Journal Pre-proof

Predicting Ride Comfort in High-Speed Rail Using Machine Learning

Amandeep Singh, Mani Entezami, Philippe Doyon-Poulin, Stephan Milosavljevic, Krishna N Dewangan, Yash Kumar Dhabi, Siby Samuel, Graeme Yeo

PII: S2949-8678(26)00029-2

DOI: https://doi.org/10.1016/j.hspr.2026.05.005

Reference: HSPR140

To appear in: *High-speed Railway*

Received date: 17 January 2026
Revised date: 10 April 2026
Accepted date: 6 May 2026

Please cite this article as: Amandeep Singh, Mani Entezami, Philippe Doyon-Poulin, Stephan Milosavljevic, Krishna N Dewangan, Yash Kumar Dhabi, Siby Samuel and Graeme Yeo, Predicting Ride Comfort in High-Speed Rail Using Machine Learning, *High-speed Railway,* (2026) doi:https://doi.org/10.1016/j.hspr.2026.05.005

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Predicting Ride Comfort in High-Speed Rail Using Machine Learning

Amandeep Singh
Department of Mathematics and Industrial Engineering
Polytechnique Montréal, Montréal, Canada, H3T 1J4
Department of System Design Engineering
University of Waterloo, Waterloo, Canada N2L 3G1
Email: amandeep.singh@polymtl.ca; a82singh@uwaterloo.ca

Mani Entezami
Department of Electronic, Electrical and Systems Engineering
University of Birmingham, Edgbaston, United Kingdom, B 15 2TT
Email: m.entezami@bham.ac.uk

Philippe Doyon-Poulin
Department of Mathematics and Industrial Engineering
Polytechnique Montréal, Montréal, Canada, H3T 1J4
Email: philippe.doyon-poulin@polymtl.ca

Stephan Milosavljevic
School of Rehabilitation Science, College of Medicine,
University of Saskatchewan, Saskatoon, S7N 2Z4 Saskatchewan, Canada
Email: stephan.milosavljevic@usask.ca

Krishna N Dewangan
Department of Agricultural Engineering
North Eastern Regional Institute of Science and Technology, Nirjuli, India, 791109
Email: knd@nerist.ac.in

Yash Kumar Dhabi
Iowa Technology Institute
The University of Iowa, Iowa, United States 52241
Email: ydhabi@uiowa.edu

Siby Samuel
Department of System Design Engineering
University of Waterloo, Waterloo, Canada N2L 3G1
Email: siby.samuel@uwaterloo.ca

Graeme Yeo
Technical Director
MoniRail Ltd
Birmingham, United Kingdom, B15 2TT
Email: graeme@monirail.co.uk

Journal Pre-proof

ACKNOLEDGEMENTS

The authors thank the Department of Mathematical and Industrial Engineering at Polytechnique Montréal for providing the computational resources and technical guidance needed for the machine learning for this study.

ABSTRACT

Ride comfort is a necessary criterion for assessing the passenger experience in next-generation high-speed rail (HSR), especially as speeds exceed 300 km/h. However, accurate comfort prediction under varying track and speed conditions remains a computational challenge when relying solely on physics-based simulation. This study proposes a surrogate ensemble meta-model trained on vibration responses generated by a validated multi-body vehicle–track interaction simulation. It uses speed, track irregularity type, and wavelength as key inputs and predicts frequency-weighted root-mean-square (RMS) acceleration in accordance with ISO 2631-1. Five ensemble learning models were compared and fused into a stacked Random Forest meta-model, which achieved a predictive R^2 of 92.1% and a mean absolute error of 0.013 m/s² on unseen cases. Learning curves and prediction intervals confirm strong generalization despite the modest dataset size (n = 345), with the model capturing nonlinear comfort deterioration between 300–400 km/h and identifying the critical 50–100 m wavelength band. This simulation-aligned machine learning surrogate provides a computationally efficient alternative to repeated multi-body simulations, enabling rapid evaluation of speed–track interaction scenarios within the validated modeling framework. Rather than replacing physics-based analysis, the proposed approach is intended to support early-stage design screening, sensitivity analysis, and large-scale scenario exploration where full simulation would be computationally prohibitive. These results establish the feasibility of integrating data-driven surrogates with validated vehicle–track models to accelerate comfort-oriented engineering assessments in HSR systems.

Keywords: High-Speed Rail, Track-Vehicle Interaction**,** Ride Comfort Prediction, Machine Learning, Ensemble Learning, Parsimony.

Journal Pre-proof

INTRODUCTION

Since the 1960s introduction of Japan's Shinkansen, high speed train velocities have steadily exceeded 300 km per hour. New systems, including HS2 in Britain at 360 km per hour, China's CR450 at 400 to 450 km per hour, Japan's ALFA X up to 400 km per hour, and upgraded fleets in France and Germany, continue this trend. As speed increases, vibration frequencies shift higher, while human sensitivity remains around 4 to 8 Hz for vertical motion [1,2]. This mismatch means passengers mainly perceive persistent low frequency vibrations, which can degrade comfort if not controlled through suspension or track design. Higher speeds also amplify vibration levels, further challenging comfort [3]. Consequently, track geometry requirements must evolve, with greater emphasis on controlling long wavelength irregularities that generate low frequency motion. Ensuring ride comfort under these conditions is a key human factors issue in modern high-speed rail (HSR). While multi body simulation models help analyze sensitivities, their computational cost limits large scale or real time use, highlighting the need for efficient surrogate models.

Comfort is an important factor influencing the choice of transportation mode. Sustainability initiatives highlight that shifting to railway transportation can reduce greenhouse gas emissions and support environmental goals [4–7]. In HSR systems, improving ride comfort is both a technical challenge and a sustainability objective. Railway manufacturers and operators now prioritize comfort enhancement. Among various comfort factors, vibrational comfort is a key factor affecting passenger experience and train dynamics, routinely monitored for safety and maintenance. With rising speeds, models that detect component degradation and early comfort decline are crucial for timely intervention [8].

Railway engineers typically evaluate ride comfort using physics-based modeling and standardized whole-body vibration metrics [9]. Car body dynamics, including rigid body modes below 3 Hz and first bending and torsional modes around 10 Hz, govern how track irregularities translate into car body accelerations. Long wavelengths of 50 to 100 m excite rigid bounce at about 1 to 2 Hz, while shorter wavelengths of 5 to 12 m excite bending at about 10 to 15 Hz, both affecting ride comfort. Human response is assessed using standards such as ISO 2631 1[11] and BS 6841 [12], which apply frequency dependent weightings. For example, ISO 2631 1 evaluates vibrations over 0.5 to 80 Hz to compute comfort metrics. Standards including UIC 513 and EN 12299 further quantify exposure using root mean square and vibration dose values. On the infrastructure side, UIC 703 and EN 13848 5 define track geometry limits for safety and comfort [13,14]. UIC 703 sets limits on curve radius, cant, and cant deficiency up to 300 km per hour, while EN 13848 5 specifies allowable irregularities up to 70 m wavelength for speeds of 300 to 360 km per hour. Beyond 70 m and 360 km per hour, guidance is limited, though Chinese standards extend to about 150 m at 350 km per hour, reflecting uncertainty. This gap highlights the need for predictive comfort models that generalize across operating conditions and support early design, operation, and policy decisions.

Traditional ride comfort evaluation in HSR relies on multi body simulation models that couple track irregularities with vehicle dynamics [1,15]. High fidelity vehicle models, including primary and secondary suspensions and flexible car body modes, simulate accelerations for given

Journal Pre-proof

track geometries. These outputs are processed using ISO 2631 1 weightings to estimate discomfort. Such models identify critical irregularity wavelengths that degrade comfort [16]. However, they are computationally intensive and require specialized expertise [17], limiting scalability for large scenario testing, real time diagnostics, and early design integration. Deterministic simulation also does not capture uncertainty or predict unseen conditions. Although sensitivity studies are possible, exploring wide ranges of speed, wavelength, and geometry is computationally demanding [18,19]. Real time comfort evaluation using full simulations is therefore impractical.

This research gap motivates the use of machine learning (ML) methods that leverage domain-specific principles from vehicle–track dynamics while achieving computational efficiency. ML techniques have been increasingly applied in the HSR domain, particularly for predictive maintenance, system monitoring, and condition assessment [20–24]. However, most existing studies focus on subsystem-level diagnostics or signal classification and do not directly address passenger-centric ride comfort prediction using standardized vibration-based metrics.

Despite these advances, limited work addresses passenger centric modeling from a human factors perspective using vibration-based standards. For example, LSTM models have been used to predict lateral car body accelerations and support maintenance decisions [8]. Other studies focus on risk diagnosis using big data and system level modeling [25], or classify ride quality from acceleration signals using transfer learning and bi directional LSTM methods [26]. However, these approaches are largely reactive and do not incorporate ISO 2631 1 based comfort metrics or vehicle track dynamic descriptors in the learning process.

This study addresses these gaps by introducing a surrogate model for ride comfort prediction using ML trained on vibration responses from a validated multi body simulation. Inputs include physically meaningful variables such as train speed, track irregularity type, and wavelength, each linked to vibration transmission and resonance related discomfort. These features preserve interpretability even beyond standard datasets. Although the dataset is simulated, it is based on validated models [1] and reflects realistic vibration behavior in modern HSR. Here, physics based refers to physically meaningful inputs and simulation derived variables rather than embedding physical equations within the model. The contribution lies in integrating established ML methods with validated simulation outputs and vibration standards to develop a domain aligned surrogate for comfort prediction.

Ensemble learning methods are well suited for this task due to their accuracy, generalization, and robustness in modeling nonlinear systems [27]. They can approximate computationally expensive simulations at much lower cost, enabling rapid evaluation across wide operating conditions such as speed, geometry, and wavelength [19]. By learning from simulation data, ensemble and stacked models capture complex relationships between track features, train speed, and comfort metrics that are difficult to model analytically [27]. Methods such as Random Forest, Extra Trees, Gradient Boosting, XGBoost, and AdaBoost learn conditional dependencies within the vehicle track passenger system [28,29]. They also reduce overfitting by combining predictions from multiple learners [30,31].

