

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE
APPLIQUÉES

PAR

AMADOU LO NIANG

PREDICTION DE L'EFFICACITE DU TRAITEMENT DE L'OSTEOSARCOME PAR
CHIMIOTHERAPIE NEOADJUVANTE, A L'AIDE D'ALGORITHME
D'APPRENTISSAGE PROFOND

MAI 2025

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

CE MÉMOIRE A ÉTÉ ÉVALUÉ PAR :

M. Fadel Touré, directeur de recherche

Département de Mathématiques et Informatique

Université du Québec à Trois-Rivières

M. Jean Sébastien Dessureault

Département de Mathématiques et Informatique

Université du Québec à Trois-Rivières

M. François Nougarou

Département de Génie Électrique et Informatique

Université du Québec à Trois-Rivières

RÉSUMÉ

L'ostéosarcome est la tumeur osseuse maligne la plus fréquente chez les enfants et les adultes. Une réponse positive à la chimiothérapie néoadjuvante (NAC) peut augmenter de manière significative les chances de survie, ce qui fait du pronostic de la réponse un facteur déterminant pour les patients. Ce pronostic repose généralement sur des données radiomiques de très haute dimension, combinées à des informations paracliniques. Les chercheurs utilisent souvent diverses techniques de réduction de dimensionnalité pour prétraiter ces données avant d'appliquer des algorithmes de classification pour prédire la réponse.

Notre objectif est de prédire cette réponse en utilisant des algorithmes récents d'apprentissage automatique (ML) sans recourir au prétraitement par réduction de dimensionnalité. Pour ce faire, nous exploitons un ensemble de données de haute dimension intégrant des caractéristiques paracliniques et radiomiques. Ces données sont ensuite traitées par des réseaux *Long Short-Term Memory* (LSTM), des *Gated Recurrent Units* (GRU), des réseaux de neurones convolutifs fédérés (FCNN) et le Tab Transformers. Les LSTM et GRU sont adaptés, car ils permettent de traiter des séquences de données tout en conservant les dépendances à long terme, un aspect essentiel pour capturer les relations complexes. Le FCNN, quant à lui, répartit l'entraînement du modèle sur plusieurs sous-réseaux tout en agrégant leurs sorties pour améliorer la précision. Enfin, les Transformers, issus des avancées récentes en traitement du langage naturel (NLP), utilisent des mécanismes d'attention pour identifier des relations complexes entre les caractéristiques.

Les modèles ainsi entraînés ont atteint une précision de 94% et une AUC de 92%, en comparaison aux 91% et 95% obtenus avec les modèles existants dans la littérature. Ces résultats ouvrent la voie à une prédiction précise de la réponse à la NAC, évitant l'étape de réduction de dimensionnalité.

ABSTRACT

Osteosarcoma is the most common malignant bone tumor in children and adults. A positive response to neoadjuvant chemotherapy (NAC) can significantly increase survival rates, making response prediction a critical factor for patients. Prognosis is typically based on high-dimensional radiomic data combined with paraclinical information. Researchers in this field often use various dimensionality reduction techniques to preprocess the data before applying classification algorithms for response prediction.

Our objective is to predict NAC response using recent machine learning (ML) algorithms without employing dimensionality reduction preprocessing. To achieve this goal, we rely on a high-dimensional dataset comprising paraclinical and radiomic features. These features serve as inputs to *Long Short-Term Memory* (LSTM) networks, *Gated Recurrent Units* (GRUs), Federated Convolutional Neural Networks (FCNN), and Tab Transformers. LSTMs and GRUs are relevant in this context, as they are designed to handle sequential data while retaining long-term dependencies, which are crucial for capturing correlations. FCNN distributes model training across multiple sub-networks while aggregating their outputs to enhance prediction accuracy. Transformers, derived from advanced Natural Language Processing (NLP) techniques, leverage attention mechanisms to capture complex relationships between features.

The trained models achieved an accuracy of 94% and an AUC of 92%, compared to 91% and 95% reported in the literature. These results pave the way for accurate NAC response prediction, bypassing the need for dimensionality reduction preprocessing.

REMERCIEMENTS

Mes remerciements s'orientent vers Dieu le ,Tout Puissant, pour m'avoir donné la force, la sagesse et la persévérance nécessaire pour l'accomplissement de ce travail. Que sa lampe continue de guider mes pas dans mes projets.

Je remercie également ma très chère maman Fambaye Seck, dont l'amour et les prières incessantes m'ont soutenu dans les moments difficiles. À mon père, que Dieu l'accueille dans son paradis ! À ma femme Mme Niang, née Fatou Faye Ndao, merci pour ta patience et ton soutien sans faille. Tu as pu gérer avec l'aide de ma très chère famille mes deux filles et notre foyer avec brio, me permettant de me consacrer à mes études. Je remercie également mes sœurs et frères ainsi que mon cousin El Hadji Mansour Mbaye qui ont mobilisé leurs fonds et soutien moral.

Je tiens à exprimer ma profonde gratitude à mon professeur Fadel Touré. Votre soutien psychologique et financier a été inestimable. Votre rigueur dans le travail m'a inculqué des valeurs de discipline et d'excellence. Vous avez été un mentor exemplaire, je suis honoré d'apprendre sous votre direction et votre fameuse expression « **FOCUS SUR LE TRAVAIL** » lorsque je traversais une zone turbulente a apporté ses fruits. Je n'oublie pas non plus le professeur Ousmane Sy dont les conseils avisés ont été précieux tout long de cette aventure académique.

Enfin, je n'oublie pas mes camarades avec qui j'ai passé de nombreuses nuits au laboratoire. Votre camaraderie et esprit de collaboration ont rendu cette expérience enrichissante.

DÉDICACES

À ma mère,

Lorsque j'étais en classe de se primaire, mon père est parti, et c'est toi qui as pris en charge seule tout le poids de la famille. Peu après, tu as dû faire face à une épreuve encore plus difficile, le cancer du sein. Malgré cette maladie, tu n'as jamais cessé de me soutenir et de m'accompagner dans mes études.

C'est dans cette période que j'ai ressenti la nécessité de me tourner vers des recherches sur le cancer, en hommage à ta force et à ton courage.

Ce mémoire est, en quelque sorte, un remerciement pour tout ce que tu as fait pour moi, pour ton amour et ta détermination qui m'ont guidée tout au long de mon parcours.

Àvec toute ma gratitude !

TABLE DES MATIÈRES

Résumé	iii
Abstract.....	iv
Remerciements.....	v
Dédicaces.....	vi
Table des matières	vii
Liste des figures	x
Liste des tableaux	xi
Liste des abréviations.....	xii
 CHAPITRE 1.....	 1
 Introduction	 1
1.1 Introduction.....	1
1.2 Contexte et problématique.....	3
1.3 Objectifs.....	4
1.4 Organisation du mémoire	5
 CHAPITRE 2.....	 6
 Revue de la littérature.....	 6
2.1 Traitement par chimiothérapie néoadjuvante (NAC)	6
2.1.1 Approches de prédiction de la réponse à la NAC	8
2.2 Présentation des réseaux de neurones profonds	10
2.2.1 Réseaux de neurones Récurrents (RNN).....	10
2.3 Les réseaux de neurones convolutifs	16

2.3.1	L'opération de convolution	16
2.3.2	Le noyau (Kernel).....	17
2.3.3	L'opération de Pooling.....	17
2.3.4	Architecture	19
2.4	Le modèle Transformers	19
2.4.1	Mécanisme d'auto-attention.....	21
2.4.2	Multi Head attention.....	23
2.4.3	Encodage positionnel	24
2.4.4	Incorporation des mots.....	25
	CHAPITRE 3	27
	Collecte, traitement des données et méthodologie	27
3.1	Donnée radiomiques	27
3.1.1	Identification de la région d'intérêt	28
3.1.2	Segmentation.....	29
3.1.3	Extraction des caractéristiques.....	30
3.2	Collecte de données	31
3.3	Applications.....	33
3.4	Méthodologie	34
3.4.1	Statistiques descriptives.....	34
3.5	Les fonctions d'activation.....	36
3.5.1	Relu	37
3.5.2	Softmax	38
3.6	Métriques d'évaluation de performance	39
3.6.1	Matrice de confusion	39
3.6.2	Exactitude.....	40
3.6.3	Précision.....	40
3.6.4	Rappel ou Sensibilité, Spécificité	41

3.6.5	F1 score.....	41
3.6.6	Courbe ROC (AUC).....	41
3.6.7	Moyenne géométrique.....	42
3.7	Outils et technologies	43
3.7.1	Google Colab	43
3.7.2	Google Drive	44
3.7.3	Python	44
3.7.4	Ressources matérielles.....	47
3.8	Expérimentations.....	47
3.8.1	Le modèle LSTM.....	48
3.8.2	Le modèle GRU	49
3.8.3	Le modèle FCNN.....	49
3.8.4	Le modèle Tab Transformers	51
	CHAPITRE 4.....	54
	Résultats et discussion.....	54
4.1	Résultats des expérimentations.	54
4.1.1	Evaluation du modèle FCNN.....	55
4.1.2	Evaluation du modèle Tab Transformers	57
4.2	Comparaison des modèles.....	59
4.3	Comparaison de notre meilleur modèle avec d'autres approches.....	60
4.4	Discussion des résultats	61
	CHAPITRE 5.....	63
	Conclusion et perspectives	63
	Références.....	64

LISTE DES FIGURES

Figure 1 : Architecture des LSTM(Olah, 2015).....	12
Figure 2 : Architecture du modèle GRU(Xu et al., 2023).....	15
Figure 3 : Calcul de la matrice pour une image RVB (Bhatt et al., 2021).....	17
Figure 4 : Pooling max avec filtre 2x2 et stride 2(Bhatt et al., 2021).....	18
Figure 5 : Pooling moyen avec filtre 2x2 et stride 2(Bhatt et al., 2021).....	18
Figure 6 : Architecture des CNN	19
Figure 7 : Architecture du modèle Transformers	21
Figure 8 : Mécanisme d’auto-attention	22
Figure 9 : Multi Head attention.....	24
Figure 10 : Habitats chez un patient atteint de glioblastome multiforme(Gillies et al., 2016).	29
Figure 11 : Processus d’extraction des caractéristiques radiomiques	30
Figure 12 : Processus de sélection des patients	32
Figure 13 : Répartition des classes.....	35
Figure 15 : Fonction d’activation ReLu.....	37
Figure 16 : Fonction d’activation softmax.....	38
Figure 17 : Courbe de variation de la fonction de perte et de précision du modèle LSTM....	48
Figure 18 : Courbes de variation de la fonction de perte et de la précision du modèle GRU	49
Figure 19 : Architecture FCNN.....	50
Figure 20 : Courbes de variation de la fonction de perte et de précision du modèle FCNN ..	51
Figure 21 : Architecture du modèle Tab Transformers	52
Figure 22 : Courbe de variation de la précision du modèle Tab Transformers	53
Figure 23 : Matrice de confusion du modèle FCNN.....	55
Figure 24 : Courbe ROC (AUC) du modèle FCNN	56
Figure 25 : Matrice de confusion du modèle Tab Transformer.....	57
Figure 26 : Courbe ROC (AUC) du modèle Tab Transformers	58

LISTE DES TABLEAUX

Tableau 1 : Matrice de confusion.....	39
Tableau 2 : Ressources matérielles	47
Tableau 3 : Rapport de classification de FCNN.....	57
Tableau 4 : : Rapport de classification de Tab Transformers	59
Tableau 5 : Récapitulation des métriques des modèles.....	59
Tableau 6 : Comparaison de notre meilleur modèle avec celui de la littérature	61

LISTE DES ABRÉVIATIONS

ACP	Analyse en Composante Principale
AUC	Area Under the Curve
CNN	Convolutional Neural Network
DL	Deep Learning
FCNN	Federate Convolutional Neural Network
FP	Faux Positif
FN	Faux Négatif
IA	Intelligence Artificielle
IRM	Imagerie par Résonance Magnétique
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
LASSO	Régression de l'opérateur de retrait et de sélection le moins absolu
LSTM	Long Short-Term Memory
ML	Machine Learning
MTV	Volume Métabolique de la Tumeur
NAC	Chimiothérapie Néoadjuvante
RNN	Recurrent Neural Network
SUV	Valeur d'Absorption Standard
SVM	Support Vecteur Machine
TLG	Glycolyse Totale de la Lésion
TEP	Tomographie par Émission de Positrons
VP	Vrai Positif
VN	Vrais Négatif

CHAPITRE 1

INTRODUCTION

1.1 Introduction

Le cancer est une maladie caractérisée par la croissance anormale et la prolifération incontrôlée des cellules cancéreuses. Ces cellules forment une tumeur et peuvent envahir ou détruire les tissus avoisinants. On parle de métastases lorsqu'une tumeur peut se développer dans n'importe quel organe ou tissu et se propager partout ailleurs dans l'organisme. Les cancers se déclinent en une multitude de types dépendamment de leurs caractéristiques et de leur site d'origine. Nous distinguons : le cancer des os, le cancer du sein, le cancer du col de l'utérus, etc. Parmi les cancers osseux, l'ostéosarcome, le sarcome d'Edwing et le chondrosarcome sont les tumeurs primitives dominantes. En effet, l'ostéosarcome est la tumeur maligne osseuse la plus fréquente chez les adolescents et les jeunes adultes (Chen et al., 2016; Desandes et al., 2016). Elle constitue le résultat de la destruction incontrôlée des cellules responsables du renouvellement osseux. Son traitement consiste généralement à subir une *Neoadjuvant Chimiotherapy* (NAC), suivie d'une intervention chirurgicale et dans certains cas une chimiothérapie adjuvante (Benjamin, 2020; Bielack et al., 2022) au cas où la biopsie décèle des séquelles.

D'un autre côté, l'Intelligence Artificielle (IA) est devenue un sujet d'actualité. Ses prouesses font régulièrement la Une à travers le *Machine Learning* (ML) pour le traitement

des manuscrits offrant des solutions innovantes pour la transcription, l'analyse et l'interprétation ; la reconnaissance faciale, le dépistage de maladies, la prédition de la réponse à un traitement, etc. L'IA peut se définir alors comme des programmes informatiques qui ont pour ambition de reproduire les schémas de perception de l'intelligence humaine. En ce sens, des tâches ordinaires effectuées par l'homme peuvent être maintenant accomplies par l'IA. Par exemple, son application dans le domaine de la santé a fourni des systèmes d'aide au diagnostic ou au dépistage, etc.

Le ML, est un sous-ensemble de l'IA qui permet à une machine d'apprendre à partir des données. Le *Deep Learning* (DL), un sous-ensemble aussi du ML particulièrement adapté aux grosses quantités de données utilisant des modèles de réseau de neurones à plusieurs couches. Le DL a permis le développement de plusieurs outils dans le domaine médical qui aident à la prédition du taux de survie ou de la réponse au traitement afin d'assurer un meilleur suivi des patients (Huang et al., 2022).

Bien plus, l'apprentissage profond a démontré une performance remarquable dans le traitement de vastes ensembles de données. C'est pourquoi nous allons nous en servir dans ce travail de recherche, pour le traitement des données radiomiques. Ces dernières sont issues de l'analyse des images médicales. C'est une traduction des données d'imagerie en données quantitatives, qualitatives ou sémantiques de haute dimension (Rizzo et al., 2018; Rogers et al., 2020)

Dans cette étude, nous allons tenter de prédire la réponse au traitement de la NAC sans réduire la dimensionnalité des données radiomiques à l'aide des modèles de DL. Pour se faire, nous allons exploiter les réseaux de neurones récurrents (LSTM et GRU), les réseaux de neurones convolutifs (CNN) et les *Transformers*. Les LSTM et GRU sont importants dans ce contexte, car ils sont conçus pour traiter des données tout en conservant les informations sur le long terme, ce qui est important pour capturer les dépendances. Le *Federate Convolutional Neural Network* (FCNN) consiste à utiliser les CNN en distribuant l'entraînement du modèle

sur plusieurs sous-réseaux, tout en agrégeant leur sortie pour améliorer la précision de la prédiction. Les Transformers sont des architectures de réseau de neurones profonds issues des techniques récentes du NLP capables de capturer des relations complexes entre les caractéristiques grâce aux mécanismes d'attentions (Vaswani et al., 2017). Ces modèles prendront en entrée des données radiomiques de haute dimensionnalité combinée à d'autres paramètres paracliniques pour tenter de prédire la réponse à la NAC sur l'ostéosarcome.

1.2 Contexte et problématique

L'ostéosarcome est la tumeur osseuse maligne primitive la plus présente chez l'enfant et l'adolescent. D'après des récentes statistiques, elle représente 60% de cas de cancers chez l'enfant et l'adolescent (Nyeko et al., 2024). Avec l'introduction de la chimiothérapie néoadjuvante et adjuvante en plus des techniques chirurgicales, des progrès sur le taux de survie sont notés avec une variation entre 65% et 75% pour les patients sans maladies métastatiques (Jean-Eric et al., 2023). Alors qu'avant, son traitement a longtemps été difficile.

