

Université de Yaoundé I
Ecole Nationale Supérieure
Polytechnique de Yaoundé
Département de Génie
Informatique



The University of Yaounde I
National Advanced School
of Engineering Yaounde
Department of Computer
Science

PREDICTION DU CANCER DU SEIN

MACHINE LEARNING II

Présenté par :

DJIFACK Line

EDO Nicolas

Flora PANSY

MBANG Belvine

TAPAMO Dimitri

TAZO Belle-Nickelle

TCHAFA Ymelda

Sous la supervision de :
Dr FIPPO

Année académique: 2024 / 2025

- **Intelligence artificielle (IA)**

Ensemble de techniques permettant aux machines d'effectuer des tâches nécessitant l'intelligence humaine, comme la perception, le raisonnement et l'apprentissage.

- **Deep Learning (DL)**

Sous-catégorie de l'apprentissage automatique utilisant des réseaux de neurones profonds pour apprendre des représentations complexes des données.

- **SMOTE (Synthetic Minority Oversampling Technique)**

Technique qui génère des exemples synthétiques pour la classe minoritaire dans un ensemble de données déséquilibré, améliorant ainsi l'équilibre des classes.

- **Wisconsin Breast Cancer Dataset (WBCD)**

Ensemble de données utilisé pour la recherche sur le cancer du sein, contenant des caractéristiques de tumeurs classées comme bénignes ou malignes.

- **AdaGradient**

Algorithme d'optimisation qui ajuste le taux d'apprentissage pour chaque paramètre en fonction des gradients précédents, permettant une convergence plus rapide et efficace dans l'entraînement des modèles de machine Learning.

Le cancer du sein demeure l'une des principales causes de mortalité chez les femmes à l'échelle mondiale, avec 670 000 décès enregistrés en 2022. Cette maladie est caractérisée par la prolifération incontrôlée de cellules mammaires anormales, entraînant la formation de tumeurs potentiellement fatales si elles ne sont pas détectées et traitées à temps. Dans ce contexte alarmant, notre étude vise à explorer le potentiel de l'intelligence artificielle et du deep learning pour améliorer la détection précoce du cancer du sein. L'objectif principal est de développer des modèles capables de classer les tumeurs en catégories bénignes ou malignes, facilitant ainsi une intervention rapide et appropriée. Pour atteindre cet objectif, nous avons adopté une méthodologie rigoureuse. Nous avons d'abord appliqué la méthode SMOTE, pour équilibrer les classes dans notre ensemble de données, générant des exemples synthétiques pour la classe minoritaire. Par la suite, notre jeu de données provenant de la Wisconsin Breast Cancer Dataset (WBCD), a été enrichi de caractéristiques significatives des échantillons, telles que la moyenne des rayons, la texture, et l'aire des cellules. Le modèle a été entraîné sur 150 époques, et ses performances ont été évaluées à l'aide de métriques clés telles que la précision, le rappel et le F1 score. Cependant, l'adoption de ces techniques nécessite une validation rigoureuse et une interopérabilité des résultats pour garantir leur fiabilité en milieu clinique. L'algorithme Adam Gradient Descent s'est imposé comme l'une des méthodes les plus efficaces pour le diagnostic du cancer du sein, offrant une précision de 99 % grâce à sa capacité à optimiser les réseaux neuronaux de manière dynamique.

Mots clés : Cancer du sein, deep learning, Adam, détection précoce.

Breast cancer remains one of the leading causes of mortality among women worldwide, with 670,000 deaths recorded in 2022. This disease is characterized by the uncontrolled proliferation of abnormal breast cells, leading to the formation of potentially fatal tumors if not detected and treated in time. In this alarming context, our study aims to explore the potential of artificial intelligence and deep learning to enhance the early detection of breast cancer. The primary objective is to develop models capable of classifying tumors into benign or malignant categories, thereby facilitating timely and appropriate intervention. To achieve this goal, we adopted a rigorous methodology. We first applied the SMOTE method to balance the classes in our dataset, generating synthetic examples for the minority class. Subsequently, our dataset, derived from the Wisconsin Breast Cancer Dataset (WBCD), was enriched with significant features of the samples, such as the mean radius, texture, and area of the cells. The model was trained over 150 epochs, and its performance was evaluated using key metrics such as accuracy, recall, and F1 score. However, the adoption of these techniques necessitates rigorous validation and interpretability of results to ensure their reliability in clinical settings. The Adam Gradient Descent algorithm has emerged as one of the most effective methods for breast cancer diagnosis, achieving an accuracy of 99% due to its ability to dynamically optimize neural networks.

Keywords: Breast cancer, deep learning, Adam, early detection.

LISTE DES TABLEAUX

Tableau 1:Résultats obtenus.....	xxii
----------------------------------	------

LISTE DES FIGURES

Figure 1: Anatomie du sein	xiii
Figure 2 : Image représentant l'importation des bibliothèques.....	xxviii
Figure 3:Image illustrant le début de l'analyse exploratoire des données	xxix
Figure 4: Matrice de confusion	xxx
Figure 5: Image représentant la structure les étiquettes de diagnostic.....	xxx
Figure 6 : Architecture du model	xxxix
Figure 7: Image des étapes d'initialisation du modèle	xxxix
Figure 8 : Image montrant les étapes d'entraînement du modèle	xxxix
Figure 9 : Graphe d'évolution de Training loss et celle de Validation loss	xxxix
Figure 10 : Graphe d'évolution de Training Accuracy et celle de Validation Accuracy.....	xxxix
Figure 11 : Evaluation du modèle	xxxix

SOMMAIRE

GLOSSAIRE.....	iii
RESUME	iv
ABSTRACT.....	v
LISTE DES TABLEAUX.....	vi
LISTE DES FIGURES.....	vii
INTRODUCTION	x
CONTEXTE & PROBLÉMATIQUE	x
OBJECTIFS	x
PLAN DU DOCUMENT.....	x
CHAPITRE 1 : PRÉSENTATION DE LA MALADIE	xii
1.1 Qu'est-ce que le cancer du sein ?.....	xiii
1.1.1 L'anatomie du sein	xiii
1.1.2 Les différents cancers du sein	xiv
1.1.3 Les lésions bénignes.....	xv
1.1.4 Les facteurs de risque de cancer du sein non liés à l'environnement.....	xv
1.1.5 Les facteurs de risque de cancer du sein liés aux habitudes de vie.....	xvii
1.1.6 Prévention du cancer du sein.	xvii
1.1.7 La Mammographie : Test de dépistage du cancer du sein	xviii
1.1.8 Diagnostic et Traitements.....	xviii
CHAPITRE 2 : ÉTAT DE L'ART	xx
2.1 Revue	xxi
2.2 Résultats Obtenus.....	xxii
2.3 Limites et Défis.....	xxii
2.4 Perspectives d'Amélioration.....	xxii
2.5 Positionnement.....	xxiii
CHAPITRE 3 : METHODOLOGIE	xxiv
3.1 Préparation du jeu de données	xxv
3.1.1 Méthode d'ajout des lignes.....	xxv
3.1.2 Méthode d'ajout des colonnes	xxv
3.2 Préparation des Données.....	xxvi
3.3. Construction et Entraînement du Modèle	xxvi
3.4. Évaluation des Performances	xxvi
CHAPITRE 4 : RESULTATS	xxvii
4.1 Description du Projet	xxviii
4.2 Importation des Bibliothèques	xxviii

4.3 Importation du Jeu de Données.....	xxviii
4.4 Analyse Exploratoire des Données	xxviii
4.5 Séparation des Caractéristiques et de la Cible	xxix
4.6 Conversion des Étiquettes.....	xxx
4.7 Division du Jeu de Données.....	xxx
4.8 Conversion en Tenseurs PyTorch.....	xxx
4.9 Création du DataLoader	xxxix
4.10 Architecture du Réseau de Neurones	xxxix
4.11 Initialisation du Modèle, Fonction de Perte et Optimiseur	xxxix
4.13 Visualisation des Résultats.....	xxxix
Courbe de Perte (Loss curve) : Figure 9	xxxix
Précision d'Entraînement et Précision de Validation : Figure 10	xxxix
4.14 Évaluation du Modèle	xxxix
Interprétation des Métriques	xxxv
CONCLUSION.....	xxxvi
REFERENCES BIBLIOGRAPHIQUES	xxxvii

CONTEXTE & PROBLÉMATIQUE

Le cancer du sein demeure une préoccupation mondiale majeure, représentant la première cause de cancer chez les femmes dans 157 pays en 2022, avec environ 670 000 décès attribués à cette maladie [1]. Malgré l'absence de facteurs de risque spécifiques chez près de la moitié des patientes, l'incidence de cette maladie augmente avec l'âge, et des disparités significatives sont observées en fonction du niveau de développement humain [2]. Ces inégalités soulignent l'urgence d'améliorer les méthodes de détection et de classification des tumeurs afin de permettre une prise en charge rapide et efficace.

