

THEME :

**Intelligence Artificielle et détection
précoce du cancer du sein chez les
femmes au Cameroun.**

Mémoire en vue de l'obtention du diplôme de :
**Master II en Management des Projets Informatiques Option Intelligence
Artificielle et Big Data**

Présenté par :
TAKENDONG TESSA ARCEL JUNIOR

Encadrant Académique :
M. ABDOURAMAN DALIL

Yaoundé, Cameroun

Année Académique : 2022-2023

DEDICACE

A Tsakem Martine

REMERCIEMENTS

Cette œuvre n'aurait jamais pu voir le jour sans l'aide, le soutien, les précieux conseils et l'encadrement de près ou de loin de certaines personnes dont nous tenons à préciser les noms :

À tout le personnel Administratif de Keyce Informatique, en reconnaissance non seulement de leurs efforts inlassables en faveur de l'épanouissement intellectuel des étudiants, mais aussi pour l'opportunité qu'ils nous offrent d'étudier dans des conditions propices.

À tout le personnel de la Cellule Informatique du MINSANTE pour l'accueil chaleureux accordé, et l'aide dans notre processus de recherche.

À notre encadreur académique, Monsieur ABDOURAMAN DALIL, pour sa précieuse orientation et son inestimable disponibilité durant toute la réalisation de ce travail.

❖ À l'ensemble du corps enseignant de Keyce Informatique, pour leur dévouement sans faille dans la mission de dispenser une éducation de qualité à tous les étudiants.

❖ À M. et Mme TAKENDONG, pour leur soutien inébranlable et leur engagement sans faille en faveur de mon succès.

❖ À M. TAKENDONG Willy Pour sa disponibilité et ses efforts inlassablement fournis.

❖ À M. et Mme BAPACK, Pour leurs conseils et leur suivi continu.

❖ Au Pasteur JOB TENDE pour sa patience et ses prières à mon endroit.

❖ À M. TONGA Jordan pour sa disponibilité et son soutien indéfectible.

❖ À YOKO Corinne pour son soutien particulier.

❖ À M. DJAMEN, Mlle AMANA L, M. MOFFO, M. KUETETIO, M. et Mme ELLE, M. et Mme NDE, sans oublier M. et Mme LEKA, pour leur soutien quotidien et leurs encouragements sans relâche, sans lesquels ce travail n'aurait pas vu le jour.

❖ À tous mes amis et frères de la MDE pour le soutien durant cette période ;

❖ À tous ceux qui ont participé de près ou de loin à l'aboutissement de ce travail ;

A vous, lecteurs et à tous ceux dont les noms n'ont pas pu être cités, reconnaissez-vous dans ce travail.

RÉSUMÉ

L'intelligence artificielle (IA) occupe une place prépondérante dans plusieurs secteurs de recherche, en particulier dans le domaine de la santé. Les anomalies mammaires touchent près d'une femme sur neuf à l'échelle mondiale et représentent une préoccupation importante au Cameroun. Cette maladie est souvent l'une des principales causes de décès des femmes, car elle est fréquemment diagnostiquée à des stades avancés. L'objectif de cette étude était d'évaluer l'impact de l'IA sur la détection précoce du cancer du sein chez les femmes au Cameroun. L'étude a mis en évidence les spécificités du cancer du sein dans le contexte camerounais, mettant l'accent sur les obstacles à sa détection, tels que le manque d'information, les problématiques infrastructurelles, ainsi que les défis économiques et sociaux.

Dans une société où la médecine est souvent réactive, il est difficile pour de nombreux Camerounais d'envisager une démarche préventive. Face à cette réalité, il est primordial de proposer une solution adaptée au contexte socio-économique, en renforçant les connaissances sur le sujet sans heurter ou marginaliser les croyances locales. Cela permettrait d'encourager même les plus sceptiques à opter pour des méthodes de détection moins onéreuses que la mammographie traditionnelle. Les résultats de cette recherche ont conduit à l'élaboration d'une solution numérique basée sur l'IA. Le premier module de cette solution informe et engage une conversation dans un langage adapté à l'audience cible. Le second module, reposant sur un CNN, a pour but de classer les anomalies détectées à partir de simples photos d'échographie mammaire, guidant ainsi la décision médicale subséquente.

Mots clés : Cancer, Précoce, IA, Seins, CNN

ABSTRACT

Artificial intelligence (AI) plays a pivotal role in various research sectors, especially in the field of health. Breast abnormalities affect approximately one in nine women globally and are a significant concern in Cameroon. This disease is often one of the leading causes of death among women, as it is frequently diagnosed in advanced stages. The aim of this study was to assess the impact of AI on early breast cancer detection in Cameroonian women. The study highlighted the peculiarities of breast cancer in the Cameroonian context, focusing on barriers to its detection, such as a lack of information, infrastructural challenges, and economic and social hurdles.

In a society where medicine is often reactive, it is challenging for many Cameroonians to consider a preventive approach. Given this reality, it's crucial to offer a solution tailored to the socio-economic context, strengthening knowledge on the subject without offending or marginalizing local beliefs. This would encourage even the most skeptical to opt for detection methods less costly than traditional mammography. The findings of this research led to the development of a digital solution based on AI. The first module of this solution educates and engages in a conversation tailored to the target audience. The second module, based on a CNN (Convolutional Neural Network), aims to classify the anomalies detected from simple breast ultrasound images, thus guiding the subsequent medical decision.

Keywords: Cancer, AI, Breasts, Early, Detection.

SOMMAIRE

DEDICACE.....	II
RÉSUMÉ.....	IV
ABSTRACT	V
SOMMAIRE.....	VI
LISTE DES ABREVIATIONS	VII
INDEX DES FIGURES	VIII
INDEX DES TABLEAUX.....	X
INTRODUCTION GENERALE.....	11
CHAPITRE 1 : CADRE THEORIQUE ET CONCEPTUEL	3
CHAPITRE 2 : METHODOLOGIE DE RECHERCHE	44
CHAPITRE 3 : PRESENTATION DE LA SITUATION (PRESENTATION DES DONNEES COLLECTEES / DES RESULTATS).....	50
CHAPITRE 4 : ANALYSE - DIAGNOSTIC DE LA SITUATION ET PROPOSITION D'INTERVENTION	56
CONCLUSION GENERALE	95
REFERENCES BIBLIOGRAPHIQUES	96

LISTE DES ABREVIATIONS

AE : Auto-Encoder.

AUC: Area Under the Curve.

CAM: Class Activation Map.

CDBN: Continue Deep Belief Network.

CNN: Convolutional Neural Network.

CRBM : Continue Restricted Boltzmann Machines.

DBN: Deep Belief Network.

DBNALRBM: Deep Belief Networks all layers Restricted Boltzmann Machines.

DDSM: Digital Database for Screening Mammography.

DNN: Deep Neural Network.

GAN: Generative Adversarial Networks.

GAP: Global Average Pooling.

HCA: Hierarchical Cluster Analysis.

IA: Intelligence Artificielle.

LSTM: Long Short-Term Memory LSTM .

MDNNMD: Multimodal Deep Neural Network Multi-Dimensional Data.

MLP : Multi-layer Perceptron.

PCA : Principal Component Analysis.

RBM : Restricted Boltzmann Machines.

RNN : Réseaux Neuronaux Récurrents. ROC Receiver Operating Characteristic.

SAE: Stacked-Auto-Encoder.

SVM: Support Vector Machine.

VAE: Variational Autoencoders

WHO: World Health Organization.

INDEX DES FIGURES

Figure 1 La structure de sein (Kelley, 2020).	6
Figure 2 : Apprentissage par renforcement (Géron, 2017).	17
Figure 3 : Les approches et les algorithmes de l'apprentissage automatique (Li, 2017).	18
Figure 4 : Le choix de l'algorithme d'apprentissage selon certains facteurs (Li, 2017) .	19
Figure 5: Un perceptron multicouche.	21
Figure 6 : Les autos-encodeuses.	23
Figure 7 : Réseaux neuronaux récurrents (RNN) (Patterson & Gibson, 2017).	31
Figure 8 : Architecture LSTM (Patterson & Gibson, 2017).	31
Figure 9 : Fonctionnement global d'un LSTM.	32
Figure 10 : Les graphes d'état des problèmes liés à l'apprentissage (Géron, 2017).	34
Figure 11 :Architecture de l'approche proposée par (Fathy & Ghoneim, 2019).	36
Figure 12 :La courbe des caractéristiques d'exploitation du récepteur (Fathy & Ghoneim, 2019).	38
Figure 13: Architecture de l'approche proposée par (Sun, Wang, & Li, 2018)	38
Figure 14 : La courbe des caractéristiques d'exploitation du récepteur (Sun, Wang, & Li, 2018).	40
Figure 15Architecture de l'approche proposée par (Xiao, W_u, Lin, & Zhao,	41
Figure 16 Repartition du genre de la population questionnée.	56
Figure 17 tranches d'âges de la population étudiée	57
Figure 18 : statistiques sur la connaissance de ce qu'est le cancer du sein 1.	57
Figure 19 : statistiques sur la connaissance de ce qu'est le cancer du sein 2.	58
Figure 20 : statistiques sur la connaissance de ce qu'est le cancer du sein 3.	59
Figure 21 : statistiques sur la connaissance de ce qu'est le cancer du sein 4.	60
Figure 22 : statistiques sur la connaissance au sujet du cancer du sein 5	60
Figure 23 : statistiques sur la connaissance au sujet du cancer du sein 6	60
Figure 24 : statistiques sur la connaissance au sujet du cancer du sein 7	61
Figure 25 : statistiques sur la connaissance au sujet du cancer du sein 8	62
Figure 26 : statistiques sur la connaissance au sujet du cancer du sein 9.	63
Figure 27 : statistiques sur la connaissance au sujet du cancer du sein 10	64
Figure 28 : statistiques sur la connaissance au sujet du cancer du sein 11	65
Figure 29 : statistiques sur la connaissance au sujet du cancer du sein 12.	66

Figure 30 : statistiques sur la connaissance au sujet du cancer du sein 13.....	67
Figure 31 : Logo jupyter et Google Colab.....	74
Figure 32 : Logo Streamlit	75
Figure 33 : Logo Tensor Flow	76
Figure 34 : Logo Open AI	76
Figure 35 : Logo GitHub	77
Figure 36 : Logo Google Drive	77
Figure 37 : intégration de l'api de openAi.....	78
Figure 38 Aperçu des images du Data set	79
Figure 39 : Architecture du model.....	80
Figure 40 Classification report : Model 1.....	81
Figure 41 : Matrice de confusion Modele 1	81
Figure 42/ Metriques (F1 score , accuracy et ROC) du model 1	81
Figure 43 : Architecture model basée sur ResNet 50	83
Figure 44 : Training loss, Val loss, Acc, Val_acc du modele 2	84
Figure 45 Classification report ; Modele 2.....	84
Figure 46 : Matrice de confusion Modèle 2	85
Figure 47 : Modele CNN basé sur VGG 16	86
Figure 48 : Metrique du modele 3 (F1 Score Roc et Accuracy	86
Figure 49 : matrice de confusion du modèle 3	87
Figure 50 :Classification repport Modèle 3	87
Figure 51 : Interface d'accueil de DePreCaS.....	88
Figure 52 : Interface de prédiction de DePreCas.....	89
Figure 53 : Exemple de prédiction avec DePreCas	90

INDEX DES TABLEAUX

Tableau 1 : Facteurs de risques de cancer du sein (Courbiere & Carcopino, 2016).	7
Tableau 2 : Comparaison entre l'apprentissage automatique et l'apprentissage en profondeur (Grossfeld, 2020).	34
Tableau 3 : Les résultats de l'approche proposée (Fathy & Ghoneim, 2019).....	37
Tableau 4 : Information sur la base de données METABRIC (Sun, Wang, & Li, 2018).....	39
Tableau 5 : les caractéristiques avant et après le processus de sélection (Sun, Wang, & Li, 2018).....	39
Tableau 6 : Tableau comparatif des modèles entraînés.	88
Tableau 7 : budget estimatif pour la réalisation de notre solution	92

INTRODUCTION GENERALE

1. Contexte général de la recherche

Le cancer du sein est le cancer le plus fréquent chez les femmes au Cameroun, avec une incidence annuelle de 15 000 nouveaux cas et une prévalence de 25 000 malades. La maladie est souvent détectée à un stade avancé, à l'âge de 45-54 ans, ce qui limite les chances de guérison. L'intelligence artificielle (IA) est une technologie émergente dont les applications touchent plusieurs domaines ne laissant pas en reste le domaine de la santé. Cette recherche explore l'utilisation de l'IA pour améliorer la détection précoce du cancer du sein chez les femmes au Cameroun.

2. Problématique de la recherche

2.1. Présentation du problème

Au Cameroun, le cancer du sein est souvent détecté à un stade avancé, ce qui réduit considérablement les chances de guérison. La plupart des femmes ne disposent pas de l'accès aux examens de dépistage du cancer du sein tels que la mammographie, en raison de leur coût élevé et de leur inaccessibilité géographique. De plus, le manque de professionnels de santé qualifiés et l'absence de programmes de dépistage organisés sont des obstacles supplémentaires à la détection précoce du cancer du sein. Cela soulève la question de savoir comment améliorer la détection précoce du cancer du sein chez les femmes au Cameroun.

2.2. Formulation du problème (questions de recherche)

Question générale de recherche : l'IA pourrait-elle être utilisée pour améliorer la détection précoce du cancer du sein chez les femmes au Cameroun ?

Questions spécifiques :

- 1- L'IA serait-elle pertinente dans la détection des maladies dites incurables ?
- 2- Une méthodologie particulière serait-elle pertinente relativement à l'utilisation de l'IA dans la détection précoce du cancer du sein chez les femmes au Cameroun ?

3. Hypothèses de recherche

Pour mener une recherche scientifique rigoureuse, les hypothèses suivantes permettront de tester la validité de notre étude :

Hypothèse générale : l'IA pourrait être utilisée pour améliorer la détection précoce du cancer du sein chez les femmes au Cameroun.

Hypothèses spécifiques :

1. L'IA revêt un caractère pertinent et intéressant qui gagnerait à être exploité notamment dans la détection des maladies dites incurables.
2. Une méthodologie particulière serait certainement indiquée dans le cadre du Cameroun, étant donné les particularités sociologiques, politiques, et technologiques à prendre à compte dans un tel contexte en matière de détection précoce du cancer du sein chez les femmes au Cameroun.

4. Objectifs de recherche

Après avoir exposé le problème et planté les hypothèses de recherche, il est nécessaire de définir les objectifs de la recherche.

Objectif général : Déterminer quelle pourrait être la contribution de l'IA dans l'amélioration de la détection précoce du cancer du sein chez les femmes au Cameroun.

Objectifs spécifiques :

- 1- Déterminer la pertinence de L'IA dans la détection des maladies dites incurables.
- 2- Vérifier quelle serait l'approche la plus pertinente relativement à l'utilisation de l'IA dans la détection précoce du cancer du sein chez les femmes au Cameroun, étant donné les réalités sociologiques, politiques, et technologiques à prendre à compte dans un tel contexte.

5. Justification de la recherche

La recherche concernant l'utilisation de l'IA pour la détection précoce du cancer du sein chez les femmes camerounaises revêt une importance capitale. D'une part, elle permettra de mieux comprendre les avantages et les limites de l'IA dans ce contexte spécifique, ainsi que les facteurs influençant son utilisation. D'autre part, cette étude pourrait avoir des

implications pratiques majeures, pouvant contribuer à la réduction du taux de mortalité lié au cancer du sein chez ces femmes.

6. Délimitation de la recherche

Cette recherche sera limitée géographiquement au Cameroun et plus précisément sur un échantillon aléatoire de la population pris comme témoin. Thématiquement notre étude ne tiendra en compte que l'usage de l'IA comme soutien des méthodes empiriques médicales traditionnellement utilisées pour la détection précoce du cancer du sein. La recherche se concentrera sur l'analyse des avantages et des limites de l'utilisation de l'IA, ainsi que des facteurs qui influencent son utilisation pour la détection précoce du cancer du sein chez les femmes au Cameroun. Les autres méthodes de dépistage, telles que la mammographie et l'auto-examen du sein, ne seront pas incluses dans cette étude.

CHAPITRE 1 : CADRE THEORIQUE ET CONCEPTUEL

1. Introduction

Le cancer est l'une des principales causes de mortalité au niveau mondial. En 2015, il a été responsable de près de 8,8 millions de décès, ce qui représente presque un décès sur six à l'échelle globale (WHO, 2018). L'Organisation mondiale de la santé (OMS) projette une augmentation préoccupante du nombre de cas, avec une estimation qui pourrait atteindre 22 millions dans les deux prochaines décennies. De plus, environ 70% des décès par cancer sont enregistrés dans des pays à revenu faible ou intermédiaire.

Le terme "cancer", remontant à l'époque d'Hippocrate, englobe aujourd'hui plus de deux cents maladies ayant des caractéristiques communes, dont la croissance non régulée des cellules. Le cancer du sein, spécifiquement, représente une préoccupation majeure, étant le cancer le plus couramment diagnostiqué chez les femmes. Selon des statistiques, une femme sur neuf à travers le monde sera diagnostiquée de ce type de cancer au cours de sa vie.

Dans le sillage de l'ère technologique moderne, l'informatique s'infiltré dans de nombreux domaines, dont la santé. L'émergence de l'intelligence artificielle (IA) a donné lieu à des avancées significatives dans le secteur médical, allant du développement de matériel médical innovant à la création de logiciels d'analyse de haute précision.

Ce chapitre se décomposera en trois volets majeurs : Une étude approfondie du cancer du sein, où nous examinerons ses origines, symptômes, facteurs de risque et les statistiques spécifiques, notamment celles relatives au Cameroun.

Une exploration détaillée de l'intelligence artificielle, couvrant ses principaux mécanismes, techniques et applications variées.

Enfin, nous nous pencherons sur les travaux antérieurs basés sur l'utilisation de l'IA pour la prédiction et la détection précoce du cancer du sein.

Nous concluons ce chapitre en soulignant l'importance de la synergie entre la médecine et l'informatique, particulièrement dans la lutte contre les maladies mortelles comme le cancer du sein.

1. Le cancer du sein

Le cancer du sein demeure le cancer le plus fréquemment diagnostiqué chez les femmes d'un point de vue mondiale, aussi bien avant qu'après la ménopause. Selon des

statistiques¹ présentées par l'organisation mondiale de la Santé, le nombre de personnes diagnostiquées a légèrement augmenté au cours des trois dernières décennies. Cependant, le taux de mortalité a régulièrement diminué durant cette même période, en raison des avancées en matière de dépistage, de diagnostic et de traitement. Néanmoins, cette baisse ne se reflète pas systématiquement dans les pays développés, principalement à cause de l'allongement de l'espérance de vie.²

Il est à noter que les hommes ne sont pas épargnés par cette maladie, bien que moins fréquemment touchés : ils comptent pour environ 1 % de l'ensemble des cas diagnostiqués.

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

Le cancer du sein débute lorsque les cellules mammaires deviennent anormales et commencent à se multiplier de manière incontrôlée. Une tumeur dite "maligne" se forme alors, composée de ces cellules cancéreuses. Ces cellules peuvent non seulement envahir et endommager le tissu avoisinant, mais aussi se déplacer et affecter d'autres parties du corps, phénomène appelé "métastase"³.

Parfois, ces changements cellulaires du sein ne mènent pas à un cancer. Au lieu de cela, ils peuvent causer des problèmes non cancéreux, tels que des anomalies bénignes comme l'hyperplasie atypique ou des kystes. De plus, ils peuvent donner lieu à des tumeurs bénignes comme les papillomes intra-canalaire.

1.2. Types de cancer de sein

On peut distinguer plusieurs types du cancer de sein, selon l'endroit de la formation des cellules. Parmi ces types nous avons :

- Carcinome canalaire.
- Carcinome lobulaire.
- Le cancer inflammatoire du sein.

¹ Statistiques actuelles du Cancer du Sein : <https://www.who.int/fr/news-room/fact-sheets/detail/breast-cancer#:~:text=Plus%20de%202%2C2%20millions,le%20cancer%20le%20plus%20courant>. Retrieved 03.07, 2023

2 Courbiere, B., & Carcopino, X. (2016). KB Gynécologie obstétrique. Paris: vernazobres-grego. Retrieved 03.07, 2023

³ Kelley, S. (2020, 01 15). Qu'est-ce que le cancer du sein? Retrieved 02.08, 2023, from Société canadienne du cancer: <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/breast/breast-cancer/?region=qc>

- La maladie de Paget du sein.
- Le cancer du sein triple négatif.

1.3. Anatomie de Sein

Le sein est une structure complexe conçue pour produire et acheminer le lait à un nouveau-né. Sa composition et sa fonction sont influencées par diverses hormones présentes tout au long de la vie d'une femme.

Composition du sein⁴

Glandes et Lobules : Les glandes du sein, organisées en lobules, sont responsables de la production de lait. Chaque sein comporte entre 15 et 25 lobules. Durant la grossesse, les hormones stimulent ces glandes pour produire du lait.

Canaux : Ce sont des tubes qui acheminent le lait des lobules jusqu'au mamelon.

Mamelon : Situé au centre de l'aréole, c'est par le mamelon que le lait est évacué. Il est composé de fibres musculaires qui, lorsqu'elles se contractent, font pointer le mamelon vers l'extérieur.

Aréole : Zone ronde autour du mamelon, sa couleur varie entre le rose et le brun. Elle abrite de petites glandes qui sécrètent une substance huileuse servant de lubrifiant pour le mamelon.

Ligaments : Ces bandes de tissu conjonctif traversent le sein et le soutiennent en se fixant aux muscles du thorax.

Le rôle principal du sein, outre l'allaitement, est influencé par les hormones féminines, en particulier l'œstrogène et la progestérone, qui fluctuent lors d'événements tels que la puberté, la grossesse ou l'allaitement.

⁴ Kelley, S. (2020, 01 15). Qu'est-ce que le cancer du sein? Retrieved 03 08, 2023, from Société canadienne du cancer: <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/breast/breast-cancer/?region=qc>

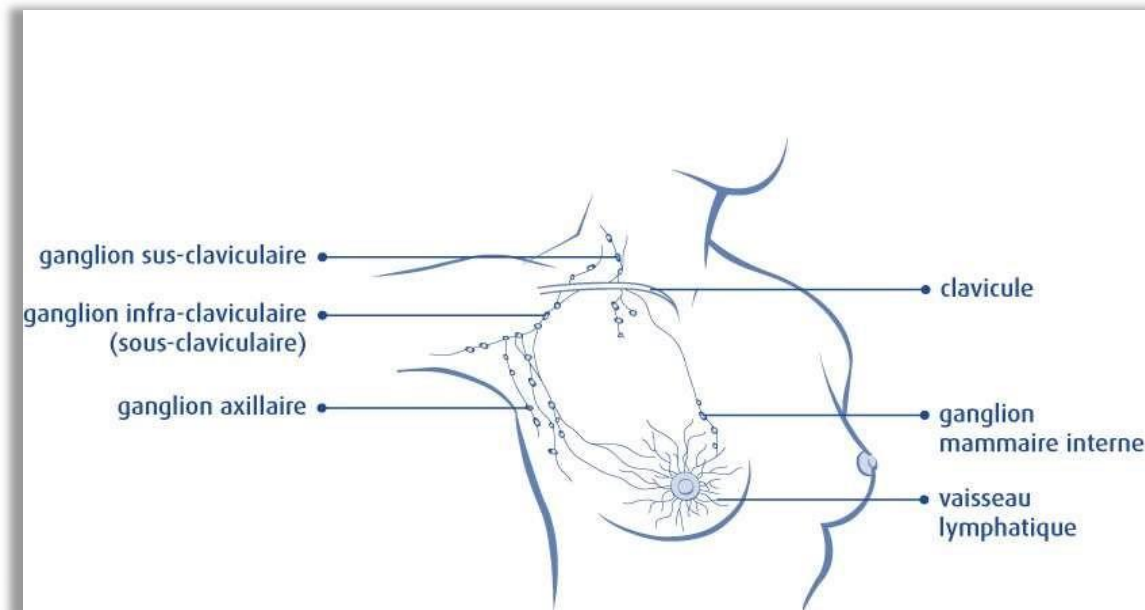


Figure 1 La structure de sein (Kelley, 2020).

1.4. Les symptômes et facteurs de risques de cancer du sein

Le cancer du sein ne pourrait présenter aucun symptôme lors de ses premiers stades. Cependant, il existe des signes distinctifs qui peuvent indiquer la présence de cette maladie. De plus, certains facteurs augmentent le risque de développer un cancer du sein.

1.4.1. Les symptômes

D'après l'OMS, les symptômes du cancer du sein varient d'une personne à une autre. Nous pouvons out de même les résumer comme suit :

- ✓ Une bosse au sein, qu'elle soit fixe ou mobile. Il s'agit du symptôme le plus fréquent, pour les femmes et les hommes.
- ✓ Des écoulements spontanés provenant du mamelon
- ✓ Une rétraction du mamelon « le mamelon est tourné vers l'intérieur ».
- ✓ Un changement d'apparence de la peau d'un sein : épaissement ou durcissement « peau d'orange », rougeur inhabituelle, chaleur, changement « desquamation » de la peau autour du mamelon.
- ✓ Un changement inhabituel de grosseur ou de la forme d'un sein.

1.4.2. Facteurs de risques

Un certain nombre d'éléments non modifiables peuvent augmenter le risque d'atteindre le cancer du sein. Ces éléments sont un mélange de causes hormonales, familiales et environnementales. Il faut noter toutefois qu'un cancer du sein peut se former dans l'absence de tous ces facteurs de risque. Ces facteurs sont : ⁵ (voir Tableau 1) :

Facteurs HORMONAUX	Puberté précoce. 1 ^{ère} grossesse tardive (> 35 ans). Absence d'allaitement. Nulliparité. Ménopause tardive (> 52 ans). Obésité.
Facteurs FAMILIAUX	Antécédents familiaux de cancer du sein Facteurs génétiques : Mutation de chromosome 17. Mutation de chromosome 13.
Facteurs HISTOLOGIQUES	Hyperplasies canalaire atypiques. Néoplasies lobulaires.
Facteurs ENVIRONNEMENTAUX	Niveau socio-économique élevé. Les pays industrialisés.

Tableau 1 : Facteurs de risques de cancer du sein (Courbiere & Carcopino, 2016).

1.5. Les traitements médicaux du cancer du sein

Le traitement du cancer du sein varie en fonction du type de cancer et de son stade d'évolution. La chirurgie est couramment utilisée pour traiter une grande majorité des cancers du sein. Cependant, avant de discuter des traitements, il est essentiel de souligner l'importance de la biopsie. La biopsie consiste à prélever un échantillon de la tumeur pour l'examiner. Les résultats de ces tests détermineront le traitement approprié. La première étape est de confirmer si la tumeur est cancéreuse et de déterminer le type de cancer. Il existe principalement cinq catégories de traitements pour le cancer du sein

⁵ Courbiere, B., & Carcopino, X. (2016). KB Gynécologie obstétrique. Paris: vernazobres-grego Retrieved 03 .07, 2023

- ✓ **Chirurgie** : Elle est souvent la première ligne de traitement et vise à enlever la tumeur.
- ✓ **Radiothérapie** : Après une mastectomie partielle, une radiothérapie est souvent recommandée pour éliminer les cellules cancéreuses résiduelles. Elle réduit le risque de réapparition de la tumeur
- ✓ **Chimiothérapie** : La chimiothérapie, qui utilise des médicaments antinéoplasiques, est couramment administrée après la chirurgie pour traiter les cancers du sein.
- ✓ **Thérapie hormonale** : Elle est souvent combinée à la chirurgie, la radiothérapie, ou la chimiothérapie.
- ✓ **Thérapie ciblée** : Cette approche implique l'administration d'un médicament qui bloque spécifiquement l'action du gène HER2, généralement par injection intraveineuse.

