

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de L'Enseignement Supérieur et de la Recherche
Scientifique
Université 20 août 1955 - Skikda
Faculté des Sciences
Département d'Informatique



MEMOIRE

En vue de l'obtention du diplôme de
Master en Informatique
Spécialité : Systèmes Informatiques

Thème

**Vision-Transformers pour la classification des images
histopathologiques du cancer du sein**

Présenté Par :

Ziat Anfal

Boussireb Abir

Encadré Par :

Dr. Hazmoune Samira

Année universitaire 2022/2023

Remerciements

“

Tout d’abord, Alhamdulillah, nous remercions Allah SWT pour Sa grandeur et pour nous avoir donné l’opportunité, la force et le courage de réaliser ce travail. Par la suite, Nous tenons à remercier spécialement notre encadreur, Dr. Samira Hazmoune, qui nous a honorés en acceptant notre encadrement et pour les efforts, les directives et les conseils qu’elle nous a donnés pour rédiger le mémoire de fin d’études. Notre gratitude cordiale est infinie envers les professeurs du département d’informatique qui ont contribué à notre formation au cours des cinq dernières années d’études. Grâce à eux, nous avons construit le premier chemin vers l’avenir à partir de leurs connaissances et de leur expérience scientifique. Nous remercions également les membres du jury qui ont accepté d’évaluer notre travail, et nous attendons avec impatience leurs conseils supplémentaires. Enfin, nos sincères remerciements seraient incomplets sans attribuer une partie à nos familles et à nos amis solidaires. Un merci spécial à nos parents, qui nous ont soutenus émotionnellement et nous ont aidés spirituellement tout au long de notre vie.

”

Dédicace

“

A mes chers parents

Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consentis pour mon instruction et mon bien être. Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours.

A mes chers frères et soeurs

Je vous remercie pour votre soutien, vous m'avez remonté le moral lorsque ma détermination flanchait. J'espère que vous appréciez le fruit de tant d'effort.

A mon binôme

Je voudrais également le dédier à mon pair pour m'avoir aidé à terminer ce travail.

Et à tous mes amis

Qui ont toujours été là quand j'avais besoin d'eux.

”

- ***Anfal***

“ **À mes chers parents**

Je dédie ce mémoire à mes chers parents qui ont été toujours à mes côtés et m'ont toujours soutenu tout au long de ces longues années d'études. En signe de reconnaissance, qu'ils trouvent ici, l'expression de ma profonde gratitude pour tout ce qu'ils ont consenti d'efforts et de moyens pour me voir réussir dans mes études.

À toute ma famille

Aucune expression ne saurait exprimer mon respect et ma considération pour votre soutien et encouragements. Je vous dédie ce travail en reconnaissance de l'amour que vous m'offrez quotidiennement et votre bonté exceptionnelle. Que Dieu le Tout Puissant vous garde et vous procure santé et bonheur.

À mon binôme

Je souhaite exprimer ma sincère reconnaissance envers mon binôme pour son assistance inestimable tout au long de ce travail.

”

- **Abir**

Résumé

Ce mémoire s'inscrit dans le cadre de la classification des images histopathologiques du cancer du sein, qui revêt une importance cruciale dans le domaine médical. Un diagnostic tardif du cancer du sein peut avoir des conséquences néfastes. Afin d'améliorer le processus de diagnostic, de nombreux chercheurs ont exploité les avantages de l'intelligence artificielle, notamment dans le domaine médical.

Les Transformers, en tant qu'architecture spécifique de réseau de neurones utilisée dans le deep learning, se sont révélés particulièrement efficaces pour les tâches impliquant des séquences de données, telles que le traitement du langage naturel (NLP). En raison du succès massif des Transformers dans le domaine du NLP, ils sont adaptés aux tâches de vision par ordinateur. Les Transformers les plus récents utilisés pour la classification d'images comprennent Vision Transformers (ViT), qui est une approche révolutionnaire dans le domaine de classification des images initialement proposée en 2020. Le ViT a démontré sa capacité à capturer les relations à long terme et les dépendances entre les différentes parties d'une image, ce qui est essentiel pour une classification précise.

Dans ce mémoire, nous présentons un modèle de transformer basé sur ViT pour la classification des images histopathologiques du cancer du sein. Afin d'améliorer l'accuracy du modèle, une série d'expérimentations sur l'ensemble de données BreakHis a été réalisée pour déterminer les hyper paramètres optimaux de l'architecture ViT. L'accuracy de la classification pour BreakHis était de 97,78%. Les résultats comparatifs avec différentes architectures CNN (Convolutional Neural Networks) et quelques travaux précédents sur la même base de données montrent la supériorité de ViT sur tous les techniques de machine learning et de deep learning classique.

Mots clés : Vision Transformer (ViT), Cancer du sein, classification des images histopathologiques.

Abstract

This master's thesis is part of the classification of histopathological images of breast cancer, which is of crucial importance in the medical field. Late diagnosis of breast cancer can have detrimental consequences. In order to improve the diagnostic process, numerous researchers have leveraged the benefits of artificial intelligence, particularly in the medical domain.

Transformers, as a specific neural network architecture used in deep learning, have proven to be particularly effective for tasks involving sequence data, such as natural language processing (NLP). Due to the massive success of Transformers in the field of NLP, they have been adapted for computer vision tasks. The most recent Transformers used for image classification include Vision Transformers (ViT), which is a groundbreaking approach in the field of image classification initially proposed in 2020. ViT has demonstrated its ability to capture long-term relationships and dependencies among different parts of an image, which is crucial for accurate classification.

In this master's thesis, we present a ViT-based transformer model for the classification of histopathological images of breast cancer. In order to improve the accuracy of the model, a series of experiments on the BreakHis dataset were conducted to determine the optimal hyperparameters of the ViT architecture. The classification accuracy for BreakHis was 97.78%. Comparative results with different Convolutional Neural Network (CNN) architectures and some previous works on the same database demonstrate the superiority of ViT over all classical machine learning and deep learning techniques.

Keywords : Vision Transformer (ViT), Breast cancer, histopathological image classification.

ملخص

هذه المذكرة هي جزء من تصنيف الصور النسيجية المرضية لسرطان الثدي ، والتي لها أهمية حاسمة في المجال الطبي. قد يكون للتشخيص المتأخر لسرطان الثدي عواقب وخيمة. من أجل تحسين عملية التشخيص ، استغل العديد من الباحثين مزايا الذكاء الاصطناعي ، خاصة في المجال الطبي.

أثبتت المحولات ، باعتبارها بنية شبكة عصبية محددة مستخدمة في التعلم العميق ، فعاليتها بشكل خاص في المهام التي تتضمن تسلسل البيانات ، مثل معالجة اللغة الطبيعية (NLP) نظراً للنجاح الهائل الذي حققته المحولات في مجال البرمجة اللغوية العصبية ، فهي مناسبة لمهام رؤية الكمبيوتر. تشمل أحدث المحولات المستخدمة لتصنيف الصور محول الرؤية (ViT) ، وهو نهج ثوري في مجال تصنيف الصور الذي تم اقتراحه لأول مرة في عام 2020. وقد أثبتت ViT قدرتها على التقاط العلاقات طويلة الأجل والتبعيات بين أجزاء مختلفة من الصورة ، وهو أمر ضروري لتصنيف دقيق.

في هذه المذكرة ، نقدم نموذج محول قائم على ViT لتصنيف الصور النسيجية المرضية لسرطان الثدي. من أجل تحسين دقة النموذج ، أجريت سلسلة من التجارب على مجموعة بيانات BreakHis لتحديد الإعدادات الفائقة المثلى لبنية ViT. كانت دقة تصنيف BreakHis 97.78%. تظهر النتائج المقارنة مع بنىات مختلفة للشبكة العصبية التلافيفية (CNN) وبعض الأعمال السابقة على نفس قاعدة البيانات تفوق ViT على جميع تقنيات التعلم الآلي الكلاسيكية والتعلم العميق.

كلمات مفتاحية : محول الرؤية (ViT) ، سرطان الثدي ، تصنيف صور الأنسجة المرضية.

Table des matières

Remerciements	I
Dédicace	II
Résumé	IV
Abstract	V
VI	ملخص
Introduction générale	1
1 Cancer du sein	3
1.1 Introduction	4
1.2 Cancer du sein	4
1.2.1 Statistiques générales	4
1.2.2 Statistiques du cancer du sein en Algérie	5
1.3 Les tumeurs du mammaires	6
1.3.1 Tumeurs bénignes	6
1.3.2 Tumeurs malignes	8
1.3.3 Classifications des cancers du sein	9
1.3.3.1 Classification TNM	9
1.3.3.2 Classification moléculaire	10
1.4 Dépistages et diagnostic du cancer du sein	11
1.4.1 Dépistages du cancer du sein	12
1.4.2 Diagnostic du cancer du sein	12
1.4.2.1 L'examen clinique	12
1.4.2.2 L'imagerie mammaire	12
1.4.2.3 Anatomopathologie	13
1.5 Imagerie mammaire	13
1.5.1 La mammographie	13
1.5.2 Échographie	13
1.5.3 IRM	14
1.5.4 L'imagerie microscopique	16
1.5.4.1 L'histopathologie	16
1.5.4.2 La biopsie	16
1.6 Classification BI RADS des anomalies	16
1.7 Conclusion	17

2	Intelligence artificielle pour le diagnostic du cancer du sein	18
2.1	Introduction	19
2.2	Quelques logiciels de l'IA pour le diagnostic des maladies cancéreuses	19
2.2.1	PathAI	19
2.2.2	Freenome	20
2.2.3	Proscia	20
2.3	Apprentissage automatique (Machine Learning)	20
2.3.1	Les différents types d'apprentissage	21
2.3.1.1	Apprentissage supervisé	21
2.3.1.2	Apprentissage non supervisé	21
2.3.1.3	Apprentissage par renforcement	21
2.3.2	Méthodes d'apprentissage automatique	22
2.3.2.1	Machines à vecteurs de support	22
2.3.2.2	K plus proche voisins	23
2.3.2.3	Arbres de décision	23
2.3.2.4	Modèles de Markov cachés	24
2.3.2.5	Réseaux de neurones artificiels (ANN)	24
2.4	Apprentissage profond (Deep Learning)	25
2.4.1	Les réseaux de neurones récurrents (RNN)	26
2.4.2	Les réseaux de neurones convolutifs (CNN)	27
2.4.2.1	Architecture globale	28
2.4.2.2	Quelques architectures des réseaux de neurones convolutifs	29
2.4.3	Les Transformers	30
2.4.3.1	Transformers pour NLP	33
2.4.3.2	Transformers pour la vision	34
2.5	Vision transformers (ViT)	35
2.5.1	Structure de Vision Transformers (ViT)	35
2.5.2	Quelques applications de Vision Transformers	37
2.6	Etat de l'art sur le diagnostic automatique du cancer du sein	38
2.6.1	Bases de données	38
2.6.1.1	Base de données (FONDÓN et al. 2018)	38
2.6.1.2	BreakHis	39
2.6.1.3	Digital Database for Screening Mammography DDSM	39
2.6.1.4	Base de données (ZABIT et al. 2020)	39
2.6.1.5	ISBI 2019 data set	40
2.6.2	Quelques travaux récents	40
2.6.2.1	FONDÓN et al. 2018	40
2.6.2.2	BARDOU et al. 2018	40
2.6.2.3	S. KHAN et al. 2019	41
2.6.2.4	Li et al. 2019	41
2.6.2.5	ALTAN 2020	41
2.6.2.6	GOUR et al. 2020	42
2.6.2.7	ZABIT et al. 2020	42
2.6.2.8	KOSARE et al. 2021	42
2.6.2.9	JIANG et al. 2021	43
2.6.2.10	ALKHATHLAN et al. 2022	43

2.6.2.11	MAYOUF et al. 2022	44
2.6.2.12	NAKACH et al. 2022	44
2.6.2.13	TUMMALA et al. 2022	44
2.6.2.14	ALOTAIBI et al. 2023	45
2.7	Conclusion	45
3	Un système basé-ViT pour la classification des images histopathologiques du cancer du sein	46
3.1	Introduction	47
3.2	Description du système de classification	47
3.2.1	Architecture générale du système	47
3.2.2	Présentation détaillée du système	48
3.2.2.1	Phase d'apprentissage	48
3.2.2.2	Phase de test	52
3.3	Expérimentation résultats et discussion	53
3.3.1	Base de données BreakHis	53
3.3.2	Ajustement des hyperparamètres	53
3.3.2.1	Taille de batch	53
3.3.2.2	Taux du Dropout	54
3.3.2.3	Le pas d'apprentissage	55
3.3.2.4	Le nombre d'époques d'apprentissage	55
3.3.3	Evaluation des performances	56
3.3.4	Comparaison des résultats	57
3.3.4.1	Avec des architectures CNN	57
3.3.4.2	Avec des travaux de la littérature	58
3.4	Conclusion	59
	Conclusion et perspectives	60
	Webographie	67
	Annexe	68
A	Outils d'implémentation	69

Table des figures

1.1	Les taux d'incidence du cancer du sein dans le monde (Web 01).	5
1.2	Cas du cancer en 2017 en Algérie selon le sexe (AKKOUCHE 2019).	6
1.3	Les cancers féminins en Algérie 2017 (AKKOUCHE 2019).	6
1.4	Image au microscope de l'ensemble de données local montrant une tumeur d'adénofibrome vu dans le facteur de grossissement 200X (Web 04).	7
1.5	Image au microscope de l'ensemble de données local montrant une tumeur Phylode vue en facteur de grossissement 200X (Web 04).	7
1.6	Image au microscope de l'ensemble de données local montrant une tumeur adénoïde vue en facteur de grossissement 200X (Web 04).	7
1.7	Image au microscope de l'ensemble de données local montrant l'adénome tubulaire vu dans le facteur de grossissement 200X (Web 04).	8
1.8	Image au microscope de l'ensemble de données local montrant un carcinome ductal invasif vu dans un facteur de grossissement de 200X (Web 04).	8
1.9	Image au microscope de l'ensemble de données local montrant le carcinome lobulaire vu dans le facteur de grossissement 200X (Web 04).	8
1.10	Image au microscope de l'ensemble de données local montrant le carcinome mucoïde vu dans le facteur de grossissement 200X (Web 04).	9
1.11	Image au microscope de l'ensemble de données local montrant le carcinome papillaire vu dans le facteur de grossissement 200X (Web 04).	9
2.1	Logo PathAI (Web 06).	19
2.2	Logo Freenome (Web 07).	20
2.3	Logo Proscia (Web 08).	20
2.4	Apprentissage par renforcement (MAZAC 2015).	22
2.5	Machine à vecteurs de support (KUMAR 2021).	23
2.6	Exemple d'illustration d'un arbre de décision (ZOUNGRANA 2020).	24
2.7	Structure générale du réseau de neurones artificiels avec deux couches cachées (MOGHADDAMNIA et al. 2009).	25
2.8	Cellule Long Short-Term Memory (LSTM)(SUEIRAS 2021).	26
2.9	Le pipeline de l'architecture générale CNN (GUO et al. 2016).	27
2.10	L'opération de convolution (GUO et al. 2016).	28
2.11	L'opération de Max pooling (GUO et al. 2016).	29
2.12	Couche Fully-Connected (GUO et al. 2016).	29
2.13	Schéma représentant les différentes étapes de scaled dotproduct attention et de multi-head attention (VASWANI et al. 2017).	31
2.14	La structure globale de l'architecture Transformer d'origine (VASWANI et al. 2017).	32
2.15	Architecture du modèle BERT (ACHEAMPONG et al. 2021).	34

2.16	Aperçu du modèle vision transformer (ViT) (DOSOVITSKIY et al. 2020).	35
3.1	Représentation schématique de l'architecture de notre système de classification.	48
3.2	Image divisées en 256 patches de taille 16x16.	50
3.3	L'architecture de l'encodeur du transformer.	51
3.4	L'architecture globale de ViT.	52
3.5	Courbes accuracy et loss obtenus.	56
3.6	La Matrice de confusion.	57
3.7	Graphique à barres illustrant la comparaison de l'accuracy sur la base de données BreakHis.	58
A.1	Bibliothèques utilisées.	72
A.2	Les opérations pour prétraitement des données.	72
A.3	La division des données.	72
A.4	Chargement d'un modèle pré-entraîné ViT.	73
A.5	Fonction d'entraînement.	73
A.6	Fonction d'accuracy.	73

Liste des tableaux

1.1	Classification des stades des cancers du sein (VERBEKE 2010).	10
1.2	Classification moléculaire des cancers du sein (PEROU et al. 2000).	11
1.3	Avantages et inconvénients des différentes techniques médicales de dépistage du cancer du sein (MEBARKI 2020).	15
1.4	Classification BI-RADS (Breast Imaging Reporting And Data System) (Web 05).	17
3.1	Résultats expérimentaux pour différentes tailles de batch.	54
3.2	Résultats expérimentaux pour déterminer le taux de Dropout.	54
3.3	Résultats expérimentaux pour déterminer le pas d'apprentissage.	55
3.4	Accuracy de notre modèle sur la base de données BreakHis.	56
3.5	Comparaison de notre ViT avec différentes architectures CNN.	57
3.6	Comparaison des performances avec l'état de l'art.	58
A.1	Caractéristiques de l'environnement.	69

Liste des abréviations

ADN	<i>Acide Désoxyribo Nucléique</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BI-RADS	<i>Breast Imaging Reporting And Data System</i>
CNN	<i>Convolutional Neural Network</i>
DeiT	<i>Data-efficient image transformers</i>
GPT	<i>Generative Pre-trained Transforme</i>
HMM	<i>Hidden Markov Models</i>
IA	<i>Intelligence artificielle</i>
IRM	<i>Imagerie par résonance magnétique</i>
ISBI	<i>International Symposium on Biomedical Imaging</i>
KNN	<i>K Nearest Neighbors</i>
TNM	<i>Tumeur Node Métastas</i>
SVM	<i>Machine à Vecteur de Support</i>
LSTM	<i>Long short-term memory</i>
MLP	<i>Multi-layer Perceptron</i>
MHA	<i>Multi-Head Attention</i>
MLM	<i>Masked Language Modeling</i>
NLP	<i>Natural Language Processing</i>

Liste des tableaux

NSP	<i>Next Sentence Prediction</i>
ReLU	<i>Rectified Linear Unit</i>
RNN	<i>Recurrent Neural Networks</i>
VGG	<i>Visual Geometry Group</i>
ViT	<i>Vision transformers</i>

Introduction générale

Contexte

Le cancer du sein est l'une des formes les plus courantes de cancer chez les femmes à travers le monde. Il représente un défi majeur en matière de santé publique et de soins médicaux. Le dépistage et le diagnostic précoces jouent un rôle essentiel dans la lutte contre cette maladie, car ils permettent une détection précoce et une intervention médicale appropriée.

Dans cette optique, les systèmes de diagnostic basés sur l'intelligence artificielle (IA) ont émergé comme des outils prometteurs pour améliorer la précision et l'efficacité du dépistage et du diagnostic du cancer du sein. L'IA, combinée à des techniques d'apprentissage automatique et d'analyse d'images médicales, permet d'extraire des informations précieuses à partir de mammographies, d'échographies et d'autres modalités d'imagerie utilisées dans le diagnostic du cancer du sein.

Depuis leur arrivée en 2017, les transformers ont montré des performances remarquables dans une variété de tâches de traitement du langage naturel (NLP), mais leur application aux tâches de vision par ordinateur est relativement récente. Le Vision Transformer (ViT), proposé en 2020, est un modèle basé sur les transformers spécifiquement conçu pour les tâches de classification d'images. Contrairement aux réseaux de neurones convolutionnels (CNN) traditionnels, ViT n'utilise pas de couches de convolution, mais traite plutôt les images sous forme de séquences de patches. Cela permet au modèle de s'adapter facilement à des tailles d'image plus grandes sans le coût computationnel des convolutions. ViT utilise également des mécanismes d'attention multi-têtes (multi-head attention) pour se concentrer sur les patches importants de l'image et apprendre des représentations d'image efficaces. ViT a atteint des performances de pointe sur plusieurs benchmarks de classification d'images, démontrant l'efficacité des modèles basés sur les transformers dans le domaine de la vision par ordinateur.

Problématique et objectifs

L'objectif principal de ce travail est de concevoir un modèle basé sur ViT, qui vise à améliorer les performances des systèmes de classification des images histopathologiques du cancer du sein pour la base de données BreakHis. Cet objectif général peut être détaillé

dans les sous-objectifs spécifiques suivants :

- Concevoir un modèle ViT pour la classification des images histopathologiques du cancer du sein en deux types : bénigne ou maligne.
- Une étude expérimentale pour déterminer les hyper paramètres optimaux, ainsi que l'effet important des hyper paramètres sur l'accuracy du modèle.
- Une étude comparative afin de démontrer l'utilité des algorithmes de ViT dans le développement de systèmes efficaces de classification des images histopathologiques du cancer du sein.

Plan de mémoire

Ce mémoire est constitué d'une introduction générale, de trois chapitres et d'une conclusion générale :

- Chapitre 1 : Il présente des généralités sur la maladie du cancer du sein. En effet, dans ce chapitre, nous donnons la définition de la maladie, les types des tumeurs mammaires, le dépistage et le diagnostic, ainsi que les outils radiologiques permettant leur détection.
- Chapitre 2 : Il donne un aperçu général des méthodes de l'apprentissage automatique ainsi que celles de l'apprentissage profond. Nous y présentons en particulier, le ViT qui est la technique utilisée dans cette étude. Nous y dressons, enfin l'état de l'art dans le domaine de diagnostic du cancer du sein.
- Chapitre 3 : Il regroupe la description de notre système, ainsi que les différentes expérimentations réalisées et l'analyse des résultats obtenus. Nous présentons, dans un premier temps, l'architecture de notre système de classification des images histopathologiques du cancer du sein. Puis, nous réalisons une série d'expérimentations afin de régler les hyper paramètres de notre modèle et d'évaluer ses performances. Enfin, nous comparons les performances du modèle conçu avec celles des architectures CNN et de certains travaux de la littérature.