L'émergence de l'intelligence artificielle (IA) et du deep learning (DL) offre des opportunités prometteuses pour transformer la manière dont nous abordons la prédiction et la classification des cancers. En particulier, l'utilisation d'algorithmes avancés tels que l'optimiseur AdaGrad permet d'adapter l'apprentissage aux spécificités des données, facilitant ainsi l'optimisation des performances des modèles [3]. Dans ce contexte, notre étude vise à développer des modèles capables de classer les tumeurs en catégories bénignes ou malignes, contribuant ainsi à la détection précoce de la maladie.

OBJECTIFS

Pour atteindre cet objectif, nous avons intégré des techniques avancées de traitement des données, notamment la méthode SMOTE (Synthetic Minority Oversampling Technique), qui permet d'équilibrer les classes dans l'ensemble de données [4]. Ce processus a été enrichi par l'ajout de caractéristiques clés des échantillons, telles que la moyenne des rayons et la texture des cellules, permettant ainsi d'améliorer la robustesse du modèle. En entraînant notre modèle sur 150 époques et en évaluant ses performances à l'aide de métriques telles que la précision, le rappel et le F1 score, nous visons à démontrer l'efficacité du deep learning dans la classification des échantillons de cancer du sein.

PLAN DU DOCUMENT

La suite de ce document est organisée comme suit :

- **Présentation de la maladie**

Dans ce chapitre, nous aborderons les caractéristiques du cancer du sein, en mettant en lumière les statistiques et l'impact de cette maladie sur la santé publique. Nous examinerons également les facteurs de risque associés au cancer du sein ainsi que les méthodes de détection actuellement utilisées, afin de mieux comprendre les enjeux liés à cette maladie.

- **État de l'art**

Ce chapitre se concentrera sur les avancées récentes en intelligence artificielle et deep learning, en présentant une revue des modèles utilisés dans la classification des cancers. Nous mettrons également l'accent sur des algorithmes spécifiques de classification, notamment l'optimiseur AdaGrad et ses diverses applications dans le domaine de la santé.

● **Méthodologie**

Nous décrirons ici les méthodes employées pour notre étude. Cela inclura la préparation des données, en mettant en avant les techniques de traitement et d'équilibrage, comme la méthode SMOTE. Nous expliquerons ensuite l'entraînement du modèle, en détaillant la configuration et le suivi des pertes. La validation du modèle sera également discutée, avec une évaluation des performances et la génération de la matrice de confusion. Enfin, nous fournirons une description de l'architecture du réseau de neurones utilisé.

● **Résultats**

Dans ce chapitre, nous présenterons les performances du modèle, en analysant les métriques pertinentes telles que la précision, le rappel et le F1 score. Nous inclurons une visualisation des résultats à travers l'interprétation de la matrice de confusion. Enfin, nous discuterons des résultats obtenus, en les comparant à d'autres études et en examinant leurs implications cliniques.

● **Conclusion**

Ce dernier chapitre proposera une synthèse des résultats obtenus, suivie de perspectives futures. Nous discuterons des suggestions pour de futures recherches et des améliorations possibles du modèle, afin de contribuer à l'avancement des méthodes de détection et de classification du cancer du sein.

CHAPITRE 1 : PRÉSENTATION DE LA MALADIE

Dans ce chapitre, nous aborderons les caractéristiques du cancer du sein, en mettant en lumière les statistiques et l'impact de cette maladie sur la santé publique. Nous examinerons également les facteurs de risque associés au cancer du sein ainsi que les méthodes de détection actuellement utilisées, afin de mieux comprendre les enjeux liés à cette maladie.

1.1 Qu'est-ce que le cancer du sein ?

Le cancer du sein est une maladie caractérisée par la croissance incontrôlée de cellules mammaires anormales qui forment alors des tumeurs. Sous le terme de cancer du sein sont regroupés plus d'une douzaine de cancer différents. Certains sont provoqués par des cellules cancéreuses dont la multiplication est favorisée par les hormones sexuelles féminines, les œstrogènes. [5]

1.1.1 L'anatomie du sein

Un sein contient divers types de tissus : [6]

- les glandes, formées de petits « sacs » (les lobules) où se trouvent les cellules productrices de lait pendant l'allaitement ;
- des canaux de lactation, qui font communiquer les lobules avec le mamelon pour permettre le passage du lait ;
- de la graisse, pour soutenir lobules et canaux ;
- de la peau qui enveloppe le tout.

Les cellules du sein sont sensibles aux hormones sexuelles (œstrogènes et progestérone, produits par les ovaires, les glandes surrénales et le tissu graisseux) ainsi qu'à d'autres hormones, par exemple la prolactine sécrétée par le cerveau. Les seins sont parcourus par des vaisseaux et des ganglions lymphatiques, qui participent à la protection de l'organisme contre les infections.

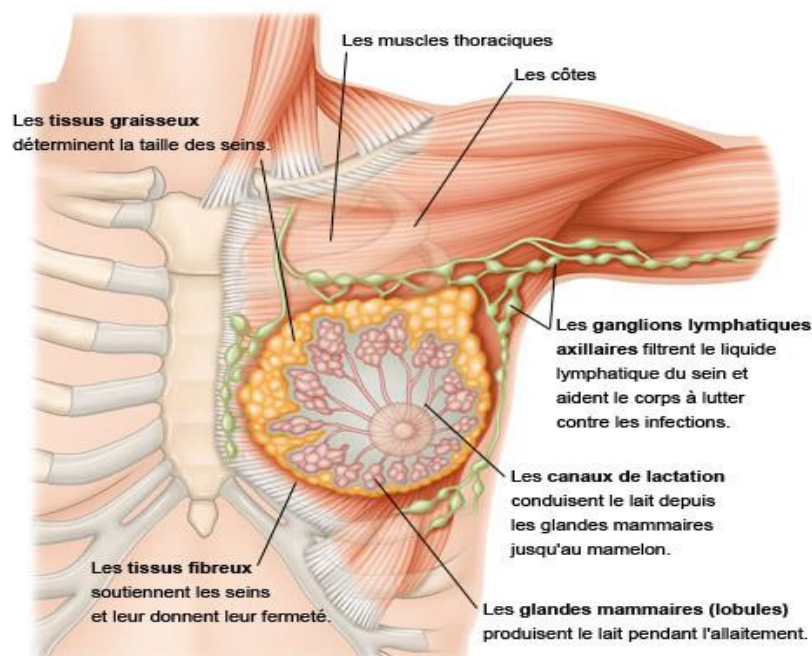


Figure 1: Anatomie du sein

1.1.2 Les différents cancers du sein

Le terme « cancer du sein » correspond à la présence de cellules anormales dans un sein, cellules qui se multiplient de façon anarchique pour former une **tumeur** maligne (un « carcinome »). Selon le type de cancer du sein, ces cellules peuvent rester confinées dans le sein ou migrer vers les ganglions avoisinants, voire le reste du corps (métastases).