2. Intelligence Artificielle

Dans le contexte de notre mémoire, il est essentiel de noter que l'intelligence artificielle (IA) s'est développée en une discipline complexe et diversifiée, composée de multiples sous-disciplines. Deux de ces branches ont récemment pris une place prépondérante : le Machine Learning et le Deep Learning. Bien qu'étroitement interconnectés, ces deux domaines présentent des caractéristiques et des applications qui les différencient de manière significative.

Le Machine Learning, en particulier l'apprentissage supervisé, propose des méthodes robustes permettant d'analyser et de prédire des données en s'appuyant sur des informations historiques. D'un autre côté, le Deep Learning puise son inspiration dans la structure du cerveau humain pour concevoir des réseaux de neurones artificiels capables de traiter des données complexes à un niveau de sophistication sans précédent.

Au sein de cette section de notre mémoire, nous entreprendrons une exploration approfondie du Machine Learning avant de plonger dans les profondeurs du Deep Learning. Cette analyse nous permettra d'examiner leurs bases fondamentales, leurs avancées récentes et leur impact considérable sur le vaste domaine de l'IA.

2.1.L'apprentissage automatique « Machine Learning »

Au cours de la dernière décennie, l'intérêt pour l'apprentissage automatique a connu une croissance significative. Malgré tout le discours positif autour de cette technologie, il existe des tensions entre les capacités réelles des machines et nos attentes⁶. L'apprentissage automatique, en tant que sous-domaine de l'intelligence artificielle (IA), se concentre sur le développement de systèmes capables d'apprendre et d'améliorer leurs performances en se basant sur les données qu'ils traitent. Les algorithmes d'apprentissage automatique sont essentiels pour optimiser, simplifier et sécuriser ce processus⁷.

⁶ Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach. Retrieved 03 .07, 2023

⁷ Clayton, R. (2019). Qu'est-ce que le Machine Learning ? Retrieved 15.08, 2023, from Oracle Algeria:

2.1.1. Définitions et types d'apprentissage automatique

➤ Définitions

La définition de l'apprentissage automatique a évolué au fil des années, reflétant les avancées et découvertes constantes dans ce domaine :

En Mars 1959, l'informaticien américain Arthur Samuel introduit la notion d' "Apprentissage automatique". Il a conçu le premier programme permettant aux ordinateurs d'apprendre à jouer aux dames et même de jouer, tout cela sans une programmation explicite⁸.

En 1997, Tom Michael Mitchell lui aussi informaticien américain a proposé une nouvelle définition de l'apprentissage automatique. Il l'a défini comme le processus par lequel un programme apprend à partir de l'expérience (E) par rapport à une classe de tâches (T), en utilisant une mesure de performance (P) ⁹.

Au fil du temps, la définition de l'apprentissage automatique s'est penché vers des niveaux de plus en plus mathématiques et statistiques. Selon les auteurs Goodfellow, Bengio et Courville en 2016, l'apprentissage automatique est essentiellement une forme de statistiques appliquées, mettant l'accent sur l'utilisation d'ordinateurs pour estimer statistiquement des fonctions complexes, avec moins d'importance accordée à démontrer des intervalles de confiance autour des fonctions en question.

Ces définitions peuvent varier en fonction de la perspective, mais elles convergent toutes vers une idée centrale : l'apprentissage automatique est l'art ou la science de programmer des ordinateurs pour qu'ils puissent apprendre à partir de données¹⁰.

➤ Types d'apprentissage automatique

Les différents types d'apprentissage automatique peuvent être définis en répondant à deux questions clés, comme le mentionne Géron en 2017 :

⁸ Géron, A. (2017). *Hands-On_Machine_Learning_with_Scikit-learn and tensorflow*. USA: O'Reilly Media. Retrieved 03 .07, 2023

⁹ Géron, A. (2017). *Hands-On_Machine_Learning_with_Scikit-learn and tensorflow*. USA: O'Reilly Media. Retrieved 03 .07, 2023

¹⁰Géron, A. (2017). *Hands-On_Machine_Learning_with_Scikit-learn and tensorflow*. USA: O'Reilly Media. Retrieved 03 .07, 2023

- 1- Est-ce que cet apprentissage dépend de la supervision humaine pendant son entraînement et son apprentissage ?
- 2- Est-ce que ce type d'apprentissage utilise une base de données fournie par l'être humain ?

En fonction des réponses à ces questions, nous pouvons classer les types d'apprentissage de la manière suivante :

✓ Apprentissage supervisé :

Si la réponse est "oui" aux deux questions, nous sommes en présence d'un apprentissage supervisé. Dans ce cas, le modèle d'apprentissage automatique est formé à l'aide de données étiquetées fournies par des humains. Ces étiquettes servent de guides pour que le modèle puisse apprendre à associer correctement les entrées aux sorties souhaitées.

✓ Apprentissage non supervisé :

Si la réponse est "non" à la première question (pas de supervision humaine) et "oui" à la deuxième question (utilisation d'une base de données humaine), nous sommes en présence d'un apprentissage non supervisé. Dans ce cas, le modèle cherche à découvrir des structures, des schémas ou des regroupements intrinsèques dans les données sans avoir d'étiquettes prédéfinies. Il explore les données pour en extraire des informations utiles de manière autonome.

✓ Apprentissage par renforcement :

Si la réponse est "non" aux deux questions (pas de supervision humaine ni utilisation d'une base de données fournie par l'être humain), alors il s'agit de l'apprentissage par renforcement. Dans ce type d'apprentissage, un agent apprend à prendre des décisions séquentielles en interagissant avec un environnement. L'agent reçoit des récompenses ou des punitions en fonction de ses actions, ce qui lui permet d'apprendre quelles actions sont les plus appropriées pour atteindre un objectif spécifique.

Ainsi, ces trois types d'apprentissage automatique, supervisé, non supervisé et par renforcement couvrent une gamme variée d'approches pour former des modèles d'IA en fonction du degré de guidance humaine et de la nature des données utilisées.

2.1.2. Apprentissage supervisé

En ce qui concerne l'apprentissage supervisé, l'algorithme est guidé par un être humain tout au long de son processus d'apprentissage, similaire à un data scientist qui montre à l'algorithme les résultats attendus. C'est aussi comparable à l'enseignement d'un enfant sur comment reconnaître des objets en leur montrant des exemples. Cet apprentissage se base sur un ensemble de données préalablement étiquetées, où les résultats souhaités sont déjà connus¹¹.

Les algorithmes d'apprentissage automatique supervisé les plus couramment utilisés sont les suivants :

➤ **Classification :**

La classification consiste à attribuer une catégorie ou une classe à une entrée quelconque, en se basant sur ses caractéristiques ou ses attributs. Par exemple, classer des e-mails comme spam ou non-spam.

➤ **Régression :**

La régression vise à prédire une valeur continue pour la variable de sortie¹². Par exemple, prédire le prix d'une maison en fonction de ses caractéristiques.

Les algorithmes les plus célèbres utilisés incluent¹³ :

¹¹ Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press . Retrieved 23 .07, 2023

¹² Dave, A. (2020, 03 07). Regression in Machine Learning. Retrieved 04 09, 2023, from Medium

¹³ .GAËL. (2019, 10 25). Machine Learning. Retrieved 04 09, 2023, from Datakeen: <https://datakeen.co/8-machine-learning-algorithms-explained-in-human-language/>

➤ **SVM (Machines à vecteurs de support) :**

Ces techniques cherchent à trouver une séparation optimale entre différentes catégories en identifiant les vecteurs de support, c'est-à-dire les points de données les plus proches de chaque catégorie. Ils peuvent être utilisés pour la classification et la régression.

➤ **Méthode des k plus proches voisins :**

Cette méthode classe un nouvel élément en se basant sur la majorité des "k" éléments existants les plus similaires. Elle est utilisée pour la classification (parfois la régression). Un exemple serait de déterminer la race d'un chien en se basant sur des caractéristiques telles que sa taille, son poids et la couleur de sa fourrure.

➤ **Naïve Bayes :**

En se basant sur le théorème de Bayes, le modèle Naïve Bayes est utilisé pour calculer la probabilité qu'une donnée appartienne à une catégorie particulière en supposant que chaque caractéristique (ou attribut) de la donnée est statistiquement indépendante des autres. Cette supposition d'indépendance est souvent appelée "naïve" parce qu'elle simplifie considérablement les calculs tout en étant une approximation.

Il est principalement utilisé pour la classification. Un exemple serait de filtrer les messages spam en se basant sur la fréquence des mots.

➤ **Arbres de décision :**

Ces modèles utilisent une structure arborescente pour prendre des décisions. Ils posent des questions sur les caractéristiques des données à chaque nœud pour aboutir à une prédiction. Ils peuvent être utilisés pour la classification et la régression. Un exemple serait de décider d'accorder ou non un prêt bancaire en fonction du revenu, de l'âge et de l'historique de crédit du demandeur.

➤ **Random Forests:**

Il s'agit des forêts aléatoires (Random Forests), ce sont une extension des arbres de décision en apprentissage automatique. Au lieu de construire un seul arbre de décision, les forêts aléatoires construisent plusieurs arbres de décision indépendants et combinent leurs prédictions pour obtenir une réponse finale, généralement plus robuste et précise. Ils sont utilisés pour la classification et la régression. Un exemple serait de prédire le prix d'une maison en se basant sur des caractéristiques telles que la taille, le nombre de chambres et l'emplacement.

➤ **Régression Logistique :**

Malgré son nom, la régression logistique est en réalité un modèle de classification couramment utilisé en apprentissage automatique. Ce modèle est utilisé pour évaluer la probabilité qu'une donnée soit affiliée à une catégorie spécifique, en particulier dans le cadre de problèmes de classification binaire (deux classes) ou de classification multiclasse (plus de deux classes). Un exemple serait d'estimer la probabilité qu'un utilisateur clique sur une publicité en ligne.

➤ **Réseaux de neurones :**

Inspirés du fonctionnement du cerveau humain, ces modèles comportent des couches de neurones interconnectés capables de reconnaître des motifs complexes dans les données. Ils sont utilisés pour la classification, la régression et le clustering, entre autres. Un exemple serait d'identifier les visages sur une photographie.

Ces algorithmes sont largement utilisés dans divers domaines pour résoudre une multitude de problèmes en exploitant la puissance de l'apprentissage automatique supervisé.

2.1.3. Apprentissage non supervisé

L'apprentissage non supervisé représente une approche autonome de l'apprentissage par les machines, où la machine apprend à découvrir des structures et des modèles complexes dans

les données sans recourir à une assistance humaine. Cette méthode se fonde sur des données non étiquetées, ne contenant aucune information spécifique sur les résultats¹⁴.

Parmi les techniques de l'apprentissage non supervisé, nous pouvons distinguer les suivantes:

➤ **Regroupement (Clustering) :**

Cette méthode d'analyse statistique vise à regrouper des données brutes en ensembles homogènes, où les éléments au sein de chaque groupe partagent des caractéristiques similaires.

➤ **Réduction de la dimension :**

L'objectif ici est de simplifier les données en réduisant leur complexité sans perdre d'informations importantes. Cela peut inclure la combinaison de plusieurs caractéristiques en une seule.

Parmi les algorithmes les plus renommés utilisés dans cette approche, nous pouvons citer les suivants (Issarane, 2019) :

➤ **K-Means :**

Un algorithme de regroupement qui réunit des données similaires dans des clusters, en utilisant un processus itératif pour produire un résultat final.

➤ **Analyse de classification hiérarchique (HCA) :**

Cette méthode crée une hiérarchie de clusters, permettant une flexibilité en termes du nombre de clusters souhaités.

¹⁴ Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press . Retrieved 03 .08, 2023

➤ **PCA (Analyse des composants principaux) :**

Cet algorithme transforme des variables liées en nouvelles variables indépendantes. Ces nouvelles variables sont les composantes principales, elles permettent au praticien de réduire ainsi la redondance de l'information et le nombre de variables.

➤ **A priori :**

Il est couramment employé dans le contexte des bases de données transactionnelles pour identifier des ensembles d'éléments qui apparaissent fréquemment et pour créer des règles qui décrivent des associations entre ces éléments.

L'apprentissage non supervisé est précieux pour découvrir des informations cachées dans les données, pour explorer des structures complexes, et pour aider à la prise de décision lorsque les résultats attendus ne sont pas clairement définis. Ces algorithmes sont largement utilisés dans des domaines tels que la segmentation de la clientèle, l'analyse de texte, et bien d'autres où la compréhension des motifs dans les données est essentielle.

2.1.4. L'apprentissage par renforcement

Avec l'apprentissage par renforcement, la machine n'a pas besoin de l'aide de l'être humain, ni en termes de supervision, ni en termes de fourniture de données.

L'apprentissage par renforcement est une approche distincte en matière de machine learning. Dans ce cadre, le système en apprentissage, désigné sous le nom d'« agent », interagit avec son environnement. L'agent a la capacité d'observer cet environnement, de décider des actions à entreprendre, puis d'exécuter ces actions. En retour de ses actions, il reçoit des récompenses ou des sanctions. Ces retours, qu'ils soient positifs (récompenses) ou négatifs (sanctions), guident l'agent dans l'élaboration de ses stratégies futures.

L'objectif de l'agent est de découvrir la meilleure stratégie, ou « politique », qui maximisera la somme des récompenses qu'il recevra sur le long terme. Cette politique sert de guide à

l'agent, lui indiquant quelle action il devrait privilégier en fonction de la situation ou de l'état actuel de l'environnement dans lequel il se trouve ¹⁵

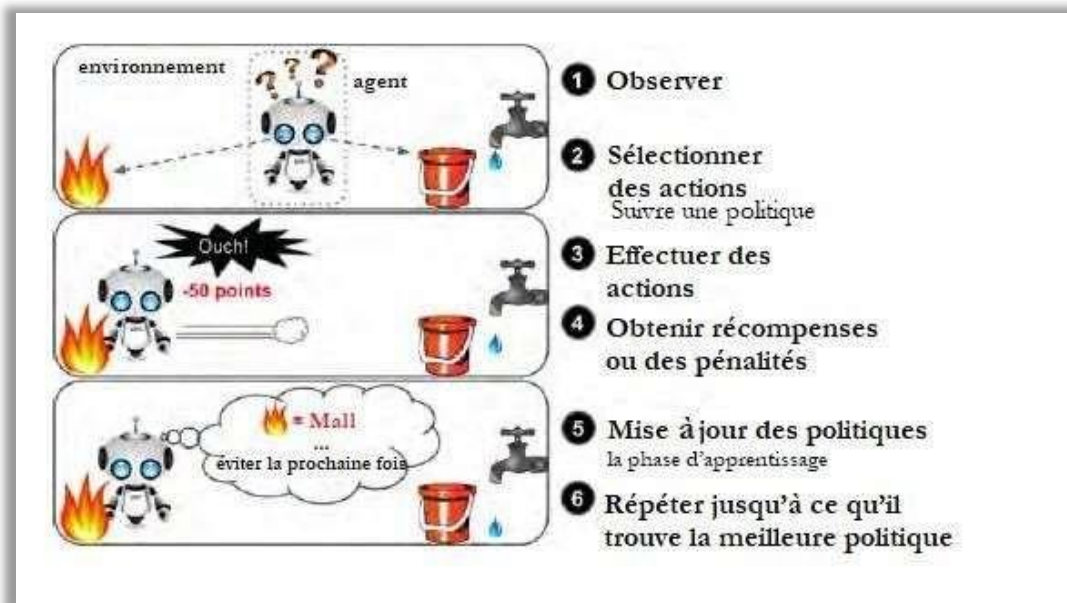


Figure 2 : Apprentissage par renforcement (Géron, 2017).

Parmi les premiers algorithmes d'apprentissage par renforcement, nous avons « Temporal différence Learning », proposé par « Richard Sutton » en 1988¹⁶. Aussi l'algorithme « Q-Learning » mis au point lors d'une thèse soutenue par « Chris Watkins » en 1989 et publié réellement en 1992¹⁷.

La figure ci-dessous exprime les différentes branches et algorithmes de l'apprentissage automatique.

¹⁵Géron, A. (2017). Hands-On_Machine_Learning_with_Scikit-learn and tensorflow. USA: O'Reilly Media. Retrieved 23 .07, 2023

¹⁶ Sutton, R. S. (1988). Learning to predict by the method of temporal differences. Machine Learning. Retrieved 15 .07, 2023

¹⁷ Watkins, C. J., & Dayan, P. (1992). Q-learning. Machine Learning. Retrieved 13 .07, 2023

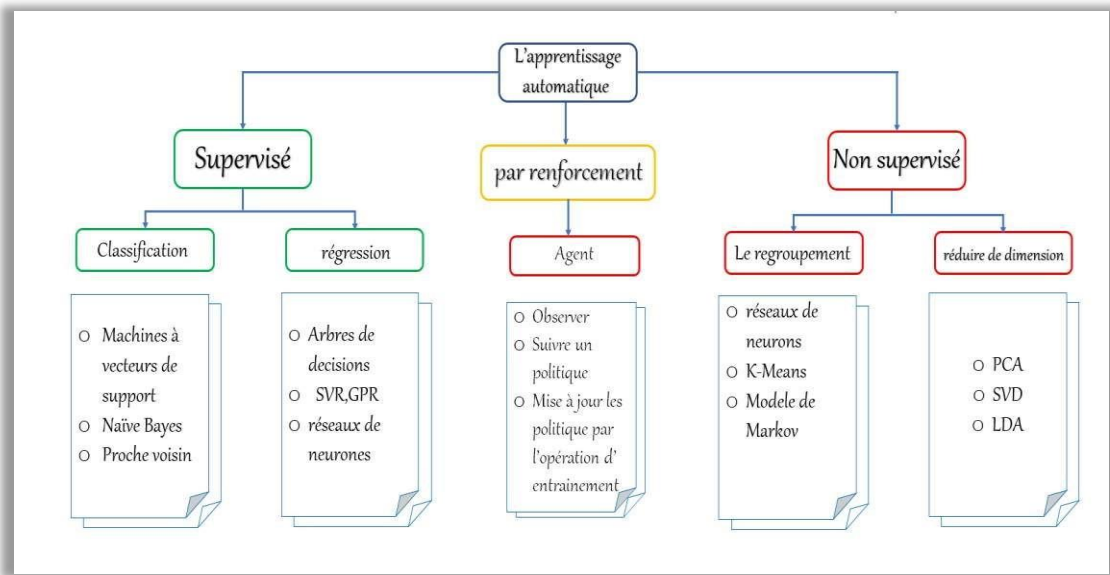


Figure 3 : Les approches et les algorithmes de l'apprentissage automatique (Li, 2017).

2.1.5. Le choix d'un type d'apprentissage automatique

Avec la présence de différents types de classifieurs pour l'apprentissage automatique, l'opération de choix d'un type est une question typique « Quel algorithme dois-je utiliser ? ».

Selon (Li, 2017), la réponse à cette question varie selon les facteurs suivants :

- La taille, la qualité et la nature des données.
- Le temps de calcul disponible.
- L'urgence de la tâche.
- Le but d'utilisation de ces données.

La figure suivante, fournir des indications sur les algorithmes à essayer en premier selon les facteurs mentionnés ci-dessus.

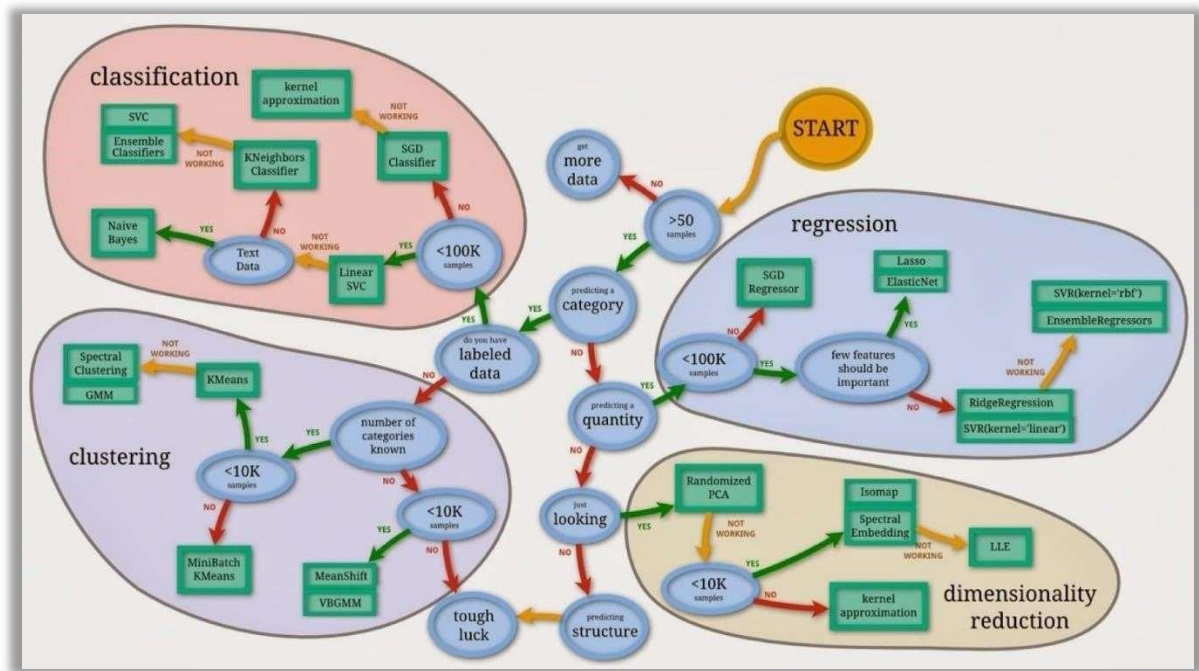


Figure 4 : Le choix de l'algorithme d'apprentissage selon certains facteurs (Li, 2017) .

2.2. De l'apprentissage automatique à l'apprentissage en profondeur

L'efficacité de l'apprentissage automatique dans la résolution d'une multitude de problèmes est indiscutable. Néanmoins, cette approche rencontre de nombreux défis. Parmi les plus prégnants, on compte la gestion du temps, la rapidité d'exécution et la précision. Si un algorithme d'apprentissage automatique produit une prédiction erronée, un ingénieur doit alors intervenir pour procéder à des corrections. Une telle intervention est le signe d'un manque d'efficacité, peut engendrer des retards et ralentir les prévisions.

Les algorithmes basiques d'apprentissage automatique présentés dans ce chapitre se montrent performants pour un large éventail de défis significatifs. Cependant, ils peinent à aborder les problématiques majeures de l'intelligence artificielle, comme la reconnaissance vocale ou la détection d'objets. L'essor de l'apprentissage en profondeur découle en partie de ces limites. Les méthodes traditionnelles, confrontées à des jeux de données volumineux, ont du mal à généraliser efficacement dans ces domaines cruciaux de l'intelligence artificielle¹⁸.

¹⁸Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press .

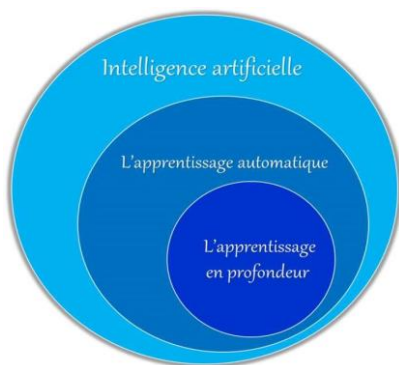
2.3. L'apprentissage profond « Deep Learning »

L'apprentissage en profondeur « Deep Learning » a été un défi à définir pour beaucoup de spécialistes dans le domaine car il a changé de forme lentement au cours de la dernière décennie. Une définition utile précise que l'apprentissage en profondeur est un réseau neuronal avec plus de deux couches¹⁹. Le problème avec cette définition est qu'elle fait écho à l'existence de ce domaine depuis les années 80 du siècle dernier, formant ainsi une grande contradiction puisque beaucoup de gens pensent que ce domaine est relativement nouveau, pour réfuter cette contradiction, il faut distinguer le moment où le domaine est apparu et celui où il a été cadré et exploité.

2.3.1. Définitions et les architectures d'apprentissage en profondeur

➤ Définitions

Le terme « apprentissage profond » a été introduit dans le domaine de l'apprentissage automatique par « Rina Dechter » en 1986, et dans les réseaux de neurones artificiels par « Igor Aizenberg » et ses collègues en 2000, dans le contexte des neurones à seuil booléen²⁰, l'apprentissage profond désigne une technique d'apprentissage d'une machine, c'est une sous-branche de l'intelligence artificielle qui vise à construire automatiquement des connaissances à partir de grandes quantités d'information. Les caractéristiques essentielles du traitement ne seront plus identifiées par un traitement humain dans un algorithme au préalable, mais directement par l'algorithme d'apprentissage profond.



¹⁹ Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach.

²⁰ Schmidhuber, J. (2015). Deep Learning. Scholarpedia, 10, 32832.

Les sous-branches de l'intelligence artificiel²¹. L'apprentissage en profondeur permet donc implicitement de répondre à des questions du type « que peut-on déduire de ces données ? » et décrire des caractéristiques parfois cachées ou des relations entre des données souvent impossibles à identifier pour l'homme.

D'après (Patterson & Gibson, 2017) l'apprentissage profond est un réseau neuronal avec un grand nombre de paramètres et de couches, l'exemple de base c'est le perceptron multicouche MLP « multi layer perceptron »

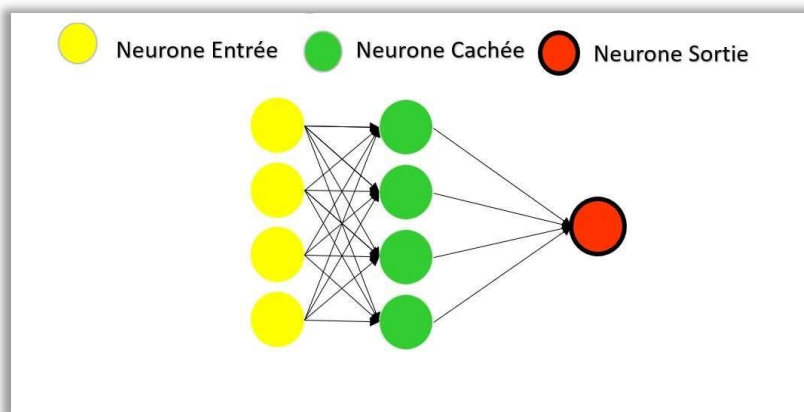


Figure 5: Un perceptron multicouche²².

Perceptrons a été inventé en 1958 au « Cornell Aviation Laboratory » par « Frank Rosenblatt » financé par le bureau de recherche navale des États-Unis, Le mot vient de verbe latin « Percipio » qui signifie en Anglaise understand ; en Français comprendre, qui montre que Le robot ou l'appareil peut apprendre et comprendre le monde extérieur²³.