Chapitre 1

Cancer du sein

1.1 Introduction

Le cancer du sein est le type de cancer le plus fréquent chez les femmes, que ce soit dans les pays développés ou en développement. Dans les pays en développement, l'incidence du cancer du sein augmente en raison de l'augmentation de l'espérance de vie, de l'urbanisation et de l'adoption de modes de vie occidentaux. Même si la prévention peut réduire les risques, elle ne permet pas d'éliminer la majorité des cancers du sein.

Dans ce chapitre nous allons discuter du cancer du sein, les types des tumeurs mammaires (bénignes, malignes), dépistage et diagnostic du cancer du sein ainsi que les outils radiologiques permettant leur détection, tels que : La mammographie, l'IRM, l'échographie et l'imagerie microscopique.

1.2 Cancer du sein

Cancer du sein n'est pas une maladie contagieuse. Contrairement à certains cancers causés par des infections, l'incidence du cancer du sein varie selon les populations. Les taux varient entre les femmes et les hommes, ainsi qu'entre les personnes de différentes races et âges. Ils varient également dans le monde entier (Web 01). Environ la moitié des cancers du sein surviennent chez des femmes qui n'ont pas d'autres facteurs de risque identifiables que le sexe (féminin) et l'âge (plus de 40 ans) (Web 02). Le cancer du sein prend naissance dans les cellules mammaires. Une tumeur cancéreuse (maligne) est un groupe de cellules cancéreuses qui peuvent envahir les tissus voisins et les détruire. Il peut également se propager (métastaser) à d'autres parties du corps.

Les cellules mammaires changent parfois, provoquant leur croissance ou leur comportement anormal. Ces changements peuvent entraîner des affections mammaires non cancéreuses (bénignes) telles que l'hyperplasie atypique et les kystes. Ils peuvent également conduire à la formation de tumeurs non cancéreuses, notamment des papillomes intra canauxaires (Web 03).

1.2.1 Statistiques générales

Le cancer du sein représente le type de cancer le plus courant chez les femmes dans le monde. Selon les données les plus récentes disponibles pour l'année 2020, on estime qu'il y a eu plus de 2 millions de nouveaux cas de cancer du sein chez les femmes à l'échelle mondiale.

Les taux de cancer du sein varient en fonction des diverses catégories de population, tels que le genre, l'ethnie, l'âge, ainsi que la localisation géographique, que ce soit aux États-Unis ou dans d'autres pays du monde. Ces différences de taux peuvent être observées tant chez les femmes que chez les hommes.

Voici quelques statistiques pour les femmes atteintes de cancer du sein recueillies aux États-Unis en 2023 :

- 297.790 nouveaux cas de cancer du sein invasif (Cela inclut les nouveaux cas de

cancer du sein primitif, mais pas les récurrences du cancer du sein).

- 55,720 nouveaux cas de carcinome ductal in situ (DCIS), un cancer du sein non invasif.
- 43 170 décès par cancer du sein.

Voici quelques statistiques pour les hommes atteints de cancer du sein recueillies aux États-Unis en 2023 :

- 2 800 nouveaux cas de cancer du sein invasif (Cela inclut les nouveaux cas de cancer du sein primitif, mais pas les récurrences du cancer du sein.)
- 530 décès par cancer du sein.

De manière générale, les taux de cancer du sein sont plus élevés dans les pays développés tels que les États-Unis, l'Angleterre et l'Australie, que dans les pays en développement tels que le Cambodge, le Népal et le Rwanda (Web 01) :

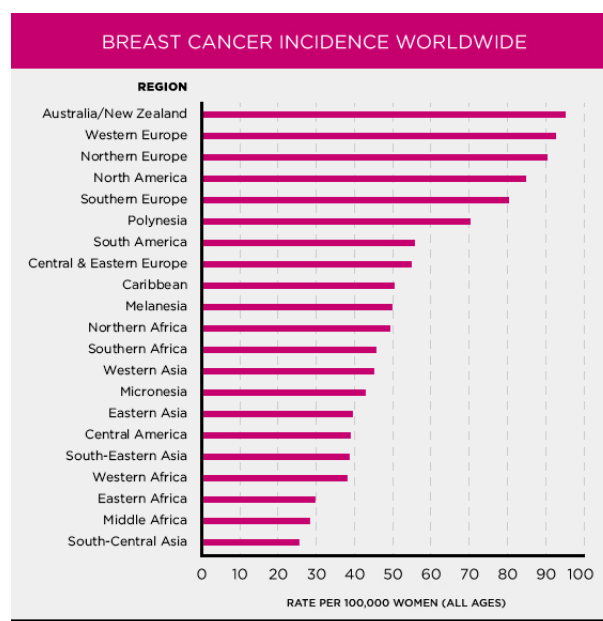


FIG. 1.1 : Les taux d'incidence du cancer du sein dans le monde (Web 01).

1.2.2 Statistiques du cancer du sein en Algérie

D'après les registres nationaux du cancer pour l'année 2017, il y a eu environ 44 000 nouveaux cas de cancer, dont 25 000 concernent des femmes et plus de 54% de ces cas sont des cancers du sein. Près de 13 625 femmes ont été diagnostiquées avec un cancer du sein, suivi du cancer colorectal, de la thyroïde et du col de l'utérus, (Voir figure 1.2 et figure 1.3).

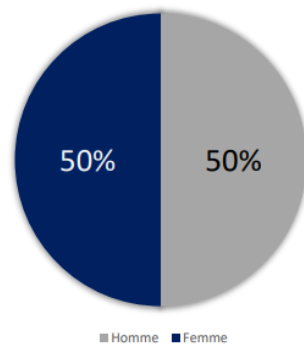


FIG. 1.2 : Cas du cancer en 2017 en Algérie selon le sexe (AKKOUICHE 2019).

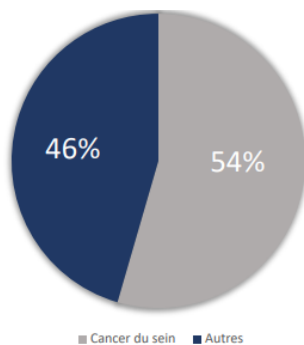


FIG. 1.3 : Les cancers féminins en Algérie 2017 (AKKOUICHE 2019).

1.3 Les tumeurs du mammaires

Les tumeurs du sein peuvent être dites bénignes, ce qui signifie qu'elles sont bien circonscrites et bien encapsulées. Ils se développent ensuite lentement localement et ne récidivent pas après excision. La tumeur bénigne du sein la plus fréquente est l'adénofibrome, le plus souvent retrouvé avant l'âge de 30 ans.

La deuxième catégorie de tumeurs mammaires, qualitativement malignes, sont au contraire mal circonscrites et non encapsulées. Ce sont des cellules à croissance rapide qui ont tendance à s'infiltrer et à détruire les tissus adjacents. La grande majorité (plus de 95% des cas) des tumeurs malignes sont des adénocarcinomes, c'est-à-dire dues à une cancérisation de l'épithélium glandulaire à partir des cellules canalaire, on dit alors carcinome canalaire, ou cellules lobaires, on dit alors Carcinome lobulaire. Les adénocarcinomes peuvent être subdivisés en fonction de leurs caractéristiques infiltration des tissus environnants (VERBEKE 2010).

1.3.1 Tumeurs bénignes

Parmi les tumeurs bénignes on trouvera :

Adénofibrome : C'est la tumeur bénigne la plus fréquente. Elle touche les femmes jeunes et est un mélange de prolifération à deux composants (NASRI 2022).

- **Composant épithélial** : Constitué de canaux de taille et de forme variables, bordé par une double couche de cellules épithéliales et myoépithéliales régulières sans atypie.
- **Conjunctive component** : Made of hyperplasic pallial tissue.

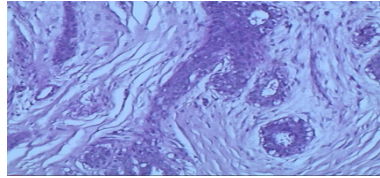


FIG. 1.4 : Image au microscope de l'ensemble de données local montrant une tumeur d'adénofibrome vu dans le facteur de grossissement 200X (Web 04).

Tumeur phyllode : Rare à croissance rapide. Mixte fibro-épithélial prolifération à prédominance conjonctive sans atypie cytonucléaire (NASRI 2022).

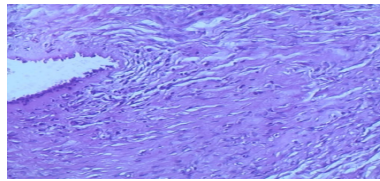


FIG. 1.5 : Image au microscope de l'ensemble de données local montrant une tumeur Phyllode vue en facteur de grossissement 200X (Web 04).

Adénose : Elle survient à tout âge, cependant elle est moins fréquente après 60 ans, elle est une hyperplasie des constituants du parenchyme mammaire (cellule épithéliale, cellule myoépithéliale du tissu conjonctif) (NASRI 2022).

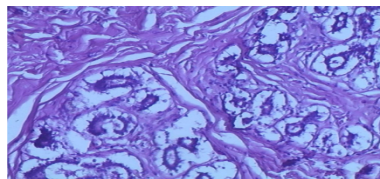


FIG. 1.6 : Image au microscope de l'ensemble de données local montrant une tumeur adénose vue en facteur de grossissement 200X (Web 04).

Adénome tubulaire : C'est une hyperplasie canalaire constituée de nombreux conduits de taille et forme variables couvertes de cellules régulières et on peut le voir (NASRI 2022).

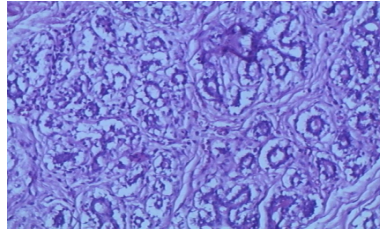


FIG. 1.7 : Image au microscope de l'ensemble de données local montrant l'adénome tubulaire tumeur vue au facteur de grossissement 200X (Web 04).

1.3.2 Tumeurs malignes

Parmi les tumeurs malignes, il existe plusieurs types histologiques :

Carcinome canalaire invasif : C'est la tumeur maligne la plus fréquente et la plus fréquente. curs de 35 à 40 ans. Il s'agit d'une prolifération maligne infiltrante constituée de tube glandulaire, cordons et touffes. Ces cellules malignes ont une hyper-vésicule noyau chromatique à contour irrégulier doté d'un ou plusieurs proéminents nucléole et présentant plusieurs figures mitotiques. La stomie tumorale est fibreuse, ponctué d'éléments inflammatoires mononucléaires (NASRI 2022).

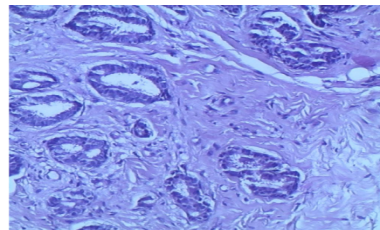


FIG. 1.8 : Image au microscope de l'ensemble de données local montrant invasive ductal carcinome vu dans un facteur de grossissement de 200X (Web 04).

Carcinome lobilaire infiltrant : Le deuxième cancer histopathologique le plus fréquent type après carcinome canalaire. Elle touche principalement les femmes âgées de 30 à 40 ans et peut parfois être bilatérale. Il est constitué de cellules non cohésives, le plus souvent de petite taille, disposées en travées, réalisant l'aspect d'une ligne séquentielle. De plus forme typique, il existe d'autres aspects architecturaux (massives,alvéolaires, tubulolobées,type circulaire) (NASRI 2022).

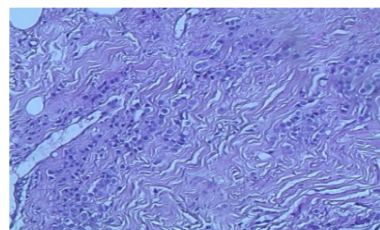


FIG. 1.9 : Image au microscope de l'ensemble de données local montrant le carcinome lobilaire vu dans le facteur de grossissement 200X (Web 04).

Carcinome mucineux : Forme histologique rare du sein, représente 1 à 6 %, il se développe chez les femmes âgées et a un bon pronostic. Constitué de cellules malignes, disposés dans un tube baigné dans une grande flaque de mucus (NASRI 2022).

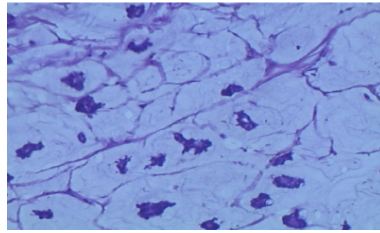


FIG. 1.10 : Image au microscope de l'ensemble de données local montrant le carcinome mucineux vu dans le facteur de grossissement 200X (Web 04).

Carcinome papillaire : Tumeur très rare qui représente 0,5 à 1 % de tous les carcinomes mas. Elle se caractérise généralement par une croissance lente et un bon pronostic. C'est une prolifération intrakystique avec excroissance papillaire, noyaux atypiques, agglutinés chromatine et pléomorphisme sévère (NASRI 2022).

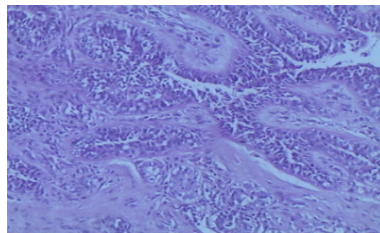


FIG. 1.11 : Image au microscope de l'ensemble de données local montrant le carcinome papillaire vu dans le facteur de grossissement 200X (Web 04).

1.3.3 Classifications des cancers du sein

Actuellement, il n'est plus considéré possible d'utiliser les termes "stades précoces" et "stades tardifs" pour décrire les différents stades d'un cancer, car ces termes reposent sur l'idée obsolète d'une progression tumorale régulière. Toutefois, il demeure important de classifier les tumeurs pour créer des groupes homogènes en termes de pronostic et de traitement. A cet effet, la classification TNM basée sur l'extension locale (Tumeur), régionale (Node ou ganglion) et générale (Métastase) de la tumeur a été développée. Néanmoins, les connaissances sur les tumeurs s'améliorant constamment, il est nécessaire de mettre à jour les classifications, comme le fait la classification moléculaire qui tient compte de l'expression de différents gènes (VERBEKE 2010).

1.3.3.1 Classification TNM

La classification TNM repose sur l'analyse de l'étendue anatomique de la tumeur, déterminée par la clinique et l'histopathologie. Elle catégorise l'atteinte locale (T0 à T4), ganglionnaire (N0 à N3) et métastatique (M0 ou M1), les chiffres associés à chaque critère

augmentant en fonction de la gravité du cancer. Ainsi, la cotation du T est déterminée par le volume de la tumeur et son extension à la cage thoracique ou à la peau. La cotation du N dépend de l'emplacement des ganglions lymphatiques par rapport à la tumeur, de leur taille, de leur nombre et de leur fixation éventuelle aux tissus environnants. Enfin, la lettre M est notée M0 en l'absence de métastases ou M1 en cas de présence de métastases, qu'elles soient uniques ou multiples et quelle que soit leur localisation (VERBEKE 2010). Ainsi on peut définir les différents stades comme dans le tableau ci-dessous :

TAB. 1.1 : Classification des stades des cancers du sein (VERBEKE 2010).

Stade	Tumeur(T)	Ganglion(N)	Métastase(M)	Description
0	Tis (in situ)	N0	M0	Cancer non invasif
I	T1	N0	M0	Tumeur < 2 cm, pas de propagation à l'extérieure du sein
II	T0/1 T2 T3	N1 N0/1 N0	M0	Tumeur de 2 à 5 cm et/ou atteinte ganglionnaire satellite mineure
III	T0/1/2 T4 Tout T	N2 N0/1/2 N3	M0	Atteinte locale importante et/ou atteinte ganglionnaire satellite majeure
IV	Tout T	Tout N	M1	Tumeur avancée localement et métastases à distance

1.3.3.2 Classification moléculaire

Grâce au développement des techniques d'analyse génomique à grande échelle telles que les biopuces à ADN, une classification plus récente a été rendue possible. Ces techniques ont permis d'établir une carte d'identité moléculaire des tumeurs, ce qui a permis d'identifier de nouveaux facteurs pronostiques et prédictifs. Ces données fournissent des informations précieuses pour la prise en charge des patients atteints de cancer du sein (VERBEKE 2010). En 2000, l'équipe de Perou (PEROU et al. 2000) a identifié cinq sous-types de carcinomes mammaires (voir tableau 1.2), qui ont depuis été associés à des valeurs pronostiques telles que l'estimation du risque de rechute ou l'évaluation du bénéfice potentiel d'un traitement spécifique. Cette découverte est considérée comme une avancée majeure dans la compréhension et la classification des cancers du sein.

TAB. 1.2 : Classification moléculaire des cancers du sein (PEROU et al. 2000).

Sous-type	Description
Luminal A	Cancers de bas grade, représentant une réceptivité hormonale importante (ER ++) et une faible prolifération. Expression des cytokératines luminales 8, 18 et du gène GATA3 impliqué dans le contrôle de la croissance et le maintien de la différenciation des tumeurs ER+. Généralement traités par hormonothérapie et de bon pronostic. P53 muté dans moins de 15% des cas.
Luminal B/C	Expression du ER moins importante et prolifération plus forte que les cancers de type « Luminal A ». De pronostics moins bons et donc généralement traités par hormonothérapie + chimiothérapie. P53 muté dans 40 à 80% des cas.
HER2	Surexpression de HER2 mais également d'autres gènes situés dans l'amplicon Erbb2 comme GRB7. Tumeurs négatives pour les récepteurs hormonaux et généralement de grade III. Associé à une surexpression de c-myc et à une forte prolifération. Pronostic défavorable mais bonne réponse aux traitements de type anthracycline, taxane et Trastuzumab (Herceptin®). P53 muté dans environ 70% des cas.
Basal-like	Tumeurs dites « triples négatives » c'est à dire n'exprimant pas les récepteurs hormonaux (ER et PgR) et ne surexprimant pas le récepteur HER2. Expression des cytokératines basales 5/6, 14 et 17, des caveolines 1 et 2, de l'EGFR et de c-kit. P53 muté dans 80% des cas et tumeurs souvent associées à une mutation de BRCA1. Pronostic le plus défavorable. Ne répondent pas aux traitements hormonaux et au Trastuzumab mais les chimiothérapies classiques peuvent aider. Présente un intérêt pour les essais cliniques d'anti-EGFR ou d'anti-angiogène.
Normal-like	Sous-type assez mal défini. Exprime des gènes connus pour être exprimé dans les cellules non-épithéliales et le tissu adipeux. Pronostic intermédiaire

1.4 Dépistages et diagnostic du cancer du sein

Malheureusement, personne n'est à l'abri du cancer du sein. Cela nécessite une grande vigilance, qui exige un dépistage, un diagnostic et un traitement.

1.4.1 Dépistages du cancer du sein

Le dépistage est une mesure préventive qui permet de détecter la présence d'une maladie avant l'apparition de tous les symptômes. Dans certains cas, il peut être judicieux de se soumettre à un test de dépistage, car le cancer du sein détecté à un stade précoce peut être traité avec succès, offrant ainsi de grandes chances de guérison. Outre l'examen de dépistage, l'auto-palpation joue un rôle essentiel dans la détection précoce d'une pathologie mammaire. Le processus de dépistage comprend trois phases cruciales : un interrogatoire suivi d'un examen clinique, un examen radiologique et, éventuellement, un examen biologique (MEBARKI 2020).

1.4.2 Diagnostic du cancer du sein

Le processus d'examen clinique pour détecter le cancer du sein commence par l'inspection et la palpation des seins et des ganglions lymphatiques. Si des signes cliniques suspects sont observés ou lors d'un dépistage régulier, une mammographie peut être réalisée. Dans les cas où la mammographie montre une lésion, une échographie peut être utilisée pour mieux caractériser la nature de la lésion, surtout dans les seins denses. Pour confirmer un diagnostic de cancer du sein, une biopsie est réalisée à partir d'un prélèvement micro ou macroscopique. Si l'image mammographique est classée ACR 4 ou ACR 5, une biopsie est systématiquement effectuée (HADJIDJ 2018). Donc la démarche conventionnelle du processeur diagnostique liée à la tumeur du sein est caractérisée par trois phases cruciales : L'examen clinique, l'imagerie mammaire et anatomopathologique.

1.4.2.1 L'examen clinique

- L'enquête médicale consiste en une série de questions posées par le médecin, afin de comprendre le contexte de la découverte, les symptômes associés, les facteurs de risque, les antécédents médicaux et les événements pertinents.
- L'examen physique, également appelé examen complet, repose sur l'observation, la palpation et l'auscultation.

1.4.2.2 L'imagerie mammaire

Il existe trois examens complémentaires classiques :

- La mammographie,
- L'échographie mammaire,
- L'imagerie par résonance magnétique.

1.4.2.3 Anatomopathologie

L'examen anapath, qui est l'abréviation de l'examen anatomopathologique, est utilisé pour analyser les cellules ou les tissus prélevés au niveau d'une anomalie mammaire, avant ou après la chirurgie. Ce type d'examen est généralement effectué par un médecin, un chirurgien ou un radiologue. L'examen commence par une inspection visuelle, puis il est complété par une ponction cytologique et une biopsie.

1.5 Imagerie mammaire

Les principales techniques d'imagerie utilisées pour le dépistage et le diagnostic du cancer du sein sont la mammographie (utilisant des rayons X), l'échographie (utilisant des ultrasons), l'IRM (utilisant la résonance magnétique) et l'imagerie microscopique (biopsie).