Selon le type de cellules à l'origine du cancer, et selon l'aspect de la tumeur, on distingue différents cancers du sein. Par exemple, un cancer du sein qui touche les cellules qui bordent les canaux sera dit « **canalaire** » et un cancer qui affecte les cellules des lobules sera dit « **lobulaire** ».

De plus, certains cancers du sein sont provoqués par des cellules cancéreuses dont la multiplication est favorisée par les hormones sexuelles féminines, les œstrogènes. Ce sont les cancers du sein dits « hormonodépendants » ou « **hormonosensibles** ». En cas de sensibilité des cellules cancéreuses aux estrogènes, le médecin peut bloquer la croissance de la tumeur en supprimant ces hormones par le biais de médicaments adaptés. [6]

● Les cancers du sein non invasifs

On parle de cancer du sein « **non invasif** » ou « **in situ** » lorsque la tumeur reste dans le tissu d'origine et n'envahit pas les tissus voisins. Ces cancers sont plus facilement traités que les cancers qui envahissent les tissus voisins (« cancers invasifs », voir ci-dessous).

Dans 90 % des cas, les cancers du sein non invasifs sont de type canalaire (ils se forment à l'intérieur des canaux de lactation), les 10 % restants étant de type lobulaire.

La maladie de Paget du mamelon est un cancer non invasif des canaux les plus proches du mamelon. Elle se traduit par des lésions du mamelon ressemblant à de l'eczéma (rougeur, plaies, suintements, croûtes, etc.) qui ne guérissent pas. Le traitement de la maladie de Paget est chirurgical.

● Les cancers du sein invasifs

Les cancers du sein sont dits « **invasifs** » ou « **infiltrants** » lorsque les cellules cancéreuses ne restent pas confinées à leur lieu d'origine et envahissent les tissus avoisinants, les ganglions locaux (dits « ganglions axillaires » et situés sous l'aisselle) et, parfois, le reste du corps. Ces cancers invasifs touchent essentiellement les canaux, rarement les lobules. Leur pronostic est moins bon que celui des cancers du sein non invasifs.

Il existe plusieurs formes rares de cancer du sein invasif dont chacune ne représente pas plus de 1 à 2 % des cas. Ces cancers ont un meilleur pronostic que les cancers du sein canaux invasifs les plus fréquents. Les formes rares de cancer du sein invasif sont les suivantes :

- Le **carcinome mucineux ou colloïde** : ce cancer du sein se développe plutôt chez les femmes de 60 à 70 ans.
- Le **carcinome papillaire** : ce cancer du sein n'est pas toujours invasif. Lorsqu'il l'est, il touche plutôt les femmes âgées.
- Le **carcinome médullaire** : ce cancer du sein invasif est plutôt diagnostiqué chez les femmes de moins de 50 ans. Chez les femmes jeunes, sa présence fait suspecter une prédisposition génétique au cancer du sein.
- Le **carcinome tubuleux** : ce cancer du sein invasif a tendance à être de petite taille et diagnostiqué chez les femmes de 55 ans et plus. Malgré sa nature invasive, il ne se propage

que rarement hors du sein atteint.

1.1.3 Les lésions bénignes

Les biopsies percutanées permettent de différencier les grosseurs dues à un cancer de celles dues à diverses maladies sans gravité. [6]

- **Le fibroadénome du sein**

Le fibroadénome est une petite masse de tissus durs qui peut être sentie à la palpation, comme une petite bille glissante. Bénin, il est fréquent chez les femmes de moins de 25 ans. Selon sa taille, il est laissé ou retiré.

- **Le kyste du sein**

Le kyste du sein est une poche remplie de liquide qui apparaît comme une masse molle et mobile à la palpation. Courant chez les femmes âgées de 35 à 50 ans, le kyste du sein est bénin et sa taille est sensible à l'influence des hormones sexuelles. Dans certains cas, il peut être nécessaire d'aspirer le liquide contenu dans le kyste ou de l'enlever.

- **Les changements fibrokystiques du sein**

Également appelés « **mastose sclérokystique** » ou « **fibrose kystique du sein** », ces changements de structure du sein sont bénins et sous influence hormonale. Très fréquents, ces changements sont observés chez 50 à 80 % des femmes, souvent au niveau des deux seins. À la palpation, ce sont des masses bosselées, parfois douloureuses, avec ou sans écoulement par le mamelon. Chez les femmes issues de famille à prédisposition génétique pour le cancer du sein, ces changements doivent amener à une surveillance régulière des seins.

La fibrose kystique du sein est parfois traitée à l'aide de médicaments à base d'hormones sexuelles, de diurétiques ou de médicaments contre la douleur. Le port d'un soutien-gorge de maintien peut être nécessaire pour soulager la douleur.

Certains facteurs de risque de cancer du sein ont été identifiés. On distingue les facteurs de risque indépendants des habitudes de vie des patientes et ceux sur lesquels chacune d'entre elles peut influencer.

1.1.4 Les facteurs de risque de cancer du sein non liés à l'environnement

Certaines personnes ont un risque plus important de cancer du sein, quelles que soient leurs habitudes de vie. Les principaux éléments qui permettent de définir les personnes à risque sont :

- **Le sexe** : plus de 99 % des cancers du sein sont diagnostiqués chez les femmes.
- **L'âge** : environ 80 % des cancers du sein sont diagnostiqués chez des femmes âgées de plus de 50 ans, 20 % chez les femmes de moins de 50 ans dont la moitié chez celles de moins de 30 ans. Si l'âge moyen du diagnostic est de 61 ans, environ un tiers des cancers du sein sont diagnostiqués après 69 ans.
- **Une prédisposition génétique familiale** (antécédents familiaux...)
- **Une prédisposition personnelle** : les femmes qui ont déjà eu un cancer du sein ont trois à quatre fois plus de risque d'en développer un autre que celles qui n'en ont jamais eu. Par exemple, un cancer du sein canalaire non invasif multiplie par huit à dix le risque d'avoir un

cancer du sein invasif. De la même manière, un cancer du sein lobulaire non invasif augmente le risque de cancer du sein pour les deux seins.

- **Certaines lésions du sein** : certaines femmes souffrent de proliférations de cellules non cancéreuses dans leurs seins (« hyperplasie épithéliale intracanalair »). Selon l'aspect de ces cellules, le risque de développer un cancer du sein est multiplié par un facteur allant de 1,5 à 5.
- **Une radiothérapie du thorax** : l'administration répétée de rayons ionisants sur le thorax (par exemple, pour traiter une maladie de Hodgkin ou pour dépister une maladie des poumons) augmente le risque de cancer du sein, surtout si cette radiothérapie a eu lieu avant l'âge de 30 ans.
- **Une exposition prolongée aux œstrogènes.**

● **Les gènes qui augmentent le risque de cancer du sein**

Le fait d'avoir une mère, une sœur ou une fille qui a eu un cancer du sein double le risque d'en développer un soi-même. Ce risque est augmenté :

- si un cancer du sein a été diagnostiqué chez deux parentes du premier degré,
- si des parentes du premier degré ont développé un cancer de l'ovaire, du côlon (gros intestin) ou du rectum,
- si une parente du premier degré a développé deux cancers du sein,
- si un parent masculin a développé ce type de cancer.

Cette **prédisposition familiale** s'explique par des mutations sur certains gènes. La recherche scientifique s'est penché sur les cas de cancers du sein dits « familiaux » où une forte prédisposition génétique pouvait être suspectée (5 à 10 % des cancers du sein). Cette recherche a mené à la découverte de divers gènes qui favorisent ce type de cancer.

Les plus fréquemment observés sont deux gènes, **BRCA 1 et BRCA2**, dont certaines mutations sont liées à une fréquence plus élevée de cancer du sein, chez la femme comme chez l'homme, mais aussi de cancer de l'ovaire chez la femme. Ces gènes codent pour des protéines qui protègent les cellules d'une transformation cancéreuse. Lors de mutation, cette transformation cancéreuse devient plus probable. Environ deux femmes sur mille portent ces mutations sur leurs gènes BRCA 1 et 2.

