability can be categorized into overall, segmental, and vertebral stability. *Overall stability*, mainly corresponding to external stability, results from the combined contributions of the rib cage, vertebral morphology, and intervertebral discs. These structures enhance load-bearing capacity and limit excessive rotational movements throughout the thoracic region [84]. The rib cage plays a pivotal role as an external stabilizer, resisting rotational forces and supporting the thoracic spine [80, 81]. *Segmental stability* depends on the coordinated interaction of adjacent vertebrae, intervertebral discs, and ligamentous structures such as the facet joint capsules, ligamentum flavum, and anterior and posterior longitudinal ligaments. These elements collectively restrict excessive flexion, extension, and lateral bending, ensuring localized stability. *Vertebral stability*, another intrinsic component, is determined by the material properties and architecture of individual vertebrae, which contribute to their load bearing and resistance capabilities.

The thoracic spine is stiffest under axial compression, with moderate flexibility in flexion and axial rotation compared to extension and lateral bending. Average stiffness values are reported as approximately 1240 N/mm for axial compression and 2.6–3.2 Nm/° for torsional and bending motions [84, 85]. Resistance to anteroposterior translation is primarily governed by facet joint orientation, with intervertebral discs and the rib cage providing additional support under axial compression [86]. In the transitional cervicothoracic zone, the sagittal plane angle of facet joints decreases, while the transverse angle increases superiorly [86]. These anatomical changes contribute to greater flexion and extension amplitudes in the cervical spine compared to the thoracic region, reflecting its increased range of motion and flexibility [84]. Conversely, in the thoracolumbar transitional zone (T11 to L1), facet joint orientation shifts from a more frontal plane alignment in the thoracic spine to a more sagittal alignment in the lumbar spine [86]. This shift decreases rotational amplitude in the lumbar spine, favoring flexion-extension over transverse rotation [84].

2.1.2.2 Loading of the Spine

Spinal loading is influenced by body position, movement, and anatomical structures. In a standing posture, the body's center of gravity maintains axial compression and postural balance. Movements that shift the center of gravity increase vertebral and disc loads, requiring greater muscle activation and stabilization [86]. The thoracic spine's load distribution is shaped by its characteristic kyphosis, which typically measures around 45° but ranges from 20° to 70° in asymptomatic individuals [87]. The mean apex of this curvature is generally located near T6 [88]. The kyphotic alignment shifts the center of gravity anteriorly, creating a lever arm that requires continuous activation of posterior muscles, such as the longissimus dorsi, to maintain an upright stance [86].

The kyphosis directs compressive forces primarily toward the anterior vertebral bodies, while posterior structures resist tensile forces [86]. The combined contributions of the rib cage, vertebral bodies, and intervertebral discs further enhance the thoracic spine's load-bearing capacity. The rib cage increases stiffness in all directions, redistributing forces evenly across the thoracic region and contributing to external stability [89, 90]. Vertebral bodies, designed to bear axial loads, achieve their strength and stiffness through the interplay of cortical and trabecular bone. Their compressive strength and stability increase caudally, correlating with higher bone mineral density [91]. Intervertebral discs play a crucial role in distributing loads by absorbing compressive forces between adjacent vertebrae. The thoracic spine, excluding the rib cage, has been shown to resist a compressive failure force of approximately 2 kN, with compressive stiffness of 300 N/mm and an energy-absorbing capacity in compression of 10 Nm [92]. These structural properties underline the thoracic spine's resilience and adaptability to functional demands.

The essential distinction between humans and other vertebrates' spine biomechanics does not lie in the basic architecture of the spine, which is relatively consistent across species, but in the unique alignment and function of the human spine. Humans possess a fully upright sagittal spinopelvic configuration, characterized by a lordotic curve originating from the pelvis, enabling bipedal locomotion and an upright posture. This arrangement aligns the body's center of gravity directly above the pelvis, optimizing balance. However, this unique human feature introduces dorsally directed shear forces in certain spine regions. These shear forces have been associated with a reduction in rotational stiffness in the affected segments [93-96]. While the spine predominantly experiences axial loads, the spatial orientation of individual vertebrae determines whether they are subjected to additional shear forces. Depending on their position, vertebrae may experience anteriorly or posteriorly directed shear loads resulting from gravitational forces and muscle activity (Figure 2.3) [94, 96].

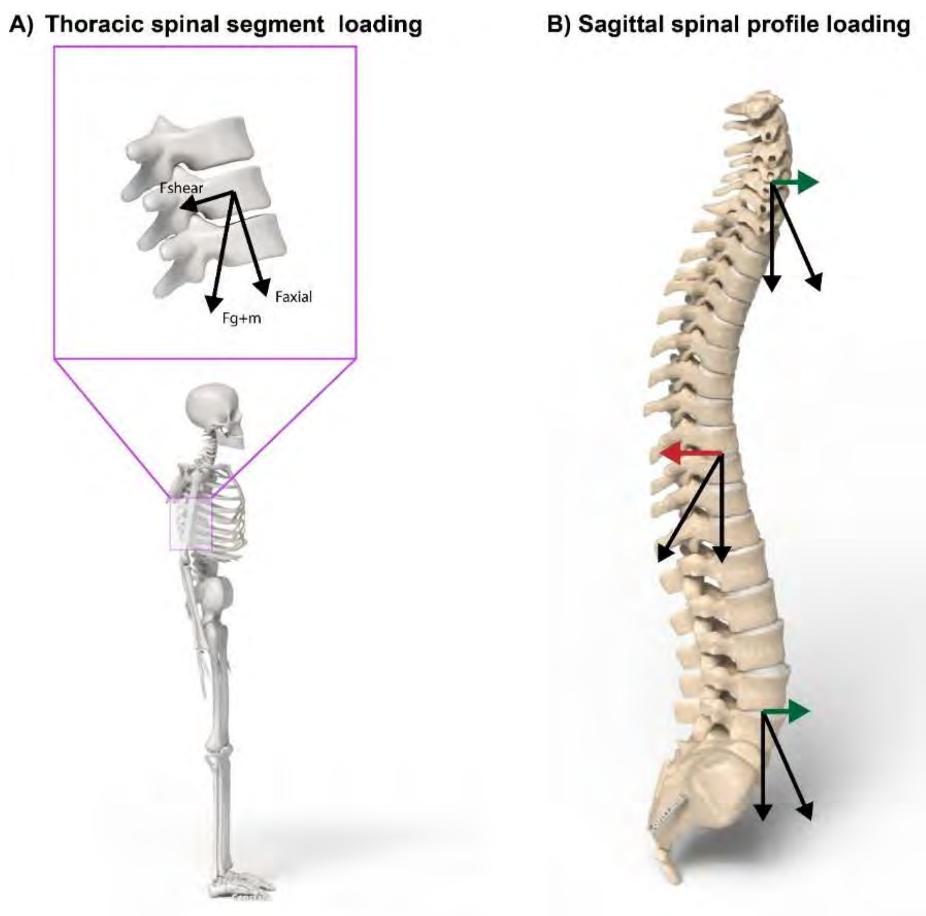


Figure 2.3 Loading of the Upright Human Spine

A) Thoracic spinal segment depicting axial loading (Faxial) and dorsal shear forces (Fshear) induced by the combined effect of gravitational and muscular forces (Fg+m).

B) Sagittal spinal profile illustrating regions subjected to anterior shear (green arrows) and ventral dorsal shear (red arrow), dependent on the individual's sagittal alignment.

2.2 Adolescent Idiopathic Scoliosis

Scoliosis is a broad term encompassing various spinal deformities characterized by changes in the shape and position of the spine, ribs, sternum, thorax, and trunk. More specifically, it is a 3D torsional deformity of the spine, involving a lateral curvature in the coronal plane, axial rotation of the vertebrae in the transverse plane, and disturbances in the natural kyphosis and lordosis of the sagittal plane [7-9]. According to the Scoliosis Research Society (SRS), scoliosis is diagnosed when a patient has a lateral curvature that exhibits a Cobb angle of at least 10° (Figure 2.4) and some degree of axial rotation of the involved vertebrae [14]. Scoliosis can affect regional anatomy, resulting in smaller pedicles on the side of the concavity, particularly in established cases. Further alterations due to scoliosis can result in restricted spine motion, early arthritis changes, and regional variability in bone mineral density.

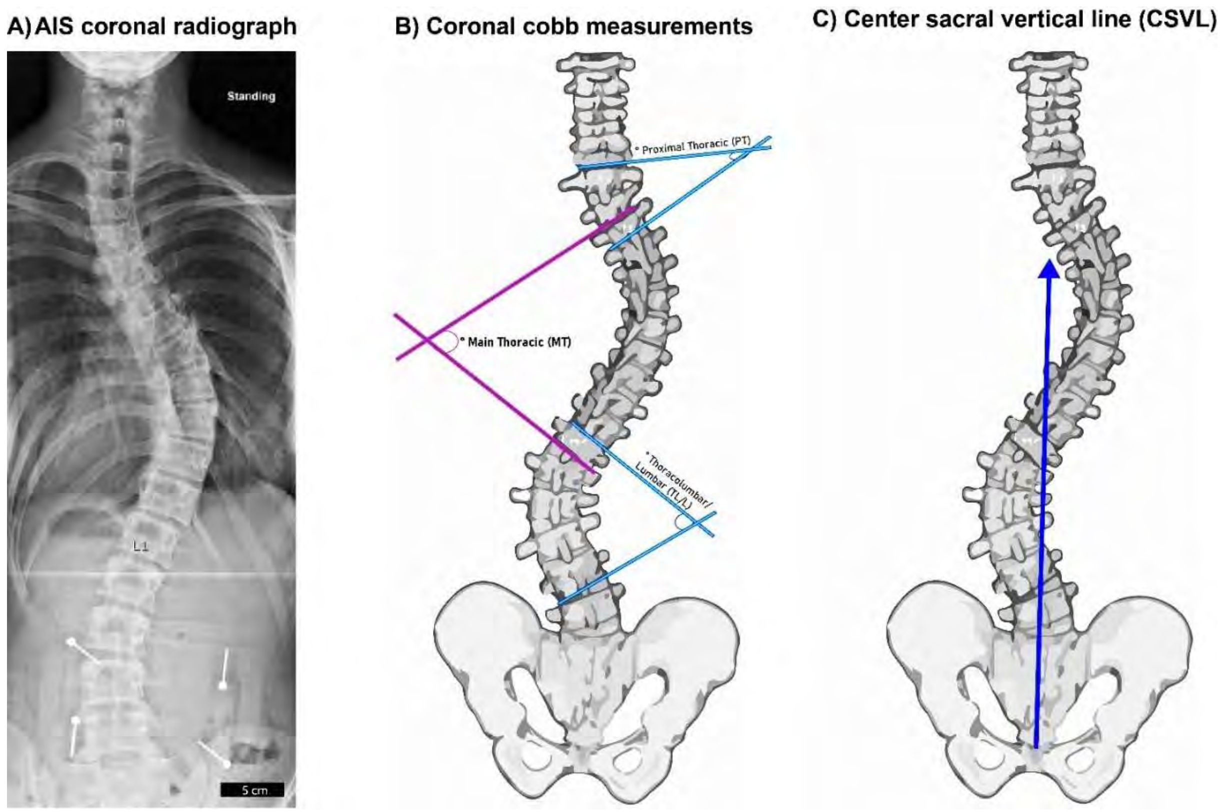


Figure 2.4 Radiographic Features of Adolescent Idiopathic Scoliosis

A) Representative radiograph demonstrating lateral curvature in the coronal plane.
B) Coronal Cobb angle measurement, used to quantify the magnitude of the coronal deformity
C) Central sacral vertical line (CSVL), a vertical line bisecting the sacrum on standing radiographs, illustrating the coronal balance between the pelvis and spine.

When no specific underlying cause, such as congenital malformations, neuropathic conditions, or myopathic disorders, can be identified, the term idiopathic scoliosis is applied [97]. The Greek term "idiopathic" implies that the disease is not linked to any physical impairment or previous medical history. Therefore, by definition, idiopathic scoliosis has an unknown etiology. However, current research suggests it may be a clinical manifestation of a syndrome with a multifactorial origin, involving rotational stability of the spine and genetics, as well as metabolic, neurogenic, and environmental factors [14, 93-95, 98]. Severe AIS cases may result in long-term complications, including physical impairment, pain, and pulmonary dysfunction [12]. The degree of deformity, curve rigidity, and apex location determine progression risks and associated morbidity. Early diagnosis and classification are essential for optimizing outcomes and reducing long-term consequences.

2.2.1 AIS Clinical Work-Up and Evaluation

The clinical work-up of AIS involves a structured approach combining physical examination and imaging to evaluate spinal deformity, assess progression risk, and guide treatment decisions. This section outlines the standard evaluation pathway, focusing on radiographic modalities and key parameters such as curve magnitude, spinal alignment, and flexibility relevant to surgical planning.

2.2.1.1 Imaging Modalities for Evaluation

Radiographic imaging is central to AIS evaluation, providing essential data on curve magnitude, spinal balance, skeletal maturity, and surgical planning. Standard posteroanterior (PA; Figure 2.5A) and lateral standing radiographs (Figure 2.5B) are the current standard for assessing scoliosis. These views quantify the coronal and sagittal spine balance and allow for the identification of potential nonidiopathic causes of spinal deformity, such as spondylolisthesis or congenital anomalies. Additional radiographic techniques, including side bending (Figure 2.5C-D), traction, and push-prone radiographs, assess curve flexibility and inform surgical planning by predicting correction potential [99-101]. Recent innovations, such as EOS imaging, have significantly reduced radiation exposure while maintaining high-quality imaging consistent with the "as low as reasonably achievable" (ALARA) principle [102, 103]. Additionally, 3D reconstructions from EOS imaging and patient-specific 3D-printed surgical guides are advancing the visualization and personalization of AIS treatment planning [104].

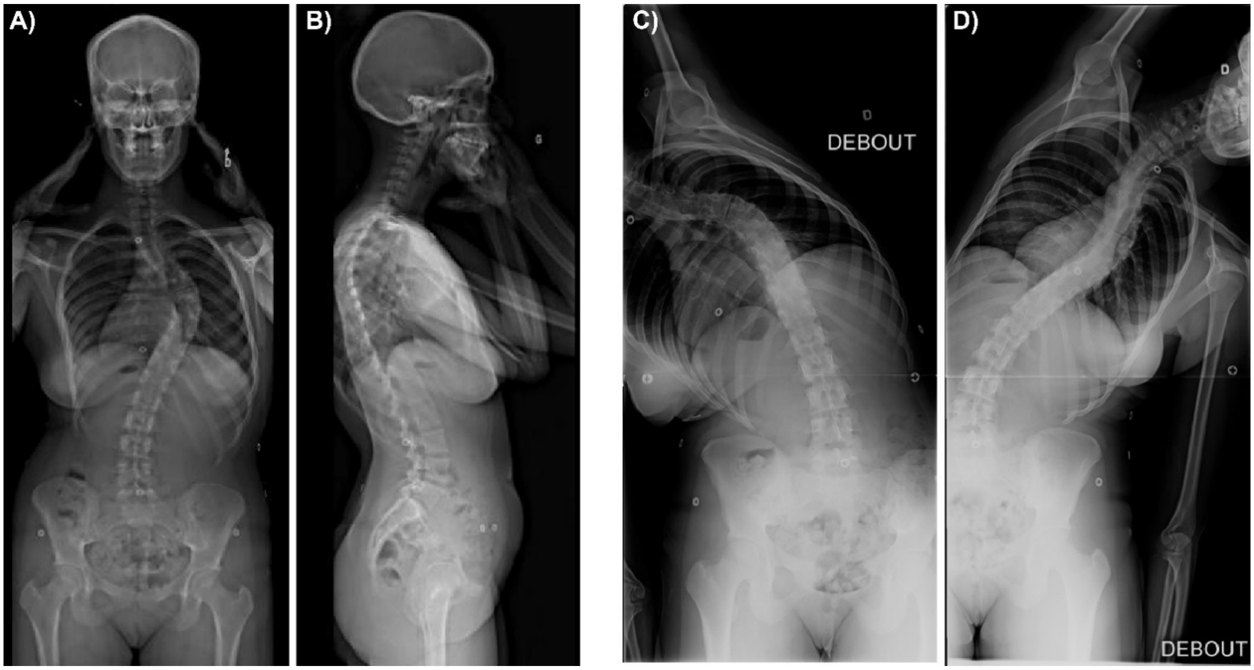


Figure 2.5 Routinely Acquired Radiographic Views for AIS Curve Evaluation
A-B) Standing posteroanterior (PA) and lateral radiographs.
C-D) PA radiographs during left (C) and right (D) side bending to assess curve flexibility.

Advanced imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI) are reserved for specific cases. CT scans are valuable for evaluating complex deformities, congenital anomalies, or pedicle size for surgical fixation, though their use is limited due to radiation exposure. MRI is preferred for identifying associated conditions such as syringomyelia, Chiari malformation, or tethered cord, which occur in 2–28% of AIS patients [105-107]. Surface topography is gaining attention as a radiation-free tool for monitoring scoliosis progression. Studies have shown that combining surface topography with radiographic imaging reduces the frequency of radiographs needed for mild curves (<25°) [108].

The Cobb angle, one of the primary metrics for quantifying spinal deformity, plays a key role in assessing both coronal and sagittal plane balance. For coronal balance, the Cobb angle measures the severity of lateral curvature by identifying the cranial and caudal end vertebrae, drawing lines along their endplates, and calculating the angle formed by the intersection of perpendicular lines

(Figure 2.4B, Figure 2.6A). Digitized methods using AI, such as spline-based systems and convolutional neural networks, have demonstrated improved accuracy and interobserver reliability compared to manual measurements [109-112]. Smartphone-based tools for Cobb angle measurement also show reliability comparable to conventional methods [113-115]. In the sagittal plane, the Cobb angle is used to quantify thoracic kyphosis (T5–T12; normal range: 10–40°; Figure 2.6B) and lumbar lordosis (L1–S1; normal range: 31–79°), which are critical parameters for understanding spinal alignment [77]. Spinal balance evaluation extends beyond angular measurements to incorporate global alignment assessments. Coronal balance between the pelvis and spine is assessed using the central sacral vertical line (CSVL; Figure 2.4C and Figure 2.6D), a vertical line on standing radiographs that bisects the sacrum. Deviation from the CSVL indicates an imbalance affecting posture and alignment. Coronal balance is also evaluated by comparing the C7 plumb line (C7PL; Figure 2.6E) to the CSVL. Balance is considered maintained if the distance between these lines is ≤2 cm (Figure 2.6F). Sagittal plane balance is determined using the sagittal vertical axis (SVA), which assesses the alignment between the C7PL and the posterior superior corner of the sacrum (Figure 2.6C).

2.2.1.2 Skeletal Maturity Assessment

Assessing skeletal maturity is vital for predicting curve progression and guiding treatment decisions. The Risser scale, based on the ossification of the iliac apophysis on coronal radiographs, is the most commonly used system in AIS patients' management and categorizes skeletal maturity from 0 (no ossification) to 5 (fully ossified) [116-118]. The Sanders Maturity Scale (SMS), which uses hand radiographs to assess skeletal maturity based on the development of hand and phalangeal bones, provides a more precise predictor of curve progression risk than the Risser stage, especially during the curve acceleration phase of growth [119]. The triradiate cartilage is another marker of skeletal immaturity, closing around the onset of menarche in girls and early puberty in boys. Its closure is often used to determine the safe timing for PSF to avoid crankshaft phenomena [120]. Automated methods using DL on EOS radiographs show promise in providing objective, efficient skeletal maturity assessments [121].

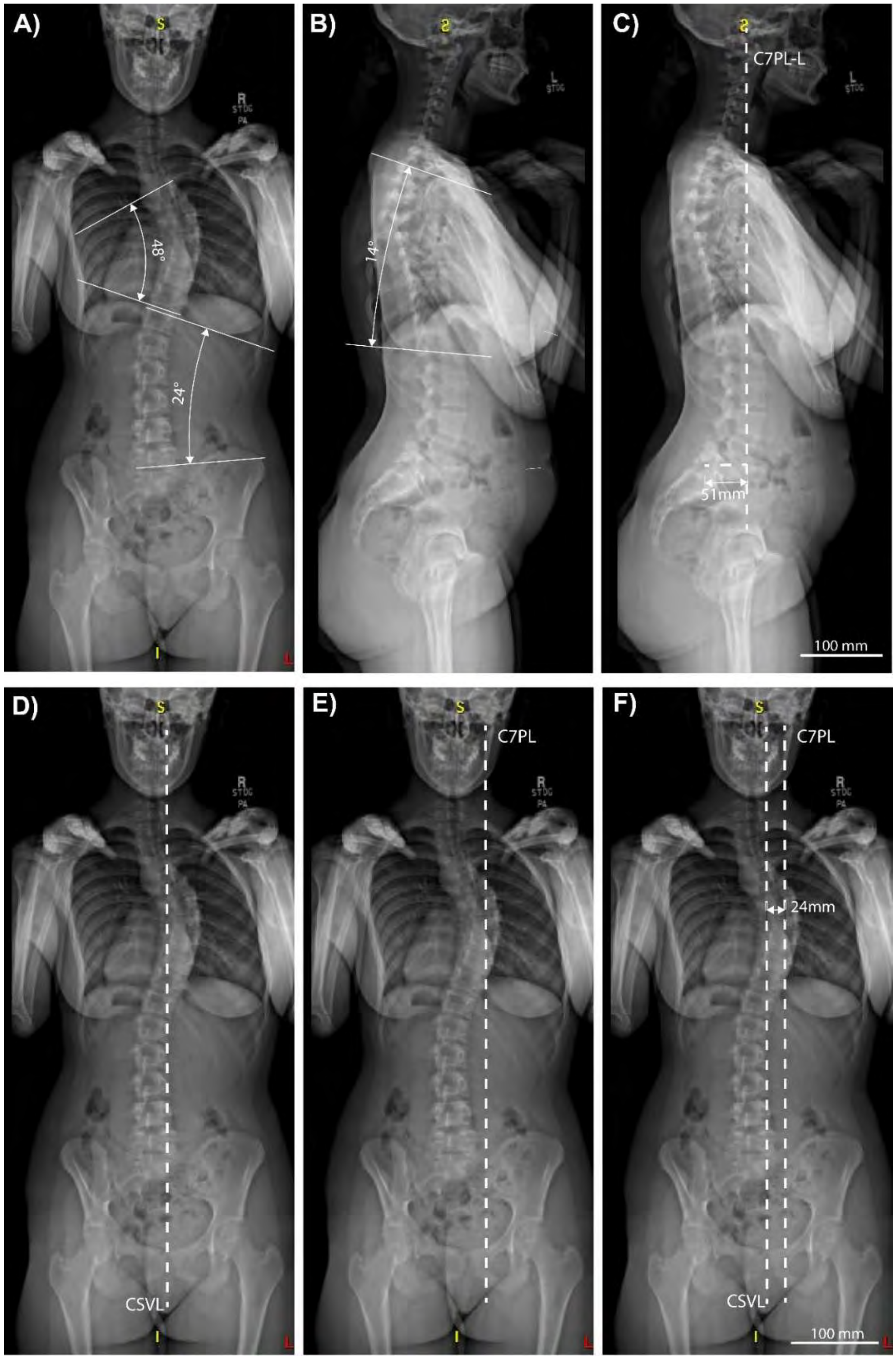


Figure 2.6 Geometric Indices for Evaluating Scoliosis in the Coronal and Sagittal Planes

A-B) Cobb angles in coronal (A) and sagittal (B) planes.
C) Sagittal balance using SVA, showing positive balance with C7PL >2 cm anterior to sacrum.
D-F) Coronal balance using CSVL (D) and C7PL (E), with coronal imbalance indicated by line distance (F) >20 mm.

2.2.2 Characterization of Scoliotic Curves

Characterizing and classifying scoliotic curves are essential in AIS management, guiding surgical intervention and planning. Classification systems define curve types, apex locations, and structural characteristics, providing a foundation for selecting vertebral levels for instrumentation and optimizing surgical outcomes [6, 18]. AIS curves are categorized by apex location, the most laterally deviated and horizontal vertebra, with types including cervical, cervicothoracic, thoracic, thoracolumbar, lumbar, and lumbosacral curves (Table 2.1) [122]. Thoracic curves are the most prevalent (48%), followed by thoracolumbar/lumbar curves (40%). Double curves are less common (9%), with double thoracic curves being the rarest (3%) [123]. Curve location significantly influences progression risk and treatment, with thoracic curves more prone to progression than lumbar curves.

Table 2.1 Curve Characterization by Apex Location and Prevalence (Suh et al. [123])

Curve Type	Location of Apex	Prevalence (%)
Cervical	C2–C6	Rare
Cervicothoracic	C7–T1	Rare
Thoracic	T2–T11	48%
Thoracolumbar	T11–L4	40% (combined thoracolumbar and lumbar)
Lumbar	L2–L4	
Lumbosacral	L5 or below	Rare; typically associated with other curves
Double curves	Multiple apexes	9%
Double Thoracic	Multiple thoracic apices	3%

2.2.2.1 Classifications based on 2D Approaches

The King-Moe classification, introduced in 1983, was the first attempt to describe AIS curve patterns and offer surgical guidelines for selective thoracic fusion based on coronal Cobb angles and curve flexibility [6]. However, poor reproducibility and limitations with modern segmental fixation led to the development of the Lenke classification system in 2001, which remains the current standard for AIS curve classification due to its improved reliability and comprehensive approach [124, 125].

The Lenke classification system defines scoliotic curves using three key components: curve type, lumbar modifier, and sagittal thoracic modifier [18]. To apply the system, upright coronal and sagittal radiographs, along with side-bending films, are obtained (Figure 2.5) to assess the *major curve* (with the greatest magnitude) and determine if accompanying curves are structural or nonstructural. Radiographic evaluation begins with identifying the proximal thoracic (PT), main thoracic (MT), and thoracolumbar/lumbar (TL/L) curves. The apex of the MT curve is typically between T3 and T11–T12, while the TL/L apex lies between T11–T12 and L4. Major curves, always considered structural, are central to surgical planning, while minor curves are assessed for flexibility. For a curve to be classified as structural, it must retain a Cobb angle ≥25° on side-bending films or exhibit significant kyphosis (≥20°) in the sagittal plane, regardless of its coronal characteristics. The *lumbar modifier* evaluates the position of the CSVL relative to the apical lumbar vertebra: A (CSVL lies between pedicles), B (CSVL touches one pedicle), and C (CSVL lies entirely outside the pedicles). The *sagittal thoracic modifier* classifies T5–T12 kyphosis as: (hypokyphotic, <10°), N (normokyphotic, 10–40°), or + (hyperkyphotic, >40°). Using the curve type, lumbar modifier, and sagittal thoracic modifier, a three-digit classification is assigned, such as 1AN. The system delineates six elemental curve types based on the location of the major curve and the structurality of accompanying minor curves (Figure 2.7) [18]. For example, Type 1 curves exhibit a major MT curve with nonstructural PT and TL/L curves, while Type 2 curves feature structural PT and MT curves with a nonstructural TL/L curve.

The Lenke system has significantly enhanced the reproducibility and reliability of curve characterization and treatment decisions, with improved intraobserver reliability (kappa = 0.92) and demonstrating greater consistency across clinical applications [126]. Despite being widely used, the Lenke classification system is solely based on a 2D approach [18], which facilitates its clinical application but also results in the separation between the 3D nature of this spinal deformity from the clinical management and surgical decision-making process [30-32]. Consequently, individualized surgical strategies often extend beyond the framework provided by the system.

Additionally, while the Lenke classification emphasizes coronal plane analysis, patients and their families are often more concerned with visible improvements in posture and symmetry, particularly the rotational aspects of the deformity that affect rib prominence and shoulder balance [127, 128]. Improvements in sagittal alignment are also important to them, as they may help prevent future problems such as neck and back pain. Surgeons increasingly recognize these concerns, although the Lenke classification system still primarily emphasizes coronal plane correction. As a result, additional clinical judgment is often required to address the 3D aspects of deformity that matter most to patients and families.

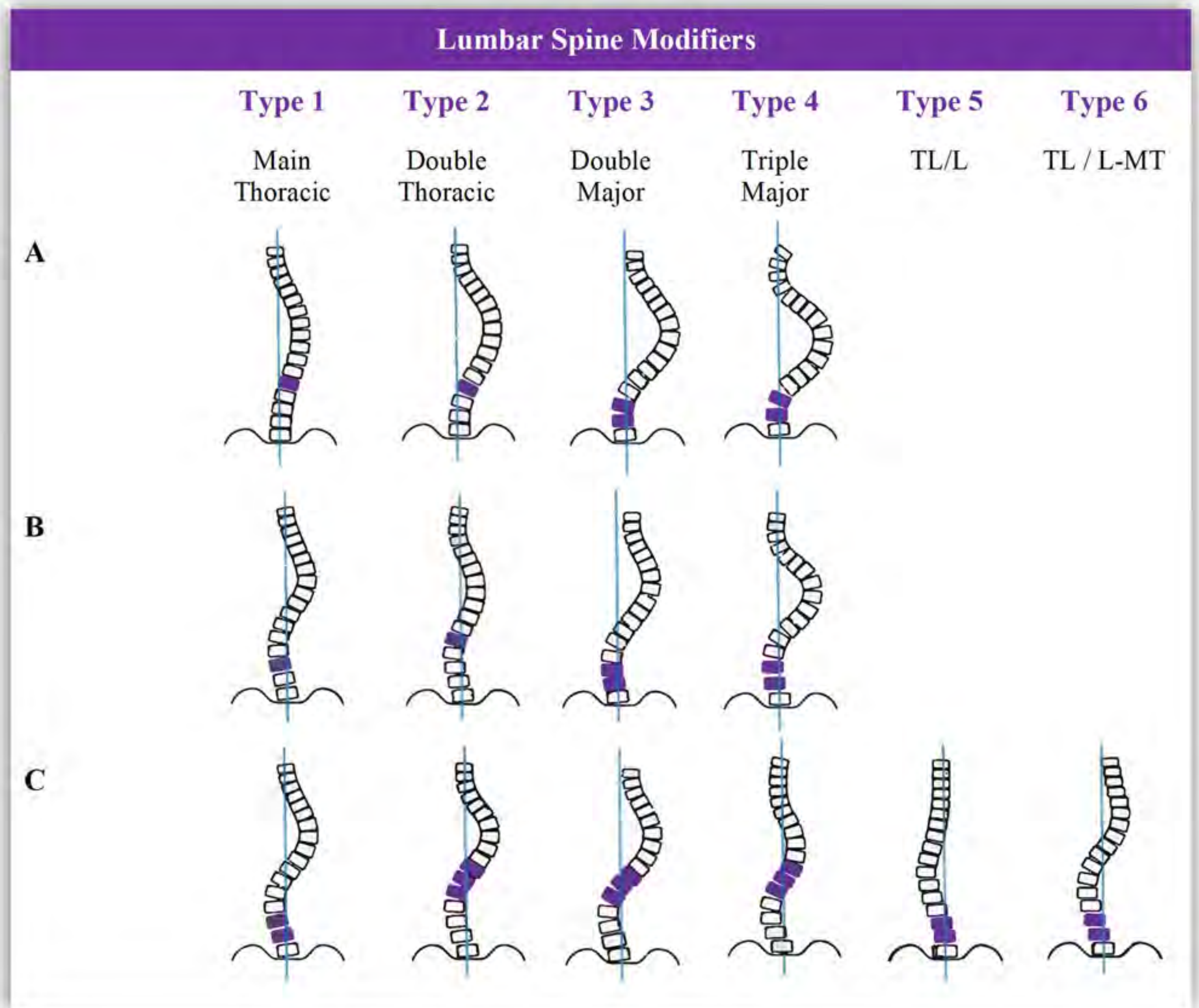


Figure 2.7 Schematic Drawings Representing the Lenke Classification of AIS Based on the Six Curve Types and the Three Lumbar Modifiers (A, B, C)

2.2.2.2 Measurements and Classifications based on 3D approaches

As outlined in the previous section, the Lenke classification, despite its widespread adoption, relies exclusively on 2D radiographic assessments, which cannot fully capture the complexities of AIS. Recognizing this limitation, the SRS initiated efforts in 2006 to develop a clinically useful 3D classification system [30-32]. Professor Ian Stokes initially outlined a 3D terminology, emphasizing the importance of parameters such as the orientation of the plane of maximum curvature (PMC) and the axial rotation of the apical vertebra (AVR) [129]. Since then, various studies, including those by Sangole et al., have identified distinct 3D subgroups within existing Lenke classifications, highlighting their impact on surgical outcomes [130]. Other studies have corroborated the existence of clinically relevant 3D subtypes within Lenke classifications, also underscoring their impact on surgical outcomes [30, 31, 130-133]. In line with this, the SRS 3D Scoliosis Committee has demonstrated that distinct 3D subtypes exist even within Lenke type 1 curves, and emphasized that recognizing these subtypes may influence surgical planning, as different 3D morphologies can require different reduction techniques and fusion strategies [134].

Three-Dimensional Reference Framework

To facilitate and standardize the evaluation of 3D deformities, the SRS Classification Committee defined global and local vertebral reference frames. Globally, the x-axis represents the anteroposterior direction, the y-axis is mediolateral, and the z-axis is upward. These axes correspond to the sagittal (x–z), coronal (y–z), and axial (x–y) planes, as viewed on standard radiographs and explained in section 2.1.1. This global framework is complemented by local reference frames for each vertebra. The x′–z′ plane divides the vertebra into medial and lateral halves, the x′–y′ plane separates the superior and inferior halves, and the z′-axis is defined by the average of perpendicular vectors to the endplates (Figure 2.8).

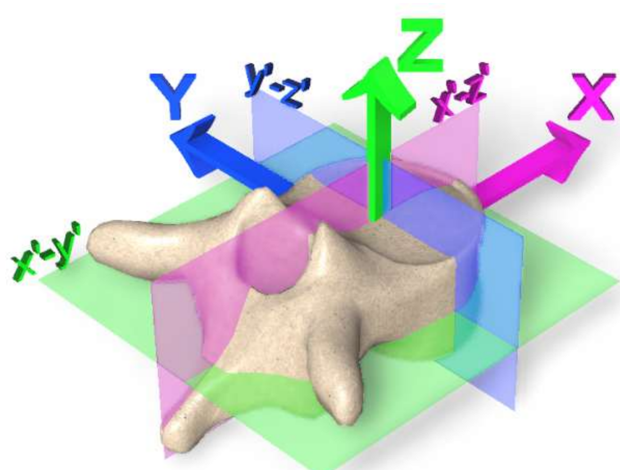


Figure 2.8 Drawings of the 3D Reference Axes by Local Vertebral Anatomic Landmarks

In healthy spines, physiological curves such as kyphosis and lordosis align closely with the x–z plane, and there is minimal rotation in the coronal plane. In scoliosis, however, vertebrae deviate from these alignments due to increasing coronal tilt and axial rotation at the curve's apex. These deviations reduce the reliability of traditional 2D radiographic assessments of kyphosis and lordosis, especially in advanced deformities. Spinal 3D reconstructions provide detailed axial plane views, enabling accurate measurement of vertebral axial rotation, which is particularly significant for evaluating the apical vertebrae [129, 135]. For right thoracic curves, this rotation appears as a clockwise twist when viewed from above.

Plane of Maximum Curvature

The PMC is an important descriptor in 3D scoliosis analysis because it defines the plane where the deformity is most pronounced (Figure 2.9). Unlike standard sagittal, coronal, and axial planes, the PMC captures the combined effect of regional rotation and lateral displacement [134]. In healthy spines, kyphosis and lordosis occur entirely within the sagittal plane, making it the default PMC. However, spinal rotation and lateral displacement in scoliosis cause the PMC to deviate from this alignment.

The PMC is constructed using three key vertebrae: the two end vertebrae that define the limits of the curve and the apical vertebra, where the deformity is most severe. By determining the centroids (midpoints) of these vertebrae, a unique 3D plane is established, allowing the curvature to be projected and quantified within this specific plane. The orientation of the PMC is measured relative to the sagittal plane and indicates the direction and extent of the apical vertebra's displacement [129, 130]. In a healthy individual, this angle is close to zero. As scoliosis progresses, the PMC typically shifts between the sagittal and coronal planes, sometimes rotating beyond the coronal plane in severe cases. The SRS introduced graphical tools like the "true da Vinci projection" to visualize these complex relationships [134].

SRS Three-Dimensional Classification System

Since it has become increasingly evident that AIS is a 3D deformity requiring a holistic approach, the SRS has been working on a 3D classification system integrating axial, coronal, and sagittal deformities into a unified framework. To ensure continuity with existing practice, this initiative builds on the widely used Lenke 2D modular system, while introducing complementary descriptors that explicitly capture transverse plane deformities, which are absent from the original framework.

Modern imaging techniques, such as synchronized upright biplanar radiography, have enabled the creation of 3D spinal reconstructions with minimal radiation exposure, providing critical insights into scoliosis's spatial complexities. Recent advances in low-dose imaging and validated self-calibration algorithms now allow accurate 3D reconstructions (<1.5 mm error) from routine radiographs, making 3D analysis accessible beyond specialized research centers [136, 137]. These reconstructions offer precise measurements of vertebral axial rotation and regional alignment within the PMC, improving the characterization of deformities and guiding surgical interventions (Figure 2.9) [138].

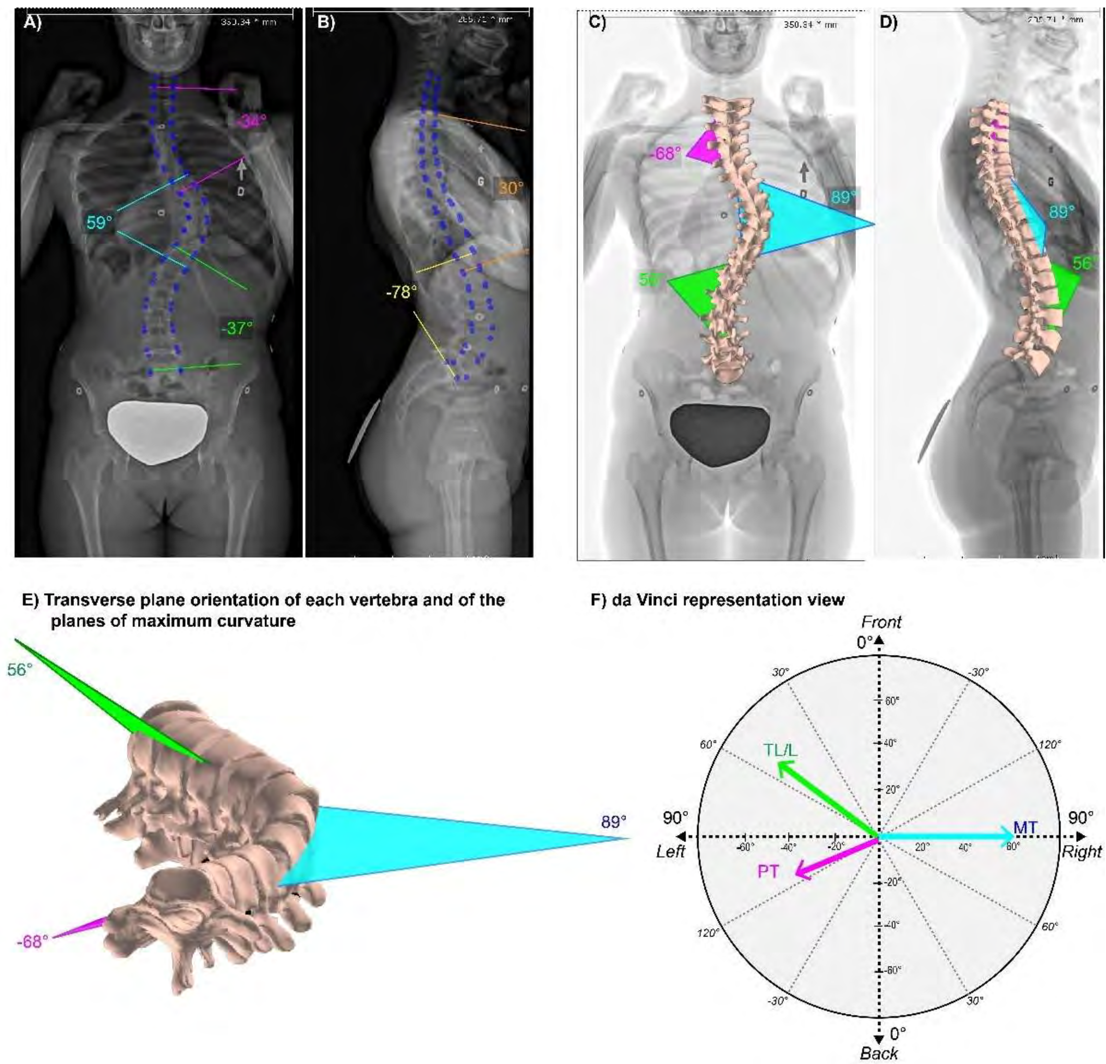


Figure 2.9 Biplanar Radiographs and 3D Reconstruction of Planes of Maximum Curvature

A-B) Coronal (A) and sagittal (B) radiographs with landmarks for 3D reconstruction. C-D) 3D views in coronal (C) and sagittal (D) planes showing PMC through TL/L, MT, and PT curves. E) Transverse "top-down" view of vertebrae and PMC. F) da Vinci projection of PMC as 2D vectors.

Research has highlighted the importance of incorporating 3D metrics into AIS classification to capture the complexities of spinal deformities better. Duong et al. used fuzzy clustering techniques to identify distinct 3D subtypes within Lenke Type 1 curves, revealing significant variability that traditional 2D methods cannot fully address [139, 140]. Further studies have explored the relationship between coronal and sagittal deformities, with Sullivan et al. and Hayashi et al. demonstrating a link between increased coronal curvature and reduced 3D thoracic kyphosis, highlighting the need to evaluate spinal deformities holistically [11, 141]. Additionally, Poncet et al. identified torsional patterns as a key intrinsic characteristic of scoliotic spines, suggesting torsion as a potential parameter for future 3D classification systems [142]. Building on this, Shen et al. proposed a robust numerical method to quantify geometric torsion in AIS and demonstrated that this index could distinguish between Lenke 1 subtypes based on twisting severity [143]. In later work, the same group contributed to a proposed SRS-endorsed 3D classification, emphasizing that integrating torsion alongside axial rotation and coronal alignment better reflects the true spatial complexity of AIS [8].

Building on these research efforts, the SRS-Lenke-Aubin 3D classification [136] recently introduced a new transverse-plane descriptor: the Orientation of the Regional Plane of Deformation (ORPD), a complementary descriptor to the AVR. ORPD captures the spatial orientation of each curve segment (PT, MT, TL/L), while AVR quantifies the degree of apical torsion. Thresholds for ORPD (<70°, 70–90°, >90°) and AVR (<10°, 10–20°, >20°) stratify clinically meaningful categories, distinguishing between preserved, flattened, or inverted sagittal profiles, and between slight, moderate, or marked torsion. These descriptors can be fully integrated into the Lenke system, producing a 4-modifier structure: Curve Type + Lumbar Modifier + Sagittal Modifier + ORPD + AVR [136].

In an extensive study using a 285-patient cohort by Aubin et al., all possible ORPD–AVR combinations were represented, underscoring the heterogeneity of AIS and demonstrating the added discriminative power of these 3D descriptors. Importantly, curves with nearly identical coronal patterns often exhibited different torsional or transverse-plane orientations, showing the limitations of 2D-only classifications [136]. Beyond their descriptive role, these indices provide actionable guidance for surgical correction trajectories, derotation maneuvers, and instrumentation strategies, bridging the gap between research-oriented metrics and operative decision-making.

Despite these advancements, 3D classifications remain predominantly research tools due to their complexity and the specialized imaging required. However, the SRS-Lenke-Aubin framework was specifically designed for clinical integration by maintaining continuity with the Lenke system and using routinely available imaging. This pragmatic approach positions it as a pivotal step toward operationalizing 3D analysis in scoliosis care, with the potential to enhance both surgical planning and long-term patient outcomes [136]. Their potential to revolutionize scoliosis management is undeniable. By offering a comprehensive understanding of spinal deformities across planes, 3D classifications promise to enhance surgical planning and patient outcomes. As imaging technologies evolve, the integration of 3D systems into routine clinical practice is likely to mark a significant milestone in scoliosis care.

2.2.3 Surgical Correction by Posterior Spinal Fusion

Approximately 10% of AIS cases necessitate medical intervention [144]. Treatment choice primarily depends on the degree of spinal curvature and the patient's remaining growth potential, as curves may progress rapidly during growth spurts. Larger curves exceeding 45° are particularly concerning, as they are associated with persistent health issues beyond adolescence. Approximately 1% of AIS patients will require surgery prior to adulthood [15]. Spinal fusion can correct spinal deformity by realigning the spine and providing immediate rigid fixation to achieve a balanced, solid vertebral fusion [18]. Over the past few decades, the PSF approach has emerged as the preferred technique due to its versatility and effectiveness [16, 17]. This method typically involves the placement of pedicle screws as vertebral anchors, which are then connected to dual rods spanning the length of the intended fusion (Figure 2.10). Despite extensive research evaluating different instrumentation strategies and 3D spinal deformity correction, no specific correction technique is applicable or successful in all AIS cases, and experienced surgeons could select multiple instrumentation strategies for the same patient [145-147].

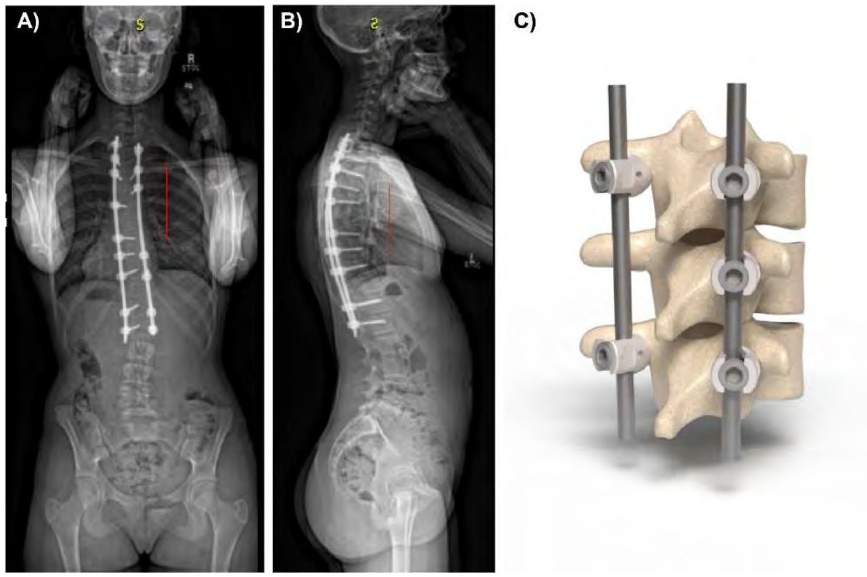


Figure 2.10 Instrumentation used for PSF Surgery for AIS Curve Correction

A-B) Posteroanterior and lateral radiographs after PSF showing instrumentation and arthrodesis. C) Schematic of thoracic segments with pedicle screws and rods, commonly used in PSF.

2.2.3.1 Surgical Planning

Preoperative planning in AIS is critical for achieving favorable postoperative outcomes and is often guided by the Lenke classification system (see section 2.2.2.1) [18]. The surgical decision-making process involves selecting the most appropriate fusion levels, the number and position of implants, and the mechanical properties of rods according to the patient-specific spinal deformity. The Lenke classification aids surgeons in distinguishing structural and compensatory (non-structural) curves, thereby determining which vertebrae should be included in the fusion [18]. Once the fusion levels are defined, the proper surgical instrumentation parameters must be identified alongside the corrective maneuvers required to realign the spine. A comprehensive discussion of these correction maneuvers is beyond the scope of this text. However, it is important to note that the success of these maneuvers is highly dependent on the instrumentation chosen. They will determine which maneuvers can be performed and their success because the implants will be under high levels of stress during the maneuvers and postoperative period.

Choice of Fusion Levels

Selection of fusion levels is generally based on the Lenke classification while accounting for the surgical objectives, anticipated deformity correction, and the surgeon's experience and evaluation of the patient-specific risks of postoperative complications [18, 148]. The number of fused levels should be minimized to preserve spinal mobility but adequately extended to prevent postoperative complications, such as malalignment or progression of unfused curves [23, 24]. Despite broad consensus on these principles, there is significant variability in surgeons' choices for the UIV and LIV [126, 145, 147, 148]. Recent literature has refined these decisions, particularly for Lenke types 1 and 3 curves. For example, Suk et al. recommend extending the fusion to the neutral vertebra when it is within two levels of the end vertebra [149]. Similarly, Cao et al. advocate ensuring that the LIV is within 1 cm of the CSVL on preoperative radiographs [150]. Additionally, studies have shown that extending the fusion to stable vertebrae bisected by the CSVL minimizes postoperative progression of unfused curves [151, 152]. The Harms Study Group Investigators recently introduced the "touched vertebra" concept (defined as the most cranial thoracolumbar or lumbar vertebra (T12-L5) that is "touched" by the CSVL on any portion of the vertebra[153]) as a Lenke classification modifier, suggesting it as an optimal LIV selection criterion [154].

Choice of Anchors

Various anchor types are available, including spinous process wires, sublaminar wires and/or bands, pedicle hooks, laminar hooks, transverse process hooks, and pedicle screws. Among these, pedicle screws are most commonly used due to their superior anchorage and facilitation of 3D deformity correction [8, 17, 21, 22]. Multiple types of pedicle screws exist, which must align with patient-specific anatomy and surgical goals. While a comprehensive review of all pedicle screw types is beyond the scope of this thesis, the following sections summarize their key parameters, surgical considerations, and insertion techniques to provide context for their application in preoperative planning. Pedicle screws may feature different head designs to accommodate various surgical objectives (Figure 2.11):

- *Monoaxial screws:* These have a fixed head, providing greater stability but requiring precise screw placement to align with the rod. They are advantageous in creating a rigid construct but may compromise pedicle placement for rod alignment, particularly in large or rigid deformities. Despite final tightening, the rod can slide more on a monoaxial screw as the rod may be oblique in the tulip head.

- *Polyaxial screws:* These screws allow for multidirectional head movement relative to the screw until the rod is locked, enabling safer pedicle placement while ensuring rod alignment. They are widely used in spinal fusions.
- *Uniplanar screws:* These screws pivot in one plane (typically sagittal), combining ease of rod engagement with strength preservation, and are specific to patients with rotational deformity as seen in scoliosis.

Recent studies showed that the use of polyaxial screws (screws that allow head rotation around the core in all directions) requires less effort during rod insertion and therefore decreases the resulting bone-screw forces but may be less effective in deformation correction [59, 155].

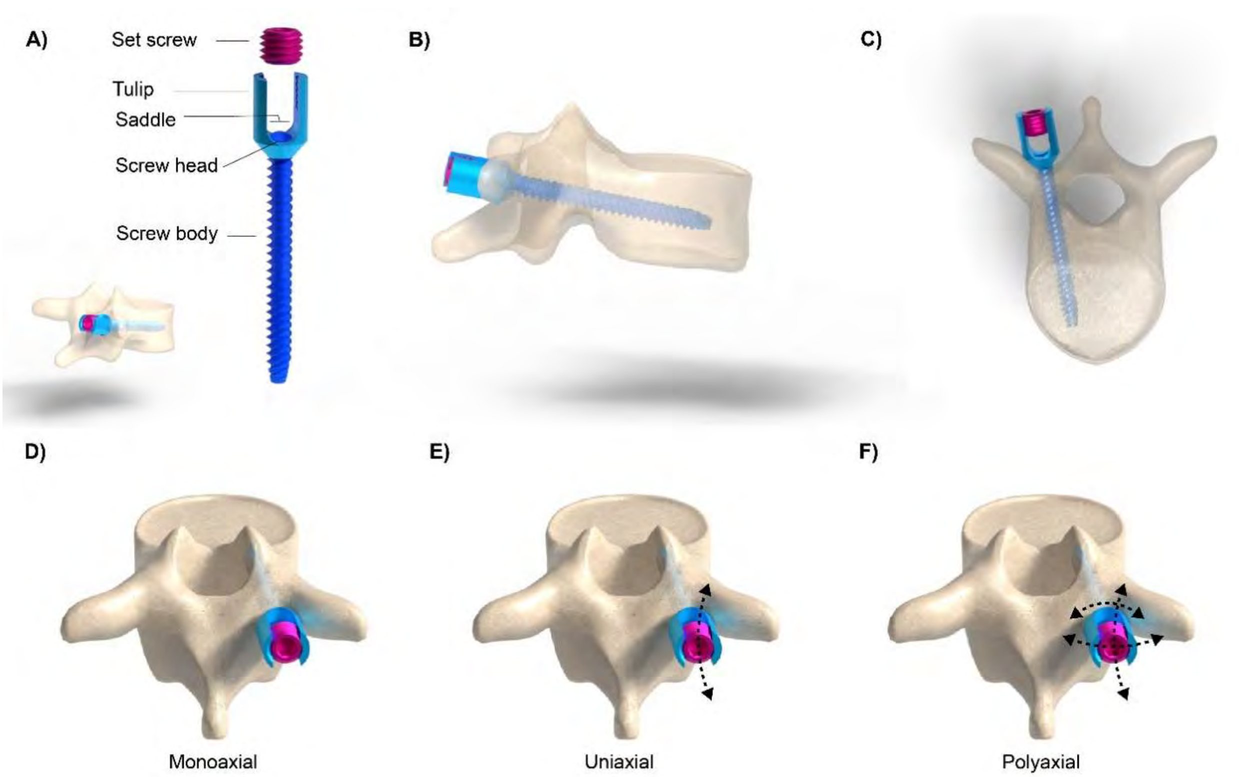


Figure 2.11 Pedicle Screw Instrumentation in PSF for AIS Correction

A) Pedicle screw anatomy and example placement in a thoracic vertebra shown in B) sagittal and C) transverse planes.
D-E) Screw types with degrees of freedom: monoaxial, uniaxial, polyaxial.

Material selection further influences screw performance, with titanium being most often preferred for its behavior and biocompatibility [17, 156]. Pedicle screw diameter and length vary widely to accommodate anatomical differences and surgical needs. Diameters typically range from 4 to 8 mm, and lengths span 20 to 100 mm, increasing in 5-mm increments. General guidelines suggest that screws should occupy 70–80% of the pedicle width and achieve 70–80% penetration of the vertebral body as seen on lateral radiographs [157, 158]. Pediatric anatomy and deformities further influence selection. AIS patients often present with asymmetry; concave pedicles are smaller due to compression, with 37% of T3–T9 concave pedicles unable to accept a 4-mm screw [159]. In contrast, convex pedicles, though larger, may require shorter screws at certain levels, such as T5 and T7–T9 [159]. In addition to pedicle fill, larger screws may have greater surface area and increased pull-out strength.

Pedicle screws can be inserted using two primary sagittal plane trajectories. When using an anatomic trajectory, the pedicle screw is inserted to follow the anatomical axis of the pedicle, directed 22° cephalocaudal. A straightforward trajectory technique is also common, in which the screw is inserted parallel to the superior endplate of the vertebral body. This trajectory offers biomechanical advantages such as higher insertion torque (0.3 vs 0.2 Nm) and pullout strength (611 vs 481 N) [160]. In the transverse plane, screws are typically inserted either along a "straight" axis parallel to the midline or along a "converging" axis, which angles inward at approximately 30° to match pedicle anatomy. “Converging” screws have an increased pullout strength by 29% [161], but constructs using “straight” screws are more stable when tested in fatigue [162].

Several other parameters, such as screw density (number of anchors per vertebral level) and distribution (their placement along the instrumented levels), have been studied in PSF surgery for AIS treatment [163, 164]. The implant density used varies among surgeons. Maximal implant density is two anchors per vertebra, resulting in a density of 2.0. While there is some radiographic evidence that spinal alignment might be improved with higher-density constructs, studies have not shown consistent improvement in intraoperative 3D spinal correction [25, 38, 164, 165] or patient outcomes [164, 166-168], and increasing the number of implants is not without risks. High-density constructs are associated with increased surgical time [169, 170], blood loss, complications [171], and costs [170, 172], while also subjecting the implants to greater mechanical stress, potentially leading to failure [25]. Therefore, anchor placement should balance correction capability with implant longevity and cost-effectiveness [25].

Choice of Rods

Rod choice, including diameter, shape, and alloy properties, is also an essential factor to consider for deformity correction [26-28]. Rod type influences the success of the corrective maneuvers and maintenance of the spinal alignment. The postoperative sagittal profile is highly dependent on the rods' biomechanical properties and the shape of the rods before insertion. The rods' properties should be chosen to have a construct's stiffness high enough for successful corrective maneuvers and sustainable deformity correction, but without being too stiff to avoid overcorrection or implant-related complications. Despite their central role in PSF instrumentation, there is no universal agreement guiding the preoperative selection of rods. Rod stiffness is highly dependent on the material stiffness (proportional to Young's modulus of elasticity) and rod diameter (increasing to the fourth power of the radius). Rods available to surgeons are typically made of stainless steel (SS), ultrahigh-strength stainless steel (UHSS), cobalt-chrome (CoCr), and titanium alloys (Ti) and have a diameter between 5.0 and 6.35mm. Rods made of SS and CoCr have a higher Young's modulus than titanium alloys [173], thus allowing smaller diameter rods to achieve similar results. However, Ti rods retain their original shape better than the other materials but cannot sustain as much corrective forces [174]. The correction technique and rod contouring are also factors related to rod selection to consider in AIS surgery. Different combinations can be used as a means of improving transverse plane deformities and restoring a normal sagittal profile. Differential rod contouring (Figure 2.12) is often used to address transverse plane deformities and restore a normal sagittal profile. This technique involves exaggerated bending on the concave side and reduced bending on the convex side before implantation. The difference in contour between the concave and convex rod before implantation influences the degree of postoperative deformity correction [44, 60].

Typically, rods are bent manually without specific measurements, relying on the surgeon's skill and intent [175, 176]. The implementation of patient-specific rods presents a viable solution to enhance sagittal balance restoration and achieve post-operative sagittal angles that align more closely with the intended goals [177, 178]. Initially applied to adults, this technique has recently been extended to adolescents as well [61, 179, 180].

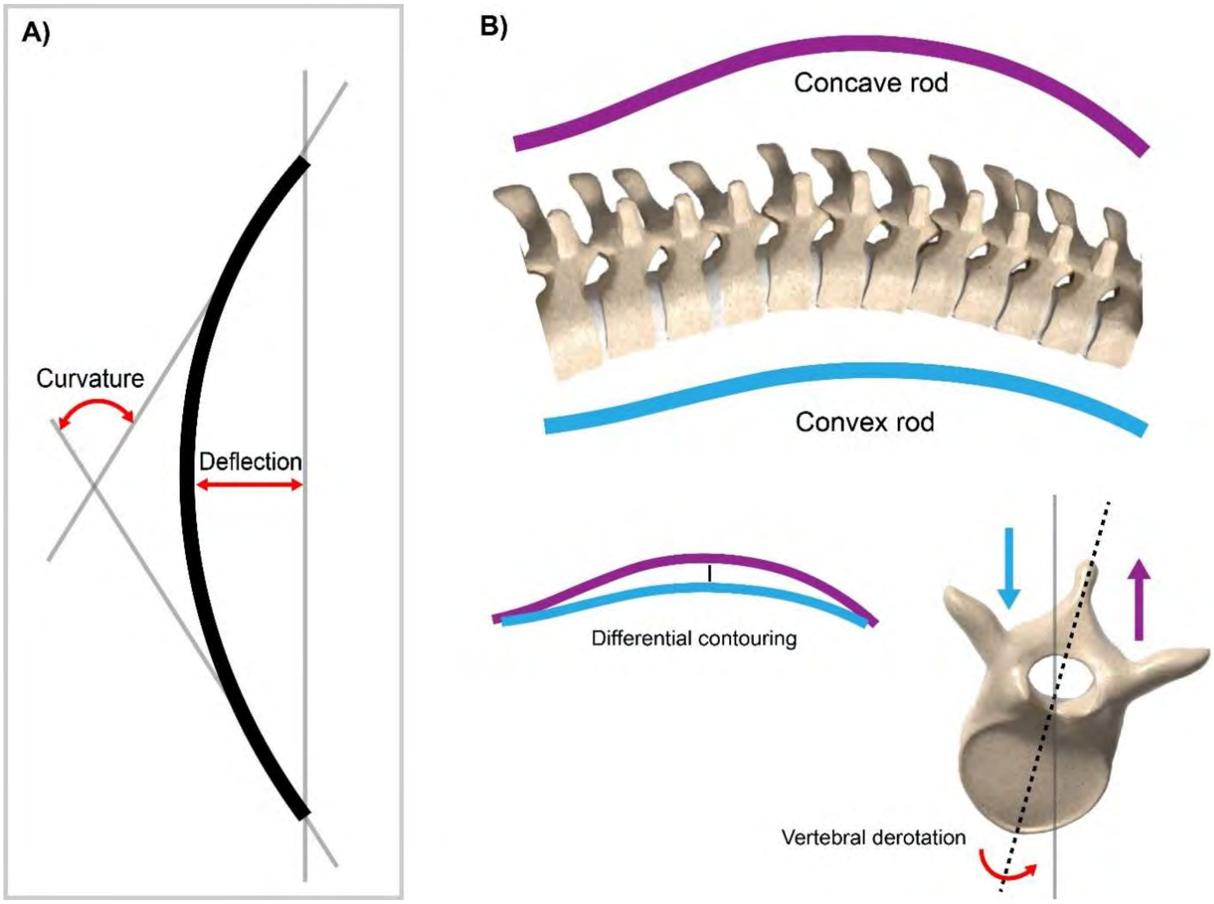


Figure 2.12 Schematic Illustration of Concepts Relevant to Differential Rod Contouring. A) Geometric parameters defining rod bending, and B) conceptual diagram of differential bending aimed at vertebral derotation in the transverse plane.

2.2.3.2 Prognosis After Surgery

Surgical outcomes for AIS have significantly improved in recent years [46] due to advancements in rigid segmental fixation techniques and a better understanding of posterior instrumentation and correction strategies [181]. However, complication rates remain relatively high, with 5% to 11.4% of AIS patients experiencing surgical complications. These complications include instrumentation failure, massive blood loss, surgical site infections, pseudoarthrosis, and neurological deficits [182-184]. Issues related to surgical instrumentation occur in up to 4.5% of AIS patients [26-28, 185] and include complications such as implant breakage, screw loosening, wear debris, neurological injuries, dural tears, and wound complications [184, 186, 187]. Various factors, including inadequate surgical planning, loss of distal fixation, and inappropriate selection of fusion levels, influence the incidence of these complications. Revision surgery may be required to address postoperative complications, particularly those arising from mechanical implant failure, such as rod breakage. Risk factors for such failures are often linked to instrumentation selection, as discussed previously [48-50]. Many elements of surgical instrumentation, often referred to as "surgeon-modifiable factors," are recognized as predictors of PSF outcomes. These include the position of the UIV and LIV [51-57], screw density and pattern [40, 58, 59], and differential rod contour or patient-specific rod contouring [60, 61].

2.3 Computer-assisted Methods for the Diagnosis and Treatment of Scoliosis

The diagnosis and treatment of scoliosis have significantly advanced with the integration of computer-assisted methods. These techniques encompass predictive models, biomechanical simulations, and AI-driven systems, all designed to enhance clinical decision-making and improve surgical outcomes. Traditionally, scoliosis assessment relied on manual measurements performed on standing coronal and lateral radiographs, a time-consuming process prone to interobserver variability. However, with the advent of digital imaging and the widespread implementation of Picture Archiving and Communication Systems (PACS) in the Digital Imaging and Communications in Medicine (DICOM) format, precise quantification of spinal curvature and pelvic parameters has become increasingly automated.

Computerized measurement tools now enable more efficient and standardized assessment of spinal deformities, allowing for enhanced preoperative planning. These systems assist surgeons in reducing the time required for planning while improving the accuracy of screw placement, rod contouring, and osteotomy angle determination [188, 189]. Additionally, advanced computational models and AI-based algorithms have been developed to predict the extent of deformity correction achievable with specific surgical strategies. While these technologies offer significant advantages in terms of precision and efficiency, their clinical adoption remains subject to factors such as validation, reliability, and accessibility. The following section explores these computer-assisted methods in detail, highlighting their applications, benefits, and current limitations.

2.3.1 Key Steps in Surgical Planning Software

Assistive planning computer software can perform many tasks relevant to AIS patients' clinical management, including radiographic measurements, instrumentation strategies, assessment of surgeon performance, and prediction of postoperative alignment. The main steps of assistive software for surgical planning generally consist of identifying preoperative parameters linked to identifying anatomical landmarks, defining postoperative alignment objectives, and performing a surgery simulation (Figure 2.13) [190].

For example, in the case of an AIS patient scheduled for PSF, surgical planning software could assist in every stage of the process, from preoperative assessment to postoperative evaluation (Table 2.2). The workflow could begin with the upload of coronal and lateral radiographs into the software. Automated landmark recognition could identify key anatomical structures such as vertebral endplates, spinous processes, and pedicles. The software could then calculate the Cobb angle, assess pelvic parameters like pelvic incidence and sacral slope, and determine vertebral rotation. Additionally, it could integrate side-bending radiographs to evaluate spinal flexibility, helping predict the degree of correction achievable during surgery.

Table 2.2 Summary of Example Key Steps for a Surgical Planning Software for AIS Using PSF

Steps	Description	Example of Software Features Used
1. Preoperative Parameter Identification	Measurement of Cobb angles, pelvic parameters, flexibility, and rotational deformities.	Automated landmark detection, vertebral segmentation, AI-based measurement tools
2. Postoperative Alignment Planning	Setting correction goals based on clinical and biomechanical targets, fusion level selection, and rod contouring objectives.	Predictive modeling, AI-based deformity simulation
3. Surgical Simulation	Planning pedicle screw placement, rod bending strategies, and osteotomy necessity.	Virtual implant simulation, 3D screws trajectory analysis, rod contour optimization, biomechanical analysis
4. Intraoperative Execution	Using navigation, robotic assistance, and patient-specific instrumentation.	Real-time fluoroscopic guidance, robotic-assisted screw placement
5. Postoperative Assessment	Evaluating surgical accuracy and monitoring patient recovery.	Long-term prediction tools, AI-based follow-up analysis

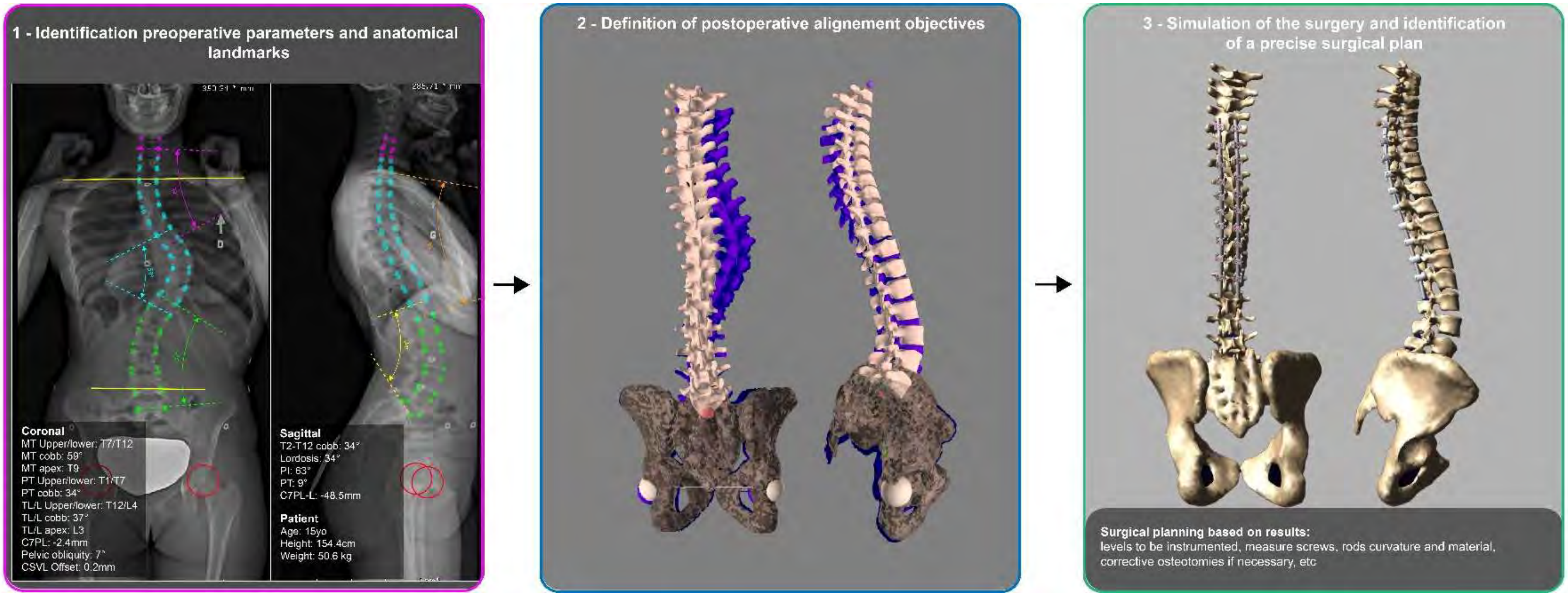


Figure 2.13 Example of Key Steps in Surgical Planning Software

This figure illustrates the critical stages involved in using surgical planning software for spinal surgery:
A) Identification of preoperative parameters and anatomical landmarks, providing a baseline for surgical planning.
B) Definition of postoperative alignment objectives to ensure optimal surgical outcomes.
C) Simulation of the surgery, allowing for a detailed and precise execution plan.

Once preoperative parameters are established, the software could be used to define postoperative alignment objectives. It might allow the surgeon to simulate different instrumentation strategies, adjusting screw density, rod contouring, and fusion levels to optimize spinal balance. For example, in a patient with a Lenke 1 curve, the system could assist in selecting the most effective fixation approach, such as segmental pedicle screw placement combined with pre-contoured rods to achieve optimal sagittal alignment. The ability to visualize different correction strategies preoperatively could help refine the surgical plan and anticipate potential challenges.

As part of instrumentation planning, the software could simulate pedicle screw placement, calculating optimal entry points, trajectories, lengths, and diameters. This could reduce variability and improve accuracy by ensuring that each screw follows the safest path based on the patient's anatomy. Similarly, the software could generate a patient-specific rod shape, providing recommendations for contouring based on predicted postoperative spinal alignment. If osteotomies are required, the system might help determine the optimal location and angle, predicting their impact on final correction.

During surgery, the preoperative plan could serve as a reference for the surgical team. If integrated with navigation technology, the software could provide real-time guidance for screw placement, confirming that the actual trajectory aligns with the planned one. Once instrumentation is secured, the surgeon could follow the planned rod rotation and compression maneuvers, ensuring that the correction matches preoperative predictions. Intraoperative imaging could be used to compare the achieved alignment with the surgical plan, allowing for last-minute adjustments if necessary.

Following surgery, the software could be used to compare preoperative and postoperative imaging, allowing for an objective evaluation of the surgical outcome. Automated measurements could assess the reduction in Cobb angle, confirm sagittal and coronal balance, and ensure that implants are positioned as intended. Over time, follow-up imaging could be uploaded to monitor spinal adaptation and fusion progression, helping to anticipate potential complications and refine future surgical planning.

This example illustrates how surgical planning software could enhance precision, improve predictability, and streamline decision-making in AIS correction using PSF. By integrating imaging analysis, virtual surgical simulation, and intraoperative guidance, such tools could help optimize deformity correction while reducing variability in complex spinal procedures.

2.3.2 Computer-Assisted Predictive Methods

Over the past decades, various software solutions have been developed to facilitate surgical planning in spine surgery, with programs such as Surgimap Spine (Nemaris, Inc.) and mediCAD (Hectec GmbH) gaining widespread adoption [35, 36]. These tools integrate preoperative imaging with computational models to assist surgeons in planning procedures. Surgimap, for instance, provides a user-friendly graphical interface that allows the automatic calculation of key surgical parameters, such as the Cobb angle, as well as the simulation of implant placements, including pedicle screws, cages, and osteotomies (Figure 2.14) [189, 191]. This assists in restoring sagittal and coronal balance preoperatively.

Figure 2.14 Example from Surgimap Preoperative Planning Software

This image demonstrates spinal balance parameters being measured and used for PSF planning. *Reproduced from Egea-Gámez et al., Rev Esp Cir Ortop Traumatol 2024;68:T73–T85, Elsevier España, with permission under CC BY-NC-ND 4.0.*

Similarly, mediCAD enables orthopedic spinal deformity planning by leveraging anatomic landmarks to track clinical measurements and predict implant positioning. This tool allows simulations of different surgical approaches and implant placement and can be used to predict pedicle screw lengths, ideal entry angles, and depths [192]. While both programs can be used for AIS preoperative planning, they primarily rely on 2D imaging, making them more suitable for degenerative conditions rather than complex 3D deformities such as scoliosis [33, 37].

Newer-generation software has sought to address the limitations of traditional 2D-based systems. UNiD Hub, part of the UNiD ASI program, integrates ML algorithms and predictive modeling (IB3D ASI) to assist with preoperative planning, intraoperative tracking, and postoperative evaluation. This program facilitates the creation of patient-specific surgical plans and the production of pre-bent rods, which have been FDA-approved for implantation [193]. Initially designed for adult spinal deformities, Medicrea's ASI system has demonstrated relevance in AIS cases [194], as its predictive modeling technology has been used to optimize rod contouring based on pelvic incidence, showing promise in improving sagittal alignment following PSF surgery [61]. Another recent addition to the field is SpineEOS (EOS Imaging), which introduces an alternative approach by generating 3D spinal reconstructions from 2D stereoradiographic images (Figure 2.15). This method provides 3D visualization without the radiation exposure associated with CT imaging [37, 195].

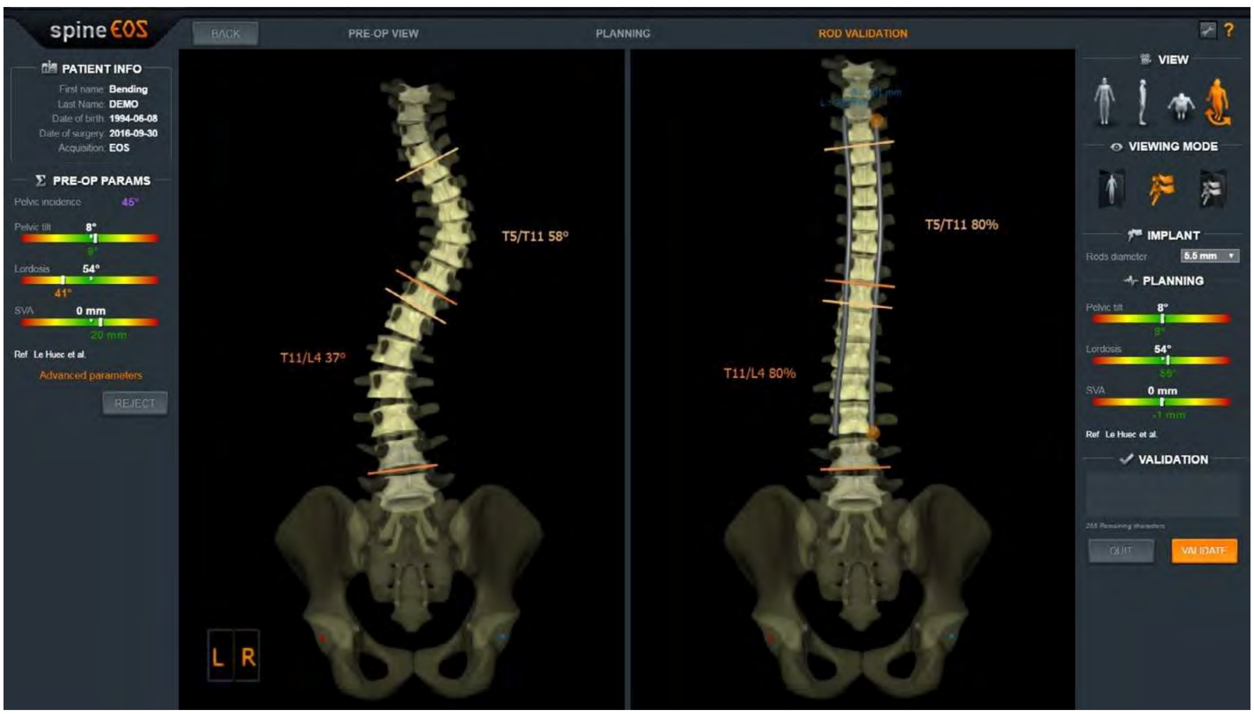


Figure 2.15 Sample Image from SpineEOS 3D Modeling Preoperative Planning Software

This image demonstrates 3D modeling and templating for an AIS patient prior to PSF surgery. Figure reproduced from Floyd et al., Semin Spine Surg 2020;32:100787. Elsevier Inc. Reproduced with permission under the STM Permissions Guidelines.

However, its workflow requires the manual identification of anatomical landmarks, which can be time-consuming despite the built-in computational tools. A clinical study using SpineEOS demonstrated improved deformity correction planning for AIS cases, but the decision-making process remained limited to rod contouring and length selection, without incorporating biomechanical analysis of patient-specific flexibility and spinal geometry [37]. Further, there are concerns with data use from these commercial systems, which may save X-ray imaging for further software optimization. Privacy rules may place increasing limitations on solutions for external 3D reconstruction and analysis of radiographs. Another planning system, Brainlab Spine Planning, focuses on high-resolution 3D reconstruction and image-guided screw trajectory planning. While more commonly used in neurosurgical and degenerative spine cases, the tools offered can apply to AIS surgery such as pedicle screw planning and segmental alignment analysis [196]. However, it does not currently incorporate biomechanical correction modeling or patient-specific rod simulations.

Given the variety of available alignment planning programs, Table 2.3 and Table 2.4 provide comparative overviews of key software used for spinal surgery planning. As digital surgical planning tools become more integrated into clinical practice, the way these platforms handle imaging data and enable surgeon oversight has direct implications for patient safety, data privacy, and regulatory compliance. While many systems offer advanced features such as cloud-based analytics and AI-driven planning, these capabilities come with varying degrees of data sharing, storage practices, and clinician control. Table 2.5 provides a comparative overview of leading spine planning software and platforms with respect to data governance, surgeon oversight, data storage, and applicable privacy regulations. It should be noted, however, that publicly available information on these aspects is often limited or inconsistently reported across company websites and user documentation.

Beyond preoperative planning, assistive planning software plays a crucial role in surgical execution by integrating imaging data into 3D modeling systems. CT scans and other imaging modalities can be uploaded into specialized software to generate virtual and physical models of a patient's spine, improving the surgeon's understanding of complex anatomical structures both preoperatively and intraoperatively [197]. These models facilitate surgical precision by enabling the creation of patient-specific instrumentation and implants through 3D printing. Custom printed drill guides tailored to the patient's anatomy can improve the accuracy of pedicle screw placement and guide complex osteotomies, reducing intraoperative variability and increasing procedural safety [198, 199]. Furthermore, custom 3D-printed interbody cages have the potential to enhance deformity correction by providing individualized support structures optimized for a patient's unique spinal alignment.

Table 2.3 Comparative Overview of Spine Alignment Planning Software [200, 201]

Software	Availability	Cost	Developer	Industry links	User base	Advantages	Limitations
Surgimap	Download	Free basic version	Independent	Used by Stryker, Globus, and others	Surgeons	Free version available; user-friendly interface	Some features require upgrades; does not incorporate biomechanical modeling
UNiD Hub	Online	Free for Medtronic users	Medtronic	Metronic exclusively	Surgeons & Medtronic team	Third-party radiological analysis available; Includes AI-driven predictive modeling; integrates with intraoperative workflow	Requires Medtronic authentication; limited to Medtronic planning; does not incorporate biomechanical modeling
SpineEOS	Online	Subscription-based	Alphatec	None	Surgeons	Integrated with EOS imaging; Generates 3D reconstructions from 2D images; avoids CT radiation exposure	Requires EOS imaging system; Requires manual landmark identification; does not incorporate biomechanical modeling

Table 2.4 Comparative Overview of Patient-Specific or Pre-bent Rods Software [200, 202, 203]

Company (Country)	Technology Type	Rod Material and Design	Rod-Screw Connection	Fixation Mechanism	Implants
NuVasive (USA)	Planning and intraoperative rod contouring using a connected bending device	Ti or CoCr, available in 6 mm or 5 mm diameter with a round profile	Top-loading connection	Polyaxial or monoaxial pedicle screws	Screws, hooks, sublaminar bands
Medicrea (France) / Medtronic (USA)	Preoperative planning and rod manufacturing	Ti or CoCr, available in 6 mm, 5.5 mm, or 3.5 mm diameters; round or derotation rod with a hybrid cross-section featuring two flat surfaces and a curved profile	Top-loading (tulip-style screws) or side-loading (dome screws) with polyaxial, monoaxial, or uniplanar pedicle screws	Various fixation options including polyaxial, derotation, and realignment connectors	Hooks, claws, sublaminar bands
SMAIO (France)	Preoperative planning and custom rod production	Ti, 6 mm or 5.5 mm diameter, round profile	Side-loading connection	Monoaxial screws	Screws, hooks, and claws

Table 2.5 Comparative Data Governance and Surgeon Oversight in Spine Planning Software [200-203]

Software / Complant	Data Use Policy	Surgeon Oversight	Data Storage Location	Regulatory/Privacy Considerations
Surgimap (Nemaris)	Collects user interaction data; stores DICOM images with potential patient identifiers.	Surgeon-driven planning; manual input and validation required.	Local or encrypted cloud storage.	HIPAA-compliant; Data may be used for software improvement
UNiD Hub (Medtronic)	Utilizes patient data for AI-driven predictive modeling; data shared within Medtronic ecosystem.	Surgeon collaborates with Medtronic team; oversight in planning and validation.	Cloud-based storage within Medtronic's infrastructure.	Requires authentication; may raise concerns over proprietary data use; Data use governed by Medtronic's policies
SpineEOS (EOS Imaging)	Collects personal data for service provision; committed to safeguarding privacy.	Surgeon inputs anatomical landmarks; manual oversight required.	Cloud-based storage managed by EOS Imaging.	Requires EOS system; Adheres to privacy policies; limited information on long-term data storage
NuVasive	Data use policy not publicly detailed; likely subject to HIPAA and other regulations.	Surgeon-driven planning; tools assist in intraoperative adjustments.	Data storage specifics not publicly detailed.	HIPAA-compliant; data use largely internal unless otherwise consented
Medicrea (France) / Medtronic	Utilizes patient data for AI-driven surgical planning; data shared within Medtronic ecosystem.	Surgeon collaborates with Medicrea team; oversight in planning and validation.	Cloud-based storage within Medtronic's infrastructure.	Data use governed by Medtronic's policies; GDPR and HIPAA compliance emphasized; may raise concerns over cloud data governance
SMAIO	Imaging data uploaded to SMAIO servers; specific consent obtained for data reuse in algorithm improvement.	Surgeon-led planning with software assistance; manual validation.	Data storage specifics not publicly detailed.	GDPR-compliant; data use governed by European standards, with opt-in for AI training

2.3.2.1 Limitations and Future Directions

Computer-assisted methods for scoliosis diagnosis and treatment have advanced clinical decision-making but face significant limitations for addressing the 3D complexity of AIS. As discussed in the previous section, current tools rely heavily on geometric-based analyses and 2D determinants, which are better suited for degenerative conditions and often fail to account for biomechanical factors such as forces, stresses, and spinal flexibility. This limitation reduces their effectiveness in guiding optimal, patient-specific instrumentation strategies and contributes to the variability in surgical outcomes. As a result, suboptimal correction or mechanical failure remains a concern that is not well addressed by available planning software. Widely used systems like Surgimap and mediCAD, though effective for radiographic measurements and surgical simulations, lack the integration of patient-specific biomechanical modeling. Even advanced systems like UNiD Hub and SpineEOS, which offer predictive modeling and 3D reconstructions, are hindered by time-consuming manual processes, such as anatomical landmark identification, and the absence of biomechanical analyses. Furthermore, key factors influencing surgical outcomes, such as rod properties (diameter, section, material, notched vs. non-notched), surgical parameters (type and density of implants, rod-screw connection, correction and release techniques), and baseline patient variates (spine stiffness, preoperative thoracic kyphosis, and patient-related factors) are not adequately accounted for. The relationship between the preoperative plan and the actual achieved alignment remains poorly understood, limiting the ability to optimize surgical corrections. These gaps limit their ability to improve postoperative alignment and reduce implant failure rates.

Future advancements must integrate 3D determinants and biomechanical simulations to enhance predictions of alignment and implant performance. Incorporating analyses of forces, stresses, and implant-bone interactions will provide more comprehensive guidance for surgical planning. Efforts should also focus on reducing the complexity and time required for anatomical landmark identification through automation and machine learning, streamlining workflows and improving usability. Dynamic, real-time modeling during surgery could further personalize strategies and reduce inter-surgeon variability. While current tools have improved preoperative planning, addressing these limitations is essential to achieve consistent and superior outcomes for AIS patients.

2.3.3 Biomechanical Simulation of PSF Surgery

2.3.3.1 Patient-Specific Biomechanical Modelling

Biomechanical simulations have become a valuable tool for preoperative planning of PSF surgery for AIS. By leveraging patient-specific computational models, these simulations allow for the evaluation of instrumentation strategies and biomechanical outcomes using 3D reconstructions of the spine. Surgeons can assess how different instrumentation configurations impact deformity correction and predict postoperative spinal alignment, tailoring surgical plans to the patient's unique anatomy [25, 40, 59, 204, 205]. Integrating computational models with 3D reconstructions from medical imaging, such as biplanar radiographs, enables accurate patient-specific spine modeling for surgical planning. These models facilitate virtual testing of multiple instrumentation strategies and correction techniques, providing biomechanical insights that would not be feasible through *in vivo* experimentation.

Simulation Techniques

Computational modeling relies primarily on finite element (FE) methods [206-209] or MB dynamic modeling, also referred to as kinetic modeling (Figure 2.16) [25, 38, 43, 58, 155, 210-212]. Both approaches enable the simulation of surgical instrumentation and the analysis of resulting spinal corrections and implant forces, but they differ in their computational methodologies and clinical applicability.

