



**Titre:** Détection automatique des complications suivant les chirurgies orthopédiques des membres inférieurs par imagerie thermique

**Auteur:** Étienne Lescarbeault

**Date:** 2024

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Lescarbeault, É. (2024). Détection automatique des complications suivant les chirurgies orthopédiques des membres inférieurs par imagerie thermique [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.  
Citation: <https://publications.polymtl.ca/61701/>

 **Document en libre accès dans PolyPublie**  
Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/61701/>  
PolyPublie URL:

**Directeurs de recherche:** Lama Séoud  
Advisors:

**Programme:** Génie informatique  
Program:

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

**Détection automatique des complications suivant les chirurgies orthopédiques  
des membres inférieurs par imagerie thermique**

**ÉTIENNE LESCARBEAULT**

Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*  
Génie informatique

Décembre 2024

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Détection automatique des complications suivant les chirurgies orthopédiques  
des membres inférieurs par imagerie thermique**

présenté par **Étienne LESCARBEAULT**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*  
a été dûment accepté par le jury d'examen constitué de :

**Farida CHERIET**, présidente

**Lama SEOUD**, membre et directrice de recherche

**Thomas HURTUT**, membre

## DÉDICACE

*À mon père, pour les sciences*

*À ma mère, pour la résilience*



## REMERCIEMENTS

C'est donc après six ans et demi que ce chapitre, ou plutôt rêve lucide (parfois non) se termine. Le chemin tracé à date ne peut être décrit sans raconter tous les détails, compter les toutes les flaques de boue comme toutes les fleurs, ou encore tous les levers et couchers de soleil où équations différentielles et problèmes d'algèbre vectorielle furent résolus.

C'est parfois dur à croire, après avoir traversé un si long parcours, qu'on retourne au camp de base pour un nouveau départ vers un sentier plus long que le dernier. Surtout lorsqu'on choisit soi-même le petit détour de deux ans. C'est pourtant ce petit bout de chemin que j'ai le plus apprécié. Après tant de lectures, de préparation, d'écriture, d'essais, de victoires et d'échecs à trouver une solution, on comprend pourquoi on appelle ça de la recherche.

Heureusement, le chemin ne s'est pas fait seul. Au travers des nombreuses rencontres, des amitiés solides se sont formées et des moments mémorables sont à jamais inscrits. Un tel sentier ne se traverse jamais en solo dès la première fois et j'ai eu la chance d'avoir une guide incroyable au cours des dernières années. Lama, merci pour tout.

C'est donc dans ces dernières lignes qu'on se rend compte du sentier qu'il reste à faire, où la vie ne fait que nous donner cette carte abstraite, sans destination ou chemin tracé d'avance, mais nous dit d'avancer à notre gré dans un parcours promis de beaux paysages. Je ne saurais autant vous remercier, amis et collègues, d'avoir aidé à son élaboration.

## RÉSUMÉ

On sait, depuis de nombreuses années, que la température surfacique du corps humain est un indicateur clé de son état de santé. L'imagerie thermique se veut un outil efficace pour visualiser les distributions d'intensités sur la peau, ouvrant de nouvelles possibilités de diagnostic. De plus, la symétrie bilatérale des températures sur les membres opposées peut être utilisée pour détecter certaines maladies. Avec l'intérêt grandissant pour les technologies de diagnostic assistées par ordinateur et l'intégration de l'apprentissage automatique profond, une modélisation efficace de ces distributions de température serait souhaitable. En apprenant les représentations de texture et en comparant leurs asymétries, nous émettons l'hypothèse que les vecteurs latents de texture entre les thermographies des parties du corps opposées permettent de décrire efficacement les différences de température qui peuvent signaler une anomalie dans le domaine thermique. Le modèle développé, baptisé TADA-SAE (*Texture Anomaly Detection using Swapping Autoencoder*), est testé sur l'ensemble de données privées OrthoPOT (*Orthopedics Post-Operative Thermograms*) pour la détection de complications post-chirurgicales suivant une arthroplastie totale du genou ou de la hanche et sur l'ensemble public DMR-IR (*Database for Mastology Research with Infrared Image*) pour la détection du cancer du sein. Un algorithme de forêt d'isolation *iForest* entraîné sur les attributs de symétrie entre les textures et des données cliniques atteint un score AUROC de 0.842 sur OrthoPOT et un score AUROC compétitif avec l'état de l'art de 0.988 sur DMR-IR tout en utilisant moins de paramètres que les méthodes traditionnelles. Ainsi, les attributs de texture et leurs différences entre les parties du corps opposées sont pertinentes pour la détection d'anomalies médicales dans le domaine thermique. Notre méthode est la première à utiliser des attributs de textures appris de manière auto-supervisée et leurs asymétries pour la détection d'anomalies médicales avec l'imagerie thermique, démontrant des résultats prometteurs qui pourraient ouvrir la voie à d'autres tâches dans le domaine médical où la symétrie bilatérale peut être exploitée.

## ABSTRACT

It has been known for many years that the surface temperature of the human body is a key indicator of its health status. Thermal imaging is an effective tool to visualize intensity distributions on the skin, opening new diagnostic possibilities. In addition, the bilateral symmetry of temperatures on opposite limbs can be used to detect certain diseases. With the growing interest in computer-aided diagnostic technologies and the integration of deep machine learning, an efficient modeling of these temperature distributions would be desirable. By learning texture representations and comparing their asymmetries, we hypothesize that the latent texture vectors between thermographs of opposite body parts can effectively describe temperature differences that may signal an anomaly in the thermal domain. The developed model, named TADA-SAE (*Texture Anomaly Detection using Swapping Autoencoder*), is tested on the private OrthoPOT (*Orthopedics Post-Operative Thermograms*) dataset for the detection of post-surgical complications following total knee or hip arthroplasty and on the public DMR-IR (*Database for Mastology Research with Infrared Image*) dataset for the detection of breast cancer. An isolation forest algorithm *iForest* trained on the symmetry attributes between textures and clinical data achieves 0.842 AUROC on OrthoPOT and a competitive 0.988 AUROC score on DMR-IR while using fewer parameters than traditional methods. Thus, texture attributes and their differences between opposite body parts are relevant for the detection of medical anomalies in the thermal domain. Our method is the first to use self-supervised learned texture attributes and their asymmetries for medical anomaly detection with thermal imaging, demonstrating promising results that could pave the way for other tasks in the medical domain where bilateral symmetry can be exploited.

## TABLE DES MATIÈRES

DÉDICACE . . . . .	iv
REMERCIEMENTS . . . . .	v
RÉSUMÉ . . . . .	vi
ABSTRACT . . . . .	vii
TABLE DES MATIÈRES . . . . .	viii
LISTE DES TABLEAUX . . . . .	x
LISTE DES FIGURES . . . . .	xi
LISTE DES SIGLES ET ABRÉVIATIONS . . . . .	xiv
LISTE DES ANNEXES . . . . .	xvi
CHAPITRE 1 INTRODUCTION . . . . .	1
CHAPITRE 2 REVUE DES CONNAISSANCES . . . . .	4
2.1 L'utilisation de l'imagerie thermique dans le domaine médical . . . . .	4
2.1.1 Principe de l'imagerie thermique . . . . .	4
2.1.2 Les applications courantes de l'imagerie thermique . . . . .	4
2.1.3 L'imagerie thermique en orthopédie . . . . .	6
2.2 L'extraction des attributs d'images en vision par ordinateur . . . . .	8
2.2.1 Les descripteurs d'images . . . . .	8
2.2.2 Les descripteurs <i>faits à la main</i> . . . . .	8
2.2.3 L'apprentissage automatique des représentations . . . . .	9
2.2.4 En imagerie médicale . . . . .	11
2.2.5 En imagerie thermique . . . . .	12
2.3 Les techniques de détection d'anomalies en vision par ordinateur . . . . .	13
2.3.1 La détection d'anomalies en vision par ordinateur . . . . .	13
2.3.2 En imagerie médicale . . . . .	14
CHAPITRE 3 RATIONNELLE ET OBJECTIFS DU PROJET . . . . .	17

CHAPITRE 4	DÉVELOPPEMENT D'UNE BASE DE DONNÉES POUR LA DÉTECTION D'ANOMALIES SUIVANT UNE CHIRURGIE ORTHOPÉDIQUE . .	19
4.1	Matériel . . . . .	19
4.1.1	Équipement . . . . .	19
4.1.2	Environnement d'acquisition . . . . .	19
4.2	Méthode . . . . .	20
4.2.1	Recrutement des patients . . . . .	20
4.2.2	Protocole d'acquisition . . . . .	21
4.2.3	Traitement des données . . . . .	23
4.2.4	Analyses des données . . . . .	24
CHAPITRE 5	ARTICLE 1 TADA-SAE: EXPLOITING BILATERAL SYMMETRY IN LEARNED TEXTURE REPRESENTATIONS FOR THERMAL MEDICAL IMAGING . . . . .	29
5.1	Abstract . . . . .	29
5.2	Introduction . . . . .	30
5.3	Related Work . . . . .	31
5.4	Methods . . . . .	32
5.4.1	TADA-SAE: Texture feature learning . . . . .	32
5.4.2	Breast cancer detection on DMR-IR . . . . .	38
5.4.3	Post-operative anomaly detection in OrthoPOT . . . . .	40
5.5	Experiments and Results . . . . .	41
5.5.1	Breast cancer detection in DMR-IR . . . . .	44
5.5.2	Post-operative anomaly detection in OrthoPOT . . . . .	46
5.6	Discussion . . . . .	48
5.7	Conclusion . . . . .	51
CHAPITRE 6	CONCLUSION . . . . .	53
6.1	Synthèse des travaux . . . . .	53
6.2	Limitations de la solution proposée . . . . .	53
6.3	Améliorations et perspectives futures . . . . .	55
RÉFÉRENCES	. . . . .	57
ANNEXES	. . . . .	70

## LISTE DES TABLEAUX

Tableau 4.1	Caractéristiques de la caméra thermique FLIR T1020 . . . . .	19
Table 5.1	Classification metrics on DMR-IR dataset . . . . .	44
Table 5.2	Ablation study of TADA-SAE on the DMR-IR dataset . . . . .	45
Table 5.3	Ablation study of TADA-SAE on the OrthoPOT dataset . . . . .	47
Table 5.4	Effect of different attributes on anomaly detection performance on OrthoPOT dataset, with surgery type (ST), weeks since surgery (WSS) and $d_{orthopot}$ . . . . .	48
Tableau B.1	Résultats de la classification par erreur d’interpolation sur l’ensemble de validation de DMR-IR . . . . .	71

## LISTE DES FIGURES

Figure 2.1	Échantillons de la base de données DMR-IR par Silva et al. [1] . . .	6
Figure 2.2	Imagerie thermique d'une opération de remplacement total du genou deux semaines après l'installation de la prothèse, tirée de Gavish et al. [2]. <b>À gauche</b> : Genou droit opéré, montrant la présence d'une cicatrice au niveau du site chirurgical. <b>À droite</b> : Image thermique démontrant l'augmentation de la température surfacique de la jambe opérée, comparée au genou controlatéral. . . . .	7
Figure 2.3	Architecture du SAE tirée de Park et al. [3] . . . . .	12
Figure 2.4	Architecture du LSAE tirée de [4]. Celle-ci est identique au SAE [3] excepté l'ajout d'une fonction de perte contrastive utilisée pour forcer le transfert de textures dans la région des poumons définie par un masque. . . . .	16
Figure 4.1	Montage de la caméra thermique FLIR T1020 sur son trépied. . . . .	20
Figure 4.2	Exemple d'images thermiques prises sur quatre vues pour un patient ayant reçu une ATG ou genou droit. . . . .	22
Figure 4.3	Exemple d'images thermiques prises sur quatre vues pour un patient ayant reçu une ATH à la hanche gauche. . . . .	23
Figure 4.4	Distribution de l'âge des patients par genre. . . . .	24
Figure 4.5	Distribution des types de chirurgie par genre. . . . .	25
Figure 4.6	Relation entre la grandeur et le poids des patients par genre. . . . .	25
Figure 4.7	Distribution de l'IMC par genre. . . . .	26
Figure 4.8	Distribution du nombre de semaines entre l'acquisition et la chirurgie selon le type d'anomalie rencontré. Celles-ci ont été diagnostiquées, sauf le cas de douleur ayant été reporté par le patient. . . . .	27
Figure 4.9	Différentiels de température moyens entre le côté opéré et controlatéral dans les ATG et ATH. . . . .	28
Figure 5.1	Overall architecture of TADA-SAE; training on the private OrthoPOT dataset is illustrated here. The structure and texture latent codes for both legs are used for reconstruction and swapping, while patches are sampled from the swapped image and the original one to ensure texture transfer. The Adaptive Discriminator Augmentations (ADA) module is added over the discriminator $D$ , which enables training with limited data. . . . .	33

Figure 5.2	Samples of normal (top row) and anomalous (bottom row) images from the DMR-IR database [1]. For samples diagnosed with breast cancer, the affected side appears warmer than the contralateral breast. . . . .	39
Figure 5.3	Inference pipeline for the two-stage breast cancer/post-operative anomaly detection method. On DMR-IR, the distance $d$ is the absolute difference between left and right texture vectors, and no clinical data are used for the classification. On the OrthoPOT dataset, due to its limited size, the distance $d$ is the dot product between the texture vectors; $d$ is concatenated with patient-specific data (dashed line) before the classification stage. . . . .	39
Figure 5.4	Thermal images showing lower limbs following a total knee arthroplasty (TKA). The operated leg (left leg) is warmer than the contralateral one around the knee. . . . .	42
Figure 5.5	Frontal view thermal images following TKA for normal (top row) and anomalous (bottom row) patients, with corresponding number of weeks since surgery (WSS). . . . .	43
Figure 5.6	Projected representations of left-right features differences from a simple autoencoder (AE) and from the disentangled texture space learned by TADA-SAE. With a small warming factor ( $WF=0.05$ ) applied to one side in healthy samples, the TADA-SAE is much more efficient at separating the samples according to the main principal components from PCA. Each axis represents one of the two principal components and its explained variance ratio. . . . .	46
Figure 5.7	Effect of texture difference principal components in OrthoPOT classification. We notice that performance decrease on AUROC as the number of principal components increases . . . . .	49
Figure A.1	Réchauffement artificiel unilatéral du sein gauche avec $R = 0.05$ . . . .	70
Figure B.1	Reconstruction et interpolation de textures dans les images thermiques des seins d'une patiente positive. Le sein droit affecté est plus chaud que le sein opposé, et le modèle arrive à intervertir les températures de ceux-ci. . . . .	72
Figure B.2	Erreur au carré entre les pixels des images reconstruites et des interpolations de texture avec le sein opposé. Dans les deux cas, les différences de température sont mises en évidence dans la région mammaire. . .	73
Figure C.1	Matrice de confusion obtenue sur l'ensemble de validation avec TADA-SAE pour la détection de complications dans les ATG. . . . .	75



Figure C.2	Matrice de confusion obtenue sur l'ensemble de validation avec TADA- SAE pour la détection de complications dans les ATH . . . . .	75
------------	---	----

## LISTE DES SIGLES ET ABRÉVIATIONS

ACP	Analyse en composantes principales
ADA	Adaptive discriminator augmentations
AE	Auto-encodeur
ATG	Arthroplastie totale du genou
ATH	Arthroplastie totale de la hanche
AUROC	Area Under receiver operating characteristic curve
AUPRC	Area Under precision-recall curve
CelebA-HQ	CelebFaces Attributes High-quality
CIUSSS	Centre intégré universitaire de santé et de services sociaux
CNN	Convolutional neural network
COVOC	COVID-19 Outcome
CXR-14	ChestX-ray14
DMR-IR	Database for Mastology Research with Infrared Image
GAN	Generative Adversarial Networks
GLCM	Grey-level co-occurrence matrices
HOG	Histogram of Oriented Gradients
I2I	Image-to-image
IEEE	Institute of Electrical and Electronics Engineers
iForest	Isolation Forest
IMC	Indice de masse corporelle
IR	Infrarouge
LBP	Local Binary Pattern
LSAE	Lung Swapping Autoencoder
LWIR	Long-wave infrared
OCSVM	One-class SVM
OrthoPOT	Orthopedics Post-operative Thermograms
PTG	Prothèse totale du genou
PTH	Prothèse totale de la hanche
ReLU	Rectified linear unit
RVB	Rouge, vert, bleu
SAE	Swapping Autoencoder
ST	Surgery type
SVM	Support vector machine

TADA-SAE	Texture Anomaly Detection from Asymmetries using Swapping Autoencoder
TBME	Transactions on Biomedical Engineering
t-SNE	t-distributed Stochastic Neighbor Embedding
TVP	Thrombose veineuse profonde
UMAP	Uniform Manifold Approximation and Projection
VAE	Variational autoencoder
WHO	World Health Organization
WSS	Weeks since surgery

**LISTE DES ANNEXES**

Annexe A	DMR-IR : Sensibilité de l'espace d'asymétries aux changements de température . . . . .	70
Annexe B	DMR-IR : Interpolation de textures . . . . .	71
Annexe C	OrthoPOT : Détection d'anomalies par type de chirurgie . . . . .	74

## CHAPITRE 1 INTRODUCTION

L'arthrose est une pathologie dégénérative causant une dégradation du cartilage au niveau des articulations, souvent associée au vieillissement [5]. Cette maladie chronique causant la destruction des tissus, souvent aux membres inférieurs, est une cause importante de douleurs et d'inconfort chez les personnes âgées. En effet, le *World Health Organisation* (WHO) [6] estime à 528 millions le nombre de personnes atteintes d'arthrose, 73% d'entre elles étant âgées de 55 ans et plus. Bien que plusieurs facteurs comme le genre, l'obésité et le mode de vie jouent un rôle sur la prévalence, le vieillissement, lié avec l'utilisation des articulations, reste toutefois un élément clé dans le développement de la maladie [7]. Outre la douleur, la perte de mobilité et la difficulté croissante à accomplir les tâches quotidiennes demandent un traitement spécialisé. Lorsque l'arthrose atteint une certaine ampleur, il peut être jugé nécessaire de recourir à une chirurgie de remplacement de l'articulation du membre inférieur [8]. Durant la procédure, un tissu artificiel est posé dans l'articulation, remplaçant le cartilage dégradé. L'opération vise donc à placer une prothèse totale de genou (PTG) ou une prothèse totale de la hanche (PTH) afin de remplacer l'articulation endommagée. Cette chirurgie, bien que commune, n'est cependant pas sans risque. Parmi ceux-ci, on recense notamment la thrombose veineuse profonde (TVP) [9] et les infections [10]. La première se manifeste par la formation d'un caillot sanguin dans les veines avoisinantes au site chirurgical. Alors que les symptômes de la TVP ne se manifestent que dans les deux tiers des cas, une étude souligne que près de 80% des embolismes pulmonaires sont dus à une thrombose veineuse dans le système vasculaire des membres inférieurs [11]. Sachant que 61% des patients ne recevant pas d'anticoagulants suivant leur chirurgie développent une TVP [9], un lien direct se trace entre l'opération chirurgicale et le risque d'embolisme pulmonaire. Les infections, quant à elles, atteignent une proportion entre 2% et 3.7% chez les patients opérés, se manifestant parfois jusqu'à un an après l'opération [10]. Ainsi, malgré l'administration de traitements préventifs, les chirurgies de remplacement des genoux et des hanches comportent plusieurs risques parfois asymptomatiques [9] pouvant dégrader la qualité de vie du patient et mettre sa santé en péril même plusieurs mois après l'intervention [12].

En plus des conséquences à court et à long terme sur la qualité de vie du patient, les complications post-chirurgicales ont aussi un effet sur le réseau de la santé. Haidar et al. [13] soulignent que la prolongation du séjour à l'hôpital, l'administration de traitements supplémentaires et la sollicitation du personnel soignant a comme effet une augmentation substantielle des coûts. Cette répercussion affecte aussi les patients, réduisant leur qualité de vie en clinique. Or, dans une étude rétrospective par Bonde et al. [14], il a été prouvé que l'apprentissage profond peut

offrir des performances supérieures aux approches classiques pour la prédiction du risque de complications post-chirurgicales dans la pratique clinique. Cela démontre la pertinence des techniques d'intelligence artificielle dans l'effort de l'évaluation de la récupération des patients, assurant un meilleur support et une meilleure qualité de vie.

À la lumière de ces enjeux, il serait souhaitable d'avoir la possibilité de recourir à des systèmes de suivi ou diagnostic assistés. Ceux-ci permettent d'amoindrir la charge des cliniciens durant le suivi post-opératoire, tout en permettant de détecter les premiers signes de complications. Ainsi, la qualité de vie des patients est mieux assurée à long terme tout en enlevant une charge considérable sur ce volet du réseau de la santé.

Dans des travaux antérieurs, des chercheurs ont constaté que l'imagerie thermique infrarouge non invasive pouvait constituer une modalité d'imagerie diagnostique pour la TVP des membres inférieurs [15], ainsi que pour la détection de l'infection et de l'inflammation des plaies dans les articulations suivant une intervention. Bien que ces travaux n'aient été jusqu'à présent que préliminaires, ils laissent entrevoir la possibilité d'une utilisation dans un système de surveillance à domicile basé sur la vision. En effet, avec les dernières avancées technologiques de capteurs infrarouges, l'imagerie thermique est désormais possible avec des capteurs infrarouges à grande longueur d'onde (LWIR) grand public et pourrait être envisagée pour la surveillance à domicile. Dans ce mémoire, nous proposons de développer un modèle d'apprentissage profond pour la détection d'anomalies post-chirurgicales à partir d'images thermiques du membre inférieur chez les patients ayant suivi une opération de remplacement des articulations du genou ou de la hanche. La combinaison d'un tel algorithme de détection de signes de complications chirurgicales avec un capteur thermique installé à domicile chez le patient permettrait de faire un suivi thermographique étendu et de détecter de manière précoce les signes de potentielles complications post-chirurgicales. Cela permettrait une prise en charge plus rapide et permettrait d'améliorer le processus de rétablissement du patient.

Dans le **CHAPITRE 2**, nous présenterons une revue de littérature et des connaissances propres à la compréhension du travail, plus spécifiquement, l'imagerie thermique et ses applications médicales, l'extraction d'attributs à partir d'images et la détection d'anomalies.

Le **CHAPITRE 3** décrit la rationnelle du projet et énonce les objectifs associés à ce dernier.

Puisqu'aucun ensemble de données public n'existe pour la détection d'anomalies post-chirurgicales dans le domaine thermique, nous discuterons du développement d'une base de données à cet effet dans le **CHAPITRE 4**. Nous aborderons ainsi les besoins de sa création et les objectifs guidant son édification. Dans ce sens, nous développerons un protocole de collecte de la base de données ainsi créée et effectuerons les analyses sur les échantillons recueillis.

Le **CHAPITRE 5** présentera le développement d’une méthode de détection d’anomalies hybride basée sur les différences de textures thermiques entre le membre opéré et controlatéral sous forme d’article soumis à la revue IEEE TBME (Transactions on Biomedical Engineering). Nous y présentons également une étude sur une base de données publique d’images thermiques pour le diagnostic du cancer du sein.

Enfin, le **CHAPITRE 6** conclura ce mémoire en faisant une synthèse des travaux, en soulignant les limitations et en proposant des pistes pour les travaux futurs.

## CHAPITRE 2 REVUE DES CONNAISSANCES

Ce chapitre traitera en détails des connaissances et de la littérature en lien avec le projet, posant les bases à l'idéation de celui-ci. D'abord, la **section 2.1** traitera de l'utilisation de l'imagerie thermique dans le milieu médical en liant son application avec les chirurgies orthopédiques. Pour suivre, la **section 2.2** tissera les liens pratiques du projet avec la médecine en discutant des méthodes d'extraction d'attributs dans les images médicales et thermiques. Enfin, la **section 2.3** fera une revue des techniques et applications pour la détection d'anomalies en vision par ordinateur.

### 2.1 L'utilisation de l'imagerie thermique dans le domaine médical

#### 2.1.1 Principe de l'imagerie thermique

Une caméra thermique a un fonctionnement de base assez intuitif. Tout comme une caméra classique, elle focalise l'énergie sous forme d'ondes électromagnétiques sur un capteur, sans émettre de rayonnement. Cependant, ces dernières sont conçues pour créer une image digitale depuis la lumière visible. C'est là la différence avec une caméra thermique. Les optiques et le capteur sont élaborés pour obtenir une image depuis les ondes infrarouges, ou la chaleur, émanant d'un corps [16]. Ainsi, l'appareil est conçu pour répondre à une longueur d'onde du spectre électromagnétique variant entre  $8\mu m$  et  $14\mu m$  [17]. Il devient ainsi possible de visualiser les intensités des rayonnements émis ou réfléchis par un objet ou, dans un contexte médical, celui du corps humain. Les rayonnements émis par celui-ci à  $27^{\circ}C$  se situent entre 2 et  $20\mu m$  [18]. Ainsi, plusieurs bases de données ont pu être érigées à des fins médicales ou pour le milieu de l'automobile telles que *Plantar Thermogram Database* pour la classification du diabète [19], *DMR-IR* pour la détection du cancer du sein [1] et *FLIR-ADAS* pour la détection d'objets sur la route [20]. Vu la diversité de ses usages, une caméra thermique doit être choisie selon plusieurs spécifications telles que sa sensibilité et sa résolution, éléments importants pouvant augmenter de manière substantielle le coût de l'appareil, variant de quelques centaines à plusieurs milliers de dollars [17].

#### 2.1.2 Les applications courantes de l'imagerie thermique

Les changements de température corporelle sont, depuis de nombreuses années, connus comme des éléments clés dans l'évaluation de l'état de santé d'une personne [21]. Une augmentation globale de celle-ci peut, par exemple, signaler une fièvre servant à combattre une infection ou



un virus. Or, avec le développement rapide des systèmes d'imagerie et des outils de mesure, il devient possible d'observer plus en détails la température surfacique du corps, en particulier sa distribution, reflétant le flux sanguin des vaisseaux superficiels. Les modalités d'imagerie en clinique se résument souvent à l'utilisation d'appareils très onéreux et peu mobiles. Par exemple, les appareils de radiographie, l'imagerie par résonance magnétique et la tomodensitométrie ne peuvent être utilisés que par un membre du personnel qualifié, engendrant des coûts de main d'œuvre supplémentaires en plus d'être difficilement transportables [22]. Cependant, l'imagerie thermique, technique non invasive et non irradiante, se veut une alternative prometteuse au diagnostic par images avec son large champ d'applications [23], tout en ouvrant de nouvelles possibilités de diagnostic.