1.5.1 La mammographie

La mammographie est une technique d'imagerie médicale qui utilise des rayons X pour produire des images de l'intérieur des seins. Elle est utilisée pour détecter des cancers de petite taille à un stade précoce, avant même que des symptômes n'apparaissent (L. WANG 2017). La mammographie est réalisée en cas de détection d'anomalie lors de l'examen clinique ou tous les deux ans pour les femmes de plus de 50 ans dans le cadre du dépistage organisé. Pour cet examen, la patiente se tient debout torse nu dans une cabine de radiologie et est positionnée devant un appareil de radiographie appelé "mammographe". Le sein est comprimé entre une plaque en plastique et un plateau porte-film, et deux radiographies sont prises pour chaque sein, l'une de face et l'autre en oblique. Bien que l'examen dure environ 10 minutes et qu'il soit indolore, il peut être inconfortable pour la patiente (WÖCKEL et al. 2018).

Le mammographe peut révéler la présence d'une tumeur sous la forme d'une masse ou de petites taches blanches qui correspondent à des microcalcifications. Les caractéristiques de ces signes radiologiques permettent de déterminer la présence éventuelle d'un cancer.

1.5.2 Échographie

L'échographie mammaire est préconisée lorsque l'examen clinique ou la mammographie détecte une anomalie. En effet, pour les femmes de moins de 30 ans, il est conseillé de réaliser une échographie des seins plutôt qu'une mammographie en raison de la densité mammaire qui peut empêcher d'obtenir des images claires.

L'échographie, utilisant des ondes sonores à haute fréquence sans danger pour le corps, permet d'obtenir des images précises de l'intérieur du sein ainsi que des lésions détectées lors de la mammographie (L. WANG 2017).

L'échographie est un examen indolore qui dure environ 10 minutes. Pour réaliser cet examen, un gel hypoallergénique est appliqué sur le sein de la patiente pour faciliter le

contact entre la sonde d'échographie et la peau. La sonde est ensuite fermement appliquée sur le sein et orientée dans toutes les directions pour permettre une observation complète de tissu mammaire et pour repérer les ganglions anormaux. Les anomalies sont détectables grâce à une différence de contraste entre ces zones et les zones saines (WÖCKEL et al. 2018).

1.5.3 IRM

Une IRM (imagerie par résonance magnétique) n'est pas systématiquement réalisée, lors du diagnostic d'un cancer du sein et, ne remplace pas une mammographie ou une échographie. Elle est utilisée lorsque les résultats obtenus avec les techniques décrites ci-dessus ne sont pas exploitables, par manque de netteté ou pour faire la différence entre les tumeurs bénignes et malignes chez les femmes à haut risque de cancer (L. WANG 2017). Les images obtenues par IRM sont produites grâce à des ondes électromagnétiques projetées sur la partie du corps examinée, offrant ainsi une grande précision pour observer l'anomalies.

Pendant l'examen, la patiente est allongée sur le ventre sur la table de l'appareil, qui est équipé de deux ouvertures pour permettre le passage des seins, entourées de bobines. Ce sont ces bobines qui reçoivent et émettent les ondes radioélectriques dans le champ magnétique de l'appareil.

Les avantages et les inconvénients des techniques médicales évoquées précédemment sont synthétisés dans le tableau 1.3.

TAB. 1.3 : Avantages et inconvénients des différentes techniques médicales de dépistage du cancer du sein (MEBARKI 2020).

Méthodes	Ionisant	Efficacité	Confort	Cout	Temps	Résolution
Mammographie	Oui	Bonne sensibilité pour les seins non denses	Non confortable	Moyen	Court	Haute qualité
Echographie	non	Bonne sensibilité pour les seins denses mais les lésions profondes sont difficiles à détecter	confortable	Moyen	Court	Résolution élevée mais la performance dépend des compétences du technicien et de l'expertise de l'opérateur.
IRM	Non	Efficace pour les patientes ayant des seins denses. Capacité de détecter de petites tumeurs et profondes. Difficile à détecter les calcifications	La patiente a peur d'être dans un espace étroit	Cher	Prend du temps (quelques heures)	Haute résolution Les images peuvent être capturées à partir d'orientations différentes.

Les limites des techniques médicales actuelles ont suscité l'intérêt d'utiliser un outil alternatif plus abordable et plus confortable pour la détection précoce et le diagnostic du cancer du sein. Au fil des années, plusieurs groupes de recherche ont proposé l'imagerie microscopique comme méthode d'imagerie alternative (EGGERTSON 2004).

1.5.4 L'imagerie microscopique

Les scientifiques utilisent les techniques d'imagerie microscopique pour améliorer leur capacité d'observation du monde microscopique. Les avancées en microscopie permettent maintenant de visualiser une grande variété de processus biologiques et de structures cellulaires (MEBARKI 2020).

1.5.4.1 L'histopathologie

L'histopathologie est le processus qui consiste à examiner les tissus, pour en savoir plus sur la maladie et la cause première de celle-ci. Les tissus sont observés au microscope, dans un laboratoire d'histopathologie. Le processus implique la réalisation d'une biopsie. Avant de commencer l'examen, le tissu est préparé, puis il est traité et ensuite analysé.

Les services d'histopathologie sont utilisés en tenant compte du type du tissu examiné ainsi que, du type d'enquête menée. L'évaluation initiale d'un cancer du sein, implique généralement un examen physique, une mammographie, une radiographie du thorax et parfois un scanner. Cependant, l'évaluation finale dépend souvent des résultats des examens pathologiques, qui sont le moyen le plus précis de déterminer l'étendue de la tumeur. Ce domaine d'étude est extrêmement crucial car il permet aux histopathologistes de comprendre et, de détecter les maladies. L'histopathologie permet de rechercher les changements dans les cellules, qui expliquent la cause réelle de la maladie du patient (SHAFIE 2019). Les pathologistes sont en mesure d'établir un diagnostic, en examinant une partie d'un tissu provenant de divers organes. L'histopathologie est essentielle, car elle permet d'élargir et de faire progresser les options de traitement. Pour la plupart des types de cancer, la biopsie est le seul moyen d'établir un diagnostic définitif (SIKANDAR 2018).

1.5.4.2 La biopsie

Une biopsie consiste à prélever des tissus d'une zone spécifique du corps afin de confirmer ou d'identifier une pathologie suspectée. Le tissu prélevé est ensuite examiné par un pathologiste, un médecin spécialisé dans le diagnostic des maladies, pour déterminer la présence éventuelle de cellules cancéreuses. Si un cancer est détecté, le pathologiste peut analyser les caractéristiques du cancer. Le rapport de la biopsie présente toutes les conclusions du pathologiste (AZAMJAH et al. 2019).

1.6 Classification BI RADS des anomalies

Le système BI-RADS (Breast Imaging-Reporting And Data System) est employé par les professionnels de santé pour évaluer les résultats d'imagerie et les classer en catégories numérotées de 0 à 6. Une évaluation de 1 ou 2 correspond à des résultats normaux ou à

des anomalies bénignes (voir tableau 1.4). Cependant, pour une évaluation de 3 ou plus, il est recommandé aux médecins et radiologues de programmer une visite de suivi ou une biopsie pour déterminer les étapes à suivre (MIAO et al. 2018).

TAB. 1.4 : Classification BI-RADS (Breast Imaging Reporting And Data System) (Web 05).

Catégorie	Imagerie	Risque de cancer	Conduite
0	Investigation incomplète		Investigation compléter ou comparer avec films antérieurs
1	Normal	0%	Retour au dépistage
2	Anomalie bénigne	0 %	Retour au dépistage
3	Anomalie probablement bénigne	>0% mais ≤2%	Suivi 6 mois
4	Anomalie demandant une biopsie	>2 % mais <95 %	Biopsie
4A	faiblement suspecte	>2 % à ≤10%	
4B	modérément suspecte	>10 % à ≤50%	
4C	très suspecte	>50% à <95%	
5	Anomalie fortement suspecte d'un cancer	≥ 95%	Biopsie
6	Cancer prouvé à la biopsie	100%	Chirurgie

1.7 Conclusion

Dans ce chapitre, nous avons d'abord introduit la maladie du cancer du sein et quelques statistiques pour les femmes atteintes du cancer du sein, ainsi que les différents types de tumeurs. Par la suite, nous avons décrit les différentes procédures de dépistage et diagnostic du cancer du sein. Après, nous avons présenté les techniques principales d'imagerie utilisées pour le dépistage et le diagnostic du cancer du sein. Et pour finir, nous avons présenté le système BI-RADS pour la classification et l'évaluation de l'imagerie.

Dans le chapitre suivant, nous montrerons, via une revue bibliographique, comment les techniques de l'intelligence artificielle peuvent être utilisées pour le diagnostic du cancer du sein.

Chapitre 2

Intelligence artificielle pour le diagnostic du cancer du sein

2.1 Introduction

Les systèmes basés sur l'intelligence artificielle (IA) sont de plus en plus utilisés pour aider au diagnostic du cancer du sein. Ces systèmes utilisent des algorithmes complexes pour analyser des images médicales et fournir des informations précieuses aux médecins. Par exemple, l'IA peut être utilisée pour analyser des mammographies et aider à détecter les signes de cancer du sein à un stade précoce. Elle peut également être utilisée pour aider à différencier les tumeurs bénignes des tumeurs malignes à partir d'images d'échographie ou d'IRM.

Dans ce chapitre, nous présenterons dans un premier temps, quelques logiciels de diagnostic des maladies cancéreuses. Ensuite, nous aborderons les différentes techniques de l'intelligence artificielle utilisées couramment pour le diagnostic du cancer du sein. Ces techniques sont classifiées en trois types principaux à savoir, le machine Learning, le Deep Learning et les Transformers et plus particulièrement les vision Transformers (ViT) que nous avons choisis comme technique de classification dans notre système de diagnostic du cancer du sein à partir des images histopathologiques. Finalement, nous exposerons un état de l'art sur le diagnostic automatique du cancer du sein.

2.2 Quelques logiciels de l'IA pour le diagnostic des maladies cancéreuses

Il existe de nombreux logiciels qui utilisent l'IA pour prédire les maladies cancéreuses. Parmi ces logiciels, citons à titre d'exemple, PathAI, Freenome et Proscia.

2.2.1 PathAI



FIG. 2.1 : Logo PathAI (Web 06).

PathAI déploie des technologies d'apprentissage automatique pour aider les pathologistes à établir des diagnostics plus précis. Les plans actuels de PathAI incluent la réduction des erreurs dans le diagnostic du cancer et développer des approches pour un traitement médical individualisé. Il est également utilisé pour la drogue découvert et utilisé par de grandes entreprises telles que Bristol-Myers Squibb, Bill Melinda Fondation Gates. Il a été développé à Cambridge, Massachusetts (Web 06).

2.2.2 Freenome



FIG. 2.2 : Logo Freenome (Web 07).

Freenome déploie l'IA lors des dépistages généraux, des tests de diagnostic et des analyses de sang visant à détecter le cancer à ses débuts et développer par la suite de nouveaux traitements. Il a été développé à San Francisco, en Californie (Web 07).

2.2.3 Proscia



FIG. 2.3 : Logo Proscia (Web 08).

Proscia est une plateforme de pathologie numérique, elle détecte des modèles dans les cellules cancéreuses à l'aide de l'IA. Le logiciel aide les laboratoires de pathologie à éliminer les goulots d'étranglement dans la gestion des données et utilise l'analyse d'images alimentée par l'IA pour connecter les points de données qui soutiennent la découverte et le traitement du cancer. Il a été développé à Philadelphie, en Pennsylvanie (Web 08).

Ces logiciels tirent parti des méthodes d'apprentissage automatique, notamment le machine learning et le deep learning.

2.3 Apprentissage automatique (Machine Learning)

L'apprentissage automatique c'est un domaine d'étude qui confère aux ordinateurs la capacité d'apprendre sans être explicitement programmés (SAMUEL 1959). Il fait partie de l'intelligence artificielle. Ces algorithmes construisent des modèles basés sur des données d'entraînement, également appelées "données d'entraînement", afin de réaliser des prédictions ou des prises de décisions sans avoir besoin d'une programmation explicite à cet effet. Les algorithmes d'apprentissage automatique sont largement utilisés dans divers domaines tels que la médecine et la vision par ordinateur (HU et al. 2020).

2.3.1 Les différents types d'apprentissage

L'apprentissage automatique peut être divisé en trois catégories principales : L'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement.

2.3.1.1 Apprentissage supervisé

L'apprentissage supervisé est la tâche d'apprentissage automatique qui consiste à apprendre une fonction qui associe une entrée à une sortie sur la base d'exemples de paires d'entrée-sortie. Il déduit une fonction à partir de données d'entraînement étiquetées, qui sont un ensemble d'exemples d'entraînement. Les algorithmes d'apprentissage supervisé sont ceux qui nécessitent une assistance externe. L'ensemble de données d'entrée est divisé en un ensemble d'entraînement et un ensemble de test. L'ensemble d'entraînement contient une variable de sortie qui doit être prédite ou classée. Tous les algorithmes apprennent certains types de schémas à partir de l'ensemble de données d'entraînement et les appliquent à l'ensemble de données de test pour la prédiction ou la classification (MAHESH 2020). Il existe deux principaux domaines dans lesquels l'apprentissage automatique supervisé est utile : les problèmes de classification et les problèmes de régression. La classification consiste à prendre une valeur d'entrée et à l'affecter à une valeur discrète. Dans les problèmes de classification, la sortie est généralement une classe ou une catégorie.

2.3.1.2 Apprentissage non supervisé

Il s'agit d'un apprentissage qui se fait sans l'utilisation d'étiquettes de données. Dans ce cas, apprend les représentations internes ou les fonctionnalités importantes afin de découvrir des relations ou une structure inconnue dans les données d'entrée. Le clustering, la réduction de la dimensionnalité et les techniques génératives sont souvent considérés comme des approches d'apprentissage non supervisé (ARULKUMARAN et al. 2017).

2.3.1.3 Apprentissage par renforcement

L'apprentissage par renforcement est un troisième paradigme d'apprentissage automatique, aux côtés de l'apprentissage supervisé et de l'apprentissage non supervisé. Un des défis qui se posent dans l'apprentissage par renforcement, et pas dans d'autres types d'apprentissage, est le compromis entre exploration et exploitation. Pour obtenir beaucoup de récompenses, un agent d'apprentissage par renforcement doit préférer les actions qu'il a déjà essayées et trouvées efficaces pour produire des récompenses. Mais pour découvrir de telles actions, il doit essayer des actions qu'il n'a pas sélectionnées auparavant. L'agent doit exploiter ce qu'il sait déjà pour obtenir des récompenses, mais il doit également explorer pour améliorer ses sélections d'actions à l'avenir. Le dilemme est que ni l'exploration ni l'exploitation ne peuvent être poursuivies exclusivement sans échouer dans la tâche. L'agent doit essayer une variété d'actions et privilégier progressivement celles qui semblent être les meilleures. Une autre caractéristique clé de l'apprentissage par renforcement est qu'il considère explicitement l'ensemble du problème d'un agent dirigé

vers un objectif interagissant avec un environnement incertain (SUTTON et al. 1998).

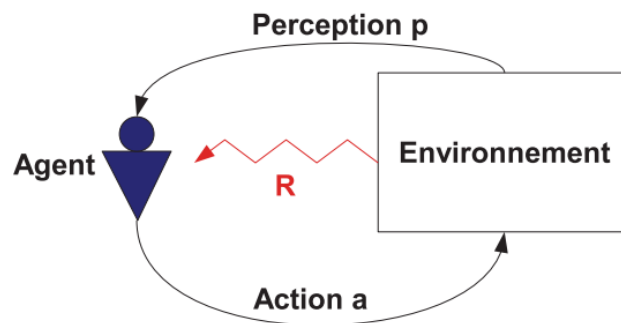


FIG. 2.4 : Apprentissage par renforcement (MAZAC 2015).

2.3.2 Méthodes d'apprentissage automatique

Il existe plusieurs méthodes d'apprentissage automatique. Dans cette section, nous présenterons de manière concise les modèles de classification les plus couramment utilisés et populaires.

2.3.2.1 Machines à vecteurs de support

Les machines à vecteurs de support (SVM) est un type spécifique de méthode d'apprentissage supervisé, sont des classifieurs particuliers qui sont basés sur le principe de maximisation de la marge. Elles effectuent une minimisation des risques structurels, qui a été introduite dans l'apprentissage automatique par Vapnik (ADANKON et al. 2009). Les SVM fonctionnent en réduisant un problème de classification ou de discrimination à un hyperplan dans un espace caractéristique où les données sont séparées en différentes classes avec une frontière maximale entre elles (ou "marge maximale"). C'est pourquoi les SVM sont également appelés séparateurs à marge large. L'idée clé est que les données doivent être linéairement séparables. Pour atteindre cela, les machines à vecteurs de support utilisent des noyaux, qui sont des fonctions mathématiques permettant de projeter et de séparer les données dans l'espace vectoriel, les "vecteurs de support" étant les données les plus proches de la frontière. L'objectif est de trouver la frontière la plus éloignée de tous les points d'entraînement, car cela permet une meilleure généralisation.

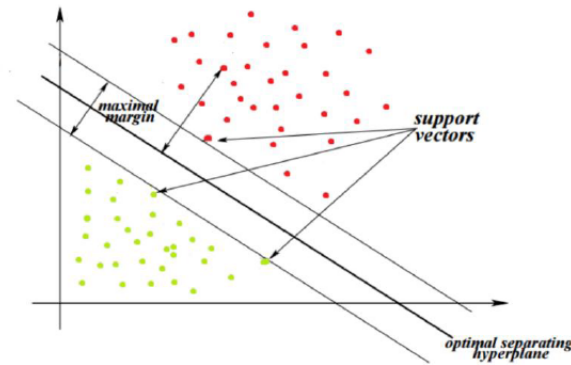


FIG. 2.5 : Machine à vecteurs de support (KUMAR 2021).

2.3.2.2 K plus proche voisins

L'algorithme KNN est l'un des algorithmes les plus simples de l'apprentissage automatique. Dans le contexte de la classification d'une nouvelle observation x , l'idée principale est de faire voter les voisins les plus proches de cette observation. La classe de x est déterminée en fonction de la classe majoritaire parmi les k voisins les plus proches de l'observation x . Ainsi, la méthode KNN est basée sur le voisinage et est non paramétrique, ce qui signifie que l'algorithme permet de réaliser une classification sans faire d'hypothèses sur la fonction $y = f(x_1, x_2, \dots, x_p)$ qui relie la variable dépendante aux variables indépendantes (MATHIEU-DUPAS 2010).

2.3.2.3 Arbres de décision

Un arbre de décision est un enchaînement hiérarchique de règles logiques ou de productions construites de manière automatique à partir d'un ensemble de données. La construction de l'arbre de décision consiste à utiliser les descripteurs, pour subdiviser progressivement l'ensemble en sous-ensembles de plus en plus fin. Les arbres de décision sont des algorithmes de classification de données reposant sur un apprentissage supervisé. Comme son nom l'indique, la structure de ces algorithmes ressemble à des arbres constitués de nœuds, de branches et de feuilles. La construction de ces arbres est réalisée à l'aide d'une base de données brute (vecteur des caractéristiques et classes) et de lois qui permettent de déterminer les variables discriminantes pour la classification efficace des futures données. Chacun des nœuds constituant l'arbre représente une règle de classification préalablement déterminée de manière récursive. Pour voir comment un arbre de décision prédit une réponse, il faut suivre les décisions dans l'arbre du nœud racine (début) jusqu'au nœud feuille qui contient la réponse. Les arbres de classification donnent des réponses nominales, tel que vrai ou faux et les arbres de régression donnent des réponses numériques. Les arbres de décision sont relativement rapides et faciles à suivre dans leur résolution ; vous pouvez voir une représentation complète du chemin parcouru de la racine à la feuille. Cela est particulièrement utile si vous devez partager les résultats avec des personnes intéressées par la manière dont une conclusion a été tirée (ZOUNGRANA 2020).

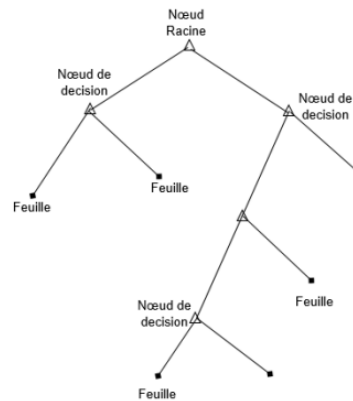


FIG. 2.6 : Exemple d'illustration d'un arbre de décision (ZOUNGRANA 2020).

2.3.2.4 Modèles de Markov cachés

Les modèles de Markov cachés (HMM) sont largement employés dans le domaine de la reconnaissance de texte pour identifier les caractères, les mots et les lignes. Ces modèles graphiques probabilistes intègrent des connaissances préalables telles que le modèle de langage et le lexique pour modéliser une séquence. À l'origine, les HMM ont été efficacement utilisés pour la reconnaissance vocale. En raison de leur succès dans ce domaine et des similitudes entre la reconnaissance de la parole et l'écriture manuscrite cursive, les HMM ont été étendus à la reconnaissance de l'écriture manuscrite en ligne et hors ligne (ALKHATEEB et al. 2011).

2.3.2.5 Réseaux de neurones artificiels (ANN)

Les réseaux de neurones artificiels (ANN) sont des architectures informatiques modélisées d'après le cerveau. Ils sont construits à partir d'une série de "neurones" organisés en couches. Ces neurones présentent un comportement global déterminé par les connexions établies entre les différents éléments de traitement et les paramètres associés au sein de l'architecture du réseau neuronal. Chaque connexion qui relie les neurones dans des couches consécutives est pondérée. Le poids w_{ij} représente la force de la connexion entre le i ème neurone d'une couche et le $j^{ième}$ neurone de la couche suivante du réseau. La structure d'un réseau neuronal se compose d'une couche "d'entrée", d'une ou plusieurs couches "cachées" et d'une couche "de sortie". Le nombre de couches cachées et le nombre de neurones dans chacune de ces couches dépendent de la complexité du système considéré. La figure 2.7 montre une architecture ANN typique avec deux couches cachées (MOGHADDAMNIA et al. 2009).

