



## **Master 2 : Modélisation et Sciences des Données**

---

### **Projet du Semestre 3**

**Applied deep learning techniques for the  
classification of multimodal breast cancer  
screening data**

#### **Projet Réalisé par :**

**Aurel Davy TCHOKPONHOUE & Ilham AARAB**

#### **Sous la Supervision de :**

**Prof. ALI IDRI**

**Asma ZIZAAN**

**Fatima Zahra NAKACH**

**ALKHAWARIZMI Department**

**2022/2023**

## Résumé

Le diagnostic précoce du cancer du sein contribue considérablement à la réduction du taux de mortalité qui lui est imputable en permettant l'adoption d'un traitement efficace. Cette étude a développé et évalué neuf modèles (DenseNet 201, Inception V3, Inception ReseNet V2, MobileNet V2, ResNet 50, VGG16 et VGG19, Xception, NASNetMobile) de Deep Learning sur les données images et cinq classificateurs (KNN, SVC, Decision Tree, Random Forest et MLP) sur les données cliniques pour ensuite créer des techniques de Late fusion multimodality combinant ces architectures monomodales. Les architectures conçues ont été évaluées en utilisant : (1) cinq critères de performance de classification (Accuracy, Precision, Recall, F1-score et AUC), (2) le test statistique de Scott Knott (SK) pour regrouper les architectures proposées et identifier le meilleur regroupement des architectures performantes, et (3) la méthode de vote du Borda count pour classer les architectures les plus performantes. Les résultats ont montré le potentiel de la combinaison des techniques d'apprentissage profond sur les données images et des classifieurs classiques sur les données cliniques pour le dépistage du cancer du sein. L'architecture de Late fusion utilisant le classificateur KNN et DenseNet201 a été la plus performante avec des valeurs de précision plus élevées atteignant 82% sur l'ensemble des données. Les résultats de cette étude recommandent l'utilisation d'architectures de fusion utilisant DenseNet 201 pour les données de mammographie et le KNN pour les données cliniques, car il a donné les meilleurs résultats pour suite à la fusion.

*Mots clés : Late fusion, multimodalité, détection du cancer du sein, Deep Learning, Réseau de neurones convolutif, Transfert Learning, Machine Learning, mammographie*

## 1. Introduction

Le cancer du sein est devenu le type de cancer le plus fréquemment diagnostiqué dans le monde [1]. En 2020, il y a eu plus de 2,26 millions de nouveaux cas de cancer du sein et près de 685 000 décès dus au cancer du sein dans le monde[1] [2]. Il est la cause la plus fréquente de décès par cancer chez les femmes[3]. En raison du manque de clarté des causes et de l'efficacité de la prévention du cancer du sein, les cas de cancer du sein n'ont pas été réduits de manière significative [4]. Cependant, le diagnostic et le traitement précoces du cancer du sein sont devenus primordiaux et indispensables pour améliorer le taux de guérison et ainsi impacter l'incidence des décès qui lui sont imputables. Par conséquent, l'établissement de méthodes de dépistage fiables est essentiel pour le diagnostic précoce du cancer du sein [5].

Ces dernières années, les méthodes d'apprentissage profond ont obtenu des résultats remarquables dans le domaine de la vision par ordinateur, ce qui a incité de nombreux chercheurs à transférer cette méthode à l'analyse des images pathologiques. Cette adoption de l'apprentissage profond à travers les différentes architectures mises en place a permis d'obtenir des performances notables. Malgré cela, la précision de la classification du cancer du sein bénin et malin en utilisant uniquement des données d'images pathologiques monomodales ne répond pas encore totalement aux exigences de la pratique clinique [9].

Bien que les images pathologiques représentent une mine d'information, les modèles ne considérant que les données de pixels comme modalité unique d'entrée ne sont pas en mesure

de contextualiser d'autres informations cliniques telles qu'il est le cas dans la pratique médicale, ce qui peut donc finalement limiter la traduction clinique. En pratique, pour procéder à un diagnostic, les médecins spécialistes et radiologues tiennent compte du contexte clinique et des informations du patient. Ces informations pertinentes et précises concernant ses symptômes actuels et ses antécédents médicaux permettent aux médecins d'interpréter les résultats d'imagerie dans le contexte clinique approprié, ce qui permet d'établir un diagnostic différentiel plus pertinent, un rapport plus utile pour les médecins et un résultat optimal pour le patient [10].

Au regard de ces limites inhérentes aux architectures monomodales, plusieurs récentes études ont commencé à explorer l'utilisation des architectures multimodales dans le domaine médical et notamment dans la détection du cancer du sein avec de résultats assez prometteur, choses qui permettra de développer des systèmes de détection et de classification automatisés pouvant utiliser avec succès à la fois les données d'imagerie médicale et les données cliniques, telles que les données démographiques du patient, les diagnostics antérieurs et les résultats de laboratoire. Ainsi, les modèles d'apprentissages multimodaux peuvent conduire à de meilleures performances, mais également à des conclusions plus pertinentes sur le plan clinique [10], [11].

Dans cette étude, l'objectif est de construire un modèle multimodal de détection du cancer du sein à partir d'images mammographiques et de données cliniques en se basant sur la méthode de Late Fusion et de faire le benchmark avec les modèles monomodaux. Parmi les modèles monomodaux, on retrouve le KNN, le SVC, la Decision Tree, le Random Forest et le MLP pour les données cliniques et les techniques de Deep Learning que sont VGG16, VGG19, InceptionV3, ResNet50, InceptionResNetV2, MobileNet, DenseNet201, Xception et NASNetMobile pour les données de mammographie.

La présente étude explore quatre questions de recherche :

**(QR1) :** Quelle est la performance globale des modèles utilisant les données monomodales dans le dépistage du cancer du sein ? Existe-t-il une architecture monomodale plus performante que les autres ?

**(QR2) :** Quel est l'impact d'une modalité de données sur les performances d'un modèle ?

**(RQ3) :** Quelle est la performance globale des modèles DL utilisant des données multimodales dans le dépistage du cancer du sein ?

**(RQ4) :** Quelle est la performance des architectures de Late fusion par rapport aux modèles monomodales ?

Les principales contributions de cette étude empirique sont les suivantes :

**(1)** le développement de neuf architectures CNN de bout en bout sur les données mammographiques et cinq modèles de machine Learning sur les données cliniques

**(2)** la préparation des données en utilisant la technique d'égalisation adaptative d'histogramme à contraste limité (CLAHE) et la normalisation d'intensité, et enfin

(3) la comparaison des résultats de chaque modèle monomodal aux modèles multimodaux et la sélection du plus performant d'entre eux à travers le test de clustering SK et de la méthode de vote Borda Count.

Notre rapport sera subdivisé en plusieurs sections. La section 2 portera sur les méthodes utilisées et leurs définitions, la section 3 sur la préparation des données, la section 4 sur le design empirique de l'étude, la section 5 sur les résultats et discussion et pour finir une section qui abordera les limites de notre étude qui sera parachever par la conclusion.

## 2. Méthodes et Définitions

Cette section présente une vue d'ensemble des algorithmes de Machine Learning et techniques de Deep Learning (DL) utilisés dans ce travail. Ensuite, nous expliquons la méthode de Late fusion multimodale proposée.

### 2.1. Algorithmes Machine Learning

Depuis les dernières décennies, un riche éventail d'algorithmes d'apprentissage automatique supervisé ont été appliqués et testés par rapport à la tâche du diagnostic du cancer du sein, contribuant ainsi à améliorer la précision de la détection du cancer du sein [19]. Les principaux types d'algorithmes utilisés dans le cadre de la détection du cancer du sein sont les suivants : K-Nearest Neighbour (KNN), Naive Bayes (NB), Support Vector Machine (SVM), les arbres de décision (DT), les forêts aléatoires (RF), les réseaux neuronaux artificiels (ANN). Nous ferons recours à ces algorithmes pour modéliser les données cliniques. Que signifient-ils concrètement ?

**KNN** : La méthode des k-proches voisins (KNN) est souvent utilisée comme référence initiale lors de l'étude d'un ensemble de données sans connaissances préalables [20]. Il s'agit d'un modèle non paramétrique et paresseux, car il n'apprend pas le modèle des données, mais classe un échantillon de test en examinant ses k plus proches voisins [19]. Les plus proches voisins du point de données de l'échantillon sont déterminés en utilisant des métriques de distance tels que la distance euclidienne, qui est la métrique la plus largement utilisée[20].

**Decision Tree** : Contrairement à KNN, l'algorithme de l'arbre de décision est un algorithme simple, mais puissant pour l'ajustement des données. Il fonctionne comme un organigramme mettant en correspondance les attributs et les valeurs des vecteurs de caractéristiques d'entrée des échantillons, créant ainsi un arbre composé de différents types de nœuds. Chaque nœud non-feuille teste l'un des attributs des vecteurs de caractéristiques, se ramifiant vers un nœud plus profond en fonction de la valeur de l'attribut. Une fois qu'un nœud feuille est atteint après plusieurs tests, une décision de classification est prise [21]. Les implémentations les plus populaires des arbres de décision utilisent des mesures d'impureté basées sur l'entropie pour générer l'arbre, comme les algorithmes d'arbre de décision ID3, C4.5 et C5.0 [19]. Un nœud atteint une valeur d'entropie de 0 lorsqu'il est "pur", c'est-à-dire qu'il ne contient que des instances d'une seule classe (uniquement des échantillons de données bénignes ou malignes).

**SVC** : Un SVC est une méthode de classification à marge maximale, qui vise à trouver l'hyperplan qui sépare le plus, deux classes, et elle se base sur des noyaux pour séparer les données non linéaires. Dans le cas de données non linéaires, les données d'apprentissage sont transposées dans une dimension supérieure où elles peuvent être séparées linéairement. Pour ce faire, on utilise l'astuce du noyau, qui fait correspondre le vecteur caractéristique d'entrées à une dimension supérieure en utilisant le produit scalaire, mais n'effectue pas la transformation, ce qui augmenterait de manière exponentielle la taille de l'espace caractéristique et, par conséquent, le temps d'apprentissage [22]. Cette astuce du noyau est ce qui a permis aux SVC de devenir l'un des modèles d'apprentissage automatique les plus utilisés dans de nombreux domaines, y compris l'analyse de l'imagerie médicale, jusqu'à aujourd'hui [19]. De nombreux noyaux peuvent être choisis pour la classification SVC, allant du noyau polynomial pour le traitement des images à la fonction de base radiale (RBF) en passant par les noyaux gaussiens pour les tâches générales lorsqu'il n'y a pas de connaissance préalable des données, influençant fortement ses performances [23].

**Réseaux neuronaux artificiels (ANN)** : Les ANN constituent la base de l'apprentissage profond contemporain et sont au cœur des techniques mentionnées à la partie concernant les architectures de Deep Learning. Inspirées à l'origine par les connexions des neurones du cerveau humain [24], les ANN correspondent à une collection de neurones (unités) qui "déclenchent" une sortie si la somme linéaire de ses entrées pondérées dépasse un seuil. Ces neurones sont placés dans des couches hiérarchiques (couches d'entrée, cachées et de sortie) qui sont connectées par des liens pondérés, ce qui conduit à un réseau neuronal entièrement connecté lorsque tous les neurones d'une couche communiquent avec tous les neurones de la couche suivante. Les neurones de la couche d'entrée acceptent les vecteurs caractéristiques  $x$ , les neurones de la couche cachée traitent les données et les neurones de la couche de sortie représentent le résultat de la classification [25]. La sortie de chaque neurone est déterminée par sa fonction d'activation, qui est généralement une fonction non linéaire telle que sigmoïde, Tanh ou ReLU (l'enchaînement de fonctions linéaires donne une fonction linéaire) [26]. Comme les algorithmes précédents, les réseaux neuronaux apprennent en minimisant la perte  $L$  entre la prédiction  $\hat{y}$  et les étiquettes  $y$ . Ceci est réalisé via l'algorithme de rétropropagation (BP), qui propage l'erreur de la couche de sortie vers les couches cachées pour chaque échantillon, permettant aux poids des liens d'être ajustés en conséquence pour minimiser cette erreur [25]. Un cycle de tous les échantillons d'entraînement est appelé une époque. La façon la plus courante de procéder est d'utiliser une fonction de perte avec la descente de gradient stochastique (SGD) [26]. Cette capacité d'apprentissage que l'on retrouve dans les ANN est ce qui leur confère leur potentiel, mais aussi leur complexité due au grand nombre d'hyperparamètres utilisés pour les ajuster finement et leur capacité à surajuster les données lorsqu'ils sont trop élaborés. En effet, les combinaisons d'hyperparamètres des réseaux neuronaux peuvent concerner la structure du réseau (nombre de couches et de neurones dans chaque couche), le taux d'apprentissage et le Momentum pour le SGD, les techniques et les paramètres de régularisation, les fonctions d'activation et les conditions d'arrêt, pour n'en citer que quelques-uns [27].

