

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
SERAPHIN GAUTIER ABELA NOMO

ANALYSE DES DONNÉES DE RÉFLECTANCE DU PROCHE INFRAROUGE
BASÉE SUR LES RÉSEAUX DE NEURONES PROFONDS

DÉCEMBRE 2024

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.

Résumé

La spectroscopie est une science qui étudie l'interaction entre la lumière et la matière. L'un de ses principaux défis concerne la qualité des données, influencée par la pureté des échantillons et la variabilité intrinsèque des mesures. Cette technique est largement employée en chimie analytique, dans l'industrie pharmaceutique et en contrôle environnemental pour détecter et quantifier la concentration des molécules. Parmi les méthodes d'analyse, la régression des moindres carrés partiels(PLSR) offre d'assez bonnes performances[1]. Toutefois, l'essor des données spectroscopiques a favorisé l'émergence d'autres modèles d'apprentissage automatique, notamment les réseaux de neurones convolutifs, qui peinent néanmoins à capturer les dépendances globales des spectres. Pour pallier cette limite, nous avons exploré les architectures des Transformers, qui ont ensuite été comparées aux modèles traditionnels. Ce mémoire évalue les performances des modèles de Transformers sur deux ensembles de données spectroscopiques. Le premier a pour objectif de déterminer la concentration en protéines des fourrages à partir de leurs spectres de réflectance. Le second issu des sciences forensiques, vise à estimer la datation relative des taches de sang à partir de leur réflectance sur différents types de substrats. Dans le cadre de nos expérimentations, l'architecture Vision Transformer a atteint une erreur quadratique moyenne(RMSE) de 0.74 pour l'estimation de la concentration en protéines, bien que le PLSR reste légèrement performant. Sur les données issues des sciences forensiques, un modèle Transformer exploitant l'attention croisée entre les caractéristiques contextuelles et spectrales a obtenu un RMSE plus prometteur de 0.19, confirmant son potentiel dans ce domaine.

Abstract

Spectroscopy is a science that studies the interaction between light and matter. One of its main challenges concerns the quality of the data, which is influenced by the purity of the samples and the intrinsic variability of the measurements. This technique is widely used in analytical chemistry, the pharmaceutical industry, and environmental monitoring to detect and quantify the concentration of molecules. Among the analysis methods, partial least squares regression (PLSR) offers fairly good performance [1]. However, the growth of spectroscopic data has fostered the emergence of other machine learning models, notably convolutional neural networks, which nevertheless struggle to capture the global dependencies of spectra. To address this limitation, we explored Transformer architectures, which were then compared to traditional models. This thesis evaluates the performance of Transformer models on two spectroscopic datasets. The first aims to determine the protein concentration in forages from their reflectance spectra. The second, originating from forensic sciences, seeks to estimate the relative dating of bloodstains based on their reflectance on different types of substrates. In our experiments, the Vision Transformer architecture achieved a root mean square error (RMSE) of 0.74 for estimating protein concentration, although PLSR remained slightly more performant. On the data from forensic sciences, a Transformer model leveraging cross-attention between contextual and spectral features obtained a more promising RMSE of 0.19, confirming its potential in this field.

Liste des abréviations

AB AdaBoost

AdamW Adaptive Moment Estimation with Weight Decay

AFNOR Association Française de Normalisation

AGoES Automatically Generating a Pre-processing Strategy

AI Artificial Intelligence

ANN Artificial Neural Network

ANNs Artificial Neural Networks

ANOVA Analyse de la Variance

API Active Pharmaceutical Ingredient

BloodNet CNN avec attention pour estimer le TSD

C3G Cyanidine-3-glucoside

CatBoost Categorical Boosting

CNN Convolutional Neural Network

CNNMulti Multi-layer CNN

COD Chemical Oxygen Demand

CUDA Compute Unified Device Architecture

CWT Continuous Wavelet Transform

D3G Delphinidine-3-glucoside

DL Deep Learning

DO Dissolved Oxygen

DT Decision Tree

EFS Emitted Fluorescence Spectra

FT-IR Fourier Transform Infrared

FT-NIR Fourier Transform Near Infrared

GA Genetic Algorithm

GA-PLS Genetic Algorithm - Partial Least Squares

GAN Generative Adversarial Network

GAs Genetic Algorithms

GFS Glucose-Fructose Syrup

GPR Gaussian Process Regression

Grad-CAM Gradient-weighted Class Activation Mapping

HSI Hyperspectral Imaging

IA Intelligence Artificielle

IC Infrared Camera

KNN k-Nearest Neighbors

KNN-LW K-Nearest Neighbors Locally Weighted PLSR

LDA Linear Discriminant Analysis

LightGBM Light Gradient Boosting Machine

LS-SVM Least Squares Support Vector Machine

LSTM Long Short-Term Memory

LWPLSR Locally Weighted Partial Least Squares Regression

MAE Mean Absolute Error

MCR-ALS Multivariate Curve Resolution - Alternating Least Squares

MIR Mid Infrared Spectroscopy

ML Machine Learning

MLP Multi-Layer Perceptron

MLR Multiple Linear Regression

MPs Microplastics

MSC Multiplicative Scatter Correction

MSE Mean Squared Error

NIR Near Infrared (Proche Infrarouge)

NNs Neural Networks

NRD Normalized Reflectance Difference

OC Organic Carbon

PCA Principal Component Analysis

PCR Principal Component Regression

PCs Principal Components

PLS Partial Least Squares

PLS-DA Partial Least Squares Discriminant Analysis

PLSR Partial Least Squares Regression

PSO Particle Swarm Optimization

RMSE Root Mean Square Error

RNN Recurrent Neural Network

RPD Ratio of Performance to Deviation

RPIQ Ratio of Performance to Interquartile Distance

SA Simulated Annealing

SAM Spectral Angle Mapper

SGD Stochastic Gradient Descent

SIMCA Soft Independent Modeling of Class Analogies

SNV Standard Normal Variate

SPIR Spectroscopie Proche Infra-Rouge

STACK Stacked AutoEncoder

SVM Support Vector Machine

SW-NIR Short-Wave Near-Infrared

TAC Total Anthocyanin Content

THZ Terahertz

TSD Time Since Deposition (Durée d'exposition d'une tache de sang)

UVE Uninformative Variable Elimination

VIP Variable Importance to Projection

Vis-NIR Visible and Near Infrared

ViT Vision Transformer

WPLSR Weighted Partial Least Squares Regression

XGBoost Extreme Gradient Boosting

Avant-propos

Je tiens à exprimer mes plus sincères remerciements à toutes les personnes qui m'ont accompagné tout au long de la réalisation de ce mémoire.

Tout d'abord, je remercie chaleureusement mon directeur de recherche, Fadel Toure, pour son encadrement, ses précieux conseils et son soutien constant. Ses orientations et ses disponibilités qui ont été déterminantes dans l'élaboration et la finalisation de ce travail.

Je tiens également à remercier Lesnoff Matthieu et Cyril Muehlethaler pour la mise à disposition des deux jeux de données utilisés dans ce projet.

Je ne saurais terminer sans pour autant adresser mes sincères remerciements à tous les étudiants avec qui j'ai eu à collaborer, plus précisément Gamkam Ma'akou, Dongmo Fofack pour leur esprit critique.

Je tiens à exprimer toute ma reconnaissance à ma famille au Cameroun, en particulier à mes parents, Nomo Jean Marie Claude et Abela Seraphine, ainsi qu'à mes frères et sœurs, pour leur amour inconditionnel. Un merci tout particulier à mon frère Ndono Nomo Denise Fulbert, pour son soutien moral et financier qui a été crucial tout au long de ce parcours.

*À mon père, NOMO Jean Marie Claude, dont la
sagesse et les conseils ont toujours été un phare dans
ma vie. À ma mère, ABELA Séraphine, dont l'amour
et les sacrifices silencieux ont fait de moi ce que je suis
aujourd'hui. À mon frère, NDO NO NOMO Denise
Fulbert, pour sa complicité et son soutien qui m'ont
aidé à avancer malgré les obstacles.*

*À mes frères et sœurs au Cameroun, votre affection et
vos encouragements, même à distance, ont été une
force précieuse. Et à mon ami, Ulrich Tedongmo, un
véritable compagnon de route, merci pour ton soutien
inébranlable et ta présence rassurante.*

*Ce mémoire est dédié à vous, ma famille et mes amis,
qui m'avez donné la force de rêver et de réaliser. Avec
tout mon amour et ma gratitude.*

*Ce que nous savons est une goutte
d'eau, ce que nous ignorons est un
océan "Issac Newton"*

Table des matières

Résumé	ii
Abstract	iii
Avant-propos	viii
Table des matières	x
Liste des tableaux	xiii
Table des figures	xv
1 Introduction	1
1.1 Problématique	1
1.2 Objectifs	3
1.3 Contributions nouvelles	4
1.4 Organisation du mémoire	4
1.5 Conclusion	5
2 Revue de littérature	6
2.1 Tendances des recherches (5 dernières années)	6
2.2 Revue analyse de littérature	8
2.2.1 Introduction	8
2.2.2 Approches statistiques et numériques	8
2.2.3 Les modèles d'apprentissage automatique	17
2.2.4 Les modèles d'apprentissage profond	23

2.3	Conclusion	27
3	Modèles d'Apprentissage	28
3.1	Concepts théoriques des CNN	28
3.1.1	Introduction au CNN	28
3.1.2	Opérations de Convolution et Stratégies de Traitement	29
3.1.3	Fonctions d'Activation, Couches Connectées et Optimisation	30
3.2	Fondements des Transformers	32
3.2.1	Principe de fonctionnement	32
3.2.2	Encodeur	33
3.2.3	Décodeur	38
3.2.4	Les Vision Transformers (ViT)	41
3.3	Quelques Métriques d'évaluation des modèles	44
3.4	Limites des modèles	46
3.5	Conclusion	47
4	Méthodologie	49
4.1	Collecte et Préparation des Données	50
4.1.1	Jeux de données Challenge	50
4.1.2	Jeux de données de criminalistique	51
4.1.3	Prétraitement des données	56
4.2	Architectures des modèles	58
4.3	Méthodes d'apprentissage	65
4.3.1	Algorithme de descente de gradient	65
4.3.2	Optimiseur AdamW	66
4.3.3	Fonction de Perte	67
4.3.4	Partition des jeux de données	67
4.4	Conclusion	68
5	Résultats et Discussions	69
5.1	Estimation des concentrations et des durées	69

5.2	Etude d'explicabilité	84
5.3	Synthèse des résultats des expérimentations	87
5.4	Conclusion	88
6	Conclusion Générale	90
	Bibliographie	93
A	Longueurs d'onde importantes	100

Liste des tableaux

3.1	Comparaison des modèles de détection de molécules	47
5.1	Tableau des valeurs expérimentales CNN	70
5.2	Les différentes valeurs du RMSE avec les Transformers	73
5.3	RMSE avec les différentes configurations des ViT	76
5.4	RMSE avec les différentes configurations de l'attention croisée	79
5.5	Résultats des RMSE avec l'attention croisée en utilisant seulement sub- strat	81
5.6	Résultats des RMSE avec toutes les caractéristiques environnementales	82
A.1	Tableau des longueurs d'onde et des valeurs originales	100

Table des figures

2.1	<i>Modèles utilisés dans les articles</i>	7
3.1	Architecture typique d'un CNN.	29
3.2	Architecture typique des Transformers.	33
3.3	<i>Architecture typique d'un encodeur</i>	34
3.4	<i>Architecture typique d'un décodeur</i>	38
3.5	<i>Architecture typique des ViT</i>	42
4.1	<i>Repartition des échantillons selon le rayonnement</i>	53
4.2	<i>Répartition des échantillons selon le rayonnement et le substrat</i>	54
4.3	<i>Analyse bivariée des données contextuelles</i>	55
4.4	<i>Encodage du substrat</i>	57
4.5	<i>Architecture Encoder</i>	59
4.6	<i>Architecture ViT Encodeur</i>	61
4.7	<i>L'attention croisée</i>	62
4.8	<i>Comparaison connexion résiduelle et cross attention</i>	63
5.1	<i>Évolution de l'erreur avec Challenge</i>	71
5.2	<i>Évolution de l'erreur du jeu de données Criminalistique</i>	71
5.3	<i>Évolution de l'erreur du jeu de données Challenge</i>	74
5.4	<i>Évolution de l'erreur du jeu de données Criminalistique</i>	74
5.5	<i>Évolution de l'erreure du jeu de donnée Challenge</i>	77
5.6	<i>Évolution de l'erreure du jeu de donnée Criminalistique</i>	77
5.7	<i>Courbe de l'attention croisée jeu de donnée Challenge</i>	79
5.8	<i>Courbe de l'attention croisée jeu de donnée Criminalistique</i>	80

5.9	<i>Résultats attention croisée avec seulement le Substrat</i>	81
5.10	<i>Résultats attention croisée avec toutes les données contextuelles</i>	82
5.11	<i>Explicabilité Transformer</i>	86
5.12	<i>comparaison architecture avec le PLSR</i>	87
5.13	<i>Comparaison des architectures avec les deux jeux de données</i>	88

Chapitre 1

Introduction

1.1 Problématique

La spectroscopie est un domaine en pleine expansion en raison de ses nombreuses applications dans divers secteurs scientifiques et industriels. Elle est notamment utilisée en agriculture, où elle permet d'estimer la teneur en protéines des fourrages, ainsi qu'en criminalistique, où elle peut jouer un rôle clé dans l'analyse rapide des taches de sang. Les méthodes traditionnelles d'analyse reposent principalement sur l'utilisation de réactifs chimiques, ce qui les rend souvent longues et fastidieuses avant d'obtenir des résultats exploitables. Pour le sang, il n'existe pas de méthode chimique standard pour mesurer sa dégradation, ce qui rend la spectroscopie particulièrement intéressante.

Avec l'augmentation des bases de données spectrales, de nouvelles approches d'analyse ont vu le jour dépendamment des problématiques visées. C'est dans ce contexte que le PLSR est utilisé pour estimer la teneur de certains composants à partir des données spectrales. Bien que le PLSR soit largement employé et efficace pour des

jeux de données linéaires et modérément complexes, il présente certaines limites dans le traitement des données hautement non linéaires, souvent rencontrées en spectroscopie. De plus, l'explicabilité des résultats des estimations reste un défi majeur, en particulier dans des domaines sensibles, comme la criminalistique.

Parallèlement, plusieurs études d'apprentissage automatique ont abordé cette problématique, en utilisant des techniques telles que les forêts aléatoires ou les machines à vecteurs de support (SVM) pour modéliser des relations plus complexes dans les données spectrales [2, 3, 4]. Cependant, ces méthodes ne capturent pas toujours les interactions complexes à longue portée présentes dans les spectres[5, 6].

L'apprentissage profond offre un potentiel immense pour surmonter ces limites. Il existe quelques approches qui utilisent l'apprentissage profond, mais elles se limitent aux architectures convolutives (CNN)[7, 8, 9]. Bien que les CNN aient montré des résultats prometteurs, ils présentent des limitations, notamment l'incapacité à capturer les contextes globaux des relations complexes entre les longueurs d'onde dans le spectre. C'est pour cette raison qu'il est nécessaire d'explorer d'autres modèles d'apprentissage profond pour améliorer les résultats déjà obtenus dans le domaine de la spectroscopie proche infrarouge.

Questions de recherche :

Cette problématique pose les questions de recherche suivantes à partir des données de spectre proche infra-rouge :

- Dans quelle mesure les modèles d'apprentissage profond permettent-ils d'améliorer l'estimation de la datation relative des taches de sang à partir des données spectroscopiques ?
- Comment les modèles d'apprentissage profond peuvent-ils être utilisés pour estimer la concentration en protéines à partir des spectres obtenus par spec-

troscopie ?

- Comment les modèles d'apprentissage profond se comparent à d'autres modèles comme la régression des moindres carrés partiels (PLSR, pour l'analyse des données spectroscopiques proche-infra-rouge ?