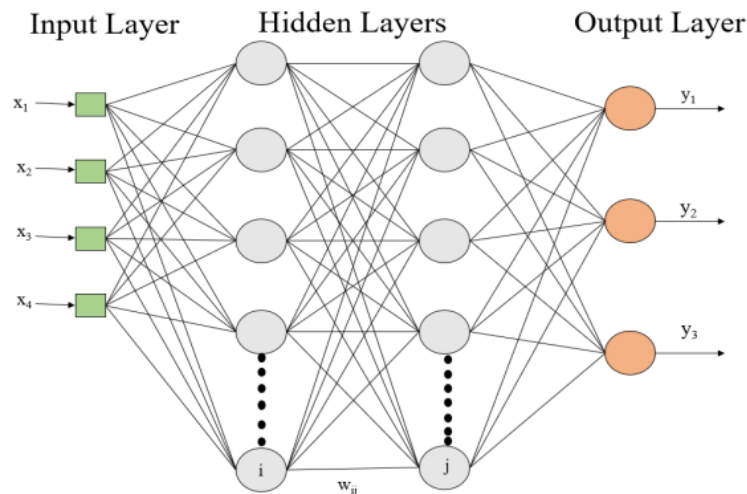


FIG. 2.7 : Structure générale du réseau de neurones artificiels avec deux couches cachées (MOGHADDAMNIA et al. 2009).

Le perceptron multi-couches (MLP)

Un modèle largement utilisé appelé perceptron multi-couches (MLP) est le type le plus courant d'ANN. Le MLP se compose d'une couche d'entrée, d'une ou plusieurs couches cachées et d'une couche de sortie. Chaque couche emploie plusieurs neurones, et chaque neurone d'une couche est connecté aux neurones de la couche adjacente avec différents poids, ce qui permet de déterminer dans quelle mesure une unité influencera l'autre. La tâche d'un MLP est d'obtenir la sortie souhaitée en fonction des entrées de certains types. Dans cette optique, l'apprentissage du perceptron est réalisé. Les entrées et les sorties désirées sont chargées dans le schéma, et l'erreur de la réponse est déterminée. Les paramètres du système, les poids dans les interconnexions, sont modifiés pendant la phase d'apprentissage de l'ANN afin de réduire la différence entre la sortie désirée et la sortie réelle. Alors que le MLP se concentre davantage sur la résolution de problèmes non linéairement séparables, la classification et les fonctions continues approximatives (FRANCO ORTELLADO et al. 2019).

2.4 Apprentissage profond (Deep Learning)

L'apprentissage profond est un sous-ensemble de l'apprentissage automatique, utilise des Réseaux de Neurones Artificiels qui sont inspirés par la biologie du cerveau humain, et cela permet de créer des algorithmes très efficaces pour résoudre les problèmes de classification. En général, le Deep Learning est un sous-domaine de l'Intelligence Artificielle qui repose sur l'utilisation de réseaux de neurones complexes et de grandes quantités de données précises, afin de permettre aux machines d'apprendre de manière autonome, comme le ferait un être humain. Le Deep Learning est considéré comme étant à la pointe de l'IA et est l'une des technologies les plus passionnantes des dernières décennies. Aujourd'hui, ces techniques d'apprentissage sont largement utilisées dans de nombreux domaines tels que classification des images (TYAGI et al. 2020).

2.4.1 Les réseaux de neurones récurrents (RNN)

Les réseaux de neurones récurrents (RNN) sont des modèles d'apprentissage profond spécialement conçus pour résoudre des problèmes liés aux séries temporelles. Les RNN excellent dans divers domaines tels que le traitement du signal, le traitement du langage naturel et la reconnaissance de la parole. Ils sont capables de prendre en compte les dépendances temporelles des données, en utilisant les sorties précédentes comme entrées pour les étapes suivantes. Cette capacité est due à leur mémoire interne qui leur permet de conserver des informations contextuelles. Cette prise en compte du contexte temporel permet aux RNN d'obtenir de meilleures performances dans des problèmes où la dimension temporelle est cruciale. Cependant, les RNN standard rencontrent des difficultés à apprendre des dépendances temporelles supérieures à quelques pas de temps discrets (environ 5 à 10) entre les événements d'entrée pertinents et les signaux cibles (BROWNLEE 2017). Le principal problème des réseaux de neurones récurrents (RNN) réside dans leur difficulté à capturer des dépendances à long terme dans les séquences. En raison de la nature récurrente de leur architecture, les RNN ont tendance à souffrir de ce que l'on appelle le "problème du gradient qui disparaît" ou le "problème du gradient explosif". Dans ce qui suit, nous présenterons brièvement le Long Short-Term Memory (LSTM) qui a été introduit comme solution à ce problème.

Le Long Short Term Memory (LSTM)

Les réseaux LSTM (Long Short-Term Memory) ont été développés pour remédier à la perte de mémoire à long terme des réseaux de neurones récurrents (RNN). Ils sont conçus pour modéliser des dépendances à longue distance dans une séquence en utilisant une mémoire interne appelée "cell state" dans chaque neurone.

Dans un LSTM, lorsque la sortie de l'unité neuronale précédente arrive à l'unité courante, elle passe par trois portes : La porte d'oubli, la porte d'entrée et la porte de sortie. Ces portes utilisent des fonctions d'activation telles que des sigmoïdes et des tangentes hyperboliques pour gérer l'information. La figure 2.8 illustre un schéma de cellule LSTM.

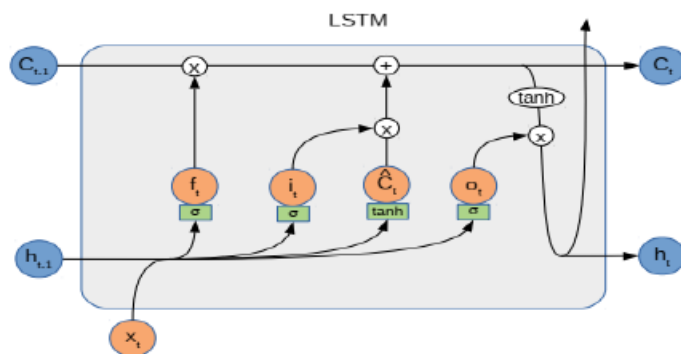


FIG. 2.8 : Cellule Long Short-Term Memory (LSTM)(SUEIRAS 2021).

La porte i , nommée porte d'entrée, apprend quelles valeurs seront mises à jour dans

le vecteur d'état. La porte f , appelée porte d'oubli, apprend les informations de l'état précédent qui peuvent être rejetées. Avec la sortie de ces deux couches, le réseau crée un vecteur de nouvelles valeurs d'état candidates c . Enfin, la porte o , nommée porte de sortie, apprend quelles informations iront à la sortie h .

Cette architecture LSTM peut également être utilisée de manière bidirectionnelle, en lisant la séquence dans l'ordre chronologique et antéchronologique. Les sorties des deux directions sont ensuite concaténées pour obtenir la sortie finale de chaque élément. Cela permet de capturer les dépendances passées et futures dans la séquence. Les réseaux LSTM sont particulièrement adaptés aux données qui nécessitent du contexte, ce qui les rend largement utilisés dans le traitement de la parole et d'autres domaines. Cependant, ils ont un temps d'apprentissage long en raison du grand nombre de paramètres à apprendre, et il est difficile de paralléliser les calculs en raison de leur nature séquentielle. De plus, dans le cas du LSTM bidirectionnel, il est nécessaire de disposer de la séquence complète pour effectuer une prédiction, ce qui peut ne pas être compatible avec des applications en temps réel (MACARY 2022).

2.4.2 Les réseaux de neurones convolutifs (CNN)

Les réseaux de neurones convolutifs (CNN) tirent leur nom de l'utilisation de l'opération mathématique de convolution. Ils sont une forme spécifique de réseaux neuronaux qui utilisent la convolution à la place de la multiplication matricielle générale dans au moins une de leurs couches. Les CNN sont considérés comme l'un des meilleurs algorithmes d'apprentissage pour effectuer la convolution, ce qui permet d'extraire des caractéristiques pertinentes à partir de données localement corrélées. Les noyaux convolutifs produisent une sortie qui est ensuite passée à une unité de traitement non linéaire, également appelée fonction d'activation. Cette fonction non linéaire facilite l'apprentissage des abstractions et intègre la non-linéarité dans l'espace des caractéristiques. Grâce à cette non-linéarité, différentes réponses génèrent différents modèles d'activations, ce qui facilite l'apprentissage des différences sémantiques dans les images (A. KHAN et al. 2020).

Sur la figure 2.9, une architecture générale de CNN pour la classification d'images est présentée couche par couche.

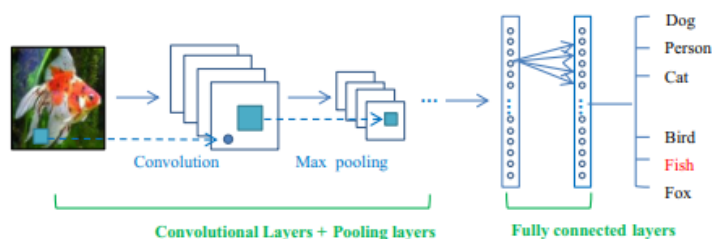


FIG. 2.9 : Le pipeline de l'architecture générale CNN (GUO et al. 2016).

2.4.2.1 Architecture globale

Le CNN se compose de trois couches neuronales principales : Couche de convolution (CONV), Couche de Pooling (POOL) et Couche Fully-Connected (FC). Différents types de couches jouent différents rôles.

- **Couche de convolution (CONV)**

Dans les couches de convolution, un réseau de neurones convolutif (CNN) utilise différents noyaux pour convoluer l'image entière ainsi que les cartes de caractéristiques intermédiaires, générant ainsi diverses cartes de caractéristiques, comme indiqué dans la figure 2.10 (ZEILER 2013).

Il y a trois principaux avantages de l'opération de convolution :

- Le mécanisme de partage de poids dans la même carte de caractéristiques réduit le nombre de paramètres,
- La connectivité locale apprend les corrélations entre les pixels voisins,
- L'invariance à l'emplacement de l'objet.

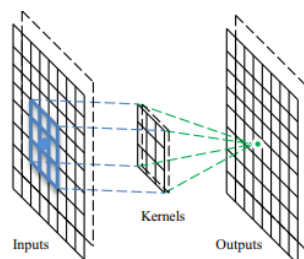


FIG. 2.10 : L'opération de convolution (GUO et al. 2016).

- **Couche de pooling (POOL)**

Une couche de pooling suit une couche de convolution et peut être utilisée pour réduire les dimensions des cartes de caractéristiques et des paramètres du réseau. Tout comme les couches de convolution, les couches de pooling sont également invariantes par translation, car leurs calculs prennent en compte les pixels voisins. Les stratégies les plus couramment utilisées sont le pooling moyen et le pooling maximum. La figure 2.11 donne un exemple d'un processus de pooling maximum. Pour des cartes de caractéristiques de 8 x 8, les cartes de sortie sont réduites à des dimensions de 4 x 4, avec un opérateur de pooling maximum de taille 2 x 2 et un décalage de 2 (GUO et al. 2016).

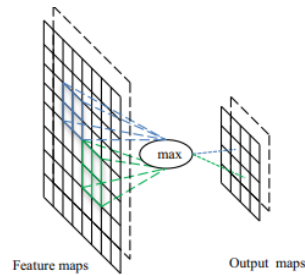


FIG. 2.11 : L'opération de Max pooling (GUO et al. 2016).

- **Couche Fully-Connected (FC)**

Couche Fully-Connected convertissant les cartes de caractéristiques 2D en un vecteur de caractéristiques 1D, pour une représentation de caractéristiques ultérieure, comme on peut le voir dans la figure 2.12. Elle fonctionne comme un réseau neuronal traditionnel et contient environ 90 % des paramètres d'un CNN. Cela nous permet de faire avancer le réseau neuronal dans un vecteur de longueur prédéfinie. Nous pourrions soit faire avancer le vecteur dans un certain nombre de catégories pour la classification d'images, (KRIZHEVSKY et al. 2017) soit le prendre comme vecteur de caractéristiques pour un traitement ultérieur (GUO et al. 2016).

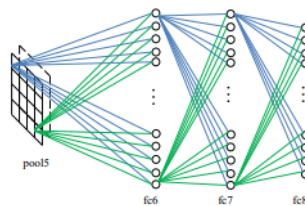


FIG. 2.12 : Couche Fully-Connected (GUO et al. 2016).

2.4.2.2 Quelques architectures des réseaux de neurones convolutifs

Diverses architectures de réseaux sont employées pour les tâches de classification d'images, parmi lesquelles VGG16, Inception Net (GoogLeNet) et ResNet se distinguent particulièrement.

- **AlexNet**

L'architecture proposée par (KRIZHEVSKY et al. 2017) a été révolutionnaire dans le domaine des réseaux de neurones à partir de 2012, notamment grâce à sa victoire dans la compétition ImageNet pour la classification d'images. Cette architecture se compose de cinq couches de convolution et trois couches de pooling. La taille des noyaux de convolution varie (11×11 , 5×5 , 3×3) en fonction de la couche. La fonction d'activation utilisée entre chaque couche est la fonction ReLU. Après le passage de l'image à travers les couches de convolution, de pooling et d'activation, une carte de caractéristiques est obtenue. Cette carte est ensuite traitée par un perceptron multicouches comprenant deux couches cachées et une couche de sortie.

- **VGG**

Cette architecture de CNN a été présentée par (SIMONYAN et al. 2014). Au lieu d'utiliser une seule couche de convolution par niveau de profondeur comme dans AlexNet, cette architecture utilise des séquences de convolutions. De plus, les filtres convolutifs sont de taille réduite par rapport à AlexNet (noyau de taille 3×3).

- **GoogLeNet**

Cette architecture neuronale convolutive a été créée par (SZEGEDY et al. 2015) et permet une réduction du temps de calcul par rapport à l'architecture VGG présentée précédemment. Pour cela, le GoogLeNet est composé de plusieurs couches appelées couches d'inception. Elles sont composées de plusieurs modules de convolution de taille 1×1 , 3×3 et 5×5 , exécutés en parallèle sur la carte de caractéristiques résultant de la couche précédente. Des filtres additionnels permettent de réduire la dimension des cartes de caractéristiques ce qui permet un gain important de temps de calcul.

- **ResNet**

Ce CNN a été introduit par (HE et al. 2016). Il permet l'apprentissage de réseaux extrêmement profonds, avec plus de 150 couches. L'apprentissage de réseaux aussi profonds présente des défis, notamment en ce qui concerne la rétropropagation du gradient. En effet, plus le réseau est profond, plus le gradient devient faible lors de la mise à jour des poids des couches inférieures (les premières couches). Par conséquent, une architecture trop profonde ne parvient pas à mettre à jour efficacement ces couches. L'idée clé développée dans ResNet est l'utilisation de connexions résiduelles, qui permettent une meilleure optimisation des réseaux profonds. Une connexion résiduelle permet à l'entrée de passer à travers deux filtres de convolution, tout en étant directement transmise aux couches suivantes.

2.4.3 Les Transformers

Transformer est une nouvelle architecture de modèle d'apprentissage profond et la norme la plus élevée pour le traitement du langage naturel (NLP), introduite dans le document "Attention is all you need" (VASWANI et al. 2017). Les transformers sont à l'origine conçus pour gérer les données d'entrée séquentielles. Grâce à l'utilisation de self-attention, ils donnent le contexte d'entrée de sorte qu'il n'a pas nécessairement besoin d'être traité dans l'ordre comme c'était dans les modèles précédents, par exemple, dans les modèles basés sur les réseaux neuronaux récurrents, ce qui permet une plus grande parallélisation et réduit donc le temps d'entraînement. L'architecture de Transformer, telle qu'elle est présente dans (VASWANI et al. 2017), a une structure encodeur-décodeur, où l'encodeur et le décodeur comprennent une série de couches dénommées *Multi-Head Attention (MHA)*, chacune suivie d'une couche de *Multi-Layer Perceptron (MLP)*. Chaque couche a une connexion résiduelle. Cette architecture prend comme entrée une série de mots, qui sont tokenisés et incorporés dans des vecteurs de taille d_{model} .

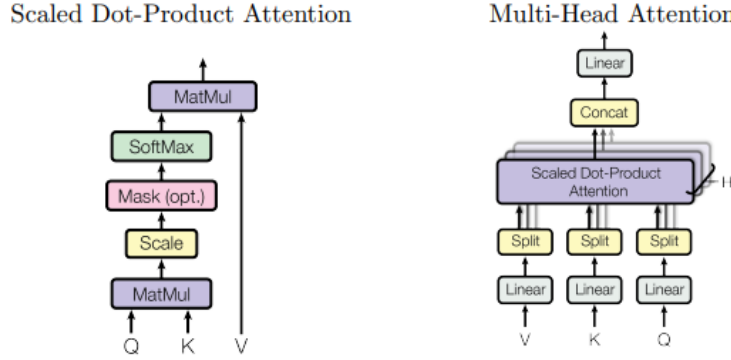


FIG. 2.13 : Schéma représentant les différentes étapes de scaled dotproduct attention et de multi-head attention (VASWANI et al. 2017).

- **Attention**

L'attention est une technique qui imite l'attention cognitive. Le but de la technique de l'attention consiste à se concentrer sur les parties importantes d'une entrée et à ignorer les informations qui ne sont pas aussi pertinentes. Une fonction d'attention mappe un vecteur de requête et un ensemble de paires de vecteurs clé-valeur à un vecteur de sortie. En particulier, l'attention portée par (VASWANI et al. 2017), appelé *Scaled Dot-Product Attention*, peut être formellement exprimé comme :

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.3)$$

où Q , K et V représentent respectivement l'ensemble de requêtes, de clés et de valeurs groupés en matrices, et d_k représente la dimension des requêtes et des clés, tandis que les valeurs ont la dimension d_v . Le produit est mis à l'échelle par $\frac{1}{\sqrt{d_k}}$ pour compenser le fait que les produits scalaires peuvent atteindre de grandes valeurs lorsque d_k est de grandes valeurs satureraient softmax et rendraient ses gradients très petits.

- **Multi-Head Attention**

Avant de passer Q , K et V à la fonction Attention, projettent linéairement, en utilisant des matrices appréhensibles, les entrées dans les vecteurs avec des dimensions d_k , d_k et d_v respectivement. En outre, au lieu de le faire une seule fois, ils le font h fois, et chaque projection est transmise simultanément à la fonction Attention, créant le soi-disant MultiHead Attention. Après le traitement parallèle, les matrices résultantes sont encadrées et projetées linéairement, en utilisant, encore une fois, une matrice appréhensible. Le traitement parallèle permet au modèle de traiter efficacement les informations provenant de différentes représentations des entrées de façon formelle.

$$MultiHead(Q, K, V) = Concat(head1, \dots, headh)W^O, \quad (2.4)$$

$$o, head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

$QW_i^Q \in \mathbb{R}^{d_{model}d_k}$, $KW_i^K \in \mathbb{R}^{d_{model}d_k}$, $VW_i^V \in \mathbb{R}^{d_{model}d_v}$ et $W^O \in \mathbb{R}^{hd_v}d_{model}$ sont les matrices appréhensibles utilisées pour projeter linéairement les entrées le résultat de

l'opération attention. Enfin, nous appelons le self-attention cas particulier où $K = V$, et – par analogie – *Multi-Head Self-Attention* (MSA) une *Multi-Head Attention* dans le cas où $K = V$.

- **Le bloc MLP et le bloc transformer global**

Après avoir calculé le self-attention pour les entrées, transmettent le résultat à un bloc Multi-Layer Perceptron (MLP) entièrement connecté avec une couche cachée. Formellement, étant donné une entrée x et des poids et biais apprenables $W1, b1, W2, b2$, et une fonction d'activation ρ , le bloc MLP peut être exprimé comme :

$$MLP(x) = \rho(xW1 + b1)W2 + b2 \tag{2.5}$$

utilisent l'unité linéaire rectifiée (ReLU) comme activation fonction. De plus, ils appliquent la normalisation de couche à la fois après les connexions résiduelles MSA et MLP (BA et al. 2016).

Pour résumer, étant donné une entrée x_l vers le l^{ime} bloc transformer, un bloc de MSA, un bloc MLP et une normalisation de couche bloc LN, le bloc transformateur global peut être exprimé comme suit :

$$\begin{aligned} x'_l &= LN(x_l + MSA(x_l)) \\ x_{l+1} &= LN(x'_l + MLP(x'_l)). \end{aligned} \tag{2.6}$$

La figure 2.14 représente la structure globale de l'architecture Transformer d'origine.

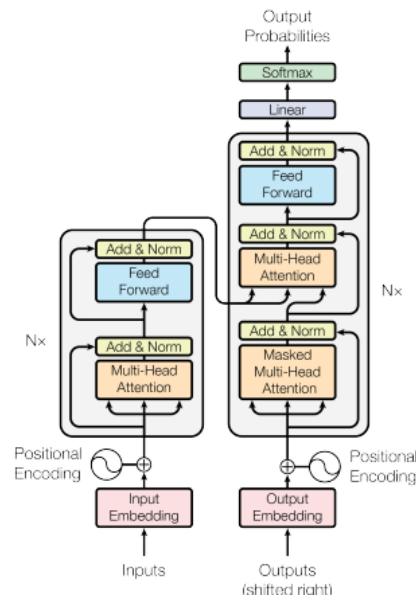


FIG. 2.14 : La structure globale de l'architecture Transformer d'origine (VASWANI et al. 2017).

- **Codage positionnel**

Nous notons que l'attention, en soi, considère l'entrée comme un ensemble et non

comme une séquence. Par conséquent, toutes les informations sur la position des entrées sont complètement perdues. Pour cette raison, ajoutent les soi-disant encodages positionnels aux incorporations d'entrée avant de les alimenter à l'encodeur et au décodeur. Les codages de position ont la même taille comme les incorporations des jetons d'entrée, c'est-à-dire, d_{model} .

Ils adoptent les fonctions sinus et cosinus de différentes fréquences, qui sont fixes (c'est-à-dire non apprises) :

$$\begin{aligned} PE_{pos,2i} &= \sin(pos/10000^{2i/d_{model}}) \\ PE_{pos,2i} &= \cos(pos/10000^{2i/d_{model}}), \end{aligned} \tag{2.7}$$

où i est la dimension et pos est la position.