Journal Pre-proof

This study evaluates five ensemble models and combines them using two meta learners, a Random Forest and a multilayer perceptron. It predicts ISO 2631 1 weighted accelerations as the ride comfort metric, aligning with international standards. Learning curves, residuals, and prediction intervals show good generalization despite limited data. Inputs reflect realistic conditions, including next generation systems such as HS2 at 360 km per hour, China CR450 at 400 to 450 km per hour, Japan ALFA X up to 400 km per hour, TGV M, Germany ICE upgrades, India Mumbai Ahmedabad at 320 km per hour, Korea HEMU 430X, and California HSR at 350 km per hour.

The proposed meta-model incorporates physically meaningful features, such as shorter wavelengths (5–12 m) that excite higher-frequency vibration modes and longer wavelengths (50–100 m) that correspond to car-body bounce motions. Including such descriptors enables realistic and bounded predictions within the validated simulation setting. The final meta-model achieves accuracy comparable to the underlying multi-body simulation while requiring only a fraction of the computational time. In this context, the model should be interpreted as a surrogate of the validated vehicle–track dynamics framework rather than a standalone predictor of real-world ride comfort across all vehicle types. Its primary engineering value lies in enabling rapid exploration of large design and operational spaces, including speed–wavelength–geometry interactions, which would otherwise require extensive simulation effort. This makes the approach particularly suitable for early-stage design evaluation, sensitivity analysis, and scenario-based screening within a defined and validated modeling configuration.

This work also sets the foundation for research involving model simplification, refinement, and deployment in digital twins and onboard systems. Future aspects of this work will explore model parsimony for low-latency implementation, incorporate additional dynamic parameters such as suspension properties, and use transfer learning to bridge simulation and field data. The long-term vision centers on scalable comfort prediction under realistic HSR conditions.

METHODS

Experimental Design

The study investigates the impact of three main features on ride comfort in HSR: speed (200, 250, 300, 350, and 400 km/h), wavelength (3 to 300 m), and track irregularities (vertical, lateral, and combined). These variables are defined within the context of a validated vehicle–track interaction model adopted from Liu et al. [1], in which the vehicle configuration, including suspension parameters, damping characteristics, and structural dynamics, remains fixed. Accordingly, the present study does not attempt to generalize across different vehicle types; instead, the proposed surrogate model represents the vibration response of this specific validated vehicle configuration. Here, wavelength refers to the one-third octave wavelength-band descriptor used in the source dataset to represent band-wise contributions to overall weighted vibration across speeds, rather than a full reconstruction of the track power spectral density (PSD). The selected wavelength range (3–300 m) spans short, medium, and long wavelength irregularities commonly considered in railway engineering. These include shorter wavelengths associated with rail joints

Journal Pre-proof

and localized track irregularities, intermediate wavelengths related to track slab behavior and support conditions, and longer wavelengths corresponding to bridge spans and large-scale track geometry variations. While the model does not explicitly encode the full PSD shape, these wavelength bands provide a structured spectral representation that captures the dominant geometry-induced excitation mechanisms influencing vehicle vibration and passenger comfort within the validated modeling framework.

Track irregularities are examined under three conditions: vertical irregularity, lateral irregularity, and all geometry combined, including vertical and lateral alignments (curvature and cant) [1]. These three configurations correspond to Original Vertical Irregularity (OVI), Original Lateral Irregularity (OLI), and All Geometry Combined (AGC), respectively. OVI and OLI represent measured deviations in the vertical and lateral components of the track geometry. AGC includes the full geometry profile, combining vertical and lateral irregularities with curvature and cant to represent more realistic operating conditions. OVI, OLI, and AGC are used here as dataset scenario categories, consistent with the source simulation study in which the vehicle model was run with each geometry component separately and also with their combination across 200–400 km/h.

The target variable is the frequency-weighted vibration exposure representing passenger discomfort. The study utilized 345 simulation-derived samples (n = 345), each representing a unique input scenario designed to capture variations in ride comfort across speed, geometry, and wavelength conditions. These samples were not synthetically generated in an abstract sense, but were obtained from a validated multi-body simulation framework reported in Liu et al. [1], where the underlying vehicle–track interaction model was calibrated and verified against measured vibration data from an in-service HSR.

Specifically, the source model demonstrated close agreement between simulated and measured acceleration responses, with frequency-weighted vibration values of 0.349 m/s² (simulated) compared to 0.356 m/s² (measured), along with consistent reproduction of dominant spectral characteristics in the 0.5–30 Hz range. As a result, the simulation outputs used in this study can be interpreted as physically grounded and experimentally validated representations of vehicle-body vibration under defined operating conditions.

Each of the 345 samples corresponds to a distinct combination of train speed, track geometry profile, and wavelength band, generated through this validated simulation framework. Although the dataset size is modest by conventional ML standards, each sample reflects a computationally intensive and physically consistent simulation scenario. The dataset was designed using a structured sampling approach to ensure coverage across the operating envelope, including variations in speed, wavelength, and track irregularity conditions.

To further support model reliability, prediction interval analysis and learning curve diagnostics were incorporated to assess generalization performance under the moderate sample size, following established practices in engineering-focused ML studies [32,33].

The study employs a three-dimensional vehicle dynamic model based on the validated formulation presented in Liu et al. [1]. The model represents the car body, bogies, and wheelsets

Journal Pre-proof

as a comprehensive multi-body system with 31 rigid degrees of freedom and 9 flexible modes associated with car-body bending and torsion. Primary and secondary suspensions are modeled using linear Kelvin–Voigt elements in the vertical, lateral, and longitudinal directions, capturing stiffness and damping characteristics governing bogie and car-body motion. Track vertical dynamics are represented by a single-degree-of-freedom mass–spring–damper system calibrated to match measured track receptance up to 80 Hz, while lateral track dynamics are treated as rigid. Wheel–rail interaction is computed using a linearized Hertzian contact model and FASTSIM-based creep forces, with the contact patch updated at each simulation step. The flexible car body is represented as a Euler–Bernoulli beam that includes the first three bending modes in the vertical and lateral directions and the first three torsional modes, with natural frequencies consistent with measurements reported in Liu et al. [1], (e.g., a first vertical bending mode near 12 Hz and a first torsional mode near 12 Hz). After calibration using accelerations measured inside an in-service passenger train at approximately 210 km/h, the model reproduced dominant rigid-body modes near 1–2 Hz and bending and torsional resonances around 10–15 Hz with good agreement.

To clarify the modeling assumptions, the underlying dynamic formulation follows the standard linearized multi-body approach widely used in HSR vibration analysis. The suspension, car-body, and wheelset equations are linearized about the nominal operating configuration, while the excitation originating from measured and extended track geometry introduces nonlinear variation into the input. Prior studies, including Liu et al. [1], show that this approach accurately reproduces vibration amplitudes and frequency content for the validated model configuration and measurement conditions. In the present study, the 300–400 km/h cases are taken from the same open, simulation-derived dataset and are used as scenario-based exploration within the model parameterization calibrated at ~210 km/h. Accordingly, the surrogate model is constructed to reproduce the MBS-simulated ISO 2631-1 frequency-weighted r.m.s. response across the provided speed range, and the 300–400 km/h results should be interpreted as simulation-defined predictions within this modeling framework. Speed-dependent effects that can grow at very high speed (e.g., aerodynamic loading and higher-frequency structural dynamics) are not separately parameterized in the released dataset and therefore remain a key source of model-form uncertainty at the upper end of the speed range.

To support the credibility of the simulation outputs, we rely on the validation results presented in Liu et al. [1], where simulated and measured acceleration spectra at the floor, seat pan, and backrest were compared. The calibrated model reproduced measured vibration spectra over the 0.5–30 Hz range with close correspondence in dominant spectral peaks and overall levels, yielding a weighted acceleration of 0.349 m/s² compared with 0.356 m/s² measured. Minor discrepancies at very low frequencies and around the first flexible mode were documented, but these had little influence on the ISO 2631-1 discomfort metric. Because the MBS model reproduced dominant spectral characteristics and weighted accelerations with demonstrated accuracy, its simulated vibration responses were treated as simulation-derived labels for surrogate model training, consistent with standard practice in railway vehicle dynamics when field testing at ultra-high speeds is not feasible.

Journal Pre-proof

The ride comfort output for each scenario was quantified using the ISO 2631-1 whole-body vibration index (frequency-weighted root-mean-square acceleration), which provides a single discomfort measure summarizing multi-axis vibrations derived from vehicle-body acceleration responses. These acceleration responses are generated by the validated multi-body vehicle–track interaction model, which inherently incorporates the effects of vehicle suspension, structural dynamics, and vibration transmission characteristics. Accordingly, the ML model does not directly map track irregularities to comfort; rather, it learns the relationship between physically simulated operating conditions and the resulting ISO-compliant vibration response, within a framework where vehicle dynamics are already embedded. The evaluation considered accelerations at multiple points (feet, seat pan, and backrest) and in longitudinal, lateral, and vertical directions, consistent with human exposure guidelines. Using this internationally recognized comfort metric ensures that the target variable is grounded in established physiological sensitivity to vibration and provides regulatory relevance. The experimental design was structured to span nonlinear operational regimes such as high-speed interactions with different track irregularity patterns, thereby exposing the ML algorithms to realistic and complex feature interactions. The resulting dataset was analyzed across defined wavelength bands, focusing on the 0.5–80 Hz range where whole-body vibration discomfort is most pronounced under ISO 2631-1.

ISO 2631-1 allows the use of frequency-weighted r.m.s. acceleration as a primary comfort indicator, but it also notes (Clause 6.2.2) that when the crest factor of the frequency-weighted vibration signal is high (commonly referenced threshold CF > 9), the r.m.s. method may underestimate the severity of shocks and auxiliary metrics such as VDV or MTVV should be considered. In the present work, the source dataset used for surrogate training provides scenario-level frequency-weighted r.m.s. values (overall weighted vibration) rather than the underlying frequency-weighted acceleration time histories. As a result, crest factor, MTVV, and VDV cannot be computed from the available data and are therefore outside the scope of this study. The proposed surrogate model should thus be interpreted as predicting ISO 2631-1 frequency-weighted r.m.s. comfort only.