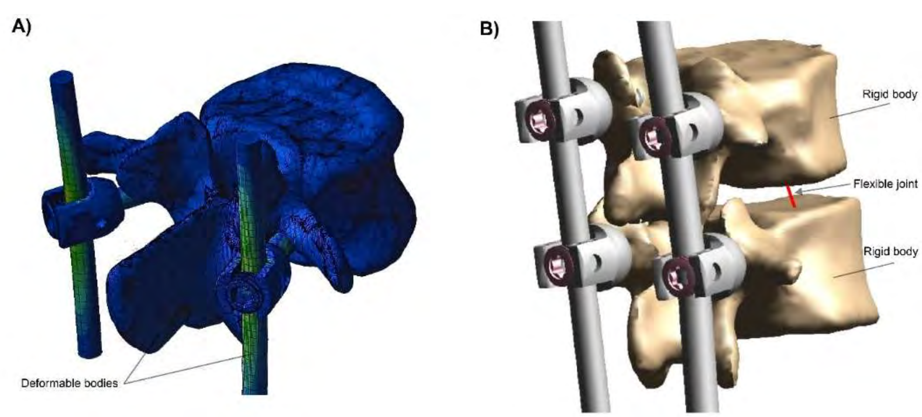


Figure 2.16 Examples of Vertebra and Surgical Instrumentation Modeling
A) Finite Element method and B) multibody approach.

FE modeling uses numerical methods to predict deformation and stress fields within solid bodies subjected to external forces. It relies on the discretization of a continuous geometry into a finite number of simpler subdomains, called elements, each governed by equations that represent material behavior. The FE method transforms the governing partial differential equations of continuum mechanics into a large system of algebraic equations, which are solved iteratively to reach equilibrium. This approach allows for the assignment of different mechanical properties to each element (e.g., cortical bone, trabecular bone, intervertebral disc, ligament, or implant), enabling detailed spatial analysis of internal forces and deformations. While FE analysis provides high spatial resolution and is well-suited for investigating stress distributions and localized phenomena, its application to scoliosis surgery is hindered by challenges related to convergence and computational complexity. The spinal system is composed of multiple interconnected vertebrae, implants, intervertebral discs, and rod-implant-vertebra joints, creating significant nonlinearities and stiffness variations that complicate FE-based solutions. Moreover, numerous state changes occur throughout the surgical procedure, such as during implant positioning, connection to the rods, and corrective maneuvers, each altering the boundary conditions and structural configuration of the model [213].

As an alternative, MB modeling has been proposed to overcome these limitations by treating the spine as a system of interconnected rigid bodies with defined joint constraints, and deformable elements. While MB models are not exempt from convergence issues, they facilitate the representation of surgical maneuvers through defined joint constraints and simplified mechanical representations of deformable components such as rods and intervertebral joints [41]. Moreover, MB modeling allows for easier representation of the numerous state changes that occur throughout the surgical procedure, such as implant positioning, connection to the rods, and corrective maneuvers, by adjusting constraints and boundary conditions dynamically within the simulation. These models provide a computationally efficient framework that yields clinically acceptable first-order estimates of both the spinal deformity correction and the forces acting on the instrumentation and spine. This makes MB modeling particularly useful for simulating the sequential steps of surgical correction and evaluating both the spinal deformity correction and the forces involved in AIS surgery.

2.3.3.2 Simulation of Spinal Instrumentation

Patient-specific biomechanical models can assist preoperative planning by predicting the biomechanical impact of various instrumentation configurations and surgical techniques (Figure 2.17) [25, 40, 59, 204, 205]. Early work by Aubin et al. (2003) pioneered the use of MB modeling to simulate surgical maneuvers in AIS, providing a framework for selecting optimal instrumentation configurations during preoperative planning [41].

Subsequent advancements incorporated patient-specific flexural deformity parameters, improved joint kinematics, and introduced flexible rods to more accurately simulate surgical procedures [214-216]. This led to the development of the "Spine Surgery Simulator," a preoperative planning tool initially validated in 10 AIS cases, demonstrating an average Cobb angle prediction error of just 1.2° compared to actual postoperative results [43].

Recent MB simulation studies have further refined instrumentation techniques for AIS patients, focusing on screw configurations, rod contouring strategies, and correction maneuvers. For instance, Wang et al. (2023) examined screw density and found that while higher density enhances transverse plane correction, it does not significantly impact coronal or sagittal alignment, emphasizing the need to prioritize correction strategies rather than solely modifying screw placement [217]. Similarly, La Barbera et al. (2021) demonstrated that optimization of screw pattern and density has a limited effect on global 3D correction, reinforcing the importance of tailored correction objectives [39]. Rod contouring in PSF for AIS patients has also been extensively analyzed using MB models. Gay et al. (2023) found that increasing concave rod contouring angles improves thoracic kyphosis and apical vertebral derotation but results in higher screw pull-out forces and reduced coronal plane correction [218]. Additionally, Wang et al. (2023) compared concave-first versus convex-first rod insertion and found no significant difference in final alignment outcomes, but confirmed that differential rod contouring angles significantly influence correction forces and vertebral rotation [219].

Despite these advancements, patient-specific MB models for AIS instrumentation are not yet fully adapted for routine clinical use. However, their credibility continues to grow, as demonstrated by a recent ASME V&V 40 validation study supporting their application in assessing proximal junctional failure risks in adult spinal deformity [42]. These findings reinforce MB modeling as a powerful tool for simulating instrumentation strategies and predicting postoperative outcomes, though further refinement is needed for widespread clinical translation.

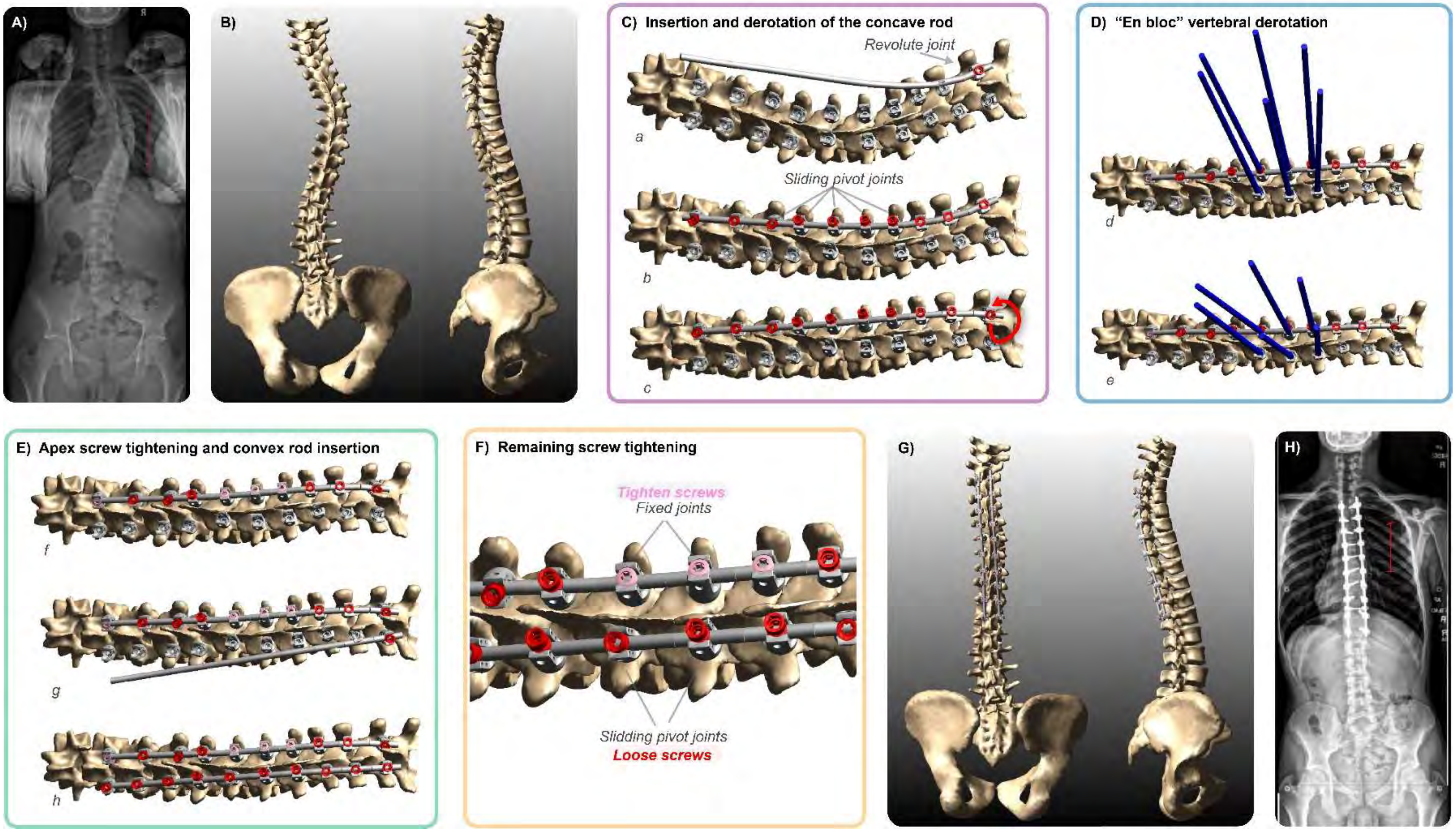


Figure 2.17 Illustration of Multibody Approach for Simulation of Spine Surgery.

A) Pre-op PA radiograph and B) 3D geometry from biplanar X-ray. C) Concave rod insertion (a–b) and torque application (c). D) En bloc derotation with apical derotators (d–e). E) Apical screw tightening (f) and convex rod insertion (g–h). F) Final tightening with fixed joints. G-H) Post-op radiograph and 3D geometry after simulation.

Optimization algorithms can be used with modeling and simulation techniques to support the surgical decision-making process. A recent optimization study developed and tested a new *in silico* patient-specific optimization method using MB modeling to identify the instrumentation strategy minimizing the 3D deformity for thoracic AIS to aid surgical planning [44]. In this study, a computerized algorithm was successfully developed using 1080 virtual intraoperative correction simulations to determine the best instrumentation parameters to achieve optimal thoracic scoliosis correction and mobility [44]. This computational patient-specific optimization employed a multivariate linear regression approach based on five instrumentation parameters (UIV, LIV, screw pattern, differential concave-convex rod curvatures, and rod stiffness) as predictors of three descriptors (thoracic Cobb angle, T4-T12 thoracic kyphosis, and thoracic apical vertebral rotation) of scoliosis correction. Afterward, an objective function rated the outcome of the simulated surgeries and produced the optimal correction strategy for each specific set of correction objectives to maximize 3D spine deformity correction while preserving motility. This study distinguished itself from previous studies by systematically investigating five instrumentation parameters, making it more suitable for accurately identifying optimal instrumentation strategies for preoperative planning in AIS correction. However, the overall computational cost of finding an optimized instrumentation strategy was approximately 36 hours, which limits the use of such an approach for each AIS patient in a routine clinical setting. Reducing the high computational time (i.e., number of simulations) required would be beneficial to implementing such an optimization method for preoperative planning of surgical strategies for patients.

2.3.3.3 Limitations and Future Directions

Despite the advancements in biomechanical simulations for preoperative planning of PSF surgery in AIS, several limitations hinder their routine clinical application. One of the main challenges is the complexity and time-intensive nature of patient-specific spine reconstruction and surgical simulation. While 3D reconstructions of the spine can be generated using biplanar radiographs instead of volumetric imaging techniques, this process remains time-intensive due to the need for precise identification of multiple anatomical landmarks. Although this method allows for patient-specific spine modeling with high accuracy, the manual or semi-automated landmark selection adds to the overall complexity and workflow burden, limiting its clinical feasibility for routine preoperative planning.

Computational cost remains a significant barrier, particularly when incorporating optimization algorithms for selecting instrumentation strategies. Recent studies using MB modeling demonstrated the feasibility of patient-specific optimization methods, but the computational time for a single optimized surgical strategy can exceed 36 hours [44]. The extensive number of instrumentation scenarios that must be simulated to determine the optimal configuration is a primary contributor to this computational burden. Reducing this high processing time would be essential for translating such methods into clinical practice, enabling real-time or near-real-time surgical planning tools.

Another critical challenge is the verification and validation (VV) of patient-specific computational models. Ensuring the accuracy of numerical predictions requires rigorous validation against clinical outcomes and experimental data. The complexity of the scoliotic spine, with its nonlinear mechanical properties and individualized response to instrumentation, makes this process particularly demanding. While deterministic numerical models provide predictions of surgical outcomes, they do not inherently identify the optimal instrumentation strategy, further necessitating iterative simulations and refinement.

Future research should focus on improving computational efficiency and enhancing the predictive capability of biomechanical models. Advancements in parallel computing, AI-assisted surrogate models, and reduced-order modeling approaches could help mitigate computational costs while maintaining accuracy. Ultimately, overcoming these limitations would facilitate the broader adoption of patient-specific computational tools for optimizing surgical strategies in AIS.

2.3.4 AI for the Diagnosis and Treatment of Spinal Deformity

Artificial intelligence, and in particular ML, is increasingly being applied in the diagnosis and surgical treatment of spinal deformities. These technologies support clinical decision-making by recognizing complex patterns in medical imaging and patient data, thereby improving diagnostic accuracy and preoperative planning. A detailed review of these developments is presented in Chapter 4, in the manuscript titled "The use of deep learning in medical imaging to improve spine care: a scoping review of current literature and clinical applications." The summary below introduces core concepts and selected applications. ML encompasses a range of approaches, including supervised, unsupervised, and reinforcement learning (Figure 2.18), that can be applied to tasks such as classification (e.g., automated grading of disc degeneration), regression (e.g., estimating anatomical landmarks), and clustering (e.g., grouping scoliosis patients into novel 3D categories) [139, 220, 221].

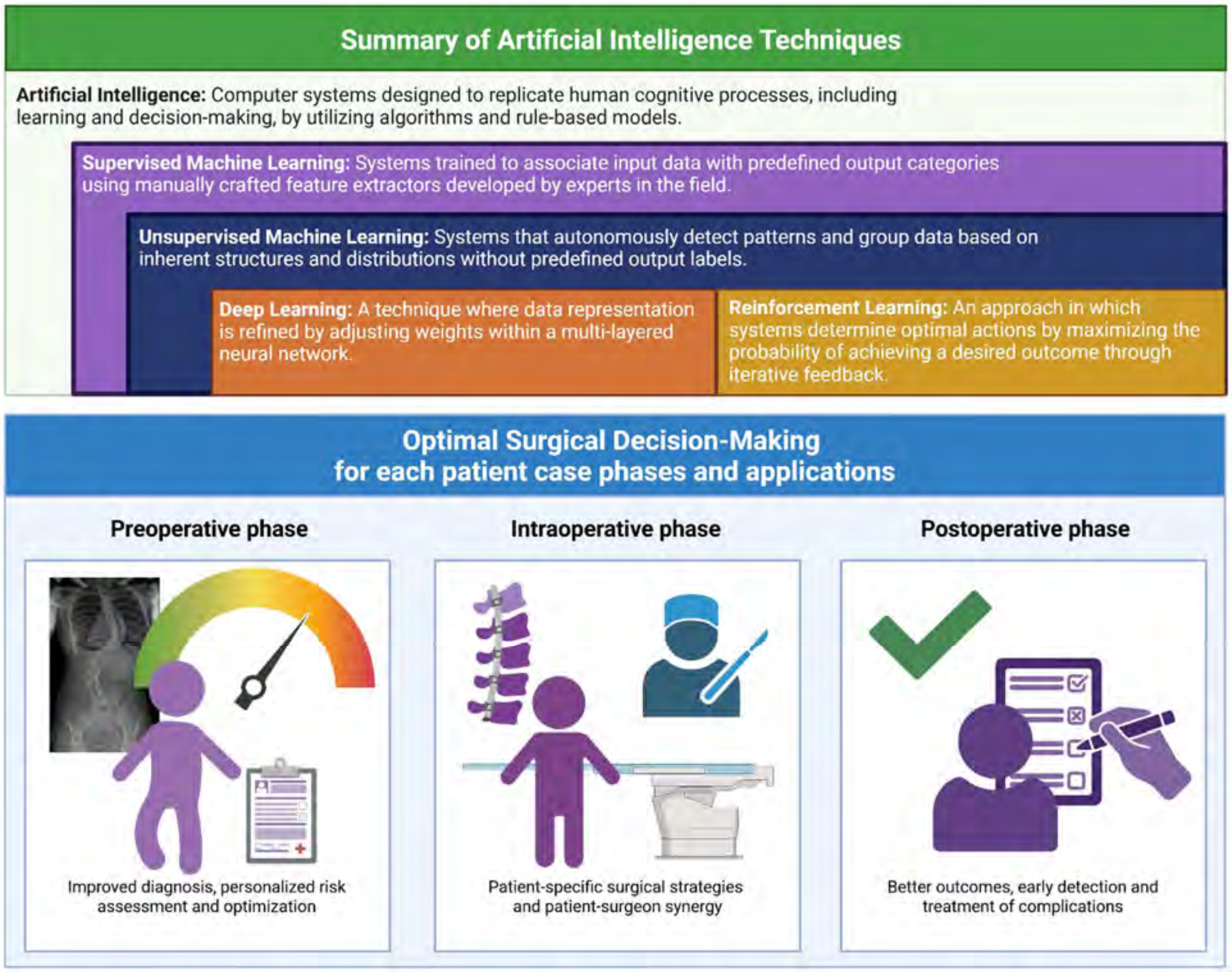


Figure 2.18 Overview of AI, Machine Learning, and Deep Learning Methods and Their Potential Applications in Surgery and Health Sciences.

Deep learning, a subset of ML, uses neural networks with multiple layers to learn high-level data representations, particularly effective in spine imaging due to its ability to learn relevant features from complex datasets with minimal manual input [222]. DL has demonstrated strong performance in scoliosis imaging tasks, including automated Cobb angle measurement [111, 112, 223-227] and identification of the Risser sign from EOS radiographs [121]. Smartphone-based and non-radiographic ML applications, such as laser scan–based and photographic scoliosis classifiers, have also achieved clinically relevant accuracy [113-115, 228, 229]. In surgical planning, ML has been applied to generate predictive models [230, 231], automate implant selection (e.g., screw diameter, angulation [232]), and assist in decision-making for complex cases. For example, Lafage et al developed a DL model that accurately predicted the UIV in 87.5% of adult deformity cases [233]. However, models tailored explicitly for instrumentation strategy in AIS are still lacking.

2.3.5 Critical Assessment on Computational and AI Approaches in Scoliosis Management

The integration of computational modeling and AI-driven approaches could mark a turning point in the diagnosis and treatment of scoliosis. These technologies have the potential to transform surgical planning, making it more precise, data-driven, and tailored to individual patients. By incorporating ML and MB biomechanical simulations, surgeons may soon be able to predict surgical outcomes with greater confidence, reducing some of the uncertainties that currently exist in scoliosis correction. Computational tools such as AI-assisted measurement systems and predictive modeling could refine surgical decision-making, helping to optimize instrumentation strategies and improve overall procedural accuracy. Beyond improving precision, these computational approaches might also play a critical role in reducing operative and postoperative risks. Detailed simulations and patient-specific biomechanical models could clarify how different instrumentation strategies affect spinal alignment and construct stability, allowing for a more proactive approach to potential complications. These technologies might help standardize procedures, enhance intraoperative decision-making, and ultimately contribute to safer, more effective surgeries if integrated into surgical workflows.

Another promising aspect of this transformation is the possibility of truly patient-specific surgical planning. Traditional methods of scoliosis correction rely heavily on the surgeon's expertise and general correction principles, but AI-assisted tools may help shift toward a more customized approach. By leveraging predictive analytics and advanced simulations, these technologies could ensure that each patient receives a surgical plan tailored to their specific spinal geometry, flexibility, and biomechanical response. This shift toward personalized surgery could lead to better alignment, fewer postoperative complications, and reduced need for revision surgeries. While these transformations are still in progress, the growing integration of AI and computational modeling could revolutionize the way scoliosis is managed in the coming years. With further refinement, these tools may become standard components of preoperative planning, intraoperative guidance, and even long-term postoperative monitoring. However, challenges remain: computational costs, validation requirements, and the need for greater clinical adoption could slow progress. Yet, as these technologies evolve, their potential to reshape scoliosis treatment is becoming increasingly apparent. If fully realized, AI and computational modeling could enhance surgical precision and redefine the standards of care for managing spinal deformities.

CHAPTER 3 RATIONALE, OBJECTIVES AND RESEARCH QUESTION

3.1 Summary of the Problems

The review of current knowledge highlighted the state and the limitations of previous work and identified the following issues related to the preoperative planning for PSF surgical instrumentation in AIS patients:

- There is a significant variability among surgeons in AIS instrumentation strategies, which is not only a matter of the surgeon's personal preference but is associated with biomechanical failures, inadequate deformity correction, and an increased need for revision surgery. The lack of a well-defined, patient-specific optimal surgical strategy may contribute to inconsistent postoperative outcomes and potentially over-treatment, as surgeons may prefer the 'less is more' approach.
- Despite this variability, AIS classification systems used to guide preoperative decision-making are still primarily based on 2D clinical assessment. These systems do not adequately capture the complex 3D nature of AIS, do not fully guide all instrumentation parameters influencing postoperative spinal alignment, and do not sufficiently reduce the risk of suboptimal correction.
- Postoperative complications, including revision surgery, occur frequently due to inadequate deformity correction, postoperative spinal misalignment, or mechanical failure of implants. Up to 13% of AIS patients need revision surgery within 5 years of initial surgical correction. Multiple "surgeon-modifiable factors" related to instrumentation, such as rod contouring, screw placement, and construct rigidity, have been identified as predictors of unfavorable outcomes. However, current planning strategies do not sufficiently integrate these biomechanical considerations.
- The computer-based preoperative planning tools currently available for scoliosis patients have weighty drawbacks limiting their use for AIS:
    - Implant arrays and correction strategies are often transmitted from surgeon to trainee through mentorship, shaped by clinical experience and individual case presentations. While these decisions may not always explicitly reference

biomechanical principles, they are grounded in experiential knowledge accumulated through years of surgical practice.

- Most assistive planning software relies on 2D determinants, making them insufficient for optimizing instrumentation in 3D deformities. Furthermore, they do not incorporate patient-specific biomechanical analysis to reduce the risk of postoperative misalignment or mechanical failure. Biomechanical simulation-based patient-specific models are currently limited in their ability to determine optimal instrumentation strategies. These models require extensive, time-consuming tasks for reconstructing the patient's spine and simulating surgical scenarios, limiting their practical use in a clinical setting.
- Deterministic numerical modeling approaches, while promising to provide a valuable first-order estimate of instrumentation correction and forces involved in the construct, suffer from high computational demands, as they require evaluating a large number of instrumentation configurations for each patient to identify an optimal strategy. Further optimized biomechanical strategies may be complex to integrate into current surgical practice, which is steeped in tradition.
- Current tools are predominantly focused on achieving 2D geometrical correction rather than addressing biomechanical stability. They fail to account for critical factors such as the forces and stresses acting on the construct, the need for optimal rod bending, the arrangement of screws, and the overall spinal balance. This lack of biomechanical integration results in suboptimal surgical planning and an increased risk of mechanical failure.

In summary, beyond the observed variability in surgical strategies, there is a critical need to improve the preoperative planning of AIS deformity correction to enhance postoperative outcomes. Current planning tools do not adequately address key biomechanical factors, including forces, stresses, and construct stability. A patient-specific, biomechanically informed approach to surgical instrumentation is necessary to decrease the risk of postoperative complications and improve long-term surgical outcomes.

3.2 Research Questions

The current state of knowledge calls for a more comprehensive determination of patient-specific instrumentation strategies to achieve safe and reliable PSF outcomes in AIS correction. This need led to the formulation of the following general and secondary research questions :

- To what extent can a hybrid approach, combining AI, a deterministic 3D biomechanical patient-specific model, and an optimization algorithm, accurately and precisely integrate key PSF parameters (UIV, LIV, screw density, and rod curvature) to optimize the planning and prediction of AIS correction based on individual patient characteristics?
    1) How does the combination of AI and deterministic biomechanical modeling of instrumentation parameters compare to standard clinical planning approaches in terms of predicted postoperative 3D scoliosis correction, residual spinal mobility, risks of implant-related biomechanical failure, and cost-effectiveness?
    2) How can a biomechanically optimized instrumentation strategy improve predicted surgical outcomes of AIS correction in terms of spinal stability, implant load distribution, and deformity correction compared to current clinical planning approaches?

3.3 Objectives

The <u>overall objective</u> of this doctoral project was to develop a tool that combines AI and physics-based deterministic model simulation to assist in PSF preoperative surgical planning for AIS patients.

More specifically, the project aimed to (Figure 3.1):

- **O1.** Develop an AI model that predicts surgical instrumentation parameters (UIV, LIV, screw density, and rod curvature) based on patient-specific preoperative information.
    - **SO1.1**: Systematically evaluate the current progress and application of deep learning in spine surgery.
    - **SO1.2**: Develop an artificial neural network using multi-task learning techniques to predict instrumentation parameters based on clinical and radiographic data.

- **O2.** Develop and validate a numerical approach that combines AI and a physics-based deterministic numerical model to simulate the biomechanical effects of different instrumentation strategies on 3D spinal correction, and assess the loads sustained by the surgical instrumentation.

- **O3.** Optimize instrumentation strategies by combining the hybrid model developed in O2 with an optimization approach to maximize spinal deformity correction while minimizing mechanical failure risks, preserving spinal mobility, and balancing surgical costs and risks.
  - **SO3.1:** Develop an optimization approach that incorporates AI-predicted instrumentation strategies (UIV, LIV, screw density, and rod curvature) while considering biomechanical and clinical constraints.
  - **SO3.2**: Compare the spinal correction and implant loading predicted by the numerical approach (SO3.1) with actual surgical instrumentation performed by surgeons.

- **O4.** Ensure the credibility of the developed computational approach by verifying, validating, and characterizing its uncertainties in accordance with established frameworks (e.g., ASME V&V 40). These verification, validation, and uncertainty quantification activities are integrated throughout the development of Objectives 1 to 3 to support their reliability in preoperative decision-making.

3.4 Hypotheses

The central premise of this work was to leverage optimized, patient-specific PSF surgical instrumentation configurations to effectively correct AIS curvatures. These configurations would be identified using a planning tool combining AI and deterministic numerical modeling.

More specifically, the research hypotheses of the project were (Figure 3.1):

- **H1.** An AI model based on multi-task artificial neural networks can generate surgically relevant instrumentation configurations (UIV, LIV, screw density, and rod characteristics) for AIS correction via PSF surgery. The model is expected to achieve a high level of accuracy (>85%; defined as the ratio of correct predictions to the total test dataset) and precision (>85%; defined as the ratio of correctly predicted positive observations to the total predicted positive observations, including both correct and incorrect), when compared to expert surgeon decisions (ground truth).

- **H2.** A combined approach using AI and a deterministic simulation model will predict surgical instrumentation configurations (UIV, LIV, screw density, and rod characteristics) that result in 3D postoperative spinal correction comparable to, or better than, those obtained from actual surgeries performed by experienced spine surgeons.
- **H3.** A planning tool combining AI, deterministic numerical model simulation, and optimization algorithms will generate instrumentation configurations that lead to improved clinical outcomes, including:
  - A greater correction of the major scoliotic curve (difference of >5°),
  - A reduction in implant usage (15% fewer implants),
  - Shorter spinal fusion constructs (at least one fewer fused level),
  - Lower biomechanical loading on the implants (average force reduction of ≥20%) compared to the configurations used in actual surgical procedures.

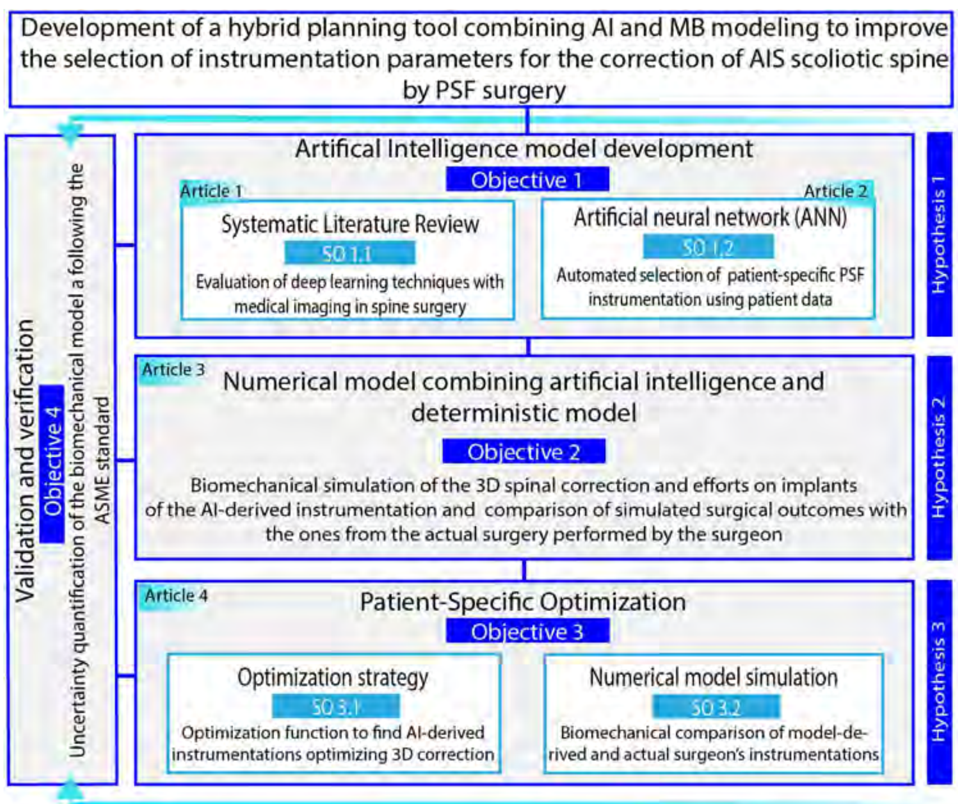


Figure 3.1 Organization of the Project according to the Research Question, Objectives, Sub-Objectives (SO), Hypotheses (H), and Resulting Articles

3.5 Methodology Outline

3.5.1 General Approach Outline

Overall, the research methodology was divided into four components to address each of the project objectives (Figure 3.1):

- The development of an original AI algorithm to predict patient-specific surgical instrumentation configurations, addressing objective 1.
- The development of a numerical model combining AI and deterministic model simulation to perform the biomechanical comparison of the AI-derived surgical instrumentation configurations proposed by the numerical approach and chosen by the surgeon, addressing objective 2.
- The development of a numerical model serving as a planning tool combining AI and deterministic numerical modeling, predicting optimized specific surgical instrumentation parameters, addressing objective 3.
- The verification and validation of the numerical approach performed throughout the previously enumerated components, addressing objective 4.

The achievement of these goals has resulted in the preparation of four scientific articles. The results from the systematic literature review are presented in *Article 1* (Section 4.1 from Chapter 4). The development of an artificial neural network capable of automatically configuring the instrumentation for surgery planning is presented in *Article 2* (Section 4.2 from Chapter 4). The combination of the AI model with the deterministic model to simulate instrumentation surgery, along with a comparison to the actual surgery, is presented in *Article 3* (Chapter 5). The final created hybrid model combines AI and simulation with a deterministic model and optimization process. This, along with the computational analysis of the optimized AI-derived surgical instrumentation configurations proposed by the numerical approach and chosen by the surgeon, leads to *Article 4* (Chapter 6).

3.5.2 Hybrid Planning Tool Development Outline

The AI algorithm was developed using patients diagnosed with Lenke 1-2 curves, representing approximately 50% of AIS cases [148], to demonstrate the feasibility of the proposed advanced patient-specific AIS scoliosis surgery planning tool. Overall, the methodology was divided into three main subcomponents: 1) automatic selection of instrumentation levels and surgical instrumentation variables, 2) biomechanical modeling of surgical instrumentation, and 3) optimization strategy. Briefly, a first DL algorithm using a neural network multitask learning model (NNML) was developed using retrospective data from patients diagnosed with AIS who underwent PSF and were previously enrolled in the Minimize Implants Maximize Outcomes (MIMO) Clinical Trial (NCT01792609). This neural network was trained to select PSF surgical instrumentation elements (UIV, LIV, screw density, and rod curvature) based on the patients' data, including the vertebrae location according to the identified landmarks on the patients' radiographs and alignment obtained from the actual surgery performed by the treating surgeon. Afterward, data from patients included in the "validation" and "performance test" subgroups during algorithm development were used to generate surgical instrumentation configurations. Their calibrated radiographs were then used to reconstruct the geometry of their spine and pelvis, enabling the digital modeling and simulation of the instrumentation proposed by the automated algorithm using a MB deterministic approach. Finally, an optimization algorithm used the deterministic model to refine the planning parameters.

CHAPTER 4 ARTIFICIAL INTELLIGENCE MODEL DEVELOPMENT

4.1 ARTICLE 1: The Use of Deep Learning in Medical Imaging to Improve Spine Care: a Scoping Review of Current Literature and Clinical Applications

To address sub-objective 1.1 (SO1.1), a systematic scoping review was conducted following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. The objective was to evaluate the current use of DL in medical imaging within spine care and to assess the scope, depth, and clinical relevance of existing studies. The review provides a comprehensive overview of the available literature, highlighting key clinical applications, model development strategies, and validation approaches.

This work led to the publication of the article titled *"The Use of Deep Learning in Medical Imaging to Improve Spine Care: a Scoping Review of Current Literature and Clinical Applications "* in the *North American Spine Society Journal* (NASSJ) in June 2023. The first author's contribution to the design, execution, writing, and revision of the manuscript is estimated at 85%.

Article 1:

Constant, C., Aubin, A-E., Maradit Kremers, H., et al. **The Use of Deep Learning in Medical Imaging to Improve Spine Care: a Scoping Review of Current Literature and Clinical Applications.** NASSJ (2023). https://doi.org/10.1016/j.xnsj.2023.100236.

Highlights

Most deep learning studies in the field of spine imaging focus on the detection and diagnosis of spinal conditions.

92% of deep learning imaging studies developed a new model, while 8% validated a pre-existing one

Deep learning in medical imaging showed promising performance in detection of imaging findings.

In the studies surveyed, implementation or demonstration of deep learning in real-world situations was rare in the studies surveyed.

Main Title: The use of deep learning in medical imaging to improve spine care: a scoping review of current literature and clinical applications

Author names: Caroline Constant[1,2,3], DMV, MSc, MENG, DACVS-LA, DECVS (caroline.constant@uzh.ch); Carl-Eric Aubin[2], Ph.D., ScD(h.c.), P.Eng., (Carl-Eric.Aubin@polymtl.ca), Hilal Maradit Kremers[1], M.D., MSc (maradit@mayo.edu), Diana V Vera Garcia[1], M.D. (veragarcia.diana@mayo.edu), Cody C Wyles[1,5], M.D. (Wyles.Cody@mayo.edu), Pouria Rouzrokh[1,4], M.D. (Rouzrokh.Pouria@mayo.edu), A. Noelle Larson[1,5], M.D (Larson.Noelle@mayo.edu).

Institutional affiliation:

[1] Orthopedic Surgery AI Laboratory, Mayo Clinic, 200 1st Street Southwest, Rochester, Minnesota, 55902, USA

[2] Polytechnique Montreal, 2500 Chem. de Polytechnique, Montréal, QC H3T 1J4, Canada

[3]AO Research Institute Davos, Clavadelerstrasse 8, CH 7270, Davos, Switzerland

[4] Radiology Informatics Laboratory, Mayo Clinic, 200, 1st Street Southwest, Rochester, Minnesota, 55902, USA

[5] Department of Orthopedic Surgery, Mayo Clinic, 200, 1st Street Southwest, Rochester, Minnesota, 55902, USA

Corresponding author: C. Constant; Department of Mechanical Engineering, Polytechnique Montreal, P.O. Box 6079, Downtown Station, Montreal, QC H3C 3A7, Canada; +41 79 910 69 76; caroline.constant@polymtl.ca

Funding disclosure(s) statement: This work was supported by the TransMedTech Institute and its main financial partner, the Apogee Canada First Research Excellence Fund, the NSERC/Medtronic Industrial Research Chair in Spine Biomechanics, Quattrone-Foderaro Grant in AI Innovation in Orthopedic Surgery 2021, Mayo Foundation, the Mayo Clinic, and the AO Foundation.

4.1.1 Abstract

Background: Artificial intelligence is a revolutionary technology that promises to assist clinicians in improving patient care. In radiology, deep learning is widely used in clinical decision aids due to its ability to analyze complex patterns and images. It allows for rapid, enhanced data and imaging analysis, from diagnosis to outcome prediction. The purpose of this study was to evaluate the current literature and clinical utilization of deep learning in spine imaging.

Methods: This study is a Scoping review and utilized the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology to review the scientific literature from 2012 to 2021. A search in PubMed, Web of Science, Embased, and IEEE Xplore databases with syntax specific for deep learning (DL) and medical imaging in spine care applications was conducted to collect all original publications on the subject. Specific data was extracted from the available literature, including algorithm application, algorithms tested, database type and size, algorithm training method, and outcome of interest.

Results: A total of 365 studies (total sample of 232 394 patients) were included and grouped into 4 general applications: diagnostic tools, clinical decision support tools, automated clinical/instrumentation assessment, and clinical outcome prediction. Notable disparities exist in the selected algorithms and the training across multiple disparate databases. The most frequently used algorithms were U-Net and ResNet. A DL model was developed and validated in 92% of included studies, while a pre-existing DL model was investigated in 8%. Of all developed models, only 15% of them have been externally validated.

Conclusions: Based on this scoping review, deep learning in spine imaging is used in a broad range of clinical applications, particularly for diagnosing spinal conditions. There is a wide variety of deep learning algorithms, database characteristics, and training methods. Future studies should focus on external validation of existing models before bringing them into clinical use.

Keywords: artificial intelligence, machine learning, deep learning, spine, imaging, clinical care, review

4.1.2 Introduction

Despite extensive research to improve spine care, spinal disorders remain prevalent.[2] A large body of evidence demonstrates the negative impact of spinal disorders on individuals and society, resulting in disabilities and considerable economic losses.[3] The annual cost of spine care has risen in the last decade,[3,4] suggesting a need for innovation to improve the care of patients with spinal conditions.

Medical imaging is critical for clinical decision-making and is integral to determining treatment indications and surgical outcomes. Various imaging modalities can be used for accurate detection and diagnosis of spinal pathologies.[5] However, despite following similar diagnostic standards, experienced radiologists can arrive at different diagnoses and measurements with error rates estimated to be 3 to 5%.[6] Much higher discrepancy rates are reported in neuroradiology, with variable reads in up to 21% of imaging studies.[7] Therefore, a key challenge in the diagnosis of spine pathology is improving the workflow to diagnose diseases quickly, automatically, and accurately. Further, medical care costs are increasing[3,4], and radiologists who train for 5 or more years after medical school represent an expensive and valuable resource. AI tools to augment the performance of radiologists and provide a low-cost tool to prevent errors hold significant promises for improved quality and cost savings. Reliable treatment planning, precise intervention, and accurate therapy are required when treating spine conditions to achieve optimal outcomes. While modern imaging techniques inform perioperative case management, [8] poor outcomes are still frequent. Thus, the development of scalable, perioperative automated assistance tools could help address this challenge.

Within health care, artificial intelligence (AI) algorithms are increasingly used for complex tasks, including remote patient monitoring, medical diagnosis and imaging, risk assessment, virtual assistance, hospital management, and drug discovery.[9] AI is a field of computer science that attempts to build enhanced "intelligence" into computer systems by implementing algorithms that apply rules to imitate reasoning and draw decisions (Figure 4.1).[10] Within AI, machine learning (ML) is a promising field for improving patient-specific spine care as it allows computers to learn without being explicitly programmed. ML can perceive important imaging trends that the average practitioner may not perceive.[11,12] To do so, ML uses provided data or previous experience to develop predictive models to determine subtle patterns and predict outcomes from a collection of statistical techniques.[11,12] In other words, ML techniques are based on available data with specific features (input data), which are used to train a machine (computer) to perform (to learn) the desired task generating a specific output (output data). The medical field has recently seen a fast improvement in ML techniques, specifically through deep learning (DL), an advanced form of machine learning capable of feature extraction to perform several tasks precisely developed to help clinicians.[13] In most circumstances, DL is based on neural networks (NN), network architectures formed of several layers, called hidden layers, containing multiple units, called artificial neurons, interconnected by mathematical relations called synapses. In each unit, a mathematical sum resulting from inputs' multiplication by a weight and inputs' summing to a bias term is processed by a linear or nonlinear activation function. In this context, the hidden layers help the network refine the input-output synapses between the units. DL attracts great interest for clinical application and image analysis in radiology partly due to its outstanding performance in image recognition, classification, and segmentation tasks.[9,14,15]

Narrative and systematic reviews have recently been completed on AI applications in the spine;[16-20] yet they did not systematically assess all published studies on the subject. Previous reviews have demonstrated that DL techniques are robust and scalable for spine care applications. However, no review has comprehensively mapped and assessed the quality of the clinical applications of DL combined with medical imaging for spinal diseases research. Thus, given the recent technological developments, we undertook a review of DL techniques for spine imaging applications to inform data scientists and clinicians on the methods and applications of big data in spine. Furthermore, we aim to highlight the challenges and limitations of DL techniques, identify gaps in the field, and outline potential opportunities for further research. Thus, this scoping review aims to broadly review and systematically evaluate the current progress in DL and how it has been applied to medical imaging and clinical applications intended for clinical spine care.

4.1.3 Materials and Methods

4.1.3.1 Search Strategy

A systematic literature review using a scoping review approach and following Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines[21] and PRISMA-ScR extension for Scoping Reviews guidelines[22] was carried out on October 15, 2021, and repeated on

January 1, 2022. The systematic research was constructed to identify studies describing DL in medical imaging for clinical spine applications. The final search terms and additional methodology details used are shown in *Appendix A – Article 1: Search Strategy.* First, the literature search was conducted through the health-related research database Medline. Next, the information technology database IEEE Xplore was searched. Lastly, databases that index both fields, including Web of Science and Embase, were searched for relevant literature.

4.1.3.2 Eligibility Criteria

Studies on DL and medical imaging dealing with applications intended for spine clinical care were selected. This review did not consider other AI methods based on fundamental ML techniques, preferring only DL-based approaches. Publications in peer-reviewed journals after 2012 were included. The beginning timepoint was selected as the first scalable convolutional neural network, significantly improving the state-of-the-art natural-image classification results.[23] Additionally, non-peered-review references published after 2012, such as case reports, proceedings or abstracts, and book chapters, were included. Studies that proposed solely technical applications without direct specific clinical implications, such as image segmentation, image quality improvement, and image reconstruction alone, were excluded. Animal experiments, reviews, correspondences, expert opinions, and editorials were also excluded. Two reviewers (CC, DVG) independently reviewed all studies, reaching a consensus on all included studies. A third reviewer resolved the disagreements in the inclusion process (NL).

4.1.3.3 Data Collection and Analysis

Data from all included studies were collected into a standardized data extraction sheet (Table 4.1). In addition, the key findings, clinical deployment, expected clinical benefits and value, economic implications, and ethical considerations of implementation in the health care system were recorded. To analyze the data, a narrative review synthesis method with descriptive statistics was selected to capture the extensive range of research investigating DL for spine clinical care. It should be noted that a meta-analysis was not appropriate for this review, given the broad range of spinal conditions, DL techniques, and types of data used in the studies identified. All retrieved manuscripts and abstracts were included in the qualitative synthesis.

4.1.4 Results

4.1.4.1 Overview of Studies Characteristics

The search strategies identified 1475 records, with 365 of these studies meeting the criteria for inclusion (Figure 4.2 and *Appendix B – Article 1: List of Included Studies*). The top three affiliated countries were the United States (17%), China (16%), and Canada (12%). Most studies were authored by multidisciplinary teams (61%), including experts from both medicine (most frequently neuro- or spine surgery and radiology) and engineering (most commonly electrical engineering, computer science, and/or data science), with the remaining articles authored by either medicine (19%) or engineering (14%) experts only, or could not be retrieved (9%). Most studies did not receive or disclose a significant contribution from an industrial partner (79%).

4.1.4.2 Study Type and Design

The study type could be identified for 352 published studies and was primarily classified as clinical research (64%; Figure 4.3). The information available in the remaining 13 studies was insufficient to classify them. The division of the studies into retrospective or prospective investigations was impossible in 96 studies (26%), meaning that we could not distinguish if the data were collected explicitly for the study or from an existing data source. Data were collected retrospectively for the remaining studies (n=269) (88%).

The study design could be identified for 318 of the published studies and was predominantly categorized as cross-sectional (47%) and descriptive studies without comparison groups (43%), and less frequently as cohort studies (9%). The information available in the remaining 47 studies (13%) was insufficient to determine the study design. For the particular case of clinical research for the diagnostic, prognostic, and predictive test accuracy studies, the top three study designs were cross-sectional studies (56%), descriptive studies without comparison groups (26%), and cohort studies (8%).

4.1.4.3 Spine Clinical Care and Imaging Focus

The clinical application of DL in spine image analysis was most frequently related to developing or validating new diagnostic tools (74%; Figure 4.4A) and primarily included studies that aimed to identify or diagnose spinal conditions. Three main themes arose from these diagnostic studies: 1) detection of diseases by developing pre-diagnosis screening tools or anomaly detection, 2) predicting the diagnosis of new patients based on a training dataset of prior diagnoses, and 3) differentiating between spinal conditions with similar imaging features or symptomatology. Other frequently identified clinical themes were clinical decision support, assessment, and outcome prediction studies, mainly aimed at predicting the progression of spinal conditions, exploring treatment possibilities, or supporting clinical opportunities for such conditions. Two main themes were identified among studies examining clinical decision support tools: 1) identifying preoperative factors to provide personalized and timely treatment or surgical interventions and 2) supporting procedures such as injections under imaging guidance and surgical navigation. Studies investigating prognosis primarily focused on using DL to predict the development of complications following spinal surgery. Other clinical applications found in the reviewed studies included public health investigations, which used large epidemiological or public datasets to monitor or screen for spinal conditions in the general population, describe average spinal measurements, or estimate disease prevalence; and clinical administration tools, which included studies that aimed at improving administrative processes in clinical work and healthcare organizations such as automated report generation. MRI was the most common imaging modality used in the reviewed studies (36%; Figure 4.4B), and the most frequently used MRI sequences were T2-weighted alone (53%), a combination of T1-weighted and T2-weighted (26%), and T1-weighted alone (13%). Most studies used DL techniques for spinal conditions or clinical care targeting a single spinal region (47%). The majority of the remaining studies investigated more than one spinal region (35%). In contrast, only a minority of studies explored DL techniques on the whole spine (14%) or did not provide enough information to identify a specific spine region (5%). The lumbar spinal segment was the most routinely studied region and was included alone or in combination with other spinal regions(s) in 79% of all investigations (Figure 4.5).

Through analysis of the data, five main domains of spinal conditions were identified, with the top 3 being inflammatory and degenerative conditions (26%), spinal deformity and alignment problems (22%), and fractures (14%; Figure 4.6). Among the studies that investigated inflammatory or degenerative spine conditions, MRI was routinely used (82%), and improvement or automation of the diagnosis was the main cited objective (89%). Automated tissue classification and measurements were the most investigated DL pipeline outcomes (67% and 18%, respectively). Among the studies targeting spinal alignment problems, radiographic images were generally used

(69%), with the main objective of improving measurement accuracy for sagittal or coronal balance (85%). Spinal alignment measurements were the most commonly investigated pipeline outcome (66%). Studies addressing spinal fractures consistently aimed to improve the accuracy and speed of diagnosis of vertebral fractures and detection of fractured levels (90%). To do so, CT scans and radiographs were the most common imaging modality used (58% and 20%, respectively), and automated vertebral status classification and fracture detection with or without specific anatomical localization were the most commonly studied pipeline outcomes (50% and 32%, respectively).

The studies targeting surgical conditions had multiple objectives, but primarily provided clinical decision support tools and automated instrumentation assessment (56% and 50%, respectively) and commonly used ultrasound or fluoroscopy as imaging modalities (33% and 27%, respectively). While less common, neoplasia was another recurrent condition investigated (8%), with most studies aimed at improving diagnosis accuracy and lesion delimitation (65%) using mainly MRI imaging (62%). To this aim, automated lesion detection or delimitation and tissue classification were the most commonly investigated pipeline outcomes (50% and 35%, respectively). Six percent of studies included in this review examined osteoporosis and mainly intended to detect bone properties abnormalities or diagnose osteoporosis automatically (80%), primarily using CT scans and DEXA as imaging modalities (48% and 33%, respectively). Within those, the primary pipeline outcomes were bone status classification and bone properties measurements, such as bone mineral density (42% and 38%, respectively). The overall publication rate for presented abstracts published as full-length articles in peer-reviewed journals was 6%.

4.1.4.4 Subjects, Images, and Datasets

Only 41% of the reviewed studies clearly reported the number of subjects and medical images included in the research. The number of subjects included was reported more often than the number of images (33% vs. 19%). Most studies included 100 to 1000 subjects or images (Figure 4.7) with a mean of 945 and 9504, respectively. When the studies targeted a particular spinal condition, subjects with a relevant positive diagnosis were regularly enrolled (73%) with a mean enrollment ratio of 79% and 71% of all subjects and images, respectively. Seventy-five percent of studies reported whether their included subjects or images had surgical implants, with most stating the exclusion of them if patients had spinal implants (75%). Only 30% of studies reported the health status of the enrolled subjects or the presence of other spinal diseases on the images included, with most studies not including data about other conditions or diseases (55%). For the studies that included the DL training phase, the mean of subjects and images in the DL development phase were 763 and 4286, respectively.

The origin of the data could be retrieved for 288 of the published studies and was primarily part of institutional data (58%), registries or clinical databases (11%; Table 4.2), public datasets (4%; Table 4.3), or a combination (4%). When institutional data, registries, or clinical databases were used, the data originated more commonly from a single center rather than multiple centers (68% vs. 32%).

4.1.4.5 DL Method and Architecture

In 84% of the included studies, images were used as input through DL computer vision tasks. In contrast, the others used extracted features from images or collected data from medical imaging results (e.g. disc height) as input. Computer vision tasks were, in general, performed by three different DL approaches 1) landmark detection, often combined with prior structure detection, 2) structure segmentation, or 3) shape model matching. To this end, the preferred DL technique was convolutional neural network (CNN) which was used in 77% of included studies. CNN techniques were investigated mainly for classification (43%), measurement tasks by structure segmentation or landmark detection (26%), and detection tasks (20%). For most classification tasks, convolutional and pooling layers were first used to extract features from the input images, followed by fully connected layers for output feature classification. Measurement tasks using landmark detection were usually performed similarly or through segmentation tasks where the fully connected layer of the CNN was replaced with up-sampling layers (encoder-decoder architecture), resulting in the conversion of landmarks into segmentable heatmaps.

Sixty-four percent of studies adapted an existing DL architecture, most commonly based on U-Net (21%) and ResNet (16%), all of which operated on image data (Figure 4.8 and Table 4.4). Studies using extracted features from images or collected data from medical imaging results as input preferably used MLP network structures, which were used in 21% of included studies. Other DL methodologies were reported, including long short-term memory layer, radial basis function network, and self-organizing map, but were only investigated in 2% of the included studies. Most studies investigated only one DL architecture (57%) with a mean ± SD of 1.8 ± 1.3 DL architectures investigated per study. Nevertheless, the performance of several DL pipelines was investigated in most studies (91%) with a mean ± SD of 2.6 ± 2.8 pipelines or models investigated per study. Sixteen percent of studies also surveyed other ML techniques as an alternative to DL in their pipelines or as comparison results, with support vector machine and K-Nearest Neighbor algorithms being the most frequently reported.

4.1.4.6 DL Training and Validation of Studies with DL Development

A DL model was developed and internally validated in 92% of included studies, while a pre-existing DL model was externally investigated in 8%. The dataset split into training, validation, and testing was mentioned in 155 (47%) development studies. When available, the mean number of subjects and images used in the DL development phase was 763 and 4286, respectively. The mean proportions of data split into the training, validation, and testing datasets were 0.75, 0.18, and 0.20, respectively. Of the development studies, 26% described the prevention of overfitting using cross-validation. Only 15% of development studies externally validated the completed DL pipeline on a data set distinct from the training dataset. Of these, 43% performed external validation using data from a completely different origin, such as from a foreign country or other hospitals. Internal validation was the only validation technique in the remaining studies (85%).

4.1.4.7 Evaluation of Performances

A large variety of performance metrics were retrieved from the studies to evaluate the similarity between the DL prediction and the ground truth (Table 4.5). For DL pipelines with classification tasks as outputs (binary as well as multi-class problems), the measurement of performance, whenever reported, generally included sensitivity and specificity of the technique (58%) and area under the curve (51%). The DL pipelines with detection tasks as outputs mainly reported the precision (48%) for localizing the object's position in the image and recall (56%) when judging whether objects belonging to certain classes appear in regions of interest. The DL pipelines with measurement tasks outputs frequently used various error calculations to evaluate the model performances, principally the mean absolute error (20%) and standard error (20%).

Only 8 peer-reviewed research articles (4%) adhered to a reporting checklist, including 3 guidelines for transparent reporting of predictive or artificial intelligence models[234-236] and one for reporting diagnostic accuracy studies [237]. A minority of studies (7%) provided the code used for model development, final trained models, or both (Table 4.6). In addition, 5% of studies used proprietary DL models or commercial prototypes.

4.1.5 Discussion

This scoping review synthesizes the recent literature on the use of DL techniques combined with medical imaging for spine clinical applications, highlighting current research and applications in clinical practice. Most of the included studies were observational clinical research and used existing datasets. For the most part, studies focused on the benefits of DL in combination with medical imaging to improve the detection and diagnosis of spinal conditions, such as inflammatory conditions, degenerative disease, and spinal deformity. Various DL methods and architectures were explored, and some studies proposed novel ones. Many DL approaches showed promising performance, demonstrating the potential of DL in the management of spinal conditions to improve the efficiency of clinical care and research.

4.1.5.1 Overall Quality of the Studies

The quality and robustness of the DL models and possible clinical implications heavily rely on the quality of the research and the input datasets, which governs the extent to which its findings can be trusted. The broad objective of this scoping review was to include studies from peer-reviewed journals and other sources such as preprint servers and conference proceedings. The peer review process is designed to critically assess the relevance of new research as well as to insure the appropriate study design and data analysis. Therefore, peer-reviewed studies are usually assumed to have at least a minimum acceptable quality. While peer-reviewed publication of completed results remains the primary goal of most medical research, a substantial number of abstracts will not be published in the medical literature. Therefore, including only articles published in peer-review journals would have limited our ability to broadly review and comprehensively map the clinical applications of DL combined with medical imaging since 30% of the included studies were obtained from outside this publication process. Nevertheless, only 6% of included conference abstracts went on to peer-reviewed publication, highlighting the need to use caution before integrating the results of these abstracts into research or clinical practice. Thus, the contribution of conference abstracts in creating a solid body of evidence required for clinical implementation remains uncertain.

Clear and transparent reporting is crucial in assessing a study's quality. Nevertheless, only 4% of the included peer-reviewed articles adhered to a reporting checklist. The quality assessment of the studies included in this review was beyond the scope of this investigation but would be an interesting area for future research. Nevertheless, the data extracted from the included studies allowed us to make a general quality assessment of the eligible studies based on checklists developed explicitly for AI in medical imaging based (CLAIM, MI-CLAIM).[28,29] Our review suggests that many studies would be classified as “incomplete” for numerous checklist items, reflecting potential methodological limitations. In many cases, the study design was incomplete; 60% of studies did not detail the cohorts' characteristics. Similar observations can be made for data and optimization items (21% did not describe the origin of the data, 47% did not report the split of the dataset) and reproducibility and transparency reporting (93% failed to provide the code used for model development or final trained models). Similar to our experience, previous medical studies focusing on AI have also noted a lack of data reporting and poor model transparency.[18,30] Implementing a standardized mandatory checklist into the DL peer-review process, as it is currently done with the STROBE checklist for human observational studies,[31] could help enhance the quality of the published studies and improve model reproducibility and comparison.[18] In turn, improving the quality of the research and robustness of the DL models may accelerate their implementation into clinical practice.

4.1.5.2 Datasets and DL reliability

Data quality and availability are significant determinants of models' performance and reliability and have been recognized as a fundamental challenge to developing DL for medical imaging.[32] In the current review, issues similar to previously reported limitations regarding the data's quality[32] were raised, including imbalanced data, lack of adequately annotated data, and limited confidence intervals. While the correct sample size required to train a DL model to perform adequately is challenging to estimate in advance, the reported datasets seem very limited compared to datasets for general computer vision tasks, which typically range from a hundred thousand to millions of annotated pictures.[33] One likely explanation for the difference in dataset size is the limited number of samples and patients currently available in the public databases for medical imaging tasks compared to public databases available for general computer vision tasks. Nevertheless, the studies included in this review commonly reported good DL performance despite potential issues related to data quantity. Still, it remains unclear how well the final DL models perform their task regarding over-fitting to their training datasets.

4.1.5.3 DL Model Performance

Most of the studies included in this review (91%) evaluated several DL pipelines or multiple types or DL models. The subsequent identification of the best model was usually based on comparing their performance using various metrics, predominantly calculated by comparing DL predictions against reference data obtained from human observers, as it is common practice in AI.[20] The predominantly used performance metrics were probabilistic measurements, including accuracy, sensitivity and specificity, and AUC. However, these metrics have considerable limitations and cannot be considered reliable in some situations frequently encountered in the reviewed studies. The main limitation of their use was the label imbalance observed in most datasets that included diseased patients. Using accuracy as an indicator of performance with an unbalanced dataset may artificially improve the performances due to this sensitivity of accuracy to the prevalence of positive diagnosis in a dataset and the tendency of this performance metric to favor the majority class.[34] These studies were then prone to positive-negative class bias and misleading models' performances in such situations due to limited pre-test probability assessment.[35] Acknowledging potential bias or study limitations is vital for accurate result interpretation, especially when claims are made regarding clinical care. Nevertheless, it is unclear if the studies included in the review accounted for the label class imbalance, specifically the number of healthy and diseased or positives and negatives in the dataset.

4.1.5.4 Clinical Implementation and Ethics

Caution is needed when developing DL methodologies for clinical practice. Before clinical implementation, external validation and replication of the DL models' performances should be completed. Compared to internal validation, external validation allows a more robust demonstration of the clinical utility of the methodology. Nevertheless, only a minority of studies (8%) investigated pre-existing DL models, and few of the remaining studies (15%), which were developing DL models, externally validated the completed DL pipeline on a dataset distinct from their training dataset. Although a tool may appear promising in a particular setting, they are unlikely to perform the same after being deployed into different spine clinical care settings, particularly if employed across different patient populations. Nevertheless, very few studies provided information or demonstrated the use of DL techniques in real-world situations, suggesting that further consideration and research are required to test such models' clinical utility and applicability. For all the reasons mentioned earlier, the field of DL combined with medical imaging for spine clinical applications does not appear ready for widespread clinical recognition and remains in its development phase. As such, DL does not replace other research or analytic approaches; instead, it can potentially add value to the available tools for spine clinical care research. Partnerships between clinicians and data science experts are essential to ensure the clinical utility of the DL models developed for healthcare.

4.1.6 Future Research Directions and Conclusions

The DL and medical imaging for spine clinical care is an emerging research field with exciting recent developments with the potential to improve patient care. This review of 365 studies showed that problem-specific DL models could significantly improve the detection and diagnosis of spinal conditions on medical imaging. While it is evident that DL is unlikely to replace radiologists or other health experts in the near future, it holds the potential to be an efficient tool to decrease the clinical burden of radiologists and clinicians. Though less frequently investigated, research into other applications of DL, such as clinical decision support, assessment, and outcome prediction, has demonstrated initial positive results. Nevertheless, the available studies on these topics are currently limited, and further research is required to identify additional benefits of DL for spine clinical care. The analysis of the included studies highlighted the following needs for further studies or improvement: 1) commitment to data and model transparency, 2) reproducibility and generalizability improvement of DL models, 3) performing external and comprehensive validation of the proposed methodologies on different datasets, and 4) establishing of DL ethics guidelines at all levels of DL development. In addition, efforts to improve research methodologies and the impact of DL on patients should be better considered. Furthermore, standard imaging protocols, agreed-upon datasets to perform DL models' benchmarking, standardized performance metrics, and unbiased accuracy indicators appeared to be lacking in the current literature. Such standards would improve the quality of AI research and allow for better clinical implementation and further advances in the field of DL for spinal imaging.

The generation of large spine datasets combined with DL tools accessible to researchers and clinicians is also needed to support the development of novel DL applications and improve the current spine clinical care models. More work is needed to define best practices with DL tools to guide clinical-decision making process for spine clinical care and to facilitate eventual clinical implementation.

4.1.7 References

1. Rohrig B, du Prel JB, Wachtlin D, et al. Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009;106:262–8 .

2. Andersson G, Frymoyer J. The adult spine: principles and practice. The adult spine: principles and practice. Philadelphia: Lippincott-Raven 1997:93–141 .

3. Lo J, Chan L, Flynn S. A systematic review of the incidence, prevalence, costs, and activity and work limitations of amputation, osteoarthritis, rheumatoid arthritis, back pain, multiple sclerosis, spinal cord injury, stroke, and traumatic brain injury in the United States: a 2019 update. Arch Phys Med Rehabil 2021;102:115–31 .

4. Diebo BG, Shah NV, Boachie-Adjei O, et al. Adult spinal deformity. Lancet 2019;394:160–72.

5. Ma VY, Chan L, Carruthers KJ. Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain. Arch Phys Med Rehabil 2014;95:986–995.e981 .

6. Kim GU, Chang MC, Kim TU, et al. Diagnostic modality in spine disease: a review. Asian Spine J 2020;14:910–20 .

7. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging 2017;8:171–82 .

8. Briggs GM, Flynn PA, Worthington M, et al. The role of specialist neuroradiology second opinion reporting: is there added value? Clin Radiol 2008;63:791–5 .

9. Sun J, Wu D, Wang Q, et al. Pedicle screw insertion: is O-arm–based navigation superior to the conventional freehand technique? A systematic review and meta–analysis. World Neurosurg 2020;144:e87–99 .

10. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. Nat Rev Cancer 2018;18:500–10 .

11. Chang M, Canseco JA, Nicholson KJ, et al. The role of machine learning in spine surgery: the future is now. Front Surg 2020;7:54 .

12. Samuel AL. Some studies in machine learning using the game of checkers. IBM J Res Dev 2000;44:206–26 .

13. Schmidhuber J. Deep learning in neural networks: an overview. Neural networks 2015;61:85–117 .

14. Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. Radiographics 2017;37:2113–31 .

15. Erickson BJ, Korfiatis P, Akkus Z, et al. Toolkits and libraries for deep learning. J Digital Imaging 2017;30:400–5 .

16. Huang J, Shlobin NA, DeCuypere M, et al. Deep learning for outcome prediction in neurosurgery: a systematic review of design, reporting, and reproducibility. Neurosurgery 2022;90:16–38 .

17. Azimi P, Yazdanian T, Benzel EC, et al. A review on the use of artificial intelligence in spinal diseases. Asian Spine J 2020;14:543–71 .

18. Smets J, Shevroja E, Hügle T, et al. Machine learning solutions for osteoporosis-a review. J Bone Miner Res 2021;36:833–51 .

19. Ong W, Zhu L, Zhang W, et al. Application of artificial intelligence methods for imaging of spinal metastasis. Cancers (Basel) 2022;14:4025 .

20. Vrtovec T, Ibragimov B. Spinopelvic measurements of sagittal balance with deep learning: systematic review and critical evaluation. Eur Spine J 2022;31:2031–45 .

21. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8:336–41 .

22. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA- ScR): checklist and explanation. Ann Intern Med 2018;169:467–73 .

23. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM 2017;60:84–90 .

24. Mathes T, Pieper D. An algorithm for the classification of study designs to assess diagnostic, prognostic and predictive test accuracy in systematic reviews. Syst Rev 2019;8:226 .

25. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Bmj 2015;350:g7594 .

26. Schwendicke F, Singh T, Lee JH, et al. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. J Dent 2021;107:103610 .

27. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18:e323 .

28. Bossuyt PM, Reitsma JB, Bruns DE, et al. [Reporting studies of diagnostic accuracy according to a standard method; the Standards for Reporting of Diagnostic Accuracy (STARD)]. Ned Tijdschr Geneeskd 2003;147:336–40 .

29. Yang J, Zhang K, Fan H, et al. Development and validation of deep learning algorithms for scoliosis screening using back images. Commun Biol 2019;2:390 .

30. Ferizi U, Honig S, Chang G. Artificial intelligence, osteoporosis and fragility fractures. Curr Opin Rheumatol 2019;31:368–75 .

31. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007;4:e296 .

32. Minaee S, Kafieh R, Sonka M, et al. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. Med Image Anal 2020;65:101794 .

33. Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 2020;43:635–40 .

34. Tecle N, Teitel J, Morris MR, et al. Convolutional Neural Network for Second Metacarpal Radiographic Osteoporosis Screening. J Hand Surg 2020;45:175–81 .

35. Branco P, Torgo L, Ribeiro RP. A survey of predictive modeling on imbalanced domains. ACM Computing Surveys (CSUR) 2016;49:1–50 .

36. Manjer J, Carlsson S, Elmståhl S, et al. The Malmö Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants. Eur J Cancer Prev 2001;10:489–99 .

37. Evaniew N, Cadotte DW, Dea N, et al. Clinical predictors of achieving the minimal clinically important difference after surgery for cervical spondylotic myelopathy: an external validation study from the Canadian Spine Outcomes and Research Network. J Neurosurg Spine 2020:1–9 .

38. Rudwaleit M, Haibel H, Baraliakos X, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. Arthritis Rheum 2009;60:717–27 .

39. Hu XJ, Chen LH, Battié MC, et al. Methodology and cohort profile for the Hangzhou Lumbar Spine Study: a study focusing on back health in a Chinese population. J Zhejiang Univ Sci B 2018;19:547–58 .

40. Lee C, Choe EK, Choi JM, et al. Health and Prevention Enhancement (H-PEACE): a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea. BMJ Open 2018;8:e019327 .

41. Stern D, Likar B, Pernu š F, et al. Parametric modelling and segmentation of vertebral bodies in 3D CT and MR spine images. Phys Med Biol 2011;56:7505–22 .

42. Ibragimov B, Likar B, Pernu š F, et al. Shape representation for efficient landmark- based segmentation in 3-d. IEEE Trans Med Imaging 2014;33:861–74 .

43. Leslie WD, Caetano PA, Macwilliam LR, et al. Construction and validation of a population-based bone densitometry database. J Clin Densitom 2005;8:25–30 .

44. Berglund G, Nilsson P, Eriksson KF, et al. Long-term outcome of the Malmö preventive project: mortality and cardiovascular morbidity. J Intern Med 2000;247:19–29 .

45. Shardell M, Parimi N, Langsetmo L, et al. Comparing analytical methods for the gut microbiome and aging: gut microbial communities and body weight in the osteoporotic fractures in men (MrOS) study. J Gerontol A Biol Sci Med Sci 2020;75:1267–75 .

46. Rikkonen T, Salovaara K, Sirola J, et al. Physical activity slows femoral bone loss but promotes wrist fractures in postmenopausal women: a 15-year follow-up of the OSTPRE study. J Bone Miner Res 2010;25:2332–40 .

47. Salovaara K, Tuppurainen M, Kärkkäinen M, et al. Effect of vitamin D(3) and calcium on fracture risk in 65- to 71-year-old women: a population-based 3-year randomized, controlled trial–the OSTPRE-FPS. J Bone Miner Res 2010;25:1487–95 .

48. Poddubnyy D, Sieper J, Akar S, et al. Characteristics of patients with axial spondyloarthritis by geographic regions: PROOF multicountry observational study baseline results. Rheumatology (Oxford) 2022;61:3299–308 .

49. Tsolinas RE, Burke JF, DiGiorgio AM, et al. Transforming Research and Clinical Knowledge in Spinal Cord Injury (TRACK-SCI): an overview of initial enrollment and demographics. Neurosurg Focus 2020;48:E6 .

50. Ishimoto Y, Yoshimura N, Muraki S, et al. Prevalence of symptomatic lumbar spinal stenosis and its association with physical performance in a population-based cohort in Japan: the Wakayama Spine Study. Osteoarthritis Cartilage 2012;20:1103–8 .

51. Modarresi S, MacDermid JC, Suh N, et al. How is the probability of reporting various levels of pain 12 months after noncatastrophic injuries associated with the level of peritraumatic distress? Clin Orthop Relat Res 2022;480:226–34 .

52. Yao J, Burns JE, Forsberg D, et al. A multi-center milestone study of clinical vertebral CT segmentation. Comput Med Imaging Graph 2016;49:16–28 .

53. Ibragimov B, Korez R, Likar B, et al. Segmentation of pathological structures by landmark-assisted deformable models. IEEE Trans Med Imaging 2017;36:1457–69 .

54. De Leener B, Fonov VS, Collins DL, et al. PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. Neuroimage 2018;165:170–9 .

55. Wu H, Bailey C, Rasoulinejad P, et al. In: Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet. Quebec, Canada: Springer; 2017. p. 127–35 .

56. Sekuboyina A, Bayat A, Husseini ME, et al. Verse: a vertebrae labelling and segmentation benchmark. arXiv org e-Print archive 2020;73:102166 . arXiv preprint arXiv: 2001.09193 .

57. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013;26:1045–57 .

58. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556 2014. https://arxiv.org/abs/1409.1556 .

59. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention – MICCAI 2015. MICCAI 2015. Lecture notes in computer science, 9351. Cham: Springer; 2015. doi: 101007/978-3-319-24574-4_28 .

60. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE Computer Society; 2016. p. 770–8 .

61. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, NV: IEEE Computer Society; 2016. p. 2818–26 .

62. Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning Thirty-first AAAI conference on artificial intelligence. San Francisco, CA: AAAI Press; 2017 .

63. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE Computer Society; 2016. p. 779–88 .

64. Redmon J, Farhadi A. Yolov3: an incremental improvement. arXiv preprint arXiv:180402767 2018. https://arxiv.org/abs/1804.02767 .

65. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–8.

66. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 2015;15:29 .

67. Popovic A, de la Fuente M, Engelhardt M, et al. Statistical validation metric for accuracy assessment in medical image segmentation. Int J Computer Assisted Radiol Surg 2007;2:169–81 .

68. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 2020. arXiv preprint arXiv:201016061 https://arxiv.org/abs/2010.16061 .

69. Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. J Med Im aging (Bellingham) 2018;5:015006 .

70. Heimann T, Bv Ginneken, Styner MA, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans Med Imaging 2009;28:1251–65 .

71. Pang S, Su Z, Leung S, et al. Direct automated quantitative measurement of spine by cascade amplifier regression network with manifold regularization. Med Image Anal 2019;55:103–15 .

72. Pang S, Leung S, Nachum IB, et al. Direct automated quantitative measurement of spine via cascade amplifier regression network. arXiv preprint arXiv:180605570 2018. https://arxiv.org/abs/1806.05570 .

73. Li YC, Chen HH, Horng-Shing Lu H, et al. Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists? Clin Orthop Relat Res 2021;479:1598–612 .

74. Derkatch S, Kirby C, Kimelman D, et al. Identification of vertebral fractures by convolutional neural networks to predict nonvertebral and hip fractures: a Registry-based cohort study of dual X-ray absorptiometry. Radiology 2019;293:405–11 .

75. Tang C, Zhang W, Li H, et al. CNN-based qualitative detection of bone mineral density via diagnostic CT slices for osteoporosis screening. Osteoporos Int 2021;32:971–9 .

76. Han Z, Wei B, Leung S, et al. Automated pathogenesis-based diagnosis of lumbar neural foraminal stenosis via deep multiscale multitask learning. Neuroinformatics 2018;16:325–37 .

77. Rzecki K, Kucyba ł a I, Gut D, et al. Fully automated algorithm for the detection of bone marrow oedema lesions in patients with axial spondyloarthritis – Feasibility study. Biocybernetics Biomed Eng 2021;41:833–53 .

78. Merali Z, Wang JZ, Badhiwala JH, et al. A deep learning model for detection of cervical spinal cord compression in MRI scans. Sci Rep 2021;11:10473 .

79. Tran VL, Lin H-Y, Liu H-W. MBNet: a multi-task deep neural network for semantic segmentation and lumbar vertebra inspection on X-ray images, in computer vision – ACCV 2020 2021;635-651.

80. Yang G, Fu X, Xu N, et al. A Landmark estimation and correction network for automated measurement of sagittal spinal parameters. In: Yang H, Pasupa K, Leung, ACS, Kwok JT, Chan JH, King I, editors. Neural information processing. ICONIP 2020. Communications in computer and information science, vol 1332. Cham: Springer; 2020.

81. Gros C, De Leener B, Badji A, et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. Neuroimage 2019;184:901–15 .

82. Ungi T, Greer H, Sunderland KR, et al. Automatic spine ultrasound segmentation for scoliosis visualization and measurement. IEEE Trans Biomed Eng 2020;67:3234–41 .

83. Wu V, Ungi T, Sunderland K, et al. Automatic segmentation of spinal ultrasound landmarks with U-net using multiple consecutive images for input, 2021.

84. Zhou J, Damasceno PF, Chachad R, et al. Automatic vertebral body segmentation based on deep learning of dixon images for bone marrow fat fraction quantification. Front Endocrinol (Lausanne) 2020;11:612 .

85. Zhang B, Yu K, Ning Z, et al. Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: a multicenter retrospective cohort study. Bone 2020;140:115561 .

86. Weber KA 2nd, Abbott R, Bojilov V, et al. Multi-muscle deep learning segmentation to automate the quantification of muscle fat infiltration in cervical spine conditions. Sci Rep 2021;11:16567 .

87. Zhang C, Wang J, He J, et al. Automated vertebral landmarks and spinal curvature estimation using non-directional part affinity fields. Neurocomputing 2021;438:280–9 .

88. Wu H, Bailey C, Rasoulinejad P, et al. In: Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet, 2017; 2017. p. 127–35 .

89. Hallinan JTPD, Zhu L, Yang K, et al. Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI. Radiology 2021;300:130–8 .

90. Chen W, Liu X, Li K, et al. A deep-learning model for identifying fresh vertebral compression fractures on digital radiography. Eur Radiol 2021;32:1496–505 .

91. Kitamura G. Hanging protocol optimization of lumbar spine radiographs with machine learning. Skeletal Radiol 2021;50:1809–19 .

92. Loffler MT, Jacob A, Scharr A, et al. Automatic opportunistic osteoporosis screening in routine CT: improved prediction of patients with prevalent vertebral fractures compared to DXA. Eur Radiol 2021;31:6069–77 .

93. De Leener B, Lévy S, Dupont SM, et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 2017;145:24–43 .

94. Lemay A, Gros C, Zhuo Z, et al. Automatic multiclass intramedullary spinal cord tumor segmentation on MRI with deep learning. Neuroimage Clin 2021;31:102766 .

95. Wolf K, Reisert M, Beltrán SF, et al. Spinal cord motion in degenerative cervical myelopathy: the level of the stenotic segment and gender cause altered pathodynamics. J Clin Med 2021;10:3788 .

96. Gibson E, Li W, Sudre C, et al. NiftyNet: a deep-learning platform for medical imaging.

Comput Methods Programs Biomed 2018;158:113–22 .

97. Tappeiner E, Pröll S, Hönig M, et al. Multi-organ segmentation of the head and neck area: an efficient hierarchical neural networks approach. Int J Comput Assist Radiol Surg 2019;14:745–54.

98. Pawlowski N, Ktena SI, Lee MC, et al. Dltk: state of the art reference implementations for deep learning on medical images. arXiv preprint arXiv:171106853 2017.

99. Milletari F, Navab N, Ahmadi SA; 2016 V-Net: fully convolutional neural networks for volumetric medical image segmentation:565–71 . arXiv: 1606.04797 .

100. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder

architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481–95 .

101. Du X, Lin TY, Jin P, et al. SpineNet: learning scale-permuted backbone for recognition and localization. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA: IEEE Xplore; 2020. p. 11589–98 .

102. Jamaludin A, Windsor R, Ather S, et al. Machine learning based berlin scoring of magnetic resonance images of the spine in patients with ankylosing spondylitis from the measure 1 study. Ann Rheum Dis 2020:40–1 .

103. Pakhomov D, Premachandran V, Allan M, et al. In: Deep residual learning for instrument segmentation in robotic surgery; 2019. p. 566–73 .