1.2 Objectifs

L'objectif principal de ce mémoire est de développer une approche basée sur l'apprentissage profond pour effectuer des estimations à partir des données spectroscopiques. Plus précisément, cette étude vise à explorer différentes architectures de Transformers, initialement conçues pour le traitement du langage naturel (NLP), et à comparer leurs performances avec celles des méthodes traditionnelles couramment utilisées en spectroscopie proche infrarouge (SPIR), notamment la PLSR.

Afin d'aborder cet objectif principal, nous l'avons découpé en trois objectifs secondaires :

- Développer plusieurs architectures de Transformers afin d'estimer la concentration en protéines des fourrages et des aliments à partir des données spectroscopiques proche-infra-rouge.
- Évaluer la capacité de ces architectures à estimer la datation relative des taches de sang, en analysant leur performance sur des données spectroscopiques proche-infra-rouge.
- Comparer les performances de ces modèles d'apprentissage profond avec celles des approches traditionnelles utilisées en spectroscopie, en particulier la régression des moindres carrés partiels (PLSR), afin de déterminer leur pertinence et leur efficacité dans ce domaine d'application.

1.3 Contributions nouvelles

Ce projet présente deux principales contributions. La principale innovation réside dans l'adaptation des Transformers pour les tâches de régression sur des données spectrales, en tirant parti des mécanismes d'attention multi-tête et de l'encodage positionnel. Cette approche exploite la capacité des Transformers à répondre aux exigences spécifiques de caractérisation des attributs d'échantillons biologiques à partir des données de réflectance.

La seconde contribution consiste en une comparaison des modèles proposés avec celles des méthodes traditionnelles en spectroscopie, notamment la régression PLSR.

1.4 Organisation du mémoire

Le chapitre 2 présente la revue de littérature sur la détection des molécules, organisé en deux parties. La première partie donne un aperçu des statistiques et tendances globales des recherches des 4 dernières années, détaillant la méthodologie employée pour identifier notre problématique. La seconde partie expose l'état de l'art des différents modèles utilisés, tout en mettant en lumière leurs limites. Le chapitre 3 introduit les concepts théoriques des modèles traditionnels, tels que les PLSR et des modèles d'apprentissage profond tels que les CNN, Transformers qui seront appliqués ultérieurement dans le chapitre 4. Ce dernier décrira la méthodologie que nous proposons pour l'estimation de la concentration de certaines molécules à l'aide des CNN et des architectures des Transformers. Enfin, le chapitre 5 analysera les résultats obtenus à la suite des expérimentations.

1.5 Conclusion

Ce chapitre a exposé la problématique, les objectifs et les contributions de ce travail. Le prochain chapitre présentera la revue de littérature afin de situer notre approche dans l'état de l'art.

Chapitre 2

Revue de littérature

Dans ce chapitre, nous examinons les travaux existants en chimiométrie, en mettant particulièrement l'accent sur les études concernant la détection/présence des molécules à l'aide des données spectroscopie proche infra-Rouge (SPIR). Nous analyserons plusieurs articles traitant de cette problématique afin de contextualiser et motiver notre sujet .

2.1 Tendances des recherches (5 dernières années)

Nous avons mené une revue systématique de la littérature scientifique en chimiométrie à partir de bases de données numériques Google Scholar, IEEE Xplore, ScienceDirect et SpringerLink. Une recherche initiale couvrant la période de 2000 à 2023 a été effectuée à l'aide des mots-clés suivants : "NIR infrared", "spectroscopy", "concentration analytics" et "detection analytics". Cette première étape a permis d'identifier environ 320 articles. Nous avons appliqué des critères de sélection : inclusion uniquement des articles évalués par les pairs, publiés entre 2020 et 2024, portant

sur des applications concrètes de la spectroscopie dans la quantification ou la détection moléculaire. Après suppression des doublons et des publications hors sujet ou redondantes, 31 articles ont été retenus. Ceux-ci ont été classés en trois catégories selon leur objectif principal (quantification, détection ou les deux), puis regroupés selon les approches méthodologiques employées, à savoir : les approches d'analyse numérique et statistique (AN/MS), les modèles d'apprentissage automatique (ML) et ceux d'apprentissage profond (DL).

La figure 2.1 présente la repartition des approches utilisées dans les articles après filtrage. Parmi les 31 articles retenus, seul 7 abordent la problématique de la détection de molécules à l'aide d'apprentissage profond. De plus, la majorité de ces études se concentrent sur l'utilisation des réseaux de neurones convolutifs. Ce qui laisse la place à l'exploration de nouveaux modèles d'apprentissage profond, tels que les Transformers.

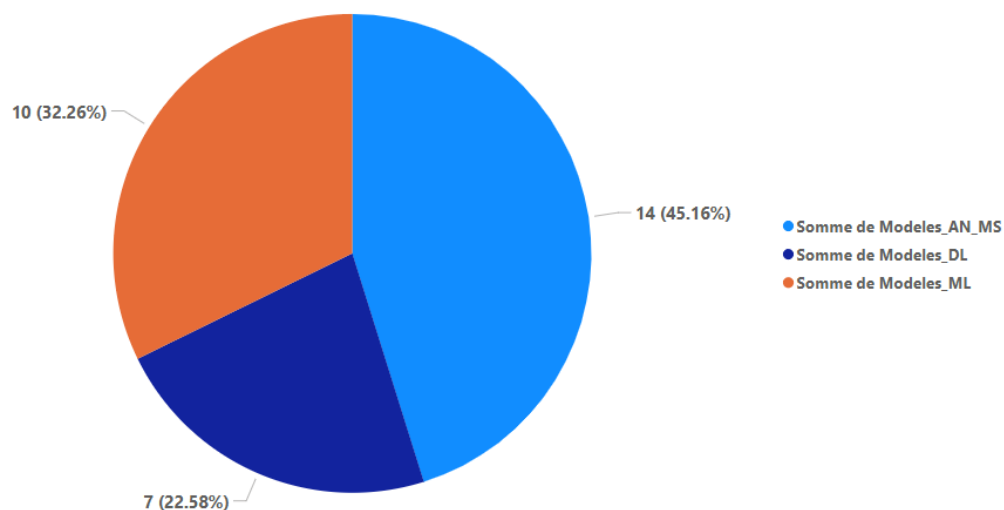


FIGURE 2.1 – *Modèles utilisés dans les articles*

Ces statistiques et tendances globales retracent l'ensemble des démarches ayant conduit à la problématique de ce mémoire, qui consiste à explorer les différentes architectures des Transformers pour la caractérisation (estimation de concentration, de la datation relative des taches de sang) de certains échantillons biologiques

2.2 Revue analyse de littérature

2.2.1 Introduction

La spectroscopie constitue la méthode privilégiée pour étudier la présence et quantifier la concentration des molécules en analysant les interactions entre la lumière et la matière. L'article fondateur de Joseph von Fraunhofer [10], publié en 1814, a mis en évidence les raies d'absorption de lumière dans le spectre solaire, marquant le début d'une ère d'analyse moléculaire approfondie. Depuis cette découverte, les capacités de caractérisation moléculaires ont été considérablement améliorées grâce à l'émergence de diverses méthodes en spectroscopie. Parmi ces méthodes, les approches d'analyse numérique et statistique ont été largement utilisées dans le cadre de plusieurs études. Par la suite, l'apprentissage automatique, puis l'apprentissage profond se sont progressivement imposés dans le domaine de la spectroscopie.

2.2.2 Approches statistiques et numériques

L'analyse des données spectroscopiques s'appuie de manière croissante sur des méthodes numériques et statistiques avancées, en réponse à la complexité et à la haute dimensionnalité des signaux obtenus. Ces approches permettent de traiter des données souvent bruitées, redondantes ou fortement corrélées, afin d'en extraire des informations pertinentes. Parmi les méthodes les plus couramment utilisées, figure la PLSR. Une technique qui intègre les fondements de la régression linéaire et de l'analyse en composantes principales (PCA) [11].

Dans [1] Lesnoff Matthieu et al. étudient l'efficacité de la régression PLS pondérée localement (LWPLSR) pour la calibration multivariée, une approche adaptée

aux données agronomiques hétérogènes, où la variabilité des échantillons engendre des relations non linéaires entre variables explicatives et réponse. LWPLSR est une extension de la PLSR pondérée (WPLSR), où chaque observation du jeu de calibration reçoit un poids en fonction de sa dissimilarité avec l'échantillon à prédire. Deux stratégies ont été comparées : LW, qui applique WPLSR à l'ensemble des observations de calibration, et KNN-LW, qui sélectionne au préalable les k plus proches voisins avant l'application du WPLSR. L'expérimentation sur trois jeux de données agronomiques a montré que LW et KNN-LW surpassaient la PLSR standard avec une MSE de 0.64, avec des performances similaires, mais un temps de calcul nettement réduit pour KNN-LW, en faisant une solution privilégiée pour les grands ensembles de données. Par ailleurs, l'étude introduit un nouvel algorithme WPLSR, basé sur l'« improved kernel #1 », qui se révèle plus rapide que l'approche NIPALS pondérée utilisée jusqu'alors. Ces résultats confirment l'intérêt de KNN-LW pour améliorer l'efficacité des calibrations multivariées tout en réduisant la complexité computationnelle dans le traitement des données agronomiques.

Shawky Eman et al.[12] développent un protocole analytique rapide et précis utilisant la spectroscopie proche infrarouge combinée à l'analyse multivariée des données pour authentifier, détecter et quantifier les principaux contaminants végétaux fréquemment utilisés pour falsifier les stigmates de safran. Parmi ces contaminants figurent le carthame, la peau de fruit de grenade, la fleur de calendula, le paprika, le curcuma, l'hibiscus, les étamines de safran et les stigmates de safran épuisés. Un modèle SIMCA (Soft Independent Modelling of Class Analogies) a été construit pour authentifier les stigmates de safran avec une sensibilité de 100%. Par ailleurs, un modèle PLS-DA (Partial Least Squares-Discriminant Analysis) a permis de discriminer correctement les échantillons de safran non altérés et ceux délibérément falsifiés, atteignant une sensibilité de 100% et une spécificité de 99%. La quantification des différents contaminants a été réalisée à l'aide de modèles de régression PLS (Partial Least Squares Regression), accompagnés de la méthode VIP (Variable Importance to Projection) pour la sélection des variables. Ces modèles ont révélé que les bandes

dans les gammes spectrales de 6000–5800 cm^{-1} , suivies de 4600–4200 cm^{-1} et de 5400–5000 cm^{-1} , étaient les plus importantes pour une prédiction précise, avec des limites de détection aussi basses que 1%. La performance des modèles a été testée à l’aide de jeux de validation internes et externes, démontrant leur fiabilité en tant qu’outil d’évaluation de la qualité du safran afin de prévenir la fraude.

Lan Weijie et al.[13] explorent l’utilisation combinée de la spectroscopie visible-proche infrarouge (Vis-NIRS), de la spectroscopie moyenne infrarouge (MIRS) et de modèles de régression des moindres carrés partiels (PLS) couplés à l’apprentissage automatique. Leur objectif est double : retracer les proportions de différentes variétés de pommes dans des purées formulées et prédire les caractéristiques de qualité des purées à partir des données spectrales des cultivars de purée initiale. Ils ont démontré que les modèles PLSR pouvaient estimer les proportions de chaque cultivar de pomme dans les mélanges de purée en utilisant les spectres MIR (avec un RMSEP inférieur à 8.1% et un RPD supérieur à 3.6), particulièrement pour la variété Granny Smith (RMSEP de 2.7% et RPD de 11.4). En outre, les auteurs ont appliqué la méthode MCR-ALS (Multivariate Curve Resolution-Alternating Least Squares) pour reconstruire les spectres des purées formulées à partir des profils de concentration. Ils ont également montré que la technique MIRS pouvait prédire de manière fiable la qualité finale des purées, en particulier en ce qui concerne la viscosité ($\text{RPD} > 4.0$), la teneur en solides solubles ($\text{RPD} = 4.1$), l’acide malique ($\text{RPD} = 4.7$) et le glucose ($\text{RPD} = 4.3$). Ainsi, les auteurs concluent que la spectroscopie infrarouge est un outil prometteur pour assurer la traçabilité des purées, et peut même être utilisée pour l’optimisation multicritères des produits finaux à partir des caractéristiques des cultivars de purée avant formulation.

Mahajan Gopal et al. [14] explorent la télédétection hyperspectrale comme alternative aux méthodes conventionnelles d’estimation des nutriments foliaires du manguiier, qui nécessitent des analyses chimiques longues et coûteuses. L’objectif de leur étude était de développer des indices spectraux spécifiques et d’appliquer des méthodes

d'analyse multivariée et d'apprentissage automatique pour modéliser la relation entre les spectres hyperspectraux et la composition nutritionnelle des feuilles. Pour cela, ils ont collecté des données spectrales dans la gamme de 350 à 1050 nm et testé différentes approches, notamment les indices spectraux normalisés et de ratio, ainsi que plusieurs modèles de régression comme la PLSR, la régression sur composantes principales et la régression par vecteurs de support (SVR). Cependant, ces modèles se sont révélés inefficaces pour prédire avec précision les nutriments foliaires. En réponse à cette limitation, ils ont combiné la PLSR avec des modèles d'apprentissage automatique, obtenant ainsi des performances améliorées. Les meilleurs résultats ont été obtenus avec le modèle Cubist ($R^2 \geq 0,91$, $RPD \geq 3,3$, $RPIQ \geq 3,71$) pour l'azote, le phosphore, le potassium et le zinc, le modèle SVR ($R^2 \geq 0,88$, $RPD \geq 2,73$, $RPIQ \geq 3,31$) pour le calcium, le fer, le cuivre et le bore, et le modèle Elastic Net ($R^2 \geq 0,95$, $RPD \geq 4,47$, $RPIQ \geq 6,11$) pour le magnésium et le soufre. Ces résultats démontrent le potentiel de la télédétection hyperspectrale pour une estimation non destructive des nutriments foliaires, offrant ainsi une solution efficace pour la gestion de la fertilisation en agriculture de précision.

Hu Chuanmin et al. [15] explorent la détection des débris marins par télédétection dans les bandes visible et proche infrarouge, en se basant sur des simulations, des analyses spectrales et des données Sentinel-2. L'étude montre que les microplastiques, même à leur densité maximale rapportée ($\sim 10^7$ particules/km²), produisent un signal 60 fois trop faible pour être détecté ($\chi \approx 0,005$ %, alors que le seuil minimum est $\chi \geq 0,2$ % pour un capteur SNR = 200). En revanche, les macroplastiques ou autres débris flottants peuvent être détectés et discriminés des algues flottantes à partir d'une couverture sous-pixellaire de $\geq 0,8$ % (Sentinel-2) ou $\geq 0,3$ % (HyspIRI), grâce à des différences spectrales dans le NIR et autour de 670 nm. Les simulations montrent aussi que la variation de résolution des bandes Sentinel-2 introduit des distorsions spectrales, rendant les analyses pixel par pixel peu fiables. En combinant la différence de réflectance (ΔR) et une moyenne de plusieurs pixels (ex. : 5×5), l'étude parvient à identifier certains types de débris, mais conclut qu'il reste difficile de distinguer

les types de déchets non-algaux, ceux-ci présentant des spectres plats similaires. Des recommandations sont faites sur l'utilisation de bandes spécifiques (560, 665, 865 nm) et d'indices (NRD, SAM) pour améliorer la détection dans un contexte opérationnel.