Feature Engineering and Data Splitting

We curated and engineered three input features capturing the physical and operational determinants of ride comfort: train speed, track wavelength, and track irregularity type. These features were selected based on explicit domain knowledge, recognizing that increasing speeds and certain wavelength bands in track geometry can induce resonant structural modes and amplify vibration exposure in HSR systems. A structured preprocessing pipeline was developed, treating speed and wavelength as numeric features and track irregularity as a categorical feature. To maintain input fidelity across scales, standardization was applied to normalize the numeric features, followed by second-degree polynomial expansion. This included the original features, their squared terms, and pairwise interaction terms. These polynomial expansions were specifically applied to the two numeric features (speed and wavelength) based on established knowledge that their interaction governs resonance behavior and comfort degradation in HSR systems [1]. In particular, second-order terms support the representation of curvature in the speed–discomfort

Journal Pre-proof

relationship and capture nonlinear amplification effects associated with specific wavelength–speed combinations. These engineered terms provide the learners with explicit access to vibration-related nonlinearities and interaction effects that would otherwise need to be inferred implicitly from limited data, improving stability and generalization while remaining aligned with vehicle–track dynamics. The categorical feature, track irregularity type, was not subjected to polynomial expansion; instead, it was one-hot encoded to retain the structure of its discrete categories (vertical, lateral, combined). This feature engineering ensured that both continuous and categorical information were represented in a form suitable for ML while preserving interpretability.

To further enhance generalization and mitigate overfitting risk under this moderate-sized dataset (n = 345), a stratified two-tier data splitting strategy was implemented. Eighty percent (80%) of the data was used exclusively for model training and internal validation, including cross-validation, model selection, and hyperparameter tuning. The remaining 20% was reserved as a held-out test set for final performance evaluation. This test set remained completely untouched throughout the training pipeline. Stratification ensured proportional representation of wavelength bands, speed ranges, and irregularity types in both partitions, maintaining distributional consistency. By enforcing this strict separation, we preserved the integrity of the test evaluation and ensured that all modeling decisions (from preprocessing to model tuning) were validated entirely within the training partition. This approach reflects best practices in ML under data-limited regimes, providing a robust measure of out-of-sample predictive performance while minimizing the risk of information leakage or overfitting.

Ensemble Models and Hypertuning

Before presenting the ensemble learners, it is important to clarify why an ML surrogate model is used despite the availability of a validated MBS. Although the physics-based model provides high-fidelity predictions, each simulation is computationally expensive and requires full dynamic evaluation of vehicle–track interaction. This limits its use for large-scale scenario exploration, real-time estimation, rapid sensitivity screening, or design optimization. The ML model therefore serves as a fast surrogate that approximates the validated simulation outputs, enabling near-instant predictions across thousands of speed–wavelength–geometry combinations that would be impractical to simulate directly. In this study, the ML outputs are explicitly trained to reproduce the simulation responses, so all performance metrics represent a direct comparison against the physics model under multiple scenarios. In this framework, the simulation provides physical accuracy while the ML model provides computational speed, making them complementary rather than competing tools.

Five ensemble models were developed to predict overall weighted vibration response using Random Forest [34], Extra Trees [35], Gradient Boosting [36], XGBoost [37], and AdaBoost [38] methods. These ensemble learners were selected based on their proven ability to capture nonlinear interactions in complex domains, including transport and mechanical systems [27,39]. Details about these ensemble models are available in Rokach's comprehensive review of ensemble learning [40].

Journal Pre-proof

Random Forest Regressor was implemented with 100 estimators, a minimum sample split of 2, minimum leaf size of 1, and a maximum depth of 30. The 'sqrt' strategy was used for feature selection at each split to manage variance and prevent overfitting. Extra Trees Regressor mirrored this setup, isolating the performance effects of its randomized splitting strategy relative to Random Forest. Gradient Boosting Regressor was configured with 500 estimators, a maximum depth of 3, minimum split of 5, and a learning rate of 0.15. These hyperparameters were identified via grid search within a cross-validation loop to optimize generalization. XGBoost Regressor used 400 estimators, max depth of 7, min child weight of 1, a subsample ratio of 0.6, and a learning rate of 0.2. All features were included in each boosting round (colsample_bytree = 1.0), and the objective was set to 'reg:squarederror'. XGBoost was chosen because it performs well on structured tabular datasets where nonlinear interactions and heterogeneous feature importance are expected. AdaBoost Regressor was deployed with 300 estimators and a conservative learning rate of 0.01 to ensure stable ensemble formation over iterative refinement.

Meta Models and Hypertuning

To enhance predictive accuracy beyond individual learners, two meta-models were developed by stacking the top-performing base learners using: (i) Random Forest Meta-Model (M-RF) [41] and (ii) Multi-Layer Perceptron Meta-Model (M-MLP) [42]. The stacking procedure involved generating out-of-fold predictions from XGBoost and Gradient Boosting using 10-fold cross-validation, which were then used as meta-features for second-level learning. M-RF used 101 estimators, max depth = 27, min samples split = 5, min samples leaf = 8, and 'sqrt' feature selection. It combined outputs from the two best base models to generate a robust aggregated prediction. M-MLP used a two-hidden-layer network (80 and 70 neurons), ReLU activation, L2 penalty of 0.000691, and constant learning rate. It was trained for a maximum of 2000 iterations, ensuring convergence without premature stopping. This neural architecture introduced nonlinear capacity at the meta-level, offering flexibility to capture residual interactions not fully represented by tree-based learners.

Model Evaluation, Validation and Interpretation

Models were evaluated based on a set of performance metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2). MAE, MSE, and RMSE quantify prediction errors, with lower values indicating superior performance. R^2 indicates the proportion of variance in the target variable explained by the model, with values closer to 1 suggesting a better fit. Given the modest dataset size (n = 345), the evaluation strategy explicitly incorporated multiple generalization diagnostics to ensure that the surrogate model remained reliable despite limited samples. In addition to the primary train–test split, 10-fold cross-validation, prediction-interval analysis, learning-curve inspection, and sensitivity-based perturbation tests were used to assess robustness under varying data volumes and feature conditions. These procedures collectively compensate for the restricted dataset size and provide insights into stability, uncertainty, and generalization.

Journal Pre-proof

After comparing the model performance of individual models and meta-models, the best-performing model was shortlisted to undergo additional validation using two key techniques: prediction interval analysis and learning curve analysis. A 95% confidence interval was computed on the test set predictions to quantify the range within which the true ride comfort values are expected to fall, with 95% certainty. This interval-based evaluation provides a practical measure of predictive uncertainty in safety-critical rail applications.

Learning curve analysis was performed using predefined arrays of training and test scores across ten different training set sizes (ranging from 10% to 100% of the data). At each increment, the model was trained multiple times (using five-fold cross-validation), and training and validation scores were recorded. Each score thus contains 50 data points (10 sizes × 5 folds). Plotting these results revealed whether the model exhibited underfitting, overfitting, or stable improvement as more data were provided, and whether additional samples would meaningfully improve performance. This combination of cross-validation and progressive-training diagnostics demonstrates that the model retains generalization capability even when trained on smaller subsets, an important requirement for moderate-sized datasets (n ~ 345).

Two explainable-AI procedures were employed to enhance interpretability: sensitivity analysis [43] and feature importance analysis [44]. Sensitivity analysis explores the model's response to gradual changes in input features. For numeric features, 100 evenly spaced values were generated within the observed ranges, while non-varied numeric features were fixed at their mean and the categorical feature ('Track') was fixed at its most frequent category. For the categorical feature, each category was evaluated while Speed and Wavelength were held at their means. These sample inputs were passed through the shortlisted base models and then into the M-RF for final predictions. This response-surface evaluation highlights how nonlinear interactions (e.g., Speed × Wavelength) influence predicted discomfort and helps identify inflection zones where ride comfort deteriorates sharply.

Feature-importance analysis determined the relative contribution of shortlisted base models (e.g., XGBoost, Gradient Boosting) to the meta-model’s final prediction. Importance scores (e.g., OOBPermutedPredictorImportance in MATLAB) were computed to quantify the influence of each learner. Bivariate feature importances were also calculated by summing the importance of pairs of features, enabling characterization of interaction effects. For instance, the interaction between Speed and Wavelength can magnify discomfort more than each factor alone, reflecting resonance-related behavior. Visualizing these combinations ensures that the meta-model captures physically meaningful vibration mechanisms. Together, these interpretability procedures verify that the model is both statistically performant and physically coherent, minimizing concerns regarding overfitting, black-box behavior, or misalignment with established vehicle-track dynamics. This combined interpretability framework also explains why the meta-model outperforms individual learners by leveraging complementary strengths of multiple modeling pathways.

RESULTS

Journal Pre-proof

Fig. 1 shows the overall trend depicting the effect of the input features on the overall weighted vibration. All axes in Fig. 1 are shown with their physical units, and the three subplots correspond to comfort variation across speed (km/h), wavelength (m), and track type, with speed groups color-coded for clarity. Descriptive analysis confirms a consistent speed-related increase in weighted vibration across all track types, indicating a progressive decline in ride comfort as train speed increases. The AGC track exhibited the maximum overall weighted vibration of 0.766 m/s^2 at 400 km/h, compared to 0.568 m/s^2 for OLI and 0.737 m/s^2 for OVI at the same speed. This pattern suggests that AGC is especially sensitive to comfort degradation in the high-speed regime, likely due to the combined influence of vertical and lateral irregularities and geometric components. In contrast, OLI maintained consistently lower vibration levels, indicating comparatively stable vibration behavior and better retention of comfort.

Regarding wavelength effects, the 50–100 m range exhibited the highest mean (0.190 m/s^2) and maximum (0.633 m/s^2) vibration levels, identifying it as a critical range associated with known car-body rigid-mode excitation and comfort degradation. In comparison, the 250–300 m wavelength range showed the lowest mean (0.013 m/s^2) and maximum (0.056 m/s^2) values, indicating minimal contribution to discomfort within the studied operating envelope. These results highlight the mid-range wavelengths as a significant contributor to ride discomfort and reinforce the importance of prioritizing monitoring and maintenance in this spectral band.