Malgré les progrès de la chirurgie et de la chimiothérapie, l'ostéosarcome reste toujours une maladie dont le traitement reste un défi. Par le biais du taux de nécrose tumorale, calculé après la chimiothérapie à l'aide d'un échantillon pathologique postopératoire utilisé comme étalon-or la réponse à la NAC est évaluée. Cette méthode est invasive et implique une chirurgie complexe (Zhang et al., 2021). Ce constat justifie, le recours aux méthodes de prédiction moins invasives qui se ferait par une analyse des données radiomiques, en exploitant des modèles de prédiction tels que : LSTM, GRU, FCNN et Transformers. L'objectif étant d'anticiper l'efficacité de la NAC, par la prédiction de la réponse du patient au traitement de l'ostéosarcome.

Au meilleur de nos connaissances et à travers l'analyse de la littérature des recherches menées dans la prédiction de la réponse à la NAC, les auteurs ont toujours eu recours à la réduction de dimensionnalité, avec le risque de perdre des informations intrinsèques à cette technique, avant d'appliquer des modèles de prédiction. Dans ce travail, nous étudierons la capacité de différents modèles de réseaux de neurones utilisant l'entièreté des données radiomiques dans la prédiction de la réponse à la NAC.

1.3 Objectifs

Les travaux présentés dans ce mémoire ont pour objectif premier, la mise en place d'une nouvelle approche de traitement des données radiomiques qui n'utilise pas les techniques de réduction de dimensionnalité entraînant inévitablement des pertes. Nous développons des approches qui combinent toutes les caractéristiques radiomiques avec les paramètres paracliniques pour alimenter des algorithmes DL. Ces modèles formés vont prédire la réponse positive ou négative de la NAC sur l'ostéosarcome.

Nous allons nous appuyer sur trois (3) architectures de réseaux de neurones profonds, largement utilisées dans la littérature à savoir, les réseaux récurrents (LSTM et GRU), les CNN et enfin les *Transformers*. L'implémentation effective de ces approches pourrait servir d'outils d'aide à la décision pour le personnel de la santé dans le but d'optimiser le traitement des patients.

En second lieu, nous allons comparer nos résultats à ceux d'une autre approche qui a utilisé le même jeu de données dans l'optique de justifier la pertinence de notre choix.

1.4 Organisation du mémoire

Ce mémoire est composé de cinq (5) chapitres dans lesquels nous traitons en profondeur les points essentiels de notre travail.

Le premier chapitre est une introduction générale du mémoire qui donne les définitions, le contexte et la problématique puis l'objectif de ce travail et enfin l'organisation du mémoire. Dans le deuxième chapitre, nous allons dans un premier temps expliquer l'impact de la NAC sur l'ostéosarcome. Par la suite, nous aborderons les différentes approches de prédition de la réponse à la NAC, et terminer par une présentation des architectures des réseaux de neurones que nous avons utilisés durant nos expérimentations. Le troisième chapitre est dédié à la collecte, au traitement des données et à la méthodologie appliquée. L'avant-dernier chapitre est consacré à la présentation des résultats de nos différentes expérimentations, suivi d'une comparaison de nos résultats avec ceux des autres chercheurs et se termine par une discussion. Enfin, le dernier chapitre présente la conclusion de ce travail et décrit les perspectives de recherches.

CHAPITRE 2

REVUE DE LA LITTÉRATURE

La prédiction de la réponse à la NAC représente un enjeu majeur pour optimiser le traitement des patients atteints de l'ostéosarcome. La complexité des données et la variabilité entre les patients constituent la principale limite des méthodes traditionnelles pour la prédiction de la réponse à la NAC.

Les réseaux de neurones en particulier les réseaux de neurones profonds ont démontré des résultats prometteurs dans le traitement des vastes ensembles de données, que ce soit la prédiction, la détection, la classification, etc.

Cette revue de la littérature présente les recherches antérieures dans la prédiction de la réponse à la NAC et met en lumière les méthodologies, puis les résultats et enfin les limites des études dans le domaine.

2.1 Traitement par chimiothérapie néoadjuvante (NAC)

La chimiothérapie néoadjuvante se définit comme une administration préopératoire d'un traitement cytotoxique (Gonçalves, 2016). Elle est appliquée dans le traitement du cancer, notamment le cancer du sein, du col de l'utérus et de l'ostéosarcome. Elle reste une intervention cruciale dans le traitement de l'ostéosarcome. Il est essentiel de souligner que la

NAC prolonge la survie et améliore la qualité de vie (Zheng et al., 2023). Entre autres, le but du traitement est de :

Faire régresser la tumeur

La régression de la tumeur par la NAC offre plusieurs avantages avant de procéder à la chirurgie, car celle-ci devient plus facile, moins invasive et peut éviter des amputations (Gharbi et al., 2023). De plus, la régression de la tumeur octroie également une possibilité à l'évaluation de la réponse de la tumeur à la NAC permettant ainsi d'adapter le traitement adéquat pour améliorer les résultats.

Traiter le micrométastatique précoce.

Les cellules micro métastatiques précoces constituent l'ensemble des cellules cancéreuses disséminées dans le corps, mais peuvent être invisibles à l'imagerie. Ainsi, la NAC permet de les traiter efficacement avant qu'elles ne gagnent du terrain et deviennent visibles. De ce fait, cette approche participe activement à la destruction des cellules micro métastatiques précoces qui sont capables de pousser la maladie vers un état plus avancé ou critique.

Favoriser la chirurgie conservatrice

La NAC favorise la chirurgie conservatrice, car elle réduit la taille de la tumeur avant l'intervention chirurgicale. Cela permet aux chirurgiens de garder les tissus sains en donnant lieu à une résection moins grave ou à une chirurgie moins invasive. Dans cette optique, cette approche pourrait permettre une intervention plus précise sur la zone tumorale.

De toute évidence, vu les avantages de la NAC, la prédition de sa réponse sera alors d'un apport capital sur le plan du traitement du patient atteint par l'ostéosarcome.

Afin de pouvoir établir un nouveau modèle prédictif efficace, il est nécessaire d'étudier les projets de recherche antérieurs ainsi que les différents modèles utilisés pour prédire la réponse à la NAC lors de ces études.

2.1.1 Approches de prédiction de la réponse à la NAC

La réponse à la NAC est un facteur déterminant sur le traitement de l'ostéosarcome, car elle permet d'avoir une orientation très importante sur l'opération chirurgicale subséquente. Dans ces circonstances, des chercheurs du domaine s'appuient sur les techniques statistiques combinées à ML pour prédire la réponse à la NAC. D'autres s'orientent vers la définition d'un pipeline qui se sert des modèles de DL pour extraire les caractéristiques radiomiques. Ces dernières sont les entrées des techniques statistiques de sélection de variables telles que la Régression de l'opérateur de retrait et de sélection le moins absolu (LASSO) et l'Analyse en Composante Principale (ACP) pour réduire la dimension. Le jeu de données obtenu à l'issue de cette opération sera passé aux algorithmes de ML pour la prédiction de la réponse à la NAC.

Certaines recherches exploitent des caractéristiques texturales combinées à l'apprentissage automatique qui utilisent les ACP pour réduire la dimensionnalité. Dans (Byun et al., 2014), les auteurs ont étudié la surveillance de la réponse précoce à la NAC dans le traitement de l'ostéosarcome. Ils ont utilisé la Tomographie par Émission de Positrons (TEP) et l'Imagerie par Résonnance Magnétique (IRM) pour évaluer les changements liés à la chimiothérapie. Les résultats montrent que la réponse histologique prédit la réponse de la NAC avec précision par TEP. Parmi les paramètres de la TEP, le Volume Métabolique de la Tumeur (MTV) et la Glycolyse Totale de la Lésion (TLG) étaient des prédicteurs indépendants de la réponse histologique. Cependant, le nombre restreint de patients (30), traités dans un seul et même établissement, constitue une limite. Par ailleurs, les seuils de paramètres suggérés par

la TEP peuvent différer en fonction des scans. En outre, les auteurs n'ont pas étudié les valeurs pronostiques des paramètres TEP et IRM après la NAC. Les investigations effectuées dans (Jeong et al., 2019) se basent sur les caractéristiques texturales pour prédire la réponse à la NAC avec une cohorte plus étendue de patients. Par le biais de la Valeur d'Absorption Standard (SUV), la TLG et le MTV, les caractéristiques quantitatives des tumeurs sont évaluées. À l'aide de l'analyse texturale des images, l'hétérogénéité de la tumeur est évaluée. Cette évaluation par la boîte à outils « Chang Gang Image Texture Analysis » a permis d'extraire plusieurs caractéristiques. L'utilisation du ACP a permis de réduire la dimension du jeu de données. Le nouveau jeu de données à dimension réduite par l'ACP est utilisé comme entrée aux algorithmes d'apprentissage machine tels que le Support Vector Machine (SVM), le Random Forest et le Gradient Boost Methods. Les auteurs ont obtenu respectivement des valeurs d'Area Under the Curve (AUC) de 0,72, 0,78 et 0,82. Ces résultats suggèrent qu'une approche basée sur le Machine Learning (ML) avec ces caractéristiques texturales peut aider à prédire la réponse à la NAC à l'aide d'images TEP de base. Cette étude présente un échantillon limité de données dû au fait que l'ostéosarcome est une maladie très rare avec moins de 1% des diagnostics de cancer (Jeong et al., 2019).

D'autres investigations utilisent des techniques de sélection des variables à partir des données radiomiques. Dans cet article (Zhang et al., 2021), les auteurs s'attaquent à la problématique de haute dimensionnalité des données radiomiques qui consiste en une projection des données d'imagerie en données quantitatives de hautes dimensions (Sun et al., 2021). Les auteurs ont appliqué la technique de réduction de dimension (LASSO) sur les données radiomiques de mille quatre cent neuf (1409) caractéristiques pour les réduire à sept (7). Le meilleur modèle formé à partir de la régression logistique avec le jeu de données à dimensionnalité réduite performe avec une exactitude de 90% et un AUC de 95%. Comme nous l'avons mentionné à propos des études précédentes, la limite principale reste toujours le manque de données quantitatives et variées. En plus, les données radiomiques et cliniques sont exploitées sans prendre en compte les informations pathologiques des patients. Dans la même optique, les

auteurs Zheng et al (Zheng et al., 2023) utilisent l'outil « Auto3DSeg » pour la segmentation des images puis se servent des réseaux de neurones profonds pour l'obtention des caractéristiques radiomiques. Parallèlement, ils effectuent manuellement l'extraction des caractéristiques. L'ensemble final de données combine des données extraites manuellement et automatiquement. Puis les techniques de sélection de variables telles que LASSO et ACP sont utilisées afin de réduire la dimensionnalité du jeu de données qui alimentera les modèles de classificateurs. Les auteurs obtiennent une précision de 93,8% et un AUC de 96,1% pour la réponse à la NAC. L'étude s'est déroulée dans un seul centre médical avec un faible nombre de cas, ce qui a pour conséquence de fortement limiter la généralisation et la validation sur des résultats de données issues de sources différentes.

2.2 Présentation des réseaux de neurones profonds

Dans les années 1980, des méthodes d'apprentissage basées sur des approches statistiques avec des classificateurs tels que la régression logistique et *Naive Bayes* ont émergé (Batutin, 2024; Manning, 1999). Cependant, ces méthodes étaient limitées dans leur capacité à traiter de longues séquences (Bengio et al., 1994). Ainsi, la nécessité de gérer des séquences de données longues et variées a mis en évidence les limites de ces approches traditionnelles. Pour pallier à ces difficultés, les réseaux de neurones récurrents ont vu le jour.

2.2.1 Réseaux de neurones Récurrents (RNN)

Issus des techniques de traitement du texte, les réseaux de neurones récurrents (RNN) sont dotés de capacité de traitement des données textuelles ou numériques avec des longueurs de séquence très variées nécessitant une capture des dépendances séquentielles (Delsinne, 2021; Elman, 1990). En effet le calcul d'une sortie y_t à l'instant t reçoit x_t l'entrée actuelle et

la sortie \mathbf{h}_{t-1} la représentation des entrées précédentes. L'état caché \mathbf{h}_t est modulé par l'ensemble des composantes linéaires ($\mathbf{x}_1, \dots, \mathbf{x}_t$) grâce à une fonction d'activation σ_h . Cette opération dans I décrite ci-dessous, permet de conserver la mémoire de la séquence (Dias et al., 2018). De plus, leur conception inclut une mémoire permettant de traiter des séquences en profondeur, contrairement aux perceptrons multicouches (MLP), aux réseaux de neurones à mémoire associative (AMN) etc.

$$\begin{cases} \mathbf{h}_t = \sigma_h(\mathbf{w}_h \mathbf{x}_t + \mathbf{U} \mathbf{h}_{t-1} + \mathbf{b}_h) \\ \quad \vdots \\ \mathbf{y}_t = \sigma_y(\mathbf{w}_y \mathbf{h}_t + \mathbf{b}_y) \end{cases} \quad (1)$$

Les paramètres de l'équation I sont σ la fonction d'activation, \mathbf{w}_h la matrice de poids associée à l'entrée actuelle \mathbf{x}_t , \mathbf{U} la matrice de poids appliquée à la sortie précédente \mathbf{h}_{t-1} puis \mathbf{b} le biais associé à la sortie ou à l'état caché.

Cependant, lors de la rétropropagation, avec un corpus assez volumineux qui demandera plus de couches pouvant entraîner la disparition du gradient montre les limites des RNN. C'est dans ce contexte, que les cellules à portes ont vu le jour dont les plus connus sont les *Long Short Term Memory* (LSTM) et sa variante simplifiée *Gated Recurrent Unit* (GRU) (Bahdanau et al., 2014).

A Long Short-Term Memory (LSTM)

Le fait que les RNN sont souvent confrontés au problème de la disparition du gradient, a conduit à la mise en place des LSTM. Il s'agit d'un autre modèle de RNN, qui se distingue par son architecture permettant de conserver des informations sur de longues séquences (Song et al., 2020). Cette capacité accrue de mémoire confère un avantage aux LSTM dans le traitement des corpus plus étendus. En outre, la performance supérieure des LSTM réside dans

leur capacité à éliminer les informations moins pertinentes lors de la transition de l'état **t-1** à l'état **t**, tout en transmettant les informations importantes à l'entrée suivante ou à la sortie finale (de la Roy et al., 2022). L'idée générale est de contrôler le flux d'information à travers des cellules récurrentes en utilisant des portes telles que la porte d'oubli (forget gate), la porte de sortie (output gate) ou la porte d'entrée (input gate). Ces portes sont utilisées pour mettre à jour l'état interne des cellules. La figure 1 illustre ce mécanisme.

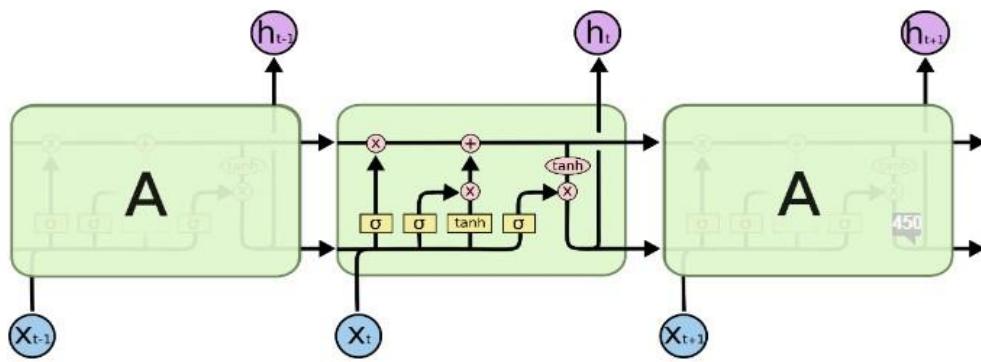


Figure 1 : Architecture des LSTM(Olah, 2015)

La question se pose initialement au sujet de la décision concernant le choix de l'information à ajouter via la porte d'entrée. Ici, le rôle de la porte d'entrée est très important du fait qu'elle sert de filtre à l'importante quantité d'information à véhiculer vers la cellule mémoire. Pour se faire, l'application de la formule ci-dessous sur l'état caché h_{t-1} et l'entrée actuelle x_t et sa valeur à l'instant t , nous obtenons la valeur de l'information à conserver i_t :

(2)

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i)$$

Mise à jour de la cellule mémoire

Une fois que la porte d'entrée décide de l'informations i_t à conserver, l'état interne de la cellule est mis à jour et est combiné à l'état cellulaire précédent $i_t * C_{t-1}$ et à l'état cellulaire $i_t * \tanh(w_c \cdot [h_{t-1}, x_t] + b_c)$. L'utilisation de \tanh garantit l'introduction de non-linéarité dans le réseau. À travers ce mécanisme, le réseau arrive à garder l'informations sur le long terme. Ainsi, l'informations actuelle C_t est obtenue par la formule 3 suivante :

$$C_t = i_t * C_{t-1} + i_t * \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

Porte d'oubli

Le réseau décide de la quantité d'informations à supprimer ou à conserver. Cette décision est traduite par la fonction d'activation sigmoïde qui renvoie un vecteur f_t de valeurs comprises entre 0 et 1. Autrement, ces valeurs indiquent quelle portion des informations de l'état de la cellule précédente doit être conservée ou supprimée pour l'état actuel.