Sechul Chun et al. [16] examinent l'usage des dispositifs analytiques portables pour la détection *in situ* des microplastiques (MPs) dans l'environnement, une alternative aux méthodes de laboratoire coûteuses, lentes et souvent non adaptées aux terrains. La revue analyse les principales méthodes classiques (FTIR, Raman, GC-MS, SEM) et décrit leurs limites en matière de portabilité, sensibilité ou préparation des échantillons. Divers dispositifs portables, comme des spectromètres NIR miniaturisés, des prototypes optiques, des spectroradiomètres ou encore des appareils basés sur les smartphones, ont été testés pour détecter et caractériser les MPs en eau, sol ou sédiments, avec des taux de récupération variant entre 90 % et 98 % selon les matériaux et contextes. Par exemple, le MicroNIR OnSite-W a permis la distinction de plusieurs types de polymères (PE, PP, PS) dans des échantillons modèles, tandis que des prototypes optiques ont détecté des MPs translucides jusqu'à 0,06 μm de rugosité. Malgré ces avancées, la détection de nanoplastiques reste hors de portée des dispositifs actuels. L'article souligne la nécessité d'adapter des technologies avancées (e.g. FLIM, DHI, TD-PTR-MS) à des formats portables, et appelle à l'intégration de l'IA et de l'IoT pour surmonter les défis liés à la résolution, la reproductibilité et l'analyse de données en conditions réelles.

Li Xue et al. [17] passent en revue les applications de la spectroscopie dans le proche infrarouge (NIR) pour l'analyse non destructive de la qualité des graines oléagineuses et des huiles comestibles. Cette technologie permet d'évaluer rapidement et sans traitement préalable des paramètres comme la teneur en huile, protéines, humidité, acides gras ou encore les composés spécifiques (glucosinolates, polyphénols, tocophérols), avec des coefficients de détermination (R^2) allant jusqu'à 0,99 pour certains acides gras ou composés nutritionnels. L'étude rapporte des modèles robustes pour la détection de l'authenticité et de la traçabilité géographique, comme la différen-

ciation d'huiles mélangées ($R^2 > 0,98$) ou l'origine de graines de sésame. En parallèle, la NIR-HSI (*hyperspectral imaging*) est utilisée pour cartographier la composition chimique à l'échelle du grain, atteignant des performances élevées pour prédire la teneur en huile ou protéines (R_p^2 jusqu'à 0,945 ; RMSEP < 0,2 %). L'article souligne l'intérêt croissant de la NIR comme méthode verte et rapide pour le contrôle qualité, mais rappelle la nécessité d'optimiser les méthodes chimiométriques pour améliorer la détection des composés en faible concentration et gérer les interférences spectrales.

Guo Zhiming et al. [18] développent une méthode rapide et non destructive basée sur la spectroscopie proche infrarouge (NIR) couplée à des algorithmes de sélection de variables pour évaluer la qualité gustative du thé matcha. L'étude se concentre sur trois indicateurs : la teneur en polyphénols totaux (TP), les acides aminés libres (FAA) et le ratio TP/FAA, qui reflète l'équilibre entre astringence et *umami*. À partir de 138 échantillons de matcha, plusieurs modèles chimiométriques ont été testés (PLS, SiPLS, SiPLS-SPA, SiPLS-GA, SiPLS-SA) après prétraitement des spectres par différentes méthodes (dont MSC, identifiée comme la plus performante). Les meilleurs résultats ont été obtenus avec les modèles SiPLS-SA ($R_c \geq 0,9775$; $R_p \geq 0,9724$; RMSEP < 0,47) et SiPLS-SPA ($R_p \geq 0,986$), utilisant seulement 9 à 18 variables spectrales sélectionnées dans des plages NIR spécifiques. Ces modèles surpassent les approches classiques en précision et efficacité, démontrant le potentiel de la NIR pour un contrôle qualité intelligent et rapide du matcha, applicable en contexte industriel.

Amanah Hanim et al. [19] évaluent la capacité des spectroscopies FT-NIR et FT-IR à estimer de manière non destructive la teneur totale en anthocyanines (TAC), ainsi que les teneurs en cyanidine-3-glucoside (C3G) et delphinidine-3-glucoside (D3G), dans des graines de soja entières. À partir de 70 variétés de soja noir, les spectres ont été collectés sur des graines individuelles et comparés aux résultats de référence obtenus par HPLC. Les modèles PLSR, optimisés par prétraitement (notamment SaViTzky-Golay pour FT-NIR et MSC pour FT-IR), ont atteint des performances élevées avec des coefficients de détermination en validation supérieurs à 0,85 pour

tous les composés (R_v^2 jusqu'à 0,90 pour C3G en FT-NIR), et des erreurs de prédiction allant de 9,4 % à 21,8 %. L'utilisation de la méthode VIP a permis de réduire de 45 à 72 % le nombre de variables spectrales sans perte significative de performance. FT-NIR s'est montré légèrement plus précis que FT-IR, bien que ce dernier offre une meilleure stabilité entre les phases de calibration et validation.

Kucharska Ambrożej et al.[20] passent en revue les applications des techniques spectroscopiques combinées à la chimiométrie pour détecter les falsifications dans les herbes et épices. Des méthodes comme l'IR, le NIR, le Raman, le FTIR ou encore le DRIFTS, couplées à des algorithmes (PCA, PLS-DA, HCA, OPLS-DA...), ont permis de détecter des adultérants dans le safran, le poivre noir, le paprika, l'ail, l'oignon, l'origan ou encore le thé. Par exemple, un modèle DRIFTS-PLS-DA a permis d'identifier des contaminants dans le safran à des niveaux de 1–3,1 % (w/w), tandis que des modèles basés sur NIR/FTIR ont détecté des fraudes dans le poivre contenant des graines de papaye ou des résidus. L'authenticité de produits comme le café, l'origan ou le thé a aussi été évaluée via des empreintes spectrales uniques et des modèles de classification atteignant jusqu'à 96 % de précision. L'étude souligne que ces approches sont rapides, non destructives et adaptées à l'analyse d'un grand nombre d'échantillons, mais requièrent des bases de données fiables et des algorithmes robustes.

Nawar Said et al.[21] étudient le potentiel de la spectroscopie visible et proche infrarouge (Vis-NIRS, 350–2500 nm) pour l'évaluation rapide des éléments potentiellement toxiques (PTEs) dans les sols agricoles contaminés, un enjeu majeur pour la gestion environnementale. L'étude visait à quantifier les concentrations de Cd, Co, Cu, Cr, Pb et Zn en combinant Vis-NIRS avec la PLSR et des méthodes de sélection de variables. Un total de 80 échantillons de sol a été collecté dans une zone polluée autour du drainage d'Al-Moheet, en Haute-Égypte, et analysé à l'aide d'un spectroradiomètre ASD FieldSpec-4. Les longueurs d'onde pertinentes ont été sélectionnées via les méthodes Simulated Annealing (SA) et Uninformative Variable Elimination (UVE), avant d'être intégrées aux modèles PLSR. Trois approches ont été comparées : FR-PLS (spectre

complet), SA-PLS (sélection par SA) et UVE-PLS (sélection par UVE). Les résultats ont montré que les modèles UVE-PLS surpassaient les autres approches, atteignant des performances optimales pour Cr ($R^2 = 0,74$, $RPD = 2,48$), Pb ($R^2 = 0,72$, $RPD = 2,03$) et Cd ($R^2 = 0,62$, $RPD = 1,86$), tandis que les prédictions de Cu, Co et Zn restaient à améliorer. Ces résultats soulignent le potentiel du modèle UVE-PLS pour l'évaluation rapide des PTEs dans les sols agricoles contaminés, avec des perspectives d'optimisation pour une application plus large en environnement aride.

Minas Ioannis et al. [22] développent des modèles de prédiction non destructifs basés sur la spectroscopie visible et proche infrarouge (Vis-NIRS) pour évaluer la qualité interne et la maturité physiologique des pêches à l'aide d'un seul scan. En utilisant un protocole combinant différentes charges de culture (*crop load*) et stades de développement, les auteurs calibrent des modèles multivariés sur 350 fruits de la variété *Sierra Rich* scannés à trois températures (0, 20, 30 °C). Les modèles prédictifs donnent d'excellentes performances pour la matière sèche (DMC : $R^2 = 0,98$; RMSEP = 0,41 %), les solides solubles (SSC : $R^2 = 0,96$; RMSEP = 0,58 %) et l'indice de différence d'absorbance (IAD : $R^2 = 0,96$; RMSEP = 0,08), mais restent insuffisants pour la fermeté de la chair (FF : $R^2 = 0,43$). Le modèle permet aussi d'évaluer en conditions réelles l'impact des pratiques culturales : des charges de culture plus légères augmentent la DMC et le SSC, tout en accélérant la maturité (IAD plus bas), notamment dans la partie supérieure du feuillage.

Han Xueqin et al. [23] développent une méthode rapide et précise pour la prédiction de la demande chimique en oxygène (COD) dans différents systèmes aquatiques, un enjeu essentiel pour le suivi de la pollution de l'eau et la gestion des ressources hydriques. L'étude combine la spectroscopie proche infrarouge (NIR) avec la PLSR afin de modéliser la COD dans les eaux usées, les eaux de surface et l'eau de mer, collectées à Guangzhou, en Chine. Trois méthodes de prétraitement ont été appliquées aux spectres pour améliorer la précision du modèle, et deux algorithmes de sélection de variables ont été évalués : Binary Gray Wolf Optimization (BGWO) et Compe-

titive Adaptive Reweighting Sampling (CARS). Les résultats montrent que ces deux méthodes améliorent la précision du modèle tout en réduisant le nombre de longueurs d'onde utilisées, surpassant la performance de PLSR seul. Le modèle CARS-PLSR a obtenu les meilleures performances, réduisant le nombre d'entrées spectrales de 96 % pour l'eau de surface et les eaux usées, et de 82 % pour l'eau de mer, tout en atteignant des coefficients de détermination (R^2) de 0,860, 0,815 et 0,692, respectivement. Ces résultats soulignent l'intérêt des algorithmes de sélection spectrale pour identifier les variables clés reliant la COD aux spectres NIR, améliorant ainsi la précision et la simplicité des modèles PLSR pour le suivi de la qualité de l'eau dans divers environnements aquatiques.

Lu Yongzong et al.[24] développent une méthode rapide et non destructive pour évaluer les lésions dues au froid (CI) chez le théier, un problème majeur affectant la qualité des thés renommés et entraînant des pertes économiques importantes. L'étude repose sur l'utilisation de la spectroscopie térahertz (THz) pour analyser l'absorbance des feuilles de thé exposées à des températures basses (4,0, 0, -2,5, -5,0, -7,5 et -10,0 °C) dans la gamme de fréquences 0,1-1,6 THz. La transformée de Fourier rapide (FFT) a été appliquée pour convertir les spectres temporels en spectres fréquentiels et améliorer le rapport signal/bruit. Afin de mieux séparer les signaux superposés dus à la diffusion causée par les cellules endommagées, deux approches ont été intégrées à la régression par moindres carrés partiels (PLSR) : la spectroscopie de corrélation bidimensionnelle (2DCOS) et l'intensité moyenne (AI), donnant naissance aux modèles 2DCOS-PLSR et AI-PLSR. L'évaluation quantitative a montré que 2DCOS-PLSR surpassait AI-PLSR, avec des coefficients de corrélation (R^2) atteignant 0,7873, 0,8305 et 0,9103, et des erreurs quadratiques moyennes (RMSE) de 0,6032, 0,5763 et 0,5221, respectivement. Bien que le modèle AI-PLSR ait obtenu des résultats proches ($R^2 = 0,7477$ à $0,8974$ et $RMSE = 0,5797$ à $0,6038$), la combinaison THz spectroscopy + 2DCOS-PLSR s'est révélée être la plus précise pour la détection des lésions causées par le froid, offrant ainsi une référence améliorée pour la sélection des intervalles spectraux et l'optimisation de la détection des dommages sur

les feuilles de thé.

2.2.3 Les modèles d'apprentissage automatique

Selon l'université de Stanford[25], l'apprentissage automatique est défini comme une famille d'algorithmes permettant aux systèmes d'apprendre à partir des données sans être explicitement programmé. Il a été adapté dans de nombreuses tâches spécifiques, en particulier dans le domaine de la spectroscopie, plusieurs chercheurs ont étudié la possibilité d'utiliser cette approche dans le domaine spectral.

Qiu Ting et al.[26] développent une méthode rapide et précise pour distinguer la racine de kudzu (KR) de la tige de kudzu (KS) et détecter d'éventuelles fraudes commerciales impliquant l'adultération de KR avec KS, une problématique impactant la qualité des compléments alimentaires et la protection des consommateurs. En combinant la spectroscopie proche infrarouge (NIR) avec quatre algorithmes — l'analyse en composantes principales (PCA), l'analyse discriminante par moindres carrés partiels (PLS-DA), la régression PLS et la machine à vecteurs de support à moindres carrés (LS-SVM) —, l'étude a permis d'évaluer l'efficacité de ces méthodes. PLS-DA a atteint une précision de classification de 100 % pour distinguer KR de KS, tandis que LS-SVM a surpassé le modèle PLS dans la prédiction du taux d'adultération, obtenant un coefficient de corrélation R_p de 0,9972 et une erreur quadratique moyenne de prédiction (RMSEP) de 2,1724, garantissant une excellente précision. Ces résultats démontrent le potentiel de la spectroscopie NIR couplée à l'apprentissage automatique comme outil pour détecter les falsifications de KR, offrant ainsi une solution fiable pour le contrôle qualité et la prévention des fraudes alimentaires.

Massei Ambra et al.[27] développent une méthode innovante basée sur la spectroscopie proche infrarouge (NIR) et l'apprentissage automatique pour prédire l'humidité résiduelle (RM) dans les produits lyophilisés, un paramètre clé influençant la stabi-

lité des principes actifs pharmaceutiques (API). Actuellement, la mesure de la RM repose sur la titration Karl-Fischer (KF), une méthode destructive et chronophage, d'où l'intérêt d'une approche alternative rapide et non invasive. L'étude compare un modèle de régression linéaire et un modèle de réseau de neurones, dont l'architecture a été optimisée pour minimiser l'erreur quadratique moyenne (MSE) et améliorer la précision des prédictions. L'analyse a pris en compte plusieurs facteurs, notamment la plage de longueurs d'onde, la forme des spectres et le type de modèle. L'approche a été testée sur différentes formulations, principalement composées de saccharose à 3 %, 6 % et 9 %, ainsi que sur des mélanges saccharose-arginine et une formulation contenant du tréhalose. Un modèle spécifique aux mélanges à 6 % de saccharose s'est révélé efficace pour prédire la RM dans des formulations similaires, mais a échoué avec des teneurs élevées en arginine. Pour surmonter cette limitation, un modèle global intégrant une proportion représentative de toutes les formulations a été développé. Les résultats démontrent que le modèle basé sur l'apprentissage automatique est plus précis et robuste que les modèles linéaires, confirmant le potentiel de cette approche pour un contrôle qualité plus rapide et efficace des produits lyophilisés.

Mohd Hilmi et al. [28] développent une méthode rapide et non destructive pour détecter l'infection à *Ganoderma boninense* chez le palmier à huile à l'aide de la spectroscopie proche infrarouge (NIR) couplée à l'analyse chimiométrique et à l'apprentissage automatique. À partir de six plants (trois sains, trois infectés), les auteurs collectent des spectres NIR (900–1700 nm) entre 16 et 32 semaines d'âge et identifient deux longueurs d'onde discriminantes (1310 et 1452 nm) liées respectivement à l'ergostérol (biomarqueur fongique) et à la teneur en eau. L'analyse HPLC confirme une corrélation élevée ($R^2 = 0,989$) entre le temps post-inoculation et la concentration d'ergostérol. La régression PLS appliquée aux spectres atteint un $R = 0,861$ pour la prédiction de l'ergostérol. Quatre algorithmes de classification (kNN, NB, SVM, DT) sont testés sur 576 spectres, avec le modèle d'arbre de décision (DT) obtenant la meilleure performance (précision = 93,1 %, F1-score = 92,6 %). L'étude conclut à la faisabilité d'un système portable de détection précoce in situ de *G. boninense*, fondé

sur des capteurs NIR et des modèles d'IA, même sur des plants jeunes en pépinière.