D'autres mutations, plus rares, ont également été observées en lien avec un risque plus élevé de développer un cancer du sein.

● **Le rôle des œstrogènes dans le cancer du sein**

Les cellules du sein sont sensibles aux œstrogènes(hormones sexuelles) et, en particulier, à la présence d'œstrogènes dans le sang tout au long de la vie. En d'autres termes, tout ce qui augmente la durée **totale** d'exposition aux œstrogènes sanguins augmente le risque de cancer du sein. Ainsi, **le risque de cancer du sein est augmenté** :

- chez les femmes qui ont eu leurs premières règles avant l'âge de douze ans ;
- chez les femmes qui ont été ménopausées après 50 ans ;
- chez les femmes qui n'ont pas eu d'enfant (les œstrogènes diminuent fortement pendant la grossesse et l'allaitement maternel) ou celles qui ont eu leur premier enfant après l'âge de 35 ans ;

- les femmes qui souffrent d'obésité (la graisse produit des œstrogènes), en particulier après la ménopause.

Les femmes qui ont eu plusieurs enfants et qui les ont allaités ont donc un risque réduit de cancer du sein.

1.1.5 Les facteurs de risque de cancer du sein liés aux habitudes de vie

Des facteurs de risque de cancer du sein dépendants des habitudes de vie ont été identifiés :

- Le surpoids et l'obésité après la ménopause ;
- le tabagisme, même lorsqu'il est passif ;
- la consommation excessive de boissons alcoolisées ;
- l'absence d'exercice physique régulier ;
- les traitements hormonaux substitutifs de la ménopause (œstrogènes et progestérone) lorsque ceux-ci sont administrés de manière continue pendant plus de cinq ans ;
- la prise de pilule contraceptive de manière continue pendant plus de quatre ans : la pilule augmente légèrement le risque de développer un cancer du sein mais ce risque accru disparaît dix ans après en avoir cessé l'usage. Cependant, la pilule diminue le risque de souffrir de cancer de l'ovaire, en particulier chez les femmes qui ont des mutations sur leurs gènes BRCA 1 et 2. Pour ces raisons, les femmes qui souffrent d'une prédisposition familiale au cancer du sein peuvent choisir la pilule comme moyen contraceptif. [6]

1.1.6 Prévention du cancer du sein.

● Peut-on prévenir la survenue d'un cancer du sein ?

Les mesures de prévention du cancer du sein sont généralement peu spécifiques et concernent tous les cancers. Parmi ces mesures, il est conseillé de ne pas fumer, d'adopter une alimentation équilibrée, de lutter contre l'embonpoint, de limiter la consommation de boissons alcoolisées et de pratiquer régulièrement une activité physique [8]. Dans certains pays peu ensoleillés, les autorités de santé recommandent la prise de suppléments de vitamine D, notamment pour les femmes âgées ou celles qui sortent peu [5]. Pour les femmes issues de familles avec une forte prédisposition génétique au cancer du sein, la prévention doit être discutée au cas par cas avec un médecin [5].

● Quels sont les symptômes du cancer du sein ?

Les symptômes du cancer du sein incluent une masse dure dans un sein, des ganglions gonflés au niveau de l'aisselle, des écoulements spontanés du mamelon, et des modifications de l'aspect du mamelon ou de la peau du sein . Lorsque le cancer s'est propagé, d'autres symptômes tels que fatigue, nausées, perte de poids et douleurs osseuses peuvent apparaître [7].

● Le carcinome inflammatoire, un cancer du sein rare

Le carcinome inflammatoire se manifeste par un sein chaud, rouge et enflé. Cette forme de cancer progresse rapidement et est plus difficile à traiter.

● **Pronostic et évolution des cancers du sein**

Les cancers invasifs peuvent envahir le sein et migrer vers les ganglions lymphatiques et d'autres organes, formant des métastases. Divers éléments influencent le pronostic, tels que l'âge de la patiente, la taille de la tumeur, et la présence de signes d'inflammation ou de métastases.

● **Les stades d'évolution du cancer du sein**

Le stade d'évolution du cancer est déterminé par une classification TNM, qui évalue la taille de la tumeur, la présence de cellules cancéreuses dans les ganglions lymphatiques, et l'existence de métastases. Les stades vont de 0 à IV, en fonction de la gravité.

1.1.7 La Mammographie : Test de dépistage du cancer du sein

La mammographie, une technique de radiographie adaptée à l'anatomie du sein, permet de dépister ou diagnostiquer le cancer. Elle consiste à prendre deux clichés du sein et est recommandée dans le cadre du programme national de dépistage organisé ou d'un dépistage individuel en cas de facteur de risque [7].

● **Résultats de la mammographie**

Les résultats de la mammographie sont classifiés selon un système ACR, allant de l'interprétation difficile à la suspicion de cancer nécessitant des examens complémentaires.

● **Impact de la mammographie sur la mortalité**

Le programme national de dépistage organisé permet la découverte précoce d'environ 15 000 cancers du sein par an en France, avec une estimation d'une réduction de 20 % du risque de décès. Cependant, le dépistage entraîne également des risques de sur diagnostic et de traitements inutiles.

● **Problèmes de sur diagnostic et de faux positifs**

Certains cancers dépistés précocement n'évolueraient pas en cancers symptomatiques, entraînant des traitements inutiles. Environ 13 femmes sur 1 000 traitées le sont pour un faux positif ou un cancer à évolution lente. De plus, 150 femmes sur 1 000 présentent un nodule non cancéreux, entraînant un stress et des examens invasifs [8]

● **Évaluation personnelle du dépistage**

Il est crucial que chaque femme discute de sa situation personnelle avec son médecin avant de décider de participer au dépistage organisé [7].

1.1.8 Diagnostic et Traitements

Lorsqu'un cancer du sein est suspecté, il est crucial de réaliser des examens complémentaires, tels que la mammographie et l'échographie, qui permettent de visualiser la présence de tumeurs. Si des anomalies sont détectées, une biopsie est effectuée pour analyser les cellules cancéreuses ; la biopsie percutanée est souvent utilisée dans ce cas, car elle permet de prélever des fragments de tumeur à l'aide d'une aiguille guidée par imagerie, garantissant ainsi un diagnostic précis.

Après confirmation d'un cancer du sein, divers tests supplémentaires sont nécessaires, notamment la recherche de récepteurs hormonaux et l'activation du gène HER2, car ces éléments influencent les décisions de traitement. Les options thérapeutiques incluent la chirurgie pour retirer la tumeur, la radiothérapie pour détruire les cellules cancéreuses restantes, ainsi que des médicaments comme la chimiothérapie, les traitements hormonaux et les thérapies ciblées. En conclusion, la prévention, le diagnostic précoce et les traitements adaptés sont essentiels pour améliorer le pronostic et la qualité de vie des patientes atteintes de cancer du sein

CHAPITRE 2 : ÉTAT DE L'ART

Ce chapitre se concentrera sur les avancées récentes en intelligence artificielle et Deep Learning, en présentant une revue des modèles utilisés dans la classification des cancers. Nous mettrons également l'accent sur des algorithmes spécifiques de classification, notamment l'optimiseur Ada-Grad et ses diverses applications dans le domaine de la santé.

2.1 Revue

L'apprentissage profond (AP) et les algorithmes d'optimisation comme Adam Gradient Descent jouent un rôle crucial dans l'amélioration des diagnostics du cancer du sein. Cette maladie demeure une menace significative pour la santé des femmes, avec un taux de détection traditionnel limité à 78 %, entraînant de nombreux faux positifs et négatifs. Pour remédier à ces défis, l'apprentissage profond est utilisé dans plusieurs applications clés.