2.4.3.1 Transformers pour NLP

Les transformers pour le NLP ont donné lieu à l'évolution de plusieurs modèles basés sur leur architecture. Les modèles abordés dans cette section comprennent le GPT (Generative Pre-trained Transformer), qui se concentre sur la génération de texte, et le BERT (Bidirectional Encoder Representations from Transformers), qui met l'accent sur les représentations d'encodeurs bidirectionnels.

- **Generative Pre-training (GPT)**

GPT développé par (RADFORD et al. 2018), est une méthode de pointe pour le traitement du langage naturel (NLP), GPT utilisé pour différentes tâches de NLP, telles que la génération de texte, la traduction automatique, la résumé de texte, etc. Il s'agit d'une pile de décodeurs de transformer. En d'autres termes, il est construit en utilisant des blocs décodeurs de transformer. Dans GPT, le module de multi-head attention avec attention croisée est supprimé du décodeur de transformer car il n'y a pas d'encodeur dans GPT. Par conséquent, les blocs décodeurs utilisés dans GPT ne comprennent que le codage de position, le module positional encoding, masked multihead self-attention et le réseau feedforward avec leurs fonctions d'addition, de normalisation de couche et d'activation. GPT est non supervisée qui permet d'être entraînée sur d'énormes quantités de données provenant d'Internet (GHOJOGH et al. 2020).

- **Bidirectional Encoder Representations from Transformers (BERT)**

Bidirectional Encoder Representations from Transformers (BERT) par (DEVLIN et al. 2018) ont utilisé les encodeurs dans un transformer comme une sous-structure pour les modèles d'entraînement préalable aux tâches de traitement du langage naturel (NLP) telles que l'analyse des sentiments, les questions-réponses, la résumé de texte etc. L'exécution de BERT pour ces tâches se fait en deux phases, à savoir l'entraînement préalable pour la compréhension du langage, et le réglage fin pour une tâche spécifique. BERT peut comprendre le langage en s'entraînant sur les mécanismes de Masked Language Modeling (MLM) et de Next Sentence Prediction

(NSP). BERT suppose un masque avec le MLM pour apprendre les contextes bidirectionnels dans les phrases. Ainsi, il prend en entrée des phrases aléatoires, masque certains mots dans les phrases et reconstruit les mots masqués à partir des textes environnants en sortie. Sa capacité à prendre en entrée deux phrases simultanément et à déterminer si la deuxième phrase vient après la première lui permet de réaliser la NSP. Cette capacité aide le modèle à maintenir des relations à longue distance entre les textes. BERT a été entraîné sur 16 Go de textes provenant des ensembles de données BooksCorpus et de l'anglais Wikipedia. Après l'entraînement préalable, le modèle est ensuite entraîné sur une tâche de NLP en effectuant un apprentissage supervisé sur un ensemble de données et en remplaçant la sortie entièrement connectée de BERT par un nouvel ensemble de couches de sortie. Le modèle BERT s'entraîne plus rapidement car les autres paramètres du modèle ne sont réglés finement que les paramètres de sortie appris à partir de zéro. BERT dispose de deux modèles, à savoir les modèles BERT-base et BERT-large (ACHEAMPONG et al. 2021). L'architecture générale du modèle BERT est illustrée dans la figure 2.15.

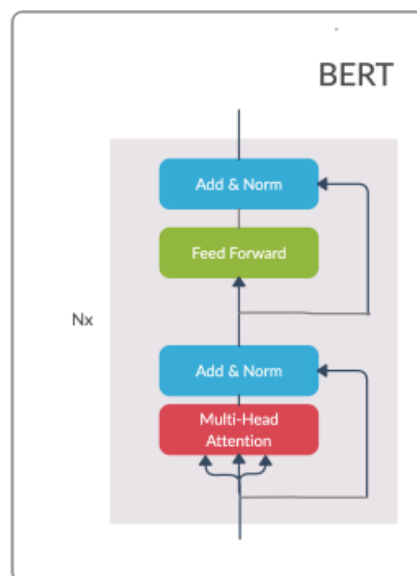


FIG. 2.15 : Architecture du modèle BERT (ACHEAMPONG et al. 2021).

2.4.3.2 Transformers pour la vision

En raison du succès massif des transformers dans le domaine du NLP, les transformers sont appliqués aux tâches de vision par ordinateur. Les transformers les plus récents utilisés pour la classification d'images comprennent Image GPT (iGPT) et Vision Transformers (ViT).

– Image GPT

Image GPT (CHEN et al. 2020), utilise GPT pour la génération d'images et entraîne le modèle sur ImageNet, mais il présente des limitations, car il nécessite une puissance de calcul élevée et offre une qualité d'image inférieure (USMAN et al. 2022).

– **Vision transformers (ViT)**

Le vision transformer est une architecture de réseau de neurones utilisée pour la classification d’images. Il permet de transformer une image en une séquence de vecteurs afin de capturer les relations à long terme entre les différentes parties de l’image. Cette approche révolutionnaire a montré des performances remarquables dans le domaine de la vision par ordinateur.

2.5 Vision transformers (ViT)

”An Image is Worth 16*16 Words : Transformers for Image Recognition at Scale” (DOSOVITSKIY et al. 2020) a été développé et publié par Neil Houlsby, Alexey Dosovitskiy, et d’autres auteurs de l’équipe de recherche de Google Brain. Il a introduit le modèle Vision Transformer, le ViT divise une image en fragments de taille fixe, les intègre correctement et inclut l’intégration de position en tant qu’entrée dans le codeur du transformateur, Sur la figure 2.16 un aperçu du modèle vision transformer.

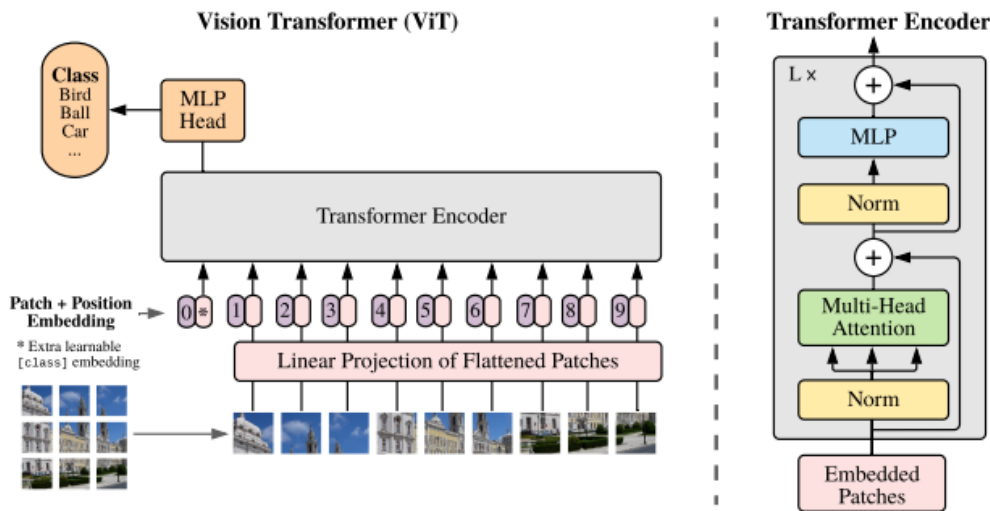


FIG. 2.16 : Aperçu du modèle vision transformer (ViT) (DOSOVITSKIY et al. 2020).

2.5.1 Structure de Vision Transformers (ViT)

La structure de l’architecture Transformer peut également être facilement adaptée aux tâches de vision par ordinateur (Dosovitskiy et al., 2021). L’architecture résultante s’appelle Vision Transformer (ViT). En particulier, il est possible de diviser les images d’entrée en patches non chevauchants, qui sont ensuite incorporés en tant que jetons. Le plus grand avantage de l’utilisation de l’attention pour les tâches de vision par ordinateur est qu’elle permet de mesurer à quel point une partie d’une image est attentive à une autre, ce qui permet de travailler davantage à un niveau

global plutôt qu'à un niveau local, comme le font les convolutions.

Il n'est pas nécessaire d'avoir une structure encodeur-décodeur : (DOSOVITSKIY et al. 2020) utilisent simplement la même structure que l'encodeur, c'est-à-dire l'incorporation des entrées, suivie d'une série de blocs composés d'un bloc MSA (Multi-Head Self-Attention) et d'un bloc MLP (Multi-Layer Perceptron). De plus, à la fin du dernier bloc, il y a ce qu'on appelle le "MLP Head".

- Couche Multi-Head Self-Attention (MSA) : Cette couche concatène linéairement toutes les sorties d'attention aux dimensions correctes. Les Multi-Head Self-Attention aident à entraîner des dépendances locales et globales dans une image.
- Couche de Multi-Layer Perceptron (MLP) : Cette couche contient deux couches avec l'unité linéaire d'erreur gaussienne (GELU).
- MLP Head : Un MLP qui prend en entrée un jeton spécial appelé "jeton [classe]" ([class] Token), résultant du dernier bloc d'attention, et le mappe à la classe prédite pour l'entrée.

Tokenisation d'entrée et encodage de position

Considérer chaque pixel comme un jeton et calculer l'attention entre chaque pixel serait impossible, car l'opération d'attention a une complexité $O(n^2)$ en termes de mémoire et d'exécution. Pour cette raison, (DOSOVITSKIY et al. 2020) divisent l'image en patchs d'entrée non superposés. Les patchs sont transformés en jetons en réduisant, de la taille des patchs, le nombre total d'entrées de l'opération d'attention. Les patchs sont intégrés en les projetant linéairement en vecteurs de dimension R_{model}^d . En pratique, la projection est efficacement mise en œuvre avec une couche de convolution ayant comme stride la taille du patch.

Une autre différence entre l'architecture proposée par (DOSOVITSKIY et al. 2020) et le Transformer original (VASWANI et al. 2017) réside dans le fait que les codages de position, au lieu d'être fixes et prédéterminés, sont apprenables.

Class token

Après la génération des jetons d'entrée, un vecteur est ajouté au début de la séquence de jetons et traité avec les autres jetons. Ce vecteur, appelé jeton [classe], est ensuite extrait des résultats du dernier bloc d'attention et est transmis à la tête de classification qui calcule la classe prédite pour l'entrée. L'état initial du vecteur (c'est-à-dire le vecteur qui est ajouté avant d'être traité) est un paramètre apprenable du modèle. Le jeton de classe est destiné à attirer l'attention sur les parties les plus pertinentes d'une image.

Il existe des modèles améliorés du Vision Transformer (ViT) qui ont été proposés par différents chercheurs, tels que SwinT (TUMMALA et al. 2022) et aussi Training data-efficient image transformers (DeiT) (TOUVRON et al. 2021). SwinT est une approche qui utilise des fenêtres décalées pour améliorer la représentation des informations spatiales dans les images. De son côté, DeiT vise à rendre l'entraînement

des modèles ViT plus efficace en utilisant des techniques d'apprentissage par transfert et de distillation de connaissances. Cette approche permet d'obtenir de bonnes performances même avec une quantité réduite de données d'entraînement.

2.5.2 Quelques applications de Vision Transformers

ViT est apparu comme une alternative compétitive aux réseaux de neurones convolutifs (CNN) qui sont actuellement à la pointe de la technologie en vision par ordinateur et largement utilisés pour différentes tâches de reconnaissance d'images.

Avec l'augmentation constante de la taille des ensembles de données et le développement continu de méthodes non supervisées et semi-supervisées, le développement de nouvelles architectures de vision qui s'entraînent plus efficacement sur ces ensembles de données devient de plus en plus essentiel. ViT est une étape préliminaire vers des architectures génériques et évolutives capables de résoudre de nombreuses tâches de vision. Par conséquent, le ViT a pris de l'importance dans les intérêts de recherche en raison de sa large applicabilité et de son évolutivité (SHAH 2022). Nous citons quelques des applications les plus importantes de ViT :

- **Classification des images** : Les CNN sont les méthodes les plus avancées pour la classification d'images, en capturant efficacement les informations locales grâce à l'utilisation de champs récepteurs restreints. Cependant, les ViT surpassent les CNN sur de grands ensembles de données en raison de leur capacité à traiter des informations globales.
- **Description d'image** : Les ViTs permettent de générer des légendes décrivant le contenu des images, offrant ainsi une catégorisation plus avancée. Ils apprennent une représentation générale des données plutôt que des étiquettes brutes, ce qui permet de générer du texte descriptif pour chaque image.
- **Segmentation des images** : DPT (DensePredictionTransformers) est un modèle de segmentation d'images développé par Intel en utilisant des vision transformers. Il atteint une précision de 49,02% mIoU pour la segmentation sémantique sur ADE20K, et améliore jusqu'à 28% les performances relatives par rapport à un réseau entièrement convolutionnel pour l'estimation de profondeur monoculaire.
- **Détection d'une anomalie** : La détection et la localisation d'anomalies d'images basées sur les transformers utilisent une approche de reconstruction et d'encodage par patch. Les réseaux transformers préservent les informations spatiales des patches intégrés, permettant ainsi de localiser les zones anormales à l'aide d'un réseau de densité de mélange gaussien.
- **Reconnaissance des actions** : Dans cet article intéressant (ARNAB et al. 2021) de l'équipe de recherche de Google, des modèles basés sur les transformers purs sont utilisés pour la classification vidéo, en s'appuyant sur leur succès récent dans la classification d'images. Le modèle extrait des jetons

spatio-temporels de la vidéo et les encode à l'aide de plusieurs couches de transformers pour reconnaître les actions. Malgré le besoin habituel de grands ensembles de données pour les modèles basés sur les transformers, les auteurs proposent des variantes efficaces du modèle qui permettent de régulariser l'entraînement et de tirer parti de modèles d'images pré-entraînés, ce qui rend possible l'entraînement sur des ensembles de données relativement petits.

- **Conduite autonome :** Lors du Tesla AI Day en 2021, Tesla a dévoilé les détails complexes du réseau de neurones utilisé dans Tesla FSD, mettant en avant un bloc de construction fascinant appelé "transformation image en BEV + fusion multi-caméras". Ce bloc intègre un module Transformer, plus précisément un module d'attention croisée, qui joue un rôle central dans le processus.

2.6 Etat de l'art sur le diagnostic automatique du cancer du sein

Cette section vise à donner un aperçu de ce que les chercheurs ont fait dans le domaine du diagnostic du cancer du sein.

2.6.1 Bases de données

Une base de données appelée également jeu de données ou dataset est une collection de données utilisées pour entraîner et tester un modèle d'apprentissage automatique. C'est un ensemble structuré ou non structuré d'échantillons de données qui sont utilisés pour résoudre un problème spécifique d'apprentissage automatique. Le jeu de données contient généralement un ensemble de caractéristiques ou de variables qui sont utilisées comme entrées pour l'algorithme d'apprentissage automatique afin de fournir une sortie ou une prédiction. Dans ce qui suit, nous allons présenter les bases de données les plus utilisées dans le domaine de diagnostic automatique du cancer du sein.

2.6.1.1 Base de données (Fondón et al. 2018)

La base de données d'images comprend un total de 156 images de coloration HE anonymes à haute résolution provenant de biopsies de tumeurs mammaires, qui ont été mises à disposition via le Bioimaging 2015 Grand Challenge. Les images ont toutes été numérisées avec les mêmes conditions d'acquisition et avec un grossissement de 200x. Tous ont une taille fixe de 2048×1536 et ont été stockés au format TIFF (Tagged Image File Format).

2.6.1.2 BreakHis

BreakHis c'est une base de données qui comprend des images microscopiques de biopsie pour des lésions bénignes et tumeurs malignes du sein avec un nombre total de 7909 images. Chaque lame de tumeurs mammaires est colorée à l'hématoxyline et l'éosine (HE). Pour obtenir les images numérisées du diapositives, un microscope système Olympus BX-50 avec un relais objectif avec un grossissement de $3,3 \times$ couplé à un appareil photo numérique couleur Samsung SCC-131AN est utilisé (BARDOU et al. 2018). Mammographie Dataset : Les images de cette base de données est tirés de la mammographie fournies par le premier hôpital de l'université médicale du Shanxi. Les lésions des glandes mammaires sont obtenues par un examen de mammographie numérique plein champ (FFDM). Un total de 2042 cas de maladie du sein sont généralement définis comme positifs pour la malignité et négatifs pour la bénigne. Après avoir été confirmé par les experts du premier hôpital de l'Université médicale du Shanxi, il y a 1011 cas négatifs et 1031 cas positifs, qui sont tous des patientes. Toutes les images de mammographie de la base de données ont été marquées par des experts (LI et al. 2019).

2.6.1.3 Digital Database for Screening Mammography DDSM

DDSM est une base de données complète de mammographies qui a été établi comme un projet collaboratif de de nombreuses universités de médecine et de génie médical de premier plan. DDSM se compose d'un nombre total de 2620 cas avec des classes normales, cancéreuses et bénignes. Les mammographies ont été scannées par les dispositifs médicaux (DBA, LUMYSIS et HOWTEK) avec différentes résolutions (ALTAN 2020).

2.6.1.4 Base de données (Zabit et al. 2020)

Zabit et al. Ont collecté au total 544 images de diapositives entières (WSI) de 80 patientes atteintes d'un cancer du sein dans le département de pathologie de l'Université Colsanitas Colombia, Bogotá, Colombie. Les coupes de tissus ont ensuite été numérisées à haute résolution ($400\times$) à l'aide d'un scanner Roche iScan HT. Ces images WSI représentant plusieurs cas de chaque patient ont été analysés à l'aide de H E, de récepteurs hormonaux, notamment ER, PR, HER2, myosine et Ki-67. Ensuite, deux pathologistes ont examiné les lames entières numériques de tissus colorés avec H E et extrait 845 zones de WSI, parmi lesquelles 408 ne sont pas des images de carcinome et 437 sont des images de carcinome. Les dimensions de ces images ont été notées comme 1278×760 pixels. Ce jeu de données est considéré comme équilibré et ses statistiques sont représentées. L'objectif principal lié à cet ensemble de données est la classification automatique de l'histopathologie du cancer du sein images, surtout les images de carcinome (ZABIT et al. 2020).

2.6.1.5 ISBI 2019 data set

Cette base de données contient 10661 images de cellules de 73 patients ont été sélectionnées, dont 7272 images de cellules lymphoblastiques B leucémiques (cellules cancéreuses) de 47 tous les patients et 3389 photos de précurseurs de lymphoïdes B (cellules normales) de 26 personnes en bonne santé. Ces cellules ont été segmentées à partir des images microscopiques, et chaque image de cellule est une image réelle après collecte. Certains bruits de coloration et erreurs d'éclairage générés pendant le processus de collecte ont été réparés dans une large mesure. La morphologie des cellules est très similaire, donc un oncologue professionnel annotera l'étiquette des images. Les étiquettes des images de cellules normales sont des échantillons positifs et les étiquettes des images de cellules cancéreuses sont des échantillons négatifs (JIANG et al. 2021).

2.6.2 Quelques travaux récents

Dans cette section, nous citons par ordre chronologique quelques travaux récents dans le domaine de diagnostic du cancer du sein.

2.6.2.1 Fondón et al. 2018

Dans ce travail, Fondon et al. Ont proposé un diagnostic assisté par ordinateur pour l'évaluation automatisée de la malignité d'échantillons de tissus mammaires basée sur le traitement d'images histologiques. Ils fournissent quatre tumeurs malignes comme sortie du système : normal, bénin, in situ et invasif. En tirant parti de techniques de traitement d'image bien établies, ils construisent un vecteur de caractéristiques pour chaque image qui sert d'entrée à un classificateur SVM (Support Vector Machine) avec un noyau quadratique. La méthode a été rigoureusement évaluée, d'abord avec une validation croisée 5 fois dans un ensemble initial de 120 images, deuxièmement avec un ensemble externe de 30 images différentes et troisièmement avec des images avec des artefacts inclus. Les niveaux de précision vont de 75,8 % sur une base de données d'images qui comprend un total de 156 images de coloration HE annotées anonymes à haute résolution provenant de biopsies de tumeurs mammaires.

2.6.2.2 Bardou et al. 2018

Dans ce travail, Bardou et al. Compèrent deux approches d'apprentissage automatique pour la classification automatique des images d'histologie du cancer du sein en bénignes et malignes. La première approche est basée sur l'extraction d'un ensemble de caractéristiques artisanales codées par deux modèles de codage (sac de mots et codage linéaire à contrainte de localité) et entraînées par des machines à vecteurs de support (SVM), tandis que la deuxième approche est basée sur les réseaux de neurones convolutionnels (CNN). Ils ont également testé expérimentalement des techniques d'augmentation de l'ensemble de données pour améliorer la précision

du réseau de neurones convolutionnels. Les résultats montrent que les réseaux de neurones convolutionnels surpassent le classifieur basé sur les caractéristiques artisanales, où ils ont obtenu une précision entre 96,15% et 98,33% sur la base de données BreakHis pour la classification binaire.

2.6.2.3 S. Khan et al. 2019

Dans ce travail, les auteures ont proposé un nouveau cadre de Deep Learning pour la détection et la classification du cancer du sein dans les images de cytologie mammaire utilisant le concept de transfer learning qui est une technique qui utilise les connaissances d'un trained model pour apprendre un autre ensemble de données. Transfer learning vise à améliorer l'apprentissage dans le domaine cible en tirant parti des connaissances du domaine source et de la tâche d'apprentissage. Dans le cadre proposé, les caractéristiques des images sont extraites à l'aide des Architectures CNN, à savoir GoogleNet, Visual Geometry Group Network (VGGNet) et Residual Networks (ResNet), qui sont introduits dans une couche entièrement connectée pour la classification des cellules malignes et bénignes à l'aide d'un classement moyen de mutualisation. Pour évaluer la performance du cadre proposé deux ensembles de données d'images microscopiques du sein sont utilisés : le premier est un ensemble de données de référence standard (BreakHis), et l'autre est développé localement à l'hôpital LRH de Peshawar, au Pakistan. L'architecture GoogleNet, VGGNet et ResNet donne individuellement une précision de classification moyenne de 93,5 %, 94,15 % et 94,35 % respectivement, tandis que le cadre proposé donne une précision de 97,525 %. Ces résultats montrent que la proposition atteint des performances élevées en termes de précision dans la détection et la classification des tumeurs du cancer du sein par rapport aux trois autres architectures.