Wilson Megan et al.[29] évaluent l'efficacité d'un spectromètre NIR de poche couplé à des algorithmes d'apprentissage automatique pour détecter des constituants endogènes et des drogues dans les ongles humains. L'étude porte sur 20 participants, dont les ongles ont été analysés "tels quels", puis artificiellement dopés avec différentes substances (cocaïne HCl, benzocaïne, procaïne, lidocaïne, diltiazem, levamisole, carbonate de calcium). Les spectres (980–1670 nm) ont été traités par des méthodes chimiométriques (MSC, dérivée première) et analysés via CWS, PCA et SOM. Les résultats montrent que certaines drogues (comme la cocaïne HCl) sont bien détectées en raison de leur forte affinité avec les protéines kératiniques, tandis que d'autres (benzocaïne, procaïne) génèrent des regroupements moins distincts. CWS a atteint des corrélations internes de $r = 0,9999$, tandis que la PCA a identifié 98,1 % de la variance sur les trois premiers axes, distinguant clairement les ongles dopés des non dopés. SOM a permis une classification encore plus fine, détectant cinq groupes dont certains reflétaient des erreurs de type I ou II.

Yu Keqiang et al.[30] étudient l'utilisation de l'imagerie hyperspectrale (HSI, 400–1000 nm) couplée à l'apprentissage automatique pour détecter le stress dû au mercure (Hg) dans le tabac (*Nicotiana tabacum*). Des plants soumis à quatre niveaux de concentration (0, 1, 3, 5 mg·L⁻¹) ont été analysés via PCA, CARS et différents modèles LS-SVM pour classer les groupes stressés et non stressés. L'analyse spectrale a révélé des différences autour des longueurs d'onde 550, 700–750 et 970 nm, liées respectivement aux pigments respiratoires, à la « red edge » végétale et à l'eau. Les modèles LS-SVM basés sur 13 longueurs d'onde sélectionnées par CARS ont atteint une précision de 100 % pour distinguer les groupes stressés et non stressés, surpassant les modèles PLS-DA (≤ 50 %). En revanche, la classification entre les trois groupes stressés a donné de moins bons résultats (précision $\leq 66,67$ %), probablement à cause de similitudes morphologiques et spectrales à faible stress. L'étude confirme le potentiel de la HSI associée à des algorithmes supervisés pour une détection précoce et non

destructive du stress métallique chez les plantes, mais souligne les limites actuelles pour discriminer entre degrés de stress faibles à modérés.

Arianti Nunik et al.[31] proposent une nouvelle approche d'optimisation automatique du prétraitement des spectres NIR, appelée *AGoES* (Automatically Generating a Pre-processing Strategy), pour améliorer la calibration multivariée et la prédiction des propriétés des échantillons complexes. Le prétraitement des spectres NIR est essentiel pour atténuer les bruits, la diffusion de la lumière et les interférences de fond, mais sa sélection repose souvent sur des approches empiriques et l'expertise des analystes. Pour pallier cette difficulté, *AGoES* s'inscrit dans une approche d'ensemble, combinant plusieurs méthodes de prétraitement et exploitant des algorithmes d'apprentissage automatique, notamment la régression par moindres carrés partiels (*PLSR*), les machines à vecteurs de support (*SVM*), le k-plus proches voisins (*k-NN*), les arbres de décision (*DT*), le boosting adaptatif (*AB*) et la régression gaussienne (*GPR*). Ces modèles sont optimisés via une validation croisée à 5 plis et une recherche par grille. L'approche a été testée sur un jeu de données public de spectres NIR afin de prédire trois paramètres des déchets organiques issus du fumier : la matière sèche (*DM*), la matière organique (*OM*) et l'azote ammoniacal (*AN*). Les résultats montrent que *SVM* associé à *AGoES* offre les meilleures performances pour *DM* ($RPD = 3,619$) et *AN* ($RPD = 2,996$), tandis que *AB* combiné à *AGoES* est le plus performant pour *OM* ($RPD = 3,185$). Cette approche non supervisée simplifie considérablement le choix du prétraitement et améliore la robustesse des analyses multivariées, rendant l'utilisation de l'apprentissage automatique plus accessible et efficace pour l'exploitation des spectres NIR.

Rocha Matheus et al. [32] explorent l'application de la spectroscopie proche infrarouge (NIR) pour le diagnostic assisté par ordinateur (*CAD*) des lésions cutanées, un enjeu majeur pour la détection précoce du mélanome, un cancer agressif responsable d'un taux de mortalité élevé. L'un des principaux obstacles à l'utilisation de l'apprentissage automatique et profond (*MDL*) en spectroscopie réside dans l'absence

de jeux de données publics. Pour combler cette lacune, les auteurs ont constitué *NIR-SC-UFES*, un nouveau jeu de données annoté, servant de référence pour la classification des lésions cutanées cancéreuses et non cancéreuses à partir de spectres NIR. Plusieurs algorithmes d'apprentissage ont été testés, notamment *XGBoost*, *CatBoost*, *LightGBM*, les machines à vecteurs de support (*SVM*), l'analyse discriminante par moindres carrés partiels (*PLS-DA*) et un réseau de neurones convolutif 1D (*1DCNN*). Les résultats expérimentaux révèlent que *LightGBM*, lorsqu'il est combiné à un pré-traitement par *Standard Normal Variate* (*SNV*), une extraction de caractéristiques et une augmentation des données via des réseaux génératifs adversariaux (*GAN*), offre les meilleures performances, avec une précision de 0,852, un rappel de 0,851, une *accuracy* équilibrée de 0,839 et un *F-score* de 0,850. Ces résultats marquent une avancée significative dans l'automatisation du tri des patients atteints de lésions cutanées *in vivo* et mettent en lumière le potentiel du *CAD* basé sur les spectres NIR pour améliorer la détection précoce du cancer de la peau.

Nturambirwe Jean et al.[33] étudient l'utilisation des données spectroscopiques proches infrarouges (*NIR*, 800–2500 nm) pour la classification des contusions sur trois cultivars de pommes (*Golden Delicious*, *Granny Smith* et *Royal Gala*), un enjeu clé pour l'évaluation rapide de la qualité des produits horticoles. L'exploitation de l'ensemble du spectre *NIR* n'est pas toujours pratique dans un contexte de production en raison de la haute dimensionnalité des données, de la colinéarité entre longueurs d'onde et de la complexité accrue des modèles. L'objectif de l'étude était donc d'optimiser la sélection des variables spectrales afin d'améliorer la robustesse des modèles tout en réduisant leur complexité. Six algorithmes d'apprentissage automatique ont été testés. Deux méthodes de sélection de variables ont permis d'identifier les longueurs d'onde les plus pertinentes pour différencier les fruits contusionnés des non contusionnés, principalement autour de 900 nm, 1300 nm, 1500 nm et 1900 nm. Les meilleurs résultats ont été obtenus avec la régression linéaire et les machines à vecteurs de support (*SVM*), en utilisant jusqu'à 40 longueurs d'onde. Ces modèles ont atteint des précisions de 0,79 à 0,86, comparables à celles obtenues en exploitant l'ensemble du

spectre. Cette approche a conduit au développement d'un cadre *open-source*, adapté à la gradation rapide des pommes selon les dommages mécaniques et possédant le potentiel d'être généralisé à d'autres défauts sur les produits frais dans le secteur agroalimentaire.

Gruszczyński Wojciech [34] *et al.* évaluent l'efficacité de la spectroscopie visible et proche infrarouge (Vis-NIR) pour l'estimation des propriétés physico-chimiques des sols et la prédiction d'un indice de qualité des sols (*Topsoil Quality Index*, TQI) en exploitant la base de données LUCAS. Trois approches ont été testées pour prédire des caractéristiques du sol, notamment la teneur en argile, le pH, le carbone organique (*SOC*), le carbonate de calcium ($CaCO_3$), l'azote (*N*) et la capacité d'échange cationique (*CEC*). Trois types de modèles d'apprentissage automatique ont été comparés : un *Stacked AutoEncoder* (*STACKAuto*), un réseau de neurones convolutif (*CNN*) et un modèle empilé de perceptrons multicouches (*MLP*). Les résultats montrent que le *CNN* a fourni les meilleures performances globales, bien que les modèles empilés (*STACKAuto*) offrent des estimations plus précises dans certains cas. Par exemple, *STACKAuto* a atteint des coefficients de détermination de $R^2 = 0.72$ pour la teneur en argile (RMSE = 3.66), $R^2 = 0.77$ pour le *SOC* (RMSE = 1.47), $R^2 = 0.76$ pour le $CaCO_3$ (RMSE = 5.64) et $R^2 = 0.79$ pour la *CEC* (RMSE = 1.56) sur un jeu de données de sols de Pologne, surpassant les modèles entraînés sur les données européennes. En revanche, le modèle *CNNMulti* a montré des performances plus équilibrées mais légèrement inférieures pour la prédiction du pH ($R^2 = 0.64$, RMSE = 0.46) et de l'azote ($R^2 = 0.59$, RMSE = 0.50). Ces résultats confirment le potentiel de la spectroscopie *Vis-NIR* pour une évaluation rapide et fiable des sols et soulignent l'importance d'optimiser les modèles selon les spécificités régionales des sols tout en exploitant des techniques avancées de *machine learning* pour améliorer la robustesse des estimations.

Lucimar Da et al. [35] étudient l'application de la spectroscopie proche infrarouge (NIR) portable et à faible coût pour l'évaluation en ligne de la fraîcheur des œufs, une

problématique majeure liée aux fraudes d'étiquetage où des œufs rassis sont vendus comme frais. L'étude a utilisé un spectromètre NIR miniature pour collecter des données spectrales, traitées avec différentes méthodes de prétraitement et d'apprentissage automatique. Deux approches ont été testées : la régression (PLS-R et SVM-R) pour prédire l'unité de Haugh (HU) et la classification (PLS-DA et SVM-C) pour distinguer les œufs frais des œufs rassis. Les modèles PLS-R et SVM-R ont montré des performances similaires, mais SVM-R sur la plage spectrale 1300–1690 nm a obtenu les meilleurs résultats avec une erreur relative de 7,32% et un RPD de 2,56. Pour la classification, PLS-DA a surpassé SVM-C, atteignant une précision de 87,0%, avec une meilleure sensibilité pour détecter les œufs rassis. Ces résultats démontrent que les spectromètres NIR portables offrent une alternative fiable et économique aux appareils de laboratoire, facilitant leur intégration dans les chaînes d'approvisionnement des œufs pour renforcer le contrôle qualité alimentaire.

2.2.4 Les modèles d'apprentissage profond

L'apprentissage profond[36] peut être défini comme une sous-catégorie de l'apprentissage automatique qui utilise des réseaux de neurones artificiels à plusieurs couches pour modéliser et approcher des problèmes de regroupement, de classification, de régressions complexes. Avec les données spectrales, les réseaux de neurones convolutifs ont été largement utilisés.

Dans [37] Zhelezova et al. proposent une méthode innovante pour la détection et l'identification des maladies des plantes, en particulier *Septoria tritici* et *Stagonospora nodorum*, jugé crucial pour la durabilité de la production agricole. Ils utilisent des données hyperspectrales pour évaluer les lésions sur les cultures de blé affectées par ces pathogènes. Dans cette recherche, le blé a été infecté artificiellement et des mesures hyperspectrales ont été réalisées à l'aide d'un hyperspectromètre VIS-NIR. Les chercheurs ont ensuite développé un modèle de réseau de neurones perceptron (PNN)

pour comparer les données des plantes saines et malades. Les résultats montrent que le PNN atteint une précision de 0,99, avec des performance tels de 0,94 en précision, 0,89 en rappel et 0,91 pour le F-score après 10 000 itérations d'apprentissage. L'exactitude de la détection des feuilles malades varie entre 88,1 % et 97,7 %, tandis que l'identification des degrés légers et moyens après maladie présente des performances moindres (38–66%). Cette approche de classification par signature hyperspectrale pourrait donc s'avérer bénéfique pour l'évaluation des lésions de STB et SNB dans des conditions de champ avant la récolte.

Zhou Hong et al. dans [38] proposent l'utilisation de la spectroscopie hyperspectrale dans le proche infrarouge (NIR) pour l'inspection non destructive des graines de blé. Ils ont constitué un vaste ensemble de données spectrales de plus de 140 000 grains de blé appartenant à 30 variétés différentes pour la classification. Afin de traiter cet ensemble de données, ils ont conçu un sélecteur de caractéristiques basé sur un réseau de neurones convolutif (CNN-FS), permettant de sélectionner les canaux spectraux les plus pertinents. Un réseau de neurones convolutifs avec attention (CNN-ATT) a également été développé pour la classification des données en une dimension. Les performances de ce modèle ont été comparées à celles d'algorithmes d'apprentissage automatique classiques, tels que la machine à vecteurs de support (SVM) et l'analyse discriminante par régression des moindres carrés partiels. Le CNN-ATT a montré de meilleures performances, atteignant une précision de 93,01 % avec l'ensemble complet de spectres et conservant une haute précision (90,20 %) avec un sous-ensemble de 60 canaux sélectionnés via le CNN-FS. Les méthodes proposées démontrent un potentiel important pour l'analyse de grands ensembles de données spectrales pour la classification automatique et non destructive de grains de blé.

Sinitisa Sergey et al. [39] explorent l'utilisation de l'apprentissage profond et de l'apprentissage par transfert pour l'analyse des spectres de fluorescence émise (EFS) en vue d'améliorer l'estimation des concentrations en solution, un défi majeur en Chimie et en surveillance de la qualité de l'eau. Les méthodes classiques de ré-

gression linéaire, bien qu'économiques et largement utilisées, sont souvent inefficaces face aux relations non linéaires complexes et aux conditions de terrain où le signal est bruité et les effets environnementaux moins contrôlés. Deux modèles ont été développés : R (module de régression) et ER (Encoder-Regression), et comparés aux méthodes linéaires et à un modèle DNN naïf. Sur des données de laboratoire de haute qualité, les modèles proposés ont largement surpassé les approches classiques, avec une erreur relative moyenne maximale de 11%, contre 184% pour les méthodes linéaires. En conditions simulées de données de faible qualité, l'erreur des modèles linéaires est montée à 263%, tandis que le modèle ER est resté stable à 9%, démontrant sa robustesse. À faibles concentrations, ER a obtenu les meilleures prédictions, avec des seuils de détection de 3,46 ppb sur les données de haute qualité et 2,4 ppb sur les données bruitées. Ces résultats confirment l'efficacité des modèles DNN et du transfert d'apprentissage pour améliorer la précision des capteurs de surveillance de l'eau, ouvrant la voie à des systèmes plus fiables et adaptés aux conditions réelles d'utilisation.

Passos Dário et al. [40] proposent une approche d'apprentissage profond pour la modélisation des données spectrales, en réponse aux défis posés par le choix de l'architecture du réseau neuronal et des hyperparamètres, qui influencent directement la robustesse des modèles. L'étude intègre des techniques avancées d'optimisation Hyperband et Bayésienne permettant d'ajuster automatiquement l'architecture neuronale et les hyperparamètres afin d'obtenir un modèle optimal. Une architecture de base est initialisée, puis modifiée dynamiquement pour améliorer les performances du modèle. Pour améliorer l'interprétabilité du réseau, un schéma de pondération des longueurs d'onde basé sur Grad-CAM (*Gradient-Weighted Class Activation Mapping*) a été mis en place. L'efficacité de cette approche a été testée sur un cas réel de classification de variétés de blé à partir de données spectrales NIR. Comparée aux approches DL et chimiométriques précédemment appliquées sur le même jeu de données, cette méthode a atteint une précision de classification de 94,9%, surpassant le meilleur résultat antérieur (93%) avec une architecture neuronale plus simple. Ces résultats démontrent que l'automatisation de l'apprentissage profond via des techniques d'optimisation avan-

cées permet non seulement d'améliorer la précision des modèles spectraux, mais aussi de réduire leur complexité, facilitant ainsi leur application dans des contextes analytiques.