**Random Forest** : La forêt aléatoire développée par Leo Breiman [4] est un groupe d'arbres de classification ou de régression non élagués, créés à partir d'une sélection aléatoire d'échantillons

de données d'apprentissage. Des caractéristiques aléatoires sont sélectionnées dans le processus d'induction. La prédiction est faite en agrégeant (vote majoritaire pour la classification ou moyenne pour la régression) des prédictions de l'ensemble. Elle présente généralement une amélioration significative des performances par rapport à un arbre de décision unique tel que C4.5. Le taux d'erreur de généralisation qu'elle produit est généralement bas comparé à la majorité des algorithmes de Machine Learning. Par ailleurs, elle est plus robuste au bruit[28].

## **2.2. Deep Learning Techniques**

La mise en œuvre de modèles d'apprentissage profond nécessite beaucoup de données pour atteindre des niveaux de performance acceptables. Ainsi, une technique d'apprentissage profond couramment utilisée lorsqu'une petite quantité de données est disponible est l'apprentissage par transfert. Elle consiste à utiliser des modèles CNN préentraînés sur de grands ensembles de données générales. Les connaissances recueillies par des CNN très performantes dans d'autres domaines généraux disposant de plus grands ensembles de données peuvent être transférées à un domaine connexe, tel que l'imagerie médicale [29].

Ces modèles sont conçus pour classer des millions d'images à travers des milliers de classes différentes et peuvent facilement être adaptés à n'importe quelle tâche de classification en remplaçant la couche de sortie dense qui effectue la prédiction réelle par une couche contenant un neurone égal au nombre de classes à prédire. [29] ont démontré comment l'apprentissage par transfert à partir d'ensembles de données de domaine général comme ImageNet pouvait être transféré au domaine des images mammographiques en utilisant des ensembles de données tels que CBISDDSM et ils obtenaient des performances atteignant des précisions de 78,4 % avec le modèle ResNet-50 [29].

Par ailleurs [30] ont montré comment l'utilisation d'architectures CNN courantes telles que VGG ou ResNet préentraînés sur ImageNet a permis d'augmenter la précision. Par exemple, la précision s'améliore de 2 à 27 % en fonction du nombre de patches utilisés.

Ce faisant, dans le cadre de notre étude nous procéderons également par Transfert Learning et nous ferons recours à plusieurs architectures telles que VGG16, VGG19, ResNet50, MobilNet, DenseNet201, InceptionV3, InceptionResNetV2, Xception, NASNetMobile,

**VGG16 et VGG19** : VGG16 (Visual Geometry Group) est une architecture de réseau neuronal à convolution (CNN) très performante pour la visualisation et la classification d'images [34]. VGG16 est attendu comme une image RVB  $224 \times 224$  avec 3 canaux en entrée. L'architecture du modèle est composée de blocs de couches de convolution de  $3 \times 3$  filtres avec un stride de 1, la même couche de padding et maxpooling de  $2 \times 2$  filtres de stride 2. Il comporte également trois couches entièrement connectées (FC) et une couche softmax pour la sortie. Toutes les couches cachées sont équipées d'une non-linéarité de type ReLU (Rectified linear unit). Les 16 couches du VGG16 font référence à 16 couches du modèle, et le VGG19 possède 19 couches.

**InceptionV3** : est une CNN de 42 couches avec une taille d'entrée par défaut fixée à  $299 \times 299$  [35]. Il possède un bloc de couches convolutives parallèles avec une taille de filtre de  $3 \times 3$ , suivi de trois modules d'initiation avec 288 filtres chacun, cinq instances des modules

d'initiation factorisés avec 768 filtres, et deux modules d'initiation avec 1280 filtres. Les sorties ont été concaténées et envoyées à une couche de sortie Softmax.

**InceptionResnetV2** : est une CNN qui contient 164 couches profondes et peut classifier des images en 1000 objets. Il s'agit d'une architecture hybride, car elle combine les connexions initiales et résiduelles. Le modèle suppose une image d'entrée de  $299 \times 299$  pixels [36].

Resnet50 : est une CNN qui a 50 couches de profondeur et une taille d'entrée d'image de  $224 \times 224$ . L'architecture de ResNet est inspirée de VGG, qui a  $3 \times 3$  filtres et suit deux règles de conception simples : (1) pour la même taille de carte de caractéristiques en sortie, les couches ont le même nombre de filtres, et (2) si la taille de la carte de caractéristiques est divisée par deux, le nombre de filtres est doublé pour préserver la complexité temporelle par couche.

**Densnet201**: est similaire à ResNet à la différence qu'il est composé de blocs denses qui sont densément connectés entre eux : Chaque couche reçoit en entrée toutes les cartes de caractéristiques de sortie des couches précédentes [37]. Un bloc dense est composé d'une normalisation par lots, d'une activation RELU, d'une convolution  $3 \times 3$ , et d'un padding zéro. Une couche de transition est composée d'une normalisation par lot, d'une convolution  $1 \times 1$  et d'une mise en commun des moyennes.

**MobilNet V2** : est une CNN basée sur un résidu inversé et conscient 54 couches ; il s'agit d'une architecture légère qui effectue une seule convolution sur chaque canal de couleur plutôt que de combiner les trois et de les aplatir.

**Xception** : Xception est une CNN qui comprend 36 couches convolutives formant la base d'extraction des caractéristiques du réseau. Les 36 couches convolutives sont structurées en 14 modules, qui sont tous entourés de connexions résiduelles linéaires autour d'eux, à l'exception du premier et du dernier module. Le modèle suppose une image d'entrée de  $299 \times 299$  pixels

**NASNetMobile** : Le NASNetMobile est un réseau particulier qui se base sur les RNN pour la construction des cellules convolutives. Les prédictions du contrôleur pour chaque cellule sont regroupées en plusieurs blocs, où chaque bloc comporte plusieurs étapes de prédiction réalisées par 5 classificateurs Softmax distincts correspondant à des choix discrets des éléments d'un bloc. Le modèle suppose une image d'entrée de  $224 \times 224$  pixels.

### 3. Préparation de Données

Cette section présente le processus de préparation de la base de données adoptée sur en ce qui concerne nos données. Elle met l'accent sur la description de la base de données, le choix de l'échantillon d'entraînement, le prétraitement de la donnée.

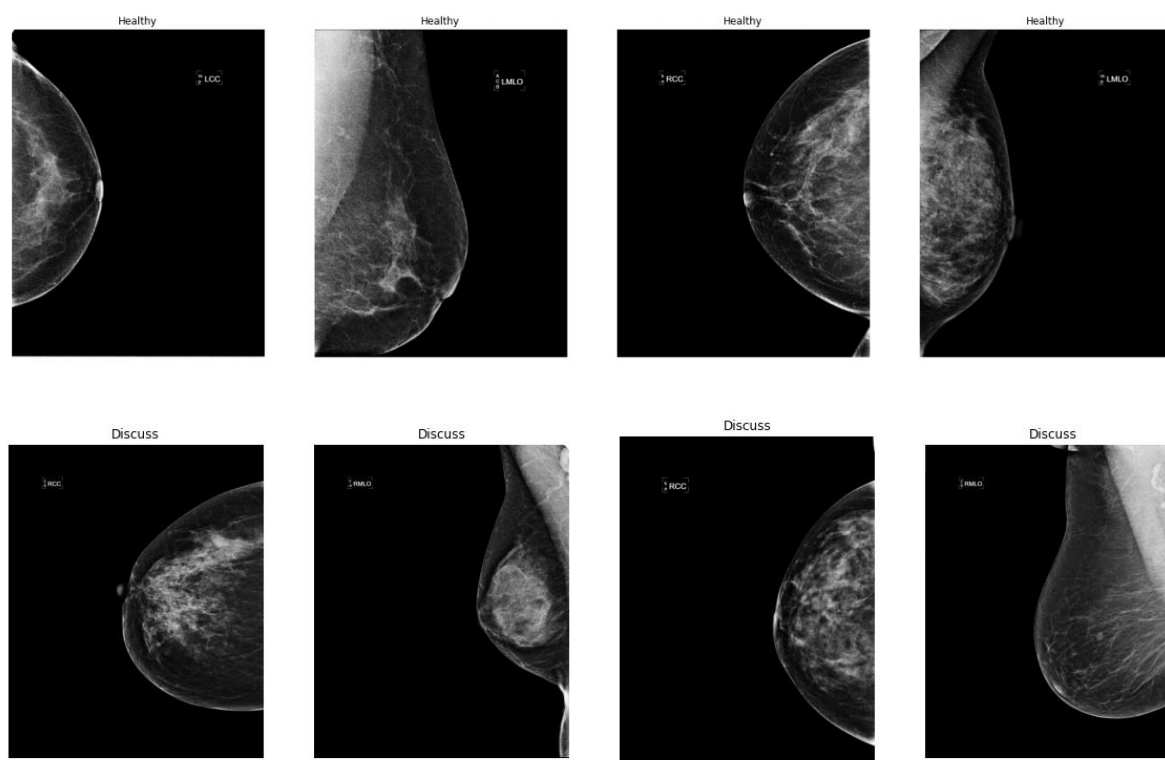
#### 3.1. Description de la base de données

Dans cette étude, nous avons utilisé les données issues d'une étude de cohorte portant sur le diagnostic du cancer du sein dont la cible était les femmes âgées de 40 et 74 ans au cours de la période de fin 2008 au 31 décembre 2015. Elles proviennent de l'hôpital universitaire

Karolinska, à Stockholm, en Suède et ont été recueillies par le chercheur principal Fredrik Strand au Karolinska Institute. Cette base de données contient non seulement des images mammographiques des patients, mais également les données cliniques de ces patientes. Elles ont été collectées afin de contribuer à des fins de recherche en IA dans le dépistage, le diagnostic et le pronostic du cancer du sein.

### - *Images pathologiques*

Dans notre base de données était inclus une sélection aléatoire de 873 femmes atteint de premier cancer du sein sur une population de 1103 femmes atteint de premier cancer du sein prenant part à l'étude de cohorte. Une autre sélection aléatoire de 10000 témoins sains au cours de la même période a été incluse et parmi celles-ci une sélection aléatoire de 7 850 cas a été incluse dans l'ensemble de données. La figure 1 ci-dessous présente quelques exemples d'images mammographiques contenues dans l'ensemble de données.



*Figure 1 : Exemple d'image contenu dans notre base de données*

### - *Clinical data*

Outre les images, l'ensemble de données contient également des données structurées sur le cancer mis en place à l'hôpital universitaire Karolinska et extraites par le centre régional de cancérologie de Stockholm-Gotland. Ces données contiennent des informations sur le moment du diagnostic et les caractéristiques du cancer. Il s'agit notamment du groupe d'âge d'appartenance des femmes, la taille de la tumeur, son histologie et les métastases des ganglions lymphatiques etc. Le tableau 1 ci-dessous présente une description exhaustive des variables incluses dans notre ensemble de donnée ainsi que leurs modalités.



Tableau 1 : Présentation des variables contenues dans la base de données clinique

<b>Variable</b>	<b>Description</b>	<b>Modalité</b>
<b>X_age</b>	Le groupe d'âge auquel appartient le patient	<b>1 = 40-55 ans au moment de la mammographie</b> <b>2 = 55+ ans au moment de la mammographie</b>
<b>X_case</b>	Le statut clinique du patient durant l'étude	<b>0 = pas de cancer pendant la période d'étude</b> <b>1 = cancer pendant la durée de l'étude</b>
<b>X_type</b>	La position de la masse	<b>1 = in situ uniquement</b> <b>2 = invasif &lt;=15 mm</b> <b>3 = invasif &gt;15 mm</b>
<b>X_lymphnode_met</b>	Métastases des ganglions lymphatiques	<b>0 = no</b> <b>1 = yes</b>
<b>Rad_timing</b>	L'instant ou l'examen de radiologie a été effectué	<b>1 = &lt;60 jours entre le dépistage et le diagnostic (cancer détecté par dépistage)</b> <b>2 = 60-729 jours (cancer d'intervalle)</b> <b>3 = 730+ jours (antécédents/autres)</b>
<b>Rad_r1</b>	La décision du radiologue 1 suite à l'examen de radiologie	<b>0 = sain</b> <b>1 = discuter</b>
<b>Rad_r2</b>	La décision du radiologue 2 suite à l'examen de radiologie	<b>0 = sain</b> <b>1 = discuter</b>
<b>Rad_recall</b>	La décision de rappeler le patient pour des analyses approfondies	<b>0 = normal</b> <b>1 = rappel</b>
<b>Rad_recall_type_right</b>	Raison du rappel pour le sein droit	<b>1 = radiologique</b> <b>2 = symptomatique</b>
<b>Rad_recall_type_left</b>	Raison du rappel pour le sein gauche	<b>1 = radiologique</b> <b>2 = symptomatique</b>
<b>Libra_breastarea</b>	Surface mammaire en cm2 estimée par le logiciel Libra	
<b>Libra_densearea</b>	Zone dense en cm2 estimée par le logiciel Libra	
<b>Libra_percentdensity</b>	Pourcentage de densité estimé par le logiciel Libra	
<b>Anon_filename</b>	Identifiant de l'image du patient	

### 3.2. *Choix de l'échantillon d'apprentissage*

L'objectif de notre étude étant de mettre en place un algorithme permettant de faire le dépistage du cancer du sein, la variable cible fut construite sur la base des décisions des radiologues ainsi un patient est dépisté cancéreux si l'un des radiologues présente des réserves par rapport au statut sain du patient.