**Diagnostic du cancer du sein :** Le cancer du sein est la forme de cancer touchant le plus de femmes dans le monde, totalisant, en 2022, 11.5% des cas dans la population globale [24]. Alors que la mammographie se veut une technique populaire pour la détection du cancer du sein, celle-ci provoque des inconforts pour de nombreuses patientes. Une plaque de compression est utilisée afin d'aplatir le sein avant la prise d'image par radiographie. Cette technique est cependant invasive et irradiante. Les images prises par ce type de modalité sont de plus assujetties à de nombreux critères de qualité, nécessitant la présence de personnel spécialisé pour la prise d'images [25]. Il est enfin souvent nécessaire de reprendre les images à raison d'un mauvais positionnement du sein ou l'absence de détails, engendrant des inconforts supplémentaires chez les patientes. Or, avec le développement rapide de l'imagerie thermique, la prise de thermogrammes permet déjà le diagnostic de cancer du sein précoce [26]. En 2014, Silva et al. [1] ont développé un ensemble de données public proposant des images thermiques de poitrines saines et comportant un cancer connu, dont quelques exemplaires sont présentés dans la figure 2.1. L'ensemble, baptisé *Database for Mastology Research with Infrared Image* (DMR-IR), les patientes sous deux protocoles, statique et dynamique avec images thermiques comportant une résolution de  $640 \times 480$  pixels. Dans le premier, cinq images sont capturées : une de face, deux latérales à  $45^\circ$  et deux latérales à  $90^\circ$ . Dans le protocole dynamique, une image est prise à chaque minute durant le refroidissement artificiel de la poitrine, résultant en une séquence de 20 images par patiente. Enfin, les données cliniques des patientes sont aussi incluses avec les images. Bien que certaines inconsistances des données cliniques dans cet ensemble aient été relevées, il n'en reste pas moins le plus utilisé en recherche à ce jour [27]. De nombreux auteurs se sont attardés à la détection du cancer du sein à partir d'images thermiques, obtenant des performances de classification dépassant 85% de précision [28–33]. Les méthodes les plus récentes emploient les réseaux de neurones convolutifs apprenant à reconnaître certains motifs dans les images prévalant à un cancer. Ainsi Mahoro et al. [33] ont, en 2024, entraîné un réseau convolutif résiduel *Resnet-50* [34] à détecter la présence

de cancer, obtenant ainsi une exactitude de 97.26% sur DMR-IR. Dans la même période, Parshionikar et Bhattacharyya [30] ont entraîné un réseau convolutif profond à l'extraction de caractéristiques sur des images thermiques et histologiques. De pair avec un algorithme de sélection d'attributs, les auteurs ont pu atteindre une exactitude presque parfaite de 99.74%.

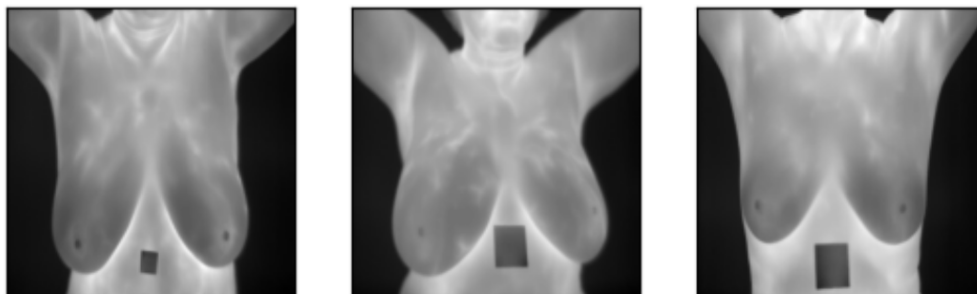


FIGURE 2.1 Échantillons de la base de données DMR-IR par Silva et al. [1]

Outre les méthodes d'apprentissage profond, les techniques traditionnelles moyennant des caractéristiques et descripteurs classiques donnent une perspective plus interprétable sur les décisions faites par un algorithme de classification. Acharya et al. [28] introduisent alors l'extraction d'attributs décrivant les textures des images, sur lesquels une machine à vecteurs de support (SVM) est entraînée à détecter la présence de cancer. Les représentations des textures sont définies par les fonctions mettant en relation les valeurs d'intensité sur les niveaux de gris de l'image thermique. Obtenant une exactitude de 88.10% sur un ensemble de données privé, les auteurs montrent la pertinence du concept de texture dans la thermographie.

### 2.1.3 L'imagerie thermique en orthopédie

**Évaluation de l'hyperthermie locale :** L'utilisation de cette technique d'imagerie voit son utilité s'étendre encore plus loin en touchant au domaine de l'orthopédie. Il a été démontré à plusieurs reprises que l'imagerie thermique se veut un outil efficace pour la détection de l'arthrose au genou [35, 36]. Les changements dans la vascularisation et l'inflammation au niveau de l'articulation causent une augmentation localisée de la température, visible à travers une caméra thermique [36]. On y voit aussi de nombreuses applications dans le suivi post-chirurgical des patients ayant reçu une chirurgie orthopédique. En effet, plusieurs travaux [37, 38] montrent l'efficacité de l'imagerie thermique dans l'évaluation de la surveillance du patient suivant une opération de remplacement complet du genou ou de la hanche dans le traitement de l'arthrose. La vasodilatation survenant après la chirurgie augmente le flux sanguin et, naturellement, une augmentation de température localisée pouvant être efficacement décrite par l'imagerie thermique [39]. La figure 2.2 montre l'augmentation de température

localisée suivant une ATG. Une méta-analyse [2] démontre par ailleurs que l'augmentation de température au site chirurgical atteint son maximum après deux semaines suivant l'opération puis décroît graduellement avec une augmentation moyenne de  $0.3^{\circ}\text{C}$  par rapport au niveau pré-chirurgical chez les patients ayant une guérison normale. Quant aux cas anormaux, il a été démontré que le membre inférieur diagnostiqué d'une TVP comportait une différence de température moyenne avec le membre opposé significative, contribuant à paver la voie vers une évaluation non invasive de la TVP [15]. En effet, la modification du flux sanguin causée par le caillot provoque des différences dans la distribution de la température entre les deux membres inférieurs [40].

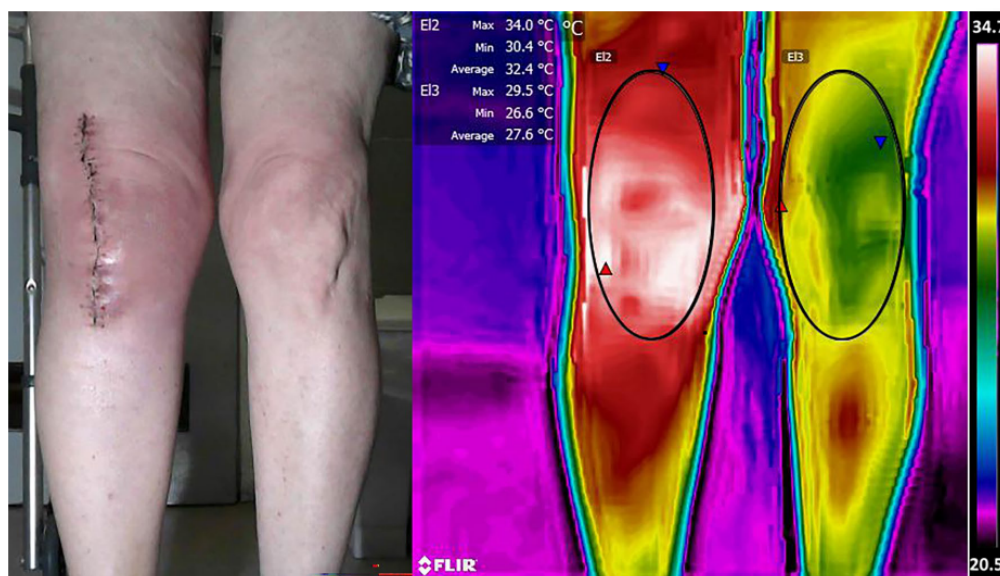


FIGURE 2.2 Imagerie thermique d'une opération de remplacement total du genou deux semaines après l'installation de la prothèse, tirée de Gavish et al. [2]. **À gauche** : Genou droit opéré, montrant la présence d'une cicatrice au niveau du site chirurgical. **À droite** : Image thermique démontrant l'augmentation de la température surfacique de la jambe opérée, comparée au genou controlatéral.

**Évolution de l'hyperthermie locale** : Durant cette même période, Zeng et al. [41] ont analysé l'évolution sur 12 mois de la température superficielle ponctuelle des genoux opérés et controlatéraux en la mettant en relation avec des biomarqueurs sanguins. Ils ont ainsi tracé les relations entre le différentiel de température entre les membres et la réponse immunitaire durant la réhabilitation, démontrant la possibilité d'utiliser le gradient de température entre les jambes gauche et droite comme critère servant à détecter les inflammations et les infections, ces dernières survenant dans 2% à 3.7% des arthroplasties totales du genou (ATG) [41]. Plus tard, Lohchab et al. [38] utilisent l'imagerie thermique en évaluant l'évolution de la température jusqu'à 90 jours après la chirurgie, montrant une corrélation entre celle-ci et la présence

de marqueurs inflammatoires, utilisés dans la détection précoce des infections. Plus récemment, l'étude de Sharma et al. [42] évalue sur une période d'un an l'évolution de la différence de température avec une large cohorte de patients ayant reçu une ATG, dont plusieurs ayant développé une infection au niveau du site chirurgical. Ils ont ainsi relevé des différences de températures plus importantes et prolongées chez ces derniers, prouvant encore une fois la pertinence de la thermographie dans la surveillance post-chirurgicale.

## 2.2 L'extraction des attributs d'images en vision par ordinateur

### 2.2.1 Les descripteurs d'images

Bien que le domaine des descripteurs d'images se veuille assez vaste, on regroupe les méthodes sous deux grandes classes, les descripteurs *locaux* et *globaux* [43]. Cette première famille regroupe les méthodes pouvant extraire des attributs aux niveaux de régions spécifiées de l'image, comme un ensemble de pixels voisins, utilisé notamment pour reconnaître un même objet dans une suite d'image. La deuxième permet plutôt de décrire une image dans son tout, permettant ainsi des comparaisons entre des images entières. Ce concept joue alors un rôle fondamental dans la classification, la détection d'anomalies et le traitement d'images médicales [44]. Il est possible de décrire les attributs de l'image à l'aide de fonctions produisant des ensembles de nombres représentant ses caractéristiques de manière souvent plus compacte. Cette information extraite est alors plus dense en information d'intérêt, permettant plus facilement la reconnaissance de motifs et la comparaison entre plusieurs images. Lorsque ces fonctions sont définies manuellement à l'aide d'analyses, on parle de descripteurs *faits à la main*. À l'opposé, les attributs extraits par des réseaux de neurones entraînés sur ces images sont des attributs *appris*, ou *représentations apprises*.

### 2.2.2 Les descripteurs *faits à la main*

En 1973, Haralick et al. [45] ont pavé le chemin des descripteurs de texture avec les matrices de cooccurrence des niveaux de gris (GLCM, *Grey-level co-occurrence matrices*). Cette méthode permet d'extraire des attributs statistiques entre les pixels voisins en opérant sur le domaine spatial de l'image. Ce premier pas reste toujours utilisé aujourd'hui, notamment en imagerie médicale [46]. On compte aussi des descripteurs caractérisant les couleurs et textures de l'image. Une technique très simple est l'histogramme de couleur. Cette méthode regroupe les pixels, encodés sur 8 bits, de sorte à créer un histogramme comptant le nombre d'occurrences pour chaque valeur d'intensité [47]. Il devient ainsi possible de comparer plusieurs histogrammes d'images entre eux afin de regrouper les images comportant un contenu simi-

laire [46]. Il a suivi, au fil des années, de nombreuses nouvelles méthodes plus efficaces telles que les histogrammes de gradients orientés (*Histogram of Oriented Gradients*, HOG) [48] et les patrons binaires locaux (*Local Binary Patterns*, LBP) [49].

### 2.2.3 L'apprentissage automatique des représentations

**Le concept de représentations** : Même si les descripteurs d'images faits à la main se veulent plus interprétables, les attributs appris par un réseau de neurones convolutif sont cependant souvent plus riches et descriptifs [43]. Or, en apprentissage profond, le réseau entraîné permet d'extraire des représentations. Celles-ci se manifestent sous forme de plusieurs niveaux d'abstraction contenant l'information utile à une tâche, telle que la classification [50]. Ces représentations, sous forme de vecteurs à plusieurs dimensions, sont alors contenues dans ce que l'on appelle l'*espace latent* [51]. Celui-ci peut aussi être visualisé à l'aide d'un algorithme de réduction de dimensionnalité, projetant ces vecteurs dans un espace de dimension moindre. Ainsi, à l'aide d'algorithmes de réduction de dimensionnalité tels que ACP [52], t-SNE [53] ou UMAP [54], on peut par exemple projeter les attributs extraits d'une couche d'un réseau de neurones.

L'apprentissage supervisé n'est heureusement pas la seule manière d'obtenir des représentations riches des images d'un ensemble de données. En effet, il serait fastidieux, voire impossible, d'étiqueter dans leur entièreté des ensembles de données très volumineux ou nécessitant des connaissances expertes. Cette problématique ouvre ainsi la voie à de nouvelles méthodes permettant l'apprentissage de représentations sur un plus grand volume de données non annotées.

**L'apprentissage auto-supervisé et les auto-encodeurs** : C'est en 1986 que le concept d'auto-encodeur a, pour la première fois, fait surface. Dans leur ouvrage, Rumelhart et al. [55] font mention d'une architecture de réseau de neurones dont le but est de reconstruire un échantillon d'entrée  $\mathcal{X}_i$ . Cette architecture peut être découpée en deux composantes, l'encodeur et le décodeur. Alors que le premier apprend à compresser l'échantillon de manière à réduire efficacement sa dimensionnalité, le décodeur est entraîné pour reconstruire l'élément depuis la quantité réduite d'information contenue dans l'espace latent [56]. Ce paradigme d'entraînement, où le réseau doit se corriger lui-même en recomposant une entrée qu'il a compressée, est qualifié d'apprentissage auto-supervisé. La *matière noire de l'intelligence artificielle* [57] permet de pré-entraîner un encodeur sur une quantité impressionnante d'images non annotées, permettant d'obtenir des attributs riches pouvant être utilisés dans une tâche en aval, comme un problème de classification sur un sous-ensemble de données annotées [58]. Outre apprendre à un réseau auto-encodeur à recomposer des images, plusieurs méthodes

d’entraînement et architectures ont su, au fil du temps, surpasser l’état de l’art de l’apprentissage complètement supervisé, notamment dans la classification d’images [59–62]. Par exemple, les techniques d’apprentissage contrastif [59, 60] visent à rapprocher les attributs pour deux images de même classe tout en éloignant les échantillons hors de celle-ci. D’autres méthodes penchent plutôt à apprendre à rapprocher les images entre elles en comparant des paires d’images augmentées pour y extraire des attributs communs [61, 62]. L’encodeur ainsi pré-entraîné offre des représentations robustes pour de nouvelles images. Il ne suffit ensuite, par exemple, que d’ajouter une tête de classification ou de segmentation afin d’accomplir la tâche en aval [58].

**Traduction image vers image :** La création d’attributs riches ne se limite pas qu’au prétexte d’une tâche en aval. Un exemple éloquent soulignant la puissance de la bonne construction d’un espace latent est le domaine de la génération d’image. Par exemple, les auto-encodeurs variationnels (VAE) de Kingma et al. [63] encodent les images sous forme de vecteurs de probabilités. Un échantillonnage aléatoire de cet espace latent permet alors de générer de nouvelles images. Dans la même période, Goodfellow et al. [64] ont proposé les réseaux antagonistes génératifs (GAN), où un réseau génératif  $\mathcal{G}$  reconstruit des images alors qu’un réseau discriminateur  $\mathcal{D}$  tente de discerner si cette image est réelle ou non, suivant les principes d’un jeu min-max dans l’entraînement conjoint de ces réseaux. De la grande versatilité de cette architecture, plusieurs cadres d’entraînement ont surgi, dont une application impressionnante est la traduction image à image (I2I) [65, 66]. Une image d’un certain domaine se retrouve alors transformée en une autre appartenant à un autre domaine, simplement en utilisant l’information de la première [67]. En 2016, Isola et al. [65] ont ainsi démontré l’application des réseaux antagonistes génératifs conditionnels [68] pour la traduction I2I d’images appariées ou alignées, avec leur architecture *Pix2Pix*. Le discriminateur doit alors discerner les vraies paires (image de l’ensemble et sa référence) des fausses paires (image générée et sa référence), tandis que le générateur essaie de traduire l’image de référence. Malgré la qualité des images générées, cette technique ne fonctionne que pour les paires d’images alignées ayant des contours et un contenu communs. Un peu plus tard, Zhu et al. [66] arrivent à entraîner un réseau à lier deux domaines d’images non appariées avec leur technique baptisée *CycleGAN*. Le réseau apprend alors à traduire des images vers celles d’un autre domaine tout en s’assurant de pouvoir reproduire l’image originale à partir de celle traduite. L’encodeur apprend éventuellement les correspondances entre les images des différents domaines, donnant lieu à de nombreuses applications d’I2I où l’acquisition d’images appariées n’est pas réalisable, notamment dans le contexte médical [69–71]. Or, ces techniques ne permettent aucun contrôle intuitif sur la traduction. Les modèles apprennent les correspondances directes entre les paires sans permettre facilement les variations [72].

**Contrôle des attributs** : C’est en 2019 que Karras et al. [73] ont conçu un modèle capable de contrôler les attributs haut niveau, ou le style, d’une image. Entraîné notamment sur le jeu de données CelebA-HQ [74], leur *StyleGAN* permet de générer des images avec un niveau fin de détails, tout en permettant le contrôle haut niveau dans l’image tel que l’âge de la personne, la couleur de cheveux et l’ajout de lunettes. Ce contrôle supplémentaire dans la génération d’images a ainsi inspiré plusieurs variantes et améliorations [3, 75–78], démontrant la puissance du désenchevêtrement de l’espace latent [79]. Or, l’entraînement d’un tel modèle nécessite une quantité importante de données, qui ne sont pas nécessairement disponibles dans un contexte médical ou spécialisé. Ainsi, Karras et al. [76] ont introduit en 2020 un module d’augmentation d’images agissant sur le réseau discriminateur seulement, évitant la fuite d’augmentations lors de la génération.

**Désenchevêtrement de l’espace latent** : Pour lier le concept de texture d’une image discuté plus tôt et la génération d’images, Park et al. [3] ont proposé, en 2020, une architecture permettant la manipulation d’images en découplant l’image en deux codes latents, un représentant la structure et l’autre la texture de l’image. Leur *Swapping Autoencoder* (SAE) rend possible la combinaison de la structure  $\mathcal{S}_a$  d’une image  $I_a$  avec la texture  $\mathcal{T}_B$  d’une image  $I_b$ , générant une image  $I_{a,b}$  comportant les éléments structuraux tels que les contours de la première, mais avec les couleurs, contrastes et luminosité de la seconde. Basée sur l’architecture de StyleGAN2 [75], le modèle est entraîné à reconstruire les images en entrée de manière antagoniste avec un discriminateur classique, tout en essayant de reproduire les imagerie d’une image de référence sur une image à reconstruire. Un discriminateur de cooccurrence est enfin utilisé pour améliorer la qualité des textures générées, en essayant de discerner les imagerie (*patches*) générés de ceux de l’ensemble d’entraînement [3]. La figure 2.3 décrit l’architecture du SAE. Le désenchevêtrement de l’espace latent permet ainsi des opérations arithmétiques entre les vecteurs de texture d’un ensemble d’images, donnant d’excellents résultats dans la traduction I2I suivant une étude sur la perception humaine évaluant la qualité des images transformées [3].

#### 2.2.4 En imagerie médicale

L’entraînement supervisé de modèles en imagerie médicale nécessite l’acquisition coûteuse et laborieuse d’étiquettes par des professionnels de la santé. Il va donc de soi que l’apprentissage auto-supervisé devienne un incontournable pour ce domaine. Ce paradigme peut alors être utilisé pour extraire des attributs d’images tirées d’une même modalité avant d’être ajusté sur un sous-ensemble annoté de plus petite taille [80]. Il devient ainsi possible d’utiliser une grande quantité d’images cliniques non annotées tout en obtenant des performances relevant de l’état

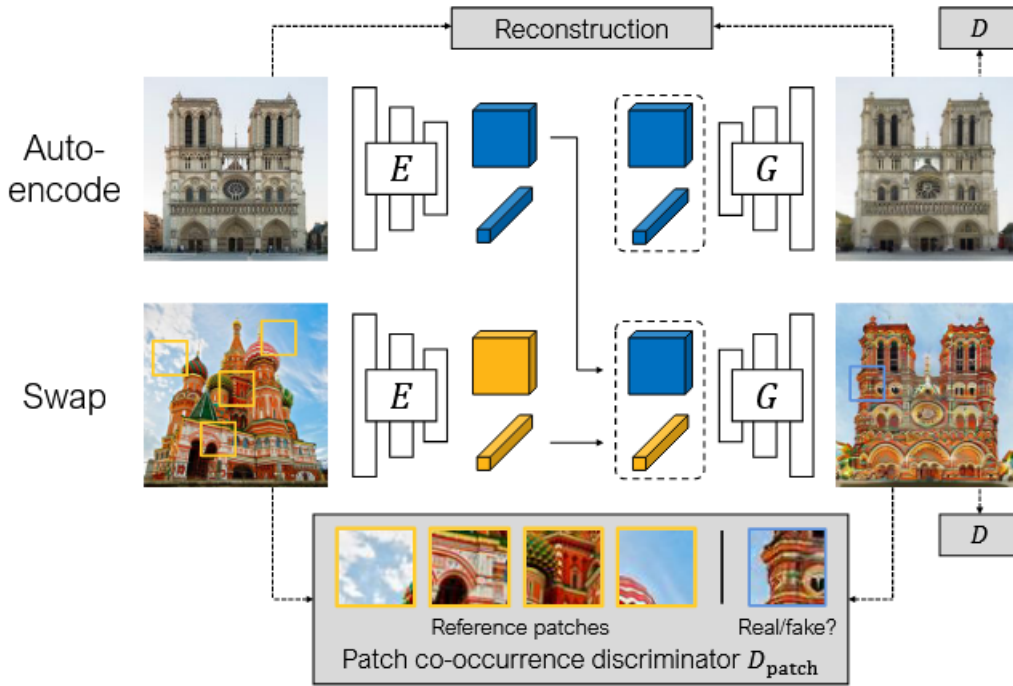


FIGURE 2.3 Architecture du SAE tirée de Park et al. [3]

de l'art en classification [81] comme en segmentation [82] d'images. Cependant, les différences marquantes de contrastes, textures, luminosités et contenus des images médicales par rapport aux images naturelles demandent la conception de méthodes adaptées à ce domaine [83]. Les images de ces ensembles d'entraînement sont souvent acquises suivant un protocole strict avec une seule modalité, engendrant un haut niveau de correspondances entre celles-ci.

### 2.2.5 En imagerie thermique

Analogiquement aux images médicales, les images thermiques diffèrent des images RVB à trois canaux sur de nombreux aspects. Les variations de contrastes, de luminosité et l'utilisation de pseudo-coloriage dans celles-ci rendent difficile l'utilisation directe de systèmes pré-entraînés sur des images dans le visible [84]. Cependant, elles ont l'avantage de révéler les structures pourtant invisibles à l'œil nu, que ce soit pour révéler une augmentation de température du corps [23] ou encore détecter les obstacles en faible condition de visibilité pour les voitures autonomes [85]. L'apprentissage auto-supervisé devient donc rapidement une solution incontournable pour tirer profit des données non étiquetées à disposition. Il est cependant courant de combiner la modalité thermique avec son image RVB correspondante.



En entraînant un réseau combinant ces deux types d’images, celui-ci peut tirer profit des avantages de chaque modalité, permettant une augmentation importante des performances par rapport à un modèle entraîné que sur l’une d’entre elles [86, 87]. Or, lorsque le caractère sensible ou identificatoire des données est un enjeu, notamment en imagerie médicale et dans le domaine visible, il peut être préférable de n’utiliser que la modalité thermique [88]. Toujours dans ce contexte, où l’on désire par exemple évaluer l’augmentation de température localisée suite à une lésion, l’introduction de la modalité RVB introduirait des biais dans l’ensemble de données [89]. La couleur de peau ou la présence de marques visibles (tatouages, cicatrices...) pourrait facilement causer une sous-représentation pour certains patients, provoquant une baisse de performance dans leur évaluation par le modèle [90]. L’imagerie thermique n’étant pas affectée par la couleur de peau, il devient plus avantageux de n’utiliser qu’une seule modalité, permettant une meilleure standardisation du jeu de données et évitant les biais [91].

## 2.3 Les techniques de détection d’anomalies en vision par ordinateur

### 2.3.1 La détection d’anomalies en vision par ordinateur

**Intuition :** La détection d’échantillons aberrants est, depuis plusieurs années, un domaine en vogue, trouvant de nombreux cas applicatifs en vidéosurveillance [92], en cybersécurité [93], en production [92] ou encore en médecine [94]. Une caméra de surveillance, par exemple, enregistrera de nombreuses images où aucun événement d’intérêt ne se produit. Or, si un vol venait à être enregistré, il pourrait être bénéfique qu’un système automatique arrive à détecter l’événement anormal. Dans cet exemple, le nombre d’images normales est alors très élevé pour une petite quantité d’images contenant l’anomalie, rendant impossible l’entraînement complètement supervisé d’un modèle d’apprentissage. Par la nature du problème, plusieurs défis sont présents dans le domaine de la détection d’anomalies par apprentissage profond, tels que la rareté des instances anormales, la diversité et complexité des anomalies et la difficulté à expliquer les résultats des modèles de détection. Ainsi, Pang et al. [95] catégorisent le problème en trois paradigmes : l’extraction d’attributs profonds, l’apprentissage de la normalité et l’apprentissage du score d’anomalie bout-à-bout.

**Extraction des attributs profonds :** Pour ce premier paradigme, on assume que les attributs à faible dimensionnalité conservent les caractéristiques représentant les anomalies. À titre d’exemple, on pourrait entraîner un auto-encodeur convolutif à extraire des représentations sur les images normales et anormales, puis utiliser celles-ci pour entraîner un classificateur sur ces attributs [96]. Cette technique nécessite toutefois une bonne quantité d’images anormales afin de permettre l’extraction efficace des attributs de celles-ci tout en

équilibrant l'ensemble d'entraînement pour le classificateur sur ces attributs.