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

Porte de sortie

La porte de sortie contrôle le dynamisme de l'informations à véhiculer vers une nouvelle entrée ou la sortie finale. Cela lui confère un rôle capital, de plus, en régulant le flux d'information tout en aidant à prévenir la perte d'information importante qui pourrait survenir. Le vecteur de sortie o_t est obtenu à partir de la formule ci-dessous :

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

En définitive, ces portes coordonnent leurs actions sur le flux d'information dans le but de permettre au réseau de sélectionner, conserver et utiliser les informations les plus importantes lors du traitement de la séquence de données.

B Gated Recurrent Unit (GRU)

Dans l'optique d'améliorer les LSTM, des chercheurs du domaine se posent les questions suivantes :

Quels sont les éléments réellement nécessaires dans l'architecture des LSTM ?

Quelles autres architectures pourraient permettre le contrôle dynamique de l'échelle de temps et d'oublier le comportement de certaines unités ?

Certaines réponses apportées par les travaux des chercheurs (Cho et al., 2014; Chrupała et al., 2015; Chung et al., 2014) sur les portes des RNN (GRU) ont contribué à cette amélioration. Le GRU est une variante spectaculaire des LSTM qui combine les portes d'oubli et d'entrée en une seule porte de mise à jour voir *figure 2* (Cho et al., 2014). Le GRU fusionne l'état de la cellule et l'état caché, ce qui aboutit à un modèle avec moins de paramètres et simplifié d'où sa popularité. Cette configuration lui permet aussi de n'avoir qu'une porte de mise à jour (update gate) et une porte de réinitialisation (reset gate) comme le montre la *figure 2* dont les valeurs sont respectivement notées \mathbf{z}_t et \mathbf{r}_t dans les équations 6 et 7.

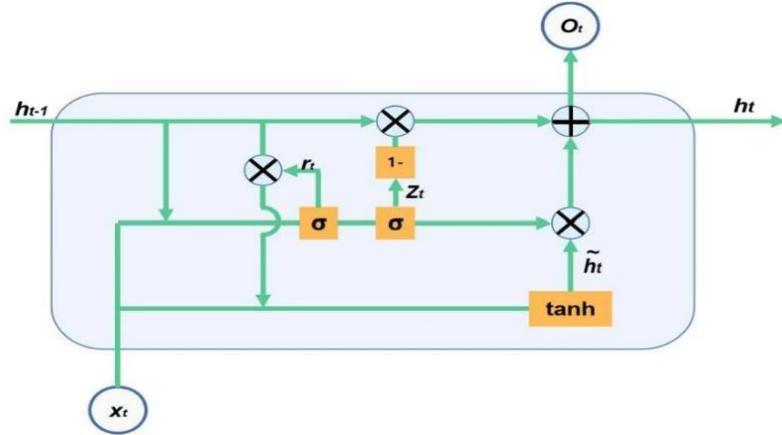


Figure 2 : Architecture du modèle GRU(Xu et al., 2023)

La porte de mise à jour (update gate) z_t :

Cette porte agit comme un intégrateur conditionnel qui peut contrôler linéairement n’importe quelle dimension. Elle peut choisir de copier la valeur existante ou de la remplacer entièrement par une nouvelle valeur. L’équation 6 permet d’effectuer la mise à jour :

(6)

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

La porte de réinitialisation (reset gate) r_t

Cette porte se concentre sur la régulation des parties de l’état utilisées pour déterminer l’état cible. Elle introduit un effet non linéaire additionnel dans la relation entre l’état antérieur et l’état futur.

(7)

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

2.3 Les réseaux de neurones convolutifs

Les réseaux convolutifs, également appelés réseaux de neurones convolutifs ou Convolutional Neural Networks (CNN) en anglais, sont devenus populaires en 1989 grâce au chercheur Yann LeCun (LeCun et al., 1989). Leur nom provient de l'opération mathématique de convolution sur laquelle ils se basent. Ces réseaux sont particulièrement adaptés au traitement de données structurées, telles que les images ou les données numériques unidimensionnelles. Ils ont révolutionné le domaine de la vision par ordinateur grâce à leur remarquable capacité à analyser les images. Ils parviennent à résoudre des problèmes complexes, notamment la classification et la détection d'objets. Les CNN ont aussi montré leur capacité à traiter les données à 1D (Bhatt et al., 2021; Vinayakumar et al., 2017). Dans (Yıldırım et al., 2018) les auteurs ont utilisé les CNN pour prédire 17 classes à partir des données numériques tabulaires issues des électrocardiogrammes (ECG). Les CNN ont, ainsi, apporté des solutions pratiques dans de nombreux secteurs, qui incluent : la médecine, l'industrie automobile, l'agriculture, etc. Leur polyvalence et leur efficacité en font des outils précieux pour de multiples applications dans ces domaines et bien d'autres.

2.3.1 L'opération de convolution

La convolution est l'opération mathématique principalement utilisée dans les CNN. Permettant d'extraire les caractéristiques des images d'entrée tout en préservant les relations spatiales entre les pixels (De Vries et al., 2016). Pour ce faire, l'image est représentée sous forme de matrice, où chaque élément correspond à la valeur d'un pixel. Cette approche facilite l'analyse locale des caractéristiques de l'image tout en conservant une structure qui respecte l'agencement spatial d'origine. La convolution est également appliquée aux données tabulaires unidimensionnelles (1D), couramment rencontrées dans les domaines tels que la santé, la

finance et l'analyse des séries temporelles. Dans ce cas, les données sont organisées sous forme de vecteurs, et des filtres de convolution 1D sont appliqués pour extraire des motifs ou des tendances dans les séries temporelles. Cette capacité à capturer les dépendances locales dans les données tabulaires rend la convolution utile pour des tâches telles que la détection d'anomalies, la classification des séries temporelles et la prédiction.

2.3.2 Le noyau (Kernel)

Le noyau ou filtre a la même profondeur que l'entrée. Il réalise une multiplication de la matrice avec les canaux de l'entrée. Il combine les résultats pour produire un canal de sortie (une matrice) (De Vries et al., 2016). Ce processus est illustré dans la figure 3 montrant le calcul de la matrice de sortie.

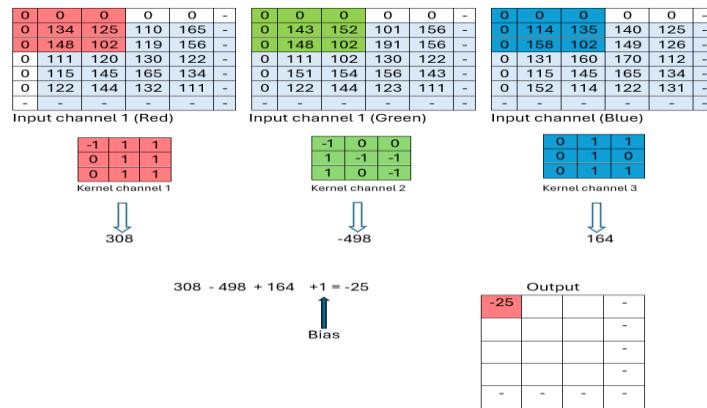


Figure 3 : Calcul de la matrice pour une image RVB (Bhatt et al., 2021)

2.3.3 L'opération de Pooling

S

Les cartes de caractéristiques ou d'activations passent par un ReLU qui remplace à son tour toutes les valeurs négatives par zéro et garde le reste. L'opération de pooling (max ou

moyenne, somme, etc.) réduit la dimensionnalité de la carte en conservant les caractéristiques les plus importantes ou moyennes ou globales. Ainsi, elle réduit le besoin en termes de ressources informatiques et le risque de sur-

apprentissage des modèles (Bhatt et al., 2021). L'extraction des caractéristiques ne modifie pas leur représentation spatiale, donc les caractéristiques restent invariantes.

Max pooling :

Le tenseur de sortie dans la couche de convolution sert d'entrée pour la couche de pooling. Dans le cas du max pooling, un noyau de taille $n*n$ ($2*2$ dans la figure 4) est glissé sur la matrice. À chaque position, la valeur maximale est sélectionnée puis insérée à l'emplacement correspondant dans la matrice de sortie.

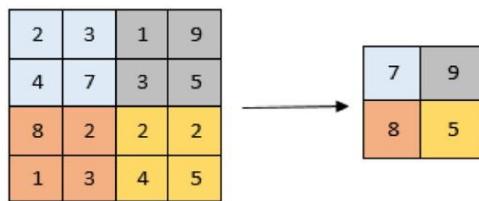


Figure 4 : Pooling max avec filtre $2x2$ et stride 2(Bhatt et al., 2021)

Moyen pooling :

Dans le cas du pooling moyen, un noyau de taille $n*n$ est glissé sur la matrice. À chaque position, la moyenne des valeurs est calculée et insérée à la position correspondante de la matrice de sortie.

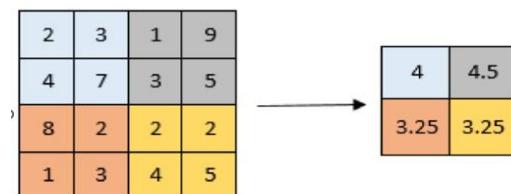


Figure 5 : Pooling moyen avec filtre $2x2$ et stride 2(Bhatt et al., 2021)

2.3.4 Architecture

En résumé, l'architecture des CNN repose sur trois types de couches essentielles, disposées de manières séquentielles et interconnectées. La couche de convolution extrait les caractéristiques importantes en appliquant des filtres aux données d'entrée. Ensuite, la couche de pooling réduit la dimensionnalité tout en conservant les informations pertinentes. Enfin, la couche entièrement connectée finalise le traitement pour aboutir aux prédictions ou à la classification des données. Toutes ces couches travaillent ensemble pour extraire, simplifier et interpréter les informations des données d'entrée.

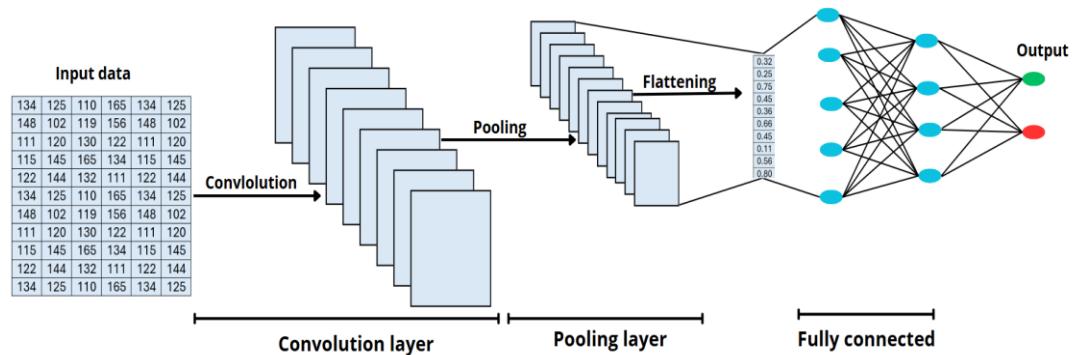


Figure 6 : Architecture des CNN

2.4 Le modèle Transformers

Rappelons que la limitation principale des RNN est due au fait que la rétropropagation temporelle avec un corpus assez volumineux nécessitant plusieurs couches, peut entraîner la disparition du gradient. Pour atténuer ce problème, les LSTM ont vu le jour afin de réduire en partie cet effet de disparition du gradient. Contrairement au RNN simples, les LSTM ont la capacité de filtrer les informations moins pertinentes et de transmettre le reste aux étapes suivantes (de la Roy et al., 2022). En vue d'améliorer les LSTM, les unités récurrentes avec

portes (GRU) sont introduites. Ces unités, avec moins de paramètres, sont particulièrement efficaces pour prédire des données séquentielles.

Bien que les LSTM et les GRU aient révolutionné le traitement des séquences, ils présentent des limites pour les longues séquences, notamment en raison de la disparition du gradient et de la perte d'information. Les Transformers ont été introduits pour pallier ces faiblesses grâce à un mécanisme d'attention, qui permet une exécution parallèle et prend en compte les relations distantes sans recourir à la récurrence. À ce sujet, la figure 7 nous montre l'architecture des Transformers qui est composée d'un bloc encodeur et d'un bloc décodeur.

Le bloc encodeur regroupe deux sous-composants principaux : le Multi-Head attention et le réseau de neurones feed-forward. Ces composants sont associés à des mécanismes de normalisation à l'aide de connexions résiduelles, ce qui stabilise et accélère l'apprentissage. Le bloc encodeur est souvent utilisé dans des tâches telles que la classification de texte, avec des modèles comme le Bidirectionnel Encoder Représentations from Transformers (BERT) (Wang et al., 2019) et ses variants, ainsi que dans la vision par ordinateur avec les Vision Transformers (ViT) (Ammar et al., 2024) entre autres.

Quant au bloc du décodeur, il comprend trois sous composants le Multi-Head attention, le réseau de feed-forward (identiques à ceux de l'encodeur) et Masked MultiHead attention qui empêche la visualisation des éléments futurs dans la séquence en cours. Cette caractéristique est essentielle dans les modèles de génération de texte, comme le Generate Pre-trained Transformer (GPT) (Hanna et al., 2024).

Dans la section suivante, nous présentons en détail tous les composants de cette architecture.

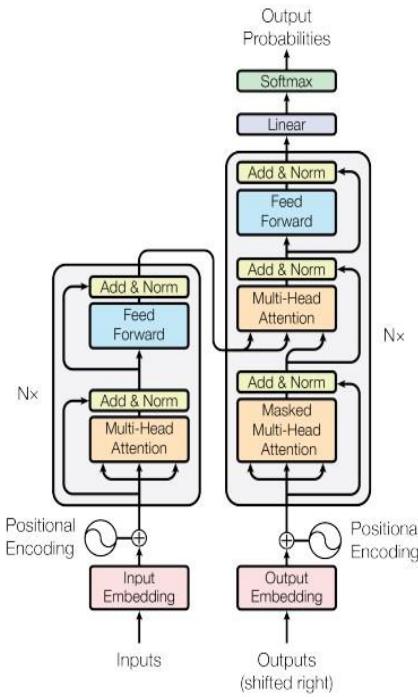


Figure 7 : Architecture du modèle Transformers

2.4.1 Mécanisme d'auto-attention

Le mécanisme d'auto-attention est une fonction qui permet à chaque élément d'une séquence de se concentrer sur d'autres éléments, même éloignés, en établissant une correspondance entre une requête (Query, Q) et un ensemble de paires clé (Key, K) / valeur (Value, V) (Vaswani et al., 2017). Ce mécanisme produit une sortie sous forme de somme pondérée des valeurs, où chaque pondération reflète l'importance de l'association entre la requête et la clé correspondante.

Voici les étapes principales :

- Calcul des scores d'alignement : On utilise une mesure de similarité, comme le produit scalaire, pour évaluer la correspondance entre chaque requête

et chaque clé. Ce score quantifie l'importance relative de chaque élément de la séquence pour la requête en cours.

- Normalisation des scores : Les scores d'alignement sont normalisés à l'aide de la fonction softmax, qui convertit les scores en probabilités. Chaque poids obtenu représente la contribution de chaque valeur au contexte de la requête.
- Repondération des valeurs : Les valeurs sont multipliées par les poids normalisés, puis additionnées pour produire une sortie pondérée. Cela donne une représentation contextuelle de la requête, synthétisant l'information la plus pertinente.

Les vecteurs Q, K, et V sont générés par des transformations linéaires appliquées à l'entrée, ce qui permet au modèle d'obtenir des représentations contextuelles uniques pour chaque élément de la séquence.

L'équation 8 traduit le mécanisme du cœur de l'auto-attention comme illustré à la figure 8.

$$\text{Attention } (Q, K, V) = \text{Softmax} \left(\frac{Qk^T}{\sqrt{dk}} \right) V \quad (8)$$

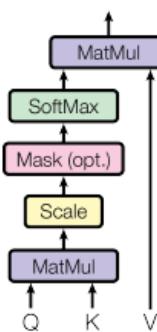


Figure 8 : Mécanisme d'auto-attention

Dans ce contexte, \mathbf{K}^T représente la transposée du vecteur \mathbf{K} , tandis que \mathbf{Q} est le vecteur de la requête et \mathbf{V} désigne le vecteur de la valeur. dk correspond à la dimension des vecteurs \mathbf{K} , \mathbf{Q} et \mathbf{V} . Enfin, **Softmax** est la fonction d'activation utilisée.