Un perceptron multicouche avec plusieurs couches cachées entre la couche d'entrée et la couche de sortie est un réseau de neurones profonds (DNN), le DNN est une fonction mathématique, qui mappe certains ensembles de valeurs d'entrée aux valeurs de sortie. La fonction est formée par la composition de nombreuses fonctions plus simples ²¹.

Certaines de ses caractéristiques²²:

²¹ Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press . © Retrieved july 08, 2023

²²Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach. Retrieved july 08, 2023

²³ Rosenblatt, F. (1958). THE PERCEPTRON: A PROBABILISTIC MODEL FOR. Psychological Review, 65 Retrieved july 08, 2023

- Plus de neurones.
- Des moyens plus complexes de connecter les couches neurones dans les réseaux neuronaux.
- Puissance de calcul.
- Extraction automatique des fonctionnalités.

L'apprentissage profond s'applique dans divers domaines (Goodfellow, Bengio, & Courville, 2016), tel que :

- L'intelligence artificielle en général.
- La reconnaissance visuelle et la comparaison de forme.
- La robotique.
- La santé et la bio-informatique.
- La sécurité.

- Les architectures d'apprentissage en profondeur

Les trois grandes architectures de réseaux profonds (selon Patterson & Gibson, 2017) :

- Réseaux de neurones pré-entraînés non supervisés.
- Réseaux neuronaux convolutionnels (CNN).
- Réseaux neuronaux récurrents (RNN).

Habituellement, de nombreuses catégories considèrent les réseaux de neurones supervisés comme appartenant au domaine d'apprentissage automatique plutôt que d'apprentissage en profondeur.

2.3.2. Réseaux de neurones pré-entraînés non supervisés

Il existe plusieurs modèles génératifs pré-entraînés non supervisés que l'on peut construire entrainer, tous ces modèles représentent les distributions de probabilités sur plusieurs variables d'une certaine façon,²⁴. Parmi ces modèles, nous pouvons citer :

- Auto-encodeurs.
- Machines Boltzmann restreintes.
- Deep Belief Network

²⁴ (Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press . Retrieved july 18, 2023

- Réseaux antagonistes génératifs (GANs).
- Auto-encodeurs

Les auto-encodeurs sont utilisés pour apprendre des représentations compressées d'ensembles de données. Ils exploitent généralement des données non étiquetées, c'est-à-dire en mode d'apprentissage non supervisé. Leur objectif est de réduire la dimensionnalité d'une base de données. La sortie d'un auto-encodeur est une reconstruction des données d'entrée sous une forme optimisée. Pour mettre à jour leurs poids, ils se basent sur la rétropropagation²⁵.

Il est à noter que les auto-encodeurs ressemblent beaucoup aux réseaux neuronaux multicouches, aussi appelés "MLP". Comme ces derniers, ils possèdent une couche d'entrée, plusieurs couches cachées et une couche de sortie. Cependant, la principale distinction entre un MLP et un auto-encodeur réside dans le fait que la couche de sortie de ce dernier a autant d'unités que sa couche d'entrée.

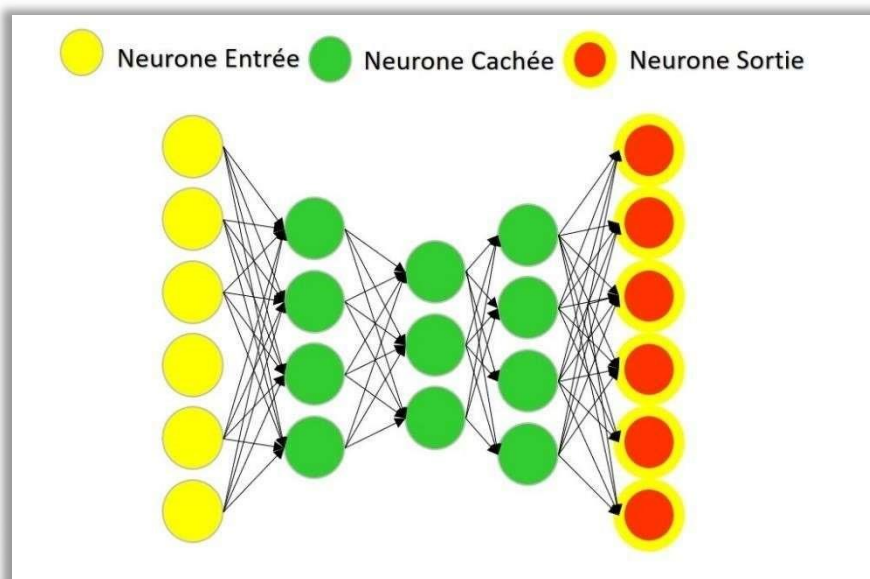


Figure 6 : Les autos-encodeuses²⁶.

Les auto-encodeurs ont de nombreuses applications dans le domaine de l'apprentissage automatique et de l'intelligence artificielle. Voici quelques-unes de leurs utilisations courantes :

²⁵ Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach.

²⁶ Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach.

Réduction de dimensionnalité : Les auto-encodeurs peuvent réduire la taille des données (comme expliqué précédemment) pour faciliter leur traitement.

Débruitage : Ils peuvent être utilisés pour nettoyer des images ou des données bruitées. En entraînant l'auto-encodeur sur des données bruitées, il apprend à reconstruire une version "propre" de ces données.

Génération d'images : Avec des variantes comme les Variational Autoencoders (VAE), ils peuvent être utilisés pour créer de nouvelles images ou contenus qui ressemblent à ceux de l'ensemble d'entraînement.

Détecteur d'anomalies : Les auto-encodeurs sont parfois utilisés pour détecter des comportements ou des données anormales. Si l'auto-encodeur est entraîné sur des données "normales", il aura du mal à reconstruire correctement une donnée "anormale", permettant ainsi sa détection.

Pré-entraînement : Ils peuvent être utilisés pour pré-entraîner les poids d'un réseau neuronal avant un apprentissage supervisé, en particulier lorsque les données étiquetées sont rares.

Représentation sémantique : Les auto-encodeurs peuvent être utilisés pour convertir des données complexes, comme des images ou du texte, en vecteurs de dimensions réduites qui capturent leur essence ou leur signification.

Amélioration de la collaboration entre les réseaux : Par exemple, dans les systèmes de recommandation, où l'auto-encodeur pourrait aider à transformer des informations d'utilisateurs en des représentations plus gérables pour des recommandations précises.

Ces applications sont juste quelques exemples de ce que les auto-encodeurs peuvent faire. Leur polyvalence et leur capacité à apprendre des représentations efficaces des données les rendent utiles dans de nombreux domaines de l'intelligence artificielle.

➤ (DBN)

Le réseau de croyances profondes (DBN - "Deep Belief Networks") se positionne comme l'une des premières architectures profondes à être efficacement entraînées sans être convolutionnelles. Cette avancée majeure a été initiée en 2006 par le chercheur canadien Geoffrey Hinton. Le DBN est structuré avec plusieurs couches de neurones cachées. Tandis que ces neurones cachés sont majoritairement binaires, ceux des couches visibles peuvent être soit binaires soit continus. Les deux couches supérieures adoptent la structure des RBM (Goodfellow, Bengio & Courville, 2016). En essence, les DBNs peuvent être vus comme

une fusion de réseaux simples et d'architectures non supervisées comme les RBM ou les auto-encodeurs. Ils utilisent une couche de RBM pour le pré-entraînement, suivie d'un réseau de neurones classique, dit "feedforward", pour l'ajustement final ou "fine-tuning"²⁷.

Bien que leur popularité ait décliné, les réseaux de croyance profonde restent un jalon important dans l'histoire de l'apprentissage profond. Leur pertinence est moins évidente aujourd'hui face à d'autres algorithmes d'apprentissage non supervisés ou génératifs, mais leur influence historique, notamment dans la modélisation d'images, demeure indéniable ²⁸

➤ Réseaux antagonistes génératifs (Gans)

Les Réseaux antagonistes génératifs GANs « Generative Adversarial Networks sont un exemple de réseau qui utilise l'apprentissage non supervisé pour entrainer deux modèles en parallèle, Les GANs sont habiles à synthétiser de nouvelles images, par l'apprentissage d'autres images. Ils peuvent être utilisés dans d'autres domaines tels que :

- Le Son .
- La Vidéo.
- La Génération des images à partir du texte descriptif.

Un aspect clé des GANs est la façon dont ils utilisent un compte de paramètres ou un ensemble de données d'entrée qui est beaucoup plus petit que la normale en ce qui concerne la quantité de données sur lesquelles nous entraînons le réseau, le réseau est obligé de représenter efficacement les données d'entraînement, ce qui le rend plus efficace pour générer des données semblables aux données d'entraînement ²⁹

2.3.3. Réseaux neuronaux convolutionnels (CNN)

Les réseaux neuronaux convolutionnels (CNNs - "Convolutional Neural Networks") sont un type spécifique de réseau neuronal conçu pour traiter des données ayant une structure

²⁷ Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach. Retrieved July 18, 2023

²⁸ Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press . Retrieved July 28, 2023

²⁹ Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach

topologique bien définie. Bien que ces réseaux puissent fonctionner en mode supervisé ou non supervisé, ils sont principalement utilisés en apprentissage supervisé ³⁰

C'est l'informaticien japonais Kunihiko Fukushima qui a posé les fondements des CNNs avec son travail sur le Neocognitron en 1980 (Vázquez, 2018). Contrairement à des réseaux de neurones entièrement connectés, les CNNs exploitent la topologie des données. Comme l'indique le terme "convolutionnel", ces réseaux utilisent une opération mathématique dénommée "convolution", représentant une forme spécifique d'opération linéaire ³¹

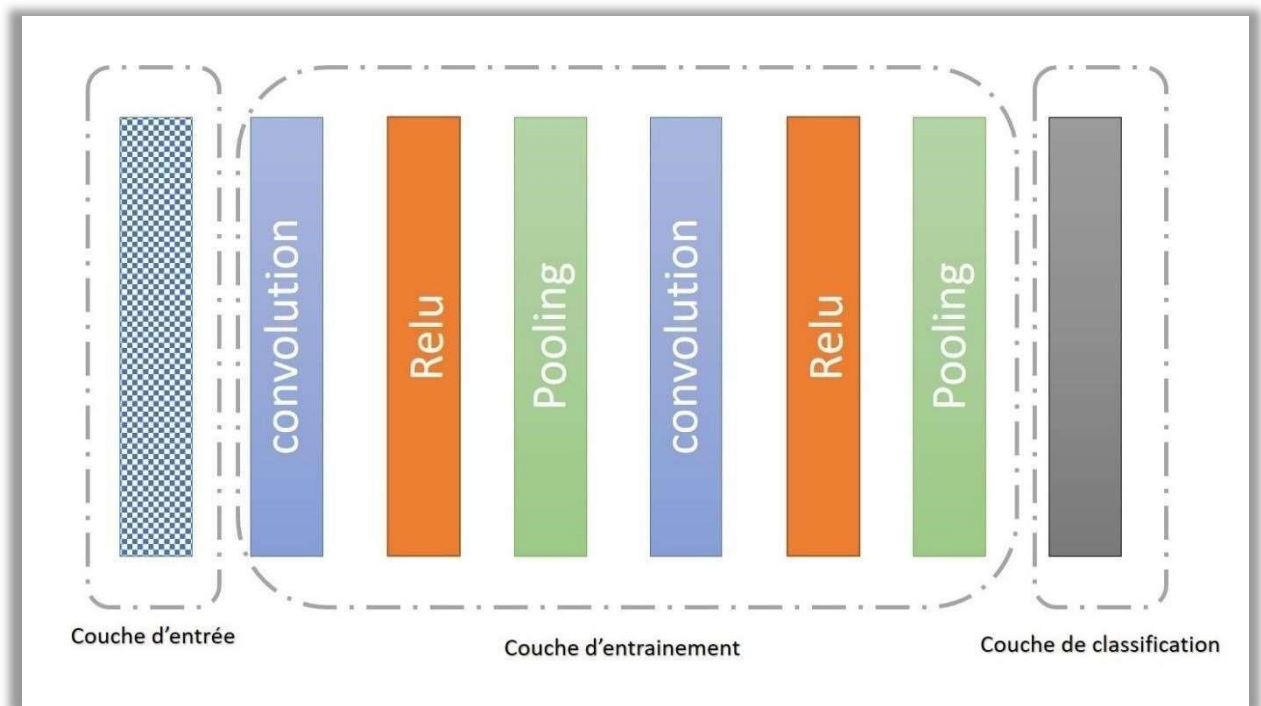
L'architecture des CNNs s'inspire de l'organisation des neurones dans le cortex visuel des animaux, où les champs récepteurs de différents neurones se chevauchent pour couvrir l'ensemble du champ visuel.

En termes de fonctionnement, les CNNs s'inspirent des processus biologiques. Ils sont constitués de multiples couches de perceptrons qui travaillent sur de petits fragments d'informations à la fois, permettant ainsi un prétraitement efficace des données ³²

³⁰ Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach Retrieved 08. 08, 2023

³¹ Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press . Retrieved 08. 08, 2023

³² Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach Retrieved 08. 08, 2023



L'architecture des réseaux de neurones convolutifs (Goodfellow, Bengio, & Courville, 2016).

➤ L'architecture des réseaux neuronaux convolutifs

L'architecture des réseaux neuronaux convolutifs (CNNs ou ConvNets) est conçue spécifiquement pour le traitement automatisé d'images, mais elle est également utilisée pour d'autres types de données séquentielles comme les signaux sonores. Voici une explication simplifiée de l'architecture typique d'un CNN :

Couche d'entrée : Reçoit l'image brute comme une matrice de pixels. Si l'image est en couleur (RGB), elle aura trois canaux correspondant à ces trois couleurs.

Couches de convolution (Conv) : Ces couches utilisent des filtres (aussi appelés noyaux) pour scanner l'image par petites zones à la fois (par exemple, 3x3 ou 5x5 pixels) et produire une carte des caractéristiques (ou feature map).

L'opération de convolution permet de détecter des caractéristiques ou des motifs locaux dans l'image, tels que des bords, des textures, ou des formes.

Couches d'activation : Après la convolution, une fonction d'activation, généralement la fonction ReLU (Rectified Linear Unit), est appliquée. Elle introduit des non-linéarités dans le modèle, permettant au réseau d'apprendre des caractéristiques complexes.

Couches de pooling (ou sous-échantillonnage) : Ces couches réduisent les dimensions spatiales (hauteur et largeur) de l'image tout en conservant ses caractéristiques importantes. Le pooling le plus courant est le "max pooling", qui ne garde que la valeur maximale d'une région de la carte des caractéristiques.

Couches entièrement connectées (FC ou Fully Connected) : Après plusieurs itérations de convolution et de pooling, l'image réduite est aplanie en un vecteur unique et traitée par une ou plusieurs couches entièrement connectées.

Ces couches permettent au réseau d'apprendre des relations non locales à partir des caractéristiques extraites précédemment.

La dernière couche entièrement connectée produit généralement la sortie du réseau, que ce soit une classification (pour les tâches de classification) ou une autre valeur (pour la régression, par exemple).

Couche de sortie : En fonction de la tâche, cette couche produira les prédictions finales du réseau. Pour une tâche de classification, elle pourrait utiliser une fonction Softmax pour produire des probabilités pour chaque classe.

Il est également courant de voir d'autres composants dans les architectures CNN modernes, comme la normalisation par lots (batch normalization), le dropout (pour éviter le surajustement) ou encore d'autres types de couches.

L'ensemble de ces éléments permet aux CNNs de reconnaître des motifs visuels avec une grande précision, du plus simple (comme les bords dans les premières couches) au plus complexe (comme les objets dans les couches plus profondes).

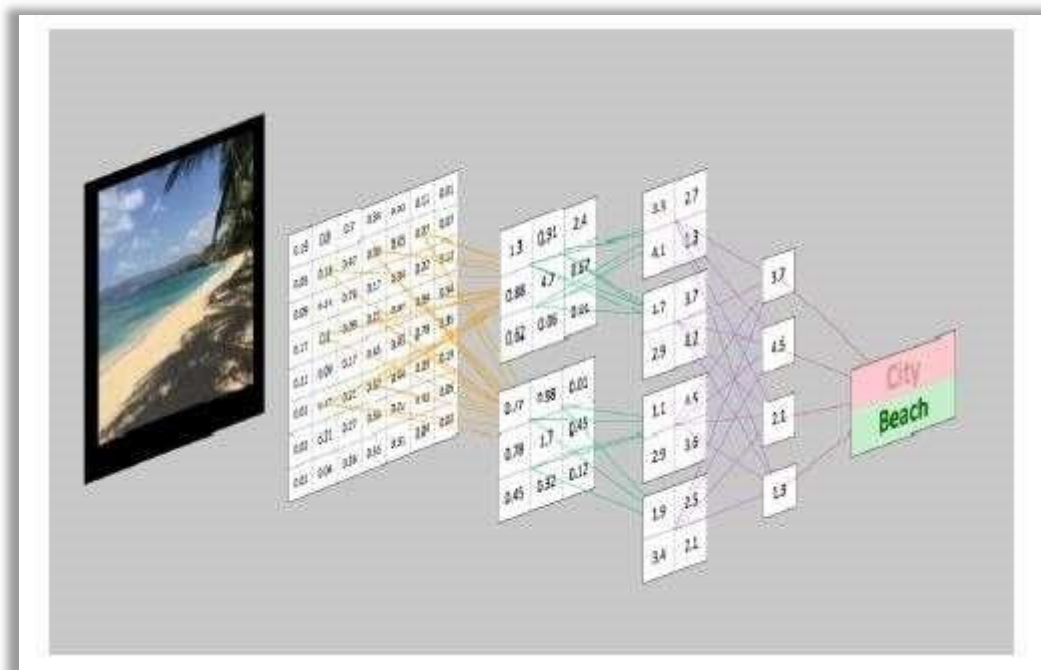
2.3.4.2. L'application des réseaux neuronaux convolutifs :

Les réseaux neuronaux convolutifs (CNNs) ont de larges applications, notamment :

La reconnaissance d'image et vidéo.

Les systèmes de recommandation.

Le traitement du langage naturel.



Reconnaissance d'une image avec CNN (Patterson & Gibson, 2017).

2.3.4. Réseaux neuronaux récurrents (RNN)

Les réseaux neuronaux récurrents (RNNs) ou "Recurrent Neural Networks" représentent une catégorie de réseaux neuronaux spécialement conçus pour traiter les données séquentielles. Ils ont été introduits par l'informaticien américain Michael Irwin Jordan en 1986³³

Le chercheur allemand Jürgen Schmidhuber compare souvent les RNNs au cerveau humain. Selon lui, le cerveau peut être vu comme un vaste réseau de neurones interconnectés en rétroaction. De manière similaire, les RNNs sont capables de traiter un flux d'entrées sensorielles et de le transformer, au fil du temps, en une séquence d'actions ou de réponses pertinentes³⁴.

³³ Vázquez, F. (2018, 04 12). A "weird" Introduction to Deep Learning. Retrieved from bbvadata&analytics: <https://www.bbvadata.com/a-weird-introduction-to-deep-learning/>

³⁴ Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach

L'une des caractéristiques distinctives des RNNs est leur capacité à modéliser la dimension temporelle. Contrairement aux réseaux neuronaux classiques, qui traitent les entrées de manière indépendante, les RNNs peuvent prendre en compte l'ordre et la séquence des informations. Cette capacité leur permet de reconnaître des motifs séquentiels et de prédire, à partir de données historiques, le scénario le plus probable qui pourrait se produire par la suite.

➤ L'architecture des réseaux neuronaux récurrents (RNN)

Les réseaux de neurones récurrents (RNNs) représentent une extension des réseaux de neurones feedforward traditionnels, avec l'ajout crucial de connexions récurrentes. Ces connexions permettent aux informations de circuler non seulement à travers les couches, comme dans un réseau feedforward, mais aussi à travers le temps, reliant les étapes temporelles adjacentes. Cette structure confère au RNN une mémoire de ses étapes précédentes, intégrant ainsi la dimension temporelle dans le modèle.

L'existence de connexions récurrentes conduit à la formation potentielle de cycles dans le réseau, ce qui signifie que l'information peut être renvoyée à des neurones antérieurs à des moments ultérieurs. Cette propriété est fondamentalement différente des architectures feedforward traditionnelles et offre une flexibilité accrue pour modéliser des dépendances séquentielles ou temporelles dans les données.

Cependant, la présence de ces connexions récurrentes et de cycles rend également l'entraînement des RNNs plus complexes, notamment à cause des problèmes comme la disparition ou l'explosion du gradient. Mais malgré ces défis, les RNNs ont montré leur efficacité dans de nombreuses applications, notamment la modélisation de séquences, la reconnaissance vocale, la génération de texte et bien d'autres domaines nécessitant une prise en compte de l'ordre temporel des données.

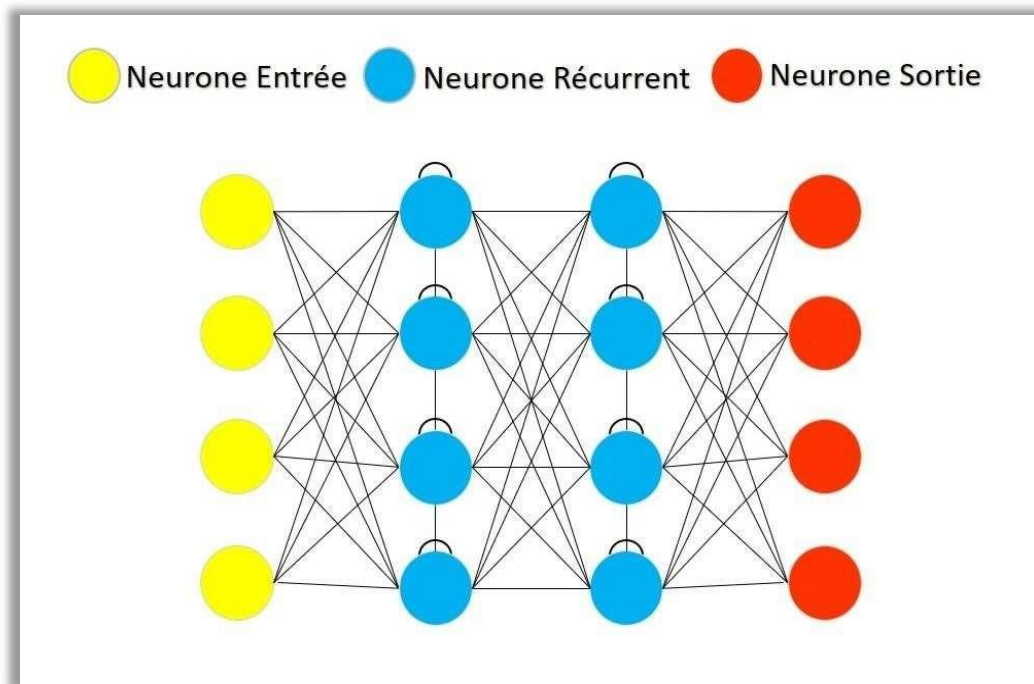


Figure 7 : Réseaux neuronaux récurrents (RNN) (Patterson & Gibson, 2017).

➤ Le réseau de neurones LSTM

L'architecture mémoire court-terme étendue LSTM « Long Short-Term Memory », est un type de RNN très utilisé en traitement du langage naturel, proposé par les informaticiens « Sepp Hochreiter » et « Jurgen Schmidhuber » en 1997 (Patterson & Gibson, 2017).

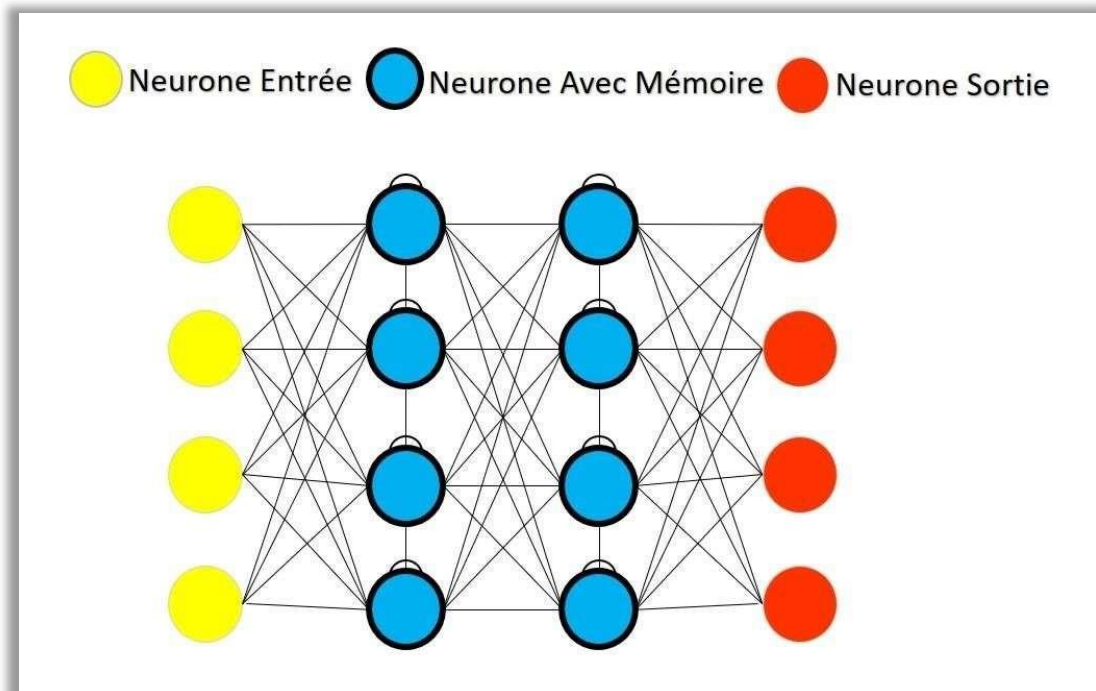


Figure 8 : Architecture LSTM (Patterson & Gibson, 2017).

L'idée principale de l'architecture LSTM est de diviser le signal entre ce qui est important à :

- Court terme à travers une couche de sortie LSTM, nommée « l'état caché ».
- Long terme, à travers un neurone qui conserve un état interne, nommé « cellule d'état », ce neurone n'est pas un neurone de sortie

Le fonctionnement global d'un LSTM peut se résumer selon les trois étapes suivantes:

Etape I	Détecter les informations pertinentes venant du passé, piochées dans une cellule d'état à travers la porte oubliée « forget gate ».
Etape II	Choisir, à partir de l'entrée courante, celles qui seront pertinentes à long terme, via l'input gate. Celles-ci seront ajoutées à la cellule d'état qui fait d'office d'une mémoire longue.
Etape III	Piocher dans la nouvelle cellule d'état les informations importantes à court terme pour générer l'état caché suivant à travers la porte de sortie « output gate ».

Figure 9 : Fonctionnement global d'un LSTM.

- L'application des réseaux neuronaux récurrents (RNN)

Les réseaux neuronaux récurrents peuvent être utilisés dans plusieurs applications ³⁵telles que :

- ✓ La transcription vocale au texte.
- ✓ Le traitement automatique du langage naturel.
- ✓ La traduction automatique.
- ✓ La génération de texte manuscrit.