La distribution de la variable cible dans notre population d'étude nous indique un fort déséquilibre de la répartition avec une ration d'environ 9 personnes saines pour une personne à état douteuse (Figure 1).

Un échantillon trop déséquilibré à un impact négative sur la construction d'un algorithme d'apprentissage robuste et est généralement source de surapprentissage des algorithmes, élément qui influence négativement la capacité de généralisation de nos modèles. De ce fait, nous avons choisir de manière aléatoire un échantillon de taille 5000 par classe. Cependant, en adoptant une approche data-centrique, qui se veut accorder un soin particulier à l'inspection des données et au choix du sous-échantillon le plus optimal pour la construction de meilleurs algorithmes d'apprentissage, nous avons éliminé de l'échantillon retenu suite à la sélection aléatoire, les données bruitées qui présentait une impacte négative sur l'apprentissage des modèles. En effet, certaines images dans l'échantillon retenu au préalable semblaient mal labellisées (Figure 2 à insérer).

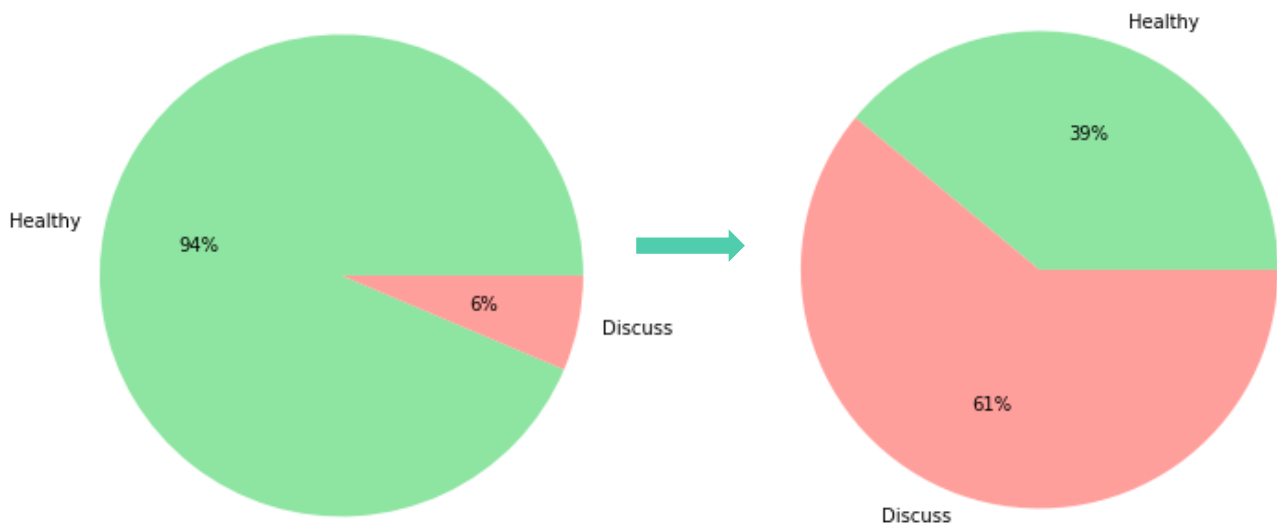


Figure 2 : Distribution obtenue suite à l'échantillon retenu pour la construction des modèles

### 3.3. *Prétraitement des données*

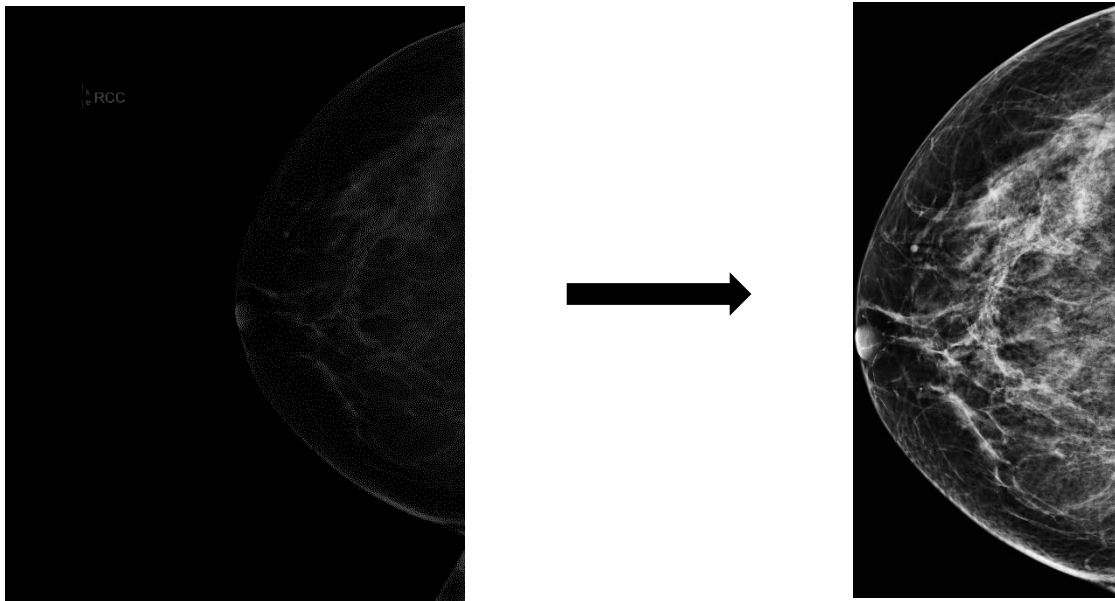
Dans cette partie, nous mettons en exergue les différentes techniques de preprocessing utilisées dans le cadre de notre étude sur les données mammographiques et sur les données cliniques des patients.

### - *Images mammographiques*

En imagerie médicale, le prétraitement est la tâche la plus capitale, il permet d'identifier les anomalies qui ne peuvent pas être reconnues en visualisant l'image directement, mais qui peuvent être détectées par des systèmes de diagnostic assistés par ordinateur. Ainsi, grâce à elle, la qualité de l'image est améliorée en supprimant les artefacts indésirables de la mammographie.

Plusieurs méthodes ont été proposées dans le cadre images mammographiques en raison de leur influence sur la détection du cancer. Parmi celles-ci, nous avons les techniques telles que le filtre médian adaptatif, le filtre moyen, le filtre moyen adaptatif, l'égalisation de l'histogramme, l'amélioration du contraste local modifié par l'histogramme, l'extraction de la région du sein, la technique d'égalisation adaptative de l'histogramme avec limitation du contraste (CLAHE) et la morphologie [12]–[14]. Dans cette étude, on a ainsi donc opté pour la technique de suppression de bruit, la détection de la région d'intérêt et le CLAHE.

En effet, les images mammographiques sont généralement caractérisées par la présence de bruit [15], ainsi le filtre médian a été employé pour enlever ce bruit. Suite à l'application du filtre médian, on applique la normalisation et le CLAHE qui nous permettent d'améliorer le contraste de l'image et donc améliorer l'effet visuel faisant ressortir plus de détail au niveau de l'image.



*Figure 3 : Format d'image obtenu suite à l'étape de prétraitement des données de mammographies*

On peut remarquer qu'au niveau de l'image obtenue comme résultat suite aux différentes techniques de preprocessing exécuté, les artefacts ont été enlevés et il ne reste que la région d'intérêt constituant la région mammaire. Par ailleurs, on peut observer une nette amélioration de l'aspect visuel de l'image ou nous observons plus de détail.

### - *Données cliniques*

Tant bien au niveau des images que des données tabulaires, le prétraitement s'avère extrêmement important. Ainsi, en ce qui concerne les données cliniques, les valeurs manquantes

ont été gérées à travers leur suppression. Cependant en ce qui concerne les variables quantitatives, une normalisation min-max a été utilisée pour pallier les différences d'échelles dont ils peuvent être sujet afin de le mettre à la même base. La normalisation pour standardiser l'échelle des variables entre 0 et 1. Face aux variables qualitatives, une one Hot Encoding a été exploitée pour celles qui n'étaient pas binaires afin de créer des variables binaires à partir de leurs modalités.

## 4. Design empirique

Cette section présente la conception empirique suivie dans cette étude, y compris : (1) les critères de performance et la méthode du 5-fold cross validation utilisée (2) le test statistique Skott Knott utilisé pour regrouper les architectures en fonction de leurs valeurs de accuracy (3) la méthode du de Borda count utilisé pour classer les architectures du meilleur cluster SK en fonction de l'accuracy, de la precision, du recall, du F1-score et de l'AUC; (4) le processus expérimental suivi pour réaliser toutes les évaluations empiriques ; (6) les abréviations utilisées pour raccourcir les noms des architectures et modèles développés ; et (7) la configuration expérimentale des classifieurs et des architectures de DL utilisées.

### 4.1. Framework proposé pour la fusion

Dans notre étude, il s'agissait d'adopter la méthode Late Fusion pour concevoir une architecture prenant en compte les deux modalités de données qu'on disposait dans notre base de données afin d'effectuer le dépistage du cancer du sein. Ainsi, il est question de combiner les modèles de Machine Learning conventionnels qu'on utilise sur les données structurées et qui seront construits pour le dépistage du cancer du sein sur la base des données cliniques des patients et les techniques de DL employé sur les données mammographiques. Plusieurs, mode d'agrégation à cet effet pouvait être employé notamment, une mode d'agrégation se basant sur la moyenne, un mode d'agrégation se basant sur la valeur maximale et un autre se basant sur la somme. Ainsi, suite aux prédictions des modèles construits sur chaque modalité de données, les probabilités prédites sont donc agrégées en fonction du mode d'agrégation retenu pour déterminer la prédiction finale. Le Framework proposé se présentait sous la forme suivante :

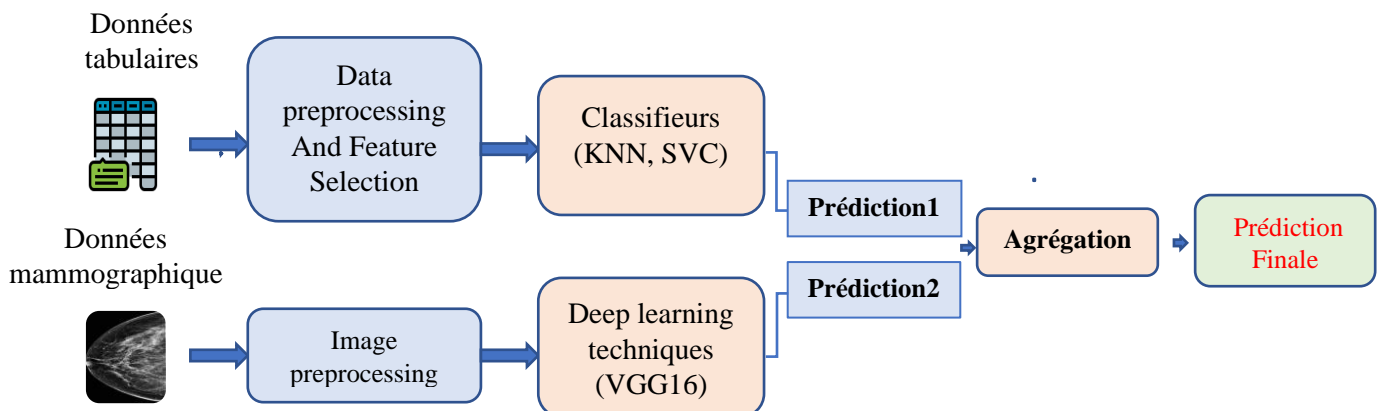


Figure 4 : Framework du Late Fusion

#### 4.2. Critère de Performance

Les mesures de performance employée dans cette étude sont l'accuracy, le rappel, la precision, le F1 score et le AUC représentant l'aire en dessous de la courbe ROC. Les terminologies suivantes sont utilisées pour définir les mesures utilisées ci-dessous : VP pour Vrai Positif (cas positif correctement prédit comme positif) ; VN pour Vrai Négatif (cas négatif correctement prédit comme négatif) ; FP pour Faux Positif (cas négatif incorrectement prédit comme positif) ; FN pour Faux Négatif (cas positif incorrectement prédit comme négatif).