NB : $\sqrt{d_k}$ est utilisé pour la normalisation, car si le produit scalaire $\mathbf{Q} \cdot \mathbf{K}^T$ est très grand le gradient peut se retrouver à des valeurs très faibles.

2.4.2 Multi Head attention

L'amélioration de la précision des Transformers provient en grande partie du mécanisme de Multi-head attention, qui permet d'exécuter le mécanisme d'attention plusieurs fois, selon le nombre de têtes d'attention (h). Plutôt que d'appliquer une seule fois la fonction d'attention, celle-ci est exécutée h fois en parallèle. Chaque tête d'attention utilise une projection linéaire différente des vecteurs de requête (Query), clé (Key) et valeur (Value), permettant ainsi d'extraire des informations variées et de capturer différents types de relations entre les éléments de la séquence.

Après ces h calculs d'attention, les valeurs de sortie de chaque tête sont concaténées, puis passent par une transformation linéaire pour produire le vecteur final. Ce vecteur représente la combinaison des différentes perspectives d'attention obtenues, offrant une représentation plus riche et nuancée de la séquence d'entrée. Cette approche multi-tête renforce la capacité des Transformers à détecter des relations complexes tout en améliorant la précision des calculs. La dimension des valeurs et des clés est :

$d_{modèle} = 512/h = dv = dk$. Ce processus est illustré dans la *figure 9*.

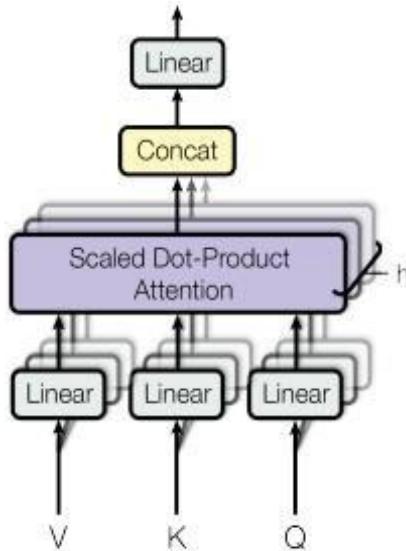


Figure 9 : Multi Head attention

2.4.3 Encodage positionnel

Les réseaux de neurones traditionnels ne reçoivent initialement aucune information sur la position des éléments dans une séquence, ce qui complique l'établissement de relations significatives, en particulier pour les longues séquences. En conséquence, le traitement de ces longues séquences peut être difficile et peu fiable pour les modèles traditionnels. Pour surmonter cette limite, la technique d'encodage positionnel associe un vecteur de position unique à chaque élément de la séquence, qui est ensuite ajouté aux embeddings (ou incorporations) des mots. Ce vecteur de position aide le modèle à repérer les relations entre les éléments adjacents et à améliorer la compréhension fine de l'ordre dans la séquence (Vaswani et al., 2017). Cette technique utilise les propriétés des fonctions trigonométriques, en appliquant des valeurs de sinus et cosinus pour encoder les positions des éléments de manière unique. Ces fonctions permettent de distinguer les positions paires des impaires

(Vaswani et al., 2017), facilitant ainsi la prise en compte de l'ordre dans la séquence selon les formules de l'équation **9** ci-dessous:

$$(9) \quad \begin{aligned} PE(Pos, 2i) &= \sin \left(\frac{Pos}{1000 \frac{2i}{d_{model}}} \right) & PE(Pos, 2i + 1) &= \cos \left(\frac{Pos}{1000 \frac{2i}{d_{model}}} \right) \end{aligned}$$

La constante **10 000** est utilisée pour l'échelle. **Pos** représente la position d'un élément dans la séquence, tandis que **i** désigne l'indice spécifique de cet élément dans la séquence. Enfin, **dmodel** fait référence à la dimension du modèle.

2.4.4 Incorporation des mots

Pour qu'une machine puisse traiter des mots, elle ne peut pas les utiliser directement sous la forme de texte brut. Il est nécessaire de transformer ces mots en une représentation compréhensible par la machine, sous la forme de vecteurs numériques denses. Cette transformation, appelée embedding ou incorporation des mots, consiste à représenter chaque mot dans un espace de dimensions réduites où les similarités et les relations entre les mots peuvent être capturées de manière sémantique.

Un embedding convertit les éléments d'une séquence en vecteurs numériques qui facilitent l'analyse des similarités sémantiques et syntaxiques entre eux, permettant au modèle d'établir des relations contextuelles et de mieux interpréter la structure du langage. Plusieurs techniques permettent de représenter les mots de manière compréhensible pour la machine, parmi lesquelles :

Bag-of-Words (BoW) : Représente chaque mot en fonction de sa présence ou absence dans le document, sans tenir compte de l'ordre.

Skip-gram : Utilisé dans les modèles de word2vec, il prédit le contexte d'un mot donné, permettant de capturer les relations contextuelles entre mots.

Continuous Bag-of-Words (CBOW) : Une autre approche de word2vec, CBOW prédit un mot cible en fonction de son contexte environnant.

Term Frequency-Inverse Document Frequency (TF-IDF) : Une technique qui pondère les mots en fonction de leur fréquence dans un document et de leur rareté dans l'ensemble des documents, renforçant les mots pertinents pour chaque document.

Nous avons choisi d'utiliser les réseaux de neurones profonds sans passer par les techniques de sélection de variables pour traiter les données radiomiques et paracliniques afin de prédire la réponse à la NAC. Pour commencer, des modèles de réseaux de neurones récurrents tels que GRU et LSTM sont présentés. Puis, les réseaux de neurones convolutifs qui sont utilisés en traitement des images et de données 1D. Enfin, les Transformers qui sont dotés d'un mécanisme d'attention pouvant traiter les relations complexes entre les éléments d'une séquence aussi longue qu'elle peut être.

En définitive, le traitement par la NAC est une approche essentielle pour le traitement de l'ostéosarcome. Elle réduit la taille de la tumeur, traite les cellules micro métastatiques précoces et favorise la chirurgie conservatrice. Les études antérieures présentent différentes approches pouvant permettre de prédire la réponse à la NAC qui constitue toujours un défi majeur sur le plan de traitement du patient.

Dans le chapitre suivant, nous présentons le processus de collecte et traitement des données afin d'étayer notre méthodologie.

CHAPITRE 3

COLLECTE, TRAITEMENT DES DONNÉES ET MÉTHODOLOGIE

Dans ce chapitre, nous allons développer des étapes phares de ce travail car les données constituent les éléments essentiels dans l'IA. Ces étapes vont permettre de réaliser une analyse fiable et rendre visible les données à traiter. En outre, l'étude des données collectées assure leur qualité et leur pertinence. Le traitement consiste à la préparation pour assurer le bon format attendu par les modèles. Enfin, la méthodologie décrit les techniques et approches pour répondre aux questions de recherches.

3.1 Donnée radiomiques

Au cours des dernières décennies, les innovations en imagerie médicale avec de nouveaux appareils et de protocoles standardisés conduisent vers l'imagerie quantitative (Lambin et al., 2012). Elle consiste à effectuer une extraction rapide d'innombrables caractéristiques quantitatives, qualitatives ou sémantiques à partir des images de tomographie par émission de positron ou par résonance magnétique. C'est grâce aux calculs à haut débit que ce processus d'extraction aboutit à la conversion des images médicales en données exploitables (Gillies et al., 2016). De plus, les images présentent des informations invisibles à l'œil nu que ces caractéristiques sont capables de faire paraître. Ceci pourrait aider la médecine

à améliorer le diagnostic, le pronostic et le traitement des patients pour bâtir une médecine de précision. Ce processus conceptuellement simple implique des étapes discrètes (Kumar et al., 2012; Lambin et al., 2012).

3.1.1 Identification de la région d'intérêt

La plupart des patients qui présentent des métastases ont plusieurs lésions dans d'autres parties du corps. De ce fait, le cœur de la radiologie en oncologie est basé sur l'identification des volumes tissulaires qui ont des valeurs pronostiques que nous retrouvons sur un ou plusieurs sites tumoraux (Gillies et al., 2016). En raison des limites spatiales et de contraste des images numériques, l'identification de ces volumes et sous volumes qui contiennent des valeurs pronostiques peut être difficile, alors que dans la pratique de la radiomique, les sous-volumes d'intérêt sont capturés et rajoutés dans l'analyse. Ainsi, la philosophie de base repose sur l'ingénierie des procédés qui consiste à capturer le maximum d'information possible (Vial et al., 2018). Ce concept d'exploitation des images pour l'identification des régions physiologiquement distinctes au sein des lésions est décrit par les auteurs Gatenby et al (Gatenby et al., 2013). Cette méthode d'exploitation des images a permis d'utiliser des images avec différents paramètres (T1, T2, FLAIR) qui peuvent être combinés pour la production de Region Of Interest (ROI) comme présenté dans la figure 9.

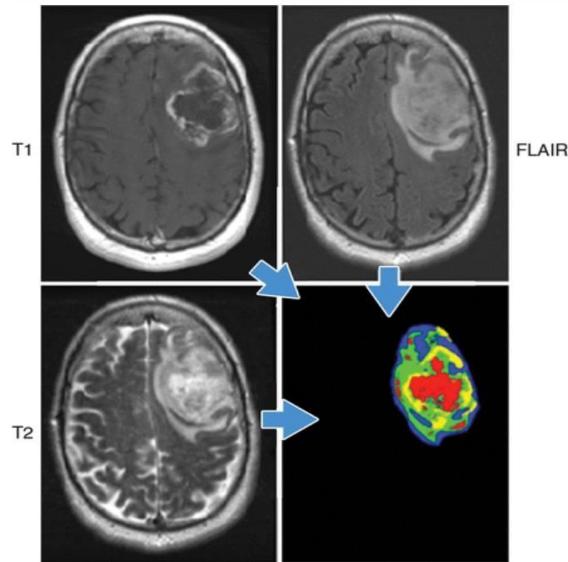


Figure 10 : Habitats chez un patient atteint de glioblastome multiforme(Gillies et al., 2016).

3.1.2 Segmentation

La segmentation constitue l'étape, la plus difficile, la plus critique et la plus controversée de la radiomique (Gillies et al., 2016). Car c'est à partir des volumes et sous volumes segmentés que les données seront extraites. Pour réussir cette tâche de délimitation des régions d'intérêt, des techniques manuelles, semi-automatiques et automatiques sont appliquées sur les images médicales. Leurs succès sont facilités par des opérations de prétraitement en amont telles que la correction des artefacts, la normalisation de l'intensité des pixels, la réduction de bruit, etc.

Dans le domaine de la vision par ordinateur appliquée à l'imagerie médicale, des solutions très performantes dans la segmentation des images sont développées. Vafaeikia et al (Vafaeikia et al., 2024) se sont servis de l'architecture du réseau U-Net dans leur recherche pour produire un modèle de segmentation des images IRM de la tumeur de gliome pédiatrique

de bas grade. Leur étude montre que, la segmentation du ROI est automatisable avec moins de ressources humaines, plus de précision et un gain énorme de temps.

3.1.3 Extraction des caractéristiques

L'extraction des caractéristiques radiomiques qui suit l'étape de la segmentation, consiste à obtenir des mesures quantitatives qui décrivent les propriétés des régions d'intérêts (ROI). Étant donné qu'il existe une diversité de formules et méthodes d'extraction de caractéristiques radiomiques disponibles, les travaux de Zwanenburg et al. (Zwanenburg et al., 2019) ont fourni des recommandations. Ces recommandations qui établissent un consensus pour le calcul des caractéristiques radiomique (Van Timmeren et al., 2020).

De plus, divers types de filtres, tels que les filtres à ondelettes ou gaussiens, sont souvent appliqués lors de l'extraction des caractéristiques. Des outils permettent d'accomplir cette tâche en appuyant simplement sur un bouton.

En résumé, l'obtention des caractéristiques radiomiques nécessite généralement trois étapes. La figure 11 montre que la première étape consiste à délimiter les régions d'intérêts (ROI) sur l'image. Ensuite, la segmentation est réalisée pour isoler la zone délimitée. Enfin, des outils sont utilisés pour extraire les caractéristiques.

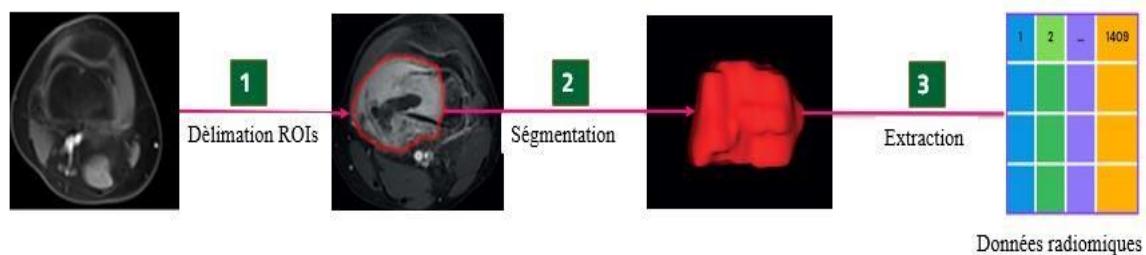


Figure 11 : Processus d'extraction des caractéristiques radiomiques

3.2 Collecte de données

Les données de ce travail sont issues d'une étude rétrospective des patients atteints d'ostéosarcome confirmés par la chirurgie et la pathologie entre janvier 2016 et mai 2020 (Zhang et al., 2021). Nous avons initialement une cohorte de 172 patients atteints d'ostéosarcome et qui ont subi une NAC et un traitement chirurgical. Dans le but de garantir la fiabilité et la richesse des données, des critères d'inclusion et d'exclusion sont établis comme suit :

Les critères d'inclusion :

1. Ostéosarcome confirmé par histopathologie ;
2. IRM réalisées dans la semaine qui précède la NAC et à la fin des deux cycles de NAC ;
3. Plus de deux (2) cycles de traitement de la NAC sont effectués à l'hôpital local.

Les critères d'exclusion :

1. Tous les patients qui n'ont pas suivi plusieurs séquences d'IRM avant ou après la NAC ;
2. Patients qui ont subi une intervention chirurgicale sans la NAC ;
3. Les patients qui n'ont pas pu réussir les deux cycles de la NAC ou achever le traitement ;
4. Les patients qui étaient incapables de fournir des données IRM complètes.

En se basant sur les critères d'inclusion et d'exclusion, 20 patients sont exclus d'après le troisième critère d'inclusion. À cela s'ajoutent 15 patients qui n'ont pas pu remplir le troisième critère d'exclusion, de même, 35 patients sont exclus par le biais du quatrième critère

d'exclusion. Finalement, 102 patients sont retenus dans cette étude. Ce processus de sélection est illustré par la *figure 12*.

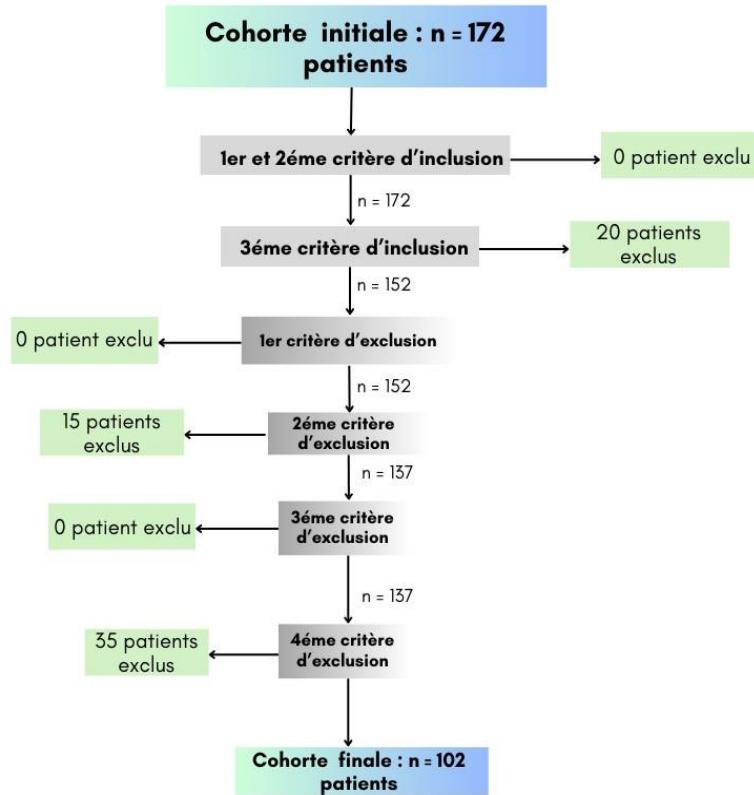


Figure 12 : Processus de sélection des patients

Après avoir appliqué les filtres de sélections sur le jeu de données initial, 102 patients sont retenus pour identifier les régions ou ROI. Le jeu de données est composé de 1409 caractéristiques quantitatives extraites à partir des images IRM et est réparti en quatre groupes.