**Apprentissage de la normalité** : L'apprentissage de la normalité est cependant plus robuste au déséquilibre trop souvent présent en détection d'anomalies. Des représentations sont extraites sur les images normales seulement, assumant que les échantillons anormaux auront des représentations capturant les régularités de l'image, mais ignorant les caractéristiques plus rares [97]. Cette technique est popularisée par les auto-encodeurs qui, entraînés à reconstruire des images normales, seront inaptes à reconstruire adéquatement une image comportant des anomalies. Les images anormales recomposées auront alors une plus grande erreur de reconstruction comparativement à leur version originale, permettant une discrimination des échantillons. De nombreuses techniques s'appuient sur ce principe, notamment les auto-encodeurs variationnels [98], les réseaux antagonistes génératifs [99] et, plus récemment, les modèles de diffusion [100]. On y compte aussi l'apprentissage d'attributs conçus pour la détection de certains types d'anomalies, où l'on retrouve des méthodes comme *One-class SVM* ou encore *Isolation Forest*. Dans les deux cas, les attributs sont appris afin de séparer de manière optimale les échantillons hors de la distribution tout en regroupant autant que possible les échantillons normaux présents en plus grand nombre.

**L'apprentissage bout-à-bout** : Enfin, dans le paradigme de l'apprentissage bout-à-bout du score d'anomalie, on cherche à apprendre un score scalaire décrivant l'anormalité des échantillons. Ainsi, on pourrait apprendre à un réseau régresseur à estimer l'éloignement dans l'espace latent entre des paires d'images normales et anormales grâce à des fonctions de perte conçues à cet effet [101]. Ces méthodes ont alors comme objectif d'estimer le score d'anomalie des images tout en apprenant leurs caractéristiques pour toutes classes confondues. Or, encore une fois, ces techniques assument une quantité suffisante d'échantillons anormaux.

### 2.3.2 En imagerie médicale

**Techniques de base** : La détection d'anomalies devient particulièrement pertinente dans le contexte de l'imagerie médicale. Dans ce domaine sous lequel un volume important de données ou d'annotations n'est que rarement disponible, il devient important de développer des algorithmes spécialisés capables de faire face à ces défis. De plus, beaucoup d'anomalies médicales apparaissent de manière très subtile sur les images, rendant très courte la distance entre le domaine des images normales et anormales. L'utilisation des auto-encodeurs en imagerie médicale n'est pas chose nouvelle. En entraînant un modèle génératif à reconstruire des images normales, celui-ci sera inapte à reconstruire adéquatement une image comportant une anomalie, même subtile [102]. Ainsi, une simple comparaison, par exemple une erreur absolue moyenne sur les pixels entre l'image en entrée et sa reconstruction, permet d'établir

un seuil discriminant. Ces techniques par reconstruction permettent alors de segmenter de manière non supervisée les anomalies, par une simple comparaison entre les images [103], les rendant plus interprétables auprès des praticiens [97]. Outre les auto-encodeurs, les GANs, suivant le même principe de détection, ont su particulièrement se démarquer dans le domaine [104–107], produisant des reconstructions de meilleure qualité que les VAEs ou auto-encodeurs simples [108]. Plus récemment, les modèles de diffusion ont érigé leur place dans la détection d’anomalies par reconstruction, produisant des images de qualité en évitant les instabilités typiques de l’entraînement de réseaux antagonistes génératifs [99, 109].

**Importance des textures** : Le concept de textures en détection d’anomalies médicales reste encore une fois pertinent. En effet, plusieurs pathologies visibles par une modalité d’imagerie quelconque peuvent se traduire par des déviations dans la texture de l’image, sans nécessairement introduire des changements structurels [4, 110]. Suivant cette logique, Zhou et al. ont introduit leur LSAE (*Lung-Swapping Autoencoder*) [4]. Basé sur l’architecture du *Swapping Autoencoder* de Park et al. [3], l’encodeur est entraîné à interchanger les éléments de structure entre plusieurs paires d’images aléatoires, causant le désenchevêtrement des représentations. Afin d’éviter les distorsions dans les reconstructions suivant l’échange, celles-ci ne sont effectuées que dans la région des poumons où les pathologies sont visibles. Cela évite, par exemple, les déformations au niveau de la structure osseuse visible sur les radiographies. La figure 2.4 illustre l’architecture du LSAE. Ainsi, une tête de classification peut être entraînée sur cet espace de textures. Testé sur CXR-14 [111] et COVOC [112] pour la classification de pathologies dans les radiographies du thorax, leur modèle obtient des performances comparables ou supérieures à l’état de l’art en plus de nécessiter moins de paramètres à l’inférence comparativement à ces méthodes. Ces résultats mettent en lumière l’importance de la texture dans la détection d’anomalies pour certaines modalités d’imagerie médicale et l’expressivité de l’espace latent construit à l’entraînement.

**Évaluation de la symétrie comme critère d’anomalie** : Un autre élément à explorer en détection d’anomalies dans les images médicales est l’utilisation de la symétrie bilatérale comme élément discriminant. En effet, il est reconnu qu’une perte de symétrie entre deux organes ou membres peut indiquer la présence d’une pathologie [113]. Par exemple, Choi et al. [114] ont développé une méthode de détection de la mastoïdite bilatérale en évaluant les similarités des cartes d’activation d’un réseau convolutif entre les vues antérieures et postérieures du mastoïde. L’utilisation de la symétrie bilatérale est plus communément utilisée dans la détection du cancer du sein, où elle permet de relever certains indicateurs pathologiques dans plusieurs modalités d’imagerie [115–118]. Shimokawa et al. [117] ont introduit, en 2022, une méthode d’entraînement par laquelle un réseau convolutif Xception [119] apprend à identifier la présence de cancer du sein unilatéral en différenciant les régions saines ayant une

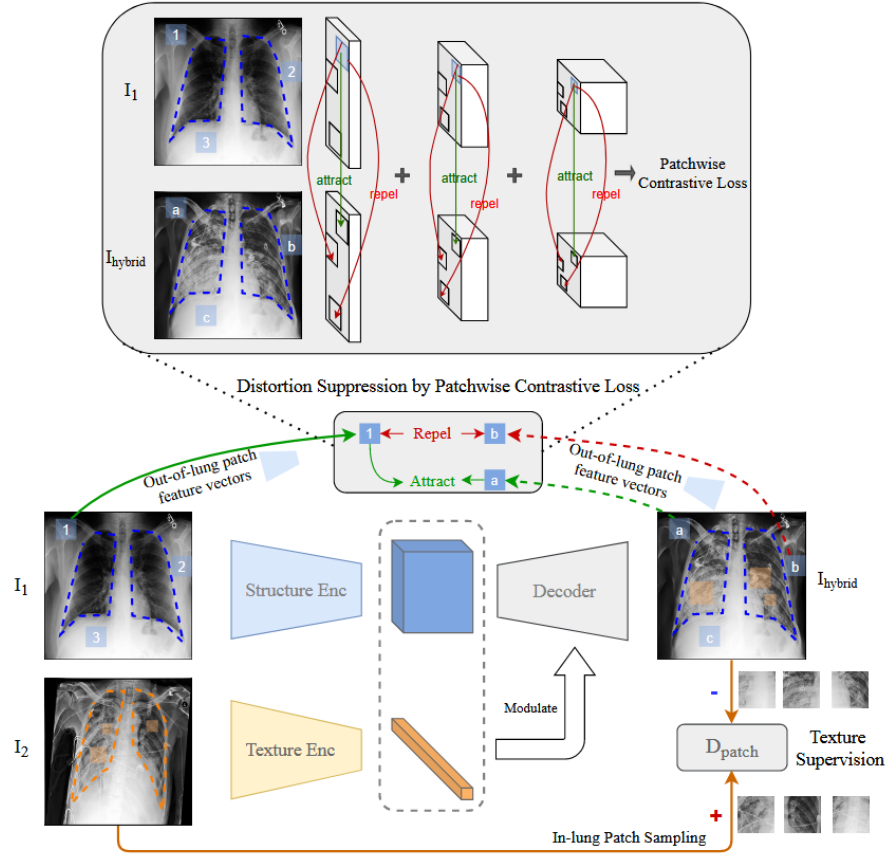


FIGURE 2.4 Architecture du LSAE tirée de [4]. Celle-ci est identique au SAE [3] excepté l'ajout d'une fonction de perte contrastive utilisée pour forcer le transfert de textures dans la région des poumons définie par un masque.

tumeur entre deux vues opposées d'une tomosynthèse mammaire numérique. Cette technique a l'avantage d'être moins sensible aux différentes densités mammaires pour chaque patiente dans l'ensemble de données, surpassant les résultats obtenus avec un classificateur Xception en réduisant le nombre de faux négatifs. Dans le domaine thermique, Dey et al. [118] ont conçu un algorithme d'extraction de caractéristiques à la main permettant de comparer les textures entre les seins gauches et droits. Malgré l'absence d'apprentissage machine dans leur travail, la méthode développée surpasse l'état de l'art sur le jeu de données DMR-IR [1], donnant une précision, sensibilité et spécificité de 96.08%, 100% et 93.57% respectivement.

### CHAPITRE 3 RATIONNELLE ET OBJECTIFS DU PROJET

À la lumière de la revue des connaissances présentée dans le chapitre précédent, la rationnelle du projet se résume dans les points suivants :

1. L'imagerie thermique est une modalité de diagnostic attrayante par son caractère portatif, non invasif, non ionisant et sans contact. Il serait envisageable d'utiliser cette modalité chez les patients, dans le confort de leur foyer, pour surveiller des changements thermiques qui pourraient signifier la présence des complications post-opératoires.
2. Des travaux antérieurs démontrent la possibilité d'utiliser le gradient de température entre les jambes gauches et droites comme critère servant à détecter les inflammations et les infections dans l'année suivant la chirurgie de remplacement. Toutefois, ces mesures sont réalisées au moyen d'un thermomètre infrarouge et non d'une image thermique, ne permettant pas une cartographie de l'information de chaleur.
3. Parmi les descripteurs d'images, ceux décrivant la texture semblent les plus prometteurs pour la détection du cancer du sein à partir d'images thermiques.
4. Plusieurs pathologies peuvent se traduire par des déviations dans la texture de l'image, sans nécessairement introduire des changements structurels. En ignorant les informations structurelles invariantes à la présence d'anomalies dans les radiographies de poumons, il devient plus facile pour un modèle comme LSAE [4] de différencier les pathologies.
5. L'utilisation de la symétrie bilatérale est efficace pour la détection du cancer du sein. La comparaison permet de définir les similitudes pour les deux côtés, établissant un référentiel utile à la détection d'anomalies.
6. Les cas de complications après chirurgie de remplacement de hanche ou de genou sont assez rares. Aborder ce problème sous l'angle de la détection d'anomalie par apprentissage de la normalité permet de pallier avec robustesse au déséquilibre entre données normales et anormales.
7. Les jeux de données de taille très réduite en imagerie thermique clinique et le manque d'étiquettes balancées justifient l'utilisation de l'apprentissage auto-supervisé et l'intégration de méthodes d'augmentation.

La littérature n'énonce en ce moment aucune méthode d'apprentissage des représentations pour la détection d'anomalies dans le suivi des arthroplasties. On en déduit qu'une méthode de détection d'anomalies dans le suivi de ces chirurgies des membres inférieurs, basée sur

l'imagerie thermique et permettant une comparaison bilatérale des attributs de texture, mériterait d'être explorée. Notons également l'absence de jeux d'images thermiques des membres inférieurs en post-opératoire.

On peut ainsi poser l'hypothèse de recherche et l'objectif général du projet :

**Hypothèse principale du projet** : La distance entre les vecteurs latents de texture entre les thermographies des membres opéré et controlatéral permet de décrire efficacement les différences de température pouvant signaler une anomalie postopératoire.

**Objectif général du projet** : Concevoir et valider un algorithme de détection d'anomalies dans les images thermiques des membres inférieurs pour le suivi des patients ayant reçu une ATG ou ATH

Cet objectif se décompose en 4 sous-objectifs :

**SO1** : Collecter un ensemble d'images thermiques de membres inférieurs de patients ayant reçu une ATG ou ATH, entre six semaines et un an après leur chirurgie,

**SO2** : Concevoir un algorithme de détection d'anomalies dans les images thermiques des membres inférieurs pour la détection de complications post-chirurgicales pour l'ATG et l'ATH,

**SO3** : Évaluer l'algorithme conçu au SO2 sur l'ensemble de données découlant de SO1 pour détecter les anomalies suivant une chirurgie de remplacement de genou ou de hanche,

**SO4** : Adapter et évaluer l'algorithme conçu au SO2 sur l'ensemble de données DMR-IR (accessible au public) pour la détection du cancer du sein pour une comparaison directe avec des méthodes existantes.

Le chapitre 4 adresse le SO1 et présente le protocole pour la collecte d'images thermiques des membres inférieurs après chirurgie orthopédique. Quant au chapitre 5, il est présenté sous la forme d'un article soumis à IEEE Transactions on BioMedical Engineering (TBME), et décrit la méthode développée en réponse à SO2 ainsi que son évaluation sur deux jeux d'images thermiques : le premier découle de SO1, il est privé et porte sur les membres inférieurs après chirurgie de remplacement (SO3), et le second est public et porte sur la détection du cancer du sein (SO4).

## CHAPITRE 4 DÉVELOPPEMENT D'UNE BASE DE DONNÉES POUR LA DÉTECTION D'ANOMALIES SUIVANT UNE CHIRURGIE ORTHOPÉDIQUE

Ce chapitre traite de la conception d'une base de données thermographiques de patients ayant reçu un ATH ou ATG entre six semaines et un an avant l'acquisition. Baptisé OrthoPOT (*Orthopedics Post-operative Thermograms*), celui-ci regroupe les images thermiques de 65 patients et patientes ainsi que des données cliniques pertinentes à l'interprétation des images.

### 4.1 Matériel

#### 4.1.1 Équipement

Afin de concevoir un ensemble de données de bonne qualité, la caméra thermique haute résolution FLIR T1020 a été utilisée pour la capture d'images infrarouges de patients opérés. Le tableau 4.1 énonce les caractéristiques de la caméra employée dans ce projet. Cette caméra

TABLEAU 4.1 Caractéristiques de la caméra thermique FLIR T1020

Dimensions	167 × 204 × 188 mm - 2100g
Résolution	1024 × 768 (px)
Champ de vision	45° × 34°
Plage de température	[-40, 2000] °C

se veut portable, simple d'utilisation tout en capturant des images en haute définition disponibles sous forme de matrice de températures (.csv) ou d'images simples (.png). Bien que cette dernière puisse aussi prendre des images dans le domaine visible, celles-ci n'ont pas été capturées dans le cadre de ce projet afin de conserver la vie privée des patients et patientes. Enfin, cette caméra peut facilement être montée sur un trépied, offrant une meilleure stabilité et une plus grande uniformité dans la capture d'images. La figure 4.1 montre la caméra installée sur son trépied.

#### 4.1.2 Environnement d'acquisition

L'ensemble de données a été conçu dans l'aile d'orthopédie de l'Hôpital Maisonneuve-Rosemont à Montréal dans la province de Québec au Canada. Une salle d'examen a été assignée afin de permettre l'accueil des participants et la prise d'images à lumière et température ambiante.



FIGURE 4.1 Montage de la caméra thermique FLIR T1020 sur son trépied.

Afin de trouver les patients éligibles, les accès aux rendez-vous planifiés et aux dossiers de patients ont été partagés. Ainsi, à l'aide d'un poste de travail attitré, il est possible de consulter les rendez-vous à venir dans l'aile d'orthopédie et correspondre le numéro de patient avec son dossier médical indiquant son historique chirurgical et ses données personnelles. Les numéros de patients éligibles à l'étude sont ainsi enregistrés avec leur date de rendez-vous, permettant leur éventuel recrutement. Les données sont de plus désidentifiées avec un code unique attribué à chaque patient.

## 4.2 Méthode

### 4.2.1 Recrutement des patients

Le recrutement de patients se fait en deux étapes. D'abord, ceux-ci sont approchés avant leur rendez-vous en leur présentant l'objectif de la recherche. Ensuite, un formulaire d'information et de consentement expliquant en détails le projet, ses retombées et le protocole d'acquisition sont présentés. Afin d'obtenir un consentement éclairé, environ 15 minutes sont laissées, si



nécessaires, pour la lecture du document. Par la suite, le patient se veut libre d'accepter et signer ou refuser. Afin qu'un patient puisse être recruté, celui-ci doit répondre à 4 critères d'éligibilité :

1. Avoir reçu une ATG ou ATH unilatérale entre 6 semaines et 1 an avant le jour de l'acquisition,
2. Avoir plus de 18 ans,
3. Être suffisamment mobile pour se tenir debout,
4. Être en mesure de signer le formulaire de consentement de manière éclairée.

## 4.2.2 Protocole d'acquisition

### Calibration de la caméra

Afin de permettre la capture d'une image nette, le patient demeurera immobile pour permettre l'ajustement de la longueur focale et du champ de vision de la caméra. De plus, celle-ci sera posée sur un trépied. Le paramètre d'émissivité  $\epsilon$  est initialisé à  $\epsilon = 0.98$  correspondant à celle du corps humain [120]. Afin d'isoler le mieux possible le patient du fond sur l'image thermique, une plage de température entre 26°C et 39°C a été choisie empiriquement. Les valeurs hors de cette plage sont ainsi mises à 0 et 255 respectivement sur l'image thermique apparaissant à l'écran, permettant une meilleure mise au point et une visualisation plus nette.

### Prise d'images

Une fois le formulaire de consentement signé par le patient, l'acquisition d'images thermiques peut débuter. Celui-ci est amené dans la salle attitrée du service d'orthopédie où le protocole d'acquisition est ré-expliqué. Un total de 10 minutes tout au plus sont nécessaires pour la capture des images, peu importe le type de chirurgie. Durant cette procédure, un total de quatre images montrant les membres inférieurs, soit une de chaque côté, sont enregistrés. Afin de capturer la plus grande superficie possible, la caméra est placée à la verticale. Le patient est donc demandé de mettre à découvert la zone opérée durant chaque prise, dévoilant aussi le côté opposé non-opéré servant de référence. Celui-ci se tient debout et immobile à un mètre de la caméra. Sur chacune des images, les deux jambes sont visibles et à découvert. La figure 4.2 montre un exemple d'images prises pour un patient ayant reçu une ATG, tandis que la figure 4.3 montre les images thermiques d'une ATH.

En raison de la différence de localisation des régions d'intérêt pour les deux types de chirurgies, les images diffèrent légèrement entre ces cas. En effet, pour une ATG, les images sont

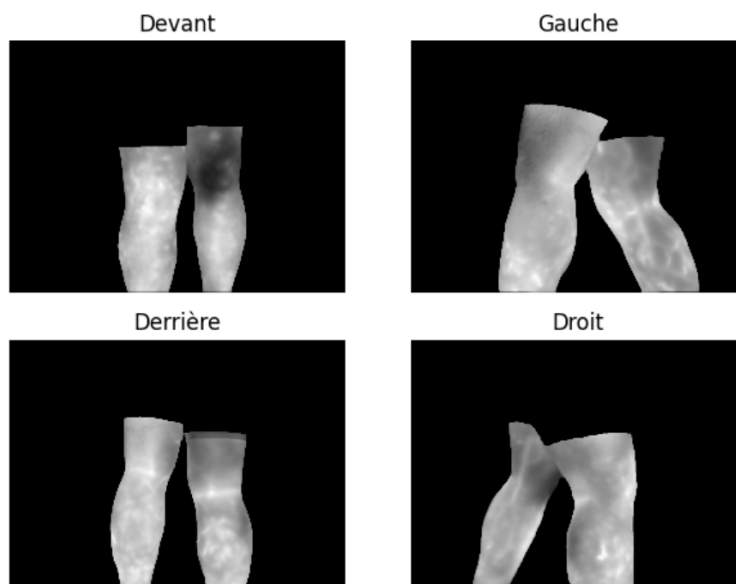


FIGURE 4.2 Exemple d'images thermiques prises sur quatre vues pour un patient ayant reçu une ATG ou genou droit.

prises d'environ 20 centimètres au-dessus de genou jusqu'à la cheville. Pour l'ATH, celles-ci montrent la hanche jusqu'à la cheville.

### Données supplémentaires recueillies

Afin de contextualiser les images thermiques des patients, plusieurs données cliniques sont conservées suivant le consentement du patient. On y compte :

1. L'âge (années),
2. Le type de chirurgie (ATG ou ATH),
3. Le côté opéré (gauche ou droit),
4. Le nombre de semaines entre l'opération et la participation à l'acquisition (sem.),
5. Le poids (kg),
6. La grandeur (m),
7. Les complications post-chirurgicales, si applicable (aucune, infection, thrombose...),
8. Le numéro de téléphone (optionnel),
9. L'adresse courriel (optionnel).

Ces deux derniers attributs sont utilisés afin d'effectuer un suivi six semaines suivant la participation. En effet, puisqu'il serait impossible d'effectuer une étude longitudinale auprès des patients par contrainte de temps, ceux-ci reçoivent automatiquement un questionnaire

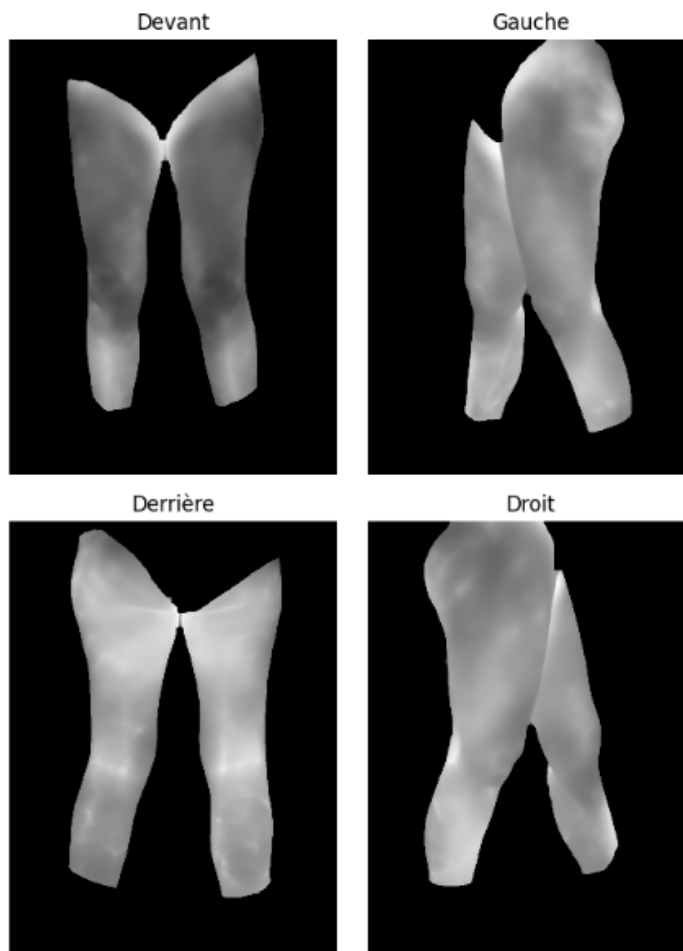


FIGURE 4.3 Exemple d'images thermiques prises sur quatre vues pour un patient ayant reçu une ATH à la hanche gauche.

électronique suivant la période prescrite à l'aide du logiciel *RedCap* fourni par l'Hôpital Maisonneuve-Rosemont. Dans celui-ci, les patients répondent à trois questions visant à déterminer si de nouveaux inconforts ou problèmes se sont développés après la visite. Cela permet ainsi de considérer le patient comme possible cas anormal.

### 4.2.3 Traitement des données

Dans son ensemble, un total de 66 patients et patientes ont pu être recrutés sur une période de 5 mois entre janvier 2024 et mai 2024, soient 30 hommes et 36 femmes. Une patiente a été enlevée du jeu de données puisque son opération a eu lieu 82 semaines avant la participation, mais a été consultée pour des douleurs au niveau du site chirurgical, représentant un cas d'étude néanmoins intéressant. Au total, 244 images ont été enregistrées, représentant en

moyenne 3.75 images par patient. En effet, vu le petit nombre de patients éligibles, les captures de patients ne pouvant se tenir debout ont été conservées dans le cas des ATG, où l'hyperthermie est visible simplement en exposant les genoux en position assise.

De ces images, des masques ont été manuellement créés avec le logiciel *Label Studio* [121] afin de définir les régions à découvert des jambes gauches et droites. Ainsi, il devient possible d'extraire les deux membres dans chaque image thermique tout en ignorant les pièces de vêtement induisant une grande variabilité dans les images.

Enfin, l'indice de masse corporelle (IMC) est calculé depuis le poids en kilogrammes et la grandeur en mètres du patient. Cet indice se mesure en  $\frac{kg}{m^2}$ . La formule utilisée est :

$$IMC = \frac{masse}{grandeur^2} \quad (4.1)$$

#### 4.2.4 Analyses des données

##### Démographie

Afin de comprendre la distribution de population, quelques analyses sont effectuées en considérant les données cliniques des patients. On affiche d'abord la distribution de l'âge des patients recrutés avec la figure 4.4.

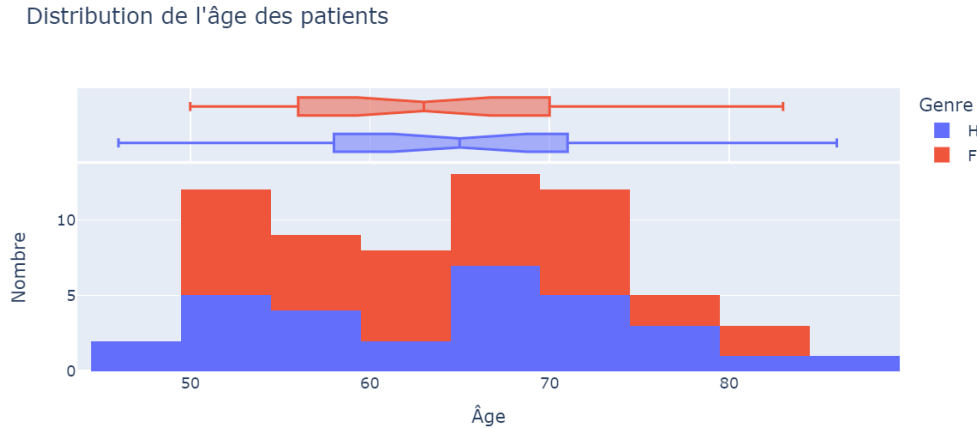


FIGURE 4.4 Distribution de l'âge des patients par genre.

On remarque un bon équilibre dans la distribution de l'âge dans les deux genres étudiés. En effet, l'âge moyen des hommes est de 64.07 ans pour 63.91 ans chez les femmes. De même manière, la distribution des types de chirurgie est montrée à la figure 4.5.