2.3.5. Le problème de sur-apprentissage(*overfitting*) et sous-apprentissage(*underfitting*)

Les facteurs déterminant le rendement d'un algorithme d'apprentissage automatique ou profond sont sa capacité à :

1. Réduire l'erreur dans la phase d'apprentissage.
2. Réduire l'écart entre le résultat d'apprentissage et l'erreur de test.

³⁵ Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach Retrieved 08. 08, 2023

Ces deux facteurs correspondent aux deux enjeux centraux d'apprentissage automatique qui sont :

- Sur-apprentissage « overfitting ».
- Sous-apprentissage « underfitting »
- **Le sur-apprentissage :**

A titre exemple, une équipe de football se comporte mal durant le passage des journées de la ligue. Par conséquent, tous les joueurs de l'équipe seront jugés par l'échec ; bien qu'il y ait des joueurs compétents et même le buteur de la ligue. Cette situation s'appelle « la généralisation ». Le problème de la généralisation est très connu chez les êtres humains, et même les machines peuvent se tomber dans ce problème. Au niveau de la Machine Learning, ce problème s'appelle le sur- apprentissage « overfitting ». Cela signifie que le modèle fonctionne sur les données d'apprentissage, mais il ne peut pas se généraliser sur d'autres données³⁶).

➤ **Le sous-apprentissage :**

Si nous prenons le même exemple cité dans la section précédente. Mais cette fois on ne connaît pas les joueurs de l'équipe aussi les supporters ne connaît même pas les résultats de l'équipe dans des nombreuses journées. Par conséquent, l'équipe est jugée qu'elle est faible et même tous ses joueurs, donc nous avons une mauvaise pensée sur l'équipe malgré le manque d'informations. Cette situation s'appelle « le sous-apprentissage ; underfitting », cela signifie que le modèle développé est faible pour apprendre la structure sous-jacente des données, vu le manque d'informations³⁷.

Pour éviter ces deux problèmes, cités ci-dessus, plusieurs solutions ont été proposées dans la littérature ³⁸Concernant la résolution du problème de sur-apprentissage, nous pouvons :

- 1- Utiliser plus des données durant la phase d'apprentissage.

³⁶ Géron, A. (2017). Hands-On_Machine_Learning_with_Scikit-learn and tensorflow. USA: O'Reilly Media. Retrieved 08. 08, 2023

³⁷ Géron, A. (2017). Hands-On_Machine_Learning_with_Scikit-learn and tensorflow. USA: O'Reilly Media. Retrieved 08. 08, 2023

³⁸ Géron, A. (2017). Hands-On_Machine_Learning_with_Scikit-learn and tensorflow. USA: O'Reilly Media. Retrieved 08. 08, 2023

2- Réduire le bruit dans les données d'apprentissage (fixer les erreurs dans la base de données).

Pour résoudre le problème de sous-ajustement, nous pouvons :

Fournir une meilleure base de données à l'algorithme d'apprentissage.

La figure suivante fournit des graphiques illustrant tous les états (Voir Figure 2.19).

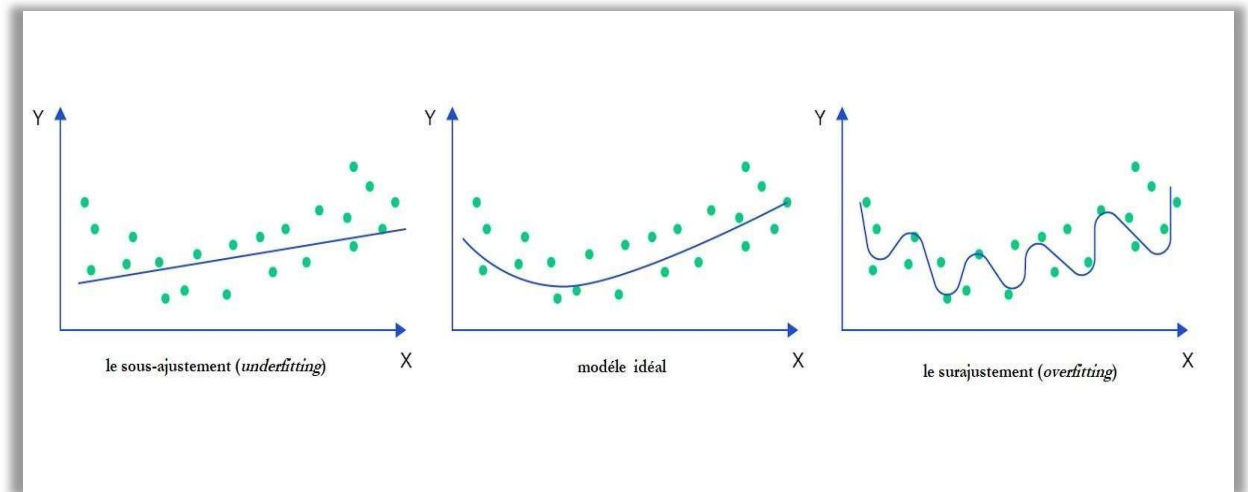


Figure 10 : Les graphes d'état des problèmes liés à l'apprentissage (Géron, 2017).

2.3.6. Comparaison entre l'apprentissage automatique et l'apprentissage en profondeur

	L'apprentissage automatique	L'apprentissage en profondeur
Dépendances des données	Une performance excellente avec des petites bases de données	Une performance excellente avec le BIG-DATA
Dépendances du matériel	Travail sur une machine faible	Besoin d'une machine forte avec a Un CPU fort
Les caractéristiques des données	Besoin de comprendre les caractéristiques des données	Ne besoin pas de comprendre les caractéristiques des données
Le temps d'exécution	Quelque minute à des heures	Des semaines

Tableau 2 : Comparaison entre l'apprentissage automatique et l'apprentissage en profondeur (Grossfeld, 2020).

3. Etat de l'art

Les sous-sections suivantes présentent quelques travaux réalisés dans le cadre de la détection du cancer du sein à partir de l'Intelligence artificielle.

3.1. Une approche d'apprentissage approfondi pour la détection de masse du cancer du sein (Fathy & Ghoneim, 2019) :

Dans ce travail, les auteurs ont utilisé respectivement l'architecture préformée « Resnet-50⁴ » et la technique « Class Activation Map » (CAM) (Voir figure 2.19), pour la classification et la localisation du cancer de sein. La technique CAM exploite la couche « Convolutional Neural Network » (CNN) avec la couche « Global Average Pooling » (GAP) pour la localisation d'objets sans aucune information supervisée sur son emplacement.

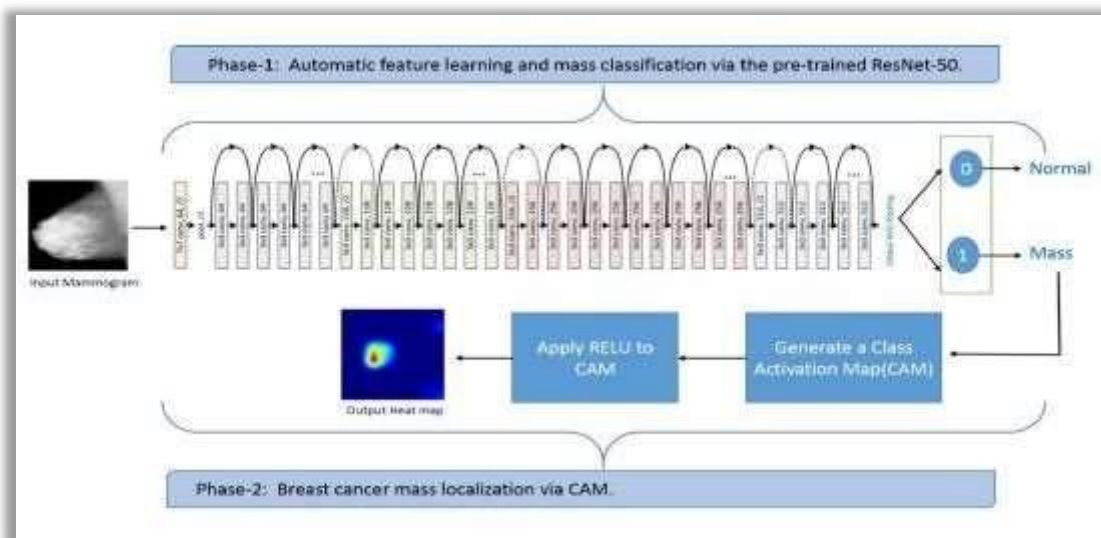


Figure 11 : Architecture de l'approche proposée par (Fathy & Ghoneim, 2019).

La base de données de la mammographie utilisée pour apprendre et évaluer le réseau de neurones proposés dans ce travail est « Digital Database for Screening Mammography » DDSM. Elle se compose de 2620 cas, catégorisés comme 695 cas normaux, 141 cas bénins sans rappel, 870 cas bénins et 914 cas malins.

La première étape de ce travail est l'étape de prétraitement, elle consiste à appliquer un ensemble de transformations pour éliminer le bruit non pertinent et les parties indésirables

en arrière-plan des mammographies. Ensuite, la couche « Convolutional Neural Network (CNN) » est appliquée pour la classification des images de la mammographie. A la fin, pour localiser la masse, les auteurs ont calculé la CAM. Pour cela, ils ont utilisé l'architecture « Resnet-50 », cette dernière utilise l'algorithme « Adam optimizer ».

Les auteurs ont utilisé 11 époques pour l'apprentissage, et les meilleurs poids sont enregistrés par les points de contrôle sur l'ensemble de validation. Ensuite, ils ont utilisé la technique (CAM) pour la classification et la reconnaissance des images par localisation de masse sans besoin de toute autre information sur son emplacement pendant l'opération d'apprentissage.

Pour éviter le problème de surajustement « overfitting » durant l'apprentissage, les auteurs ont appliqué les méthodes d'augmentation de données suivantes sur l'ensemble des données :

- Rotation aléatoire entre 0 et 180.
- Basculement horizontal.
- Décalage arbitraire de la hauteur (dans la fraction 0,1).
- Décalage arbitraire de la largeur (dans la fraction 0,1).
- Basculement vertical et zoom arbitraire.

Selon les résultats expérimentaux, l'approche proposée a atteint 96 % de la surface sous la courbe (AUC) des caractéristiques d'exploitation du récepteur (ROC).

La classification	Recognition
96 % AUC avec une sensibilité de 99,8 % et une spécificité de 82,1 %.	La sensibilité était de 93,67 % avec une moyenne de 0,122 faux positif par image.

Tableau 3 :Les résultats de l'approche proposée (Fathy & Ghoneim, 2019)

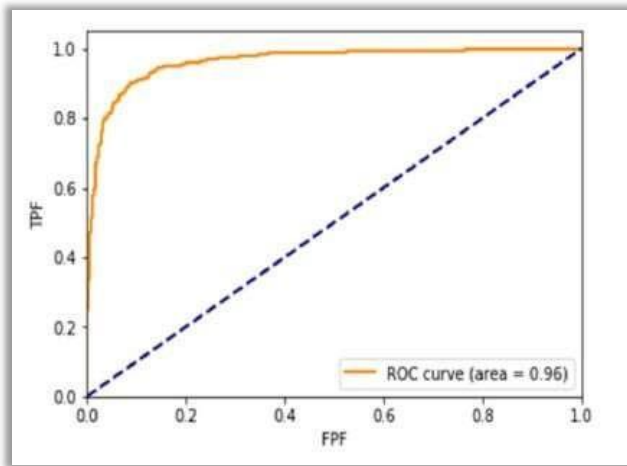


Figure 12 :La courbe des caractéristiques d'exploitation du récepteur (Fathy & Ghoneim, 2019).

3.2.Un réseau neuronal profond mul modal pour la prévision du cancer du sein humain en intégrant des données mul dimensionnelle (Sun, Wang, & Li, 2018)

Dans ce travail, les auteurs proposent une nouvelle architecture, basée sur la fusion des données multidimensionnelles pour la prédiction du cancer du sien, Cela signifie l'utilisation de différentes données pour un seul but.

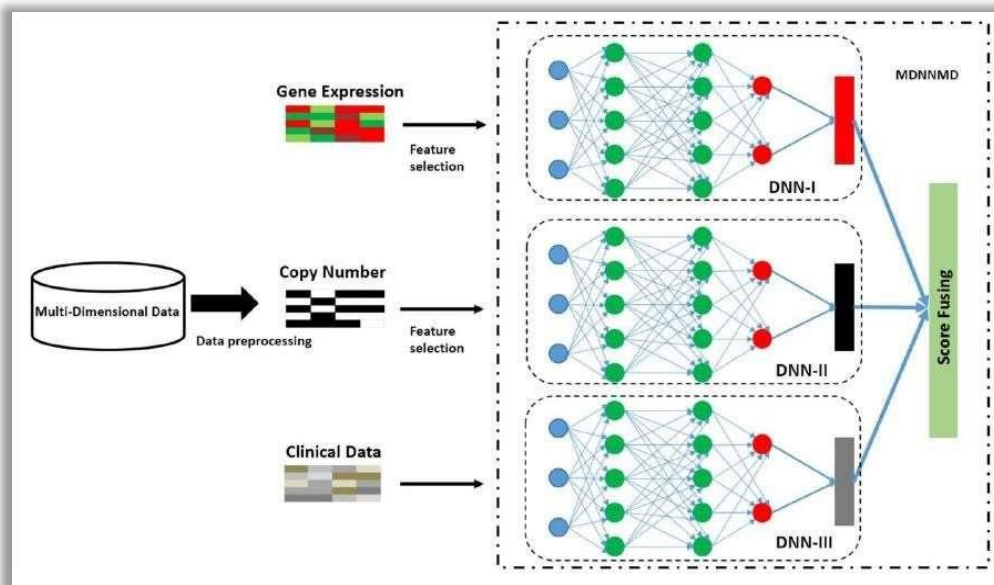


Figure 13: Architecture de l'approche proposée par (Sun, Wang, & Li, 2018)

L'approche MDNNMD « Multimodal Deep Neural Network by integrating Multi-dimensional Data » intègre les données multidimensionnelles suivantes :

- Expression génétique (processus biochimiques).
- Altérations du nombre de copies (CNA) (la structure chromosomique).

➤ Données cliniques.

Pour évaluer l'approche proposée, les auteurs ont utilisé la base de données METABRIC « Molecular Taxonomy of Breast Cancer International Consortium », le tableau suivant contient des informations sur cette base de données (Voir tableau 2.3).

Période (année)	5
Total de la populations	1980
Survivants de longue date	1489
Survivants de courte durée	491
Âge médian dans le diagnostic	61
Survie moyenne (mois)	125.1

Tableau 4 : Information sur la base de données METABRIC (Sun, Wang, & Li, 2018).

Pour la sélection des caractéristiques et pour éviter le problème de « malédiction de la dimensionnalité », ils ont utilisé l'algorithme mRMR pour réduire la dimension de l'ensemble de données sans perte d'informations, le tableau suivant montre le nombre des caractéristiques avant et après le processus de sélection (Voir tableau 5).

Catégorie	N° des caractéristiques avant	N° des caractéristiques après
Expression génétique	27	25
Altérations du nombre de copies (CNA).	24368	400
Données cliniques	26298	200

Tableau 5 : les caractéristiques avant et après le processus de sélection (Sun, Wang, & Li, 2018).

Dans la deuxième étape de ce travail, les auteurs ont utilisé un modèle de réseau neuronal profond (DNN) pour chaque type de données pour la prédiction du cancer de sein. Dans la phase finale, les résultats de prédiction obtenus à partir des trois 3 modèles sont fusionnés, pour obtenir le résultat final de l'approche MDNNMD.

Les résultats montrent que la méthode proposée atteint une meilleure performance par rapport aux méthodes de prédiction avec des données unidimensionnelles (Voir figure 2.22).

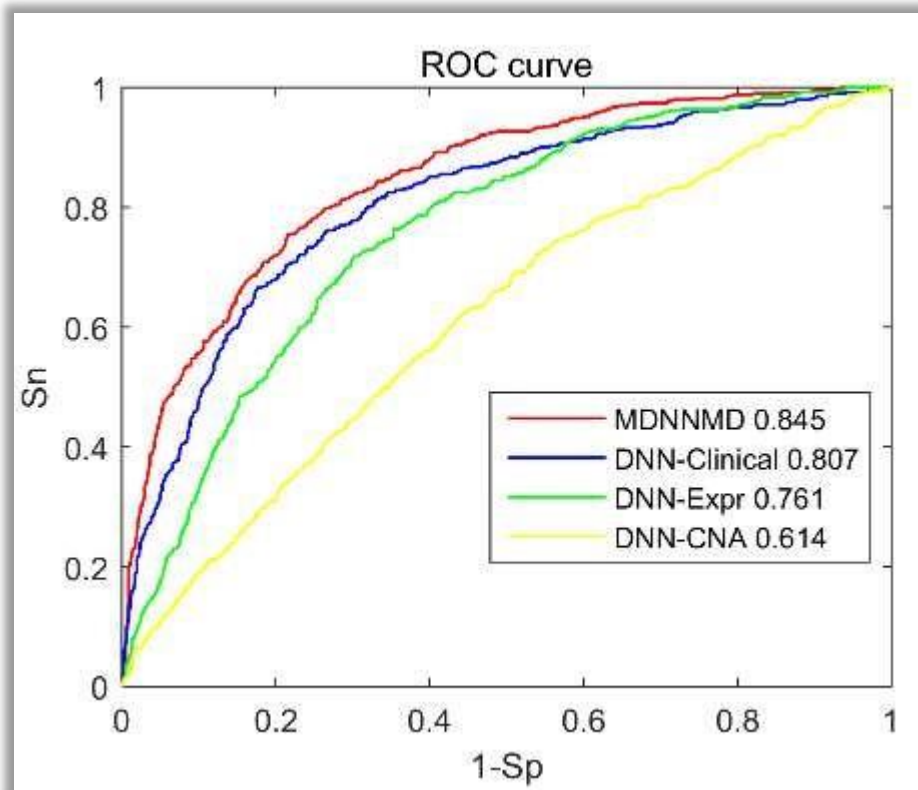


Figure 14 : La courbe des caractéristiques d'exploitation du récepteur (Sun, Wang, & Li, 2018).

3.3.Diagnostic du cancer du sein à l'aide d'un algorithme d'extraction de fonctionnalités non supervisé basé sur l'apprentissage en profondeur³⁹

Dans ce travail, les auteurs proposent une nouvelle architecture (Voir figure 2.23), qui intègre un algorithme d'extraction de fonctionnalités basé sur l'apprentissage en profondeur non supervisé. Ils ont empilé les auto-encodeurs avec un modèle de machine de support vectoriel (SAE-SVM) pour le diagnostic du cancer de sein. L'utilisation des auto-encodeurs

³⁹ Xiao, Y., Wu, J., Lin, Z., & Zhao, X. (2018). Breast Cancer Diagnosis Using an Unsupervised Feature. the 37th Chinese Control Conference, (p. 6). Wuhan, China. Retrieved 03 08, 2023

empilés (SAE) avec l'algorithme de pré-entraînement « greedy layer-Wise » qui améliore la mise à jour des poids, pour capturer les informations essentielles et extraire les caractéristiques nécessaires des données originales. Ensuite, un modèle de machine de vecteur (SVM) est utilisé pour classer les échantillons avec de nouvelles caractéristiques, tel que les tumeurs peuvent être malignes ou bénignes.

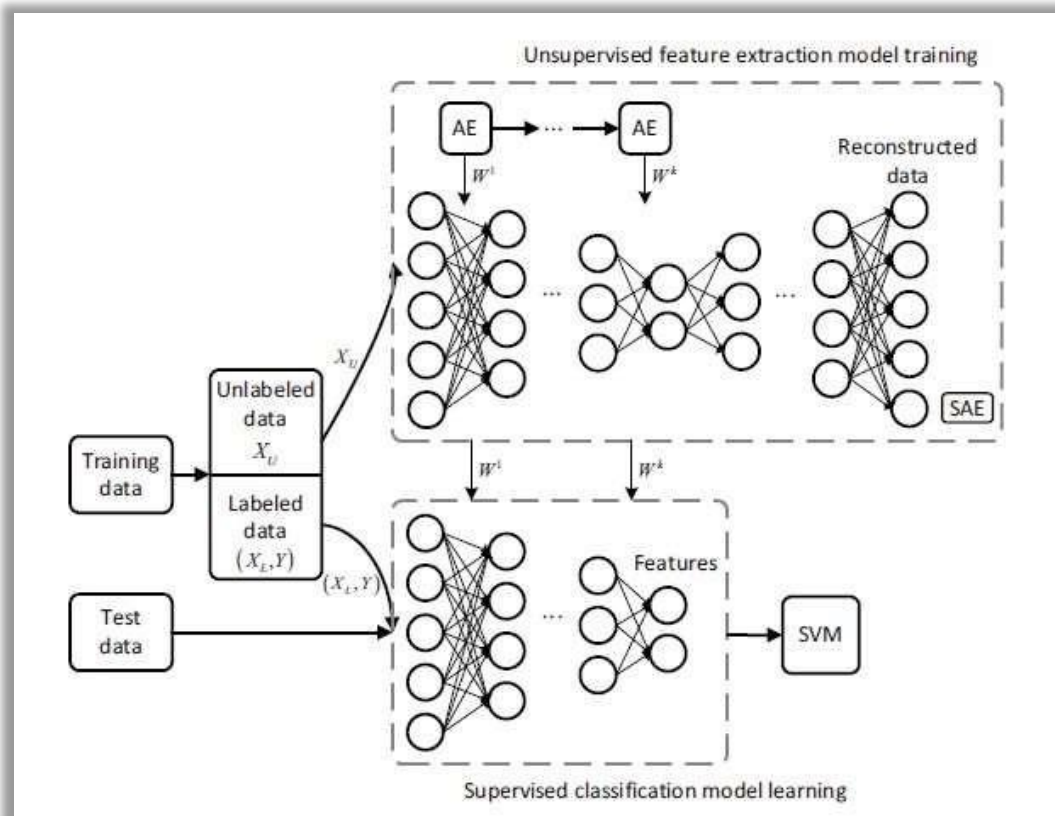


Figure 15 Architecture de l'approche proposée par (Xiao, W_u, Lin, & Zhao,

Ce travail a été validé sur la base de données WDBC « Breast Cancer Wisconsin Diagnostic », qui contient 30 caractéristiques pour le diagnostic du cancer de sein.

Cette approche se compose de deux étapes, l'étape de pré-entraînement non supervisée basée sur un auto-encodeur empilé (SAE) et l'étape d'apprentissage supervisé basé sur une machine de support vectoriel (SVM). Durant la première étape, les auteurs ont utilisé l'algorithme « greedy layer- Wise » pour l'apprentissage du modèle, cet algorithme permet l'extraction des caractéristiques pour utiliser les poids de cette phase dans deuxième étape. Une machine de support vectoriel SVM a été appliquée pour la classification.

Pour évaluer la performance de cette approche, les auteurs ont comparé les résultats obtenus avec les résultats trouvés par les travaux de la littérature. La comparaison montre que la méthode (SAE-SVM) proposée améliore la précision à 98,25 % et surpasse les méthodes décrites (Voir tableau 3.2).

Approche	Précision (%)	Dimension de la caractéristique
ACO-SVM	95.96	15
GA-SVM	97.19	18
K-SVM	97.38	6
PCA-SVM	96.84	-
PSO-SVM	97.37	17
SAE-SVM	98.25	15

Tableau 3.2 Diagnostique de précision des différentes méthodes avec la méthode SAESVM (Xiao, Wu, Lin, & Zhao, 2018).

Cette approche atteint 97.4 % de la surface sous la courbe des caractéristiques d'exploitation du récepteur (ROC), ce résultat est meilleur par rapport aux SVM et K-SVM (Voir figure 2.24).

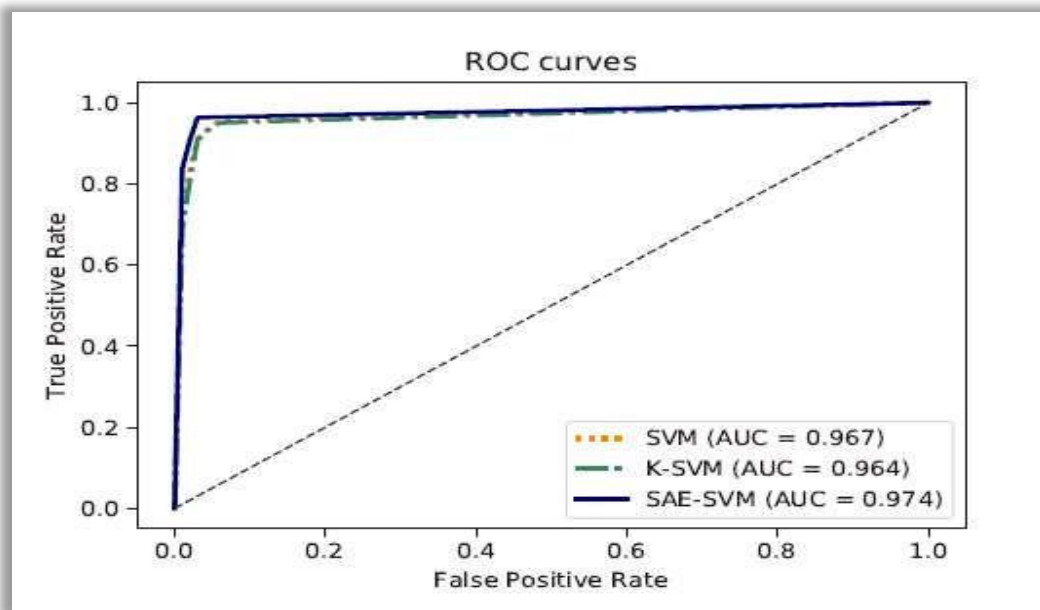


Figure : La courbe des caractéristiques d'exploitation du récepteur (Xiao, Wu, Lin, & Zhao, 2018).

Conclusion

En synthèse, le cancer du sein, malgré sa prévalence alarmante, offre une lueur d'espoir à travers l'intersection de la médecine et de l'informatique. La technologie, en particulier l'intelligence artificielle, s'est révélée être un allié précieux dans la quête de solutions plus rapides, plus précises et plus accessibles pour diagnostiquer et traiter cette maladie. Comprendre les origines, les symptômes et les facteurs de risque du cancer du sein, ainsi que les avancées récentes de l'IA dans ce domaine, est crucial pour saisir l'ampleur et l'importance de cette synergie.

La fusion entre l'informatique et la médecine, surtout dans des maladies aussi répandues et dévastatrices que le cancer du sein, présente une opportunité majeure pour améliorer la santé globale et la qualité de vie des patients. La présente étude, tout en mettant en lumière cette synergie, établit également une base solide pour l'étude suivante.

En transition vers le chapitre suivant, nous nous concentrerons sur la proposition d'une méthodologie innovante basée sur l'IA pour la détection précoce du cancer du sein chez les femmes au Cameroun. Cette démarche cherchera non seulement à tirer parti des avancées technologiques de l'IA, mais aussi à les adapter aux spécificités et aux besoins du contexte camerounais.

CHAPITRE 2 : METHODOLOGIE DE RECHERCHE

Introduction

La méthodologie de recherche permet de structurer et d'organiser la manière dont une étude est menée. Pour aborder un sujet aussi délicat et pertinent que l'utilisation de l'IA pour la détection précoce du cancer du sein chez les femmes au Cameroun, il est essentiel de définir avec précision la nature de la recherche, les variables en jeu, ainsi que les outils utilisés.