4.1.8 Tables

Table 4.1 Extracted Characteristics from the Studies Included in the Review

Characteristic	Description
(1) Author's data	• Country of the authors' affiliations: the affiliation country from the majority of the authors or the corresponding author • Authors' fields of expertise: health fields, data science fields, or both • Status with industry: if one of more authors was affiliated with an industrial partner
(2) Year	• The year it was published based on Medline, IEEE Xplore, Web of Science, or Embased databases
(3) Study type and design	• Study type: classification of primary studies into basic, clinical and epidemiological research; and subclassification into interventional or observational[238] • Study design: classification of studies into retrospective or prospective nature and further categorization into study design being cross-sectional, cohort, descriptive, case-control, or case-series type study designs using a described classification algorithms[239]
(4) Area of spine care focus	• Clinical application type: either diagnostic tools, clinical decision support tools, automated clinical/instrumentation assessment, clinical outcome prediction, combined or others • Studied anatomy: either cervical, thoracic, lumbar, sacral, combiner • Studied disease: if there was specific pathology that the study targeted • Studied surgery: if there were particular surgeries or procedures that the study targeted • Study imaging modality: either standard radiograph, dual-energy x-ray absorptiometry (**DEXA**), CT, MRI including which sequence(s), US, interventional imaging (fluoroscopy, O-arm, guided navigation), combined, or others

Table 4.1 Extracted characteristics from the studies included in the review (cont'd)

(5) Number of subjects and images included	• Overall size: number of subjects, images (before data augmentation) , or both included in the complete study • Overall diseased subjects: number of patients, images, or both included in the general study diagnosed with at least one spine condition • Categorized size: overall subjects, images, or both included in the general study binned into categories of <100, 100-1000, 1000-10000, 10000-100000, and >100000
(6) Size of the dataset used for DL development and validation	• Dataset size: number of subjects, images, or both included in the DL development phase (when applicable) • Dataset diseased subjects: number of patients, images, or both included in the DL development phase diagnosed with at least one spine condition (when applicable) • Categorized dataset size: dataset subjects: number of patients, images, or both included in the DL development phase (when applicable) binned into categories of <100, 100-1000, 1000-10000, 10000-100000, and >100000
(7) Origin of the dataset	• Either single-center, multicentric, public registry or dataset, synthetic images, or combined
(8) Dataset availability	• Availability and information of origin and access of datasets from publicly available and part of a clinical registry or database datasets
(9) DL method and architecture used	• DL methodology: either CNN, long short-term memory networks (LSTM), Recurrent neural network (RNN), Generative Adversarial Networks (GAN), Radial basis function networks (RBFN), Multilayer Perceptrons (MLP), Self-organizing maps (SOM), deep belief network (DBM), restricted Boltzmann machines (RBM), or other • DL task type: classification, regression, segmentation, object detection, image generation, or other • DL architecture: architecture and backbone family (ex: DenseNet, VGGs, etc) • Number of pipeline(s) and DL architecture(s) used or tested • Other ML techniques used or tested in the study
(10) DL training and validation	• Training: split of the dataset into training, validation, testing, and use of cross-validation • External validation: if external validation of the completed DL pipeline was performed
(12) Evaluation of performances	• Performance metrics used to validate their pipeline • If they used external data to validate their pipeline

Table 4.2 Registries and Clinical Databases Used in Studies on Deep Learning in Medical Imaging for Spine Care with Corresponding Summary When Available

Registry / Database name	Official Title	Responsible Party, Owner or Sponsor	Resources
ALSPAC[60]	Avon Longitudinal Study of Parents and Children	University of Bristol, Bristol, United Kingdom	http://www.bristol.ac.uk/alspac/researchers/access/
AO CSM-I	Surgical Treatment of Cervical Spondylotic Myelopathy	AO Innovation Translation Center (AO Clinical Investigation and Publishing Documentation), Dübendorf, Switzerland	https://clinicaltrials.gov/ct2/show/NCT00565734
AO CSM-NA	AOSPINE Assessment of Surgical Techniques for Treating Cervical Spondylotic Myelopathy	AOSpine North America Research Network, Pennsylvania, USA	https://clinicaltrials.gov/ct2/show/NCT00285337
CSORN[61]	Canadian Spine Outcomes and Research	Canadian Spine Society, Markdale, Canada	https://www.csorncss.ca/
Genodisc Project	Disc-degeneration linked pathologies: novel biomarkers and diagnostics for targeting treatment and repair	University of Oxford, Oxford, United Kingdom	https://cordis.europa.eu/project/id/201626/reporting
GESPIC[62]	German Spondyloarthritis Inception Cohort	Charite University, Berlin, Germany	https://clinicaltrials.gov/ct2/show/NCT01277419
Hangzhou lumbar spine study[63]	Hangzhou Lumbar Spine Study: a study focusing on back health in a Chinese population	First Affiliated Hospital of Zhejiang University, Hangzhou, China	*Not found*

Table 4.2 Registries and Clinical Databases Used in Studies on Deep Learning in Medical Imaging for Spine Care with Corresponding Summary When Available (cont'd)

H-PEACE[64]	Health and prevention enhancement	Seoul National University Hospital, Seoul, South Korea	http://en-healthcare.snuh.org/HPEACEstudy
LumbSeg[65,66]	Lumbar vertebra segmentation CT image datasets	*Not found*	*Not found*
Manitoba BMD Registry[67]	The Manitoba BMD Registry	Manitoba Bone Density Program Committee, Manitoba, Canada	https://www.gov.mb.ca/health/primarycare/providers/chronicdisease/bonedensity/research.html
MDCS[60]	Malmö Diet and Cancer Study	University of Lund, University Hospital, Malmo, Sweden	Not found
MIDICAM	Cervical spondylotic myelopathy: Application of spinal diffusion-based microstructural imaging (DMI) and phase-contrast MRI	University Medical Center Neurozentrum, Freiburg im Breisgau, Germany	https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00012962 German registry of clinical trials, number: DRKS00012962
MPP[68]	Malmö Preventive Medicine Project	University of Lund, University Hospital, Malmo, Sweden	*Not found*
MrOs[69]	Osteoporotic Fractures in Men Study (MrOS) from Database of Genotypes and Phenotypes (dbGaP)	University of California San Francisco, California, USA	https://sfcc.ucsf.edu/news/osteoporotic-fractures-men-mros-study-group-publish-two-articles
MySPINE	Functional prognosis simulation of patient-specific spinal treatment for clinical use	Fundacio Institut de Bioenginyeria de Catalunya, Barcelona, Spain	https://cordis.europa.eu/project/id/269909

Table 4.2 Registries and Clinical Databases Used in Studies on Deep Learning in Medical Imaging for Spine Care with Corresponding Summary When Available (cont'd)

NFBC1966	Northern Finland Birth Cohort 1966	University of Oulu, Oulu, Finland	http://www. oulu.fi/nfbc/
NHANES 2011-2012	National Health and Nutrition Examination Survey 2011-2012 Database	National Center for Health Statistics, Maryland, USA	https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/overview.aspx?BeginYear=2011
NHANES II	Second National Health and Nutrition Examination Survey Database	National Center for Health Statistics, Maryland, USA	https://wwwn.cdc.gov/nchs/nhanes/nhanes2/default.aspx
OSTPRE[70]	Kuopio Osteoporosis Risk Factor and Prevention Study	Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland	https://sites.uef.fi/kmru/ostpre/
OSTPRE-FPS[71]	OSTPRE Fracture Prevention Study	Kuopio University Hospital, Kuopio, Findland	https://clinicaltrials.gov/ct2/show/NCT00592917
PROOF[72]	Multicountry Registry of Clinical Characteristics	AbbVie, Cham, Switzerland	http://www.chictr.org.cn/showprojen.aspx?proj=10022
ROAD	Research on Osteoarthritis/Osteoporosis Against Disability	University of Tokyo, Tokyo, Japan	*Not found*
SCI database	*Not found*	Orange image diagnostic center, USA	*Not found*
SDSG	Spinal Deformity Study Group database	*Not found*	*Not found*

Table 4.2 Registries and Clinical Databases Used in Studies on Deep Learning in Medical Imaging for Spine Care with Corresponding Summary When Available (cont'd)

TRACK-SCI[69,73]	Transforming Research and Clinical Knowledge in Spinal Cord Injury	University of California, San Francisco, California, USA	https://clinicaltrials.gov/ct2/show/NCT04565366
TwinsUK registry	TwinsUK registry	King's College London, United Kingdom	www.twinsuk.ac.uk
Wakayama Spine study[74]	Wakayama Spine study	Wakayama MedicalUniversity School of Medicine, Wakayama, Japan	*Not found*
Whiplash[75]	Neuromuscular Mechanisms Underlying Poor Recovery From Whiplash Injuries	Northwestern University, Ilinois, USA	ClinicalTrials.gov Identifier: NCT02157038

Table 4.3 Publicly Available Datasets Used in the Studies Focusing on Deep Learning in the Field of Medical Imaging for Spine Care Investigated in This Review with the Corresponding Summary when Available

Public Dataset Name	Official Title	Responsible Party, Owner or Sponsor	Resources
[76]**CSI 2014 workshop dataset**	Localization and identification challenge of the CSI 2014 Workshop.	University of Washington in St Louis, Missouri, USA	*The link provided in the publication is no longer working (http://research.microsoft.com/spine/)*
[77]**DeepLesion**	*Not found*	*Not found*	https://www.kaggle.com/datasets/kmader/nih-deeplesion-subset
[78]**ILSVRC**	ImageNet Video dataset (ILSVRC)	Stanford Vision Lab, Stanford University, Princeton University	https://image-net.org/challenges/LSVRC/index.php
IoMT Spine Dataset	Internet of Medical Things (IoMT) platform Spine Dataset	Not found	http://spineweb.digitalimaginggroup.ca/spineweb.
MICCAI[79,80]	Testing set A of the MICCAI Challenge on Vertebral Fracture Analysis	Medical Image Computing and Computer Assisted Intervention Workshop & Challenge (MICCAI 2016)	Upon request at: http://spineweb.digitalimaginggroup.ca/dataset.php
MS Annotated Spine CT Database	MS Annotated Spine CT Database	*Not found*	No longer working (http://research.microsoft.com/en-us/projects/spine/)
PAM50[78,81]	Unbiased multimodal MRI template of the spinal cord and the brainstem	Polytechnique Montreal, Canada and Aix-Marseille Université, France	https://github.com/sct-data/PAM50/releases/download/r20191029/20191029_pam50.zip

Table 4.3 Publicly Available Datasets Used in the Studies Focusing on Deep Learning in the Field of Medical Imaging for Spine Care Investigated in This Review with the Corresponding Summary when Available (cont'd)

SpineWeb Dataset 16	Dataset 16: 609 spinal anterior-posterior x-ray images	Stanford Vision Lab, Stanford University, Princeton University London Health Sciences Center, Ontario, Canada	https://image-net.org/challenges/LSVRC/index.phpUp on request (http://spineweb.digitalimaginggroup.ca/dataset.php)
SpiSeg[76]	Spine and vertebrae segmentation datasets	University of California, Irvine, School of Medicine, California, USA	Upon request at: http://spineweb.digitalimaginggroup.ca/dataset.php
TCIA[81]	Cancer Imaging Archive	Washington University School of Medicine, Missouri, USA and Frederick National Laboratory for Cancer Research, Frederick, MD, USA	No longer working (http://research.microsoft.com/en-us/projects/spine/)https://cloud.google.com/healthcare-api/docs/resources/public-datasets/tcia
UCI Vertebral Column Dataset	UCI Repository of Machine Learning Databases	UC Irvine Machine Learning Repository	http://archive.ics.uci.edu/ml/machine-learning-databases/00212/
VerSe'20[80]	Large Scale Vertebrae Segmentation Challenge	Medical Image Computing and Computer Assisted Intervention (MICCAI 2020)	https://zenodo.org/record/3759104#.Y3UE0nbMInI
xVertSeg.v1[77]	Segmentation and Classification of Fractured Vertebrae	University of Ljubljana, Faculty of Electrical Engineering, Slovenia	http://lit.fe.uni-lj.si/xVertSeg/

Table 4.4 Overview of Common CNN Architectures used in the Studies Focusing on Deep Learning in the Field of Medical Imaging for Spine Care Investigated in this Review with the Variants Reported for the Spine and Corresponding Short Summary

Backbone	Year	Description	Variants Reported for the Use in Spine	Ressources
AlexNet[40]	2012	An architecture developed for image classification that launched DL development. The network contains a set of convolutional and max-pooling layers ended by 3 fully connected layers.	AlexNet. *Note: also used in in some of the R-CNN architectures*	https://github.com/deep-diver/AlexNet
VGG[82]	2014	Backbone widely used for computer vision and computer sciences tasks. VGG is composed of convolutional, max-pooling, and fully connected layers. It uses smaller kernels to create deeper networks.	VGG-11, VGG-16, VGG-19, VGG-M, VGG-Net, VGG-Net16, FCN-VGG-16	https://github.com/pytorch/vision/blob/master/torchvision/models/vgg.py
U-Net[83]	2015	Fully convolutional network Popular architecture for biomedical image semantic segmentation. It comprises a contracting path ("traditional" CNN) and an expansive path.	3D U-Net, BiLuNet, Co-U-Net, DC-U-Net, Deeplab V3+, Deep-U-Net, fuse-U-Net, MDR2-U-Net, MDR2-U-Net, Residual U-Netstacked hourglass network (SHN)	https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/u-net-release-2015-10-02.tar.gz
ResNet[84]	2015	Backbone widely used for object detection and image segmentation that introduced skip connected. ResNet is made of residual NN consisting of skip-connections or recurrent units between blocks of pooling and convolutional layers. Many versions with different deepness.	FR-ResNet, Multi ResNet, ResNet-12, ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-ST-50, ResNet-XT-50. *Note: also used in some of the Faster R-CNN architectures.*	https://github.com/pytorch/vision/blob/master/torchvision/models/resnet.py

Table 4.4 Overview of Common CNN Architectures used in the Studies Focusing on Deep Learning in the Field of Medical Imaging for Spine Care Investigated in this Review with the Variants Reported for the Spine and Corresponding Short Summary (cont'd)

Inception-v2-v3[85]	2015	Network with a low number of parameters to be able to be used on lower-performance machines. It was modified from Inception V1 (blocks of inception containing sets of convolutional layers) by replacing nn convolutional kernels into 3x3 or 1x1 using a concatenation method.	Inception-V3	https://github.com/weiaicunzai/pytorchcifar100/ blob/master/models/inceptionv3.py
Inception-ResNet-V2[86]	2015	ResNet and inception architecture combine using skip-connections between blocks of layers, called residual connections.	Inception-ResNet-V2, Inception-ResNet-V3. *Note: also used in some Faster R-CNN architectures.*	https://github.com/zhulf0804/Inceptionv4_and_Inception-ResNetv2.PyTorch
YOLO[87,88]	2015	YOLO, short for "You Only Look Once" is from a series of object detection models capable of detecting multiple objects simultaneously. It is based on CNN and is used to predict classes as well as object localization.	YOLOV2, YOLOV3, HoloYOLO	https://github.com/ultralytics/yolov5
DenseNet[84,89]	2018	CNN built on the idea of ResNet but with a lower number of connections of L/(L+1)/2, with L being the number of layers. The feature maps of all previous layers are used as input in the next layer, making DenseNet well-suited for smaller datasets.	DenseNet-14, DenseNet-22, DenseNet-26, DenseNet-48, DenseNet-121, DenseNet-169, DenseNet-201, DMML-Net, FC-DenseNet, BMDC-Net	https://github.com/liuzhuang13/DenseNet

Table 4.5 List of the Main Performance Metrics with Corresponding Equations Reported in the Studies Focusing on Deep Learning in the Field of Medical Imaging for Spine Care Investigated in this Review

Notes: true positive (TP), true negative (TN), false positive (FP), false negative (FN), and A and B represent the ground truth and predicted segmentation

Metric group	Abbr.	Name	Equation	Description	% Use
Probabilistic	Se	Sensitivity[90,91]	$Se = \frac{TP}{TP + FN}$	True positive detection capabilities. Math equivalent: Recall, TPR	49%
	Sp	Specificity[90,91]	$Sp = \frac{TN}{TN + FP}$	Capabilities for correctly identifying true negative classes. Math equivalent: TNR	41%
	Acc	Accuracy[90]	$Acc = \frac{TP + TN}{TP + TN + FP + FN}$	Total number of correct predictions, compared to the total number of predictions	48%
	ROC	Receiver Operator Characteristics[90]	*Line plot showing performance with different discrimination thresholds*	Line plot of the diagnostic ability of a classifier through TPR against FPR	10%
	AUC	Area under Receiver Operator Characteristics[92]	$AUC = \frac{1}{2}\left(\frac{FP}{FP + TN} + \frac{FN}{FN + TP}\right)$	Area under the simple trapezoid	36%

Table 4.5 List of the Main Performance Metrics with Corresponding Equations Reported in the Studies Focusing on Deep Learning in the Field of Medical Imaging for Spine Care Investigated in this Review (cont'd)

	PPV	Positive Predictive Value[93]	$PPV = \frac{TP}{TP + FP}$	Amount of true diagnosis with respect to true positive test. Math equivalent: precision	11%
	KAP	Cohen Kappa Coefficient[90]	$KAP = \frac{(TN + FN) - fc}{(TP + TN + FN + FP) - f}$	Measure of agreement between annotated and predicted classifications or predicted and ground truth segmentation	
F-measure based metrics	DSC	Dice Similarity Coefficient or Sorensen-Dice Index [90,91,93]	$DSC = F1 = \frac{2TP}{2TP + FP + FN}$	Amount of pixel overlap over the total number of pixels in predicted and ground truth segmentation	24%
	F1	F1 score			
	JACC	Jaccard Index[90,91,93]	$JAC = IoU = \frac{TP}{TP + FP + FN}$	Amount of pixel overlap divided by their union	8%
	IoU	Intersection-over-Union			
Spatial overlap and distance	PREC	Precision[93]	$PREC = \frac{\lvert A \cap B\rvert}{\lvert B\rvert}$	Amount of predicted pixel overlap over with respect to ground truth	14%

Table 4.5 List of the Main Performance Metrics with Corresponding Equations Reported in the Studies Focusing on Deep Learning in the Field of Medical Imaging for Spine Care Investigated in this Review (cont'd)

AHD	Average Hausdorff distance/Max. Symmetric Surface Distance[90,94]	$d(A,B) = \frac{1}{N}\sum_{a \in A} \min_{b \in B} \|\|a - b\|\|$ $AHD\ (A,B) = \max\ (d(A,B), d(B,A))$ in which A and B represent the ground truth and predicted segmentation, respectively, and ||a-b|| represents a distance function like Euclidean distance	Spatial distance average over predicted and ground truth points	4%

Table 4.6 A Short List of Available Code or DL Platforms used in the Methodology or Provided as a Result in the Published Studies Investigated in This Review Focusing on DL in the Field of Medical Imaging for Applications Intended for Spine Clinical Care

Name	Summary	Implementation
SpineCube[95]	Intelligent agent for diagnosing scoliosis and evaluating the severity of scoliosis.	https://github.com/js3611/Deep-MRI-Reconstruction
[96,97]	Direct automated quantitative measurement of the spine by cascade amplifier regression network with manifold regularization	https://github.com/pangshumao/CARN
[98]	Testing DL model for automated detection of vertebral fractures of the lumbar spine	https://links.lww.com/CORR/A505
VFADL[99]	Source code for automated identification of vertebral fractures at VF assessment performed with dual-energy x-ray absorptiometry	https://github.com/DougUC/VFADL-PUBLIC/blob/master/VFADL.ipynb
BMDC-Net[100]	Method for the qualitative detection of BMD (normal bone mass, low bone mass, and osteoporosis) via diagnostic CT slices	https://github.com/tangchao1010/classification-of-BMD
DMML-Net[101]	Deep multiscale multitask learning network to directly localize all lumbar organs with bounding boxes and grade all lumbar organs with crucial differential diagnoses (normal and abnormal).	https://github.com/zhyhan/DMML-Net/tree/master
[102]	Fully automated algorithm for the detection of bone marrow edema lesions in patients with axial spondyloarthritis	https://github.com/krzysztofrzecki/bone-marrow-oedema-detection
[103]	DL model for detection of cervical spinal cord compression in MRI scans.	https://github.com/zamirmerali/dcm-mri

Table 4.6 A Short List of Available Code or DL Platforms used in the Methodology or Provided as a Result in the Published Studies Investigated in This Review Focusing on DL in the Field of Medical Imaging for Applications Intended for Spine Clinical Care (cont'd)

MBNET[104]	Multi-Task Deep Neural Network with supervised learning applied for two tasks, semantic segmentation and parameter inspection for the diagnosis of lumbar vertebrae	https://github.com/LuanTran07/BiLUnet-Lumbar-Spine
LEN-LCN[105]	Implementation code of automated Landmark Estimation and Correction Network to estimate landmarks on lateral X-rays.	https://github.com/LuanTran07/BiLUnet-Lumbar-Spine
DeepSeg[106]	Fully automatic framework for segmentation of the spinal cord and intramedullary multiple sclerosis lesions from conventional MRI data using CNN	https://github.com/spinalcordtoolbox/spinalcordtoolbox/tree/master/spinalcordtoolbox/deepseg
[107,108]	Scripts for image segmentation using CNN to segment bones in US images automatically	https://github.com/SlicerIGT/aigt
VerteSeg[109]	Code for automatic segmentation of vertebrae from sagittal IDEAL (Iterative Decomposition of water and fat with Echo Asymmetric and Least-squares estimation) spine MR images	https://github.com/zhoji/verteseg
[110]	Deep CNN model to classify osteopenia and osteoporosis using lumbar spine X-ray images.	https://github.com/zhang-de-lab/zhang-lab/tree/master/osteoporosis
[111]	Model developed using NiftyNet for neck muscle segmentation.	https://github.com/kennethaweberii/Neck_Muscle_Segmentation
[112,113]	Automatic landmark estimation and spinal curvature estimation for adolescent idiopathic scoliosis	https://github.com/zc402/Scoliosis

Table 4.6 A Short List of Available Code or DL Platforms used in the Methodology or Provided as a Result in the Published Studies Investigated in This Review Focusing on DL in the Field of Medical Imaging for Applications Intended for Spine Clinical Care (cont'd)

SpineAI[114]	Implementation code to automatically detect and classify lumbar spinal stenosis on MRI images.	https://github.com/NUHS-NUS-SpineAI/SpineAI-Detect-Classify-LumbarMRI-Stenosis
[115]	Model for identifying fresh VCF from digital radiography.	https://github.com/TXVision/DR_Fracture_Classification
[116]	Code for lumbar spine hanging protocol label lumbar spine views/positions, detect hardware and rotate the lateral views to straighten the image.	https://github.com/GeneKitamura/L_spine_hanging_protocol
Anduin[80,117]	Freely available research tool to segment vertebrae in a CT scan and to assess various bone measures in clinical CT.	anduin.bonescreen.de
Spinal Cord Toolbox[118,119]	Open-source set of command-line tools dedicated to the processing and analysis of spinal cord MRI data.	https://github.com/spinalcordtoolbox/spinalcordtoolbox
Nora Imaging[120]	Web-based framework using CNN for medical image analysis	http://www.nora-imaging.org
NiftyNet[121]	Open-source CNN platform for medical image analysis.	http://niftynet.io
Modified NiftyNet[122]	Monai and NiftyNet version of the code to generate the multi-organ segmentation of the head and neck area	https://github.com/elitap/NiftyNet

Table 4.6 A Short List of Available Code or DL Platforms used in the Methodology or Provided as a Result in the Published Studies Investigated in This Review Focusing on DL in the Field of Medical Imaging for Applications Intended for Spine Clinical Care (cont'd)

DLTK[123]	NN toolkit written in Python to enable fast prototyping with a low entry threshold for medical imaging	https://github.com/DLTK/DLTK
V-Net[124]	3D image segmentation based on a volumetric, fully CNN.	https://github.com/faustomilletari/VNet
SegNet[125]	Deep fully CNN architecture for semantic pixel-wise segmentation	http://mi.eng.cam.ac.uk/projects/segnet/
SpineNet[126]	CNN backbone with scale-permuted intermediate features and cross-scale connections learned on an object detection task by Neural Architecture Search.	https://github.com/lucifer443/SpineNet-Pytorch/tree/a7059eff295dcee16d719b381f80af8eb3fe42f6
SpineTK[127]	Code to train a network for doing MR, CT, and X-ray image annotation, including landmark annotation of six keypoints on individual vertebral bodies for vertebral height measurement	https://github.com/abhisuri97/SpineTK
[128]	Source code for deep residual learning for multi-class robotic tool segmentation	https://github.com/warmspringwinds/tf-image-segmentation

4.1.9 Figures

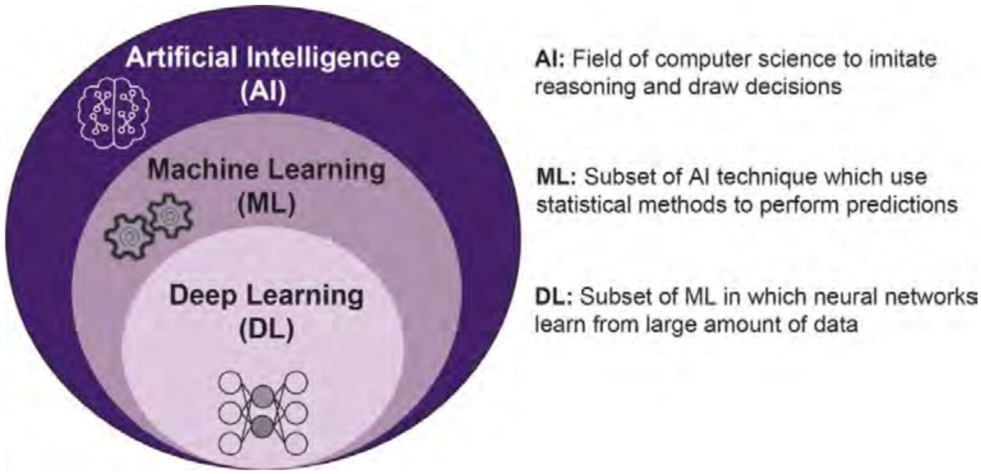


Figure 4.1 Overview of Artificial Intelligence, Machine Learning and Deep Learning

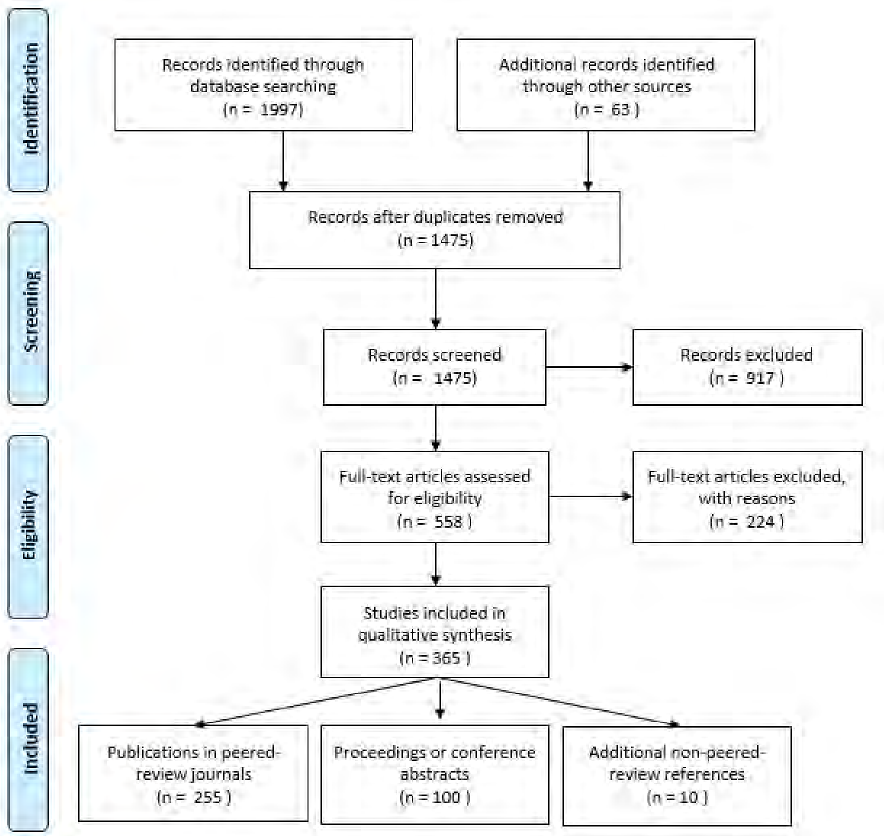


Figure 4.2 The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Statement Flowchart of the Process Performed for the Review of the Current State-of-the-Art Progress and Utilization of DL in the Field of Medical Imaging for Spine Care

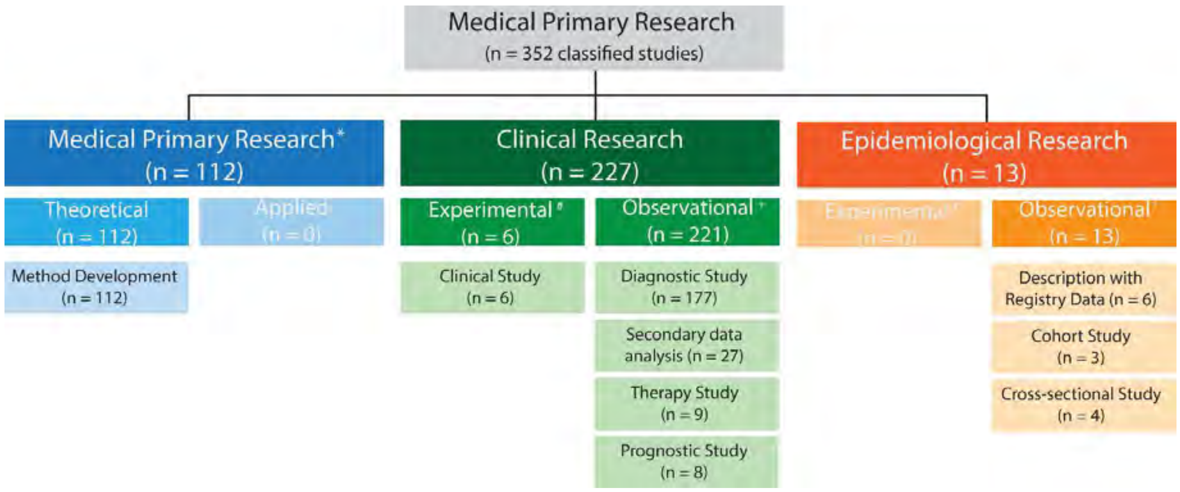


Figure 4.3 Classification of the Types of Studies From 352 Published Studies Focusing on DL in the Field of Medical Imaging for Spine Care According to the Classification Schemes for Studies in Medical Research by Rohrig et al.[1]

* , sometimes known as experimental research; #, analogous term to interventional; +, analogous term to noninterventional or nonexperimental

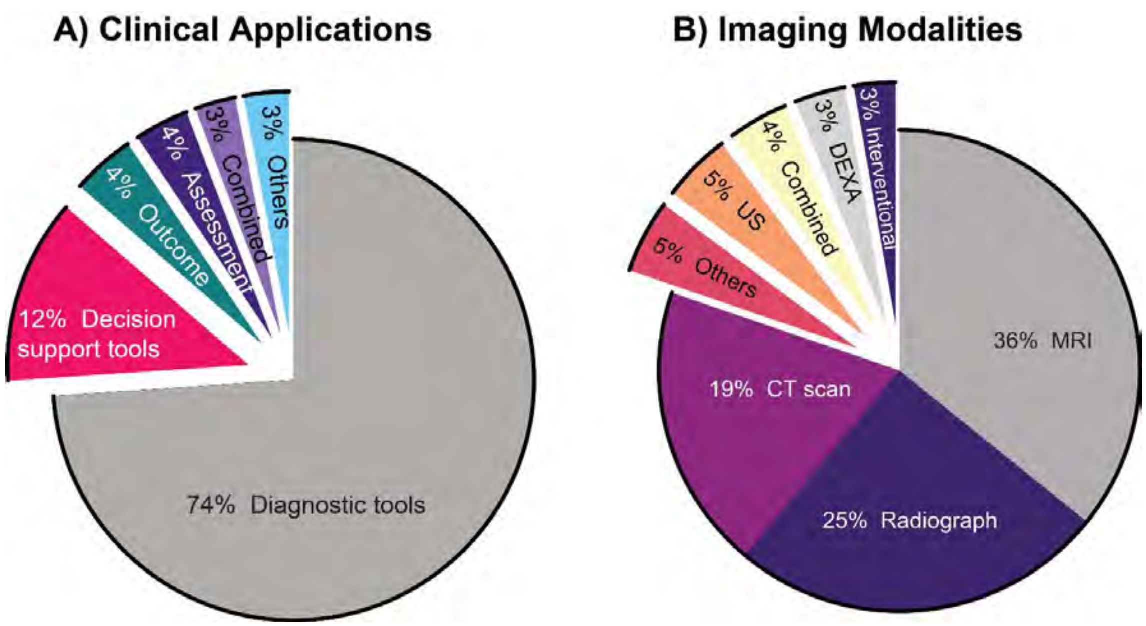


Figure 4.4 Clinical applications (A) and Imaging Modalities (B) related to the Published Studies Investigated in this Review Focusing on DL in the Field of Medical Imaging for Applications Intended for Spine Clinical Care.

The clinical application types included diagnostic tools, clinical decision support tools, automated clinical/instrumentation assessment, clinical outcome prediction, combined or others. The imaging modalities used included magnetic resonance imaging (MRI), radiograph, computed tomography (CT scan), ultrasound (US), dual-energy x-ray absorptiometry (DEXA), intervention imaging (fluoroscopy, O-arm, guided navigation), and others.

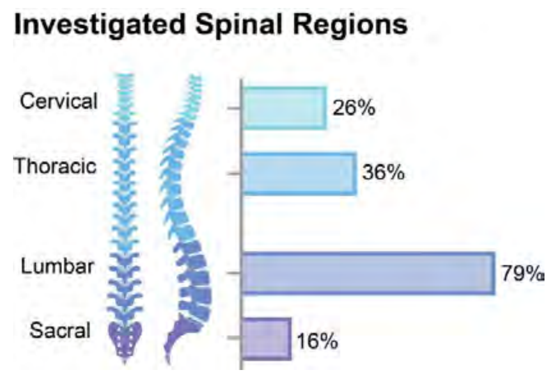


Figure 4.5 Distribution of the Frequency of Investigation of Spinal Regions Targeted by the Published Studies Investigated in this Review Focusing on DL in the Field of Medical Imaging for Applications Intended for Spine Clinical Care

Note: sum is not equal to 100% as 49% of reviewed studies investigated multiple spinal regions.

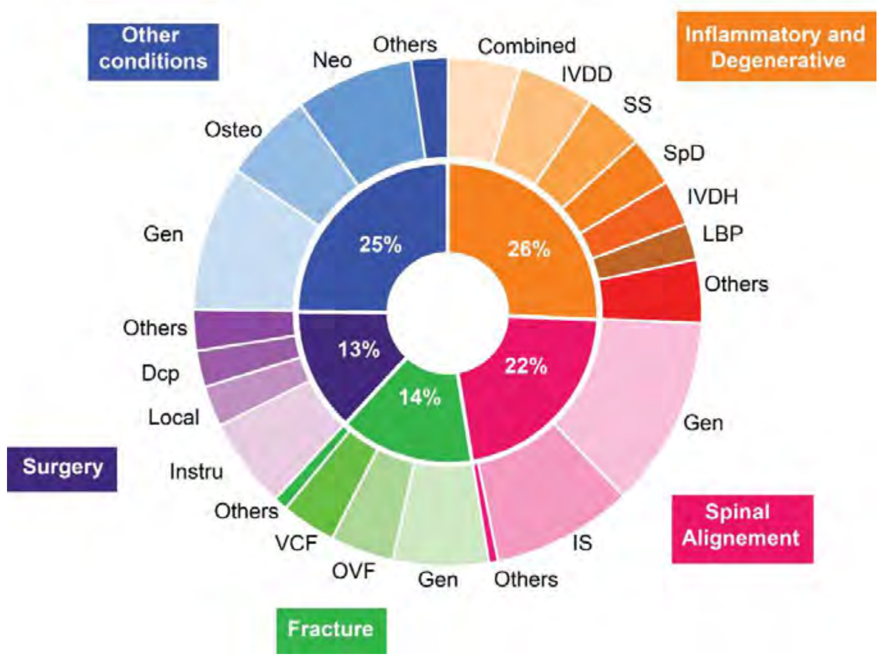


Figure 4.6 Spinal Diseases and Conditions Examined in the Published Studies Investigated in This Review Focusing on DL in the Field of Medical Imaging for Applications Intended for Spine Clinical Care

Inflammatory and degenerative conditions included intervertebral disc degeneration (IVDD), spinal stenosis (SS), spondylitis and spondyloarthritis (SpD), intervertebral disc herniation (IVDH), and lower back pain complex (LBP). Spinal alignment conditions included measurement applicable to general alignment problems (Gen) and idiopathic scoliosis (IS). Fracture assessment included general vertebral bone assessment (Gen), osteoporotic vertebral fracture (OVF), and vertebral compression fracture. Investigation targeting surgical procedures included spinal instrumentation (Instru), local analgesia or anesthesia procedures (Local), and spinal decompression surgery (DCP). Among other conditions, general spinal assessment (Gen), osteoporosis (Osteo), and neoplastic diseases (Neo) have been studied.

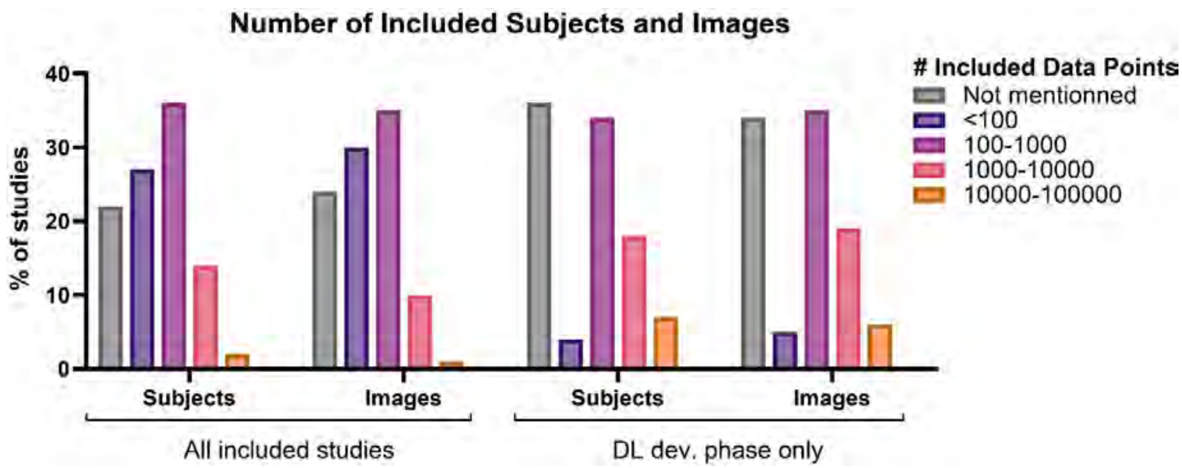


Figure 4.7 Number of Subjects and Images Included in the Published Studies Investigated in this Review Focusing on DL in the Field of Medical Imaging for Applications Intended for Spine Clinical Care

The number of data points is categorized and presented for the overall subjects and images comprised in all included studies and for the subjects and images comprised only in the DL development phase of the studies, when applicable.

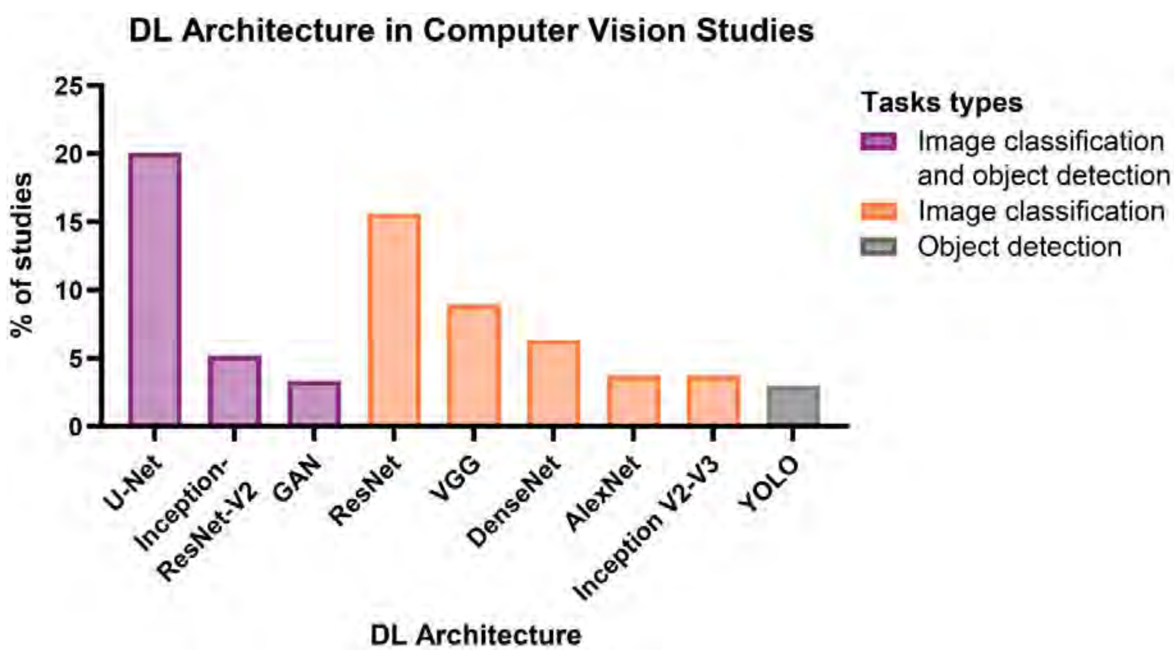


Figure 4.8 Distribution of the Most Commonly Investigated DL Network Architectures in the Studies Included in the Review Investigating DL Network Structure for Computer Vision Tasks in Medical Imaging for Spine Care (n = 307 Studies)

4.2 ARTICLE 2: Neural Network-Based Multi-Task Learning to Assist Planning of Posterior Spinal Fusion Surgery for Adolescent Idiopathic Scoliosis

To address Sub-objective 1.2 (SO1.2), a deep learning model using a multitask neural network architecture was developed to support surgical planning for PSF surgery in AIS patients. The model was trained to predict key surgical decisions, including UIV, LIV, screw density, and rod curvature, based on standard preoperative clinical and radiographic parameters. Its performance was compared to single-task models and evaluated using both internal and external datasets. This study contributes to the field by (1) introducing a novel AI-based planning tool for PSF surgery, (2) demonstrating improved performance of multitask over single-task neural networks for surgical parameter prediction, and (3) showing that such models can generalize across institutions and surgeons.

This work led to the publication of the article titled "*Neural Network-Based Multi-Task Learning to Assist Planning of Posterior Spinal Fusion Surgery for Adolescent Idiopathic Scoliosis*" in the *Spine Deformity Journal* in June 2025. The first author's contribution to the conception, development, experimentation, analysis, and writing of the manuscript is evaluated at over 80%.

Article 2:

Constant, C., Larson, A.N., Polly, D.W. et al. **Neural Network-Based Multi-Task Learning to Assist Planning of Posterior Spinal Fusion Surgery for Adolescent Idiopathic Scoliosis.** Spine Deform (2025). https://doi.org/10.1007/s43390-025-01125-9

Highlights

First study to develop and validate a multitask neural network (NNML) for AIS fusion planning.

The final model was able to mimic expert surgeon decisions and could help reducing variability in planning.

NNML outperformed single-task models with high accuracy for vertebrae and rod predictions.

External validation confirmed robustness across different patients and surgeons.

Article Title: Neural network-based multi-task learning to assist planning of posterior spinal fusion surgery for adolescent idiopathic scoliosis

Author names:

Caroline Constant[1,2,3,4], DMV, MSc, MENG, DACVS-LA, DECVS (caroline.constant@uzh.ch), A. Noelle Larson[1] M.D (Larson.Noelle@mayo.edu), David W. Polly, Jr. [5], MD (pollydw@umn.edu), Carl-Eric Aubin[2,3], Ph.D., ScD(h.c.), P.Eng., (Carl-Eric.Aubin@polymtl.ca), And Minimize Implants Maximize Outcomes Study Group

Institutional affiliation:

[1] Department of Orthopedic Surgery, Mayo Clinic, 200 1st Street Southwest, Rochester, Minnesota, 55905, USA

[2] Polytechnique Montréal, 2500 Chemin de Polytechnique, Montréal, H3T 1J4, Canada

[3] Centre Hospitalier Universitaire (CHU) Sainte-Justine, 3175 ch. Côte Sainte-Catherine, Montréal H3T 1C5, Canada

[4] AO Research Institute Davos, Clavadelerstrasse 8, CH 7270, Davos, Switzerland

[5] Department of Orthopedic Surgery, University of Minnesota, Minneapolis, MN

Institution at which the work was performed: 1, 2, 3

Corresponding author: C. Constant; Department of Mechanical Engineering, Polytechnique Montreal, P.O. Box 6079, Downtown Station, Montreal, QC H3C 3A7, Canada; +41 79 910 69 76; caroline.constant@polymtl.ca

Funding sources: This work was supported by the TransMedTech Institute and its main financial partner, the Canada First Research Excellence Fund, the NSERC/Medtronic Industrial Research Chair in Spine Biomechanics, Quattrone-Foderaro Grant in AI Innovation in Orthopedic Surgery 2021, Mayo Foundation, the Mayo Clinic, and the AO Foundation. The MIMO Study Group was financially supported by the Orthopedic Research and Education Foundation and Scoliosis Research Society.

Competing Interests: The authors declare no competing interests related to this manuscript. They have no financial, personal, or professional relationships within the past three years, or beyond this period, that could be perceived as influencing the research conducted or the preparation of this manuscript.

Ethical Approval : This study received Institutional Review Board (IRB) approval through the respective committees overseeing the MIMO Clinical Trial (NCT01792609) and research projects at the authors' institutions, and complies with ethical standards in accordance with the Declaration of Helsinki.

Consent: all patients provided informed consent for both clinical trial participation and subsequent image analysis. The authors affirm that patients provided informed consent for publication of the radiographs in relevant figures.

Authorship: All authors confirm their substantial contributions to the conception, design, data acquisition, analysis, or software creation for this work; they have critically revised the manuscript, approved the final version for publication, and accept accountability for the work's integrity. All authors consented to submission and obtained required institutional permissions.

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Caroline Constant and the group MIMO. The first draft of the manuscript was written by Caroline Constant and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

4.2.1 Abstract

Purpose: Posterior spinal instrumentation and fusion (PSF) is the gold standard for severe adolescent idiopathic scoliosis (AIS), yet instrumentation strategies vary widely, often leading to suboptimal results. Deep learning's potential in AIS planning is underexplored.

Methods: This study trained and validated an artificial neural network multitask learning model (NNML) using preoperative clinical and radiographic data from 189 AIS patients with Lenke 1A and 2A curves enrolled in the MIMO Clinical Trial (NCT01792609). The model mimics experienced spine surgeons' decision-making for selecting the upper and lower instrumented vertebrae (UIV, LIV), determining rod curvature, and predicting screw density based on the study's randomized allocation. Models were trained with data from 179 patients, utilizing 10-fold cross-validation, and externally validated on 10 patients from a separate hospital and surgeons outside the training set. For UIV and LIV selection, accuracy within the top two predictions was used as a classification performance metric, ensuring that other clinically relevant alternatives were considered.

Results: The ANNML, which comprised 83 inputs and multiple hidden layers, led to significant gains over ST-NN and proved more robust during the internal validation (loss 6.2 vs. 9.3; $p \leq 0.01$). It showed 82-95% and 80-100% accuracy for UIV and LIV predictions and 70-90% accuracy for predicting the rod curvatures ±5°. The RMSE for the screw density and rod curvature predictions was 0.2-0.3 and 3.7-5.6°, respectively.

Conclusion: An ANNML can better use the features of relevant AIS patients for mixed task prediction pertinent to PSF surgery planning than ST-NN. In addition, ANNML was capable of mimicking experienced spine surgeons' decision-making process when designing the instrumentation.

Keywords: Posterior spinal fusion (PSF), adolescent idiopathic scoliosis (AIS), deep learning, artificial intelligence, preoperative planning, neural network multitask learning model

4.2.2 Introduction

Idiopathic scoliosis is a pathologic curvature of the spine of unknown etiology associated with a spinal deviation in the coronal plane with a Cobb angle greater than 10° in magnitude [1]. The most common form in children is adolescent idiopathic scoliosis (AIS) and has a prevalence of up to 5% between 10 and 18 years of age [2-4], The current gold standard for surgical management of AIS cases is posterior spinal instrumentation and fusion (PSF) with pedicle screws. Although modern pedicle screws facilitate spinal deformity correction, there is not a standardized approach regarding surgical correction maneuvers, number of screws, and array of screws used for AIS cases. Thus, there is heterogeneity of instrumentation strategies even among experienced surgeons, and a lack of information regarding which approach produces optimal long-term results [5, 6]. While guidelines such as the Lenke classification guide the surgical decision-making process of preoperative PSF surgery planning for AIS patients [7, 8], suboptimal results have been reported following surgery performed, and surgeons frequently deviate from Lenke classification recommendations [9]. Problems related to surgical instrumentation affect up to 4.5% of AIS cases [10]. These findings suggest a lack of standardization and the potential for improvement in preoperative PSF deformity correction planning. Such efforts could decrease the risk of unfavorable postoperative outcomes and achieve improved alignment in long-term outcomes.

In the past decades, different software allowing surgical simulations have increased in popularity for preoperative planning in spine surgery. Machine learning (ML), a field of artificial intelligence (AI), enables predictive modeling using large datasets, allowing for data-driven decision-making in personalized surgical planning [11]. While ML-based models have been explored for outcome prediction in AIS, no prior study has developed a neural network specifically designed to predict instrumentation configurations for AIS patients. The successful development of AI-based planning algorithms has the potential to significantly improve patient-specific AIS instrumentation strategies. These models can provide decision support for less experienced surgeons, assisting them in selecting optimal surgical parameters based on real-world clinical data. By integrating large-scale clinical datasets into preoperative planning, AI-driven models can also reduce variability in surgical strategies, addressing inconsistencies in instrumentation selection among surgeons. Additionally, enhancing reproducibility and standardization in AIS surgery could lead to more consistent and predictable patient outcomes.

The aim of this study was to train and validate an artificial neural network-based multitask learning model (NNML) using preoperative clinical and radiographic data to mimic the decision-making process of experienced spine surgeons in selecting the upper and lower instrumented vertebrae (UIV, LIV), the number of pedicle screws over the instrumented segment (screw density), and rod contour used for the surgical correction of AIS deformity. While UIV, LIV, and rod contour were modeled based on experienced decisions, screw density was predicted according to the randomized allocation in the MIMO study dataset. By combining expert knowledge with AI-driven decision-making, this NNML model could represent an important step toward AI-assisted, patient-specific preoperative planning in AIS surgery, enhancing standardization and ultimately improving surgical precision and long-term patient outcomes.

4.2.3 Materials and Methods

4.2.3.1 Workflow Overview

This study developed an NNML network to predict surgical instrumentation parameters (UIV, LIV, screw density, and rod curvatures) using vertebrae corners' coordinates from coronal plane radiographs, preoperative patient parameters, and postoperative spine alignment. Unlike traditional models that perform one task at a time, this NNML model predicts multiple surgical parameters simultaneously. The methodology framework had three stages: network development, performance comparison with single-task neural networks (ST-NN), and external testing on an independent dataset (Figure 4.9 and Figure 4.10). The study included 189 patients, with 179 for development and 10 for external testing (Figure 4.9A).

4.2.3.2 AIS Patient Sample and Data Collection

After institutional review board approval, a retrospective review was conducted on a multicenter database of AIS patients from the MIMO Clinical Trial (NCT01792609) [12]. Patients were treated by experienced spine surgeons from 14 high-volume hospitals. Inclusion criteria were patients aged 10–18 years with AIS, Lenke 1A or 2A curve pattern, and a Cobb angle of 45°-65°, recommended for PSF surgery. Patients with mechanical failure of implants were excluded from the study. All patients provided consent for the clinical trial and subsequent image analysis.

Demographic, radiographic, and surgical characteristics were collected from the trial's database (Table 4.7). Cobb angle measurements were obtained for proximal thoracic (PT), main thoracic (MT), and thoracolumbar/lumbar (TL/L) regions. Rod curvatures were reconstructed from postoperative radiographs [13], and rod contouring was calculated as the difference between convex and concave rods (differential bending). Postoperative spinal alignment was measured on radiographs taken 3 months post-PSF surgery or up to 1 year if unavailable. Features used for the NNML model are shown in Figure 4.10. Remaining variables described the study population. All numerical features were standardized using the Min-Max normalization formula [14].

For model development, UIV, LIV, and rod curvature were determined based on the decision-making of experienced spine surgeons. Screw density was based on the randomized allocation in the MIMO study, where patients were assigned to either a high-density (>1.8 screws per level fused) or low-density (<1.4 screws per level fused) group. However, within each assigned category, the total number of implants and their distribution across instrumented levels were chosen at the surgeon's discretion.

4.2.3.3 ***Stage 1 - NNML Architecture and Development***

The NNML model was designed with three main components: shared layers, task-specific output layers (for regression and classification), and an optimization function. Shared layers extract common patterns from the input data, while task-specific layers focus on individual predictions (e.g., UIV, LIV, rod curvature, and screw density).

The model used Rectified Linear Unit (ReLU) activation functions for numerical predictions (screw density, rod curvatures), and Softmax functions for categorical predictions for categorical predictions (UIV, LIV). The Adam optimization algorithm was applied for learning. For regression tasks, the Mean Squared Error (MSE) loss function was used, while classification predictions were optimized using the multi-class cross-entropy (CE) loss function. A combined loss function was used to balance the learning of multiple tasks. To prevent overfitting, the model was trained using 10-fold cross-validation, with random dropout and batch normalization applied to improve generalization [15]. A range of hyperparameters was tested using a grid search approach, and the final model was selected based on validation performance. For comparison, five single-task neural networks (ST-NN), each predicting only one surgical parameter, were developed using the same architecture for direct performance comparison.

4.2.3.4 Stage 2 - Evaluation of the Models' Performances and Comparison Metrics

To assess model robustness, three different cross-validation strategies (10-fold, 5-fold, and 3-fold) were used to test performance across multiple data splits. For each partition ran, the validation subsets from the model development dataset were used to evaluate prediction performance using common neural network metrics [11] presented in *Appendix C – Article 2: Mathematical Formulations and Loss Function Definitions.* The final NNML model, which outperformed the ST-NN models, was first evaluated on the internal test subset before undergoing external validation (Stage 3).

Classification accuracy was assessed both for the highest-ranked prediction and within the top two predictions. Accuracy within the top two predictions was used to evaluate whether the actual surgical choice was among the model's two highest-ranked outputs. This metric accounts for cases where the model does not make an exact match but still provides a clinically reasonable alternative that could closely aligns with the surgeon's final decision. Classification accuracy was calculated for the highest and top two predictions. For regression tasks, accuracy was assessed with an acceptable variation of ±0.2 for screw density and ±5° for rod curvatures. The number of vertebral levels between predicted and actual UIV and LIV was reported as median and range.

4.2.3.5 Stage 3 - External Testing of the Final Model Performances

To evaluate real-world performance, the final fully trained NNML model was tested on an independent dataset that had never been used during training. This external dataset included patients from a different hospital and surgeries performed by surgeons not involved in the training phase. The same performance metrics as in Stage 2 were used.

4.2.3.6 Statistical Analyses

Descriptive statistics characterized the study sample's demographic, radiographic, and surgical parameters. Preoperative and postoperative radiographic parameters were compared using paired t-tests. Model performance metrics were tested for normal distribution (Shapiro-Wilk) and homogeneity of variance (Levene). One-way ANOVA and multivariate analysis for repeated measures assessed differences between model development and external validation datasets and the effect of multitask versus single-task models on predictions.

4.2.4 Results

4.2.4.1 Data Summary

Of the 915 patients screened, 637 were not eligible and 67 declined (Figure 4.11). From 211 patients with radiographs in the MIMO database, 189 (162 female, 27 male) operated by 22 surgeons were included. Of the 23 excluded, 21 had missing data, and 2 required revision surgery for implant removal. Eight included patients (6.2%) required revision surgeries for seroma, wound infection, or wound dehiscence but remained in the study.

The most frequent Lenke classification was type 1 (n=162, 85.7%) with a neutral thoracic modifier (n=182, 96.3%) and the median Risser sign was 4 (range 0-5). Most patients underwent fusion of 10 vertebral levels (n=68, 36.0%) from T4 (n=86, 45.5%) to L1 (n=82, 43.4%). PSF was performed in 114 cases (60.3%) using 5.5 mm diameter Cobalt Chrome rods.

Postoperatively, coronal deformities significantly improved ($p \leq 0.001$ for all curves) with mean ± SD reductions of 64±18%, 71±13%, and 43±19% for PT, MT, and TL/L curves, respectively. Comparing the model development dataset (n=179) to the external testing dataset (n=10), weight, height, and proximal thoracic Cobb angle were significantly different ($p=0.017$, $p=0.033$, $p=0.036$). Postoperative PT and MT Cobb angles were significantly greater in the external dataset ($p=0.017$, $p=0.020$), associated with lower thoracic curve correction compared to the model development dataset (72±13% vs 55±19%, $p=0.026$). Further details on the data summary are provided in *Appendix D – Article 2: Supplementary Cohort Data.*

4.2.4.2 Final NNML Architecture

The final NNML comprised 83 inputs and 5 outputs (Figure 4.10). Based on the hyperparameter grid search results, a total of 3 hidden shared neurons with 3 and 4 hidden tasks specific neurons for the classification and regression tasks, respectively, were the optimal parameters for the model. The network was trained with a batch size of 4 for up to 250 epochs while monitoring the loss function, halting the training process once the loss stopped improving and a dropout rate of 0.01.

4.2.4.3 Models' Performances during Cross-Validation Experiments

The multivariate analysis for the models robustness showed a statistically significant difference between the cross-validations groups (10-fold, 5-fold, and 3-fold cross-validations) on the combined categorical loss and accuracy measurements, $F(4, 260)=2.8$, $p=0.026$, Wilks' $\Lambda=0.919$ and the combined regression loss and error measurements, $F(6, 404)=4.7$, $p <0.001$, Wilks' $\Lambda=0.873$. The pairwise comparisons indicate that a significant difference exists between the 10-fold and 3-fold cross-validations' categorical loss values for the NNML model ($p=0.014$, Table 4.8) and the regression tasks' RMSE for the ST-NN models ($p=0.010$, Table 4.9). Nevertheless, both models were robust, with fairly competitive classification and regression performances even with fewer data resources for training.

The multivariate analysis for the models' performances showed a strong statistical difference in the combined performance metrics between the NNML and ST-NN models, with the NNML model performing significantly better, $F(24, 260)=146.9$, $p <0.001$, Wilks' $\Lambda=0.010$. The lower overall loss for all tasks combined obtained by the NNML compared to ST-NN ($p <0.001$) also suggests better generalization capacities of NNML to new patients' data. The pairwise comparisons of the classification and regression results obtained after the cross-validation for the training and internal testing datasets from the NNML and compared with the ST-NN models are presented in Figure 4.12**.**

4.2.4.4 Final Model Performances and External Testing

Based on the comparisons from the last sections, we established that the general performances of the NNML model outperformed the ST-NN regardless of the training and validation dataset splits and was, therefore, retained. Table 4.10 summarizes the performance of the final NNML model developed over the test subset from the model development dataset (internal testing, Stage 2) and external performance testing dataset (Stage 3). Accuracy of categorical predictions was reached in 82% and 86% of patients during internal testing and 70% and 80% during external testing for UIV and LIV, respectively. Accuracy increased to 95%, 80%, 97%, and 100% when the accuracy within the 2 best model predictions was calculated for UIV and LIV, respectively. The median number of vertebral levels between the predicted and actual UIV and LIV was 0 and ranged from -3 to 2 and -2 to 2 levels, respectively. When different than the ground truth, most of the model predictions were within 1 vertebral level of the actual surgery performed, as shown in the confusion matrices in Figure 2.13A.

Screw density prediction showed the lowest accuracy across all tasks, with the scatter plot depicting the relationship between the model predictions and the actual surgery performed showing an overall absence of a discernible linear or monotonic pattern across the entire range of predictions (Figure 2.13B). Contrarywise, the model consistently approximated the rod's curvature ground truth values with a high degree of accuracy that reached 90% and 88% of cases during internal testing and 70% and 80% during external testing for the concave and differential rod curvatures, respectively. The relatively tight clustering of data points within a ±5° corridor of acceptable variation seen on the scatter plots depicting the rod curvatures predictions suggests a strong linear relationship between the model's predictions and the actual ground truth, indicating that the model reliably captures the underlying patterns in the rod data (Figure 2.13B). Examples of the fully trained final model's prediction performances on clinical cases are shown in Figure 2.14.

4.2.5 Discussion

This study presents a neural network-based multitask learning model using preoperative clinical and radiographic data to predict key surgical instrumentation parameters for AIS correction. The NNML approach showed increased reliability compared to multiple single-task neural networks and demonstrated good to high accuracy in identifying the upper and lower instrumented vertebrae and rod curvatures. One of the key clinical implications of this model is its potential use as a decision-support tool for surgeons, particularly those with less experience in complex spinal instrumentation planning. AIS surgery requires balancing multiple competing factors and this model could assist in surgical decision-making by integrating multiple parameters simultaneously, providing data-driven recommendations based on real-world surgical outcomes. Moreover, the ability to standardize preoperative planning through AI-based models could help reduce variability between surgeons, leading to more reproducible surgical outcomes. By leveraging AI, the model may enhance preoperative decision-making, allowing for greater consistency in patient-specific instrumentation strategies. Future refinements and validation on larger, diverse datasets are necessary before clinical application, but this study represents an important first step toward AI-assisted preoperative planning in AIS surgery.

4.2.5.1 ***Multitask versus Single-Learning Models***

Multitask learning models aim to learn multiple tasks simultaneously. Initially, it was hypothesized that multitask learning models could outperform single-task models based on the observation that humans usually learn clusters of related tasks instead of isolated ones [16]. This study indicates that an NNML model can significantly outperform ST-NN models for predicting surgical instrumentation parameters for AIS correction. The proposed NNML architecture's ability to allow parameter sharing across tasks improved the model's predictions, suggesting that predicting multiple related surgical instrumentation parameters allowed the model to exploit correlations. Similar findings showed that multitask learning models are superior to single-task learning models in health applications [17, 18].

4.2.5.2 UIV and LIV Prediction Accuracy

Our model demonstrated high accuracy for LIV prediction (80% to 100%), but more discordance existed between model predictions and surgeons' decisions regarding UIV, with a UIV selection accuracy of 70% to 80%. The surgeons' decisions for UIV and LIV selection are usually based on much more data than the model's inputs. One possible reason for lower UIV accuracy is the exclusion of parameters usually considered by the surgeon, such as radiographic and clinical shoulder imbalance [19]. Adding these parameters might improve UIV prediction accuracy. Instrumentation strategies vary among surgeons, meaning differences between model predictions and surgical decisions may reflect variability in experienced surgeons’ decision-making process rather than actual model errors. The accuracy metric for UIV and LIV selection in this study was based on a tolerance of ±1 vertebral level, meaning the model was considered correct if its top prediction differed by one level from the actual surgical choice. While this approach accounts for clinically relevant alternatives, it remains uncertain whether an adjacent vertebral level is always an optimal or functionally equivalent choice. Further validation studies is needed to assess whether the UIV and LIV selected by the model for each patients can produce comparable postoperative outcomes or impact spinal alignment.

4.2.5.3 Rod Curvatures and Limitations in the Developed Model

Managing spinal contour and balance in the sagittal plane is crucial for AIS patients, who typically have a hypokyphotic thoracic spine [20]. Since rod curvatures significantly influence postoperative thoracic kyphosis, rods are bent with a curved profile representative of the desired sagittal contour and a straight profile in the coronal plane to correct scoliotic deformities [21]. The model predicted concave and differential rod curvatures with a difference of less than 5° compared to postoperative rod curvatures. While biomechanical studies show that larger alterations in differential rod contouring affect 3D deformity correction [22, 23], the clinical significance of a 5° difference in rod curvature is likely clinically irrelevant. Nevertheless, the model currently predicts postoperative rod curvatures based on available data. The curvature of rods needed before insertion could be estimated by adding 16 to 20° to the predicted concave rod curvature due to elasto-plastic deformation and spine flexibility [21, 24].

4.2.5.4 External Validation on an External Dataset

External validation on an independent dataset gave some certainty on the model's ability to generalize well on new data and avoid overfitting. The model performed slightly better on the internal test subset compared to the external performance testing dataset, but the difference remained relatively small, suggesting good generalizability. In addition, the model performed well on the external dataset despite different population characteristics, such as preoperative weight, height, and proximal thoracic curve, suggesting its robustness. A model trained on one population and effective on another with different characteristics supports its transferability, reducing the need for extensive data retraining model in every new patient population or setting.

4.2.5.5 Screw Density and its Role in the Model

Unlike UIV, LIV, and rod curvature, which were based on the decision-making of experienced spine surgeons, screw density in this study was determined by the randomized allocation of patients in the MIMO study rather than surgeon preference. The MIMO trial assigned patients to either a high-density (HD, >1.8 screws per level fused) or low-density (LD, <1.4 screws per level fused) group. However, within each assigned category, surgeons still had discretion in the total number and distribution of implants across the instrumented levels. While this dataset structure allowed the model to learn from real-world variations in instrumentation, it does not necessarily reflect surgeon-driven decision-making for screw placement. Future studies incorporating datasets where screw density is determined solely by experienced surgeons' decisions could enhance the model's ability to capture surgeon-driven planning nuances.

4.2.5.6 Limitations

The dataset was from the MIMO clinical trial, where screw density was assigned by randomization into high- and low-density groups rather than reflecting individual surgeon preference. Although surgeons determined the total number and distribution of screws within each density category, this structure does not fully capture surgeon-driven decision-making regarding implant density. Future work using datasets where screw density is solely determined by experienced surgeon planning could improve the model's ability to mimic real-world surgical decision-making. Additionally, the sample size of 189 patients for model development was based on previous studies [25]. Expanding the dataset to larger, more diverse patient populations could help refine the model's algorithm and improve its generalizability. Additionally, incorporating datasets from non-randomized clinical settings would enhance model robustness and applicability across different surgical strategies. The medium-sized sample and relatively complex model could have affected performance. Since surgical instrumentation was based on human decisions, we cannot rule out suboptimal strategies. The model does not incorporate preoperative flexibility of the scoliotic spine, a key aspect in surgical planning [26]. Incorporating this parameter could improve prediction accuracy.

4.2.6 Conclusion

This study developed an accurate, externally validated NNML model that mimics experienced surgeons' decision-making in selecting instrumented vertebrae, pedicle screws, and rod curvatures for AIS patients. The proposed model is a first step in providing spine surgeons with a more comprehensive identification of patient-specific instrumentation strategies to achieve reproducible AIS outcomes. This study is original in using an artificial neural network-based multitask learning model to predict surgical instrumentation elements and deep learning for preoperative planning of PSF instrumentation. The NNML model can serve as a foundation for future tools to improve AIS and spinal surgery.

4.2.7 References

1. Dunn J, Katz DE, Roberts DW, Dormans JP, Weinstein SL (2018) Screening for adolescent idiopathic scoliosis: evidence report and systematic review for the US Preventive Services Task Force. JAMA 319(2):173–187. https:// doi. org/ 10. 1001/ jama.2017.19342

2. Weinstein SL, Dolan LA, Wright JG, Dobbs MB (2013) Effects of bracing in adolescents with idiopathic scoliosis. N Engl J Med 369(16):1512–1521. https://doi.org/10.1056/NEJMoa1307337

3. Thomas JJ, Nelson SC, Boan C, Herring JA (2018) Does school screening affect scoliosis curve magnitude at presentation to a pediatric orthopedic clinic? Spine Deform 6(4):403–408. https:// doi.org/10.1016/j.jspd.2018.01.005

4. Negrini S, Donzelli S, Aulisa AG, Czaprowski D, Schreiber S, de Mauroy JC et al (2014) Actual evidence in the medical approach to adolescents with idiopathic scoliosis. Eur J Phys Rehabil Med 50(1):87–92

5. Le Navéaux F, Larson AN, Labelle H (2018) Significant variabil ity in surgeons' preferred correction maneuvers and instrumen tation strategies when planning adolescent idiopathic scoliosis surgery. Scoliosis Spinal Disord 13:21. https:// doi. org/ 10. 1186/ s13013-018-0169-8

6. Schlager B, Großkinsky M, Ruf M, Wiedenhöfer B, Akbar M, Wilke HJ (2024) Range of surgical strategies for individual ado lescent idiopathic scoliosis cases: evaluation of a multi-centre survey. Spine Deform 12(1):35–46. https:// doi. org/ 10. 1007/ s43390-023-00756-0

7. King HA, Moe JH, Bradford DS, Winter RB (1983) The selection of fusion levels in thoracic idiopathic scoliosis. J Bone Joint Surg Am 65(9):1302–1313

8. Lenke LG, Betz RR, Harms J, Bridwell KH, Clements DH, Lowe TG et al (2001) Adolescent idiopathic scoliosis: a new classifica tion to determine extent of spinal arthrodesis. J Bone Joint Surg Am 83(8):1169–1181

9. Pasha S, Mac-Thiong JM (2020) Defining criteria for optimal lumbar curve correction following the selective thoracic fusion surgery in Lenke 1 adolescent idiopathic scoliosis: developing a decision tree. Eur J Orthop Surg Traumatol 30(3):513–522. https://doi.org/10.1007/s00590-019-02569-7

10. Coe JD, Arlet V, Donaldson W, Berven S, Hanson DS, Mudiyam R et al (2006) Complications in spinal fusion for adolescent idi opathic scoliosis in the new millennium. A report of the Scolio sis Research Society Morbidity and Mortality Committee. Spine 31(3):345–349. https:// doi. org/ 10. 1097/ 01. brs. 00001 97188. 76369.13

11. Constant C, Aubin CE, Kremers HM, Garcia DVV, Wyles CC, Rouzrokh P, Larson AN (2023) The use of deep learning in medi cal imaging to improve spine care: a scoping review of current literature and clinical applications. N Am Spine Soc J 15:100236. https://doi.org/10.1016/j.xnsj.2023.100236

12. Larson AN, Upasani VV, Cahill PJ, Alanay A, Miyanji F, Spon seller PD et al (2023) The effect of implant density on adolescent idiopathic scoliosis fusion: Results of the Minimize Implants Maximize Outcomes randomized clinical trial. J Bone Joint Surg Am 105(14):1249–1257. https://doi.org/10.2106/JBJS. 22.00752

13. Wang X, Aubin CE, Arnoux PJ, Petit Y, Zeng L, Labelle H et al (2023) Concave rod first vs. convex rod first in AIS instrumen tation with differential rod contouring: computer modeling and simulations based on ten AIS surgical cases. Spine Deform 11(6):1317–1324. https://doi.org/10.1007/s43390-023- 00755-1

14. Dharamvir AKV (2020) Data normalization techniques on intru sion detection for dataset applications. Int J Adv Sci Technol 29(7s):5083–5093

15. Rodriguez JD, Perez A, Lozano JA (2009) Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell 32(3):569–575. https:// doi. org/ 10. 1109/ TPAMI.2009.187

16. Caruana R (1993) Multitask learning: a knowledge-based source of inductive bias. In: Proceedings of the tenth international Spine Deformity conference on machine learning, Amherst, MA: Morgan Kauf mann. p. 41–48.

17. Liu M, Cheng D, Yan W, Liu H, Zhang D (2019) Joint classifi cation and regression via deep multi-task multi-channel learn ing for Alzheimer's disease diagnosis. IEEE Trans Biomed Eng 66(5):1195–1206. https://doi.org/10.1109/TBME.2018.2863263

18. Ding M, Liu M, Utriainen M, Cheng D, Zhang D (2018) Deep Insight: multi-task multi-scale deep learning for mental disorder diagnosis. British machine vision conference (BMVC). BMVA Press, Newcastle

19. Elfiky TA, Rosenberg JB, Michael KW, Samdani AF, Betz RR, Bastrom TP et al (2011) The proximal thoracic curve in adolescent idiopathic scoliosis: surgical strategy and management outcomes. Glob Spine J 1(1):27–36. https:// doi. org/ 10. 1055/s- 0031- 12747 83

20. Newton PO, Yaszay B, Upasani VV, Bastrom TP, Betz RR, Lenke LG et al (2010) Preservation of thoracic kyphosis is critical to maintain lumbar lordosis in the surgical treatment of adolescent idiopathic scoliosis. Spine 35(14):1365–1370. https://doi.org/10. 1097/BRS.0b013 e3181c2a82b

21. Cidambi KR, Glaser DA, Bastrom TP, Newton PO (2012) Post operative changes in spinal rod contour in adolescent idiopathic scoliosis: an in vivo deformation study. Spine 37(18):1566–1572. https://doi.org/10.1097/BRS.0b013e318252ccbe

22. Le Navéaux F, Aubin CE, Parent S, Newton PO, Labelle H (2017) 3D rod shape changes in adolescent idiopathic scoliosis instru mentation: how much does it impact correction? Eur Spine J 26(6):1676–1683. https://doi.org/10.1007/s00586-017-4958-1

23. Wang X, Aubin CE, Arnoux PJ, Petit Y, Zeng L, Labelle H (2016) How does differential rod contouring contribute to 3-dimensional correction and affect the bone-screw forces in adolescent idio pathic scoliosis instrumentation? Clin Biomech (Bristol, Avon) 39:115–121. https://doi.org/10.1016/j.clinbiomech.2016.09.011

24. Salmingo RA, Cheng PG, Nolte LP, Reyes M, Zheng G (2014) Influence of implant rod curvature on sagittal correction of scolio sis deformity. Spine J 14(8):1432–1439. https:// doi. org/ 10. 1016/j. spinee.2013.08.038

25. Lafage R, Elysee J, Ames CP, Passias PG, Bess S, Shaffrey CI et al (2021) Predictive model for selection of upper treated vertebra using a machine learning approach. World Neurosurg 146:e225 e232. https://doi.org/10.1016/j.wneu.2020.10.110

26. Zhang M, Li S, Wang L, Yang C, Li X (2023) Correlation between supine flexibility and postoperative correction in adolescent idi opathic scoliosis. BMC Musculoskelet Disord 24(1):126. https:// doi.org/10.1186/s12891-023-06146-

4.2.8 Tables

Table 4.7 Table Summary of Data Collected from Dataset

Demographic and clinical data	- Height - Weight - Age
Radiographic	- Vertebral corner localization - Lenke classification - Cobb angle measurements (PT, MT, TL/L) - Upper and lower vertebrae - Stable vertebrae - Thoracic kyphosis (T2-T12) - Risser sign
Surgical	- Upper instrumented vertebra - Lower instrumented vertebra - Screw density - Rod curvatures (concave side and concave/convex side differential bending) - Type of screws and rods used
Postoperative spinal alignment	- Cobb angle measurements (PT, MT, TL/L) - Thoracic kyphosis (T2-T12)

Table 4.8 Loss and Accuracy Measure Comparison for the Classification Tasks Between Training and Validation Using the Neural Network-Based Multitask Learning (NNML) and Single-Task Neural Network (ST-NN) Models

*Performance metrics' mean and standard deviation (Std dev) of both models were calculated over 3 experiments: 10-fold, 5-fold, and 3-fold cross-validations subsets. * indicates a significant difference between the models with p<0.05*

		Training				**Validation**			
		NNML		**ST-NN**		**NNML**		**ST-NN**	
Training/ Validation Split	Metrics	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Experiment 1: 90/10	Loss	0.28	0.06	0.32	0.05	0.52*	0.17	0.49	0.21
	Accuracy	0.95	0.03	0.86	0.08	0.81	0.05	0.79	0.06
Experiment 2: 80/20	Loss	0.35	0.04	0.38	0.06	0.58	0.08	0.47	0.11
	Accuracy	0.94	0.02	0.90	0.06	0.79	0.02	0.79	0.06
Experiment 3: 70/30	Loss	0.28	0.03	0.31	0.05	0.77*	0.30	0.50	0.23
	Accuracy	0.95	0.02	0.93	0.03	0.77	0.04	0.75	0.03

Table 4.9 Loss and Errors Comparison for the Regression Tasks Between Training and Validation Using the Neural Network-Based Multitask Learning (NNML) and Single-Task Neural Network (ST-NN) Models

*Performance metrics' mean and standard deviation (Std dev) of both models were calculated over 3 experiments: 10-fold, 5-fold, and 3-fold cross-validations subsets. * indicates a significant difference between the models with p<0.05*

		Training				Validation			
		NNML		ST-NN		NNML		ST-NN	
Training/ Validation Split	Metrics	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Experiment 1: 90/10	Loss	1.44	0.90	2.27	1.70	1.97	1.29	2.64	1.87
	RMSE	1.69	1.11	3.02	2.24	2.78	1.83	3.58*	2.48
	MAE	1.12	0.72	2.14	1.67	1.97	1.29	2.64	1.87
Experiment 2: 80/20	Loss	1.58	0.98	2.68	2.13	2.10	1.35	2.93	2.15
	RMSE	1.84	1.16	3.33	2.65	3.00	1.97	3.98	2.83
	MAE	1.22	0.75	2.42	2.04	2.10	1.35	2.93	2.15
Experiment 3: 70/30	Loss	1.51	0.93	2.82	2.04	2.21	1.41	3.21	2.15
	RMSE	1.90	1.19	3.63	2.55	3.17	2.05	7.10*	1.21
	MAE	1.28	0.79	2.62	1.91	2.21	1.41	3.21	2.15

Table 4.10 Performance Measures of the Final Fully Trained Neural Network-Based Multitask Learning (NNML) Applied over 2 Independent Datasets

Performance metrics of the trained and optimized model developed were calculated over the test subset from the model development dataset (internal testing; n=36 patients) and external performance testing on the independent dataset (n=10 patients).

		Model Development – Internal Testing	**Model Performance – External Testing**
Upper instrumented vertebra (UIV)	Accuracy	**0.82**	**0.70**
	Accuracy within 2 best predictions	**0.95**	**0.80**
	Precision	0.86	0.62
	F1-score	0.85	0.51
Lower instrumented vertebra (LIV)	Accuracy	**0.86**	**0.80**
	Accuracy within 2 best predictions	**0.97**	**1.00**
	Precision	0.88	0.75
	F1-score	0.88	0.75
Screw Density	Accuracy ±0.2	**0.59**	**0.40**
	Root means square error (RMSE)	0.24	0.30
	mean absolute error (MAE)	0.20	0.24
	mean absolute percentage error (MAPE)	13.12	15.00
	explained variance (EV)	0.21	-0.15
Concave Rod Curvature	Accuracy ±5°	**0.90**	**0.70**
	RMSE	3.81	5.57
	MAE	2.60	4.35
	MAPE	17.55	20.72
	EV	0.71	-0.13
Differential Rod Curvature	Accuracy ±5°	**0.88**	**0.80**
	Root mean square error (RMSE)	3.69	5.60
	mean absolute error (MAE)	2.57	4.33
	mean absolute percentage error (MAPE)	57.48	45.42
	explained variance (EV)	0.74	0.34

4.2.9 Figures

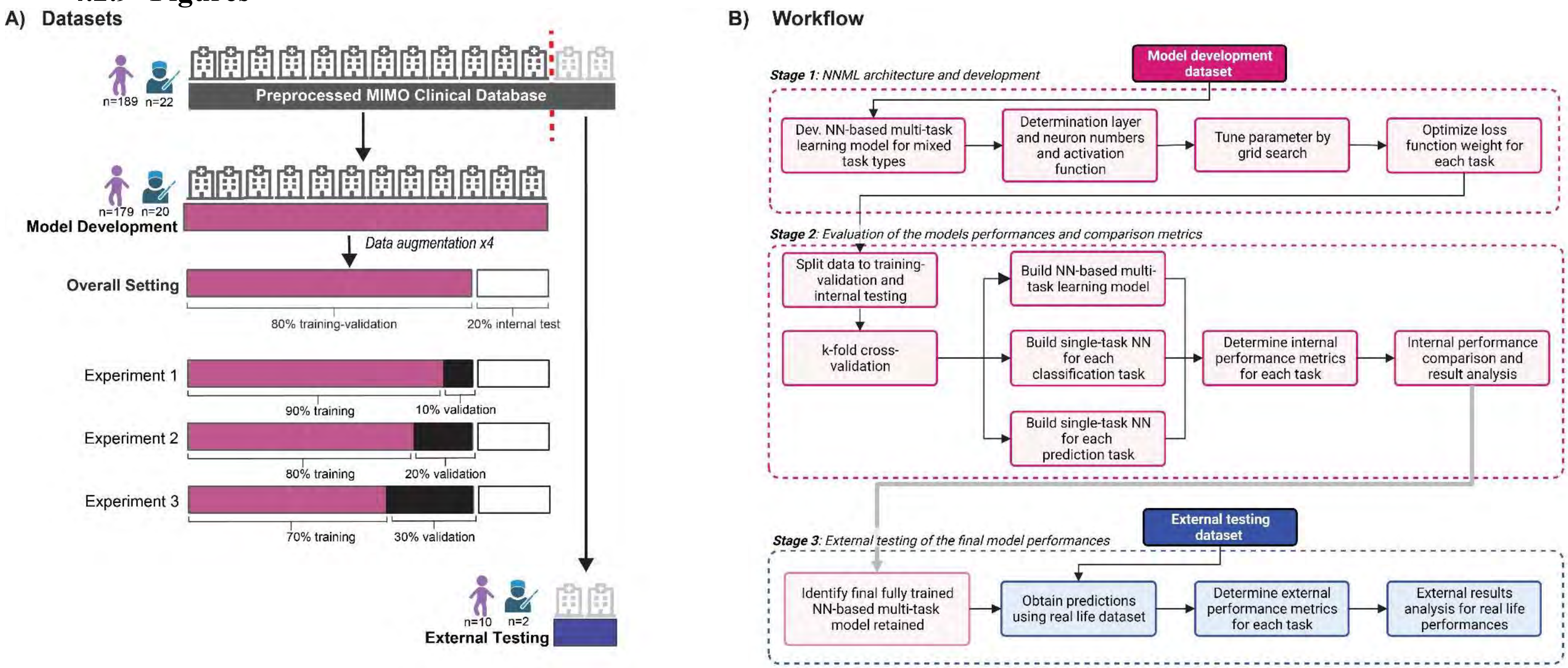


Figure 4.9 Overview of the Study Methodology

A) Flowchart of the study dataset assignment detailing the number of patients included and stratified into the 2 study sub-groups: model development, including the training and testing dataset settings and external performance testing, kept completely independent from model development.

B) Workflow of the research framework depicting the 3 main sequential stages from neural network-based multitask learning model (NNML) development to external performance testing of the final NNML network on a real-life dataset.

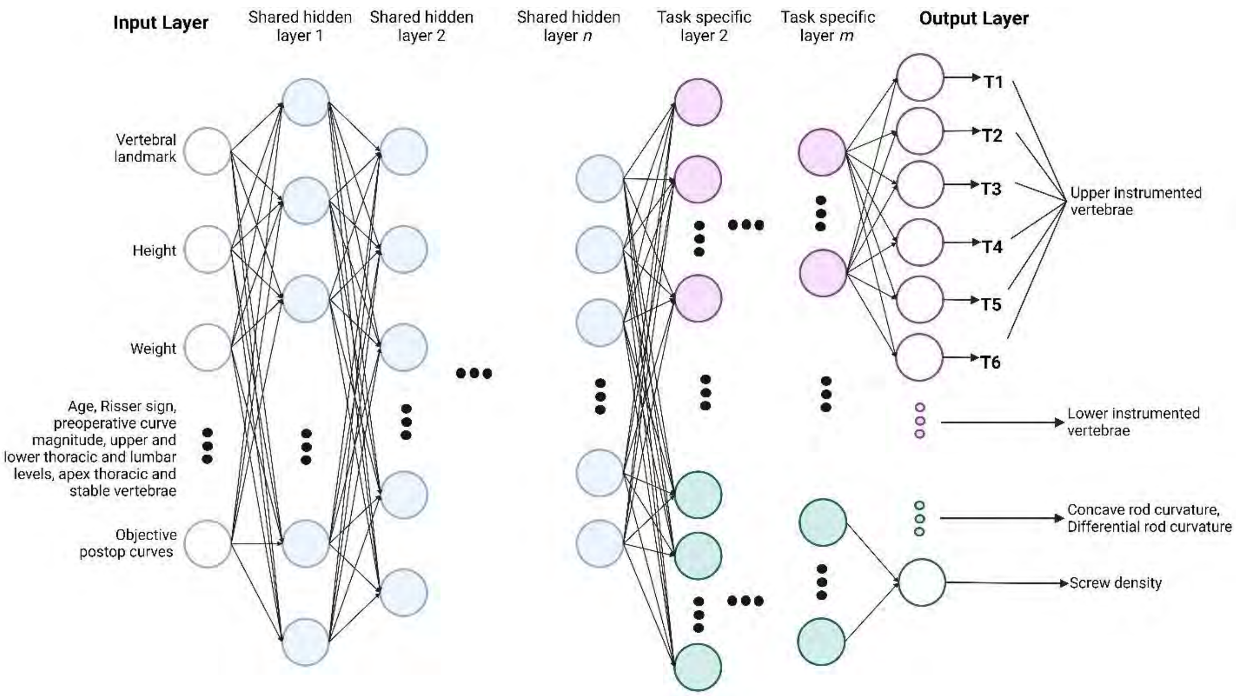


Figure 4.10 Neural Network-Based Multitask Learning Model using Preoperative Clinical And Radiographic Data from AIS Patients for PSF Surgical Instrumentation Prediction

The developed conceptual model comprises 77 inputs with shared and task-specific layers constructed to simultaneously give all 5 outputs from the two task types (regression and classification) using only one model with shared features by optimizing task-specific functions in one step.

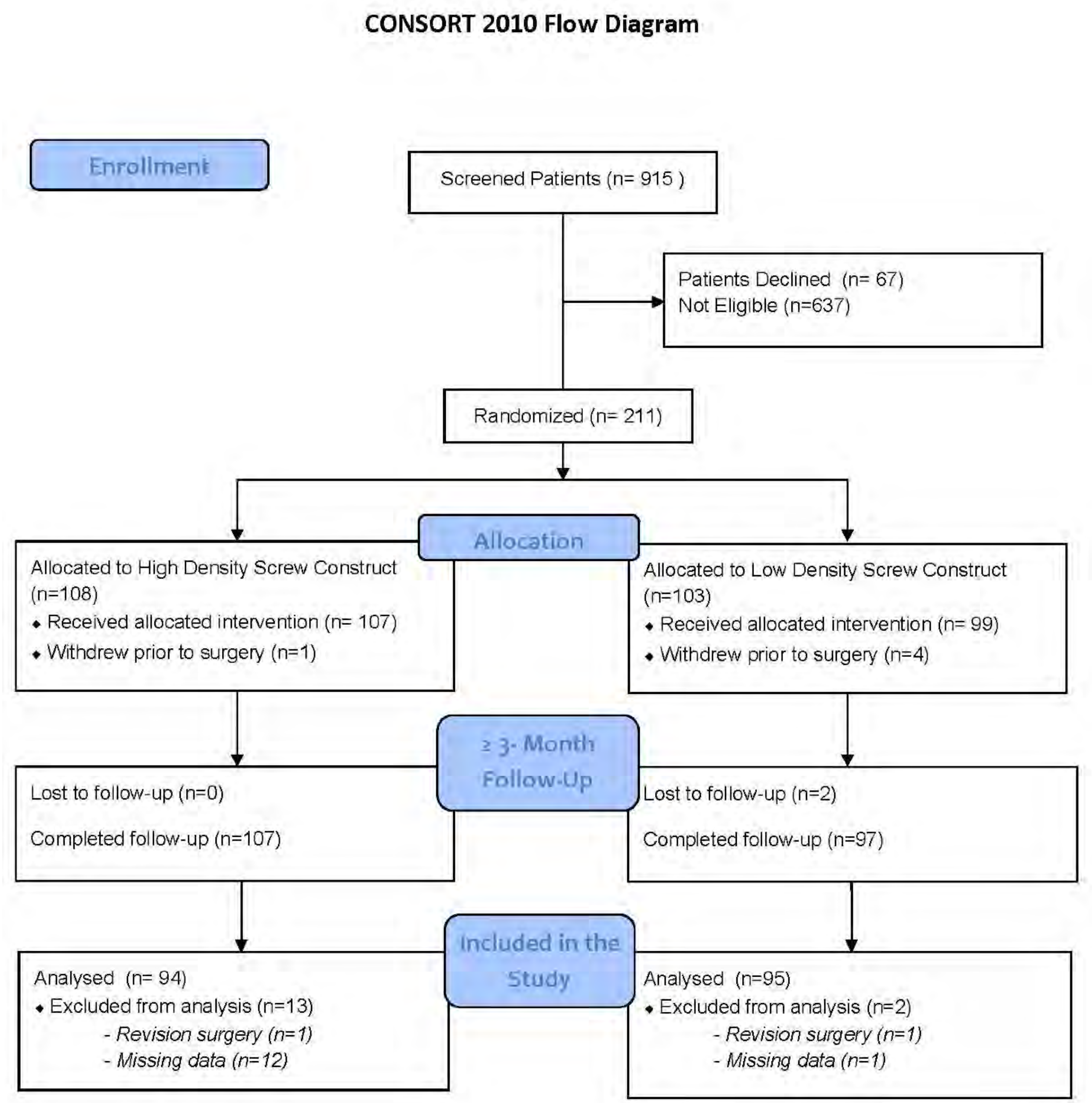


Figure 4.11 Study Enrollment and Treatment of the Patients

Between February 2013 and August 2017, a total of 915 patients were screened. Of those, 211 patients were eligible, consented, and underwent randomization and were included in the intention-to-treat population. There was no significant difference in loss to follow-up between the groups.

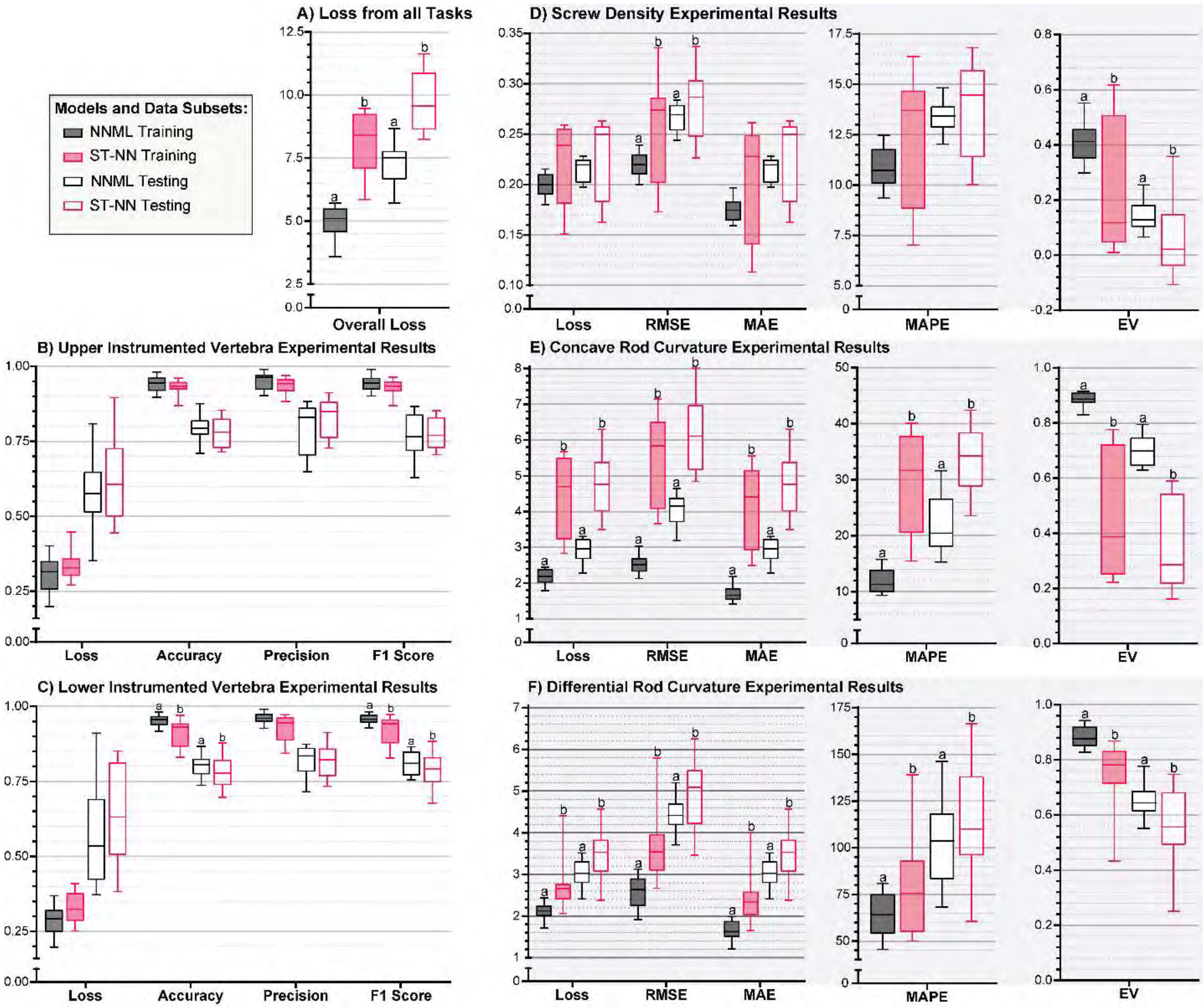


Figure 4.12 Performance Metrics Comparison Between Training and Testing Obtained during Model Development Stages (N=179 Patients) using the Neural Network-Based Multitask Learning (NNML) and Single-Task Neural Network (ST-NN) Models

Performance metrics of both models were calculated over all folds (n=17 folds) from 3 experiments: 10-fold, 5-fold, and 3-fold cross-validations subsets. The training results (grey) were obtained during the training process using the validation subsets (8%, 16%, or 24% of model development dataset based on the experiment) and testing results (pink) using the internal testing subset (20% of the model development dataset).

In the box plots, the box's middle line indicates the mean, the box's boundaries are the 25th and 75th percentiles, and the whiskers are the 10th and 90th percentiles.