On compte un total de 41 ATG et 24 ATH. Dans ce premier cas, 23 femmes et 18 hommes

Distribution des types de chirurgie

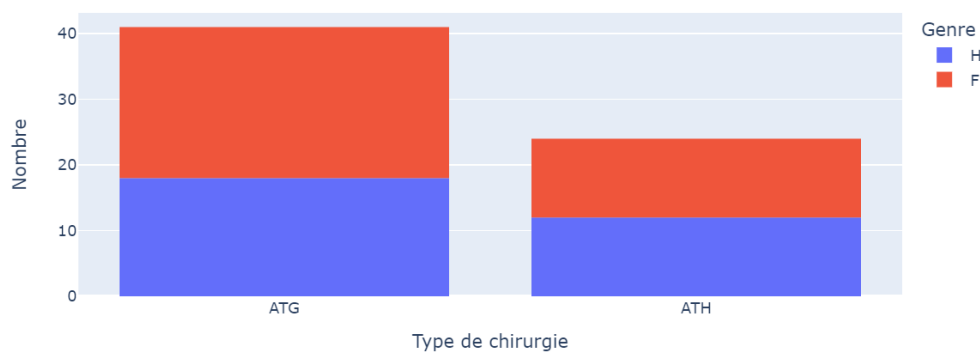


FIGURE 4.5 Distribution des types de chirurgie par genre.

y sont représentés. Pour ce qui est des ATH, la proportion y est un peu plus équilibrée avec exactement 12 hommes et 12 femmes.

On peut de plus visualiser les distributions de masse corporelle et de grandeur chez les hommes et les femmes, ainsi que leur relation telle que montrée par la figure 4.6.

Relation entre la grandeur et le poids

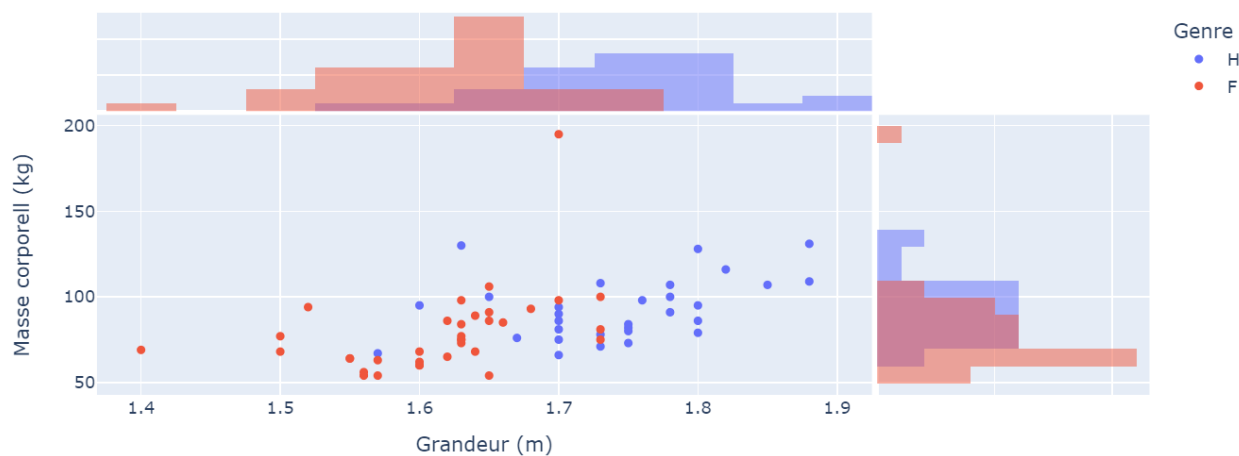


FIGURE 4.6 Relation entre la grandeur et le poids des patients par genre.

Tel qu'espéré, le graphique montre une tendance linéaire claire entre la grandeur du patient et sa masse, peu importe le genre. On remarque aussi que les hommes sont en moyenne de plus grande taille et plus lourds que les femmes. De ces données, il devient facile de calculer

l'IMC des patients et patientes et évaluer leur distribution à l'aide de l'histogramme en figure 4.7.

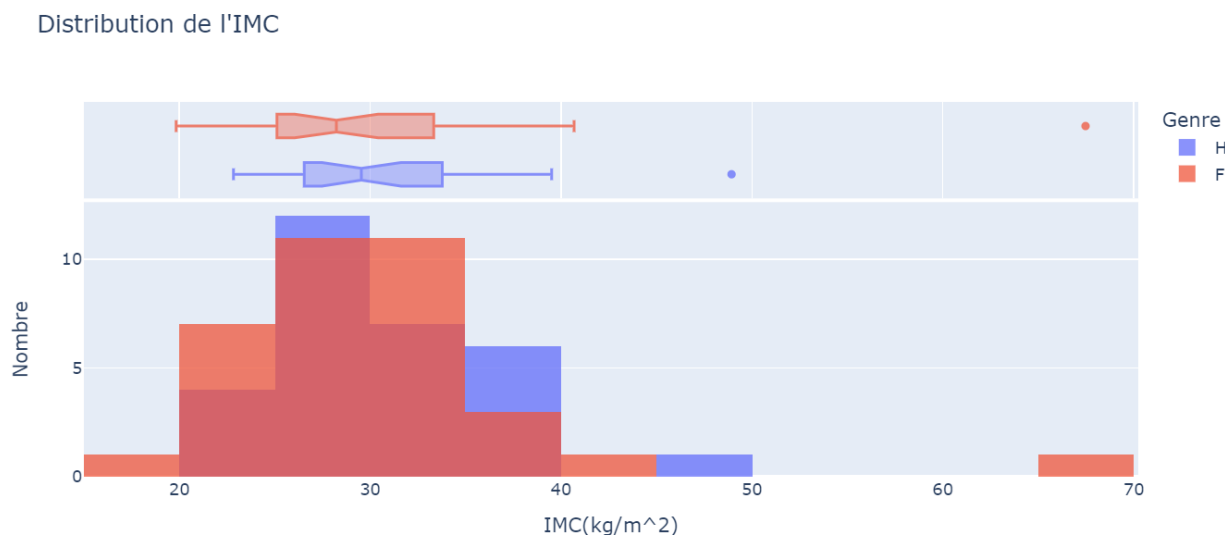


FIGURE 4.7 Distribution de l'IMC par genre.

Les indices de masse corporelle chez les hommes et les femmes suivent une distribution assez similaire. Ces premiers ont un IMC moyen de 30.57 pour 30.16 chez le sexe opposé.

En ce qui a trait aux anomalies, on peut comprendre leur distribution au travers du nombre de semaines depuis leur chirurgie à l'aide de l'histogramme présenté à la figure 4.8.

Les distributions d'anomalies sur le nombre de semaines depuis la chirurgie suivent approximativement celle des échantillons normaux. On recense deux cas d'infection après 7 et 14 semaines, un cas de douleur intense reporté par le patient après cinq semaines, une bursite trochantérienne à la 24<sup>e</sup> semaine et un cas de prothèse non cimentée nécessitant une révision à la 32<sup>e</sup> semaine. On peut effectuer une analyse sommaire de ces échantillons en calculant les différences de moyennes de température entre le côté opéré et non-opéré pour les mettre en relation avec le nombre de semaines depuis la chirurgie. La figure 4.9 permet de comprendre comment les différences d'intensités globales entre les deux membres s'illustrent en fonction du temps.

Dans les cas des ATG, on observe une légère diminution du différentiel de température au fil du temps. Or, ces différences de moyennes entre le côté opéré et controlatéral ne décrivent pas la distribution de température et la localisation de l'hyperthermie. On observe une tendance similaire avec les ATH, mais un biais important est soulevé. En effet, les deux anomalies sont trouvées après 24 et 32 semaines, représentant des valeurs aberrantes faciles à détecter dans

Distribution des semaines depuis la chirurgie et des anomalies

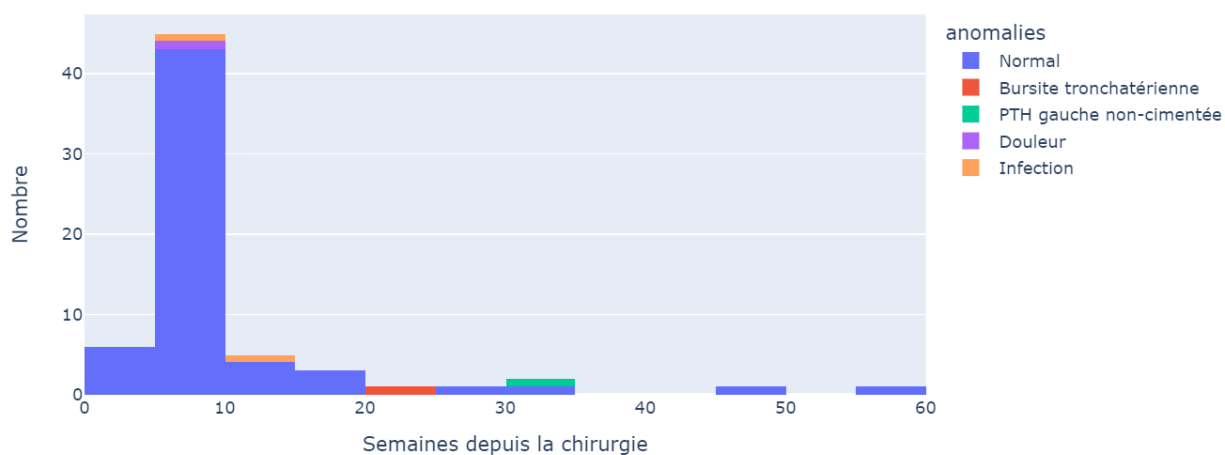


FIGURE 4.8 Distribution du nombre de semaines entre l'acquisition et la chirurgie selon le type d'anomalie rencontré. Celles-ci ont été diagnostiquées, sauf le cas de douleur ayant été reporté par le patient.

l'utilisation de données thermiques ou cliniques autres. Ce biais important est ainsi à prendre en considération dans l'évaluation du système de détection d'anomalies.

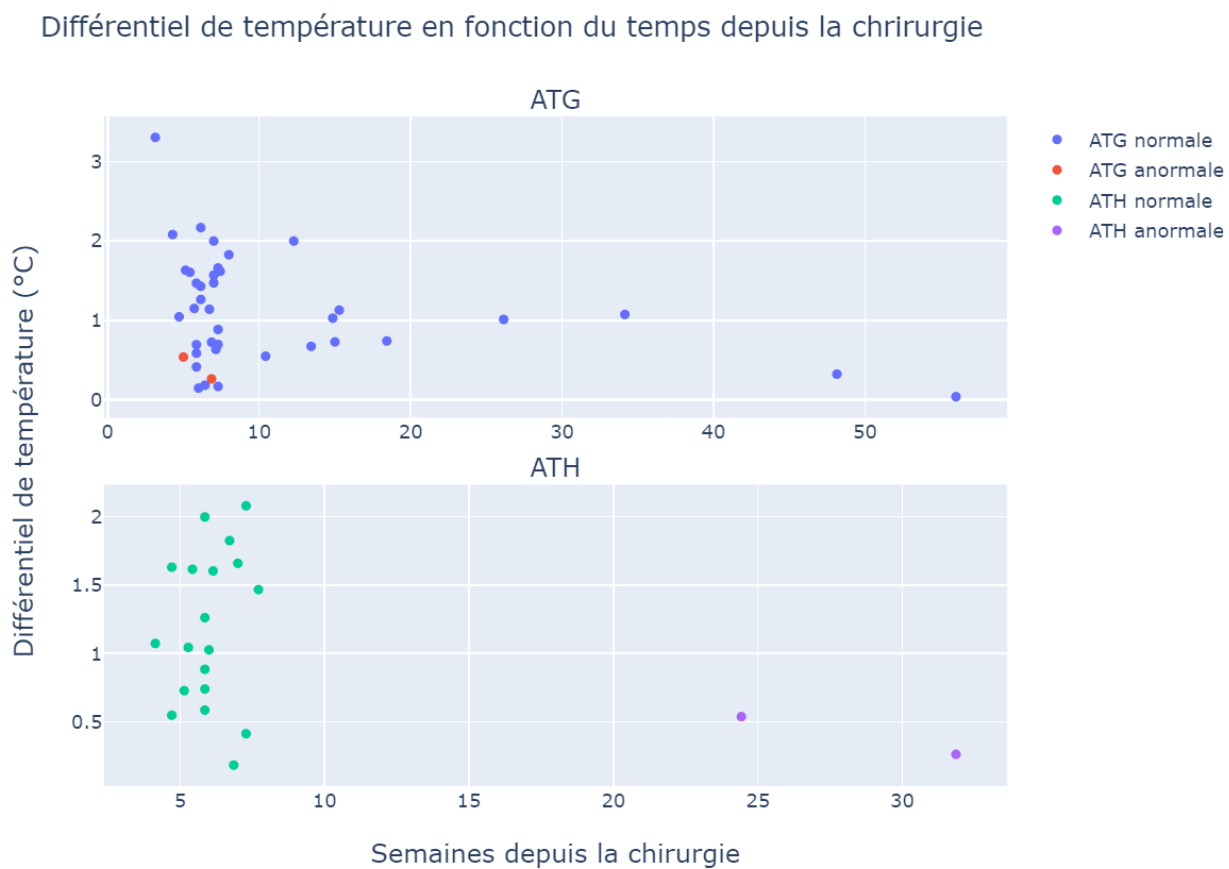


FIGURE 4.9 Différentiels de température moyens entre le côté opéré et controlatéral dans les ATG et ATH.



## CHAPITRE 5    ARTICLE 1 TADA-SAE: EXPLOITING BILATERAL SYMMETRY IN LEARNED TEXTURE REPRESENTATIONS FOR THERMAL MEDICAL IMAGING

Ce chapitre présente, sous la forme d’un article soumis le 18 novembre 2024 à la revue *Transactions on Biomedical Engineering* (TBME), la méthode développée dans le cadre de cette maîtrise. L’élaboration de la méthode, les expérimentations, l’analyse, la recherche bibliographique ainsi que la rédaction de l’article ont été complétées dans le cadre de ce mémoire. L’annexe A explique en détail la méthodologie utilisée pour évaluer la sensibilité aux différentiels de température sur DMR-IR. L’annexe B définit une expérience supplémentaire utilisant les capacités de génération du modèle pour la détection d’anomalies sur DMR-IR. Enfin, l’annexe C montre les capacités de détection d’anomalies de TADA-SAE par type de chirurgie.

**Auteurs:** Étienne Lescarbeault, Manuela Kunz, Mina Waheed Wahba Morcos, Lama Séoud

### 5.1 Abstract

*Objective:* The skin surface temperature is a well known indicator of a patient’s health. Thermal imaging is an effective tool for visualizing its distribution of intensities, opening new possibilities of diagnosis. Moreover, deviations from bilateral symmetry in temperature between opposite body parts can be used to detect pathologies. With the ever-growing interest in computer-aided diagnosis tools and machine learning techniques, efficient modeling of these distributions is desirable. *Methods:* By learning texture representations and comparing their asymmetries, we hypothesize that texture latent vectors between opposite body parts imaged by thermography allow to efficiently describe the temperature differences that can signal an anomaly in the thermal domain. Our TADA-SAE (Texture Anomaly Detection from Asymmetries using Swapping Autoencoder) model is tested on two datasets: the publicly available DMR-IR (Database for Mastology Research with Infrared Image) dataset for breast cancer detection; and the private OrthoPOT (Orthopedics Post-Operative Thermograms) dataset for post-operative complications following total arthroplasty on the lower limbs. *Results:* Our approach achieves a competitive 0.988 AUROC on DMR-IR while using fewer parameters than previous methods. Furthermore, an isolation forest trained on symmetry features learned by TADA-SAE combined with clinical data achieves 0.842 AUROC on OrthoPOT. *Conclusion:* Learned texture features and their asymmetries are relevant for medical anomaly detection in the thermal domain. *Significance:* Our method is the first

to use learned texture features and asymmetry for anomaly detection in medical thermal imaging, and shows promising results which could transfer to other tasks in the medical field where bilateral symmetry can be leveraged.

## 5.2 Introduction

The temperature of the human body is a well known indicator of health [21]. A significant increase, for instance, may indicate that the immune system is fighting a virus or an infection. It has been used for centuries as a clinical diagnostic measurement. With the development of thermal imaging and infrared sensors, surface skin temperature has become more descriptive, providing a spatial distribution of temperatures. With their high portability, ease of use, and contactless and non-ionizing properties, thermal cameras are taking their place in the medical field with numerous applications by offering a cheaper alternative to standard medical imaging modalities [23]. Lahiri et al. [18] provide an extensive review of thermographic medical applications, including the diagnosis of breast cancer, diabetes-related neuropathy and peripheral vascular disorders, and the detection of problems in dermatology, kidney transplants and fever screening. Our paper focuses on two specific applications: breast cancer diagnosis and detection of complications after total arthroplasty of the lower limbs.

While mammography is the gold standard modality for breast cancer diagnosis, it requires compressing and flattening the breast before capturing the radiography, which induces discomfort for the patients [26]. Thermal imaging offers a promising alternative, being non-irradiating and without necessitating physical contact with the patient. It has been widely explored as an alternative imaging modality for breast cancer diagnosis. The availability of the *DMR-IR* (Database for Mastology Research with Infrared Image) public dataset [1] has allowed many research groups to propose thermal image analysis in that field [28, 29, 31–33, 118, 122, 123].

In the field of orthopedics, previous works have addressed the usefulness of thermal imaging in patient monitoring after total knee arthroplasty (TKA) or total hip arthroplasty (THA) [15, 38]. The post-operative recovery following these procedures is not without risks. A common complication is the development of an infection at the surgery site, affecting between 2% and 3.7% of patients and manifesting up to one year after the intervention [10]. Another possible complication, while rarer, is deep venous thrombosis (DVT) [9]. Thermal imaging has been explored as a diagnosis tool for systemic or prosthetic joint infections after TKA and THA [2] and thrombosis [40]. Due to improved surgical techniques and prophylaxis methods, serious postoperative complications after joint arthroplasty are rare, but they still pose a real risk, especially for older adults undergoing this surgery [124]. Therefore, continuous

monitoring and early detection can improve the long-term outcome for these patients.

While thermal imaging has been considered for both breast cancer diagnosis and detection of arthroplasty post-operative complications, no unified image processing method exists to analyze skin temperature distribution. However, one major and common feature of these two applications is bilateral symmetry. This symmetry can be exploited in terms of *texture*, i.e. the contrasts and intensities in the image. To effectively harness this information, it must be separated from the image’s *structure* information (shape and content). In breast cancer detection, a temperature increase in one breast can signal the presence of a tumor. Being able to compare textures instead of structures is particularly relevant, since most women have some degree of physiological breast shape asymmetry. Some studies even report as many as 70% to 88% [125] of women with such shape asymmetry. Meanwhile, an extended and intense hyperthermia can signal a post-operative complication. Therefore, image descriptors able to efficiently measure temperature differences between symmetric body parts are needed. In this work, we introduce Texture Anomaly Detection from Asymmetries using Swapping Autoencoder (TADA-SAE), a new deep-learning based pipeline to extract and compare texture features in medical thermal images.

The contributions of this paper are the following: (i) We propose new features that are learned in a self-supervised manner to describe the texture information in thermal images. (ii) We evaluate the ability of left-right texture asymmetries to classify healthy versus malignant unilateral breast cancer in the publicly available DMR-IR dataset [1]. (iii) We show how learned texture features can enable anomaly detection following a knee or hip replacement surgery on a proprietary dataset. The code is available through this link: <https://github.com/EtienneLescarbeault/TADA-SAE>.

### 5.3 Related Work

Features extraction from medical thermal images has been well documented, especially in breast cancer detection. Several studies have found that texture features could describe temperature variations over the warmer skin region affected by the underlying breast cancer [29,31–33]. Texture can be described through hand-made features such as GLCM [126] (*Grey-Level Co-occurrence matrices*), HOG (Histogram of Oriented Gradients) [48] and LBP (Local Binary Patterns) [49]. Acharya et al. [28] used co-occurrence and run-length matrices with a support vector machine (SVM) classifier. Later, [123] used representation learning along with hand-crafted texture features for breast cancer classification. Moreover, it was found that texture differences between left and right breasts could also be used [122]. Recently, [118] used hand-crafted texture features and their asymmetries between left and right breasts to

achieve state-of-the-art classification abilities on DMR-IR.

While hand-made texture features are interpretable, it has been shown that data-driven techniques like feature extraction from convolutional neural networks (CNNs) can help extract richer features which are more descriptive for a downstream task like classification [43]. In deep learning, the concept of self-supervised learning harnesses the possibility for neural networks to extract image features without the need for labels. One method using deep learning for texture features extraction is the swapping autoencoder (SAE) [3]. Based on StyleGan2 [75], its encoder learns to disentangle the structure of an image from its texture, allowing the generator to create image hybrids with the content (structure) of one image and the appearance (texture) of another. The structure tensor extracts spatial information of the image, while the texture vector modulates the generator weights. Although it was initially conceived for image manipulation, this architecture found its way into the medical field in the form of the lung swapping autoencoder (LSAE) for chest x-ray pathology classification [4]. The relevance of texture features in medical and thermal imaging as well as continual improvements in the field of self-supervised learning pave the way for more data-driven methods, ultimately leading to better classification algorithms for computer-aided diagnosis.

## 5.4 Methods

In this section, we start by presenting TADA-SAE, a model and training strategy to learn texture features from thermal images (section 5.4.1). Then, we present two medical applications through which TADA-SAE is evaluated: the first one on a public dataset for breast cancer detection (section 5.4.2) and the second one on a private dataset for post-operative anomaly detection in arthroplasty surgeries (section 5.4.3).

### 5.4.1 TADA-SAE: Texture feature learning

To learn texture features from thermal images, we propose TADA-SAE, whose overall architecture is illustrated in Figure 5.1. It is inspired by the SAE [3] and LSAE [4] architectures. From training pairs of thermal images  $I_a$  and  $I_b$ , along with their binary masks, we want to extract disentangled structure and texture vectors. These image pairs, during training, are the left and right body parts from the same patient or two different patients. This is done using a self-supervised texture swapping task between  $I_a$  and  $I_b$ . The model must also create accurate reconstructions and be able to mix the texture and structure latent codes between images to generate hybrid images. The building blocks of the architecture as well as the training strategy are detailed hereafter.

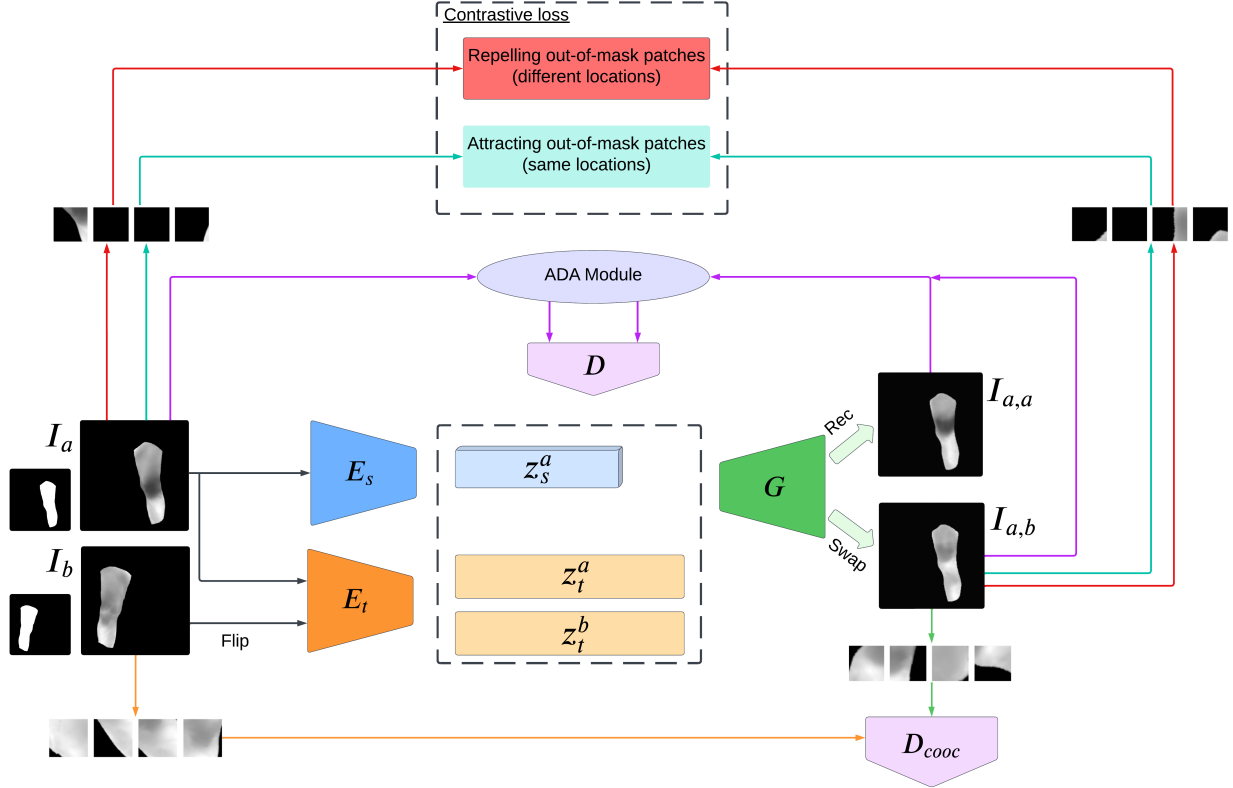


Figure 5.1 Overall architecture of TADA-SAE; training on the private OrthoPOT dataset is illustrated here. The structure and texture latent codes for both legs are used for reconstruction and swapping, while patches are sampled from the swapped image and the original one to ensure texture transfer. The Adaptive Discriminator Augmentations (ADA) module is added over the discriminator  $D$ , which enables training with limited data.

## Encoder

The encoder has a total of 2.6M parameters and has two branches, one encoding the structure and the second encoding textures. The structure branch  $E_s$  comprises four consecutive ResNet [127] blocks, downsampling an input image to produce  $z_s$ , a disentangled structure vector of size  $16 \times 4 \times 4$ , storing the spatial information of the image. The texture encoding branch  $E_t$  is obtained similarly, with the addition of an average pooling layer followed by a dense layer. The output tensor is then flattened to produce  $z_t$ , the disentangled texture vector of size  $1 \times 1024$ . The encoder takes a pair of input images  $I_a$  and  $I_b$ , and for each of these, both branches of the encoder will create latent structure and texture vectors, noted respectively  $z_s^a$ ,  $z_t^a$ , and  $z_s^b$ ,  $z_t^b$ . We keep the structure and texture vectors of  $I_a$ , while only the texture vector of  $I_b$  is retained for the swapping.

## Generator

The generator  $G$ , based on StyleGAN2 [75], is trained to produce two image hybrids: from  $z_s^a$  and  $z_t^a$  it generates  $I_{a,a}$ , a reconstruction of the original  $I_a$ , and from  $z_s^a$  and  $z_t^b$  it creates the swapped image  $I_{a,b}$  with the structure of  $I_a$  and texture of  $I_b$ . It is built with four residual blocks and four upsampling blocks, with a final convolution to generate an image. Since  $E_s$  produced the tensor  $z_s$  containing spatial information, it can be decoded to reconstruct the original input structure. On the other hand, the texture vector  $z_t$  is used at every decoder layer to modulate the latters' parameters. Following the generator implementation of StyleGAN2, this includes layer noise injection, weight demodulation, bilinear filtering, equalized learning rate, minibatch variance adjustment and leaky ReLU activation functions [75].