1. Nature de la recherche

1.1 Présentation des méthodes de recherche

Une méthode de recherche se définit comme étant « l'ensemble des procédures et des démarches précises adoptées pour arriver à un résultat. En science, la méthode est primordiale, et les procédés utilisés lors d'une recherche en déterminent les résultats ». Il existe plusieurs méthodes de recherche en fonction du type de résultat visé, de la nature des données manipulées et du cadre dans lequel la recherche est menée. Une recherche peut donc être :

Expérimentale : Cette méthode implique une intervention active du chercheur. Il modifie ou introduit une variable (connue sous le nom de variable indépendante) et observe l'effet sur une autre variable (variable dépendante). Par exemple, tester l'efficacité d'un nouveau médicament sur des patients.

Non expérimentale : Dans cette approche, le chercheur n'effectue aucune manipulation. Il se contente d'observer les variables telles qu'elles sont naturellement. Par exemple, étudier la relation entre le régime alimentaire et l'incidence du cancer du sein.

Quasi-expérimentale : Cette recherche est semblable à l'expérimentale, mais elle n'a pas de groupe témoin randomisé. Par exemple, comparer l'efficacité de deux méthodes d'éducation sur le dépistage du cancer du sein dans deux régions distinctes du Cameroun.

Exploratoire : Adoptée lorsqu'on a peu d'informations sur le sujet. Elle permet d'identifier les patterns, les idées ou les hypothèses plutôt que de tester une hypothèse établie. Cette méthode serait utile lorsqu'on cherche à comprendre les attitudes envers le dépistage du cancer du sein dans une région peu étudiée du Cameroun.

Descriptive : Elle décrit les caractéristiques d'un phénomène ou la relation entre deux ou plus de variables. Par exemple, décrire les taux de dépistage du cancer du sein par région au Cameroun.

Corrélationnelle : Cherche à comprendre et mesurer les relations entre deux variables sans influencer ou interférer avec elles. Par exemple, étudier la corrélation entre l'âge et l'incidence du cancer du sein.

Explicative : Va au-delà de la simple description pour expliquer les raisons ou les causes sous-jacentes. Par exemple, pourquoi certains groupes ethniques au Cameroun ont des taux plus élevés d'incidence du cancer du sein.

Qualitative : S'appuie sur des données non numériques comme les entretiens, les observations et les analyses textuelles. Elle permet une compréhension profonde des attitudes, comportements et expériences des individus.

Quantitative : Utilise des données numériques et des méthodes statistiques pour comprendre et interpréter des phénomènes. Par exemple, utiliser des statistiques pour déterminer l'efficacité d'une campagne de sensibilisation.

Mixte : Combinaison des approches qualitatives et quantitatives, offrant une perspective complète. C'est souvent recommandé lorsque l'on veut à la fois une vue d'ensemble chiffrée et des détails contextuels.

1.2 Justification de la nature de la recherche

Dans le contexte camerounais, une approche mixte serait la plus adaptée étant donné que nous souhaitons avoir une vue d'ensemble et contextuelle pour le Cameroun. Par

conséquent, cette étude se concentrera principalement sur une méthodologie qualitative et quantitative avec un support d'analyse documentaire. La raison est que la recherche qualitative permettra de plonger plus profondément dans les spécificités culturelles, sociales et économiques du Cameroun qui pourraient influencer la perception et la gestion du cancer du sein. L'analyse documentaire, d'autre part, permettra de situer cette étude par rapport aux travaux antérieurs et d'identifier les éventuelles lacunes ou domaines nécessitant une investigation supplémentaire.

2. Variables de recherche

2.1 Définition conceptuelle

Variable : Toute caractéristique, nombre ou quantité qui peut être mesurée ou comptée.

Variable indépendante : La cause présumée dans une relation de cause à effet. Elle est manipulée ou catégorisée par le chercheur pour observer son effet sur une autre variable. Pour notre étude : l'intelligence artificielle.

Variable dépendante : L'effet ou le résultat que le chercheur est intéressé à observer. Elle dépend de la variable indépendante. Pour notre étude : Le nombre de cas de cancer du sein détecter précocement grâce à l'IA.

2.2 Limites et Difficultés

- Limites :

Absence d'organe de surveillance : Il n'existe actuellement pas d'organe au Cameroun qui surveille spécifiquement l'évolution du cancer du sein, ce qui limite l'accès à des statistiques fiables.

Manque de statistiques exactes : Le manque d'une organisation dédiée signifie également l'absence de statistiques précises sur le cancer du sein, que ce soit par région ou pour le pays dans son ensemble.

Absence de base de données : Il n'y a pas de base de données nationale contenant des clichés numériques de mammographies et d'échographies mammaires. Cette absence limite la capacité à réaliser diverses études dans le domaine.

Difficultés :

Manque d'information : Il y a un déficit notable d'information parmi la population concernant la santé mammaire.

Perception culturelle : Le cancer du sein est souvent perçu comme un sujet tabou au sein de la communauté camerounaise, rendant la sensibilisation et la communication sur le sujet particulièrement délicates.

Intégration de l'IA : La mise en œuvre de l'intelligence artificielle pour la détection précoce pourrait rencontrer des résistances dues à un manque de compréhension ou de confiance envers cette technologie.

2.3 Utilisation des variables

Variable indépendante - L'Intelligence Artificielle (IA) :

L'utilisation de l'IA dans le domaine médical, et spécifiquement pour la détection précoce du cancer du sein, est au cœur de notre étude. Cette technologie représente un potentiel significatif pour améliorer la précision et l'efficacité de la détection, notamment en analysant rapidement et avec précision de grandes quantités d'images médicales.

L'IA sera examinée à travers différents algorithmes et méthodes utilisés pour la détection précoce, et nous explorerons comment ces techniques peuvent être adaptées et intégrées au contexte camerounais.

Variable dépendante - Nombre de cas de cancer du sein détectés précocement grâce à l'IA:

Le nombre de cas détectés précocement est un indicateur direct de l'efficacité de l'IA dans le processus de détection. Une détection précoce peut augmenter considérablement les chances de survie et diminuer le coût et la complexité des traitements.

En utilisant cette variable, nous cherchons à déterminer si l'introduction de l'IA pourrait avoir un impact significatif sur la détection précoce du cancer du sein au Cameroun, en tenant compte des ressources disponibles et des spécificités culturelles et socio-économiques du pays.

3. Outils de recherche

3.1. Présentation d'outils de recherche

Questionnaire : C'est un instrument de recherche composé d'une série de questions conçues pour recueillir des informations auprès des répondants. Les questionnaires peuvent être administrés en personne, par téléphone, par courrier ou en ligne.

Exemple : Un questionnaire Google-Forms peut être utilisé pour recueillir des données sur la perception des professionnels de santé camerounais concernant l'utilisation de l'IA pour la détection précoce du cancer du sein. Les résultats peuvent ensuite être facilement analysés et interprétés à l'aide de statistiques descriptives.

Entretiens : Ce sont des conversations structurées entre le chercheur et le participant, avec des questions prédéfinies, visant à recueillir des informations détaillées sur un sujet spécifique.

Exemple : Des entretiens individuels peuvent être menés avec des oncologues pour discuter de la faisabilité et des défis de l'introduction de l'IA dans le processus de détection du cancer du sein au Cameroun.

Analyse documentaire : C'est une méthode de recherche qui implique l'examen et l'analyse de documents écrits, que ce soit sous forme imprimée ou électronique. Les documents peuvent inclure des articles, des rapports, des archives, etc.

Exemple : Analyse des études précédentes réalisées sur le cancer du sein pour obtenir un aperçu des tendances, des défis et des opportunités dans le domaine.

Observations : C'est une méthode où le chercheur observe directement les comportements, événements ou conditions et leurs contextes. Les observations peuvent être structurées (avec des critères prédéfinis) ou non structurées.

Exemple : Observation dans les hôpitaux pour voir comment les mammographies sont actuellement réalisées et comment les patients réagissent aux diagnostics.

Expérimentation : C'est une méthode de recherche dans laquelle une ou plusieurs variables sont manipulées pour observer l'effet sur une autre variable, dans des conditions contrôlées.

Exemple : Un essai pilote pourrait être mené pour tester l'efficacité d'un nouvel algorithme IA pour la détection du cancer du sein, comparativement aux méthodes traditionnelles.

3.2 Justification du choix des outils

Pour cette recherche spécifique sur l'IA et la détection précoce du cancer du sein au Cameroun, le questionnaire (via Google-Forms) est choisi pour sa capacité à atteindre un grand nombre de personnes rapidement et à moindre coût. L'analyse documentaire est cruciale pour comprendre le contexte actuel et les études antérieures menées dans ce domaine, garantissant que notre recherche ajoute une valeur nouvelle et unique.

Conclusion

La méthodologie proposée fournit une structure robuste pour explorer et décrire l'utilisation et l'impact de l'IA sur la détection précoce du cancer du sein au Cameroun. Avec une combinaison d'approches qualitatives et quantitatives, cette recherche aspire à fournir des insights profonds et des recommandations pratiques pour améliorer la santé des femmes camerounaises

CHAPITRE 3 : PRESENTATION DE LA SITUATION (PRESENTATION DES DONNEES COLLECTEES / DES RESULTATS)

Introduction

La prévention et le dépistage du cancer du sein demeurent l'un des défis majeurs du système de santé au Cameroun. Les avancées technologiques, notamment dans le domaine de l'intelligence artificielle (IA), ouvrent de nouvelles perspectives pour améliorer les méthodes de dépistage. Cependant, il est essentiel de comprendre le contexte local, les connaissances, les perceptions et les attitudes des individus pour s'assurer que ces innovations répondent aux besoins réels de la population. Dans cette optique, notre étude a cherché à combiner une recherche empirique au niveau du terrain et une analyse documentaire approfondie, en se focalisant sur le Ministère de la Santé comme lieu de recherche initiale. Cet effort vise à fournir un tableau détaillé et nuancé des défis, des opportunités et du niveau de sensibilisation au cancer du sein au Cameroun. La présentation qui suit détaille le site de notre recherche, les données collectées et l'analyse de ces informations pour éclairer les parties prenantes et orienter les futures interventions.

1. Présentation du site de recherche

1.1. Lieu de Recherche Initiale : Le Ministère de la Santé

Le Ministère de la Santé Publique est responsable de l'élaboration et de la mise en œuvre de la politique du Gouvernement en matière de santé publique.

A ce titre, il est chargé :

- ✓ d'assurer l'organisation, la gestion et le développement des formations sanitaires publiques;
- ✓ d'assurer le contrôle technique des formations sanitaires privées ;
- ✓ de veiller à l'extension de la couverture sanitaire du Territoire ;
- ✓ de veiller au développement des actions de prévention et de lutte contre les épidémies et des pandémies ;
- ✓ de la médecine préventive ;
- ✓ de veiller à la qualité des soins et à l'amélioration du plateau technique des formations sanitaires publiques et privées ;

- ✓ d'assurer la promotion des infrastructures sanitaires en liaison avec les Administrations concernées ;

Il assure la liaison entre le Gouvernement et l'Organisation Mondiale de la Santé (OMS) ainsi qu'avec les organismes internationaux relevant de son domaine de compétence en relation avec le Ministère des Relations Extérieures.

Lors de la conception initiale de notre recherche, notre choix s'est porté sur le Ministère de la Santé, compte tenu de sa pertinence et de son rôle central dans les politiques et initiatives liées à la santé au Cameroun. Nous avons eu l'honneur d'être acceptés comme stagiaires, ce qui nous a permis d'approfondir notre compréhension des défis et des opportunités associés au dépistage précoce du cancer du sein dans le pays.

1.2. Défi Rencontré : Absence de Plateforme de Recherche Adéquate

Malgré l'opportunité offerte par le Ministère de la Santé, nous avons été confrontés à un obstacle majeur : l'absence d'une plateforme ou d'une base de données centralisée permettant de garantir l'effectivité et l'exhaustivité de notre étude. Cette situation nous a conduit à repenser notre approche de collecte de données et de notre recherche de manière générale.

1.3. Adaptation : Échantillonnage Aléatoire

Face à cette contrainte, nous avons opté pour une méthodologie d'échantillonnage aléatoire. Cette approche nous a permis de sélectionner un échantillon représentatif de la population, garantissant ainsi que les données collectées reflètent une variété d'expériences, de croyances et de connaissances liées au cancer du sein et à son dépistage. Les personnes interrogées proviennent de divers horizons socio-économiques, garantissant ainsi la diversité et la représentativité de notre échantillon.

Bien que le parcours initial ait subi des modifications en raison des défis rencontrés, cette adaptation a renforcé la rigueur et la pertinence de notre étude. Le recours à un échantillonnage aléatoire nous a non seulement permis de surmonter les obstacles initiaux, mais a également enrichi la portée et la profondeur de notre analyse. De plus, nous avons

complété notre approche empirique par une recherche documentaire approfondie, ce qui nous a permis de contextualiser nos résultats dans le cadre des connaissances actuelles sur le sujet. Cette combinaison d'approches méthodologiques a assuré une compréhension holistique du dépistage précoce du cancer du sein au Cameroun, tout en apportant une contribution significative à la littérature existante.

2. Données

2.1. Données collectées via le questionnaire

Objectif général : Analyser les connaissances, les perceptions et les attitudes des individus au Cameroun concernant le dépistage précoce du cancer du sein à l'aide de l'IA.

2.1.1 *Démographie :*

a. Sexe :

Objectif : Identifier la répartition de l'échantillon selon le sexe.

Donnée espérée : Connaître le pourcentage de femmes par rapport aux hommes, ce qui est crucial car le cancer du sein affecte principalement les femmes.

b. Tranche d'âge :

Objectif : Comprendre la répartition des répondants par âge.

Donnée espérée : Identifier l'âge des individus les plus conscients et concernés par le cancer du sein.

c. Statut :

Objectif : Évaluer la répartition matrimoniale des répondants.

Donnée espérée : Analyser si le statut matrimonial influence la conscience ou la prise en charge du cancer du sein.

d. Zone de résidence :

Objectif : Comprendre la répartition géographique des répondants.

Donnée espérée : Évaluer si les ressources et la sensibilisation varient entre les zones urbaines et rurales.

2.1.2 *Connaissances sur le cancer du sein:*

Objectif : Évaluer le niveau de connaissance sur le cancer du sein et ses symptômes parmi les répondants.

Donnée espérée : Identifier les lacunes dans la connaissance et les domaines nécessitant une sensibilisation accrue.

2.1.3 Connaissance et accessibilité du dépistage du cancer du sein :

Objectif : Évaluer la sensibilisation, l'accessibilité et l'abordabilité du dépistage du cancer du sein.

Donnée espérée : Découvrir les obstacles potentiels à un dépistage précoce et la volonté des individus à investir dans le dépistage.

2.1.4 Communication sur le cancer du sein :

Objectif : Comprendre comment le sujet est discuté au niveau familial et communautaire.

Donnée espérée : Identifier les facteurs culturels ou sociaux qui pourraient entraver la sensibilisation et le dépistage.

2.1.5 Connaissances sur l'IA :

Objectif : Évaluer la familiarité avec l'IA et sa perception comme outil potentiel dans le dépistage du cancer du sein.

Donnée espérée : Mesurer la volonté d'adopter de nouvelles technologies pour la santé et identifier les obstacles potentiels à leur acceptation.

2.2. L'Analyse Documentaire :

2.2.1. Cartographie de l'État de l'Art :

Objectif : Identifier et analyser les recherches, études et interventions antérieures, tant au niveau national qu'international, concernant l'utilisation de l'IA pour le dépistage et la prédiction du cancer du sein.

2.2.2. Identification des Gaps Technologiques :

Objectif : Analyser les technologies d'IA existantes utilisées dans la détection précoce du cancer du sein et identifier les éventuelles lacunes technologiques, en particulier dans le contexte camerounais.

2.2.3. Évaluation de la Pertinence Socio-culturelle :

Objectif : Explorer les obstacles socio-culturels potentiels, les croyances et les attitudes qui pourraient influencer l'adoption de l'IA pour le dépistage du cancer du sein au Cameroun.

2.2.4. Compréhension de l'Infrastructure Technologique :

Objectif : Évaluer l'état actuel des infrastructures technologiques du Cameroun, en particulier dans le secteur médical, pour déterminer leur capacité à soutenir et intégrer des

solutions basées sur l'IA.

2.2.5. Mise en Lumière des Opportunités et Défis :

Objectif : Identifier les opportunités potentielles que l'IA pourrait offrir pour améliorer le dépistage du cancer du sein au Cameroun, tout en reconnaissant et en analysant les défis potentiels à son adoption.

2.2.6. Contribution au Discours Scientifique :

Objectif : Assurer que la recherche actuelle apporte une contribution significative à la littérature existante, évitant ainsi une redondance tout en identifiant de nouvelles zones d'interrogation ou des angles d'approche innovants.

2.2.7. Fondation pour la Recherche Future :

Objectif : Établir une base solide à partir de laquelle des recherches futures peuvent être lancées, en identifiant des domaines nécessitant des études plus approfondies ou des zones où l'IA pourrait être appliquée de manière innovante pour le dépistage précoce du cancer du sein.

3. Présentation des résultats.

La structure des résultats de notre étude a été organisée de manière systématique pour garantir une compréhension claire et une interprétation pertinente des données collectées.

3.1. Résultats de la Collecte des Données

Les données recueillies auprès des répondants sont représentées sous la forme d'un tableau à deux dimensions. Les colonnes du tableau reprennent les différentes questions posées aux participants, tandis que les lignes reflètent les réponses individuelles de chaque soumission de formulaire. Cet agencement offre une vue panoramique des réponses et permet d'identifier rapidement des tendances, des similitudes ou des divergences dans les réponses des participants. Pour faciliter la consultation, un extrait de ce tableau est fourni à l'Annexe 2.3.

3.2. Résultats de l'Analyse Documentaire

Notre analyse documentaire a abouti à une collection de notes et de condensés, rassemblant les informations clés et les découvertes pertinentes issues de diverses sources documentaires. Ces résultats ont été essentiels pour établir des connexions entre les tendances actuelles, les défis, les opportunités et les données collectées sur le terrain. Une partie significative de ces résultats a été intégrée dans la section "État de l'Art", qui sert de fondement théorique à notre étude, mettant en lumière les avancées, les lacunes et les implications des recherches antérieures sur le dépistage précoce du cancer du sein, en particulier dans le contexte camerounais.

La combinaison des résultats de la collecte de données et de l'analyse documentaire offre une perspective équilibrée et multidimensionnelle sur le sujet. Tandis que les données recueillies auprès des participants fournissent un aperçu direct et actuel des connaissances et des perceptions, l'analyse documentaire permet de positionner ces informations dans un cadre plus large, renforçant ainsi la robustesse et la pertinence de notre étude.

Conclusion

En évoluant au sein du Ministère de la Santé et en naviguant à travers les défis inhérents à la recherche, nous avons été amenés à reconnaître la complexité et la diversité des expériences, des croyances et des connaissances liées au cancer du sein et à son dépistage au Cameroun. Grâce à une méthodologie adaptée et à une analyse documentaire approfondie, nous avons pu obtenir une vue d'ensemble enrichie de la situation. La convergence des données empiriques et des recherches antérieures nous a permis de jeter un éclairage précieux sur le paysage actuel du dépistage du cancer du sein au Cameroun, en soulignant à la fois les obstacles et les opportunités. Alors que nous nous tournons vers l'avenir, ces informations serviront de fondation à la mise en œuvre de solutions technologiques innovantes, en veillant à ce qu'elles soient pertinentes, acceptées et efficaces pour la population camerounaise.

CHAPITRE 4 : ANALYSE - DIAGNOSTIC DE LA SITUATION ET PROPOSITION D'INTERVENTION

Introduction

La santé, bien qu'elle soit un droit fondamental de l'homme, présente de nombreux défis en termes d'accès, de compréhension et d'intervention. Parmi les nombreuses maladies qui préoccupent nos sociétés actuelles, le cancer du sein se démarque comme une pathologie nécessitant une attention particulière, surtout dans les pays en voie de développement tels que le Cameroun. L'importance du dépistage précoce, associée à la nécessité d'une sensibilisation approfondie et à l'adoption d'innovations technologiques, forme le cœur de cette analyse. A travers le prisme d'une enquête minutieuse et d'une étude documentaire, ce document vise à cerner la situation actuelle, tout en proposant des solutions adaptées pour améliorer le dépistage et la sensibilisation au cancer du sein au Cameroun.

1. Présentation et analyse de la situation

1.1 Analyse des résultats du questionnaire

1.1.1 Analyse

Notre étude a été menée sur un échantillon de 286 volontaires, choisis aléatoirement. Les graphiques ci-après dépeignent les proportions des réponses des participants aux questions que nous avons estimées pertinentes au regard de notre étude, accompagnées d'un commentaire orienté selon les objectifs de notre recherche. Cependant, l'analyse graphique complète du questionnaire est à retrouver en annexe.

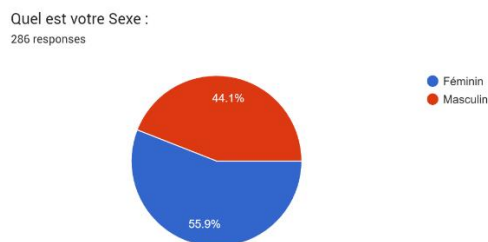


Figure 16 Repartition du genre de la population questionnée.

La majorité des répondants à ce questionnaire étaient des femmes, ce qui est pertinent étant donné que le cancer du sein concerne principalement cette population. Cependant, avoir près de la moitié des répondants masculins montre également l'intérêt des hommes pour ce sujet, soit pour eux-mêmes (les hommes peuvent également être touchés par le cancer du sein, bien que cela soit moins courant), soit par préoccupation pour les femmes de leur entourage.

Bien vouloir choisir votre tranche d'âge:
286 responses

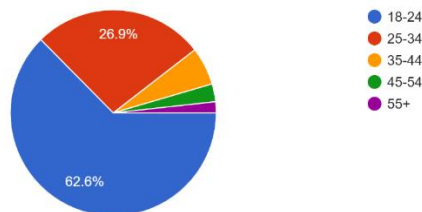


Figure 17 tranches d'âges de la population étudiée

Une large majorité des répondants appartient à la tranche d'âge des 18-24 ans, suivie des 25-34 ans. Cette prédominance de jeunes répondants peut indiquer que ces groupes d'âge sont plus engagés ou accessibles pour ce type d'enquête, ou qu'ils sont plus conscients des enjeux liés à la technologie et à la santé. Il est cependant important de noter que le groupe le plus à risque de développer un cancer du sein (45 ans et plus) est sous-représenté dans l'échantillon. Cette disparité pourrait avoir un impact sur la manière dont les résultats du questionnaire sont interprétés, en particulier lors de l'évaluation des connaissances et des attitudes envers le dépistage du cancer du sein et l'utilisation de l'IA.

Savez-vous ce qu'est le cancer du sein?
286 responses

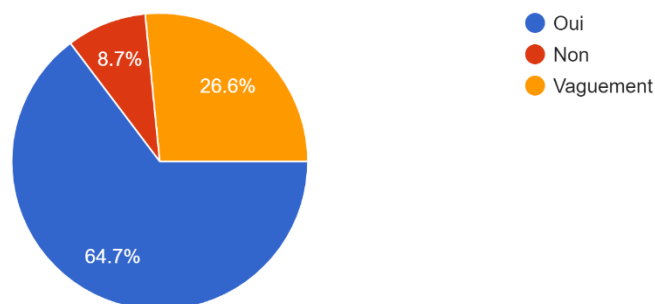


Figure 18 : statistiques sur la connaissance de ce qu'est le cancer du sein 1.

Plus de la moitié de l'échantillon est bien informée sur le cancer du sein, ce qui montre un certain niveau d'éducation et d'information sur le sujet. Cela pourrait faciliter l'acceptation de nouvelles méthodes de dépistage, comme l'IA.

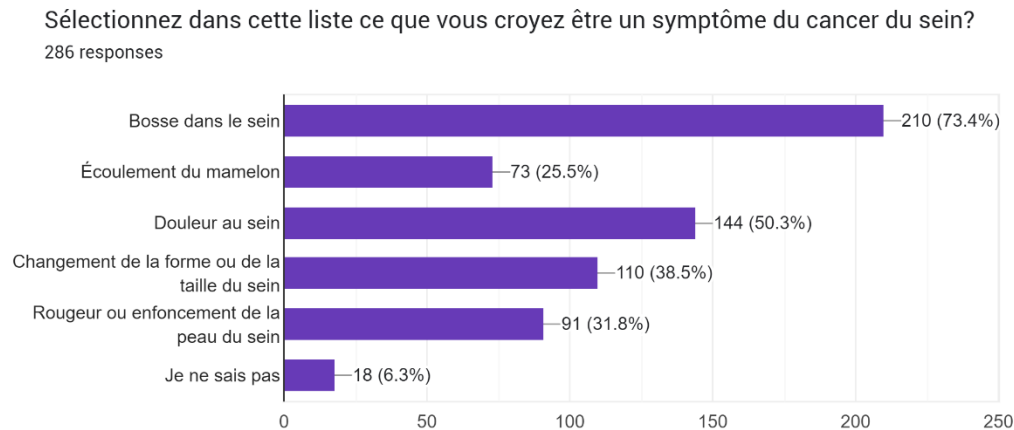


Figure 19 : statistiques sur la connaissance de ce qu'est le cancer du sein 2.

Le symptôme le plus couramment identifié comme étant associé au cancer du sein est la présence d'une "Bosse dans le sein", avec 73.4% des répondants ayant sélectionné cette option. Cela indique une assez bonne connaissance de l'un des signes les plus courants du cancer du sein.

Il est également encourageant de constater que seulement 6.3% des répondants ont déclaré ne pas connaître les symptômes du cancer du sein, indiquant une sensibilisation relativement élevée parmi les participants.

Cependant, moins de la moitié des participants ont identifié des symptômes tels que le "Changement de la forme ou de la taille du sein" et la "Rougeur ou enfoncement de la peau du sein", ce qui suggère que, bien que la sensibilisation soit présente, il y a encore des lacunes dans la connaissance complète des différents symptômes.

La compréhension de la perception des symptômes par le public est essentielle pour évaluer l'efficacité potentielle des interventions basées sur l'IA. Si la majorité des personnes ne sont pas conscientes de l'ensemble des symptômes, l'efficacité de l'IA pourrait être limitée à la détection des symptômes les plus couramment reconnus, à moins qu'une campagne de sensibilisation ne soit mise en place en parallèle.

Selon vous Laquelle/Lesquelles de ces méthodes est un examen de dépistage classique du cancer du sein?

286 réponses

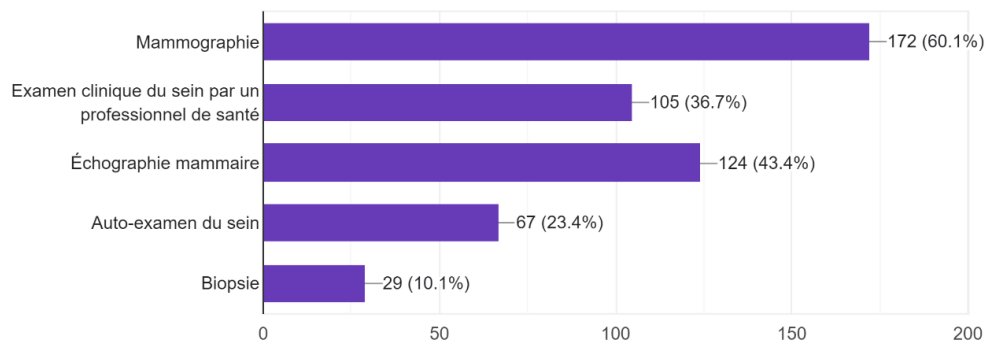


Figure 20 : statistiques sur la connaissance de ce qu'est le cancer du sein 3.