Premièrement, la **classification d'images** est une application majeure. Les réseaux de neurones, tels que ResNet et DenseNet, sont employés pour déterminer si une image est normale ou anormale, atteignant des précisions allant jusqu'à 99 % dans certaines études [9]. Deuxièmement, la **détection d'objets** joue un rôle essentiel. Des modèles comme YOLO et R-CNN permettent de localiser des anomalies dans les images médicales, facilitant ainsi le diagnostic précoce [10]. Enfin, la **segmentation**, réalisée par des modèles comme U-Net, est cruciale pour délimiter les contours des lésions, ce qui est essentiel pour la planification chirurgicale [9].

Cependant, pour que ces applications soient pleinement efficaces, une optimisation adéquate est nécessaire. C'est ici qu'intervient l'**algorithme Adam Gradient Descent**, qui se révèle particulièrement efficace pour améliorer les performances des modèles de deep learning. Adam se distingue par sa rapidité de convergence et sa capacité à résoudre le problème des gradients disparus. Cet algorithme a permis d'atteindre une précision de 98,24 % pour le diagnostic du cancer du sein, surpassant d'autres modèles comme les forêts aléatoires et les SVM [11].

Pour mieux comprendre le fonctionnement d'Adam, il est important de noter qu'il combine deux méthodologies de descente de gradient : le **momentum** et le **RMSPProp**. Le momentum accélère l'algorithme de descente en prenant en compte la moyenne exponentiellement pondérée des gradients, ce qui permet une convergence plus rapide vers les minima. En parallèle, Adam ajuste les taux d'apprentissage pour chaque paramètre individuellement, ce qui le rend bien adapté aux problèmes avec des gradients épars ou du bruit dans les données.

Ainsi, l'intégration de l'apprentissage profond avec l'optimisation par Adam Gradient Descent représente une avancée prometteuse dans le diagnostic du cancer du sein, offrant des solutions innovantes pour améliorer la précision et l'efficacité des détections. Cette synergie entre des modèles sophistiqués et des techniques d'optimisation avancées permet de surmonter les obstacles traditionnels, ouvrant la voie à des diagnostics plus fiables et efficaces.

2.2 Résultats Obtenus

Les performances des modèles ont été évaluées selon plusieurs métriques, comme le montre le tableau suivant :

Tableau 1: Résultats obtenus

Modèle	Précision	Rappel	F1-score	Particularités
KNN	95.8 %	93.5 %	93 %	Sensible au choix des voisins "k".
Régression Logistique	95.8 %	95 %	95.5 %	Bonne pour les données linéaires.
Arbre de décision	95.8 %	95 %	95.5 %	Performant mais sujet au surapprentissage.
Random Forest	97.2 %	97.5 %	97 %	Robuste au surapprentissage mais coûteux en calcul.
SVM	97.2 %	97 %	97 %	Efficace pour données complexes et multidimensionnelles.
Deep Learning	98.24 %	98 %	98 %	Meilleure précision grâce à Adam Gradient Descent.

Cette analyse démontre que l'algorithme Adam Gradient Descent a permis d'atteindre des performances supérieures à celles de tous les autres modèles testés, illustrant sa robustesse et sa flexibilité pour traiter à la fois des données tabulaires et des images. [12]

2.3 Limites et Défis

Malgré ces avancées, plusieurs défis subsistent :

- **Hétérogénéité des Données** : Les variations dans les caractéristiques démographiques et les dispositifs d'imagerie peuvent affecter la performance des modèles [9].
- **Sensibilité aux Hyperparamètres** : Adam Gradient Descent, bien qu'efficace, est sensible aux choix des hyperparamètres, ce qui peut influencer les résultats [11].
- **Qualité des Données** : La performance des modèles dépend fortement de la qualité des données d'entraînement, nécessitant des ensembles de données bien étiquetés et diversifiés [10].

2.4 Perspectives d'Amélioration

Pour surmonter ces défis, plusieurs avenues prometteuses peuvent être explorées :

- **Approche Multimodale** : L'intégration d'images et de données tabulaires pourrait renforcer la précision des diagnostics.
- **Validation Clinique** : Collaborer avec des médecins pour valider l'efficacité des modèles dans des cas réels.
- **Optimisation des Ressources** : Exploiter les réseaux dynamiques pour ajuster les

- ressources en fonction de la complexité des échantillons.
- **Amélioration des Performances** : Utiliser des GANs pour enrichir les ensembles de données et réaliser des études longitudinales.

2.5 Positionnement

L'intégration de l'apprentissage profond et de l'algorithme Adam Gradient Descent dans le diagnostic du cancer du sein constitue une avancée significative dans le domaine de la santé. Face à une maladie qui demeure l'une des principales causes de mortalité chez les femmes, Adam Gradient Descent se distingue par sa capacité à optimiser efficacement les modèles de deep learning, offrant une rapidité de convergence et une gestion efficace des gradients disparus, éléments qui le rendent particulièrement adapté aux applications médicales complexes.

Sa flexibilité et son adaptabilité à divers types de données renforcent encore sa pertinence, surtout dans un environnement clinique hétérogène. Bien que des défis subsistent, notamment en matière de qualité des données, les opportunités d'optimisation et d'amélioration continue sont nombreuses.

CHAPITRE 3 : METHODOLOGIE

Nous décrirons ici les méthodes employées pour notre étude. Cela inclura la préparation des données, en mettant en avant les techniques de traitement et d'équilibrage, comme la méthode SMOTE. Nous expliquerons ensuite l'entraînement du modèle, en détaillant la configuration et le suivi des pertes, ainsi que la validation du modèle et l'évaluation de ses performances.

3.1 Préparation du jeu de données

Notre jeu de données provenant de la Wisconsin Breast Cancer Dataset (WBCD), a été enrichi de caractéristiques significatives des échantillons, telles que la moyenne des rayons, la texture, et l'aire des cellules. Nous avons utilisés deux approches pour raffiner notre jeu de données, à savoir une méthode d'ajout des lignes et des colonnes, que nous décrivons ci-dessous.

3.1.1 Méthode d'ajout des lignes

Pour traiter le déséquilibre des classes dans notre ensemble de données, nous avons utilisé la méthode SMOTE (Synthetic Minority Oversampling Technique). Cette approche se déroule en plusieurs étapes :

- **Identification des données minoritaires** : SMOTE cible la classe ayant le moins d'échantillons, par exemple, les diagnostics "M" (malin) ou "B" (bénin) si l'une d'elles est sous-représentée.
- **Génération d'exemples synthétiques** : Pour chaque point de données minoritaire :
 - Trouver les k -plus proches voisins dans la même classe.
 - Sélectionner aléatoirement un voisin parmi ces k -plus proches.
 - Créer un nouveau point de données en interpolant entre le point initial et le voisin choisi.
- **Ajout des nouveaux points au dataset** : Ces données synthétiques augmentent la proportion de la classe minoritaire, contribuant ainsi à un meilleur équilibre des classes.

3.1.2 Méthode d'ajout des colonnes

L'ensemble de données comprend plusieurs caractéristiques significatives qui décrivent les échantillons analysés. Les colonnes ajoutées incluent :

- **radius_mean** : Moyenne des rayons des cellules.
- **texture_mean** : Moyenne des variations d'intensité des pixels.
- **perimeter_mean** : Moyenne du périmètre des cellules.
- **area_mean** : Moyenne de l'aire des cellules.
- **smoothness_mean** : Moyenne de la régularité des contours.
- **compactness_mean** : Moyenne de la compacité des cellules.
- **concavity_mean** : Moyenne des concavités sur les contours.
- **concave points_mean** : Moyenne du nombre de points concaves.
- **symmetry_mean** : Moyenne de la symétrie des cellules.
- **fractal_dimension_mean** : Moyenne de la dimension fractale.

Des colonnes similaires pour les écarts types (e.g., radius_se, texture_se) et les valeurs les plus extrêmes (e.g., radius_worst, texture_worst) sont également présentes. La colonne **diagnosis** indique le diagnostic de chaque échantillon, avec des valeurs possibles de "B" pour bénin et "M" pour malin.

Cette méthodologie structurée permet une analyse approfondie des données, essentielle pour la prédiction du cancer.