** Different superscript within the same metric indicates a significant difference between the models with $p<0.05$*

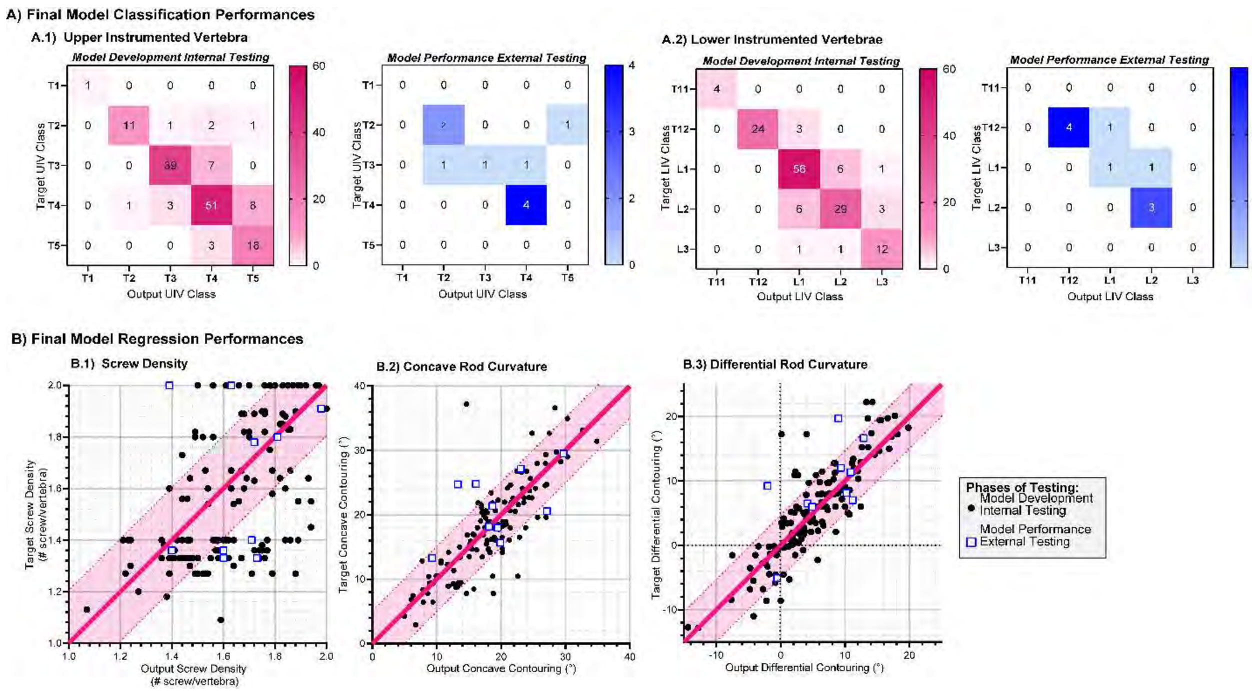


Figure 4.13 Final Classification And Regression Performances of the Fully Trained Neural Network-Based Multitask Learning (NNML) Applied over two Independent Datasets

(A) Classification results: Confusion matrices for the upper (UIV, A.1) and lower (LIV, A.2) instrumented vertebrae, calculated over the test subset from the model development dataset (internal testing; pink, n=36 patients) and external performance testing on the independent dataset (blue; n=10 patients), a dataset never seen by the model before this stage to represent real-life performance. The target class was obtained from the ground truth (actual surgery performed; y-axis) and the output class (x-axis) by the trained and optimized model. Numbers in the square charts represent incidence, with the shade of color indicating the relative abundances of the corresponding vertebrae in each dataset.

(B) Regression results: Scatter plots showing the relationship between the target and output values over the test subset from the model development dataset (internal testing; black circles, n=36 patients) and external performance testing on the independent dataset (blue squares; n=10 patients). The target values were obtained from the ground truth (actual surgery performed; y-axis), and the output values (x-axis) from the trained and optimized model. The pink line indicates perfect agreement, and the shaded corridors indicate variation from the ground truth considered acceptable (±0.2 and ±5° for screw density and rod curvatures, respectively).

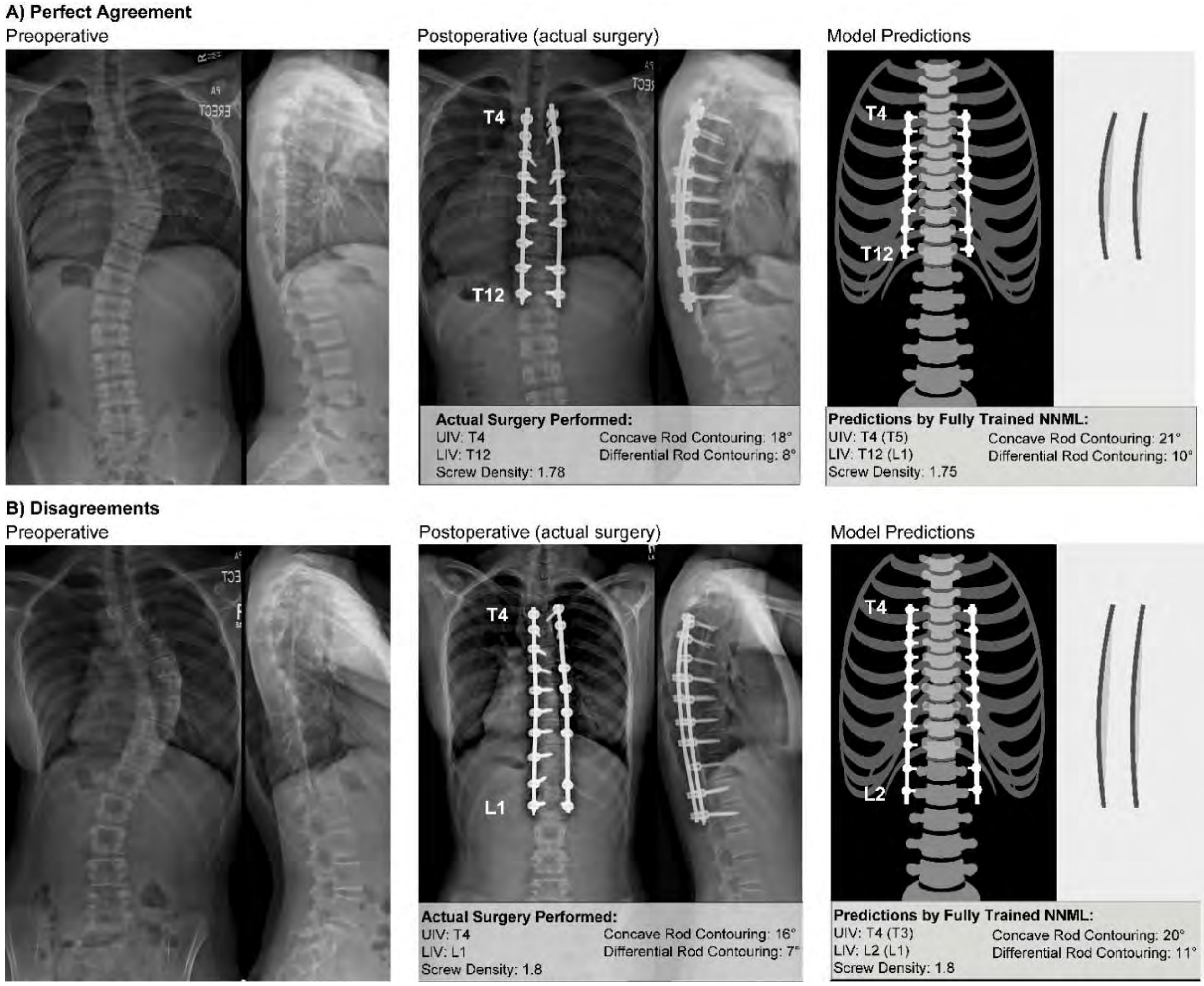


Figure 4.14 Examples of Pre- And Postoperative Radiographic Images and Instrumentation Predictions Obtained from the Fully Trained Neural Network-Based Multitask Learning (NNML)

The 2 clinical cases were part of the external performance testing on the independent dataset to represent the real-life performances and selected as examples for the "best" (A) and "worst" (B) case scenarios of agreement between the model's predictions and the ground truth (actual surgery performed).

4.3 Complementary Methodological Aspects

This section complements the methodological description in Article 2 by detailing foundational components of the neural network architecture used in our NNML model for PSF planning in AIS. While Article 2 emphasized high-level architecture and clinical outcomes, the following subsections describe the architectural components, activation functions, loss functions, and learning algorithms underpinning the network's performance. These components are further illustrated and summarized in *Appendix E – Article 2: NNML Model Architecture*. In addition, *Appendix C* provides a comprehensive list of all equations used in this section, and *Appendix D* contains supplementary graphical representations of the AIS patient population included in the NNML development phase but not shown in the final manuscript.

4.3.1.1 Architecture of the Neural Network Used

The NNML developed in this study integrates two key types of artificial neural networks: Multilayer Perceptrons (MLP) and Recurrent Neural Networks (RNN), specifically Long Short-Term Memory (LSTM) units. These were strategically combined to support multimodal input processing and the simultaneous prediction of multiple surgical parameters, including UIV and LIV, rod curvatures, and screw density (see *Appendix E* for more detailed topographic information on the final architecture). This hybrid architecture was selected to combine the strength of MLP in handling heterogeneous tabular data with the ability of LSTM to capture spatial dependencies along the vertebral column, an inherently ordered anatomical structure.

Multilayer Perceptrons

MLP forms the computational backbone of the shared and task-specific branches in our NNML model. MLPs are a class of feedforward artificial neural networks composed of multiple layers of interconnected artificial neurons. Each neuron computes a weighted sum of its inputs, adds a bias term, and passes the result through a non-linear activation function, enabling the network to learn complex, non-linear relationships between inputs and outputs. MLPs were chosen for their effectiveness in handling structured data and non-linear relationships, such as those present between clinical variables (e.g., Cobb angle, age) and surgical decisions regarding the instrumentation. Their simplicity and adaptability made them ideal for the shared and task-specific branches of our multitask model. As shown in Figure 4.15A, the MLP component propagates input signals forward and updates weights using backpropagation during training, forming the foundation of hierarchical feature learning in our model.

In our model, fully connected MLP layers were used to process both numerical and categorical data derived from standardized clinical and radiographic parameters, such as Cobb angles and patient demographics. The input layer receives the feature vector $X = \{x_1,\ x_2, \dots , x_n\}$. Each hidden layer performs a series of transformations using learned weights and biases, followed by an activation function. The final output layer produces predictions $Y = \{y_1,\ y, \dots , y_m\}$, corresponding to either categorical variables (UIV and LIV) or continuous variables (rod curvature and screw density). The total number of trainable parameters in an MLP depends on its architecture and can be estimated as follows [240]:

$$n \cdot h_1 + \sum_{k=1}^{N-1} h_k \cdot h_{k+1} + h_N \cdot m$$

where n is the number of inputs, h_k is the number of neurons in the h_k-th hidden layer, and m is the number of output neurons. As this expression indicates, network complexity and training time grow substantially with increasing depth and layer width.

In our final optimized NNML architecture, the MLP components included 37,050 total parameters, of which 37,047 were trainable. These were distributed across shared dense layers and task-specific branches, which were responsible for regression and classification outputs. The remaining 3 non-trainable parameters stemmed from the normalization layer applied before the first dense layer. This configuration provided sufficient learning capacity while preserving model tractability and avoiding overfitting.

Recurrent Neural Networks

To model the sequential and spatial organization of the spine, we incorporated a RNN layer, specifically a LSTM unit, at the input stage of our NNML model. This LSTM layer processed the ordered sequence of 68 vertebral coordinates (x, y), allowing the network to recognize spatial patterns along the spine. Unlike MLPs, which treat each input independently, RNNs retain memory of previous inputs, making them especially effective for data with a natural order [241], such as vertebral landmarks. Each LSTM unit employs gating mechanisms (input, forget, and output) to determine which information to retain or discard at each time step. These mechanisms enable the network to learn how individual vertebrae interact with one another across the spine. The output sequence from the LSTM layer was concatenated with patient-specific clinical and radiographic features, then passed into downstream dense layers.

As shown in Figure 4.15B, this recurrent structure allows each output to be fed into the next time step, effectively learning ordered anatomical dependencies not captured by standard MLPs. The LSTM unit was chosen explicitly over basic RNNs due to its superior ability to retain long-term dependencies and avoid vanishing gradient problems, which are important when modeling relationships along the entire spine. This architectural choice ensured that spatial patterns among vertebral coordinates could be learned efficiently, improving the prediction of location-based decisions such as UIV and LIV. To our knowledge, this is the first time a RNN has been applied to capture vertebral patterns for surgical planning or any other type of prediction.

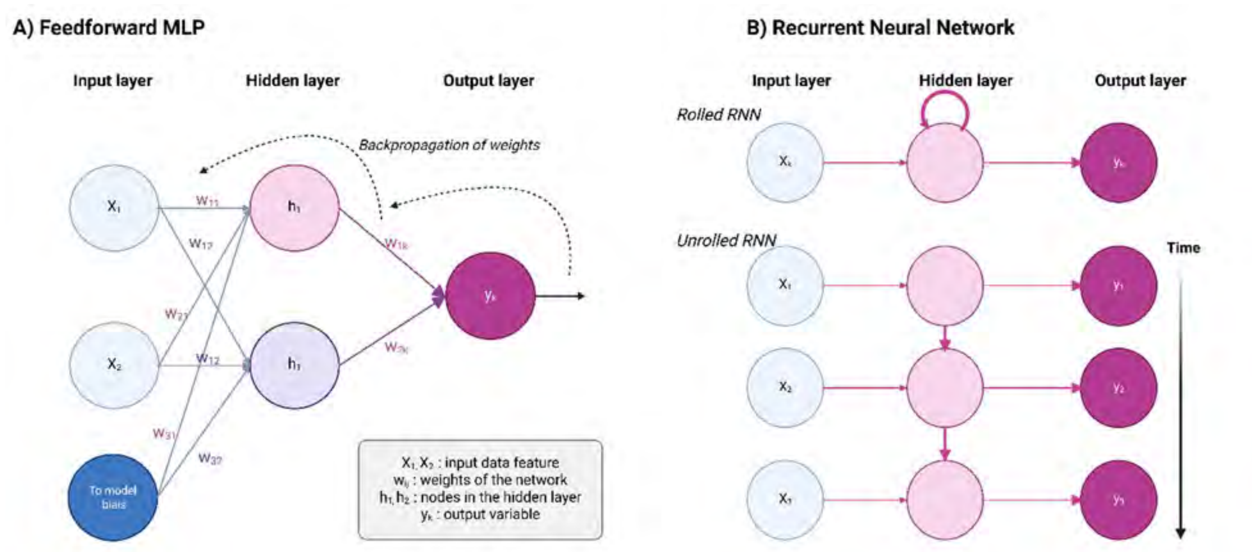
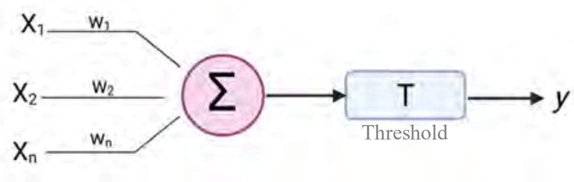


Figure 4.15 General Architectures of the Neural Networks used in Article 2

A) Structure of a single hidden-layer MLP, highlighting the feedforward and backpropagation steps. B) Recurrent Neural Network (RNN) for sequential data.

4.3.1.2 Perceptron and Activation Function

Neurons in the NNML perform calculations based on a perceptron model. A perceptron computes a weighted sum of its inputs and applies an activation function. In the classical case, this output is binary: 1 if the sum exceeds a threshold, 0 otherwise (Figure 4.16).



$$y = \begin{cases} 1, if \ \overbrace{\sum_i w_i x_i}^{\text{Weighted sum}} - \underbrace{T}_{\text{Threshold}} > 0 \\ 0, otherwise \end{cases}$$

Figure 4.16 Perceptron Neuron Model and Threshold Logic

Perceptron neuron model illustrating the basic structure and functioning (left) and threshold logic demonstrating the activation function based on input summation (right).

Rectified Linear Unit for Shared and Numerical Layers

Rectified Linear Unit (ReLU) was used for all hidden layers in the shared and regression pathways. It is defined as:

$$R(z) = \max(0, z)$$

where *z* represents the input to the ReLU (R) function. This simple function outputs the input value directly if it is positive and zero otherwise (Figure 4.17). ReLU introduces nonlinearity while being computationally efficient, reducing training time and helping to avoid overfitting compared to sigmoid or tanh activations. It also generally performs well without extensive tuning [242]. ReLU was selected because it accelerates convergence during training, helps mitigate vanishing gradient issues, and has demonstrated better empirical performance in DL models for both medical and non-medical applications [242].

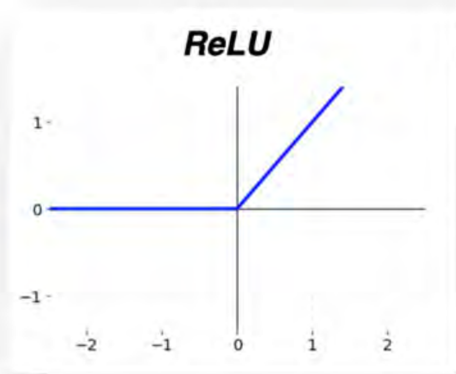


Figure 4.17 Graphical Representation of the ReLU Function.
This figure shows the ReLU function used in both shared and numerical prediction layers, illustrating its behavior and impact on neural network activation.

Softmax for Categorical Layers

Softmax was used in output layers for categorical classification (UIV and LIV). It converts a raw score vector into a probability distribution:

$$\sigma(z)_a = \frac{e^{z_a}}{\sum e^{z_a}}$$

where *z* is the input vector to the Softmax function and *a* is an index that specifies the input vector's element. In addition, e^{za} is the exponential of the input z for class a, and the denominator is the sum of the exponentials of the inputs for all classes in the dataset (Figure 4.18). This ensures that probabilities for all classes sum to 1. The class with the highest probability becomes the prediction. In our model, each neuron corresponded to a vertebral level. Softmax was selected because it is the standard activation function for multi-class classification problems, such as predicting UIV and LIV, where each vertebral level corresponds to a discrete class [243].

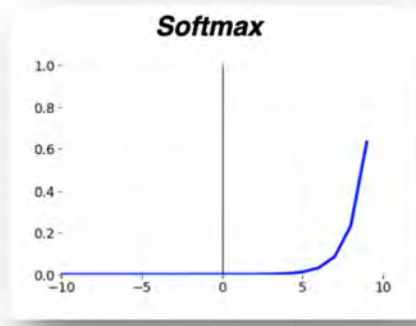


Figure 4.18 Graphical Representation of the Softmax Function.

This figure shows the Softmax function used in categorical prediction layers, illustrating its behavior and impact on neural network activation.

4.3.1.3 Loss Functions

The NNML model used task-specific loss functions to optimize prediction performance across both regression and classification tasks.

Mean Squared Error

For regression outputs, namely screw density and rod curvature predictions, the model employed the Mean Squared Error (MSE) loss function. MSE is calculated as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where N is the number of data points in the dataset, yi and $\hat{y}_i$ are the actual (true) and predicted values for each data point, respectively. In addition, $\frac{1}{N}\sum_{i=1}^{N}$ represents the mean operator for all N data points in the dataset. This function penalizes large errors more heavily than small ones, making it particularly suitable when minimizing the impact of outliers is critical [219]. MSE is widely used in regression tasks due to its stability and differentiability, which are essential for gradient-based learning. MSE was chosen for rod curvature and screw density tasks due to its sensitivity to larger deviations, ensuring accurate modeling of continuous surgical parameters for promoting realistic predictions.

Categorical Cross-Entropy

For classification tasks, specifically the prediction of UIV and LIV instrumented vertebrae, the model used categorical cross-entropy (CE) loss, defined as:

$$CE = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

Where N is the number of data points in the dataset, y_i is the actual (true) value in a one-hot encoded vector format for each data point, and $\hat{y}_i$ is the predicted probability distribution for each data point. In addition, $-\frac{1}{N}\sum_{i=1}^{N}$ represents the negative mean operator applied over the log probabilities for all N data points in the dataset. CE quantifies the distance between the predicted and actual (true) distributions, penalizing incorrect classifications with high confidence. Only one correct class is assumed per data point, and the loss encourages the network to assign high probability to this true class. CE was chosen as it is the standard loss for classification problems where outputs represent mutually exclusive classes, such as UIV and LIV.

Combined Loss Function

To enable simultaneous optimization across all tasks, a combined loss function was constructed by weighting the loss contributions from each task. This multi-objective formulation is written as:

$$L_{total} = \sum_{t}^{T} \beta(MSE_t) + \alpha(CE_t)$$

where T is the total number of data points for which the loss is being calculated, t is the total number of tasks, β and α are weighting coefficients (hyperparameters) for the MSE and CE loss components, respectively. MSE_t is the Mean Squared Error loss for the t-th task, and CE_t is the cross-entropy for the t-th task. In addition, $\sum_{t}^{T}$ represents the summation over all tasks and data points. The weighting coefficients for the classification (β) and the regression (α) loss functions are hyperparameters that control the relative influence of each task during training. They were optimized during model development to balance task performance and minimize overfitting [244]. The weighting of the combined loss terms was selected to ensure that no task dominated the learning process. This design choice was necessary because multitask learning can otherwise suffer from task imbalance, leading to poor generalization on one or more outputs.

4.3.1.4 Backpropagation

Backpropagation is the core learning algorithm used to update model weights. It computes the gradient of the loss function with respect to each weight in the network via the chain rule, allowing for efficient parameter updates using gradient descent.

Each iteration involves:

- **Forward pass:** Computing the output of the network given current weights.
- **Loss computation:** Calculating the error between prediction and ground truth using MSE or CE.
- **Gradient computation:** Deriving partial derivatives of the loss with respect to weights.
- **Weight update:** Adjusting weights using the learning rate and gradient.

In our model, ReLU and Softmax were chosen for their differentiability, which is essential for this process. The optimization algorithm used was Adaptive Moment Estimation (Adam), which adjusts learning rates per parameter. Adam is computationally efficient, typically converges faster than classical Stochastic Gradient Descent, and performs well without extensive hyperparameter tuning [245].

The model was implemented in Python 3.9.13 using TensorFlow 2.10.0, developed by Google Brain, and distributed via Anaconda, Inc. The Adam optimizer used corresponds to TensorFlow's built-in implementation (tf.keras.optimizers.Adam) and includes advanced features such as adaptive learning rates and weight decay. The training and optimization of the developed NNML model were based on 10-fold cross-validation [246]. The validation dataset was divided into 10 folds: 9 folds for training and the remaining 1-fold for validation. The final error was estimated by averaging the errors committed in each fold [246]. A random dropout technique and batch normalization were used as a regularization method during the network training process to reduce overfitting [247].

The Adam optimizer was chosen for its computational efficiency, adaptive learning rate adjustment, and lower sensitivity to initial hyperparameter settings, which made it ideal for training a complex model with multiple loss functions. A 10-fold cross-validation was used to ensure robust model evaluation and avoid performance inflation due to random data splits, particularly in medium-sized datasets. Random dropout and batch normalization were implemented as regularization techniques to reduce overfitting and improve generalization, based on their well-documented success in DL applications [246, 247].

CHAPTER 5 HYBRID NUMERICAL MODEL COMBINING ARTIFICIAL INTELLIGENCE AND DETERMINISTIC MODELING

5.1 ARTICLE 3 : AI-Derived vs. Surgeon-Performed Instrumentation in Adolescent Idiopathic Scoliosis: A Biomechanical Simulation Analysis

To address Sub-objective 2 (SO2), this study developed and evaluated a novel methodological approach that combines data-informed surgical planning with patient-specific deterministic biomechanical simulations for PSF surgery in AIS. While previous work has proposed predictive tools for instrumentation planning, this is the first study to assess their biomechanical performance in individual patients using validated multibody models. The primary contribution lies in integrating planning strategies derived from clinical data with detailed biomechanical evaluation, offering a framework to support surgical decision-making. By comparing predicted and surgeon-performed instrumentation strategies, the study demonstrates the feasibility of a hybrid planning method that enables patient-specific construct selection and sets the stage for future optimization using simulation.

The resulting article, titled "*AI-Derived vs. Surgeon-Performed Instrumentation in Adolescent Idiopathic Scoliosis: A Biomechanical Simulation Analysis*", was submitted to Spine Deformity in September 2025. The first author's contribution to the study design, simulations, data analysis, and manuscript preparation is estimated at over 80%.

Article 3: Constant, C., Larson, A.N., Polly, D.W. et al. **AI-Derived vs. Surgeon-Performed Instrumentation in Adolescent Idiopathic Scoliosis: A Biomechanical Simulation Analysis.**

Highlights

First biomechanical evaluation of AI-derived vs. surgeon-performed instrumentation in AIS

AI-derived plans achieved comparable overall 3D correction to surgeons, with better restoration of thoracic kyphosis. Surgeon plans provided slightly greater coronal Cobb correction (~3–4°).

AI strategies consistently used fewer screws and shorter fusion constructs without increasing implant loads.

"Best AI" configurations matched or exceeded surgeons' correction scores in 77% of patients.

Article Title: AI-Derived vs. Surgeon-Performed Instrumentation in Adolescent Idiopathic Scoliosis: A Biomechanical Simulation Analysis

Author names:
Caroline Constant[1,2,3], DMV, MSc, MENG, DACVS-LA, DECVS
(caroline.constant@aofoundation.org);
A. Noelle Larson[1] M.D (Larson.Noelle@mayo.edu),

David W. Polly, Jr. [4], MD (pollydw@umn.edu),

Carl-Eric Aubin[2,3], Ph.D., ScD(h.c.), P.Eng., (Carl-Eric.Aubin@polymtl.ca)

And Minimize Implants Maximize Outcomes Study Group

Institutional affiliation:

[1] Department of Orthopedic Surgery, Mayo Clinic, 200 1st Street Southwest, Rochester, Minnesota, 55905, USA

[2] Polytechnique Montréal, 2500 Chemin de Polytechnique, Montréal, H3T 1J4, Canada

[3] Centre Hospitalier Universitaire Sainte-Justine, 3175 ch. Côte Sainte-Catherine, Montréal H3T 1C5, Canada

[4] Department of Orthopedic Surgery, University of Minnesota, Minneapolis, MN

Corresponding author: C. Constant; Polytechnique Montréal, 2500 Chemin de Polytechnique, Montréal, H3T 1J4, Canada; +41 79 910 69 76; caroline.constant@polymtl.ca

Funding sources: This work was supported by the TransMedTech Institute and its main financial partner, the Canada First Research Excellence Fund, an NSERC-Alliance university-industry grant with Medtronic, Quattrone-Foderaro Grant in AI Innovation in Orthopedic Surgery 2021, Mayo Foundation, the Mayo Clinic, and the AO Foundation. The MIMO Study Group was financially supported by the Orthopedic Research and Education Foundation and Scoliosis Research Society.

Competing Interests: The authors declare no competing interests related to this manuscript. They have no financial, personal, or professional relationships within the past three years, or beyond this period, that could be perceived as influencing the research conducted or the preparation of this manuscript.

Ethical Approval : This study received Institutional Review Board (IRB) approval through the respective committees overseeing the MIMO Clinical Trial (NCT01792609) and research projects at the authors' institutions, and complies with ethical standards in accordance with the Declaration of Helsinki.

Consent: all patients provided informed consent for both clinical trial participation and subsequent image analysis. The authors affirm that patients provided informed consent for publication of the radiographs in relevant figures.

Authorship: All authors confirm their substantial contributions to the conception, design, data acquisition, analysis, and/or software creation for this work; they have critically revised the manuscript, approved the final version for publication, and accept accountability for the work's integrity. All authors consented to submission and obtained required institutional permissions.

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Caroline Constant and the group MIMO. The first draft of the manuscript was written by Caroline Constant and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

5.1.1 Abstract

Background Context: Posterior spinal fusion (PSF) using pedicle screws is the standard approach for managing adolescent idiopathic scoliosis (AIS) with curves >45°, yet significant variability in instrumentation strategies among surgeons persists. Current planning guidelines lack patient-specific detail and 3D optimization, contributing to inconsistent outcomes. Recent advances in artificial intelligence (AI) offer new opportunities for enhancing surgical planning.

Purpose: To biomechanically evaluate the 3D spinal correction and forces experienced by spinal implants of AI-derived instrumentation plans compared to surgeon-performed instrumentations in AIS patients undergoing PSF.

Study design: Biomechanical Modeling and Evaluation of AI-Generated vs. Surgeon-Selected Instrumentation Strategies in Adolescent Idiopathic Scoliosis.

Patient Sample: Thirty-five AIS patients (Lenke 1/2, aged 12–19) treated surgically by PSF, with data sourced from the MIMO Clinical Trial and affiliated hospitals.

Outcome Measures: Main thoracic (MT) Cobb angle, thoracic kyphosis (TK), apical vertebral rotation (AVR), simulated screw pullout forces, number of screws, and levels fused.

Methods: Patient-specific 3D biomechanical models of the spine were reconstructed from preoperative radiographs using validated geometric algorithms and multibody dynamics in MSC Adams. For each patient, nine instrumentation strategies, based on AI predictions (AI-generated) or surgical records (Surgeon-Performed), were simulated using consistent surgical steps. Instrumentation biomechanics and spinal corrections were analyzed across configurations, with statistical comparisons using repeated-measures ANOVA and Friedman tests.

Results: Surgeon-performed instrumentation patterns achieved greater Cobb angle correction ($p < .001$) but required more screws and fused levels. AI-generated plans produced higher thoracic kyphosis ($p = .008$) and matched or exceeded overall correction in 77% of patients when optimized. “Best AI” strategies reduced screw count ($p = .048$) without compromising simulated correction quality or increasing implant loads.

Conclusions/ clinical significance: Biomechanical modeling of AI-derived surgical plans demonstrates their potential to achieve competitive 3D deformity correction with fewer implants.

These findings support the integration of AI and personalized 3D modeling into preoperative planning for AIS.

keywords: spine, posterior spinal fusion, adolescent idiopathic scoliosis, surgery, clinical care, artificial intelligence, machine learning, deep learning, modeling, 3D, multi-body model

5.1.2 Introduction

The current standard approach for the surgical management of adolescent idiopathic scoliosis (AIS) cases with curves greater than 45° to 50° is posterior spinal instrumentation and fusion (PSF) using pedicle screws [1-4]. The primary goal of this surgery is to achieve a balanced and corrected spine in the three anatomical planes, tailored to the unique characteristics of each patient's curve, while minimizing the number of levels fused. Pedicle screws, considered the state-of-the-art instrumentation constructs, play a critical role in facilitating spinal deformity correction [5-7]. However, no standardized approach exists regarding the optimal number and configuration of screws in AIS cases. Significant variability in instrumentation strategies among expert surgeons persists, and there is limited evidence on which approach yields the best long-term outcomes [8-10]. High-density constructs are frequently considered standard of care despite evidence to the contrary [11].

Preoperative planning for PSF surgery in AIS patients is often guided by the Lenke classification system [12, 13]. However, even with these guidelines, suboptimal outcomes have been reported, and surgeons frequently deviate from Lenke recommendations [1, 14-18]. Moreover, key factors such as rod contouring and screw density, which are crucial for postoperative spinal alignment, are not addressed by current treatment algorithms [19-21]. Complications with surgical instrumentation have been reported in up to 4.5% of AIS cases [22-25]. Considerable variability in strategies among experts[9, 10] highlights the need for standardization and improvement in preoperative planning for PSF deformity correction. In recent years, surgical simulation software has gained popularity for spine surgery planning[26, 27], but these tools are rarely used for AIS patients and typically rely on 2D determinants, which are more suited for degenerative conditions than for the complex 3D deformities characteristic of AIS [28]. Additionally, these tools are often limited in their scope, focusing primarily on rod contour and length [28].

Machine learning, a subset of artificial intelligence (AI), offers promising potential in enhancing care for AIS patients by assisting in patient-specific implant planning. Recently, a neural network-based multi-task learning model (NNML) was developed to leverage preoperative clinical and radiographic data to mimic expert surgeons' decision-making in selecting the upper and lower instrumented vertebrae (UIV, LIV), determining screw density, and shaping of rod contour for surgical correction of AIS deformities [29]. The NNML was trained and validated on data from 179 AIS patients with Lenke 1A and 2A curves (45-65°), followed by external validation on 10 patients from different hospitals and surgeons. The model, with 83 input features and multiple hidden layers, demonstrated robust performance: 82-95% and 80-100% accuracy for UIV and LIV predictions, 70-90% accuracy for rod curvatures within ±5°, and RMSE values of 0.2-0.3 for screw density and 3.7-5.6° for rod curvature. These promising results suggest that the model could enhance PSF planning by replicating expert surgeons' decision-making using patient-specific data, despite being a feasibility study. However, the biomechanical performance of the instrumentation parameters predicted by the NNML for individual patients remains unexamined, raising uncertainty about whether these predictions could improve 3D scoliosis correction or potentially introduce new risks.

This study aims to biomechanically evaluate 3D spinal correction and mechanical loads imposed on spinal instrumentation from AI-generated surgical instrumentation plans based on this NNML model, and compare them to the actual instrumentation strategy chosen by surgeons.

5.1.3 Materials and Methods

5.1.3.1 Workflow Overview

In this study, a validated numerical deterministic biomechanical modeling[30, 31] was used to assess the performance of surgical instrumentation strategies for 35 AIS patients (Figure 5.1), from two approaches: 1) AI-derived configurations were obtained using a trained and validated NNML to predict UIV, LIV, screw density, and rod curvatures[29], 2) Surgeon-derived configurations represented the actual surgeries performed. For each patient, a multibody biomechanical model was created using 3D coordinates of key spinal and pelvic landmarks, and flexibility films. Instrumentation simulations kept all parameters consistent, except for UIV, LIV, screw density, and rod curvatures, which varied between AI and surgeon strategies.

5.1.3.2 AIS Patient Sample and Data Collection

After institutional review board approval, 35 Lenke 1-2 AIS cases were selected from the multicenter database of AIS patients from the MIMO Clinical Trial (NCT01792609, n=32 patients)11. or associated hospitals (n=12 patients). Of these from the MIMO Clinical Trial, 15 were randomly chosen from the patients assigned to the "performance testing" subset used during the development of the NNML, and 8 were from the external validation subset[29], resulting in a total of 20 patients for external performance testing (n=8 + 12). Demographic, radiographic, and surgical characteristics were collected from the hospitals and the trial's databases (Table 5.1). Rod curvatures were reconstructed from postoperative radiographs[31], with rod contouring quantified by the differential bending, defined as the difference between convex and concave rods (differential bending). Postoperative spinal alignment was assessed using radiographs taken 3 months post-PSF surgery, or up to 2 years postoperatively if the 3-month radiographs were unavailable. Features were used as inputs for the NNML model and used to describe the study population. Geometric indices were computed from the reconstructed preoperative 3D spinal models to provide detailed 3D anatomical insights.

5.1.3.3 Surgical Instrumentation Configurations

For each patient, nine surgical instrumentation configurations were modeled and biomechanically compared. One represented the actual instrumentation performed by the surgeon, as documented in the surgical report and confirmed by postoperative radiographs for validation purposes, while the other eight were derived from NNML predictions (Figure 5.2).

To generate the NNML-derived configurations, the same inputs used during model development and validation were applied, assuming the surgeon's intended postoperative outcomes were achieved. Rod curvatures and screw density were directly obtained from the NNML predictions. Since the NNML did not specify screw positions, these were determined based on the top five screw patterns most commonly accepted by experienced spine surgeons, as reported in the literature

(Figure 5.3) [9] . We then evaluated the best and second-best model predictions for UIV and LIV, generating four simulations: (1) best UIV and LIV, (2) second-best UIV and LIV, (3) best UIV and second-best LIV, and (4) second-best UIV and best LIV ("AI instrumentation"; Figure 5.2). To optimize simulated outcomes, NNML predictions were repeated with postoperative targets set 25% better than the surgeon's results by reducing postoperative Cobb angles and improving T4-T12 kyphosis. Rod curvatures and screw density were derived, generating another four simulations ("AI-adapted instrumentation"; Figure 5.2).

5.1.3.4 Patient-Specific Geometric and Biomechanical Models of the Spine

First, a geometric model of each patient's spine was created using preoperative posteroanterior (PA) and lateral radiographs. A previously developed 3D reconstruction technique was employed to transform 2D coordinates of 14 anatomical landmarks per vertebra into 3D coordinates using self-calibration and optimization algorithms [32]. The vertebrae and pelvic models were registered using these 3D coordinates with a free-form deformation technique [32]. The accuracy of the reconstructed pedicles and vertebral bodies was reported as 1.6 ± 1.1 mm and 1.2 ± 0.8 mm, respectively [33].

The geometric spine model was then implemented in MSC Adams 2019 software (MSC Software, Santa Ana, California) to create patient-specific multibody biomechanical models. Vertebrae from T1 to L5 were modeled as rigid bodies connected via flexible joints representing intervertebral discs, facet joints, and ligaments. Initial mechanical properties were based on published cadaveric data[34, 35] and adjusted for patient-specific geometry and curve flexibility (Figure 5.4) [36]. This representation defined the nonlinear relationships between displacements and reaction forces/moments for six degrees of freedom between each pair of vertebrae. Boundary conditions were imposed with the first thoracic vertebra modeled as a spherical slider joint to allow free rotation and translation, and the pelvis fixed to prevent movement or rotation.

5.1.3.5 Surgical Instrumentation Simulation

The pedicle screws were modeled as rigid bodies with a spherical joint between the head and shaft to simulate polyaxial screws. The connection between the screw shaft and vertebral bone was modeled as a flexible joint with an interface stiffness represented by a non-linear spring, restricted according to a stiffness matrix from cadaveric data [30, 31]. The rods were modeled using the finite segment method as flexible beams divided into rigid bodies connected by massless joints, with stiffness based on Timoshenko theory, considering only the modulus of elasticity [37]. Per surgical reports, all rods were modeled as 5.5 mm in diameter, with Cobalt-chrome properties (Young's modulus of 220 GPa, yield strength 793 MPa).

5.1.3.6 Posterior Spinal Fusion Surgery Simulation.

Instrumentation biomechanics were studied consistently across all models, except for variations in UIV, LIV, screw density and pattern, and rod curvatures. The surgical procedure and correction maneuvers were simulated as follows:

Pedicle screws were inserted based on 3D coordinates of the vertebral pedicles identified on preoperative radiographs [38]. The first rod, placed on the concave side of the deformity, was positioned and engaged in the screw heads. Rod insertion was achieved by creating displacement constraints between the rod and each screw head until the rod was fully seated. Set-screws were inserted, and cylindrical joints were used to model the rod-screw connection kinematics. Once the rod was in place, the simulated reduction forces were removed. Set screws were inserted, and cylindrical joints were used to model the kinematics of the rod-screw connection. A fixed joint was modeled between the distal screw and the rod to simulate the tightening of the set screw, after which the simulated reduction forces were removed.

En bloc derotation was applied bilaterally on the apical and periapical screws available, using torque-controlled derotation simultaneously on the screws at the three apical levels. Incremental vertebral derotation torque was applied until the postoperative apical vertebral rotation (AVR) angle was reached or until a maximum torque of 5 Nm was achieved. Once derotation was completed, fixed joints were modeled between the screws and the corresponding rod, simulating set screw tightening. Derotation torques were then removed, allowing the spine-rod construct to "spring back" as the surgeon disengages the derotation devices.

Finally, the convex-side rod was placed, aligned with the sagittal plane, and attached using the same procedure as the concave rod, ensuring both rods were fully seated and secured.

The computational spine and instrumentation models were previously validated using pre- and postoperative data from 35 AIS cases, including implant specs, rod contours, osteotomies, radiographs, and flexibility tests. Simulations accurately reproduced Cobb angles within 5° in coronal and sagittal planes. Given this context, the models are considered credible for this study [39-42].

5.1.3.7 Statistical Analysis

All statistical analyses were conducted using IBM SPSS Statistics (Version 29.0.1.0, IBM Corp). Descriptive statistics were provided to characterize the study population. Assumptions for parametric analysis were assessed using the Shapiro-Wilk test for normality and Levene's test for homogeneity of variances. When these assumptions were met, the effects of instrumentation strategy (AI, AI-adapted, and surgeon-performed) on continuous outcomes, including MT Cobb angle, TK, AVR, and average screw pullout force calculated across the apex, the level above, and the level below, were compared using repeated-measures ANOVA. Violations of sphericity were corrected using the Greenhouse-Geisser adjustment, and Bonferroni correction was applied for post hoc pairwise comparisons. Partial eta squared (η^2) was reported as a measure of effect size. For outcomes violating parametric assumptions or involving ordinal variables (number of levels fused, number of screws), non-parametric Friedman tests were used to assess differences across strategies, followed by Wilcoxon signed-rank tests for pairwise comparisons where appropriate.

In addition to global comparisons, the AI strategy yielding the highest weighted 3D correction score (50% coronal, 25% sagittal, 25% axial) was identified for each patient. These "best AI" configurations were directly compared to surgeon-performed instrumentations to assess whether further refining of AI-derived plans could match or exceed surgical outcomes.

The study was powered based on expected differences in postoperative Cobb angle, based on a previous study comparing optimized surgical strategies[39]. With a power of 80% and $\alpha = 0.05$, a minimum of 30 participants was required to detect statistically significant differences in primary outcomes. To ensure robustness and account for potential data variability, 35 patients were included.

5.1.4 Results

5.1.4.1 Patient Characteristics and Surgical Instrumentations

Descriptive statistics for the 35 included patients are summarized in Table 5.2. The mean age at surgery was 16 years (range: 12–19), with the majority of patients presenting Lenke 1A or 1B patterns. Baseline radiographic parameters, including MT Cobb angle, TK, and AVR, are provided in Table 5.2.

Implant configuration comparisons revealed significant differences in the number of fused vertebral levels ($\chi^2(2) = 7.91$, $p = .019$) and the number of pedicle screws used ($\chi^2(2) = 8.00$, $p=.018$) across the instrumentation strategies. Surgeon-performed plans instrumentations employed more screws (19.9 ± 3.6) than both AI (18.1 ± 2.7) and AI-adapted (17.9 ± 2.4) configurations ($p = .031$ and $p=.012$, respectively). Screw density was also greater in the surgery group (1.8 ± 0.3 screws/vertebra) compared to AI (1.65 ± 0.2) and AI-adapted (1.7 ± 0.2) strategies ($p = .008$ and $p = .042$, respectively). The average number of fused vertebrae was higher in the surgical group compared to the AI instrumentations by approximately one level ($p = .023$; Figure 5.5).

5.1.4.2 3D Deformity Correction and Forces Experienced by Spinal Implants

Across the full cohort, the average preoperative MT Cobb angle was 57° ± 7° (range: 40–68°), TK was 31° ± 16° (range: 1–64°), and thoracic AVR was –15° ± 7° (range: –31 to –4°). Instrumentation strategy significantly influenced both MT Cobb correction ($F(1.48, 50.19) = 13.80$, $p < .001$, $\eta^2 = 0.289$) and TK ($F(1.13, 38.52) = 7.37$, $p = .008$, $\eta^2 = 0.178$; Figure 5.6), with no differences in AVR ($p = .063$, $\eta^2 = 0.090$) or average pullout force ($p = .091$, $\eta^2 = 0.070$). Post hoc comparisons revealed that surgeon-performed instrumentations resulted in significantly greater Cobb angle correction than both AI (mean difference = 4.19°, $p < .001$) and AI-adapted (mean difference = 3.36°, $p = .003$) configurations. Similarly, MT correction percentages were higher with the surgeon-performed instrumentations compared to AI (mean difference = 0.074, $p < .001$) and AI-adapted configurations (0.056, $p = .006$; Figure 5.6A). Postoperative thoracic kyphosis was greater in AI-derived plans than with surgeon instrumentation (AI: +4.00°, $p = .015$; AI-adapted: +3.64°, $p = .047$; Figure 5.6B).

5.1.4.3 Comparison of Best AI Strategy vs Surgeon Instrumentation

For each patient, the AI-derived configuration yielding the highest composite correction score was identified. This "best AI strategy" outperformed surgeon instrumentation in 27 of 35 patients (77%) based on this score. When compared directly, individual MT Cobb angle ($F(1,34) = 1.37$, $p = .251$, $\eta^2 = 0.039$), apical vertebral rotation ($F(1,34) = 1.61$, $p = .214$, $\eta^2 = 0.045$), or average pullout force ($F(1,34) = 3.00$, $p = .091$, $\eta^2 = 0.081$) did not significantly differ between "best AI" and surgeon instrumentations. However, thoracic kyphosis was significantly greater in the "best AI"-derived constructs ($F(1,34) = 7.93$, $p = .008$, $\eta^2 = 0.189$). In addition, implant configuration analysis revealed that "best AI" strategies utilized fewer pedicle screws on average (18.8 ± 3.2 vs. 19.9 ± 3.6, $p = .048$), although there were no differences in the number of fused vertebrae ($p = .251$) or screw density ($p = .077$).

5.1.5 Discussion

This study biomechanically evaluated the performance of AI-derived versus surgeon-performed posterior spinal instrumentation strategies for adolescent idiopathic scoliosis using patient-specific numerical simulations. For the first time, the biomechanical performance of the instrumentation parameters predicted by a neural network-based multi-task learning model was assessed on an individual patient basis. The results demonstrated that AI-derived instrumentation strategies achieved comparable overall 3D deformity correction to surgeon-performed plans, although surgeon-derived strategies achieved slightly better coronal curve correction. Importantly, AI-derived plans did not result in increased mechanical loads on the spinal implants. These results suggest that AI-assisted planning tools may support efforts to standardize and personalize surgical strategies for AIS, although further clinical validation is required.

5.1.5.1 3D Deformity Correction and Clinical Significance

Despite established systems such as the Lenke classification [12, 13], substantial variability in AIS surgical planning persists, which may contribute to a 4.5% complication rate [9, 10]. Current classifications do not provide guidance on rod contouring or implant density. Addressing this variability, our study provides new evidence that a NNML model can approximate surgeon decision-making by generating instrumentation strategies that achieve similar sagittal and transverse plane corrections for Lenke 1 and 2 curves without adversely affecting mechanical performance. By providing patient-specific instrumentation plans, AI-assisted approaches could contribute in the future to standardizing surgical planning and potentially reducing variability-related complications. Although surgeon-performed plans achieved slightly better coronal Cobb angle reduction, averaging 3–4° more than AI-derived plans, the clinical significance of this small difference remains uncertain given the multifactorial nature of long-term outcomes.

AIS is now well recognized as a complex 3D deformity involving not only coronal curvature but also vertebral rotation and sagittal misalignment[5, 43, 44], making the ability of AI-derived plans to support improved 3D correction particularly relevant. In this study's cohort, AI plans restored thoracic kyphosis more effectively than surgeon-performed plans, an important finding given the established association between sagittal alignment and better functional outcomes. Suboptimal correction in any plane has been linked to decreased quality of life, increased pain, spinal decompensation, and adjacent segment degeneration [45-49]. Moreover, achieving an appropriate sagittal alignment may reduce the risk of rod fracture and improve pulmonary function [50-52]. Considering that normal thoracic kyphosis ranges around 37 ± 10° in adolescents[53], it is notable that AI-derived plans restored or maintained thoracic kyphosis within this range on average (31° for AI vs. 27° for surgeon-performed plans). While the clinical significance of a 4° mean difference remains unclear, surgeon-performed strategies resulted in a higher incidence of postoperative hypokyphosis, with 25% of patients having kyphosis below 20°, and some below 10°, which was not observed in any AI-derived plan. Nevertheless, the 27° average kyphosis achieved in the surgeon-performed plans was slightly better than values reported in prior PSF-Hooks cohorts (22.6° at 1 year)[54], suggesting above-average correction in this cohort. These results suggest that AI-assisted planning may help address some of the under correction of sagittal alignment in AIS.

While these findings support the potential of AI-assisted planning, they also highlight clinical factors that remain difficult to capture in simulation. Intraoperative elements such as muscle dissection, Ponte osteotomies, rod overbending and unbending, and screw purchase can all influence correction but are not explicitly modeled in this study. Surgeon-to-surgeon variability in surgical technique likely plays a meaningful role in outcomes and remains challenging to replicate computationally. Similarly, ongoing debate around fusion level selection such as the role of intersegmental rotation for LIV selection illustrates how surgical decision-making continues to evolve [55]. While correction maneuvers were standardized to allow fair comparisons, this may have limited the full correction potential of AI-derived plans, particularly in the axial plane, where derotation torques were capped based on clinical outcomes. Incorporating intraoperative variability and procedural nuance into future models could help bridge the gap between virtual planning and real-world execution.

5.1.5.2 Implant Density and Fusion Levels

In the present study, AI-derived instrumentation strategies consistently used fewer screws and exhibited lower implant densities compared to surgeon-performed plans, yet achieved similar 3D deformity correction in most simulations. These findings are consistent with the results of the recent MIMO randomized clinical trial, which demonstrated that low-density constructs provided equivalent coronal correction to high-density constructs in Lenke 1A patients [11]. Prior studies also suggest that increasing implant density beyond a certain threshold does not significantly improve deformity correction and may instead increase surgical costs, operative time, and complication risk [41, 56, 57]. Our results reinforce the potential of AI-assisted planning to favor more efficient constructs without compromising biomechanical performance, supporting a shift away from the historical emphasis on high-density fixation in AIS surgery.

Furthermore, AI-derived strategies led to shorter fusion constructs compared to surgeon-performed plans. While our study did not assess clinical outcomes, shorter fusions are known to reduce hospital stay, lower costs, and most importantly preserve spinal mobility and function [58-62]. A careful balance remains essential to avoid compromising curve correction or stability, but AI-assisted fusion planning could represent a valuable tool for tailoring the extent of fusion on a patient-specific basis.

5.1.5.3 Future Directions

Although we explored an "AI-adapted" strategy by setting postoperative targets 25% better than the original surgical outcomes to drive improved Cobb angle reduction and thoracic kyphosis restoration, this approach did not lead to significantly superior 3D correction compared to the original AI-derived plans. Both AI and AI-adapted strategies exhibited comparable biomechanical behavior and correction profiles relative to surgeon-performed plans, indicating that simply adjusting the target outcomes in the current neural network framework was insufficient to optimize surgical planning further. In an additional analysis, we identified a "best-AI" configuration for each patient based on the highest weighted 3D correction score (50% coronal, 25% sagittal, 25% axial). While these best AI plans often matched or exceeded the surgeons' 3D correction scores, the methodology remained a selection exercise rather than a true optimization. Moving forward, the integration of AI-derived initial configurations with iterative deterministic optimization represents a promising avenue. Multibody models, while offering robust predictions, do not inherently identify optimal solutions. Combining AI predictions with optimization algorithms could generate personalized constructs achieving superior 3D correction with fewer levels fused, fewer implants, and lower mechanical loads. Such a hybrid AI-biomechanical optimization platform could substantially improve patient-specific planning for AIS deformity correction.

5.1.5.4 Limitations

Several limitations should be acknowledged. First, the evaluation relied on deterministic multibody biomechanical simulations rather than clinical outcomes, and thus, the clinical significance of the observed differences in deformity correction and mechanical performance remains to be established. Although our models were validated for AIS correction, they included necessary simplifications, such as rigid vertebrae and soft tissue properties derived from cadaveric data, which may not fully capture in vivo complexities. Screw positions for AI-derived plans were based on common patterns rather than predicted by the neural network, introducing potential variability. The study focused only on Lenke 1 and 2 curves with moderate severity (45–65° Cobb angles), limiting generalizability to more severe deformities or other curve types. Moreover, most patients were enrolled in the MIMO clinical trial, which randomized implant density strategies and may not fully represent individual surgeon preferences. Simulations also reflected only the immediate postoperative biomechanics, without modeling long-term biological processes such as bone fusion, implant fatigue, or adjacent segment degeneration. Lastly, although adequately powered for primary outcomes, some subgroup analyses may have lacked sufficient power to detect slight differences. Future research should include diverse curve types and clinical follow-up to validate AI-assisted surgical planning further.

5.1.6 Conclusion

In summary, this study provided the first biomechanical evaluation of AI-predicted surgical instrumentation strategies for AIS using patient-specific multibody simulations. While surgeon-performed plans achieved slightly superior coronal corrections, AI-derived strategies matched or exceeded overall 3D deformity correction in most cases, restored thoracic kyphosis more effectively, and maintained comparable mechanical safety. Importantly, AI plans achieved these results with fewer screws and shorter constructs, suggesting opportunities for more efficient, patient-specific surgical planning. These findings highlight the potential of AI-assisted preoperative tools to support surgical decision-making, particularly for less experienced surgeons. Building on this foundation, future integration of AI models with biomechanical optimization could further enhance surgical planning and deformity correction in AIS.

5.1.7 References

1. Kim, S.-S., J.-H. Kim, and S.-I. Suk, Effect of Direct Vertebral Rotation on the Uninstrumented Lumbar Curve in Thoracic Adolescent Idiopathic Scoliosis. Asian Spine J, 2017. 11(1): p. 127-137.

2. de Kleuver, M., et al., Optimal surgical care for adolescent idiopathic scoliosis: an international consensus. European Spine Journal, 2014. 23(12): p. 2603-2618.

3. Weinstein, S.L., et al., Adolescent idiopathic scoliosis. The Lancet, 2008. 371(9623): p. 1527-1537.

4. Cheng, J.C., et al., Adolescent idiopathic scoliosis. Nature Reviews Disease Primers, 2015. 1(1): p. 15030.

5. Shen, J., et al., Towards a new 3D classification for adolescent idiopathic scoliosis. Spine Deform, 2020. 8(3): p. 387-396.

6. Udoekwere, U.I., et al., Effect of Lowest Instrumented Vertebra on Trunk Mobility in Patients With Adolescent Idiopathic Scoliosis Undergoing a Posterior Spinal Fusion. Spine Deformity, 2014. 2(4): p. 291-300.

7. Ohashi, M., et al., The Benefits of Sparing Lumbar Motion Segments in Spinal Fusion for Adolescent Idiopathic Scoliosis Are Evident at 10 Years Postoperatively. Spine, 2020. 45(11): p. 755-763.

8. Aubin, C.E., H. Labelle, and O.C. Ciolofan, Variability of spinal instrumentation configurations in adolescent idiopathic scoliosis. Eur Spine J, 2007. 16(1): p. 57-64.

9. Le Naveaux, F., et al., Significant variability in surgeons' preferred correction maneuvers and instrumentation strategies when planning adolescent idiopathic scoliosis surgery. Scoliosis Spinal Disord, 2018. 13: p. 21.

10. Schlager, B., et al., Range of surgical strategies for individual adolescent idiopathic scoliosis cases: evaluation of a multi-centre survey. Spine Deformity, 2023.

11. Larson, A.N., et al., The Effect of Implant Density on Adolescent Idiopathic Scoliosis Fusion: Results of the Minimize Implants Maximize Outcomes Randomized Clinical Trial. J Bone Joint Surg Am, 2023.

12. King, H.A., et al., The selection of fusion levels in thoracic idiopathic scoliosis. J Bone Joint Surg Am, 1983. 65(9): p. 1302-13.

13. Lenke, L.G., et al., Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. J Bone Joint Surg Am, 2001. 83(8): p. 1169-81.

14. Pasha, S. and J.M. Mac-Thiong, Defining criteria for optimal lumbar curve correction following the selective thoracic fusion surgery in Lenke 1 adolescent idiopathic scoliosis: developing a decision tree. Eur J Orthop Surg Traumatol, 2020. 30(3): p. 513-522.

15. Pasha, S., et al., Timing of Changes in Three-Dimensional Spinal Parameters After Selective Thoracic Fusion in Lenke 1 Adolescent Idiopathic Scoliosis: Two-Year Follow-up. Spine Deform, 2017. 5(6): p. 409-415.

16. Takahashi, J., et al., Selective thoracic fusion in adolescent idiopathic scoliosis: factors influencing the selection of the optimal lowest instrumented vertebra. Spine (Phila Pa 1976), 2011. 36(14): p. 1131-41.

17. Ramo, B.A. and B.S. Richards, Repeat Surgical Interventions Following “Definitive” Instrumentation and Fusion for Idiopathic Scoliosis: Five-Year Update on a Previously Published Cohort. Spine, 2012. 37(14): p. 1211-1217.

18. Clements, D.H., et al., Did the Lenke classification change scoliosis treatment? Spine (Phila Pa 1976), 2011. 36(14): p. 1142-5.

19. Clements, D.H., et al., Correlation of scoliosis curve correction with the number and type of fixation anchors. Spine (Phila Pa 1976), 2009. 34(20): p. 2147-50.

20. Lowenstein, J.E., et al., Coronal and sagittal plane correction in adolescent idiopathic scoliosis: a comparison between all pedicle screw versus hybrid thoracic hook lumbar screw constructs. Spine (Phila Pa 1976), 2007. 32(4): p. 448-52.

21. Ofiram, E., et al., Cervical degenerative changes in idiopathic scoliosis patients who underwent long fusion to the sacrum as adults: incidence, severity, and evolution. J Orthop Traumatol, 2009. 10(1): p. 27-30.

22. Coe, J.D., et al., Complications in Spinal Fusion for Adolescent Idiopathic Scoliosis in the New Millennium. A Report of the Scoliosis Research Society Morbidity and Mortality Committee. Spine, 2006. 31(3): p. 345-349.

23. Bagchi, K., et al., Hardware complications in scoliosis surgery. Pediatr Radiol, 2002. 32(7): p. 465-75.

24. Albers, H.W., et al., Comparison of single- and dual-rod techniques for posterior spinal instrumentation in the treatment of adolescent idiopathic scoliosis. Spine (Phila Pa 1976), 2000. 25(15): p. 1944-9.

25. Wattenbarger, J.M., B.S. Richards, and J.A. Herring, A comparison of single-rod instrumentation with double-rod instrumentation in adolescent idiopathic scoliosis. Spine (Phila Pa 1976), 2000. 25(13): p. 1680-8.

26. Hernandez, D., et al., Computer-assisted Orthopaedic Surgery. Orthopaedic Surgery, 2017. 9(2): p. 152-158.

27. Floyd, E., et al., A review of preoperative planning technologies for spinal deformity correction. Seminars in Spine Surgery, 2020. 32(2): p. 100787.

28. Ferrero, E., et al., Preliminary experience with SpineEOS, a new software for 3D planning in AIS surgery. Eur Spine J, 2018. 27(9): p. 2165-2174.

29. Constant, C., Neural network-based multi-task learning to assist planning of posterior spinal fusion surgery for adolescent idiopathic scoliosis. Spine Deformity, 2025.

30. Wang, X., et al., Biomechanical Computational Study of Pedicle Screw Position and Density in Adolescent Idiopathic Scoliosis Instrumentation. Spine (Phila Pa 1976), 2023. 48(20): p. 1436-1445.

31. Wang, X., et al., Concave rod first vs. convex rod first in AIS instrumentation with differential rod contouring: computer modeling and simulations based on ten AIS surgical cases. Spine Deformity, 2023. 11(6): p. 1317-1324.

32. Cheriet, F., et al., A Novel System for the 3-D Reconstruction of the Human Spine and Rib Cage From Biplanar X-Ray Images. IEEE Transactions on Biomedical Engineering, 2007. 54(7): p. 1356-1358.

33. Delorme, S., et al., Assessment of the 3-d reconstruction and high-resolution geometrical modeling of the human skeletal trunk from 2-D radiographic images. IEEE Trans Biomed Eng, 2003. 50(8): p. 989-98.

34. Panjabi, M.M., R.A. Brand Jr, and A.A. White III, Three-dimensional flexibility and stiffness properties of the human thoracic spine. Journal of biomechanics, 1976. 9(4): p. 185-192.

35. Gardner-Morse, M.G. and I.A. Stokes, Structural behavior of human lumbar spinal motion segments. Journal of biomechanics, 2004. 37(2): p. 205-212.

36. Petit, Y., C.É. Aubin, and H. Labelle, Patient-specific mechanical properties of a flexible multi-body model of the scoliotic spine. Medical and Biological Engineering and Computing, 2004. 42(1): p. 55-60.

37. Valembois, R., P. Fisette, and J. Samin, Comparison of various techniques for modelling flexible beams in multibody dynamics. Nonlinear Dynamics, 1997. 12(4): p. 367-397.

38. Lehman Jr, R.A., et al., Straight-forward versus anatomic trajectory technique of thoracic pedicle screw fixation: a biomechanical analysis. Spine, 2003. 28(18): p. 2058-2065.

39. La Barbera, L., et al., In silico patient-specific optimization of correction strategies for thoracic adolescent idiopathic scoliosis. Clinical Biomechanics, 2021. 81: p. 105200.

40. Le Navéaux, F., et al., 3D rod shape changes in adolescent idiopathic scoliosis instrumentation: how much does it impact correction? European Spine Journal, 2017. 26(6): p. 1676-1683.

41. Wang, X., et al., Biomechanical effect of pedicle screw distribution in AIS instrumentation using a segmental translation technique: computer modeling and simulation. Scoliosis Spinal Disord, 2017. 12: p. 13.

42. Le Navéaux, F., et al., How does implant distribution affect 3D correction and bone-screw forces in thoracic adolescent idiopathic scoliosis spinal instrumentation? Clin Biomech (Bristol, Avon), 2016. 39: p. 25-31.

43. Chung, N., et al., Spinal phantom comparability study of Cobb angle measurement of scoliosis using digital radiographic imaging. J Orthop Translat, 2018. 15: p. 81-90.

44. Lenke, L.G., et al., Rationale behind the current state-of-the-art treatment of scoliosis (in the pedicle screw era). Spine (Phila Pa 1976), 2008. 33(10): p. 1051-4.

45. Edgar, M.A. and M.H. Mehta, Long-term follow-up of fused and unfused idiopathic scoliosis. J Bone Joint Surg Br, 1988. 70(5): p. 712-6.

46. Moskowitz, A., et al., Long-term follow-up of scoliosis fusion. J Bone Joint Surg Am, 1980. 62(3): p. 364-76.

47. Youn, M.S., et al., Relationship between cervical sagittal alignment and health-related quality of life in adolescent idiopathic scoliosis. Eur Spine J, 2016. 25(10): p. 3114-3119.

48. Lonner, B.S., et al., Multivariate analysis of factors associated with kyphosis maintenance in adolescent idiopathic scoliosis. Spine (Phila Pa 1976), 2012. 37(15): p. 1297-302.

49. Hwang, S.W., et al., Cervical sagittal plane decompensation after surgery for adolescent idiopathic scoliosis: an effect imparted by postoperative thoracic hypokyphosis. J Neurosurg Spine, 2011. 15(5): p. 491-6.

50. Smith, J.S., et al., Assessment of symptomatic rod fracture after posterior instrumented fusion for adult spinal deformity. Neurosurgery, 2012. 71(4): p. 862-7.

51. Johnston, C.E., et al., Correlation of Preop Curve Severity with Pulmonary Function Tests (PFTs) in AIS: Paper #98. Spine Journal Meeting Abstracts, 2009. 10: p. 122-123.

52. Winter, R.B., W.W. Lovell, and J.H. Moe, Excessive thoracic lordosis and loss of pulmonary function in patients with idiopathic scoliosis. J Bone Joint Surg Am, 1975. 57(7): p. 972-7.

53. Ghandhari, H., et al., Assessment of Normal Sagittal Alignment of the Spine and Pelvis in Children and Adolescents. BioMed Research International, 2013. 2013(1): p. 842624.

54. Sucato, D.J., et al., Restoration of thoracic kyphosis after operative treatment of adolescent idiopathic scoliosis: a multicenter comparison of three surgical approaches. Spine (Phila Pa 1976), 2008. 33(24): p. 2630-6.

55. Seo, S.H., et al., Selection of Optimal Lower Instrumented Vertebra for Adolescent Idiopathic Scoliosis Surgery. Neurospine, 2023. 20(3): p. 799-807.

56. Le Naveaux, F., et al., Implant distribution in surgically instrumented Lenke 1 adolescent idiopathic scoliosis: does it affect curve correction? Spine (Phila Pa 1976), 2015. 40(7): p. 462-8.

57. Delikaris, A., et al., Implant Density at the Apex Is More Important Than Overall Implant Density for 3D Correction in Thoracic Adolescent Idiopathic Scoliosis Using Rod Derotation and En Bloc Vertebral Derotation Technique. Spine (Phila Pa 1976), 2018. 43(11): p. E639-e647.

58. Yoshihara, H., C. Paulino, and D. Yoneoka, Predictors of Increased Hospital Stay in Adolescent Idiopathic Scoliosis Patients Undergoing Posterior Spinal Fusion: Analysis of National Database. Spine Deformity, 2018. 6(3): p. 226-230.

59. Fletcher, N.D., et al., Clinical and economic implications of early discharge following posterior spinal fusion for adolescent idiopathic scoliosis. Journal of Children's Orthopaedics, 2014. 8(3): p. 257-263.

60. Hengartner, A.C., et al., Impact of a quality improvement initiative and monthly multidisciplinary meetings on outcomes after posterior spinal fusion for adolescent idiopathic scoliosis. Spine Deformity, 2024. 12(4): p. 1043-1051.

61. Harris, A.B., et al., Factors associated with extended length of stay in patients undergoing posterior spinal fusion for adolescent idiopathic scoliosis. Spine Deformity, 2020. 8(2): p. 187-193.

62. Engsberg, J.R., et al., Prospective Comparison of Gait and Trunk Range of Motion in Adolescents With Idiopathic Thoracic Scoliosis Undergoing Anterior or Posterior Spinal Fusion. Spine, 2003. 28(17): p. 1993-2000.

5.1.8 Tables

Table 5.1 Summary of Data Collected from Dataset

Demographic and clinical data	- Height - Weight - Age
Radiographic	- Vertebral corner localization - Lenke classification - Cobb angle measurements (PT, MT, TL/L) - Upper and lower vertebrae - Stable vertebrae - Thoracic kyphosis (T2-T12) - Risser sign
Surgical	- Upper instrumented vertebra - Lower instrumented vertebra - Screw density - Rod curvatures (concave side and concave/convex side differential bending) - Type of screws and rods used
Postoperative spinal alignment	- Cobb angle measurements (PT, MT, TL/L) - Thoracic kyphosis (T2-T12)

Table 5.2 Descriptive Statistics of the Patients Included in this Study

Category	Model Development (n=15)	External Performance Testing (n=20)
Height (mean ± SD, range)	159 ± 5cm (146-169)	166 ± 10 cm (154 – 190)
Weight (mean ± SD, range)	54 ± 10 kg (40 – 75)	59 ± 13 kg (40 – 92)
Age (mean ± SD, range)	15.7 ± 1.6 years (12.8 – 19)	15.6 ± 1.7 years (12.8 – 19)
Risser Sign (median, range)	4 (2 – 5)	4 (2 – 5)
Lenke Classification (1/2 breakdown)	1: n= 13 2: n= 2	1: n= 17 2: n= 3
Preoperative Cobb angle (mean ± SD, range)		
Proximal Thoracic (PT)	30 ± 7° (16 – 38)	31 ± 8° (13 – 46)
Main thoracic (MT)	53 ± 7° (40 – 66)	59 ± 6° (48 – 68)
Thoracolumbar / Lumbar (TL/L)	28 ± 9° (14 – 43)	35 ± 5° (27-46)
End Vertebrae of main thoracic curve (median, range)		
Upper end vertebra	T6 (T4-T8)	T6 (T2-T7)
Lower end vertebra	T12 (T11-L2)	T12 (T10-L2)
Apical Vertebrae of main thoracic curve (median, range)	T9 (T8-T11)	T9 (T7-T11)
Preoperative Thoracic Kyphosis T4-T12 (mean ± SD, range)	34 ± 17° (1 – 64)	30 ± 14° (8 – 63°)
Preoperative Apical Vertebra Rotation (mean ± SD, range)	-13 ± 6° (-22 - -4)	-17° ± 7° (-30 - -5°)

5.1.9 Figures

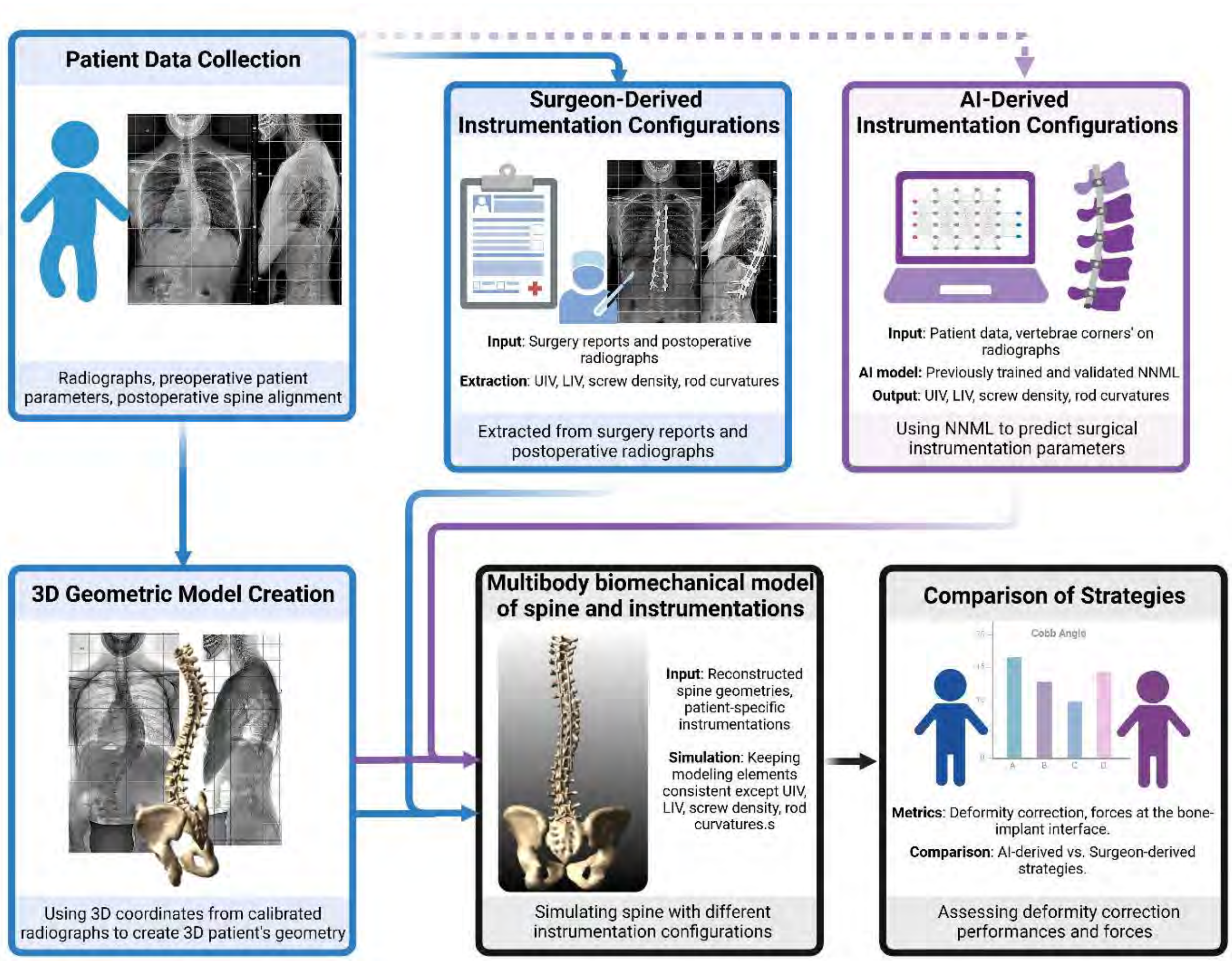


Figure 5.1 Schematic Overview of the Study Workflow Comparing AI-Derived and Surgeon-Derived Instrumentation Configurations for AIS Spinal Surgery

The process included patient data collection, extraction of surgical parameters from both AI model (AI-derived) and surgeon reports along with postoperative radiographs (surgeon-derived), creation of 3D geometric model of the spine and patient-specific multibody biomechanical models, and performance comparison of deformity correction in all anatomical planes and forces at the bone-implant interface.

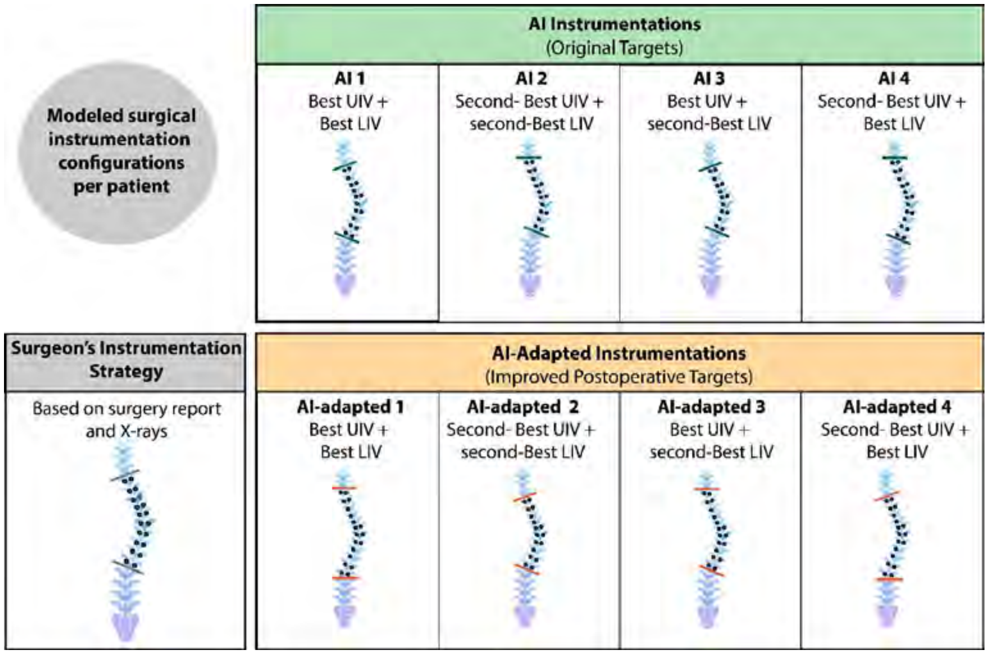


Figure 5.2 Overview of the Nine Instrumentation Configurations Modeled per Patient

Example of surgeon-derived strategy representing the actual instrumentation performed by the surgeon as documented in the surgery report and postoperative radiographs (grey), four AI-derived strategies based on best/second-best predictions of upper and lower instrumented vertebrae (UIV, LIV, green), and four AI-adapted strategies using improved postoperative targets to enhance deformity correction (orange). Screw locations and rod curvatures were generated from the NNML predictions and standardized patterns derived from clinical practice

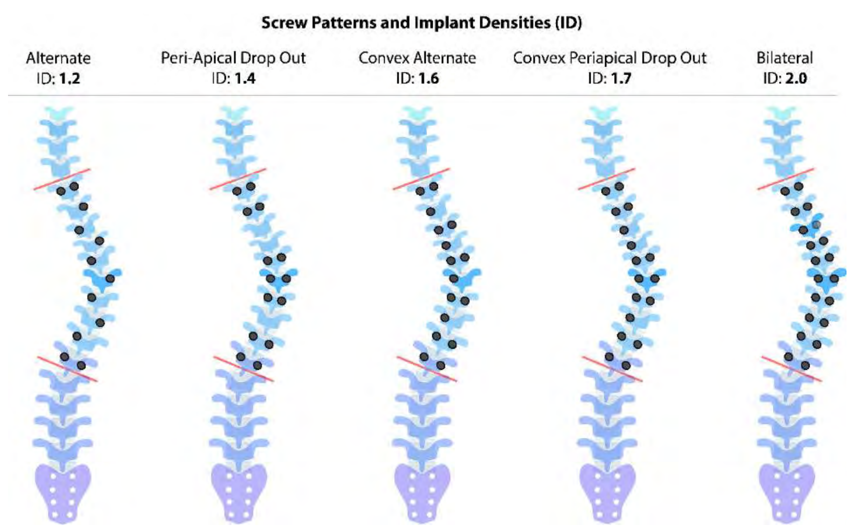


Figure 5.3 Screw Patterns and Corresponding Implant Densities (ID) used for Modeling of Surgical Instrumentation Derived from NNML Predictions (UIV = T4, LIV = L1)

Implant densities (ID) were predicted for each patient, and screw positions were then selected based on these densities, following the top five screw patterns preferred by experienced spine surgeons.

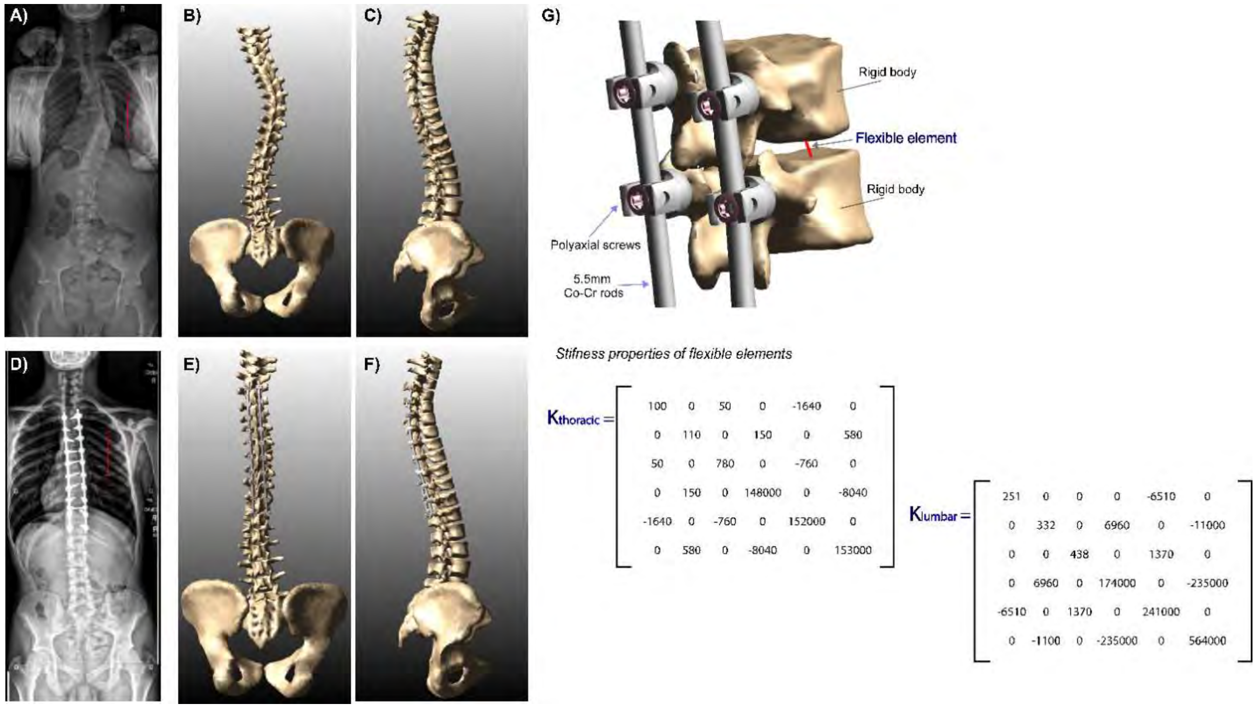


Figure 5.4 Preoperative Radiographs and Reconstructed Biomechanical Model

Coronal preoperative (A) and (D) postoperative standing radiographs of the patient. Reconstructed 3D patient-specific biomechanical models pre- (B-C) and postoperatively (E-F), with illustration of an instrumented 3D thoracic spinal unit (G) with details of initial stiffness matrices for the thoracic (KT1–T7, KT7–T12) and lumbar (KT12–L2, KL2–S1) regions based on Panjabi et al., 1976 (Kthoracic) and Gardner-Morse et al., 2004 (Klumbar).

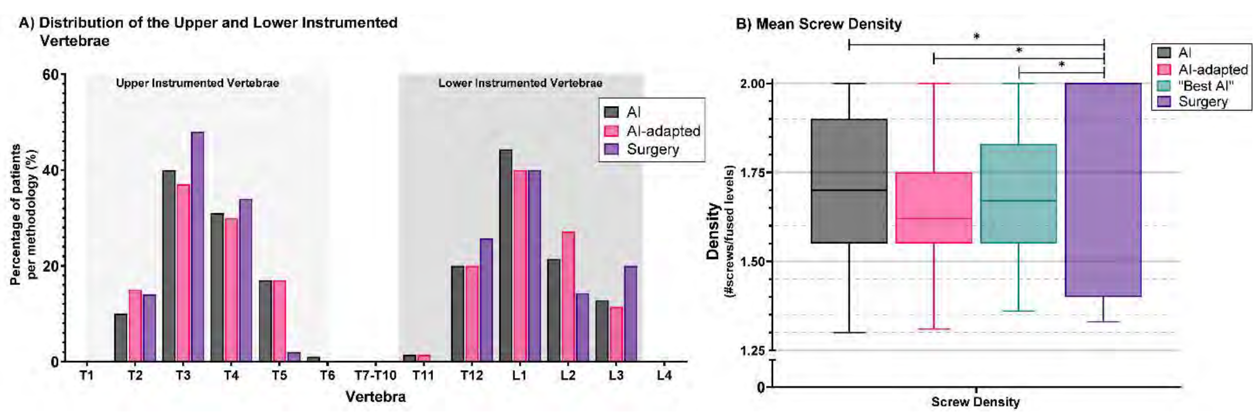


Figure 5.5 Instrumentation Characteristics Across Surgical Strategies

(A) Distribution of UIV (left) and LIV (right) for each strategy, expressed as the percentage of patients receiving instrumentation at each level. (B) Screw density per patient. Data are presented for AI-derived and surgeon-performed instrumentation strategies. In the box plots, the middle line represents the mean, the box boundaries represent the 25th and 75th percentiles, and the whiskers indicate the min and max. * *indicates a statistically significant difference ($p < 0.05$) between groups.*

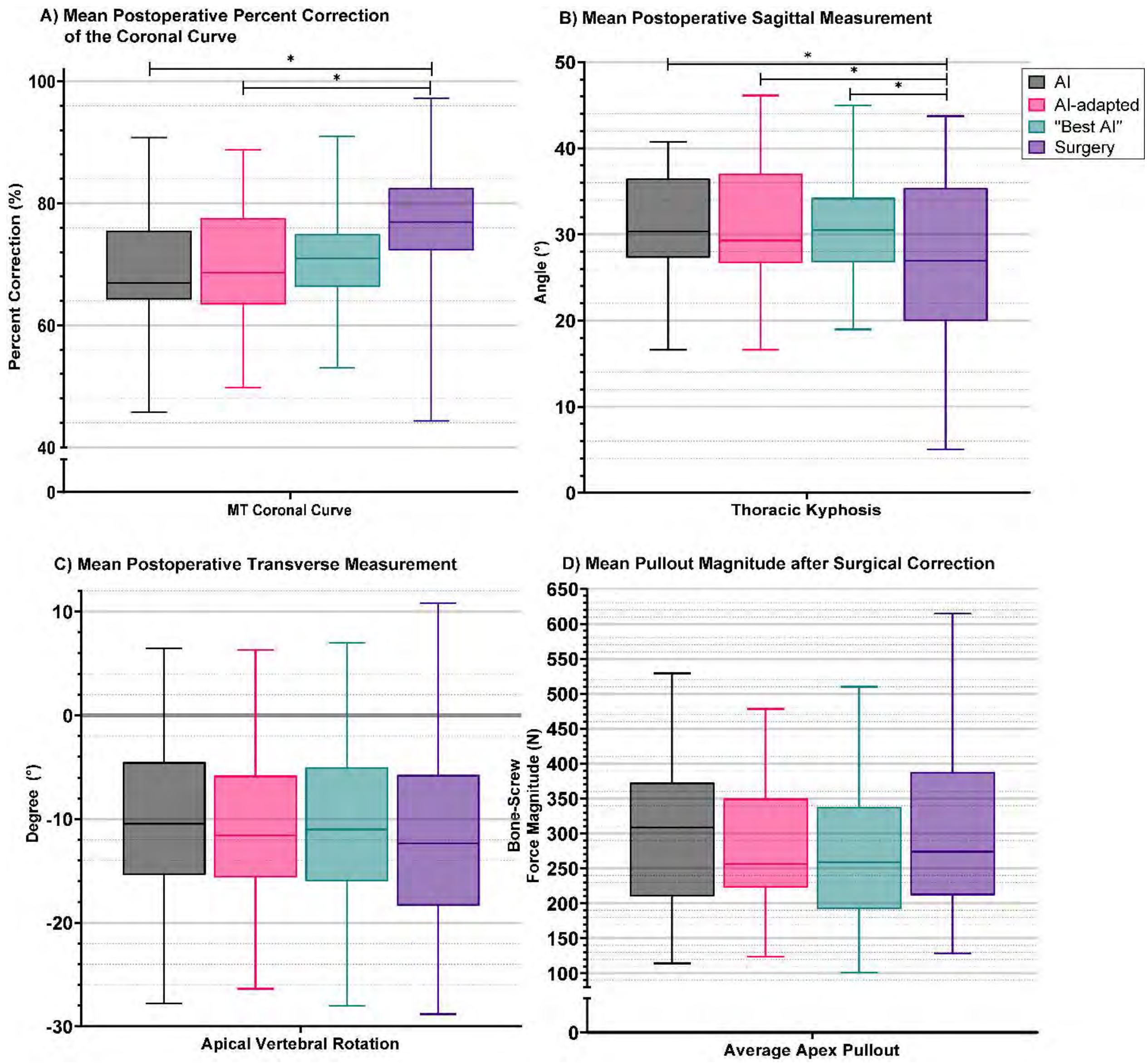


Figure 5.6 Postoperative 3D Deformity Correction and Forces Experienced by Spinal Implants

Data are presented for (A) mean percent correction of the main thoracic (MT) coronal curve , (B) T4-T12 thoracic kyphosis, (C) thoracic apical vertebral rotation (AVR), and (D) mean pullout force experienced by the pedicle screws at the apex and level above and below, stratified by AI-derived and surgeon-performed instrumentation strategies. In the box plots, the middle line represents the mean, the box boundaries represent the 25th and 75th percentiles, and the whiskers indicate the min and max. * *indicates a significant difference between the AI-derived and surgeon-performed instrumentation ($p<0.05$).*

5.2 Complementary Methodological Aspects - Surgical Calibration and Patient-Specific Modeling Workflow

A comprehensive biomechanical modeling framework was implemented to support this article's simulation-based comparison of instrumentation strategies. Details on the development of the personalized multibody spine model, instrumentation components, and simulation of surgical maneuvers are provided below for clarity and conciseness. These complementary methods form the technical foundation underlying the biomechanical evaluation described in the study. Verification, validation, and uncertainty quantification (VVUQ) of this computational modeling framework are presented separately in Chapter 6, Section 6.2, as part of the *Credibility Assessment of the Computational Model Developed as a Medical Device*, within the final framework that includes optimization.

5.2.1 Surgical Calibration Process

To ensure that each surgical simulation accurately reflected the clinical correction achieved during surgery, a manual calibration process was applied to three parameters that are difficult to measure intraoperatively and therefore not documented in surgical reports, yet remain critical to simulation fidelity. These parameters were rod curvature, rod rotation angle, and derotation force, which can be extrapolated from postoperative radiographs and 3D reconstructions. The adjustments were performed manually and iteratively, rather than through automated optimization, until the differences between simulated and clinical postoperative values for MT thoracic Cobb angle, T2–T12 kyphosis, and AVR were considered clinically negligible (≤5°).