**Accuracy** : représente le taux de bonne prédiction du modèle. Sa formule se présente comme suit :

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN}$$

**Recall et Precision** : La Precision et le Recall respectivement nous permettent d'avoir une idée de la proportion d'individus positifs réels correctement identifiée par le modèle et la proportion d'individu négatif réel identifiée par le modèle. Les formules sont respectivement :

$$Recall = \frac{VP}{VP + FN} \quad Precision = \frac{VP}{VP + FP}$$

**F1-Score** : Le F1 score combine la Precision et le Recall à travers leur moyenne harmonique et indique à quel point le modèle est précis :

$$F1\ score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

**AUC** : Le AUC représentant l'aire sous la courbe ROC est une valeur numérique qui nous indique comment le modèle se comportera dans différentes situations. Elle nous renseigne sur la capacité prédictive du modèle. Elle peut être calculée par :

$$AUC = \frac{\sum_i R_i(I_p) - \frac{I_p(I_p + 1)}{2}}{I_p + I_n}$$

Où  $I_p$  et  $I_n$  représente le nombre d'images positive et négative du cancer du sein et  $R_i$  le classement de la  $i$ ème image

#### 4.3. Méthodes d'analyse statistique

Suite à l'obtention des performances obtenues pour les divers algorithmes construits, deux tests statistiques non paramétriques ont été utilisés pour comparer les performances des modèles et également leur attribuer un rang. Il s'agit du Scoot Knott test et du Borda Count.

**Scott Knott (SK)** : C'est un algorithme de clustering hiérarchique utilisé comme outil d'analyse exploratoire des données. Il est habituellement utilisé dans l'analyse de la variance (ANOVA). Il a été proposé par Scott et Knott en 1974 [16] et permet d'identifier des groupes sans chevauchement sur la base de comparaisons multiples des moyennes de traitement lorsque le coefficient du test de Fisher se trouve significatif [17]. C'est l'algorithme le plus fréquemment utilisé en raison de sa simplicité et de sa robustesse. Dans cette étude, le test SK a été effectué sur la base des valeurs de la l'accuracy.

**Borda Count** : C'est une méthode de vote qui permet de sélectionner un seul gagnant. Cette technique consiste à attribuer des points aux candidats en fonction de leur classement : 1 point pour le dernier choix, 2 points pour l'avant-dernier choix, et ainsi de suite jusqu'au sommet. Les valeurs des points pour tous les rangs sont totalisées, et le candidat ayant le plus grand nombre de points est le gagnant [18]. Dans la présente étude, la technique de comptage de Borda a été utilisée pour déterminer les algorithmes de Machine Learning les plus performants sur les données cliniques des patients, pour déterminer les architectures de Deep Learning les plus performant sur les images mammographiques, pour déterminer les architectures de fusion la plus performante et enfin les architectures les plus performantes globalement, tout cela à partir de cinq critères de performances auxquels un poids uniforme a été attribué.

#### **4.4.    *Processus expérimental***

La figure 5 ci-dessus présente le processus expérimental de ce projet qui consiste en 5 étapes. En effet, elle débute par :

- Estimation des métriques telles que le Accuracy, le Recall, la Précision, le F1 Score et le ROCAUC pour tous les modèles construits sur chaque modalité de données, que ce soit les données cliniques des patients (KNN, DT, SVC, RF, ANN) ou les données de mammographie (VGG16, VGG19, INCEPV3, DNET201, XCEP, INRENV2, MBNET, RNET50, NANET) et pour les modèles de fusion obtenus suite aux croisements des différentes méthodes simples avec les modes d'agrégation (avg, sum, max) pour les deux modalités de données en utilisant un 5-fold validation croisée.
- Un clustering de tous les modèles construits sur chaque modalité de données et de fusion obtenus par méthode d'agrégation en utilisant le test de Scott Knott sur l'accuracy et identification des meilleurs clusters SK pour chaque modalité de données et pour la fusion.
- Classement des modèles du meilleur cluster SK de chaque groupe de modèles estimés sur chaque modalité de données et ceux de la fusion par méthode d'agrégation sur l'ensemble des modalités de données, en utilisant la méthode du Borda count sur la base des cinq métriques de performances. Sélection par la suite des deux meilleurs modèles sur chaque modalité de données et pour la fusion par méthode d'agrégation.
- Appliquer le test SK basé sur l'accuracy des meilleurs modèles sur chaque modalité de données et pour la fusion (obtenus à l'étape 3).
- Classer les meilleurs modèles de l'étape 4 en utilisant le Borda count sur la base des cinq critères de performance

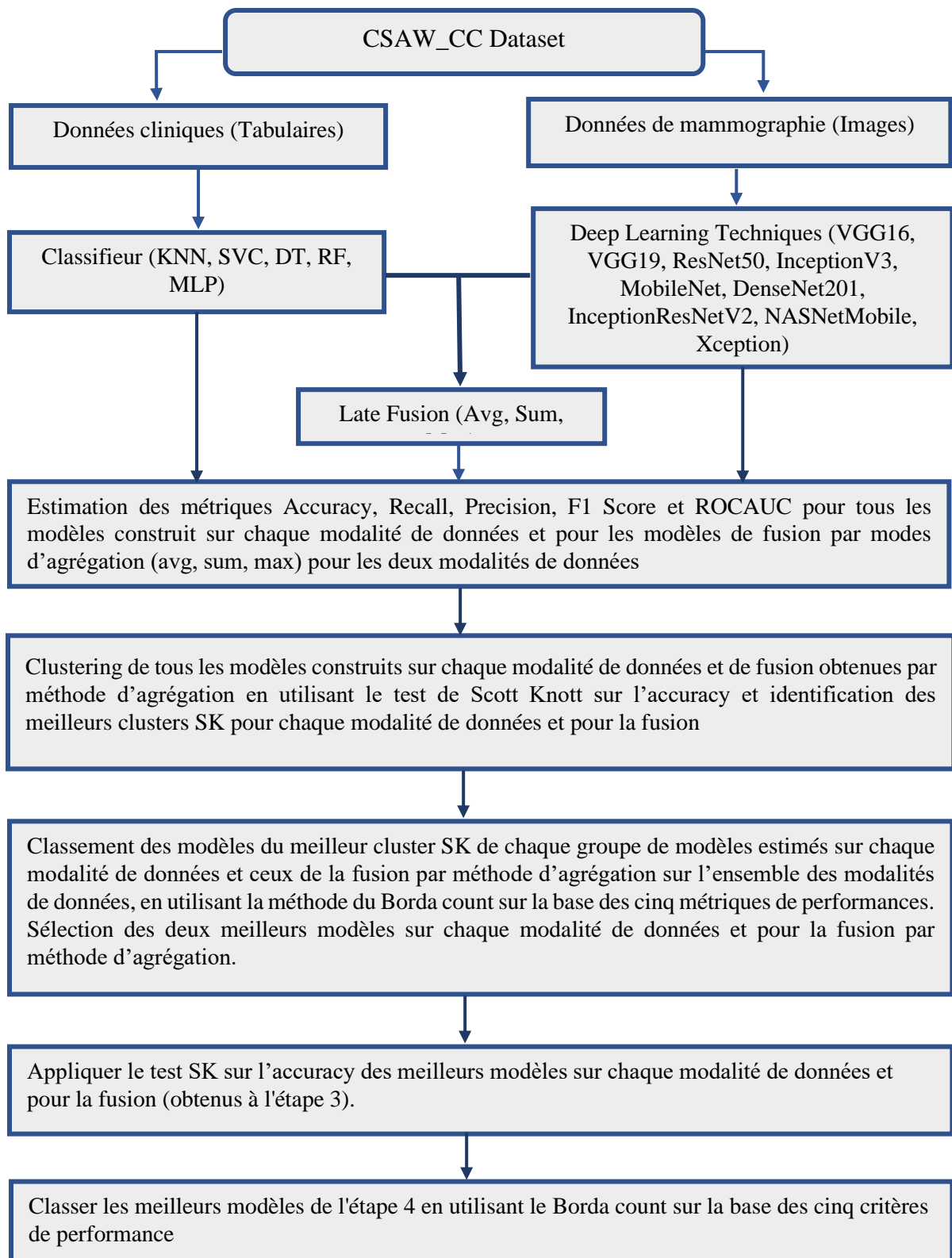


Figure 5 : experimental process

#### **4.5. Configuration de l'expérience**

Pour concevoir les architectures de fusion, quatre classifieurs (KNN, DT, SVC, RF et MLP) et neuf techniques DL à travers le Transfert Learning (VGG16, VGG19, Inception V3, DenseNet 201, ResNet 50, Inception ResNet V2, MobileNet V2, Xception et NASNetMobile) ont été respectivement utilisés pour les données cliniques des patients et leur donnée mammographiques. Les configurations expérimentales suivantes ont été utilisées sur l'échantillon de données considérées :

- Les images de l'échantillon de données contenant les images mammographiques ont été redimensionnées à  $224 \times 224$  pixels indépendamment de la technique de DL utilisé pour le Transfert Learning sachant qu'on a adopté la méthode du End-To-End donc juste la dernière couche de classification a été adaptée à la tâche qu'on veut exécuter.
- Pour les données cliniques, nous avons effectué un Grid Search pour trouver les meilleurs hyperparamètres de chaque classifieur utilisé sur un petit espace de recherche. Ainsi pour le KNN les paramètres retenus sont la distance de Manhattan et  $K=12$  pour le nombre de voisins. Pour le SVC, les configurations par défaut de Sckit-Learn s'avéraient être les meilleures sur l'espace de recherche considéré. Pour la DT, les paramètres retenus sont : l'entropie comme critère, une profondeur maximale de 7 et le nombre minimum de feuilles par nœuds. Pour le RF, nous avons retenu comme critère l'entropie, une profondeur maximale de 10 et 200 comme nombre d'estimateurs. Par contre les configurations par défaut de la bibliothèque Scikit-Learn du langage de programmation Python ont été retenues pour le MLP.

#### **4.6. Abréviations**

Nous avons songé à raccourcir le nom des techniques de Machine Learning et de Deep Learning utilisées, nous suivons les règles de dénomination suivantes. Nous abrégeons le nom de chaque technique comme indiqué dans le tableau A.1 de l'annexe A pour chaque modèle entraîné en monomodalité.

Par contre au niveau des modèles de fusion, nous avons utilisé l'abréviation du modèle de Deep Learning construit sur les données de mammographie couplée à l'aide de '+' au nom du modèle de Machine Learning construit sur les données cliniques auquel est ajoutée la méthode d'agrégation utilisée. Un exemple désignant la fusion de la InceptionResNetV2 et de KNN au moyen de la méthode d'agrégation somme s'écrit 'INREN+KNN+sum'

### **5. Résultats et Discussion**

Cette section présente et discute les résultats des évaluations empiriques des modèles sur chaque modalité de données : les classificateurs sur les données cliniques et les techniques de DL sur les données de mammographie. Les performances des modèles ont été évaluées en utilisant cinq critères de performance. Tout d'abord, les performances ont été évaluées en se basant sur l'accuracy des différents modèles construits sur chaque modalité de données (QR1). Ensuite, l'évaluation de l'impact des données multimodales sur le dépistage du cancer du sein (QR2). Après cela, les architectures de fusion seront évaluées pour mettre en exergue leur performance



(QR3). Enfin, les meilleures architectures sur chaque modalité de données et pour la fusion ont été comparées afin d'identifier l'architecture la plus performante toutes les modalités de données confondues (QR4).

Pour y arriver, Python 3.7 a été utilisé avec les deux Frameworks DL Keras et Tensorflow pour l'entraînement des architectures de DL, et le framework Scikit-Learn pour l'apprentissage automatique. R 4.0.2 a été utilisé pour l'évaluation statistique et les tests effectués.

**(QR1) : Quelle est la performance globale des modèles utilisant les données monomodales dans le dépistage du cancer du sein ? Existe-t-il une architecture monomodale plus performante que les autres ?**

- *Données cliniques des patients*

Plusieurs modèles de classification ont été construits sur les données cliniques des patients notamment le KNN, la SVC, Decision Tree, le Random Forest et le MLP classier. Suite à la réalisation du test de Scott Knott sur l'accuracy des divers modèles construit, on remarquait que tous les modèles appartenaient au même cluster (voir figure), ainsi donc de façon globale il n'existait pas de différence significative entre les performances de ces modèles.