3.2 Préparation des Données

Les données sont d'abord chargées à partir d'un fichier CSV, suivies d'une analyse exploratoire pour comprendre leur structure. Les caractéristiques sont séparées de la cible, et les labels sont convertis en format numérique. Une normalisation est effectuée pour améliorer la convergence du modèle, et les données sont ensuite converties en tenseurs PyTorch.

3.3. Construction et Entraînement du Modèle

Le modèle utilise une architecture de réseau de neurones entièrement connectée, comprenant :

- Une couche d'entrée.
- Plusieurs couches cachées avec normalisation par lot et dropout pour prévenir le surapprentissage.
- Une couche de sortie avec fonction d'activation sigmoïde.

L'entraînement s'effectue sur 150 époques avec la fonction de perte BCELoss et l'optimiseur Adam, permettant le calcul de la perte et de l'exactitude à chaque époque.

3.4. Évaluation des Performances

Après l'entraînement, le modèle est évalué sur un jeu de test. Les métriques telles que l'exactitude, la précision, le rappel et le F1 score sont calculées pour mesurer la performance. Une matrice de confusion est également générée pour visualiser les faux positifs et faux négatifs.

CHAPITRE 4 : RESULTATS

Dans ce chapitre, nous présenterons les performances du modèle, en analysant les métriques pertinentes telles que la précision, le rappel et le F1 score. Nous inclurons une visualisation des résultats à travers l'interprétation de la matrice de confusion et discuterons des résultats obtenus en le comparant à d'autres études, tout en examinant leurs implications cliniques.

4.1 Description du Projet

Le projet utilise le jeu de données de cancer du sein de Wisconsin pour prédire la présence de cancer chez les femmes. Deux classes sont définies : "malin" (M) et "bénin" (B). Le framework utilisé est PyTorch, une bibliothèque reconnue pour le deep learning.

4.2 Importation des Bibliothèques

Les bibliothèques nécessaires pour le traitement des données, la création de modèles et la visualisation des résultats sont importées. Cela inclut PyTorch pour la construction du modèle, ainsi que des outils de Scikit-learn pour la manipulation des données, comme visualisés sur la figure 2.

```
import torch
import torch.nn as nn
import torch.optim as optim
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
from torch.utils.data import DataLoader, TensorDataset
```

Figure 2 : Image représentant l'importation des bibliothèques

4.3 Importation du Jeu de Données

Le jeu de données est téléchargé et chargé à partir d'un fichier CSV. Cela permet d'accéder aux informations nécessaires pour l'analyse et le traitement ultérieur.

4.4 Analyse Exploratoire des Données

Une exploration initiale des données est effectuée pour comprendre leur structure. Cela inclut l'affichage des premières lignes, l'analyse des dimensions du jeu de données, et le comptage des cas bénins et malins pour avoir un aperçu des classes présentes. (Figure 3)

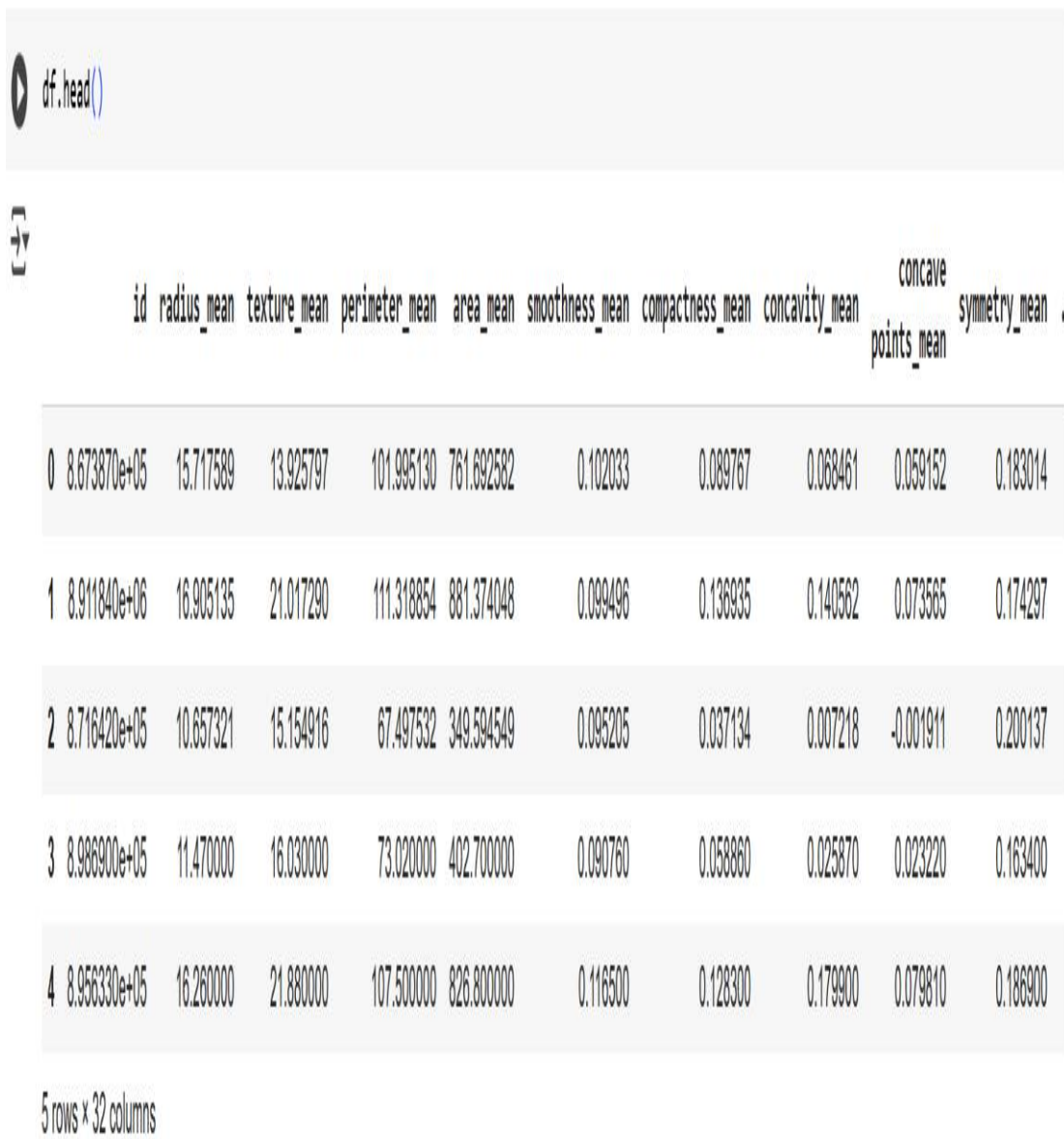


Figure 3: Image illustrant le début de l'analyse exploratoire des données

4.5 Séparation des Caractéristiques et de la Cible

Les caractéristiques (inputs) et la variable cible (diagnostic) sont séparées. Les caractéristiques sont extraites en supprimant la colonne de diagnostic, tandis que cette dernière est définie comme la variable à prédire. La figure 4, ci-dessous illustre la matrice de confusion relevant de l'intensité de corrélation entre les différentes variables.