The correlation between speed and vibration exposure (0.2599 m/s^2) further reinforces the positive association between operational velocity and dynamic excitation. Track-specific wavelength–vibration correlations showed more complex patterns. For OVI and AGC, the negative correlations (OVI: −0.2709; AGC: −0.1569) indicate that within these track conditions, longer wavelengths tend to reduce vibration amplification. In contrast, OLI showed a weak positive correlation (0.0789), implying that for lateral irregularities alone, longer wavelengths may slightly increase vibration under certain operating conditions. These findings highlight the need for track-specific comfort strategies, particularly in balancing the influence of short and long wavelengths when designing or maintaining high-speed infrastructure.

Journal Pre-proof

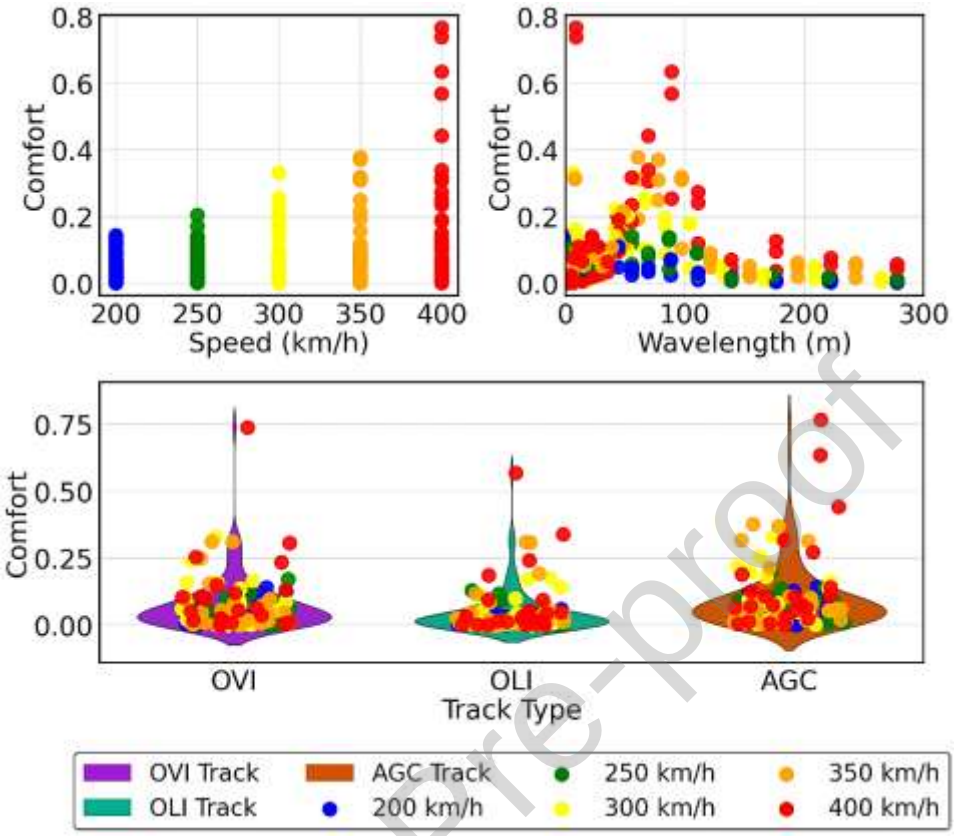

Fig. 1 Effects of train speed (km/h), track irregularity wavelength (m), and track type (OVI, OLI, AGC) on overall weighted vibration (m/s² r.m.s.) based on simulation-derived comfort values. Panels show comfort variation across speed, wavelength, and track type, with color-coded points representing individual speed groups

Table 1 the results of the three-way ANOVA. Train speed significantly affected ride comfort ($F(4, 315) = 6.76$, $p < 0.001$), contributing 6.7% of the variance ($\eta^2 = 0.067$). Wavelength also had a significant effect ($F(1, 315) = 6.17$, $p < 0.05$), explaining 1.6% of the variance. Track irregularity was likewise significant ($F(2, 315) = 8.37$, $p < 0.001$), accounting for 4.3% of the variance. While the Speed × Wavelength interaction was not significant ($F(4, 315) = 0.11$, $p > 0.05$, $\eta^2 = 0.001$), the Wavelength × Track Irregularity interaction was significant ($F(2, 315) = 3.48$, $p < 0.05$), contributing 1.8% of the variance. This interaction effect confirms that the influence of wavelength on comfort is modulated by the nature of track irregularities.

Table 1 Three-way ANOVA for the effect of speed, wavelength and track irregularity on comfort

Features	Sum Sq.	df	F-Value	P-Value	Partial Eta Square
Speed	0.264473	4.0	6.755682	0.000032	0.067190
Wavelength	0.060405	1.0	6.171972	0.013498	0.016185
Track Irregularity	0.163863	2.0	8.371423	0.000287	0.042722
Speed*Wavelength	0.004361	4.0	0.111392	0.978481	0.001186

Journal Pre-proof

Speed*Track Irregularity	0.019885	8.0	0.253965	0.979620	0.005386
Wavelength*Track Irregularity	0.068060	2.0	3.477065	0.032089	0.018199
Speed*Wavelength*Track Irregularity	0.007774	8.0	0.099283	0.999221	0.002113

All model performance metrics reported in this section represent direct comparisons between the ML predictions and the vibration responses produced by the validated physics-based MBS model. For each of the 345 scenarios, the surrogate model was evaluated against the corresponding simulation output, ensuring that the R^2, MAE, and RMSE values quantify the accuracy of the ML surrogate in reproducing the deterministic simulation-derived responses under varying combinations of speed, wavelength, and track irregularity.

Fig. 2 compares the performance of the ensemble models and the two meta-models in predicting ride comfort. Among the ensemble learners, XGBoost emerged as the best-performing model, explaining 90.35% of the variance in ride comfort. Its low mean squared error (MSE: 0.049%) and root mean squared error (RMSE: 2.211%) reflect high accuracy and minimal deviations from the simulation labels, while an MAE of 1.567% indicates strong precision in predicting comfort changes. XGBoost's strength lies in its capacity to capture nonlinear relationships among speed, wavelength, and irregularity type, enabling consistent reproduction of the simulation responses. Random Forest explained 70.59% of the variance, demonstrating moderate predictive capability, but with higher variability in its outputs (RMSE 3.861%, MAE 2.475%). These elevated error values suggest that Random Forest does not fully represent the interaction effects embedded in the simulation dataset, potentially underestimating combined speed–wavelength–track influences on ride comfort.

Extra Trees performed poorly, explaining only 38.89% of the variance. With an RMSE of 5.566% and an MAE of 2.893%, this model showed substantially larger errors, indicating limited generalization and insufficient representation of vibration-related trends in the data. Gradient Boosting, however, performed strongly, explaining 80.66% of the variance, with a relatively low MSE (0.098%) and RMSE (3.132%). An MAE of 1.984% further illustrated competent predictive capabilities, making it a viable but secondary choice compared to XGBoost. AdaBoost explained 43.49% of the variance, accompanied by a relatively high MSE (0.287%) and RMSE (5.353%), indicating lower predictive accuracy than the top-performing models. An MAE of 3.672% suggests that its predictions frequently deviated from the simulation outputs, limiting its practical usefulness for comfort prediction.

To quantify the benefit of the stacked architecture relative to the strongest single learner, XGBoost was compared directly with the Random Forest based meta model that fuses XGBoost and Gradient Boosting predictions. XGBoost achieved $R^2 = 0.9035$ with RMSE = 2.211% and MAE = 1.567%, while the RF meta model achieved $R^2 = 0.9200$ with RMSE = 2.089% and MAE = 1.432%. This corresponds to an increase of 0.0165 in explained variance and absolute reductions of 0.122 and 0.135 percentage points in RMSE and MAE, respectively, indicating a measurable but modest gain over the single model baseline under the same evaluation protocol.

Journal Pre-proof

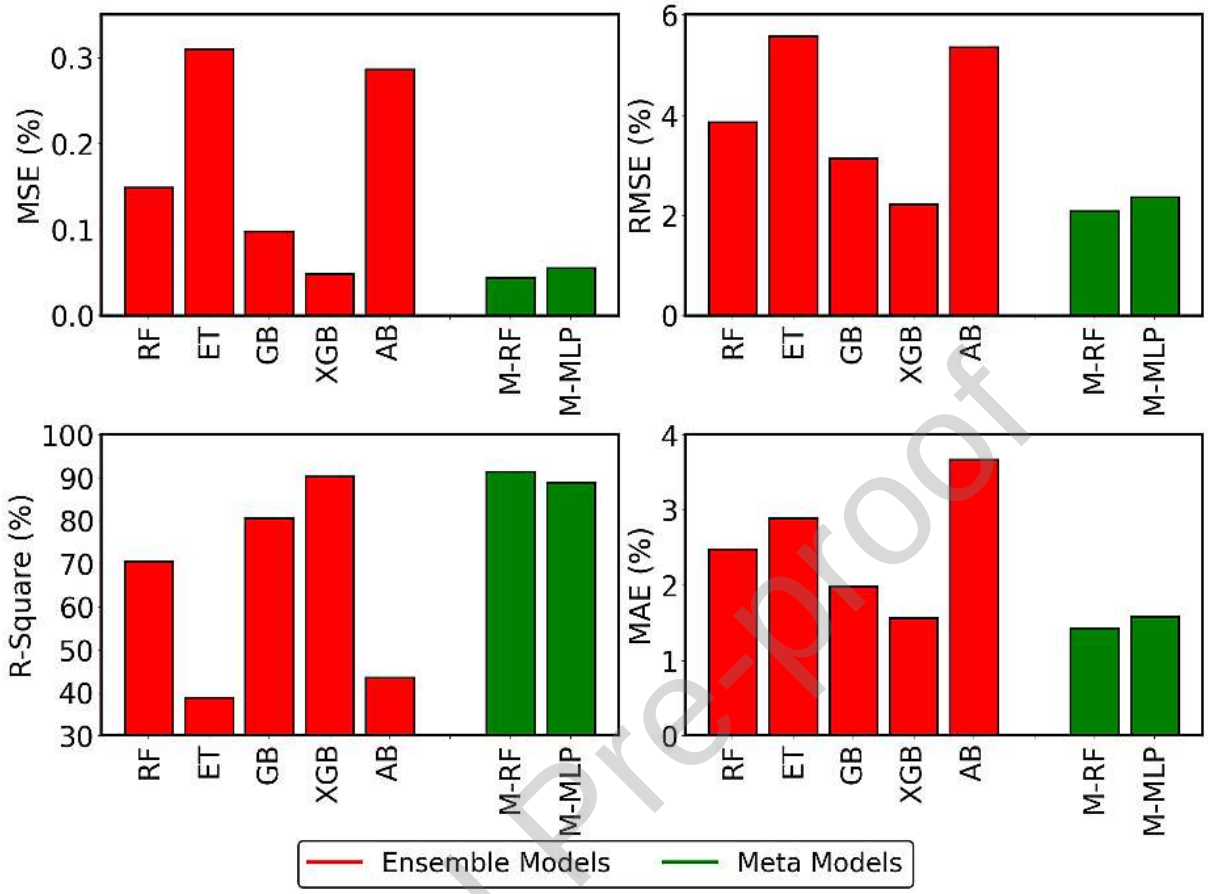

Fig. 2 Comparison of ensemble and meta-model performance using R^2, MAE (%), and RMSE (%). All values represent direct error between ML predictions and MBS outputs for 345 scenarios.