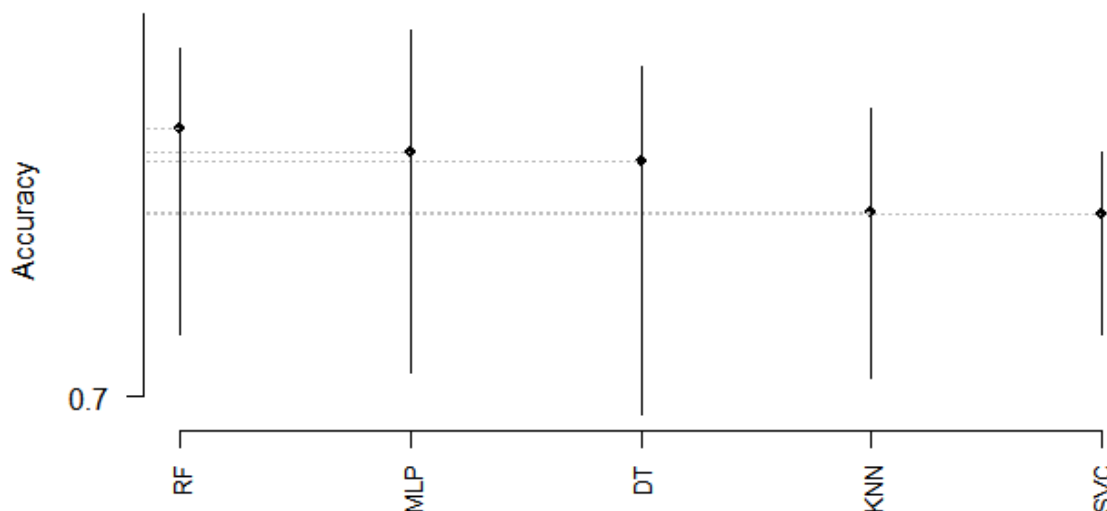


Figure 6 : Résultat du Scott Knott test sur les classifieurs

Cependant, en passant à la méthode du Borda Count pour effectuer un classement des divers modèles sur la base des cinq métriques calculées, on remarque que le modèle Random Forest a obtenu le meilleur classement avec 19 points pour le Borda count et est caractérisé par une accuracy de 74,13%, un recall de 82,06%, un F1 score de 77,36% et un AUC de 73,36%. Vient en deuxième position le modèle MLP avec 17 points pour le Borda count et qui est caractérisé par une accuracy de 73,77% et un Recall de 85,89% et F1-score de 75,87% et une AUC de 74,36%. En dernière position par contre venait le modèle KNN qui avait le moins performé de tous les modèles avec 12 points pour le Borda Count et une accuracy de 72,84%, un Recall de 8,20%, F1-score de 75,63% et un AUC de 72,73%.

Tableau 2 : Résultat du borda count sur les classifieur

Model	Accuracy	Recall	Precision	F1_score	AUC	Borda count Rang
RandomForest	0,7413	0,8206	0,7355	0,7756	0,7336	1
MLP	0,7377	0,8590	0,6805	0,7588	0,7428	2
DecisionTree	0,7364	0,8077	0,7432	0,7740	0,7270	3
SVC	0,7281	0,9266	0,6020	0,7289	0,7632	4
KNN	0,7285	0,8320	0,6936	0,7564	0,7273	5

De ces analyses on pourrait donc conclure que le Random Forest performait mieux sur les données cliniques que tous les autres modèles construits sur les données cliniques.

#### - *Données de mammographies*

Plusieurs techniques de DL ont été appliquées à travers le Transfert Learning sur les données de mammographies des patients. Il s'agit notamment du VGG16, VGG19, Inception V3, DenseNet 201, ResNet 50, Inception ResNet V2, MobileNet V2, Xception et NASNetMobile. Suite à la réalisation du test de Scott Knott sur l'accuracy des divers modèles construit, on constate la formation de quatre clusters. Le premier cluster contenant les meilleurs modèles en termes de accuracy est uniquement constitué du modèle DenseNet202. Le deuxième cluster quant à lui prenait en compte VGG16 et VGG19, ensuite le troisième cluster composé de ResNet50, MobileNet et NaSNetMobile puis viens en dernière position le cluster constitué de InceptionResNetV2, InceptionV3 et Xception.

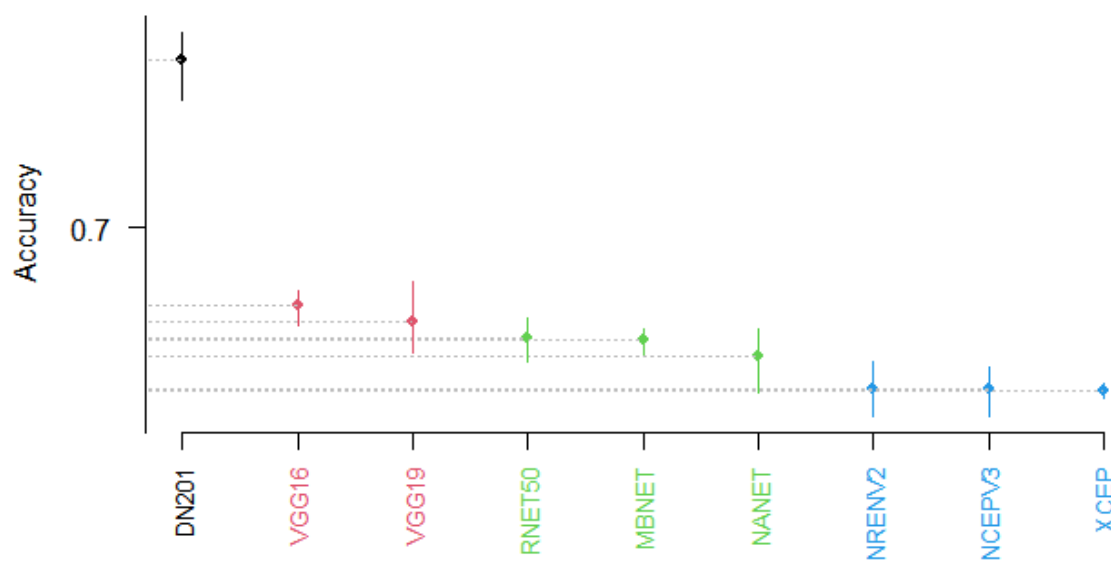


Figure 7 : Résultat du Scott Knott test sur les architectures de DL

Suite au test de Scoot Knott, en se référant au classement de ces modèles sur la base de la méthode du Borda Count sur les diverses métriques calculées, on remarque que les modèles venant en tête sont le modèle de DenseNet201 et le modèle VGG16. En effet, le modèle

DenseNet201 a totalisé 39 points et était notamment caractérisé par des performances qui surpassaient de loin celles de tous les autres modèles, une accuracy de 76,78%, un recall de 79,98%, une precision de 82.30%, un F1-Score de 81.10% et un AUC de 75,65%. Le modèle VGG16 venant en deuxième position avec 35 points a eu les performances suivantes : une accuracy de 66,87%, un recall de 68,86%, une precision de 83.29%, un F1-Score de 75.31% et un AUC de 65,39%. Le modèle ayant le moins performé sur les données image était le modèle Xception qui se trouvait caractérisé par une accuracy de 63,49%, un recall de 65,64%, une precision de 84,01%, un F1-score de 73,63% et un AUC de 61,13%.

Tableau 3 : Résultat du borda count sur les architectures de DL

Model	Accuracy	Recall	Precision	F1_score	AUC	Borda count Rang
DenseNet201	0,7670	0,7998	0,8230	0,8110	0,7565	1
VGG16	0,6687	0,6886	0,8330	0,7532	0,6539	2
VGG19	0,6626	0,6803	0,8491	0,7529	0,6549	3
MobileNet	0,6549	0,6727	0,8460	0,7483	0,6399	4
ResNet50	0,6557	0,6825	0,8168	0,7418	0,6393	5
InceptionResNetV2	0,6357	0,6558	0,8575	0,7396	0,6231	6
NASNetMobile	0,6487	0,6757	0,8181	0,7377	0,6332	7
InceptionV3	0,6353	0,6558	0,8500	0,7382	0,6160	8
Xception	0,6349	0,6564	0,8401	0,7363	0,6113	9

De ces analyses on pourrait donc conclure que l'architecture DenseNet201 performait mieux sur les données mammographiques que toutes les autres architectures construites sur les données mammographiques.

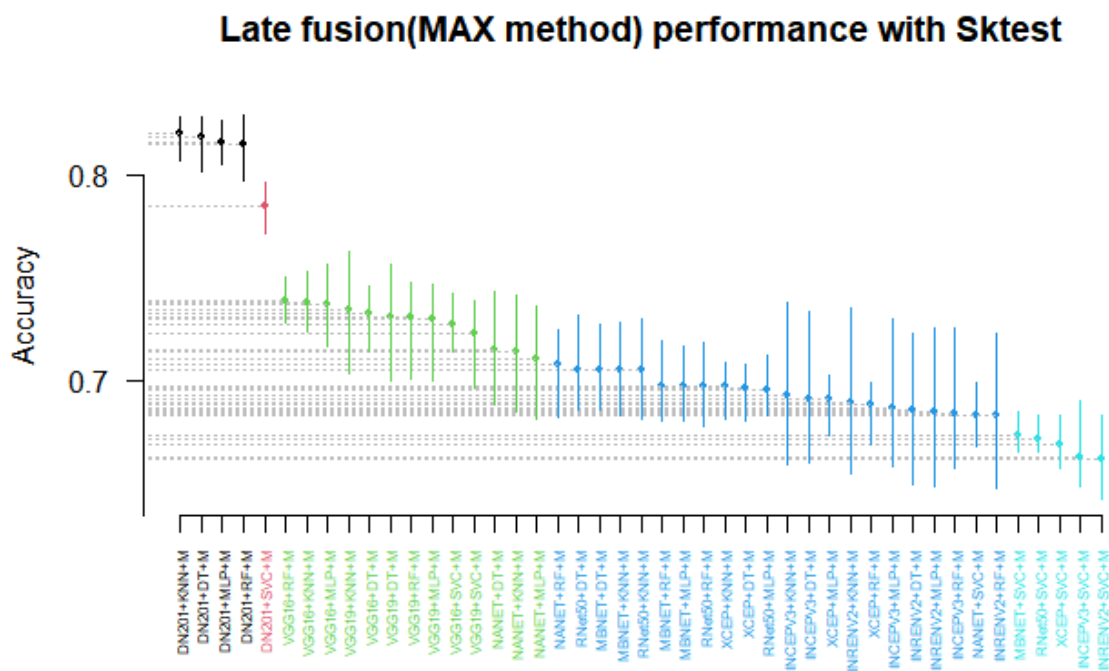
**(QR2) : Quel est l'impact d'une modalité de données sur les performances de diagnostic du cancer du sein ?**

La réponse à cette question réside dans la comparaison des meilleurs modèles obtenus suite sur chaque modalité de données. En effet, le meilleur modèle retenu sur les données cliniques des patients était le Random Forest avec des performances notables telles qu'un accuracy de 74,13%, un recall de 82,06%, un F1 score de 77,36% et un AUC de 73,36%. Cependant, sur les données mammographiques, le meilleur modèle était celui du DenseNet201 qui surpassait de loin en termes de performance tous les autres modèles de Deep Learning utilisé. Il avait obtenu les performances suivantes : une accuracy de 76,78%, un recall de 79,98%, une precision de 82.30%, un F1-Score de 81.10% et un AUC de 75,65%. Sur la base de ces observations on peut remarquer les données mammographiques ont un plus fort pouvoir discriminatoire dans le dépistage du cancer du sein comparé aux données sur le contexte clinique des patients. En effet, elle s'avère présentée sur toutes les métriques d'évaluations de performance, les meilleurs résultats.

**(RQ3) : Quelle est la performance globale des modèles DL utilisant des données multimodales dans le dépistage du cancer du sein ?**

Trois modes de Late Fusion ont été adoptés dans cette étude. La méthode de fusion « max » qui repose sur la technique d'agrégation maximale pour laquelle la valeur maximale des probabilités prédites pour chaque classe par les modèles conçus sur chaque modalité de données sont faites pour obtenir la prédiction finale, la méthode de fusion « sum » dans lequel qui consiste à effectuer la sommation des probabilités prédites pour chaque modalité de données pour déterminer la prédiction finale et la méthode de fusion « avg » qui repose sur la technique d'agrégation moyenne ou la moyenne des probabilités prédites pour chaque classe pour chaque modalité de données sont utilisées pour effectuer la prédiction finale.

Pour chaque mode de fusion, on a remarqué que le test de Scott Knott nous formait toujours les mêmes clusters. En effet, indépendamment de la méthode de fusion six clusters ont été formés. Le meilleur cluster est constitué des modèles de fusion obtenue suite à la combinaison du DenseNet et du KNN, du DenseNet201 et de la Decision Tree, du DenseNet201 et du MLP, et du DenseNet et du RandomForest. Par contre, le deuxième cluster est constitué uniquement de la combinaison entre DenseNet201 et SVC.