La majorité des répondants (60.1%) ont identifié la mammographie comme un examen de dépistage classique, ce qui est positif car c'est l'une des méthodes les plus couramment utilisées pour le dépistage du cancer du sein. Moins de la moitié des répondants ont identifié l'échographie mammaire et l'examen clinique du sein comme des méthodes de dépistage, ce qui pourrait indiquer un manque de sensibilisation à ces méthodes particulières au Cameroun. Seulement 23.4% ont reconnu l'auto-examen du sein comme une méthode de dépistage, malgré le fait qu'il soit largement promu comme une méthode de dépistage préliminaire pour les femmes, notamment dans les régions où les ressources médicales peuvent être limitées. La biopsie est correctement reconnue par seulement 10.1% des répondants comme une méthode de dépistage. Cependant, il est important de noter qu'une biopsie est généralement réalisée après qu'une anomalie ait été détectée par d'autres méthodes de dépistage pour confirmer un diagnostic de cancer.

Selon vous, combien coûte en moyenne un examen du sein au Cameroun ?

286 réponses

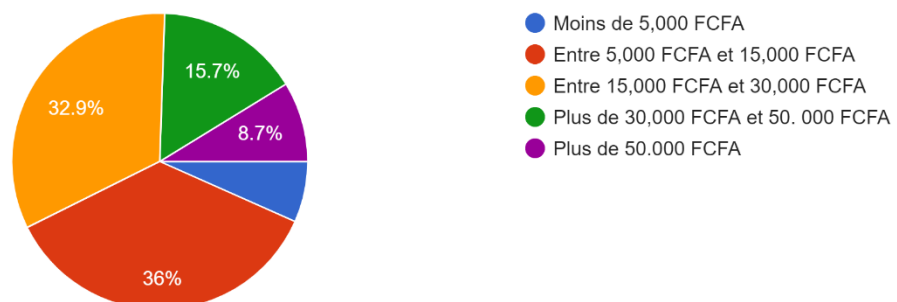


Figure 21 : statistiques sur la connaissance de ce qu'est le cancer du sein 4.

Une grande partie des répondants (68.9% combinés) estiment que le coût moyen d'un examen du sein au Cameroun se situe entre 5,000 FCFA et 30,000 FCFA. Seuls 6.6% pensent que le coût est inférieur à 5,000 FCFA, ce qui suggère que la perception est que les examens du sein ne sont pas nécessairement bon marché. Environ 24.4% des répondants estiment que le coût pourrait dépasser 30,000 FCFA, indiquant une variabilité dans la perception du coût de ces examens.

Combien seriez-vous prêt à dépenser pour un examen du sein même en l'absence de symptôme ?

286 réponses

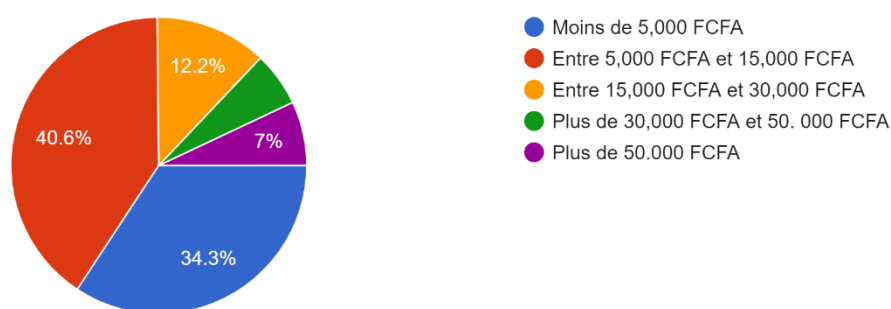


Figure 22 : statistiques sur la connaissance au sujet du cancer du sein 5 .

Les résultats montrent que la majorité des répondants (75%) seraient prêts à dépenser jusqu'à 15,000 FCFA pour un examen du sein, même s'ils ne présentent pas de symptômes. Cependant, la volonté de dépenser diminue considérablement pour les tranches de prix plus élevées. Cela suggère que le coût est un facteur déterminant dans la décision de subir un examen, surtout lorsque aucun symptôme n'est perçu.

Comment qualifieriez-vous la communication sur le cancer du sein et des anomalies mammaires dans votre famille ?

286 réponses

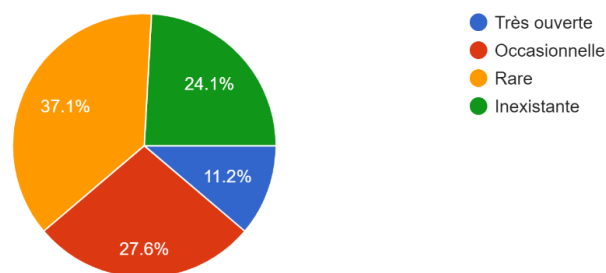


Figure 23 : statistiques sur la connaissance au sujet du cancer du sein 6 .

La majorité des répondants (37.1%) estime que la communication sur le cancer du sein dans leur famille est rare. Cette donnée est suivie de près par les répondants qui considèrent que la communication est occasionnelle (27.6%). Seuls 11.2% des personnes interrogées ont indiqué que la discussion sur le cancer du sein est très ouverte au sein de leur famille, ce qui souligne un manque flagrant de dialogue ouvert sur ce sujet. D'autre part, un pourcentage significatif (24.1%) considère que la communication à ce sujet est totalement inexistante dans leur famille.

Dans votre communauté (entourage), comment percevez-vous la discussion sur le cancer du sein ?

286 responses

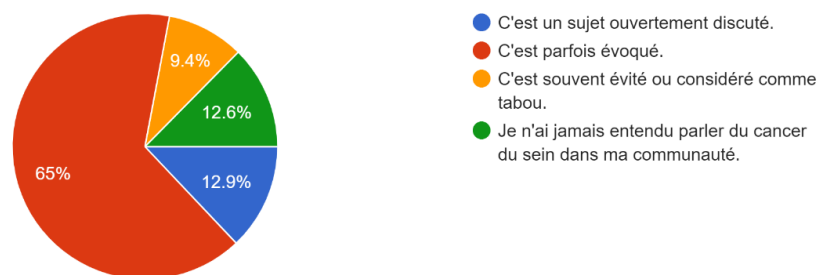


Figure 24 : statistiques sur la connaissance au sujet du cancer du sein 7 .

La majorité des répondants (65%) considèrent que le cancer du sein est un sujet occasionnellement évoqué dans leur communauté. Toutefois, il est préoccupant de noter qu'un nombre non négligeable de personnes (9.4%) considèrent que ce sujet est souvent évité ou même considéré comme tabou, et 12.6% disent qu'ils n'ont jamais entendu parler du cancer du sein dans leur communauté. Cela suggère que, bien que le cancer du sein soit un sujet de discussion pour certains, il reste un sujet sensible ou même ignoré pour d'autres. L'absence de discussion ouverte sur le cancer du sein pourrait entraver la sensibilisation et la détection précoce, rendant les femmes plus vulnérables aux stades avancés de la maladie.

Selon vous, quels sont les principaux facteurs de risque du cancer du sein? (Plusieurs réponses possibles)

286 réponses

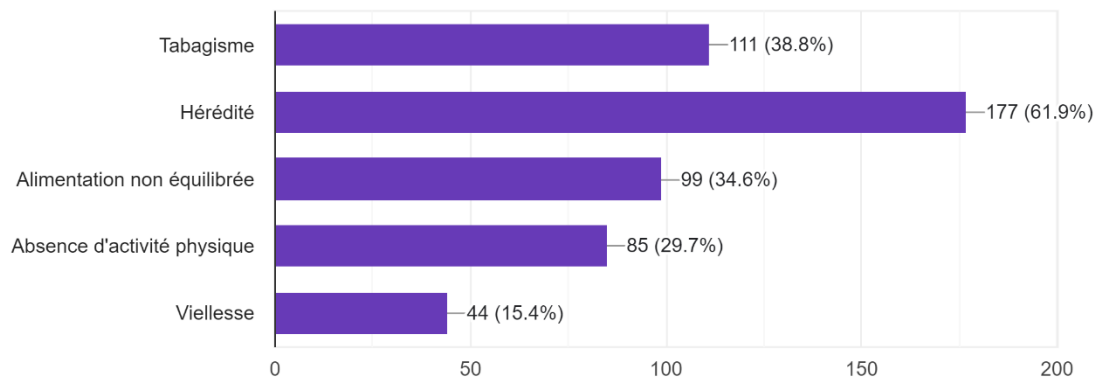


Figure 25 : statistiques sur la connaissance au sujet du cancer du sein 8 .

Selon les répondants, l'hérédité est considérée comme le principal facteur de risque du cancer du sein, avec une majorité de 61,9% des participants le reconnaissant. Cela indique une prise de conscience de la corrélation entre la génétique et le risque de cancer du sein. Le tabagisme est également identifié comme un facteur de risque significatif par 38,8% des participants, suivis de près par l'alimentation non équilibrée à 34,6%. Ces chiffres démontrent une sensibilisation aux effets néfastes du tabagisme et de la malnutrition sur la santé générale et le risque de cancer. L'absence d'activité physique est reconnue par près d'un tiers (29,7%) des répondants comme un facteur de risque. L'importance de l'exercice régulier dans la prévention de diverses maladies, y compris le cancer, semble donc être assez bien comprise. Enfin, la vieillesse, bien que scientifiquement reconnue comme un facteur de risque pour de nombreux types de cancers, n'est identifiée que par 15,4% des participants. Cela pourrait indiquer une méconnaissance de la corrélation entre l'âge et l'augmentation du risque, ou peut-être une concentration des répondants sur des facteurs de risque modifiables. Ces informations peuvent aider à orienter les campagnes de sensibilisation, mettant l'accent sur les facteurs de risque moins compris tout en renforçant la connaissance sur ceux déjà identifiés par le public.

Selon vous quel tranche d'age est la plus susceptible d'avoir le cancer du sein ?

286 responses

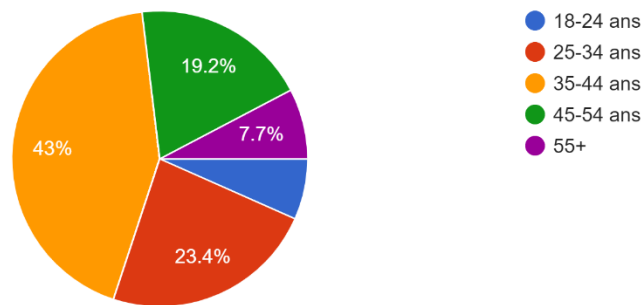


Figure 26 : statistiques sur la connaissance au sujet du cancer du sein 9.

Les perceptions des répondants semblent suggérer que la majorité (43%) croit que la tranche d'âge 35-44 ans est la plus susceptible d'avoir le cancer du sein. Cela est suivi par ceux qui croient que c'est la tranche d'âge 25-34 ans, à 23.4%. Il est intéressant de noter que seulement 19.2% des répondants considèrent la tranche d'âge 45-54 ans comme étant la plus à risque, alors que, dans de nombreuses régions du monde, le risque de cancer du sein augmente effectivement avec l'âge, en particulier après la ménopause. Les perceptions concernant les tranches d'âge les plus jeunes (18-24 ans) et les plus âgées (55+ ans) sont les moins communes, avec respectivement 6.6% et 7.7% des répondants les sélectionnant. Il est possible que ces perceptions soient influencées par les cas médiatisés de cancer du sein chez les femmes plus jeunes ou par les expériences personnelles des répondants. Cependant, cela indique également un besoin de sensibilisation et d'éducation sur le fait que le risque de cancer du sein augmente généralement avec l'âge.

Connaissez-vous dans votre entourage une personne souffrant ou ayant souffert d'une anomalie mammaire ?

286 responses

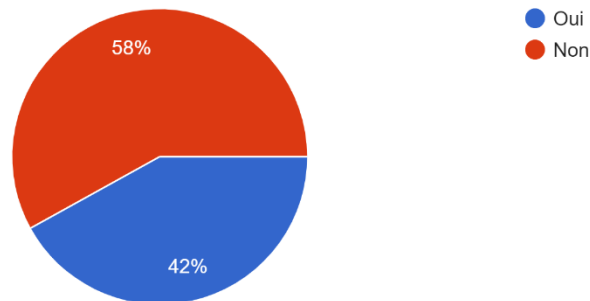


Figure 27 : statistiques sur la connaissance au sujet du cancer du sein 10 .

Il est révélateur que près de la moitié (42%) des répondants connaissent quelqu'un dans leur entourage qui a souffert d'une anomalie mammaire. Cela pourrait signifier que le cancer du sein ou les anomalies mammaires sont relativement courants, du moins parmi les connaissances des répondants. Cependant, cela contraste avec le fait que certains répondants estiment que le sujet du cancer du sein n'est pas souvent discuté dans leur communauté.

La disparité entre la prévalence perçue des anomalies mammaires et le niveau de discussion autour du sujet suggère qu'il y a un manque de sensibilisation et d'éducation autour de la maladie. De plus, cela pourrait signifier que, malgré la familiarité de certains avec la maladie, le sujet reste tabou ou est évité dans certaines discussions ou cercles.

Il est donc primordial de promouvoir une culture de discussion ouverte et éducative autour de la santé mammaire pour garantir que ceux qui sont touchés ou à risque soient bien informés et soutenus.

Connaissez-vous dans votre entourage ou parmi vos connaissances des personnes décédées à cause du cancer du sein ?

286 responses

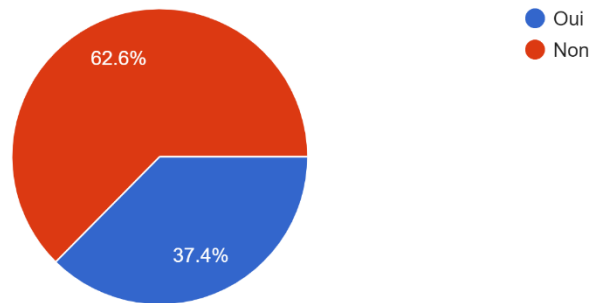


Figure 28 : statistiques sur la connaissance au sujet du cancer du sein 11 .

Il est profondément troublant de constater que plus d'un tiers (37,4%) des répondants connaissent personnellement quelqu'un qui est décédé des suites du cancer du sein. Cette statistique, combinée au taux élevé de personnes connaissant quelqu'un avec une anomalie mammaire, souligne la gravité du problème du cancer du sein au Cameroun. Malgré cela, le fait que le sujet du cancer du sein ne soit pas couramment discuté dans de nombreuses communautés, comme l'ont indiqué les répondants précédemment, est préoccupant. Cela montre qu'il pourrait y avoir une déconnexion entre l'impact réel de la maladie et la sensibilisation du public à son égard. Il est donc essentiel d'intensifier les efforts de sensibilisation et de détection précoce pour réduire le nombre de décès dus au cancer du sein. Une sensibilisation accrue pourrait non seulement aider à détecter et traiter la maladie à un stade précoce, mais aussi à offrir un soutien émotionnel et psychologique aux personnes touchées.

Avez-vous déjà subi un examen du sein ?

286 responses

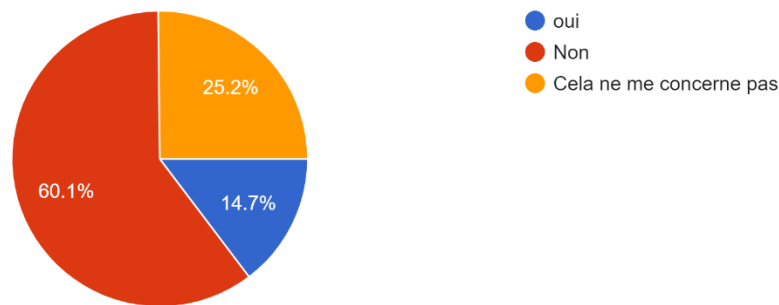


Figure 29 : statistiques sur la connaissance au sujet du cancer du sein 12.

Une majorité significative des répondants (60,1%) n'a jamais subi d'examen du sein, malgré le fait que le cancer du sein soit apparemment assez courant au sein de la communauté, comme le suggèrent les données précédentes. Cela souligne un écart flagrant entre la nécessité d'un dépistage et la réalité du dépistage actuel parmi les répondants. Il est également intéressant de noter que 25,2% des répondants ont indiqué que cela ne les concernait pas. Cette réponse pourrait être attribuée à plusieurs facteurs. Certains pourraient croire que le risque de cancer du sein ne les concerne pas, en raison de leur sexe, de leur âge ou d'autres facteurs démographiques. Toutefois, il est important de noter que le cancer du sein peut toucher tout le monde, y compris les hommes, bien que cela soit plus rare. Cette notion erronée du "cela ne me concerne pas" pourrait également découler d'un manque d'éducation et de sensibilisation à la maladie. Une minorité, seulement 14,7% des participants, a déclaré avoir déjà subi un examen du sein. Cela indique la nécessité d'intensifier les campagnes de sensibilisation sur l'importance du dépistage précoce et régulier, compte tenu de la prévalence du cancer du sein.

Seriez-vous ouvert à l'usage d'une technologie non traditionnelle pour la consultation et le diagnostic des anomalies mammaires.?

286 responses

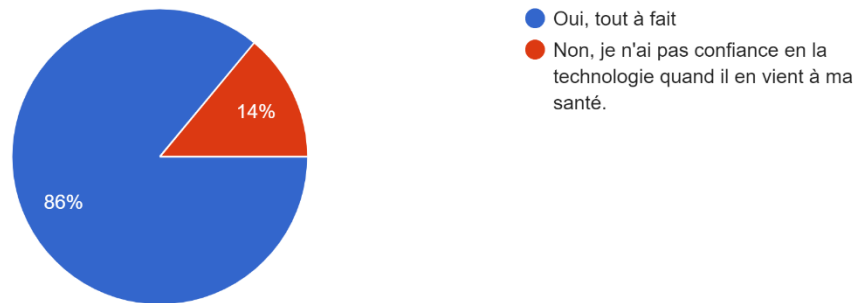


Figure 30 : statistiques sur la connaissance au sujet du cancer du sein 13.

Un impressionnant 86% des répondants se montrent ouverts à l'utilisation de technologies non traditionnelles pour la consultation et le diagnostic des anomalies mammaires. Cela montre une acceptation assez élevée de l'innovation technologique dans le secteur de la santé, même pour des problématiques aussi sensibles que le cancer du sein.

1.1.2 Discussion:

✓ Éducation et Sensibilisation:

Une proportion significative des participants semble manquer d'une compréhension approfondie du cancer du sein, en dépit d'une familiarité générale avec le sujet. En effet, bien que 64.7% prétendent connaître le cancer du sein, cela ne se traduit pas nécessairement par une connaissance des symptômes ou des méthodes de dépistage. Ceci est confirmé par le fait que près de 42.7% des répondants n'ont pas été sensibilisés au cancer du sein à l'école ou dans d'autres cadres éducatifs.

✓ Barrières Sociales et Culturelles:

La nature taboue du cancer du sein est évidente à travers plusieurs indicateurs du questionnaire. 65% des participants indiquent que le cancer du sein est souvent évité ou considéré comme tabou dans leur communauté. Ces statistiques suggèrent une réticence culturelle à discuter ouvertement du sujet, ce qui peut entraver la diffusion d'informations

vitales. De plus, la fréquence limitée de discussions sur le sujet au sein des familles renforce cette notion.

✓ Conscience des Risques et Expérience Personnelle:

Bien que la majorité des participants ne connaissent pas personnellement quelqu'un ayant des anomalies mammaires, un impressionnant 42% connaît quelqu'un qui est décédé du cancer du sein. Cela indique une prise de conscience des dangers du cancer du sein, mais cela pourrait également renforcer la peur et le sentiment d'évitement autour du sujet.

✓ Obstacles Financiers:

L'inquiétude concernant le coût du dépistage est manifeste. Bien que 34.3% soient prêts à payer entre 5,000 FCFA et 15,000 FCFA pour un examen, une large majorité perçoit le dépistage comme peu accessible. Cela suggère que, même si certains sont disposés à payer, ils pourraient considérer le coût comme prohibitif ou estimer qu'il n'y a pas de services abordables à proximité.

✓ Technologie et Innovation:

Malgré un manque général de connaissances et les barrières sociales entourant le cancer du sein, il est intéressant de noter qu'une grande majorité des participants (79.4%) est au courant de l'IA. De plus, 86% des participants sont ouverts à l'utilisation de l'IA dans le domaine de la santé. Cela pourrait indiquer une opportunité pour introduire des solutions technologiques moins coûteuses et plus accessibles pour combattre le manque de sensibilisation et les barrières financières.

La population interrogée présente un contraste évident entre une conscience des dangers du cancer du sein et des barrières culturelles, sociales et financières qui entravent un engagement actif pour le dépistage et la prévention. Les efforts futurs devraient se concentrer sur des méthodes de sensibilisation ciblées et rapprochées qui tiennent compte de ces barrières tout en explorant l'opportunité offerte par la technologie pour rendre le dépistage plus accessible et abordable.

1.2.Synthèse de l'analyse documentaire

Notre analyse documentaire s'est basée principalement sur la consultation d'une vingtaine d'articles scientifiques relatifs aux méthodes de dépistage du cancer du sein, à l'intelligence artificielle, et plus spécifiquement à l'application de cette dernière dans la détection des anomalies mammaires. Nos sources principales étaient les ressources disponibles sur Google Scholar, ainsi que les synthèses d'articles provenant de journaux en ligne tels que AJR et Thema-Radiologie. Il est à noter que nous n'avons pas approfondi notre analyse sur les spécificités médicales du cancer du sein, mais nous avons privilégié les connaissances essentielles dans ce domaine pour proposer une solution adaptée au contexte camerounais.

Selon une étude réalisée par "Cameroun Tribune", mentionnée dans une édition du même quotidien, une mammographie coûterait en moyenne 50 000 francs au Cameroun. Sachant que le SMIG est de 48 000, cette information suscite immédiatement des interrogations quant à la proposition d'une solution basée sur la recommandation des mammographies pour la détection précoce du cancer du sein, d'autant plus que cet examen devrait être répété tous les deux ans en moyenne.

Dans le cadre de nos recherches, nous nous sommes penchés sur plusieurs travaux, dont un intitulé "Contribution de l'échographie dans le dépistage et le diagnostic des cancers du sein : une expérience de Yaoundé, Cameroun". Cette étude a démontré, sur un échantillon de 103 femmes, qu'il était possible de détecter une anomalie mammaire via une simple échographie mammaire. Conduite dans un contexte socio-économique qui tenait compte des faibles revenus du Camerounais moyen, cette étude a révélé comme principal obstacle la capacité du personnel médical à interpréter correctement toutes les échographies. Au final, l'échographie a permis de détecter 32,25 % de lésions chez des femmes dont l'examen clinique ne révélait aucune anomalie. De ces lésions suspectes, 14 ont été identifiées, parmi lesquelles 8 ont été confirmées comme étant cancéreuses à l'anatomopathologie.

Un autre article notable est celui du Dr. Berg intitulé "Toward AI-supported US Triage of Women with Palpable Breast Lumps in a Low-Resource Setting". Avec ses collaborateurs, en se basant sur la WBD, elle tente d'apporter une solution à moindre coût pour la détection du cancer dans les zones avec un accès médical limité. La conclusion majeure de leurs recherches est que, en combinant l'échographie à l'intelligence artificielle, il est possible de diagnostiquer à temps un certain nombre de cas d'anomalies mammaires. En particulier, les

anomalies dont l'échographie révèle un comportement suspect pourraient bénéficier d'une évaluation différente, approfondie, et adaptée, optimisant ainsi les ressources disponibles et offrant une chance accrue de détection précoce pour les patientes concernées.

2. Intervention proposée et justification

Sur la base des données récoltées et des analyses faites, la nécessité d'une solution adaptée au contexte camerounais est évidente. Les barrières financières, le manque de sensibilisation et la réticence sociale à aborder ouvertement les questions liées au cancer du sein sont autant de défis majeurs à surmonter. Une intervention technologique bien pensée et axée sur l'utilisateur peut offrir une solution viable.

2.1 Description de la solution :

Nous proposons la mise en place d'une application web dénommée "DePreCaS". Cette application sera composée de deux modules principaux :

Un assistant virtuel spécialisé pour répondre en temps réel aux questions des utilisateurs concernant le cancer du sein, les anomalies mammaires, les procédures de dépistage et d'autres sujets connexes. Ce chat bot serait formé à partir de données fiables, avec des réponses conçues pour être à la fois informatives et rassurantes, tout en tenant compte des réalités culturelles et sociales du Cameroun.

Un algorithme de pointe intégrée pour l'analyse d'images échographiques mammaires. Les utilisateurs pourront télécharger leurs échographies mammaires pour obtenir un pré-diagnostic. Si une anomalie est détectée, l'application conseillera une consultation médicale pour un examen approfondi.

1.1. Justification de la solution :

- ✓ Accessibilité financière : Le coût prohibitif d'une mammographie au Cameroun

rend difficile le dépistage régulier pour une grande majorité. Cependant, comme discuté précédemment, l'échographie mammaire est une alternative moins coûteuse et efficace, surtout lorsqu'elle est combinée à l'intelligence artificielle pour l'interprétation.

- ✓ Éducation et sensibilisation : La méconnaissance du cancer du sein et des méthodes de dépistage est répandue. Le module chatbot offrirait une plateforme accessible et anonyme pour que les personnes obtiennent des informations essentielles sans crainte du jugement social.
- ✓ Optimisation des ressources médicales : Le CNN peut traiter un grand nombre d'images en un temps réduit, permettant ainsi de filtrer efficacement les cas qui nécessitent une attention médicale. Cela permettrait de réduire la pression sur les établissements médicaux et de concentrer les ressources sur les cas nécessitant une intervention.
- ✓ Intégration technologique : Avec l'adoption croissante des smartphones et de la technologie mobile au Cameroun, une application mobile est une plateforme idéale pour toucher un large public.

En conclusion, "DePreCaS" pourrait combler le fossé entre les besoins médicaux et les défis socio-économiques rencontrés par de nombreuses Camerounaises en matière de dépistage du cancer du sein. Cette application serait non seulement une plateforme d'information et de sensibilisation, mais aussi un outil puissant de diagnostic préliminaire.

2.2 Objectifs de l'intervention

2.2.1 Objectif général

Faciliter et accélérer le dépistage précoce du cancer du sein au Cameroun à travers une approche technologique intégrée.

2.2.2 Objectifs spécifiques

✓ Éducation et sensibilisation :

Mettre en place un Chatbot intégré à l'application pour éduquer, informer, et répondre aux questions des utilisateurs concernant le cancer du sein et les méthodes de dépistage, tout en abordant et démystifiant les tabous associés.

✓ Détection précoce à l'aide de la technologie :

Intégrer un CNN de pointe à l'application pour analyser rapidement les échographies mammaires, identifier les anomalies potentielles et fournir une première évaluation.