5.2.1.1 1. Rod Curvature Adjustment

To determine the pre-instrumentation rod contouring, the curvature of the concave rod was incrementally increased in the simulation. This was done by applying a multiplicative factor to the Y-coordinates of the rod's centerline points, while maintaining the curvilinear abscissa by adjusting the X-coordinates to preserve the total arc length (difference < 1 mm). As a starting point, the simulated concave rod was assigned a curvature approximately 20% greater than that observed postoperatively. This factor was then iteratively adjusted, typically ranging between +10% and

+30%, until the simulated postoperative thoracic kyphosis matched the clinical value within a tolerance of ±5°. Most patients required an increase of around 20% to achieve agreement between simulated and clinical outcomes. See Figure 5.7 for examples of curvature enhancement through multiplicative factors.

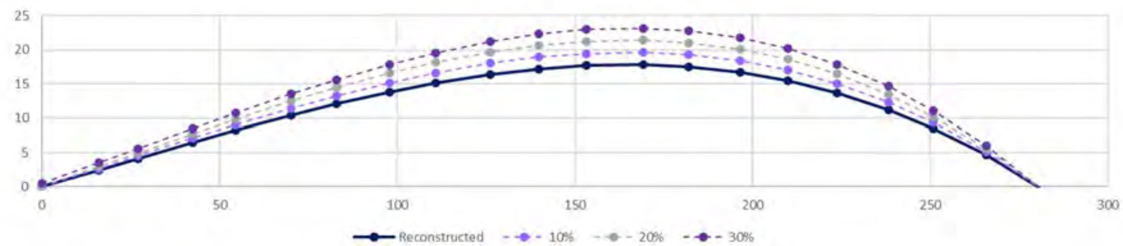


Figure 5.7 Examples of Concave Rod Curvature Enhancement

Rod profiles were incrementally adjusted by applying multiplicative factors of +10%, +20%, and +30% to the Y-coordinates of the reconstructed postoperative rod, simulating intraoperative deformation. The arc length was preserved by modifying the X-coordinates accordingly.

5.2.1.2 Rod Rotation Angle Adjustment

Rod rotation is a corrective maneuver used in PSF surgery for AIS, where a pre-contoured rod is inserted into the concave side of the curve and rotated to achieve a more physiological coronal alignment. Although the theoretical rod rotation angle is often cited as 90°, surgeons commonly rotate the rod beyond this angle intraoperatively to compensate for the elastic "spring-back" effect, ultimately achieving an in-situ rod orientation close to 90° post-correction [205]. This practice aligns with recent biomechanical findings by Pillot et al. (2025), who reported that rod orientations exceeding the sagittal plane (i.e., off-centre by 20–30°) significantly enhance coronal correction without compromising AVR or TK [248]. In our simulation framework, rod rotation angles were manually adjusted between 80° and 100° for most patients. In some cases, lower angles (e.g., 75°) were tested to refine alignment and achieve a clinically acceptable match to postoperative outcomes (±5° Cobb angle). However, values as low as 70° are rarely used in current surgical practice and are not typically recommended due to the risk of under correction. This calibration aimed to align simulated correction with the actual postoperative Cobb angle while preserving the standardization of surgical maneuvers across all cases.

It is important to note that this maneuver does not necessarily reflect current surgical practice. Many surgeons today prefer to insert the contoured rod directly in its sagittal profile and allow each screw to engage the rod individually, avoiding global rod rotation altogether. Nevertheless, for the patients in our cohort, many of whom were drawn from the MIMO database, which includes surgeries performed between February 2013 and August 2017, rod rotation was the correction strategy that most reliably reproduced the clinical outcomes during our surgical simulation. This finding likely reflects common practice at the time among contributing centers. Because all surgical steps were applied consistently and iteratively across both AI-derived and surgeon-performed simulations, the comparative results remain robust and clinically meaningful, despite possible variations in surgical technique that may be encountered today. However, it is essential to acknowledge that this standardization, although necessary to avoid introducing bias, may have underestimated the full potential of the coronal correction offered by AI-derived instrumentation strategies. Surgical maneuvers were iterated to reproduce the surgeon-performed correction, even when alternative steps, such as applying greater rod rotation than was used clinically, might have yielded more favorable outcomes.

5.2.1.3 Apical Vertebral Derotation Calibration

En bloc apical vertebral derotation is a corrective maneuver in which a torque is applied simultaneously to a block of vertebrae, typically the apical vertebra and the two levels above and below, to reduce AVR and restore more physiological alignment. In this study, the maneuver was applied to all available pedicle screws at the apical and peri-apical levels (i.e., apical ±1 level) on both concave and convex sides. The derotation torque was incrementally increased during surgical calibration until the simulated AVR matched the clinical postoperative AVR within a ±5° threshold. An alternative technique frequently used in clinical practice is segmental derotation, in which individual vertebrae are rotated one at a time using instrumentation attached to the pedicle screws. Segmental derotation may be a more current practice, allowing for more localized correction; however, it involves multiple sequential maneuvers, typically one per level, and is thus more complex to simulate. This is primarily due to the greater number of surgical steps that must be individually implemented and controlled in the simulation environment, resulting in increased modeling time, simulation duration, and post-processing workload compared to a single en bloc derotation step.

In preliminary modeling trials, en bloc derotation produced simulated corrections that more closely matched the postoperative AVR in our cohort. This approach was therefore selected as the standardized derotation method across all cases. From a modeling perspective, en bloc derotation also offers advantages in terms of computational efficiency. As a single multi-level maneuver, it is less time-consuming to simulate than segmental derotation, which would require modeling three separate rotational steps per patient. While this difference is negligible at the scale of the current study, it becomes critical in optimization frameworks, such as the one applied in Chapter 6 (Article 4), where over 900 surgical configurations were simulated per patient. Standardizing en bloc derotation thus supports both methodological consistency and scalability for future optimization.

However, as with rod rotation and other standardized maneuvers, this choice may have constrained the full correction potential of the AI-derived strategies. The applied derotation torque was either limited to reproduce the postoperative AVR within ±5°, or capped at a maximum of 5 Nm to reflect physiological constraints. As a result, a more aggressive axial correction that could have been achieved with the AI-derived instrumentation was not pursued to maintain consistency with surgeon-performed cases and avoid introducing bias.

5.2.2 Automated Rod Generation for Simulation Based on AI-Derived Instrumentation

To simulate the AI-derived instrumentation strategies presented in the current chapter (Article 3), a semi-automated pipeline was developed to generate patient-specific rods with limited user input. This process utilized preoperative 3D reconstructions to define the rod length and deflection, and integrated the curvature predicted by the NNML, as described in Chapter 4 (Article 2).

5.2.2.1 Rod Length and Deflection Estimation

Rod length and deflection were derived from the preoperatively reconstructed 3D vertebral geometry for each patient. Vertebral level indices were standardized (e.g., T1 = 3, L5 = 19), and geometry was extracted from a standard coordinate system where Z represents craniocaudal, X anteroposterior, and Y mediolateral directions (Figure 2.8), with vertebral centers of mass used to limit operator variability. Specifically, the rod length was calculated as 1.05 times the Z-axis distance (vertical or craniocaudal direction) between the vertebra one level above the upper instrumented vertebra (UIV–1) and one level below the lower instrumented vertebra (LIV+1). This 5% extension was determined empirically through trial and error to ensure that, once the coronal deformity was corrected during simulation, the rod remained sufficiently long to span from UIV to LIV along the now-straightened spine.

Rod deflection was not measured directly from vertebral offsets, but instead computed from rod length and the tangent angle at the instrumented endpoints. For each rod, the deflection D was obtained using:

$$D = \frac{L}{4} \circ \tan\frac{\theta}{2}$$

where L is the rod length and θ the included predicted angle between tangents (in radians). For the concave rod, deflection was set to 120% of the base value, while the convex rod was assigned the calculated value. This approach reflects the effect of elasto-plastic deformation during rod insertion as described earlier and was selected instead of directly altering the predicted rod curvature, which was retained as an output from the NNML model.

In the NNML model, rod curvature was predicted based on postoperative outcomes. By adjusting deflection instead of curvature, we preserved the predicted rod curvature as a fixed outcome. We avoided introducing variability that would compromise its role as a reference across patients and simulations. This distinction is particularly important for future refinement of the modeling framework, where the ability to back-calculate pre-contoured rod shapes from predicted postoperative geometry will be essential for simulation-based optimization strategies. While this simplified approach does not capture all surgical and patient-specific factors, it provides a practical and reproducible method for generating paired rod profiles in a standardized way.

5.2.2.2 Curve Generation from AI-Derived Rod Curvature

Once rod length and deflection were estimated from the preoperative 3D geometry, rod curvature was obtained directly from the NNML model. These values were combined in a Python-based script to generate smooth, continuous rod geometries. Each rod was modeled as a quadratic curve described by the equation:

$$y = ax^2 + bx + c$$

where the coefficients a, b, and c were calculated to satisfy the following constraints:

- The curve passes through both endpoints (x = 0 and x = rod length)
- The curve reaches a maximum vertical deflection at the midpoint
- The maximum deflection corresponds to the calculated rod deflection, corresponding to the value derived from rod length and tangent angle.

A custom Python script was written to automate this process. The script reads rod parameters (length, deflection, curvature, and identifier) from an Excel sheet, fits a quadratic curve using symbolic equation solving (*sympy*), and generates evenly spaced coordinate points along the curve. These points represent the rod shape in 2D and are exported as .xlsx files, ready to be integrated into the simulation environment. Each rod is created in a dedicated output file, with separate handling for concave and convex rods. The script also includes functionality to optionally import the generated rod data directly into the patient-specific simulation file using an auxiliary script, further streamlining the workflow. This process allowed the rapid and consistent generation of AI-informed rod geometries that were plausible according to the NNML model and tailored to each patient's anatomy. It also enabled the high-throughput simulation required for the parametric comparisons and optimization work presented in Chapter 6 (Article 4).

CHAPTER 6 PATIENT-SPECIFIC OPTIMIZATION

6.1 ARTICLE 4: Optimizing AIS Surgery with a Digital Twin Framework Integrating AI and Personalized Biomechanical Modeling

To address Sub-objective 3.1 (SO3.1), this study developed an original optimization framework combining AI-based planning with patient-specific biomechanical simulation to guide preoperative planning in AIS surgery. Building on the NNML predictive model and MB simulation workflow introduced in Articles 2 and 3, the approach systematically evaluated over 900 surgical strategies per patient using a custom optimization function that balanced 3D correction targets with clinical constraints. This work represents the first attempt to apply large-scale, patient-specific optimization of PSF constructs by integrating artificial intelligence with deterministic modeling. It shows that such a strategy can identify constructs that outperform 3D correction of both standard surgical and AI-derived plans, while minimizing the number of vertebral levels fused and screw used.

The resulting article, titled "*Optimizing AIS Surgery with a Digital Twin Framework Integrating AI and Personalized Biomechanical Modeling*" was submitted to The Spine Journal in September 2025. The first author contributed approximately 80% to the design, implementation, data analysis, and manuscript preparation.

Article 4: Constant, C., Larson, A.N., Polly, D.W. et al. **Optimizing AIS Surgery with a Digital Twin Framework Integrating AI and Personalized Biomechanical Modeling**

Highlights

First patient-specific digital twin for AIS integrating AI predictions, biomechanical simulation, and multi-objective optimization. Applied to 20 patients with 19,440 simulated configurations.

Optimized plans achieved greater 3D correction (~+3° MT Cobb, +2.8° AVR) while maintaining thoracic kyphosis.

Optimization plan had significantly reduced implant burden without increasing screw pullout forces : ~4 fewer screws and 2 fewer fused levels.

Optimization plans favored low density screw patterns (convex-alternate, periapical dropout) and more distal UIV (often T5).

Article Title: Optimizing AIS Surgery with a Digital Twin Framework Integrating AI and Personalized Biomechanical Modeling

Author names:
Caroline Constant[1,2,3], DMV, MSc, MENG, DACVS-LA, DECVS (caroline.constant@uzh.ch);
A. Noelle Larson[1] M.D (Larson.Noelle@mayo.edu),
David W. Polly, Jr. [4], MD (pollydw@umn.edu),
Carl-Eric Aubin[2,3], Ph.D., ScD(h.c.), P.Eng., (Carl-Eric.Aubin@polymtl.ca)
And Minimize Implants Maximize Outcomes Study Group

Institutional affiliation:

[1] Department of Orthopedic Surgery, Mayo Clinic, 200 1st Street Southwest, Rochester, Minnesota, 55905, USA

[2] Polytechnique Montréal, 2500 Chemin de Polytechnique, Montréal, H3T 1J4, Canada

[3] Centre Hospitalier Universitaire Sainte-Justine, 3175 ch. Côte Sainte-Catherine, Montréal H3T 1C5, Canada

[4] Department of Orthopedic Surgery, University of Minnesota, Minneapolis, MN

Corresponding author: C. Constant; Department of Mechanical Engineering, Polytechnique Montreal, P.O. Box 6079, Downtown Station, Montreal, QC H3C 3A7, Canada; +41 79 910 69 76; caroline.constant@polymtl.ca

Funding sources: This work was supported by the TransMedTech Institute and its main financial partner, the Canada First Research Excellence Fund, NSERC (Alliance grant with Medtronic), Quattrone-Foderaro Grant in AI Innovation in Orthopedic Surgery 2021, Mayo Foundation, and the Mayo Clinic. The MIMO Study Group was financially supported by the Orthopedic Research and Education Foundation and Scoliosis Research Society.

Competing Interests: The authors declare no competing interests related to this manuscript. They have no financial, personal, or professional relationships within the past three years, or beyond this period, that could be perceived as influencing the research conducted or the preparation of this manuscript.

Ethical Approval : This study received Institutional Review Board (IRB) approval through the respective committees overseeing the MIMO Clinical Trial (NCT01792609) and research projects at the authors' institutions, and complies with ethical standards in accordance with the Declaration of Helsinki.

Consent: all patients provided informed consent for both clinical trial participation and subsequent image analysis. The authors affirm that patients provided informed consent for publication of the radiographs in relevant figures.

Authorship: All authors confirm their substantial contributions to the conception, design, data acquisition, analysis, and/or software creation for this work; they have critically revised the manuscript, approved the final version for publication, and accept accountability for the work's integrity. All authors consented to submission and obtained required institutional permissions.

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Caroline Constant and the group MIMO. The first draft of the manuscript was written by Caroline Constant and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

6.1.1 Abstract

Background Context: Posterior spinal fusion (PSF) with pedicle screw instrumentation is the primary treatment for adolescent idiopathic scoliosis (AIS) for curves exceeding 45°. However, the biomechanical implications of different instrumentation strategies remain poorly understood. Current surgical planning relies on radiographic-based and empirical guidelines that lack integration of patient-specific biomechanics and 3D deformity characteristics. These limitations may affect the consistency and quality of surgical correction. Advances in artificial intelligence (AI) and biomechanical modeling now offer promising avenues to objectively evaluate and optimize instrumentation strategies on a case-by-case basis.

Purpose: To develop and validate a patient-specific optimization framework that integrates AI-predicted instrumentation strategies and deterministic biomechanical simulation to plan instrumentation configurations maximizing 3D deformity correction and biomechanical outcomes while minimizing instrumentation-related mechanical stress.

Study design: AI-derived instrumentation combined with patient-specific biomechanical modeling and optimization vs. Surgeon-Selected Instrumentation Strategies in AIS.

Patient Sample: Twenty AIS patients (Lenke 1/2, aged 12–19) treated surgically by PSF, with data sourced from the MIMO Clinical Trial and affiliated hospitals.

Outcome Measures: Simulated main thoracic (MT) Cobb angle, thoracic kyphosis (TK), apical vertebral rotation (AVR), screw pullout, number of screws, and levels fused, as well as related measurements derived from actual post-op radiographs.

Methods: A patient-specific digital twin of the spine was generated using validated 3D reconstruction from preoperative biplanar radiographs and a multibody dynamic model implemented in MSC Adams. AI-derived instrumentation parameters (upper and lower instrumented vertebrae (UIV, LIV), number of screws, rod curvature) were simulated within algorithmically defined ranges, generating 972 configurations per patient. In total, 19,440 simulations were analyzed for the 20 cases. For each case, two instrumentation strategies were evaluated: one derived from a multi-objective optimization algorithm integrating 3D spinal correction, minimal screw pullout forces, and preservation of spinal mobility (unfused vertebrae), and the other reflecting a surgeon-informed design. Both strategies were simulated biomechanically and compared using repeated-measures two-way ANOVA.

Results: Across all patients, the optimized configurations achieved similar or improved 3D correction compared to surgeon-performed instrumentation. The highest mean coronal correction improvement was 3° greater than that of surgeon-performed constructs (95% CI [0.58, 5.81], $p = .019$). Optimized plans also achieved greater axial derotation (mean AVR improvement: 2.8°, 95% CI [0.30, 5.67], $p = .05$) while also restoring thoracic kyphosis within acceptable clinical ranges, similar to surgeon-performed plans (29.2° ± 6.4° vs. 27.4° ± 8.9°, $p = .372$).

Optimized strategies used significantly fewer pedicle screws (mean reduction of 4 screws per construct and fewer fused levels (median reduction of 2 levels) than surgeon-performed plans ($p < 0.001$). All configurations respected peri-apical screw pullout force thresholds, suggesting biomechanical feasibility.

Conclusions/ clinical significance:

Integrating AI-derived instrumentation with a patient-specific digital twin framework, combining deterministic biomechanical modeling and optimization, may enhance surgical planning for AIS by improving 3D correction, refining level selection and screw patterns, and maintaining implant loads within functionally acceptable ranges. This study supports the clinical potential of digital twin–based planning tools for spinal deformity surgery.

keywords: adolescent idiopathic scoliosis, posterior spinal fusion, artificial intelligence, biomechanical modeling, digital twin, multi-body model, optimization

6.1.2 Introduction

Adolescent idiopathic scoliosis (AIS) is a complex three-dimensional (3D) spinal deformity affecting up to 5% of adolescents aged 10 to 18 years[1, 2]. For curves exceeding 45°, posterior spinal fusion (PSF) with pedicle screw instrumentation remains the primary treatment[3-5]. However, despite widespread use of the Lenke classification to determine fusion levels, substantial variability persists in surgical planning, particularly in the number and placement of screws, the choice of fusion levels, and rod contouring[6-8]. This variability combined with high-quality evidence on long-term outcomes, may contribute to suboptimal correction, inconsistent results, and avoidable complications[3, 9-13].

While the Lenke classification offers a structured framework for level selection, adherence to the indicated levels does not guarantee optimal results. In addition, surgeons frequently deviate from recommended levels [14, 15]. Moreover, this classification does not provide detailed guidance on other key aspects of surgical design, such as screw patterns, rod properties, or construct strategy. In recent years, digital tools for preoperative planning have emerged, but they largely rely on user experience and manual interpretation[16, 17]. Artificial intelligence (AI) and machine learning now offer new opportunities for data-driven, personalized planning by learning from large clinical datasets[18]. Still, most AI-based systems rely on 2D radiographs and often overlook key biomechanical factors, such as load sharing, implant stress, and construct stability[16]. They also focus heavily on rod contouring, with limited insight into how screw configuration or rod properties affect outcomes.

To address these gaps, we previously developed a neural network model using multitask learning (NNML) to predict key surgical instrumentation parameters from preoperative clinical and radiographic data[19]. The model accurately predicted upper and lower instrumented vertebrae (UIV, LIV) and rod curvatures, with up to 95% accuracy for vertebral level selection. Subsequent work applied validated deterministic patient-specific biomechanical models to compare these AI-derived strategies with those chosen by surgeons[20]. While surgeon-selected constructs yielded greater coronal correction, AI-derived configurations showed superior sagittal alignment and, in 77% of cases, at least one AI configuration matched or exceeded overall 3D correction, often with fewer screws and shorter fusion spans[20]. However, attempts to refine these AI predictions through modified targets remained limited to static selection among pre-generated options, without true optimization or incorporation of biomechanical considerations. These limitations highlight the need for integrated, patient-specific digital twins; models that not only predict but also optimize surgical strategies by combining AI with biomechanical simulation.

In this study, we developed and tested a hybrid framework combining AI-based prediction, deterministic biomechanical simulation, and algorithmic optimization. We hypothesized that this integrated approach would yield personalized instrumentation strategies that outperform conventional constructs in terms of 3D correction, implant efficiency, and biomechanical safety. By systematically varying surgical parameters around the AI-predicted configuration (UIV, LIV, screw number, rod curvature) and exploiting deterministic biomechanical modeling, we applied multi-objective optimization to identify patient-specific constructs that balance correction, fusion length, and screw forces. These optimized configurations were biomechanically evaluated and compared to surgeon-performed strategies, demonstrating the added value of digital twin-based planning in AIS surgery.

6.1.3 Materials and Methods

6.1.3.1 Overview of the Digital Twin-Based Optimization Framework

The following outlines the overall structure of the patient-specific optimization workflow (Figure 6.1), which integrates AI-based prediction, 3D biomechanical modeling, and algorithmic optimization to design tailored instrumentation strategies for AIS patients. To evaluate the applicability of the approach, a retrospective test group of 20 AIS patients was selected, for whom preoperative radiographs, flexibility films, and surgical reports were collected to support model construction and simulation, as detailed in the section below.

The workflow consisted of three main stages:

1. **AI-based prediction of initial instrumentation:** A previously developed NNML model was used to predict key surgical parameters (UIV, LIV, screw number, rod curvature) based on preoperative clinical and radiographic data.

2. **Generation and biomechanical simulation of alternative configurations:** Around the AI-predicted construct, 972 variations were generated by modifying UIV, LIV, screw pattern, and rod contour. Each configuration was simulated using a validated deterministic, patient-specific biomechanical model to estimate postoperative 3D correction and screw pullout forces [21-24].

3. **Multi-objective optimization:** An optimization algorithm was applied to identify the configuration that best balanced spinal correction, motion preservation, and biomechanical safety.

To assess the clinical relevance of the proposed workflow, the optimized configurations were biomechanically compared to the actual surgeon-performed instrumentation. This comparative analysis served to quantify the potential benefits of the digital twin-based framework in improving construct efficiency and correction quality. The following sections detail each stage of the workflow, including patient data preparation, AI-based prediction, biomechanical simulation, optimization strategy, and comparative evaluation.

6.1.3.2 AIS Patient Sample and Data Collection

Following institutional review board approval, 20 AIS patients with Lenke type 1 or 2 curves were retrospectively selected from the MIMO Clinical Trial (NCT01792609) or affiliated clinical sites[25]. Demographic, radiographic, and surgical data were obtained from institutional and trial databases (Table 6.1). Rod curvatures were reconstructed from postoperative radiographs [26]. and quantified using differential bending, defined as the angular difference between convex and concave rods. Postoperative spinal alignment was assessed using radiographs taken approximately 3 months after surgery, or the closest available imaging within the first two operative years. In addition, pre- and postoperative 3D spinal reconstructions were used to calculate geometric indices describing each patient's spinal deformity and surgical correction.

6.1.3.3 AI-Based Prediction of Initial Instrumentation using a NNML Model (Stage 1)

A previously developed and validated NNML model by Constant et al. was used to predict key surgical instrumentation parameters: UIV, LIV, rod curvature, and screw density[19]. The model was trained on 179 AIS patients using 83 preoperative clinical and radiographic features, with multitask outputs optimized for classification and regression. These input features included vertebral corner coordinates extracted from biplanar radiographs, demographic information (age, sex, height, weight), curve classification, Risser sign, and coronal and sagittal alignment parameters such as Cobb angle, TK, lumbar lordosis, and AVR.

For each patient in the current study, the same input features used during model training were collected preoperatively to generate NNML-based predictions of UIV, LIV, rod curvature, and screw density. Because the NNML model does not define individual screw positions, five common screw patterns corresponding to predicted densities were predefined based on literature and expert input (Figure 6.2)[7]. These AI-derived constructs were used as reference configurations for subsequent biomechanical simulation and optimization.

6.1.3.4 Biomechanical Modeling and Simulation of Alternative Configurations (Stage 2)

A geometric model of each patient's spine was reconstructed using preoperative posteroanterior (PA) and lateral radiographs. Three-dimensional reconstructions were performed using a previously validated method based on self-calibration and optimization algorithms, which converts the 2D positions of 14 anatomical landmarks per vertebra into corresponding 3D coordinates[27]. Using these coordinates, vertebrae and pelvic structures models were registered through a free-form deformation approach[27]. Reported reconstruction accuracy for vertebral bodies and pedicles is 1.2 ± 0.8 mm and 1.6 ± 1.1 mm, respectively[28]. The resulting 3D geometric spine model generated patient-specific multibody biomechanical models in MSC Adams 2019 (MSC Software, Santa Ana, CA). Vertebrae from T1 to L5 were modeled as rigid segments linked by flexible joints representing intervertebral discs, ligaments, and facet articulations. Mechanical behavior between each spinal segment was defined using nonlinear relationships between displacements and reaction forces/moments across six degrees of freedom, based on previously published cadaveric data and adjusted based on individual spinal geometry and flexibility[29-31]. The pelvis was fixed in space to prevent translation or rotation, while T1 was constrained using a spherical slider joint to permit free rotation and translation.

Polyaxial pedicle screws were modeled as rigid bodies with a spherical joint at the screw head to replicate polyaxial functionality. The interface between the screw shaft and vertebral body was represented by a flexible joint incorporating a nonlinear spring element, with stiffness properties defined via a cadaver-derived stiffness matrix. Spinal rods were modeled using a discrete flexible beam approach, dividing each rod into small consecutive segments each carrying two local reference frames shared with its neighboring segments at the connecting points. Timoshenko beam elements were respectively defined between the adjacent local reference frames corresponding to the geometry of each segment and to the elastic modulus and Poisson's ratio of the rod. All rods were assigned a 5.5 mm diameter and Cobalt-Chrome material properties (Young's modulus: 220 GPa; yield strength: 793 MPa), consistent with surgical reports.

Posterior spinal fusion simulations were conducted using a standardized protocol across all patients, with variations introduced only in UIV, LIV, screw number and pattern, and rod curvature. Surgical correction was modeled through a translation maneuver, consistent with clinical practice. This simulation framework was previously validated in 35 AIS cases using detailed pre- and postoperative data; in a similar context of application, simulated corrections reproduced coronal and sagittal Cobb angles within 5°, supporting the adequacy and validity of the modeling tool for biomechanical evaluation[21-24].

Pedicle screws were positioned using 3D coordinates extracted from preoperative radiographs[32]. The concave rod was inserted first by progressively constraining it to each screw head until fully seated. Set screws were then applied, with cylindrical joints used to simulate rod-screw interactions. A fixed joint at the distal screw represented final set screw tightening, after which simulated reduction forces were removed.

Bilateral en bloc derotation was performed by applying torque-controlled rotation simultaneously to apical and periapical screws over three vertebral levels. Torque was incrementally increased until the target apical vertebral rotation (AVR) was achieved. Fixed joints were then applied between the screws and rod, simulating set screw tightening, and the torques were released to allow the construct to elastically recoil as occurs intraoperatively. The convex rod was then inserted using the same method, and all remaining screws were secured.

Instrumentation parameters were systematically varied around the AI-predicted configuration to generate a range of plausible surgical scenarios while remaining within clinically acceptable margins. Variation limits were selected based on expert consensus and typical ranges observed in AIS surgical planning: ±1 level for UIV and LIV (reflecting reasonable alternatives in fusion extent), screw patterns corresponding to ±0.2 variation in implant density (to span commonly applied construct densities), and ±10° rod curvature (representing variation of rod contouring achievable intraoperatively). This approach resulted in 972 surgical instrumentation configurations per patient, all biomechanically simulated and compared, for a total of 19,440 simulations across the 20 patients.

Additionally, the actual surgeon-performed instrumentation was biomechanically simulated for each patient based on surgical reports and postoperative radiographs, then compared to the optimized AI-derived configurations using standardized biomechanical outputs, including simulated Cobb angle, TK, AVR, screw pullout forces, number of screws, and fusion length.

To ensure validity, the biomechanical model of the actual surgeon-performed instrumentation was verified by comparing simulated postoperative alignment to 3D reconstructions obtained from postoperative radiographs. Simulated values within 5° of the reconstructed postoperative values were considered acceptable.

6.1.3.5 Multi-Objective Optimization (Stage 3)

An objective function (ϕ) was defined to evaluate the performance of each simulated configuration based on predicted postoperative 3D spinal correction and preserved mobility (number of unfused vertebrae). It included four weighted terms reflecting key surgical objectives, adapted from prior formulations [21, 33, 34]:

$$\phi = W_1 \cdot \left(\frac{Cobb}{Cobb_0}\right)^2 + W_2 \cdot \left(\frac{TK - TK_n}{TK_0 - TK_n}\right)^2 + W_3 \cdot \left(\frac{AVR}{AVR_0}\right)^2 + W_4 \cdot \left(\frac{N_{fused}}{N_0}\right)^2$$

Cobb, TK, and AVR refer to postoperative main thoracic Cobb angle, thoracic kyphosis, and apical vertebral rotation, each normalized to its preoperative value (subscript “0”). TK_n denotes the normative kyphotic target based on a normo-kyphotic range[35]: $TK_n = 20°$ if $TK < 20°$, $TK_n = TK$ if $20° \leq TK \leq 40°$, and $TK_n = 40°$ if $TK > 40°$. Mobility preservation was assessed via the number of fused vertebrae (N_{fused}), normalized to the total number of thoracolumbar vertebrae available ($N_0 = 17$). To ensure implant safety, a hard constraint was implemented on the maximum pullout force experienced by any pedicle screw in each simulation. This constraint was stratified by spinal region, with screws placed from T1–T8 limited to 425 N and screws from T9–T12 limited to 645 N. These thresholds represent approximately 80% of experimentally reported average pullout strengths for each region [36] and can be modified by the user based on clinical considerations. Configurations exceeding these thresholds were excluded from optimization. Minimizing ϕ

identified the configuration that best balanced correction in all three anatomical planes, mobility preservation, and implant safety for each patient. A lower ϕ score indicated greater correction with fewer fused segments and reduced bone-screw pullout.

To refine selection among the best-performing configurations, a second-stage optimization was conducted on the 20 instrumentation scenarios with the lowest ϕ scores. These same variables used in the optimization algorithm were normalized using min–max scaling applied only within the top 20 group to preserve local variation. A secondary composite score was calculated by applying the same weighting factors (W_1–W_4) as in the primary optimization, ensuring consistent prioritization of coronal correction, sagittal alignment, derotation, and motion preservation. This approach allowed for a clinically guided refinement, emphasizing scenarios with the lowest Cobb angle, TK values closest to 40°, lowest AVR, and minimal fusion length, among otherwise equivalently high-performing configurations.

To explore different surgical priorities in the multi-objective optimization process, three weighting scenarios were tested for the four terms in the objective function (W_1–W_4). Scenario 1 applied equal weighting across all components, assigning 25% to each of the four objectives. This scenario served as a neutral reference, assuming no prioritization of any particular surgical goal. Scenario 2 represented a correction-focused strategy in which 80% of the total weight was distributed across the three anatomical correction planes, emphasizing the coronal plane. Specifically, W_1 (coronal) was set at 50%, W_2 (sagittal) at 20%, and W_3 (transverse) at 10%, while the remaining 20% for mobility preservation (W_4). This scenario was designed to reflect a hypothetical prioritization of 3D correction, especially in the coronal plane, and was arbitrarily defined for this study. Scenario 3 incorporated weighting based on expert consensus, using averaged priorities assigned by 11 experienced spinal deformity surgeons as reported in previous survey-based studies[37, 38]. In these surveys, surgeons were asked to assign relative importance to the various correction objectives when planning optimal instrumentation for AIS. The raw averaged importance values resulting in the following distribution: W_1 = 37% (coronal), W_2 = 28% (sagittal), W_3 = 19% (transverse), W_4 = 16% (mobility). This third scenario aimed to approximate real-world surgical decision-making by integrating clinical judgment into the optimization framework.

6.1.3.6 Statistical Analyses

All statistical analyses were conducted using IBM SPSS Statistics (Version 29.0.1.0, IBM Corp). Descriptive statistics (means, standard deviations, and ranges) were used to characterize the study population and summarize radiographic and simulation-derived parameters.

To evaluate the clinical relevance of the optimization framework, each patient's instrumentation configuration with the lowest objective function value (ϕ) was identified under each weighting scenario. These optimized configurations were compared to the actual surgeon-performed instrumentation to assess potential improvements in 3D spinal correction, bone-screw pullout force, and number of fused vertebrae.

A repeated-measures multivariate analysis of variance (MANOVA) was performed to assess the effect of instrumentation strategy (S1, S2, S3, surgeon-performed) on five dependent variables: MT Cobb, TK, AVR, average screw pullout force calculated across the apex ± 1 level, and number of fused vertebrae. Assumptions of multivariate normality and homogeneity of variance-covariance matrices were assessed using the Shapiro–Wilk test and Box's M test, respectively. When a significant multivariate effect was detected, univariate repeated-measures ANOVAs were performed for each outcome variable. Mauchly's test assessed sphericity, and Greenhouse–Geisser corrections applied when necessary. Pairwise comparisons between strategies were conducted using least significant difference (LSD) tests for multiple comparisons correction to maintain sensitivity given the limited number of planned comparisons. Effect sizes were reported as partial eta squared (η^2), and confidence intervals (CIs) for mean differences were provided to estimate effect precision. If test assumptions were violated, non-parametric alternatives (Friedman test with post hoc Wilcoxon signed-rank tests) was used for the affected outcomes. All statistical tests were two-tailed, and significance set at $p < 0.05$.

6.1.4 Results

6.1.4.1 Patient Characteristics

Descriptive statistics for the 20 included patients are summarized in Table 6.2. The mean age at surgery was 16 years (range: 13–19), with the most cases representing Lenke 1A or 1B patterns.

Across all patients, the preoperative MT Cobb angle averaged 59° ± 6°, with thoracic AVR of –15° ± 7° and T4–T12 kyphosis of 30° ±14°.

6.1.4.2 Implant Configuration Differences

Implant configuration comparisons revealed significant differences across instrumentation strategies in both the number of fused vertebral levels ($\chi^2(3) = 33.894$, $p < .001$) and the number of pedicle screws ($F(1.778, 33.776) = 31.39$, $p < .001$, $\eta^2 = 0.623$; Figure 6.3). Surgeon-performed plans involved the longest constructs (median = 11 levels, range = 9–13; mean rank = 3.55), whereas the equal weighting scenario (S1) used the fewest levels (median = 9, range = 7–12; mean rank = 1.58), favoring greater preservation of spinal mobility ($p < .001$). All optimized strategies required significantly fewer screws than the surgeon-performed constructs (20.5 ± 3.0 screws, 95% CI: [19.10, 21.90]; Figure 6.3C). S1 used the fewest (15.1 ± 1.7, 95% CI: [14.33, 15.87], $p < .001$), followed by S2 (16.7 ± 1.9, 95% CI: [15.80, 17.60], $p = .004$) and S3 (16.4 ± 2.0, 95% CI: [15.40, 17.30], $p < .001$). Among all optimized strategies, convex alternate screw patterns were most common (CA, 47%; n=28/60 cases) followed by convex periapical dropout (CPAD, 37%; n = 22/60). Bilateral (B, 8%, n = 5/60) and peri-apical dropout (PAD, 8%, n = 5/60) were rarely selected as optimal strategies. The equal weighting scenario (S1) and coronal correction-focused scenario (S2) yielded the highest proportions of CA patterns (60%, n=12/20 and 50%, n=10/20, respectively), while S3 often prioritized CPAD (55%, n=11/20). UIV and LIV selection varied across strategies (Figure 6.3A). Optimized configurations most frequently chose T5 (45%; n = 27/60) and T4 (20%; n=12/60) as UIV, while T3 was favored in surgeon-performed plans (50%, n=10/20), with T5 used only once (5%). Across all strategies, L1 was the most frequent LIV (40%, n=32/80). Optimized constructs ended as proximally as T11 in 12% of cases (n=7/60), whereas surgeon-performed constructs ended above T12 in only one case (5%).

6.1.4.3 3D Deformity Correction and Forces Experienced by Spinal Implants

Instrumentation strategy significantly influenced Cobb angle correction ($F(1.914, 36.360)=5.48$, $p = .009$, $\eta^2 = 0.224$, Figure 6.4A). The correction-focused strategy (S2) achieved the greatest improvement among optimized configurations, significantly outperforming S1 (mean difference = -3.7°, 95% CI: [-5.72, -1.59], p =.002) and surgeon-performed constructs (-3.2°, 95% CI: [-5.81, -

0.58], p = .019). No statistically significant differences were observed in postoperative T4–T12 kyphosis and AVR across strategies (p = .372 and .077, respectively, Figure 6.4BC), although AVR was marginally greater in surgeon-performed plans compared to S1 (mean difference = -2.8°, 95% CI: [-5.67, 0.30], p = .05). Apex pullout forces did not differ significantly (p = .942), suggesting comparable biomechanical loading on screws across all approaches. A representative patient example is presented in Figure 6.5 to illustrate differences in instrumentation characteristics and simulated outcomes between the actual surgery and each of the scenarios used to identified patient-specific optimization strategies.

6.1.5 Discussion

This study presents a novel framework for optimizing PSF instrumentation in AIS using a combination of AI-predicted configurations and patient-specific deterministic biomechanical simulation, essentially functioning as a digital twin of the patient's spine and surgical construct. It differs from prior optimization studies such as those by La Barbera [21] and Majdouline [33], which primarily focused on a limited number of target variables, by integrating a broader set of clinically relevant inputs through AI prediction, subsequently refined via deterministic simulation. Moreover, it distinguishes itself from purely AI-based planning approaches, such as those described by Lafage [39] or the earliest NNML approach by Constant used in this study [19] by incorporating biomechanical simulations to assess the mechanical feasibility of each construct and support multi-objective optimization. By evaluating hundreds of configurations per patient and applying multi-objective optimization, this digital twin approach identified instrumentation strategies that achieved comparable or superior 3D correction relative to surgeon-performed constructs, while also reducing implant burden and preserving spinal mobility.

6.1.5.1 3D Deformity Correction and Surgical Prioritization

AIS is inherently a three-dimensional spinal deformity, and efforts to optimize surgical correction must address all three planes[2]. In this study, surgeon-performed instrumentation achieved similar results to the "expert-informed" optimization scenario (S3) in all three planes. Clinical intuition and operative technique of experienced scoliosis surgeons were well-aligned with S3's weighted correction priorities. On average, the coronal-focused optimization scenario (S2) achieved even greater Cobb angle correction, suggesting that algorithmic optimization may further improve outcomes when maximal correction in a specific plane is desired.

Interestingly, the equal-weighting scenario (S1), which treated all correction goals equally, outperformed the surgeon-performed constructs in reducing apical vertebral rotation and did not induce any postoperative hypokyphosis (TK range 20–39°), in contrast to surgeon-performed plans where hypokyphosis occurred in 15% of cases (TK range 6.3–38°). This supports prior modeling work showing that when sagittal plane correction is prioritized explicitly, or at least not underemphasized compared to coronal correction, better kyphosis restoration is possible without necessarily increasing implant stress or construct length[34]. While our framework was adapted to each patient's anatomy and deformity, it did not incorporate individual preferences or priorities such as a desire for improved rib hump appearance or trunk balance. These results nonetheless highlight the potential of such an approach to further customized surgical planning to be more effectively tailored to patient-specific goals whether that means prioritizing coronal correction for cosmetic reasons, improving kyphosis in hypokyphotic patients, or targeting AVR reduction in patients with pronounced rib hump or trunk imbalance.

Yet, even with optimization, no simulated configuration achieved complete correction in all three planes. None of the patients had a modeled plan that simultaneously reached >95% coronal correction, thoracic kyphosis within 20–40°, and near-zero AVR. This reinforces the reality that 3D deformity correction in AIS involves trade-offs. Despite exploring nearly 20,000 configurations, the "perfect" plan remained elusive in some patients. These findings mirror clinical experience, where achieving optimal correction across all dimensions is often limited by mechanical constraints, curve stiffness, and anatomic variability constraints including unknown or poorly characterized curve parameters[4, 9]. They also underscore the challenge of achieving sufficient controllable degrees of motion with current instrumentation systems to allow individualized vertebra-level adjustments and regional correction of scoliotic segments. Moreover, the long-term stability of simulated corrections remains uncertain, particularly regarding risks of adding-on or loss of correction. Ultimately, the surgeon's role in setting priorities, guided by clinical judgment, patient-specific factors, and shared decision-making, remains central.

6.1.5.2 Model Design Considerations and Implications

To ensure credible and unbiased comparison across strategies, the surgeon-performed instrumentation was first simulated and validated by confirming close alignment between simulated and actual postoperative 3D reconstructions. This validation step reinforced the model's clinical relevance and served as a foundation for simulating all optimized configurations under identical surgical conditions, therefore minimizing confounding from modeling variability and attributing outcome differences to instrumentation strategy alone. However, applying uniform maneuvers across all constructs may have constrained the optimization potential, particularly for axial correction. For instance, derotation torque was capped at the clinical AVR, potentially limiting achievable correction in constructs that could have tolerated more. While this standardization enhanced model credibility, it may have underestimated the full benefit of certain optimized strategies. Future work should examine the impact of tailoring surgical maneuvers to the mechanical properties of each optimized construct to better align planning with patient-specific biomechanics.

6.1.5.3 Implant Density and Fusion Levels

In this study, optimized plans used fewer pedicle screws than surgeon-performed instrumentation while achieving similar 3D correction. This aligns with prior evidence suggesting that low-density constructs can provide adequate correction in selected AIS cases[25, 34]. Reducing implant density may help decrease costs, operative time, and risks associated with screw placement, which are considerations of growing importance given rising scoliosis surgery expenditures[40-42]. Nevertheless, the MIMO study reported no significant differences in operative time or blood loss between high- and low-density constructs, suggesting that potential efficiency gains from reduced screw counts may be more limited than often assumed[25]. Optimized strategies also tended to produce shorter fusion constructs, particularly under equal-weighting (S1), with a quite remarkable median decrease of 2 levels instrumented. Although not directly evaluated, shorter fusions have been associated with reduced hospitalization costs, preserved mobility, and fewer complications[43-47]. These findings suggest that patient-specific simulation may support more efficient instrumentation strategies, provided correction and construct stability remain uncompromised. However, with accelerated discharge pathways such as the Fletcher protocol in place[48], further reductions in length of hospital stay may be unlikely to result from changes in fusion length or implant density alone.

6.1.5.4 Limitations, Clinical Relevance and Future Directions

While not yet applied in real-time clinical decision-making, this framework was developed and tested under conditions closely mirroring surgical practice, reinforcing its relevance as a foundation for future decision-support tools. Its clinical integration remains to be validated in prospective studies. Rather than prescribing a single optimal solution, it could help surgeons navigate multiple construct options with varying trade-offs, tailored to individual anatomy and treatment goals. With integration into clinical workflows, such tools could enhance personalization and reduce outcome variability in complex or uncertain cases. Further, our aim is to provide a selection of possible surgical plans to the surgeon. Introduction of new technology for surgical care of patients is best done in an incremental fashion with significant room for surgeon input as the first iterations are validated. AI and deterministic surgical planning tools should serve to augment surgeon decision-making. Eventually such tools can provide a positive feedback loop as surgeons and AI models can learn from past models.

Several limitations remain. Several factors that can influence surgical decision-making and outcomes in AIS were not explicitly modeled in this study. For example, posterior column osteotomies, which are increasingly used in larger or stiffer curves or other techniques focused on anterior column lengthening to address three-plane deformity were not incorporated. Parameters such as thoracic kyphosis targets tailored to pelvic incidence, shoulder balance optimization, and prediction of compensatory lumbar curve behavior were also not included. The framework did not optimize individual screw size, screw head, thread pitch, or screw trajectory; instead, it relied on predefined, commonly used patterns. Nevertheless, the approach developed would allow such optimization and could be explored in a subsequent study. It also simplified certain surgical realities and excluded biological factors like fusion capacity or bone quality. These aspects remain areas of active investigation, and future work could integrate them into optimization frameworks. Finally, although only three optimization scenarios were tested, each involving extensive simulation and comparison, the framework is flexible and could support patient-specific prioritization, which is an important direction for future refinement.

6.1.6 Conclusion

This study presents an original patient-specific framework combining AI-based predictions with deterministic biomechanical simulation and multi-objective optimization to support surgical planning in AIS. In most cases, the optimized instrumentation strategies achieved comparable or better 3D deformity correction than surgeon-performed constructs, while using fewer implants and preserving more spinal mobility. Pullout forces at the apex were also comparable across strategies, indicating that these gains may be achieved without increasing local biomechanical loading on the implants.

By identifying configurations not readily apparent through standard planning, this approach highlights how surgical priorities, such as coronal correction, kyphosis restoration, or shorter fusion, can be tailored to individual patients. While this tool is not a replacement for clinical expertise, it is intended to support, not supplant, the surgeon's judgment. It may offer particular value in complex cases, when planning goals are uncertain, or as a decision aid for less experienced clinicians. Further clinical validation is needed, but these findings suggest that integrating AI guidance with biomechanical modeling could help move scoliosis surgery toward more personalized and data-informed planning.

6.1.7 References

1. Thomas, J.J., et al., Does school screening affect scoliosis curve magnitude at presentation to a pediatric orthopedic clinic? Spine deformity, 2018. 6(4): p. 403-408.

2. Negrini, S., et al., Actual evidence in the medical approach to adolescents with idiopathic scoliosis. 2014.

3. Kim, S.-S., J.-H. Kim, and S.-I. Suk, Effect of Direct Vertebral Rotation on the Uninstrumented Lumbar Curve in Thoracic Adolescent Idiopathic Scoliosis. Asian Spine J, 2017. 11(1): p. 127-137.

4. de Kleuver, M., et al., Optimal surgical care for adolescent idiopathic scoliosis: an international consensus. European Spine Journal, 2014. 23(12): p. 2603-2618.

5. Weinstein, S.L., et al., Adolescent idiopathic scoliosis. The Lancet, 2008. 371(9623): p. 1527-1537.

6. Aubin, C.E., H. Labelle, and O.C. Ciolofan, Variability of spinal instrumentation configurations in adolescent idiopathic scoliosis. Eur Spine J, 2007. 16(1): p. 57-64.

7. Le Naveaux, F., et al., Significant variability in surgeons' preferred correction maneuvers and instrumentation strategies when planning adolescent idiopathic scoliosis surgery. Scoliosis Spinal Disord, 2018. 13: p. 21.

8. Schlager, B., et al., Range of surgical strategies for individual adolescent idiopathic scoliosis cases: evaluation of a multi-centre survey. Spine Deformity, 2023.

9. Pasha, S. and J.M. Mac-Thiong, Defining criteria for optimal lumbar curve correction following the selective thoracic fusion surgery in Lenke 1 adolescent idiopathic scoliosis: developing a decision tree. Eur J Orthop Surg Traumatol, 2020. 30(3): p. 513-522.

10. Pasha, S., et al., Timing of Changes in Three-Dimensional Spinal Parameters After Selective Thoracic Fusion in Lenke 1 Adolescent Idiopathic Scoliosis: Two-Year Follow-up. Spine Deform, 2017. 5(6): p. 409-415.

11. Takahashi, J., et al., Selective thoracic fusion in adolescent idiopathic scoliosis: factors influencing the selection of the optimal lowest instrumented vertebra. Spine (Phila Pa 1976), 2011. 36(14): p. 1131-41.

12. Ramo, B.A. and B.S. Richards, Repeat Surgical Interventions Following "Definitive" Instrumentation and Fusion for Idiopathic Scoliosis: Five-Year Update on a Previously Published Cohort. Spine, 2012. 37(14): p. 1211-1217.

13. Clements, D.H., et al., Did the Lenke classification change scoliosis treatment? Spine (Phila Pa 1976), 2011. 36(14): p. 1142-5.

14. King, H.A., et al., The selection of fusion levels in thoracic idiopathic scoliosis. J Bone Joint Surg Am, 1983. 65(9): p. 1302-13.

15. Lenke, L.G., et al., Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. J Bone Joint Surg Am, 2001. 83(8): p. 1169-81.

16. Ferrero, E., et al., Preliminary experience with SpineEOS, a new software for 3D planning in AIS surgery. Eur Spine J, 2018. 27(9): p. 2165-2174.

17. Lafage, R., et al., Self-learning computers for surgical planning and prediction of postoperative alignment. Eur Spine J, 2018. 27(Suppl 1): p. 123-128.

18. Constant, C., et al., The use of deep learning in medical imaging to improve spine care: A scoping review of current literature and clinical applications. N Am Spine Soc J, 2023. 15: p. 100236.

19. Constant, C., et al., Neural network-based multi-task learning to assist planning of posterior spinal fusion surgery for adolescent idiopathic scoliosis. Spine Deformity, 2025.

20. Constant, C., et al., AI-Derived vs. Surgeon-Performed Instrumentation in Adolescent Idiopathic Scoliosis: A Biomechanical Simulation Analysis. 2025: The Spine Journal.

21. La Barbera, L., et al., In silico patient-specific optimization of correction strategies for thoracic adolescent idiopathic scoliosis. Clinical Biomechanics, 2021. 81: p. 105200.

22. Le Navéaux, F., et al., 3D rod shape changes in adolescent idiopathic scoliosis instrumentation: how much does it impact correction? European Spine Journal, 2017. 26(6): p. 1676-1683.

23. Wang, X., et al., Biomechanical effect of pedicle screw distribution in AIS instrumentation using a segmental translation technique: computer modeling and simulation. Scoliosis Spinal Disord, 2017. 12: p. 13.

24. Le Navéaux, F., et al., How does implant distribution affect 3D correction and bone-screw forces in thoracic adolescent idiopathic scoliosis spinal instrumentation? Clin Biomech (Bristol, Avon), 2016. 39: p. 25-31.

25. Larson, A.N., et al., The Effect of Implant Density on Adolescent Idiopathic Scoliosis Fusion: Results of the Minimize Implants Maximize Outcomes Randomized Clinical Trial. J Bone Joint Surg Am, 2023.

26. Wang, X., et al., Concave rod first vs. convex rod first in AIS instrumentation with differential rod contouring: computer modeling and simulations based on ten AIS surgical cases. Spine Deformity, 2023. 11(6): p. 1317-1324.

27. Cheriet, F., et al., A Novel System for the 3-D Reconstruction of the Human Spine and Rib Cage From Biplanar X-Ray Images. IEEE Transactions on Biomedical Engineering, 2007. 54(7): p. 1356-1358.

28. Delorme, S., et al., Assessment of the 3-d reconstruction and high-resolution geometrical modeling of the human skeletal trunk from 2-D radiographic images. IEEE Trans Biomed Eng, 2003. 50(8): p. 989-98.

29. Panjabi, M.M., R.A. Brand Jr, and A.A. White III, Three-dimensional flexibility and stiffness properties of the human thoracic spine. Journal of biomechanics, 1976. 9(4): p. 185-192.

30. Gardner-Morse, M.G. and I.A. Stokes, Structural behavior of human lumbar spinal motion segments. Journal of biomechanics, 2004. 37(2): p. 205-212.

31. Petit, Y., C.É. Aubin, and H. Labelle, Patient-specific mechanical properties of a flexible multi-body model of the scoliotic spine. Medical and Biological Engineering and Computing, 2004. 42(1): p. 55-60.

32. Lehman Jr, R.A., et al., Straight-forward versus anatomic trajectory technique of thoracic pedicle screw fixation: a biomechanical analysis. Spine, 2003. 28(18): p. 2058-2065.

33. Majdouline, Y., et al., Computer simulation for the optimization of instrumentation strategies in adolescent idiopathic scoliosis. Med Biol Eng Comput, 2009. 47(11): p. 1143-54.

34. La Barbera, L., A.N. Larson, and C.E. Aubin, Correction objectives have higher impact than screw pattern and density on the optimal 3D correction of thoracic AIS: a biomechanical study. Spine Deform, 2021. 9(3): p. 655-664.

35. Sucato, D.J., et al., Restoration of thoracic kyphosis after operative treatment of adolescent idiopathic scoliosis: a multicenter comparison of three surgical approaches. Spine (Phila Pa 1976), 2008. 33(24): p. 2630-6.

36. Liljenqvist, U., et al., Pullout strength of pedicle screws versus pedicle and laminar hooks in the thoracic spine. Acta Orthop Belg, 2001. 67(2): p. 157-63.

37. Majdouline, Y., et al., Scoliosis correction objectives in adolescent idiopathic scoliosis. J Pediatr Orthop, 2007. 27(7): p. 775-81.

38. Majdouline, Y., et al., Preoperative assessment and evaluation of instrumentation strategies for the treatment of adolescent idiopathic scoliosis: computer simulation and optimization. Scoliosis, 2012. 7(1): p. 21.

39. Lafage, R., et al., Predictive Model for Selection of Upper Treated Vertebra Using a Machine Learning Approach. World Neurosurg, 2021. 146: p. e225-e232.

40. Vigneswaran, H.T., et al., Surgical treatment of adolescent idiopathic scoliosis in the United States from 1997 to 2012: an analysis of 20, 346 patients. Journal of Neurosurgery: Pediatrics, 2015. 16(3): p. 322-328.

41. Baky, F.J., et al., Predictors of cost for posterior spinal fusion in adolescent idiopathic scoliosis. Spine Deform, 2020. 8(3): p. 421-426.

42. Larson, A.N., et al., What would be the annual cost savings if fewer screws were used in adolescent idiopathic scoliosis treatment in the US? J Neurosurg Spine, 2016. 24(1): p. 116-23.

43. Yoshihara, H., C. Paulino, and D. Yoneoka, Predictors of Increased Hospital Stay in Adolescent Idiopathic Scoliosis Patients Undergoing Posterior Spinal Fusion: Analysis of National Database. Spine Deformity, 2018. 6(3): p. 226-230.

44. Fletcher, N.D., et al., Clinical and economic implications of early discharge following posterior spinal fusion for adolescent idiopathic scoliosis. Journal of Children's Orthopaedics, 2014. 8(3): p. 257-263.

45. Hengartner, A.C., et al., Impact of a quality improvement initiative and monthly multidisciplinary meetings on outcomes after posterior spinal fusion for adolescent idiopathic scoliosis. Spine Deformity, 2024. 12(4): p. 1043-1051.

46. Harris, A.B., et al., Factors associated with extended length of stay in patients undergoing posterior spinal fusion for adolescent idiopathic scoliosis. Spine Deformity, 2020. 8(2): p. 187-193.

47. Engsberg, J.R., et al., Prospective Comparison of Gait and Trunk Range of Motion in Adolescents With Idiopathic Thoracic Scoliosis Undergoing Anterior or Posterior Spinal Fusion. Spine, 2003. 28(17): p. 1993-2000.

48. Fletcher, N.D., et al., Use of a Novel Pathway for Early Discharge Was Associated With a 48% Shorter Length of Stay After Posterior Spinal Fusion for Adolescent Idiopathic Scoliosis. Journal of Pediatric Orthopaedics, 2017. 37(2).

6.1.8 Tables

Table 6.1 Summary of Data Collected from Dataset

Demographic and clinical data	- Height (cm) - Weight (kg) - Age
Radiographic	- Vertebral corner localization - Lenke classification - Cobb angle measurements (PT, MT, TL/L) - Upper and lower end vertebrae - Stable vertebrae - Thoracic kyphosis (T2-T12) - Risser sign
Surgical	- Upper instrumented vertebra - Lower instrumented vertebra - Screw density - Rod curvatures (concave side and concave/convex side differential bending) - Type of screws and rods used
Postoperative spinal alignment	- Cobb angle measurements (PT, MT, TL/L) - Thoracic kyphosis (T2-T12)

Table 6.2 Descriptive Statistics of the Patients Included in this Study

Category	Population (n=20)
Height (mean ± SD, range)	166 ± 10 cm (154 – 190)
Weight (mean ± SD, range)	59 ± 13 kg (40 – 92)
Age (mean ± SD, range)	16 ± 1.7 years (13 – 19)
Risser Sign (median, range)	4 (2 – 5)
Lenke Classification (1/2 breakdown)	1: n= 17 2: n= 3
Preoperative Cobb angle (mean ± SD, range)	
Proximal Thoracic (PT)	31 ± 8° (13 – 46)
Main thoracic (MT)	59 ± 6° (48 – 68)
Thoracolumbar / Lumbar (TL/L)	35 ± 5° (27-46)
End Vertebrae of main thoracic curve (median, range)	
Upper end vertebra	T6 (T2-T7)
Lower end vertebra	T12 (T10-L2)
Apical Vertebrae of main thoracic curve (median, range)	T9 (T7-T11)
Preoperative Thoracic Kyphosis T4-T12 (mean ± SD, range)	30 ± 14° (8 – 63°)
Preoperative Apical Vertebra Rotation (mean ± SD, range)	-17° ± 7° (-30 - -5°)

6.1.9 Figures

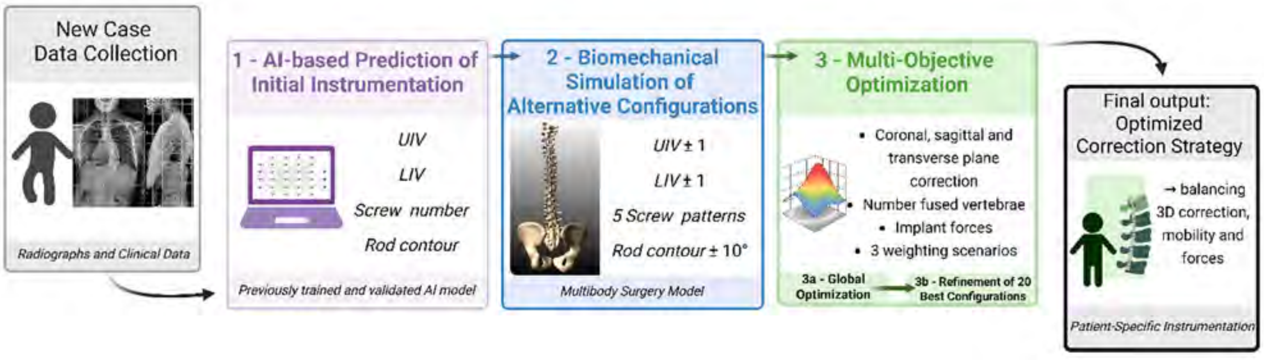


Figure 6.1 Overview of the Patient-Specific Optimization Workflow

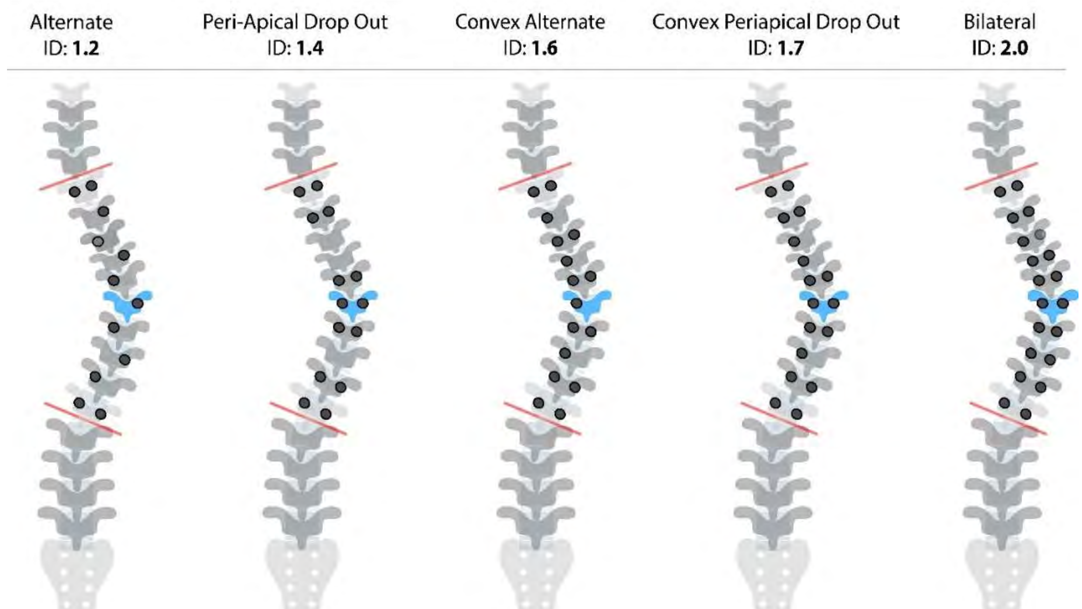


Figure 6.2 Screw Patterns and Corresponding Implant Densities (ID) used for Modeling AI-Derived Surgical Instrumentation

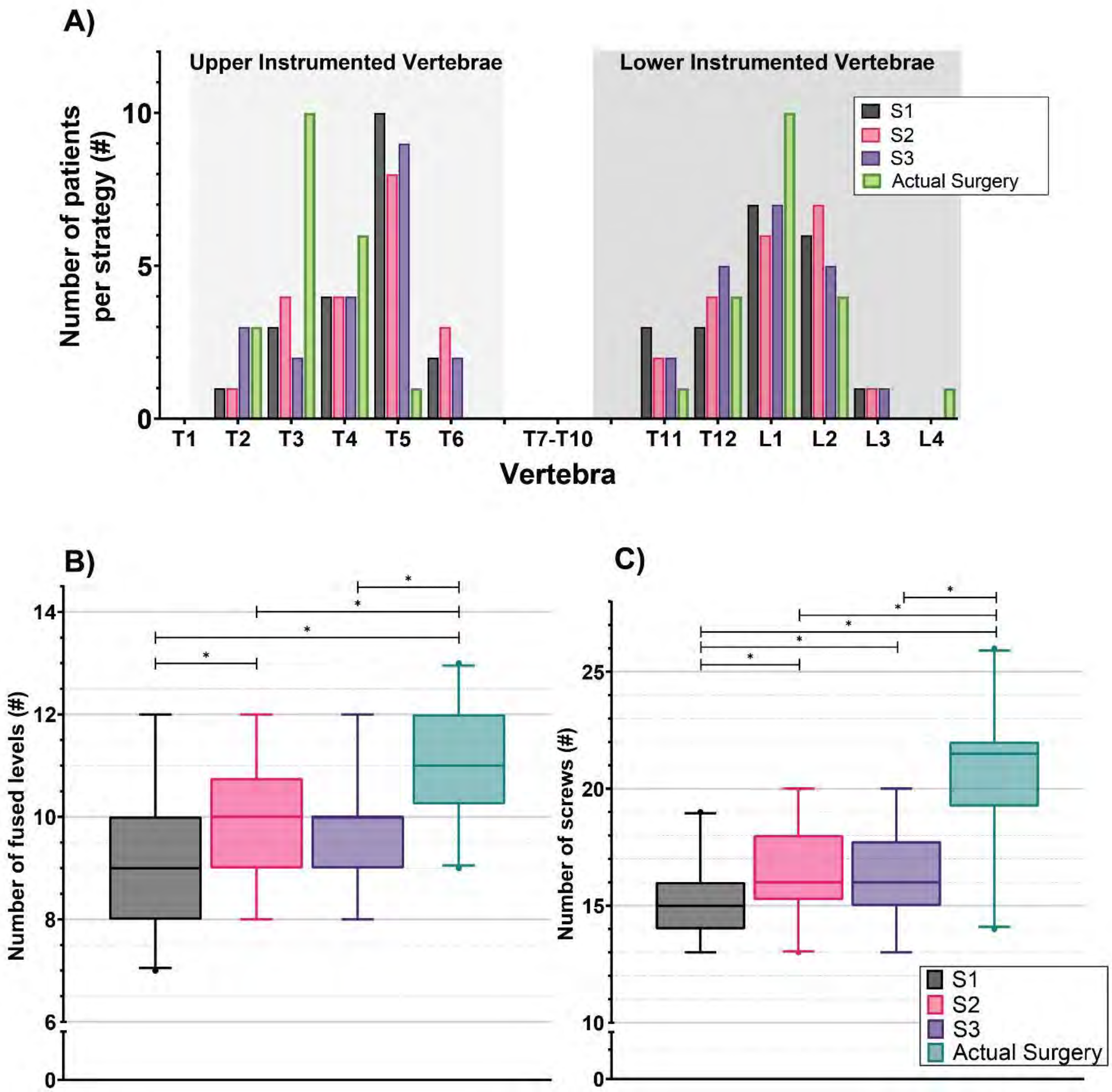


Figure 6.3 Instrumentation Characteristics Across Surgical Strategies

(A) Distribution of Upper Instrumented Vertebrae (UIV, left) and Lower Instrumented Vertebrae (LIV, right) for each strategy, expressed as the number of patients receiving instrumentation at each level. (B) Number of fused levels per patient. (C) Number of pedicle screws per patient. Data are presented for three optimized strategies and the actual surgeon-performed instrumentation: the equal weighting scenario (S1), the coronal correction-focused scenario (S2), the expert clinician-based scenario (S3), and the surgeon-performed strategy (Surgery).

In the box plots, the middle line represents the mean, the box boundaries represent the 25th and 75th percentiles, and the whiskers indicate the 5th and 95th percentiles. *Asterisks (*) denote statistically significant differences ($p < 0.05$) between groups.*

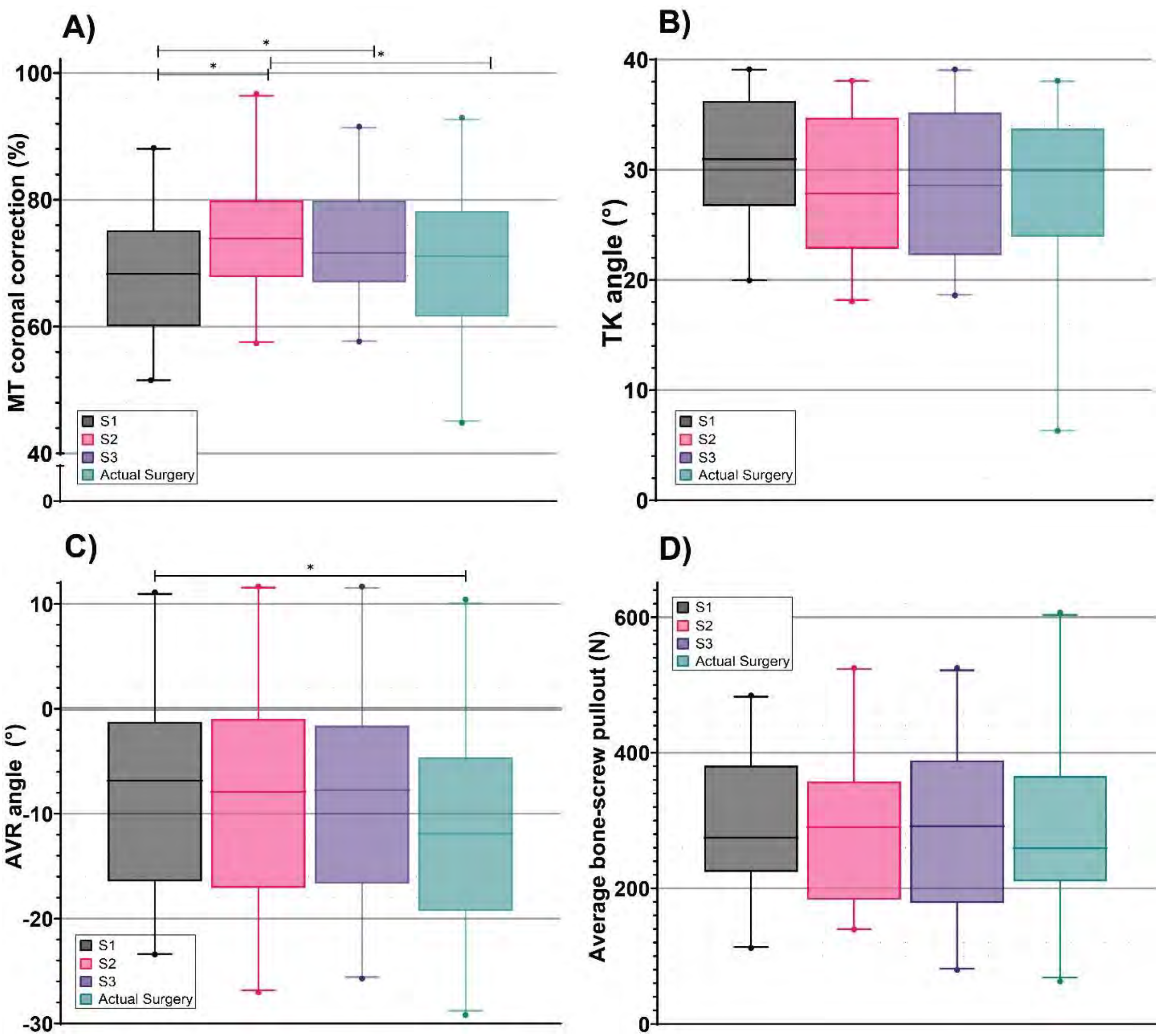


Figure 6.4 Postoperative 3D Deformity Correction and Forces Experienced by Spinal Screws

Data are presented for (A) mean percent correction of the main thoracic (MT) coronal curve, (B) T4–T12 thoracic kyphosis, (C) thoracic apical vertebral rotation (AVR), and (D) mean pullout force experienced by the apical pedicle screws, stratified by three AI-optimized strategies and the actual surgeon-performed instrumentation: the equal weighting scenario (S1), the coronal correction-focused scenario (S2), the expert clinician-based scenario (S3), and the surgeon-performed strategy (Surgery).

In the box plots, the middle line represents the mean, the box boundaries represent the 25th and 75th percentiles, and the whiskers indicate the 5th and 95th percentiles. *Asterisks (*) denote statistically significant differences ($p < 0.05$) between groups.*

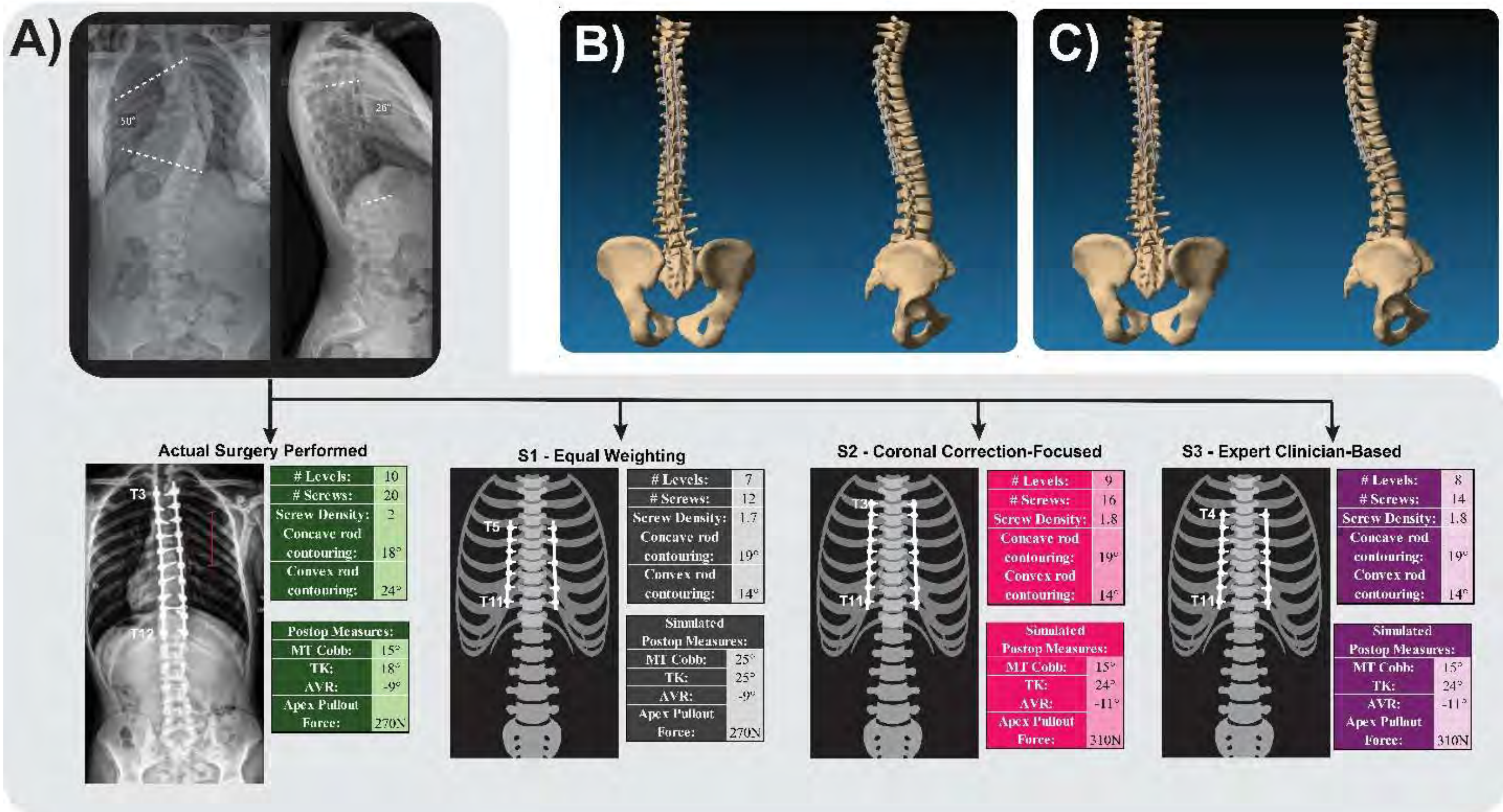


Figure 6.5 Patient-Specific Example Comparing Optimized and Surgeon-Performed Instrumentation Strategies

(A) Preoperative coronal and sagittal radiographs of one representative patient (top row), and (bottom row) actual postoperative plan and simulated outcomes for the three scenarios of optimized strategies: equal weighting (S1), coronal correction-focused (S2), and expert clinician-based (S3). For each scenario, the instrumented levels, screw density, rod curvature, and simulated postoperative alignment and screw pullout force are shown. (B) 3D rendering of the simulated surgical outcome using the actual instrumentation performed by the surgeon and (C) of the simulated outcome for the coronal correction-focused strategy (S2), highlighting the possibility in this case to achieve similar 3D correction while preserving more spinal mobility by fusing one fewer level and using four fewer screws than the surgeon's original plan.

6.2 Complementary Methodological Aspects - Regulatory Context and Credibility of the CM&S Framework

6.2.1 Regulatory Status and International Alignment

This section assesses the credibility of the computational modeling and simulation (CM&S) framework developed in this thesis, which supports surgical planning for PSF in AIS patients. The CM&S community and regulatory bodies have issued several key guidelines to support the credible development of such models. Among these, the ASME V&V 10 standard offers general principles for verification and validation in computational mechanics, while ASME V&V 40 provides a risk-informed framework tailored to medical applications [249, 250].

In the United States, these standards are supported by guidance from the U.S. Food and Drug Administration (FDA), which promotes the use of computational models in regulatory submissions when credibility is appropriately demonstrated [251-253]. In Canada, Health Canada's Medical Devices Directorate (MDD) accepts computer models and modeling data in device applications and encourages adherence to international standards, particularly for innovative or high-risk technologies [254]. However, based on Canada's current guidance for Software as a Medical Device (SaMD), the CM&S framework presented in this thesis does not meet the definition of a regulated medical device. It satisfies all four of Health Canada's exclusion criteria: (1) it does not acquire or analyze medical images or physiological signals; (2) it displays and compares standard clinical metrics such as Cobb angles and alignment parameters; (3) it is intended only to support rather than drive or replace clinical decision-making; and (4) it does not replace the clinical judgment of healthcare professionals, who maintain complete authority over surgical planning [255]. As such, the model and optimization framework we developed is not subject to medical device licensing requirements under Canadian law [254, 255].

Nonetheless, given the increasing global convergence in regulatory expectations and the potential for future clinical use, the model's credibility was evaluated using the ASME V&V 40 framework, with relevant verification, validation, and uncertainty quantification (VVUQ) activities conducted in accordance with ASME V&V 10 guidelines. This structured, transparent, and risk-based approach ensures scientific rigor, and we believe it will broaden acceptance within both clinical and research communities. This emphasis on international best practices is particularly appropriate given that this project was conducted in collaboration with institutions in both Canada and the United States, each operating under distinct regulatory environments. These partnerships underscore the importance of aligning with globally recognized standards to ensure consistency, reproducibility, and readiness for clinical translation.

6.2.2 Credibility Assessment of the Computational Model Developed as a Medical Device

This section presents a structured credibility assessment of the CM&S framework developed to support surgical planning for PSF surgery in AIS patients. The computational model was first introduced in Chapter 5 (Article 3), where it was used to simulate and compare surgeon-performed versus AI-derived instrumentation strategies through patient-specific multibody simulations. In that context, the model underwent VVUQ following ASME principles to ensure the reliability of its outputs related to spinal alignment and implant loading [250]. In the current chapter (Article 4), the same model was integrated into a larger optimization framework that systematically evaluated over 900 construct configurations per patient to select an optimized instrumentation strategy. This integration demonstrated the model's scalability and potential as a clinical decision-support tool. Because this final application directly informs surgical planning decisions, a formal credibility assessment following the ASME V&V 40 standard was conducted [249]. The work presented in this section thus combines the VVUQ procedures established in Chapter 5 with the extended application presented in this current chapter to evaluate the credibility of the complete optimization framework as a medical device. While the optimization step builds on the validated simulation core, both components are interdependent and are assessed here as a single, clinically relevant pipeline.

6.2.2.1 Risk-Informed Credibility Assessment Framework

According to ASME, computational modeling refers to the numerical implementation of a mathematical model to solve clinical or engineering problems. To evaluate and communicate the credibility of the developed CM&S framework, this thesis adopts the terminology and structure outlined in the ASME V&V 40 (2018) standard [249]. This risk-informed approach defines a stepwise process for model evaluation, starting from the question of interest and context of use (COU), and guiding the selection of appropriate VVUQ activities (Figure 6.6). In this thesis, verification activities focused on confirming the numerical implementation of the MB simulation model and the stability of solver behavior. Validation assessed the model's ability to reproduce real-world postoperative outcomes in AIS patients, and uncertainty quantification explored the impact of input variability on predicted outputs. Each of these components is addressed in the following sections, with their associated credibility factors and results summarized in dedicated tables.

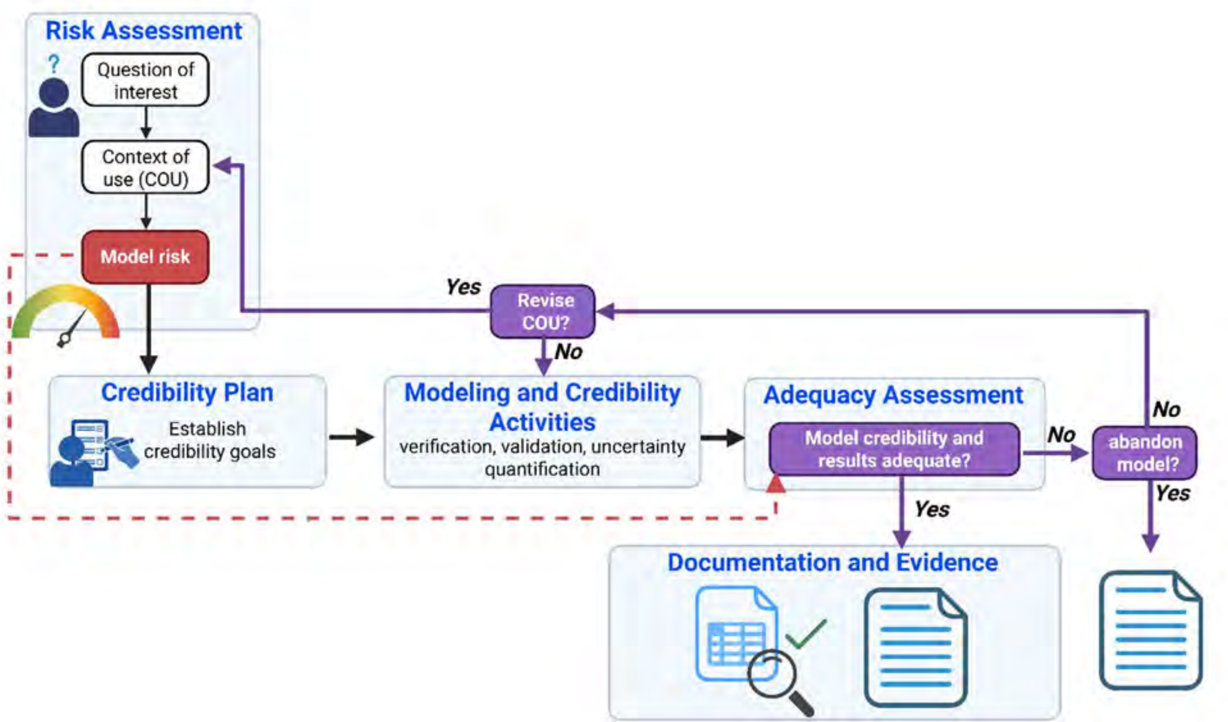


Figure 6.6 Establishing Model Credibility for Verification, Validation, and Uncertainty Quantification

This diagram illustrates the process of establishing model credibility, which guides the verification, validation, and uncertainty quantification activities.

6.2.2.2 Context of Use of the Model

The COU defines the specific role and scope of a computational model in supporting a clinical or engineering decision. A well-defined COU is essential for framing the purpose, scope, and clinical integration of a computational model. It directly informs its risk classification and the extent of VVUQ required. According to ASME V&V 40, computational models in healthcare typically fall into three roles: embedded in a medical device, used as valid scientific evidence, or serving as standalone decision-support tools [249]. The CM&S framework developed in this thesis aligns with the third category. Thus, its COU is defined as highlighted below.

Context of Use of the Model

"To support surgical planning in AIS by generating patient-specific biomechanical predictions of postoperative alignment and implant loading, and by selecting optimized posterior spinal fusion constructs based on predefined clinical and mechanical objectives."

This classification, as a decision-support tool that includes optimization, sets the model's intended use and underpins the credibility assessment approach presented in the following sections.

6.2.2.3 Model Risk Assessment

Model risk refers to the combination of model influence and decision consequence, and is a key consideration in determining the appropriate level of rigor for VVUQ activities. Model influence reflects the degree to which simulation outputs contribute to the clinical decision. In contrast, decision consequence refers to the potential severity of harm if a decision informed by the model were incorrect.

In this thesis, a computational framework was developed as a clinical decision support system for PSF in AIS. Given the model COU, the *model's influence* is considered medium: its recommendations meaningfully contribute to preoperative decision-making but remain advisory. The *decision consequence* is classified as low, since any suboptimal outcome arising from model-informed planning would likely result in reduced surgical efficiency or less effective correction, rather than irreversible harm; ultimately, the surgeon retains full authority over the final clinical decision. Together, these assessments yield an overall **model risk classification of medium–low**, which determines the appropriate credibility goals and justifies the level of detail in the VVUQ activities presented in the following sections. A visual summary of the influence–consequence interaction is provided in Figure 6.7.

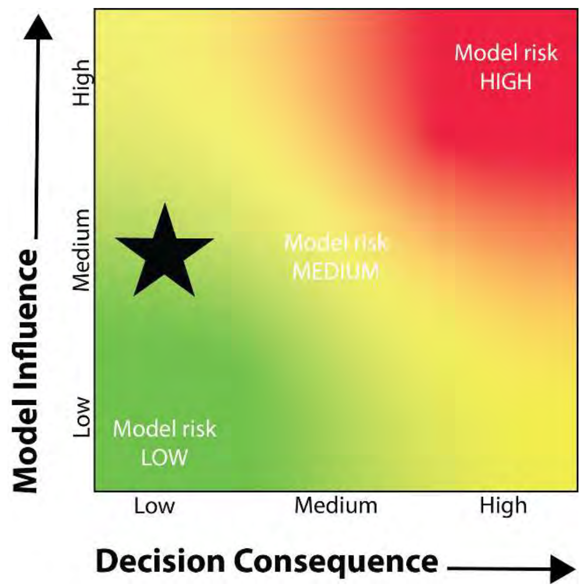


Figure 6.7 Graphical Representation of Model Risk Assessment
The overall model risk classification was evaluated to be medium–low (star).

6.2.2.4 Credibility Factor Analysis

To evaluate model performance against the 23 credibility factors defined in ASME V&V 40, a structured assessment was conducted across four domains: verification, validation, uncertainty quantification, and applicability (complete list of factors and their grading is provided in *Appendix F - ASME V&V 40 Credibility Factors*). These factors are designed to determine whether a model meets the credibility standards appropriate for its intended clinical use. Each factor includes graded levels of rigor, ranging from (a) low credibility, to (b) basic credibility, which is typically sufficient for low-risk models, to (c) moderate credibility, suitable for medium-risk models, and up to (d) high credibility, expected for high-risk or influential models. Given the model's medium-low risk classification, the target levels were set accordingly: at least level (b) or (c) for factors with three rigor levels, and (c) or (d) for those with four. This grading ensured the model was assessed proportionally to its role in clinical decision-making while accounting for its current stage of development. To enable direct comparison across factors with varying grading schemes, a 12-point normalization scale was applied following published methodologies [256]. This mapped two-level factors to values of 1 and 6; three-level factors to 1, 6, and 12; and four-level factors to 1, 4, 8, and 12. As shown in Figure 6.8 and detailed in the supporting tables (Tables 6.3-6.6), most factors met or exceeded the desired credibility thresholds. Several reached level (c), indicating moderate credibility; others achieved level (b), which is acceptable for models not yet in clinical deployment. A few factors, such as discretization error and numerical code verification, were rated lower (or b), which highlight the model's early development stage and its limitations.