Wójcik Szymon et al. [41] développent une approche basée sur les réseaux de neurones récurrents (RNN) pour estimer la teneur en sirop de glucose-fructose (GFS) ajouté aux jus de pomme, une problématique clé pour la détection des falsifications alimentaires. L'étude s'appuie sur des mesures voltammétriques effectuées avec une électrode quadruple en iridium, appliquées à 17 jus de pomme d'origines variées, répartis en six groupes avec des ajouts de 0 à 50 % de GFS. Les méthodes classiques de régression linéaire multivariée ont montré des performances limitées, avec un R^2 ne dépassant pas 0,86 et une erreur quadratique moyenne de prédiction (RMSEP) entre 6,3 et 10,8 %. En revanche, l'approche basée sur les RNN avec couche Long Short-Term Memory (LSTM) a significativement amélioré la précision, réduisant le RMSEP à 1,9–2,1 %, soit environ trois fois inférieur aux modèles de calibration linéaire, avec un R^2 supérieur à 0,98. Ces résultats démontrent que l'apprentissage profond permet de développer un modèle universel et performant pour la quantification du GFS dans les jus de pomme, offrant ainsi un outil robuste pour le contrôle qualité et la lutte contre la fraude alimentaire.

Shang Hui et al.[42] exploitent la spectroscopie proche infrarouge (NIR) pour analyser *in situ* les modifications moléculaires associées à la cancérisation, en utilisant une sonde NIR développée en laboratoire. Les spectres diffusés ont été collectés à différents emplacements des zones cancéreuses et paracancéreuses, permettant d'établir un modèle discriminant basé sur un réseau de neurones convolutifs 1D (1D-CNN). Après optimisation de l'architecture du réseau, le modèle a atteint une précision de classification de 94,67 %, avec un rappel/sensibilité de 95,33 %, une spécificité de 94,00 %, une précision de 94,08 % et un score F1 de 0,9470, surpassant nettement les méthodes traditionnelles comme le *K-Nearest Neighbor* (KNN : 88,34 % de précision, 98,21 % de sensibilité, 76,11 % de spécificité, 83,59 % de précision et un F1-score de

0,9031) et l'analyse discriminante de Fisher (FDA : 90,00 % de précision, 96,43 % de sensibilité, 81,82 % de spécificité, 87,10 % de précision et un F1-score de 0,9153). Ces résultats indiquent que l'application du 1D-CNN permet de différencier efficacement les tissus cancéreux et paracancéreux du sein, offrant ainsi une méthode intelligente pour le repérage clinique, le diagnostic et le traitement du cancer du sein.

Li Huiyu et al. [43] développent BloodNet, un réseau de neurones convolutifs avec attention, destiné à estimer le temps écoulé depuis le dépôt d'une tache de sang (TSD) à partir d'images. Un ensemble de 50 000 photos, prises sur 29 jours, a été utilisé pour entraîner le modèle à classer les TSD en cinq intervalles. BloodNet repose sur seize blocs résiduels intégrant des modules d'attention spatiale et par canal. Les résultats montrent une précision de 99,44 % en validation interne et un AUC de 0,975 sur un jeu de test externe. Une comparaison avec d'autres architectures (ResNet, SEnet, Swin Transformer) montre que BloodNet obtient de meilleures performances sur plusieurs métriques. Une méthode complémentaire basée sur la spectroscopie Raman et un modèle GBDT a également été testée, avec des résultats inférieurs, notamment pour les durées de 21 et 28 jours. Malgré ses performances, le modèle ne prend pas en compte les données environnementales et s'appuie exclusivement sur l'analyse visuelle des images pour estimer le TSD. Malgré ses performances, le modèle ne prend pas en compte les données environnementales et s'appuie exclusivement sur l'analyse visuelle des images pour estimer le TSD.

2.3 Conclusion

Dans ce chapitre, nous avons présenté les différents travaux existants en soulignant les modèles utilisés. Le prochain chapitre sera consacré à l'explication du fonctionnement des principaux modèles, ainsi que leurs limites et la justification du choix de notre modèle.

Chapitre 3

Modèles d'Apprentissage

Dans cette section, nous allons présenter les bases théoriques des principaux modèles d'apprentissage automatique utilisés dans le cadre de notre travail. L'objectif est de fournir une compréhension approfondie des concepts qui sous-entendent les méthodes employées pour la caractérisation à partir des données spectrales. Nous abordons ainsi deux catégories de modèles : les réseaux de neurones convolutifs (CNN) et les Transformers, en accordant une attention particulière sur les différentes configurations de leurs architectures qui seront utilisées pour la tâche spécifique de régression linéaire au chapitre 4.

3.1 Concepts théoriques des CNN

3.1.1 Introduction au CNN

Le développement des CNN a considérablement transformé les approches de traitement des images et de reconnaissance visuelle. Contrairement aux architectures tra-

ditionnelles de réseaux de neurones, les CNN exploitent des opérations de convolution qui permettent d'extraire des caractéristiques locales pertinentes tout en réduisant la complexité computationnelle.

Le CNN se compose généralement d'une couche d'entrée, suivie d'une ou plusieurs couches de convolution alternant avec des couches de pooling, puis d'une ou plusieurs couches entièrement connectées, et enfin d'une couche de sortie.

La figure 3.1 illustre un CNN typique montrant ces composants clés.

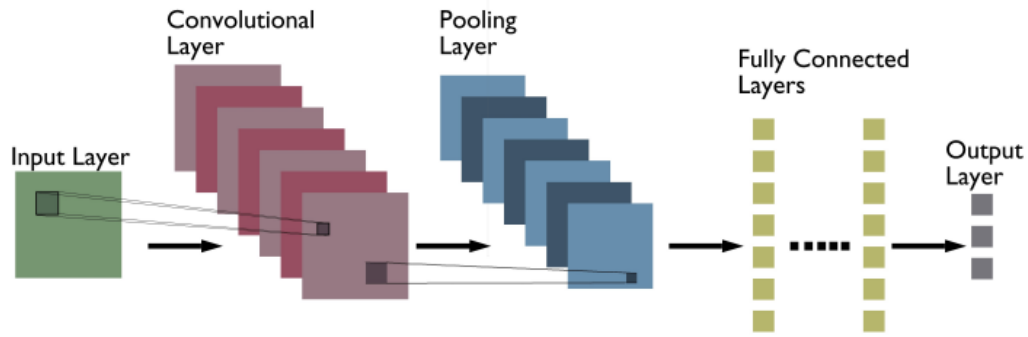


FIGURE 3.1 – Architecture typique d'un CNN.

Source : Yazdani Abyaneh et al. [44].

3.1.2 Opérations de Convolution et Stratégies de Traitement

La convolution est une opération centrale dans les CNN. Elle permet d'extraire des caractéristiques locales de l'image en appliquant un filtre qui se déplace sur l'image. L'opération de convolution est définie par la formule suivante :

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (3.1)$$

Où $I(i, j)$ est l'image d'entrée, $K(m, n)$ est le noyau de convolution (filtre), et $S(i, j)$ est la sortie.

Le *stride* (S) détermine de combien le filtre se déplace à chaque étape de la convolution, tandis que le *padding* consiste à ajouter des bordures autour de l'image pour conserver les dimensions de sortie :

$$\text{Dimension de sortie} = \frac{N - F + 2P}{\sum_m \sum_n I(i - m, j - n)K(m, n)} + 1 \quad (3.2)$$

N est la dimension de l'image, F est la taille du filtre, P est le padding, et S est le pas de translation de la fenêtre (stride).

Le pooling réduit la dimensionnalité des cartes de caractéristiques tout en conservant les informations essentielles, souvent utilisé après la convolution. Le max-pooling est défini par :

$$S_{\text{pool}}(i, j) = \max\{S(i, j), S(i + 1, j), S(i, j + 1), S(i + 1, j + 1)\} \quad (3.3)$$

3.1.3 Fonctions d'Activation, Couches Connectées et Optimisation

Les fonctions d'activation introduisent de la non-linéarité dans le réseau. La fonction *ReLU* (Rectified Linear Unit) est souvent utilisée et est définie comme :

$$f(x) = \max(0, x) \quad (3.4)$$

Elle active les neurones ayant une sortie positive.

Les couches entièrement connectées sont utilisées pour la classification. Elles ap-

pliquent un produit matriciel sur les caractéristiques extraites :

$$y = W \cdot x + b \quad (3.5)$$

Où W est la matrice de poids, x est le vecteur d'entrée, et b est le biais.

La fonction de perte mesure la différence entre la prédiction et la vérité terrain. Par exemple, l'entropie croisée est utilisée pour la classification :

$$\ell = -(y \log(p) + (1 - y) \log(1 - p)) \quad (3.6)$$

La descente de gradient permet de mettre à jour les poids selon la règle :

$$W := W - \eta \frac{\partial L}{\partial W} \quad (3.7)$$

Où η est le taux d'apprentissage.

La rétropropagation dans les CNN calcule les gradients à travers le réseau pour ajuster les poids en remontant couche par couche, à partir de la couche de sortie jusqu'aux premières couches convolutives.

Après avoir exploré les CNN et leur rôle essentiel dans les tâches de vision par ordinateur, il est maintenant pertinent d'introduire une approche émergente qui se révèle prometteuse : les Transformers. Grâce à leur mécanisme d'auto-attention, ces modèles sont envisagés comme une piste intéressante dans des domaines comme la spectroscopie. La prochaine section portera sur les Vision Transformers. Nous y expliquerons leur fonctionnement, avant de les implémenter dans le chapitre 4, dédié à la méthodologie.

3.2 Fondements des Transformers

Les modèles des Transformers ont été introduits par Vaswani et al. [45], il s'agit d'une architecture qui repose exclusivement sur des mécanismes d'attention, supprimant l'usage des convolutions et des récurrences utilisées jusque là dans des modèles traditionnels comme les RNN et les LSTM. Grâce à l'attention multi-tête, les Transformers peuvent capturer des relations à longue distance dans les séquences et traiter les données en parallèle, améliorant ainsi l'efficacité et les performances, notamment pour des tâches comme la traduction automatique.

3.2.1 Principe de fonctionnement

Le modèle Transformer repose sur une architecture conçue pour traiter des séquences à l'aide de mécanismes d'attention, permettant de capturer efficacement les relations entre les éléments d'une séquence. Cette approche favorise un traitement parallèle des données tout en gérant de manière optimale les dépendances à longue portée, surpassant les modèles séquentiels traditionnels.

L'architecture illustrée par la Figure 3.2, met en lumière les principaux composants de ce modèle.

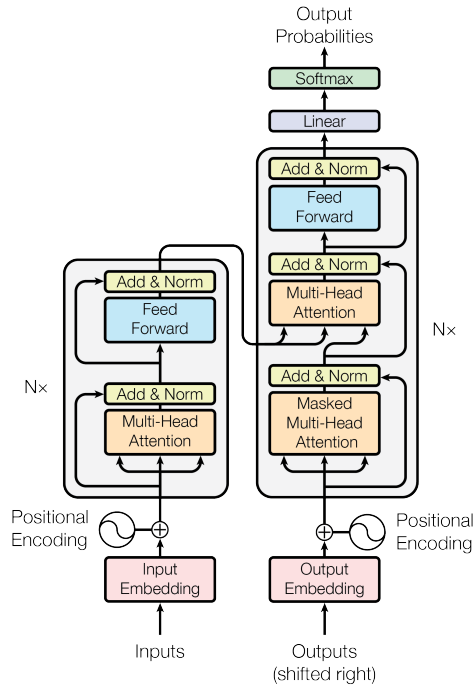


FIGURE 3.2 – Architecture typique des Transformers.

Source : Vaswani Ashish et al. [45].

Cette architecture des Transformers est composée de deux principales parties : l'encodeur et le décodeur.

3.2.2 Encodeur

C'est un composant clé dans l'architecture des Transformers. Il permet d'encoder les données d'entrée en une représentation compacte, permettant au modèle de comprendre et d'extraire les relations entre les éléments d'entrée[45].

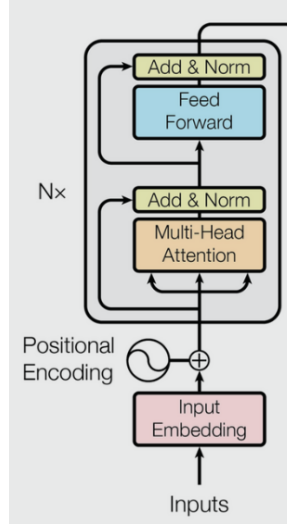


FIGURE 3.3 – Architecture typique d'un encodeur

Source : Vaswani Ashish et al. [45].

- **Input Embeddings**

Chaque mot ou élément de la séquence d'entrée est représenté par un vecteur appelé embedding. La séquence peut alors être vue comme une matrice X où chaque ligne est le vecteur d'embedding associé à un élément de la séquence d'entrée :

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (3.8)$$

$x_i \in \mathbb{R}^d$ est le vecteur d'embedding de dimension d pour le i -ième élément de la séquence.

- **Ajout encodages positionnels**

Le Transformer étant invariant à l'ordre des éléments d'entrée, il est néces-

saire d'ajouter des informations de position au moyen d'encodages positionnels, chaque position étant représentée par un code unique. :

$$Z_0 = X + PE \quad (3.9)$$

Où PE est la matrice des encodages positionnels, de même dimension que X .

- **Multi-Head Self-Attention**

L'auto-attention permet à chaque élément de la séquence de "s'attarder" sur les autres éléments. Pour chaque tête d'attention, on projette les entrées en trois espaces : les requêtes Q , les clés K , et les valeurs V .

Nous utilisons des projections linéaires :

$$Q = Z_0 W_Q, \quad K = Z_0 W_K, \quad V = Z_0 W_V \quad (3.10)$$

Où W_Q, W_K, W_V sont des matrices de poids pour les projections, de dimensions $\mathbb{R}^{d \times d_k}$, $\mathbb{R}^{d \times d_k}$, et $\mathbb{R}^{d \times d_v}$ respectivement.

Ensuite, l'attention est calculée en utilisant le produit scalaire entre les requêtes et les clés, suivi d'une normalisation via une fonction softmax :

$$\text{Attention}(Q, K, V) = \text{Softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (3.11)$$

Cela donne une pondération des valeurs en fonction de la similarité des

requêtes et des clés.

Pour un mécanisme multi-têtes, ces calculs sont effectués h fois (une pour chaque tête), puis les résultats sont concaténés et projetés à nouveau dans l'espace original via une matrice de poids W_O :

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O \quad (3.12)$$

- **Add & Norm**

Les résultats de l'attention multi-têtes sont additionnés avec l'entrée Z_0 via une connexion résiduelle, puis une normalisation est appliquée :

$$Z_1 = \text{LayerNorm}(Z_0 + \text{MultiHead}(Q, K, V)) \quad (3.13)$$

- **Feed Forward Network**

Chaque vecteur passe ensuite par un réseau de neurones feed-forward, généralement constitué de deux couches linéaires avec une activation non-linéaire entre les deux (comme ReLU) :

$$\text{FFN}(z) = \max(0, zW_1 + b_1)W_2 + b_2 \quad (3.14)$$

Où $W_1 \in \mathbb{R}^{d \times d_{ff}}$ et $W_2 \in \mathbb{R}^{d_{ff} \times d}$ sont les matrices de poids, et b_1, b_2 sont les biais. La dimension d_{ff} est généralement plus grande que d .

- **Add & Norm**

Une connexion résiduelle est à nouveau rajoutée et est suivie d'une norma-

lisation :

$$Z_2 = \text{LayerNorm}(Z_1 + \text{FFN}(Z_1)) \quad (3.15)$$

- **Sortie de l'encodeur**

Le vecteur Z_2 représente la sortie de l'encodeur pour chaque élément de la séquence. Il est ensuite transmis aux différentes couches suivantes de l'encodeur ou utilisé directement pour une tâche spécifique, telle que la classification ou la génération de texte.