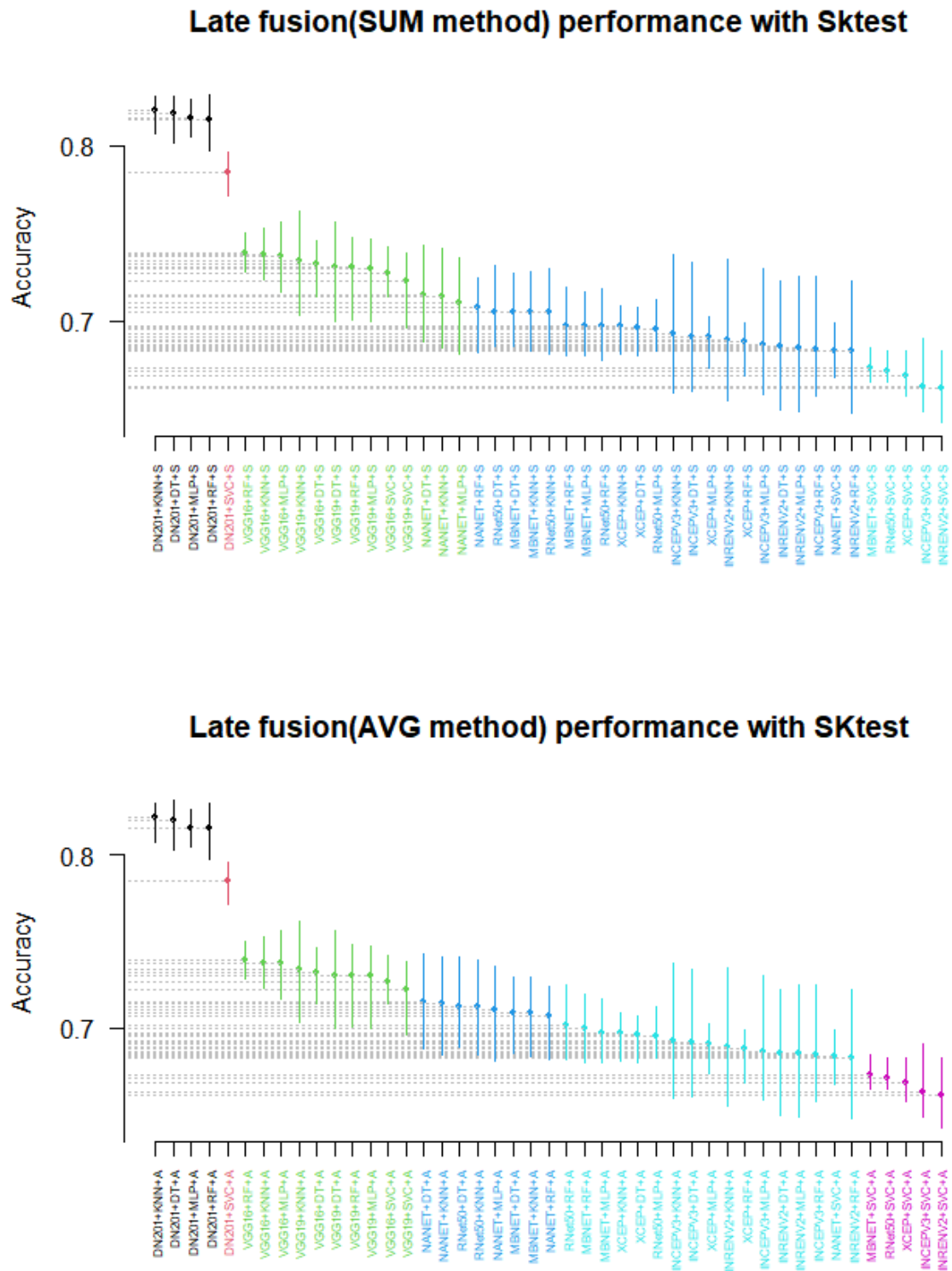


Figure 8 : Résultat du Scott Knott test sur les architectures de Late fusion par mode d'agrégation

Quand on analyse les performances des modèles du meilleur cluster déterminé par le Scott Knott test, on remarque néanmoins une légère différence des performances des modèles quand

on passe d'un mode d'agrégation à un autre même si globalement le test d'ANOVA nous indique que cette différence de performance observée lorsqu'on passe d'un mode d'agrégation à un autre n'est pas significatif.

En utilisant la méthode d'agrégation « max », la méthode Borda count classe en première et deuxième positions la fusion obtenue suite à la combinaison des modèles DenseNet201 et KNN et la combinaison des modèles DenseNet201 et Decision Tree. En effet, le modèle obtenu suite à la combinaison du DenseNet201 et KNN a obtenu 19 points avec les performances tels qu'une accuracy de 82%, un recall de 82,05%, une precision de 90,16%, un F1-score de 85,89% et un AUC de 82,05. Ce qui nous montre que le modèle KNN indépendamment de sa faible performance comparée à celle des autres modèles sur les données cliniques à une grande force de décision quand il s'agit de la fusion. La combinaison du DenseNet201 et de la Decision Tree par contre avec 16 points a eu les performances suivantes : une accuracy de 81,87%, un recall de 82,11%, une precision de 89,75%, un F1-score de 85,75 et un AUC de 81,79%.

Tableau 4 : Résultat du borda count sur le mode de fusion max

Model	Accuracy	Recall	Precision	F1_score	AUC	Borda count Rang
DNET201+KNN+max	0,8200	0,8205	0,9016	0,8589	0,8205	1
DNET201+DT+max	0,8187	0,8211	0,8975	0,8575	0,8179	2
DNET201+MLP+max	0,8157	0,8185	0,8957	0,8552	0,8148	3
DNET201+RF+max	0,8145	0,8187	0,8929	0,8540	0,8131	4

En passant à la méthode d'agrégation « sum », la méthode Borda count classe en première et deuxième positions la fusion obtenue suite à la combinaison des modèles DenseNet201 et KNN et la combinaison des modèles DenseNet201 et Decision Tree. En effet, le modèle obtenu suite à la combinaison du DenseNet201 et KNN a obtenu 19 points avec les performances tels qu'une accuracy de 82%, un recall de 82,05%, une precision de 90,16%, un F1-score de 85,89% et un AUC de 82,05. La combinaison du DenseNet201 et de la Decision Tree par contre avec 16 points a eu les performances suivantes : une accuracy de 81,87%, un recall de 82,11%, une precision de 89,75%, un F1-score de 85,75% et un AUC de 81,79%.

Tableau 5 : Résultat du borda count sur le mode de fusion sum

Model	Accuracy	Recall	Precision	F1_score	AUC	Borda count Rang
DNET201+KNN+sum	0,8200	0,8205	0,9016	0,8589	0,8205	1
DNET201+DT+sum	0,8187	0,8211	0,8975	0,8575	0,8179	2
DNET201+MLP+sum	0,8157	0,8185	0,8957	0,8552	0,8148	3
DNET201+RF+sum	0,8145	0,8187	0,8929	0,8540	0,8131	4

Quand a la méthode d'agrégation « avg » présentait de meilleures performances pour les mêmes modèles que les méthodes d'agrégation « max » et « sum ». À son niveau, la méthode Borda count classe en première et deuxième positions la fusion obtenue suite à la combinaison des modèles DenseNet201 et KNN et la combinaison des modèles DenseNet201 et Decision Tree également. En effet, le modèle obtenu suite à la combinaison du DenseNet201 et KNN a obtenu

19 points avec les performances tels qu'une accuracy de 82,15%, un recall de 82,09%, une precision de 90,41%, un F1-score de 86,03% et un AUC de 82,25. La combinaison du DenseNet201 et de la Decision Tree par contre avec 16 points a eu les performances suivantes : une accuracy de 82%, un recall de 82,12%, une precision de 90%, un F1-score de 85,87% et un AUC de 81,98%.

Tableau : Résultat du borda count sur le mode de fusion avg

Model	Accuracy	Recall	Precision	F1_score	AUC	Borda count Rang
DNET201+KNN+avg	0,8215	0,8209	0,9041	0,8603	0,8225	1
DNET201+DT+avg	0,8200	0,8212	0,9000	0,8587	0,8198	2
DNET201+MLP+avg	0,8157	0,8185	0,8957	0,8552	0,8148	3
DNET201+RF+avg	0,8157	0,8190	0,8948	0,8551	0,8146	4

**(RQ4) : Quelle est la performance des architectures de Late fusion par rapport aux modèles monomodal ?**

À la suite des résultats obtenus sur chaque modalité de données et la fusion, un test de Scott Knott a été effectué sur les meilleurs modèles observés. Ainsi suite au test de Scott Knott sur ces derniers, quatre clusters ont été formés et le meilleur cluster était notamment constituer que des modèles de fusion retenue. Le deuxième était constitué du modèle de DenseNet201 obtenu sur les données de mammographies, le troisième des modèles de Random Forest et MLP obtenu sur les données cliniques. On peut donc constater que les modèles de fusion sont généralement meilleurs et plus performants que les modèles construit sur chaque modalité de données à elle seul. Par ailleurs, le modèle de fusion faisant se basant sur la méthode d'agrégation moyenne donnait généralement de meilleures performances et celle faisant intervenir le DenseNet201 et le KNN était la plus performante.

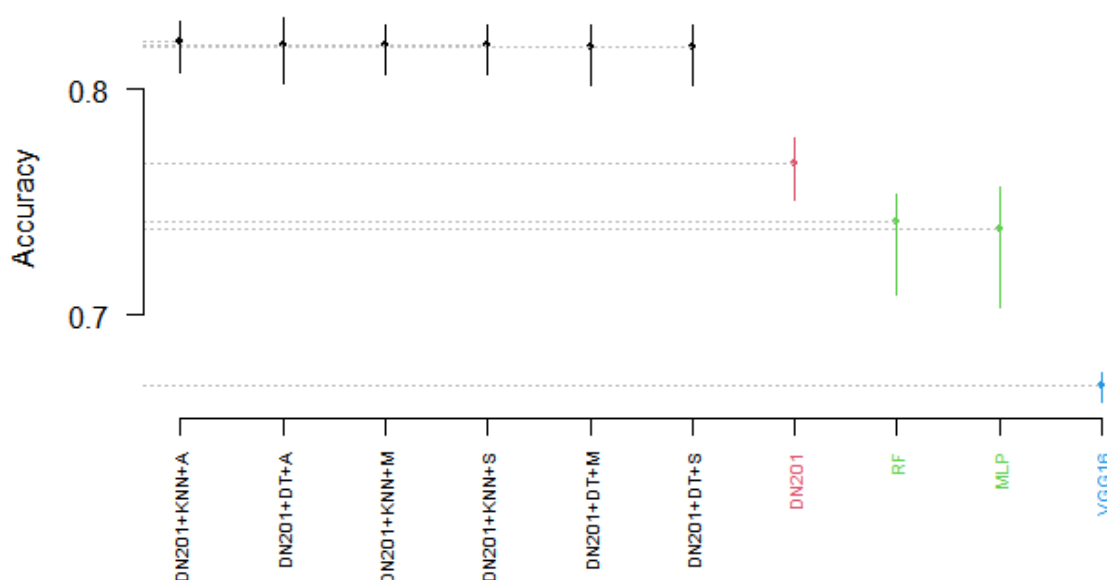


Figure 17 : Résultat du Scoot Knott test pour la comparaison globale

## 6. Limitations

Une limite de notre projet réside dans l'insuffisance de données concernant le contexte clinique du patient au niveau de la base de données tabulaire. En effet, la présence d'informations relatives aux antécédents familiaux de cancer du sein des patients ou encore par rapport à d'autres facteurs de risque aurait probablement permis d'améliorer les performances de nos modèles. De plus, bien que nous ayons utilisé des techniques d'apprentissage profond en faisant du Transfert Learning de bout en bout dans notre analyse, la complexification des architectures à travers l'ajout d'additionnelle couche aurait peut-être permis d'obtenir de meilleures performances. Une autre limite est que l'ensemble de données utilisé dans ce projet est relativement petit, ce qui peut ne pas être représentatif de l'ensemble de la population. Cela peut limiter la capacité de généralisation de nos modèles lorsqu'ils seront appliqués à des ensembles de données plus importants. Enfin, nous nous sommes concentrés sur des données images de type mammographiques, d'autres modalités d'images auraient peut-être permis de parvenir à de bien meilleures performances, chose que les prochaines études pourront explorer.

## 7. Conclusion

Dans cette étude, nous avons proposé une approche d'apprentissage profond pour la classification des données de dépistage multimodal du cancer du sein. L'objectif était de se baser sur la méthode du Late fusion pour développer une architecture de Deep Learning robuste sur les données multimodales et par prolongement de comparer leur performance aux performances des modèles construits sur chaque modalité de données. Ainsi, cinq critères de performance ont été utilisés dans le processus d'évaluation empirique, de même que le test statistique SK et la méthode de vote du Borda count pour évaluer et classer les neuf modèles DL formés sur les données de mammographie et les cinq classifieurs formés sur les données cliniques. Les résultats par rapport aux questions de recherche sont les suivants :

**(QR1) : Quelle est la performance globale des modèles utilisant les données monomodales dans le dépistage du cancer du sein ? Existe-t-il une architecture monomodale plus performante que les autres ?** Les évaluations empiriques ont montré que la meilleure technique DL sur les données de mammographie était le DenseNet201 et le meilleur modèle sur les données cliniques était le Random Forest. Ils ont notamment obtenu les meilleures performances en termes de accuracy, de recall, de precision, de F1-score et de AUC.

**(QR2) : Quel est l'impact d'une modalité de données sur les performances de diagnostic du cancer du sein ?** Les modèles utilisant les données tabulaires globalement présentaient de bonnes performances comparées aux modèles basés sur les données mammaires. Cependant, la meilleure performance a été obtenue sur les données mammographiques notamment avec le DenseNet201 qui même dans d'autres études se montrait efficace pour les tâches relatives aux domaines de la santé et particulièrement dans le dépistage du cancer du sein.