✓ Réduction des barrières économiques et géographiques :

Proposer une solution mobile accessible, qui minimise la nécessité de consultations coûteuses et fréquentes, permettant ainsi à un plus grand nombre de femmes, y compris dans des zones reculées, d'accéder à des informations et des évaluations préliminaires sur leur santé mammaire.

3. Composantes de l'intervention envisagée

Comme mentionné plus haut notre intervention consistera à la mise en place d'une solution numérique basé sur l'intelligence artificielle et ayant comme composantes principales deux modules indépendants à savoir :

- Un chat bot c'est à dire un assistant virtuel spécialisé pour répondre en temps réel aux questions des utilisateurs concernant le cancer du sein, les anomalies mammaires, les procédures de dépistage et d'autres sujets connexes. Ce chat bot serait formé à partir de données fiables, avec des réponses conçues pour être à la fois informatives et rassurantes, tout en tenant compte des réalités culturelles et sociales du Cameroun.
- Un module de diagnostic par Deep Learning basé sur un CNN (Réseau Neuronal Convolutif) de pointe intégrée pour l'analyse d'images échographiques mammaires. Les utilisateurs pourront télécharger leurs échographies mammaires pour obtenir un pré-diagnostic. Si une anomalie est détectée, l'application conseillera une consultation médicale pour un examen approfondi.

4. Stratégies d'action

4.1. Analyse des Besoins:

4.1.1 Identification des utilisateurs cibles:

Lors de l'identification des utilisateurs cibles, les groupes suivants sont à considérer :

Patients : Individus qui cherchent des informations sur le cancer du sein, les anomalies mammaires et les procédures de dépistage. Ils pourraient aussi vouloir utiliser le module de diagnostic pour obtenir un pré-diagnostic à partir de leurs échographies.

Professionnels de santé : Médecins, radiologues et infirmières qui pourraient recommander l'application à leurs patients ou l'utiliser comme outil pour des procédures de soin préliminaires ou complémentaires.

Organismes de sensibilisation au cancer : Ces groupes peuvent utiliser l'application pour informer et éduquer la population sur le cancer du sein.

Éducateurs et étudiants : Ceux qui souhaitent obtenir des informations à des fins éducatives.

Proches des patients : Familles et amis des personnes touchées par le cancer du sein qui cherchent à mieux comprendre la maladie.

4.2. Recueil des exigences fonctionnelles et non fonctionnelles

1.2.1 Exigences fonctionnelles

Ce sont les fonctionnalités que l'application doit offrir. Pour DePreCaS nous aurons :

- **Interface de chatbot**: L'application doit permettre aux utilisateurs de poser des questions et de recevoir des réponses en temps réel.
- **Module de diagnostic par Deep Learning** : Les utilisateurs doivent pouvoir télécharger leurs échographies mammaires et obtenir un pré-diagnostic.
- **Notifications** : En cas d'anomalie détectée, l'application devrait envoyer une notification à l'utilisateur lui conseillant une consultation médicale.
- **Base de connaissances** : L'application devrait avoir une section FAQ où les utilisateurs peuvent trouver des réponses à des questions courantes.
- **Multilinguisme** : Étant donné les différentes langues parlées au Cameroun,

l'application pourrait offrir plusieurs options de langue.

1.2.2 Exigences non fonctionnelles

Ce sont les caractéristiques que l'application doit posséder pour garantir une bonne expérience utilisateur et une performance optimale.

- **Performance:** L'application doit être rapide et réactive.
- **Sécurité:** Toutes les données téléchargées par les utilisateurs pour le diagnostic doivent être protégées et confidentielles.
- **Compatibilité:** L'application doit être compatible avec différents appareils et systèmes d'exploitation.
- **Scalabilité:** L'infrastructure doit être capable de gérer un grand nombre d'utilisateurs simultanément.
- **Intuitivité:** L'interface doit être simple et facile à utiliser.
- **Accessibilité:** L'application devrait être accessible aux personnes handicapées.
- **Fiabilité:** L'application doit être stable et ne pas présenter de bugs fréquents.
- Avec une compréhension claire des utilisateurs cibles et des exigences, vous pourrez mieux concevoir et développer une solution qui répond aux besoins des utilisateurs.

4.3. Sélection de la technologie

4.3.1 Environnement de Développement :

Afin de développer notre solution nous utiliserons principalement deux environnements Jupiter et Colab.



Figure 31 : Logo jupyter et Google Colab

Description : D'après Wikipédia Jupyter est une application web utilisée pour programmer. Jupyter permet de réaliser des calepins ou notebooks, c'est-à-dire des programmes contenant à la fois du texte, simple ou enrichi typographiquement et

sémantiquement grâce au langage à balises simplifié Mark down, et du code, lignes sources et résultats d'exécution. Google Colab est Une plateforme basée sur le cloud qui fournit un environnement Jupyter pour le développement et l'exécution de scripts Python, notamment dans les domaines du Machine Learning et du Deep Learning.

Raison de l'utilisation : Google Colab Offre un accès gratuit à des GPU et TPU, facilitant l'entraînement de modèles d'apprentissage profond. L'intégration directe avec TensorFlow en fait un choix idéal pour développer le module de diagnostic. Jupiter quant à lui es propice pour les librairies difficiles à manipuler sur Colab

4.3.2 *Application et Interface:*

Afin de faciliter la réalisation de notre application nous allons nous concentrer sur les algorithmes et laisser Streamlit gérer le reste.



Figure 32 : Logo Streamlit

Description: C'est un outil pour transformer des scripts Python en applications web partageables.

Raison de l'utilisation: Facilité d'utilisation et intégration simple avec Python. Idéal pour créer des interfaces interactives pour des applications d'IA sans se plonger dans des détails complexes de développement web.

4.3.3 Framework d'Apprentissage Profond:



Figure 33 : Logo Tensor Flow

Description: C'est un framework open source développé par Google pour l'apprentissage automatique et l'apprentissage profond.

Raison de l'utilisation: Reconnu pour sa flexibilité et sa performance, il est particulièrement adapté aux CNN, ce qui est essentiel pour l'analyse d'images échographiques mammaires.

4.3.4 Développement de Chatbot:



Figure 34 : Logo Open AI

Description: Open AI API nous permettra d'accéder à des modèles d'IA avancés, tels que GPT (Generative Pre-trained Transformer).

Raison de l'utilisation: Avec cette API, le développement du chatbot sera simplifié, tout en offrant des capacités de génération de texte de haute qualité. Cela permet de construire un chatbot informatif et réactif avec moins d'efforts en matière de formation et de développement.

4.3.5 Outils supplémentaires:



Figure 35 : Logo GitHub

Version Control avec Git: Étant donné que nous travaillons sur un projet technologiquement avancé, l'utilisation d'un système de contrôle de version comme Git (via GitHub ou GitLab) est vivement recommandée pour gérer les mises à jour et la collaboration.



Figure 36 : Logo Google Drive

Stockage des données: En utilisant Google Colab, Google Drive est le choix naturel pour le stockage des données.

4.4. Développement du Chatbot:

4.4.1 Collecte de données:

Sources potentielles :

- ✓ Interviews avec des professionnels de santé spécialisés dans le domaine du cancer du sein.
- ✓ Forums et communautés en ligne centrés sur le cancer du sein ou la santé en général.
- ✓ FAQ des organisations de santé reconnues.
- ✓ Retours et interrogations fréquentes de patients ou de personnes concernées.

- ✓ Procédés cliniques d'interventions médicales

4.4.2 Formation du Chatbot:

- ✓ Formation :

Nous Utiliserons le modèle GPT via l'API OpenAI pour former le chatbot afin qu'il réponde de manière précise et pertinente.

```
openai.api_key = st.secrets["openai_api_key"]

if "openai_model" not in st.session_state:
    st.session_state["openai_model"] = "gpt-3.5-turbo"
```

Figure 37 : intégration de l'api de openAi

4.4.3 Intégration culturelle :

- ✓ Localisation linguistique :

Bien que le français soit l'une des langues officielles du Cameroun, il est essentiel de s'assurer que le langage utilisé est clair, simple et compréhensible pour une large audience. Il sera préférable pour nous d'éviter le jargon médical complexe. Utilisez plutôt des termes couramment utilisés et compréhensibles par le grand public.

- ✓ Sensibilité culturelle :

Nous intégrerons des informations spécifiques au Cameroun, comme les centres de traitement disponibles, les initiatives locales de sensibilisation, etc étant conscient des croyances et des pratiques culturelles liées à la santé et à la maladie.

- ✓ Engagement communautaire :

Nous impliquerons la communauté locale dans le processus de développement en organisant des sessions de feedback avec des utilisateurs potentiels pour s'assurer que le chatbot répond bien à leurs besoins et respecte leurs sensibilités tout en gardant ces étapes à l'esprit et en mettant l'accent sur l'intégration culturelle, nous serons en mesure de développer un chatbot qui non seulement fournit des informations précises sur le cancer du sein, mais qui est aussi adapté et sensible aux besoins et aux réalités du public camerounais.

4.5.Développement du module diagnostique

Pour la réalisation du module diagnostique nous allons réaliser un CNN le plus optimal possible afin de minimiser les erreurs de prédiction. Ains nos métriques d'évaluations seront l'accuracy⁴⁰, le loss⁴¹, la matrice de confusion⁴², le classification report⁴³ et le AUC.

4.5.1 Acquisition des données :

Le MinSanté ne possédant pas de base de données des échographies mammaires réalisées par les différents hôpitaux et centres radiologiques, nous nous orientons vers un ensemble de données public d'échographies mammaires. Les données collectées initialement comprennent des images d'échographie mammaire de femmes âgées de 25 à 75 ans. Ces données ont été collectées en 2018. Le nombre de patientes est de 600. L'ensemble de données se compose de 780 images avec une taille moyenne d'image de 500 x 500 pixels. Les images sont au format PNG. Elles sont classées en trois catégories : normales, bénignes et malignes et chaque échographie possède un ou plusieurs masques⁴⁴.

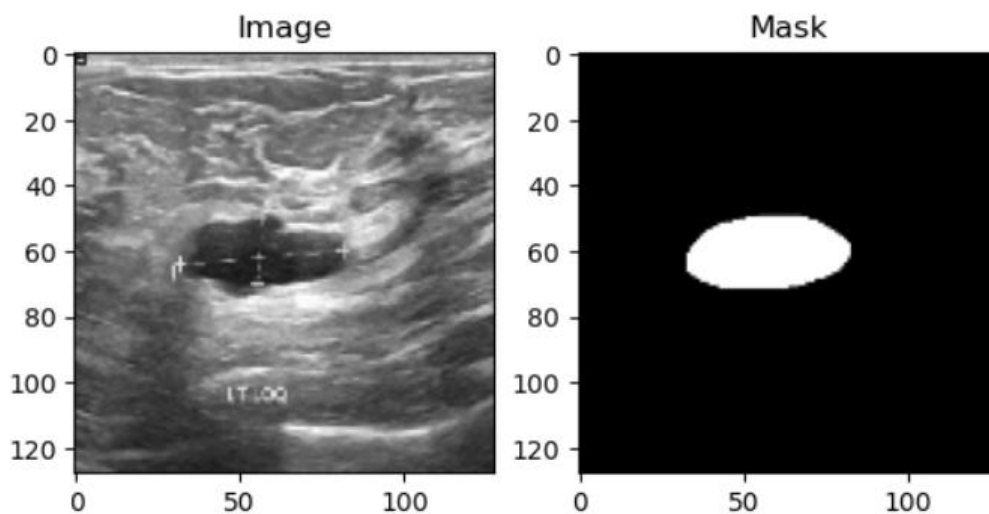


Figure 38 Aperçu des images du Data set

⁴⁰ Pour les tâches de classification, il s'agit du rapport entre le nombre de prédictions correctes et le nombre total de prédictions.

⁴¹ Indicateur Typiquement utilisé pendant l'entraînement pour surveiller la convergence.

⁴² Une table qui montre les prédictions correctes et incorrectes pour chaque classe. Utile pour voir où le modèle se trompe le plus.

⁴³ C'est un outil d'évaluation qui présente les principales métriques de classification pour chaque classe d'un problème de classification. (F1_score, Accuracy, Précision, Recall)

⁴⁴ Dans le contexte de la vision par ordinateur, un "masque" pour une image est une image, généralement binaire, qui est utilisée pour extraire, masquer ou mettre en évidence certaines parties d'une autre image

4.5.2 Prétraitement des images

Quel que soit le modèle à créer, il est essentiel d'effectuer un certain nombre d'opérations préalables afin de préparer les données nécessaires pour l'entraînement. Ces opérations passent généralement par l'épuration du dataset, la normalisation des données pour faciliter la convergence dans certains cas, ou même la data augmentation en cas de données insuffisantes.

4.5.3 Conception, formation et validation du modèle

➤ Modèle N°1

Notre premier modèle est un basé sur un CNN simple dont l'architecture est la suivante :

```
#Initialisation du CNN
classifier = Sequential()

#Convolution
classifier.add(Conv2D(16,(3,3),input_shape=(64,64,3),activation="relu"))

#Max Pooling
classifier.add(MaxPooling2D(pool_size=(2,2)))

#Flattening
classifier.add(Flatten())

#Full connection
classifier.add(Dense(units=128,activation="relu"))
classifier.add(Dense(units=3,activation="softmax"))
```

Figure 39 : Architecture du model

Après entrainement de notre modèle, nous obtenons les métriques d'évaluation suivantes :

	precision	recall	f1-score	support
benign	0.58	0.71	0.64	223
malignant	0.35	0.28	0.31	106
normal	0.14	0.07	0.10	67
accuracy			0.49	396
macro avg	0.35	0.36	0.35	396
weighted avg	0.44	0.49	0.46	396

Figure 40 Classification report : Model 1.

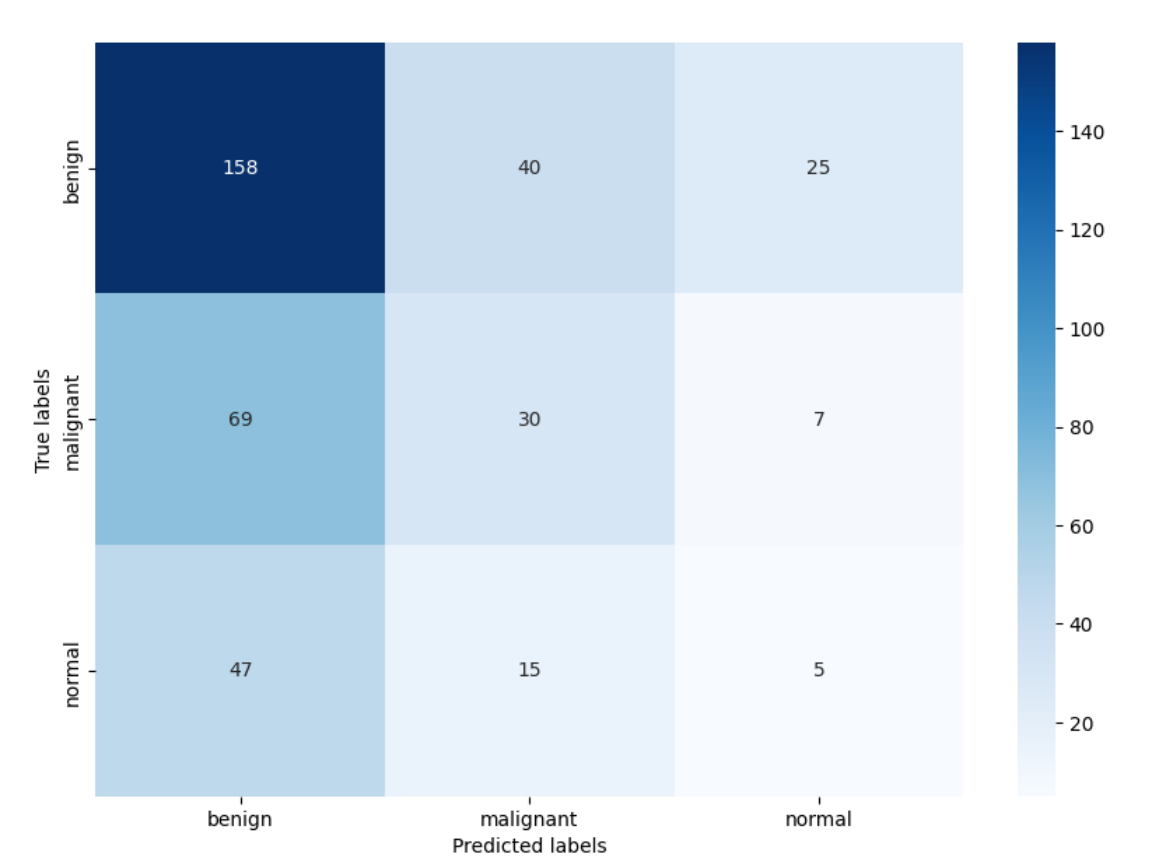


Figure 41 : Matrice de confusion Modele 1

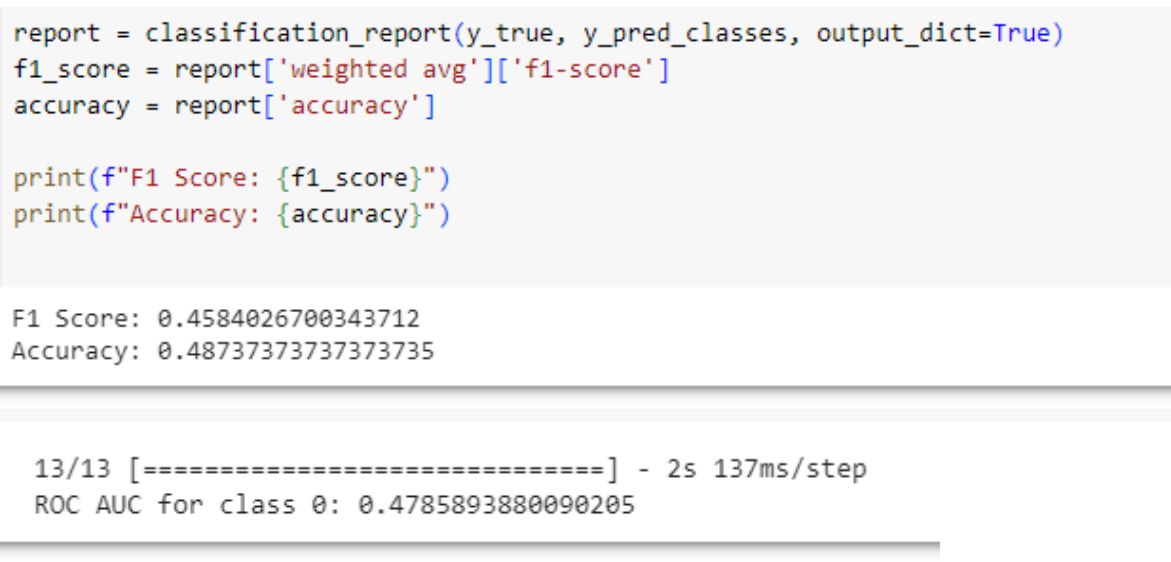


Figure 42/ Métriques (F1 score , accuracy et ROC) du model 1

Commentaire : Avec les métriques étudiées nous remarquons que ce modèle n’est pas très efficace malgré l’optimisation des hyperparamètres.

➤ Modèle N°2

Aux vues de l'inefficacité du premier modèle nous tentons une approche basée sur un réseau pré entraîné, notre choix se porte dans un premier temps sur le modèle ResNet50.

L'architecture du modèle résultant est donc le suivant :

```
[51] base_model = tf.keras.applications.ResNet50(include_top=False, weights="imagenet", input_tensor=None, input_shape=img_shape, pooling=None)
    for layer in base_model.layers:
        layer.trainable = False

model = Sequential()
model.add(base_model)

model.add(Flatten())
model.add(BatchNormalization())
model.add(Dense(256, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(3, activation='softmax'))

model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['acc'])
```

Figure 43 : Architecture model basée sur ResNet 50

Après training sur le dataset mentionné plus haut et optimisation des hyperparamètres nous obtenons comme métrique d'évaluation les éléments suivants :

✓ *Accuracy, F1_score et ROC*

```
19/19 [=====] - 40s 2s/step - loss: 0.3271 - acc: 0.8977
3/3 [=====] - 3s 819ms/step - loss: 0.6008 - acc: 0.7164
4/4 [=====] - 7s 1s/step - loss: 0.6670 - acc: 0.7436
4/4 [=====] - 7s 2s/step
Train accuracy = 0.8976510167121887
Validation accuracy = 0.7164179086685181
Test accuracy = 0.7435897588729858
f1_measure = 0.7381599548266214
roc_area = 0.8851472725161549
```

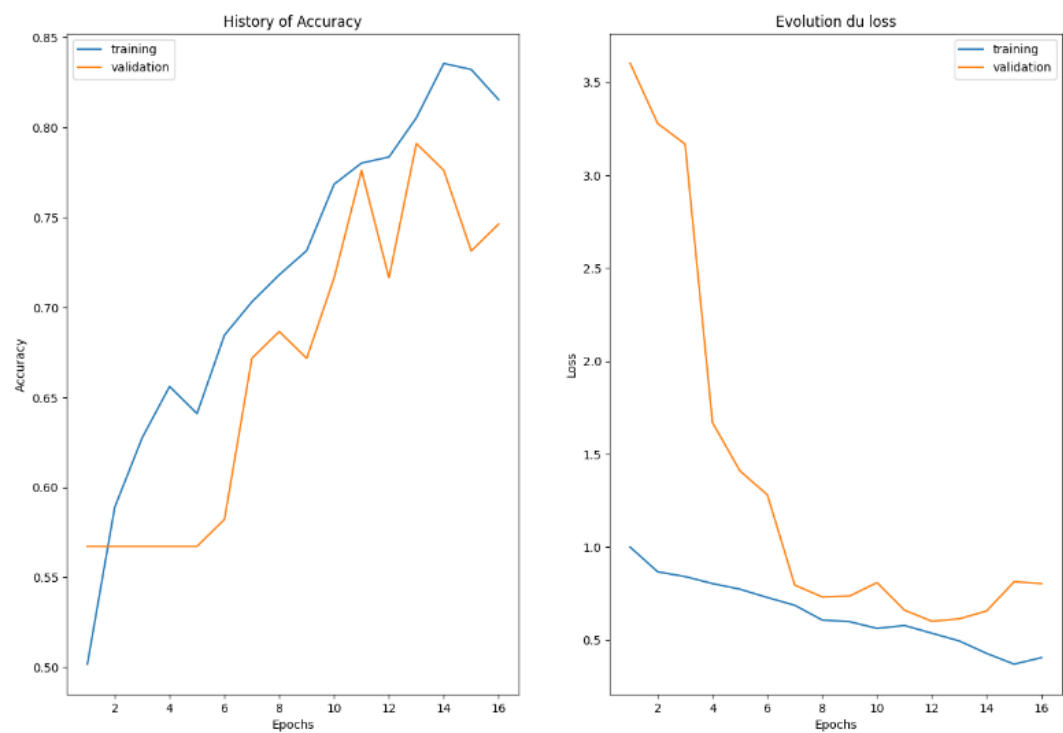


Figure 44 : Training loss, Val loss, Acc, Val_acc du modele 2

✓ *Classification report*

	precision	recall	f1-score	support
0	0.75	0.86	0.80	58
1	0.71	0.68	0.69	37
2	0.80	0.55	0.65	22
accuracy			0.74	117
macro avg	0.75	0.69	0.71	117
weighted avg	0.75	0.74	0.74	117

Figure 45 Classification report ; Modele 2

✓ *Confusion Matrix*

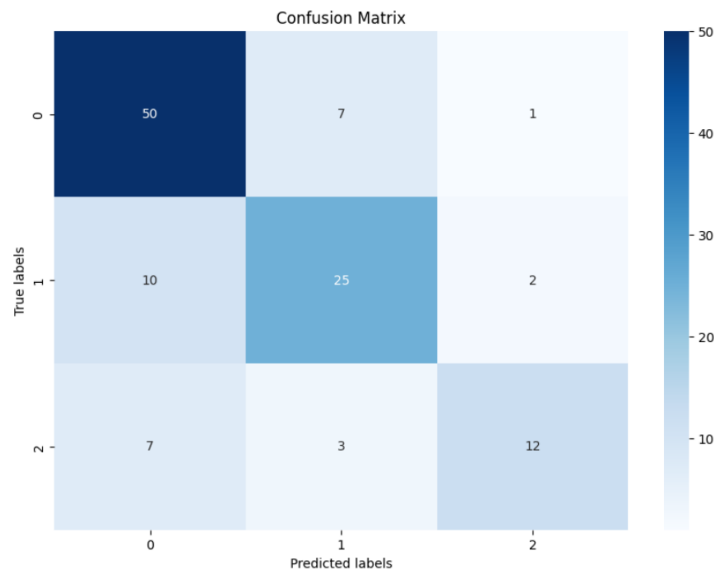


Figure 46 : Matrice de confusion Modèle 2

Commentaire : Nous remarquons une évolution notoire dans les différents paramètres d'évaluation rendant ce modèle pré entraîné de loin meilleur que le précédent.

➤ Modèle N°3 :

Dans la mise en place de cette 3^{ème} architecture CNN nous nous sommes pencherons sur la structure des données et de leurs masques en particulier afin de maximiser l'usage de leurs caractéristiques.

Architecture du Réseau.