- **Synthèse des opérations**

En résumé, chaque couche de l'encodeur peut être décrite par les étapes suivantes :

$$Z_{\text{out}} = \text{LayerNorm}(Z_{\text{in}} + \text{MultiHead}(Q, K, V)) \quad (3.16)$$

$$Z_{\text{out-final}} = \text{LayerNorm}(Z_{\text{out}} + \text{FFN}(Z_{\text{out}})) \quad (3.17)$$

Le décodeur prend en entrée les représentations produites par l'encodeur et les utilise pour générer une séquence de sortie. Il est utilisé dans des tâches telles que la traduction automatique ou la génération de texte, où l'objectif est de produire une séquence à partir d'une autre.

3.2.3 Décodeur

Le décodeur est la partie qui génère la séquence de sortie à partir des informations fournies par l'encodeur.

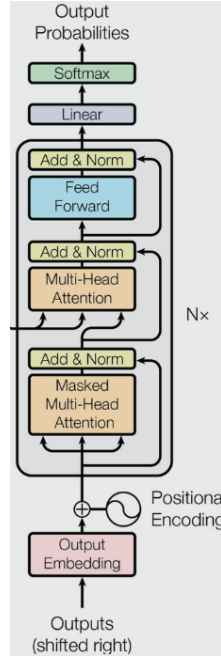


FIGURE 3.4 – Architecture typique d'un décodeur

Source : Vaswani Ashish et al. [45].

A chaque étape, le décodeur prend les tokens déjà générés ainsi que la sortie de l'encodeur pour générer un nouveau token de sortie. Selon les étapes suivantes :

- **Embeddings des tokens de sortie** : Les tokens déjà générés sont convertis en vecteurs d'embedding, et des embeddings de position sont ajoutés. Si la séquence de sortie partielle est donnée par y_1, y_2, \dots, y_{t-1} , alors :

$$\mathbf{Y}_{\text{emb}} = [\mathbf{E}(y_1), \mathbf{E}(y_2), \dots, \mathbf{E}(y_{t-1})] + \mathbf{P} \quad (3.18)$$

$$\mathbf{Y}_{\text{emb}} = [\mathbf{E}(y_1), \mathbf{E}(y_2), \dots, \mathbf{E}(y_{t-1})] + \mathbf{P} \quad (3.19)$$

Où :

$\mathbf{E}(y_i)$ est l'embedding du token y_i .

\mathbf{P} est l'embedding de position ajouté à chaque vecteur.

- **Self-attention masquée** : Le décodeur applique une **self-attention masquée** sur les tokens déjà générés, de sorte que chaque position ne peut voir que les positions précédentes et elle-même. La matrice de scores d'attention $\mathbf{A}_{\text{masked}}$ est calculée ainsi :

$$\mathbf{A}_{\text{masked}} = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}} + \mathbf{M} \right) \quad (3.20)$$

Où :

$$\mathbf{Q} = \mathbf{W}_Q \mathbf{Y}_{\text{emb}},$$

$$\mathbf{K} = \mathbf{W}_K \mathbf{Y}_{\text{emb}},$$

$$\mathbf{V} = \mathbf{W}_V \mathbf{Y}_{\text{emb}}.$$

La sortie de cette attention est obtenue par :

$$\mathbf{Z}_1 = \mathbf{A}_{\text{masked}} \mathbf{V} \quad (3.21)$$

- **Attention croisée avec la sortie de l'encodeur** : Par la suite, le décodeur applique une **attention croisée** avec la sortie de l'encodeur. Si la sortie de

l'encodeur est donnée par \mathbf{H}_{enc} , l'attention croisée est calculée ainsi :

$$\mathbf{A}_{\text{cross}} = \text{softmax} \left(\frac{\mathbf{Q}_d \mathbf{K}_e^\top}{\sqrt{d_k}} \right) \quad (3.22)$$

Où :

$$\mathbf{Q}_d = \mathbf{W}_Q \mathbf{Z}_1 \text{ (requêtes du décodeur),}$$

$$\mathbf{K}_e = \mathbf{W}_K \mathbf{H}_{\text{enc}},$$

$$\mathbf{V}_e = \mathbf{W}_V \mathbf{H}_{\text{enc}} \text{ (clés et valeurs de l'encodeur)}$$

.

La sortie de cette attention croisée est :

$$\mathbf{Z}_2 = \mathbf{A}_{\text{cross}} \mathbf{V}_e \quad (3.23)$$

- **Couches feed-forward** : Comme pour l'encodeur, la sortie du décodeur après l'attention croisée est passée par une couche feed-forward :

$$\mathbf{F}(\mathbf{Z}_2) = \text{ReLU}(\mathbf{Z}_2 \mathbf{W}_1 + b_1) \mathbf{W}_2 + b_2 \quad (3.24)$$

- **Génération de sortie** : À l'étape finale de la génération, le décodeur produit un **token de sortie** en appliquant une couche linéaire suivie d'une fonction softmax pour obtenir des probabilités sur le vocabulaire :

$$y_t = \operatorname{argmax}(\operatorname{softmax}(\mathbf{Z}_2 \mathbf{W}_{\text{out}} + b_{\text{out}})) \quad (3.25)$$

Où y_t est le token de sortie prédit à l'étape (t).

Le décodeur dans les Transformers est la partie du modèle qui génère une séquence de sortie à partir d'une séquence partiellement produite et des informations provenant de l'encodeur. À chaque étape, le décodeur utilise les tokens déjà générés, applique une auto-attention masquée et une attention croisée avec la sortie de l'encodeur pour intégrer les informations de la séquence d'entrée. Ensuite, les données passent par des couches feed-forward et produisent un nouveau token à chaque étape, jusqu'à générer la séquence complète.

Cette structure des Transformers a inspiré de nombreuses variantes, dont les Vision Transformers (ViT), qui appliquent ces concepts d'attention à des images découpées en patches pour traiter des données visuelles.

3.2.4 Les Vision Transformers (ViT)

Les Vision Transformers constituent une architecture de réseau neuronal profond qui adapte le mécanisme d'auto-attention des Transformers, initialement conçu pour le traitement du langage naturel, au domaine de la vision par ordinateur. Au lieu de traiter des séquences de mots, les ViT traitent des séquences de patches d'images.

Cette architecture a été proposée pour la première fois par Dosovitskiy Alexey et al. en 2021 [46]. La figure 3.5 illustre son principe de fonctionnement.

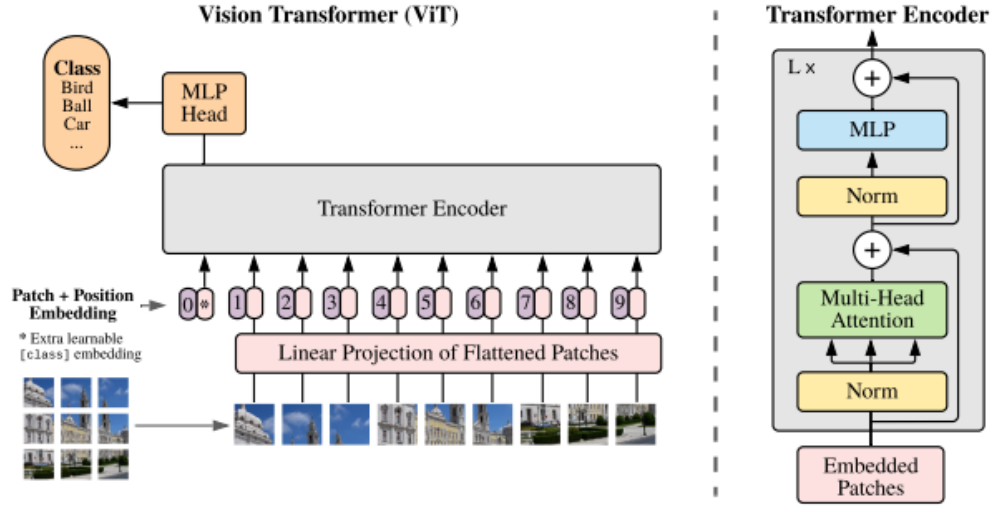


FIGURE 3.5 – Architecture typique des ViT

Source : Dosovitskiy Alexey et al. [46].

- **Division en Patches :**

L'image d'entrée $I \in \mathbb{R}^{H \times W \times C}$, où H est la hauteur, W est la largeur, et C est le nombre de canaux, est divisée en petits patches de taille $P \times P$. Chaque patch est transformé en un vecteur $p_i \in \mathbb{R}^{P^2 \cdot C}$.

L'image devient alors une séquence de patches :

$$X = [p_1, p_2, \dots, p_N], \quad \text{avec } N = \frac{H \cdot W}{P^2} \quad (3.26)$$

- **Projection Linéaire des Patches :**

Les patches sont projetés dans un espace de dimension D via une projection linéaire :

$$z_i = W_p \cdot p_i + b_p \quad (3.27)$$

Où $W_p \in \mathbb{R}^{D \times P^2 \cdot C}$ est une matrice de projection et b_p est un biais.

Ainsi, chaque patch projeté devient $z_i \in \mathbb{R}^D$.

- **Ajout des Embeddings de Position :**

Des embeddings de position $e_{\text{pos}} \in \mathbb{R}^{N \times D}$ sont ajoutés pour conserver l'ordre

des patches :

$$z'_i = z_i + e_{\text{pos}}(i) \quad (3.28)$$

- **Encodage par Transformer :**

Les patches projetés sont ensuite envoyés dans un encodeur Transformer. À chaque couche, deux opérations principales sont effectuées :

- **Attention Multi-têtes :**

Chaque vecteur z'_i est transformé en vecteurs de requêtes Q , de clés K , et de valeurs V :

$$Q = W_Q \cdot z'_i, \quad K = W_K \cdot z'_i, \quad V = W_V \cdot z'_i \quad (3.29)$$

où $W_Q, W_K, W_V \in \mathbb{R}^{D \times D}$ sont des matrices de poids.

L'attention entre les patches est calculée par :

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{D}} \right) V \quad (3.30)$$

- **MLP :**

Les sorties du mécanisme d'attention sont passées à travers un MLP (Multi-Layer Perceptron) :

$$z''_i = \text{MLP}(\text{Attention}(Q, K, V)) \quad (3.31)$$

Ces étapes sont répétées L fois à travers L couches du *Transformer encodeur*.

- **Classification :**

Un token spécial, appelé *class token* z_0 , est ajouté au début de la séquence de patches. Ce token capture l'information globale de l'image.

Après les couches du *Transformer encoder*, le *class token* est passé à un perceptron multicouche (MLP) pour prédire la classe :

$$\hat{y} = \text{MLP}(z_0) \quad (3.32)$$

Dans ce mémoire, nous testons différentes architectures de Transformers pour l'estimation à partir des données spectrales.

3.3 Quelques Métriques d'évaluation des modèles

Dans l'ensemble des modèles illustrés précédemment, plusieurs métriques sont utilisées pour évaluer leurs performances. Voici une présentation des principales métriques avec leurs formules.

- **Erreur quadratique moyenne (MSE - Mean Squared Error) :**

La MSE mesure la moyenne des carrés des écarts entre les valeurs prédites et les valeurs réelles, utilisée principalement pour les tâches de régression. L'équation est la suivante :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.33)$$

Où y_i représente la valeur réelle, \hat{y}_i la valeur prédite et n le nombre d'exemples.

- **La racine carrée de l'erreur quadratique moyenne (RMSE - Root Mean Squared Error) :**

La RMSE est la racine carrée de la MSE, permettant une interprétation des erreurs dans la même unité que les valeurs cibles. L'équation est donnée par :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.34)$$

- **Coefficient de détermination (R^2)** : Le coefficient R^2 mesure la proportion de la variance des données expliquée par le modèle. Il s'exprime par l'équation suivante :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.35)$$

Où \bar{y} est la moyenne des valeurs réelles. Un R^2 proche de 1 indique une bonne capacité prédictive du modèle.

- **Précision (Accuracy)** :

Cette métrique est utilisée principalement pour les tâches de classification et mesure le pourcentage de prédictions correctes. Son équation est donnée par :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.36)$$

Où TP est le nombre de vrais positifs, TN le nombre de vrais négatifs, FP les faux positifs, et FN les faux négatifs.

- **F1-score** :

Le F1-score est la moyenne harmonique de la précision et du rappel³, particulièrement utile lorsque les classes sont déséquilibrées. Son équation est donnée par :

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.37)$$

3. le rappel est la capacité d'identifier toutes les instances d'une classe particulière.

$$\text{avec : Precision} = \frac{TP}{TP+FP} \text{ et Recall} = \frac{TP}{TP+FN}.$$

Ces métriques d'évaluation présentent des avantages et des limites en fonction des types de modèles et de données utilisés. Par exemple, la MSE et le R^2 sont généralement mieux adaptés aux modèles de régression, tandis que la précision et le F1-score sont couramment employés pour les modèles de classification.

Dans ce travail, étant donné que l'objectif est de faire des estimations à l'aide des données spectrales, le problème est formulé comme un problème de régression. Ainsi, la métrique choisie pour évaluer les performances du modèle présenté dans le Chapitre 5, sera le RMSE (Root Mean Square Error). Ce choix se justifie principalement par son utilisation fréquente pour la comparaison avec des modèles existants. De plus, Draxler R. et al. [47] ont démontré que cette métrique offre des performances satisfaisantes par rapport à d'autres indicateurs, renforçant ainsi sa pertinence pour ce type de problématique.

3.4 Limites des modèles

Dans le domaine spectral, divers modèles analytiques sont couramment utilisés, chacun avec ses propres caractéristiques et avantages. Cependant, il est essentiel de reconnaître que ces modèles présentent également des limites significatives qui peuvent influencer leur efficacité. Pour mieux comprendre ces défis, nous avons élaboré un tableau comparatif qui met en lumière les lacunes des modèles de régression des moindres carrés, de l'apprentissage automatique et des réseaux de neurones convolutifs. Le tableau 3.1 révèle les faiblesses inhérentes à chaque approche, mais également les métriques les plus fréquemment utilisées pour évaluer leur performance, offrant une vue d'ensemble précieuse pour orienter nos choix méthodologiques dans la recherche.

TABLE 3.1 – Comparaison des modèles de détection de molécules

Modèle	Inconvénients	Métriques les plus utilisées
Régression des moindres carrés	Sensibilité aux valeurs aberrantes, Capacité limitée à capturer des relations non linéaires, nécessite une sélection préalable des variables.[48].	Erreur quadratique moyenne (MSE), Coefficient de détermination (R^2).
Apprentissage Automatique(SVM)	Difficulté à gérer des ensembles de données très grands, Risque de surapprentissage sans régularisation[49].	Précision, Rappel (Recall), F1-score, AUC-ROC.
Réseau de neurones convolutifs	Nécessité de grandes quantités de données, Coût computationnel élevé pour l'entraînement, Problème de contexte[50].	Précision, Erreur de classification, MSE pour les tâches de régression.

Compte tenu des limitations des modèles traditionnels, tels que la régression des moindres carrés et les réseaux de neurones convolutifs, notamment leur difficulté à capturer des relations complexes et à longue portée dans les données spectrales, nous avons opté pour les modèles basés sur les architectures des Transformers. Ces derniers, grâce à leurs mécanismes d'auto-attention sont particulièrement adaptés à l'analyse de données séquentielles et permettent de modéliser des dépendances complexes entre différentes régions du spectre. Le chapitre suivant sera consacré à une présentation détaillée de ces modèles, en insistant sur leur fonctionnement et leurs avantages spécifiques pour notre application.