Groupe 1 : 18 caractéristiques qui exposent les statistiques de premier ordre

Ces caractéristiques décrivent la distribution des intensités de voxels (Mayerhoefer et al., 2020) de façon quantitative sans tenir compte de la relation spatiale entre les pixels (Lambin et al., 2017).

Groupe 2 : 14 caractéristiques basées sur la forme et la taille

Elles présentent des informations concernant les changements irréguliers de la tumeur. Car des chercheurs ont montré que les tumeurs avec un rapport surface-volume plus élevé peuvent présenter des formes plus irrégulières avec un degré de malignité plus élevé vis-à-vis des masses rondes (Haralick et al., 1973).

Groupe 3 : 75 caractéristiques de texture de second ordre

Ce sont des caractéristiques couramment utilisées qui se servent des statistiques d'intensité de second ordre pour analyser les variations des attributs des images. Leur objectif est de mesurer l'étendue des irrégularités dans les tissus (Vial et al., 2018).

Groupe 4 : 1302 caractéristiques de filtrage d'ordre supérieur

Elles constituent les statistiques de premier ordre et de texture. L'application de filtres Lapaciens, logarithmiques, exponentiels et ondelettes sur les images a permis d'obtenir ces caractéristiques (Zhang et al., 2021).

3.3 Applications

En neuro-oncologie, les auteurs (Lambin et al., 2017; Mohy-ud-Din & Rathore, 2020) ont cité diverses applications possibles avec l'exploitation des données radiomiques. Entre autres, par le biais des données radiomiques, nous pouvons identifier des biomarqueurs à partir desquels nous pouvons prédire :

La sélection du traitement : il consiste à sélectionner le plan de traitement le plus adéquat pour un patient. Il considère le degré et le stade de la tumeur de même que la santé et les préférences générales du patient.

La survie du patient : elle mesure généralement le pourcentage des patients qui sont encore en vie pendant un certain temps après le diagnostic.

Réponse au traitement : l'efficacité du traitement pour contrôler ou réduire la croissance d'un cancer.

Identification d'une région récidive : il consiste à identifier la récurrence du cancer dans une zone après le traitement.

3.4 Méthodologie

Dans cette partie, nous présentons la statistique descriptive des données ainsi que les outils technologiques utilisés notamment Python avec les bibliothèques spécialisées pour la manipulation des données et l'implémentation des modèles. Ces derniers nécessitent des fonctions d'activations et des métriques pour mesurer l'efficacité des modèles.

3.4.1 Statistiques descriptives

Notre jeu de données se compose de 1409 caractéristiques radiomiques, du stade chirurgical et de 102 individus répartis en deux classes. Pour obtenir une représentation plus précise, nous étudierons l'équilibrage des classes et l'analyse de la variance des caractéristiques.

A Équilibrage des classes

En apprentissage supervisé, l'équilibrage des classes est un facteur déterminant lors de l'entraînement du modèle, car il peut impacter sur les performances lors de l'entraînement, plus probablement à la phase de test. Dans ce travail de recherche, nous avons deux classes (réponse positive = 1 et réponse négative = 0) dont leur répartition d'après la *figure 13* indique une légère prédominance entre les classes. Cette prédominance ne nécessite pas de retraitre les données faisant appel à des techniques spécifiques de traitement de déséquilibre entre elles. Signalons à ce propos que, la classe de la réponse négative représente 53,92% du jeu de données et que celle de la réponse positive représente 46,08%. Cette légère prédominance de 7,84% de la classe négative pourrait influencer certains modèles lors de l'apprentissage ou du test. Dans la suite de notre démarche, nous veillerons sur son impact lors de l'analyse des performances des modèles sur l'ensemble de tests.

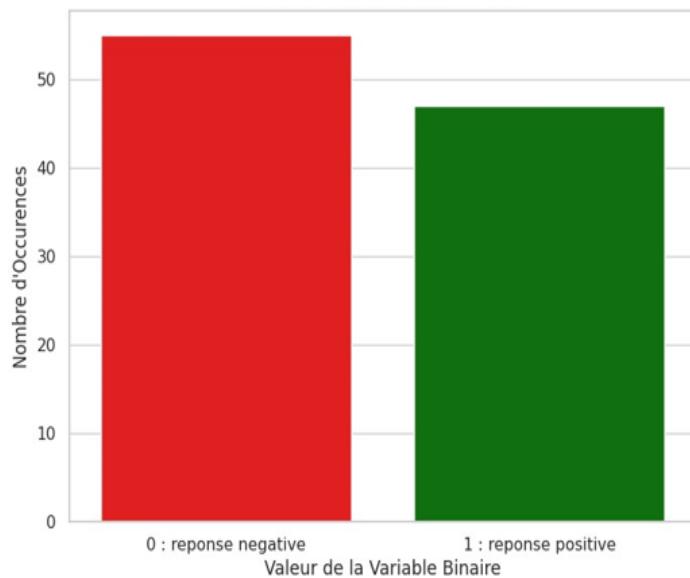


Figure 13 : Répartition des classes

B Caractéristiques à variance nulle

D'un point de vue statistique, une variance nulle signifie que toutes les valeurs d'une variable sont identiques pour toutes les observations. Ces observations sont inutiles pour le processus de modélisation, car elles n'apportent aucune information distinctive permettant de différencier les observations. Leur présence peut non seulement alourdir le modèle, mais aussi augmenter la complexité de manière inutile, sans contribuer à améliorer les performances prédictives.

Dans notre étude, les données initiales comprennent 1409 caractéristiques radiomiques combinées avec le stade chirurgical de la maladie pour 102 observations. En appliquant un filtre pour détecter les variables à variance nulle, nous avons identifié 161 caractéristiques dont les valeurs sont constantes sur l'ensemble des observations. En d'autres mots, ces variables n'apportent aucune information pour les modèles d'apprentissage automatique.

Par conséquent, nous avons retiré ces variables du jeu de données finales, afin de simplifier le modèle et de concentrer l'analyse sur des caractéristiques potentiellement plus informatives et pertinentes pour la prédiction de la réponse au traitement de l'ostéosarcome par la NAC.

3.5 Les fonctions d'activation

Les données passent dans un réseau qui effectuera leur transformation par le biais des fonctions d'activation (linéaires ou non) (Bhatt et al., 2021). Ainsi, la production des résultats à partir de ces données est faite grâce au processus de forward propagation (Goodfellow, 2016). De ce fait, les fonctions d'activation sont utilisées en fonction de la tâche comme la régression, la classification binaire ou la classification multiclasse en fonction de la sortie

souhaitée. Dans la suite, nous allons présenter les fonctions d'activation non linéaires les plus utilisées dans le DL.

3.5.1 ReLU

La fonction d'activation ReLU (Rectified Linear Unit) est largement utilisée dans les réseaux de neurones profonds en raison de sa simplicité et de son efficacité. ReLU est obtenue par la formule ci-dessous :

(10)

$$ReLU(x) = \max(0, x)$$

Cette formule dans l'équation (10) confère le rôle de filtre des valeurs supérieures ou égales à zéro (voir figure 15) à la fonction. Car toutes les valeurs négatives sont converties à zéro et le reste est maintenu. ReLU est moins coûteuse en termes de calcul, de plus, elle aide à l'atténuation du gradient faisant converger l'entraînement plus rapidement.

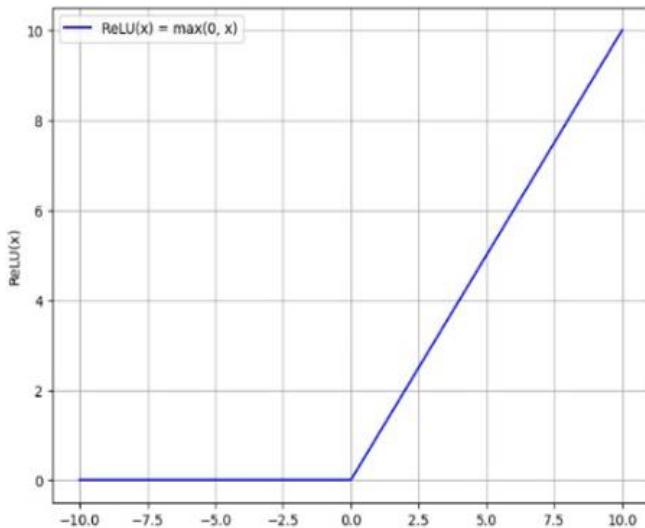


Figure 14 : Fonction d'activation ReLU.

3.5.2 Softmax

C'est une fonction que nous utilisons souvent dans les dernières couches de réseaux de neurones pour la classification binaire ou multi classes. La sortie de la fonction Softmax est un vecteur de probabilités, où chaque élément du vecteur représente la probabilité d'appartenance d'une classe donnée, et la somme de toutes les probabilités dans le vecteur est égale à 1. Notons que dans les tâches de classification binaire sans relation d'ordre, *Softmax* est plus appropriée, car sa sortie ne fournit aucune relation d'ordre entre les classes contrairement à sigmoïde. De ce fait, au cours de ce travail, nous allons l'utiliser pour classifier les patients qui ont donné une réponse positive ou négative à la NAC. Par le biais de l'équation 12, nous obtenons sa sortie $\sigma(z)_j$:

(11)

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

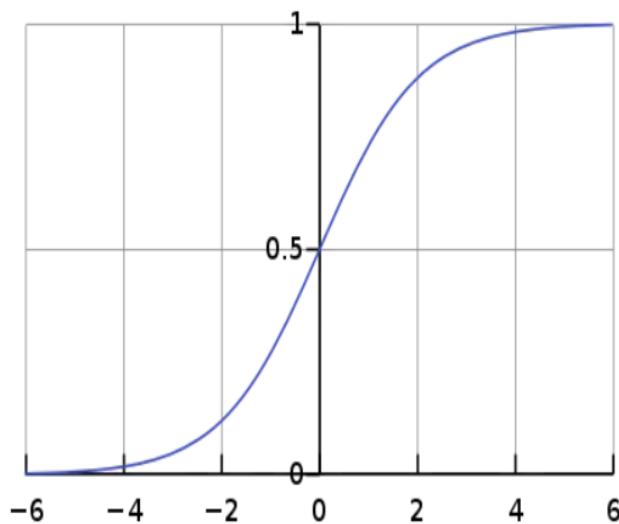


Figure 15 : Fonction d'activation softmax

3.6 Métriques d'évaluation de performance

Les mesures discriminantes pour la performance de classification sont très importantes lors de l'entraînement des classificateurs. Elles sont aussi importantes que l'algorithme lui-même. De plus, elles représentent la première étape clé d'une expérimentation réussie (Gu et al., 2009). Dans notre étude, nous nous basons généralement sur les métriques contenues dans le module *metrics* de *sklearn* (matrice de confusion, précision, rappel, F1 score et l'exactitude), la courbe ROC (AUC) et la moyenne géométrique.

3.6.1 Matrice de confusion

La matrice de confusion est un tableau essentiel pour l'évaluation de la performance des modèles de Machine Learning dans les tâches de classification. Elle offre une visualisation sur la performance des modèles qui comptabilisent les prédictions qui sont correctes ou incorrectes en fonction des classes réelles et prédites. Dans notre étude, nous avons deux classes : la réponse positive et négative. D'après le tableau ci-dessous, les colonnes représentent les classes prédites et les lignes les classes réelles.

Tableau 1 : Matrice de confusion

Classe	Prédite : réponse négative	Prédite : réponse positive
Réelle : réponse positive	Vrai Négatif : VN	Faux Positif : FP
Réelle : réponse négative	Faux Négatif : FN	Vrai Positif : VP

VP : Nombre de fois où le modèle a correctement prédit la classe positive.

VN : Nombre de fois où le modèle a correctement prédit la classe négative.

FP : Nombre de fois où le modèle a prédit la classe **positive**, alors que la classe réelle est **négative**.

FN : Nombre de fois où le modèle a prédit la classe **négative**, alors que la classe réelle est **positive**.

3.6.2 Exactitude

L'exactitude est la proportion des prédictions exactes sur l'ensemble des données de test.

(12)

$$\text{Exactitude} = \frac{VP + VN}{FP + VP + FN + VN}$$

3.6.3 Précision

Elle mesure de façon plus robuste la performance du modèle pour des classes spécifiques. La formule ci-dessous est établie pour avoir sa valeur à partir de la matrice de confusion.

(13)

$$\text{Précision} = \frac{VP}{VP + FP}$$

3.6.4 Rappel ou Sensibilité, Spécificité

Le rappel et la spécificité permettent d'évaluer séparément et respectivement les cas positifs et négatifs d'un classificateur. Ils sont souvent utilisés dans le domaine biomédical (Sokolova et al., 2006) et sont calculés à partir des formules ci-dessous :

(14)

$$\text{Rappel ou Sensibilité} = \frac{VP}{VP + FN} \quad ; \quad \text{Spécificité} = \frac{VN}{VP + FN}$$

3.6.5 F1 score

Elle résulte de la combinaison de la précision et du rappel, fournissant une évaluation équilibrée de la performance du modèle (Sokolova et al., 2006).

(15)

$$F1 \text{ score} = 2 * \frac{\text{Précision} \cdot \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

3.6.6 Courbe ROC (AUC)

La courbe ROC illustre graphiquement la performance d'un classificateur binaire en représentant le taux de vrais positifs par rapport au taux de faux positifs. L'aire sous la courbe (AUC) quantifie cette performance globale, avec une valeur de 1 qui indique une discrimination parfaite et 0,5 une discrimination aléatoire. Ces indicateurs combinés offrent une évaluation visuelle et numérique complète de la capacité du modèle à distinguer les classes. Ils permettent ainsi une compréhension approfondie des forces et des limites du

classificateur dans sa tâche de discrimination. Les travaux de (Nahm, 2022) ont donné des intervalles d'appréciation de la valeur de l'AUC selon sa valeur.

(16)

$$AUC = \frac{(Sensibilité + Spécificité)}{2}$$

Area Under the Curve (AUC)	Interprétation
$0.9 \leq AUC$	- Excellent
$0.8 \leq AUC < 0.9$	- Bien
$0.7 \leq AUC < 0.8$	- Passable
$0.6 \leq AUC < 0.7$	- Faible
$0.5 \leq AUC < 0.6$	- Insuffisant

3.6.7 Moyenne géométrique

La moyenne géométrique, comme le F1-score, équilibre équitablement sensibilité et spécificité. Elle favorise les classificateurs performants sur ces deux critères. Un classificateur avec un grand écart entre sensibilité et spécificité aura une moyenne géométrique faible. Cette approche pénalise les modèles dominants sur une seule mesure, offrant ainsi une évaluation plus juste et globale de la performance. Contrairement aux matrices classiques qui privilégient un seul aspect, elle garantit un équilibre entre la détection des cas positifs (sensibilité) et la distinction des cas négatifs (spécificité). Cette méthode est particulièrement utile dans des domaines comme la médecine ou la sécurité, où les faux positifs et négatifs peuvent entraîner des conséquences significatives. Elle est obtenue grâce à l'équation 18 ci-dessous :

(18)

$$\text{Moyenne géométrique (G-mean)} = \sqrt{\text{Sensibilité} * \text{Spécificité}}$$

Dans le but d'évaluer la valeur de la moyenne géométrique par intervalle, Boucher et Badri (Boucher & Badri, 2018) se sont basés sur des intervalles d'appréciation définis comme suit.

Moyenne géométrique (G-mean)	Interprétation
$0.9 \leq \text{G-mean}$	Excellent
$0.8 \leq \text{G-mean} < 0.9$	Bien
$0.7 \leq \text{G-mean} < 0.8$	Passable
$0.6 \leq \text{G-mean} < 0.7$	Faible
$0.5 \leq \text{G-mean} < 0.6$	Insuffisant

3.7 Outils et technologies

Ce travail de recherche nécessite l'utilisation de certains outils technologiques pour garantir son bon déroulement, qu'il s'agisse du développement des modèles, de leur entraînement ou de leur évaluation. Ces outils technologiques offrent plusieurs avantages qui combinent puissance, flexibilité et simplicité surtout dans le cadre collaboratif. Ces outils ont facilité nos expérimentations. Dans la suite de ce document, nous présentons brièvement chacun de ces outils.

3.7.1 Google Colab

Google Colab est un outil collaboratif basé sur le cloud offert par Google qui permet aux développeurs d'accéder à un environnement de développement et une collection de

bibliothèques simplifiés en Python, spécifiquement destinés à l'apprentissage machine (Google, 2025b). Pour les chercheurs qui n'ont pas accès à des machines de hautes performances, Google Colab offre un accès gratuit aux CPU, GPU et TPU avec une mémoire de 15 Go. Cette dernière permet de lancer l'entraînement des modèles qui ne nécessitent pas beaucoup de ressources. Cependant, si le modèle devient complexe et très gourmand en ressources, Google Colab offre une version payante qui permet d'augmenter les ressources selon tes besoins réels. De plus, il est directement connecté à un système de gestion de fichier (Google Drive) de façon sécuritaire, ce qui rend encore plus facile la collaboration avec d'autres chercheurs. Enfin, il ne requiert aucune configuration au préalable pour l'environnement de développement qui intègre presque toutes les bibliothèques de base du Python.