Notre modèle sera principalement basé sur le modèle pré entraîné VGG16 qui VGG16 est un modèle de réseau de neurones convolutifs (CNN) qui a été développé et entraîné par l'Université d'Oxford's Visual Geometry Group (VGG). Il a été introduit en 2014 et est l'un des modèles populaires utilisés dans le domaine de la reconnaissance d'images.


```
[40] base_model = tf.keras.applications.VGG16(  
    include_top=False,  
    weights="imagenet",  
    input_tensor=None,  
    input_shape=img_shape,  
    pooling=None  
)  
  
for l in base_model.layers:  
    l.trainable = False  
  
[41] model = Sequential()  
model.add(base_model)  
  
model.add(Flatten())  
model.add(BatchNormalization())  
model.add(Dense(512,activation='relu'))  
model.add(Dropout(0.5))  
model.add(Dense(256,activation='relu'))  
model.add(Dropout(0.5))  
model.add(Dense(128,activation='relu'))  
model.add(Dropout(0.5))  
model.add(Dense(64,activation='relu'))  
model.add(Dropout(0.5))  
model.add(Dense(3,activation='softmax'))  
  
model.compile(optimizer='adam',loss='categorical_crossentropy',metrics=['acc'])
```

Figure 47 : Modèle CNN basé sur VGG 16

```
Train accuracy = 0.9983221292495728  
Validation accuracy = 0.9701492786407471  
Test accuracy = 0.9829059839248657  
f1_measure = 0.9829059829059829  
roc_area = 0.9993538194411942
```

Figure 48 : Metrique du modele 3 (F1 Score Roc et Accuracy)

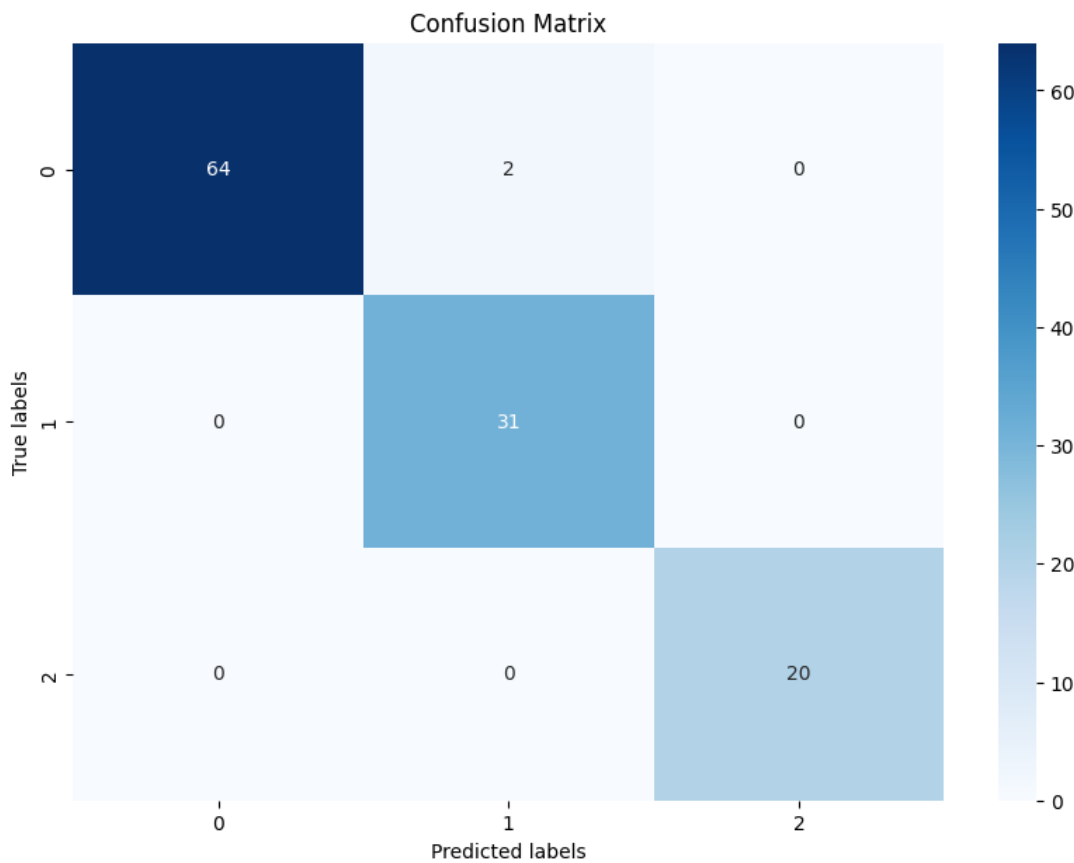


Figure 49 : matrice de confusion du modèle 3

	precision	recall	f1-score	support
0	1.00	0.97	0.98	66
1	0.94	1.00	0.97	31
2	1.00	1.00	1.00	20
accuracy			0.98	117
macro avg	0.98	0.99	0.98	117
weighted avg	0.98	0.98	0.98	117

Figure 50 :Classification repport Modèle 3

4.5.4 Choix du meilleur modèle :

Le tableau suivant fait un récapitulatif des modèles entrainés ainsi que leurs métriques d'évaluation.

N°	Accuracy	F1_Score	ROC
CNN simple	0.48	0.45	0.47

Modèle avec pour couche de base ResNet 50	0.74	0.73	0.88
Modèle avec pour couche de base VGG16	0.98	0.98	0.99

Tableau 6 : Tableau comparatif des modèles entraînés.

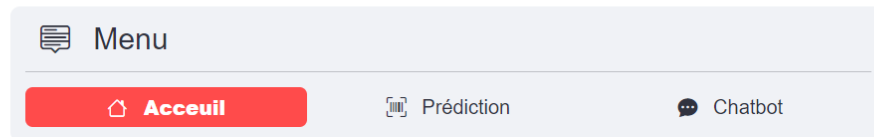
Etant donné ces résultats notre choix se porte sur le modèle N°3 vu ses performances.

4.6. Interface utilisateur (UI):

Nous Utilisons Streamlit pour créer une interface utilisateur interactive et attrayante. Streamlit facilite la conception d'applications axées sur les données et est particulièrement adapté aux applications basées sur l'IA.

Ainsi Intégrons des widgets, comme des curseurs ou des boutons, pour permettre aux utilisateurs de télécharger des échographies mammaires ou de poser des questions au chatbot.

DePreCas



Bienvenue à DePreCaS, votre solution dédiée à la détection du cancer du sein au Cameroun. Plus qu'une simple plateforme, nous sommes votre assistant pour répondre à toutes vos questions relatives à cette maladie. Conçue tant pour les patients que pour les professionnels de santé et les proches des malades, DePreCaS est là pour vous informer, vous guider et vous soutenir dans cette lutte contre le cancer du sein. Comment pouvons-nous vous aider aujourd'hui ?

Figure 51 : Interface d'accueil de DePreCaS

DePreCas

Menu

Acueil

Prédiction

Chatbot

Prédiction

Bien vouloir importer l'echographie

Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG, PNG

Browse files

Figure 52 : Interface de prédiction de DePreCas

Prédiction

Bien vouloir importer l'échographie



Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG, PNG

Browse files



malignant (100).png 290.2KB

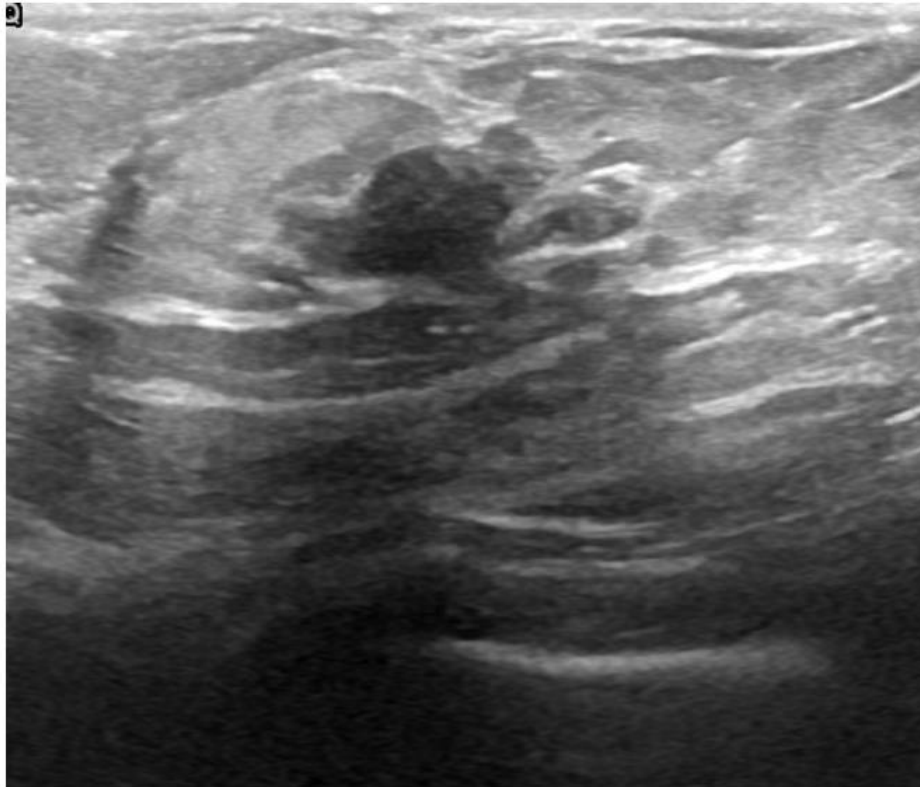


Image téléchargée.

Malign

Figure 53 : Exemple de prédiction avec DePreCas

4.7. Sécurité:

Afin de garantir la sécurité et la confidentialité des données médicales manipulées sur notre application nous utiliserons le cryptage SSL/TLS pour protéger les données en transit entre le client et le serveur. Le stockage des données quant à lui se fera dans un environnement sécurisé et crypté.

4.8. Déploiement:

Pour l'hébergement nous Considérons des options telles que Google Cloud, AWS, ou Azure, qui offrent des solutions robustes pour héberger des applications basées sur l'IA.

4.9.Lancement:

Nous devons nous Assurer que l'application est optimisée, testée, et prête pour un déploiement à grande échelle.

4.10. Maintenance et mises à jour:

Il est important de planifiez une surveillance continue pour identifier rapidement tout problème éventuel. Ainsi nous prévoyez des mises à jour régulières pour intégrer les améliorations et répondre aux feedbacks des utilisateurs.

5. Faisabilité

5.1. Faisabilité Économique:

ELEMENTS DU BUDGET	NOMBRE DE MOIS	QUANTITE	COUT MENSUEL (en FCFA)	COUT TOTAL (en FCFA)
Coûts de main-d'œuvre				
Ingénieurs en logiciel				
Phase de recherche et developpement	7	1	200000	1400000
Phase de test	3	1	200000	600000
Chercheurs en IA				
Phase de recherche et developpement	7	2	300000	4200000
Phase de test	3	1	200000	600000
Expert en santé	4	1	480000	1920000
Gestionnaire de projet	12	1	70000	840000
Alternants et stagiaires	12	5	20000	1200000
Coûts matériels et logiciels				
Serveurs cloud	12	2	78000	1872000
Licence des logiciels	12	3	5000	180000

Ordinateurs	1	8	150000	1200000
Carte Graphiques (GPU)	1	3	25000	75000
Disques durs externe	1	3	15000	45000
Coûts liés aux données				
Collecte et annotation des images	2	1	15000	30000
Stockage des données	12	1	8000	96000
Sécurité des données	12	1	10000	120000
Formation				
Formation du personnel	5	5	20000	500000
Réserve budgétaire (10 %)				
Risques et imprévus	1	0,1	14878000	1487800
Coût total estimé				16365800

Tableau 7 : budget estimatif pour la réalisation de notre solution

Ressources nécessaires : Le développement d'une application nécessite non seulement du temps mais aussi des ressources financières. Cela inclut les coûts d'hébergement pour le serveur, l'achat de données pour former le CNN (si elles ne sont pas disponibles gratuitement), les licences pour certains logiciels ou API, et les coûts de marketing pour promouvoir l'application.

Retour sur investissement : Bien que le coût initial puisse être élevé, la solution pourrait potentiellement attirer l'attention des investisseurs, des institutions médicales ou des organisations gouvernementales. Elle pourrait aussi servir comme prototype pour des solutions plus vastes à l'avenir, offrant un potentiel de monétisation.

Financements et Subventions : Il pourrait être possible de rechercher des subventions universitaires, des financements de départ, ou des partenariats avec des ONG pour couvrir certains des coûts initiaux.

5.2.Faisabilité Sociale:

Acceptation par la communauté Étant donné le coût prohibitif des mammographies au Cameroun, il y a une forte probabilité que la solution soit bien accueillie par les utilisateurs cibles, surtout si elle est offerte à un coût abordable ou gratuitement.

Éducation et sensibilisation Une partie de la faisabilité sociale serait d'assurer une sensibilisation suffisante à l'importance de la détection précoce et comment l'application peut aider dans ce processus. Cette sensibilisation pourrait nécessiter des fonds pour les campagnes.

Considérations culturelles : Le tabou social associé à la discussion sur le cancer du sein devrait être pris en compte. Des campagnes de sensibilisation pour encourager l'adoption de l'application pourraient nécessiter des fonds supplémentaires.

5.3.Faisabilité Technique

Expertise en IA : Bien que nous ayons les compétences nécessaires en IA, le développement d'un CNN efficace et optimisé nécessite des logiciels ou des services (Puissance de calcul) payants.

Accès aux données : L'achat ou l'accès à une base de données d'échographies mammaires coûterait de l'argent si elles ne sont pas disponibles gratuitement.

Infrastructure : L'hébergement, le stockage et la sécurité nécessitent des fonds, surtout pour maintenir la confidentialité des données des utilisateurs.

5.4.Faisabilité Environnementale

Impact sur l'environnement : Une application numérique a généralement un impact minimal sur l'environnement. Cependant, l'électricité et les ressources utilisées pour les serveurs et les dispositifs des utilisateurs doivent être prises en compte.

Durabilité : À long terme, une application basée sur le cloud serait durable, mais elle implique des coûts récurrents d'hébergement.

Consommation d'énergie : L'énergie utilisée pour faire fonctionner l'application, pour la formation du CNN, et pour les appareils des utilisateurs serait un coût à prendre en compte.

Conclusion

L'analyse approfondie de la situation actuelle au Cameroun révèle un paysage contrasté en matière de sensibilisation et de dépistage du cancer du sein. Si d'un côté, il y a une prise de conscience des risques associés à cette maladie, de l'autre, des barrières socioculturelles, éducatives et financières persistent. L'innovation technologique se présente cependant comme une lueur d'espoir, offrant la possibilité de transcender ces obstacles et de fournir des solutions accessibles et efficaces. En tenant compte des réalités et des défis propres au contexte camerounais, les interventions basées sur la technologie et l'éducation peuvent ouvrir la voie à une prévention et à un dépistage plus efficace, changeant ainsi potentiellement la trajectoire de nombreuses vies. En fin de compte, l'objectif est clair : assurer que chaque femme au Cameroun ait accès aux informations nécessaires et aux moyens de dépistage afin de garantir une meilleure santé mammaire.

CONCLUSION GENERALE

. Notre recherche avait pour objectif de déterminer si l'intelligence artificielle peut être utilisée pour améliorer la détection précoce du cancer du sein chez les femmes au Cameroun. À travers l'enquête menée auprès de la population, nous avons identifié un certain nombre de problèmes entravant la détection précoce de cette maladie. La première barrière identifiée est d'ordre économique, suivie du contexte socioculturel qui rend le sujet tabou et conduit à une mauvaise information des populations. Ceci accentue le taux élevé de cas détectés tardivement.

Grâce à une approche à la fois quantitative et qualitative, nous avons élaboré et validé une solution. Cette solution, que nous avons nommée DePreCaS, est une application intelligente basée sur un réseau de neurones convolutifs (CNN) qui permet de prédire le cancer du sein à partir d'échographies mammaires. Elle présente donc l'avantage de réduire les coûts et de surmonter la barrière économique.

Le deuxième obstacle, relatif au contexte socioculturel, est adressé par un autre volet de notre solution : un assistant virtuel intelligent. Cet assistant est conçu pour discuter aisément du cancer du sein, tant avec les personnes suspectées d'en être atteintes qu'avec les professionnels de la santé, afin d'aider à établir des protocoles de soins appropriés, entre autres.

Notre solution offre donc une réponse pertinente aux besoins de la société camerounaise. Néanmoins, nous reconnaissons qu'il y a des améliorations possibles. À l'avenir, il serait intéressant d'intégrer une prédiction basée sur les antécédents médicaux et familiaux des patients.

REFERENCES BIBLIOGRAPHIQUES

- [1]. GAËL. (2019, 10 25). Machine Learning., from Datakeen: <https://datakeen.co/8-machine-learning-algorithms-explained-in-human-language/>
- [2]. Issarane, H. (2019, 02 09). Apprentissage Non Supervisé., from Le DataScientist: <https://le-datascientist.fr/apprentissage-non-supervise>
- [3]. Li, H. (2017, 04 24). Which ML Algorithms to Use?, from The Eponymous Pickle: <http://eponymouspickle.blogspot.com/2017/04/which-ml-algorithms-to-use.html>
- [4]. Rouse, M. (2019, 07 03). Que signifie Réseaux de neurones récurrents? - Definition IT de Whatis.fr2019. from LeMagIT: <https://www.lemagit.fr/definition/Reseaux-de-neurones-recurrents>
- [5]. University of Wisconsin . (2020). Discover UW, from wisc: <https://www.wisc.edu>
- [6]. Algérie presse service. (2019, 10 23). Cancer du sein : le dépistage précoce de nouveau préconisé. from Algérie presse service: <http://www.aps.dz/sante-science-technologie/96199-cancer-du-sein-le-depistage-precoce-de-nouveau-preconise>
- [7]. Bremme, L. (2015, 08 25). Définition : Qu'est-ce que le Big Data. from lebigdata: <https://www.lebigdata.fr/definition-big-data>
- [8]. Cayla, B. (2018, 09 07). La star des algorithmes de ML : XGBoost. Retrieved 05 18, 2020, from datacorner: <https://www.datacorner.fr/xgboost/>
- [9]. Centre international de recherche sur le cancer. (2018, 09 18). Dernières données mondiales sur le cancer. Retrieved from iarc: https://www.iarc.fr/wp-content/uploads/2018/09/pr263_F.pdf
- [10]. Claw, S. (2020, 06 23). Python. Retrieved from TkInter: <https://wiki.python.org/moin/TkInter>
- [11]. Clayton, R. (2019). Qu'est-ce que le Machine Learning ? Retrieved Mars 08, 2020, from Oracle Algeria:
- [12]. <https://www.oracle.com/dz/artificial-intelligence/what-is-machine-learning.html>
- [13]. Courbiere, B., & Carcopino, X. (2016). KB Gynécologie obstétrique. Paris: vernazobres-grego.
- [14]. Dave, A. (2020, 03 07). Regression in Machine Learning. Retrieved , from Medium:
- [15]. <https://medium.com/datadriveninvestor/regression-in-machine-learning-296caae933ec>
- [16]. Deginé, P. (1855). Chirurgie. Paris: V. Masson.
- [17]. developers.google. (2020, 02 10). Classification : ROC et AUC. Retrieved from

- developers.google: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=fr>
- [18]. Dua, D., & Graff, C. (2017). Breast Cancer Wisconsin (Diagnostic) Data Set. Retrieved, from UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+%28diagnostic%29>
- [19]. dupré, x. (2020, 04 05). La classification. Retrieved from Xavierdupre: http://www.xavierdupre.fr/app/mlstatpy/helpsphinx/c_ml/rn/rn_3_clas.html
- [20]. Fathy, W., & Ghoneim, A. (2019). A Deep Learning Approach for Breast Cancer Mass Detection.
- [21]. International Journal of Advanced Computer Science and Applications, 10(1).
- [22]. Géron, A. (2017). Hands-On_Machine_Learning_with_Scikit-learn and tensorflow. USA: O'Reilly Media.
- [23]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press .
- [24]. Grossfeld, B. (2020, 01 23). Deep learning vs machine learning: a simple way to understand the difference. Retrieved from zendesk blog: <https://www.zendesk.com/blog/machine-learning-and-deeplearning/>
- [25]. Grossfeld, B. (2020, 01 23). Le programme du Master en intelligence artificielle. Retrieved, from Unidistance: https://unidistance.ch/intelligence-artificielle/master/modules/?gclid=Cj0KCQjwj7v0BRDOARIsAGh37io1ERDQK4WbBncrriozBfS6GtD7utp2VGD2nvFABJQDjxZBHPnrWJMaAh0XEALw_wcB
- [26]. Gupta, N. (2019, 10 29). Why is Python Used for Machine Learning? Retrieved 04 29, 2020, from Hacker
- [27]. Noon: <https://hackernoon.com/why-python-used-for-machine-learning- u13f922ug>
- [28]. Ibrahim, A. A.-H., Hashad, I. A., Nigm El-Deen , S., & Maher, A. (2015). Robust Breast Cancer Diagnosis on Four Different Datasets Using Multi-Classifiers Fusion. International Journal of Engineering Research and, 4(3).
- [29]. Institut National Du Cancer. (2018). Anatomie du sein. Retrieved 02 13, 2020, from INSTITUT NATIONAL DU CANCER: <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Anatomie-du-sein>
- [30]. Kelley, S. (2020, 01 15). Qu'est-ce que le cancer du sein? Retrieved 02 13, 2020, from Société canadienne du cancer: <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/breast/breast-cancer/?region=qc>
- [31]. Lahmer, M. (2018). Tumeurs du sein. Faculté de médecine. Constantine: université de

- [32]. Passeport-sante. (2019). Les traitements médicaux du cancer du sein. Retrieved from [passeportsante:
https://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer-sein-pm-traitements-medicaux-du-cancer-du-sein](https://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer-sein-pm-traitements-medicaux-du-cancer-du-sein)
- [33]. Patterson, J., & Gibson, A. (2017). Deep Learning : A Practitioner's Approach.
- [34]. Beijing: O'Reilly Media. Prevost, L. (2019). Modèles Connexionnistes Apprentissage
- [35]. Rocca, J. (2020, 03 28). Understanding Variational Autoencoders (VAEs). Retrieved 07 04, 2020, from Medium: <https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73>
- [36]. Rosenblatt, F. (1958). THE PERCEPTRON: A PROBABILISTIC MODEL FOR. Psychological Review, 65.
- [37]. Rtbtf Tendence. (2020, 01 04). Diagnostic du cancer du sein: l'intelligence artificielle peut faire mieux que l'humain. Retrieved, from Rtbtf Tendence: https://www.rtbtf.be/tendance/bien-etre/sante/detail_diagnostic-du-cancer-du-sein-l-intelligence-artificielle-peut-faire-mieux-que-l-humain?id=10399031
- [38]. Sahu, B., Mohanty, S., & Rout, S. K. (2018, 07 13). A Hybrid Approach for Breast Cancer Classification and Diagnosis. ICST Transactions on Scalable Information Systems.
- [39]. Schmidhuber, J. (2015). Deep Learning. Scholarpedia, 10, 32832.
- [40]. Sun, D., Wang, M., & Li, A. (2018). A Multimodal Deep Neural Network for Human
- [41]. Breast Cancer Prognosis Prediction by Integrating Multi-Dimensional Data. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 16(3), 841850.
- [42]. Sutton, R. S. (1988). Learning to predict by the method of temporal differences. Machine Learning.
- [43]. The Pallets Projects. (2010). Flask. Retrieved, from Flask palletsprojects: <https://flask.palletsprojects.com/en/1.1.x/>
- [44]. Vázquez, F. (2018, 04 12). A “weird” Introduction to Deep Learning. Retrieved from [bbva data&analytics: https://www.bbva.com/a-weird-introduction-to-deep-learning/](https://www.bbva.com/a-weird-introduction-to-deep-learning/)
- [45]. W. Ruddon, R. (2007). CANCER BIOLOGY (4 ed.). Ann Arbor, Michigan, Usa: Oxford University Press.
- [46]. Watkins, C. J., & Dayan, P. (1992). Q-learning. Machine Learning.

- [47]. Werfelli, O. (2015). présentation générale des réseaux de neurones artificiels. Cours. Retrieved 04 05, 2020, from <https://fr.slideshare.net/OussamaWerfelli/rseaux-de-neurones-artificiels>
- [48]. WHO. (2018, 09 15). Cancer. Retrieved, from Organisation mondiale de la santé: <https://www.who.int/fr/news-room/fact-sheets/detail/cancer>
- [49]. Xiao, Y., Wu, J., Lin, Z., & Zhao, X. (2018). Breast Cancer Diagnosis Using an Unsupervised Feature. the 37th Chinese Control Conference, (p. 6). Wuhan, China. Retrieved 03 21, 2020
- [50]. Yarbrow, H. C., Wujcik, D., & Gobel, B. H. (2011). Cancer nursing: principles and practice (éd. 7).
- [51]. Burlington, Massachusetts, États-Unis: Jones & Bartlett Learning.
- [52]. Yum, J. (2019, 05 19). What is the deep neural network known as “ResNet-50”? Retrieved, from quora: <https://www.quora.com/What-is-the-deep-neural-network-known-as-ResNet-50>
- [53]. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief. 2020 Feb;28:104863. DOI: 10.1016/j.dib.2019.104863.

TABLE DES MATIERES

DEDICACE.....	II
RÉSUMÉ.....	IV
ABSTRACT	V
SOMMAIRE.....	VI
LISTE DES ABREVIATIONS	VII
INDEX DES FIGURES	VIII
INDEX DES TABLEAUX.....	X
INTRODUCTION GENERALE.....	11
1. Contexte général de la recherche.....	11
2. Problématique de la recherche.....	11
2.1. Présentation du problème	11
2.2. Formulation du problème (questions de recherche).....	11
3. Hypothèses de recherche	12
4. Objectifs de recherche	12
5. Justification de la recherche	12
6. Délimitation de la recherche.....	13
CHAPITRE 1 : CADRE THEORIQUE ET CONCEPTUEL	3
1. Introduction	3
1. Le cancer du sein	3
1.1. Qu'est-ce que le cancer du sein ?	4
1.2. Types de cancer de sein.....	4
1.3. Anatomie de Sein	5
1.4. Les symptômes et facteurs de risques de cancer du sein.....	6
1.5. Les traitements médicaux du cancer du sein	7
2. Intelligence Artificielle.....	9

2.1. L'apprentissage automatique « Machine Learning »	9
2.2. De l'apprentissage automatique à l'apprentissage en profondeur.....	19
2.3. L'apprentissage profond « Deep Learning ».....	20
3. Etat de l'art	36
3.1. Une approche d'apprentissage approfondi pour la détection de masse du cancer du sein (Fathy & Ghoneim, 2019) :	36
3.2. Un réseau neuronal profond mul modal pour la prévision du cancer du sein humain en intégrant des données mul dimensionnelle (Sun, Wang, & Li, 2018).....	38
3.3. Diagnostic du cancer du sein à l'aide d'un algorithme d'extraction de fonctionnalités non supervisé basé sur l'apprentissage en profondeur.....	40
Conclusion	43
CHAPITRE 2 :	44
METHODOLOGIE DE RECHERCHE	44
Introduction	44
1. Nature de la recherche	44
1.1 Présentation des méthodes de recherche	44
1.2 Justification de la nature de la recherche	45
2. Variables de recherche.....	46
2.1 Définition conceptuelle	46
2.2 Limites et Difficultés.....	46
2.3 Utilisation des variables	47
3. Outils de recherche	48
3.1. Présentation d'outils de recherche.....	48
3.2 Justification du choix des outils	49
Conclusion.....	49
CHAPITRE 3 :	50
PRESENTATION DE LA SITUATION (PRESENTATION DES DONNEES COLLECTEES / DES RESULTATS)	50

Introduction	50
1. Présentation du site de recherche.....	50
2. Données	52
3. Présentation des résultats.....	54
Conclusion.....	55
CHAPITRE 4 :	56
ANALYSE - DIAGNOSTIC DE LA SITUATION ET PROPOSITION D'INTERVENTION	56
Introduction	56
1. Présentation et analyse de la situation	56
1.1 Analyse des résultats du questionnaire.....	56
1.2. Synthèse de l'analyse documentaire	68
2. Intervention proposée et justification	70
2.1 Description de la solution :.....	70
2.2 Objectifs de l'intervention	71
3. Composantes de l'intervention envisagée	72
4. Stratégies d'action	73
4.1. Analyse des Besoins:.....	73
4.2. Recueil des exigences fonctionnelles et non fonctionnelles	73
4.3. Sélection de la technologie.....	74
4.4. Développement du Chatbot:.....	77
4.5. Développement du module diagnostique	79
4.6. Interface utilisateur (UI):.....	88
4.7. Sécurité:.....	90
4.8. Déploiement:	91
4.9. Lancement:	91
4.10. Maintenance et mises à jour:	91

5. Faisabilité	91
5.1. Faisabilité Économique:	91
5.2. Faisabilité Sociale:	92
5.3. Faisabilité Technique	93
5.4. Faisabilité Environnementale	93
Conclusion	94
CONCLUSION GENERALE	95
REFERENCES BIBLIOGRAPHIQUES	96