**(RQ3) : Quelle est la performance globale des modèles DL utilisant des données multimodales dans le dépistage du cancer du sein ?** Les modèles les mieux classés pour la Late fusion ont obtenu en moyenne un score de 84% dans les cinq métriques pour l'ensemble



des modalités de données indépendamment du mode d'agrégation employé. En particulier, le modèle DenseNet201 combiné au KNN a été la meilleure technique de Late fusion DL pour l'ensemble des données, tandis que la Late fusion entre InceptionResNetV2 et SVC a été la moins bonne. Ce qui nous indique que bien que le KNN performait de façon moindre sur les données cliniques étaient en mesure d'apporter plus de contributions lors de la fusion.

**(RQ4) : Quelle est la performance des architectures de Late fusion par rapport aux modèles monomodalité ?** Par rapport aux modèles basés sur la monomodalité, la fusion tardive a obtenu les meilleurs résultats avec une accuracy de 82% par rapport aux meilleurs modèles formés sur chaque modalité de données où Random Forest nous donnait une accuracy de 74% sur les données cliniques et DenseNet201, une accuracy de 76,7% sur les données de mammographie.

Les prochaines études pourront explorer d'autres méthodes de fusion telles que la Early fusion ou la jointe fusion afin de voir les éventuelles améliorations qu'elles pourront provoquer où encore s'accroître sur l'utilisation d'autres types de données images telles que les données histopathologiques. Il pourra également s'agir d'explorer beaucoup plus de modalités de données notamment les données génomiques.

## Annexe

Table A.1 : Abréviation des modèles conçus sur les données de mammographie

Modèle	Abréviations
VGG16	VGG16
VGG19	VGG19
ResNet50	RNET50
InceptionV3	INCEPV3
InceptionResNetV2	INRENV2
MobileNet	MBNET
DenseNet201	DNET201
Xception	XCEP
NASNetMobile	NANET

Table A.2 : Abréviation des modèles conçus sur les données de cliniques

Modèle	Abréviations
K-Nearest Neighbors	KNN
Decision Tree	DT
Support Vector Classifier	SVC
Multi Layer Perceptron	MLP
Random Forest	RF

Table B.1 : Performance des architectures de fusion en utilisant le mode « avg »

Model	Accuracy	Recall	Precision	F1_score	AUC
VGG16+SVC+avg	0,7270	0,7250	0,8892	0,7983	0,7309
VGG16+KNN+avg	0,7375	0,7369	0,8851	0,8039	0,7394
VGG16+DT+avg	0,7323	0,7324	0,8833	0,8004	0,7336
VGG16+RF+avg	0,7392	0,7357	0,8926	0,8063	0,7442
VGG16+MLP+avg	0,7374	0,7362	0,8873	0,8043	0,7401
VGG19+SVC+avg	0,7226	0,7193	0,8982	0,7974	0,7354
VGG19+KNN+avg	0,7342	0,7319	0,8935	0,8035	0,7421
VGG19+DT+avg	0,7308	0,7287	0,8938	0,8015	0,7396
VGG19+RF+avg	0,7308	0,7277	0,8966	0,8021	0,7405
VGG19+MLP+avg	0,7302	0,7280	0,8944	0,8013	0,7395
RNET50+SVC+avg	0,6715	0,6898	0,8407	0,7563	0,6604
RNET50+KNN+avg	0,7123	0,7084	0,9000	0,7919	0,7220
RNET50+DT+avg	0,7125	0,7095	0,8972	0,7915	0,7211
RNET50+RF+avg	0,7021	0,7030	0,8882	0,7837	0,7072
RNET50+MLP+avg	0,6953	0,7006	0,8761	0,7773	0,6969
INRENV2+SVC+avg	0,6617	0,6685	0,8929	0,7619	0,6696
INRENV2+KNN+avg	0,6891	0,6842	0,9221	0,7832	0,7206
INRENV2+DT+avg	0,6857	0,6816	0,9186	0,7806	0,7087
INRENV2+RF+avg	0,6828	0,6794	0,9196	0,7792	0,7093
INRENV2+MLP+avg	0,6853	0,6803	0,9242	0,7814	0,7183
INCEPV3+SVC+avg	0,6630	0,6688	0,8907	0,7625	0,6632
INCEPV3+KNN+avg	0,6930	0,6837	0,9289	0,7866	0,7201
INCEPV3+DT+avg	0,6919	0,6837	0,9245	0,7851	0,7147
INCEPV3+RF+avg	0,6845	0,6791	0,9199	0,7802	0,7043
INCEPV3+MLP+avg	0,6866	0,6799	0,9236	0,7820	0,7092
XCEP+SVC+avg	0,6687	0,6729	0,8876	0,7650	0,6660
XCEP+KNN+avg	0,6970	0,6884	0,9184	0,7865	0,7151
XCEP+DT+avg	0,6960	0,6886	0,9146	0,7853	0,7107
XCEP+RF+avg	0,6885	0,6831	0,9118	0,7806	0,7010
XCEP+MLP+avg	0,6911	0,6838	0,9171	0,7831	0,7070
DNET201+SVC+avg	0,7847	0,8069	0,8497	0,8275	0,7770
DNET201+KNN+avg	0,8215	0,8209	0,9041	0,8603	0,8225
DNET201+DT+avg	0,8200	0,8212	0,9000	0,8587	0,8198
DNET201+RF+avg	0,8157	0,8190	0,8948	0,8551	0,8146
DNET201+MLP+avg	0,8157	0,8185	0,8957	0,8552	0,8148
MBNET+SVC+avg	0,6734	0,6818	0,8712	0,7639	0,6689
MBNET+KNN+avg	0,7085	0,6987	0,9193	0,7933	0,7269
MBNET+DT+avg	0,7085	0,6993	0,9165	0,7928	0,7246
MBNET+RF+avg	0,7000	0,6935	0,9112	0,7870	0,7139
MBNET+MLP+avg	0,6974	0,6927	0,9062	0,7845	0,7091
NANET+SVC+avg	0,6834	0,6935	0,8656	0,7682	0,6826
NANET+KNN+avg	0,7147	0,7086	0,9081	0,7949	0,7293
NANET+DT+avg	0,7155	0,7099	0,9056	0,7949	0,7287

NANET+RF+avg	0,7075	0,7041	0,9019	0,7895	0,7205
NANET+MLP+avg	0,7109	0,7058	0,9059	0,7923	0,7247

Table B.2 : Performance des architectures de fusion en utilisant le mode « max »

Model	Accuracy	Recall	Precision	F1_score	AUC
VGG16+SVC+max	0,7270	0,7250	0,8892	0,7983	0,7309
VGG16+KNN+max	0,7375	0,7369	0,8851	0,8039	0,7394
VGG16+DT+max	0,7323	0,7324	0,8833	0,8004	0,7336
VGG16+RF+max	0,7392	0,7357	0,8926	0,8063	0,7442
VGG16+MLP+max	0,7374	0,7362	0,8873	0,8043	0,7401
VGG19+SVC+max	0,7226	0,7193	0,8982	0,7974	0,7354
VGG19+KNN+max	0,7342	0,7319	0,8935	0,8035	0,7421
VGG19+DT+max	0,7308	0,7287	0,8938	0,8015	0,7396
VGG19+RF+max	0,7308	0,7277	0,8966	0,8021	0,7405
VGG19+MLP+max	0,7302	0,7280	0,8944	0,8013	0,7395
RNET50+SVC+max	0,6715	0,6898	0,8407	0,7563	0,6604
RNET50+KNN+max	0,7053	0,7058	0,8886	0,7857	0,7100
RNET50+DT+max	0,7055	0,7068	0,8858	0,7852	0,7093
RNET50+RF+max	0,6972	0,7011	0,8802	0,7793	0,6988
RNET50+MLP+max	0,6953	0,7006	0,8761	0,7773	0,6969
INRENV2+SVC+max	0,6617	0,6685	0,8929	0,7619	0,6696
INRENV2+KNN+max	0,6891	0,6842	0,9221	0,7832	0,7206
INRENV2+DT+max	0,6857	0,6816	0,9186	0,7806	0,7087
INRENV2+RF+max	0,6828	0,6794	0,9196	0,7792	0,7093
INRENV2+MLP+max	0,6853	0,6803	0,9242	0,7814	0,7183
INCEPV3+SVC+max	0,6630	0,6688	0,8907	0,7625	0,6632
INCEPV3+KNN+max	0,6926	0,6836	0,9283	0,7862	0,7191
INCEPV3+DT+max	0,6915	0,6836	0,9239	0,7848	0,7137
INCEPV3+RF+max	0,6842	0,6789	0,9193	0,7798	0,7033
INCEPV3+MLP+max	0,6866	0,6799	0,9236	0,7820	0,7092
XCEP+SVC+max	0,6687	0,6729	0,8876	0,7650	0,6660
XCEP+KNN+max	0,6970	0,6884	0,9184	0,7865	0,7151
XCEP+DT+max	0,6960	0,6886	0,9146	0,7853	0,7107
XCEP+RF+max	0,6885	0,6831	0,9118	0,7806	0,7010
XCEP+MLP+max	0,6911	0,6838	0,9171	0,7831	0,7070
DNET201+SVC+max	0,7847	0,8069	0,8497	0,8275	0,7770
DNET201+KNN+max	0,8200	0,8205	0,9016	0,8589	0,8205
DNET201+DT+max	0,8187	0,8211	0,8975	0,8575	0,8179
DNET201+RF+max	0,8145	0,8187	0,8929	0,8540	0,8131
DNET201+MLP+max	0,8157	0,8185	0,8957	0,8552	0,8148
MBNET+SVC+max	0,6734	0,6818	0,8712	0,7639	0,6689
MBNET+KNN+max	0,7053	0,6975	0,9140	0,7905	0,7211
MBNET+DT+max	0,7053	0,6981	0,9112	0,7900	0,7189
MBNET+RF+max	0,6977	0,6927	0,9075	0,7849	0,7102
MBNET+MLP+max	0,6974	0,6927	0,9062	0,7845	0,7091

NANET+SVC+max	0,6834	0,6935	0,8656	0,7682	0,6826
NANET+KNN+max	0,7145	0,7085	0,9078	0,7947	0,7290
NANET+DT+max	0,7153	0,7099	0,9053	0,7947	0,7284
NANET+RF+max	0,7075	0,7041	0,9019	0,7895	0,7205
NANET+MLP+max	0,7109	0,7058	0,9059	0,7923	0,7247

Table B.3 : Performance des architectures de fusion en utilisant le mode « sum »

Model	Accuracy	Recall	Precision	F1_score	AUC
VGG16+SVC+sum	0,7270	0,7250	0,8892	0,7983	0,7309
VGG16+KNN+sum	0,7375	0,7369	0,8851	0,8039	0,7394
VGG16+DT+sum	0,7323	0,7324	0,8833	0,8004	0,7336
VGG16+RF+sum	0,7392	0,7357	0,8926	0,8063	0,7442
VGG16+MLP+sum	0,7374	0,7362	0,8873	0,8043	0,7401
VGG19+SVC+sum	0,7226	0,7193	0,8982	0,7974	0,7354
VGG19+KNN+sum	0,7342	0,7319	0,8935	0,8035	0,7421
VGG19+DT+sum	0,7308	0,7287	0,8938	0,8015	0,7396
VGG19+RF+sum	0,7308	0,7277	0,8966	0,8021	0,7405
VGG19+MLP+sum	0,7302	0,7280	0,8944	0,8013	0,7395
RNET50+SVC+sum	0,6715	0,6898	0,8407	0,7563	0,6604
RNET50+KNN+sum	0,7053	0,7058	0,8886	0,7857	0,7100
RNET50+DT+sum	0,7055	0,7068	0,8858	0,7852	0,7093
RNET50+RF+sum	0,6972	0,7011	0,8802	0,7793	0,6988
RNET50+MLP+sum	0,6953	0,7006	0,8761	0,7773	0,6969
INRENV2+SVC+sum	0,6617	0,6685	0,8929	0,7619	0,6696
IN- RENV2+KNN+sum	0,6891	0,6842	0,9221	0,7832	0,7206
INRENV2+DT+sum	0,6857	0,6816	0,9186	0,7806	0,7087
INRENV2+RF+sum	0,6828	0,6794	0,9196	0,7792	0,7093
INRENV2+MLP+sum	0,6853	0,6803	0,9242	0,7814	0,7183
INCEPV3+SVC+sum	0,6630	0,6688	0,8907	0,7625	0,6632
INCEPV3+KNN+sum	0,6926	0,6836	0,9283	0,7862	0,7191
INCEPV3+DT+sum	0,6915	0,6836	0,9239	0,7848	0,7137
INCEPV3+RF+sum	0,6842	0,6789	0,9193	0,7798	0,7033
INCEPV3+MLP+sum	0,6866	0,6799	0,9236	0,7820	0,7092
XCEP+SVC+sum	0,6687	0,6729	0,8876	0,7650	0,6660
XCEP+KNN+sum	0,6970	0,6884	0,9184	0,7865	0,7151
XCEP+DT+sum	0,6960	0,6886	0,9146	0,7853	0,7107
XCEP+RF+sum	0,6885	0,6831	0,9118	0,7806	0,7010
XCEP+MLP+sum	0,6911	0,6838	0,9171	0,7831	0,7070
DNET201+SVC+sum	0,7847	0,8069	0,8497	0,8275	0,7770
DNET201+KNN+sum	0,8200	0,8205	0,9016	0,8589	0,8205
DNET201+DT+sum	0,8187	0,8211	0,8975	0,8575	0,8179
DNET201+RF+sum	0,8145	0,8187	0,8929	0,8540	0,8131
DNET201+MLP+sum	0,8157	0,8185	0,8957	0,8552	0,8148
MBNET+SVC+sum	0,6734	0,6818	0,8712	0,7639	0,6689