Building on these performance measures, the best-performing models (XGBoost and Gradient Boosting) were integrated into meta-models. The Random Forest–based meta-model outperformed all individual learners, explaining 92% of the variance in ride comfort, with a very low MSE (0.044%) and RMSE (2.089%). This result indicates that combining XGBoost and Gradient Boosting predictions through a Random Forest based meta model can provide a small additional reduction in error within this dataset. The MLP-based meta-model also performed competitively, explaining 88.99% of the variance. Its MSE (0.056%) and RMSE (2.362%) indicate strong overall accuracy despite being slightly less precise than the RF meta-model (MAE 1.579%). The observation that the individual XGBoost model slightly outperformed the MLP meta-model suggests that the neural architecture may require larger datasets or more specialized tuning to match the efficiency of optimized ensemble learners in this application. Fig. 3 shows regression fit plots for ensemble and meta-models; the Random Forest meta-model displayed the highest concentration of predictions near the one-to-one line, indicating the strongest alignment with simulation-derived comfort values.

Journal Pre-Proof

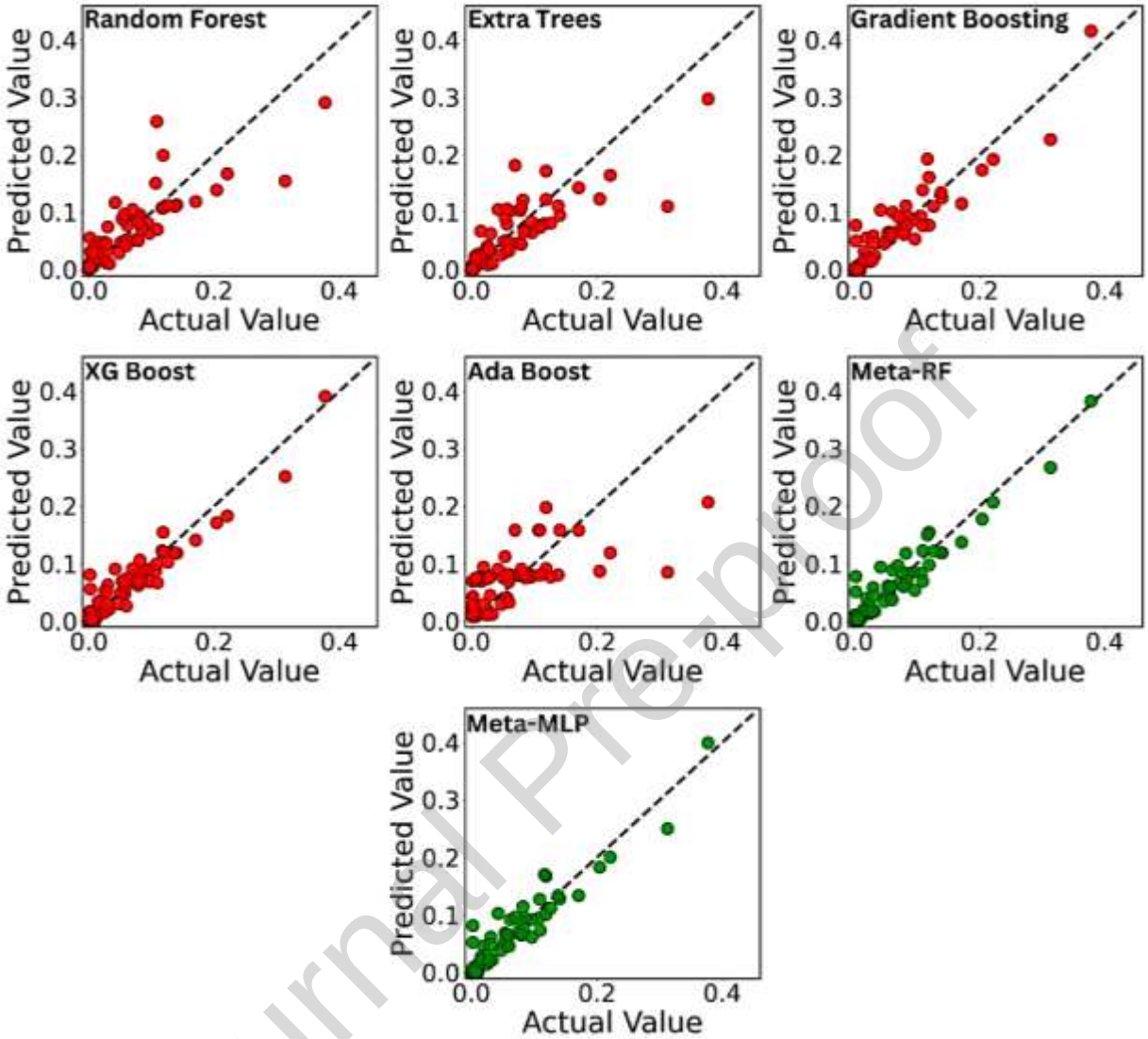

Fig. 3 Regression fit between predicted overall weighted vibration (m/s², r.m.s.) and the corresponding values generated by the validated MBS model for each learner. The dashed 1:1 line represents perfect agreement; dispersion around this line reflects the surrogate model's accuracy across the full range of vibration magnitudes.

To further examine the wider scatter visible in Fig. 3 at low vibration levels, a piecewise error analysis was performed using bins defined by the true overall weighted vibration values (< 0.05 m/s², 0.05 to 0.15 m/s², and ≥ 0.15 m/s²). For the Random Forest based meta model, the piecewise errors were: < 0.05 m/s² (RMSE = 3.581%, MAE = 1.607%), 0.05 to 0.15 m/s² (RMSE = 3.751%, MAE = 2.720%), and ≥ 0.15 m/s² (RMSE = 4.968%, MAE = 4.199%). Although absolute errors remain in a moderate range, the wider scatter in the very low vibration region indicates reduced local discrimination for fine separation of low-amplitude cases. Because the true values in this range occupy a narrow interval with low variance, variance-normalized metrics are highly sensitive to small deviations. Therefore, local performance in this region is interpreted

Journal Pre-proof

primarily using absolute error metrics. This should be considered when interpreting the model for early warning or minor fault detection applications, whereas the present surrogate remains suitable for global comfort trend analysis and scenario screening within the current simulation-derived dataset.

Fig. 4 illustrates model validation through prediction-interval analysis and learning-curve diagnostics. The prediction intervals were relatively narrow for many data points, indicating high confidence and stable predictive behaviour across much of the dataset. Predicted values ranged from near zero to above 0.3, demonstrating that the surrogate-maintained accuracy across both low- and moderate-discomfort regimes. The distribution of the 95% prediction intervals showed that most predictions fell within reasonable proximity to the simulation-derived values, allowing practitioners to identify input conditions where uncertainty increases and additional validation may be required.

Learning-curve analysis supported these results. As the training portion increased from 10% to 100%, the mean training score rose from 0.832 to 0.928, and the mean test score increased from 0.816 to 0.898. This parallel improvement confirms that the model benefited from additional training data and was able to learn progressively more detailed patterns. At the largest training size, the model reached a final training score of 0.928 (±0.0233) and a test score of 0.898 (±0.0265), indicating consistent generalization and suggesting that further data collection could support incremental performance gains.

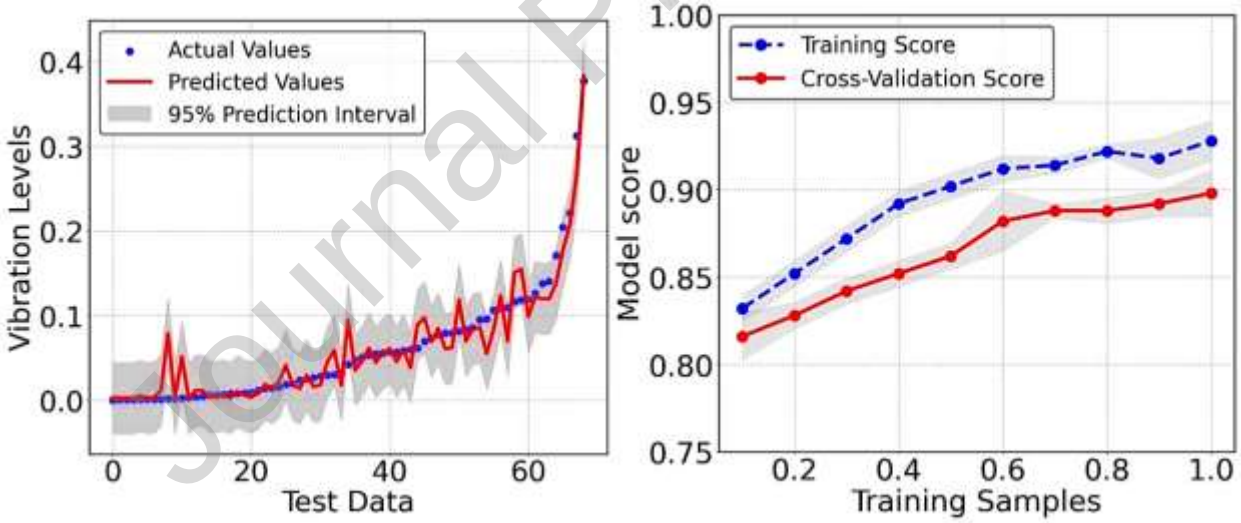

Fig. 4 (Left) 95% prediction intervals for ML-predicted vibration values compared with simulation derived values. (Right) Learning curves showing mean training and validation scores across ten training-set sizes (10%–100%). Error bars denote ±1 SD across 5-fold cross-validation