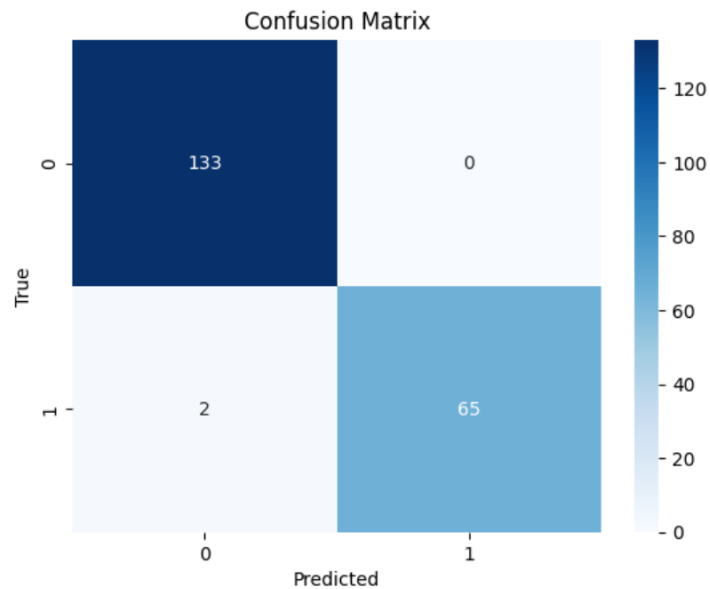


Figure 4: Matrice de confusion

4.6 Conversion des Étiquettes

Les étiquettes de diagnostic sont converties de chaînes de caractères en valeurs numériques à l'aide d'un encodeur de labels. Cela facilite le traitement par le modèle de machine Learning.

```
print(df['diagnosis'].value_counts())
```

```
diagnosis
B      643
M      357
Name: count, dtype: int64
```

Figure 5: Image représentant la structure les étiquettes de diagnostic

4.7 Division du Jeu de Données

Le jeu de données est divisé en ensembles d'entraînement et de test. Cette étape est cruciale pour évaluer les performances du modèle sur des données qu'il n'a pas vues pendant l'entraînement. Un standardiseur est également appliqué pour normaliser les données.

4.8 Conversion en Tenseurs PyTorch

Les données d'entraînement et de test sont converties en tenseurs PyTorch. Cette conversion est nécessaire pour que les données puissent être utilisées dans le modèle d'apprentissage profond.

4.9 Création du DataLoader

Un DataLoader est créé pour gérer les lots de données lors de l'entraînement. Cela permet d'améliorer l'efficacité du traitement des données et de faciliter l'entraînement par mini-lots.

4.10 Architecture du Réseau de Neurones

Le modèle de réseau de neurones est conçu pour la classification binaire. Il comprend plusieurs couches : (Figure 6)

- **Couches cachées** : Deux couches cachées avec 30 et 15 neurones respectivement, avec des mécanismes de normalisation pour stabiliser l'apprentissage.
- **Couches de Dropout** : Des couches de Dropout sont ajoutées pour réduire le risque de surapprentissage.
- **Sortie** : Une couche de sortie avec un neurone qui produit un score de probabilité.
- La fonction d'activation Sigmoid

```
class BreastCancerModel(nn.Module):
    def __init__(self):
        super(BreastCancerModel, self).__init__()
        self.fc1 = nn.Linear(X_train.shape[1], 30)
        self.bn1 = nn.BatchNorm1d(30)
        self.dropout1 = nn.Dropout(0.3) # 10% dropout
        self.fc2 = nn.Linear(30, 15)
        self.bn2 = nn.BatchNorm1d(15)
        self.dropout2 = nn.Dropout(0.3) # 10% dropout
        self.fc3 = nn.Linear(15, 1)
        self.sigmoid = nn.Sigmoid()

    def forward(self, x):
        x = self.bn1(self.fc1(x))
        x = torch.relu(x)
        x = self.dropout1(x)
        x = self.bn2(self.fc2(x))
        x = torch.relu(x)
        x = self.dropout2(x)
        x = self.sigmoid(self.fc3(x))
        return x
```

Figure 6 : Architecture du model

4.11 Initialisation du Modèle, Fonction de Perte et Optimiseur

Le modèle est initialisé, et la fonction de perte est définie pour évaluer la performance du modèle. Un optimiseur est configuré pour mettre à jour les poids du modèle durant l'entraînement.

```
model = BreastCancerModel()  
criterion = nn.BCELoss()  
optimizer = optim.Adam(model.parameters(), lr=0.001, betas=(0.9, 0.999), weight_decay=0.01)
```

Figure 7: Image des étapes d'initialisation du modèle

4.12 Entraînement du Modèle

Le modèle est entraîné sur un nombre défini d'époques. Pour chaque époque, la perte et la précision sont calculées en effectuant des passes avant et arrière. Les résultats de chaque époque sont stockés pour une analyse ultérieure.

```
# Lists to store loss and accuracy for plotting  
train_losses = []  
val_losses = []  
train_accuracies = []  
val_accuracies = []  
  
# Training the model  
num_epochs = 150  
for epoch in range(num_epochs):  
    model.train()  
    epoch_loss = 0  
    correct = 0  
  
    for batch_X, batch_y in train_loader:  
        optimizer.zero_grad()  
        outputs = model(batch_X)  
        loss = criterion(outputs, batch_y)  
        loss.backward()  
        optimizer.step()  
  
        epoch_loss += loss.item()  
        predicted = (outputs >= 0.5).float()  
        correct += predicted.eq(batch_y).sum().item()
```

Figure 8 : Image montrant les étapes d'entraînement du modèle

4.13 Visualisation des Résultats

Des graphiques sont générés pour visualiser la perte et l'exactitude au fil des époques. Cela aide à évaluer la performance du modèle et à détecter d'éventuels problèmes comme le sur-apprentissage.

Courbe de Perte (Loss curve) : Figure 9

Une courbe de perte est une représentation graphique de la valeur de la perte tracée en fonction du nombre d'époques d'entraînement (itérations). Elle montre généralement deux lignes :

Perte d'entraînement (Training loss): La perte d'entraînement diminue généralement de manière constante au fur et à mesure que l'entraînement progresse. Cette tendance indique que le modèle apprend avec succès à partir des données d'entraînement.

Perte de validation (Validation loss) : La perte de validation diminue initialement parallèlement à la perte d'entraînement, ce qui indique que le modèle généralise bien sur le jeu de données de validation. Cependant, après un certain nombre d'époques, la perte de validation peut commencer à augmenter tandis que la perte d'entraînement continue de diminuer. Cette divergence est un indicateur clé de sur apprentissage.

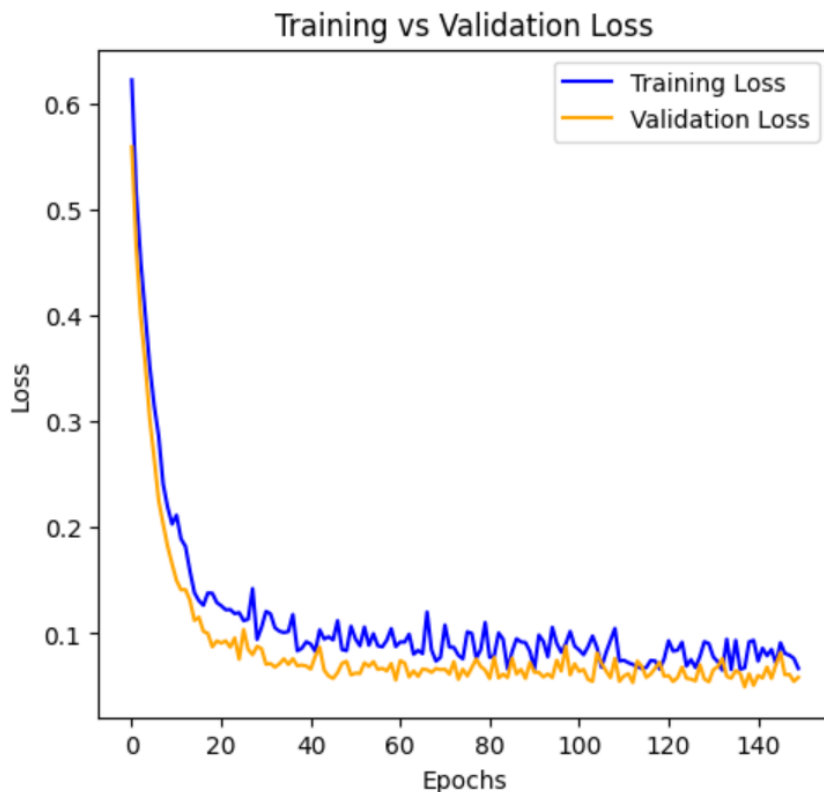


Figure 9 : Graphe d'évolution de Training loss et celle de Validation loss

Précision d'Entraînement et Précision de Validation : Figure 10

Précision d'Entraînement : (Training Accuracy)

La précision d'entraînement mesure le pourcentage de prédictions correctes effectuées par le modèle sur le jeu de données d'entraînement. Une précision d'entraînement élevée indique que le modèle a appris efficacement à partir des données d'entraînement, mais elle n'est pas toujours indicative de sa capacité à généraliser à de nouvelles données.