3.7.2 Google Drive

Google Drive est une plateforme de stockage offrant la possibilité d'organiser, de conserver et de partager des fichiers à travers le cloud avec un niveau de sécurité élevé(Google, 2025a)s. Dans ce travail, cet outil nous a facilité la gestion des données, la sauvegarde des notebooks et des modèles. De plus, son accessibilité pendant l'exécution des scripts a facilité l'entraînement de nos modèles depuis Google Colab.

3.7.3 Python

Au cours de ce travail, nous avons utilisé Python pour le développement, l'entraînement ou l'évaluation des modèles. En raison de sa simplicité la quasi-totalité des modèles d'apprentissage machine sont développés avec Python. Les bibliothèques Python, développées par la communauté, proposent des outils dédiés à l'apprentissage automatique tels

que PyTorch et TensorFlow, ainsi qu'au traitement des données tabulaires avec des bibliothèques comme Pandas et NumPy(Python, 2025). L'intérêt de ces bibliothèques est que toutes les fonctions de base sont déjà intégrées.

Pandas

Pandas est une bibliothèque open source dédiée à la manipulation et à l'analyse de données, largement utilisée en science de données(Pandas, 2025). Elle offre des structures flexibles permettant la manipulation des grandes quantités de données. Pandas permet d'effectuer des opérations complexes comme le filtrage, la fusion, l'agrégation, etc. C'est une bibliothèque très utilisée dans le prétraitement des données en apprentissage automatique.

NumPy

NumPy ou Numerical Python est une bibliothèque open source pour les calculs scientifiques, offrant un support pour les tableaux multidimensionnels ainsi qu'une large gamme de fonctions mathématiques(NumPy, 2025). Grâce à sa capacité à effectuer des opérations vectorisées, NumPy permet d'effectuer des calculs rapides et efficaces. C'est une base pour les autres bibliothèques telles que Pandas, SciPy et Matplotlib qui s'appuient sur ces structures de données performantes. NumPy est largement utilisée en apprentissage automatique, car elle permet de manipuler efficacement les matrices et les tenseurs qui sont les éléments clés des réseaux de neurones.

One Hot Encoding

Le One Hot Encoding est une technique utilisée en apprentissage automatique pour la transformation des données catégoriques en vecteurs binaires. Il permet ainsi aux modèles de les traiter numériquement sans l'introduction d'ordre artificiel. Pour ce faire, chaque catégorie est représentée par un seul vecteur où une seule position est activée avec 1 et les autres sont désactivées par 0 (Aurélien, 2019).

C'est une méthode de présentation du label catégoriel pour les problématiques de classification dans lesquelles une indépendance des catégories est obligatoire. Cependant, le One Hot Encoding peut poser des problèmes lorsqu'il existe plusieurs catégories qui vont accroître la dimensionnalité nécessitant plus d'espace mémoire. Dans ce cas, les embeddings constituent une alternative (Gu & Sung, 2021).

Tensorflow

Tensorflow est une bibliothèque d'apprentissage machine créée en 2015 par Google. Elle est largement utilisée pour la recherche et les applications industrielles (Abadi et al., 2016). Son architecture repose sur des graphes computationnels statiques, optimisés pour un déploiement à grande échelle. Bien qu'elle ait été d'abord considérée comme moins flexible que PyTorch, Tensorflow a su évoluer avec l'introduction de l'exécution en mode Eager qui permet un calcul plus dynamique (Suhan et al., 2021). Un des principaux atouts de Tensorflow est son écosystème solide, comprenant Tensorflow Serving pour les déploiements en production et Tensorflow Lite pour les appareils mobiles (Ilame, 2024). Grâce à son support pour les GPU et TPU, Tensorflow donne des performances optimales pour des modèles complexes tels que les LSTM, les CNN et les Transformers

PyTorch

PyTorch est une bibliothèque open source d'apprentissage machine développée par Facebook en 2016 comme évolution de Torch. PyTorch a révolutionné la recherche grâce à ses graphes computationnels dynamiques (Paszke et al., 2019). Ce mécanisme permet d'ajuster les modèles en temps réel facilitant le débogage et l'expérimentation. PyTorch est doté d'une fonction d'autodifférenciation automatique, appelée Autograd. Cette dernière simplifie le calcul des gradients en rendant l'entraînement des modèles plus intuitifs. PyTorch est utilisé dans divers domaines, tels que le traitement d'images avec les CNN et le NLP avec des LSTM et des Transformers offrant une grande flexibilité (Ketkar et al., 2021). De plus, il dispose d'un

support natif pour l'accélération GPU via CUDA, ce qui améliore significativement les performances des calculs intensifs.

3.7.4 Ressources matérielles

Aujourd'hui, les ressources matérielles participent activement dans le processus d'entraînement des modèles. Car pour des modèles complexes nécessitant une bonne capacité en termes de ressources, un environnement simple ne pourra jamais le supporter. De plus, les modèles DL utilisent beaucoup de données, surtout les Transformers. En raison de cela, ils nécessitent des environnements capables de supporter leur exécution. Le tableau 2 affiche les ressources de Google Colab et celle de la machine CUDA de l'UQTR.

Tableau 2 : Ressources matérielles

Google Colab		Machine CUDA UQTR	
RAM GPU	15 Go	Processeur	20 cœurs partagés
RAM système	12.7 Go	Mémoire	100 GB
Espace utilisateur	78.2 Go	Disque SSD	500 GB

3.8 Expérimentations

Cette section détaille les expérimentations menées au cours de cette étude. Pour l'entraînement des modèles LSTM, GRU et FCNN, les données sont réparties à hauteur de 70% pour l'entraînement et 30% pour le test. En revanche, le modèle TabTransformer est entraîné avec 90% des données, réservant les 10% restants pour le test. Nous examinons les courbes d'apprentissage, en particulier celles représentant la fonction de perte et la précision,

afin d'évaluer les performances des modèles. Par la suite, nous procédon à une interprétation pour mieux comprendre les comportements observés.

3.8.1 Le modèle LSTM

Au cours de l'entraînement du modèle LSTM, nous notons que la *figure 17* de la fonction de perte sur les données d'entraînement décroît progressivement au cours des époques, contrairement à la courbe de validation croît. Cela démontre que le modèle n'arrive pas à performer sur les données qu'il n'a jamais vues. Ce qui montre que le modèle fait surapprentissage. En revanche, la précision sur les données d'entraînement accroît considérablement.

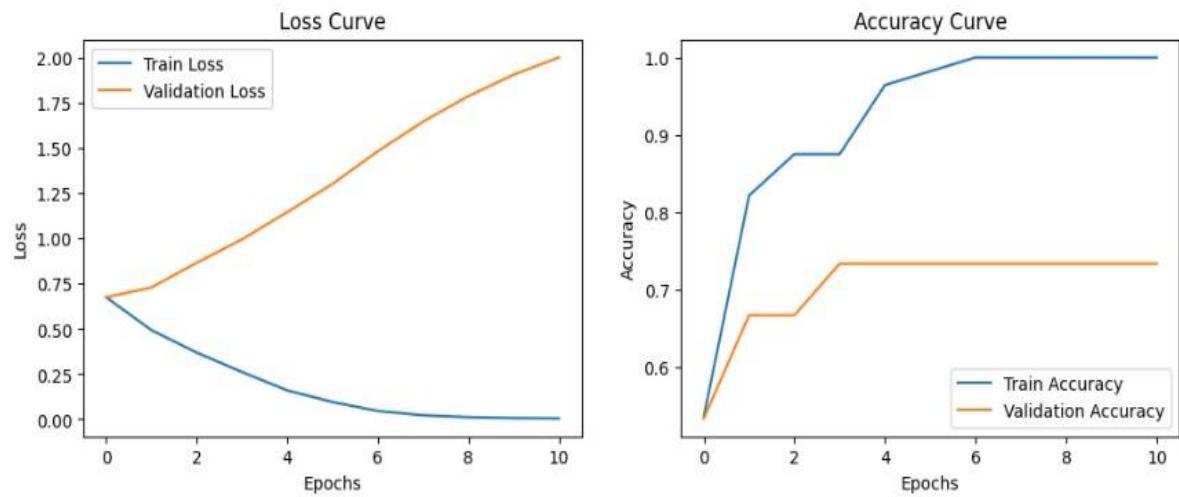


Figure 16 : Courbe de variation de la fonction de perte et de précision du modèle LSTM

3.8.2 Le modèle GRU

La *figure 18* affiche les courbes d'apprentissage de la fonction de perte et de précision du modèle GRU. La courbe de la fonction de perte montre une décroissance rapide et continue sur les données d'entraînement et se stabilise à la septième époque. Après elle atteint une valeur stable qu'elle va maintenir sur le reste des époques. Cela indique que le modèle parvient à minimiser les erreurs sur les données d'entraînement. Concernant la validation elle s'accroît progressivement comparé à celle de l'entraînement. La courbe de la précision montre une croissance très rapide sur les données d'entraînement jusqu'à atteindre la valeur 1. Cependant, celle de la validation croît et stagne à 0.75. On note un décalage significatif entre les deux courbes et une ressemblance à celle du LSTM. Ces évolutions traduisent un surapprentissage. Ainsi, GRU n'est pas capable de généraliser.

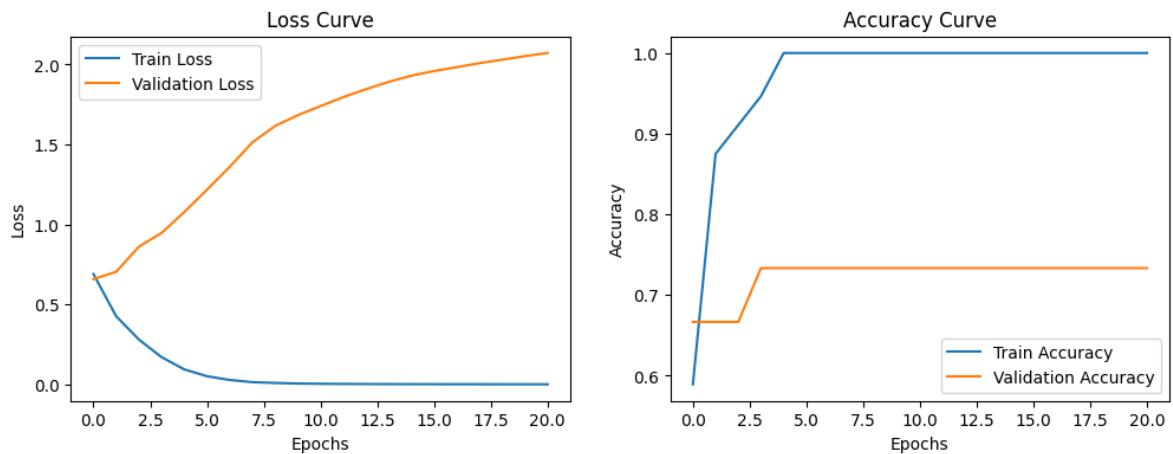


Figure 17 : Courbes de variation de la fonction de perte et de la précision du modèle GRU

3.8.3 Le modèle FCNN

Le modèle FCNN basé sur les CNN est composé de quatre sous-réseaux de convolutions de la *figure 19* spécialisés qui analysent différents groupes de caractéristiques

contenues dans notre jeu de données. Notons que le groupe des caractéristiques de filtrage d'ordre supérieur contient 161 variables de variance nulle. Ces dernières sont exclues lors du processus de prétraitement des données. Les sorties de ces sous-réseaux sont combinées et servent d'entrée à un réseau de classification avec une fonction d'activation softmax. Cette méthode pourrait favoriser une meilleure intégration et capturer de différents éléments du jeu de données. Cet apprentissage regroupé peut contribuer à améliorer le taux de réussite de la classification finale (Mehta et al., 2023). Toutefois, elle implique une plus grande complexité et exige davantage d'efforts en matière d'optimisation et de gestion des sous-réseaux individuels. Dans l'ensemble, cette stratégie semble constituer une solution robuste pour des jeux de données à haute dimension.

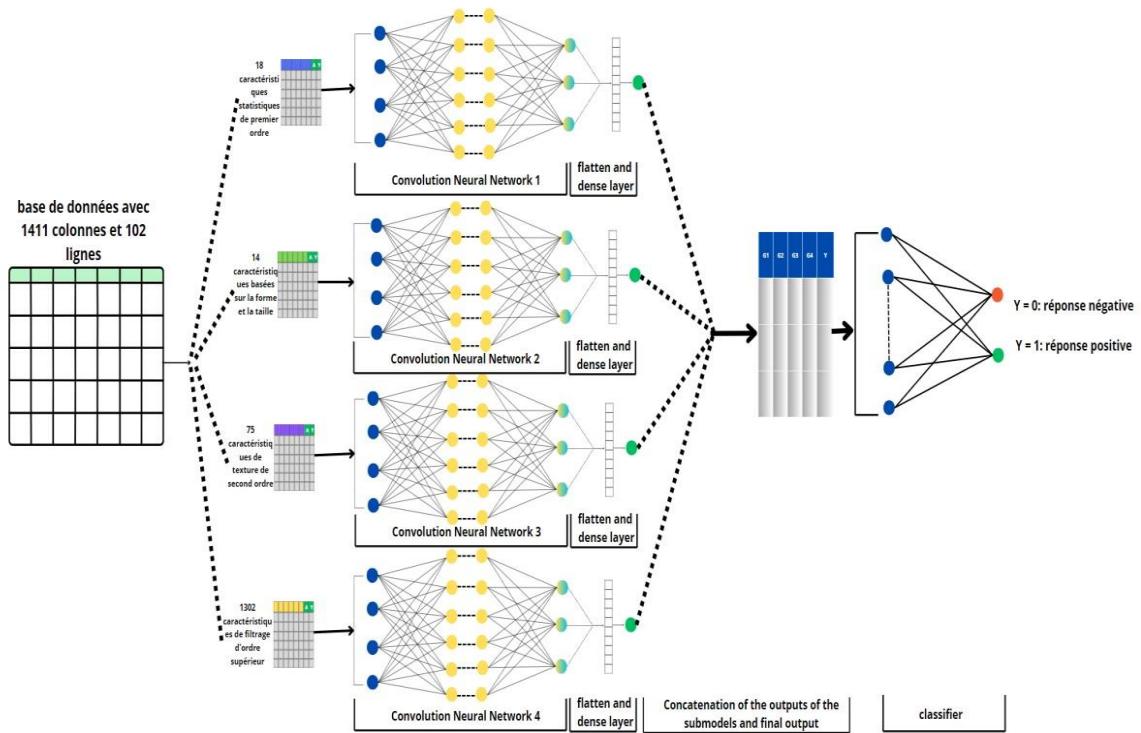


Figure 18 : Architecture FCNN

La *figure 20* montre la variation des fonctions de perte et de précision lors de l’entraînement du modèle FCNN. Nous constatons que la fonction de perte décroît de façon régulière sur les données d’entraînement et de validation au cours des époques. Ce qui montre que le modèle arrive à réduire l’erreur lors de son apprentissage et sa validation. De plus, la courbe de précision du modèle accroît progressivement et atteint presque la valeur 1. La courbe de validation présente des tendances similaires à celle de l’entraînement en termes de variation, mais elle reste légèrement inférieure. En définitive, on note que les variations des courbes n’affichent pas un signe de surapprentissage ou sous-apprentissage. Donc, le modèle a de belles performances et s’adapte bien aux données.