2.6.2.4 Li et al. 2019

Li et al. Ont développé un modèle amélioré de réseau de neurones, appelé DenseNet-II, pour la classification précise et efficace des images de mammographie bénignes et malignes. Les images sont prétraitées pour éviter les interférences de lumière et le sur-apprentissage. Le modèle DenseNet-II remplace la première couche convolutive de DenseNet par la structure Inception. Les performances des différents modèles sont comparées, et les résultats montrent que le modèle DenseNet-II est plus performant que les autres modèles. La précision moyenne atteint 94,55% sur mammographie Dataset, ce qui prouve "selon les auteurs", la bonne généralisation et la robustesse du modèle.

2.6.2.5 Altan 2020

Dans ce travail, Altan et al. Ont propos une architecture de réseau de neurones convolutif(CNN) pour la classification de cas de cancer du sein sur des mammographies. Les mammographies sont des outils efficaces pour le diagnostic du cancer du sein, et les systèmes assistés par ordinateur basés sur l'apprentissage en profond

(DL) peuvent améliorer l'exactitude des évaluations en détectant des pathologies précoces. L'objectif du travail est d'évaluer l'applicabilité de différentes architectures CNN pour améliorer la capacité d'apprentissage des modèles DL pour le diagnostic opératoire du cancer du sein. Les performances de classification ont été évaluées en utilisant la base de données DDSM et le modèle proposé a atteint des taux de performance élevés de 92,84%, 95,30% et 96,72% pour l'exactitude, la sensibilité, la spécificité et la précision, respectivement.

2.6.2.6 Gour et al. 2020

Dans cette recherche, les chercheurs ont développé une approche automatisée pour le diagnostic des tumeurs cancéreuses du sein à l'aide d'images histologiques. Dans l'approche proposée, ils ont conçu un réseau neuronal convolutif à 152 couches basé sur l'apprentissage résiduel, nommé ResHist pour la classification des images histopathologiques du cancer du sein. Le modèle ResHist apprend des caractéristiques riches et discriminantes des images et des classifications histopathologiques en classes bénignes et malignes. De plus, pour améliorer les performances du modèle développé, ils ont conçu une augmentation de données. Le modèle ResHist proposé obtient une précision de 84,34% et un score F1 de 90,49% pour la classification d'images histopathologiques (BreakHis), et aussi, cette approche atteint une précision de 92,52 % et un score F1 de 93,45 % lorsque l'augmentation des données est utilisée.

2.6.2.7 Zabit et al. 2020

Les auteurs ont présente une approche d'apprentissage profond en ensemble pour la classification définitive des images d'histopathologie du cancer du sein en non carcinome et carcinome en utilisant leur propre ensemble de données collecté. Ils ont entraîné quatre modèles différents basés sur les architectures pré-entraînées VGG16 et VGG19. Initialement, ils ont suivi des opérations de validation croisée à 5 plis sur chacun des modèles individuels, à savoir les modèles VGG16 entièrement entraînés, VGG16 ajustés finement, VGG19 entièrement entraînés et VGG19 ajustés finement. Ensuite, ils ont Suivis une stratégie d'ensemble en prenant la moyenne des probabilités prédites et ils ont constaté que l'ensemble des modèles VGG16 ajustés finement et VGG19 ajustés finement offrait des performances de classification compétitives, notamment pour la classe carcinome. L'ensemble des modèles VGG16 ajustés finement et VGG19 a offert une sensibilité de 97,73% pour la classe carcinome sur leur propre ensemble de données collecté et une précision globale de 95,29%.

2.6.2.8 Kosare et al. 2021

Dans ce travail, Kosare et al. Ont développe un algorithme approfondi capable de détecter avec précision les tests de mammographie du cancer du sein à l'aide d'une approche d'apprentissage "de bout en bout" utilisant soit un ensemble de données d'apprentissage cliniquement défini, soit simplement le statut du cancer (étiquettes)

de l'image entière. Le diagnostic assisté par ordinateur proposé se compose de quatre parties : la reconstruction de mammographies, l'extraction de caractéristiques à l'aide d'un réseau sophistiqué, la détection de la qualité et enfin la fragmentation de la qualité à l'aide d'un réseau de neurones entièrement connecté (FC-NN). Ils ont testé sur l'ensemble de données BreakHis et ont atteint une précision de 88,7 %, 85,3 % et 88,4 % pour les données agrandies 40X, 100X, 200X et 400X, respectivement.

2.6.2.9 Jiang et al. 2021

Ce travail propose un modèle d'ensemble ViT-CNN pour classer les images de cellules cancéreuses et les images de cellules normales afin d'aider au diagnostic de la leucémie aiguë lymphoblastique. Le modèle d'ensemble ViT-CNN est un modèle d'ensemble qui combine le modèle de vision transformer et le modèle de réseau neuronal convolutif (CNN). Le modèle d'ensemble ViT-CNN peut extraire les caractéristiques des images de cellules de deux manières complètement différentes pour obtenir de meilleurs résultats de classification. De plus, l'ensemble de données utilisé dans ce travail est un ensemble de données déséquilibré et présente une certaine quantité de bruit, et ils ont proposé une méthode d'amélioration des données d'échantillonnage aléatoire à amélioration différentielle (DERS), créons un nouvel ensemble de données équilibré et utilisons la méthode symétrique fonction de perte d'entropie croisée pour réduire l'impact du bruit dans l'ensemble de données. La précision de classification du modèle d'ensemble ViT-CNN sur l'ensemble de test ISBI 2019 a atteint 99,03%, et il est prouvé par une comparaison expérimentale que le modèle ViT-CNN est le meilleur modèle que les autres modèles.

2.6.2.10 Alkathlan et al. 2022

Le travail de recherche proposé vise à développer une méthode pour prédire et classer le cancer du sein à un stade précoce. Dans cette recherche, trois modèles sont développés et leurs performances sont comparées les unes aux autres. Le premier modèle a été construit à l'aide de l'un des algorithmes d'apprentissage automatique (SVM), le deuxième modèle a été construit à l'aide d'un algorithme d'apprentissage en profond (CNN), et le troisième modèle combine les CNN avec une technique d'apprentissage par transfert pour offrir de meilleurs résultats. L'ensemble de données est fourni par la BC Histopathological Image Classification (BreakHis). Tous les modèles sont entraînés sur l'ensemble d'entraînement avec deux catégories principales : tumeur bénigne et tumeur maligne. La catégorie des tumeurs malignes est divisée en sous-ensembles de tumeurs carcinomes invasives et de tumeurs carcinomes in situ. De plus, les tumeurs carcinomateuses invasives sont classées en grade 1, grade 2 ou grade 3, le grade 3 étant le plus élevé et le plus agressif. Les résultats montrent que la précision de la classification des images de biopsie à l'aide de SVM est de 92 %, la précision de CNN est de 94 % et la précision de CNN utilisant la technique d'apprentissage par transfert est de 97 %.

2.6.2.11 Mayouf et al. 2022

Ce travail examine les aspects méthodologiques de la procédure d'entraînement des réseaux neuronaux pour la classification d'images médicales. La question de recherche concerne la conjecture selon laquelle feeding un réseau avec des bases de données pour l'augmentation de magnification leverages s'appuie sur des connaissances de haut niveau. Cette étude confirme cette hypothèse par le biais d'une expérience menée sur un ensemble de données d'images histopathologiques du cancer du sein. Les résultats présentés soulignent l'importance de l'ordre dans lequel les données sont introduites dans le réseau neuronal pendant la phase d'entraînement. Des expériences approfondies menées sur l'ensemble de données BreakHis démontrent que l'apprentissage incrémental atteint une précision de 98,76% pour la classification binaire (bénignes et malignes).

2.6.2.12 Nakach et al. 2022

Ce travail vise à étudier l'apprentissage d'ensemble et l'apprentissage par transfert pour la classification binaire des images histologiques sur les valeurs de facteur de grossissement de l'ensemble de données BreakHis. Les ensembles homogènes proposés sont mis en œuvre à l'aide d'une architecture hybride qui combine : (1) trois des techniques d'apprentissage en profond les plus récentes pour l'extraction de caractéristiques : DenseNet_201, MobileNet_V2 et Inception_V3, et (2) quatre des méthodes de boosting les plus populaires pour la classification : AdaBoost (ADB), Gradient Boosting Machine (GBM), LightGBM (LGBM) et XGBoost (XGB) avec Decision Tree (DT) comme apprenant de base. L'étude a évalué et comparé : (1) un ensemble d'ensembles de boosting conçus avec la même architecture hybride et un nombre différent d'arbres (50, 100, 150 et 200) ; (2) différentes méthodes de renforcement et (3) le classificateur DT unique avec les meilleurs ensembles de boosting. Les évaluations empiriques utilisées : quatre critères de performance de classification (exactitude, rappel, précision et score F1), la validation croisée quintuple, le test statistique de Scott Knott pour sélectionner le meilleur cluster des modèles les plus performants et le système de vote Borda Count pour classer le meilleur des plus performants. Le meilleur ensemble de boosting a atteint une valeur de précision de 92,52% sur la base de données (BreakHis) et il a été construit en utilisant XGB avec 200 arbres et Inception_V3 comme extracteur de caractéristiques.

2.6.2.13 Tummala et al. 2022

Dans ce travail, Tummala et al. ont examiné la capacité d'un ensemble de modèles SwinT pour la classification en deux classes de bénignes par rapport à malignes, ainsi que la classification en huit classes de quatre sous-types bénignes et quatre sous-types malignes, en utilisant un ensemble de données BreakHis disponible publiquement contenant 7909 images d'histopathologie acquises à différents facteurs de zoom de 40×, 100×, 200× et 400×. L'ensemble de modèles SwinT (comprenant les versions tiny, small, base et large) a démontré une précision moyenne de test de 96,0% pour la classification en huit classes et de 99,6% pour la classification en deux classes.

2.6.2.14 Alotaibi et al. 2023

Dans ce travail, Alotaibi et al. ont proposé une approche qui est un modèle d'ensemble de deux types de modèles pré-entraînés de vision transformer, à savoir Vision Transformer (ViT) et Data-Efficient Image Transformer (DeiT). Le modèle ViT-DeiT proposé classe les images histopathologiques du cancer du sein en huit catégories, dont quatre sont classées comme bénignes, tandis que les autres sont classées comme malignes. Le jeu de données public BreakHis a été utilisé pour évaluer le modèle proposé. Les résultats expérimentaux ont montré une accuracy de 98,17%, une précision de 98,18%, un rappel de 98,08% et un score F1 de 98,12%.

2.7 Conclusion

Dans ce chapitre nous avons d'abord présenté quelques logiciels de l'intelligence artificielle (IA) pour le diagnostic des maladies cancéreuses. Et par la suite nous avons donné un aperçu général sur les différentes techniques de l'IA telle que le machine learning et le deep learning ainsi que les Transformers et ses différents types. Enfin, nous avons présenté certaines réalisations récentes dans le domaine de diagnostic du cancer du sein.

Dans le chapitre suivant, nous utiliserons les Vision Transformers pour concevoir un système de classification des images histopathologiques du cancer du sein.

Chapitre 3

Un système basé-ViT pour la
classification des images
histopathologiques du cancer du
sein

3.1 Introduction

L'analyse histopathologique des échantillons de tissus est une méthode clé pour diagnostiquer le cancer du sein. Cependant, cette méthode repose principalement sur l'expertise des pathologistes pour interpréter les images histologiques, ce qui peut être sujet à des erreurs de diagnostic. Ainsi, améliorer la précision et la rapidité du diagnostic pour les images histologiques du cancer du sein est crucial pour obtenir de meilleurs résultats cliniques. L'objectif principal de notre travail est de construire un modèle adapté pour la classification des images histopathologiques avec une faible marge d'erreur. Par conséquent, ce chapitre présente un système de classification d'images histopathologiques du cancer du sein basé sur un modèle de Vision Transformers (ViT).

Ce chapitre est divisé en deux sections principales. Dans la première, nous abordons les différentes étapes de la mise en œuvre de notre modèle ViT. Dans la deuxième section, nous présentons et les expérimentations menées et les résultats obtenus.

3.2 Description du système de classification

Cette section est consacrée à la description de notre système de classification des images histopathologiques du cancer du sein. Avant, d'abordes les détails du système, nous donnons d'abord son architecture générale.

3.2.1 Architecture générale du système

L'architecture du système de classification des images histopathologiques suit des étapes essentielles. Le schéma bloc illustré dans la figure 3.1 montre cette architecture.

Notre démarche commence par l'utilisation d'une base d'apprentissage qui contient les images nécessaires pour l'entraînement du modèle ViT. Ensuite, nous procédons à l'étape de prétraitement des données pour améliorer la qualité des images. Par la suite, en se basant sur la technique de fine-tuning, nous entraînons le modèle sur une partie des données d'apprentissage traitées (l'autre partie est réservée pour la validation). Ce modèle généré sera utilisé ultérieurement (durant la phase de test ou de classification) pour classer des nouvelles images en deux catégories : Celles représentant une tumeur maligne et celles appartenant à la catégorie des tumeurs bénignes.

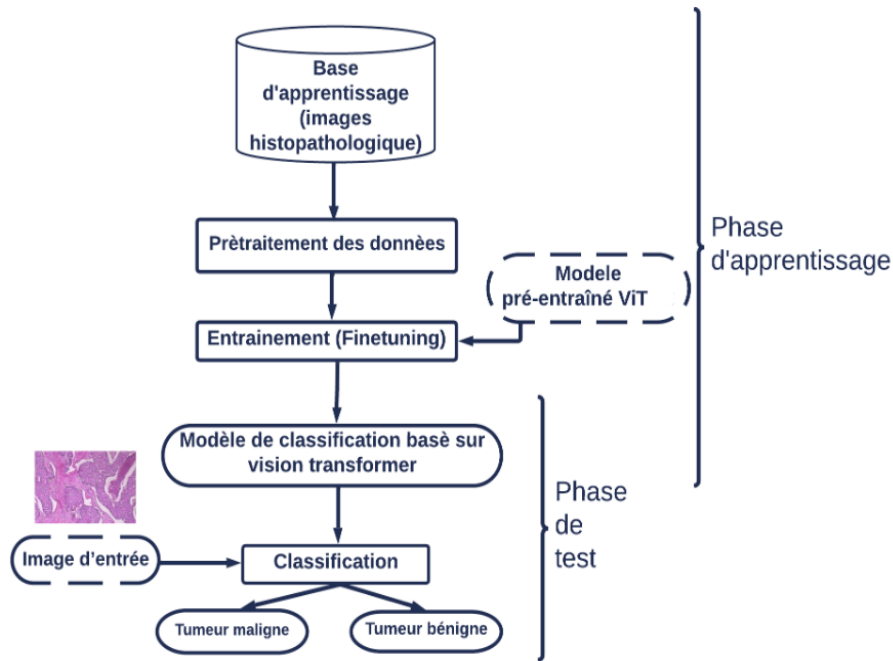


FIG. 3.1 : Représentation schématique de l'architecture de notre système de classification.

Après avoir présenté l'architecture globale de notre système, nous abordons en détail, dans la sous-section suivante, les différentes étapes nécessaires pour la mise en place de ce système.

3.2.2 Présentation détaillée du système

Afin de réaliser notre modèle, nous devons disposer de bases de données représentant les objets à étudier. Nous avons utilisé la base de données BreakHis, qui est composée de 7 909 images microscopiques cliniquement représentatives de tissus tumoraux du sein, collectées auprès de 82 patients. La base de données BreakHis sera présentée en détail plus loin dans la sous-section 3.1.

Dans ce qui suit, nous décrivons en détail les différentes phases constituant notre système.

3.2.2.1 Phase d'apprentissage

L'apprentissage consiste à générer le modèle ViT, qui sera par la suite utilisé pour la classification des images histopathologiques du cancer du sein.

- **Prètraitement des données** : Dans le domaine de l'apprentissage automatique, la qualité des données est d'une importance capitale, car elle influence directement le succès du projet. Par conséquent, le prétraitement des données est généralement considéré comme la première étape à effectuer dans un projet.

Dans notre système, nous réalisons des opérations de prétraitement appliquées aux images afin de les préparer pour l'entraînement du modèle. Ces opérations incluent des rotations horizontales et verticales aléatoires, l'application d'un filtre gaussien, le redimensionnement de l'image, le recadrage centré, la conversion en tenseur et la normalisation des valeurs des pixels. Ces opérations visent à améliorer la diversité des données, à préparer les images pour l'entraînement du modèle et à normaliser les valeurs des pixels pour une meilleure convergence de l'apprentissage.

- **Division des données** : Nous avons divisé cette base de données en trois sous-ensembles : l'ensemble d'entraînement, l'ensemble de validation et l'ensemble de test (train/val/test).
 - * L'ensemble d'entraînement (train) : Est un échantillon de l'ensemble de données utilisé pendant la phase d'entraînement du modèle. Il est utilisé pour ajuster les paramètres du modèle afin qu'il puisse classer correctement des nouvelles données jamais vues.
 - * L'ensemble de validation (validation) : Est un échantillon de l'ensemble de données utilisé pendant la phase de validation. Il est utilisé pour évaluer les performances du modèle à différents stades de l'entraînement et pour ajuster les hyperparamètres du modèle, tels que le taux d'apprentissage afin d'optimiser ses performances en évitant les problèmes d'overfitting et d'underfitting.
 - * L'ensemble de test (test) : Est un échantillon de l'ensemble de données utilisé pour tester et évaluer le modèle final. Il est utilisé pour mesurer les performances du modèle sur des données indépendantes et inconnues, afin de déterminer son efficacité dans des situations réelles.
- **Création des data loaders** : Dans cette étape, nous créons les dataloaders pour charger les données en lots (batch) pendant l'entraînement et l'évaluation du modèle. Cela facilite le traitement par lots des images et améliore l'efficacité du modèle.
- **Entraînement** : Dans un système de classification d'images histopathologiques utilisant le modèle ViT, l'objectif principal est de développer un système capable de classer différentes caractéristiques et motifs présents dans les images histopathologiques.

Les étapes principales de l'entraînement de notre modèle ViT sont les suivantes :

Initialisation du modèle ViT : Tout d'abord, nous initialisons le modèle ViT qui a été pré-entraîné sur l'ensemble de données ImageNet. Puis nous procédons à la réentraînement des dernières couches sur un ensemble de données plus petit et plus spécifique BreakHis, c'est ce qu'on appelle fine-tuning. Le réentraînement des dernières couches en fine-tuning permet au modèle de s'adapter aux particularités de la tâche spécifique tout en préservant les représentations générales de l'image apprises lors de pré-entraînement sur un

ensemble de données plus larges. De cette manière, les ViT sont capables de généraliser aux nouvelles tâches de vision par ordinateur (image histopathologique dans notre cas) sans nécessiter d'entraînement sur des ensembles de données massifs et coûteux.

Extraction des caractéristiques : Dans notre architecture ViT, le modèle ViT joué un rôle crucial dans l'extraction des caractéristiques à partir des images.

L'architecture générale de ViT se compose de deux parties principales : l'encodeur de patches (patch encoder) et l'encodeur transformer (transformer encoder).

- * **L'encodeur de patches :** L'encodeur de patches est une succession de deux couches : couche patch embedding et couche projection linéaire qui prétraitent l'image en entrée.

Couche patch embedding : Les images sont divisées en 256 patches de taille 16x16 (voir la figure 3.2), les patches sont non superposés qui sont ensuite incorporés en tant que jetons (Tokens).



FIG. 3.2 : Image divisées en 256 patches de taille 16x16.

Couche projection linéaire : Les patches sont passés en projection linéaire dans un espace vectoriel de dimension réduite pour obtenir une séquence de vecteurs représentant chacun un patch. Le ViT prenant en entrée des images en 2D, il est nécessaire d'aplatir l'image en une séquence 1D puis de faire une projection linéaire sur chaque séquence de patches avant de les alimenter dans une architecture transformer.

Cette technique permet de réduire le nombre de dimensions et de paramètres en entrée, ce qui facilite le traitement de chaque patch par les couches de l'encodeur transformer. Suite à cette projection, les vecteurs représentant les patches sont ensuite traités par les différentes couches de transformer, où des relations complexes entre les différents patches sont apprises, ainsi que les caractéristiques globales et les relations spatiales entre les patches. La représentation finalement obtenue permet alors de réaliser la tâche de classification

d'images.

- * **L'encodeur transformer** : L'encodeur transformer reçoit les vecteurs des patches et les fait passer à travers un ensemble de couches. Les couches d'encodeur transformer sont une pile de plusieurs couches transformer qui sont responsables d'apprendre les relations entre les patches de l'image. Chaque couche de Transformer se compose de deux sous-couches : Une couche de self-attention multi-tête (Multi-Head Attention) et une couche de feedforward. La couche de self-attention multi-tête permet au modèle d'apprendre les relations entre les patches à différentes échelles, tandis que la couche de feedforward apporte des fonctions non linéaires à chaque représentation de patch.

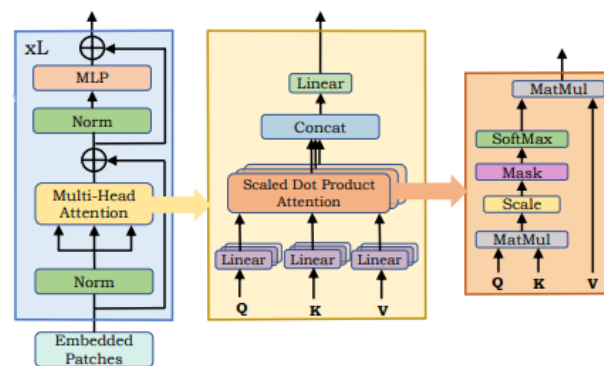


FIG. 3.3 : L'architecture de l'encodeur du transformer.