## Discriminator

Given the original input image  $I_a$ , its reconstruction  $I_{a,a}$  and its swapped version  $I_{a,b}$ , the goal of the discriminator network  $D$  is to distinguish reconstructed  $I_{a,a}$  and swapped images  $I_{a,b}$  in the training set. The discriminator network  $D$  also follows the implementation guidelines introduced by StyleGAN [73]. It consists of downsampling residual blocks with a binary classification head.

## Co-occurrence discriminator

A co-occurrence discriminator model is introduced to enforce that patches from the swapped image  $I_{a,b}$  be identical to patches at the same position from the original input image  $I_b$ ,

which ensures proper texture transfer. To accomplish this, a group of  $N = 4$  reference patches  $patches_{ref}$  are sampled by cropping the image to between  $1/8$  and  $1/4$  of its original size. The patches are then randomly scaled and rotated in the region of interest defined by the binary mask of the original input image. We thereby sample four positions across the region of interest (e.g., the mask identifying the leg), and compare the patches from the original input image  $patch_{ref}$  with those from the swapped image  $patches_{swap}$  at these positions. The co-occurrence discriminator  $D_{cooc}$  then tries to differentiate the original patches from the swapped ones, ensuring that the textures between both images become as identical as possible at the level of these patches. The co-occurrence discriminator itself is constituted of downsampling residual blocks followed by average pooling and four linear layers with learning rate equalization, producing binary classification outputs.

### Self-supervised training

During training, any combination of individual body parts for texture swapping is permitted. This is useful to efficiently learn to transfer texture between fairly different structures. However, it was found empirically that adding horizontal flips to one side (e.g., the right side) led to more stability during the training process. Indeed, more structural correspondences are present between the two inputs if both body parts have the same orientation.

The training objective of the model can be decomposed into four components: reconstruction, adversarial, texture supervision and distortion suppression. We define the overall training objective as:

$$\mathcal{L} = \mathcal{L}_{recon} + \lambda_1 \mathcal{L}_G + \lambda_2 \mathcal{L}_{tex} + \lambda_3 \mathcal{L}_{supp} \quad (5.1)$$

**Image reconstruction** Since we want the swapping autoencoder to reconstruct the input image given its own structure and texture latent codes, the reconstruction loss function can simply be described as:

$$\mathcal{L}_{recon} = \mathbb{E}_{I_a \sim X} ||G(z_s^a, z_t^a) - I_a||_1 \quad (5.2)$$

**Adversarial training** Following the adversarial process for GAN training by Goodfellow et al. [64], we define the adversarial loss for the generator  $G$  as:

$$\mathcal{L}_G = \mathbb{E} \left[ -\log(D(G(z_s^a, z_t^a))) - \log(D(G(z_s^a, z_t^b))) \right] \quad (5.3)$$

The generator will try to fool the discriminator  $D$  with both reconstructed and swapped images, ensuring realism in the generation.

**Texture supervision** To ensure that the texture codes are decoupled from structure, we force the patches of both  $I_b$  and  $I_{a,b}$  to be as similar as possible in the masked region. This is done by training a patch discriminator network  $D_{cooc}$  to differentiate the sampled patches from  $I_b$  and those of the hybrid  $I_{a,b}$ , while  $G$  tries to fool  $D_{cooc}$ . We can define this objective function as:

$$\mathcal{L}_{tex} = \mathbb{E}_{I_b \sim X} [-\log(D_{cooc}(\tau_{in}(I_b), \tau_{in}(I_{a,b})))] \quad (5.4)$$

with  $\tau_{in}$  being the method that samples patches with random scaling and rotations over the region of interest defined by the mask of  $I_a$ , the source structure image.

**Distortion suppression** Since the texture features are only relevant in the masked region, we want to prevent the texture transfer from distorting or introducing artifacts outside the region of interest, including borders and out-of-mask image regions. We also want to learn features based only on information from exposed skin areas. To achieve this, we use the patch-wise contrastive loss introduced in [128] and used in [4]. Here, a structure feature vector  $q_i$  is sampled from the  $h^{th}$  layer of the structure encoder, fed with the swapped image (re-encoded). Then, from the encoded structure image  $I_a$ , we sample a set of  $N$  features  $\{p\}_N$  with  $p^+$  corresponding to the same position as  $q_i$  and  $\{p^-\}_{N-1}$  being those at different positions. Ultimately, we want to attract  $q_i$  with  $p^+$  while repelling  $p^-$ . This can be written as a negative cross-entropy loss (NCE):

$$l_{nce}(q_i, \{p\}_N) = -\log \left( \frac{e^{(q_i \cdot p^+ / \alpha)}}{e^{(q_i \cdot p^+ / \alpha)} + \sum_{j=1}^{N-1} e^{(q_i \cdot p_j^- / \alpha)}} \right) \quad (5.5)$$

The parameter  $\alpha = 0.07$  is the temperature, set as in [4]. The final structural distortion suppression objective is:

$$\mathcal{L}_{supp} = \mathbb{E} \left[ \sum_h^H \sum_i^N l_{nce}(\tau_{out}^i(z_{s,h}^{a,b}), \{\tau_{out}(z_{s,h}^a)\}_N) \right] \quad (5.6)$$

with  $\tau_{out}$  being a sampling function, taking  $N$  attributes from the structure vectors. These spatial attributes are located outside the mask of  $I_a$ .



**Training parameters** The coefficients for the loss terms are set to  $\lambda_1 = 0.5$ ,  $\lambda_2 = 1$  and  $\lambda_3 = 1$ , following the hyperparameters suggested by [3] and [4]. The networks are optimized by Adam [129] with  $\beta_1 = 0.5$  and  $\beta_2 = 0.99$  and a learning rate of 0.001. We train the TADA-SAE swapping task on one NVIDIA RTX3080 for 400,000 steps with a batch size of four and unit gradient clipping. The learning rates for all networks of the GAN model are exponentially decayed with  $\gamma = 0.99$ , which is updated every 1,000 steps. Moreover, R1 regularization [130] is applied to both discriminators every 16 steps to stabilize the training procedure.

**Data augmentations** Efficient training of GANs needs considerable amounts of data to produce realistic generations as well as properly disentangled latent spaces. Popular image augmentation techniques such as Gaussian blur, Gaussian noise, affine transformation and color jitter can be applied directly on the images. This has been proved to boost performance in various tasks such as image classification, object segmentation and in the medical field over multiple modalities [131]. Since this mechanism is powerful and versatile, it may be tempting to use it directly in image generation models. However, it has been shown that GAN generation then risks emulating the augmentations’ distribution [132,133]. To alleviate this problem, Karras et al. introduced an augmentation scheme named ADA (Adaptive Discriminator Augmentations) acting only on the discriminator network [76]. A diverse set of affine, geometric and color transformations are adaptively generated based on the discriminator’s accuracy throughout training. The augmentation heuristic is defined as:

$$r = \mathbb{E}[\text{sign}(D_{\text{train}})] \quad (5.7)$$

meaning that the heuristic is equivalent to  $D$ ’s accuracy. The value of  $r$  ranges from 0 to 1, with  $r = 0$  meaning no overfitting and  $r = 1$  for a discriminator with perfect classification. The more accurate  $D$  is, the more augmentations are applied to both generated and original images, up to a maximum sampling probability set to  $p_{\text{max}} = 0.8$ . Using the previously computed heuristic, the augmentation probability at step  $i$  is described as:

$$p_i = \min(p_{i-1} + \gamma(r_i - r_{\text{target}}), p_{\text{max}}) \quad (5.8)$$

with  $r_{\text{target}}$  being the minimum heuristic value over which the augmentation probability will increase by a factor set to  $\gamma = 0.0001$ . Ultimately, the ADA method prevents the discriminator from overfitting while making the GAN model able to generalize over small datasets of only few thousands images [76]. In our method, it was sufficient to apply the

ADA augmentations to the discriminator network only since random scaling and rotations are already applied to the co-occurrence discriminator.

#### 5.4.2 Breast cancer detection on DMR-IR

##### Dataset

The Database for Mastology Research with Infrared Image (DMR-IR) used in this project comprises breast thermal images of 57 patients collected by the Computing Institute of the Fluminense Federal University in Brazil [1]. This dataset, comprising 1102 images in total, is widely used for breast cancer classification. Each patient had a set of thermograms (maximum 20) collected according to a dynamic protocol: a thermal image was taken every 15 seconds after artificial cooling of the breasts. A total of 38 patients were diagnosed with early or late stage unilateral breast cancer; the condition can be observed in the thermal images as localized temperature increases over the affected regions. The remaining 19 were labeled as normal. Images were acquired using a FLIR SC620 thermal camera with a resolution of  $640 \times 480$  pixels, a sensitivity of  $0.04^\circ\text{C}$  and a range from  $-40^\circ\text{C}$  to  $500^\circ\text{C}$ . Figure 5.2 shows normal (healthy) and anomalous (cancerous) samples from the DMR-IR database.

##### Two-stage breast cancer detection

Figure 5.3 shows the two-stage inference pipeline developed for both breast cancer detection and post-operative anomaly detection. Inspired by previous work that explored the idea of asymmetry for breast cancer detection [118], we propose to divide breast thermograms into left and right images ( $I_l$  and  $I_r$  respectively) using a U-Net [134] to extract their corresponding masks. It is then possible to compute the disentangled texture feature vectors using TADA-SAE for each side ( $z_t^l$  and  $z_t^r$  respectively) and to quantify the asymmetry using a distance metric  $d(z_t^l, z_t^r)$  between the two vectors. Note that, similarly to the training stage, left side images are flipped horizontally to ensure alignment with the corresponding right side images before being passed to the frozen texture encoder.

Since the presence of hyperthermia on one breast can reveal the presence of a tumor, the comparison of left and right feature vectors is straightforward. Considering the size of the DMR-IR dataset, we found that an absolute difference between both  $z_t^l$  and  $z_t^r$  vectors is sufficient for classification. The distance function is:

$$d_{dmrir}(z_t^l, z_t^r) = ||z_t^l - z_t^r|| \quad (5.9)$$

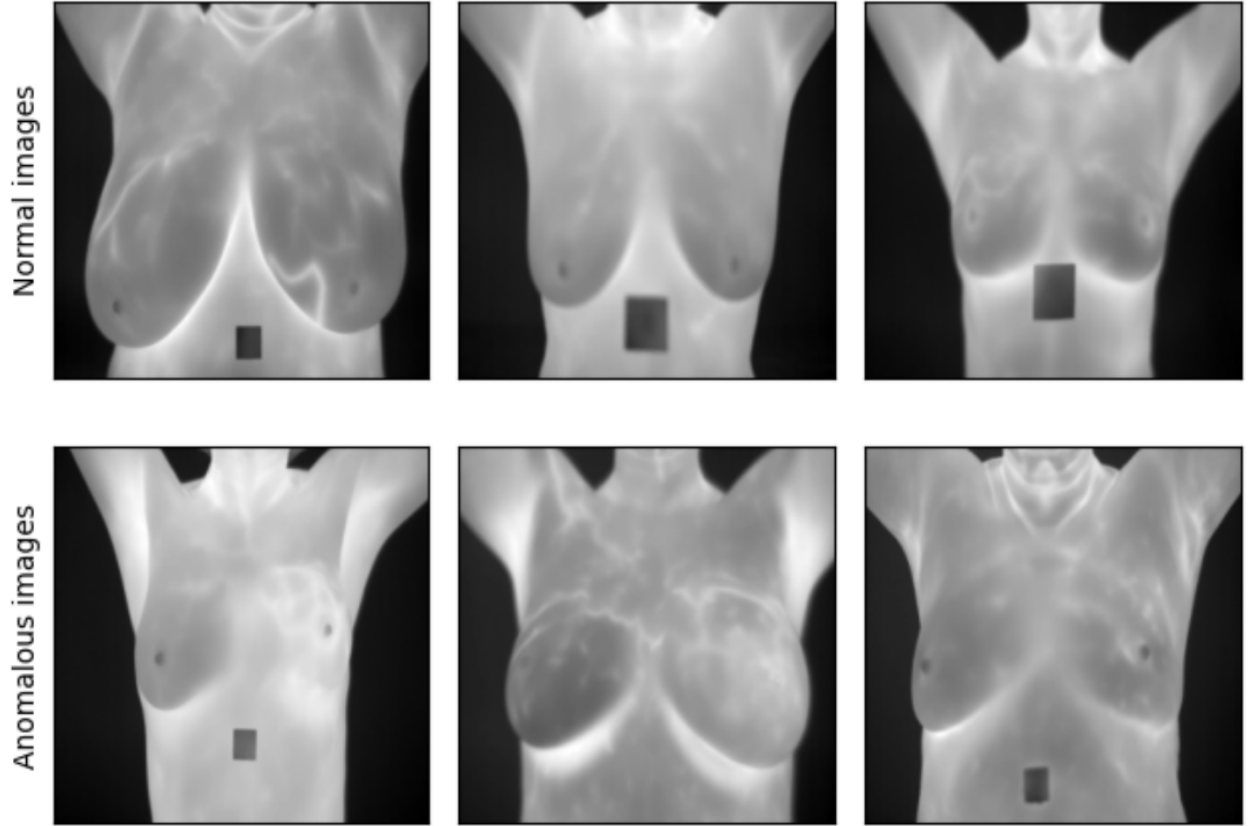


Figure 5.2 Samples of normal (top row) and anomalous (bottom row) images from the DMR-IR database [1]. For samples diagnosed with breast cancer, the affected side appears warmer than the contralateral breast.

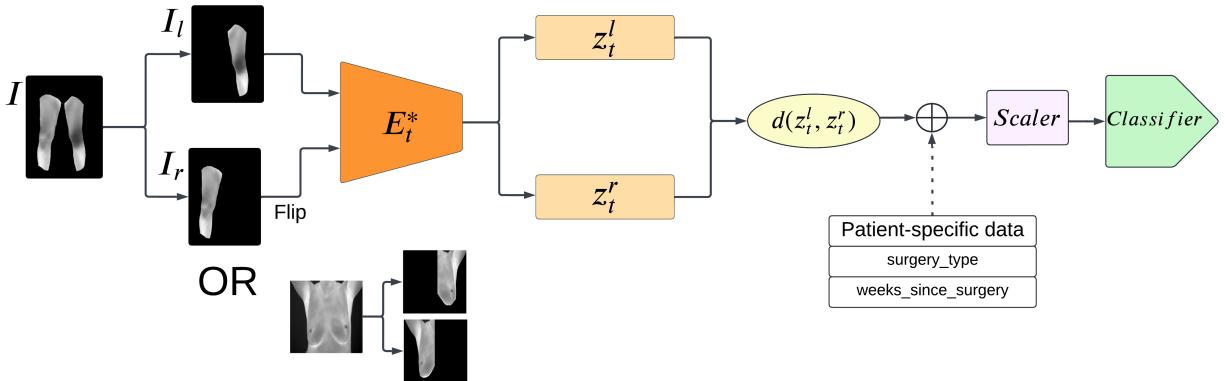


Figure 5.3 Inference pipeline for the two-stage breast cancer/post-operative anomaly detection method. On DMR-IR, the distance  $d$  is the absolute difference between left and right texture vectors, and no clinical data are used for the classification. On the OrthoPOT dataset, due to its limited size, the distance  $d$  is the dot product between the texture vectors;  $d$  is concatenated with patient-specific data (dashed line) before the classification stage.

which produces a distance vector quantifying feature asymmetry in the texture space.

The final step of the inference pipeline consists of a classification model, trained with supervision over features asymmetry. We chose an SVM classifier with a Gaussian kernel for this task due to its ability to work with limited data and interpretability.

### 5.4.3 Post-operative anomaly detection in OrthoPOT

#### Dataset

In the absence of any publicly available datasets of thermal images of the lower limbs after TKA and THA, we conducted our own data collection, with the approval of the ethics committees of the CIUSSS de l’Est de l’Île de Montréal (#2024-3501) and Polytechnique Montreal (CER-2324-70-D). 65 patients, attending their six-weeks to one-year postoperative follow-up visit at Maisonneuve-Rosemont Hospital in Montréal, were recruited over the span of five months. Thermal images were acquired using a FLIR T1020 camera with a resolution of  $1024 \times 768$  and a thermal precision of 0.02 degrees Celsius. When possible, a total of four images per patients were taken (front, left, back, and right), with each picture showing both lower limbs while standing up. Each view shows both legs from the ankle to the hip, regardless of the surgery type. The camera was positioned about 1.5m from the patient. Both legs were manually annotated in each image using the Label Studio [121] program to generate segmentation masks of exposed skin areas, excluding any clothing. An example is provided in Figure 5.4. While raw temperature matrices from the camera were kept, the range was clipped between  $26.00^{\circ}\text{C}$  and  $39.00^{\circ}\text{C}$  to optimize background removal and maximize the intensity range on the skin surface regions, thus easing the segmentation process. Moreover, patient-specific and clinical data such as age, gender, weight, height, surgery type (ST), surgery location, number of weeks since surgery (WSS) at the time of acquisition and the abnormality if present, were recorded.

The resulting Orthopedics Post-Operative Thermograms (OrthoPOT) dataset comprises a total of 244 thermal images of lower limbs of 65 patients. The cohort includes 30 male and 35 female patients (average age:  $64.18 \pm 9.47$  years old); 41 (resp. 24) of them underwent TKA (resp. THA). Among these patients, four were considered as showing post-operative complications and their images were labeled as anomalous: one reported intense pain after five weeks, one had an infection over the operated area, one was diagnosed with trochanteric bursitis and one had an uncemented prosthesis that needed revision. Figure 5.5 shows thermal images (frontal view) for both normal and anomalous cases with the corresponding WSS.

The key difference between OrthoPOT and DMR-IR is that in the former, both normal

and anomalous images will have a temperature increase over the operated region [2, 42], whereas in the latter, only malignant cases will reveal an asymmetry in temperature between left and right breasts [118]. Therefore, it is important for the anomaly detection model to contextualize these post-operative images with temperature increases. This justifies the need to consider patient-specific data on top of thermal images for this application (see Figure 5.3). For instance, severe hyperthermia lasting several months after a surgery could be predictive of an underlying complication.

### Two-stage post-operative anomaly detection

The inference pipeline for post-operative anomaly detection follows the same steps as for breast cancer detection (section 5.4.2) with a few modifications. For inference, only images showing the operated and contralateral region are kept. Therefore, only the frontal view with both limbs is considered for TKA patients, whereas the left and right lateral views, showing the hip areas, are used for THA. Since OrthoPOT is much smaller than DMR-IR, it is more effective to reduce the dimensionality of the texture asymmetry vector by considering another definition for the distance  $d_{orthopot}(z_t^l, z_t^r)$ . Different distance metrics were tested: Euclidean distance, cosine similarity and dot product between  $z_t^l$  and  $z_t^r$ .

The distance value between left and right texture vectors is concatenated with patient-specific data before classification. While the number of weeks since surgery is predominant in understanding the temperature differences between limbs [2, 41], we also include the one-hot encoded surgery type (TKA or THA) for classification on the OrthoPOT dataset. The clinical attributes (for OrthoPOT only) and thermal asymmetry features are then scaled using a robust scaler [135], which is less sensitive to outliers, by centering the data points on the median and dividing them by the interquartile range. Due to the high class imbalance found in the OrthoPOT dataset, we opted for an anomaly detection strategy. More specifically, we used an isolation forest (iForest) [136] with 200 estimators as the classifier for this task.

## 5.5 Experiments and Results

To assess the classification and anomaly detection abilities of the method on both datasets, we employed 10-fold cross-validation, with each fold repeated with 10 different seeds. We evaluated the models using the following metrics: F1 score, precision, recall and AUROC (area under receiver operating characteristic curve). For OrthoPOT in particular, we also considered AUPRC (area under precision-recall curve) due to the high class imbalance.

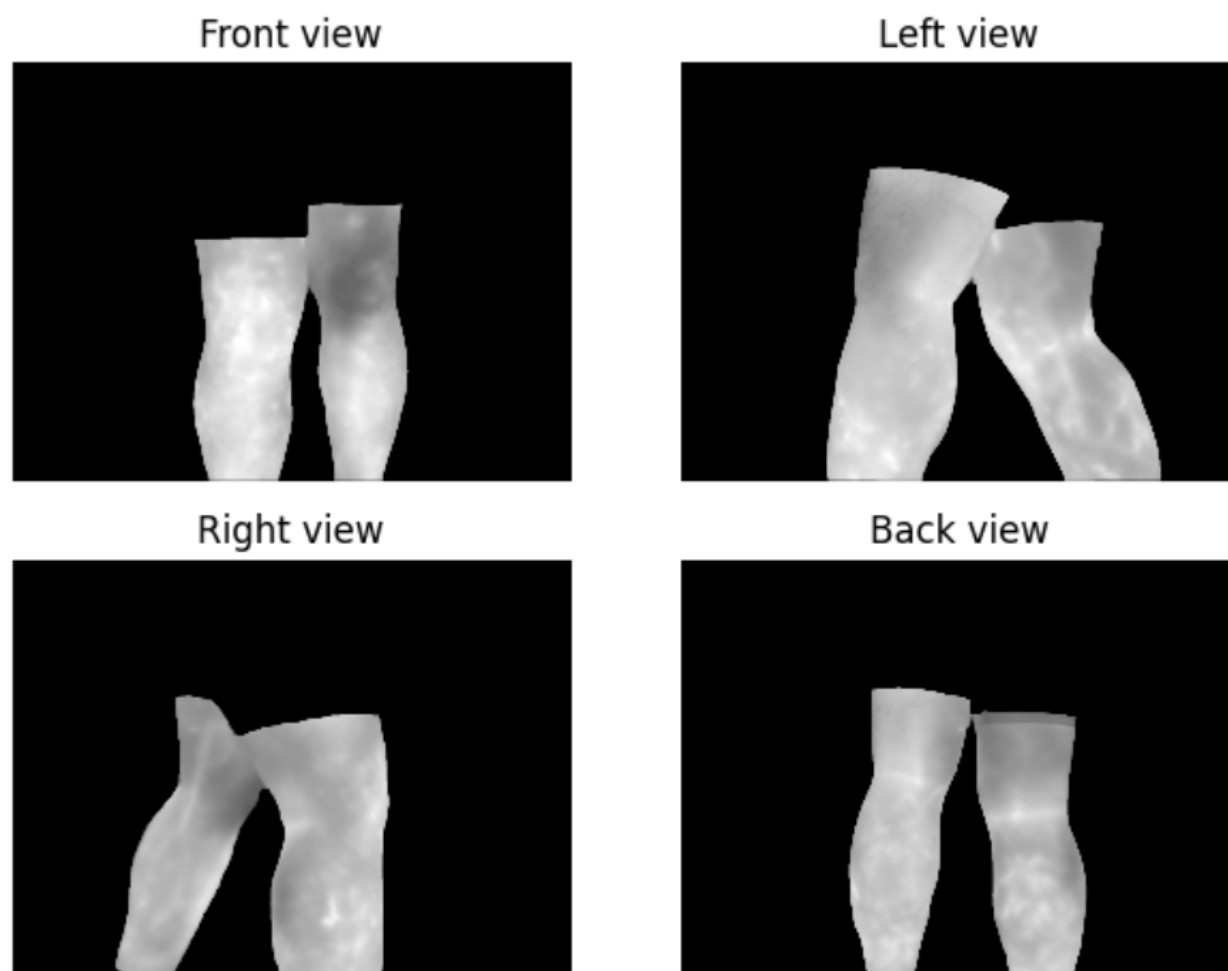


Figure 5.4 Thermal images showing lower limbs following a total knee arthroplasty (TKA). The operated leg (left leg) is warmer than the contralateral one around the knee.

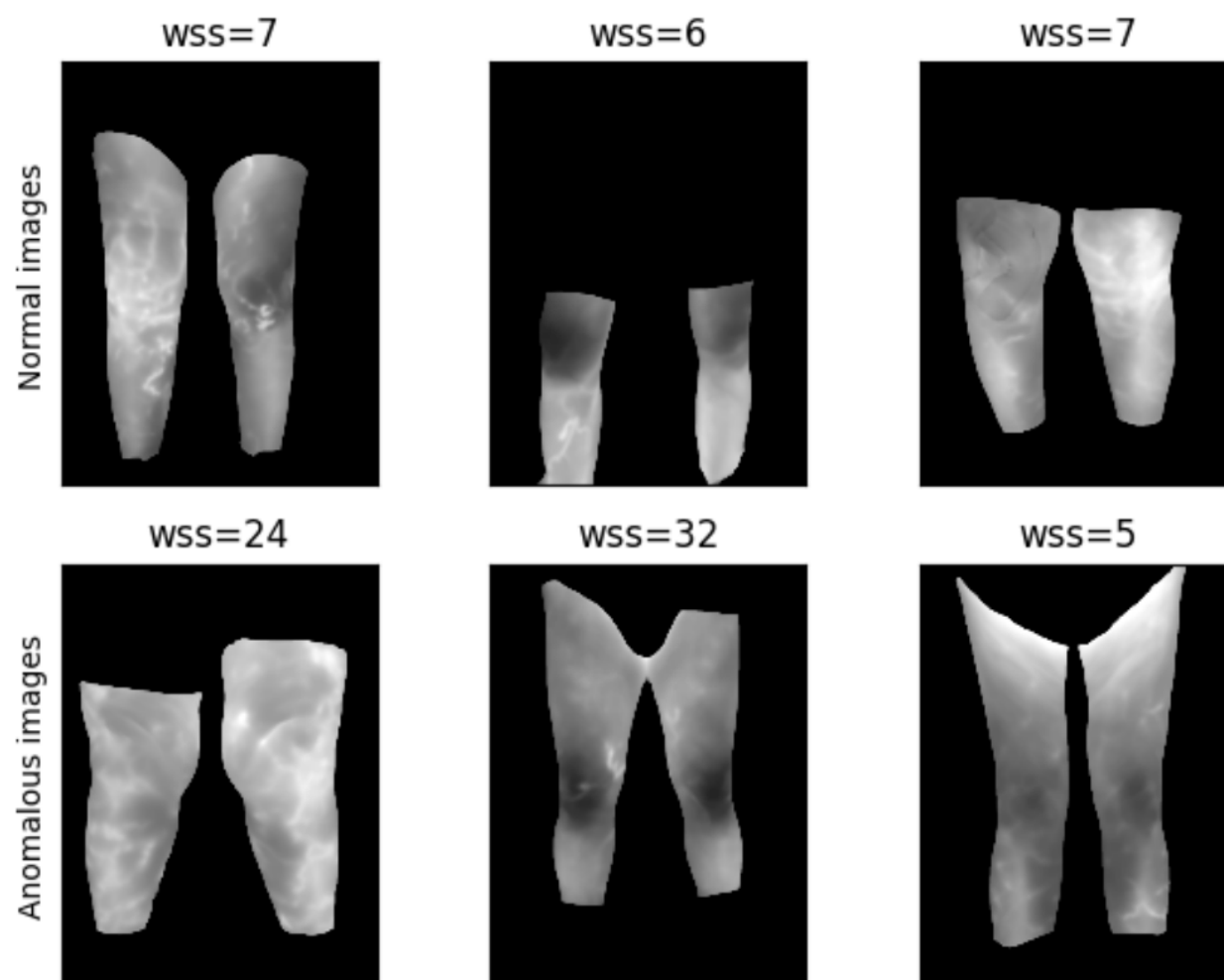


Figure 5.5 Frontal view thermal images following TKA for normal (top row) and anomalous (bottom row) patients, with corresponding number of weeks since surgery (WSS).