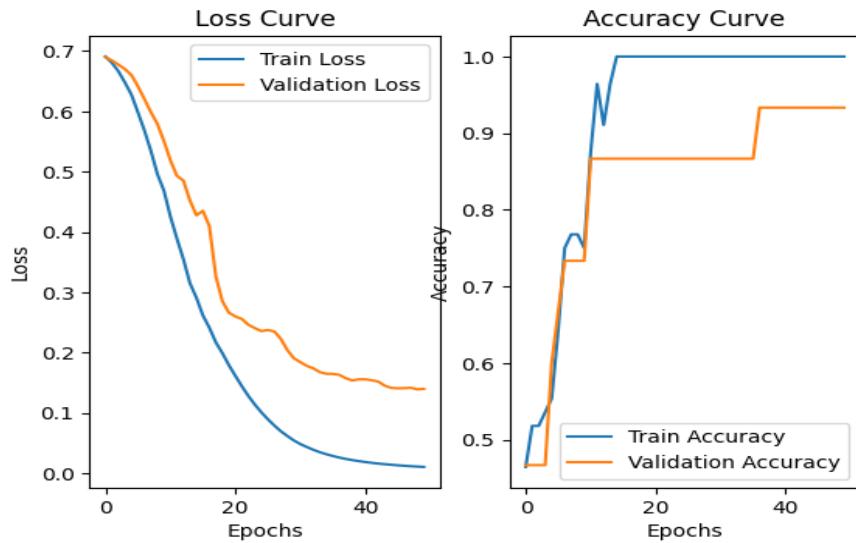


Figure 19 : Courbes de variation de la fonction de perte et de précision du modèle FCNN

3.8.4 Le modèle Tab Transformers

Le Tab Transformers est une extension de l’architecture du modèle Transformers (voir figure 21) pour le traitement de données numériques tabulaires (Huang et al., 2020). Il s’inspire du succès des Transformers, qui ont révolutionné le domaine du NLP grâce à leur mécanisme

d'auto-attention qui parvient à capturer des relations complexes entre les variables. Dans notre cas, cette architecture s'adapte aux données radiomiques. D'abord, les données sont normalisées via Layer Normalisation pour stabiliser leur distribution facilitant le processus d'entraînement. Ensuite, elles passent par un bloc de Transformers qui intègre des couches d'attention multi-têtes permettant l'apprentissage des relations complexes entre les variables. Enfin, le perceptron multi couches produit la prédiction finale qui sera comparée par la cible grâce à la fonction de perte qui permet d'ajuster les paramètres du modèle lors de l'entraînement. L'entraînement est lancé sur une machine CUDA avec 90% des données pour l'entraînement, la validation et 10% pour le test. Nous signalons ici que plusieurs expérimentations ont été effectuées sur la base d'une séparation de 70 et 30% du jeux de données. Mais le modèle n'était pas aussi performant ce qui nous amène à augmenter les données d'entraînement et diminuer les données de tests.

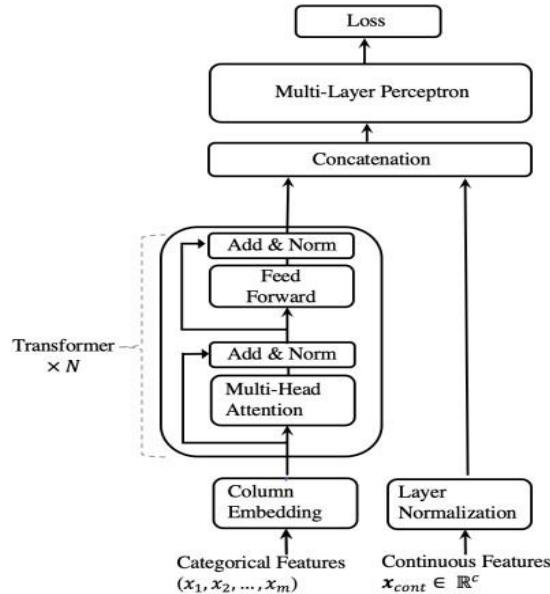


Figure 20 : Architecture du modèle Tab Transformers

La courbe de précision de la *figure 22* sur les données d'entraînement croît graduellement jusqu'au voisinage de 0.95. Ce qui montre que le modèle arrive à ajuster les paramètres lors de son entraînement. Cependant, la précision sur les données de validation a augmenté au cours de l'entraînement jusqu'à 0,82. Après ce point culminant, elle décroît de façon continue. En résumé, les courbes révèlent que Tab Transformers performe bien sur les données d'entraînement. En ce qui a trait aux données de validation, le modèle, suggère une possibilité de surapprentissage à partir de l'époque 7. Tab Transformers n'arrive pas à bien généraliser.

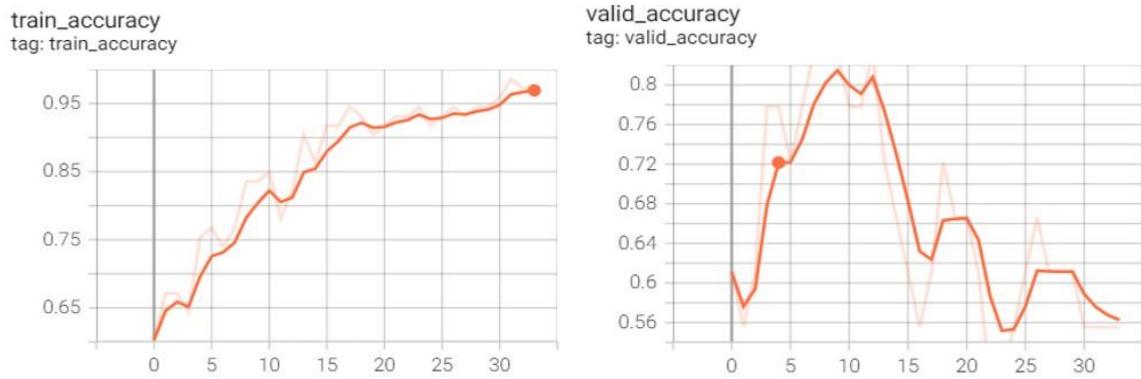


Figure 21 : Courbe de variation de la précision du modèle Tab Transformers

CHAPITRE 4

RÉSULTATS ET DISCUSSION

Dans cette partie, nous exposons les résultats des diverses expérimentations. Pour rappel, notre objectif consiste à prédire l'efficacité de la NAC sur l'ostéosarcome à l'aide des données radiomiques et paracliniques sans passer par les techniques de réduction de dimensionnalité. L'ensemble des données radiomiques et paracliniques servent ensuite d'entrée pour nos modèles d'apprentissage profond. Nous détaillons graduellement les résultats obtenus en fonction du type de réseau neurone profond exploité. Par la suite, nous procéderons à une comparaison des résultats obtenus.

4.1 Résultats des expérimentations.

Après avoir mené diverses expérimentations, nous évaluons les performances de nos modèles. Cette évaluation nous permet de mesurer leur efficacité et identifier les points d'amélioration potentiels. Pour cela, nous examinons plusieurs métriques de performance. Nous commençons par analyser la matrice de confusion, qui donne une vue d'ensemble sur les prédictions correctes et incorrectes. Ensuite, nous étudions la courbe ROC et la valeur AUC, qui mesurent la capacité du modèle à distinguer entre les classes. Nous allons également discuter le rapport de classification, qui fournit des détails sur la précision, le rappel et le F1 score pour la classe négative et positive. Enfin, nous prenons en compte la moyenne géométrique, afin de mieux évaluer l'équilibre du modèle sur la prédiction des classes. Nous

signalons que les résultats des modèles LSTM et GRU seront exclus car ils sont en surapprentissage.

4.1.1 Evaluation du modèle FCNN

La matrice de confusion de la *figure 23* montre que le FCNN arrive à prédire correctement toutes les réponses négatives. Cependant, pour la réponse positive, sur 12 observations, le modèle a prédit 2 observations comme réponse négative alors que c'est la classe positive. Dans sa globalité. Nous notons que sur 31 observations, 29 ont été bien prédites. Cela montre que le modèle arrive à faire de bonnes prédictions, quel que soit la classe.

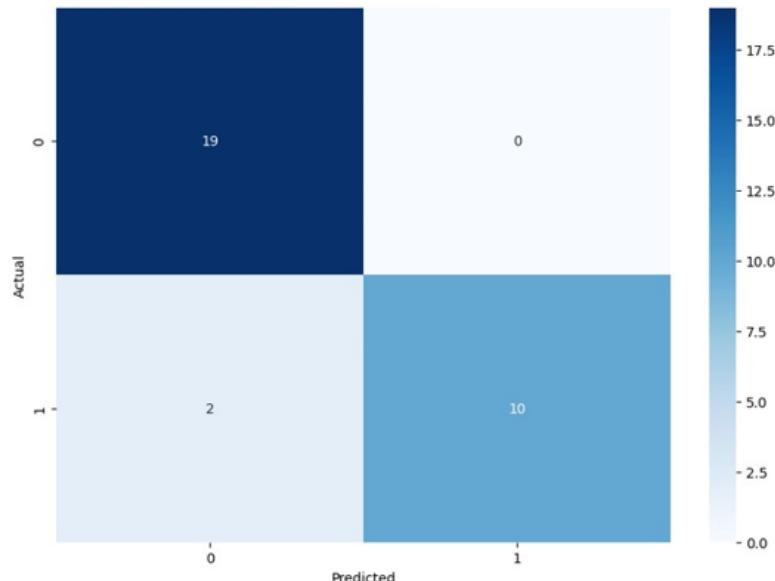


Figure 22 : Matrice de confusion du modèle FCNN

La courbe ROC de la *figure 24* a obtenu un AUC de 92%. Cette valeur montre que le modèle a une bonne capacité discriminatoire entre les classes, ce qui confirme encore que le modèle arrive à bien classifier.

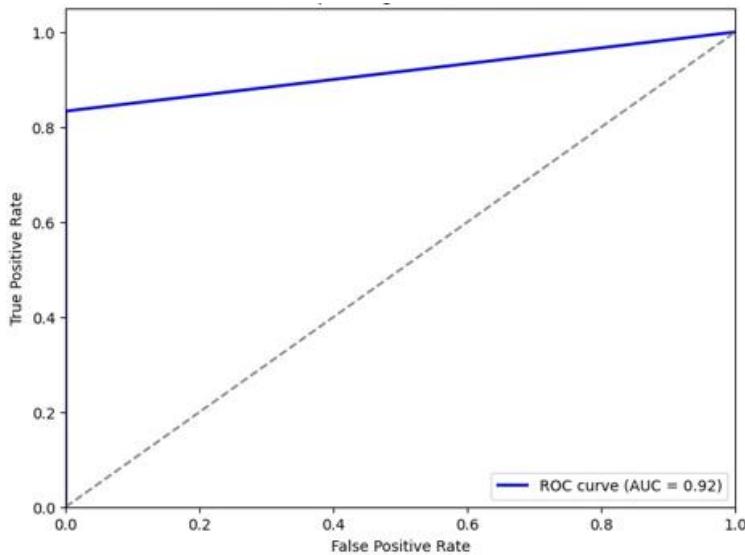


Figure 23 : Courbe ROC (AUC) du modèle FCNN

Le tableau 5 affiche que FCNN a obtenu une précision de 90% pour la classe négative au traitement. Pour la classe positive, le modèle a obtenu une précision de 100 %. Le F1 score résultant de la combinaison de la précision et du rappel donne respectivement 95% et 91% pour les réponses négatives et positives. En conséquence, le modèle a une précision globale de 94%. De plus, sa moyenne géométrique est de 91% indiquant une bonne performance prédictive. Ces performances encourageantes montrent que le modèle FCNN prédit bien les patients ayant répondu négativement ou positivement à la NAC.

En définitive, les performances élevées sur l'ensemble des métriques prouvent que le modèle arrive à bien prédire les réponses négatives et positives avec une marge d'erreur très petite.

Tableau 3 : Rapport de classification de FCNN

Modèle	Classe	Précision	Rappel	F1 score	G-mean	Support
FCNN	Réponse négative	90%	100%	95%	91%	19
	Réponse positive	100%	83%	91%		12
	Accuracy	94%				31

4.1.2 Evaluation du modèle Tab Transformers

La *figure 25* montre que Tab Transformers arrive à prédire 7 observations sur 8 pour la réponse positive. La réponse négative à son tour a obtenu 2 bonnes prédictions sur 3. Ces résultats relativement modestes confirment que le modèle arrive à faire prédictions acceptables. En revanche, vu le nombre d'observations réservé pour l'ensemble de tests, nous notons un déséquilibre entre les classes. Cela pourrait suggérer, dans le futur, de s'orienter vers un jeu de données plus vaste qui assure un bon équilibre entre les classes.

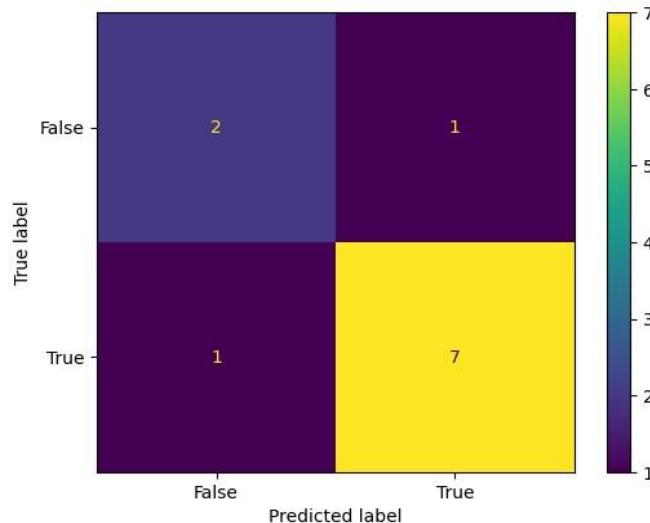


Figure 24 : Matrice de confusion du modèle Tab Transformer

Le modèle a obtenu un AUC de 0,77, ce qui montre une capacité discriminatoire de 77 % pour différencier les classes. Cette valeur est supérieure à 0,5 indiquant que le modèle est significativement meilleur qu'un classificateur aléatoire.

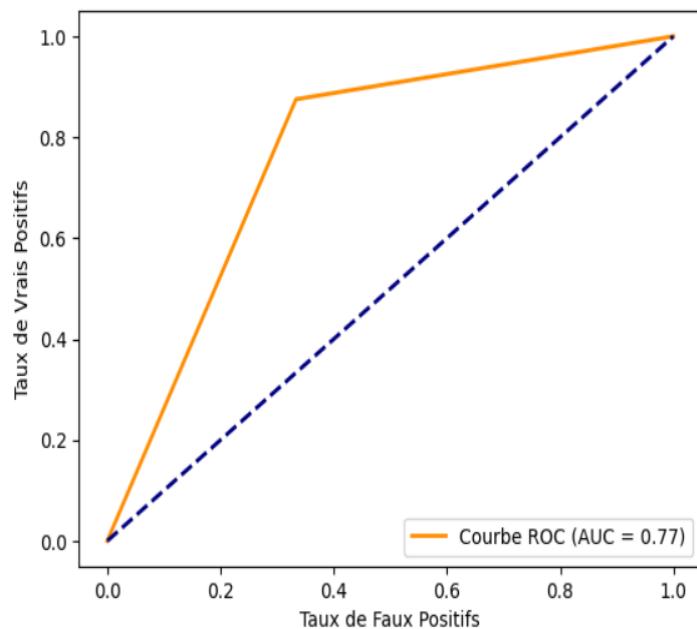


Figure 25 : Courbe ROC (AUC) du modèle Tab Transformers

Le modèle Tab Transformers a affiché pour la réponse négative, une précision, un rappel et un F1 score tous égaux à 67%. Cela indique une performance modérée de sa prédiction pour cette classe. Cependant, la classe positive obtient un score de 88% sur toutes ses métriques montrant une différence de 21% par rapport à la réponse négative. Ainsi, sa précision globale est de 82%. Ces résultats pourraient être dus au fait que le support de la classe négative est de 3 tandis que celui de la classe positive est de 8 comme le montre le *tableau 6*. Sa moyenne géométrique est égale à 76 %, elle indique une bonne classification.

Tableau 4 : Rapport de classification de Tab Transformers

Modèles	Classe	Précision	Rappel	F1 score	G-mean	Support
Tab Transformers	Réponse négative	67%	67%	67%	76%	3
	Réponse positive	88%	88%	88%		8
	Accuracy	82%				11

4.2 Comparaison des modèles

Dans le cadre de notre étude qui consiste à prédire la réponse positive ou négative de la NAC sur l'ostéosarcome, nous avons effectué différentes expérimentations suivant 2 modèles (FCNN et Tab Transformers). Ces expérimentations avaient pour but de trouver la meilleure approche dans la prédiction de la réponse à la NAC. Dans le tableau 7, nous avons comparé tous les modèles en nous basant sur les métriques.

Tableau 5 : Récapitulation des métriques des modèles

Modèles	Accuracy	AUC	G-mean
FCNN	94%	92%	91 %
Tab Transformers	82%	77%	76 %

Le modèle FCNN se distingue par ses performances remarquables, affichant une précision de 94%, une AUC de 92% et une moyenne géométrique de (G-mean) de 91%. Ce modèle

démontre une excellente capacité à équilibrer la prédiction des différentes classes, ce qui le positionne comme le meilleur modèle.

Le modèle Tab Transformers a également montré des performances qui semblent prometteuses bien qu’inférieures à celles du FCNN. Sa précision globale est de 82%, une AUC de 77% et une moyenne géométrique G-mean de 76% reflétant un déséquilibre sur la prédiction entre les classes. Cela pourrait indiquer une sensibilité accrue aux variations des données. De plus, l’échantillon de test onze (11) est très petit comparé aux autres modèles trente-trois (33). Cela suggère que son évaluation demande plus d’attention. Par précaution, l’idéal serait d’évaluer dans un jeu de données plus large.