Pour optimiser les paramètres du modèle et améliorer la vitesse de l'entraînement, la sortie de l'encodeur transformer est d'abord transmise à une couche de normalisation de lot (batch normalization). Ensuite, une couche de classification linéaire (linear classifier) convertit la forme des données en vecteur unidimensionnel, suivi d'une couche Dropout qui est une technique de régularisation permettant de réduire le sur-apprentissage en empêchant le modèle de devenir trop dépendant de certaines caractéristiques de l'ensemble de données. Le bloc MLP dans la partie classification comporte une couche Dense avec 100 neurones, suivi par Dropout. Enfin, la sortie finale est produite grâce à une couche Dense avec 2 neurones, permettant la classification binaire de l'image en tumeur maligne ou tumeur bénigne. La sortie de cette couche est une probabilité qui représente la confiance du modèle dans sa classification. En d'autres termes, si la valeur de sortie de cette couche est proche de 1, le modèle est très sûr que l'image est une tumeur maligne, et si elle est proche de 0, il pense que c'est une tumeur bénigne.

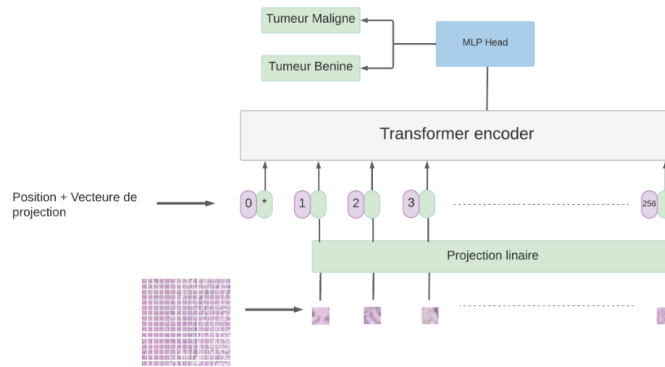


FIG. 3.4 : L'architecture globale de ViT.

Ajustement des paramètres du modèle : Pour entraîner notre modèle, nous utilisons une boucle d'entraînement qui itère sur les données d'entraînement pendant plusieurs époques. Pour chaque lot de données, nous effectuons les étapes suivantes :

- * Nous passons les images à travers le modèle pour obtenir les prédictions,
- * Nous calculons la perte en comparant les prédictions avec les étiquettes réelles à l'aide de la fonction de perte de cross-entropy qui est la fonction la plus adaptée aux problèmes de classification binaire (tumeur maligne ou bénigne),
- * Nous réalisons la rétropropagation (backpropagation) pour calculer les gradients et ajuster les poids du modèle. En ce qui concerne l'algorithme d'optimisation, nous sélectionnons l'optimiseur Adam. Nous définissons le pas d'apprentissage comme un hyperparamètre d'optimisation pour contrôler la vitesse et la direction des mises à jour des poids du modèle,
- * Nous mettons à jour les poids en utilisant l'algorithme d'optimisation choisi.

Après chaque époque d'entraînement, nous évaluons les performances du modèle sur l'ensemble de validation. Nous calculons l'accuracy pour évaluer la qualité des prédictions du modèle et prévoir le problème d'overfitting. Pour améliorer les performances du modèle nous avons ajusté les hyperparamètres en fonction des résultats de la validation, cette étape sera présentée dans la section 3.4.

3.2.2.2 Phase de test

Une fois l'entraînement terminé, nous évaluons le modèle final sur l'ensemble de test pour obtenir une estimation impartiale de ses performances sur des données inconnues. Chaque image de l'ensemble de test est passée dans le modèle pour obtenir des prédictions de classification. Les prédictions sont sous forme de probabilités pour chaque classe. Nous calculons les métriques

d'évaluation (accuracy) et analysons les résultats obtenus pour évaluer la performance et la capacité de généralisation du modèle.

3.3 Expérimentation résultats et discussion

Dans cette section, l'ensemble de données BreakHis, qui est utilisé pour l'apprentissage et l'évaluation du ViT est d'abord présentée. Ensuite, des expérimentations sont réalisées pour déterminer la combinaison optimale d'hyperparamètres. Enfin, les meilleurs résultats expérimentaux obtenus sont comparés aux ceux de différentes architectures CNN (InceptionV3, VGG19), ainsi qu'aux ceux de l'état de l'art du problème du diagnostic du cancer du sein.

3.3.1 Base de données BreakHis

La recherche scientifique n'est équitable que si les techniques appliquées sont basées sur les mêmes données. Étant donné que les différentes méthodes de collecte de données peuvent influencer la comparaison entre les différentes études, une base de données publique brésilienne a été publiée pour faciliter l'évaluation, notamment parce que la plupart de ces travaux sur les images histopathologiques du cancer du sein sont réalisés sur de petits ensembles de données. (SPANHOL et al. 2015) est une revue de l'ensemble de données BreakHis. Il est composé de 7 909 images microscopiques cliniquement représentatives de tissus tumoraux du sein, collectées auprès de 82 patients à l'aide de différents facteurs de grossissement ($40\times$, $100\times$, $200\times$ et $400\times$). À ce jour, il contient 2 480 échantillons bénignes et 5 429 échantillons malignes. L'ensemble de données comprend actuellement quatre types histologiques distincts de tumeurs bénignes du sein : adénose (A), fibroadénome (F), tumeur phyllode (PT) et adénome tubulaire (TA); ainsi que quatre tumeurs malignes (cancer du sein) : carcinome canalaire (DC), carcinome lobulaire (LC), carcinome mucineux (MC) et carcinome papillaire (PC).

3.3.2 Ajustement des hyperparamètres

Dans cette partie, nous présentons les résultats expérimentaux du modèle ViT pour la classification des images histopathologiques. Les expérimentations sont menées afin d'évaluer les performances de l'architecture et d'identifier les composants et les hyperparamètres qui permettent d'obtenir les meilleurs résultats. Les hyperparamètres étudiés sont la taille de batch, le taux de dropout, le pas d'apprentissage et le nombre d'époques.

3.3.2.1 Taille de batch

La taille de batch est un hyperparamètre utilisé dans l'apprentissage des modèles de machine learning. Il définit le nombre d'échantillons (images) qui sont regroupés

et traités ensemble lors de l'entraînement et de l'évaluation du modèle. Le tableau 3.1 montre les résultats obtenus des expérimentations réalisées pour analyser l'effet de la taille de batch sur le l'accuracy en utilisant différentes tailles de batch.

TAB. 3.1 : Résultats expérimentaux pour différentes tailles de batch.

Taille batch	Train Accuracy	Test accuracy
128	97,69%	97,17%
64	97,99%	97,70%
32	97,28%	97,69%
16	96,95%	97,12%
8	96,80%	97,23%

D'après le tableau 3.1, il est observable que les résultats obtenus pour la taille de batch sont assez similaires. Cependant, le meilleur accuracy est obtenu avec la taille 64 avec 97,99% sur les données d'apprentissage et 97,70% sur les données de test.

3.3.2.2 Taux du Dropout

L'idée principale du Dropout est de désactiver aléatoirement certains neurones du réseau de neurones pendant l'apprentissage. Cela force le modèle à apprendre à partir de différentes combinaisons de neurones et à ne pas dépendre excessivement de certains neurones spécifiques. Ainsi, le Dropout agit comme une forme de régularisation efficace, réduisant le risque de sur-apprentissage (Overfitting) et améliorant la capacité de généralisation du modèle. Dans cette étude, notre objectif principal est d'avoir un modèle ViT qui fonctionne bien à la fois sur les données d'apprentissage utilisées pour l'entraînement (la base d'apprentissage) et sur les nouvelles données sur lesquelles le modèle sera utilisé pour faire des prédictions (la base de test). Nous voulons que le modèle apprenne à partir des exemples connus et généralise à partir de ces exemples vers de nouveaux exemples inconnus. Cette expérimentation consiste à trouver le taux de la régularisation du Dropout à utiliser pour réduire l'overfitting et améliorer la généralisation de notre modèle ViT. Les résultats de cette expérimentation pour différents taux de Dropout sont présentés dans le tableau 3.2.

TAB. 3.2 : Résultats expérimentaux pour déterminer le taux de Dropout.

Dropout	Train Accuracy	Test Accuracy
0,3	98,00%	97,78%
0,5	97,34%	97,08%
0,7	95,63%	96,22%
0,9	90,58%	93,61%

Comme le montrent les résultats présentés dans le tableau 3.2, il est clairement observable que le Dropout de 0.3 est la meilleure mesure. Cette mesure signifie que seulement 30% de neurones sont éliminés à chacune des époques.

Il est également remarquable que l'accuracy sur les données d'apprentissage (données connues) 98% et l'accuracy sur les données de test (nouvelles données) est de 97,78%, ce qui signifie que notre modèle ViT possède une capacité de généralisation excellente.

3.3.2.3 Le pas d'apprentissage

Le pas d'apprentissage fait référence au pas auquel les poids du modèle sont mis à jour lors de l'entraînement et il peut être ajusté de différentes manières pour obtenir de meilleures performances de l'apprentissage. Les résultats obtenus en termes de l'accuracy pour différentes pas d'apprentissage sont reportés dans le tableau 3.3.

TAB. 3.3 : Résultats expérimentaux pour déterminer le pas d'apprentissage.

Pas d'apprentissage	Train Accuracy	Test accuracy
0.1	75,05%	88,02%
0.01	94,19%	90,36%
0.001	98,37%	97,78%
0.0001	93,00%	93,55%
0.00001	89,71%	89,41%

D'après les observations du tableau 3.3, on peut conclure que le pas d'apprentissage optimal pour obtenir les meilleurs résultats est de 0.001 (98,37% sur les données d'apprentissage et 97.78% sur les données de test). En examinant les autres valeurs du taux d'apprentissage, on constate que lorsqu'on s'éloigne de cette valeur, qu'il soit plus petit ou plus grand, les performances diminuent.

3.3.2.4 Le nombre d'époques d'apprentissage

Pour étudier le comportement de notre modèle durant la phase d'apprentissage en fonction du nombre d'époques, plusieurs expérimentations sont effectuées. Les résultats sont illustrés par les courbes de la figure 3.5.

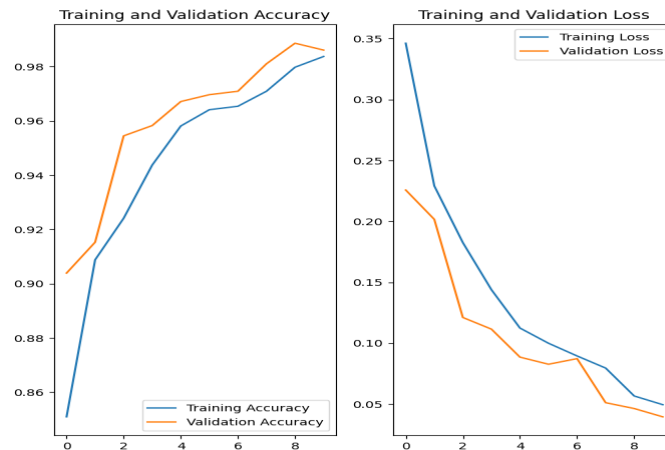


FIG. 3.5 : Courbes accuracy et loss obtenus.

Au cours de l'apprentissage, l'accuracy sur les données d'entraînement progresse rapidement en fonction du nombre d'époques avec une tendance logarithmique, atteignant un score de 0,85 avant de converger vers un score de 0,98. Quant à l'accuracy de validation, elle commence par un score de 0,90 et connaît une augmentation progressive pour atteindre un score de 0,98 à la fin de l'apprentissage (10 époques). En outre, la fonction de perte (Loss) démontre une diminution exponentielle à la fois pour les données d'entraînement et de validation, ce qui indique que le modèle s'améliore en réduisant l'erreur.

Ces résultats suggèrent que le système de classification est de bonnes performances sur les données d'entraînement et de validation, ce qui est un indicateur positif prouvant sa capacité à généraliser avec précision sur de nouvelles données.

3.3.3 Evaluation des performances

Le tableau 3.4 présente les performances en termes d'accuracy de notre modèle sur la base de données BreakHis. Les résultats du tableau confirment bien les constats des expérimentations précédentes en ce qui concerne la bonne généralisation sur les données de test jamais vues.

TAB. 3.4 : Accuracy de notre modèle sur la base de données BreakHis.

Images Histopathologiques	Train accuracy	Test accuracy	Validation accuracy
Taille de la base	6327	791	791
Accuracy	98,37%	97,78%	98,61%

Afin de détecter les erreurs de confusion potentielles de notre système, nous avons procédé au calcul de la matrice de confusion, comme illustré dans la figure 3.6. D'après cette matrice, la classe des images malignes a été correctement prédite pour 542 images, tandis que 10 images ont été étiquetées comme bénignes. Quant à la classe des images bénignes, le modèle a réussi à identifier 231 images correctement,

mais 8 images ont été faussement identifiées comme malignes. Dans l'ensemble, ces résultats indiquent que le modèle est globalement performant même s'il y a quelques confusions. Les faux positifs et les faux négatifs peuvent être liés à la similarité des caractéristiques visuelles des exemples confondus, rendant difficile leur distinction, même pour les experts humains.

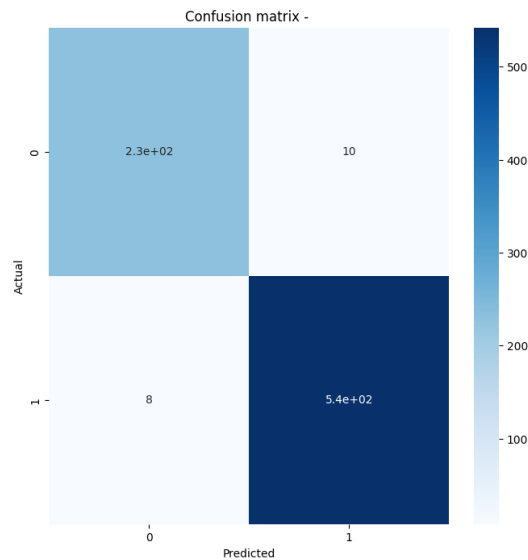


FIG. 3.6 : La Matrice de confusion.

3.3.4 Comparaison des résultats

Dans ce qui suit, nous comparerons les résultats de notre système avec ceux de deux architectures CNN ainsi qu'à ceux issus des travaux existants dans la littérature de la base de données BreakHis.

3.3.4.1 Avec des architectures CNN

Pour des fins de comparaison, nous avons implémenté, en plus du modèle ViT, deux architectures différentes de CNN : InceptionV3 et VGG19. Les résultats comparatifs sont résumés dans le tableau 3.5.

TAB. 3.5 : Comparaison de notre ViT avec différentes architectures CNN.

Modèles	Accuracy	Val accuracy	Loss	Val loss
InceptionV3	86,12%	88,06%	32,47%	30,70%
VGG19	68,57%	68,36%	62,26%	62,41%
ViT	97,78%	98,37%	4,59%	4,70%

Le tableau 3.5 montre que le modèle ViT nous a donné la meilleure valeur d'accuracy (97,78%), ce qui nous permet de dire que c'est la meilleure approche à adopter pour résoudre le problème de classification des images histopathologiques du cancer du sein.

3.3.4.2 Avec des travaux de la littérature

Dans le but d'évaluer les performances de notre système par rapport à d'autres études existantes, nous avons effectué une comparaison de son accuracy avec certains travaux récents menés sur la base de données BreakHis. Les résultats de cette comparaison sont rapportés dans le tableau 3.6.

TAB. 3.6 : Comparaison des performances avec l'état de l'art.

Système	Technique	Accuracy
GOUR et al. 2020	CNN	92,52 %
ALKHATHLAN et al. 2022	SVM	92 %
MAYOUF et al. 2022	CNN	98,76%
NAKACH et al. 2022	hybride	92,52%
TUMMALA et al. 2022	SwinT	99,6%
ALOTAIBI et al. 2023	ViT-DeiT	98,17%
Notre système, 2023	ViT	97,78%

Et pour plus de lisibilité, nous avons présenté les résultats obtenus graphiquement dans la figure 3.7. Les systèmes sont triés de gauche à droite du moins performant au plus performant.

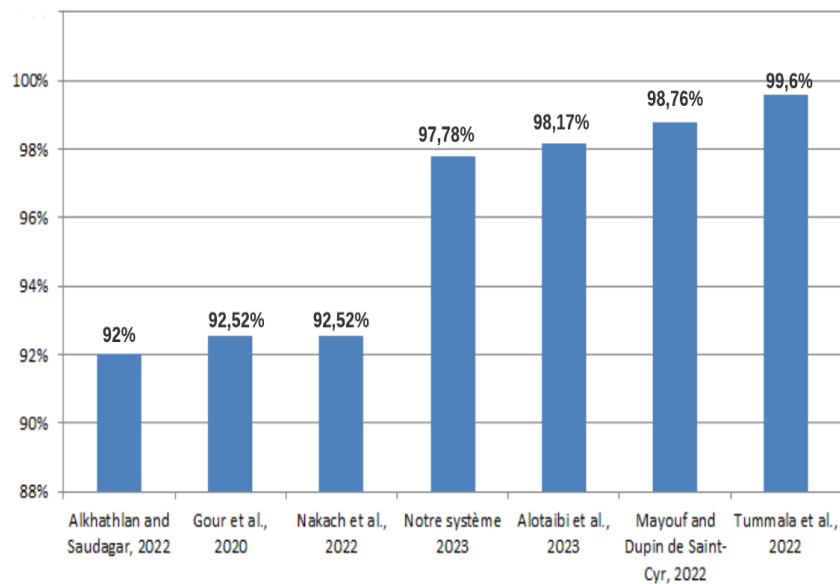


FIG. 3.7 : Graphique à barres illustrant la comparaison de l'accuracy sur la base de données BreakHis.

D'après le graphique, il est remarquable que le meilleur résultat est obtenu pour Swin Transformer (TUMMALA et al. 2022) avec une accuracy de 99,6%. Il est, également, intéressant de remarquer que notre système donne des résultats compétitifs très encourageants par rapport aux meilleurs résultats rapportés avec une accuracy de

97,78%. De plus, à partir des résultats, nous pouvons constater que les techniques basées sur les Transformers donnent toujours des performances meilleures que celles des CNN et des algorithmes d'apprentissage classique.

Ces techniques, le plus souvent utilisées avec des Vision transformer, sont capables de se classer en tête du classement absolu dans l'ensemble de données BreakHis et pour la classification des images en général.

3.4 Conclusion

Dans ce chapitre nous avons présenté les différentes étapes de mise en œuvre de notre modèle ViT réalisant la classification des images histopathologiques du cancer du sein. De plus, nous avons analysé l'effet de différentes paramétrisations sur l'accuracy du modèle. Pour cette tâche, une série d'expérimentations a été réalisée, ce qui nous a permis de fixer les valeurs optimales des hyper paramètres de notre modèle ViT. Nous l'avons évalué sur la base de données BreakHis et avons comparé ses performances avec d'autres travaux récents. Les résultats expérimentaux ont montré que notre système basée-transformers est efficace pour la classification des images histopathologiques du cancer du sein et prometteur pour une utilisation clinique future.

Conclusion et perspectives

Conclusion générale

Bilan

Les systèmes de diagnostic du cancer du sein jouent un rôle crucial dans la détection précoce, la précision du diagnostic, la prise de décision médicale, la réduction des délais de traitement et l'élargissement de l'accès aux soins de santé. Leur utilisation permet d'améliorer les résultats cliniques et de sauver des vies en fournissant des informations précises et en facilitant la mise en place de traitements appropriés.

Notre travail s'inscrit dans le cadre de la classification des images histopathologiques du cancer du sein. Nous nous intéressons en particulier, à améliorer les performances des systèmes de classification des images histopathologiques du cancer du sein pour la base de données BreakHis. La difficulté de ce sujet a conduit de nombreux chercheurs à entreprendre de nombreux travaux pour aborder le problème de la classification. Cependant, la recherche demeure loin d'atteindre le niveau de la capacité humaine dans ce domaine.

Dans ce cadre, notre focus s'est concentré sur la classification des images histopathologiques du cancer du sein en utilisant les ViTs qui sont une technique très récente de l'intelligence artificielle qui a révolutionné le domaine de la vision par ordinateur.

Afin d'explorer les capacités du modèle ViT, une étude expérimentale a été menée dans le but est d'identifier les hyper paramètres offrant les meilleurs résultats en termes d'accuracy. À partir de cette étude, nous avons pu définir un système extrêmement performant, atteignant un taux de reconnaissance de 97,78% pour la base de données BreakHis.

Afin d'évaluer le système basé-ViT proposé, nous l'avons d'abord comparé, en termes d'accuracy, avec différentes architectures CNN, à savoir, InceptionV3 et VGG19. Ensuite, nous avons comparé les résultats obtenus avec ceux de certains travaux récents utilisant la même base de données. À la suite de cette comparaison, nous avons pu conclure que notre système se positionne favorablement parmi les meilleurs, démontrant ainsi l'efficacité de ViT dans la classification des images histopathologiques du cancer du sein.

Perspectives

En ce qui concerne nos perspectives pour les travaux futurs, de nombreuses perspectives sont envisageables pour l'amélioration et la poursuite de nos travaux, dont les plus importantes sont les suivantes :

- Nous envisageons d'améliorer la généralisation du modèle de classification des images histopathologiques du cancer du sein en utilisant le data augmentation.

- En plus de ViT qui a déjà démontré son efficacité, il serait intéressant d’explorer d’autres modèles basés transformers tels que ImageGPT. Ces modèles peuvent apporter de nouvelles perspectives en termes de représentation et de capacités de classification, ouvrant ainsi de nouvelles possibilités pour améliorer la précision du diagnostic.
- Une autre perspective intéressante serait de développer des modèles capables de classifier les sous-types spécifiques de tumeurs du sein. En identifiant les différentes caractéristiques des sous-types de tumeurs, il serait possible d’obtenir des diagnostics plus précis, ce qui pourrait avoir un impact significatif sur les options de traitement et les résultats pour les patients.