### 5.5.1 Breast cancer detection in DMR-IR

#### Baseline and literature comparisons

To evaluate the classification abilities of TADA-SAE on DMR-IR, we first compared it to popular image classification architectures as baselines trained from scratch, and to results reported in the literature on the same dataset. These results are shown in Table 5.1.

Table 5.1 Classification metrics on DMR-IR dataset. Results from [29,31,33,118] are reported as is. Methods marked with (\*) may not be comparable with ours due to undisclosed train-test split strategy or to different validation methodology.

Model	Params. (M)	Accuracy	F1	Prec.	Recall	AUROC
Small CNN	0.5	0.859	0.838	0.875	0.857	0.950
ResNet-18	11.1	0.846	0.803	0.983	0.702	0.894
Inception-v3	27.1	0.755	0.775	0.726	0.900	0.913
Zuluaga-Gomez et al. [29]	11.1	0.90	0.91	0.91	0.90	0.90
Dey et al. [118] (*)	–	0.961	0.966	0.934	1.000	–
Silva et al. [31] (*)	–	0.986	0.978	0.986	0.971	0.99
Mahoro et al. [33] (*)	66.3	0.984	0.964	0.964	0.964	0.997
TADA-SAE (ours)	2.6	0.962	0.971	0.994	0.950	0.988

Among the considered baseline models, a small CNN with only 0.5M parameters appears to perform the best, showing how model size affects the results. When comparing to previous work, it is worth noting that the reported results are not always comparable to our own due to undisclosed train-test split strategy or to different validation methodology. In [29], the authors used segmented breasts and data augmentations to train a CNN classifier and performed Bayesian hyperparameter optimization. Similar to our work, Dey et al. [118] used asymmetry in hand-crafted texture features and a hard voting algorithm for classification. Silva et al. [31] trained a K-Star classifier to detect anomalies based on image sequences. Recently, [33] used a transformer-based approach for breast segmentation and trained an EfficientNet model [137] over these images. Our TADA-SAE achieves competitive results with an SVM classifier, while needing considerably fewer parameters than the other approaches.

#### Ablation study

We then performed an ablation study of the TADA-SAE architecture adapted to the DMR-IR dataset. A convolutional autoencoder (AE) of 3.8M parameters was trained to reconstruct full breast images and left-right regions. The extracted encoder features were then used to train the classifier from full image features (full im.) or from asymmetry using  $d_{dmr\text{ir}}$  (left-



right). To show the efficiency of texture asymmetry, an LSAE model was trained to extract texture features from full thermal images, hence without left-right comparison. These models were compared with the TADA-SAE architecture, as shown in Table 5.2.

Table 5.2 Ablation study of TADA-SAE on the DMR-IR dataset

Encoder	Classifier	Accuracy	F1	Precision	Recall	AUROC
AE full im.	SVM	0.701	0.605	0.923	0.450	0.508
AE left-right	SVM	0.733	0.737	0.670	0.896	0.739
LSAE full im.	SVM	0.758	0.802	0.670	<b>1.000</b>	0.539
LSAE full im.	Linear	0.882	0.850	0.939	0.806	0.965
TADA-SAE	Linear	0.902	0.914	0.858	0.997	<b>0.999</b>
TADA-SAE	SVM	<b>0.962</b>	<b>0.971</b>	<b>0.994</b>	0.950	0.988

The simple convolutional autoencoder (AE) trained on full images does not perform left-right comparison nor produce a disentangled latent space. The trained SVM could not perform effective feature classification, as shown by the 0.508 AUROC. Performance increased when AE left-right feature differences were computed prior to classification, showing the benefit of asymmetry analysis. To understand the effect of latent space disentanglement only, the LSAE was trained on full images and detection was performed with linear and SVM classifiers, increasing accuracy. By combining the strengths of texture features and the left-right comparison of learned attributes, the TADA-SAE with an SVM classifier gave the best classification performances overall.

### Sensitivity to thermal differences

To further analyze the sensitivity of the differences between disentangled texture features in breast cancer detection, we simulated unilateral hyperthermia by taking healthy patients and increasing pixel values over a breast region by a uniform warming factor, effectively raising the temperature of a single breast. The pixels of the images were scaled between 0 and 1. We then performed a principal component analysis (PCA) on the absolute differences between left and right texture vectors from TADA-SAE and on the left-right feature vector differences from the convolutional autoencoder (AE), as shown in Figure 5.6. We chose a warming factor  $W_F$  of 0.05 (applied to a randomly chosen side on each patient).

The effect of disentanglement is clearly shown in the PCA projections. For the AE, the temperature increase is barely reflected in the first two principal components. On the other hand, the texture feature differences from TADA-SAE reveal an important shift of the samples after the temperature increase.

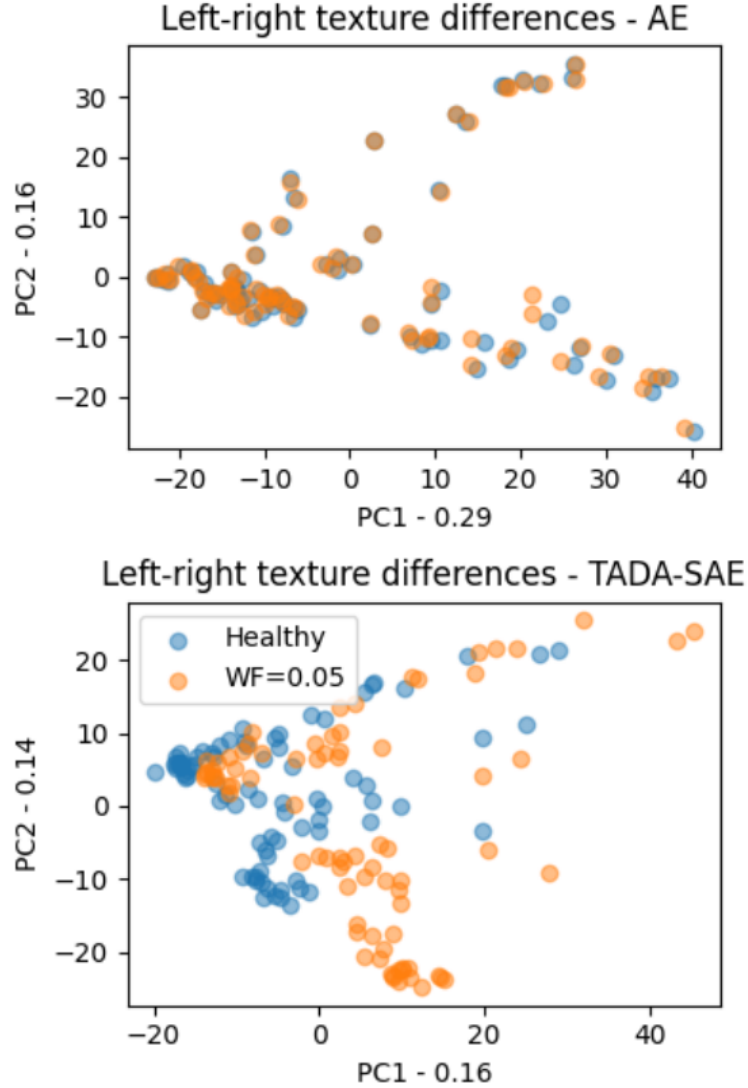


Figure 5.6 Projected representations of left-right features differences from a simple autoencoder (AE) and from the disentangled texture space learned by TADA-SAE. With a small warming factor ( $WF=0.05$ ) applied to one side in healthy samples, the TADA-SAE is much more efficient at separating the samples according to the main principal components from PCA. Each axis represents one of the two principal components and its explained variance ratio.

### 5.5.2 Post-operative anomaly detection in OrthoPOT

#### Ablation study

In every fold, the classification model was trained on 51 normal and two anomalous patients, then evaluated on six normal and two anomalous patients. We trained baseline models and TADA-SAE with multiple distance functions  $d_{orthopot}$  in the texture latent space to

quantify the bilateral asymmetry. As baselines, we use the difference of average temperatures over the mask region (mean diff.), fed a one-class SVM (OCSVM) [138] and isolation forest (iForest) [136] along with clinical data. Results are gathered in Table 5.3.

Table 5.3 Ablation study of TADA-SAE on the OrthoPOT dataset

Model	Dist. Metric	F1	Precision	Recall	AUROC	AUPRC
OCSVM	Mean diff.	0.700	0.695	0.845	0.741	0.557
iForest	Mean diff.	0.780	0.833	0.800	0.795	<b>0.707</b>
AE g-d	Euclidean	0.759	0.857	0.745	0.740	0.646
TADA-SAE	Absolute diff.	0.463	0.507	0.555	0.505	0.333
TADA-SAE	PC 1	0.792	<b>0.861</b>	0.810	0.788	0.705
TADA-SAE	PC 1,2	0.762	0.800	0.815	0.769	0.647
TADA-SAE	PC 1,2,3	0.741	0.764	0.810	0.733	0.611
TADA-SAE	Euclidean	0.780	0.829	0.815	0.789	0.688
TADA-SAE	Cosine sim.	0.786	0.823	0.830	0.799	0.681
TADA-SAE	Dot product	<b>0.809</b>	0.792	<b>0.900</b>	<b>0.842</b>	0.681

We started by training two popular anomaly detection algorithms as baselines, namely an OCSVM and an iForest, on average temperature differences along with the WSS and ST attributes. The iForest classifier showed performance increases on all metrics except recall. As a third baseline, we trained a classic autoencoder (AE) to reconstruct individual limbs. An iForest trained on the Euclidean distance between these features plus relevant patient data yielded an increase in precision but a decrease in other metrics. Following this, we trained the same classifier on the absolute difference of texture vectors from TADA-SAE with clinical data. However, this experiment revealed the inefficiency of using high-dimensional (1,024) texture feature vectors with this small dataset of 244 images. In fact, when using only the three first principal components (PC 1,2,3) of the texture asymmetry space, we found that performance decreased as we increased the number of principal components used. The TADA-SAE was then tested with several distance functions, producing scalar difference values that were concatenated with clinical data for anomaly detection. Overall, the dot product between texture vectors gave the best F1 score, recall and AUROC.

### Importance of clinical data and distance measurement

We evaluated the respective roles of clinical data and thermal imaging information for anomaly detection. We removed individual attributes and trained the iForest algorithm under each condition. Results are reported in Table 5.4.

The clinical data (WSS and ST), are shown to be crucial in complementing the thermal

Table 5.4 Effect of different attributes on anomaly detection performance on OrthoPOT dataset, with surgery type (ST), weeks since surgery (WSS) and  $d_{orthopot}$

Attributes	F1	Precision	Recall	AUROC	AUPRC
$d_{orthopot}$ +ST	0.689	0.597	0.925	0.718	0.491
$d_{orthopot}$ +WSS	0.646	0.709	0.670	0.604	0.506
WSS + ST	0.754	0.849	0.760	0.774	0.678
$d_{orthopot}$	0.635	0.663	0.735	0.635	0.517
all	0.809	0.792	0.900	0.842	0.681

asymmetry. By removing WSS ( $d_{orthopot}$ +ST), one can achieve classification even if temporal context is omitted. This could be due to biases in the dataset and to the small amount of data available for evaluation. We also found that surgery type is important for the model to recognize normal distance values for TKA and THA, which have different levels of temperature distributions, since frontal and lateral images were used for TKA and THA respectively ( $d_{orthopot}$ +WSS). We then removed the distance attribute to test how the model performed on clinical data only (WSS+ST). Neither of these attributes should have any incidence on post-operative outcome, yet non-random classification was nevertheless possible. Likewise, When removing both clinical attributes, the iForest algorithm trained only on the scalar product  $d_{orthopot}$  was able to achieve better than random predictions. These results can be ascribed to the dataset’s small dataset size and potential biases. Ultimately, combining all of these attributes leads to substantial increases in the AUROC and F1 scores.

### Dimensionality effect

As shown in Table 5.3, the small number of samples makes the direct use of full-length vectors from the texture space inefficient. To illustrate the curse of dimensionality in OrthoPOT, we plot, in Figure 5.7, the performance of the model as a function of the number of principal components taken from the vectors in the texture space. Model performance decreases steadily with increasing dimension of the principal component vector, and falls to the level of random guessing at 20 principal components.

## 5.6 Discussion

While our approach for texture feature extraction is based on LSAE, multiple adaptations were made to achieve efficient anomaly detection. First, the self-supervised training for texture swapping is done with individual body parts, which were previously extracted using their segmentation masks, ensuring that the texture features only contain relevant informa-

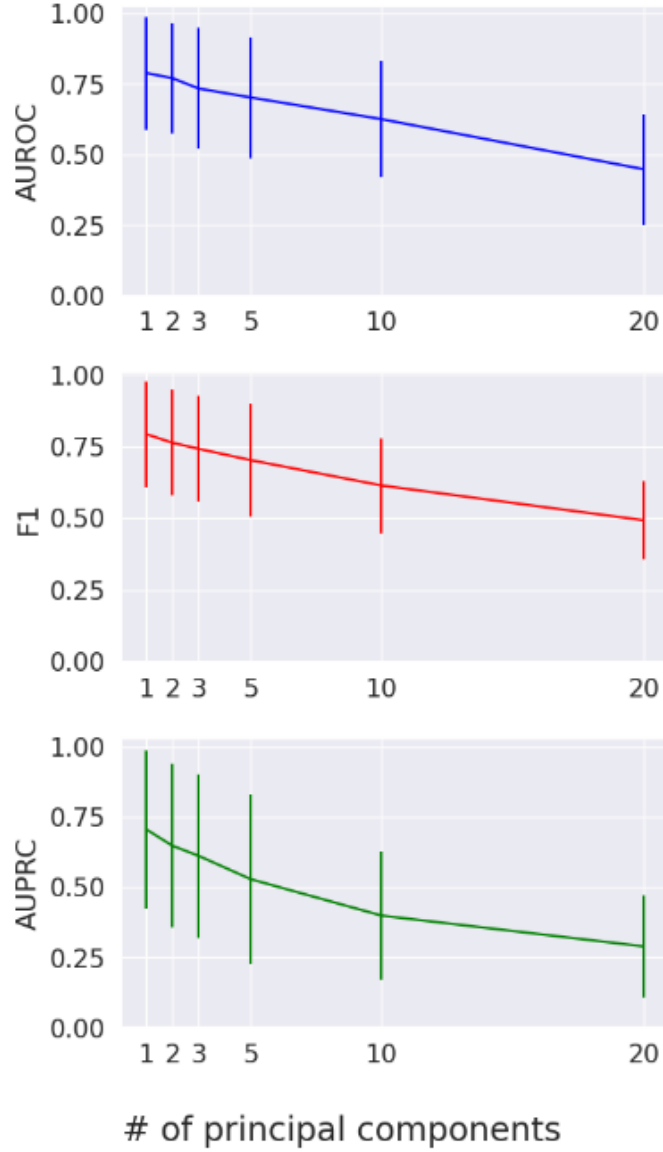


Figure 5.7 Effect of texture difference principal components in OrthoPOT classification. We notice that performance decrease on AUROC as the number of principal components increases

tion. To enable this self-supervised training paradigm to efficiently work on small datasets, an Adaptive Discriminator Augmentations module was added to the discriminator. This allows us to use data augmentations on that module without leaking into the generations (and learned features), a problem that affects generative models that use standard image augmentation techniques. Finally, we compare left and right texture vectors from opposite body parts of the same patient to perform anomaly detection with popular classification methods like OCSVM and iForest.

The competitive results obtained on DMR-IR show how the texture differences capture temperature variations between the two sides, leading to efficient classification using basic algorithms. Since similar temperature distributions and intensities will produce similar texture vectors regardless of the underlying structures containing these values, an absolute difference between left and right sides can describe temperature changes in the breasts. Compared to state-of-the-art methods using hand-crafted features, the symmetry vector of length 1,024 was found descriptive enough to achieve competitive results, as shown in Table 5.1. Competitive results were achieved by an isolation forest on DMR-IR despite the small size of the texture encoder, yielding only 2.6M parameters for the inference model.

For OrthoPOT, dimensionality reduction operations were necessary to boost the classification performance when compared to simple differences of temperature means between operated and contralateral limbs (mean diff.). The first approach was to use the principal components of vectors in the texture space, thus drastically reducing the dimensionality. However, as the number of principal components increases, performances decreases, showing that low-dimensional vector projections were not meaningful enough to achieve decent classification. Therefore, various distance metrics were tested to obtain scalar values which could be better suited at quantifying the degree of asymmetry between sides. The Euclidean distance, measuring the length of the line segment between points in feature space, did not lead to performance gains in terms of precision, AUROC and AUPRC compared to an iForest trained with temperature means differences over the masks, but it did increase recall. On the other hand, cosine similarity, which ignores the vector magnitudes, was found to increase AUROC and AUPRC. Ultimately, the dot product between the vectors was the most descriptive distance metric, by considering both vector angles and magnitudes. The last row of results in Table 5.3 indeed shows substantial increases in F1, recall and AUROC compared to feature differences with classic autoencoder and TADA-SAE with PCA dimensionality reduction.

However, our experiments summarized in Table 5.4 highlight potential issues related to the small number of patients and data biases in OrthoPOT. For instance, it is possible to achieve 0.774 AUROC without using  $d_{orthopot}$ . An example of bias pertains to the THA patients: both anomalous cases in that group were acquired 24 and 32 weeks following surgery, whereas all the normal cases were acquired between six and 10 weeks post-THA, making detection obvious for this group when using the WSS attribute. Nevertheless, including  $d_{orthopot}$  increases AUROC, F1 and recall by a significant margin compared to only using clinical attributes (WSS + ST), showing the relevance of left-right texture comparisons. Moreover, some identified anomalies were prospective based on patient feedback using a survey sent six weeks following the acquisition. Note that comparisons with the literature for post-operative anomaly detection were not possible due to the lack of public datasets and studies tackling similar

problems.

As shown on both the DMR-IR and OrthoPOT datasets, extracting texture features and evaluating their asymmetry provides rich information to distinguish normal and abnormal samples. Since bilateral symmetry is a universal feature of the human body, one could extend this method to other imaging modalities and organs. As demonstrated by [4], texture features are also efficient for chest x-ray classification. In the present work, we show how the approach can be applied effectively to the thermal domain. However, it is important to recall the differences between the DMR-IR and OrthoPOT scenarios. In the former, normality can be described by an uniformity in the temperatures between the two breasts. In the case of OrthoPOT, since higher skin temperature is a normal post-surgical manifestation, every patient is likely to have a prior hyperthermia over the operated limb. The anomaly detection, in that case, must be able to link the observed temperature increase to the various factors affecting it, most notably the time since surgery, as explained from a clinical standpoint in [2, 41]. Furthermore, since the underlying concept of our method does not depend on the classification method used after texture features extraction, one could extend TADA-SAE to new applications with any appropriate classification algorithm suited to the specific task.

Another important factor to consider is the explainability of the developed solution, particularly in the medical field [139]. While explainable methods can be used for classification after feature extraction, the extracted features themselves cannot intuitively characterize the temperature variations and intensities contained in the textures, let alone the differences between opposite sides. However, the work of [140] shows how generative AI, especially using StyleGAN2, can be used to interpret the effects of different components of the style vector on the generated images. Such an approach could be well suited to gain insights on how texture vectors encode the temperature information; this could be addressed in future work.

## 5.7 Conclusion

This paper proposes TADA-SAE, a method for anomaly detection that uses disentangled texture features from thermal images and leverages a fundamental feature of the human body, namely bilateral symmetry. In several medical contexts, symmetry of surface temperature between opposite body parts characterizes a normal state, while asymmetry signals potential pathology. The disentangled texture vectors learned by TADA-SAE from corresponding left and right sides were found to effectively describe subtle but significant temperature differences. A classification module can then be trained on the difference (distance) between the texture vectors to predict the presence of anomalies. Our approach achieves competitive results on the public DMR-IR dataset for breast cancer classification. On a private dataset

of thermal images of patients following TKA or THA, our method yields promising results by combining texture representations with selected clinical data to detect post-operative complications. Its lightweight and intuitive inference framework make TADA-SAE a versatile method for anomaly detection in medical images.

### **Compliance with Ethical Standards**

The authors declare that they have no conflicts of interest. The data collection and the research project were approved by the research ethics boards of the CIUSSS de l’Est de l’Île de Montréal (#2024-3501) and Polytechnique Montreal (CER-2324-70-D).

### **Acknowledgments**

We want to sincerely thank all the participants who contributed to the OrthoPOT dataset, as well as the medical and administrative staff at Hôpital Maisonneuve-Rosemont (Montréal, QC, Canada) who made this project possible. This project is funded in part by the Government of Canada, through the National Research Council’s Aging in Place Challenge Program, and a Natural Sciences and Engineering Research Council of Canada’s discovery grant (RGPIN-2020-06695). This research was enabled in part by support provided by Calcul Québec and the Digital Research Alliance of Canada. The authors would also like to thank Philippe Debanné for editing this manuscript.



## CHAPITRE 6 CONCLUSION

Ce chapitre conclut le travail présenté dans ce mémoire. D’abord, une synthèse des travaux accomplis dans le cadre de cette maîtrise sera décrite. Ensuite, les limitations de la solution proposée seront discutées pour terminer avec les perspectives futures que ce projet a permis d’ouvrir.

### 6.1 Synthèse des travaux

Dans ce projet, deux éléments majeurs ont été développés afin de détecter les complications postopératoires dans les chirurgies de remplacement total du genou et de la hanche.

D’abord, un ensemble de données comportant 66 patients (61 cas normaux, 5 anomalies) a été érigé. De pair avec l’Hôpital Maisonneuve-Rosemont de Montréal, un total de quatre images thermiques montrant les membres inférieurs des patients ayant été opérés entre six semaines et un an ont été récoltées sur une période de cinq mois. En plus de ces images, l’ensemble de données baptisé OrthoPOT (*Orthopedics Post-operative Thermograms*) comporte des attributs propres au patient associés à son ensemble d’images thermiques. Sur ces acquisitions, des annotations définissant sous forme de masque des jambes gauches et droites des patients ont été rajoutées afin de définir plus clairement les régions d’intérêt, en l’occurrence la peau exposée, démontrant le profil thermique de la région opérée et controlatérale.

En s’inspirant des techniques de détection du cancer du sein par imagerie thermique et du désenchevêtrement de l’espace latent par apprentissage auto-supervisé, une méthode de détection d’anomalies a été développée. Le TADA-SAE (Texture Anomaly Detection from Asymmetries using Swapping AutoEncoder) permet de décrire les asymétries dans les attributs de texture dans les membres opposés. Celui-ci a d’abord été testé sur l’ensemble public DMR-IR pour la détection de cancer du sein unilatéral par imagerie thermique. Les résultats convaincants avoisinant ceux de l’état de l’art ont par la suite poussé l’adaptation à la détection d’anomalies post-chirurgicales, le cœur du projet. De pair avec les attributs cliniques des patients, la méthode s’est démontrée efficace pour la détection de complications sur OrthoPOT.

### 6.2 Limitations de la solution proposée

Malgré la simplicité et l’efficacité de la méthode d’inférence pour la détection du cancer du sein et des complications post-chirurgicales, l’entraînement du SAE menant au désenchevê-

trement de l'espace latent n'est pas sans contrainte. Afin d'arriver à une solution satisfaisante, il a été nécessaire d'entraîner l'architecture sur 400000 itérations pour une durée de plus de 60 heures sur une carte graphique NVIDIA 3080. Pour une taille de lot de quatre images de dimensions  $256 \times 256$  pixels, un total de six gigaoctets de mémoire vidéo est ainsi nécessaire. Il va de soi que la préparation de ce système est lourde en termes de calcul, pouvant rendre difficile son adaptation pour d'autres cas d'utilisation.

De plus, en raison de la petite quantité de données disponibles, surtout dans OrthoPOT, l'apprentissage de la normalité peut être fortement biaisé par les déséquilibres dans l'ensemble de données. Par exemple, les patients présentant des cas d'anomalies sont souvent connus des médecins au préalable et se présentent à leur consultation suivant un horaire irrégulier, souvent près d'un an après l'intervention. Pourtant, nécessaires aux analyses, cette majorité d'anomalies ayant eu une consultation plus tardive dans l'ensemble de données constitue un biais non négligeable pouvant poser un problème à l'utilisation du système dans le monde réel. Pour ce faire, il faudrait élargir la taille de l'ensemble de données et pousser les analyses en conséquence pour en tirer des conclusions plus solides et définitives. Enfin, plusieurs de ces anomalies sont prospectives et proviennent des retours des patients, rendant difficile l'identification définitive de ces anomalies. Or, le manque d'études adaptées à la réalité clinique dans ce domaine rend difficile la comparaison avec les méthodes existantes et la détection d'anomalies diagnostiquées.

Dans un même ordre d'idées, l'apprentissage de la normalité basé notamment sur le nombre de semaines depuis l'intervention ne permet pas de définir adéquatement l'évolution du patient, dépendant d'un nombre de facteurs bien plus grand que les données cliniques disponibles. L'absence de données longitudinales ne permet ainsi pas de modéliser les déviations possibles pouvant survenir durant le rétablissement du patient, facilitant ainsi la détection d'infections ou de thromboses.

Bien que les résultats sur OrthoPOT semblent prometteurs, les biais présents dans l'ensemble de données et l'absence de collecte longitudinale ne permettent pas de conclure sur l'efficacité de TADA-SAE dans le monde réel pour la détection de complications post-opératoires. Cependant, tel que démontré avec l'ensemble DMR-IR pour la détection du cancer du sein par imagerie thermique, l'évaluation des asymétries entre les vecteurs de texture se présente comme une méthode efficace pour la détection d'anomalies lorsque la symétrie bilatérale du corps humain peut être utilisée comme base de normalité.