Nous signalons que la variation des données de test entre le Tab Transformers et les autres modèles est motivée par le fait que les Transformers ont besoin de plus de données pendant leur entraînement. Cela a peut-être amélioré les performances du modèle, ce qui pourrait orienter vers les Transformers dans le futur si toutefois l’étude est portée sur un jeu de données plus vaste.

Les résultats peuvent ouvrir la voie vers une application clinique. Un tel outil pourra permettre d’anticiper sur les réponses de la NAC.

4.3 Comparaison de notre meilleur modèle avec d’autres approches

Après avoir évalué les performances de nos différents modèles, nous répondrons à la deuxième question des objectifs de ce travail. Le tableau 8 compare les résultats de notre meilleur modèle FCNN à ceux du modèle de régression logistique développé par Zhang et al.(Zhang et al., 2021).

Tableau 6 : Comparaison de notre meilleur modèle avec celui de la littérature

Modèle	Technique appliquée	ACC	AUC
FCNN	Supprimer les colonnes de variance nulle	94%	92%
Logistique régression	LASSO	91%	95%

En comparaison avec nos modèles, les métriques présentées dans le Tableau 8 montrent que le modèle FCNN présente une amélioration de 3% sur la précision par rapport au modèle de la régression logistique. Cependant, une dégradation de 3% sur la valeur de l'AUC est observée. Le modèle formé à partir de la régression logistique a une meilleure capacité à distinguer les classes. Notre modèle le surpasse sur les prédictions correctes dans sa globalité. En définitive, l'importance relative accordée à chaque modèle dépendra de l'application souhaitée

4.4 Discussion des résultats

L'introduction de la chimiothérapie néoadjuvante dans le traitement de l'ostéosarcome a considérablement amélioré le taux de survie des patients atteints d'ostéosarcome. Dans notre étude, nous avons abordé une nouvelle approche qui s'avère prometteuse pour la prédition de la réponse au traitement. Récemment Zheng et al., (Zheng et al., 2023), ont proposé une méthode qui combine diverses techniques pour extraire des caractéristiques quantitatives de données manuellement suivies de l'utilisation du réseau de neurones ResNet18. Les résultats obtenus sont prometteurs, avec une précision de 93,8% et une AUC de 0,961, surpassant ceux de (Zhang et al., 2021), probablement en raison de l'utilisation d'images améliorées en une seule étape, contre des améliorations dynamiques dans l'étude de Zhang et al (Zhang et al., 2021).

D'autres chercheurs comme (Chen et al., 2021), ont aussi démontré la pertinence des signatures radiomiques dérivées de l'IRM, obtenant une AUC de 0,882 sur la cohorte

principale et de 0,842 sur la cohorte de validation, confirmant la robustesse prédictive de leur modèle.

Ces résultats montrent que notre approche, intégrant toutes les données radiomiques et le stade chirurgical sans recours à des méthodes de sélection de variables, semble efficace. En d'autres mots, elle paraît offrir une précision prédictive améliorée.

Cependant, notre étude présente plusieurs limites. D'abord, le nombre d'observations est insuffisant pour généraliser nos résultats. Ensuite, le processus de collecte présente un risque de biais de sélection lors de la collecte des données, ce qui invite à interpréter nos conclusions avec prudence. En dernier lieu, le fait qu'un seul centre ait été impliqué et que les paramètres des appareils varient d'un lieu à l'autre signifie que les résultats pourraient ne pas être reproductibles partout. Nous estimons effectuer une plus large étude avec des données issues de plusieurs centres de traitement, pour permettre de tendre vers un modèle plus général et robuste.

CHAPITRE 5

CONCLUSION ET PERSPECTIVES

Notre recherche s'est concentrée sur la prédiction de la réponse à la chimiothérapie néoadjuvante en utilisant des réseaux de neurones profonds. Quatre (4) modèles, incluant LSTM, GRU, Tab Transformers et FCNN, ont été entraînés sur un ensemble de données combinant le stade chirurgical et l'ensemble des caractéristiques radiomiques, après filtrage des colonnes à variance nulle. Les modèles LSTM, GRU et Tab Transformers ont directement traité l'ensemble de données, tandis que FCNN utilise des sous-réseaux spécifiques à chaque sous-groupe. En évitant la sélection de variables, notre approche a généré des résultats prometteurs, susceptibles d'aider les professionnels de santé à personnaliser le traitement des patients.

Notre meilleur modèle montre une capacité à prédire la réponse des patients à la NAC, avec une efficacité comparée à ceux des travaux de Zhang et al. (Zhang et al., 2021). Ces résultats peuvent ouvrir la voie à une prise de décision ciblée, renforçant le principe du traitement adapté au bon patient et au bon moment.

Pour améliorer et généraliser ces résultats, nous envisageons d'incorporer davantage de sources de données afin d'augmenter le nombre d'observations, notamment avec les Transformers. Nous projetons également de développer un support décisionnel multifactoriel intégrant données radiomiques, génomiques et cliniques, pour affiner les prédictions et offrir des recommandations encore plus adaptées. Par ailleurs, l'utilisation des algorithmes génétiques pourrait constituer une piste de recherche prometteuse pouvant toujours utiliser la méthode de sélection de variables.

RÉFÉRENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., & Isard, M. (2016). {TensorFlow}: a system for {Large-Scale} machine learning. 12th USENIX symposium on operating systems design and implementation (OSDI 16),
- Ammar, L. B., Gasmi, K., & Ltaifa, I. B. (2024). ViT-TB: ensemble learning based ViT model for tuberculosis recognition. *Cybernetics and Systems*, 55 (3), 634-653.
- Aurélien, G. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. o'reilly.
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Batutin, A. (2024). *Maîtriser l'histoire de la PNL : d'hier à aujourd'hui*.
<https://shelf.io/blog/master-nlp-history-from-then-to-now/>
- Benjamin, R. S. (2020). Adjuvant and neoadjuvant chemotherapy for osteosarcoma: a historical perspective. *Current Advances in Osteosarcoma: Clinical Perspectives: Past, Present and Future*, 1-10.
- Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K., & Ghayvat, H. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics*, 10 (20), 2470.
- Bielack, S. S., Blattmann, C., Borkhardt, A., Csóka, M., Hassenpflug, W., Kabíčková, E., Kager, L., Kessler, T., Kratz, C., & Kühne, T. (2022). Osteosarcoma and causes of death: A report of 1520 deceased patients from the Cooperative Osteosarcoma Study Group (COSS). *European journal of cancer*, 176, 50-57.
- Boucher, A., & Badri, M. (2018). Software metrics thresholds calculation techniques to predict fault-proneness: An empirical comparison. *Information and Software Technology*, 96, 38-67.
- Byun, B. H., Kong, C.-B., Lim, I., Kim, B. I., Choi, C. W., Song, W. S., Cho, W. H., Jeon, D.-G., Koh, J.-S., & Lee, S.-Y. (2014). Early response monitoring to neoadjuvant chemotherapy in osteosarcoma using sequential 18 F-FDG PET/CT and MRI. *European journal of nuclear medicine and molecular imaging*, 41, 1553-1562.
- Chen, H., Zhang, X., Wang, X., Quan, X., Deng, Y., Lu, M., Wei, Q., Ye, Q., Zhou, Q., & Xiang, Z. (2021). MRI-based radiomics signature for pretreatment prediction of pathological response to neoadjuvant chemotherapy in osteosarcoma: a multicenter study. *European Radiology*, 31, 7913-7924.
- Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., Bray, F., Jemal, A., Yu, X. Q., & He, J. (2016). Cancer statistics in China, 2015. *CA: a cancer journal for clinicians*, 66 (2), 115-132.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

- Chrupała, G., Kádár, A., & Alishahi, A. (2015). Learning language through pictures. *arXiv preprint arXiv:1506.03694*.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- de la Roy, E. d. R., Recht, T., Zemmari, A., Bourreau, P., Houchat, K., & Mora, L. (2022). Détection d'ouvertures de fenêtres au moyen de réseaux de neurones artificiels LSTM.
- De Vries, H., Memisevic, R., & Courville, A. C. (2016). Deep Learning Vector Quantization. *ESANN*,
- Delsinne, T. (2021). Diagnostic écologique des tourbières et des forêts de la RNN Chastreix-Sancy par les syrphes (méthode Syrph the Net). *Rapport de la Société d'Histoire Naturelle Alcide-d'Orbigny pour le Syndicat Mixte du Parc Naturel Régional des Volcans d'Auvergne*.
- Desandes, E., Lacour, B., & Clavel, J. (2016). Cancer in adolescents and young adults in France: Epidemiology and pathways of care. *Bulletin du Cancer*, 103 (12), 957-965.
- Dias, C.-E., De Gabriac, C. G. d. F., Guigue, V., & Gallinari, P. (2018). RNN et modèle d'attention pour l'apprentissage de profils textuels personnalisés. CORIA,
- Elman, J. L. (1990). Finding structure in time. *Cognitive science*, 14 (2), 179-211.
- Gatenby, R. A., Grove, O., & Gillies, R. J. (2013). Quantitative imaging in cancer evolution and ecology. *Radiology*, 269 (1), 8-14.
- Gharbi, M. A., Zendeoui, A., Tborbi, A., Bouzidi, R., Ezzaouia, K., & Nefiss, M. (2023). Conservative surgical management of surface osteosarcoma using 3D printing technology: An unusual case report and literature review. *International Journal of Surgery Case Reports*, 113, 109086.
- Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: images are more than pictures, they are data. *Radiology*, 278 (2), 563-577.
- Gonçalves, A. (2016). Chimiothérapie néo-adjuvante des cancers du sein HER2-positifs et triple-négatifs. *Bulletin du Cancer*, 103 (6), S76-S89.
- Goodfellow, I. (2016). Deep learning. In: MIT press.
- Google. (2025a). <https://support.google.com/drive/answer/2424384?hl=fr&co=GENIE.Platform%3DDESKTOP>
- Google. (2025b). *Google Colab*. <https://colab.google/notebooks/>
- Gu, B., & Sung, Y. (2021). Enhanced reinforcement learning method combining one-hot encoding-based vectors for CNN-based alternative high-level decisions. *Applied Sciences*, 11 (3), 1291.
- Gu, Q., Zhu, L., & Cai, Z. (2009). Evaluation measures of the classification performance of imbalanced data sets. Computational Intelligence and Intelligent Systems: 4th International Symposium, ISICA 2009, Huangshi, China, October 23-25, 2009. Proceedings 4,
- Hanna, M., Liu, O., & Variengien, A. (2024). How does GPT-2 compute greater-than?: Interpreting mathematical abilities in a pre-trained language model. *Advances in neural information processing systems*, 36.

- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* (6), 610-621.
- Huang, C., Dai, Y., Chen, Q., Chen, H., Lin, Y., Wu, J., Xu, X., & Chen, X. (2022). Development and validation of a deep learning model to predict survival of patients with esophageal cancer. *Frontiers in Oncology*, 12, 971190.
- Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*.
- Jean-Eric, K. K., Régis, A. A. J., Natacha, K. A. A., Zolopégué, S., Leopold, K. K., & Michel, K. (2023). Aspects épidémiologiques et thérapeutiques des ostéosarcomes en milieu tropical/Epidemiological and therapeutic aspects of osteosarcomas in a tropical environment.
- Jeong, S. Y., Kim, W., Byun, B. H., Kong, C.-B., Song, W. S., Lim, I., Lim, S. M., & Woo, S.-K. (2019). Prediction of chemotherapy response of osteosarcoma using baseline 18F-FDG textural features machine learning approaches with PCA. *Contrast media & molecular imaging*, 2019b.
- Ketkar, N., Moolayil, J., Ketkar, N., & Moolayil, J. (2021). Introduction to pytorch. *Deep learning with python: learn best practices of deep learning models with PyTorch*, 27-91.
- Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., Forster, K., Aerts, H. J., Dekker, A., & Fenstermacher, D. (2012). Radiomics: the process and the challenges. *Magnetic resonance imaging*, 30 (9), 1234-1248.
- Lambin, P., Leijenaar, R. T., Deist, T. M., Peerlings, J., De Jong, E. E., Van Timmeren, J., Sanduleanu, S., Larue, R. T., Even, A. J., & Jochems, A. (2017). Radiomics: the bridge between medical imaging and personalized medicine. *Nature reviews Clinical oncology*, 14 (12), 749-762.
- Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R. G., Granton, P., Zegers, C. M., Gillies, R., Boellard, R., & Dekker, A. (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*, 48 (4), 441-446.
- LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2.
- Manning, C. D. (1999). *Foundations of statistical natural language processing*. The MIT Press.
- Mayerhoefer, M. E., Materka, A., Langs, G., Häggström, I., Szczypiński, P., Gibbs, P., & Cook, G. (2020). Introduction to radiomics. *Journal of Nuclear Medicine*, 61 (4), 488-495.
- Mehta, S., Kukreja, V., & Gupta, A. (2023). Next-Generation Wheat Disease Monitoring: Leveraging Federated Convolutional Neural Networks for Severity Estimation. 2023 4th International Conference for Emerging Technology (INCET),
- Nahm, F. S. (2022). Receiver operating characteristic curve: overview and practical use for clinicians. *Korean journal of anesthesiology*, 75 (1), 25-36.
- NumPy, E. (2025). <https://numpy.org/>

- Nyeko, R., Geriga, F., Angom, R., Kambugu, J. B., & Van Heerden, J. (2024). The management of osteosarcoma in children and adolescents in a resource-limited setting: quality improvement considerations to improve treatment outcomes. *BMC cancer*, 24 (1), 1-8.
- Olah, C. (2015). Comprendre les réseaux LSTM. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Pandas. (2025). <https://pandas.pydata.org/>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., & Antiga, L. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Python. (2025). <https://www.python.org/>
- Rizzo, S., Botta, F., Raimondi, S., Origgi, D., Fanciullo, C., Morganti, A. G., & Bellomi, M. (2018). Radiomics: the facts and the challenges of image analysis. *European radiology experimental*, 2, 1-8.
- Rogers, W., Thulasi Seetha, S., Refaee, T. A., Lieverse, R. I., Granzier, R. W., Ibrahim, A., Keek, S. A., Sanduleanu, S., Primakov, S. P., & Beuque, M. P. (2020). Radiomics: from qualitative to quantitative imaging. *The British journal of radiology*, 93 (1108), 20190948.
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. Australasian joint conference on artificial intelligence,
- Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L., & Cheng, Z. (2020). Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *Journal of Petroleum Science and Engineering*, 186, 106682.
- Suhan, A., Libenzi, D., Zhang, A., Schuh, P., Saeta, B., Sohn, J. Y., & Shabalin, D. (2021). LazyTensor: combining eager execution with domain-specific compilers. *arXiv preprint arXiv:2102.13267*.
- Sun, R., Lerousseau, M., Henry, T., Carré, A., Leroy, A., Estienne, T., Niyoteka, S., Bockel, S., Rouyar, A., & Andres, É. A. (2021). Intelligence artificielle en radiothérapie: radiomique, pathomique, et prédition de la survie et de la réponse aux traitements. *Cancer/Radiothérapie*, 25 (6-7), 630-637.
- Vafaeikia, P., Wagner, M. W., Hawkins, C., Tabori, U., Ertl-Wagner, B. B., & Khalvati, F. (2024). MRI-based end-to-end pediatric low-grade glioma segmentation and classification. *Canadian Association of Radiologists Journal*, 75 (1), 153-160.
- Van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H., & Baessler, B. (2020). Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights into imaging*, 11 (1), 91.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Vial, A., Stirling, D., Field, M., Ros, M., Ritz, C., Carolan, M., Holloway, L., & Miller, A. A. (2018). The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. *Translational Cancer Research*, 7 (3).

- Vinayakumar, R., Soman, K., & Poornachandran, P. (2017). Applying convolutional neural network for network intrusion detection. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI),
- Wang, W., Bi, B., Yan, M., Wu, C., Bao, Z., Xia, J., Peng, L., & Si, L. (2019). Structbert: Incorporating language structures into pre-training for deep language understanding. *arXiv preprint arXiv:1908.04577*.
- Yıldırım, Ö., Pławiak, P., Tan, R.-S., & Acharya, U. R. (2018). Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Computers in biology and medicine*, 102, 411-420.
- Zhang, L., Ge, Y., Gao, Q., Zhao, F., Cheng, T., Li, H., & Xia, Y. (2021). Machine learning-based radiomics nomogram with dynamic contrast-enhanced MRI of the osteosarcoma for evaluation of efficacy of neoadjuvant chemotherapy. *Frontiers in Oncology*, 11, 758921.
- Zheng, F., Yin, P., Liang, K., Wang, Y., Hao, W., Hao, Q., & Hong, N. (2023). Fusion Radiomics-Based Prediction of Response to Neoadjuvant Chemotherapy for Osteosarcoma. *Academic Radiology*.
- Zwanenburg, A., Leger, S., Vallières, M., & Löck, S. (2019). The image biomarker standardisation initiative—IBSI 0.0. 1dev documentation 2019. *ProQuest Number: INFORMATION TO ALL USERS*.