Bibliographie

- ACHEAMPONG, Francisca Adoma, Henry NUNOO-MENSAH et Wenyu CHEN (2021). “Transformer models for text-based emotion detection : a review of BERT-based approaches”. In : *Artificial Intelligence Review*, p. 1-41.
- ADANKON, Mathias M et Mohamed CHERIET (2009). “Model selection for the LS-SVM. Application to handwriting recognition”. In : *Pattern Recognition* 42.12, p. 3264-3270.
- AKKOUICHE, Salima (2019). “Cancer du sein : Les inquiétantes statistiques des spécialistes”. In : *algerie360*.
- ALKHATEEB, Jawad H et al. (2011). “Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition”. In : *knowledge-based systems* 24.5, p. 680-688.
- ALKHATHLAN, Lina et Abdul Khader Jilani SAUDAGAR (2022). “Predicting and Classifying Breast Cancer Using Machine Learning”. In : *Journal of Computational Biology* 29.6, p. 497-514.
- ALOTAIBI, Amira et al. (2023). “ViT-DeiT : An Ensemble Model for Breast Cancer Histopathological Images Classification”. In : *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*. IEEE, p. 1-6.
- ALTAN, Gokhan (2020). “Deep learning-based mammogram classification for breast cancer”. In : *International Journal of Intelligent Systems and Applications in Engineering* 8.4, p. 171-176.
- ARNAB, Anurag et al. (2021). “Vivit : A video vision transformer”. In : *Proceedings of the IEEE/CVF international conference on computer vision*, p. 6836-6846.
- ARULKUMARAN, Kai et al. (2017). “A brief survey of deep reinforcement learning”. In : *arXiv preprint arXiv :1708.05866*.
- AZAMJAH, Nasrindokht, Yasaman SOLTAN-ZADEH et Farid ZAYERI (2019). “Global trend of breast cancer mortality rate : a 25-year study”. In : *Asian Pacific journal of cancer prevention : APJCP* 20.7, p. 2015.
- BA, Jimmy Lei, Jamie Ryan KIROS et Geoffrey E HINTON (2016). “Layer normalization”. In : *arXiv preprint arXiv :1607.06450*.
- BARDOU, Dalal, Kun ZHANG et Sayed Mohammad AHMAD (2018). “Classification of breast cancer based on histology images using convolutional neural networks”. In : *Ieee Access* 6, p. 24680-24693.
- BROWNLEE, Jason (2017). “A gentle introduction to long short-term memory networks by the experts”. In : *Machine Learning Mastery* 19.
- CHEN, Mark et al. (2020). “Generative pretraining from pixels”. In : *International conference on machine learning*. PMLR, p. 1691-1703.

- DEVLIN, Jacob et al. (2018). “Bert : Pre-training of deep bidirectional transformers for language understanding”. In : *arXiv preprint arXiv :1810.04805*.
- DOSOVITSKIY, Alexey et al. (2020). “An image is worth 16x16 words : Transformers for image recognition at scale”. In : *arXiv preprint arXiv :2010.11929*.
- EGGERTSON, Laura (2004). “MRIs more accurate than mammograms but expensive”. In : *CMAJ* 171.8, p. 840-840.
- FONDÓN, Irene et al. (2018). “Automatic classification of tissue malignancy for breast carcinoma diagnosis”. In : *Computers in biology and medicine* 96, p. 41-51.
- FRANCO ORTELLADO, Blas Manuel et al. (2019). “Applications of artificial neural networks in three agro-environmental systems : microalgae production, nutritional characterization of soils and meteorological variables management”. In.
- GHOJOGH, Benyamin et Ali GHODSI (2020). “Attention mechanism, transformers, BERT, and GPT : Tutorial and survey”. In.
- GOUR, Mahesh, Sweta JAIN et T SUNIL KUMAR (2020). “Residual learning based CNN for breast cancer histopathological image classification”. In : *International Journal of Imaging Systems and Technology* 30.3, p. 621-635.
- GUO, Yanming et al. (2016). “Deep learning for visual understanding : A review”. In : *Neurocomputing* 187, p. 27-48.
- HADJIDJ, Ismahan (2018). “Analyse des Images Mammographiques pour l’Aide à la Détection du Cancer du Sein.” Thèse de doct.
- HE, Kaiming et al. (2016). “Deep residual learning for image recognition. CVPR. 2016”. In : *arXiv preprint arXiv :1512.03385*.
- HU, Junyan et al. (2020). “Voronoi-based multi-robot autonomous exploration in unknown environments via deep reinforcement learning”. In : *IEEE Transactions on Vehicular Technology* 69.12, p. 14413-14423.
- JIANG, Zhencun et al. (2021). “Method for diagnosis of acute lymphoblastic leukemia based on ViT-CNN ensemble model”. In : *Computational Intelligence and Neuroscience* 2021.
- KHAN, Asifullah et al. (2020). “A survey of the recent architectures of deep convolutional neural networks”. In : *Artificial intelligence review* 53, p. 5455-5516.
- KHAN, SanaUllah et al. (2019). “A novel deep learning based framework for the detection and classification of breast cancer using transfer learning”. In : *Pattern Recognition Letters* 125, p. 1-6.
- KOSARE, Neha, Gayatri BHOYAR et Neetu GYANCHANDANI (2021). “Breast Cancer Detection using Deep Learning”. In.
- KRIZHEVSKY, Alex, Ilya SUTSKEVER et Geoffrey E HINTON (2017). “Imagenet classification with deep convolutional neural networks”. In : *Communications of the ACM* 60.6, p. 84-90.
- KUMAR, N (2021). *Introduction to Support Vector Machines (SVMs)*.
- LI, Hua et al. (2019). “Benign and malignant classification of mammogram images based on deep learning”. In : *Biomedical Signal Processing and Control* 51, p. 347-354.
- MACARY, Manon (2022). “Analyse de données massives en temps réel pour l’extraction d’informations sémantiques et émotionnelles de la parole”. Thèse de doct. Le Mans Université.

- MAHESH, Batta (2020). “Machine learning algorithms-a review”. In : *International Journal of Science and Research (IJSR).[Internet]* 9, p. 381-386.
- MATHIEU-DUPAS, Eve (2010). “Algorithme des k plus proches voisins pondérés et application en diagnostic”. In : *42èmes Journées de Statistique*.
- MAYOUF, Mouna Sabrine et Dupinde SAINT-CYR (2022). “Curriculum Incremental Deep Learning on BreakHis DataSet”. In : *Proceedings of the 2022 8th International Conference on Computer Technology Applications*, p. 35-41.
- MAZAC, Sebastien (2015). “Approche décentralisée de l'apprentissage constructiviste et modélisation multi-agent du problème d'amorçage de l'apprentissage sensorimoteur en environnement continu : application à l'intelligence ambiante”. Thèse de doct. Université Claude Bernard-Lyon I.
- MEBARKI, Menseur (2020). “Détection De Pathologies Mammaires Pour L'aide À L'interprétation”. Mém. de mast. Université Saad Dahleb - Blida.
- MIAO, Shumei et al. (2018). “Extraction of BI-RADS findings from breast ultrasound reports in Chinese using deep learning approaches”. In : *International journal of medical informatics* 119, p. 17-21.
- MOGHADDAMNIA, A et al. (2009). “Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques”. In : *Advances in Water Resources* 32.1, p. 88-97.
- NAKACH, Fatima-Zahrae, Hasnae ZEROUAOUI et Ali IDRI (2022). “Hybrid deep boosting ensembles for histopathological breast cancer classification”. In : *Health and Technology*, p. 1-18.
- NASRI (2022). “Breast Cancer Histopathological Image Classification using Deep Learning”. Mém. de mast. University of Constantine.
- PEROU, Charles M et al. (2000). “Molecular portraits of human breast tumours”. In : *nature* 406.6797, p. 747-752.
- RADFORD, Alec et al. (2018). “Improving language understanding by generative pre-training”. In.
- SAMUEL, Arthur L (1959). “Machine learning”. In : *The Technology Review* 62.1, p. 42-45.
- SHAFIE, Abdulkadir Hassan (2019). *Histopathology Made easy, 1st Editi.*
- SIKANDAR, Arbab (2018). “Histopathology : An old yet important technique in modern science”. In : *Histopathology-An Update*, p. 1.
- SIMONYAN, Karen et Andrew ZISSERMAN (2014). “Very deep convolutional networks for large-scale image recognition”. In : *arXiv preprint arXiv :1409.1556*.
- SPANHOL, Fabio A et al. (2015). “A dataset for breast cancer histopathological image classification”. In : *Ieee transactions on biomedical engineering* 63.7, p. 1455-1462.
- SUEIRAS, Jorge (2021). “Continuous Offline Handwriting Recognition using Deep Learning Models”. In : *arXiv preprint arXiv :2112.13328*.
- SUTTON, Richard S, Andrew G BARTO et al. (1998). *Introduction to reinforcement learning*. T. 135. MIT press Cambridge.
- SZEGEDY, Christian et al. (2015). “Going deeper with convolutions”. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 1-9.

- TOUVRON, Hugo et al. (2021). “Training data-efficient image transformers & distillation through attention”. In : *International conference on machine learning*. PMLR, p. 10347-10357.
- TUMMALA, Sudhakar, Jungeun KIM et Seifedine KADRY (2022). “BreaST-Net : Multi-Class Classification of Breast Cancer from Histopathological Images Using Ensemble of Swin Transformers”. In : *Mathematics* 10.21, p. 4109.
- TYAGI, Amit Kumar et G REKHA (2020). “Challenges of applying deep learning in real-world applications”. In : *Challenges and applications for implementing machine learning in computer vision*. IGI Global, p. 92-118.
- USMAN, Mohammad, Tehseen ZIA et Ali TARIQ (2022). “Analyzing transfer learning of vision transformers for interpreting chest radiography”. In : *Journal of digital imaging* 35.6, p. 1445-1462.
- VASWANI, Ashish et al. (2017). “Attention is all you need”. In : *Advances in neural information processing systems* 30.
- VERBEKE, Stéphanie (2010). “Étude des voies de signalisation du récepteur p75^{ntr} impliquées dans la croissance des cellules de cancer du sein”. Thèse de doct. Lille 1.
- WANG, Lulu (2017). “Early diagnosis of breast cancer”. In : *Sensors* 17.7, p. 1572.
- WÖCKEL, A et al. (2018). “Früherkennung, Diagnostik, Therapie und Nachsorge des primären Mammakarzinoms”. In : *Dtsch Arztebl Int* 115.18, p. 316-323.
- ZABIT et al. (2020). “Breast cancer histopathology image classification using an ensemble of deep learning models”. In : *Sensors* 20.16, p. 4373.
- ZEILER, Matthew D (2013). “Hierarchical convolutional deep learning in computer vision”. Thèse de doct. New York University.
- ZOUNGRANA, Wend-Benedo Arnaud Bienvenue (2020). “Application des algorithmes d’apprentissage automatique pour la détection de défauts de roulements sur les machines tournantes dans le cadre de l’Industrie 4.0”. Thèse de doct. Université du Québec à Chicoutimi.

Webographie

Web 01 : Les taux d'incidence du cancer du sein dans le monde. Retrieved June 8 ,2023. website : <https://www.komen.org/breast-cancer>

Web 02 : Breast cancer. Retrieved June 8 ,2023. website : <https://www.who.int/news-room/factsheets/detail/breast-cancer>

Web 03 : Qu'est-ce que le cancer du sein?. Retrieved June 8 ,2023. website : <https://cancer.ca/fr/cancer-information/cancer-types/breast/what-is-breast-cancer>

Web 04 : kaggle - Breakhis Datasets . Retrieved June 8 ,2023. website : <https://www.kaggle.com/datasets/ambarish/breakhis>

Web 05 : Classification BI-RADS. Retrieved June 8 ,2023. website : <http://www.depistagesein.ca/classification-bi-rads/>

Web 06 : Pathai - Improving Patient Outcomes with AI-Powered Pathology. Retrieved June 8 ,2023. website : <https://www.pathai.com>

Web 07 : proscia - The data to fight cancer is in this image. Retrieved June 8 ,2023. website : <https://proscia.com>

Web 08 : Freenome - Decoding the means to cancer's end Retrieved June 8 ,2023. website : <https://www.freenome.com>

Web 09 : Colaboratory. Retrieved June 8 ,2023. website : <https://research.google.com/colaboratory/faq.html?hl=fr>

Web 10 : Python. Retrieved June 8 ,2023. <https://www.datarockstars.ai/glossary/python/>

Web 11 : la bibliothèque Python la plus utilisée en Data Science. Retrieved June 8 ,2023. website : <https://datascientest.com/numpy/>

Web 12 : Pandas. Retrieved June 8 ,2023. website : <https://datascientest.com/pandas-python-data-science>

Web 13 : A SCIENTIFIC COMPUTING FRAMEWORK FOR LUJIT. Retrieved June 8 ,2023. website : <http://torch.ch/>

Web 14 : Matplotlib. Retrieved June 8 ,2023. website : <https://datascientest.com/matplotlib-tout-savoir>

Web 15 : Seaborn. Retrieved June 8 ,2023. website : <https://datascientest.com/seaborn-tout-savoir>

Annexe

Annexe A

Outils d'implémentation

A.1 Introduction

Dans cette annexe, nous présentons d'abord les outils de développements, puis nous décrivons les différentes étapes pour la mise en œuvre du système de classification des images histopathologiques du cancer du sein.

A.2 La plateforme et environnements de développement

Afin d'obtenir des résultats concrets et évolutifs dans notre étude, nous avons fait un usage intensif de Google Colab, également connu sous le nom de Colab. Colab est un service cloud proposé par Google qui permet de reproduire l'environnement Jupyter Notebook dans le cloud. Il met à disposition de nombreux utilisateurs des processeurs graphiques (GPUs) puissants, ce qui est particulièrement avantageux pour ceux qui ne peuvent pas se permettre de former des projets d'apprentissage automatique à distance (Web 09). Le tableau ci-dessous présentent les caractéristiques de l'environnement :

TAB. A.1 : Caractéristiques de l'environnement.

CPU Model	2-core Xeon 2.2GHz
RAM	25 GB
Disque dur	225 GB

A.2.1 Présentation du langage Python

Python est un langage de programmation open source appartenant à la catégorie des langages interprétés. Il permet aux développeurs de se concentrer sur les actions qu'ils réalisent plutôt que sur les différentes méthodes pour les exécuter. Comparé

aux langages compilés, Python offre aux programmeurs un gain de temps considérable. Il est largement utilisé, apprécié et très demandé dans le domaine de l'informatique. Python est un langage à typage dynamique, ce qui signifie que le type d'une variable peut être modifié. Les annotations sont utilisées pour spécifier les types des arguments et des valeurs de retour des fonctions. Python est couramment utilisé pour le développement web, l'analyse de données, l'intelligence artificielle, les réseaux de neurones, le calcul scientifique et d'autres domaines de l'informatique avancée. La programmation est sans aucun doute le principal domaine d'utilisation de ce langage, qui convient aussi bien aux projets simples qu'aux projets complexes (Web 10).

A.3 Les bibliothèques utilisées

Ces bibliothèques ont été utilisées pour la mise en œuvre et l'analyse des modèles de classification des images histopathologiques du cancer du sein basés sur les Transformers, ainsi que pour la manipulation des données et l'évaluation des résultats.

A.3.1 NumPy

NumPy, qui est l'abréviation de "Numerical Python", est une bibliothèque utilisée dans la programmation scientifique en Python, principalement pour manipuler des données numériques. Elle offre des structures de données multidimensionnelles sous forme de tableaux, ainsi qu'un ensemble d'outils intégrés pour faciliter l'implémentation en Python. NumPy combine essentiellement les avantages du langage C et de Python, en traitant les données numériques sous forme de tableaux pour effectuer des opérations multidimensionnelles et des manipulations de données (Web 11).

A.3.2 Pandas

Pandas est une bibliothèque Python utilisée pour l'analyse de données. Son développement a été motivé par la nécessité d'avoir un outil puissant et flexible pour l'analyse quantitative, ce qui lui a permis de devenir l'une des bibliothèques Python les plus populaires. Pandas bénéficie d'une communauté de contributeurs très active, qui contribue régulièrement à son amélioration et à son évolution (Web 12).

A.3.3 TorCh

Torch est un cadre de calcul scientifique avec un large soutien pour les algorithmes d'apprentissage automatique qui privilégie les GPU. Il est facile à utiliser et efficace. L'objectif de Torch est d'offrir une flexibilité maximale et une grande rapidité dans la construction de vos algorithmes scientifiques, tout en rendant le processus

extrêmement simple. Torch est accompagné d'un vaste écosystème de packages axés sur la communauté dans les domaines de l'apprentissage automatique, de la vision par ordinateur, du traitement du signal, du traitement parallèle, de l'image, de la vidéo, de l'audio et du réseau, entre autres, et repose sur la communauté Lua. Au cœur de Torch se trouvent les bibliothèques populaires de réseaux neuronaux et d'optimisation, qui sont simples à utiliser tout en offrant une flexibilité maximale pour la mise en œuvre de topologies de réseaux neuronaux complexes. Vous pouvez construire des graphes arbitraires de réseaux neuronaux et les paralléliser sur des CPU et des GPU de manière efficace (Web 13).

A.3.4 Matplotlib

Matplotlib est largement reconnue comme la bibliothèque de visualisation et d'exploration de données la plus populaire. Elle offre une large gamme d'outils permettant de créer des graphiques de base, tels que des graphiques linéaires, des nuages de points, des histogrammes, des graphiques à barres et des graphiques circulaires. Matplotlib constitue la fondation de nombreuses autres bibliothèques de visualisation. C'est une bibliothèque de traçage conçue spécifiquement pour le langage de programmation Python et son extension numérique NumPy. En utilisant Matplotlib, les utilisateurs peuvent visualiser des modèles, des tendances et des corrélations qui pourraient ne pas être détectés en examinant simplement les données textuelles (Web 14).

A.3.5 Seaborn

Seaborn est construit sur Matplotlib. Utilisé pour dessiner des graphiques statistiques attrayants et informatifs. Seaborn est aussi un support spécialisé pour les variables de catégorie pour montrer les observations. Seaborn dispose d'outils pour sélectionner des palettes de couleurs qui révèlent des motifs cachés dans les données (Web 15). Principales caractéristiques de la raison pour laquelle nous utilisons Seaborn :

- Fonctionnalité : Utilise moins de syntaxe et propose des thèmes par défaut simples et intéressants,
- Flexibilité : Fournit les thèmes par défaut les plus utilisés,
- Gestion de plusieurs figures : Automatise la création de plusieurs figures pouvant entraîner des problèmes de mémoire insuffisante.

A.4 Etapes d'implémentation

Voici les étapes pour mettre en œuvre le système de classification des images histopathologiques du cancer du sein.

A.4.1 Importation des bibliothèques

Tout d'abord, nous allons importer tous les modules dont nous aurons besoin pour entraîner notre modèle. La figure A.1 montre un morceau de code permettant d'importer les bibliothèques nécessaires.

```
1 import numpy as np
2 import pandas as pd
3 import torch
4 import matplotlib.pyplot as plt
5 import seaborn as sns
```

FIG. A.1 : Bibliothèques utilisées.

A.4.2 Prétraitement des données

Les données d'image ne peuvent pas être introduites directement dans le modèle, nous devons donc effectuer certaines opérations et traiter les données avant d'être introduites dans le modèle ViT.

```
def get_data_loaders(data_dir, batch_size, train = False):
    if train:
        transform = transforms.Compose([ #normalisation
            transforms.RandomHorizontalFlip(),
            transforms.RandomVerticalFlip(),
            transforms.RandomApply(transforms=[
                transforms.GaussianBlur(kernel_size=(5, 9), sigma=(0.1, 5))], p=0.2),
            transforms.Resize(256),
            transforms.CenterCrop(224),
            transforms.ToTensor(),
            transforms.Normalize((0.485, 0.456, 0.406), (0.229, 0.224, 0.225)),
```

FIG. A.2 : Les opérations pour prétraitement des données.

A.4.3 La division des données

Nous avons divisé cette base de données en trois sous-ensembles, une pour l'entraînement, une pour la validation et une pour le test (train/val/test). La division de ces données a été effectuée à l'aide du code présenté ci-dessous :

```
all_data = datasets.ImageFolder(data_dir, transform=transform)
train_data_len = int(len(all_data)*0.80)
valid_data_len = int((len(all_data) - train_data_len)/2)
test_data_len = int(len(all_data) - train_data_len - valid_data_len)
```

FIG. A.3 : La division des données.

A.4.4 Création du modèle ViT

Pour créer un modèle ViT pré-entraîné prêt à être utilisé pour des tâches de classification d'images, où il est capable de prendre des images en entrée, extraire des caractéristiques visuelles et effectuer des prédictions sur les classes d'images. On utilise la fonction présentée dans la figure ci-dessous.

```
model = timm.create_model('vit_base_patch16_224', pretrained=True)
```

FIG. A.4 : Chargement d'un modèle pré-entraîné ViT.

A.4.5 Apprentissage

Cette ligne de code exécute l'entraînement de notre modèle.

```
model_ft = train_model(model, criterion, optimizer, exp_lr_scheduler,  
                       num_epochs=10)
```

FIG. A.5 : Fonction d'entraînement.

A.4.6 Test

Nous avons 791 images dans notre ensemble de données qui seront utilisées pour évaluer la qualité de notre modèle. Les données de test n'ont pas été impliquées dans l'apprentissage, il s'agit donc de nouvelles données pour notre modèle. La précision obtenue est environ 97.78%.

```
test(model_ft)
```

FIG. A.6 : Fonction d'accuracy.

A.5 Conclusion

Nous avons présenté dans cette annexe les différents aspects du développement de notre système. Nous avons décrit précisément les outils et les étapes d'implémentation de notre système et avons montré quelques captures d'écrans de notre code python.