### 6.3 Améliorations et perspectives futures

Plusieurs perspectives sont envisageables dans la poursuite de ces travaux. Inscrit dans le cadre du projet *Aging in Place* visant le suivi à distance de la réadaptation post-arthroplastie, il serait nécessaire de concevoir un ensemble de données permettant des analyses longitudinales par patient afin de dresser une normalité plus spécifique pour chacun. Tel que discuté dans la section précédente, de nombreux facteurs influencent la guérison, et par le fait même la baisse de l’hyperthermie locale caractérisant le degré d’inflammation. De pair avec la méthode développée, il deviendrait ainsi possible de décrire avec plus de précision l’évolution du patient et permettre de signaler les déviations anormales dans les attributs extraits sur plusieurs échantillons pris à intervalles réguliers.

Dans l’optique de concevoir sur le long terme un système de suivi à distance pour la réhabilitation suivant une arthroplastie totale, un système de captation d’imagerie thermique à domicile implique l’utilisation de caméras de résolution moindre que celle utilisée dans ce projet, afin de réduire les coûts. Sachant que le TADA-SAE a été entraîné sur des images thermiques sous-échantillonnées de  $256 \times 256$  pixels, des méthodes de super-résolution pourraient être implémentées et analysées afin de rendre possible l’utilisation de caméras de grade commercial et comprendre l’impact de la résolution sur les performances du système.

Il serait de plus envisageable de réduire la complexité du modèle LSAE, moelle épinière de la solution développée dans ce projet. En effet, son implémentation a été utilisée sans modifications architecturales majeures. Il serait alors envisageable de poursuivre les tests de manière à réduire en taille le générateur et les discriminateurs, permettant un entraînement plus rapide, efficace et mieux adapté aux ensembles de données de plus petite taille.

Enfin, l’architecture développée reste tout de même peu explicable, prenant des allures de boîte noire dans les décisions prises par le classificateur. Bien que l’extraction d’attributs de texture pour évaluer les asymétries puisse sembler intuitive et compréhensible, il n’en reste pas moins que le dernier mot revient au classificateur final, en l’occurrence une machine à vecteurs de support ou un classificateur linéaire. Il est en effet difficile de comprendre la nature des faux positifs, surtout dans le cas d’OrthoPOT où la mesure d’asymétrie et les données cliniques sont fusionnées. Par exemple, un patient plus âgé que la moyenne ou portant une marque visible sur un membre inférieur pourrait provoquer une classification positive sans toutefois être en présence d’un cas anormal. Ainsi, il serait bénéfique d’analyser les effets des éléments du vecteur de textures extraits des images pour comprendre leur impact sur la décision du classificateur, permettant éventuellement une meilleure calibration du modèle pour que celui-ci agisse de manière interprétable dans le monde réel [141].

Enfin, la symétrie bilatérale étant une caractéristique universelle chez les êtres vivants, il est envisageable que la méthode développée soit utilisable pour la détection d'autres pathologies où l'uniformité entre les paires gauche-droite peut être utilisée comme descripteur de normalité.

## RÉFÉRENCES

- [1] L. F. Silva, D. C. M. Saade, G. O. Sequeiros, A. C. Silva, A. C. Paiva, R. S. Bravo et A. Conci, “A new database for breast research with infrared image,” *Journal of Medical Imaging and Health Informatics*, vol. 4, n<sup>o</sup>. 1, p. 92–100, mars 2014.
- [2] L. Gavish, L. Kandel, G. Rivkin, S. D. Gertz et O. Hoffer, “Natural history of changes in knee skin temperature following total knee arthroplasty : A systematic review and meta-analysis,” *Scientific Reports*, vol. 13, n<sup>o</sup>. 1, p. 6810, avr. 2023.
- [3] T. Park, J.-Y. Zhu, O. Wang, J. Lu, E. Shechtman, A. A. Efros et R. Zhang, “Swapping autoencoder for deep image manipulation,” déc. 2020, arXiv :2007.00653 [cs]. [En ligne]. Disponible : <http://arxiv.org/abs/2007.00653>
- [4] L. Zhou, J. Bae, H. Liu, G. Singh, J. Green, A. Gupta, D. Samaras et P. Prasanna, “Lung swapping autoencoder : Learning a disentangled structure-texture representation of chest radiographs,” janv. 2022, arXiv :2201.07344 [cs, eess]. [En ligne]. Disponible : <http://arxiv.org/abs/2201.07344>
- [5] S. Grässel et D. Muschter, “Recent advances in the treatment of osteoarthritis,” *F1000Research*, vol. 9, p. 325, mai 2020.
- [6] W. H. Organisation, “Osteoarthritis,” <https://www.who.int/news-room/fact-sheets/detail/osteoarthritis>, juill. 2023, accessed : 2024-08-19.
- [7] Y. Zhang et J. M. Jordan, “Epidemiology of osteoarthritis,” *Clinics in Geriatric Medicine*, vol. 26, n<sup>o</sup>. 3, p. 355–369, août 2010.
- [8] J. N. Katz, K. R. Arant et R. F. Loeser, “Diagnosis and treatment of hip and knee osteoarthritis : A review,” *JAMA*, vol. 325, n<sup>o</sup>. 6, p. 568, févr. 2021.
- [9] K. Tsuda, M. Takao, J. Kim, H. Abe, N. Nakamura et N. Sugano, “Asymptomatic deep venous thrombosis after elective hip surgery could be allowed to remain in place without thromboprophylaxis after a minimum 2-year follow-up,” *The Journal of Arthroplasty*, vol. 35, n<sup>o</sup>. 2, p. 563–568, févr. 2020.
- [10] T. Bongartz, C. S. Halligan, D. R. Osmon, M. S. Reinalda, W. R. Bamlet, C. S. Crowson, A. D. Hanssen et E. L. Matteson, “Incidence and risk factors of prosthetic joint infection after total hip or knee replacement in patients with rheumatoid arthritis,” *Arthritis Care & Research*, vol. 59, n<sup>o</sup>. 12, p. 1713–1720, déc. 2008.
- [11] J. S. Lee, T. Moon, T. H. Kim, S. Y. Kim, J. Y. Choi, K. B. Lee, Y. J. Kwon, S. H. Song, S. H. Kim, H. O. Kim, H. K. Hwang, M. J. Kim et Y. K. Lee, “Deep vein thrombosis

- in patients with pulmonary embolism : Prevalance, clinical significance and outcome,” *Vascular Specialist International*, vol. 32, n°. 4, p. 166–174, déc. 2016.
- [12] N. Walter, M. Rupp, K. Hierl, M. Koch, M. Kerschbaum, M. Worlicek et V. Alt, “Long-term patient-related quality of life after knee periprosthetic joint infection,” *Journal of Clinical Medicine*, vol. 10, n°. 5, p. 907, févr. 2021.
- [13] S. Haidar, R. Vazquez et G. Medic, “Impact of surgical complications on hospital costs and revenues : retrospective database study of medicare claims,” *Journal of Comparative Effectiveness Research*, vol. 12, n°. 7, p. e230080, juill. 2023.
- [14] A. Bonde, K. M. Varadarajan, N. Bonde, A. Troelsen, O. K. Muratoglu, H. Malchau, A. D. Yang, H. Alam et M. Sillesen, “Assessing the utility of deep neural networks in predicting postoperative surgical complications : a retrospective study,” *The Lancet Digital Health*, vol. 3, n°. 8, p. e471–e485, août 2021.
- [15] F. Deng, Q. Tang, G. Zeng, H. Wu, N. Zhang et N. Zhong, “Effectiveness of digital infrared thermal imaging in detecting lower extremity deep venous thrombosis : Effectiveness of digital infrared thermal imaging,” *Medical Physics*, vol. 42, n°. 5, p. 2242–2248, avr. 2015.
- [16] T. FLIR, “How do thermal cameras work ?” <https://www.flir.com/discover/rd-science/how-do-thermal-cameras-work/?srsltid=AfmBOoqZbV-QWf8N6UXMUeLI2IDI2Ih9MQ1W-RFqiwRrYYdV17757C62>, juill. 2020, accessed : 2024-08-23.
- [17] —, “Thermal camera specs you should know before buying,” <https://www.flir.com/discover/professional-tools/thermal-camera-specs-you-should-know-before-buying/>, déc. 2019, accessed : 2024-08-23.
- [18] B. Lahiri, S. Bagavathiappan, T. Jayakumar et J. Philip, “Medical applications of infrared thermography : A review,” *Infrared Physics and Technology*, vol. 55, n°. 4, p. 221–235, juill. 2012. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.infrared.2012.03.007>
- [19] D. A. Hernandez-Contreras, H. Peregrina-Barreto, J. d. J. Rangel-Magdaleno et F. J. Renero-Carrillo, “Plantar thermogram database for the study of diabetic foot complications,” *IEEE Access*, vol. 7, p. 161296–161307, 2019. [En ligne]. Disponible : <http://dx.doi.org/10.1109/ACCESS.2019.2951356>
- [20] T. FLIR, “Free teledyne flir thermal dataset for algorithm training,” <https://www.flir.in/oem/adas/adas-dataset-form/>, accessed : 2024-09-11.
- [21] E. F. J. Ring, “The historical development of thermometry and thermal imaging in medicine,” *Journal of Medical Engineering & Technology*, vol. 30, n°. 4, p. 192–198, janv. 2006.

- [22] A. Elliott, “Medical imaging,” *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 546, n°. 1–2, p. 1–13, juill. 2005.
- [23] D. Kesztyüs, S. Brucher, C. Wilson et T. Kesztyüs, “Use of infrared thermography in medical diagnosis, screening, and disease monitoring : A scoping review,” *Medicina*, vol. 59, n°. 12, p. 2139, déc. 2023.
- [24] W. I. A. for Research in Cancer, “Data visualization tools for exploring the global cancer burden in 2022,” <https://gco.iarc.fr/today/en>, 2022, accessed : 2024-08-26.
- [25] K. Feigin, “Quality assurance in mammography : An overview,” *European Journal of Radiology*, vol. 165, p. 110935, août 2023. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.ejrad.2023.110935>
- [26] N. Thakur, P. Kumar et A. Kumar, “A systematic review of machine and deep learning techniques for the identification and classification of breast cancer through medical image modalities,” *Multimedia Tools and Applications*, vol. 83, n°. 12, p. 35 849–35 942, sept. 2023.
- [27] J. Pérez-Martín et R. Sánchez-Cauce, “Quality analysis of a breast thermal images database,” *Health Informatics Journal*, vol. 29, n°. 1, p. 146045822311537, janv. 2023.
- [28] U. R. Acharya, E. Y. K. Ng, J.-H. Tan et S. V. Sree, “Thermography Based Breast Cancer Detection Using Texture Features and Support Vector Machine,” *Journal of Medical Systems*, vol. 36, n°. 3, p. 1503–1510, juin 2012.
- [29] J. Zuluaga-Gomez, Z. Al Masry, K. Benagoune, S. Meraghni et N. Zerhouni, “A CNN-based methodology for breast cancer diagnosis using thermal images,” *Computer Methods in Biomechanics and Biomedical Engineering : Imaging & Visualization*, vol. 9, n°. 2, p. 131–145, mars 2021.
- [30] S. Parshionikar et D. Bhattacharyya, “An enhanced multi-scale deep convolutional orchard capsule neural network for multi-modal breast cancer detection,” *Healthcare Analytics*, vol. 5, p. 100298, juin 2024.
- [31] L. Silva, F. Seixas, C. Fontes, D. Muchaluat-Saade et A. Conci, “A Computational Method for Breast Abnormality Detection Using Thermographs,” dans *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. Rochester, MN, USA : IEEE, juill. 2020, p. 469–474.
- [32] S. Tello-Mijares, F. Woo et F. Flores, “Breast Cancer Identification via Thermography Image Segmentation with a Gradient Vector Flow and a Convolutional Neural Network,” *Journal of Healthcare Engineering*, vol. 2019, p. 1–13, nov. 2019.

- [33] E. Mahoro et M. A. Akhloufi, “Breast cancer classification on thermograms using deep CNN and transformers,” *Quantitative InfraRed Thermography Journal*, vol. 21, n<sup>o</sup>. 1, p. 30–49, janv. 2024.
- [34] K. He, X. Zhang, S. Ren et J. Sun, “Deep residual learning for image recognition,” 2015. [En ligne]. Disponible : <https://arxiv.org/abs/1512.03385>
- [35] A. Arfaoui, M. A. Bouzid, H. Pron, R. Taiar et G. Polidori, “Application of Infra-red Thermography as a Diagnostic Tool of Knee Osteoarthritis,” *Journal of Thermal Science and Technology*, vol. 7, n<sup>o</sup>. 1, p. 227–235, 2012.
- [36] V. Lohchab, P. Rathod, P. K. Mahapatra, V. Bachhal et A. Hooda, “Non-invasive assessment of knee osteoarthritis patients using thermal imaging,” *IET Science, Measurement & Technology*, vol. 16, n<sup>o</sup>. 4, p. 242–249, 2022.
- [37] C. L. Romanò, R. D’Anchise, M. Calamita, G. Manzi, D. Romanò et V. Sansone, “Value of digital telethermography for the diagnosis of septic knee prosthesis : a prospective cohort study,” *BMC Musculoskeletal Disorders*, vol. 14, n<sup>o</sup>. 1, janv. 2013. [En ligne]. Disponible : <http://dx.doi.org/10.1186/1471-2474-14-7>
- [38] V. Lohchab, J. Singh, P. Mahapatra, V. Bachhal, A. Hooda, K. Jindal et M. Dhillon, “Thermal imaging in total knee replacement and its relation with inflammation markers,” *Mathematical Biosciences and Engineering*, vol. 18, n<sup>o</sup>. 6, p. 7759–7773, 2021.
- [39] C. Windisch, S. Brodt, E. Roehner et G. Matziolis, “Regional differences in temperature course after knee arthroplasty,” *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 24, n<sup>o</sup>. 8, p. 2686–2691, août 2016.
- [40] S. Kacmaz, E. Ercelebi, S. Zengin et S. Cindoruk, “The use of infrared thermal imaging in the diagnosis of deep vein thrombosis,” *Infrared Physics & Technology*, vol. 86, p. 120–129, nov. 2017.
- [41] Y. Zeng, W. Feng, X. Qi, J. Li, J. Chen, L. Lu, P. Deng, J. Zeng et F. Li, “Differential knee skin temperature following total knee arthroplasty and its relationship with serum indices and outcome : A prospective study,” *The Journal of International Medical Research*, vol. 44, n<sup>o</sup>. 5, p. 1023–1033, oct. 2016.
- [42] R. Sharma, J. Werle, M. Clark, S. Puloski, R. Kuchinad, K. Johnston, J. I. Stewart, G. Abelseth, R. Ng, J. Kubik, A. Meldrum, J. N. Powell, R. Sharma, M. J. Monument, D. Joly, N. Desy, E. Batuyong, S. A. Timmermann et E. Rendall, “Skin temperature following total knee arthroplasty : A longitudinal observational study,” *The Journal of Arthroplasty*, p. S088354032400576X, juin 2024.
- [43] P. Napoletano, *Hand-Crafted vs Learned Descriptors for Color Texture Classification*. Springer International Publishing, 2017, p. 259–271. [En ligne]. Disponible : [http://dx.doi.org/10.1007/978-3-319-56010-6\\_22](http://dx.doi.org/10.1007/978-3-319-56010-6_22)



- [44] L. Armi et S. Fekri-Ershad, “Texture image analysis and texture classification methods - a review,” 2019.
- [45] R. M. Haralick, K. Shanmugam et I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, n°. 6, p. 610–621, 1973.
- [46] F. Bianconi, A. Fernández, F. Smeraldi et G. Pascoletti, “Colour and texture descriptors for visual recognition : A historical overview,” *Journal of Imaging*, vol. 7, n°. 11, p. 245, nov. 2021.
- [47] R. Bello-Cerezo, F. Bianconi, F. D. Maria, P. Napoletano et F. Smeraldi, “Comparative evaluation of hand-crafted image descriptors vs. off-the-shelf cnn-based features for colour texture classification under ideal and realistic conditions,” *Applied Sciences*, 2019. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:85508982>
- [48] N. Dalal et B. Triggs, “Histograms of oriented gradients for human detection,” dans *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 1. IEEE, 2005, p. 886–893. [En ligne]. Disponible : <http://dx.doi.org/10.1109/cvpr.2005.177>
- [49] M. Pietikäinen et G. Zhao, *Two decades of local binary patterns*. Elsevier, 2015, p. 175–210. [En ligne]. Disponible : <http://dx.doi.org/10.1016/b978-0-12-802806-3.00009-9>
- [50] Y. Bengio, A. C. Courville et P. Vincent, “Unsupervised feature learning and deep learning : A review and new perspectives,” *CoRR*, vol. abs/1206.5538, 2012. [En ligne]. Disponible : <http://arxiv.org/abs/1206.5538>
- [51] F. Leeb, S. Bauer, M. Besserve et B. Schölkopf, “Exploring the latent space of autoencoders with interventional assays,” 2023. [En ligne]. Disponible : <https://arxiv.org/abs/2106.16091>
- [52] I. Jolliff, *Principal Component Analysis*, ser. Springer Series in Statistics. Springer, 1986. [En ligne]. Disponible : <https://link.springer.com/book/10.1007/b98835>
- [53] L. van der Maaten et G. Hinton, “Visualizing data using t-sne,” *Journal of Machine Learning Research*, vol. 9, n°. 86, p. 2579–2605, 2008. [En ligne]. Disponible : <https://jmlr.org/papers/v9/vandermaaten08a.html>
- [54] L. McInnes, J. Healy et J. Melville, “Umap : Uniform manifold approximation and projection for dimension reduction,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/1802.03426>
- [55] D. E. Rumelhart, G. E. Hinton et R. J. Williams, *Learning internal representations by error propagation*. Cambridge, MA, USA : MIT Press, 1986, p. 318–362.

- [56] D. Bank, N. Koenigstein et R. Giryes, “Autoencoders,” 2021. [En ligne]. Disponible : <https://arxiv.org/abs/2003.05991>
- [57] Y. LeCun et I. Misra, “Self-supervised learning : The dark matter of intelligence,” <https://ai.meta.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/>, mars 2021, accessed : 2024-09-11.
- [58] R. Balestrieri, M. Ibrahim, V. Sobal, A. Morcos, S. Shekhar, T. Goldstein, F. Bordes, A. Bardes, G. Mialon, Y. Tian, A. Schwarzschild, A. G. Wilson, J. Geiping, Q. Garrido, P. Fernandez, A. Bar, H. Pirsiavash, Y. LeCun et M. Goldblum, “A cookbook of self-supervised learning,” 2023. [En ligne]. Disponible : <https://arxiv.org/abs/2304.12210>
- [59] K. He, H. Fan, Y. Wu, S. Xie et R. Girshick, “Momentum contrast for unsupervised visual representation learning,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/1911.05722>
- [60] T. Chen, S. Kornblith, M. Norouzi et G. Hinton, “A simple framework for contrastive learning of visual representations,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/2002.05709>
- [61] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. H. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. D. Guo, M. G. Azar, B. Piot, K. Kavukcuoglu, R. Munos et M. Valko, “Bootstrap your own latent : A new approach to self-supervised learning,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/2006.07733>
- [62] X. Chen et K. He, “Exploring simple siamese representation learning,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/2011.10566>
- [63] D. P. Kingma et M. Welling, “Auto-encoding variational bayes,” *CoRR*, vol. abs/1312.6114, 2013. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:216078090>
- [64] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville et Y. Bengio, “Generative adversarial networks,” 2014. [En ligne]. Disponible : <https://arxiv.org/abs/1406.2661>
- [65] P. Isola, J.-Y. Zhu, T. Zhou et A. A. Efros, “Image-to-image translation with conditional adversarial networks,” 2018. [En ligne]. Disponible : <https://arxiv.org/abs/1611.07004>
- [66] J.-Y. Zhu, T. Park, P. Isola et A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/1703.10593>
- [67] S. Tripathy, J. Kannala et E. Rahtu, “Learning image-to-image translation using paired and unpaired training samples,” 2018. [En ligne]. Disponible :

<https://arxiv.org/abs/1805.03189>

- [68] M. Mirza et S. Osindero, “Conditional generative adversarial nets,” 2014. [En ligne]. Disponible : <https://arxiv.org/abs/1411.1784>
- [69] K. Armanious, C. Jiang, S. Abdulatif, T. Kustner, S. Gatidis et B. Yang, “Unsupervised medical image translation using cycle-medgan,” dans *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, sept. 2019. [En ligne]. Disponible : <http://dx.doi.org/10.23919/EUSIPCO.2019.8902799>
- [70] P. Iacono et N. Khan, “Structure preserving cycle-gan for unsupervised medical image domain adaptation,” 2023. [En ligne]. Disponible : <https://arxiv.org/abs/2304.09164>
- [71] C. Czobit et R. Samavi, “Cyclegan models for mri image translation,” 2024. [En ligne]. Disponible : <https://arxiv.org/abs/2401.00023>
- [72] A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman et A. Courville, “Augmented cyclegan : Learning many-to-many mappings from unpaired data,” 2018. [En ligne]. Disponible : <https://arxiv.org/abs/1802.10151>
- [73] T. Karras, S. Laine et T. Aila, “A style-based generator architecture for generative adversarial networks,” mars 2019, arXiv :1812.04948 [cs, stat]. [En ligne]. Disponible : <http://arxiv.org/abs/1812.04948>
- [74] T. Karras, T. Aila, S. Laine et J. Lehtinen, “Progressive growing of gans for improved quality, stability, and variation,” 2018. [En ligne]. Disponible : <https://arxiv.org/abs/1710.10196>
- [75] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen et T. Aila, “Analyzing and improving the image quality of stylegan,” mars 2020, arXiv :1912.04958 [cs, eess, stat]. [En ligne]. Disponible : <http://arxiv.org/abs/1912.04958>
- [76] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen et T. Aila, “Training generative adversarial networks with limited data,” oct. 2020, arXiv :2006.06676 [cs, stat]. [En ligne]. Disponible : <http://arxiv.org/abs/2006.06676>
- [77] A. Sauer, K. Schwarz et A. Geiger, “Stylegan-xl : Scaling stylegan to large diverse datasets,” 2022. [En ligne]. Disponible : <https://arxiv.org/abs/2202.00273>
- [78] D. Epstein, T. Park, R. Zhang, E. Shechtman et A. A. Efros, “Blobgan : Spatially disentangled scene representations,” 2022. [En ligne]. Disponible : <https://arxiv.org/abs/2205.02837>
- [79] X. Liu, P. Sanchez, S. Thermos, A. Q. O’Neil et S. A. Tsaftaris, “Learning disentangled representations in the imaging domain,” *Medical Image Analysis*, vol. 80, p. 102516, août 2022.

- [80] H. Hojjati, T. K. K. Ho et N. Armanfard, “Self-supervised anomaly detection in computer vision and beyond : A survey and outlook,” *Neural Networks*, vol. 172, p. 106106, avr. 2024.
- [81] M. Nielsen, L. Wenderoth, T. Sentker et R. Werner, “Self-supervision for medical image classification : State-of-the-art performance with 100 labeled training samples per class,” *Bioengineering*, vol. 10, n°. 8, p. 895, juill. 2023.
- [82] A. Kalapos et B. Gyires-Tóth, “Self-supervised pretraining for 2d medical image segmentation,” 2022. [En ligne]. Disponible : <https://arxiv.org/abs/2209.00314>
- [83] L. Morra, L. Piano, F. Lamberti et T. Tommasi, “Bridging the gap between natural and medical images through deep colorization,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/2005.10589>
- [84] P. Chatterjee et A. Zaman, “Thermal face image classification using deep learning techniques,” dans *2023 International Conference on Computational Science and Computational Intelligence (CSCI)*. Las Vegas, NV, USA : IEEE, déc. 2023, p. 1208–1213. [En ligne]. Disponible : <https://ieeexplore.ieee.org/document/10590085/>
- [85] B. Miethig, A. Liu, S. Habibi et M. v. Mohrenschildt, “Leveraging thermal imaging for autonomous driving,” dans *2019 IEEE Transportation Electrification Conference and Expo (ITEC)*, 2019, p. 1–5.
- [86] K. Yang, W. Xiang, Z. Chen, J. Zhang et Y. Liu, “A review on infrared and visible image fusion algorithms based on neural networks,” *Journal of Visual Communication and Image Representation*, vol. 101, p. 104179, 2024. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S1047320324001342>
- [87] U. Thirunavukkarasu, S. Umapathy, V. Ravi et T. J. Alahmadi, “Tongue image fusion and analysis of thermal and visible images in diabetes mellitus using machine learning techniques,” *Scientific Reports*, vol. 14, n°. 1, juin 2024. [En ligne]. Disponible : <http://dx.doi.org/10.1038/s41598-024-64150-0>
- [88] A. N. Wilson, K. A. Gupta, B. H. Koduru, A. Kumar, A. Jha et L. R. Cenkeramaddi, “Recent advances in thermal imaging and its applications using machine learning : A review,” *IEEE Sensors Journal*, vol. 23, n°. 4, p. 3395–3407, févr. 2023.
- [89] M. Benčević, M. Habijan, I. Galić, D. Babin et A. Pižurica, “Understanding skin color bias in deep learning-based skin lesion segmentation,” *Computer Methods and Programs in Biomedicine*, vol. 245, p. 108044, mars 2024. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.cmpb.2024.108044>
- [90] T. P. Pagano, R. B. Loureiro, F. V. N. Lisboa, G. O. R. Cruz, R. M. Peixoto, G. A. de Sousa Guimarães, L. L. dos Santos, M. M. Araujo, M. Cruz, E. L. S.