3.5 Conclusion

Dans ce chapitre, nous avons examiné les principes fondamentaux des différents modèles d'apprentissage exploités dans ce travail. Nous avons détaillé le fonctionne-

ment des réseaux de neurones convolutifs, ainsi que l'approche basée sur les architectures des Transformers. En plus de leur mécanisme de fonctionnement, nous avons mis en évidence les métriques couramment utilisées pour évaluer leurs performances, ainsi que les limitations spécifiques associées à chaque modèle. Par ailleurs, nous avons souligné les avantages des Transformers, qui seront employés dans la suite de ce travail.

Le prochain chapitre sera consacré à notre méthodologie. Nous y décrirons l'implémentation concrète du modèle des Transformers, en expliquant les étapes clés de son entraînement, de son évaluation et des stratégies d'optimisation mises en œuvre pour atteindre des résultats pertinents.

Chapitre 4

Méthodologie

Ce chapitre expose la méthodologie adoptée pour la mise en œuvre et l'évaluation des modèles étudiés précédemment, en précisant les choix techniques et les étapes expérimentales.

Nos expérimentations ont été menées sur un ordinateur équipé d'une carte graphique NVIDIA de type GeForce RTX 3050 avec une version 11.6 de CUDA, d'une RAM de 16 GB et de 12 CPUs de 2.5 GHz. Les scripts utilisent Python 3.9 et la bibliothèque PyTorch : [51].

Nous allons dans un premier temps présenter la collecte et la préparation des données, une étape essentielle pour garantir un entraînement efficace du modèle. Nous présenterons ensuite les différentes architectures de Transformers, les paramètres d'entraînement retenus, ainsi que les stratégies d'optimisation adoptées pour maximiser les performances.

4.1 Collecte et Préparation des Données

Dans cette section, nous présentons les jeux de données utilisés pour notre modèle. Le premier jeu de données Challenge2018 provient du défi du congrès Chemometrics2018 de Paris en 2018, ces données sont disponibles dans JchemoData[52]. Ce challenge visait à estimer les concentrations en protéines des fourrages et aliments pour les animaux. Le second jeu de données est un ensemble de données issu des sciences forensiques, dans lequel la cible à estimer correspond à la datation relative des taches de sang.

4.1.1 Jeux de données Challenge

Ce jeu de données provient du challenge organisé en 2018 lors de la conférence ECML-PKDD. Il est composé d'un fichier de données de spectre avec des observations identifiées et d'un fichier des concentrations pour chacune des observations.

Le fichier de données contenant les spectres infrarouges se compose de 4075 observations et de 680 caractéristiques. Chaque ligne représente un échantillon, et chaque colonne correspond à une longueur d'onde spécifique. Les valeurs indiquent l'absorption de la lumière infrarouge par les échantillons à ces longueurs d'onde. Les échantillons sont des produits laitiers.

Le fichier contenant les concentrations regroupe pour chaque échantillon de lait, les mesures de référence obtenues en laboratoire (matière grasse, protéines, etc.). Ces mesures servent de valeurs cibles pour entraîner un modèle à estimer ces concentrations chimiques à partir des spectres infrarouges fournis dans l'autre fichier.

Les différentes colonnes de ce fichier sont :

- **typ** (*chaîne de caractères*) : identifie le type d'échantillon.
- **label** nom de l'échantillon biologique.
- **conc** (*valeur numérique réelle*) : indique la concentration des produits laitiers, qui constitue la variable cible à prédire.
- **test** partitionne les données en indiquant si l'échantillon fait partie des données d'entraînement ou de test.

Dans cette étude, nous tenterons d'estimer la concentration (conc) des produits laitiers(notre cible) et la colonne "test" nous permet de diviser le jeu de données en ensembles d'entraînement et de validation. Ce jeu de données contient 4075 observations et 682 variables.

- **Analyse descriptive Challenge**

Le jeu de données Challenge avait déjà été préalablement standardisé, avec une moyenne égale à 0 et un écart type égal à 1. Il possède une propriété "conc" qui représente la caractéristiques cible à estimer, ainsi qu'une propriété "test" qui permet de distinguer les observations appartenant à l'ensemble d'entraînement de celles de l'ensemble de validation. Les autres caractéristiques correspondent aux réflectances.

4.1.2 Jeux de données de criminalistique

Ce jeu de données provenant du laboratoire de criminalistique de l'université du Québec à Trois-Rivières contient des réflectances proche infrarouge de plusieurs échantillons de sang exposés à l'air libre, sous différentes conditions de température, d'humidité et de rayonnement, sur divers substrats et à différents stades de dégradation indiquant leurs durées d'exposition.

- **Caractéristiques spectrales :**

- **Spectres** : Données spectrales représentées par des valeurs numériques réelles, correspondant aux mesures d'absorption ou de réflectance à différentes longueurs d'onde. Elles contiennent des informations essentielles sur les propriétés chimiques et physiques de l'échantillon.
- **Substrat** : Variable catégorielle décrivant le type de surface (béton, coton, céramique, plancher, tapis) sur laquelle l'échantillon est appliqué.
- **Rayonnement** : Variable binaire indiquant la présence (1) ou l'absence (0) de rayonnement durant l'exposition de l'échantillon.
- **Humidité** : Taux d'humidité ambiant durant l'exposition de l'échantillon, pouvant affecter les mesures et accélérer les phénomènes de dégradation de l'échantillon.
- **Variable cible (target)** :
 - **Nombre d'heures** : Variable cible représentant la datation relative des taches de sang.

Ce jeu de donnée est constitué de 1997 observations et de 130 caractéristiques.

- **Analyse descriptive du jeu de données**

Parmi les variables principales, on trouve le type de substrat, dont la répartition est la suivante : Céramique (409 échantillons), Coton (400), Tapis (400), Béton (400) et Plancher (388). La température ne prend que trois valeurs distinctes (15, 20 et 25 °C), se répartissant en deux catégories : 10–20 °C pour 1447 échantillons et 20–30 °C pour 550 échantillons. L'humidité varie entre 25 % et 70 %, avec trois classes observées : 20–40 % (550 échantillons), 40–60 % (537) et 60–80 % (910). Le rayonnement, codé de façon binaire (0/1), se situe uniquement dans la catégorie 0–200 W/m² pour l'ensemble des échantillons. Enfin, la variable Heures présente des valeurs allant de 0,5 h à 72 h, avec une moyenne de 16,95 h.

Le graphique figure4.1 presente la répartition des observations selon, le rayonnement, nous constatons que dans 75 % des cas, ces observations ont été faites en présence de rayonnement.

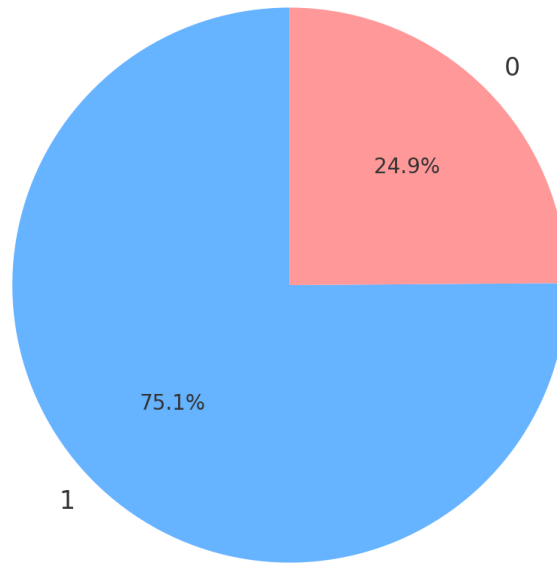


FIGURE 4.1 – *Repartition des échantillons selon le rayonnement*

La figure 4.2 illustre la répartition des observations selon le type de substrat et la présence (R :1) ou l'absence (R :0) de rayonnement. On constate que les substrats avec rayonnement (R :1) sont majoritaires, les plus représentés étant la céramique (15,5 %), le coton (15 %), le tapis (15 %), le béton (15 %) et le plancher (14,5 %). Les substrats sans rayonnement (R :0) représentent une plus petite part, chaque catégorie tournant autour de 5 % : céramique (5 %), béton (5 %), tapis (5 %), coton (5 %) et plancher (4,9 %).

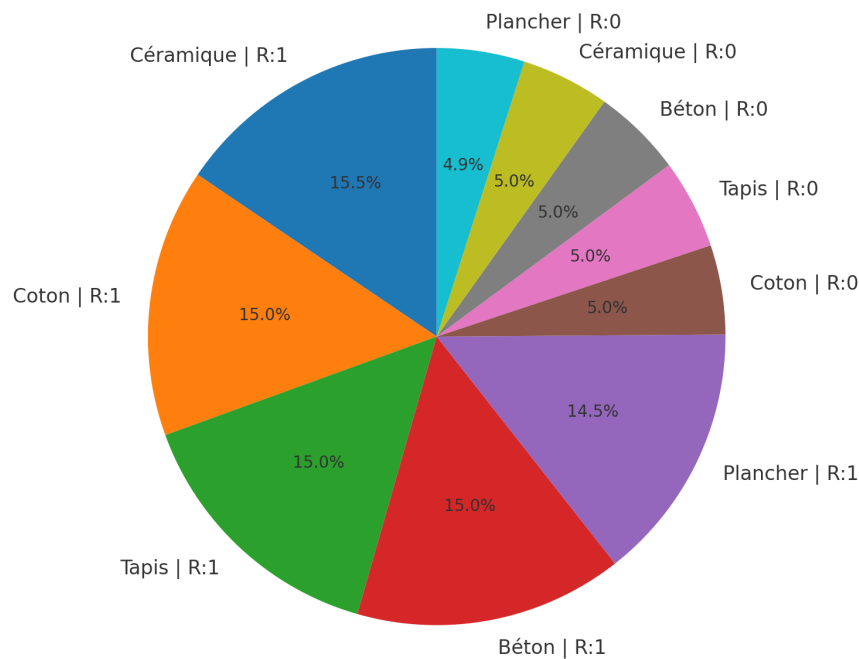


FIGURE 4.2 – Répartition des échantillons selon le rayonnement et le substrat

- **Analyse bivariée des variables en fonctions de la durée :**

L'analyse bivariée est une méthode d'analyse statistique qui examine la relation entre deux variables. Contrairement à l'analyse univariée, qui observe une seule variable à la fois, l'analyse bivariée vise à identifier et à comprendre les liens potentiels entre deux variables, qu'elles soient quantitatives ou qualitatives. Notre objectif est d'évaluer comment les variables contextuelles (substrat, rayonnement, humidité, température) se répartissent en fonction des différentes classes de durée d'exposition, qui constituent notre variable cible. Concrètement, il s'agit de vérifier si l'on dispose d'un nombre comparable d'échantillons pour chaque durée selon les conditions expérimentales : par exemple, en présence ou en absence de rayonnement, ou encore selon les différents niveaux de température, d'humidité et de type de substrat.

La figure 4.3 met en lumière la répartition des durées d'exposition en fonction des différentes variables, avec une attention particulière portée à la présence du substrat

comme facteur dans l'analyse. Lorsque l'on examine la courbe correspondant aux substrats, on constate des écarts significatifs dans la moyenne et la dispersion des durées selon le type de surface (notamment entre plancher, céramique et coton), avec des variabilités inégales. Cette hétérogénéité montre que la durée d'exposition n'est pas répartie de manière aléatoire entre les substrats : certains, comme le coton ou le tapis, sont associés à des durées très similaires. Cela suggère une dépendance entre substrat et durée d'exposition. En revanche, les autres variables comme l'humidité, la température ou le rayonnement montrent également des variations dans les datations relatives des taches de sang, mais ces variations semblent moins systématiquement liées à des catégories fixes.

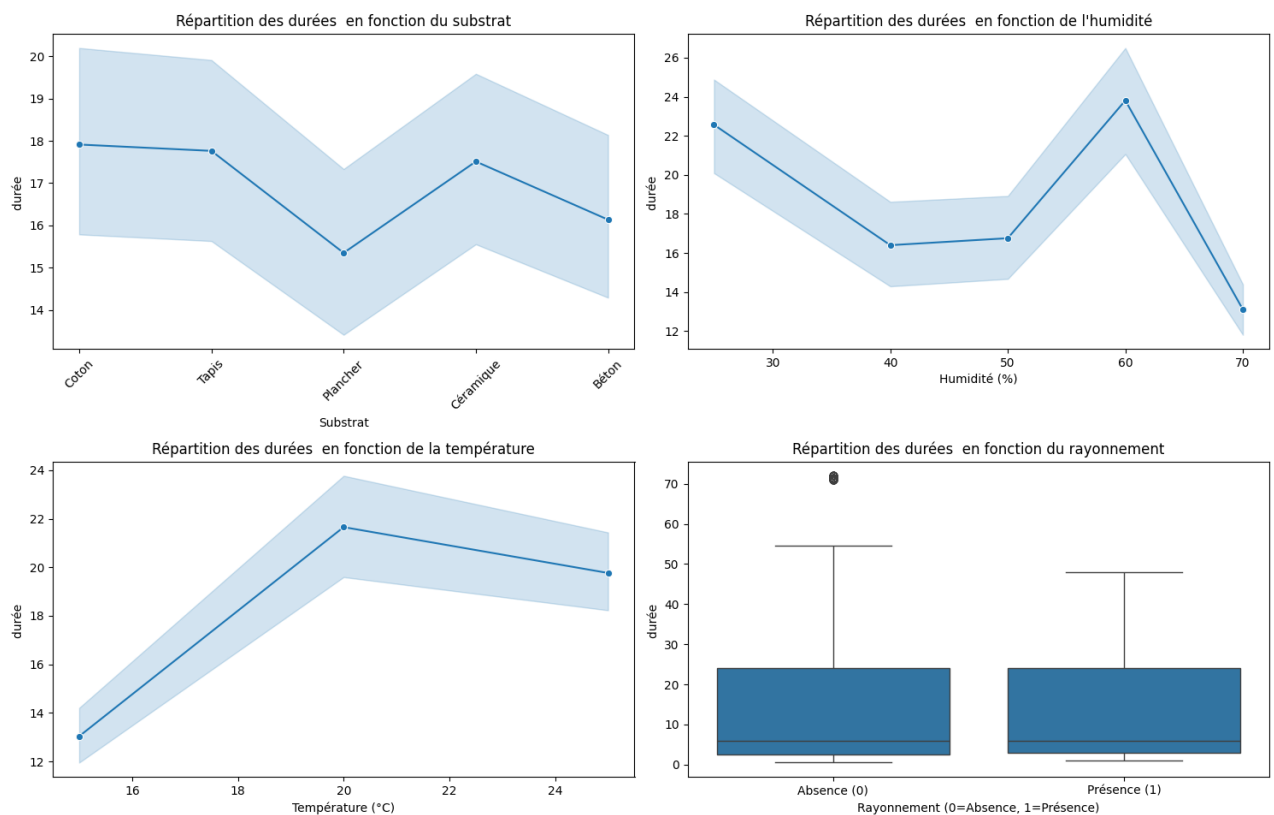


FIGURE 4.3 – Analyse bivariée des données contextuelles

En analysant la répartition des durées en fonction du rayonnement dans la figure 4.3, les distributions observées pour les conditions de présence et d'absence de rayonnement apparaissent très proches. Le diagramme en boîte indique des médianes

similaires (environ 5–6 heures), ainsi qu’une variabilité comparable s’étendant de 0 à 25 heures. La seule différence est la présence d’une valeur aberrante d’environ 70 heures dans la condition sans rayonnement (représentée par le point noir). Contrairement aux autres facteurs environnementaux comme la température et l’humidité qui montrent des tendances claires, le rayonnement semble être un facteur peu déterminant dans le processus de dégradation du sang dans cette étude.

4.1.3 Prétraitement des données

Le prétraitement des données spectroscopiques est une étape essentielle pour garantir la qualité et la fiabilité des analyses ultérieures. En effet, les spectres bruts contiennent souvent des bruits dus aux conditions expérimentales, au bruit de fond et aux particularités des instruments de mesure. Ces éléments peuvent masquer et fausser les informations pertinentes. Le prétraitement vise donc à réduire ces artefacts et à rendre les données plus homogènes et exploitables. Cette étape permet d’obtenir des résultats plus précis, en assurant que les variations observées sont bien représentatives des caractéristiques de l’échantillon étudié et non d’effets externes.