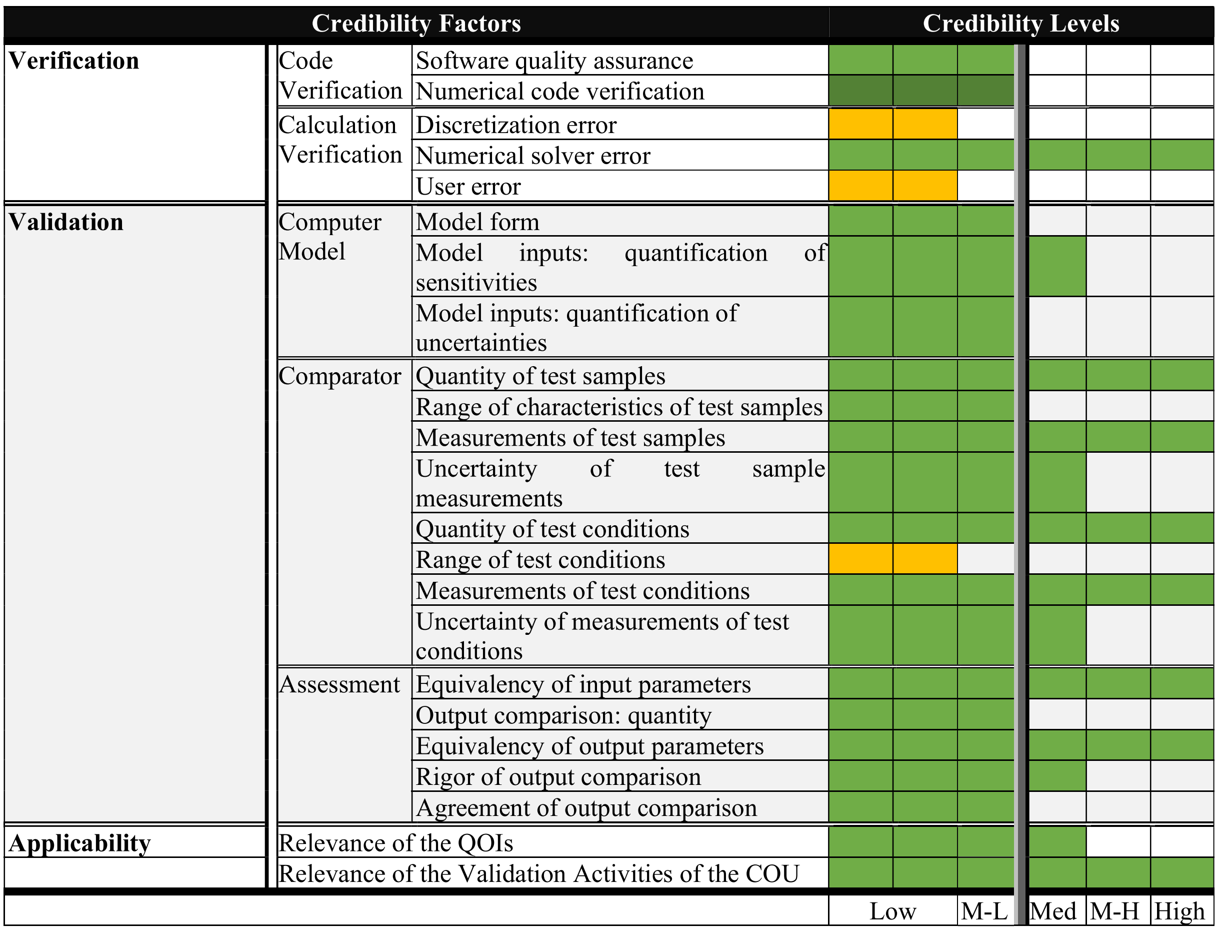


Figure 6.8 Summary of Credibility Factors and Goals Selected for this Study

Colors show alignment with credibility goals. Green meets or exceeds the threshold; orange falls slightly below. The bold line marks the minimum required level for the model’s medium-low (M-L) risk.

6.2.2.5 Verification

Verification assesses whether the computational model was implemented and executed correctly. For this thesis, all simulations were performed using MSC Adams 2019, a widely used commercial software for multibody dynamics. While no formal software quality assurance procedures, such as software metrics tracking, simulations followed standardized workflows and consistently generated reproducible outputs. This level of quality control corresponds to a basic but acceptable credibility level (b) for software quality assurance and code verification, as summarized in Table 6.3.

Numerical solver behavior was evaluated in a targeted sensitivity study (*Appendix G – Solver Parameter Sensitivity Study*) using 20 representative simulation cases (10 based on postoperative patient constructs and 10 on AI-predicted strategies). Six solver parameters were perturbed around their default values, including residual tolerance, stability factor, imbalance limit, time and angular limits, and maximum iterations. Across all perturbations, no statistically significant differences were observed for MT Cobb angle, TK, AVR, or mean pullout force (all $p > 0.05$). Mean deviations were generally negligible (≤1° for Cobb and AVR; ≤3.5 N for pullout). The only notable case was TK, which exhibited a deviation of approximately 6° under the most extreme stability settings. Since the default solver stability (0.001) was retained in all analyses, this deviation is not considered relevant to the proposed computer model and does not impact the study conclusions. The reduced iteration number, capped at 25, applied for large-scale optimization in Chapter 6 (Article 4), also introduced only minimal deviations. Together, these findings demonstrate that the solver configuration provides numerically stable and decision-invariant outputs, supporting the assignment of credibility level (c) for numerical solver error. Full results are detailed in *Appendix G*.

Nevertheless, discretization error and user error scored below the target level for medium-low risk models. Discretization effects were not explicitly quantified, and while user inputs were reviewed and tested for reproducibility, no formal input logging or error-checking tools were implemented. These limitations reflect the early development stage of the framework, and are acknowledged as areas for refinement in future iterations.

Table 6.3 Goals for Credibility Factors associated with Verification.

Credibility Factor	Selected Grade	Justification Summary
Code Verification		
Software Quality Assurance (SQA)	b	Documented research workflows; no formal anomaly tracking or software metrics.
Numerical Code Verification (NCV)	b	Used validated tools; no exact solution or formal grid study.
Calculation Verification		
Discretization Error	a/b	Observed stable behavior; discretization error not explicitly quantified.
Numerical Solver Error (NSE)	c	Sensitivity analysis confirmed minimal solver impact (*Appendix G – Solver Parameter Sensitivity Study*)
Use Error	b	Inputs/outputs verified by the practitioner, along with reproducibility checks.

6.2.2.6 Validation and Applicability

Validation assesses whether the computational model accurately reflects the clinical outcomes it aims to simulate. For this CM&S framework, validation efforts focused on its ability to reproduce the postoperative spinal alignment following PSF in AIS patients. Two complementary strategies were employed: (1) comparison between simulated and actual surgical outcomes, and (2) sensitivity analysis to assess robustness across input variations. Credibility factor scores related to validation are summarized in the following paragraphs, which provide a structured overview of key components assessed, along with the assigned credibility levels and a brief justification for each. Additionally, the following paragraphs elaborate on the datasets, evaluation strategies, and findings from the validation activities.

Cohorts and Comparators

The validation of the model relied on two complementary patient cohorts. The first included 35 AIS patients used to compare simulated versus actual postoperative outcomes (dataset from Chapter 5). The second cohort consisted of 20 AIS patients, in whom the robustness of optimized constructs was tested through targeted sensitivity analyses (dataset from Chapter 6). In both cases, quantitative comparisons focused on clinically relevant alignment metrics such as MT Cobb, TK, AVR, and pullout forces.

Both cohorts reflected a typical adolescent AIS population. Patients were 13–19 years old, with skeletal maturity ranging from Risser 2 to 5, and presented Lenke types 1 and 2. Curve severities spanned moderate to severe deformities (MT Cobb 48–68°, PT Cobb 13–46°, TL/L Cobb 27–46°), accompanied by variable sagittal profiles (TK 8–63°) and a broad range of apical vertebral rotations (–30° to –5°). Anthropometric measures also reflected typical adolescents, with heights ranging from 154 to 190 cm and weights between 40 and 92 kg. This ensured that the cohorts included representative clinical cases, although extreme deformities and atypical Lenke curve types were not explicitly targeted. The test conditions covered a broad range of clinically relevant surgical strategies. Simulations included UIV levels from T2 to T6, LIV from T11 to L4, fusion lengths spanning 7 to 13 vertebrae, screw densities ranging from 1.2 to 2.0, and rod curvatures ranging from 0° to 40°. This range of simulated constructs reflects the planning options most often used in AIS correction and supports the framework's applicability to typical clinical decision-making.

Model Form and Assumptions

The framework builds on prior validation work, where simulated corrections were compared in 35 AIS cases with detailed pre- and postoperative radiographic data and shown to reproduce coronal and sagittal Cobb angles within 5° of surgical outcomes, supporting its adequacy for biomechanical evaluation under analogous surgical conditions [25, 44, 58, 257]. Although certain model assumptions, such as stiffness properties and boundary conditions, were adopted from these earlier validation works, their influence was not explicitly re-evaluated in the present study. Given the comparable context, however, similar behavior can reasonably be expected.

Three-dimensional reconstructions were obtained using a previously validated self-calibration and optimization method [258], with reported accuracy of 1.2 ± 0.8 mm for vertebral bodies and 1.6 ± 1.1 mm for pedicles [137]. These error margins define the resolution of extracted radiographic parameters. As the same reconstruction workflow was applied here, comparable accuracy and repeatability are anticipated without the need for additional measurement studies.

Postoperative Outcome Validation

The first validation strategy assessed whether the simulation framework could reproduce the spinal alignment achieved in surgery, using data from 35 AIS patients (dataset presented in Article 3, Chapter 5). For each case, patient-specific instrumentation was reproduced in the 3D MB model, and predicted alignment was compared with the actual postoperative radiographs. Results demonstrated close agreement between simulated and observed outcomes. Mean differences were small and consistently within a clinically meaningful ±5° threshold, with no statistically significant discrepancies detected (Figure 6.9). With 35 patients, the analysis provided more than 90% statistical power to confirm equivalence for MT Cobb, establishing strong confidence in the accuracy of the framework. These findings support that the model produces biomechanically realistic and clinically acceptable predictions of postoperative alignment. Detailed results, including per-patient comparisons and statistical outputs, are presented in *Appendix H - Postoperative Outcome Validation.*

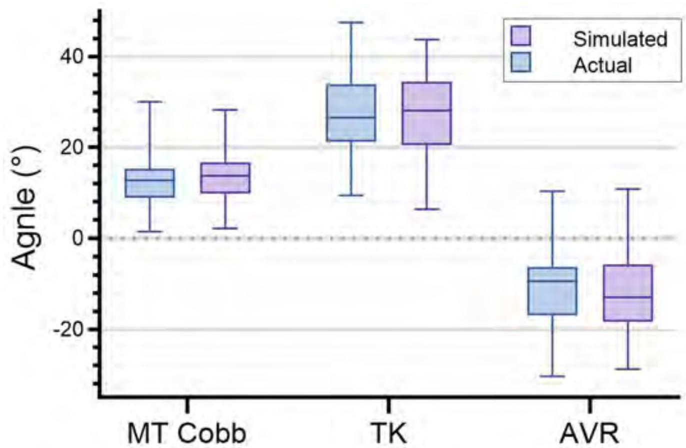


Figure 6.9 Comparison of Simulated and Actual Postoperative 3D Alignment Outcomes

Boxplots show MT Cobb, TK, and AVR values; horizontal lines indicate medians, boxes show interquartile ranges, and whiskers denote data spread. No significant differences were observed (all $p > 0.32$), and all means fell within the ±5° clinical equivalence margin.

Sensitivity to Surgical Inputs

The robustness of the framework to variations in surgical instrumentation was tested in 20 patients (dataset from Article 3 presented in Chapter 6) with 280 simulations exploring clinically relevant perturbations in rod curvature (±10°), fusion levels (UIV/LIV ±1 vertebra), and screw distributions (five patterns, density 1.2–2.0). Results showed distinct effects across parameters (Table 6.4). Rod curvature changes had a limited influence on MT Cobb and AVR but produced deviations of up to 9° in TK, confirming that sagittal alignment is most responsive to rod contouring. Fusion level selection was the most influential factor, driving the most considerable changes in all three alignment metrics and occasionally increasing implant loading by more than 100 N, consistent with its recognized impact on clinical outcomes. Screw distribution had minimal effect on 3D correction but substantially modulated implant forces: high-density constructs increased pullout in most cases, while lower-density strategies reduced loads by nearly 60% without compromising correction. However, recent computational analyses suggest that when screws are well aligned, higher screw densities can actually lower average bone–screw forces, highlighting that construct design and alignment are critical determinants of load distribution [217]. Across these instrumentation variations, the model maintained stable and realistic predictions for both spinal alignment and implant loading, demonstrating reliability under diverse planning conditions. Full results are detailed in *Appendix I - Sensitivity Analysis of Instrumentation*.

Table 6.4 Sensitivity Analysis Results for Surgical Input Variations - 20 AIS Patients, 280 Simulations

Variation tested	MT Cobb (°)	TK (°)	AVR (°)	Screw pullout
Rod curvature (±10°)	±1.6	±9.1	±1.8	Stable
Fusion levels (UIV/LIV ±1)	±14.0	±9.2	±3.1	↑ >100 N in 3 cases
Screw pattern (5 configs)	±3.3	±2.0	±1.8	Higher density ↑ forces in 74% of cases; low-density (1.2) ↓ forces ~59% while maintaining correction

Implications for Credibility

As indicated in Table 6.5, the developed numerical model generally met expectations for a medium-low risk decision-support tool. Moderate scores were achieved for model form and input sensitivities, as several assumptions (e.g., boundary conditions, material properties) were adopted from prior validated models without re-evaluation in this study. The validation datasets used spanned a representative range of clinical cases and surgical strategies, and comparisons were based on quantitative analyses. However, some aspects, such as uncertainty propagation and validation in extreme-case scenarios, were not fully addressed, consistent with the framework's feasibility status. Overall, outcome validation and sensitivity analyses described earlier demonstrated that the model delivers biomechanically realistic, decision-relevant outputs for AIS surgical planning, supporting its credibility at the medium–low risk level defined by ASME V&V 40.

Table 6.5 Goals for Credibility Factors associated with Validation

Credibility Factor	Selected Grade	Justification Summary
		Computational Model
Model Form	b	Model assumptions (e.g., stiffness, boundary conditions) were adopted from earlier validated AIS simulation frameworks [25, 44, 58, 257]. Their influence was not re-evaluated here, but given the analogous surgical context, similar behavior is expected.
Model Inputs: Quantification of Sensitivities	b/c	Sensitivity analyses were performed on selected key inputs, including rod curvature and screw pattern (*Appendix I - Sensitivity Analysis of Instrumentation*).
Model Inputs: Quantification of Uncertainties	b	Expected key uncertainties (e.g., preop curvature, flexibility) were identified and assessed, though propagation to outputs was not fully implemented.
		Comparator
Comparator Quantity of Test Samples	c	An outcome validation study was conducted, involving 35 patients, each with pre- and postoperative measurements of MT Cobb, TK, and AVR, providing statistical validity (*Appendix H - Postoperative Outcome Validation*).
Range of Characteristics of Test Samples	b/c	The cohort represented typical AIS patients (age 13–19, Risser 2–5, Lenke 1–2), spanning MT Cobb angles of 48–68°, TK angles of 8–63°, and AVR angles of –30° to –5°; extreme curve types were not explicitly targeted.
Measurements of Test Samples	c	Key radiographic parameters (MT Cobb, TK, AVR) were extracted pre- and post-operatively.
Uncertainty of Test Sample Measurements	c	Instrument accuracy and repeatability were established in prior studies [137, 258], with reported reconstruction errors of ~1–2 mm for vertebrae and pedicles; no new repetition analysis was performed.
Quantity of Test Conditions	c	More than four construct strategies were evaluated per patient across diverse test conditions.
Range of Test Conditions	b	Simulations covered clinically relevant planning scenarios (UIV T2–T6, LIV T11–L4, 7–13 fused levels, screw density 1.2–2.0, rod curvature 0–40°).
Measurements of Test Conditions	c	Construct variations spanned both standard and extreme surgical scenarios.

Table 6.5 Goals for Credibility Factors associated with Validation (cont'd)

Uncertainty of Test Sample Measurements	c	Instrument accuracy and repeatability are supported by prior validation studies [137, 258], though not re-assessed here.
Assessment		
Equivalency of Input Parameters	c	Patient-specific models matched clinical radiographs; types and ranges of input parameters were equivalent.
Output Comparison: Quantity	b	Multiple outputs (MT Cobb angle, TK, AVR, pullout) were compared between simulation and post-operative results.
Output Comparison: Equivalency of Output Parameters	c	Compared output types were equivalent between simulations and clinical measurements.
Output Comparison: Rigor of Output Comparison	c	Simulation results were compared with postoperative radiographic measurements using quantitative metrics, but uncertainties were not propagated.
Output Comparison: Agreement of Output Comparison	b	An agreement was reached for the majority of comparisons, with deviations below the 5° clinical relevance threshold in most cases, but not all.

6.2.2.7 *Applicability to the COU*

The applicability to the COU domain addresses the alignment between the validation data, outputs, and conditions, and the model's real-world role in clinical decision-making. Rather than reassessing accuracy, this component focuses on whether the types of patients, surgical scenarios, and output metrics used during validation truly represent the setting in which the model is intended to operate. In this work, validation relied on two AIS patient cohorts (n = 35 and n = 20) spanning a representative adolescent clinical population, with diverse curve severities, Lenke types, and sagittal profiles. Surgical conditions tested included a broad range of UIV/LIV levels, rod curvatures, and screw densities, reflecting the planning decisions typically encountered in posterior spinal fusion for AIS. The outputs evaluated (MT Cobb, TK, AVR, implant pullout) correspond directly to clinically relevant decision parameters and surgical goals.

As summarized in Table 6.6, the validation efforts addressed the full scope of the model's intended application, both in terms of output relevance and the diversity of planning conditions covered. This alignment strengthens the translational relevance of the model and supports its potential use in future exploratory clinical studies.

Table 6.6 Credibility Factors Associated with Applicability to the COU

Credibility Factor	Selected Grade	Justification Summary
Relevance of the QOIs	c	The quantities of interest used in validation (e.g., MT Cobb angles, TK, AVR) were the same as those evaluated within the COU.
Relevance of the Validation Activities to the COU	d	Validation activities directly mirrored the COU, with simulations and comparisons covering the full range of surgical constructs and patient anatomies intended for clinical use.

6.2.3 Conclusions on Credibility Assessment

The comprehensive credibility assessment activities support the conclusion that the CM&S framework developed in this thesis meets expectations for a medium-low risk decision-support tool in the context of AIS surgical planning. Verification confirmed correct model implementation and stable solver performance, while validation showed that the model can reproduce credible postoperative spinal alignment across a range of real clinical scenarios. Sensitivity analyses further demonstrated that outputs remained robust under varied surgical inputs. The validation strategy was purposefully aligned with the model's intended use, focusing on clinically relevant outputs and planning decisions. Although some elements, such as discretization error and uncertainty propagation, did not reach the target credibility levels, this is consistent with the framework's feasibility stage. In summary, the model shows adequate levels of credibility for exploratory use in surgical planning research. Further refinements in usability, uncertainty quantification, and prospective validation will be required before clinical implementation is feasible.

CHAPTER 7 GENERAL DISCUSSION

The work presented in this thesis aims to address a longstanding challenge in the surgical management of AIS: the need to more holistically account for 3D spinal deformities and both patient- and instrumentation-specific biomechanical characteristics when planning PSF surgical instrumentation. More specifically, this work set out to articulate and investigate the following research question: *To what extent can a hybrid approach, combining AI, a deterministic 3D biomechanical patient-specific model, and an optimization algorithm, accurately and precisely integrate key PSF parameters (upper and lower instrumented vertebrae, screw density, and rod curvature) to optimize the planning and prediction of AIS correction based on individual patient characteristics?*

This question was explored through a series of complementary studies presented in Chapters 4 through 6. The first part of the work addressed our first objective by reviewing how deep learning has been applied in spine imaging and surgical decision-making. Building on this, an original neural network model was developed to predict surgeon-like instrumentation strategies using preoperative clinical and radiographic data. To meet the second objective, these AI-generated instrumentations were tested *in silico* using a multibody biomechanical model that simulated their effect on spinal alignment and mechanical loading. The third objective was addressed by implementing an optimization method to identify instrumentation configurations that could improve 3D correction while reducing implant use and fusion length, without compromising biomechanical safety. Finally, the fourth objective involved applying the ASME V&V40:2018 framework to verify, validate, and quantify the uncertainty of the computational model, thereby strengthening the credibility of both the model and the framework, and considering potential implications for regulatory readiness in the context of preoperative surgical planning.

The results obtained provide compelling evidence of the proposed framework's proof of concept. The AI model demonstrated high predictive performance (precision and accuracy greater than 85% for most predicted outcomes) for instrumentation parameters, supporting our first hypothesis. When integrated into biomechanical simulations, AI-derived constructs achieved comparable or superior 3D correction compared to surgeon-performed strategies in most cases, particularly in restoring thoracic kyphosis and axial derotation, thereby supporting our second hypothesis.

Optimization further identified instrumentation plans delivering similar or improved 3D correction with fewer implants and shorter fusions, while maintaining comparable implant loading, thus generally supporting our third hypothesis. These collective advances led to four peer-reviewed journal articles and one peer-reviewed conference presentation.

This general discussion reflects on how the results collectively address the initial research questions and objectives, while also considering their broader implications for clinical practice. The chapter begins by examining the potential of the proposed hybrid framework to improve the consistency, efficiency, and personalization of AIS surgical planning. It then offers a critical appraisal of the methodological limitations inherent to the AI and modeling approaches used. Finally, it discusses the translational challenges of progressing from *in silico* validation to clinical implementation, including regulatory, ethical, and data governance considerations that are essential for the safe, transparent, and responsible adoption of computer-model-based decision support systems in spine surgery.

7.1 Implications of Results for Clinical Practice

7.1.1 Bridging the Gaps from Available Tools

The work presented in Chapters 4 and 5 demonstrates how this thesis's findings have surpassed the limitations of existing tools by directly integrating predictive modeling with biomechanical simulation in a unified workflow. While systems like UNiD or EOSsim, discussed in Chapter 2, illustrate current capabilities and constraints, the contribution here lies in bridging those two worlds: generating patient-specific surgical strategies through AI prediction, followed by rigorous evaluation with validated physics-based multibody simulations. This approach enabled the comparison of multiple plausible constructs for the same patient, grounded in both anatomical and biomechanical considerations, rather than relying on static rules or evaluating only a single predefined plan.

By embedding these predictions within a simulation environment, the work expanded the planning process from descriptive to exploratory, allowing the assessment of how variations in construct design could influence alignment, mechanical loading, and other surgical objectives. However, at this stage, the exploration was limited to pre-generated configurations, without dynamic adaptation to competing goals. Addressing this gap, Chapter 6 introduced a multi-objective optimization framework capable of refining AI-predicted strategies according to both clinical and biomechanical priorities. Instead of beginning with a fixed plan, surgeons can now explore how different instrumentation choices perform under varying objectives, such as maximizing 3D correction, minimizing implant burden, or preserving motion segments, thereby transforming the process into an interactive and patient-specific exercise. Results showed that this integrated and optimized approach can produce instrumentation plans that match or exceed expert strategies in both correction quality and biomechanical feasibility, highlighting its potential as a next-generation planning paradigm.

This optimization approach builds on earlier efforts in simulation-based planning for AIS, such as those by Majdouline et al. (2009) and La Barbera et al. (2021), which showed that patient-specific biomechanical models and optimization algorithms can help predict postoperative alignment and assess the effects of surgical parameters such as fusion level and rod stiffness [39, 44, 259]. However, these earlier approaches typically required the manual specification of candidate constructs and did not integrate AI predictive modeling. The framework developed in this thesis advances the field by combining AI-based instrumentation prediction, deterministic biomechanical simulation, and multi-objective optimization in a single workflow. By using AI-predicted constructs as starting points and exploring variations around them, as introduced in Chapter 6, computation time for patient-specific optimization was reduced from over 36 hours per case [44] to less than 6 hours per patient, in hidden computation time on a standard, non-specialized computer.

7.1.2 Opening the Door to Shared Decision-Making

Beyond its technical capabilities, the planning framework developed in Chapter 6 offers the potential for a cultural shift in AIS surgical care, opening the possibility for surgeons, patients, and families to take a more active role in preoperative discussions, an aspect not directly addressed in this thesis but considered a promising perspective for future work. Rather than producing a single, fixed recommendation, the system enables the surgeon to explore multiple biomechanically viable strategies, each with a distinct balance of correction, invasiveness, and implant use. This flexibility makes the trade-offs between surgical options visible, concrete, and discussable, and also opens the door for a planning approach that is both transparent and adaptable. By assigning weights to different surgical objectives, clinicians can tailor plans to reflect not only anatomical correction goals but also broader priorities such as mobility preservation. Rather than dictating a single "best solution", the framework supports the generation of several biomechanically viable options, each reflecting a different balance of trade-offs. This allows surgeons to communicate, for example, how one option may offer greater rotational correction but require a longer fusion, while another might preserve more spinal mobility at the expense of minor aesthetic compromise. In turn, the framework supports a shift from a one-way surgeon decision to a collaborative process, enabling clinicians to explain the biomechanical and correction consequences of different strategies and for patients to voice their preferences. Having such a tool is especially relevant in AIS, where clinical and personal priorities often intersect. While radiographic alignment remains central, many patients and their families care more about posture, rib symmetry, and physical confidence, factors that are not always prioritized in traditional planning frameworks [127, 128]. Although not directly addressed in this thesis, these patient-centered outcomes could be incorporated in future developments, aligning surgical strategies with both anatomical correction and individual values. Ultimately, personalized care means matching constructs not only to anatomy but also to the patient's goals, and the ability to compare and discuss multiple valid, simulation-grounded options moves this vision closer to reality.

7.1.3 Cost and Risk Management

In this thesis, the AI-based strategy presented in Chapter 4 was designed to emulate expert surgical judgment and often generated constructs with fewer implants, without increasing simulated mechanical risk. These early results suggested potential for more cost-efficient planning. The integration of the optimization framework in Chapter 6 strengthened this evidence, showing that surgical burden, both in implant density and fusion length, could be reduced while maintaining or even improving 3D correction and biomechanical integrity. On average, optimized constructs achieved up to 3° greater major curve correction, with 21% fewer pedicle screws (a mean reduction of 4 screws per patient) and a potential of two fewer fused levels compared to surgeon-performed strategies. These findings are notable given that postoperative complications such as implant failure, suboptimal alignment, and revision surgery remain a substantial burden in AIS treatment, with 7–13% of patients requiring reoperation within five years [46, 47]. Many such events are linked to modifiable surgical factors, including fusion length, screw density, and rod configuration, that influence both outcomes and resource use [40, 48-61].

Although no formal cost analysis was performed, the implications are notable. Literature suggests that low-density constructs can reduce implant costs by approximately $1200 per vertebral level compared to high-density strategies [260], with implants accounting for roughly 25% of total AIS hospital costs and screw density being an independent cost predictor [261]. Using a conservative range of $600–$1000 per pedicle screw [172], the observed screw reduction equates to a direct implant cost saving of $2400–$4000 per patient. These results align with projections by Larson et al., who estimated that reducing 3.2 screws per patient could yield $11–$20 million in annual U.S. savings, or 4–7% of national AIS hospitalization costs [172]. Our reduction of 4 screws per patient falls above this range, underscoring the potential economic impact if replicated in clinical practice.

Clinically, shorter constructs are linked to better postoperative mobility, reduced pain, and lower complication rates, as well as shorter surgeries and hospital stays [262-266]. These results challenge the assumption that high-density, “over-instrumented” constructs inherently produce better outcomes, instead suggesting that patient-specific, simulation-informed strategies could reduce both patient burden and system costs. While this study did not experimentally confirm that lower instrumentation burden decreases complication rates, a critical step for future research, the convergence of correction quality, mechanical safety, and reduced surgical burden observed here supports further investigation of optimization as a means to enhance surgical value.

7.2 Methodological Limitations

The specific limitations of the developed models and datasets used have been addressed in Chapters 4 to 6, alongside each technical development (e.g., reliance on a narrow dataset, exclusion of specific clinical decision criteria from AI training, and validation limited to simulations). Here, broader issues are considered, focusing on the generalizability, fairness, and clinical readiness of AI- and simulation-based planning tools. A key limitation concerns the generalizability of the neural network. It was trained on a relatively small, homogeneous cohort from the MIMO trial, primarily female patients with Lenke type 1 curves and limited ethnic diversity. This narrow training base may bias predictions and limit performance in more complex or diverse populations. While biomechanical simulation helps mitigate this risk by allowing construct performance to be tested across various scenarios, the initial strategies remain anchored to the AI-generated baseline outputs, so any embedded bias may persist.

The reliance on retrospective data also presents challenges. Historical surgical strategies often reflect local norms rather than universally accepted, evidence-based standards. As a result, the neural network may encode preferences that lack broad clinical justification, particularly if certain instrumentation styles were overrepresented. Here, the integration of biomechanical simulation and optimization is a crucial countermeasure, enabling clinicians to review and refine AI suggestions rather than accepting them at face value. The dataset's age is another factor: many surgical plans are over a decade old, predating current instrumentation, hardware, and surgical objectives. Comparisons between model- and surgeon-selected constructs may therefore reflect outdated practices. To remain clinically relevant, the framework will need evaluation with recent, prospectively collected cases and periodic retraining to stay aligned with evolving surgical approaches.

Finally, all results in this thesis are derived from biomechanical simulations. While rigorously validated simulations provide robust insights into construct performance, they cannot fully capture intraoperative variability, individual biological responses, or long-term outcomes. Whether simulated advantages, such as reduced implant burden or improved alignment, translate into better patient outcomes remains to be confirmed through prospective clinical studies. These limitations define the pathway from proof of concept to clinical application: expanding and diversifying datasets, updating and refining models, and conducting prospective validation will be essential for the safe and effective integration into surgical practice.

7.3 Clinical Integration Challenges

Despite promising simulation results, translating the proposed planning framework into routine clinical practice entails several practical challenges that go beyond predictive model accuracy. These include technical, institutional, and human factors that must be addressed to enable successful adoption. A key hurdle is workflow integration. Although computational runtime has been reduced significantly, from over 36 hours to under 6 hours per case [44], the process still relies on specialized software, imaging inputs, and computational resources that are not widely available in clinical environments. Additionally, the framework currently operates in a research environment without direct connections to EHRs or PACS. As highlighted in recent reviews, interoperability gaps between clinical systems limit the scalability of AI-based tools [267, 268]. Establishing standardized data formats and secure interfaces with clinical IT systems is a critical first step.

Another challenge is demonstrating clinical utility. While simulation results suggest the potential for reducing implant burden and improving spinal alignment, the actual impact on surgical workflow, decision-making, and patient outcomes in real-world settings remains untested. As seen with many digital health tools, innovations often falter not due to technical performance but from a lack of demonstrated relevance in practice [269]. Surgeon acceptance is also essential. The framework may propose construct strategies that depart from conventional norms in terms of fusion length or implant density, challenging ingrained habits. Given that surgical planning often rests on anatomical classification and years of tacit experience [6, 18], any AI- and simulation-based system must be accurate, explainable, and intuitive. Surgeons need to understand not only the recommendation but also the reasoning behind it. There is also an educational gap. Few surgeons are trained to interpret the outputs of machine learning or biomechanical simulations. This lack of knowledge can lead to either uncritical acceptance of algorithmic suggestions or unwarranted skepticism. Early engagement, user training, and emphasis on interpretability have been highlighted as key to the adoption of such numerical models [268, 269].

Nevertheless, a notable strength of this thesis lies in the interdisciplinary nature of the development team, which involves engineers, spine surgeons, and veterinary scientists, ensuring clinical relevance and responsiveness to user needs. Scaling implementation will require extending this engagement to IT staff, administrators, and patients. Experiences from prior digital health projects demonstrate that insufficient stakeholder involvement during development often results in a poor fit and limited uptake [270, 271]. In summary, the future success of this planning system will depend not only on its technical merits but also on thoughtful clinical integration, a clear demonstration of added value, and continuous co-development with end-users. Without these elements, the framework risks remaining a research tool rather than becoming part of routine surgical practice.

7.4 Policy and Ethical Considerations

7.4.1 Developing Policies and Guidelines

The clinical adoption of decision-support technologies depends not only on performance and usability, but also on regulatory credibility and international alignment. The planning framework developed in this thesis, which combines AI prediction, biomechanical simulation, and multi-objective optimization, falls between multiple regulatory categories. Under Health Canada's current SaMD guidance, it meets all four exclusion criteria and, consequently, is not subject to medical device licensing [255], though such a designation may become necessary as it nears clinical use. Nevertheless, broader regulatory challenges remain. While standards such as ASME V&V 40 guide the validation of static computational models, they do not fully address adaptive, data-driven tools incorporating machine learning. This has practical implications: unlike deterministic systems, the planning framework enables surgeons to explore multiple options shaped by trade-offs among correction, implant economy, and mechanical risk. Such flexibility, while clinically valuable, complicates the definition of model validity, especially if regulations focus narrowly on matching a predefined ground truth. For exploratory decision-support tools, evaluation must consider both clinical utility and accuracy. Transparency is equally important. Although biomechanical simulation enhances interpretability compared to black-box AI, the system still lacks a straightforward explanation for why specific strategies are favored. Results show that varying optimization weights can yield distinct, yet mechanically sound constructs. To support shared decision-making and build trust, clearer standards are needed to define what qualifies as an "explainable" surgical plan. No such regulatory benchmarks currently exist, representing an important challenge for future work.

To address these gaps proactively, VVUQ activities were conducted in accordance with the ASME V&V 10 and 40 frameworks, as detailed in Chapter 6. These efforts provide a structured foundation for model credibility and position the developed framework for eventual clinical integration. The collaborative development between Canada and the U.S. illustrates the value of aligning with international best practices to ensure scientific rigor and cross-border applicability. In summary, this thesis presents an advanced hybrid model that highlights the need for updated guidelines that align with the rapid evolution of surgical AI and facilitate its responsive clinical application. While regulatory gaps may slow adoption, this work demonstrates that such approaches are both feasible and relevant, and that closing these gaps is essential for safely integrating next-generation decision-support tools into surgical practice.

7.4.2 Importance of Patient Consent and Data Privacy

This thesis used anonymized, pre-consented data from the MIMO clinical trial. However, broader clinical deployment will require consent processes that are more explicit, transparent, and informative. Patients should understand not only that algorithmic tools may contribute to surgical planning, but also how these systems operate and influence decision-making. AI in healthcare raises legitimate concerns about data privacy, secondary use, and control of ownership [270]. Many patients are willing to share data to improve care, but may object to its use in developing commercial tools. This distinction, between academic research and commercial deployment, is ethically and socially significant. As the system evolves toward clinical application, transparent consent and clear boundaries about data use, ownership, and benefit sharing will be critical. Technical measures, such as federated learning or synthetic datasets, can help mitigate privacy risks while preserving performance; however, they cannot replace the need for transparent governance and accountability.

Accountability is another open issue. If a model-informed plan contributes to a poor outcome, legal responsibility will remain with the surgeon, yet ethical responsibility should be shared. Developers, researchers, and institutions involved in building and promoting such tools must be transparent and accountable. In this work, all experiments were conducted within a research lab certified under ISO 13485 standards, ensuring traceability, risk management, and regulatory compliance. Maintaining rigorous quality systems will be vital as the tool transitions to clinical use, not only to meet regulatory requirements, but also to demonstrate a sustained commitment to patient safety, transparency, and shared responsibility.

CHAPTER 8 CONCLUSION AND RECOMMENDATIONS

This thesis directly addressed its central research question: whether a hybrid approach combining AI, deterministic 3D biomechanical modeling, and optimization can accurately and precisely integrate key posterior spinal fusion parameters (upper and lower instrumented vertebrae, screw density, and rod curvature) to plan and predict AIS correction based on patient-specific factors. Traditional surgical planning often relies heavily on surgeon experience and heuristic, rule-based classifications, which may overlook the biomechanical complexity and variability of AIS. The original framework developed in this thesis supports a more systematic and patient-specific planning process grounded in computational modeling.

The first objective tested AI's ability to replicate surgeon planning decisions using only preoperative data. The trained neural network predicted instrumentation parameters with over 85% accuracy, demonstrating that data-driven planning is feasible. Patient-specific biomechanical simulations confirmed that AI-derived constructs achieved comparable or superior 3D corrections to those planned by surgeons, particularly in sagittal alignment and vertebral derotation. However, coronal correction was slightly lower in some cases. Adding the optimization step further improved performance: optimized strategies used on average 21% fewer screws and up to two fewer fused levels, while maintaining correction goals and often outperforming surgeon plans in 3D metrics. Implant loading did not decrease, indicating room for further refinement.

These findings demonstrate that the hybrid approach can both replicate expert planning and enhance it through systematic exploration of biomechanically viable alternatives. The proposed framework bridges data-driven prediction and biomechanical reasoning, two approaches often viewed as distinct, into a single, clinically meaningful process, enabling evaluation of multiple tailored strategies beyond static, rule-based paradigms. Methodologically, it unites validation, simulation, and multi-objective reasoning into a potentially translatable pipeline. Nonetheless, reliance on a retrospective, relatively homogeneous dataset limits generalizability, and all results remain *in silico*. While simulations were comprehensively validated in accordance with ASME V&V 40 standards, prospective clinical trials will be necessary to confirm whether the observed benefits translate to real-world outcomes. Clinical adoption will also depend on usability, surgeon trust, and workflow integration. This work also raises regulatory and ethical considerations. As a hybrid AI-simulation decision-support tool, it occupies a regulatory gray area. Responsible deployment will require addressing algorithmic bias, strong data governance, and clear definitions of shared responsibility between clinicians and developers.

Beyond its technical achievements, this thesis contributes a novel conceptual and methodological framework for integrating AI with comprehensively validated biomechanical simulation to support patient-specific surgical planning. The ability to fuse data-driven predictions with biomechanical reasoning and to optimize instrumentation through multi-objective strategies opens a new paradigm in digital health tools for spine surgery. This hybrid approach moves beyond static planning models toward an adaptive, simulation-based process that can evolve in response to patient needs, surgeon priorities, and data availability. In doing so, this work addresses critical gaps in current clinical workflows by offering a reproducible, explainable, and regulation-conscious pipeline that can be embedded into shared decision-making practices. The framework's modularity also makes it adaptable to other spinal conditions or surgical techniques, offering a solid foundation for broader clinical translation and future innovation in the sector.

To ensure successful clinical adoption and societal benefit, several key recommendations emerge from this work:

i. Prospective Clinical Validation: A multi-center, prospective study would be needed to evaluate the real-world impact of the proposed framework on surgical outcomes, resource utilization (e.g., fusion levels, implant density), and patient-reported outcomes. This will allow for quantification of clinical benefit and usability across diverse care settings.

ii. Human Factors and Integration in Workflow: To build surgeon trust and promote effective use, the interface should be co-designed with end-users and integrated into standard preoperative planning workflows. Usability testing and iterative improvements must ensure that the tool supports, rather than disrupts, clinical decision-making.

iii. Regulatory and Ethical Readiness: Given its nature as a hybrid AI-simulation device, the framework should be advanced toward regulatory clearance through continued compliance with ASME V&V 40 and ISO 13485 practices. In parallel, clear governance structures must be established to address accountability, data privacy, and explainability in clinical deployment.

In conclusion, this thesis provides a comprehensive proof-of-concept that answers its research question: a hybrid framework can integrate AI-based prediction with deterministic, patient-specific biomechanical modeling to optimize key PSF parameters and support individualized, efficient, and biomechanically grounded planning strategies. Although further validation, regulatory adaptation, and clinical integration are needed, the results indicate a strong foundation for next-generation tools that can assist spine surgeons and AIS patients in navigating complex surgical decisions. With continued interdisciplinary collaboration and refinement, the proposed framework could evolve into a practical, widely adoptable solution.

REFERENCES

[1] J. Dunn *et al.*, "Screening for Adolescent Idiopathic Scoliosis: Evidence Report and Systematic Review for the US Preventive Services Task Force," vol. 319, n°. 2, p. 173-187, Jan 9 2018.

[2] S. L. Weinstein *et al.*, "Effects of bracing in adolescents with idiopathic scoliosis," vol. 369, n°. 16, p. 1512-1521, 2013.

[3] J. J. Thomas *et al.*, "Does school screening affect scoliosis curve magnitude at presentation to a pediatric orthopedic clinic?," vol. 6, n°. 4, p. 403-408, 2018.

[4] S. Negrini *et al.*, "Actual evidence in the medical approach to adolescents with idiopathic scoliosis," 2014.

[5] J. Cobb, "Outline for the study of scoliosis," vol. 5, p. 261-275, 1948.

[6] H. A. King *et al.*, "The selection of fusion levels in thoracic idiopathic scoliosis," vol. 65, n°. 9, p. 1302-13, Dec 1983.

[7] N. Chung *et al.*, "Spinal phantom comparability study of Cobb angle measurement of scoliosis using digital radiographic imaging," vol. 15, p. 81-90, Oct 2018. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S2214031X18300974?via%3Dihub

[8] J. Shen *et al.*, "Towards a new 3D classification for adolescent idiopathic scoliosis," vol. 8, n°. 3, p. 387-396, Jun 2020. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s43390-020-00051-2.pdf

[9] L. G. Lenke *et al.*, "Rationale behind the current state-of-the-art treatment of scoliosis (in the pedicle screw era)," vol. 33, n°. 10, p. 1051-4, May 1 2008.

[10] T. B. Sullivan *et al.*, "More severe thoracic idiopathic scoliosis is associated with a greater three-dimensional loss of thoracic kyphosis," vol. 8, n°. 6, p. 1205-1211, Dec 2020. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s43390-020-00149-7.pdf

[11] T. B. Sullivan *et al.*, "Thoracic Idiopathic Scoliosis Severity Is Highly Correlated with 3D Measures of Thoracic Kyphosis," vol. 99, n°. 11, p. e54, Jun 7 2017.

[12] S. L. Weinstein *et al.*, "Health and function of patients with untreated idiopathic scoliosis: a 50-year natural history study," vol. 289, n°. 5, p. 559-67, Feb 5 2003. [En ligne]. Disponible: https://jamanetwork.com/journals/jama/articlepdf/195893/joc21444_559_567_1610576436.54376.pdf

[13] S. L. Weinstein *et al.*, "Adolescent idiopathic scoliosis," vol. 371, n°. 9623, p. 1527-1537, 2008, 1527. [En ligne]. Disponible: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(08)60658-3/fulltext

[14] J. C. Cheng *et al.*, "Adolescent idiopathic scoliosis," vol. 1, n°. 1, p. 15030, 2015/09/24 2015. [En ligne]. Disponible: https://doi.org/10.1038/nrdp.2015.30

https://www.nature.com/articles/nrdp201530

[15] W.-W. Chau, B. K.-W. Ng et A. L.-H. Hung, "Health-related quality of life (HRQOL) of adolescent idiopathic scoliosis (AIS) patients from surgery to after 30 years using SRS-22 questionnaire," vol. 8, nº. 5, p. 951-956, 2020/10/01 2020. [En ligne]. Disponible: https://doi.org/10.1007/s43390-020-00132-2

https://link.springer.com/content/pdf/10.1007/s43390-020-00132-2.pdf

[16] S.-S. Kim, J.-H. Kim et S.-I. Suk, "Effect of Direct Vertebral Rotation on the Uninstrumented Lumbar Curve in Thoracic Adolescent Idiopathic Scoliosis," vol. 11, nº. 1, p. 127-137, 2 2017. [En ligne]. Disponible: https://doi.org/10.4184/asj.2017.11.1.127

http://www.asianspinejournal.org/journal/view.php?number=192

https://www.asianspinejournal.org/upload/pdf/asj-11-127.pdf

[17] M. de Kleuver *et al.*, "Optimal surgical care for adolescent idiopathic scoliosis: an international consensus," vol. 23, nº. 12, p. 2603-2618, 2014/12/01 2014. [En ligne]. Disponible: https://doi.org/10.1007/s00586-014-3356-1

https://link.springer.com/content/pdf/10.1007/s00586-014-3356-1.pdf

[18] L. G. Lenke *et al.*, "Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis," vol. 83, nº. 8, p. 1169-81, Aug 2001.

[19] S. D. Glassman *et al.*, "The impact of positive sagittal balance in adult spinal deformity," vol. 30, nº. 18, p. 2024-9, Sep 15 2005.

[20] F. Schwab *et al.*, "Adult spinal deformity-postoperative standing imbalance: how much can you tolerate? An overview of key parameters in assessing alignment and planning corrective surgery," vol. 35, nº. 25, p. 2224-31, Dec 1 2010.

[21] U. I. Udoekwere *et al.*, "Effect of Lowest Instrumented Vertebra on Trunk Mobility in Patients With Adolescent Idiopathic Scoliosis Undergoing a Posterior Spinal Fusion," vol. 2, nº. 4, p. 291-300, 2014/07/01/ 2014. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S2212134X1400063X

[22] M. Ohashi *et al.*, "The Benefits of Sparing Lumbar Motion Segments in Spinal Fusion for Adolescent Idiopathic Scoliosis Are Evident at 10 Years Postoperatively," vol. 45, nº. 11, p. 755-763, 2020. [En ligne]. Disponible: https://journals.lww.com/spinejournal/Fulltext/2020/06010/The_Benefits_of_Sparing_Lumbar_Motion_Segments_in.15.aspx

[23] S. I. Suk *et al.*, "Selective thoracic fusion with segmental pedicle screw fixation in the treatment of thoracic idiopathic scoliosis: more than 5-year follow-up," vol. 30, nº. 14, p. 1602-9, Jul 15 2005.

[24] R. H. Cho *et al.*, "Which Lenke 1A curves are at the greatest risk for adding-on... and why?," vol. 37, nº. 16, p. 1384-90, Jul 15 2012.

[25] F. Le Navéaux *et al.*, "How does implant distribution affect 3D correction and bone-screw forces in thoracic adolescent idiopathic scoliosis spinal instrumentation?," vol. 39, p. 25-

31, Nov 2016. [En ligne]. Disponible: https://www.clinbiomech.com/article/S0268-0033(16)30130-9/fulltext

[26] K. Bagchi *et al.*, "Hardware complications in scoliosis surgery," vol. 32, nº. 7, p. 465-75, Jul 2002. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00247-002-0659-x.pdf

[27] H. W. Albers *et al.*, "Comparison of single- and dual-rod techniques for posterior spinal instrumentation in the treatment of adolescent idiopathic scoliosis," vol. 25, nº. 15, p. 1944-9, Aug 1 2000.

[28] J. M. Wattenbarger, B. S. Richards et J. A. Herring, "A comparison of single-rod instrumentation with double-rod instrumentation in adolescent idiopathic scoliosis," vol. 25, nº. 13, p. 1680-8, Jul 1 2000.

[29] A. Pan *et al.*, "Adjacent segment degeneration after lumbar spinal fusion compared with motion-preservation procedures: a meta-analysis," vol. 25, nº. 5, p. 1522-1532, May 2016. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00586-016-4415-6.pdf

[30] S. Pasha et K. Baldwin, "Surgical outcome differences between the 3D subtypes of right thoracic adolescent idiopathic scoliosis," vol. 28, nº. 12, p. 3076-3084, Dec 2019. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00586-019-06145-4.pdf

[31] S. Pasha *et al.*, "A hierarchical classification of adolescent idiopathic scoliosis: Identifying the distinguishing features in 3D spinal deformities," vol. 14, nº. 3, p. e0213406, 2019. [En ligne]. Disponible: https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0213406&type=printable

[32] P. O. Newton *et al.*, "Defining the "Three-Dimensional Sagittal Plane" in Thoracic Adolescent Idiopathic Scoliosis," vol. 97, nº. 20, p. 1694-701, Oct 21 2015.

[33] S. D. Glassman *et al.*, "Correlation of radiographic parameters and clinical symptoms in adult scoliosis," vol. 30, nº. 6, p. 682-8, Mar 15 2005.

[34] F. Schwab *et al.*, "A clinical impact classification of scoliosis in the adult," vol. 31, nº. 18, p. 2109-14, Aug 15 2006.

[35] D. Hernandez *et al.*, "Computer-assisted Orthopaedic Surgery," vol. 9, nº. 2, p. 152-158, 2017. [En ligne]. Disponible: https://onlinelibrary.wiley.com/doi/abs/10.1111/os.12323

https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/os.12323?download=true

[36] E. Floyd *et al.*, "A review of preoperative planning technologies for spinal deformity correction," vol. 32, nº. 2, p. 100787, 2020/06/01/ 2020. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S1040738320300046

[37] E. Ferrero *et al.*, "Preliminary experience with SpineEOS, a new software for 3D planning in AIS surgery," vol. 27, nº. 9, p. 2165-2174, Sep 2018. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00586-018-5591-3.pdf

[38] L. La Barbera, A. N. Larson et C. E. Aubin, "How do spine instrumentation parameters influence the 3D correction of thoracic adolescent idiopathic scoliosis? A patient-specific biomechanical study," vol. 84, p. 105346, Apr 2021. [En ligne]. Disponible: https://www.clinbiomech.com/article/S0268-0033(21)00076-0/fulltext

[39] L. La Barbera, A. N. Larson et C. E. Aubin, "Correction objectives have higher impact than screw pattern and density on the optimal 3D correction of thoracic AIS: a biomechanical study," vol. 9, nº. 3, p. 655-664, May 2021. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/33501602

[40] A. Delikaris *et al.*, "Implant Density at the Apex Is More Important Than Overall Implant Density for 3D Correction in Thoracic Adolescent Idiopathic Scoliosis Using Rod Derotation and En Bloc Vertebral Derotation Technique," vol. 43, nº. 11, p. E639-e647, Jun 1 2018.

[41] C. E. Aubin *et al.*, "Biomechanical modeling of posterior instrumentation of the scoliotic spine," vol. 6, nº. 1, p. 27-32, Feb 2003. [En ligne]. Disponible: https://www.tandfonline.com/doi/abs/10.1080/1025584031000072237

[42] M. Lopez Poncelas *et al.*, "Credibility assessment of patient-specific biomechanical models to investigate proximal junctional failure in clinical cases with adult spine deformity using ASME V&V40 standard," p. 1-11, Aug 24 2021. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/34427119

[43] C. E. Aubin *et al.*, "Preoperative planning simulator for spinal deformity surgeries," vol. 33, nº. 20, p. 2143-52, Sep 15 2008.

[44] L. La Barbera *et al.*, "In silico patient-specific optimization of correction strategies for thoracic adolescent idiopathic scoliosis," vol. 81, p. 105200, 2021/01/01/ 2021. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S0268003320303193

[45] M. Chang *et al.*, "The Role of Machine Learning in Spine Surgery: The Future Is Now," vol. 7, nº. 54, 2020-August-21 2020. [En ligne]. Disponible: https://www.frontiersin.org/article/10.3389/fsurg.2020.00054

[46] H. T. Vigneswaran *et al.*, "Surgical treatment of adolescent idiopathic scoliosis in the United States from 1997 to 2012: an analysis of 20, 346 patients," vol. 16, nº. 3, p. 322-328, 2015.

[47] B. A. Ramo et B. S. Richards, "Repeat Surgical Interventions Following “Definitive” Instrumentation and Fusion for Idiopathic Scoliosis: Five-Year Update on a Previously Published Cohort," vol. 37, nº. 14, p. 1211-1217, 2012. [En ligne]. Disponible: https://journals.lww.com/spinejournal/Fulltext/2012/06150/Repeat_Surgical_Interventions_Following.5.aspx

[48] J. C. Dick et C. A. Bourgeault, "Notch sensitivity of titanium alloy, commercially pure titanium, and stainless steel spinal implants," vol. 26, nº. 15, p. 1668-72, Aug 1 2001.

[49] T. Q. Nguyen *et al.*, "The fatigue life of contoured cobalt chrome posterior spinal fusion rods," vol. 225, nº. 2, p. 194-8, Feb 2011. [En ligne]. Disponible: https://journals.sagepub.com/doi/10.1243/09544119JEIM763

[50] C. E. Johnston, 2nd *et al.*, "Mechanical consequences of rod contouring and residual scoliosis in sublaminar segmental instrumentation," vol. 5, n°. 2, p. 206-16, 1987. [En ligne]. Disponible: https://onlinelibrary.wiley.com/doi/10.1002/jor.1100050207

[51] S. Pasha, S. Shah et P. Newton, "Machine Learning Predicts the 3D Outcomes of Adolescent Idiopathic Scoliosis Surgery Using Patient-Surgeon Specific Parameters," vol. 46, n°. 9, p. 579-587, May 1 2021.

[52] D. L. Skaggs *et al.*, "Assessment of Lowest Instrumented Vertebra Tilt on Radiographic Measurements in Lenke "C" Modifier Curves Undergoing Selective Thoracic Fusion in Adolescent Idiopathic Scoliosis," vol. 4, n°. 2, p. 125-130, 2016/03/01/ 2016. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S2212134X15002270

[53] S. Pasha et J. M. Mac-Thiong, "Defining criteria for optimal lumbar curve correction following the selective thoracic fusion surgery in Lenke 1 adolescent idiopathic scoliosis: developing a decision tree," vol. 30, n°. 3, p. 513-522, Apr 2020. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/31760495

https://link.springer.com/content/pdf/10.1007/s00590-019-02596-z.pdf

[54] S. Pasha *et al.*, "Relationships Between the Axial Derotation of the Lower Instrumented Vertebra and Uninstrumented Lumbar Curve Correction: Radiographic Outcome in Lenke 1 Adolescent Idiopathic Scoliosis With a Minimum 2-Year Follow-up," vol. 38, n°. 4, 2018. [En ligne]. Disponible: https://journals.lww.com/pedorthopaedics/Fulltext/2018/04000/Relationships_Between_the_Axial_Derotation_of_the.7.aspx

[55] J. Schulz *et al.*, "Optimal radiographical criteria after selective thoracic fusion for patients with adolescent idiopathic scoliosis with a C lumbar modifier: does adherence to current guidelines predict success?," vol. 39, n°. 23, p. E1368-73, Nov 1 2014.

[56] J. H. Cho *et al.*, "Disc Wedge and Vertebral Body Tilt Angle Below Lower Instrumented Vertebra After Posterior Correction and Fusion in Patients With a Structural Thoracolumbar/Lumbar Curve: A Minimum 5-Year Follow-up," vol. 44, n°. 24, p. E1436-e1442, Dec 15 2019.

[57] X. Qin *et al.*, "Selecting the Last Substantially Touching Vertebra as Lowest Instrumented Vertebra in Lenke type 2A-R and 2A-L Curves," vol. 45, n°. 5, p. 309-318, Mar 1 2020.

[58] X. Wang *et al.*, "Biomechanical effect of pedicle screw distribution in AIS instrumentation using a segmental translation technique: computer modeling and simulation," vol. 12, p. 13, 2017. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/28428982

https://scoliosisjournal.biomedcentral.com/track/pdf/10.1186/s13013-017-0120-4.pdf

[59] X. Wang *et al.*, "Correction Capability in the 3 Anatomic Planes of Different Pedicle Screw Designs in Scoliosis Instrumentation," vol. 30, n°. 4, p. E323-e330, May 2017.

[60] S. Seki *et al.*, "Differential Rod Contouring is Essential for Improving Vertebral Rotation in Patients With Adolescent Idiopathic Scoliosis: Thoracic Curves Assessed With Intraoperative CT," vol. 43, n°. 10, p. E585-e591, May 15 2018.

[61] F. Solla *et al.*, "Patient-specific rods for thoracic kyphosis correction in adolescent idiopathic scoliosis surgery: Preliminary results," vol. 106, nº. 1, p. 159-165, Feb 2020. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/31757655

https://www.sciencedirect.com/science/article/pii/S1877056819303494?via%3Dihub

[62] M. A. Edgar et M. H. Mehta, "Long-term follow-up of fused and unfused idiopathic scoliosis," vol. 70, nº. 5, p. 712-6, Nov 1988.

[63] A. Moskowitz *et al.*, "Long-term follow-up of scoliosis fusion," vol. 62, nº. 3, p. 364-76, Apr 1980.

[64] M. S. Youn *et al.*, "Relationship between cervical sagittal alignment and health-related quality of life in adolescent idiopathic scoliosis," vol. 25, nº. 10, p. 3114-3119, Oct 2016. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/26963762

https://link.springer.com/content/pdf/10.1007/s00586-016-4488-2.pdf

[65] B. S. Lonner *et al.*, "Multivariate analysis of factors associated with kyphosis maintenance in adolescent idiopathic scoliosis," vol. 37, nº. 15, p. 1297-302, Jul 1 2012. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/22228329

[66] S. W. Hwang *et al.*, "Cervical sagittal plane decompensation after surgery for adolescent idiopathic scoliosis: an effect imparted by postoperative thoracic hypokyphosis," vol. 15, nº. 5, p. 491-6, Nov 2011. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/21800955

[67] J. S. Smith *et al.*, "Assessment of symptomatic rod fracture after posterior instrumented fusion for adult spinal deformity," vol. 71, nº. 4, p. 862-7, Oct 2012.

[68] C. E. Johnston *et al.*, "Correlation of Preop Curve Severity with Pulmonary Function Tests (PFTs) in AIS: Paper #98," vol. 10, p. 122-123, 2009. [En ligne]. Disponible: https://journals.lww.com/spinejournalabstracts/Fulltext/2009/10003/Correlation_of_Preop_Curve_Severity_with_Pulmonary.98.aspx

[69] R. B. Winter, W. W. Lovell et J. H. Moe, "Excessive thoracic lordosis and loss of pulmonary function in patients with idiopathic scoliosis," vol. 57, nº. 7, p. 972-7, Oct 1975.

[70] S. Pasha *et al.*, "Timing of Changes in Three-Dimensional Spinal Parameters After Selective Thoracic Fusion in Lenke 1 Adolescent Idiopathic Scoliosis: Two-Year Follow-up," vol. 5, nº. 6, p. 409-415, Nov 2017.

[71] J. Takahashi *et al.*, "Selective thoracic fusion in adolescent idiopathic scoliosis: factors influencing the selection of the optimal lowest instrumented vertebra," vol. 36, nº. 14, p. 1131-41, Jun 15 2011.

[72] D. H. Clements *et al.*, "Correlation of scoliosis curve correction with the number and type of fixation anchors," vol. 34, nº. 20, p. 2147-50, Sep 15 2009.

[73] J. E. Lowenstein *et al.*, "Coronal and sagittal plane correction in adolescent idiopathic scoliosis: a comparison between all pedicle screw versus hybrid thoracic hook lumbar screw constructs," vol. 32, nº. 4, p. 448-52, Feb 15 2007.

[74] E. Ofiram *et al.*, "Cervical degenerative changes in idiopathic scoliosis patients who underwent long fusion to the sacrum as adults: incidence, severity, and evolution," vol. 10, nº. 1, p. 27-30, Mar 2009. [En ligne]. Disponible: https://jorthoptraumatol.springeropen.com/track/pdf/10.1007/s10195-008-0044-0.pdf

[75] D. H. Clements *et al.*, "Did the Lenke classification change scoliosis treatment?," vol. 36, nº. 14, p. 1142-5, Jun 15 2011.

[76] F. Galbusera, "Chapter 1 - The Spine: Its Evolution, Function, and Shape," dans *Biomechanics of the Spine*, F. Galbusera et H.-J. Wilke, Édit.: Academic Press, 2018, p. 3-9.

[77] M. P. Steinmetz, T. E. Mroz et J. C. Wang, *AO Spine Textbook: Comprehensive Overview on Surgical Management of the Spine*: Jaypee Brothers Medical Publishers Pvt. Limited, 2020.

[78] J. C. Le Huec *et al.*, "Equilibrium of the human body and the gravity line: the basics," vol. 20, nº. 5, p. 558, 2011/08/02 2011. [En ligne]. Disponible: https://doi.org/10.1007/s00586-011-1939-7

[79] M. E. Kunkel *et al.*, "Morphometric analysis of the relationships between intervertebral disc and vertebral body heights: an anatomical and radiographic study of the human thoracic spine," vol. 219, nº. 3, p. 375-387, 2011. [En ligne]. Disponible: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-7580.2011.01397.x

[80] M. M. PANJABI *et al.*, "Thoracic Human Vertebrae Quantitative Three-Dimensional Anatomy," vol. 16, nº. 8, p. 888-901, 1991. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/1991/08000/thoracic_human_vertebrae_quantitative.6.aspx

[81] A. A. White et M. M. Panjabi, *Clinical Biomechanics of the Spine*: Lippincott, 1990.

[82] F. Galbusera et H.-J. Wilke, *Biomechanics of the spine: Basic concepts, spinal disorders and treatments*: Academic Press, 2018.

[83] M. Boakye *et al.*, *Biomechanics of Spine Stabilization*: Thieme Medical Publishers, 2025.

[84] I. Busscher *et al.*, "Biomechanical Characteristics of Different Regions of the Human Spine: An: In Vitro: Study on Multilevel Spinal Segments," vol. 34, nº. 26, p. 2858-2864, 2009. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2009/12150/biomechanical_characteristics_of_different_regions.6.aspx

[85] M. M. Panjabi, R. A. Brand, Jr. et A. A. White, 3rd, "Three-dimensional flexibility and stiffness properties of the human thoracic spine," vol. 9, nº. 4, p. 185-92, 1976.

[86] C. Liebsch et H.-J. Wilke, "Chapter 3 - Basic Biomechanics of the Thoracic Spine and Rib Cage," dans *Biomechanics of the Spine*, F. Galbusera et H.-J. Wilke, Édit.: Academic Press, 2018, p. 35-50.

[87] P. Roussouly *et al.*, "Classification of the Normal Variation in the Sagittal Alignment of the Human Lumbar Spine and Pelvis in the Standing Position," vol. 30, nº. 3, p. 346-353,

2005. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2005/02010/classification_of_the_normal_variation_in_the.16.aspx

[88] J. J. W. Boyle, N. Milne et K. P. Singer, "Influence of age on cervicothoracic spinal curvature: An ex vivo radiographic survey," vol. 17, nº. 5, p. 361-367, 2002/06/01/ 2002. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S026800330200030X

[89] E. M. Mannen *et al.*, "Mechanical Contribution of the Rib Cage in the Human Cadaveric Thoracic Spine," vol. 40, nº. 13, p. E760-E766, 2015. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2015/07010/mechanical_contribution_of_the_rib_cage_in_the.6.aspx

[90] R. I. Watkins *et al.*, "Stability Provided by the Sternum and Rib Cage In the Thoracic Spine," vol. 30, nº. 11, p. 1283-1286, 2005. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2005/06010/stability_provided_by_the_sternum_and_rib_cage_in.10.aspx

[91] K. Singer *et al.*, "Prediction of thoracic and lumbar vertebral body compressive strength: correlations with bone mineral density and vertebral region," vol. 17, nº. 2, p. 167-174, 1995/08/01/ 1995. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S8756328295001654

[92] N. Yoganandan *et al.*, "Biomechanical Effects of Laminectomy on Thoracic Spine Stability," vol. 32, nº. 4, p. 604-610, 1993. [En ligne]. Disponible: https://journals.lww.com/neurosurgery/fulltext/1993/04000/biomechanical_effects_of_laminectomy_on_thoracic.17.aspx

[93] J.-W. M. Kouwenhoven *et al.*, "Effects of Dorsal Versus Ventral Shear Loads on the Rotational Stability of the Thoracic Spine: A Biomechanical Porcine and Human Cadaveric Study," vol. 32, nº. 23, p. 2545-2550, 2007. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2007/11010/effects_of_dorsal_versus_ventral_shear_loads_on.6.aspx

[94] R. M. Castelein, "Pre-existent rotation of the normal spine at different ages and its consequences for the scoliotic mechanism," dans *Research into Spinal Deformities 8*: IOS Press, 2012, p. 20-25.

[95] M. Janssen, J.-W. M. Kouwenhoven et R. M. Castelein, "The role of posteriorly directed shear loads acting on a pre-rotated growing spine: a hypothesis on the pathogenesis of idiopathic scoliosis," dans *Research into Spinal Deformities 7*: IOS Press, 2010, p. 112-117.

[96] M. M. A. Janssen *et al.*, "Differences in Male and Female Spino-Pelvic Alignment in Asymptomatic Young Adults: A Three-Dimensional Analysis Using Upright Low-Dose Digital Biplanar X-rays," vol. 34, nº. 23, p. E826-E832, 2009. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2009/11010/differences_in_male_and_female_spino_pelvic.26.aspx

[97] S. KLEINBERG, "THE OPERATIVE TREATMENT OF SCOLIOSIS," vol. 5, nº. 3, p. 631-645, 1922. [En ligne]. Disponible: https://doi.org/10.1001/archsurg.1922.01110150184008

[98] R. G. Burwell *et al.*, "Adolescent idiopathic scoliosis (AIS): a multifactorial cascade concept for pathogenesis and embryonic origin," vol. 11, nº. 1, p. 8, 2016/01/30 2016. [En ligne]. Disponible: https://doi.org/10.1186/s13013-016-0063-1

[99] G. Cheh *et al.*, "The reliability of preoperative supine radiographs to predict the amount of curve flexibility in adolescent idiopathic scoliosis," vol. 32, nº. 24, p. 2668-72, Nov 15 2007.

[100] K. D. Luk *et al.*, "Assessment of scoliosis correction in relation to flexibility using the fulcrum bending correction index," vol. 23, nº. 21, p. 2303-7, Nov 1 1998.

[101] B. J. Davis *et al.*, "Traction radiography performed under general anesthetic: a new technique for assessing idiopathic scoliosis curves," vol. 29, nº. 21, p. 2466-70, Nov 1 2004.

[102] P. L. Khong *et al.*, "ICRP publication 121: radiological protection in paediatric diagnostic and interventional radiology," vol. 42, nº. 2, p. 1-63, Apr 2013.

[103] B. Ilharreborde *et al.*, "EOS microdose protocol for the radiological follow-up of adolescent idiopathic scoliosis," vol. 25, nº. 2, p. 526-31, Feb 2016.

[104] B. Garg *et al.*, "Outcome and safety analysis of 3D-printed patient-specific pedicle screw jigs for complex spinal deformities: a comparative study," vol. 19, nº. 1, p. 56-64, Jan 2019.

[105] B. S. Richards *et al.*, "Right thoracic curves in presumed adolescent idiopathic scoliosis: which clinical and radiographic findings correlate with a preoperative abnormal magnetic resonance image?," vol. 35, nº. 20, p. 1855-60, Sep 15 2010.

[106] J. Qiao *et al.*, "Indication for preoperative MRI of neural axis abnormalities in patients with presumed thoracolumbar/lumbar idiopathic scoliosis," vol. 22, nº. 2, p. 360-6, Feb 2013.

[107] M. Inoue *et al.*, "Preoperative MRI analysis of patients with idiopathic scoliosis: a prospective study," vol. 30, nº. 1, p. 108-14, Jan 1 2005.

[108] P. Knott *et al.*, "Multicenter Comparison of 3D Spinal Measurements Using Surface Topography With Those From Conventional Radiography," vol. 4, nº. 2, p. 98-103, Mar 2016.

[109] K. G. Shea *et al.*, "A comparison of manual versus computer-assisted radiographic measurement. Intraobserver measurement variability for Cobb angles," vol. 23, nº. 5, p. 551-5, Mar 1 1998.

[110] R. T. Morrissy *et al.*, "Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error," vol. 72, nº. 3, p. 320-7, Mar 1990.

[111] J. Liu *et al.*, "The measurement of Cobb angle based on spine X-ray images using multi-scale convolutional neural network," Jul 12 2021. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/34251603

[112] M. H. Horng *et al.*, "Cobb Angle Measurement of Spine from X-Ray Images Using Convolutional Neural Network," vol. 2019, p. 6357171, 2019. [En ligne]. Disponible: https://downloads.hindawi.com/journals/cmmm/2019/6357171.pdf

[113] J. Qiao *et al.*, "Reliability analysis of a smartphone-aided measurement method for the Cobb angle of scoliosis," vol. 25, n°. 4, p. E88-92, Jun 2012.

[114] F. Jacquot *et al.*, "Measuring the Cobb angle with the iPhone in kyphoses: a reliability study," vol. 36, n°. 8, p. 1655-60, Aug 2012. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00264-012-1579-5.pdf

[115] E. Ketenci İ *et al.*, "Reliability of 2 Smartphone Applications for Cobb Angle Measurement in Scoliosis," vol. 13, n°. 1, p. 67-70, Mar 2021. [En ligne]. Disponible: https://ecios.org/Synapse/Data/PDFData/0157CIOS/cios-13-67.pdf

[116] J. C. Risser, "The Iliac apophysis; an invaluable sign in the management of scoliosis," vol. 11, p. 111-9, 1958.

[117] A. Coussement, *Radiopédiatrie en pratique courante*: Elsevier Masson, 2001.

[118] J. H. Hacquebord et S. S. Leopold, "In brief: The Risser classification: a classic tool for the clinician treating adolescent idiopathic scoliosis," vol. 470, n°. 8, p. 2335-8, Aug 2012.

[119] K. M. Neal, E. D. Shirley et G. M. Kiebzak, "Maturity Indicators and Adolescent Idiopathic Scoliosis: Evaluation of the Sanders Maturity Scale," vol. 43, n°. 7, p. E406-E412, 2018. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2018/04010/maturity_indicators_and_adolescent_idiopathic.12.aspx

[120] J. O. Sanders, D. G. Little et B. S. Richards, "Prediction of the crankshaft phenomenon by peak height velocity," vol. 22, n°. 12, p. 1352-6; discussion 1356-7, Jun 15 1997.

[121] E. Magnide *et al.*, "Automatic bone maturity grading from EOS radiographs in Adolescent Idiopathic Scoliosis," vol. 136, p. 104681, Jul 23 2021. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/34332349

[122] Scoliosis Research Society et W. G. o.-D. C. a. t. T. Committee. (2022) Revised Glossary of Terms - 3-D Classification and Terminology. [En ligne]. Disponible: https://www.srs.org/professionals/online-education-and-resources/glossary/revised-glossary-of-terms

[123] S. W. Suh *et al.*, "Idiopathic scoliosis in Korean schoolchildren: a prospective screening study of over 1 million children," vol. 20, n°. 7, p. 1087-94, Jul 2011.

[124] L. G. Lenke *et al.*, "Intraobserver and interobserver reliability of the classification of thoracic adolescent idiopathic scoliosis," vol. 80, n°. 8, p. 1097-106, Aug 1998.

[125] R. J. Cummings *et al.*, "Interobserver reliability and intraobserver reproducibility of the system of King et al. for the classification of adolescent idiopathic scoliosis," vol. 80, n°. 8, p. 1107-11, Aug 1998.

[126] L. G. Lenke *et al.*, "Multisurgeon assessment of surgical decision-making in adolescent idiopathic scoliosis: curve classification, operative approach, and fusion levels," vol. 26, nº. 21, p. 2347-53, Nov 1 2001.

[127] B. K.-W. Ng et W.-W. Chau, "Changes of Shoulder Balance, Sagittal Alignments, and Curve Correction in the Treatment of Lenke 1 and 2 Adolescent Idiopathic Scoliosis Using a Three-Dimensional-Based Correction Strategy in Correlation to Health-Related Quality of Life Using the Scoliosis Research Society-22 Questionnaire," vol. 12, nº. 1, 2020. [En ligne]. Disponible: https://journals.lww.com/jotr/fulltext/2020/12010/changes_of_shoulder_balance,_sagittal_alignments,.2.aspx

[128] H. R. Lee *et al.*, "Shoulder balance in Lenke type 2 adolescent idiopathic scoliosis: correlations among radiological indices, cosmetic indices, and patient-reported outcomes," vol. 40, nº. 6, p. 700-707, 01 Jun. 2024 2024. [En ligne]. Disponible: https://thejns.org/spine/view/journals/j-neurosurg-spine/40/6/article-p700.xml

[129] I. A. Stokes, "Three-dimensional terminology of spinal deformity. A report presented to the Scoliosis Research Society by the Scoliosis Research Society Working Group on 3-D terminology of spinal deformity," vol. 19, nº. 2, p. 236-48, Jan 15 1994.

[130] A. P. Sangole *et al.*, "Three-dimensional classification of thoracic scoliotic curves," vol. 34, nº. 1, p. 91-9, Jan 1 2009.

[131] I. A. F. Stokes, A. P. Sangole et C.-E. Aubin, "Classification of scoliosis deformity three-dimensional spinal shape by cluster analysis," *Spine (Phila Pa 1976),* vol. 34, nº. 6, p. 584-590, 2009. [En ligne]. Disponible: https://pubmed.ncbi.nlm.nih.gov/19282737

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2664249/

[132] S. Kadoury et H. Labelle, "Classification of three-dimensional thoracic deformities in adolescent idiopathic scoliosis from a multivariate analysis," vol. 21, nº. 1, p. 40-9, Jan 2012. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00586-011-2004-2.pdf

[133] S. Pasha *et al.*, "Discovering the association between the pre- and post-operative 3D spinal curve patterns in adolescent idiopathic scoliosis," vol. 9, nº. 4, p. 1053-1062, Jul 2021. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s43390-020-00276-1.pdf

[134] H. Labelle *et al.*, "Seeing the spine in 3D: how will it change what we do?," vol. 31, nº. 1 Suppl, p. S37-45, Jan-Feb 2011.

[135] O. Gille *et al.*, "Reliability of 3D reconstruction of the spine of mild scoliotic patients," vol. 32, nº. 5, p. 568-73, Mar 1 2007.

[136] C. Aubin *et al.*, "The SRS-Lenke-Aubin 3D Classification of Adolescent Idiopathic Scoliosis," non publié.

[137] S. Delorme *et al.*, "Assessment of the 3-d reconstruction and high-resolution geometrical modeling of the human skeletal trunk from 2-D radiographic images," vol. 50, nº. 8, p. 989-98, Aug 2003. [En ligne]. Disponible: https://ieeexplore.ieee.org/document/1213851/

[138] J. M. Lombardi et L. G. Lenke, "6 Classification of Adolescent Idiopathic Scoliosis for Surgical Intervention", dans *Idiopathic Scoliosis*, 2. Edition^{e} éd., P. O. Newton et H. S. Group, édit. New York: Thieme Medical Publishers, Inc., 2022, http://www.thieme-connect.de/products/ebooks/lookinside/10.1055/b-0041-184082

[139] L. Duong, F. Cheriet et H. Labelle, "Three-dimensional classification of spinal deformities using fuzzy clustering," vol. 31, n^{o}. 8, p. 923-930, 2006.

[140] L. Duong *et al.*, "Three-dimensional subclassification of Lenke type 1 scoliotic curves," vol. 22, n^{o}. 2, p. 135-43, Apr 2009.

[141] K. Hayashi *et al.*, "Three-dimensional analysis of thoracic apical sagittal alignment in adolescent idiopathic scoliosis," vol. 34, n^{o}. 8, p. 792-7, Apr 15 2009.

[142] P. Poncet, J. Dansereau et H. Labelle, "Geometric torsion in idiopathic scoliosis: three-dimensional analysis and proposal for a new classification," vol. 26, n^{o}. 20, p. 2235-43, Oct 15 2001.

[143] J. Shen *et al.*, "Geometric Torsion in Adolescent Idiopathic Scoliosis: A New Method to Differentiate Between Lenke 1 Subtypes," vol. 42, n^{o}. 9, p. E532-E538, 2017. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2017/05010/geometric_torsion_in_adolescent_idiopathic.11.aspx

[144] S. Negrini *et al.*, "2011 SOSORT guidelines: Orthopaedic and Rehabilitation treatment of idiopathic scoliosis during growth," vol. 7, n^{o}. 1, p. 3, Jan 20 2012. [En ligne]. Disponible: https://scoliosisjournal.biomedcentral.com/track/pdf/10.1186/1748-7161-7-3.pdf

[145] C. E. Aubin, H. Labelle et O. C. Ciolofan, "Variability of spinal instrumentation configurations in adolescent idiopathic scoliosis," vol. 16, n^{o}. 1, p. 57-64, Jan 2007. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00586-006-0063-6.pdf

[146] F. Le Naveaux *et al.*, "Significant variability in surgeons' preferred correction maneuvers and instrumentation strategies when planning adolescent idiopathic scoliosis surgery," vol. 13, p. 21, 2018. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/30324149

https://scoliosisjournal.biomedcentral.com/track/pdf/10.1186/s13013-018-0169-8.pdf

[147] M. Robitaille, C. E. Aubin et H. Labelle, "Intra and interobserver variability of preoperative planning for surgical instrumentation in adolescent idiopathic scoliosis," vol. 16, n^{o}. 10, p. 1604-14, Oct 2007. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00586-007-0431-x.pdf

[148] L. G. Lenke *et al.*, "Curve prevalence of a new classification of operative adolescent idiopathic scoliosis: does classification correlate with treatment?," vol. 27, n^{o}. 6, p. 604-11, Mar 15 2002.

[149] S. I. Suk *et al.*, "Determination of distal fusion level with segmental pedicle screw fixation in single thoracic idiopathic scoliosis," vol. 28, n^{o}. 5, p. 484-91, Mar 1 2003.

[150] K. Cao *et al.*, "Selection of lower instrumented vertebra in treating Lenke type 2A adolescent idiopathic scoliosis," vol. 39, n^{o}. 4, p. E253-61, Feb 15 2014.

[151] P. Parisini *et al.*, "Selective thoracic surgery in the Lenke type 1A: King III and King IV type curves," vol. 18 Suppl 1, n°. Suppl 1, p. 82-8, Jun 2009. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00586-009-0990-0.pdf

[152] Y. Wang *et al.*, "Distal adding-on phenomenon in Lenke 1A scoliosis: risk factor identification and treatment strategy comparison," vol. 36, n°. 14, p. 1113-22, Jun 15 2011.

[153] L. Lenke *et al.*, "Radiographic results of selecting the touched vertebra as the lowest instrumented vertebra in Lenke 1A AIS curves at a minimum five-year follow up," communication présentée à Scoliosis Research Society 49th Annual Meeting & Course, 2014, p. 10-13.

[154] E. C. Beauchamp *et al.*, "Selecting the "Touched Vertebra" as the Lowest Instrumented Vertebra in Patients with Lenke Type-1 and 2 Curves: Radiographic Results After a Minimum 5-Year Follow-up," vol. 102, n°. 22, p. 1966-1973, 2020. [En ligne]. Disponible: https://journals.lww.com/jbjsjournal/Fulltext/2020/11180/Selecting_the__Touched_Vertebra__as_the_Lowest.6.aspx

[155] X. Wang *et al.*, "Biomechanical analysis of corrective forces in spinal instrumentation for scoliosis treatment," vol. 37, n°. 24, p. E1479-87, Nov 15 2012.

[156] H. Arima *et al.*, "Expert consensus on surgical treatment for adolescent idiopathic scoliosis in Japan," vol. 26, n°. 5, p. 765-773, Sep 2021. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/32951958

[157] S. I. Suk *et al.*, "Pedicle screw instrumentation in adolescent idiopathic scoliosis (AIS)," vol. 21, n°. 1, p. 13-22, Jan 2012. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/21874625

https://link.springer.com/content/pdf/10.1007/s00586-011-1986-0.pdf

[158] G. F. Solitro *et al.*, "Currently Adopted Criteria for Pedicle Screw Diameter Selection," vol. 13, n°. 2, p. 132-145, Apr 2019. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/31131212

https://www.ijssurgery.com/content/ijss/13/2/132.full.pdf

[159] K. Takeshita *et al.*, "Diameter, Length, and Direction of Pedicle Screws for Scoliotic Spine: Analysis by Multiplanar Reconstruction of Computed Tomography," vol. 34, n°. 8, p. 798-803, 2009. [En ligne]. Disponible: https://journals.lww.com/spinejournal/Fulltext/2009/04150/Diameter,_Length,_and_Direction_of_Pedicle_Screws.9.aspx

[160] R. A. Lehman Jr *et al.*, "Straight-forward versus anatomic trajectory technique of thoracic pedicle screw fixation: a biomechanical analysis," vol. 28, n°. 18, p. 2058-2065, 2003.

[161] J. W. Barber *et al.*, "Biomechanical study of lumbar pedicle screws: does convergence affect axial pullout strength?," vol. 11, n°. 3, p. 215-20, Jun 1998.

[162] W. Sterba *et al.*, "Biomechanical analysis of differing pedicle screw insertion angles," vol. 22, n°. 4, p. 385-91, May 2007. [En ligne]. Disponible: https://www.clinbiomech.com/article/S0268-0033(06)00219-1/fulltext

[163] F. Le Naveaux *et al.*, "Implant distribution in surgically instrumented Lenke 1 adolescent idiopathic scoliosis: does it affect curve correction?," vol. 40, nº. 7, p. 462-8, Apr 1 2015. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/25608243

[164] N. J. Bharucha *et al.*, "Low-density versus high-density thoracic pedicle screw constructs in adolescent idiopathic scoliosis: do more screws lead to a better outcome?," vol. 13, nº. 4, p. 375-81, Apr 2013. [En ligne]. Disponible: https://www.thespinejournalonline.com/article/S1529-9430(12)00442-1/fulltext

[165] C. J. Hwang *et al.*, "Minimum 5-year follow-up results of skipped pedicle screw fixation for flexible idiopathic scoliosis," vol. 15, nº. 2, p. 146-50, Aug 2011. [En ligne]. Disponible: https://thejns.org/spine/abstract/journals/j-neurosurg-spine/15/2/article-p146.xml

[166] A. Howard *et al.*, "Improvement in quality of life following surgery for adolescent idiopathic scoliosis," vol. 32, nº. 24, p. 2715-8, Nov 15 2007.

[167] S. Donaldson *et al.*, "Surgeon reliability in rating physical deformity in adolescent idiopathic scoliosis," *Spine (Phila Pa 1976),* vol. 32, nº. 3, p. 363-367, 2007/02// 2007. [En ligne]. Disponible: http://europepmc.org/abstract/MED/17268269

https://doi.org/10.1097/01.brs.0000253605.71168.2e

[168] A. N. Larson *et al.*, "Does higher anchor density result in increased curve correction and improved clinical outcomes in adolescent idiopathic scoliosis?," vol. 39, nº. 7, p. 571-8, Apr 1 2014.

[169] B. S. Lonner *et al.*, "Thoracic pedicle screw instrumentation: the learning curve and evolution in technique in the treatment of adolescent idiopathic scoliosis," vol. 34, nº. 20, p. 2158-64, Sep 15 2009.

[170] J. W. Kemppainen *et al.*, "Evaluation of Limited Screw Density Pedicle Screw Constructs in Posterior Fusions for Adolescent Idiopathic Scoliosis," vol. 4, nº. 1, p. 33-39, Jan 2016.

[171] L. Y. Carreon *et al.*, "Non-neurologic complications following surgery for adolescent idiopathic scoliosis," vol. 89, nº. 11, p. 2427-32, Nov 2007.

[172] A. N. Larson *et al.*, "What would be the annual cost savings if fewer screws were used in adolescent idiopathic scoliosis treatment in the US?," vol. 24, nº. 1, p. 116-23, Jan 2016. [En ligne]. Disponible: https://thejns.org/downloadpdf/journals/j-neurosurg-spine/24/1/article-p116.pdf

[173] M. Wedemeyer *et al.*, "Titanium versus stainless steel for anterior spinal fusions: an analysis of rod stress as a predictor of rod breakage during physiologic loading in a bovine model," vol. 32, nº. 1, p. 42-8, Jan 1 2007.

[174] H. Serhan *et al.*, "Would CoCr rods provide better correctional forces than stainless steel or titanium for rigid scoliosis curves?," vol. 26, nº. 2, p. E70-4, Apr 2013.

[175] J. P. Sardi *et al.*, "Accuracy of Rod Contouring to Desired Angles With and Without a Template: Implications for Achieving Desired Spinal Alignment and Outcomes," vol. 13, nº. 2, p. 425-431, Mar 2023.

[176] F. Solla *et al.*, "Patient-specific Rods for Surgical Correction of Sagittal Imbalance in Adults: Technical Aspects and Preliminary Results," vol. 32, nº. 2, p. 80-86, Mar 2019.

[177] S. Prost *et al.*, "“Patient-specific” rods in the management of adult spinal deformity. One-year radiographic results of a prospective study about 86 patients," vol. 66, nº. 3, p. 162-167, 2020/06/01/ 2020. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S0028377020300576

[178] S. Prost *et al.*, "Adult Spinal Deformities: Can Patient-Specific Rods Change the Preoperative Planning into Clinical Reality? Feasibility Study and Preliminary Results about 77 Cases," vol. 2020, p. 6120580, 2020.

[179] C. Barton *et al.*, "Early Experience and Initial Outcomes With Patient-Specific Spine Rods for Adult Spinal Deformity," vol. 39, nº. 2, p. 79-86, Mar-Apr 2016.

[180] S. S. Jabbouri *et al.*, "Pre-contoured patient-specific rods result in superior immediate sagittal plane alignment than surgeon contoured rods in adolescent idiopathic scoliosis," vol. 10, nº. 2, p. 177-189, Jun 21 2024.

[181] B. S. Lonner *et al.*, "Evolution of Surgery for Adolescent Idiopathic Scoliosis Over 20 Years: Have Outcomes Improved?," vol. 43, nº. 6, p. 402-410, 2018. [En ligne]. Disponible: https://journals.lww.com/spinejournal/Fulltext/2018/03150/Evolution_of_Surgery_for_Adolescent_Idiopathic.10.aspx

[182] C. E. Bartley *et al.*, "Perioperative and Delayed Major Complications Following Surgical Treatment of Adolescent Idiopathic Scoliosis," vol. 99, nº. 14, p. 1206-1212, 2017. [En ligne]. Disponible: https://journals.lww.com/jbjsjournal/Fulltext/2017/07190/Perioperative_and_Delayed_Major_Complications.7.aspx

[183] W. H. Chung *et al.*, "Severe Lenke 1 and 2 adolescent idiopathic scoliosis had poorer perioperative outcome, higher complication rate, longer fusion and higher operative cost compared to non-severe scoliosis," vol. 31, nº. 4, p. 1051-1059, Apr 2022. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/35066683

https://link.springer.com/content/pdf/10.1007/s00586-022-07118-w.pdf

[184] A. R. Hariharan *et al.*, "Complications following surgical treatment of adolescent idiopathic scoliosis: a 10-year prospective follow-up study," Apr 30 2022. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/35488969

https://link.springer.com/content/pdf/10.1007/s43390-022-00508-6.pdf

[185] J. D. Coe *et al.*, "Complications in Spinal Fusion for Adolescent Idiopathic Scoliosis in the New Millennium. A Report of the Scoliosis Research Society Morbidity and Mortality Committee," vol. 31, nº. 3, p. 345-349, 2006. [En ligne]. Disponible: https://journals.lww.com/spinejournal/Fulltext/2006/02010/Complications_in_Spinal_Fusion_for_Adolescent.15.aspx

[186] S. Botolin, C. Merritt et M. Erickson, "Aseptic Loosening of Pedicle Screw As a Result of Metal Wear Debris in a Pediatric Patient," vol. 38, nº. 1, p. E38-E42, 2013. [En ligne]. Disponible:

https://journals.lww.com/spinejournal/Fulltext/2013/01010/Aseptic_Loosening_of_Pedicle_Screw_As_a_Result_of.23.aspx

[187] L. D. E. D. M. Smals *et al.*, "Post-marketing surveillance on safety and efficacy of posterior spinal correction and fusion with the CD Horizon Solera instrumentation for adolescent idiopathic scoliosis. A retrospective cohort study," vol. 8, 2021. [En ligne]. Disponible: https://doi.org/10.1016/j.xnsj.2021.100085

https://www.nassopenaccess.org/article/S2666-5484(21)00037-8/pdf

[188] A. Mason *et al.*, "The accuracy of pedicle screw placement using intraoperative image guidance systems: A systematic review," vol. 20, nº. 2, p. 196-203, 2014. [En ligne]. Disponible: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84893337997&doi=10.3171%2f2013.11.SPINE13413&partnerID=40&md5=6c615fba932098e101a62718a7c52a2e

[189] I. Nemaris, "Surgimap User Guide (Version 2.2.15 and higher), Revision 4.4," éd: Nemaris, Inc., 2019.