- de Oliveira, I. Winkler et E. G. S. Nascimento, “Bias and unfairness in machine learning models : a systematic literature review,” 2022. [En ligne]. Disponible : <https://arxiv.org/abs/2202.08176>
- [91] J. Verstockt, S. Verspeek, F. Thiessen, W. A. Tjalma, L. Brochez et G. Steenackers, “Skin cancer detection using infrared thermography : Measurement setup, procedure and equipment,” *Sensors*, vol. 22, n<sup>o</sup>. 9, p. 3327, avr. 2022. [En ligne]. Disponible : <http://dx.doi.org/10.3390/s22093327>
- [92] W. Sultani, C. Chen et M. Shah, “Real-world anomaly detection in surveillance videos,” 2019. [En ligne]. Disponible : <https://arxiv.org/abs/1801.04264>
- [93] D. A. Bierbrauer, A. Chang, W. Kritzer et N. D. Bastian, “Cybersecurity anomaly detection in adversarial environments,” 2021. [En ligne]. Disponible : <https://arxiv.org/abs/2105.06742>
- [94] I. Zingman, B. Stierstorfer, C. Lempp et F. Heinemann, “Learning image representations for anomaly detection : Application to discovery of histological alterations in drug development,” *Medical Image Analysis*, vol. 92, p. 103067, févr. 2024. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.media.2023.103067>
- [95] G. Pang, C. Shen, L. Cao et A. V. D. Hengel, “Deep learning for anomaly detection : A review,” *ACM Computing Surveys*, vol. 54, n<sup>o</sup>. 2, p. 1–38, mars 2022.
- [96] S.-C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung et A. S. Chaudhari, “Self-supervised learning for medical image classification : a systematic review and implementation guidelines,” *npj Digital Medicine*, vol. 6, n<sup>o</sup>. 1, avr. 2023. [En ligne]. Disponible : <http://dx.doi.org/10.1038/s41746-023-00811-0>
- [97] Y. Cai, W. Zhang, H. Chen et K.-T. Cheng, “Medianomaly : A comparative study of anomaly detection in medical images,” 2024. [En ligne]. Disponible : <https://arxiv.org/abs/2404.04518>
- [98] H. H. Nguyen, C. N. Nguyen, X. T. Dao, Q. T. Duong, D. P. T. Kim et M.-T. Pham, “Variational autoencoder for anomaly detection : A comparative study,” 2024. [En ligne]. Disponible : <https://arxiv.org/abs/2408.13561>
- [99] M. Esmaili, A. Toosi, A. Roshanpoor, V. Changizi, M. Ghazisaeedi, A. Rahmim et M. Sabokrou, “Generative adversarial networks for anomaly detection in biomedical imaging : A study on seven medical image datasets,” *IEEE Access*, vol. 11, p. 17906–17921, 2023.
- [100] J. Wyatt, A. Leach, S. M. Schmon et C. G. Willcocks, “Anoddpm : Anomaly detection with denoising diffusion probabilistic models using simplex noise,” dans *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CV-PRW)*, juin 2022, p. 649–655.

- [101] G. Pang, C. Shen, H. Jin et A. van den Hengel, “Deep weakly-supervised anomaly detection,” 2023. [En ligne]. Disponible : <https://arxiv.org/abs/1910.13601>
- [102] A. A. Neloy et M. Turgeon, “A comprehensive study of auto-encoders for anomaly detection : Efficiency and trade-offs,” *Machine Learning with Applications*, vol. 17, p. 100572, sept. 2024. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.mlwa.2024.100572>
- [103] C. Baur, S. Denner, B. Wiestler, N. Navab et S. Albarqouni, “Autoencoders for unsupervised anomaly segmentation in brain mr images : A comparative study,” *Medical Image Analysis*, vol. 69, p. 101952, avr. 2021. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.media.2020.101952>
- [104] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth et G. Langs, “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” mars 2017, arXiv :1703.05921 [cs]. [En ligne]. Disponible : <http://arxiv.org/abs/1703.05921>
- [105] S. Akcay, A. Atapour-Abarghouei et T. P. Breckon, “Ganomaly : Semi-supervised anomaly detection via adversarial training,” nov. 2018, arXiv :1805.06725 [cs]. [En ligne]. Disponible : <http://arxiv.org/abs/1805.06725>
- [106] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs et U. Schmidt-Erfurth, “f-anogan : Fast unsupervised anomaly detection with generative adversarial networks,” *Medical Image Analysis*, vol. 54, p. 30–44, mai 2019.
- [107] S. Park, K. H. Lee, B. Ko et N. Kim, “Unsupervised anomaly detection with generative adversarial networks in mammography,” *Scientific Reports*, vol. 13, n°. 1, févr. 2023. [En ligne]. Disponible : <http://dx.doi.org/10.1038/s41598-023-29521-z>
- [108] S. Bond-Taylor, A. Leach, Y. Long et C. G. Willcocks, “Deep generative modelling : A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, n°. 11, p. 7327–7347, nov. 2022. [En ligne]. Disponible : <http://dx.doi.org/10.1109/TPAMI.2021.3116668>
- [109] J. Wolleb, F. Bieder, R. Sandkuhler et P. C. Cattin, “Diffusion models for medical anomaly detection,” 2022.
- [110] M. Monemian, M. Irajpour et H. Rabbani, “A review on texture-based methods for anomaly detection in retinal optical coherence tomography images,” *Optik*, vol. 288, p. 171165, oct. 2023. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.ijleo.2023.171165>
- [111] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri et R. M. Summers, “Chestx-ray8 : Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification

- and localization of common thorax diseases,” dans *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, juill. 2017. [En ligne]. Disponible : <http://dx.doi.org/10.1109/CVPR.2017.369>
- [112] L. Zhou, J. Bae, H. Liu, G. Singh, J. Green, D. Samaras et P. Prasanna, *Chest Radiograph Disentanglement for COVID-19 Outcome Prediction*. Springer International Publishing, 2021, p. 345–355. [En ligne]. Disponible : [http://dx.doi.org/10.1007/978-3-030-87234-2\\_33](http://dx.doi.org/10.1007/978-3-030-87234-2_33)
- [113] L. Hogeweg, C. I. Sánchez, P. Maduskar, R. H. Philipsen et B. van Ginneken, “Fast and effective quantification of symmetry in medical images for pathology detection : Application to chest radiography,” *Medical Physics*, vol. 44, n<sup>o</sup>. 6, p. 2242–2256, mai 2017. [En ligne]. Disponible : <http://dx.doi.org/10.1002/mp.12127>
- [114] D. Choi, L. Sunwoo, S.-H. You, K. J. Lee et I. Ryoo, “Application of symmetry evaluation to deep learning algorithm in detection of mastoiditis on mastoid radiographs,” *Scientific Reports*, vol. 13, n<sup>o</sup>. 1, p. 5337, avr. 2023.
- [115] A. Mencattini, M. Salmeri et P. Casti, “Bilateral asymmetry identification for the early detection of breast cancer,” dans *2011 IEEE International Symposium on Medical Measurements and Applications*, vol. 53. IEEE, mai 2011, p. 613–618. [En ligne]. Disponible : <http://dx.doi.org/10.1109/MeMeA.2011.5966746>
- [116] S. Hennessey, E. Huszti, A. Gunasekura, A. Salleh, L. Martin, S. Minkin, S. Chavez et N. F. Boyd, “Bilateral symmetry of breast tissue composition by magnetic resonance in young women and adults,” *Cancer Causes and Control*, vol. 25, n<sup>o</sup>. 4, p. 491–497, janv. 2014. [En ligne]. Disponible : <http://dx.doi.org/10.1007/s10552-014-0351-0>
- [117] D. Shimokawa, K. Takahashi, D. Kurosawa, E. Takaya, K. Oba, K. Yagishita, T. Fukuda, H. Tsunoda et T. Ueda, “Deep learning model for breast cancer diagnosis based on bilateral asymmetrical detection (bilad) in digital breast tomosynthesis images,” *Radiological Physics and Technology*, vol. 16, n<sup>o</sup>. 1, p. 20–27, nov. 2022. [En ligne]. Disponible : <http://dx.doi.org/10.1007/s12194-022-00686-y>
- [118] A. Dey, E. Ali et S. Rajan, “Bilateral symmetry-based abnormality detection in breast thermograms using textural features of hot regions,” *IEEE Open Journal of Instrumentation and Measurement*, vol. 2, p. 1–14, 2023.
- [119] F. Chollet, “Xception : Deep learning with depthwise separable convolutions,” 2016. [En ligne]. Disponible : <https://arxiv.org/abs/1610.02357>
- [120] I. Fernandez-Cuevas, J. C. Bouzas Marins, J. Arnaiz Lastras, P. M. Gomez Carmona, S. Pinonosa Cano, M. A. Garcia-Concepcion et M. Sillero-Quintana, “Classification of factors influencing the use of infrared thermography in humans : A review,”

- Infrared Physics and Technology*, vol. 71, p. 28–55, juill. 2015. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.infrared.2015.02.007>
- [121] M. Tkachenko, M. Malyuk, A. Holmanyuk et N. Liubimov, “Label Studio : Data labeling software,” 2020-2024, open source software available from <https://github.com/HumanSignal/label-studio>. [En ligne]. Disponible : <https://github.com/HumanSignal/label-studio>
- [122] S. V. Francis et M. Sasikala, “Automatic detection of abnormal breast thermograms using asymmetry analysis of texture features,” *Journal of Medical Engineering and Technology*, vol. 37, n<sup>o</sup>. 1, p. 17–21, nov. 2012. [En ligne]. Disponible : <http://dx.doi.org/10.3109/03091902.2012.728674>
- [123] M. Abdel-Nasser, A. Moreno et D. Puig, “Breast cancer detection in thermal infrared images using representation learning and texture analysis methods,” *Electronics*, vol. 8, n<sup>o</sup>. 1, p. 100, janv. 2019. [En ligne]. Disponible : <http://dx.doi.org/10.3390/electronics8010100>
- [124] D. A. Story, “Postoperative complications in elderly patients and their significance for long-term prognosis,” *Current Opinion in Anaesthesiology*, vol. 21, n<sup>o</sup>. 3, p. 375–379, juin 2008. [En ligne]. Disponible : <http://dx.doi.org/10.1097/aco.0b013e3282f889f8>
- [125] V. Médard de Chardon, T. Balaguer, B. Chignon-Sicard, T. Ihrai et E. Lebreton, “Les asymétries constitutionnelles en chirurgie d’augmentation mammaire esthétique : incidence, satisfaction et applications chirurgicales,” *Annales de Chirurgie Plastique Esthétique*, vol. 54, n<sup>o</sup>. 4, p. 340–347, août 2009. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.anplas.2009.01.034>
- [126] R. M. Haralick, K. Shanmugam et I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, n<sup>o</sup>. 6, p. 610–621, nov. 1973. [En ligne]. Disponible : <http://dx.doi.org/10.1109/tsmc.1973.4309314>
- [127] K. He, X. Zhang, S. Ren et J. Sun, “Deep residual learning for image recognition,” dans *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, p. 770–778.
- [128] T. Park, A. A. Efros, R. Zhang et J.-Y. Zhu, “Contrastive learning for unpaired image-to-image translation,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/2007.15651>
- [129] D. P. Kingma et J. Ba, “Adam : A method for stochastic optimization,” 2014. [En ligne]. Disponible : <https://arxiv.org/abs/1412.6980>
- [130] K. Roth, A. Lucchi, S. Nowozin et T. Hofmann, “Stabilizing training of generative adversarial networks through regularization,” 2017. [En ligne]. Disponible : <https://arxiv.org/abs/1705.09367>



- [131] C. Shorten et T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, n°. 1, juill. 2019. [En ligne]. Disponible : <http://dx.doi.org/10.1186/s40537-019-0197-0>
- [132] H. Zhang, Z. Zhang, A. Odena et H. Lee, “Consistency regularization for generative adversarial networks,” 2019. [En ligne]. Disponible : <https://arxiv.org/abs/1910.12027>
- [133] Z. Zhao, S. Singh, H. Lee, Z. Zhang, A. Odena et H. Zhang, “Improved consistency regularization for gans,” 2020. [En ligne]. Disponible : <https://arxiv.org/abs/2002.04724>
- [134] O. Ronneberger, P. Fischer et T. Brox, “U-net : Convolutional networks for biomedical image segmentation,” dans *Medical image computing and computer-assisted intervention–MICCAI 2015 : 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, p. 234–241.
- [135] L. B. de Amorim, G. D. Cavalcanti et R. M. Cruz, “The choice of scaling technique matters for classification performance,” *Applied Soft Computing*, vol. 133, p. 109924, janv. 2023. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.asoc.2022.109924>
- [136] F. T. Liu, K. M. Ting et Z.-H. Zhou, “Isolation forest,” dans *2008 Eighth IEEE International Conference on Data Mining*. IEEE, déc. 2008, p. 413–422. [En ligne]. Disponible : <http://dx.doi.org/10.1109/icdm.2008.17>
- [137] M. Tan et Q. V. Le, “Efficientnet : Rethinking model scaling for convolutional neural networks,” 2019. [En ligne]. Disponible : <https://arxiv.org/abs/1905.11946>
- [138] P. Perera, P. Oza et V. M. Patel, “One-class classification : A survey,” 2021. [En ligne]. Disponible : <https://arxiv.org/abs/2101.03064>
- [139] B. H. van der Velden, H. J. Kuijf, K. G. Gilhuijs et M. A. Viergever, “Explainable artificial intelligence (xai) in deep learning-based medical image analysis,” *Medical Image Analysis*, vol. 79, p. 102470, juill. 2022. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.media.2022.102470>
- [140] O. Lang, D. Yaya-Stupp, I. Traynis, H. Cole-Lewis, C. R. Bennett, C. R. Lyles, C. Lau, M. Irani, C. Semturs, D. R. Webster, G. S. Corrado, A. Hassidim, Y. Matias, Y. Liu, N. Hammel et B. Babenko, “Using generative ai to investigate medical imagery models and datasets,” *eBioMedicine*, vol. 102, p. 105075, avr. 2024. [En ligne]. Disponible : <http://dx.doi.org/10.1016/j.ebiom.2024.105075>
- [141] N. van der Vleuten, T. Radusinović, R. Akkerman et M. Reksoprodjo, “[Re] Explaining in Style : Training a GAN to explain a classifier in StyleSpace,” *ReScience C*, vol. 8, n°. 2, p. 42, mai 2022. [En ligne]. Disponible : <https://zenodo.org/record/6574709/files/article.pdf>

## ANNEXE A DMR-IR : SENSIBILITÉ DE L'ESPACE D'ASYMÉTRIES AUX CHANGEMENTS DE TEMPÉRATURE

Pour comprendre l'efficacité des différences de vecteurs de textures contenus à modéliser les changements de température, effectuée une analyse en composantes principales sur ceux provenant de cas sains de l'ensemble de test et on les compare à leur copie dont un seul côté a été artificiellement réchauffé en ajoutant une valeur  $R$  entre 0 et 1. Cette expérience est effectuée sans désenchevêtrement avec l'auto-encodeur AE et avec TADA-SAE. Une différence marquée dans les représentations avec et sans augmentation de température permettra de tirer des conclusions sur la sensibilité des différences de vecteurs de texture et comprendre l'efficacité de la classification par la tête linéaire et la SVM. On choisit un réchauffement  $R = 0.05$  sur les valeurs des pixels normalisées entre 0 et 1. La figure A.1 montre la subtilité de l'augmentation artificielle de la température que les modèles devront détecter.

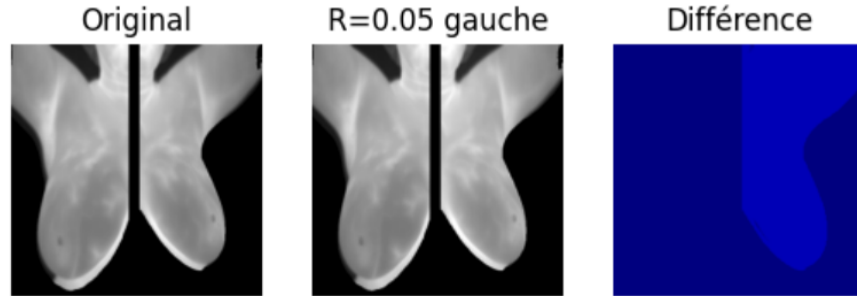


FIGURE A.1 Réchauffement artificiel unilatéral du sein gauche avec  $R = 0.05$ .

## ANNEXE B DMR-IR : INTERPOLATION DE TEXTURES

Sachant que la température du membre opéré diminue graduellement avec le temps, on aimerait utiliser les capacités de génération d’images du *Swapping Autoencoder* afin d’intervertir les températures des côtés sains et positifs. Il devient ainsi possible de comparer l’image thermique du sein avec son homologue ayant reçu les nouvelles textures du sein opposé, et vice-versa. Une différence absolue entre les deux images permet de déterminer à quel point les images diffèrent les unes des autres. On assume ainsi que les différences entre les images avec et sans interpolation d’une patiente auront des différences moins marquées que pour une patiente ayant une hyperthermie due au cancer.

D’abord, la figure B.1 montre comment le modèle de SAE arrive à reconstruire et interpoler les textures de deux images de seins gauche et droit provenant de la même patiente. Celle-ci a été diagnostiquée avec une tumeur au sein droit. On observe ainsi comment la région plus chaude a été transférée du sein droit au sein gauche, et ce premier est maintenant sans hyperthermie. On peut visualiser ces différences en calculant l’erreur au carré pour chaque pixel entre chaque sein tel que montré à la figure B.2.

On relève ainsi les différences de température entre les reconstructions et interpolations pour chaque sein. Bien qu’une petite région au-dessus de l’épaule contienne une plus grande erreur due au transfert de textures, la majorité des différences thermiques importantes se trouvent au niveau du mamelon pour cette patiente. Suivant ce principe, on peut calculer les erreurs au carré moyennes pour chaque image de chaque patiente, puis seuiller celles-ci pour obtenir une classification. Les résultats sont énoncés dans le tableau B.1.

TABLEAU B.1 Résultats de la classification par erreur d’interpolation sur l’ensemble de validation de DMR-IR

<b>F1</b>	<b>Préc.</b>	<b>Sens.</b>	<b>AUROC</b>
0.766	0.684	0.870	0.703

Bien qu’il soit possible de classer les exemplaires avec cette méthode, ses résultats sont fortement inférieurs à ceux de TADA-SAE avec SVM sur toutes les métriques. En effet, cette technique utilisant les interpolations assume que les deux côtés sont normalement parfaitement symétriques, ce qui en réalité n’est pas vraiment le cas. Les petites structures veineuses visibles dans l’image thermique sont vues comme des erreurs lors de l’interpolation. Il est donc nécessaire de recourir à une technique moins naïve qu’une méthode par seuillage pour une détection efficace.

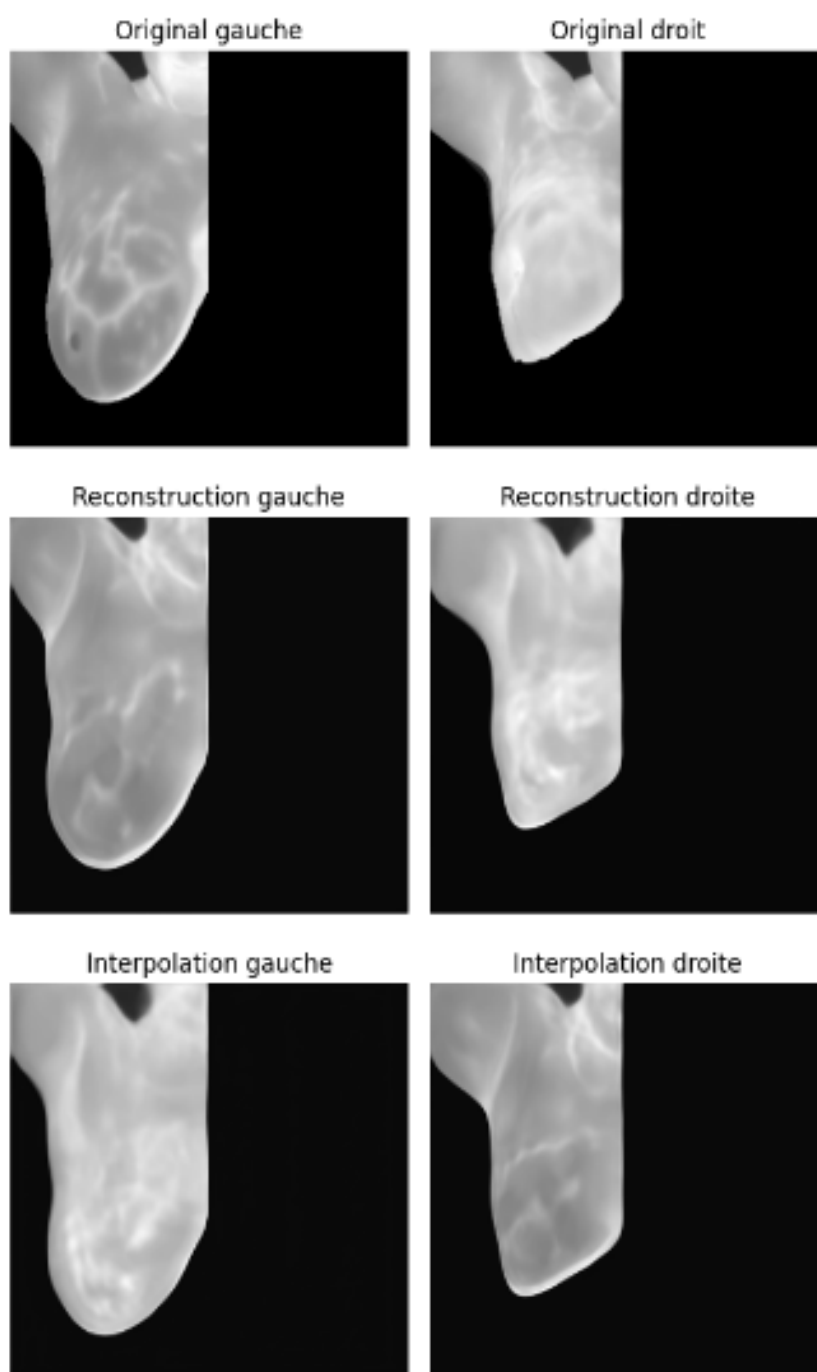


FIGURE B.1 Reconstruction et interpolation de textures dans les images thermiques des seins d'une patiente positive. Le sein droit affecté est plus chaud que le sein opposé, et le modèle arrive à intervertir les températures de ceux-ci.

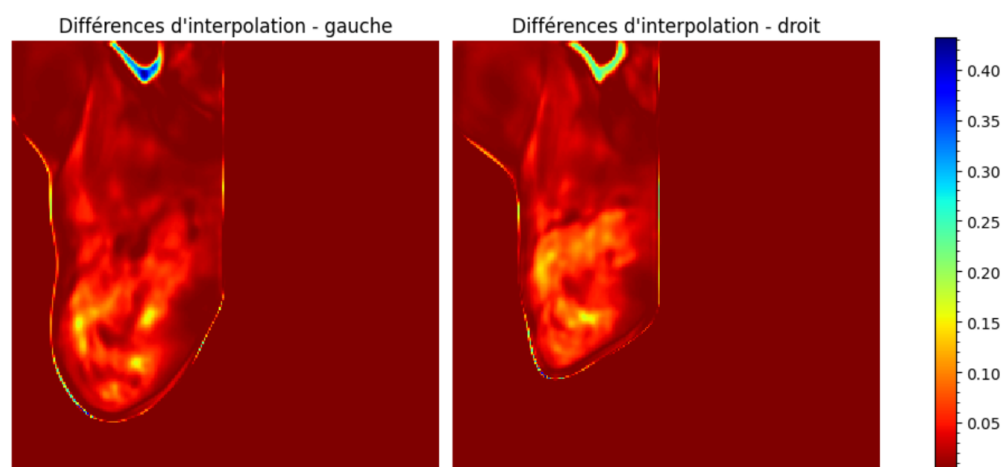


FIGURE B.2 Erreur au carré entre les pixels des images reconstruites et des interpolations de texture avec le sein opposé. Dans les deux cas, les différences de température sont mises en évidence dans la région mammaire.

## ANNEXE C ORTHOPOT : DÉTECTION D'ANOMALIES PAR TYPE DE CHIRURGIE

Sachant que l'ensemble de données comporte des ATH et ATG, on tente de déterminer les capacités de détection d'anomalies sur ces deux cas se manifestant de manière différente sur l'image thermique. En effet, suivant une ATH, l'augmentation de température se manifeste au niveau de la hanche, vue de côté par la caméra. Pour les ATG, l'hyperthermie apparaît plutôt au niveau du genou opéré, près du centre de l'image. Afin de guider le processus d'isolation de l'algorithme *iForest*, on conserve un exemplaire anormal du type de chirurgie concerné pour son entraînement. Comme stratégie de validation, on exécute 10 entraînements et validations sur le type de chirurgie concerné avant de moyenner les vrais et faux positifs puis vrais et faux négatifs dans une matrice de confusion. Les figures C.1 et C.2 montrent les matrices de confusion de TADA-SAE pour la détection d'anomalies dans les ATH et ATG respectivement.

Les résultats obtenus pour cette matrice de confusion montrent que le modèle arrive très bien à identifier les cas anormaux dans les ATG. Cependant, celui-ci est très sensible et a un nombre de faux positifs moyen élevé sans faux négatif. Néanmoins, le modèle réussit à classer correctement la majorité des cas négatifs.

Le modèle réussit d'autre part à détecter correctement la présence d'anomalies dans tous les cas d'ATH de l'ensemble de validation. Bien que ces performances semblent impressionnantes, elles sont à prendre avec un grain de sel. En effet, le biais relevé à la figure 4.9 peut expliquer de telles performances.

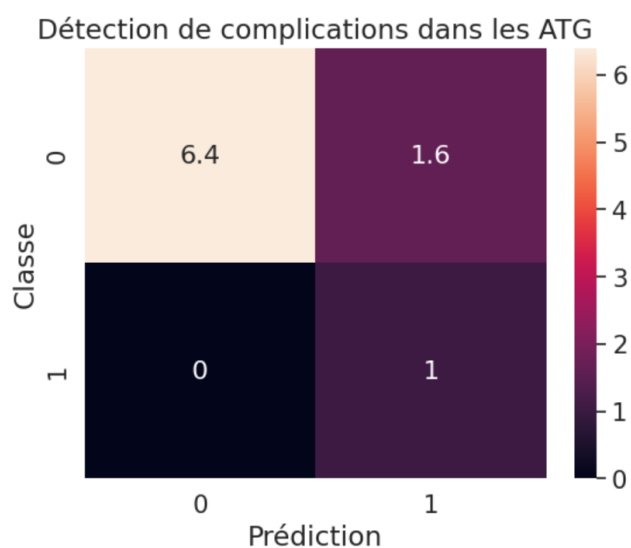


FIGURE C.1 Matrice de confusion obtenue sur l'ensemble de validation avec TADA-SAE pour la détection de complications dans les ATG.

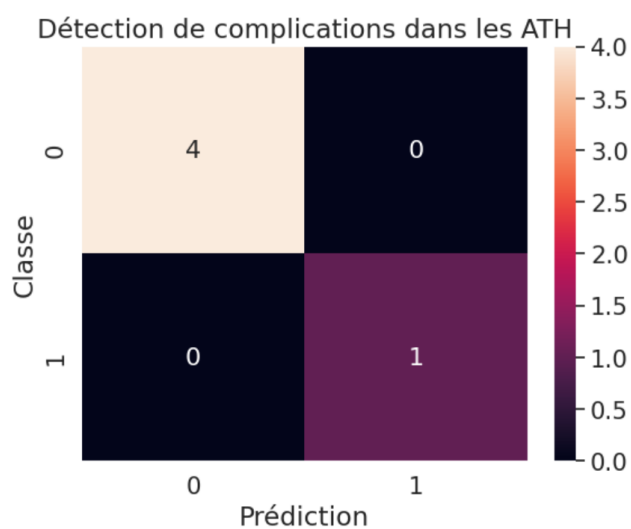


FIGURE C.2 Matrice de confusion obtenue sur l'ensemble de validation avec TADA-SAE pour la détection de complications dans les ATH