MBNET+KNN+sum	0,7053	0,6975	0,9140	0,7905	0,7211
MBNET+DT+sum	0,7053	0,6981	0,9112	0,7900	0,7189
MBNET+RF+sum	0,6977	0,6927	0,9075	0,7849	0,7102
MBNET+MLP+sum	0,6974	0,6927	0,9062	0,7845	0,7091
NANET+SVC+sum	0,6834	0,6935	0,8656	0,7682	0,6826
NANET+KNN+sum	0,7145	0,7085	0,9078	0,7947	0,7290
NANET+DT+sum	0,7153	0,7099	0,9053	0,7947	0,7284
NANET+RF+sum	0,7075	0,7041	0,9019	0,7895	0,7205
NANET+MLP+sum	0,7109	0,7058	0,9059	0,7923	0,7247

## Référence bibliographique

- [1] ‘Breast cancer’. <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (accessed Dec. 09, 2022).
- [2] ‘Breast cancer – IARC’. <https://www.iarc.who.int/cancer-type/breast-cancer/> (accessed Dec. 05, 2022).
- [3] H. Zerouaoui and A. Idri, ‘Deep hybrid architectures for binary classification of medical breast cancer images’, *Biomedical Signal Processing and Control*, vol. 71, p. 103226, Jan. 2022, doi: 10.1016/j.bspc.2021.103226.
- [4] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, ‘Cancer statistics, 2021’, *Ca Cancer J Clin*, vol. 71, no. 1, pp. 7–33, 2021.
- [5] J. Song *et al.*, ‘Multiview multimodal network for breast cancer diagnosis in contrast-enhanced spectral mammography images’, *Int J CARS*, vol. 16, no. 6, pp. 979–988, Jun. 2021, doi: 10.1007/s11548-021-02391-4.
- [6] E. A. Mohamed, E. A. Rashed, T. Gaber, and O. Karam, ‘Deep learning model for fully automated breast cancer detection system from thermograms’, *PLOS ONE*, vol. 17, no. 1, p. e0262349, Jan. 2022, doi: 10.1371/journal.pone.0262349.
- [7] M. Jiang, S. Lei, J. Zhang, L. Hou, M. Zhang, and Y. Luo, ‘Multimodal Imaging of Target Detection Algorithm under Artificial Intelligence in the Diagnosis of Early Breast Cancer’, *J Healthc Eng*, vol. 2022, p. 9322937, Jan. 2022, doi: 10.1155/2022/9322937.
- [8] N. M. ud din, R. A. Dar, M. Rasool, and A. Assad, ‘Breast cancer detection using deep learning: Datasets, methods, and challenges ahead’, *Computers in Biology and Medicine*, vol. 149, p. 106073, Oct. 2022, doi: 10.1016/j.combiomed.2022.106073.
- [9] R. Yan *et al.*, ‘Richer fusion network for breast cancer classification based on multimodal data’, *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 134, Apr. 2021, doi: 10.1186/s12911-020-01340-6.
- [10] S.-C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren, ‘Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines’, *npj Digit. Med.*, vol. 3, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41746-020-00341-z.
- [11] J. N. Acosta, G. J. Falcone, P. Rajpurkar, and E. J. Topol, ‘Multimodal biomedical AI’, *Nat Med*, vol. 28, no. 9, Art. no. 9, Sep. 2022, doi: 10.1038/s41591-022-01981-2.

- [12] D. N. Ponraj, M. E. Jenifer, P. Poongodi, and J. S. Manoharan, 'A survey on the preprocessing techniques of mammogram for the detection of breast cancer', *Journal of emerging trends in computing and information sciences*, vol. 2, no. 12, pp. 656–664, 2011.
- [13] K. J. McLoughlin, P. J. Bones, and N. Karssemeijer, 'Noise equalization for detection of microcalcification clusters in direct digital mammogram images', *IEEE transactions on medical imaging*, vol. 23, no. 3, pp. 313–320, 2004.
- [14] M. Talha, G. B. Sulong, and A. Jaffar, 'Preprocessing digital breast mammograms using adaptive weighted frost filter', *Biomedical Research-tokyo*, 2016, Accessed: Jan. 10, 2023. [Online]. Available: <https://www.semanticscholar.org/paper/Preprocessing-digital-breast-mammograms-using-frost-Talha-Sulong/1874888fe67cfae07f818507499ca73c1ccd7e9f>
- [15] M. Chevalier, F. Leyton, M. Nogueira, M. Oliveira, T. A. da Silva, and J. Emilio, 'Image Quality Requirements for Digital Mammography in Breast Cancer Screening', *Imaging of the Breast - Technical Aspects and Clinical Implication*, 2012, Accessed: Jan. 10, 2023. [Online]. Available: [https://www.academia.edu/85825883/Image\\_Quality\\_Requirements\\_for\\_Digital\\_Mammography\\_in\\_Breast\\_Cancer\\_Screening](https://www.academia.edu/85825883/Image_Quality_Requirements_for_Digital_Mammography_in_Breast_Cancer_Screening)
- [16] K. J. Worsley, 'A Non-Parametric Extension of a Cluster Analysis Method by Scott and Knott', *Biometrics*, vol. 33, no. 3, pp. 532–535, 1977, doi: 10.2307/2529369.
- [17] E. Jelihovschi and J. Faria, 'ScottKnott: A Package for Performing the Scott-Knott Clustering Algorithm in R', *TEMA (São Carlos)*, vol. 15, Mar. 2014, doi: 10.5540/tema.2014.015.01.0003.
- [18] P. Emerson, 'The original Borda count and partial voting', *Soc Choice Welf*, vol. 40, no. 2, pp. 353–358, Feb. 2013, doi: 10.1007/s00355-011-0603-9.
- [19] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, 'Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis', *Designs*, vol. 2, no. 2, Art. no. 2, Jun. 2018, doi: 10.3390/designs2020013.
- [20] L. Peterson, 'K-nearest neighbor', *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009, doi: 10.4249/scholarpedia.1883.
- [21] S. L. Salzberg, 'C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993', *Mach Learn*, vol. 16, no. 3, pp. 235–240, Sep. 1994, doi: 10.1007/BF00993309.
- [22] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., 2019.
- [23] S. Amari and S. Wu, 'Improving support vector machine classifiers by modifying kernel functions', *Neural Networks*, vol. 12, no. 6, pp. 783–789, Jul. 1999, doi: 10.1016/S0893-6080(99)00032-5.
- [24] F. B. Fitch, 'Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. Bulletin of mathematical biophysics, vol. 5 (1943), pp. 115–133.', *The Journal of Symbolic Logic*, vol. 9, no. 2, pp. 49–50, Jun. 1944, doi: 10.2307/2268029.
- [25] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Pearson, 2021.
- [26] G. Litjens *et al.*, 'A survey on deep learning in medical image analysis', *Med Image Anal*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

- [27] '1.17. Neural network models (supervised)', *scikit-learn*. [https://scikit-learn/stable/modules/neural\\_networks\\_supervised.html](https://scikit-learn/stable/modules/neural_networks_supervised.html) (accessed Jan. 24, 2023).
- [28] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, 'Random Forests and Decision Trees', *International Journal of Computer Science Issues(IJCSI)*, vol. 9, Sep. 2012.
- [29] L. G. Falconi, M. Perez, and W. G. Aguilar, 'Transfer Learning in Breast Mammogram Abnormalities Classification With Mobilenet and Nasnet', *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 109–114, Jun. 2019, doi: 10.1109/IWSSIP.2019.8787295.
- [30] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, 'Deep Learning to Improve Breast Cancer Detection on Screening Mammography', *Sci Rep*, vol. 9, no. 1, p. 12495, Aug. 2019, doi: 10.1038/s41598-019-48995-4.
- [31] J. Li and N. M. Allinson, 'A comprehensive review of current local features for computer vision', *Neurocomputing*, vol. 71, no. 10, pp. 1771–1787, Jun. 2008, doi: 10.1016/j.neucom.2007.11.032.
- [32] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, 'Gradient-based learning applied to document recognition', *Proceedings of the Institute of Radio Engineers*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.
- [33] 'Computer Vision: Algorithms and Applications | SpringerLink'. <https://link.springer.com/book/10.1007/978-3-030-34372-9> (accessed Jan. 24, 2023).
- [34] K. Simonyan and A. Zisserman, 'Very Deep Convolutional Networks for Large-Scale Image Recognition'. arXiv, Apr. 10, 2015. doi: 10.48550/arXiv.1409.1556.
- [35] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, 'Rethinking the Inception Architecture for Computer Vision'. arXiv, Dec. 11, 2015. doi: 10.48550/arXiv.1512.00567.
- [36] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, 'Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning'. arXiv, Aug. 23, 2016. doi: 10.48550/arXiv.1602.07261.
- [37] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, 'Densely Connected Convolutional Networks'. arXiv, Jan. 28, 2018. doi: 10.48550/arXiv.1608.06993.
- [38] S.-C. Huang, A. Pareek, R. Zamanian, I. Banerjee, and M. P. Lungren, 'Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection', *Sci Rep*, vol. 10, no. 1, Art. no. 1, Dec. 2020, doi: 10.1038/s41598-020-78888-w.
- [39] A. Kline *et al.*, 'Multimodal machine learning in precision health: A scoping review', *npj Digit. Med.*, vol. 5, no. 1, Art. no. 1, Nov. 2022, doi: 10.1038/s41746-022-00712-8.
- [40] G. Hamed, M. A. E.-R. Marey, S. E.-S. Amin, and M. F. Tolba, 'Deep Learning in Breast Cancer Detection and Classification', in *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, Cham, 2020, pp. 322–333. doi: 10.1007/978-3-030-44289-7\_30.
- [41] B. B. Zimmermann *et al.*, 'Multimodal breast cancer imaging using coregistered dynamic diffuse optical tomography and digital breast tomosynthesis', *JBO*, vol. 22, no. 4, p. 046008, Apr. 2017, doi: 10.1117/1.JBO.22.4.046008.
- [42] 'Health and Well-Being'. <https://www.who.int/data/gho/data/major-themes/health-and-well-being> (accessed Dec. 05, 2022).

- [43] E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, 'Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review', *Expert Systems with Applications*, vol. 167, p. 114161, Apr. 2021, doi: 10.1016/j.eswa.2020.114161.
- [44] N. Arya and S. Saha, 'Multi-modal advanced deep learning architectures for breast cancer survival prediction', *Knowledge-Based Systems*, vol. 221, p. 106965, Jun. 2021, doi: 10.1016/j.knosys.2021.106965.
- [45] A. A. A. Figueiredo, J. G. do Nascimento, F. C. Malheiros, L. H. da Silva Ignacio, H. C. Fernandes, and G. Guimaraes, 'Breast tumor localization using skin surface temperatures from a 2D anatomic model without knowledge of the thermophysical properties', *Computer Methods and Programs in Biomedicine*, vol. 172, pp. 65–77, Apr. 2019, doi: 10.1016/j.cmpb.2019.02.004.
- [46] K. Chikamai, S. Viriri, and J. R. Tapamo, 'Combining Feature Methods for Content-Based Classification of Mammogram Images', *International Journal of Computers Communications & Control*, vol. 8, no. 4, p. 499, 2013, Accessed: Jan. 06, 2023. [Online]. Available: [https://www.academia.edu/11350978/Combining\\_Feature\\_Methods\\_for\\_Content\\_Based\\_Classification\\_of\\_Mammogram\\_Images](https://www.academia.edu/11350978/Combining_Feature_Methods_for_Content_Based_Classification_of_Mammogram_Images)
- [47] S. Thawkar, 'Automatic Detection and Classification of Masses in Digital Mammograms', Accessed: Jan. 06, 2023. [Online]. Available: [https://www.academia.edu/33454583/Automatic\\_Detection\\_and\\_Classification\\_of\\_Masses\\_in\\_Digital\\_Mammograms](https://www.academia.edu/33454583/Automatic_Detection_and_Classification_of_Masses_in_Digital_Mammograms)
- [48] B. Halalli, A. Makandar, B. Halalli, and A. Makandar, *Computer Aided Diagnosis - Medical Image Analysis Techniques*. IntechOpen, 2017. doi: 10.5772/intechopen.69792.