[190] R. Lafage *et al.*, "Self-learning computers for surgical planning and prediction of postoperative alignment," vol. 27, nº. Suppl 1, p. 123-128, Feb 2018.

[191] W. Wu *et al.*, "Reliability and reproducibility analysis of the Cobb angle and assessing sagittal plane by computer-assisted and manual measurement tools," vol. 15, p. 33, Feb 6 2014. [En ligne]. Disponible: https://bmcmusculoskeletdisord.biomedcentral.com/track/pdf/10.1186/1471-2474-15-33.pdf

[192] H. G. mediCAD. (2017) mediCAD SPINE 3D: Sagittal Balance, Automatic Measurement, Spondylodesis, Implants, Osteotomy. [En ligne]. Disponible: https://www.hectec.de/content/images/demo/new_images/3d/2017_mediCAD-SPINE3D_EN_05-2017_Web.pdf

[193] Medtronic, "Clinical Brief – UNiD™ ASI Adaptive Spine Intelligence," 2020. [En ligne]. Disponible: https://ccevent.site/wp-content/uploads/2021/03/UNiD%E2%84%A2-Rod-Clinical-Brief.pdf

[194] J. L. Clement *et al.*, "Surgical increase in thoracic kyphosis increases unfused lumbar lordosis in selective fusion for thoracic adolescent idiopathic scoliosis," vol. 28, nº. 3, p. 581-589, Mar 2019. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/30206693

[195] J. Rehm *et al.*, "3D-modeling of the spine using EOS imaging system: Inter-reader reproducibility and reliability," vol. 12, nº. 2, p. e0171258, 2017. [En ligne]. Disponible: https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0171258&type=printable

[196] Brainlab. (Brainlab Surgery Products) [En ligne]. Disponible: https://www.brainlab.com/surgery-products//

[197] P.-P. Girod *et al.*, "Asymmetric pedicle subtractionosteotomy (aPSO) guided by a 3D-printed model to correct a combined fixed sagittal and coronal imbalance," vol. 40, nº. 4, p.

689-693, 2017/10/01 2017. [En ligne]. Disponible: https://doi.org/10.1007/s10143-017-0882-4

[198] W. Luo, C. Yu et L. Hu, "Solid-State Electrolyte," dans *Encyclopedia of Inorganic and Bioinorganic Chemistry*, p. 1-9.

[199] P. A. J. Pijpker *et al.*, "Three-Dimensional Planning and Use of Individualized Osteotomy-Guiding Templates for Surgical Correction of Kyphoscoliosis: A Technical Case Report," vol. 119, p. 113-117, 2018. [En ligne]. Disponible: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85052073837&doi=10.1016%2fj.wneu.2018.07.219&partnerID=40&md5=a97e5701b6ffbad0ee6b981ea1169487

[200] Medtronic. Medtronic UNiD™ HUB. [En ligne]. Disponible: https://platformous.medicrea.com/Account/

[201] I. Nemaris. Surgimap. [En ligne]. Disponible: https://www.surgimap.com/

[202] I. NuVasive. Featured Offerings - Bendini®. [En ligne]. Disponible: https://www.nuvasive.com/procedures/featured-offerings/bendini/

[203] S.M.A.I.O. (2021) K-ROD – Custom-Made Rods: Instructions for Use (Rev. A). [En ligne]. Disponible: https://smaio.com/e-ifu/archives/rev_a/KROD-IMP-IFU-UE-EN_Rev_a.pdf

[204] X. Wang *et al.*, "How does differential rod contouring contribute to 3-dimensional correction and affect the bone-screw forces in adolescent idiopathic scoliosis instrumentation?," vol. 39, p. 115-121, 2016/11/01/ 2016. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S026800331630153X

https://www.clinbiomech.com/article/S0268-0033(16)30153-X/fulltext

[205] L. Boyer, "Analyse biomécanique de techniques de dérotation vertébrale pour la correction 3D de la scoliose lors de la chirurgie d'instrumentation postérieure du rachis," Ecole Polytechnique, Montreal (Canada), 2017.

[206] C. E. Aubin, V. Goussev et Y. Petit, "Biomechanical modelling of segmental instrumentation for surgical correction of 3D spinal deformities using Euler-Bernoulli thin-beam elastic deformation equations," vol. 42, nº. 2, p. 216-21, Mar 2004. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/BF02344634.pdf

[207] V. Lafage *et al.*, "3D finite element simulation of Cotrel-Dubousset correction," vol. 9, nº. 1-2, p. 17-25, 2004. [En ligne]. Disponible: https://www.tandfonline.com/doi/abs/10.3109/10929080400006390

[208] N. Cobetto, C.-E. Aubin et S. Parent, "Surgical Planning and Follow-up of Anterior Vertebral Body Growth Modulation in Pediatric Idiopathic Scoliosis Using a Patient-Specific Finite Element Model Integrating Growth Modulation," vol. 6, nº. 4, p. 344-350, 2018/07/01/ 2018. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S2212134X17304161

[209] N. Cobetto, C.-E. Aubin et S. Parent, "Anterior Vertebral Body Growth Modulation: Assessment of the 2-year Predictive Capability of a Patient-specific Finite-element

Planning Tool and of the Growth Modulation Biomechanics," vol. 45, nº. 18, p. E1203-E1209, 2020. [En ligne]. Disponible: https://journals.lww.com/spinejournal/Fulltext/2020/09150/Anterior_Vertebral_Body_Growth_Modulation_.20.aspx

[210] C. E. Aubin *et al.*, "Instrumentation Strategies to Reduce the Risks of Proximal Junctional Kyphosis in Adult Scoliosis: A Detailed Biomechanical Analysis," vol. 3, nº. 3, p. 211-218, May 2015. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/abs/pii/S2212134X14002664?via%3Dihub

[211] M. Cammarata *et al.*, "Biomechanical risk factors for proximal junctional kyphosis: a detailed numerical analysis of surgical instrumentation variables," vol. 39, nº. 8, p. E500-7, Apr 15 2014.

[212] X. Wang *et al.*, "Biomechanical modeling and analysis of a direct incremental segmental translation system for the instrumentation of scoliotic deformities," vol. 26, nº. 6, p. 548-55, Jul 2011. [En ligne]. Disponible: https://www.clinbiomech.com/article/S0268-0033(11)00033-7/fulltext

[213] F. Poulin *et al.*, "[Biomechanical modeling of instrumentation for the scoliotic spine using flexible elements: a feasibility study]," vol. 52, nº. 8, p. 761-7, 1998.

[214] Y. Petit, C. É. Aubin et H. Labelle, "Patient-specific mechanical properties of a flexible multi-body model of the scoliotic spine," vol. 42, nº. 1, p. 55-60, 2004/01/01 2004. [En ligne]. Disponible: https://doi.org/10.1007/BF02351011

https://link.springer.com/content/pdf/10.1007/BF02351011.pdf

[215] F. Luce, *Contribution à l'amélioration du réalisme d'un modèle multi-corps flexibles de chirurgie pour le traitement de la scoliose*: École Polytechnique de Montréal, 2004.

[216] R. Stirbu, *Modélisation multi-corps flexibles de chirurgies d'instrumentation du rachis scoliotique*: École Polytechnique de Montréal, 2004.

[217] X. Wang *et al.*, "Biomechanical Computational Study of Pedicle Screw Position and Density in Adolescent Idiopathic Scoliosis Instrumentation," vol. 48, nº. 20, p. 1436-1445, Oct 15 2023.

[218] M. Gay *et al.*, "Biomechanical analysis of rod contouring in posterior spinal instrumentation and fusion for 3D correction of adolescent idiopathic scoliosis," vol. 11, nº. 6, p. 1309-1316, Nov 2023.

[219] X. Wang *et al.*, "Concave rod first vs. convex rod first in AIS instrumentation with differential rod contouring: computer modeling and simulations based on ten AIS surgical cases," vol. 11, nº. 6, p. 1317-1324, 2023/11/01 2023. [En ligne]. Disponible: https://doi.org/10.1007/s43390-023-00727-5

[220] H. Wu *et al.*, "Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet," communication présentée à International Conference on Medical Image Computing and Computer-Assisted Intervention, 2017, p. 127-135.

[221] I. Castro-Mateos *et al.*, "Intervertebral disc classification by its degree of degeneration from T2-weighted magnetic resonance images," vol. 25, nº. 9, p. 2721-7, Sep 2016. [En ligne]. Disponible: https://link.springer.com/content/pdf/10.1007/s00586-016-4654-6.pdf

[222] J. Schmidhuber, "Deep learning in neural networks: An overview," vol. 61, p. 85-117, 2015.

[223] J. Zhang *et al.*, "Automatic Cobb measurement of scoliosis based on fuzzy Hough Transform with vertebral shape prior," vol. 22, nº. 5, p. 463-72, Oct 2009. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/18516643

https://link.springer.com/content/pdf/10.1007/s10278-008-9127-y.pdf

[224] H. Wu *et al.*, "Automated comprehensive Adolescent Idiopathic Scoliosis assessment using MVC-Net," vol. 48, p. 1-11, Aug 2018. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/29803920

[225] Y. Pan *et al.*, "Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays," vol. 28, nº. 12, p. 3035-3043, Dec 2019. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/31446493

https://link.springer.com/content/pdf/10.1007/s00586-019-06115-w.pdf

[226] J. Zhang *et al.*, "Computer-Aided Cobb Measurement Based on Automatic Detection of Vertebral Slopes Using Deep Neural Network," vol. 2017, p. 9083916, 2017. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/29118806

https://downloads.hindawi.com/journals/ijbi/2017/9083916.pdf

[227] P. Bernstein *et al.*, "Radiographic scoliosis angle estimation: spline-based measurement reveals superior reliability compared to traditional COBB method," vol. 30, nº. 3, p. 676-685, Mar 2021. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/32856177

[228] K. Watanabe, Y. Aoki et M. Matsumoto, "An Application of Artificial Intelligence to Diagnostic Imaging of Spine Disease: Estimating Spinal Alignment From Moire Images," vol. 16, nº. 4, p. 697-702, Dec 2019. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/31905459

https://www.e-neurospine.org/upload/pdf/ns-1938426-213.pdf

[229] J. Yang *et al.*, "Development and validation of deep learning algorithms for scoliosis screening using back images," vol. 2, nº. 1, p. 390, 2019/10/25 2019. [En ligne]. Disponible: https://doi.org/10.1038/s42003-019-0635-8

https://www.nature.com/articles/s42003-019-0635-8.pdf

[230] A. P. Mc *et al.*, "Middle-Column Gap Balancing and Middle-Column Mismatch in Spinal Reconstructive Surgery," vol. 12, nº. 2, p. 160-171, Apr 2018. [En ligne]. Disponible: https://www.ijssurgery.com/content/ijss/12/2/160.full.pdf

[231] J. Mikulka *et al.*, "Pediatric Spine Segmentation and Modeling Using Machine Learning," communication présentée à 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2019, p. 1-5.

[232] K. B. Siemionow *et al.*, "Autonomous lumbar spine pedicle screw planning using machine learning: A validation study," vol. 12, n°. 3, p. 223-227, Jul-Sep 2021. [En ligne]. Disponible: https://www.jcvjs.com/article.asp?issn=0974-8237;year=2021;volume=12;issue=3;spage=223;epage=227;aulast=Siemionow

[233] R. Lafage *et al.*, "Predictive Model for Selection of Upper Treated Vertebra Using a Machine Learning Approach," vol. 146, p. e225-e232, Feb 2021.

[234] G. S. Collins *et al.*, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement," vol. 350, p. g7594, Jan 7 2015.

[235] F. Schwendicke *et al.*, "Artificial intelligence in dental research: Checklist for authors, reviewers, readers," vol. 107, p. 103610, Apr 2021.

[236] W. Luo *et al.*, "Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View," vol. 18, n°. 12, p. e323, Dec 16 2016.

[237] P. M. Bossuyt *et al.*, "[Reporting studies of diagnostic accuracy according to a standard method; the Standards for Reporting of Diagnostic Accuracy (STARD)]," vol. 147, n°. 8, p. 336-40, Feb 22 2003.

[238] B. Rohrig *et al.*, "Types of study in medical research: part 3 of a series on evaluation of scientific publications," vol. 106, n°. 15, p. 262-8, Apr 2009. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/19547627

[239] T. Mathes et D. Pieper, "An algorithm for the classification of study designs to assess diagnostic, prognostic and predictive test accuracy in systematic reviews," vol. 8, n°. 1, p. 226, Sep 3 2019. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/31481098

[240] R. Qaddoura *et al.*, "A Multi-Layer Classification Approach for Intrusion Detection in IoT Networks Based on Deep Learning," vol. 21, n°. 9, p. 2987, 2021. [En ligne]. Disponible: https://www.mdpi.com/1424-8220/21/9/2987

[241] D. P. Mandic et J. Chambers, *Recurrent neural networks for prediction: learning algorithms, architectures and stability*: John Wiley & Sons, Inc., 2001.

[242] J. Cao *et al.*, "Randomly translational activation inspired by the input distributions of ReLU," vol. 275, p. 859-868, 2018/01/31/ 2018. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S0925231217315485

[243] L. He *et al.*, "Neural network-based multi-task learning for inpatient flow classification and length of stay prediction," vol. 108, 2021, 107483.

[244] A. Kendall, Y. Gal et R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," communication présentée à Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, p. 7482-7491.

[245] D. P. Kingma et J. Ba, "Adam: A method for stochastic optimization," 2014.

[246] J. D. Rodriguez, A. Perez et J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," vol. 32, n°. 3, p. 569-575, 2009.

[247] N. Srivastava *et al.*, "Dropout: a simple way to prevent neural networks from overfitting," vol. 15, nº. 1, p. 1929-1958, 2014.

[248] C. Pillot *et al.*, "Biomechanical analysis of 3D correction and bone-screw forces as a function of rod insertion sequence and orientation relative to the sagittal plane in adolescent idiopathic scoliosis instrumentation," vol. 128, p. 106618, Jul 15 2025.

[249] ASME, "ASME V&V 40-2018: Assessing credibility of computational modeling through verification and validation: application to medical devices," 2018.

[250] ASME., "ASME V&V 10-2019: Standard for Verification and Validation in Computational Solid Mechanics," communication présentée à, 2020.

[251] U. S. F. a. D. Administration, "Reporting of Computational Modeling Studies in Medical Device Submissions: Guidance for Industry and Food and Drug Administration Staff," éd. Silver Spring, MD: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, 2016.

[252] U. S. F. a. D. Administration, "Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions: Guidance for Industry and Food and Drug Administration Staff," éd. Silver Spring, MD: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, 2023.

[253] A. P. Baumann *et al.*, "Assessing the use of finite element analysis for mechanical performance evaluation of intervertebral body fusion devices," vol. 4, nº. 1, 2021. [En ligne]. Disponible: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103141074&doi=10.1002%2fjsp2.1137&partnerID=40&md5=1f44c5c7368ba33336988c24a1a62cd5

[254] H. Canada, "Guidance Documents – Medical Devices," éd. Ottawa, Canada: Government of Canada, Health Canada, 2025.

[255] H. Canada, "Guidance Document – Software as a Medical Device (SaMD): Definition and Classification," éd. Ottawa, Canada: Government of Canada, Health Canada, 2013.

[256] S. Nagaraja *et al.*, "Establishing finite element model credibility of a pedicle screw system under compression-bending: An end-to-end example of the ASME V&V 40 standard," vol. 225, p. 74-88, 2024/05/01/ 2024. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S1046202324000665

[257] F. Le Navéaux *et al.*, "3D rod shape changes in adolescent idiopathic scoliosis instrumentation: how much does it impact correction?," vol. 26, nº. 6, p. 1676-1683, 2017/06/01 2017. [En ligne]. Disponible: https://doi.org/10.1007/s00586-017-4958-1

[258] F. Cheriet *et al.*, "A Novel System for the 3-D Reconstruction of the Human Spine and Rib Cage From Biplanar X-Ray Images," vol. 54, nº. 7, p. 1356-1358, 2007. [En ligne]. Disponible: https://ieeexplore.ieee.org/document/4237330/

[259] Y. Majdouline *et al.*, "Computer simulation for the optimization of instrumentation strategies in adolescent idiopathic scoliosis," vol. 47, nº. 11, p. 1143-54, Nov 2009. [En ligne]. Disponible: https://www.ncbi.nlm.nih.gov/pubmed/19669822

https://link.springer.com/content/pdf/10.1007/s11517-009-0509-1.pdf

[260] C. Chotigavanichaya *et al.*, "Comparative effectiveness of different pedicle screw density patterns in spinal deformity correction of small and flexible operative adolescent idiopathic scoliosis: inverse probability of treatment weighting analysis," vol. 32, nº. 6, p. 2203-2212, 2023/06/01 2023. [En ligne]. Disponible: https://doi.org/10.1007/s00586-023-07615-6

[261] F. J. Baky *et al.*, "Predictors of cost for posterior spinal fusion in adolescent idiopathic scoliosis," vol. 8, nº. 3, p. 421-426, 2020/06/01 2020. [En ligne]. Disponible: https://doi.org/10.1007/s43390-020-00053-0

[262] H. Yoshihara, C. Paulino et D. Yoneoka, "Predictors of Increased Hospital Stay in Adolescent Idiopathic Scoliosis Patients Undergoing Posterior Spinal Fusion: Analysis of National Database," vol. 6, nº. 3, p. 226-230, 2018/05/01/ 2018. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S2212134X17303854

[263] N. D. Fletcher *et al.*, "Clinical and economic implications of early discharge following posterior spinal fusion for adolescent idiopathic scoliosis," vol. 8, nº. 3, p. 257-263, 2014. [En ligne]. Disponible: https://journals.sagepub.com/doi/abs/10.1007/s11832-014-0587-y

[264] A. C. Hengartner *et al.*, "Impact of a quality improvement initiative and monthly multidisciplinary meetings on outcomes after posterior spinal fusion for adolescent idiopathic scoliosis," vol. 12, nº. 4, p. 1043-1051, 2024/07/01 2024. [En ligne]. Disponible: https://doi.org/10.1007/s43390-024-00859-2

[265] A. B. Harris *et al.*, "Factors associated with extended length of stay in patients undergoing posterior spinal fusion for adolescent idiopathic scoliosis," vol. 8, nº. 2, p. 187-193, 2020/04/01 2020. [En ligne]. Disponible: https://doi.org/10.1007/s43390-019-00008-0

[266] J. R. Engsberg *et al.*, "Prospective Comparison of Gait and Trunk Range of Motion in Adolescents With Idiopathic Thoracic Scoliosis Undergoing Anterior or Posterior Spinal Fusion," vol. 28, nº. 17, p. 1993-2000, 2003. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2003/09010/prospective_comparison_of_gait_and_trunk_range_of.15.aspx

[267] C. S. Kruse *et al.*, "Challenges and Opportunities of Big Data in Health Care: A Systematic Review," vol. 4, nº. 4, p. e38, 2016/11/21 2016. [En ligne]. Disponible: http://medinform.jmir.org/2016/4/e38/

https://doi.org/10.2196/medinform.5359

http://www.ncbi.nlm.nih.gov/pubmed/27872036

[268] J. He *et al.*, "The practical implementation of artificial intelligence technologies in medicine," vol. 25, nº. 1, p. 30-36, 2019/01/01 2019. [En ligne]. Disponible: https://doi.org/10.1038/s41591-018-0307-0

[269] T. Davenport et R. Kalakota, "The potential for artificial intelligence in healthcare," vol. 6, nº. 2, p. 94-98, 2019/06/01/ 2019. [En ligne]. Disponible: https://www.sciencedirect.com/science/article/pii/S2514664524010592

[270] B. Khan *et al.*, "Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector," p. 1-8, Feb 8 2023.

[271] P. Phan *et al.*, "Computer algorithms and applications used to assist the evaluation and treatment of adolescent idiopathic scoliosis: a review of published articles 2000–2009," vol. 20, nº. 7, p. 1058-1068, 2011/07/01 2011. [En ligne]. Disponible: https://doi.org/10.1007/s00586-011-1699-4

[272] D. A. Glaser, J. Doan et P. O. Newton, "Comparison of 3-Dimensional Spinal Reconstruction Accuracy: Biplanar Radiographs With EOS Versus Computed Tomography," vol. 37, nº. 16, 2012. [En ligne]. Disponible: https://journals.lww.com/spinejournal/fulltext/2012/07150/comparison_of_3_dimensional_spinal_reconstruction.8.aspx

[273] X. Wang *et al.*, "Biomechanical comparison of alternative densities of pedicle screws for the treatment of adolescent idiopathic scoliosis," vol. 21, nº. 6, p. 1082-1090, 2012/06/01 2012. [En ligne]. Disponible: https://doi.org/10.1007/s00586-011-2089-7

https://link.springer.com/content/pdf/10.1007/s00586-011-2089-7.pdf

APPENDIX A - ARTICLE 1: SEARCH STRATEGY

The broad search terms were selected to allow us to determine the scope and coverage of the body of literature available on the topic and to clearly indicate the volume of literature and studies available, along with a detailed overview of its focuses. In order to do so, the search terms included three parts, one to have anatomical structures related to the spine and spinal conditions, another to target the DL component of the studies, and, finally, a part to search for relevant imaging modalities.

The search terms from the titles, abstracts, and subject headings of identified primary studies and previous systematic reviews in related fields will be extracted and used to develop a search scheme presented in Table S.1 below.

The search terms from the titles, abstracts, and subject headings of identified primary studies and previous systematic reviews in related fields were extracted and used to develop the search scheme. Search terms were also identified and checked using the PubMed PubReMiner word frequency analysis tool. The relevance of the chosen Mesh terms and keywords were verified using the Yale MeSH Analyzer. Finally, a publications test set of 3 studies previously identified as necessary was used to test against the proposed search strategies.

A citation search of the identified articles in the included studies was also be performed. Additional searches were done for gray literature with Google Scholar, Proquest Dissertation and Theses Global, hand searching of reference lists, and conferences abstract from relevant conferences.

Table A.1 Search Formula used for the 4 Databases Included in the Systematic Review

Database	**Search strategy**
PubMed	((“Spine” [MeSH] OR “back”[MeSH] OR “Zygapophyseal Joint”[MeSH] OR “Spinal Diseases”[MeSH] OR “Sciatica”[MeSH] or "Spinal Injuries"[MeSH] OR “Laminectomy”[MeSH] OR “Cementoplasty”[MeSH] OR “Diskectomy”[MeSH] OR “Intervertebral Disc Chemolysis”[MeSH] OR “Laminoplasty”[MeSH] OR “Osteotomy”[MeSH] OR “Spinal Fusion”[MeSH] OR “Spinal Puncture”[MeSH] OR “Foraminotomy”[MeSH] OR “Neuroendoscopy”[MeSH] OR “Total Disc Replacement”[MeSH] OR "Pedicle Screws”[MeSH] OR Spine[tiab] OR Spina*[tiab] OR "degenerative disc"[tiab] OR “vertebr*”[tiab] OR “scoliosis”[tiab] OR "disc degeneration"[tiab] OR "Disc Degradation"[tiab] OR "disc disease"[tiab] OR "intervertebral disc"[tiab])) AND ((("Machine Learning"[MeSH] OR "Neural Networks, Computer"[MeSH] OR "naive bayes"[tiab] OR "bayesian learning"[tiab] OR "neural network*"[tiab] OR "random forest"[tiab] OR "deep learning"[tiab] OR "machine prediction"[tiab] OR "machine intelligence"[tiab] OR "generative adversarial networks"[tiab] OR "Hierarchical Learning"[tiab] OR "computer vision"[tiab] OR "computational intelligence"[tiab] OR "computational learning"[tiab] OR "computer reasoning"[tiab] OR "machine learning"[tiab] OR "reinforcement learning"[tiab] OR "convolutional network*"[tiab] OR "artificial intelligence"[tiab] OR “Self Organizing MAP"[tiab] OR "Self-Organizing MAP"[tiab] OR "AutoEncoder"[tiab] OR "CNN"[tiab] OR "GAN"[tiab] OR "GANN"[tiab])) OR (("convolute"[All Fields] OR "convoluted"[All Fields] OR "convolutes"[All Fields] OR "convoluting"[All Fields] OR "convolution"[All Fields] OR "convolutional"[All Fields] OR "convolutions"[All Fields] OR "convolutive"[All Fields]) AND ("neural networks, computer"[MeSH Terms] OR ("neural"[All Fields] AND "networks"[All Fields] AND "computer"[All Fields]) OR "computer neural networks"[All Fields] OR ("neural"[All Fields] AND "network"[All Fields]) OR "neural network"[All Fields]))) AND (("Diagnostic Imaging”[MeSH] OR ”Image Processing, Computer-Assisted”[MeSH] OR ”Imaging" OR "Radiograph*" OR "x?ray" OR "Tomograph*" OR "Magnetic Resonance" OR "MR?image*" OR "MRI" OR "MRA"[tiab] OR "CT?Scan*" OR "Ultrasonograph*" OR "Ultrasound*" OR "PET?Scan" OR "c-arm" OR "fluoroscop*" OR "arthrogram*" OR "arthrograph*" OR "venogram*" OR "venograph*" OR "cone?beam CT" OR "image-guided adaptive radiation therapy" OR "IGART"[tiab])) AND (("2012/01/01"[Date - Publication] : "3000"[Date - Publication])) NOT ((systematic review[pt] OR review[pt]))

Table A.1 Search Formula used for the 4 Databases Included in the Systematic Review (cont'd)

Web of Science	(((TS=("Spine" OR "Spinal*" OR "vertebr*" OR "Intervertebral Disc*" OR "intervertebral disk*" OR "Lumbosacr*" OR "Sacrococcy*" OR "Annulus Fibrosus" OR "Nucleus Pulposus" OR "Sacrum" OR "Epidural" OR "Zygapophyseal Joint*" OR "facet joint*" OR "spinous process*" OR "transverse process*" OR "lamina" OR "Posterior Longitudinal Ligament*" OR "Spinal Injur*" OR "Sciatica" OR "low* back pain" OR "degenerative disc*" OR "scoliosis" OR "disc degeneration" OR "Disc Degradation" OR "disc disease*" OR "disc herniat*" OR "Disc Displacement" OR "Kyphosis" OR "Lordosis" OR "scoliosis" OR "Osteochondrosis" OR "Scheuermann Disease" OR "Osteophytosis" OR "Hyperostosis" OR "Spondylitis" OR "Discitis" OR "Spondylarthritis" odor "Tuberculosis" OR "Spondylosis" OR "Spondylolisthesis" OR "Spondylarthropathies" OR "Neurosurg*" OR "Laminectom*" OR "Cementoplast*" OR "Diskectom*" OR "discectomy" OR "micro?discectomy" OR "Laminoplast*" OR "Spin* Fusion" OR "Spinal Puncture" OR "Foraminotom*" OR "Neuroendoscop*" OR "Disc Replacement" OR "Pedicle Screw*" OR "discectom*" OR "interbod* fusion" OR "foraminotomy" OR "posterior decompression" OR "Posterior Gutter Fusion" OR "Posterolateral gutter fusion" OR "Transforaminal" OR "artificial disc" OR "ALIF" OR "PLIF" OR "TLIF" OR "XLIF" OR "endif" OR "ACDF")) AND TS=((("machine learning" OR "naive bayes" OR "bayesian learning" OR "neural network*" OR "random forest" OR "deep learning" "generative adversarial networks" OR "Hierarchical Learning" OR "reinforcement learning" OR "convolutional network*" OR "Self Organizing MAP" OR "Self-Organizing MAP" OR "AutoEncoder" OR "CNN" OR "GAN" OR "GANN")))) AND TS=((("Diagnostic Imaging" OR "Imaging" OR "Radiograph*" OR "x?ray" OR "Tomograph*" OR "Magnetic Resonance" OR "MR?image*" OR "MRI" OR "CT?Scan*" OR "Ultrasonograph*" OR "Ultrasound*" OR "PET?Scan" OR "c-arm" OR "fluoroscop*" OR "arthrogram*" OR "arthrograph*" OR "venogram*" OR "venograph*")))) AND PY=(2012-2021)
IEEE Xplore	Search 1 (index terms): ("Index Terms":"Spine" OR "Index Terms":"vertebra" OR "Index Terms":"vertebrae" OR "Index Terms":"Spinal Disease" OR "Index Terms":"Spine Surgery" OR "Index Terms":"Spinal Surgery") AND ("Index Terms":CNN OR "Index Terms":"Convolutional neural network" OR "Index Terms":"deep learning") AND ("Index Terms":"Diagnostic imaging" OR "medical imaging" OR "Index Terms":"Radiography" OR "Index Terms":"computed tomography" OR "Index Terms":"Magnetic Resonance")

Table A.1 Search Formula used for the 4 Databases Included in the Systematic Review (cont'd)

Search 2 (abstract terms):

("Abstract":“Spine” OR "Abstract":“Spinal” OR "Abstract":"vertebrae" OR "Abstract":"vertebra" OR "Abstract":“Intervertebral" OR "Abstract":“Lumbosacral” OR "Abstract":“Sacrococcygeal” OR "Abstract":“Annulus Fibrosus” OR "Abstract":“Nucleus Pulposus” OR "Abstract":“Sacrum” OR "Abstract":“Epidural” OR "Abstract":“Zygapophyseal Joint” OR "Abstract":"facet joint" OR "Abstract":"spinous process" OR "Abstract":"transverse process" OR "Abstract":"lamina" OR "Abstract":“Posterior Longitudinal Ligament” OR "Abstract":"Spinal Injur*" OR "Abstract":“Sciatica” OR "Abstract":"low* back pain" OR "Abstract":"degenerative disc" OR "Abstract":“scoliosis” OR "Abstract":"disc degeneration" OR "Abstract":"Disc Degradation" OR "Abstract":"disc disease" OR "Abstract":"disc herniation" OR "Abstract":"Disc Displacement” OR "Abstract":“Kyphosis” OR "Abstract":“Lordosis” OR "Abstract":“scoliosis” OR "Abstract":“Osteochondrosis” OR "Abstract":“Scheuermann Disease” OR "Abstract":“Osteophytosis” OR "Abstract":“Hyperostosis” OR "Abstract":“Spondylitis” OR "Abstract":“Discitis” OR "Abstract":“Spondylarthritis” OR "Abstract":“Tuberculosis” OR "Abstract":“Spondylosis” OR "Abstract":“Spondylolisthesis” OR "Abstract":“Spondylarthropathies” OR "Abstract": “Neurosurgery" OR "Abstract":“Laminectomy” OR "Abstract":“Cementoplasty” OR "Abstract":“Diskectomy” OR "Abstract": "micro?discectomy" OR "Abstract":“Laminoplasty” OR "Abstract":“Spin* Fusion” OR "Abstract":“Spinal Puncture” OR "Abstract":“Foraminotomy” OR "Abstract":“Neuroendoscopy” OR "Abstract":“Disc Replacement” OR "Abstract":"Pedicle Screw” OR "Abstract":"discectomy" OR "Abstract":"interbody fusion" OR "Abstract":"foraminotomy" OR "Abstract":"posterior decompression" OR "Abstract":"Posterior Gutter Fusion" OR "Abstract":"Posterolateral gutter fusion" OR "Abstract":"Transforaminal" OR "Abstract":"artificial disc" OR "Abstract":"ALIF" OR "Abstract":"PLIF" OR "Abstract":"TLIF" OR "Abstract":"XLIF" OR "Abstract":"ETDIF" OR "Abstract":"ACDF") AND ("Abstract":"machine learning" OR "Abstract":"naive bayes” OR "Abstract":"bayesian learning” OR "Abstract":"neural network*” OR "Abstract":"random forest” OR "Abstract":"deep learning” "generative adversarial network” OR "Abstract":"Hierarchical Learning” OR "Abstract":"reinforcement learning” OR "Abstract":"convolutional network” OR "Abstract":“Self Organizing MAP” OR "Abstract":"Self-Organizing MAP” OR "Abstract":"AutoEncoder” OR "Abstract":"CNN” OR "Abstract":"GAN” OR "Abstract":"GANN") AND ("Abstract":"medical imaging" OR "Abstract":"Radiography" OR "Abstract":"x?ray" OR "Abstract":"Tomography" OR "Abstract":"Magnetic Resonance" OR "Abstract":"MR?image*" OR "Abstract":"MRI" OR "Abstract":"CT?Scan*" OR "Abstract":"Ultrasonography" OR "Abstract":"Ultrasound*" OR "Abstract":"PET?Scan" OR "Abstract":"c-arm" OR "Abstract":"fluoroscopy" OR "Abstract":"cone?beam CT" OR "Abstract":"image-guided adaptive radiation therapy" OR "Abstract":"IGART")

Table A.1 Search Formula used for the 4 Databases Included in the Systematic Review (cont'd)

Embase	('spine'/exp/mj OR 'spine disease'/exp/mj OR 'spine surgery'/exp/mj) AND ('machine learning'/exp OR 'deep learning algorithm'/exp OR 'deep learning model'/exp) AND ('radiodiagnosis'/exp OR 'computer assisted diagnosis'/exp OR 'nuclear magnetic resonance'/exp OR 'image guided biopsy'/exp OR 'imaging and display'/exp)

APPENDIX B - ARTICLE 1: LIST OF INCLUDED STUDIES

1. Ahammad SH, Rajesh V, Rahman MZU, Lay-Ekuakille A. A Hybrid CNN-Based Segmentation and Boosting Classifier for Real Time Sensor Spinal Cord Injury Data. *IEEE SENSORS JOURNAL.* 2020;20(17):10092-10101.
2. Al-Helo S, Alomari RS, Ghosh S, et al. Compression fracture diagnosis in lumbar: a clinical CAD system. *Int J Comput Assist Radiol Surg.* 2013;8(3):461-469.
3. Al-Kafri AS, Sudirman S, Hussain A, et al. Boundary Delineation of MRI Images for Lumbar Spinal Stenosis Detection lough Semantic Segmentation Using Deep Neural Networks. *IEEE ACCESS.* 2019;7:43487-43501.
4. Al-Kafri AS, Sudirman S, Hussain A, et al. Boundary Delineation of MRI Images for Lumbar Spinal Stenosis Detection Through Semantic Segmentation Using Deep Neural Networks. *IEEE Access.* 2019;7:43487-43501.
5. Alharbi RH, Alshaye MB, Alkanhal MM, Alharbi NM, Alzahrani MA, Alrehaili OA. Deep Learning Based Algorithm For Automatic Scoliosis Angle Measurement. 2020.
6. Alsiddiky A, Awwad W, Bakarman K, Fouad H, Mahmoud NM. Magnetic resonance imaging evaluation of vertebral tumor prediction using hierarchical hidden Markov random field model on Internet of Medical Things (IOMT) platform. *MEASUREMENT.* 2020;159.
7. Amasya H, Cesur E, Yıldırım D, Orhan K. Validation of cervical vertebral maturation stages: Artificial intelligence vs human observer visual analysis. *Am J Orthod Dentofacial Orthop.* 2020;158(6):e173-e179.
8. Amasya H, Yildirim D, Aydogan T, Kemaloglu N, Orhan K. Cervical vertebral maturation assessment on lateral cephalometric radiographs using artificial intelligence: comparison of machine learning classifier models. *Dentomaxillofac Radiol.* 2020;49(5):20190441.
9. Andrew J, DivyaVarshini M, Barjo P, Tigga I, Ieee. Spine Magnetic Resonance Image Segmentation Using Deep Learning Techniques. 2020.
10. Ansari S, Sajjad F, Zia ul Q, Naveed N, Shafi I, Ieee. Diagnosis of Vertebral Column Disorders Using Machine Learning Classifiers. 2013.
11. Antico M, Little JP, Jennings H, et al. Deep Learning-Based Automatic Segmentation for Reconstructing Vertebral Anatomy of Healthy Adolescents and Patients With Adolescent Idiopathic Scoliosis (AIS) Using MRI Data. *IEEE ACCESS.* 2021;9:86811-86823.
12. Aouad T, Lopez-Medina C, Molto A, et al. Development of a Deep Learning Algorithm to Predict Positive MRI of the Sacroiliac Joints According to the ASAS Definition in Patients with Recent Axial Spondyloarthritis. Paper presented at: Arthritis and Rheumatology2020.
13. Arti P, Agrawal A, Adishesh A, Lahari VM, Niranjana KB. Convolutional Neural Network Models for Content Based X-Ray Image Classification. 2019.
14. Aubert B, Vazquez C, Cresson T, Parent S, De Guise J, Ieee. AUTOMATIC SPINE AND PELVIS DETECTION IN FRONTAL X-RAYS USING DEEP NEURAL NETWORKS FOR PATCH DISPLACEMENT LEARNING. 2016.
15. Aubert B, Vazquez C, Cresson T, Parent S, de Guise JA. Toward Automated 3D Spine Reconstruction from Biplanar Radiographs Using CNN for Statistical Spine Model Fitting. *IEEE Trans Med Imaging.* 2019;38(12):2796-2806.
16. Auloge P, Cazzato RL, Ramamurthy N, et al. Augmented reality and artificial intelligence-based navigation during percutaneous vertebroplasty: a pilot randomised clinical trial. *Eur Spine J.* 2020;29(7):1580-1589.

17. Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S. Use of artificial neural networks to decision making in patients with lumbar spinal canal stenosis. *Journal of Neurosurgical Sciences.* 2017;61(6):603-611.

18. Baka N, Leenstra S, van Walsum T. Ultrasound Aided Vertebral Level Localization for Lumbar Surgery. *IEEE Trans Med Imaging.* 2017;36(10):2138-2147.

19. Bakhous C, Aubert B, Vazquez C, Cresson T, Parent S, De Guise J. Automatic pedicles detection using convolutional neural network in a 3D spine reconstruction from biplanar radiographs. 2018.

20. Bakhous C, Vazquez C, Cresson T, Parent S, De Guise J, Ieee. A FULLY AUTOMATIC 3D RECONSTRUCTION OF SCOLIOTIC SPINE FROM BIPLANAR RADIOGRAPHS IN A SUSPENSION FRAMEWORK. 2019.

21. Bar A, Wolf L, Amitai OB, Toledano E, Elnekave E. Compression Fractures Detectionon CT. 2017.

22. Barreiro MD, Nogueira-Barbosa MH, Rangayyan RM, et al. SEMIAUTOMATIC CLASSIFICATION OF INTERVERTEBRAL DISC DEGENERATION IN MAGNETIC RESONANCE IMAGES OF THE SPINE. 2014.

23. Barreiro MdS, Nogueira-Barbosa MH, Rangayyan RM, Reis RdM, Pereyra LC, Azevedo-Marques PM. Semiautomatic classification of intervertebral disc degeneration in magnetic resonance images of the spine. Paper presented at: 5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC); 26-28 May 2014, 2014.

24. Bernstein P, Metzler J, Weinzierl M, Seifert C, Kisel W, Wacker M. Radiographic scoliosis angle estimation: spline-based measurement reveals superior reliability compared to traditional COBB method. *Eur Spine J.* 2021;30(3):676-685.

25. Beulah A, Sharmila TS, Pramod VK. Disc bulge diagnostic model in axial lumbar MR images using Intervertebral disc Descriptor (IdD). *MULTIMEDIA TOOLS AND APPLICATIONS.* 2018;77(20):27215-27230.

26. Beulah A, Sharmila TS, Pramod VK. Degenerative disc disease diagnosis from lumbar MR images using hybrid features. *VISUAL COMPUTER.* 2021.

27. Biswas JK, Dey S, Karmakar SK, Roychowdhury A, Datta S. Design of Patient Specific Spinal Implant (Pedicle Screw Fixation) using FE Analysis and Soft Computing Techniques. *Curr Med Imaging.* 2020;16(4):371-382.

28. Bressem KK, Vahldiek JL, Adams L, et al. Deep learning for detection of radiographic sacroiliitis: achieving expert-level performance. *Arthritis Res Ther.* 2021;23(1):106.

29. Cai D, Wang Z, Liu Y, Zhang Q, Han X, Liu W. Automatic Path Planning for Navigated Pedicle Screw Surgery Based on Deep Neural Network. 2019.

30. Cai Y, Landis M, Laidley DT, Kornecki A, Lum A, Li S. Multi-modal vertebrae recognition using Transformed Deep Convolution Network. *Comput Med Imaging Graph.* 2016;51:11-19.

31. Castro-Mateos I, Hua R, Pozo JM, Lazary A, Frangi AF. Intervertebral disc classification by its degree of degeneration from T2-weighted magnetic resonance images. *Eur Spine J.* 2016;25(9):2721-2727.

32. Castro-Zunti R, Park EH, Choi Y, Jin GY, Ko SB. Early detection of ankylosing spondylitis using texture features and statistical machine learning, and deep learning, with some patient ageanalysis Riel. *COMPUTERIZED MEDICAL IMAGING AND GRAPHICS.* 2020;82.

33. Chae DS, Nguyen TP, Park SJ, Kang KY, Won C, Yoon J. Decentralized convolutional neural network for evaluating spinal deformity with spinopelvic parameters. *Comput Methods Programs Biomed.* 2020;197:105699.

34. Chang CY, Buckless C, Yeh KJ, Torriani M. Automated detection and segmentation of sclerotic spinal lesions on body CTs using a deep convolutional neural network. *Skeletal Radiol.* 2021.

35. Chen B, Xu Q, Wang L, Leung S, Chung J, Li S. An Automated and Accurate Spine Curve Analysis System. *IEEE Access.* 2019;7:124596-124605.

36. Chen CC, Fang YH. Using Bi-planar X-Ray Images to Reconstruct the Spine Structure by the Convolution Neural Network. 2020.

37. Chen HY, Hsu BW, Yin YK, et al. Application of deep learning algorithm to detect and visualize vertebral fractures on plain frontal radiographs. *PLoS One.* 2021;16(1):e0245992.

38. Chen W, Liu X, Li K, et al. A deep-learning model for identifying fresh vertebral compression fractures on digital radiography. *Eur Radiol.* 2021.

39. Chianca V, Cuocolo R, Gitto S, et al. Radiomic Machine Learning Classifiers in Spine Bone Tumors: A Multi-Software, Multi-Scanner Study. *Eur J Radiol.* 2021;137:109586.

40. Chmelik J, Jakubicek R, Jan J, et al. Fully Automatic CAD System for Segmentation and Classification of Spinal Metastatic Lesions in CT Data. 2019.

41. Chmelik J, Jakubicek R, Walek P, et al. Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. *Med Image Anal.* 2018;49:76-88.

42. Cho BH, Kaji D, Cheung ZB, et al. Automated Measurement of Lumbar Lordosis on Radiographs Using Machine Learning and Computer Vision. *Global Spine J.* 2020;10(5):611-618.

43. Cho SM, Kim YG, Jeong J, Kim I, Lee HJ, Kim N. Automatic tip detection of surgical instruments in biportal endoscopic spine surgery. *Computers in Biology and Medicine.* 2021;133.

44. Choi R, Waaname K, Fujita N, et al. Measurement of vertebral rotation from moire image for screening of adolescent idiopathic scoliosis. *IIEEJ Transactions on Image Electronics and Visual Computing.* 2018;6(2):56-64.

45. Cina A, Bassani T, Panico M, et al. 2-step deep learning model for landmarks localization in spine radiographs. *Sci Rep.* 2021;11(1):9482.

46. Clement A, Whyne C, Hardisty M. Automated segmentation of osteoblastic vertebral metastasis: a radiomics approach. *COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING-IMAGING AND VISUALIZATION.* 2021.

47. Colombo T, Mangone M, Agostini F, et al. Supervised and unsupervised learning to classify scoliosis and healthy subjects based on non-invasive rasterstereography analysis. *PLoS One.* 2021;16(12):e0261511.

48. De Guio F, Ahmed EH, Michelet F, Shevroja E, Lamy O, Hans D. A deep learning artificial intelligence (AI) method to improve spine segmentation in AP spine DXA scans and subsequently its BMD and TBS accuracy. Journal of Bone and Mineral Research; 2019.

49. Deng Q, Huang Q. Detecting Vertebra Landmarks From Ultrasound Image Using Single Shot MultiBox Detector. 2019.

50. Derkatch S, Kirby C, Kimelman D, Jozani MJ, Davidson JM, Leslie WD. Identification of Vertebral Fractures by Convolutional Neural Networks to Predict Nonvertebral and Hip Fractures: A Registry-based Cohort Study of Dual X-ray Absorptiometry. *Radiology.* 2019;293(2):405-411.

51. Devikanniga D, Raj RJS. Classification of osteoporosis by artificial neural network based on monarch butterfly optimisation algorithm. *HEALTHCARE TECHNOLOGY LETTERS.* 2018;5(2):70-75.

52. Doerr SA, Uneri A, Huang YX, et al. Data-Driven Detection and Registration of Spine Surgery Instrumentation in Intraoperative Images. 2021.

53. Durand WM, Lafage R, Hamilton DK, et al. Artificial intelligence clustering of adult spinal deformity sagittal plane morphology predicts surgical characteristics, alignment, and outcomes. *Eur Spine J.* 2021;30(8):2157-2166.

54. Ebrahimzadeh E, Fayaz F, Nikravan M, Ahmadi F, Dolatabad MR. TOWARDS AN AUTOMATIC DIAGNOSIS SYSTEM FOR LUMBAR DISC HERNIATION: THE SIGNIFICANCE OF LOCAL SUBSET FEATURE SELECTION. *BIOMEDICAL ENGINEERING-APPLICATIONS BASIS COMMUNICATIONS.* 2018;30(6).

55. Erjiang E, Wang T, Yang L, et al. Machine Learning Can Improve Clinical Detection of Low BMD: The DXA-HIP Study. *J Clin Densitom.* 2021;24(4):527-537.

56. Ernst P, Hille G, Hansen C, Tonnies K, Rak M. A CNN-Based Framework for Statistical Assessment of Spinal Shape and Curvature in Whole-Body MRI Images of Large Populations. 2019.

57. Esfandiari H, Newell R, Anglin C, Street J, Hodgson AJ. A deep learning framework for segmentation and pose estimation of pedicle screw implants based on C-arm fluoroscopy. *Int J Comput Assist Radiol Surg.* 2018;13(8):1269-1282.

58. Esfandiari H, Weidert S, Kövesházi I, Anglin C, Street J, Hodgson AJ. Deep learning-based X-ray inpainting for improving spinal 2D-3D registration. *Int J Med Robot.* 2021;17(2):e2228.

59. Faleiros MC, Nogueira-Barbosa MH, Dalto VF, et al. Machine learning techniques for computer-aided classification of active inflammatory sacroiliitis in magnetic resonance imaging. *Adv Rheumatol.* 2020;60(1):25.

60. Fan G, Liu H, Wang D, et al. Deep learning-based lumbosacral reconstruction for difficulty prediction of percutaneous endoscopic transforaminal discectomy at L5/S1 level: A retrospective cohort study. *Int J Surg.* 2020;82:162-169.

61. Fan G, Liu H, Wu Z, et al. Deep Learning-Based Automatic Segmentation of Lumbosacral Nerves on CT for Spinal Intervention: A Translational Study. *AJNR Am J Neuroradiol.* 2019;40(6):1074-1081.

62. Fan X, Zhang X, Zhang Z, Jiang Y. Deep Learning on MRI Images for Diagnosis of Lung Cancer Spinal Bone Metastasis. *Contrast Media Mol Imaging.* 2021;2021:5294379.

63. Fan XJ, Zhang XY, Zhang ZB, Jiang YF. Deep Learning-Based Identification of Spinal Metastasis in Lung Cancer Using Spectral CT Images. *SCIENTIFIC PROGRAMMING.* 2021;2021.

64. Fang L, Liu J, Mao R. Automatic Segmentation and 3D Reconstruction of Spine Based on FCN and Marching Cubes in CT Volumes. 2018.

65. Fang Y, Li W, Chen X, et al. Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks. *Eur Radiol.* 2021;31(4):1831-1842.

66. Feng SX, Liu BB, Zhang Y, Zhang XY, Li YH. Two-Stream Compare and Contrast Network for Vertebral Compression Fracture Diagnosis. *IEEE TRANSACTIONS ON MEDICAL IMAGING.* 2021;40(9):2496-2506.

67. Franke J, Grover P, Siebenwirth J, Caspari C, Dreischarf M, Putzier M. Can artificial intelligence support or even replace physicians in measuring the sagittal balance? A validation study on preoperative and postoperative images of 170 patients. Brain and Spine; 2021/01/01/, 2021.

68. Frighetto-Pereira L, Metzner GA, Azevedo-Marques PM, et al. Recognition of vertebral compression fractures in magnetic resonance images using statistics of height and width. 2016.

69. Frighetto-Pereira L, Rangayyan RM, Metzner GA, Azevedo-Marques PM, Nogueira-Barbosa MH. Recognition of vertebral compression fractures in magnetic resonance images using measures of distances to the centroid. *International Journal of Computer Assisted Radiology and Surgery.* 2016;11(1):S123-S124.

70. Frighetto-Pereira L, Rangayyan RM, Metzner GA, de Azevedo-Marques PM, Nogueira-Barbosa MH. Shape, texture and statistical features for classification of benign and malignant vertebral compression fractures in magnetic resonance images. *Comput Biol Med.* 2016;73:147-156.

71. Galbusera F, Bassani T, Brayda-Bruno M, et al. Artificial neural networks as a tool for the automated analysis of medical images of the spine. Paper presented at: European Spine Journal2015.

72. Galbusera F, Bassani T, Costa F, Brayda-Bruno M, Zerbi A, Wilke HJ. Artificial neural networks for the recognition of vertebral landmarks in the lumbar spine. *COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING-IMAGING AND VISUALIZATION.* 2018;6(4):447-452.

73. Galbusera F, Cina A, Bassani T, Panico M, Sconfienza LM. Automatic Diagnosis of Spinal Disorders on Radiographic Images:

Leveraging Existing Unstructured Datasets With Natural Language Processing. *Global Spine J.* 2021:21925682211026910.

74. Galbusera F, Niemeyer F, Bassani T, Sconfienza LM, Wilke HJ. Estimating the three-dimensional vertebral orientation from a planar radiograph: Is it feasible? *J Biomech.* 2020;102:109328.

75. Galbusera F, Niemeyer F, Wilke HJ, et al. Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. *Eur Spine J.* 2019;28(5):951-960.

76. Gao F, Liu S, Zhang X, Wang X, Zhang J. Automated Grading of Lumbar Disc Degeneration Using a Push-Pull Regularization Network Based on MRI. *J Magn Reson Imaging.* 2021;53(3):799-806.

77. Gao F, Wu M. Deep Learning-Based Denoised MRI Images for Correlation Analysis between Lumbar Facet Joint and Lumbar Disc Herniation in Spine Surgery. *J Healthc Eng.* 2021;2021:9687591.

78. Gaonkar B, Beckett J, Villaroman D, et al. Quantitative Analysis of Neural Foramina in the Lumbar Spine: An Imaging Informatics and Machine Learning Study. *Radiol Artif Intell.* 2019;1(2):180037.

79. Gaonkar B, Villaroman D, Beckett J, et al. Quantitative Analysis of Spinal Canal Areas in the Lumbar Spine: An Imaging Informatics and Machine Learning Study. *AJNR Am J Neuroradiol.* 2019;40(9):1586-1591.

80. Gaweł D, Główka P, Kotwicki T, Nowak M. Automatic Spine Tissue Segmentation from MRI Data Based on Cascade of Boosted Classifiers and Active Appearance Model. *Biomed Res Int.* 2018;2018:7952946.

81. Ghosh S, Chaudhary V, Dhillon G. Exploring the utility of axial lumbar MRI for automatic diagnosis of intervertebral disc abnormalities. 2013.

82. Goedmakers C, Lak A, Duey A, et al. Predicting adjacent segment disease with deep learning on pre-operative cervical MRI for patients undergoing single-level anterior cervical discectomy and fusion. Paper presented at: Journal of Neurosurgery2021.

83. Goedmakers CMW, Lak AM, Duey AH, et al. Deep Learning for Adjacent Segment Disease at Preoperative MRI for Cervical Radiculopathy. *Radiology.* 2021;301(3):664-671.

84. Gong H, Liu JH, Li S, Chen B. Axial-SpineGAN: simultaneous segmentation and diagnosis of multiple spinal structures on axial magnetic resonance imaging images. *PHYSICS IN MEDICINE AND BIOLOGY.* 2021;66(11).

85. González G, Washko GR, Estépar RSJ. Deep learning for biomarker regression: application to osteoporosis and emphysema on chest CT scans. Proc SPIE Int Soc Opt Eng; 2018.

86. Greer H, Gerber S, Niethammer M, et al. SCOLIOSIS SCREENING AND MONITORING USING SELF CONTAINED ULTRASOUND AND NEURAL NETWORKS. *Proc IEEE Int Symp Biomed Imaging.* 2018;2018:1500-1503.

87. Grigorieva I, Vyunnik N, Kolpinsky G. The Construction of an Individualized Spinal 3D Model Based on the X-ray Recognition. 2018.

88. Gros C, De Leener B, Badji A, et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *Neuroimage.* 2019;184:901-915.

89. Guerroumi N, Playout C, Laporte C, Cheriet F, Ieee. AUTOMATIC SEGMENTATION OF THE SCOLIOTIC SPINE FROM MR IMAGES. 2019.

90. Guha I, Nadeem SA, You CY, et al. Deep Learning Based High-Resolution Reconstruction of Trabecular Bone Microstructures from Low-Resolution CT Scans using GAN- CIRCLE. 2021.

91. Hallinan J, Zhu L, Yang K, et al. Deep Learning Model for Automated Detection and Classification of Central Canal, Lateral Recess, and Neural Foraminal Stenosis at Lumbar Spine MRI. *Radiology.* 2021;300(1):130-138.

92. Han Z, Wei B, Leung S, Nachum IB, Laidley D, Li S. Automated Pathogenesis-Based Diagnosis of Lumbar Neural Foraminal Stenosis via Deep Multiscale Multitask Learning. *Neuroinformatics.* 2018;16(3-4):325-337.

93. Han Z, Wei B, Mercado A, Leung S, Li S. Spine-GAN: Semantic segmentation of multiple spinal structures. *Med Image Anal.* 2018;50:23-35.

94. Han Z, Wei B, Xi X, Chen B, Yin Y, Li S. Unifying neural learning and symbolic reasoning for spinal medical report generation. *Med Image Anal.* 2021;67:101872.

95. Haubold J, Demircioglu A, Theysohn JM, et al. Generating virtual short tau inversion recovery (Stir) images from t1-and t2-weighted images using a conditional generative adversarial network in spine imaging. *Diagnostics.* 2021;11(9).

96. He Z, Wang Y, Qin X, et al. Classification of neurofibromatosis-related dystrophic or nondystrophic scoliosis based on image features using Bilateral CNN. *Med Phys.* 2021;48(4):1571-1583.

97. Henmar S, Simonsen EB, Berg RW. What are the gray and white matter volumes of the human spinal cord? *J Neurophysiol.* 2020;124(6):1792-1797.

98. Hetherington J, Lessoway V, Gunka V, Abolmaesumi P, Rohling R. SLIDE: automatic spine level identification system using a deep convolutional neural network. *Int J Comput Assist Radiol Surg.* 2017;12(7):1189-1198.

99. Hetherington J, Pesteie M, Lessoway VA, Abolmaesumi P, Rohling RN. Identification and Tracking of Vertebrae in Ultrasound Using Deep Networks with Unsupervised Feature Learning. 2017.

100. Hille G, Steffen J, Dünnwald M, Becker M, Saalfeld S, Tönnies KD. Spinal Metastases Segmentation in MR Imaging using Deep Convolutional Neural Networks. *ArXiv.* 2020;abs/2001.05834.

101. Hopkins BS, Weber KA, 2nd, Kesavabhotla K, Paliwal M, Cantrell DR, Smith ZA. Machine Learning for the Prediction of Cervical Spondylotic Myelopathy: A Post Hoc Pilot Study of 28 Participants. *World Neurosurg.* 2019;127:e436-e442.

102. Horng MH, Kuok CP, Fu MJ, Lin CJ, Sun YN. Cobb Angle Measurement of Spine from X-Ray Images Using Convolutional Neural Network. *Comput Math Methods Med.* 2019;2019:6357171.

103. Hsieh CI, Zheng K, Lin C, et al. Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. *Nat Commun.* 2021;12(1):5472.

104. Hu X, Chen M, Pan J, Liang L, Wang Y. Is it appropriate to measure age-related lumbar disc degeneration on the mid-sagittal MR image? A quantitative image study. *Eur Spine J.* 2018;27(5):1073-1081.

105. Huang J, Shen H, Wu J, et al. Spine Explorer: a deep learning based fully automated program for efficient and reliable quantifications of the vertebrae and discs on sagittal lumbar spine MR images. *Spine J.* 2020;20(4):590-599.

106. Huang KT, Silva MA, See AP, et al. A computer vision approach to identifying the manufacturer and model of anterior cervical spinal hardware. *J Neurosurg Spine.* 2019:1-7.

107. Huang Y, Uneri A, Jones CK, et al. 3D vertebrae labeling in spine CT: an accurate, memory-efficient (Ortho2D) framework. *Phys Med Biol.* 2021;66(12).

108. Huber FA, Stutz S, De Martini IV, et al. Perfomance of machine-learning for central lumbar spinal stenosis detection and grading. Paper presented at: Skeletal Radiology2018.

109. Huo X, Wang H, Shao K, Jing J, Tian D, Cheng L. Automatic generation of pedicle contours in 3D vertebral models. *Comput Biol Med.* 2020;116:103565.

110. Hussain D, Han SM. Computer-aided osteoporosis detection from DXA imaging. *Comput Methods Programs Biomed.* 2019;173:87-107.

111. Imran AAZ, Huang C, Tang H, et al. Fully-Automated Analysis of Scoliosis from Spinal X-Ray Images. 2020.

112. Iriondo C, Pedoia V, Majumdar S. Lumbar intervertebral disc characterization through quantitative MRI analysis: An automatic voxel-based relaxometry approach. *Magn Reson Med.* 2020;84(3):1376-1390.

113. Ishimoto Y, Jamaludin A, Cooper C, et al. Could automated machine-learned MRI grading aid epidemiological studies of lumbar

spinal stenosis? Validation within the Wakayama spine study. *BMC Musculoskelet Disord.* 2020;21(1):158.

114. İstanbullu M, Aydin M, Benveniste R, Uçan ON, Jennane R. Early diagnosis of osteoporosis using Artificial Neural Networks and Support Vector Machines. 2012.

115. Ito S, Ando K, Kobayashi K, et al. Automated Detection of Spinal Schwannomas Utilizing Deep Learning Based on Object Detection From Magnetic Resonance Imaging. *Spine (Phila Pa 1976).* 2021;46(2):95-100.

116. Iyer S, Sowmya A, Blair A, White C, Dawes L, Moses D. A Novel Approach to Vertebral Compression Fracture Detection Using Imitation Learning and Patch Based Convolutional Neural Network. 2020.

117. Jakubicek R, Chmelik J, Jan J, Ourednicek P, Lambert L, Gavelli G. Learning-based vertebra localization and labeling in 3D CT data of possibly incomplete and pathological spines. *Comput Methods Programs Biomed.* 2020;183:105081.

118. Jakubicek R, Chmelik J, Ourednicek P, Jan J. Deep-learning-based fully automatic spine centerline detection in CT data. Annu Int Conf IEEE Eng Med Biol Soc; 2019.

119. Jamaludin A, Fairbank J, Harding I, et al. Identifying Scoliosis in Population-Based Cohorts: Automation of a Validated Method Based on Total Body Dual Energy X-ray Absorptiometry Scans. *Calcif Tissue Int.* 2020;106(4):378-385.

120. Jamaludin A, Kadir T, Zisserman A. Self-supervised Learning for Spinal MRIs. 2017.

121. Jamaludin A, Kadir T, Zisserman A. SpineNet: Automated classification and evidence visualization in spinal MRIs. *Med Image Anal.* 2017;41:63-73.

122. Jamaludin A, Lootus M, Kadir T, et al. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur Spine J.* 2017;26(5):1374-1383.

123. Jamaludin A, Windsor R, Ather S, et al. Machine Learning-based Berlin Scoring of Magnetic Resonance Images of the Spine in Patients with Ankylosing Spondylitis: Analysis of Data from a Phase 3 Trial with Secukinumab. Paper presented at: ARTHRITIS & RHEUMATOLOGY2020.

124. Jamaludin A, Windsor R, Ather S, et al. MACHINE LEARNING BASED BERLIN SCORING OF MAGNETIC RESONANCE IMAGES OF THE SPINE IN PATIENTS WITH ANKYLOSING SPONDYLITIS FROM THE MEASURE 1 STUDY. Paper presented at: ANNALS OF THE RHEUMATIC DISEASES2020.

125. Jamaludin M. *Automated analysis of spinal MRI using deep learning*, University of Oxford; 2017.

126. Jia P, Zhu S, Guo L. The Clinical Value of CT Scans for the Early Diagnosis and Treatment of Spinal Fractures and Paraplegia. *JOURNAL OF HEALTHCARE ENGINEERING.* 2021;2021.

127. Jiang X, Ren M, Shuang X, et al. Multiparametric MRI-Based Radiomics Approaches for Preoperative Prediction of EGFR Mutation Status in Spinal Bone Metastases in Patients with Lung Adenocarcinoma. *J Magn Reson Imaging.* 2021;54(2):497-507.

128. Joseph F, Wolfendale H, Holmes A, Hyatt R. Improving the detection and management of vertebral fractures (VF) in an acute trust: Audit of clinical features of patients with undiagnosed osteoporotic VF on CT scan. Paper presented at: Osteoporosis International2018.

129. Julio AG, Hubert R, Jürgen W. Intraoperative registration of 2D C-arm Images with Preoperative CT data in computer assisted spine surgery: motivation to use convolutional neural networks for initial pose generator. Paper presented at: Biomedizinische Technik2019.

130. Kaddioui H, Duong L, Joncas J, et al. Convolutional Neural Networks for Automatic Risser Stage Assessment. *Radiol Artif Intell.* 2020;2(3):e180063.

131. Kalderstam J, Sadik M, Edenbrandt L, Ohlsson M. Analysis of regional bone scan index measurements for the survival of patients with prostate cancer. *BMC Med Imaging.* 2014;14:24.

132. Karhade AV, Shah AA, Bono CM, et al. Development of machine learning algorithms for prediction of mortality in spinal epidural abscess. *Spine J.* 2019;19(12):1950-1959.

133. Kashiwagi N, Tanaka H, Yamashita Y, et al. Applicability of deep learning-based reconstruction trained by brain and knee 3T MRI to lumbar 1.5T MRI. *Acta Radiol Open.* 2021;10(6):20584601211023939.

134. Kausch L, Thomas S, Kunze H, et al. Toward automatic C-arm positioning for standard projections in orthopedic surgery. *Int J Comput Assist Radiol Surg.* 2020;15(7):1095-1105.

135. Ketcha MD, Marrama M, Souza A, et al. Sinogram + image domain neural network approach for metal artifact reduction in low-dose cone-beam computed tomography. J Med Imaging (Bellingham); 2021.

136. Ketola JHJ, Inkinen SI, Karppinen J, Niinimaki J, Tervonen O, Nieminen MT. T-2-weighted magnetic resonance imaging texture as predictor of low back pain: A texture analysis-based classification pipeline to symptomatic and asymptomatic cases. *Journal Orthopaedic Research.* 2020.

137. Khan A, Iliescu D, Hines E, Hutchinson C, Sneath R. Neural Network Based Spinal Age Estimation Using Lumbar Spine Magnetic Resonance Images (MRI). 2013.

138. Khan AA, Iliescu DD, Hines EL, Hutchinson CE, Sneath RJS. Classification of Age-Related Changes in Lumbar Spine with the help of MRI Scores. 2013.

139. Kim DH, Jeong JG, Kim YJ, Kim KG, Jeon JY. Automated Vertebral Segmentation and Measurement of Vertebral Compression Ratio Based on Deep Learning in X-Ray Images. *J Digit Imaging.* 2021;34(4):853-861.

140. Kim K, Kim S, Lee YH, Lee SH, Lee HS. Performance of the deep convolutional neural network based magnetic resonance image scoring algorithm for differentiating between tuberculous and pyogenic spondylitis. *Sci Rep.* 2018;8(1):13124.

141. Kim KC, Cho HC, Jang TJ, Choi JM, Seo JK. Automatic detection and segmentation of lumbar vertebrae from X-ray images for compression fracture evaluation. *Comput Methods Programs Biomed.* 2021;200:105833.

142. Kim KC, Yun HS, Kim S, Seo JK. Automation of Spine Curve Assessment in Frontal Radiographs Using Deep Learning of Vertebral-Tilt Vector. *IEEE Access.* 2020;8:84618-84630.

143. Kim KH, Do WJ, Park SH. Improving resolution of MR images with an adversarial network incorporating images with different contrast. *Med Phys.* 2018;45(7):3120-3131.

144. Kim S, Jang H, Hong S, Hong YS, Bae WC, Hwang D. Fat-saturated image generation from multi-contrast MRIs using generative adversarial networks with Bloch equation-based autoencoder regularization. *Med Image Anal.* 2021;73:102198.

145. Kirilov NK, Kirilova EK. Classifying Dual-Energy X-ray Absorptiometry Images Using Machine Learning. 2021.

146. Kirschke J, Löffler M, El-Husseini M, et al. Automated detection of osteoporotic vertebral fractures using a conditioned variational auto-encoder and probabilistic point cloud reconstructions. Paper presented at: Neuroradiology2019.

147. Kitamura G. Hanging protocol optimization of lumbar spine radiographs with machine learning. *Skeletal Radiol.* 2021;50(9):1809-1819.

148. Koh J, Chaudhary V, Dhillon G. Disc herniation diagnosis in MRI using a CAD framework and a two-level classifier. *Int J Comput Assist Radiol Surg.* 2012;7(6):861-869.

149. Kök H, Acilar AM, İzgi MS. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. *Prog Orthod.* 2019;20(1):41.

150. Kök H, Izgi MS, Acilar AM. Determination of growth and development periods in orthodontics with artificial neural network. *Orthod Craniofac Res.* 2020.

151. Kök H, İzgi MS, Acılar AM. Evaluation of the Artificial Neural Network and Naive Bayes Models Trained with Vertebra Ratios for

Growth and Development Determination. *Turk J Orthod.* 2021;34(1):2-9.

152. Kokabu T, Kanai S, Kawakami N, et al. An algorithm for using deep learning convolutional neural networks with three dimensional depth sensor imaging in scoliosis detection. *Spine J.* 2021;21(6):980-987.

153. Kokabu T, Kawakami N, Uno K, et al. Three-dimensional depth sensor imaging to identify adolescent idiopathic scoliosis: a prospective multicenter cohort study. *Sci Rep.* 2019;9(1):9678.

154. Kokabu T, Kawakami N, Uno K, et al. An algorithm for using deep learning convolutional neural networks with 3-dimensional depth sensor imaging in scoliosis detection. Paper presented at: 36th Annual Meeting of the North American Spine Society2021.

155. Kolanu N, Silverstone EJ, Ho BH, et al. Clinical Utility of Computer-Aided Diagnosis of Vertebral Fractures From Computed Tomography Images. *J Bone Miner Res.* 2020;35(12):2307-2312.

156. Kolarik M, Burget R, Riha K, Bartusek K. Suitability of CT and MRI Imaging for Automatic Spine Segmentation Using Deep Learning. 2021.

157. Kong SH, Ahn D, Kim BR, et al. A Novel Fracture Prediction Model Using Machine Learning in a Community-Based Cohort. *JBMR Plus.* 2020;4(3):e10337.

158. Kong SH, Kim JH, Lee JW, et al. Development of Spine Radiography-Based Fracture Prediction Model Using Convolutional Neural Network. *JOURNAL OF BONE AND MINERAL RESEARCH.* 2020;35:207-207.

159. Kónya S, Natarajan TS, Allouch H, Nahleh KA, Dogheim OY, Boehm H. Convolutional neural network-based automated segmentation and labeling of the lumbar spine X-ray. *J Craniovertebr Junction Spine.* 2021;12(2):136-143.

160. Koprowski R. Automatic analysis of the trunk thermal images from healthy subjects and patients with faulty posture. *Comput Biol Med.* 2015;62:110-118.

161. Korez R, Putzier M, Vrtovec T. Computer-assisted measurement of sagittal pelvic alignment parameters from radiographic images. *Zdravniški Vestnik.* 2018;87(11/12):519-529.

162. Korez R, Putzier M, Vrtovec T. Automated Measurement of Pelvic Incidence from X-Ray Images. 2019; Cham.

163. Korez R, Putzier M, Vrtovec T. A deep learning tool for fully automated measurements of sagittal spinopelvic balance from X-ray images: performance evaluation. *Eur Spine J.* 2020;29(9):2295-2305.

164. Krishnaraj A, Barrett S, Bregman-Amitai O, et al. Simulating Dual-Energy X-Ray Absorptiometry in CT Using Deep-Learning Segmentation Cascade. *J Am Coll Radiol.* 2019;16(10):1473-1479.

165. Kuang X, Cheung JP, Wu H, Dokos S, Zhang T. MRI-SegFlow: a novel unsupervised deep learning pipeline enabling accurate vertebral segmentation of MRI images. Annu Int Conf IEEE Eng Med Biol Soc; 2020.

166. Kurachka KS, Tsalka IM. Vertebrae detection in X-ray images based on deep convolutional neural networks. 2017.

167. Lafage R, Ang B, Alshabab BS, et al. Predictive Model for Selection of Upper Treated Vertebra Using a Machine Learning Approach. *World Neurosurg.* 2021;146:e225-e232.

168. Lang N, Zhang Y, Zhang E, et al. Differentiation of spinal metastases originated from lung and other cancers using radiomics and deep learning based on DCE-MRI. *Magn Reson Imaging.* 2019;64:4-12.

169. Lee JH, Han IH, Kim DH, et al. Spine Computed Tomography to Magnetic Resonance Image Synthesis Using Generative Adversarial Networks : A Preliminary Study. *J Korean Neurosurg Soc.* 2020;63(3):386-396.

170. Lee KH, Choi ST, Lee GY, Ha YJ, Choi SI. Method for diagnosing the bone marrow edema of sacroiliac joint in patients with axial spondyloarthritis using magnetic resonance image analysis based on deep learning. *Diagnostics.* 2021;11(7).

171. Lee S, Choe EK, Kang HY, Yoon JW, Kim HS. The exploration of feature extraction and machine learning for predicting bone density from simple spine X-ray images in a Korean population. *Skeletal Radiol.* 2020;49(4):613-618.

172. Lee SJ, Lee HJ. Spinal stenosis grading in magnetic resonance imaging using deep convolutional neural networks. Paper presented at: Annual Meeting of the North American Spine Society2020.

173. Lee YH, Kim S, Suh JS, Hwang D. Learning Radiologist's Step-by-Step Skill for Cervical Spinal Injury Examination: Line Drawing, Prevertebral Soft Tissue Thickness Measurement, and Swelling Detection. *IEEE ACCESS.* 2018;6:55492-55500.

174. Lehnen NC, Haase R, Faber J, et al. Detection of Degenerative Changes on MR Images of the Lumbar Spine with a Convolutional Neural Network: A Feasibility Study. *Diagnostics (Basel).* 2021;11(5).

175. Lein GA, Nechaeva NS, Mammadova GM, Smirnov AA, Statsenko MM. Automation analysis x-ray of the spine to objectify the assessment of the severity of scoliotic deformity in idiopathic scoliosis: A preliminary report. *Pediatric Traumatology, Orthopaedics and Reconstructive Surgery.* 2020;8(3):317-326.

176. Lemay A, Gros C, Zhuo Z, et al. Automatic multiclass intramedullary spinal cord tumor segmentation on MRI with deep learning. *Neuroimage Clin.* 2021;31:102766.

177. Lewandrowsk IK, Muraleedharan N, Eddy SA, et al. Feasibility of Deep Learning Algorithms for Reporting in Routine Spine Magnetic Resonance Imaging. *Int J Spine Surg.* 2020;14(s3):S86-s97.

178. Lewandrowski KU, Muraleedharan N, Eddy SA, et al. Artificial Intelligence Comparison of the Radiologist Report With Endoscopic Predictors of Successful Transforaminal Decompression for Painful Conditions of the Lumber Spine: Application of Deep Learning Algorithm Interpretation of Routine Lumbar Magnetic. *Int J Spine Surg.* 2020;14(s3):S75-s85.

179. Lewandrowski KU, Muraleedharan N, Eddy SA, et al. Reliability Analysis of Deep Learning Algorithms for Reporting of Routine Lumbar MRI Scans. *Int J Spine Surg.* 2020;14(s3):S98-s107.

180. Li H, Luo H, Huan W, et al. Automatic lumbar spinal MRI image segmentation with a multi-scale attention network. *Neural Comput Appl.* 2021:1-14.

181. Li Q, Du Z, Yu H. Grinding trajectory generator in robot-assisted laminectomy surgery. *Int J Comput Assist Radiol Surg.* 2021;16(3):485-494.

182. Li X, Dou Q, Chen H, et al. 3D multi-scale FCN with random modality voxel dropout learning for Intervertebral Disc Localization and Segmentation from Multi-modality MR Images. *Med Image Anal.* 2018;45:41-54.

183. Li Y, Zhang Y, Zhang E, et al. Differential diagnosis of benign and malignant vertebral fracture on CT using deep learning. *Eur Radiol.* 2021.

184. Li YC, Chen HH, Horng-Shing Lu H, Hondar Wu HT, Chang MC, Chou PH. Can a Deep-learning Model for the Automated Detection of Vertebral Fractures Approach the Performance Level of Human Subspecialists? *Clin Orthop Relat Res.* 2021;479(7):1598-1612.

185. Lin L, Tao X, Pang S, et al. Multiple Axial Spine Indices Estimation via Dense Enhancing Network With Cross-Space Distance-Preserving Regularization. *IEEE J Biomed Health Inform.* 2020;24(11):3248-3257.

186. Lin L, Tao X, Yang W, et al. Quantifying Axial Spine Images Using Object-Specific Bi-Path Network. *IEEE J Biomed Health Inform.* 2021;25(8):2978-2987.

187. Liu J, Yuan C, Sun X, Sun L, Dong H, Peng Y. The measurement of Cobb angle based on spine X-ray images using multi-scale convolutional neural network. *Phys Eng Sci Med.* 2021;44(3):809-821.

188. Liu X, Han C, Lin ZY, et al. Semi-automatic quantitative analysis of the pelvic bony structures on apparent diffusion coefficient maps based on deep learning: establishment of reference ranges. *QUANTITATIVE IMAGING IN MEDICINE AND SURGERY.* 2022.

189. Liu Y, Sui X, Liu C, Kuang X, Hu Y. Automatic Lumbar Spine Tracking Based on Siamese Convolutional Network. *J Digit Imaging.* 2020;33(2):423-430.

190. Liu Y, Sui XB, Sun YC, Liu CW, Hu Y. Siamese Convolutional Networks for Tracking the Spine Motion. 2017.

191. Löffler MT, Jacob A, Scharr A, et al. Automatic opportunistic osteoporosis screening in routine CT: improved prediction of patients with prevalent vertebral fractures compared to DXA. *Eur Radiol.* 2021;31(8):6069-6077.

192. Löffler MT, Sollmann N, Burian E, et al. Opportunistic Osteoporosis Screening Reveals Low Bone Density in Patients With Screw Loosening After Lumbar Semi-Rigid Instrumentation: A Case-Control Study. *Front Endocrinol (Lausanne).* 2020;11:552719.

193. Lorenz C, Vogt N, Bornert P, Brosch T. Spine centerline extraction and efficient spine reading of MRI and CT data. 2018.

194. Lu J-T, Pedemonte S, Bizzo BC, et al. DeepSPINE: Automated Lumbar Vertebral Segmentation, Disc-level Designation, and Spinal Stenosis Grading Using Deep Learning. *ArXiv.* 2018;abs/1807.10215.

195. Luximon D, Neylon J, Lamb J. Automatic detection of vertebral body misalignment errors in cone-beam CT guided radiotherapy using a convolutional neural network. *Medical Physics.* 2021;48(6):e145.

196. Lyu J, Bi X, Banerjee S, et al. Dual-task ultrasound spine transverse vertebrae segmentation network with contour regularization. *Computerized Medical Imaging and Graphics.* 2021;89.

197. Lyu J, Ling SH, Banerjee S, et al. Ultrasound volume projection image quality selection by ranking from convolutional RankNet. *Comput Med Imaging Graph.* 2021;89:101847.

198. Ma S, Huang Y, Che X, Gu R. Faster RCNN-based detection of cervical spinal cord injury and disc degeneration. *J Appl Clin Med Phys.* 2020;21(9):235-243.

199. Magalhães Tenorio AP, Calil Faleiros M, Raniery Ferreira Junior J, et al. Radiomics association of MRI texture features with spondyloarthritis and sacroiliitis. Paper presented at: International Journal of Computer Assisted Radiology and Surgery2018.

200. Magnide E, Tchaha GW, Joncas J, et al. Automatic bone maturity grading from EOS radiographs in Adolescent Idiopathic Scoliosis. *Comput Biol Med.* 2021;136:104681.

201. Makaremi M, Lacaule C, Mohammad-Djafari A. Deep Learning and Artificial Intelligence for the Determination of the Cervical Vertebra Maturation Degree from Lateral Radiography. *ENTROPY.* 2019;21(12).

202. Maki S, Furuya T, Horikoshi T, et al. A Deep Convolutional Neural Network With Performance Comparable to Radiologists for Differentiating Between Spinal Schwannoma and Meningioma. *Spine (Phila Pa 1976).* 2020;45(10):694-700.

203. Mandel W, Oulbacha R, Roy-Beaudry M, Parent S, Kadoury S. Image-Guided Tethering Spine Surgery With Outcome Prediction Using Spatio-Temporal Dynamic Networks. *IEEE Trans Med Imaging.* 2021;40(2):491-502.

204. Mbarki W, Bouchouicha M, Tshienda FT, Moreau E, Sayadi M. Herniated Lumbar Disc Generation and Classification Using Cycle Generative Adversarial Networks on Axial View MRI. *ELECTRONICS.* 2021;10(8).

205. McCoy DB, Dupont SM, Gros C, et al. Convolutional Neural Network-Based Automated Segmentation of the Spinal Cord and Contusion Injury: Deep Learning Biomarker Correlates of Motor Impairment in Acute Spinal Cord Injury. *AJNR Am J Neuroradiol.* 2019;40(4):737-744.

206. Merali Z, Wang JZ, Badhiwala JH, Witiw CD, Wilson JR, Fehlings MG. A deep learning model for detection of cervical spinal cord compression in MRI scans. *Sci Rep.* 2021;11(1):10473.

207. Messina C, Piodi LP, Grossi E, et al. Artificial neural network analysis of bone quality DXA parameters response to teriparatide in fractured osteoporotic patients. *PLoS One.* 2020;15(3):e0229820.

208. Mezghani N, Phan P, Mitiche A, Labelle H, de Guise JA. A Kohonen neural network description of scoliosis fused regions and their corresponding Lenke classification. *Int J Comput Assist Radiol Surg.* 2012;7(2):257-264.

209. Miao Y, Sun Y, Li S, Zhang P, Yang Y, Hu Y. Spinal Neoplasm Image Inpainting with Deep Convolutional Neutral Networks [*]. 2019.

210. Miura M, Maki S, Miura K, et al. Automated detection of cervical ossification of the posterior longitudinal ligament in plain lateral radiographs of the cervical spine using a convolutional neural network. *Sci Rep.* 2021;11(1):12702.

211. Mohamed EI, Meshref RA, Abdel-Mageed SM, Moustafa MH, Badawi MI, Darwish SH. A Novel Morphological Analysis of DXA-DICOM Images by Artificial Neural Networks for Estimating Bone Mineral Density in Health and Disease. *J Clin Densitom.* 2019;22(3):382-390.

212. Mohanty K, Yousefian O, Karbalaeisadegh Y, Ulrich M, Grimal Q, Muller M. Artificial neural network to estimate micro-architectural properties of cortical bone using ultrasonic attenuation: A 2-D numerical study. *Comput Biol Med.* 2019;114:103457.

213. Monchka BA, Kimelman D, Lix LM, Leslie WD. Feasibility of a generalized convolutional neural network for automated identification of vertebral compression fractures: The Manitoba Bone Mineral Density Registry. *Bone.* 2021;150:116017.

214. Moraes DR, Casati JPB, Rodrigues ELL, Ieee. Analysis of Polynomial Behavior of the C3 Cervical Concavity to Bone Age Estimation Using Artificial Neural Networks. 2013.

215. Mourad R, Kolisnyk S, Kim J, Rusakov A, Lebl D. Performance of Artificial Intelligence in Determining Candidacy for Lumbar Stenosis Surgery. 2021.

216. Muehlematter UJ, Mannil M, Becker AS, et al. Vertebral body insufficiency fractures: detection of vertebrae at risk on standard CT images using texture analysis and machine learning. *Eur Radiol.* 2019;29(5):2207-2217.

217. Murata K, Endo K, Aihara T, et al. Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. *Sci Rep.* 2020;10(1):20031.

218. Murata K, Endo K, Aihara T, et al. Use of residual neural network for the detection of ossification of the posterior longitudinal ligament on plain cervical radiography. *Eur Spine J.* 2021;30(8):2185-2190.

219. Mustapha A, Hussain A, Samad SA, Zulkifley MA, Diyana Wan Zaki WM, Hamid HA. Design and development of a content-based medical image retrieval system for spine vertebrae irregularity. *Biomed Eng Online.* 2015;14:6.

220. Nagarajan MB, Checefsky WA, Abidin AZ, et al. Characterizing Trabecular Bone structure for Assessing Vertebral Fracture Risk on Volumetric Quantitative Computed Tomography. Paper presented at: Proc SPIE Int Soc Opt Eng2015.

221. Nguyen TP, Chae DS, Park SJ, Kang KY, Yoon J. Deep learning system for Meyerding classification and segmental motion measurement in diagnosis of lumbar spondylolisthesis. *Biomedical Signal Processing and Control.* 2021;65.

222. Nicolaes J, Raeymaeckers S, Robben D, et al. Automated detection of vertebral fractures in CT using 3D convolutional neural networks. Bone Reports; 2020.

223. Niemeyer F, Galbusera F, Tao Y, Kienle A, Beer M, Wilke HJ. A Deep Learning Model for the Accurate and Reliable Classification of Disc Degeneration Based on MRI Data. *Invest Radiol.* 2021;56(2):78-85.

224. Nikravan M, Ebrahimzadeh E, Izadi MR, Mikaeili M. TOWARD A COMPUTER AIDED DIAGNOSIS SYSTEM FOR LUMBAR DISC HERNIATION DISEASE BASED ON MR IMAGES ANALYSIS. *BIOMEDICAL ENGINEERING-APPLICATIONS BASIS COMMUNICATIONS.* 2016;28(6).

225. Nissinen T, Suoranta S, Saavalainen T, et al. Detecting pathological features and predicting fracture risk from dual-energy X-ray absorptiometry images using deep learning. *Bone Rep.* 2021;14:101070.

226. Ong T, Copeland R, Thiam CN, Cerda Mas G, Marshall L, Sahota O. Integration of a vertebral fracture identification service into a fracture liaison service: a quality improvement project. *Osteoporos Int.* 2021;32(5):921-926.

227. Orosz L, Haines CM, Thomson A, et al. 74. Novel artificial intelligence algorithm can accurately and independently measure spinopelvic parameters. The Spine Journal; 2021/09/01/, 2021.

228. Ost K, Jacobs WB, Evaniew N, Cohen-Adad J, Anderson D, Cadotte DW. Spinal Cord Morphology in Degenerative Cervical Myelopathy Patients; Assessing Key Morphological Characteristics Using Machine Vision Tools. *J Clin Med.* 2021;10(4).

229. Oulbacha R, Kadoury S. MRI to C-arm spine registration through Pseudo-3D CycleGANs with differentiable histograms. *Medical Physics.* 2020;47(12):6319-6333.

230. P G, J S, C C, M D, M P, J F. Can artificial intelligence support or even replace physicians

in measuring the sagittal balance? – a validation study on

preoperative and postoperative images of 170 patients. Paper presented at: 15th German Spine Congres; 2020/11/01, 2020.

231. Page J, Moser F, Maya M, Prasad R, Pressman B. Validation of a Deep Learning Based Algorithm to Diagnose Vertebral Compression Fractures. Arthritis and Rheumatology; 2020.

232. Paliwal M, Weber KA, 2nd, Smith AC, et al. Fatty infiltration in cervical flexors and extensors in patients with degenerative cervical myelopathy using a multi-muscle segmentation model. *PLoS One.* 2021;16(6):e0253863.

233. Pan Q, Zhang K, He L, et al. Automatically Diagnosing Disk Bulge and Disk Herniation With Lumbar Magnetic Resonance Images by Using Deep Convolutional Neural Networks: Method Development Study. *JMIR Med Inform.* 2021;9(5):e14755.

234. Pan Y, Chen Q, Chen T, et al. Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays. *Eur Spine J.* 2019;28(12):3035-3043.

235. Pan Y, Shi D, Wang H, et al. Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening. *Eur Radiol.* 2020;30(7):4107-4116.