Fig. 5 presents the sensitivity analysis of the three input features (Speed, Wavelength, Track Type). The figure shows the variation in predicted vibration levels when each feature is systematically varied across its operational range, with Speed shown in km/h, Wavelength in meters, and Track Type treated as a categorical variable. At lower speeds (≤200 km/h), the predicted vibration level remained relatively stable (0.034–0.049 m/s²), indicating marginal

Journal Pre-proof

comfort deterioration. However, between 200 and 250 km/h, noticeable increases in predicted vibration (0.079–0.097 m/s²) emerged, followed by a sharp rise near 300 km/h (~0.15 m/s²). A further substantial increase appeared between 300 and 350 km/h (~0.25 m/s²), with the most pronounced changes above 350 km/h (0.33–0.36 m/s²). By 400 km/h, the predicted vibration peaked at 0.365 m/s², indicating significantly reduced comfort at ultra-high speeds.

Short wavelengths (1–10 m) exhibited elevated vibration levels, with a peak around 5–6 m, reflecting their strong influence on ride comfort. As wavelengths increased into the medium range (10–40 m), vibration levels diminished but displayed fluctuations around 40–65 m, consistent with localized resonances associated with vehicle–track dynamics. Beyond 70 m, overall vibrations decreased, although minor fluctuations persisted. The track-irregularity sensitivity further showed that OVI produced a notable predicted level (~0.152 m/s²), OLI produced moderate levels (~0.105 m/s²), and AGC generated the highest (~0.236 m/s²), indicating that combined vertical–lateral irregularities impose the largest comfort penalties. These sensitivity patterns highlight which speed–wavelength combinations present greater discomfort risk and confirm the importance of targeted control over both operational speed and mid-range wavelength disturbances to maintain optimal ride comfort.

Fig. 6 shows the relative importance of XGBoost, Gradient Boosting, and their combined interaction within the Random Forest-based meta-model. The combined feature ('XGB + GB') achieved the highest importance score (0.98), demonstrating that integrating predictions from both algorithms substantially enhances the meta-model's accuracy. Individually, XGBoost contributed an importance score of 0.58, while Gradient Boosting contributed 0.51, indicating that both learners were influential, with XGBoost offering a marginal advantage due to its stronger capability to model nonlinearities and interaction effects. The large difference between the combined and individual importance values illustrates how ensemble stacking leverages complementary strengths, capturing patterns that may not be fully represented by a single algorithm. In practical terms, this synergy suggests that design, diagnostic, or maintenance decisions guided by the meta-model are likely to be more reliable than those based solely on XGBoost or Gradient Boosting.

Journal Pre-proof

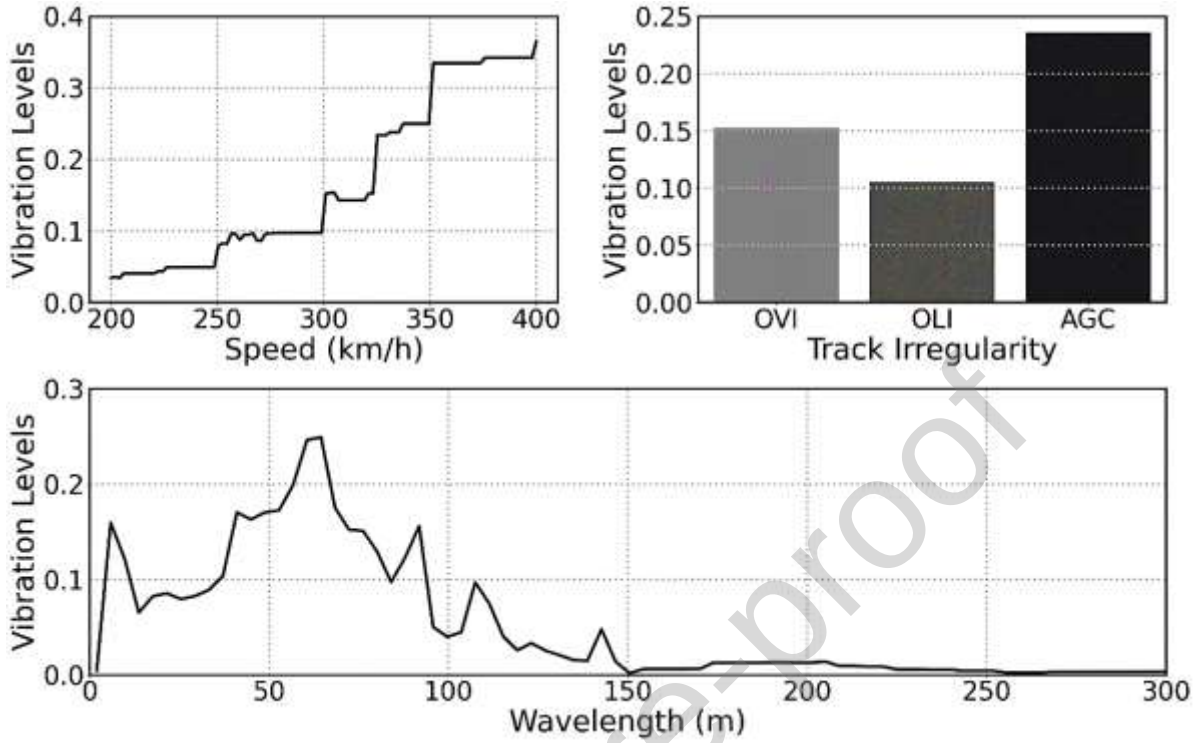

Fig. 5 Sensitivity analysis of the trained model showing the effect of Speed (km/h), Track Irregularity type (OVI, OLI, AGC), and Wavelength (m) on predicted vibration levels (m/s² r.m.s.). Each panel isolates the influence of one input while the remaining features are held at representative values.

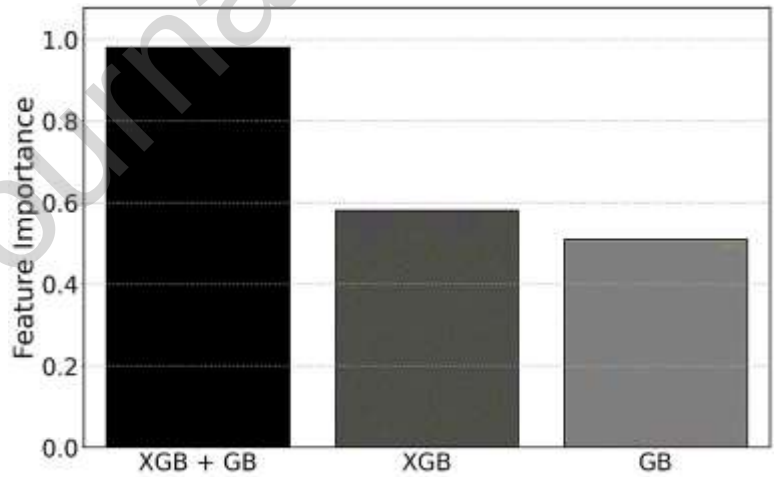

Fig. 6 Relative importance of XGBoost, Gradient Boosting, and their combined contribution (XGB + GB) within the RF-based meta-model. Higher bars indicate greater contribution to predicting overall weighted vibration (m/s² r.m.s.)

DISCUSSION

The present study addresses the challenge of evaluating ride comfort in HSR systems under expanding operational regimes, particularly in the 300–400 km/h range where existing standards

Journal Pre-proof

and empirical data remain limited. Rather than proposing a standalone predictive model for direct field deployment, this work develops a simulation-aligned ML surrogate that reproduces the outputs of a validated vehicle–track interaction model. This distinction is important from an engineering perspective: the surrogate does not replace physics-based analysis but enables rapid and scalable approximation of its results. As such, the contribution of this work lies in enhancing the usability of validated dynamic models by significantly reducing computational cost, thereby supporting large-scale scenario exploration, early-stage design assessment, and sensitivity analysis across key operating variables.

Within this framework, the ML model functions as a computational accelerator for the underlying multi-body simulation. While each high-fidelity simulation captures detailed vehicle–track dynamics, it is not practical to evaluate thousands of combinations of speed, wavelength, and geometry in a design or planning context. The surrogate model addresses this limitation by delivering near-instant predictions that closely approximate simulation outputs within the validated parameter space. This enables engineers to identify critical operating regimes, such as speed thresholds and wavelength bands associated with comfort degradation, and to prioritize mitigation strategies more efficiently. The approach therefore provides practical engineering value as a decision-support tool, particularly in early-stage analysis and comparative evaluation scenarios.

The findings confirm that speeds exceeding 300 km/h introduce threshold-dependent comfort degradation, with the 300–400 km/h band marking a nonlinear transition zone where ride quality significantly deteriorates. This highlights the need for speed-targeted interventions to maintain comfort at high speeds [1]. Analysis of track irregularity types indicates that combined alignment deviations (AGC) exert the most substantial impact at higher speeds, reinforcing the need to prioritize multi-axis irregularity correction in maintenance protocols. The 50–100 m wavelength range was found to influence comfort prominently, with localized resonances likely amplifying discomfort within this band. Maintenance strategies that focus on this range may yield more efficient outcomes than traditional methods, which have historically emphasized shorter-wavelength defects [15]. The results thus point to the need for technological advancements in detecting and correcting long-wavelength irregularities, potentially reshaping future track monitoring systems. Moreover, track-specific differences in vibration response emphasize that comfort preservation must follow a context-aware, infrastructure-specific strategy, with implications for both maintenance planning and new HSR line design. Collectively, these findings support a shift toward precision maintenance guided by predictive modeling.