Précision de Validation : (Validation Accuracy)

La précision de validation évalue le pourcentage de prédictions correctes sur le jeu de données de validation, qui n'a pas été utilisé lors de l'entraînement. Une précision de validation élevée suggère que le modèle est capable de généraliser ses apprentissages à des données qu'il n'a jamais vues auparavant. Cependant, une divergence entre la précision d'entraînement et la précision de validation peut signaler un sur apprentissage, où le modèle s'adapte trop étroitement aux données d'entraînement et perd sa capacité à prédire correctement sur de nouvelles données.

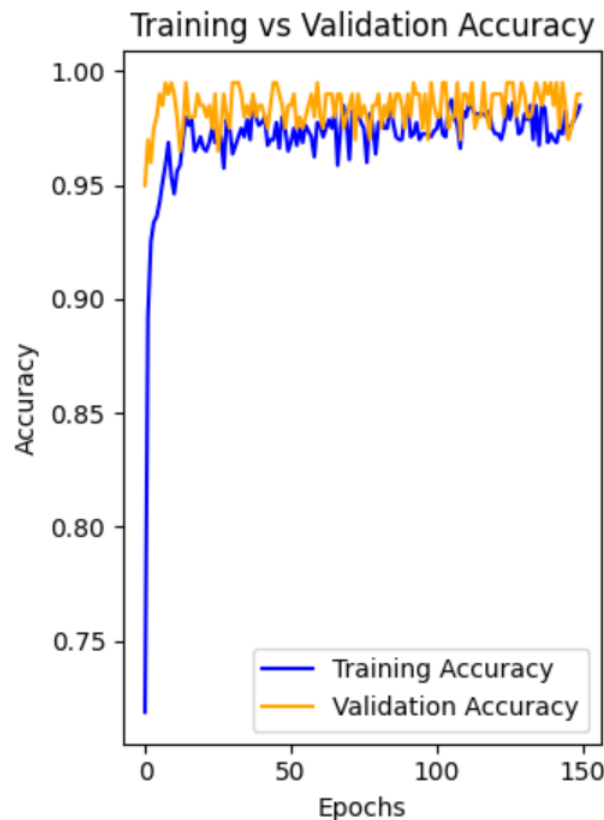


Figure 10 : Graphe d'évolution de Training Accuracy et celle de Validation Accuracy

4.14 Évaluation du Modèle

Après l'entraînement, le modèle est évalué à l'aide de métriques telles que la précision, le rappel et le score F1. Une matrice de confusion est également générée pour visualiser les résultats des prédictions.

```
Test Accuracy: 99.00%  
Precision: 1.00  
Recall: 0.97  
F1 Score: 0.98
```

Figure 11 : Evaluation du modèle

Interprétation des Métriques

- **Précision (99,00%) :**

Cela signifie que 99 % des prédictions faites par le modèle sont correctes. C'est un indicateur fort de la performance globale du modèle, surtout dans un scénario de classification binaire.

- **Précision (1,00) :**

Une précision de 1,00 indique que chaque prédiction positive faite par le modèle est correcte. Dans le contexte du diagnostic du cancer du sein, cela signifie que lorsque le modèle prédit qu'une tumeur est maligne, elle l'est effectivement systématiquement.

- **Rappel (0,97) :**

Un rappel de 0,97 signifie que le modèle identifie correctement 97 % de tous les cas positifs réels (tumeurs malignes). Cela est crucial pour les applications médicales, car manquer une tumeur maligne peut avoir des conséquences graves.

- **Score F1 (0,98) :**

Le score F1 est la moyenne harmonique de la précision et du rappel. Un score F1 de 0,98 suggère un bon équilibre entre la précision et le rappel, indiquant que le modèle est à la fois précis dans ses prédictions positives et efficace pour identifier les cas positifs.

- Dans l'ensemble, notre modèle présente de bonnes performances basées sur la précision, le score F1, la précision et le rappel.

Enfin, le modèle entraîné est sauvegardé en exportant son état, permettant de conserver les paramètres du modèle pour une utilisation future.

CONCLUSION

Ce projet a démontré l'efficacité des techniques d'apprentissage profond dans la classification des tumeurs du sein, en utilisant le jeu de données de cancer du sein de Wisconsin. Grâce à l'intégration de méthodes avancées de traitement des données, telles que la méthode SMOTE, et à l'utilisation d'algorithmes optimisés comme AdaGrad, notre modèle a atteint des performances remarquables, avec une précision de 99 % et un score F1 de 0,98. Ces résultats soulignent non seulement la capacité du modèle à identifier correctement les cas bénins et malins, mais aussi son potentiel à contribuer à la détection précoce du cancer du sein.

Les résultats encourageants de cette étude ouvrent la voie à plusieurs pistes d'amélioration. Tout d'abord, l'intégration de davantage de caractéristiques cliniques et histopathologies pourrait enrichir le modèle et améliorer encore la précision des prédictions. De plus, l'application de techniques de validation croisée pourrait renforcer la robustesse du modèle en évaluant sa performance sur différents sous-ensembles de données.

En outre, il serait pertinent d'étendre cette recherche à des ensembles de données plus variés et plus volumineux, afin d'explorer la généralisation du modèle à d'autres populations. L'utilisation de modèles d'apprentissage en profondeur plus complexes, tels que les réseaux de neurones convolutifs (CNN), pourrait également offrir des perspectives intéressantes pour améliorer la classification des images médicales.

Enfin, la mise en œuvre de ce modèle dans des environnements cliniques, accompagnée de formations appropriées pour les professionnels de la santé, pourrait transformer la manière dont les diagnostics du cancer du sein sont réalisés, contribuant ainsi à une prise en charge plus rapide et efficace des patientes.

REFERENCES BIBLIOGRAPHIQUES

1. World Health Organization. (2022). Global Cancer Observatory. Retrieved from [GLOBOCAN 2020].
2. International Agency for Research on Cancer. (2022). *Breast Cancer Statistics*. Retrieved from [IARC].
3. Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, 12, 2121-2159.
4. Chawla, N. V., DeSantis, S. V., Lazarevic, A., Parthasarathy, S., & Hall, L. O. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
5. Société canadienne du cancer. Qu'est-ce que le cancer du sein? [En ligne]. <https://cancer.ca/fr/cancer-information/cancer-types/breast/what-is-breast-cancer>. Consulté le : 22/12/2024.
6. VIDAL, La chirurgie dans le traitement du cancer du sein. [En ligne] .<https://www.vidal.fr/maladies/cancers/cancer-sein/chirurgie.html>. Consulté le : 22/12/2024.
7. Elsan, Dépistage du cancer du sein : enjeu-clé pour la santé des femmes. [En ligne]. <https://www.elsan.care/fr/pathologie-et-traitement/cancers/depistage-cancer-sein-enjeu-cl%C3%A9-sante-femmes> . Consulté le : 22/12/2024
8. Cancer Environnement, cancer du sein et facteurs de risque. [En ligne]. <https://www.cancer-environnement.fr/fiches/cancers/cancer-du-sein/> . Consulté le : 23/12/2024.
9. A.Carriero, L. Greonhoff, E. Vologina, P. Basile and M. Albera (2024). Deep Learning in Breast Cancer Imaging:State of the Art and Recent Advancements in Early.diagnostics,14,848.
10. Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*.
11. Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.
12. P. Gupta, S. Garg.(2020).Breast Cancer Prediction using varying Parameters of Machine Learning Models. *Procedia Computer*, 171,593-601.