236. Pang S, Leung S, Ben Nachum I, Feng Q, Li S. Direct Automated Quantitative Measurement of Spine via Cascade Amplifier Regression Network. 2018; Cham.

237. Pang S, Su Z, Leung S, et al. Direct automated quantitative measurement of spine by cascade amplifier regression network with manifold regularization. *Med Image Anal.* 2019;55:103-115.

238. Peng L, Zhang G, Zuo H, Lan L, Zhou X. Surgical Design Optimization of Proximal Junctional Kyphosis. *J Healthc Eng.* 2020;2020:8886599.

239. Pesteie M, Abolmaesumi P, Ashab HA, et al. Real-time ultrasound image classification for spine anesthesia using local directional Hadamard features. *Int J Comput Assist Radiol Surg.* 2015;10(6):901-912.

240. Pesteie M, Lessoway V, Abolmaesumi P, Rohling R. Automatic Midline Identification in Transverse 2-D Ultrasound Images of the Spine. *Ultrasound Med Biol.* 2020;46(10):2846-2854.

241. Pesteie M, Lessoway V, Abolmaesumi P, Rohling RN. Automatic Localization of the Needle Target for Ultrasound-Guided Epidural Injections. *IEEE Trans Med Imaging.* 2018;37(1):81-92.

242. Petibon Y, Fahey F, Cao XH, et al. Detecting lumbar lesions in Tc-99m-MDP SPECT by deep learning: Comparison with physicians. *MEDICAL PHYSICS.* 2020.

243. Petraikin A, Belaya Z, Gombolevsky V, et al. Assessment ofvertebralbodydeformityforthe diagnosis of vertebral fractures in a routine chest ct scan using artificial intelligence. Paper presented at: Osteoporosis International2020.

244. Phan P, Mezghani N, Wai EK, de Guise J, Labelle H. Artificial neural networks assessing adolescent idiopathic scoliosis: comparison with Lenke classification. *Spine J.* 2013;13(11):1527-1533.

245. Popescu D, Amza CG, Laptoiu D, Amza G. Competitive Hopfield Neural Network Model for Evaluating Pedicle Screw Placement Accuracy. *STROJNISKI VESTNIK-JOURNAL OF MECHANICAL ENGINEERING.* 2012;58(9):509-516.

246. Porto L, Rohling R, Ieee. IMPROVING INTERPRETABILITY OF 2-D ULTRASOUND OF THE LUMBAR SPINE. 2020.

247. Prasetio RT, Riana D, Ieee. A Comparison of Classification Methods in Vertebral Column Disorder with the Application of Genetic Algorithm and Bagging. 2015.

248. Raghavendra U, Bhat NS, Gudigar A, Acharya UR. Automated system for the detection of thoracolumbar fractures using a CNN architecture. *FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE.* 2018;85:184-189.

249. Raja G, Mohan J. Hybrid Techniques for MRI Spine Images Classification. *JOURNAL OF SCIENTIFIC & INDUSTRIAL RESEARCH.* 2020;79(9):829-832.

250. Rastegar S, Vaziri M, Qasempour Y, et al. Radiomics for classification of bone mineral loss: A machine learning study. *Diagn Interv Imaging.* 2020;101(9):599-610.

251. Ren M, Yang H, Lai Q, et al. MRI-based radiomics analysis for predicting the EGFR mutation based on thoracic spinal metastases in lung adenocarcinoma patients. *Med Phys.* 2021;48(9):5142-5151.

252. Rhee DJ, Cardenas CE, Elhalawani H, et al. Automatic detection of contouring errors using convolutional neural networks. *Med Phys.* 2019;46(11):5086-5097.

253. Roller BL, Boutin RD, O'Gara TJ, et al. Accurate prediction of lumbar microdecompression level with an automated MRI grading system. *Skeletal Radiol.* 2021;50(1):69-78.

254. Roth HR, Lu L, Liu J, et al. Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation. *IEEE Trans Med Imaging.* 2016;35(5):1170-1181.

255. Roth HR, Yao J, Lu L, Stieger J, Burns JE, Summers RM. Detection of Sclerotic Spine Metastases via Random Aggregation of Deep Convolutional Neural Network Classifications. *ArXiv.* 2014;abs/1407.5976.

256. Rothstock S, Weiss HR, Krueger D, Paul L. Clinical classification of scoliosis patients using machine learning and markerless 3D surface trunk data. *Med Biol Eng Comput.* 2020;58(12):2953-2962.

257. Rueckel J, Sperl JI, Kaestle S, et al. Reduction of missed thoracic findings in emergency whole-body computed tomography using artificial intelligence assistance. *Quant Imaging Med Surg.* 2021;11(6):2486-2498.

258. Russell E, Fower E, Javaid K. Opportunistic identification of vertebral fractures from cross-sectional imaging and impact on fracture liaison services. Paper presented at: Journal of Bone and Mineral Research2017.

259. Rzecki K, Kucybała I, Gut D, et al. Fully automated algorithm for the detection of bone marrow oedema lesions in patients with axial spondyloarthritis – Feasibility study. *Biocybernetics and Biomedical Engineering.* 2021;41(2):833-853.

260. Salehi E, Khanbare S, Yousefi H, Sharpasand H, Sheyjani OS, Ieee. Deep Convolutional Neural Networks for Automated Diagnosis of Disc Herniation on Axial MRI. 2019.

261. Salehi E, Yousefi H, Rashidi H, Ghanaatti H, Ieee. Automatic Diagnosis of Disc Herniation In Two-dimensional MR Images with Combination of Distinct Features Using Machine Learning Methods. 2019.

262. Salehinejad H, Ho E, Lin HM, et al. Deep Sequential Learning For Cervical Spine Fracture Detection On Computed Tomography Imaging. 2021.

263. Samanu NS, Zulkifley MA, Hussain A, Ieee. Semi-Automatic Spine Extraction for Disc Space Narrowing Diagnosis. 2015.

264. Schaffert R, Weiß M, Wang J, Borsdorf A, Maier A. Learning-based correspondence estimation for 2-D/3-D registration. In: *Bildverarbeitung für die Medizin 2020.* Springer; 2020:222-228.

265. Schwartz JT, Cho BH, Tang P, et al. Deep Learning Automates Measurement of Spinopelvic Parameters on Lateral Lumbar Radiographs. *Spine (Phila Pa 1976).* 2021;46(12):E671-e678.

266. Schwartz JT, Valliani AA, Arvind V, et al. Identification of Anterior Cervical Spinal Instrumentation Using A Smartphone Application Powered by Machine Learning. Paper presented at: Spine (Phila Pa 1976)2021.

267. Seo H, Hwang J, Jeong T, Shin J. Comparison of Deep Learning Models for Cervical Vertebral Maturation Stage Classification on Lateral Cephalometric Radiographs. *J Clin Med.* 2021;10(16).

268. Seo JW, Lim SH, Jeong JG, Kim YJ, Kim KG, Jeon JY. A deep learning algorithm for automated measurement of vertebral body compression from X-ray images. *Sci Rep.* 2021;11(1):13732.

269. Sha G, Wu J, Yu B. Detection of Spinal Fracture Lesions Based on Improved Faster-RCNN. 2020.

270. Sha G, Wu J, Yu B. Detection of Spinal Fracture Lesions based on Improved Yolov2. 2020.

271. Sha G, Wu J, Yu B. Spinal fracture lesions segmentation based on U-net. 2020.

272. Shen H, Huang J, Zheng Q, et al. A Deep-Learning-Based, Fully Automated Program to Segment and Quantify Major Spinal Components on Axial Lumbar Spine Magnetic Resonance Images. *Phys Ther.* 2021;101(6).

273. Shin Y, Han K, Lee YH. Temporal Trends in Cervical Spine Curvature of South Korean Adults Assessed by Deep Learning System Segmentation, 2006-2018. *JAMA Netw Open.* 2020;3(10):e2020961.

274. Shinde JV, Joshi YV, Manthalkar RR. Multidomain Feature Level Fusion for Classification of Lumbar Intervertebral Disc Using Spine MR Images. *IETE JOURNAL OF RESEARCH.* 2020.

275. Shinde JV, Joshi YV, Manthalkar RR, Joshi. Machine learning-based approach for segmentation of intervertebral disc degeneration from lumbar section of spine using MRI images. *Bio-Algorithms and Med-Systems.* 2021.

276. Shioji M, Yamamoto T, Ibata T, Tsuda T, Adachi K, Yoshimura N. Artificial neural networks to predict future bone mineral density and bone loss rate in Japanese postmenopausal women. *BMC Res Notes.* 2017;10(1):590.

277. Siemionow K, Forsthoefel C, Foy M, Gawel D, Luciano C. Autonomous lumbar spine pedicle screw planning using machine learning: A validation study. *Journal of Craniovertebral Junction and Spine.* 2021;12(3):223-227.

278. Siemionow K, Luciano C, Forsthoefel C, Aydogmus S. Autonomous image segmentation and identification of anatomical landmarks from lumbar spine intraoperative computed tomography scans using machine learning: A validation study. *J Craniovertebr Junction Spine.* 2020;11(2):99-103.

279. Small JE, Osler P, Paul AB, Kunst M. CT Cervical Spine Fracture Detection Using a Convolutional Neural Network. *AJNR Am J Neuroradiol.* 2021;42(7):1341-1347.

280. Staartjes VE, de Wispelaere MP, Vandertop WP, Schröder ML. Deep learning-based preoperative predictive analytics for patient-reported outcomes following lumbar discectomy: feasibility of center-specific modeling. *Spine Journal.* 2019;19(5):853-861.

281. Steffen J, Hille G, Becker M, Saalfeld S, Tonnies K. Automatic Segmentation of Necrosis Zones after Radiofrequency Ablation of Spinal Metastases. 2020.

282. Sudo H, Kokabu T, Abe Y, et al. Automated noninvasive detection of idiopathic scoliosis in children and adolescents: A principle validation study. *Scientific Reports.* 2018;8(1):17714.

283. Suri A, Jones BC, Ng G, et al. A deep learning system for automated, multi-modality 2D segmentation of vertebral bodies and intervertebral discs. *Bone.* 2021;149:115972.

284. Tajdari M, Maqsood A, Li H, Saha S, Sarwark JF, Liu WK. Artificial intelligence data-driven 3D model for AIS. *Stud Health Technol Inform.* 2021;280:141-145.

285. Tajdari M, Pawar A, Li H, et al. Image-based modelling for Adolescent Idiopathic Scoliosis: Mechanistic machine learning analysis and prediction. *Computer Methods in Applied Mechanics and Engineering.* 2021;374.

286. Takaki T, Murakami S, Watanabe R, Aoki T, Fujibuchi T. Calculating the target exposure index using a deep convolutional neural network and a rule base. *Phys Med.* 2020;71:108-114.

287. Tan WK, Hassanpour S, Heagerty PJ, et al. Comparison of Natural Language Processing Rules-based and Machine-learning Systems to Identify Lumbar Spine Imaging Findings Related to Low Back Pain. *Acad Radiol.* 2018;25(11):1422-1432.

288. Tan Z, Yang K, Sun Y, et al. An Automatic Scoliosis Diagnosis and Measurement System Based on Deep Learning. 2018.

289. Tang C, Zhang W, Li H, et al. CNN-based qualitative detection of bone mineral density via diagnostic CT slices for osteoporosis screening. *Osteoporos Int.* 2021;32(5):971-979.

290. Tang H, Pei XB, Huang SL, Li X, Liu C. Automatic Lumbar Spinal CT Image Segmentation With a Dual Densely Connected U-Net. *IEEE ACCESS.* 2020;8:89228-89238.

291. Tenório APM, Faleiros MC, Junior JRF, et al. A study of MRI-based radiomics biomarkers for sacroiliitis and spondyloarthritis. *Int J Comput Assist Radiol Surg.* 2020;15(10):1737-1748.

292. Thawait SK, Kim J, Klufas RA, et al. Comparison of four prediction models to discriminate benign from malignant vertebral compression fractures according to MRI feature analysis. *AJR Am J Roentgenol.* 2013;200(3):493-502.

293. Thies M, Zäch JN, Gao C, et al. A learning-based method for online adjustment of C-arm Cone-beam CT source trajectories for artifact avoidance. *Int J Comput Assist Radiol Surg.* 2020;15(11):1787-1796.

294. Thomsen FSL, Delrieux CA, Pisula JI, et al. Noise reduction using novel loss functions to compute tissue mineral density and trabecular bone volume fraction on low resolution QCT. *Comput Med Imaging Graph.* 2020;86:101816.

295. Thong W, Parent S, Wu J, Aubin CE, Labelle H, Kadoury S. Three-dimensional morphology study of surgical adolescent idiopathic scoliosis patient from encoded geometric models. *Eur Spine J.* 2016;25(10):3104-3113.

296. Tirindelli M, Victorova M, Esteban J, et al. Force-Ultrasound Fusion: Bringing Spine Robotic-US to the Next "Level". *IEEE ROBOTICS AND AUTOMATION LETTERS.* 2020;5(4):5661-5668.

297. Tomita N, Cheung YY, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Comput Biol Med.* 2018;98:8-15.

298. Toufani H, Vard A, Adibi I. A pipeline to quantify spinal cord atrophy with deep learning: Application to differentiation of MS and NMOSD patients. *Phys Med.* 2021;89:51-62.

299. Tran VL, Lin H-Y, Liu H-W. MBNet: A Multi-task Deep Neural Network for Semantic Segmentation and Lumbar Vertebra Inspection on X-Ray Images. 2021; Cham.

300. Tsai JY, Hung IY, Guo YL, et al. Lumbar Disc Herniation Automatic Detection in Magnetic Resonance Imaging Based on Deep Learning. *Front Bioeng Biotechnol.* 2021;9:708137.

301. Tseng LP, Pei YC, Chen YS, Hou TH, Ou YK. Choice between Surgery and Conservative Treatment for Patients with Lumbar Spinal Stenosis: Predicting Results through Data Mining Technology. *APPLIED SCIENCES-BASEL.* 2020;10(18).

302. Tumko V, Kim J, Drayer B. Neural network model for the detection and classification of central canal stenosis of the lumbar spine in MRI images. 2021.

303. Ulivieri FM, Piodi LP, Grossi E, et al. The role of carboxy-terminal cross-linking telopeptide of type I collagen, dual x-ray absorptiometry bone strain and Romberg test in a new osteoporotic fracture risk evaluation: A proposal from an observational study. *PLoS One.* 2018;13(1):e0190477.

304. Ulivieri FM, Rinaudo L, Messina C, et al. Bone Strain Index predicts fragility fracture in osteoporotic women: an artificial intelligence-based study. *Eur Radiol Exp.* 2021;5(1):47.

305. Ulivieri FM, Rinaudo L, Piodi LP, et al. Bone strain index as a predictor of further vertebral fracture in osteoporotic women: An artificial intelligence-based analysis. *PLoS One.* 2021;16(2):e0245967.

306. Ulmert D, Kaboteh R, Fox JJ, et al. A novel automated platform for quantifying the extent of skeletal tumour involvement in prostate cancer patients using the Bone Scan Index. *Eur Urol.* 2012;62(1):78-84.

307. Unal Y, Polat K, Erdinc Kocer H. Pairwise FCM based feature weighting for improved classification of vertebral column disorders. *Comput Biol Med.* 2014;46:61-70.

308. Unal Y, Polat K, Kocer HE, Hariharan M. Detection of abnormalities in lumbar discs from clinical lumbar MRI with hybrid models. *APPLIED SOFT COMPUTING.* 2015;33:65-76.

309. Ungi T, Greer H, Sunderland KR, et al. Automatic Spine Ultrasound Segmentation for Scoliosis Visualization and Measurement. *IEEE Trans Biomed Eng.* 2020;67(11):3234-3241.

310. Upadhyay U, Singhal B, Singh M, Ieee. Spinal Stenosis Detection in MRI using Modular Coordinate Convolutional Attention Networks. 2019.

311. Varçin F, Erbay H, Çetin E, İ Ç, Kültür T. Diagnosis of Lumbar Spondylolisthesis via Convolutional Neural Networks. 2019.

312. Varçın F, Erbay H, Çetin E, Çetin İ, Kültür T. End-To-End Computerized Diagnosis of Spondylolisthesis Using Only Lumbar X-rays. *J Digit Imaging.* 2021;34(1):85-95.

313. Vergari C, Skalli W, Gajny L. A convolutional neural network to detect scoliosis treatment in radiographs. *Int J Comput Assist Radiol Surg.* 2020;15(6):1069-1074.

314. von Atzigen M, Liebmann F, Hoch A, et al. HoloYolo: A proof-of-concept study for marker-less surgical navigation of spinal rod implants with augmented reality and on-device machine learning. *Int J Med Robot.* 2021;17(1):1-10.

315. Voter AF, Larson ME, Garrett JW, Yu JJ. Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Cervical Spine Fractures. *AJNR Am J Neuroradiol.* 2021;42(8):1550-1556.

316. Wang H, Zhang T, Cheung KM, Shea GK. Application of deep learning upon spinal radiographs to predict progression in adolescent idiopathic scoliosis at first clinic visit. *EClinicalMedicine.* 2021;42:101220.

317. Wang J, Fang Z, Lang N, Yuan H, Su MY, Baldi P. A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks. *Comput Biol Med.* 2017;84:137-146.

318. Wang L, Xu Q, Leung S, Chung J, Chen B, Li S. Accurate automated Cobb angles estimation using multi-view extrapolation net. *Med Image Anal.* 2019;58:101542.

319. Wang X, Yang W, Weinreb J, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. *Sci Rep.* 2017;7(1):15415.

320. Wang X, Zheng F, Xiao R, et al. Comparison of image quality and lesion diagnosis in abdominopelvic unenhanced CT between reduced-dose CT using deep learning post-processing and standard-dose CT using iterative reconstruction: A prospective study. *Eur J Radiol.* 2021;139:109735.

321. Wang Z, Qin J, Huang J, Wang Y, Li J. Automatic diagnosis of disc herniation based on DenseNet fusion model. 2020.

322. Wardani SK, Sigit R, Setiawardhana, Putri SMS, Yunitasari D. Measurement of Spinal Curvature for Scoliosis Classification. 2018.

323. Weber KA, 2nd, Abbott R, Bojilov V, et al. Multi-muscle deep learning segmentation to automate the quantification of muscle fat infiltration in cervical spine conditions. *Sci Rep.* 2021;11(1):16567.

324. Weber KA, Smith AC, Wasielewski M, et al. Deep Learning Convolutional Neural Networks for the Automatic Quantification of Muscle Fat Infiltration Following Whiplash Injury. *Sci Rep.* 2019;9(1):7973.

325. Weng CH, Wang CL, Huang YJ, et al. Artificial Intelligence for Automatic Measurement of Sagittal Vertical Axis Using ResUNet Framework. *J Clin Med.* 2019;8(11).

326. Wilson B, Gaonkar B, Yoo B, et al. Predicting Spinal Surgery Candidacy From Imaging Data Using Machine Learning. *Neurosurgery.* 2021;89(1):116-121.

327. Wolf K, Reisert M, Beltrán SF, et al. Focal cervical spinal stenosis causes mechanical strain on the entire cervical spinal cord tissue - A prospective controlled, matched-pair analysis based on phase-contrast MRI. *Neuroimage Clin.* 2021;30:102580.

328. Wolf K, Reisert M, Beltrán SF, et al. Spinal Cord Motion in Degenerative Cervical Myelopathy: The Level of the Stenotic Segment and Gender Cause Altered Pathodynamics. *J Clin Med.* 2021;10(17).

329. Won D, Lee HJ, Lee SJ, Park SH. Spinal Stenosis Grading in Magnetic Resonance Imaging Using Deep Convolutional Neural Networks. *Spine (Phila Pa 1976).* 2020;45(12):804-812.

330. Wu H, Bailey C, Rasoulinejad P, Li S. Automatic Landmark Estimation for Adolescent Idiopathic Scoliosis Assessment Using BoostNet. Paper presented at: Medical Image Computing and Computer Assisted Intervention − MICCAI 2017; 2017//, 2017; Cham.

331. Wu H, Bailey C, Rasoulinejad P, Li S. Automated comprehensive Adolescent Idiopathic Scoliosis assessment using MVC-Net. *Med Image Anal.* 2018;48:1-11.

332. Wu P, Huang T, Juan C. Automatic Spine Vertebra segmentation in CT images using Deep Learning. 2019.

333. Wu Q, Nasoz F, Jung J, Bhattarai B, Han MV. Machine Learning Approaches for Fracture Risk Assessment: A Comparative Analysis of Genomic and Phenotypic Data in 5130 Older Men. *Calcif Tissue Int.* 2020;107(4):353-361.

334. Wu V, Ungi T, Sunderland K, Pigeau G, Schonewille A, Fichtinger G. Automatic segmentation of spinal ultrasound landmarks with U-net using multiple consecutive images for input. 2021.

335. Xia WY, Fortin M, Ahn J, et al. Automatic Paraspinal Muscle Segmentation in Patients with Lumbar Pathology Using Deep Convolutional Neural Network. 2019.

336. Xiao P, Zhang T, Dong XN, Han Y, Huang Y, Wang X. Prediction of trabecular bone architectural features by deep learning models using simulated DXA images. *Bone Rep.* 2020;13:100295.

337. Xiong X, Wang J, Hu S, Dai Y, Zhang Y, Hu C. Differentiating Between Multiple Myeloma and Metastasis Subtypes of Lumbar Vertebra Lesions Using Machine Learning-Based Radiomics. *Front Oncol.* 2021;11:601699.

338. Xue Z, Rajaraman S, Long R, Antani S, Thoma G. Gender Detection from Spine X-Ray Images Using Deep Learning. 2018.

339. Yabu A, Hoshino M, Tabuchi H, et al. Using artificial intelligence to diagnose fresh osteoporotic vertebral fractures on magnetic resonance images. *Spine J.* 2021.

340. Yacoub B, Kabakus IM, Schoepf UJ, et al. Performance of an Artificial Intelligence-Based Platform Against Clinical Radiology Reports for the Evaluation of Noncontrast Chest CT. *Acad Radiol.* 2021.

341. Yamada K, Nagahama K, Abe Y, Hyugaji Y, Takahata M, Iwasaki N. Morphological analysis of Kambin's triangle using 3D CT/MRI fusion imaging of lumbar nerve root created automatically with artificial intelligence. *Eur Spine J.* 2021;30(8):2191-2199.

342. Yang C, Jiang MY, Chen MJ, Fu MQ, Li JY, Huang QH. Automatic 3-D Imaging and Measurement of Human Spines With a Robotic Ultrasound System. *IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.* 2021;70.

343. Yang G, Fu X, Xu N, Zhang K, Wu J. A Landmark Estimation and Correction Network for Automated Measurement of Sagittal Spinal Parameters. 2020; Cham.

344. Yang HS, Kim KR, Kim S, Park JY. Deep Learning Application in Spinal Implant Identification. *Spine (Phila Pa 1976).* 2021;46(5):E318-e324.

345. Yang J, Zhang K, Fan H, et al. Development and validation of deep learning algorithms for scoliosis screening using back images. *Commun Biol.* 2019;2:390.

346. Yang L, Wei L, Yinlong Z, Haibo A, Jindong T. Automatic lumbar vertebrae detection based on feature fusion deep learning for partial occluded C-arm X-ray images. Annu Int Conf IEEE Eng Med Biol Soc; 2016.

347. Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O. Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network. *Eur Radiol.* 2020;30(6):3549-3557.

348. Yeh YC, Weng CH, Huang YJ, Fu CJ, Tsai TT, Yeh CY. Deep learning approach for automatic landmark detection and alignment analysis in whole-spine lateral radiographs. *Sci Rep.* 2021;11(1):7618.

349. Yilmaz EB, Mader AO, Fricke T, Peña JA, Meyer C, Glüer CC. Towards explainable AI: Can Visualizations via Saliency Maps Support Understanding a Neural Network's Decision on 3D CT Vertebral Fracture Discrimination? Journal of Bone and Mineral Research; 2020.

350. Yu H, Fu Y, Wei Y, et al. A Novel Framework for 3D-2D Vertebra Matching. 2019.

351. Yu S, Tan KK. Classification of Lumbar Ultrasound Images with Machine Learning. 2014.

352. Zhang B, Yu K, Ning Z, et al. Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study. *Bone.* 2020;140:115561.

353. Zhang C, Wang J, He J, Gao P, Xie G. Automated vertebral landmarks and spinal curvature estimation using non-directional part affinity fields. *Neurocomputing.* 2021;438:280-289.

354. Zhang J, Li H, Lv L, Zhang Y. Computer-Aided Cobb Measurement Based on Automatic Detection of Vertebral Slopes Using Deep Neural Network. *Int J Biomed Imaging.* 2017;2017:9083916.

355. Zhang M, Gong H, Zhang K. Prediction of lumbar vertebral strength of elderly men based on quantitative computed tomography images using machine learning. *Osteoporos Int.* 2019;30(11):2271-2282.

356. Zhang M, Kim L, Cheong R, et al. Deep-learning artificial intelligence model for automated detection of cervical spine fracture on computed tomography (CT) imaging. *Journal of Neurosurgery.* 2019;131(1):19.

357. Zhang T, Li YF, Cheung JPY, Dokos S, Wong KYK. Learning-Based Coronal Spine Alignment Prediction Using Smartphone-Acquired Scoliosis Radiograph Images. *IEEE ACCESS.* 2021;9:38287-38295.

358. Zhang T, Zhu C, Lu Q, Liu J, Diwan A, Cheung JPY. A novel tool to provide predictable alignment data irrespective of source and image quality acquired on mobile phones: what engineers can offer clinicians. *Eur Spine J.* 2020;29(3):387-395.

359. Zhang X, Li Y, Liu Y, et al. Automatic spinal cord segmentation from axial-view MRI slices using CNN with grayscale regularized active contour propagation. *Comput Biol Med.* 2021;132:104345.

360. Zhao S, Wu X, Chen B, Li S. Automatic spondylolisthesis grading from MRIs across modalities using faster adversarial recognition network. *Med Image Anal.* 2019;58:101533.

361. Zhou CA, Chan HP, Hadjiiski LM, Dong Q. Deep learning based risk stratification for treatment management of multiple myeloma with sequential MRI scans. 2021.

362. Zhou J, Damasceno PF, Chachad R, et al. Automatic Vertebral Body Segmentation Based on Deep Learning of Dixon Images for Bone Marrow Fat Fraction Quantification. *Front Endocrinol (Lausanne).* 2020;11:612.

363. Zhou JX, Zhou CA, Chan HP, Hadjiiski LM, Dong Q. Deep learning based similarity-consistency abnormality detection (SCAD) model for classification of MRI patterns of multiple myeloma (MM) infiltration. 2021.

364. Zhou X, Feng T, Zhan HC, Yuan J, Cheng Q, Wang XD. Analysis of photoacoustic signals of Hyperosteogeny and Osteoporosis. 2019.

365. Zhou Y, Liu Y, Chen Q, Gu G, Sui X. Automatic Lumbar MRI Detection and Identification Based on Deep Learning. *J Digit Imaging.* 2019;32(3):513-520.

APPENDIX C - ARTICLE 2: MATHEMATICAL FORMULATIONS AND LOSS FUNCTION DEFINITIONS

Equation 1 - Min-Max normalization formula

$$X_a = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where X represents the original value of a data point in the dataset, X_{min} and X_{max} are the minimum and maximum values in the dataset, respectively, and X_a is the normalized after applying the Min-Max normalization process.

Equation 2 - Rectified Linear Unit (ReLU) activation function

$$R(z) = \max(0, z)$$

where z represents the input to the ReLU (R) function.

Equation 3 – Softmax activation function

$$\sigma(z)_a = \frac{e^{z_a}}{\sum e^{z_a}}$$

Where z is the input vector to the Softmax function and a is an index that specifies the input vector's element. In addition, e^{za} is the exponential of the input z for class a, and the denominator is the sum of the exponentials of the inputs for all classes in the dataset

Equation 4 - Mean Squared Error loss function for regression tasks

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Where N is the number of data points in the dataset, yi and ŷi are the actual (true) and predicted values for each data point, respectively. In addition, $\frac{1}{N} \sum_{i=1}^{N}$ represent the mean operator for all N data points in the dataset.

Equation 5 - multi-class cross-entropy loss function for classification tasks

$$CE = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

Where N is the number of data points in the dataset, yi is the actual (true) value in a one-hot encoded vector format for each datapoint, and ŷi is the predicted probability distribution for each data point, In addition, $-\frac{1}{N} \sum_{i=1}^{N}$ represent the negative mean operator applied over the log probabilities for all N data points in the dataset.

Equation 6 - NNML model combined loss functions

$$L_{total} = \sum_{t}^{T} \beta(MSE_t) + \alpha(CE_t)$$

Where T is the total number of data points for which the loss is being calculated, t is the total number of tasks, β and α are weighting coefficients (hyperparameters) for the MSE and CE loss components, respectively, MSE_t is the Mean Squared Error loss for the t-th task, and CE_t is the cross-entropy for the t-th tasks. In addition, $\sum_{t}^{T}$ represent the summation over all tasks and data points.

Equation 7: Root mean square error

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

Where N is the number of data points in the dataset, yi and y Ί are the actual (true) and predicted values for each data point, respectively. In addition, $\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}$ represent the mean operator for all N data points in the dataset.

Equation 8: Mean absolute error

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

Where N is the number of data points in the dataset, yi and y Ί are the actual (true) and predicted values for each data point, respectively. In addition, $\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}$ represent the mean operator for all N data points in the dataset.

Equation 9: Mean absolute percentage error

$$MAPE = \frac{100}{N}\sum_{i=1}^{N} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \%$$

Where N is the number of data points in the dataset, yi and y Ί are the actual (true) and predicted values for each data point, respectively. In addition, $\frac{\mathbf{100}}{\boldsymbol{N}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}$ represent the averaging of the absolute percentage errors for all N data points in the dataset and multiplying by 100 to convert it into a percentage.

Equation 10: Explained variance

$$EV = 1 - \frac{Var(\hat{\boldsymbol{y}} - \boldsymbol{y})}{Var(\boldsymbol{y})}$$

Where var() is the variance of the residuals and y and yˆare the actual (true) and predicted values.

Equation 11: Root mean squared log error

$$RMSLE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(y_i) - \log(\hat{y}_i))^2}$$

Where N is the number of data points in the dataset, yi and yˆI are the actual (true) and predicted values for each data point, respectively. In addition, $\frac{1}{N}\sum_{i=1}^{N}$ represent the mean operator for all N data points in the dataset.

Equation 12: Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP and TN are the true positives and negatives classified by the model, and FP and FN are the false positives and negatives classified by the model.

Equation 13: Specificity

$$Sp = \frac{TN}{TN + FP}$$

Where TN are the true negatives classified by the model and FP are the false positives classified by the model.

Equation 14: Sensitivity

$$Se = \frac{TP}{TP + FN}$$

Where TP are the true positives classified by the model and FN are the false negatives classified by the model.

Equation 15: Precision

$$Precision = \frac{TP}{TP + FP}$$

Where TP are the true positives classified by the model and FP are the false positives classified by the model.

Equation 16: Recall

$$Recall = \frac{(Se \times N_{TP}) + (Sp \times N_{TN})}{N_{TP} + N_{TN}}$$

Where Se is the sensitivity, SP is the specificity, N_{TP} is the count of true positives classified by the model, and N_{TN} is the count of true negatives classified by the model.

Equation 17: F1 score

$$F1\ score = \frac{2 \times precision \times recall}{precission + recall}$$

APPENDIX D - ARTICLE 2: SUPPLEMENTARY COHORT DATA: NNML DEVELOPMENT POPULATION

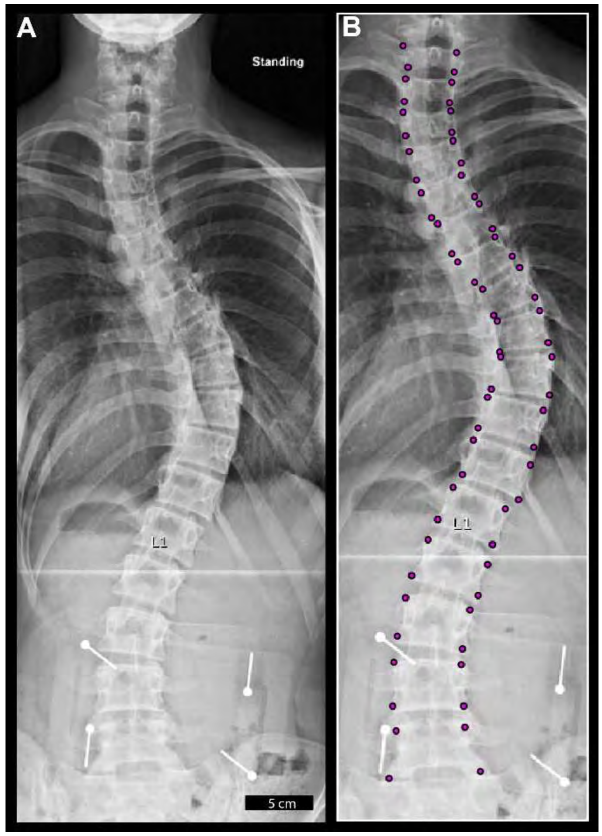


Figure D.1 Preoperative PA Radiographic Images of a Lenke 1 AIS Patient (A) with Manually Identified Vertebral Landmarks (B, red dots)

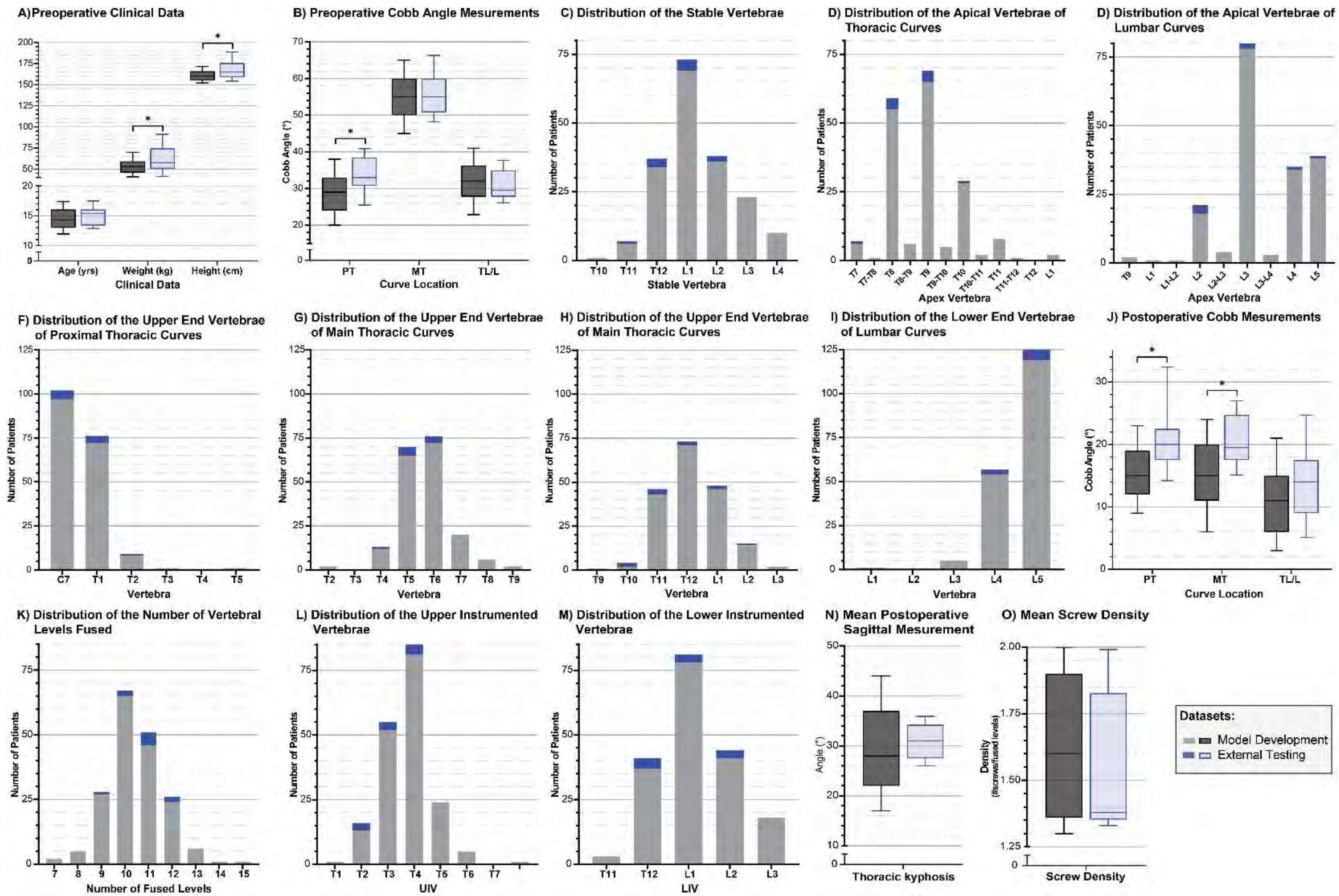


Figure D.2 Descriptive Statistics of the Patients Included in this Study

The data was stratified for each parameter according to the patients' respective datasets: model development (grey; n=179) and external performance testing (blue; n=10). In the box plots, the box's middle line indicates the mean, the box's boundaries are the 25th and 75th percentiles, and the whiskers are the 10th and 90th percentiles. * indicates a significant difference between the datasets with $p<0.05$

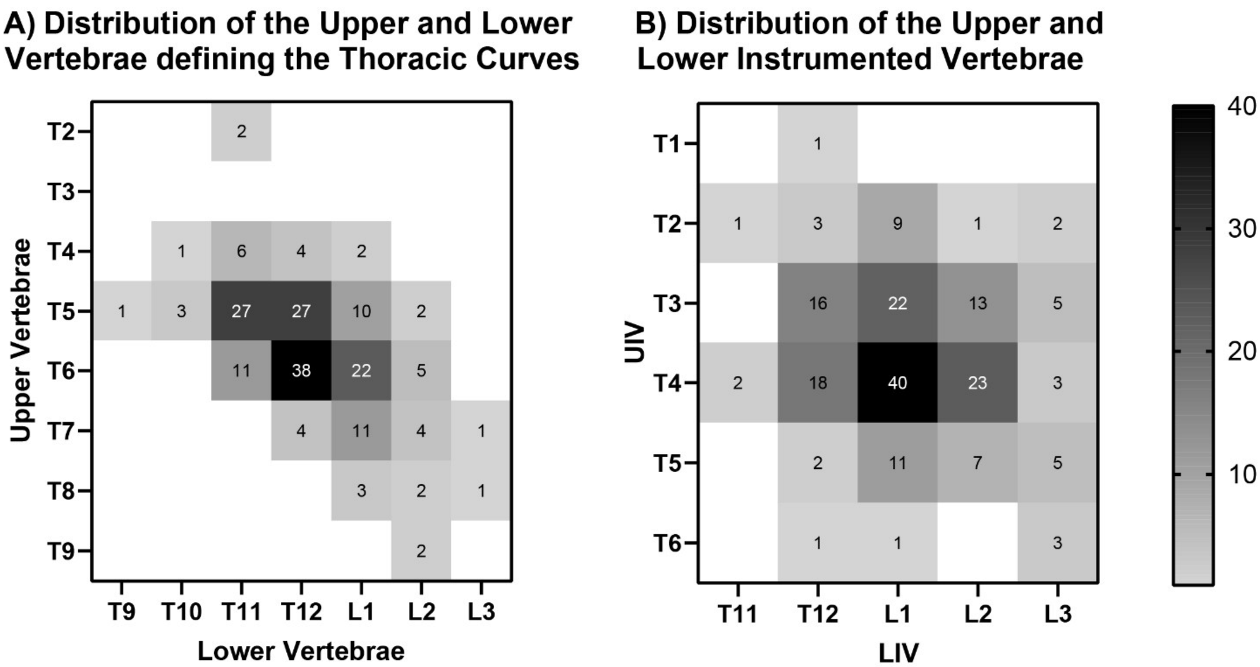


Figure D.3 Distribution of the Patient's Upper and Lower End Vertebrae defining the Thoracic Curve and Instrumented Vertebrae Included in this Study

Heat map showing the upper and lower end vertebrae defining the thoracic curves (A) and upper and lower instrumented vertebrae according to the surgery performed (B). The numbers in the square charts represent the incidence with the shade of grey indicating the relative abundances of the corresponding vertebrae in the patients included in this study.

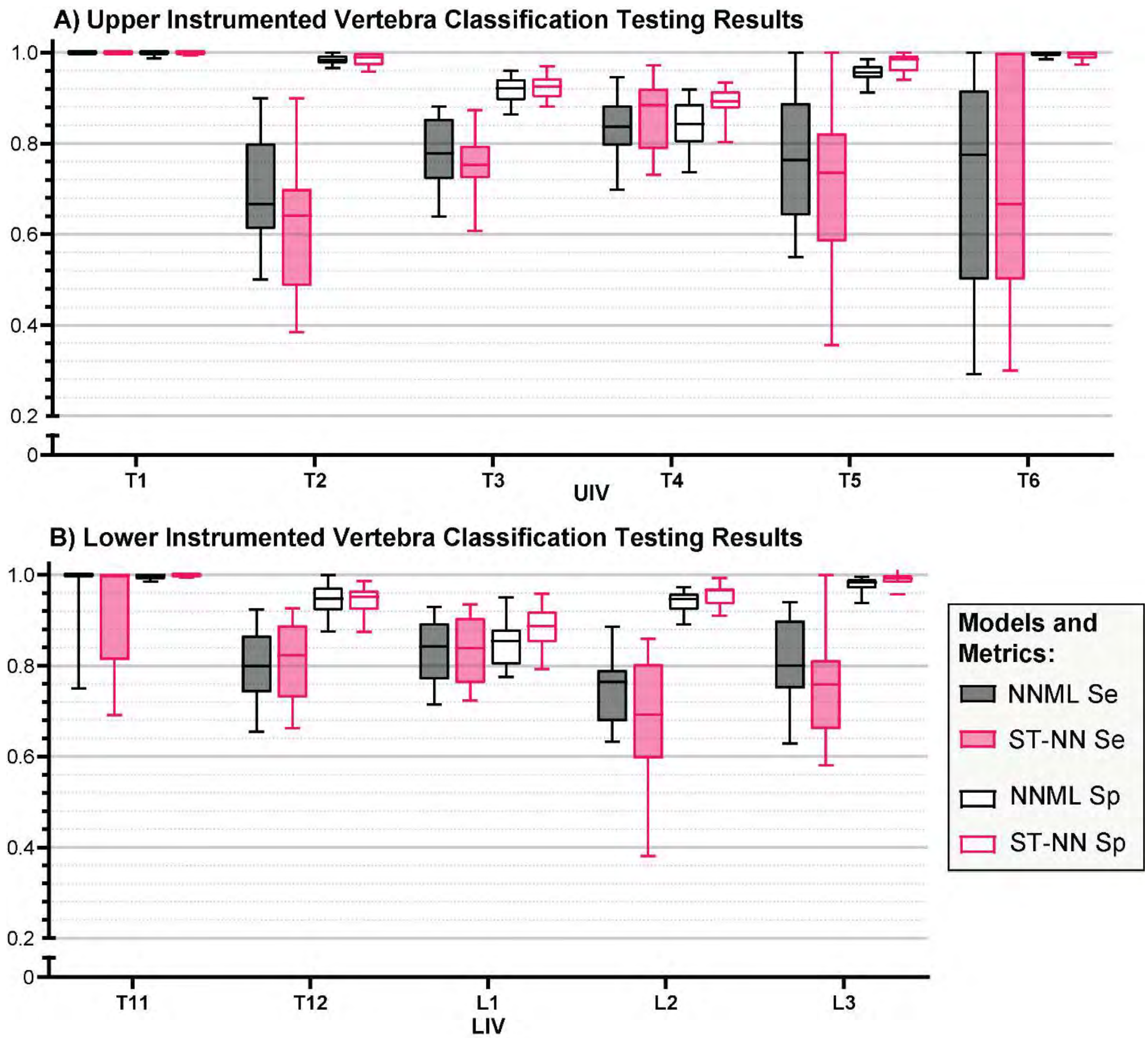


Figure D.4 Sensitivity (Se) and specificity (Sp) of UIV and LIV Classification obtained during Model Development Stages using the NNML and ST-NN Models.

Performance metrics of NNML (grey) and ST-NN (pink) models were calculated for each vertebral level included in the internal testing subset (n=36 patients) from the model development dataset over all folds (n=17 folds) from 3 experiments: 10-fold, 5-fold, and 3-fold cross-validations subsets.

In the box plots, the box's middle line indicates the mean, the box's boundaries are the 25th and 75th percentiles, and the whiskers are the 10th and 90th percentiles.

APPENDIX E - ARTICLE 2: NNML MODEL ARCHITECTURE

Model Architecture Summary

Layer (type)	Output Shape	Param #	Connected to
vertebra_landmarks_x (InputLayer)	(None, 68, 1)	0	-
vertebra_landmarks_y (InputLayer)	(None, 68, 1)	0	-
xy_by_vert (Concatenate)	(None, 68, 2)	0	vertebra_landmarks_x[0][0], vertebra_landm...
radiographic parameters (InputLayer)	(None, 8)	0	-
clinical parameters (InputLayer)	(None, 3)	0	-
vert_features (LSTM)	(None, 17)	1360	xy_by_vert[0][0]
Patient_parameters (Concatenate)	(None, 11)	0	radiographic parameters[0][0], clinical pa...
target postop (InputLayer)	(None, 4)	0	-
Combined (Concatenate)	(None, 32)	0	vert_features[0][0], Patient_parameters[0]...
Normalization (Normalization)	(None, 32)	3	Combined[0][0]
Dense1 (Dense)	(None, 50)	1650	Normalization[0][0]
dropout_10 (Dropout)	(None, 50)	0	Dense1[0][0]
Dense2 (Dense)	(None, 75)	3825	dropout_10[0][0]
dropout_11 (Dropout)	(None, 75)	0	Dense2[0][0]
Dense3 (Dense)	(None, 75)	5700	dropout_11[0][0]
dropout_12 (Dropout)	(None, 75)	0	Dense3[0][0]
Dense4N (Dense)	(None, 25)	1900	dropout_12[0][0]
Dense4C (Dense)	(None, 25)	1900	dropout_12[0][0]
dropout_16 (Dropout)	(None, 25)	0	Dense4N[0][0]
dropout_13 (Dropout)	(None, 25)	0	Dense4C[0][0]
Dense5N (Dense)	(None, 100)	2600	dropout_16[0][0]
Dense5C (Dense)	(None, 50)	1300	dropout_13[0][0]
dropout_17 (Dropout)	(None, 100)	0	Dense5N[0][0]
dropout_14 (Dropout)	(None, 50)	0	Dense5C[0][0]

Model Architecture Summary

Dense6N (Dense)	(None, 100)	10100	dropout_17[0][0]
Dense6C (Dense)	(None, 75)	3825	dropout_14[0][0]
dropout_18 (Dropout)	(None, 100)	0	Dense6N[0][0]
dropout_15 (Dropout)	(None, 75)	0	Dense6C[0][0]
dropout_19 (Dropout)	(None, 100)	0	dropout_18[0][0]
UIV (Dense)	(None, 17)	1292	dropout_15[0][0]
LIV (Dense)	(None, 17)	1292	dropout_15[0][0]
Density (Dense)	(None, 1)	101	dropout_19[0][0]
Concave (Dense)	(None, 1)	101	dropout_19[0][0]
Differential (Dense)	(None, 1)	101	dropout_19[0][0]

APPENDIX F - COMPLEMENTARY METHODOLOGICAL ASPECTS FOR CREDIBILITY ASSESSMENT OF THE COMPUTATIONAL MODEL: ASME V&V 40 CREDIBILITY FACTORS

Credibility factors are summarized from ASME V&V 40 with gradations from lowest to highest levels of credibility. The credibility activity level selected for this study is highlighted in bold.

F.1 Goals for credibility factors associated with verification.

Software Quality Assurance (SQA)
a. Very little or no SQA procedures were specified or followed
b. SQA procedures were specified and documented
c. SQA procedures were specified and documented; the software anomaly list and the software development environment are fully understood and the impact on the COU is analyzed and documented; quality metrics are tracked

Numerical Code Verification (NCV)
a. NCV was not performed
b. The numerical solution was compared to an accurate benchmark solution
c. Discretization error was quantified by comparison to an exact solution, and a grid convergence study demonstrated that the numerical solution asymptotically approached the exact solution as the discretization was refined.
d. In addition to (c), the observed order of accuracy was quantified and compared to the theoretical order of accuracy.

Discretization Error
a. No grid convergence performed
b. Applicable grid convergence analyses were performed and their respective convergence behaviors were observed to be stable, but the discretization error was not estimated.
c. Applicable grid convergence analyses were performed and discretization error was estimated.

Numerical Solver Error (NSE)
a. No solver parameter sensitivity was performed.
b. No solver parameter sensitivity was performed. Solver parameters were established based on values from a previously verified computational model.
c. Problem-specific sensitivity study was performed on solver parameters, confirming that changes in simulation results due to changes in the solver parameters were negligible relative to the model accuracy goal.

Use Error
a. Inputs and outputs were not verified.
b. Key inputs and outputs were verified by the practitioner.
c. Key inputs and outputs were verified by internal peer review.
d. Key inputs and outputs were verified by reproducing simulations as part of an external peer review.

F.2 Goals for credibility factors associated with validation

<u>Model Form</u>
a. Influence of model form assumptions was not explored.
b. Influence of expected key model form assumptions was explored.
c. Comprehensive evaluation of model form assumptions was conducted.

<u>Model Inputs: Quantification of Sensitivities</u>
a. Sensitivity analysis was not performed.
b. Sensitivity analysis on expected key parameters was performed.
c. Comprehensive sensitivity analysis was performed.

<u>Model Inputs: Quantification of Uncertainties</u>
a. Uncertainties were not identified.
b. Uncertainties on expected key inputs were identified and quantified, but were not propagated to quantitatively assess the effect on the simulation results.
c. Uncertainties on all inputs were identified and quantified, and were propagated to quantitatively assess the effect on the simulation results.

<u>Quantity of Test Samples</u>
a. A single sample was used.
b. Multiple samples were used, but not enough to be statistically relevant.
c. A statistically relevant number of samples were used.

<u>Range of Characteristics of Test Samples</u>
a. One or more samples with a single set of characteristics were included.
b. Samples representing a range of characteristics near nominal were included.
c. Samples representing the expected extreme values of the parameters were included.
d. Samples representing the entire range of parameters were included.

<u>Measurements of Test Samples</u>
a. Test samples were not measured and/or characterized.
b. One or more key characteristics of the test samples were measured.
c. All key characteristics of the test samples were measured.

<u>Uncertainty of Test Sample Measurements</u>
a. Samples were not characterized or were characterized with gross observations, and measurement uncertainty was not addressed.
b. Uncertainty analysis incorporated instrument accuracy only.
c. Uncertainty analysis incorporated instrument accuracy and repeatability (i.e., statistical treatment of repeated measurements).
d. Uncertainty analysis incorporated a comprehensive uncertainty quantification, which included instrument accuracy, repeatability, and other conditions affecting the measurements.

Quantity of Test Conditions
a. A single test condition was examined.
b. Multiple (two to four) test conditions were examined.
c. More than four test conditions were examined.

Range of Test Conditions
a. A single test condition was examined.
b. Test conditions representing a range of conditions near nominal were examined.
c. Test conditions representing the expected extreme conditions were examined.
d. Test conditions representing the entire range of conditions were examined.

Measurements of Test Conditions
a. Test conditions were qualitatively measured and/or characterized.
b. One or more key characteristics of the test conditions were measured.
c. All key characteristics of the test conditions were measured.

Uncertainty of Test Sample Measurements
a. Test conditions were not characterized or were characterized with gross observations; measurement uncertainty was not addressed.
b. Uncertainty analysis incorporated instrument accuracy only.
c. Uncertainty analysis incorporated instrument accuracy and repeatability (i.e., statistical treatment of repeated measurements).
d. Uncertainty analysis incorporated a comprehensive uncertainty quantification, which included instrument accuracy, repeatability, and other conditions affecting the measurements.

Equivalency of Input Parameters
a. The types of some inputs were dissimilar.
b. The types of all inputs were similar, but the ranges were not equivalent.
c. The types and ranges of all inputs were equivalent.

Output Comparison: Quantity
a. A single output was compared.
b. Multiple outputs were compared.

Output Comparison: Equivalency of Output Parameters
a. Types of outputs were dissimilar.
b. Types of outputs were similar.
c. Types of outputs were equivalent.

Output Comparison: Rigor of Output Compa rison
a. Visual comparison was performed.
b. Comparison was performed by determining the arithmetic difference between computational results and experimental results.
c. Uncertainty in the output of the computational model or the comparator was used in the output comparison.
d. Uncertainties in the output of the computational model and the comparator were used in the output comparison.

Output Comparison: Agreement of Output C omparison
a. The level of agreement of the output comparison was not satisfactory for key comparisons.
b. The level of agreement of the output comparison was satisfactory for key comparisons, but not all comparisons.
c. The level of agreement of the output comparison was satisfactory for all comparisons.

F.3 Goals for credibility factors associated with applicability to the COU.

<u>Relevance of the QOIs</u>

a. The QOIs from the validation activities were related, though not identical, to those for the COU.

b. A subset of the QOIs from the validation activities were identical to those for the COU.

c. The QOIs from the validation activities were identical to those for the COU.

<u>Relevance of the Validation Activities to the COU</u>.

a. There was no overlap between the ranges of the validation points and the COU

b. There was partial overlap between the ranges of the validation points and the COU

c. The COU encompassed some of the validation points

d. The COU encompassed all validation and the validation points spanned the entire COU space.

APPENDIX G - COMPLEMENTARY METHODOLOGICAL ASPECTS FOR CREDIBILITY ASSESSMENT OF THE COMPUTATIONAL MODEL: SOLVER PARAMETER SENSITIVITY STUDY

To support the verification of the computational model under the Numerical Solver Error (NSE) criterion of the ASME V&V 40 (2018) framework, a targeted sensitivity analysis was conducted to evaluate the influence of numerical solver parameters on key simulation outcomes. This activity was essential to demonstrate that the selected solver settings do not introduce significant variability in the biomechanical predictions, thereby contributing to the model's overall credibility.

Scope and aim. Under ASME V&V 40, Numerical Solver Error (NSE) was examined to confirm that the equilibrium solver settings used in the multibody (MB) model do not materially influence the study's decision-making outputs within the stated Context of Use (COU): comparative evaluation and optimization of patient-specific posterior spinal fusion (PSF) strategies in AIS. Simulations were run in MSC Adams 2019 using the validated, patient-specific MB pipeline described in Articles 3–4 and Chapters 5–6.

Design. A targeted sensitivity analysis varied six solver parameters around their defaults. The following solver parameters were tested for their potential impact on model outputs:

- **Error tolerance (error)**: Maximum allowable residual force error per iteration.
- **Stability threshold (stability)**: Level of damping threshold or numerical stability for equilibrium convergence.
- **Force imbalance limit (imbalance)**: Permissible unbalanced force/moment magnitude.
- **Time limit (tlimit)**: Maximum time allocated for solver convergence.
- **Angular limit (alimit)**: Maximum angular deviation tolerated between iterations.
- **Maximum iterations (maxit)**: Upper limit on the number of equilibrium solver iterations.

Each parameter (except *stability*) was perturbed ±20% and simulations were rerun under these conditions, except for the parameter *stability*, for which a more appropriate logarithmic variation was applied due to its scale. Specifically, simulations were conducted with values of 1e-5 and 1e-2, compared to the default 1e-3. *Maxit* additionally included 25 iterations to reflect the accelerated screening setting used during early optimization screening runs.

The test set comprised 20 representative cases: 10 derived from actual postoperative patient configurations and 10 from AI-predicted instrumentation strategies included in the optimization study. The primary outcome measures analyzed were postoperative coronal Cobb angle, TK, AVR, and average implant-related pullout forces. Statistical comparison using a mixed-effects model for paired repeated measures. Multiple comparisons within each outcome variable were conducted between all solver parameter conditions using Tukey's method to adjust for multiplicity. All outcomes are summarized as mean ± SD, with statistical significance set at $p < 0.05$.

Findings. Across all parameters and ranges, no condition differed significantly from the default setting for any outcome (all $p > 0.05$; Table G.1; Figure G.1). Mean deviations were generally negligible (≤1° for Cobb and AVR; ≤3.5 N for pullout). The only exception was TK, which showed a deviation approaching 6° under the most extreme perturbation of the *stability* factor (1e-5–1e-2). However, this occurred only in this stress-test condition; the validated default setting (*stability* = 0.001) was retained for all main analyses. Given this, and considering that <5° deviation would be regarded as acceptable variation for TK, the observed change is not expected to impact the study's conclusions.

Table G.1 Solver Parameter Sensitivity Table

P-values indicate the statistical difference between default and perturbed solver settings for each outcome variable

Parameter	***Description***	***Default***	***Tested Range***	***Cobb Angle difference (°)***	***TK difference (°)***	***AVR difference at T8 (°)***	***Pullout Force difference (N)***
error	*Residual tolerance*	*0.1*	*0.08–0.12*	*0.02 (p=0.98)*	*0.00 (p>0.99)*	*0.00 (p>0.99)*	*0.01 (p=0.99)*
Stability	*Damping factor*	*0.001*	*1e-5, 1e-2*	*0.96 (p>0.99)*	*5.70 (P=0.99)*	*0.80 (p=0.98)*	*1.98 (p=0.99)*
Imbalance	*Force imbalance limit*	*10*	*8–12*	*0.00 (p>0.99)*	*0.00 (p>0.99)*	*0.00 (p>0.99)*	*0.00 (P=0.99)*
Tlimit	*Max time (s)*	*10*	*8–12*	*0.00 (p>0.99)*	*0.00 (p>0.99)*	*0.00 (p>0.99)*	*0.00 (p>0.99)*
Alimit	*Max angle (°)*	*5*	*4–6*	*0.49 (p=0.97)*	*4.14 (p=0.99)*	*0.59 (p=0.99)*	*2.02 (p=0.84)*
Maxit	*Max iterations*	*250*	*25, 150–350*	*0.15 (p=0.09)*	*0.32 (p=0.09)*	*0.17 (p=0.09)*	*3.49 (p=0.99)*

For the parameter maxit, in addition to ±20% variations from the default of 250 iterations, a fourth test case of 25 iterations was included. This setting reflects the early configuration used in the large-scale optimization process (Chapter 6, Article 4), where simulation speed was prioritized to enable screening of over 900 surgical construct configurations per patient. At that stage, instrumentation strategies failing to converge within 25 iterations were automatically excluded from further analysis. This resulted in the removal of average of less than 2 non-converging AI-derived instrumentation options per patient, helping streamline the optimization workflow without compromising on model credibility.

A)

Scenario	error	stability	imbalance	tlimit	alimit	maxit
Default	0.1	0.001	10	10	5	250
1	**0.08**	0.001	10	10	5	250
2	**0.12**	0.001	10	10	5	250
3	0.1	**0.00001**	10	10	5	250
4	0.1	**0.01**	10	10	5	250
5	0.1	0.001	**8**	10	5	250
6	0.1	0.001	**12**	10	5	250
7	0.1	0.001	10	**8**	5	250
8	0.1	0.001	10	**12**	5	250
9	0.1	0.001	10	10	**4**	250
10	0.1	0.001	10	10	**6**	250
11	0.1	0.001	10	10	5	**25**
12	0.1	0.001	10	10	5	**150**
13	0.1	0.001	10	10	5	**350**

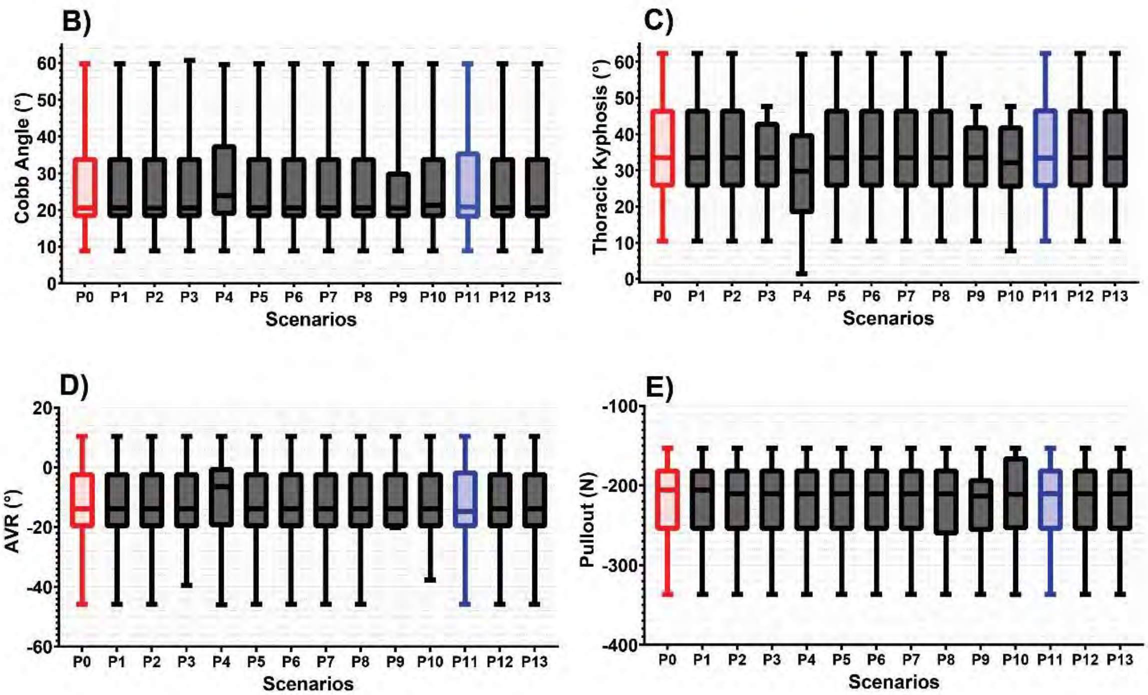


Figure G.1 Solver Parameter Sensitivity Analysis across 14 Test Scenarios

(A) Overview of solver parameter perturbations (P0–P13), where P0 corresponds to the validated default solver configuration and P11 to the reduced-iteration setting used during the optimization screening phase. (B–E) Distribution of key outcome measures under each condition: (B) postoperative MT Cobb angle, (C) thoracic kyphosis, (D) apical vertebral rotation, and (E) mean implant pullout force.

Boxplots display the 5th–95th percentile range with median values; red boxes (P0) indicate the default solver settings used for all final analyses, and blue boxes (P11) indicate the accelerated solver configuration applied during optimization screening. No significant differences were detected between P0 and any perturbed solver condition (all $p > 0.05$).

Interpretation. Varying residual tolerance, stability factor, imbalance limit, time and angular limits, and iteration cap produced no statistically significant changes in simulated MT Cobb, TK, AVR, or average pullout forces compared with the default solver configuration (all $p > 0.05$). The absolute mean deviations observed under the most extreme tests (log-scaled stability; 25-iteration cap) were numerically small and critically subordinate to the model's validated accuracy for postoperative spinal alignment outcomes in this application. In other words, the numerical solution procedure is not a dominant source of variability relative to the model-measurement agreement already established for this MB framework.

From a COU perspective (patient-specific comparison and optimization of constructs rather than prediction of absolute loads in isolation), this stability is decisive: rankings among candidate strategies, and thus optimization decisions, are robust to reasonable solver perturbations. The use of a 25-iteration cap during the screening phase did not alter conclusions, because (i) only a very small number of non-convergent AI-derived options per patient were excluded at that stage, and (ii) all shortlisted strategies were re-evaluated with default solver settings for final selection and for every analysis reported in Article 4 (Chapter 6).

Implications for credibility. These results support NSE at **level c** (ASME V&V 40) for the COU. Practically, they justify (i) the use of the default solver settings for all primary analyses in Article 3 (comparative biomechanics; default solver settings in Table G.2) and for the formal comparison between AI-derived optimized instrumentation and the actual surgical constructs in Article 4 (patient-specific optimization), and (ii) the temporary 25-iteration cap for screening of large configuration sets (reduced-time solver settings presented in Table G.2).

Table G.2 Solver Parameter Settings used during the Initial Optimization Phase for Rapid Screening of AI-derived Instrumentation Strategies

Parameter	***Description***	***Value used***
error	*Residual tolerance*	*0.1*
stability	*Damping factor*	*0.001*
imbalance	*Force imbalance limit*	*10.0*
tlimit	*Max time (s)*	*10.0*
alimit	*Max angle (°)*	*5.0*
maxit	*Max iterations*	*25*

Conclusion. The solver parameter sensitivity study demonstrates that the Adams-based equilibrium solver configuration used in this thesis provides numerically stable and decision-invariant outputs for the key clinical and biomechanical metrics that drive the developed digital-twin workflow. Accordingly, NSE credibility is established at level c for the specified COU, supporting the standardized solver settings adopted throughout (Articles 3 and 4 presented in Chapters 5 and 6). This supports the credibility of the model's verification activities and justifies the use of the standardized solver parameters used in the main study population. Together with prior validation of the MB pipeline and the optimization results (972 configurations/patient; 20 patients), these findings strengthen the overall credibility of the CM&S framework for regulatory-aligned assessment under ASME V&V 40-2018 and its intended future translational path.

APPENDIX H - COMPLEMENTARY METHODOLOGICAL ASPECTS FOR CREDIBILITY ASSESSMENT OF THE COMPUTATIONAL MODEL: POSTOPERATIVE OUTCOME VALIDATION

To support the validation of the computational model under the Output Comparison: Agreement of Output Comparison criterion of the ASME V&V 40 (2018) framework, a postoperative outcome validation study was performed. This activity aimed to determine whether the MB simulation framework can reproduce the spinal alignment achieved after PSF in AIS patients. By comparing predicted versus actual surgical outcomes, the analysis tested the accuracy of the framework in replicating clinically relevant corrections.

Scope and Aim. Within ASME V&V 40, outcome validation addresses whether model outputs are equivalent to the clinical measurements they seek to represent. The specific aim here was to evaluate the agreement between simulated and actual postoperative values of MT Cobb, TK, and AVR in AIS patients. The analysis further sought to establish whether deviations remained within a clinically meaningful threshold (±5°), thereby confirming that the framework can generate predictions suitable for patient-specific evaluation and optimization of PSF constructs.

Design. The analysis used a cohort of 35 AIS patients (Article 3 dataset), each with complete pre- and postoperative biplanar radiographs. Patient-specific 3D MB models were reconstructed using the same previously validated method based on self-calibration and optimization algorithms as described in Chapter 5. For each patient, the simulation reproduced the exact instrumentation and maneuvers applied by the surgeon. Alignment outcomes were evaluated using three clinically relevant radiographic parameters: MT Cobb angle, TK, and AVR. A ±5° equivalence margin was adopted as the threshold of clinically meaningful difference, reflecting the established accuracy of 3D reconstructions from biplanar radiographs [272]. Statistical analysis was performed using two-way repeated-measures ANOVA with patient matching to compare simulated versus actual outcomes.

Findings. Simulation predictions closely matched the actual postoperative results. Across all three measured parameters, average differences were small and consistently fell within the predefined ±5° margin of clinical relevance (Table H.1). Standard deviations also remained inside this threshold, further supporting the accuracy of the model. Statistical testing confirmed the absence of significant differences between simulated and surgical results (MT Cobb $p = 0.5893$; TK $p = 0.9852$; AVR $p = 0.3228$). With a cohort of 35 patients, the study achieved approximately 94% power to demonstrate that deviations in MT Cobb remained within the equivalence boundary, providing strong confidence in the model's predictive capacity. Full individual patient results are provided in Table H.1.

Table H.1 Actual and Simulated Postoperative Outcomes

	MT Cobb angle			TK			AVR		
Patients	Actual	Simulated	Difference	Actual	Simulated	Difference	Actual	Simulated	Difference
Patient 1	9	8	0	25	28	3	-24	-21	3
Patient 2	11	16	5	23	19	-5	-7	-7	0
Patient 3	13	12	-1	31	27	-4	-17	-16	1
Patient 4	6	2	-4	42	38	-4	1	-2	-3
Patient 5	11	12	2	9	15	5	-14	-15	0
Patient 6	12	10	-2	31	34	3	-24	-19	5
Patient 7	13	11	-1	23	23	0	-7	0	7
Patient 8	6	8	1	19	16	-3	-5	-7	-2
Patient 9	19	14	-5	31	36	5	-11	-13	-3
Patient 10	18	17	-2	40	44	4	-8	-14	-6
Patient 11	7	11	4	37	39	1	-11	-15	-4
Patient 12	14	16	1	34	29	-5	-14	-19	-5
Patient 13	15	17	2	24	26	2	-17	-18	-1
Patient 14	14	18	4	21	20	-1	-25	-22	4
Patient 15	22	17	-5	32	31	-1	6	4	-2
Patient 16	1	6	5	20	20	0	-6	-4	2
Patient 17	14	17	3	27	25	-1	10	11	1
Patient 18	9	9	0	27	32	5	-6	-6	1
Patient 19	15	14	-1	24	26	1	-30	-29	2
Patient 20	15	11	-4	17	20	2	-11	-11	0
Patient 21	12	16	4	21	17	-3	-17	-21	-4
Patient 22	14	17	3	9	6	-3	-9	-11	-2
Patient 23	29	28	-1	27	32	5	-17	-20	-2
Patient 24	13	14	2	32	38	5	-9	-10	-2
Patient 25	21	24	2	47	44	-4	-6	-12	-6
Patient 26	15	14	-2	33	29	-3	-17	-19	-1
Patient 27	3	7	4	36	31	-5	-2	-1	1
Patient 28	6	10	4	18	17	0	-15	-17	-2
Patient 29	9	10	0	18	24	6	-12	-14	-3
Patient 30	3	8	5	29	25	-4	-17	-13	5
Patient 31	12	12	0	26	28	2	-16	-10	6
Patient 32	30	25	-5	26	23	-3	-9	-13	-4
Patient 33	17	15	-2	34	35	1	-9	-14	-4
Patient 34	14	19	5	25	30	6	-5	-3	1
Patient 35	10	11	0	41	37	-4	2	-2	-4
Mean ± SD	13 ± 6	14 ± 5	**1 ± 3**	27 ± 9	28 ± 8	**0 ± 4**	-11 ± 8	-12 ± 8	**-1 ± 3**
Mean difference (absolute value)			**2.6**			**3.1**			**2.8**

Implications for Credibility. This validation provides strong evidence supporting a grade b credibility level for the Output Comparison: Agreement of Output Comparison factor. The close agreement between simulated and observed outcomes enhances confidence in the model's predictive ability. Importantly, this establishes that deviations observed in subsequent sensitivity analyses and optimization studies (Articles 3 and 4 presented in Chapters 5 and 6) are attributable to true differences in surgical input parameters rather than systematic bias in the underlying model.

Conclusion. The outcome validation demonstrated that the MB simulation framework achieves biomechanically realistic and clinically robust predictions of postoperative alignment in AIS patients. This strengthens its credibility for use as a medium–low risk decision-support tool in evaluating surgical strategies and supports its application in optimization workflows.

APPENDIX I - COMPLEMENTARY METHODOLOGICAL ASPECTS FOR CREDIBILITY ASSESSMENT OF THE COMPUTATIONAL MODEL: SENSITIVITY ANALYSIS OF INSTRUMENTATION

To support the validation of the computational model under the Model Inputs: Quantification of Sensitivities criterion of the ASME V&V 40 (2018) framework, a targeted sensitivity analysis was conducted to examine the robustness of the MB model and optimization framework to clinically relevant variations in surgical instrumentation parameters. This activity was essential to demonstrate that surgical planning decisions derived from the framework remain stable when subjected to reasonable perturbations in rod curvature, fusion levels, and screw patterns.

Scope and aim. Under ASME V&V 40, Model Input Sensitivities were assessed to ensure that clinically plausible variations in key surgical parameters do not disproportionately affect the comparative or optimization decisions within the model's COU: patient-specific evaluation and optimization of PSF constructs in AIS. The analysis specifically tested whether input variations propagate to outcome changes that might alter decision-making (e.g., ranking of strategies or interpretation of construct performance).

Design. Sensitivity analyses were performed on the 20 AIS patients included in the optimization study (Article 4, Chapter 6). Three categories of instrumentation inputs were perturbed:

- **Rod curvature:** ±10° relative to the AI-predicted baseline profile, applied independently to both concave and convex rods.
- **Fusion levels (UIV/LIV):** Upper and lower instrumented vertebrae shifted ±1 level to simulate small variations in fusion extent.
- **Screw patterns:** Five commonly described literature-based distributions were tested per patient, spanning fully instrumented (density of 2.0) to low-density constructs (density ~1.2) as described in Article 4.

When one instrumentation parameter was varied, all others were held constant to isolate its specific influence. In particular, for the screw pattern analysis, both rod curvature and fusion levels were fixed at their default values to ensure that observed effects reflected screw distribution alone. In all tests, surgical maneuvers and boundary conditions were held constant, and the validated MB workflow (Sections 5.2 and 6.2) was applied. This yielded 280 simulations (14 variations × 20 patients). The primary outcome measures analyzed were postoperative coronal MT Cobb angle, TK, AVR, and average implant-related pullout forces. Statistical comparison using a mixed-effects model for paired repeated measures. Multiple comparisons within each outcome variable were conducted between the different instrumentation conditions using Tukey's method to adjust for multiplicity. All outcomes are summarized as mean ± SD, with statistical significance set at $p < 0.05$.

Findings.

Rod curvature sensitivity (Figure I.1). Varying rod curvature by ±10° resulted in only mild deviations in MT Cobb (±1.6 ± 8.8°) and AVR (±1.8 ± 0.6°). In contrast, thoracic kyphosis was more sensitive, with average changes of ±9.1 ± 2.0°. Bone-screw pullout remained stable across conditions (variation 77 ± 23 N), and no unacceptable force increases were observed.

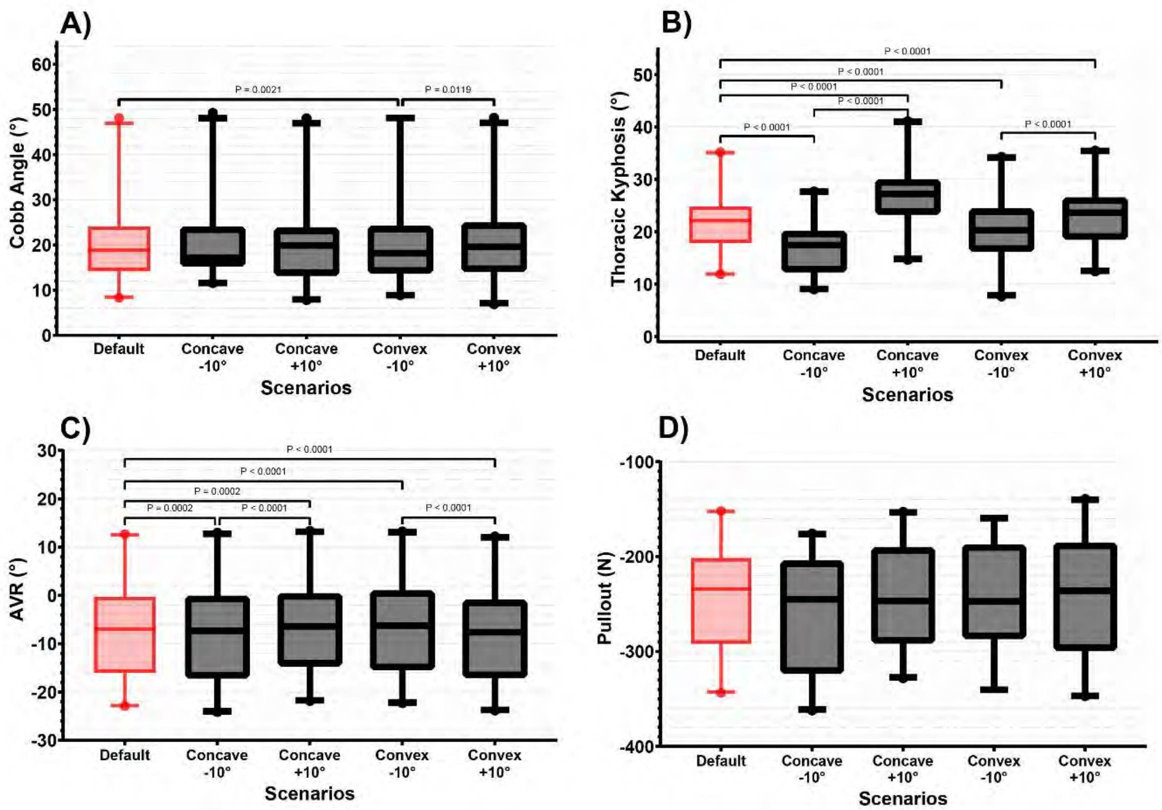


Figure I.1 Impact of Rod Curvature Variation on Simulation Outcomes

P-values indicate whether increasing or decreasing the rod curvature by 10° is statistically different than default AI-predicted rod curvature.

Fusion level sensitivity (Figure I.2).

Adjusting UIV or LIV by ±1 vertebra had the largest influence across outcomes: MT Cobb ±14.0 ± 8.6°, TK ±9.2 ± 4.8°, AVR ±3.1 ± 4.8°. While most cases remained within acceptable ranges, three instances showed >100 N increases in screw pullout force, highlighting a potential trade-off when modifying fusion extent.

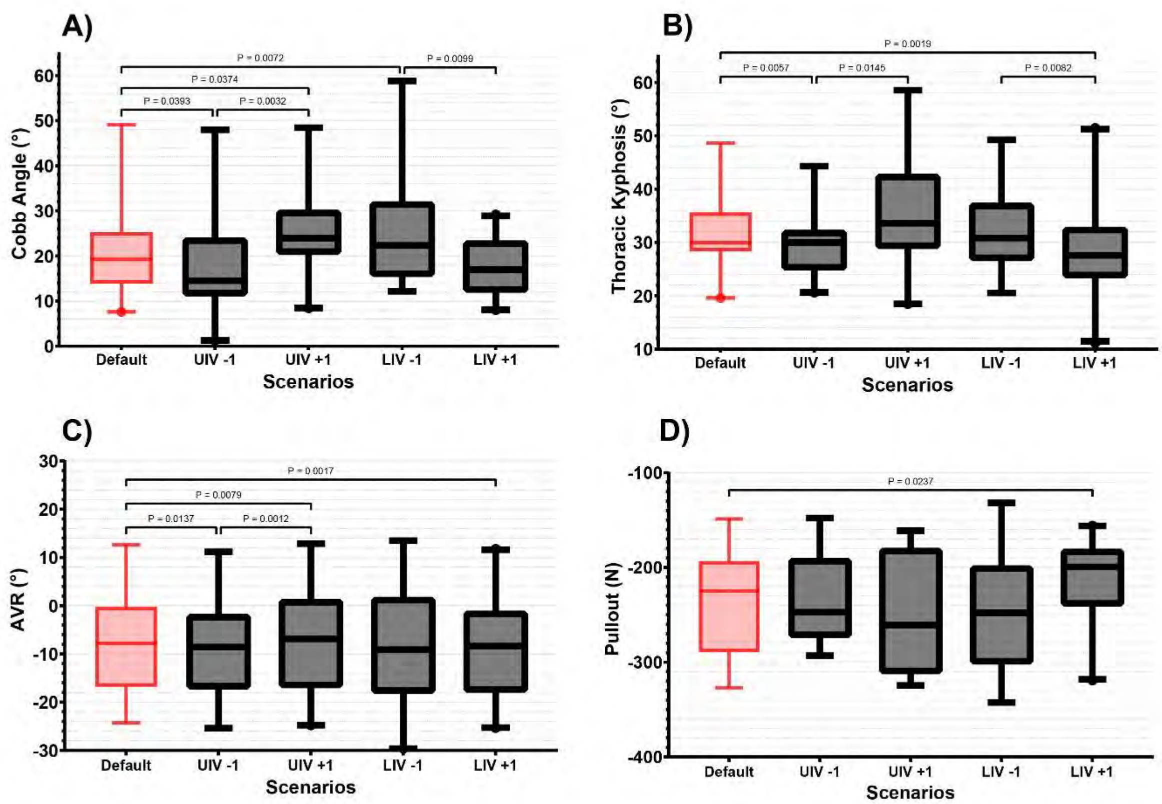


Figure I.2 Effect of ±1 Vertebra Change in UIV and LIV on Simulation Outcomes

P-values indicate whether adjusting the UIV or LIV by ±1 vertebra is statistically different than default AI-predicted UIV and LIV.

Screw pattern sensitivity (Figure I.3).

Testing five screw distributions showed alignment corrections remained largely stable: MT Cobb ±3.3 ± 2.1°, TK ±2.0 ± 1.3°, AVR ±1.8 ± 1.3°. Importantly, rod curvature and fusion levels were fixed during this analysis to isolate the impact of screw distribution. Pullout forces varied substantially with implant density: high-density constructs increased loads in 74% of cases, whereas alternate low-density configurations (1.2 density) reduced pullout by 59% while preserving correction.

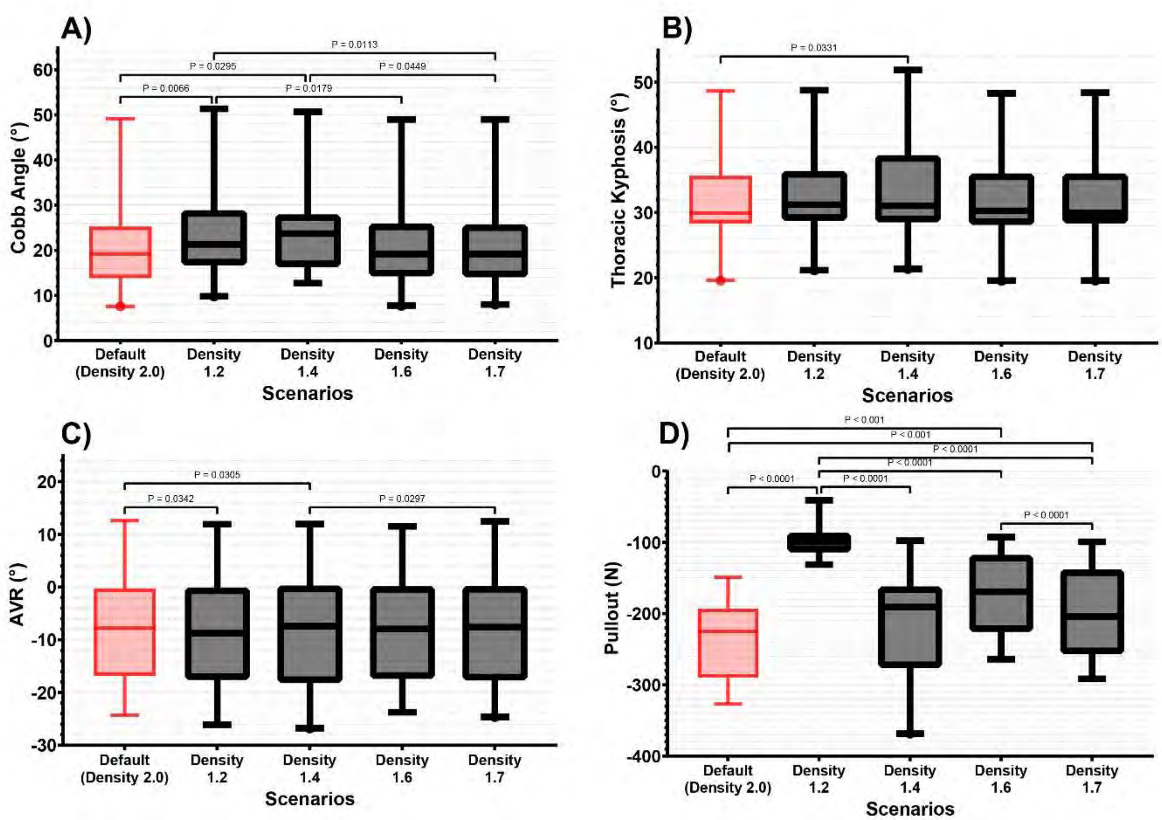


Figure I.3 Effect of Screw Pattern Variability on Simulation Outcomes

P-values indicate whether using different screw patterns is statistically different than default AI-predicted pattern.

Interpretation. The sensitivity analysis demonstrates distinct effects of rod curvature, fusion levels, and screw distribution on postoperative outcomes.

Rod curvature variations had little effect on coronal and axial alignment but produced deviations of up to 9° in thoracic kyphosis, a clinically relevant change. This confirms that rod contouring primarily influences sagittal correction while coronal and axial balance remain relatively stable [38, 218]. Fusion level adjustments were the most influential on 3D correction, causing the largest deviations across all three alignment metrics and occasionally increasing pullout forces by >100 N. This agrees with the known impact of UIV/LIV selection on alignment and mechanical loading, reinforcing why careful level selection is emphasized in clinical guidelines [18, 148, 153, 154, 248]. Screw distribution, tested with rod curvature and fusion levels held constant, produced only minor changes in 3D alignment but substantially affected implant forces. High-density constructs increased pullout, whereas low-density strategies reduced loads by 60% without compromising correction, consistent with prior biomechanical findings [38, 273].

Overall, the framework behaved in line with established biomechanical principles: sagittal correction is sensitive to rod contouring, fusion levels dominate alignment, and screw density mainly modulates implant loading [38, 218, 248, 273]. Importantly, the magnitude of observed deviations exceeded the ±5° accuracy of the model and thus represent clinically meaningful outcomes, supporting robust comparative and optimization analyses under realistic input variations.

Implications for credibility. These results support a credibility level of **b/c** for Model Input Sensitivities under ASME V&V 40. They justify the robustness of the optimization framework when applied to clinically relevant perturbations in rod curvature, fusion extent, and screw distribution. Importantly, the findings highlight that optimization decisions are primarily driven by patient-specific curve characteristics and surgical strategy rather than being dominated by small variations in input parameters.

Conclusion. The instrumentation sensitivity study confirms that the Adams-based MB simulation framework yields biomechanically realistic and clinically robust predictions across a representative range of surgical input variations. Fusion level adjustments emerged as the most critical parameter, while rod curvature and screw distribution exerted smaller, clinically acceptable influences. These results strengthen the overall validation and applicability of the framework for its COU and align with the medium-low risk credibility target established in this thesis.