Given these operational demands, there is a strong need for computational models that deliver accurate and interpretable comfort predictions across a range of operating conditions. Five ensemble learners were evaluated, among which XGBoost consistently outperformed Random Forest, Extra Trees, Gradient Boosting, and AdaBoost. Its superior performance highlights the importance of capturing nonlinear, interaction-driven dynamics when modeling speed–wavelength–track relationships. The study also showed that raw features alone were insufficient: initial models using only speed, wavelength, and track type exhibited suboptimal predictive

Journal Pre-proof

accuracy. To address this, domain-informed feature transformations were incorporated, including second-degree polynomial expansions of Speed and Wavelength, grounded in rail dynamics literature describing resonance amplification and wavelength–speed coupling. These transformations revealed hidden nonlinearities and encoded threshold transitions observed at high speeds. Additionally, standard scaling equalized the influence of input magnitudes, ensuring balanced learning. Together, these steps pushed the best model's predictive performance beyond 90% R^2, establishing a strong basis for physically meaningful comfort prediction. Nonetheless, the surrogate is limited by using only three primary input descriptors, which capture major components of ride dynamics but do not explicitly represent additional contributors such as suspension parameters, curvature, vehicle–track interaction variability, and passenger load. It is important to note, however, that the effects of vehicle suspension and dynamic response are implicitly embedded within the simulation-derived outputs used for training, since the underlying multi-body model already incorporates these physical mechanisms. Future work can extend this framework by explicitly including vehicle parameters and predicting intermediate vehicle-body acceleration responses prior to comfort evaluation, as suggested by the reviewer, to further enhance physical interpretability and generalizability.

Parsimony considerations were preliminarily explored to ensure that improvements in prediction were not the result of unnecessary model complexity. Simpler configurations, such as models trained with raw features or without engineered terms, produced lower predictive power and poorer generalization. This confirmed that the added nonlinear transformations and stacked architecture were functionally necessary rather than redundant. Balancing model expressiveness with simplicity remains important for deployment, particularly under real-time or embedded-system constraints. Future research may apply quantitative parsimony assessments such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), parameter sparsity, inference latency, and memory usage to further evaluate trade-offs between accuracy and simplicity. Such evaluation can help identify minimal but high-performing architectures suitable for settings such as real-time onboard diagnostics or embedded systems.

Building upon XGBoost's strong baseline performance, the meta-models were designed to evaluate whether combining two strong tree-boosting learners could further reduce residual error; while XGBoost and Gradient Boosting are both boosting-based methods, the observed improvement of the RF meta-model over XGBoost ($\Delta R^2 = +0.0165$) indicates a measurable accuracy gain, and the trade-off between this gain and added architectural complexity is noted for engineering deployment. The enhanced performance of the Random Forest-based meta-model demonstrates that tree-based aggregators can effectively fuse predictions from diverse base models such as XGBoost and Gradient Boosting. By contrast, the MLP meta-model performed slightly below XGBoost, suggesting that neural architectures may require larger or richer datasets to capture residual patterns not already modeled by tree-based methods.

Beyond predictive accuracy, model robustness, interpretability, and transparency were emphasized as essential for integration into planning and operations. Prediction-interval and learning-curve analyses confirmed that the meta-model generalized well and benefited from

Journal Pre-proof

additional training data. Final training and test scores (0.92 ± 0.0233 and 0.898 ± 0.0265) demonstrated progressive learning and resistance to overfitting. In addition, the piecewise error analysis showed reduced local discrimination in the very low vibration range (< 0.05 m/s²), even though the errors remained small in absolute terms. This reflects difficulty in resolving closely spaced low-amplitude cases rather than failure of global trend prediction. This should be considered when interpreting the present surrogate for early warning or minor fault detection on good track conditions. Sensitivity analysis revealed critical discomfort thresholds (e.g., 300–400 km/h, 5–10 m wavelengths, AGC irregularities), while feature-importance analysis confirmed that stacked predictions (XGB + GB) contributed more (0.98) than individual learners (XGB: 0.58, GB: 0.51). This synergistic gain across diverse learners shows that ensemble stacking enhances pattern capture beyond what a single model can achieve.

Collectively, these results support a transition toward compact, interpretable, and domain-aligned multi-model pipelines for next-generation comfort prediction. The meta-model's architecture remains lightweight while achieving high fidelity, making it suitable for scenario exploration, adaptive speed planning, and predictive maintenance in HSR. Future research can build on this foundation by including additional physical parameters, exploring advanced sequence models, and strengthening transferability through field data or laboratory-based human response studies. In this way, transparent ML can help bridge simulation-based vehicle–track dynamics with passenger experience objectives as systems progress toward speeds exceeding 400 km/h.

CONCLUSIONS

This study presents a simulation-aligned ML framework for predicting ISO 2631-1 ride comfort metrics based on outputs generated by a validated vehicle–track interaction model. The results demonstrate that ensemble-based surrogate models can accurately reproduce vibration–comfort relationships across a range of operating conditions, including the identification of critical speed thresholds and wavelength bands associated with comfort degradation. Among the ensemble learners evaluated, XGBoost produced the strongest individual performance, while a stacked Random Forest meta-model combining XGBoost and Gradient Boosting predictions achieved the highest overall accuracy, with test R^2 values near 0.92. Sensitivity analysis and feature-importance findings confirmed that the surrogate captured physically interpretable relationships across speed, wavelength, and irregularity type. Cross-validation, prediction intervals, and learning curves supported stable generalization within the simulated domain. Importantly, the proposed model is intended as a computational surrogate of the validated simulation rather than a universal predictor of real-world ride comfort. Its primary contribution lies in enabling rapid scenario evaluation, sensitivity analysis, and large-scale design exploration within a defined modeling framework, thereby enhancing the practical usability of physics-based simulation in HSR engineering.

The study faced constraints and identified several directions for future improvement, including the incorporation of additional physical inputs such as suspension parameters, curvature, vehicle–track interaction, and passenger load. The dataset consists of 345 simulation-derived samples, multiple measures, including cross-validation, prediction-interval estimation, learning-curve analysis and perturbation-based sensitivity tests, were applied to strengthen generalization.

Journal Pre-proof

Broader datasets would further support transferability to new operating conditions. Additional methodological exploration, including Support Vector Regression, Gaussian Process Regression, recurrent networks, Temporal Convolutional Networks, or Transformer-based models, may provide further gains as datasets expand or diversify. A further limitation is that the dataset provides only aggregated ISO frequency-weighted r.m.s. values, not the time-domain acceleration signals needed to evaluate crest factor and compute VDV/MTVV for shock-dominated vibration. Therefore, the present results address comfort prediction under the r.m.s.-based ISO metric and do not quantify shock sensitivity. If weighted time histories become available, future work can enable explicit crest-factor screening and training of surrogate models for VDV/MTVV in addition to r.m.s. Generalization to railway networks with different track irregularity PSD shapes (e.g., different bandwidths/spectral slopes) is not established by the present dataset and would require adding PSD-descriptor inputs (e.g., band powers or spectral slope parameters) to the surrogate model. A further limitation is that local predictive discrimination is weaker in the very low vibration range (< 0.05 m/s²), which may limit the current model's suitability for threshold sensitive early warning of minor faults without additional low amplitude focused training and expanded datasets.

It should also be noted that the stacked model uses two boosting based base learners, so the gain over XGBoost is modest and comes with added complexity. For engineering deployment, XGBoost remains a strong single model option, while future work can reduce complexity by using a simpler meta learner such as ridge regression or by adding more diverse base learners when larger datasets are available.

Techniques such as SHAP or LIME [43] offer additional interpretability and may support regulators and engineers in understanding factors influencing predicted discomfort. Human-centred studies can offer complementary validation: laboratory tests using vibration platforms or motion simulators could expose participants to model-predicted vibration environments and collect subjective and physiological responses. Such developments may help bridge engineering-based comfort predictions with human experience as future HSR systems move toward operating speeds that exceed 400 km/h.

An important extension of this work is to develop multi-stage ML frameworks that explicitly incorporate vehicle parameters and predict vehicle-body acceleration prior to comfort computation, which would further align data-driven models with physics-based comfort evaluation principles.

SOURCE DATA

The source data used in this study can be accessed at https://osf.io/ucmfa/ and https://eprints.soton.ac.uk/432605/ to support reproducibility and further research.

REFERENCES

[1] C. Liu, D. Thompson, M.J. Griffin, M. Entezami, Effect of train speed and track geometry on the ride comfort in high-speed railways based on ISO 2631-1, Proceedings of the Institution of

Journal Pre-proof

Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 234 (2020) 765–778. https://doi.org/10.1177/0954409719868050.
[2] C. Chang, X. Ding, Z. Sun, Y. Yu, L. Zhang, A survey on the mechanism and countermeasures of low-frequency swaying of high-speed trains caused by aerodynamic loads, Engineering Applications of Artificial Intelligence 126 (2023) 107162. https://doi.org/10.1016/j.engappai.2023.107162.
[3] M.J. Griffin, Discomfort from feeling vehicle vibration, Vehicle System Dynamics 45 (2007) 679–698. https://doi.org/10.1080/00423110701422426.
[4] T. Canada, Transportation 2030: Green and Innovative Transportation, ACK 13294803 (2019). https://tc.canada.ca/en/corporate-services/transportation-2030-green-innovative-transportation (accessed March 25, 2025).
[5] D.C. Vock, From Promises to Progress for High-Speed Rail, American Planning Association (n.d.). https://www.planning.org/planning/2024/nov/from-promises-to-progress-for-high-speed-rail/ (accessed March 25, 2025).
[6] orbiteers, Reducing Carbon Emissions - Why High Speed Rail, High Speed Rail Alliance (2022). https://www.hsrail.org/blog/how-will-hsr-reduce-carbon/ (accessed March 25, 2025).
[7] Z. Litvina, Smart and affordable rail services in the EU: a socio-economic and environmental study for High-Speed in 2030 and 2050, Europe's Rail (2023). https://rail-research.europa.eu/publications/smart-and-affordable-rail-services-in-the-eu-a-socio-economic-and-environmental-study-for-high-speed-in-2030-and-2050/ (accessed March 25, 2025).
[8] P. Garrido Martínez-Llop, J. de D. Sanz Bobi, M. Olmedo Ortega, Time consideration in machine learning models for train comfort prediction using LSTM networks, Engineering Applications of Artificial Intelligence 123 (2023) 106303. https://doi.org/10.1016/j.engappai.2023.106303.
[9] J. Wang, X. Xiao, L. Peng, J. Wang, Y. He, X. Sheng, Modelling metro-induced environmental vibration by combining physical-numerical and deep learning methods, Mechanical Systems and Signal Processing 220 (2024) 111687. https://doi.org/10.1016/j.ymssp.2024.111687.
[10] W. Zhai, S. Zhu, Track Design, Dynamics and Modelling, in: Handbook of Railway Vehicle Dynamics, Second Edition, 2nd ed., CRC Press, 2019.
[11] ISO 2631-1:1997, Mechanical vibration and shock — Evaluation of human exposure to whole-body vibration — Part 1: General requirements, (2010). https://www.iso.org/standard/7612.html (accessed August 12, 2023).
[12] BS 6841:1987, Guide to measurement and evaluation of human exposure to whole-body mechanical vibration and repeated shock, Https://Www.En-Standard.Eu (n.d.). https://www.en-standard.eu/bs-6841-1987-guide-to-measurement-and-evaluation-of-human-exposure-to-whole-body-mechanical-vibration-and-repeated-shock/ (accessed July 29, 2024).
[13] BS EN 13848-5, Railway applications - Track - Track geometry quality - Part 5: Geometric quality levels - Plain line, switches and crossings | GlobalSpec, (2017). https://standards.globalspec.com/std/10257419/en-13848-5 (accessed July 29, 2024).
[14] F. Xu, China High-Speed Railway: Country's Golden Name Card, in: F. Xu (Ed.), The Belt and Road: The Global Strategy of China High-Speed Railway, Springer, Singapore, 2018: pp. 3–17. https://doi.org/10.1007/978-981-13-1105-5_1.
[15] T. Wang, Q. Luo, J. Liu, G. Liu, H. Xie, Method for slab track substructure design at a speed of 400 km/h, Transportation Geotechnics 24 (2020) 100391. https://doi.org/10.1016/j.trgeo.2020.100391.
[16] X. Xiao, Xu ,HanWen, Yang ,Yi, Chen ,Peng, Q. and Hu, Analysis of the influence of track irregularity on high-speed train ride comfort, Vehicle System Dynamics 62 (2024) 1658–1685. https://doi.org/10.1080/00423114.2023.2250888.
[17] M. Spiryagin, Edelmann ,Johannes, Klinger ,Florian, C. and Cole, Vehicle system dynamics in digital twin studies in rail and road domains, Vehicle System Dynamics 61 (2023) 1737–1786. https://doi.org/10.1080/00423114.2023.2188228.

Journal Pre-proof

[18] Nitish, A.K. Singh, Dynamic modeling and ride comfort evaluation of railway vehicle under random track irregularities: A case study of a Linke-Hofmann-Busch coach, Journal of Engineering Research 12 (2024) 984–993. https://doi.org/10.1016/j.jer.2023.08.017.
[19] J. Gao, X. Wang, Comparative analysis of ride comfort evaluation indices of high-speed vehicles based on a vehicle-seat-human body coupled dynamics model, Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-Body Dynamics 238 (2024) 507–523. https://doi.org/10.1177/14644193241291468.
[20] K. Zhong, J. Wang, S. Xu, C. Cheng, H. Chen, Overview of fault prognosis for traction systems in high-speed trains: A deep learning perspective, Engineering Applications of Artificial Intelligence 126 (2023) 106845. https://doi.org/10.1016/j.engappai.2023.106845.
[21] S.R. Bose, V.S. Kumar, C. Sreekar, In-situ enhanced anchor-free deep CNN framework for a high-speed human-machine interaction, Engineering Applications of Artificial Intelligence 126 (2023) 106980. https://doi.org/10.1016/j.engappai.2023.106980.
[22] M. Peng, H. Tang, Y. Kou, Adversarial and self-adaptive domain decomposition physics-informed neural networks for high-order differential equations with discontinuities, Engineering Applications of Artificial Intelligence 145 (2025) 110156. https://doi.org/10.1016/j.engappai.2025.110156.
[23] M. Zhang, W. Xu, J. Mo, Z. Xiang, Z. Zhou, Monitoring the deterioration state of braking friction performance of high-speed train under imbalanced data, Engineering Applications of Artificial Intelligence 143 (2025) 109939. https://doi.org/10.1016/j.engappai.2024.109939.
[24] T. Wang, X. Li, Y. Lu, L. Dong, F. Shi, Z. Lin, An efficient thermal comfort prediction method for indoor airflow environment using a CFD-based deep learning model, Building and Environment 267 (2025) 112246. https://doi.org/10.1016/j.buildenv.2024.112246.
[25] Q. Hu, X. Guan, X. Wu, Risk diagnosis model for high-speed rail safety operation in big-data environment, Journal of Traffic and Transportation Engineering (English Edition) 12 (2025) 12–22. https://doi.org/10.1016/j.jtte.2023.03.003.
[26] A. Kasimu, W. Zhou, Q. Meng, Y. Wang, Z. Wang, Q. Zhang, Y. Peng, Performance evaluation of pretrained deep learning architectures for railway passenger ride quality classification, Alexandria Engineering Journal 118 (2025) 194–207. https://doi.org/10.1016/j.aej.2025.01.007.
[27] A. Singh, N. Nawayseh, S. Rakheja, Ensemble modeling for predicting head vibration based on driving seating conditions: Towards adaptive seating systems, Engineering Applications of Artificial Intelligence 145 (2025) 110174. https://doi.org/10.1016/j.engappai.2025.110174.
[28] Y. Cao, T.A. Geddes, J.Y.H. Yang, P. Yang, Ensemble deep learning in bioinformatics, Nat Mach Intell 2 (2020) 500–508. https://doi.org/10.1038/s42256-020-0217-y.
[29] Y. Wang, Y. Pang, T. Xue, S. Zhang, X. Song, Ensemble learning based hierarchical surrogate model for multi-fidelity information fusion, Advanced Engineering Informatics 60 (2024) 102535. https://doi.org/10.1016/j.aei.2024.102535.
[30] A.A. Khan, O. Chaudhari, R. Chandra, A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation, Expert Systems with Applications 244 (2024) 122778. https://doi.org/10.1016/j.eswa.2023.122778.
[31] N. Arsov, M. Pavlovski, L. Basnarkov, L. Kocarev, Generating highly accurate prediction hypotheses through collaborative ensemble learning, Sci Rep 7 (2017) 44649. https://doi.org/10.1038/srep44649.
[32] A. Singh, N. Nawayseh, P. Doyon-Poulin, S. Milosavljevic, K.N. Dewangan, Y. Kumar, S. Samuel, Comparative analysis of classical and ensemble models for predicting whole body vibration induced lumbar spine stress. A case study of agricultural tractor operators, International Journal of Industrial Ergonomics 108 (2025) 103775. https://doi.org/10.1016/j.ergon.2025.103775.
[33] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics 378 (2019) 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.

Journal Pre-proof

[34] L. Breiman, Random Forests, Machine Learning 45 (2001) 5–32. https://doi.org/10.1023/A:1010933404324.
[35] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach Learn 63 (2006) 3–42. https://doi.org/10.1007/s10994-006-6226-1.
[36] J.H. Friedman, Stochastic gradient boosting, Computational Statistics & Data Analysis 38 (2002) 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2.
[37] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA, 2016: pp. 785–794. https://doi.org/10.1145/2939672.2939785.
[38] Y. Freund, R.E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, Journal of Computer and System Sciences 55 (1997) 119–139. https://doi.org/10.1006/jcss.1997.1504.
[39] A. Singh, N. Nawayseh, P. Doyon-Poulin, S. Milosavljevic, S. Rakheja, Y. Kumar, K.N. Dewangan, C. Trask, S. Samuel, Multi-model machine learning for predicting tractor operator discomfort caused by whole-body vibration, Computers and Electronics in Agriculture 243 (2026) 111375. https://doi.org/10.1016/j.compag.2025.111375.
[40] L. Rokach, Pattern Classification Using Ensemble Methods, World Scientific Publishing Co., Inc., USA, 2010.
[41] D. Carneiro, M. Guimarães, M. Carvalho, P. Novais, Using meta-learning to predict performance metrics in machine learning problems, Expert Systems 40 (2023) e12900. https://doi.org/10.1111/exsy.12900.
[42] F. Rustam, A.A. Reshi, I. Ashraf, A. Mehmood, S. Ullah, D.M. Khan, G.S. Choi, Sensor-Based Human Activity Recognition Using Deep Stacked Multilayered Perceptron Model, IEEE Access 8 (2020) 218898–218910. https://doi.org/10.1109/ACCESS.2020.3041822.
[43] A. Holzinger, A. Saranti, C. Molnar, P. Biecek, W. Samek, Explainable AI Methods - A Brief Overview, in: A. Holzinger, R. Goebel, R. Fong, T. Moon, K.-R. Müller, W. Samek (Eds.), xxAI - Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers, Springer International Publishing, Cham, 2022: pp. 13–38. https://doi.org/10.1007/978-3-031-04083-2_2.
[44] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, R. Ranjan, Explainable AI (XAI): Core Ideas, Techniques, and Solutions, ACM Comput. Surv. 55 (2023) 194:1-194:33. https://doi.org/10.1145/3561048.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Highlights

- Physics-guided meta-model predicts ISO 2631-1 ride comfort with 92% accuracy
- Model uses speed, wavelength, and track class to capture key vibration dynamics
- Ensemble stacking improves prediction beyond individual tree-based learners
- Identifies 300–400 km/h as critical speed zone for discomfort escalation
- Initiates a scalable and interpretable approach for future real-time deployment

Journal Pre-proof