Dans notre cas, nous avons effectué deux principaux prétraitements sur le jeu de données issu des sciences forensiques. Le premier concerne les données catégorielles : nous avons utilisé `Label Encoder` de Scikit-learn¹, qui consiste à transformer les colonnes catégorielles en colonnes numériques en attribuant un entier unique à chaque catégorie.

L’encodage a été effectué de la manière suivante au niveau de la colonne `substrat`, comme le montre la figure 4.4.

1. <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>

Le choix de l’encodage par labels pour la colonne substrat se justifie par le fait que, dans nos expérimentations, cette approche a fourni des résultats d’estimation similaires à ceux obtenus avec un encodage one-hot. Les Transformers apprennent efficacement les représentations pertinentes quel que soit le format de codage, dès lors que les informations des tokens différenciant les substrats sont préservées. En effet, chaque entier issu du label encoding est uniquement utilisé comme indice d’accès à un vecteur d’embedding, ce qui élimine toute interprétation ordinale et permet au modèle de construire lui-même les représentations appropriées.

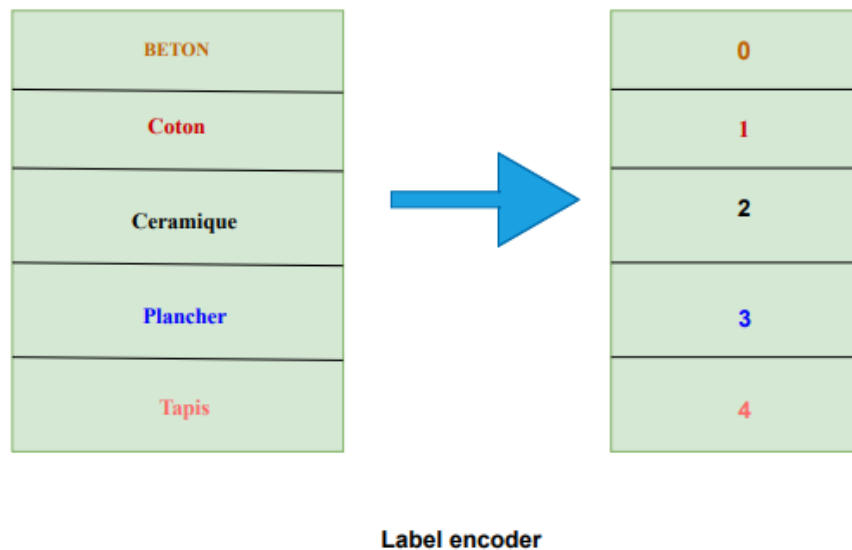


FIGURE 4.4 – Encodage du substrat

La colonne rayonnement ayant juste deux catégories a subi un simple encodage binaire.

Le second prétraitement est la standardisation des données. La standardisation consiste à transformer les valeurs d’une variable pour qu’elle soit de moyenne 0 et d’écart-type 1. Cette transformation est particulièrement utile dans les contextes où les variables ont des échelles différentes, car elle les ramène toutes à une échelle comparable.

$$z = \frac{x - \mu}{\sigma} \quad (4.1)$$

où :

- x : la valeur initiale de la variable,
- μ : la moyenne de la variable,
- σ : l'écart-type de la variable.

Sur le jeu de données du challenge, la standardisation avait déjà été effectuée.

Nous allons dans la section suivante, introduire l'architecture de notre modèle.

4.2 Architectures des modèles

Dans un premier temps, nous avons étudié trois architectures des Transformers : les Transformers utilisant uniquement la partie encodeur, les Vision Transformers (ViT) également basés sur un encodeur, et enfin l'architecture complète des Transformers intégrant à la fois un encodeur et un décodeur pour permettre le mécanisme d'attention croisée (cross-attention).

A. Transformer avec l'encodeur

Notre première architecture utilise seulement la partie encodeur. On veut examiner si l'utilisation du mécanisme d'auto-attention permet de capturer efficacement les relations complexes entre les longueurs d'onde dans les données spectrales et si cela induit à de meilleures performances en termes de précision et de généralisation par rapport aux méthodes classiques, telles que la régression PLSR, les CNN et autres architectures des Transformers.

Cette architecture est illustrée comme suit :

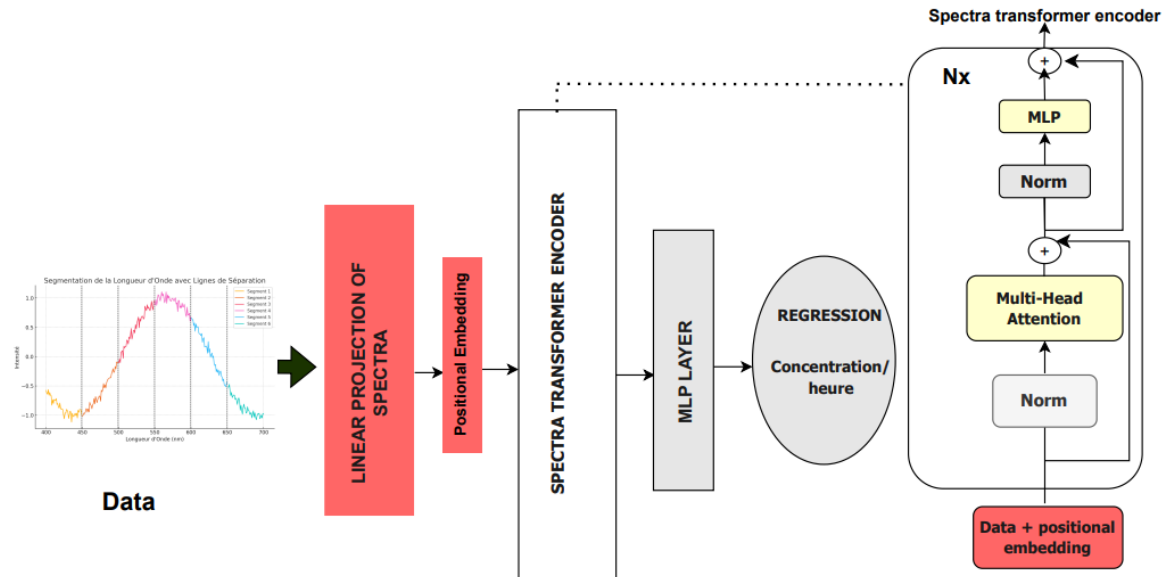


FIGURE 4.5 – *Architecture Encoder*

1. Données d'entrée (Data) : L'entrée du modèle est constituée de spectres, qui sont des séries de valeurs représentant une distribution spectrale sur différentes longueurs d'onde. Ces données sont traitées sous forme séquentielle.

2. Projection linéaire des spectres (Linear Projection of Spectra) : Une transformation linéaire est appliquée pour ajuster la représentation des spectres à l'encodeur du Transformer. Cette projection permet de transformer les données d'origine dans un espace de dimension adaptée au modèle.

3. Ajout d'un encodage positionnel (Positional Embedding) : À cause de la nature séquentielle des spectres, un encodage positionnel est ajouté. Cela permet de préserver l'ordre et la structure de la séquence, ce qui aide le modèle à capturer les dépendances entre les longueurs d'onde.

4. L'encodeur : Les données transformées passent ensuite dans l'encodeur, qui traite la séquence à travers plusieurs couches successives :

- **Normalisation (Norm)** : Stabilise l'apprentissage et améliore la convergence.
- **Multi-Head Attention** : Capture les relations complexes entre les différentes longueurs d'onde.
- **MLP (Multi-Layer Perceptron)** : Applique des transformations non linéaires pour extraire des caractéristiques plus abstraites.

Ces opérations sont répétées trois fois, car le modèle utilise uniquement trois couches.

5. Passage par la couche MLP : Cette dernière couche affine les représentations extraites par l'encodeur afin de les adapter à la régression.

6. Régression (Regression - Concentration/Heure) La sortie du modèle permet d'estimer :

- Une concentration pour le jeu de données challenge.
- Une durée d'exposition des taches de sang pour le jeu de données issu des sciences forensiques.

B. Architecture des ViT

La figure 4.6 présente l'architecture des ViT avec l'encodeur.

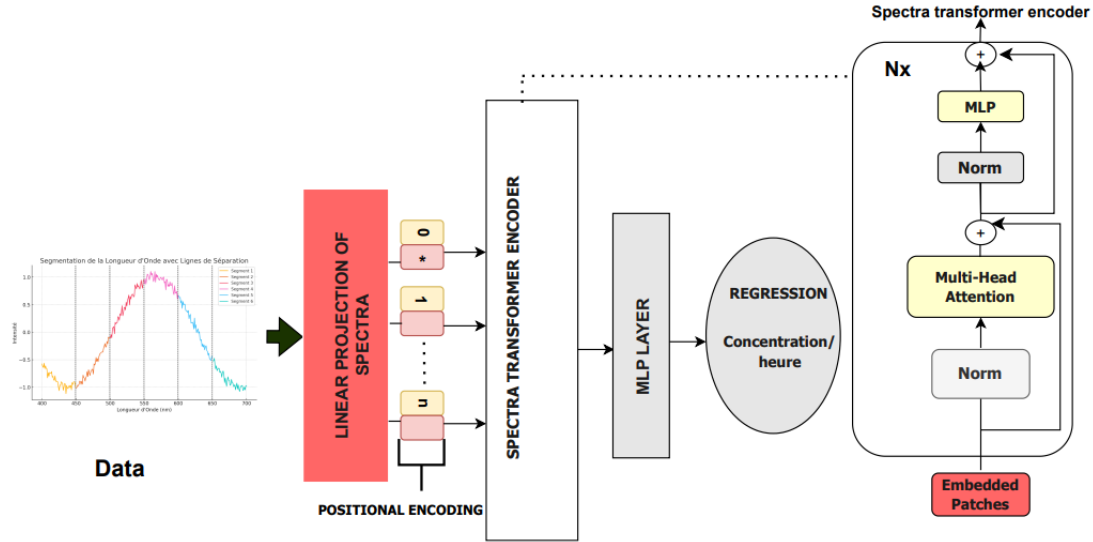


FIGURE 4.6 – Architecture ViT Encodeur

L'architecture de ce modèle est assez similaire à la précédente à la seule différence que les données sont divisées en patches et un encodage positionnel est associé à chaque patch pour conserver les informations de position de la séquence et ce sont les patches qui sont passés dans notre modèle.

Dans cette architecture, nous souhaitons expérimenter si le regroupement des données en patches permet d'associer des réflectances qui, ensemble, pourraient avoir plus d'importance qu'individuellement. Cependant, ce regroupement pourrait également introduire du bruit en combinant des réflectances sans liens significatifs entre elles, ce qui pourrait nuire aux performances du modèle.

C. Architecture Complète des Transformers

La dernière architecture que nous avons testée est une version complète des Transformers qui utilise en plus de l'auto-attention, l'attention croisée.

Cette architecture est inspiré de l'architecture de Zhao Chongjia et al.[53].

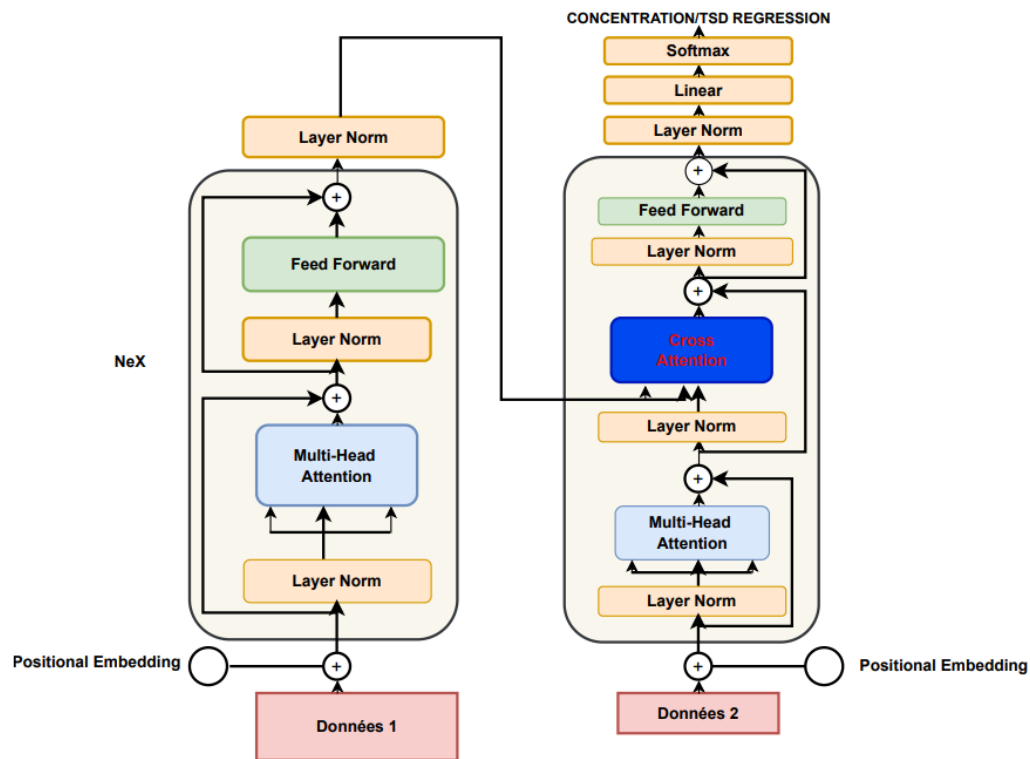


FIGURE 4.7 – L'attention croisée

Dans cette architecture, nous visons à étudier l'impact de l'utilisation complète des Transformers, en exploitant notamment le mécanisme d'attention croisée entre l'encodeur et le décodeur. L'une des approches envisagées consiste à faire transiter les "Données 1" (par exemple, les données spectrales) par l'encodeur, tandis que les "Données 2" (par exemple, les données environnementales) seraient traitées par le décodeur. Une autre expérimentation pertinente serait d'examiner l'effet du croisement

entre ces deux ensembles, notamment l'intégration des informations environnementales (substrat, humidité, température et rayonnement) avec les données spectrales. Cette démarche pourrait s'avérer particulièrement intéressante dans le contexte des données issues des sciences forensiques.

Luo Qiwu et al. [54] mettent en exergue l'avantage de l'attention croisée sur les connexions résiduelles classiques qui sont habituellement utilisées dans les Transformers.

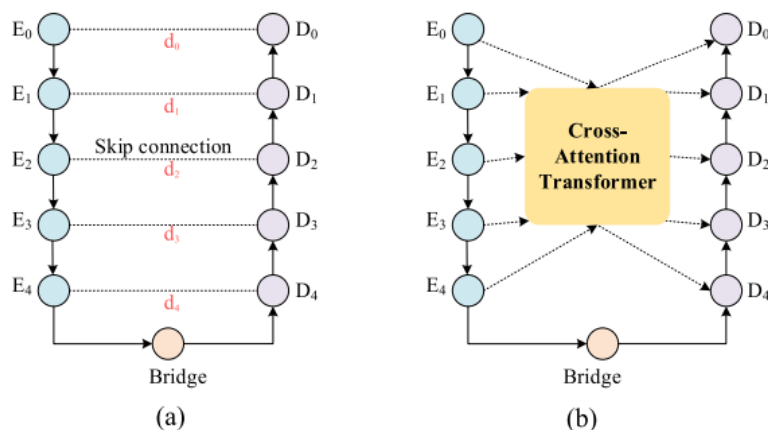


FIGURE 4.8 – *Comparaison connexion résiduelle et cross attention*

Source : Luo Qiwu et al. [54].

La figure (a) représente la connexion résiduelle classique. Dans cette configuration, chaque couche de l'encodeur (E_1, E_2, \dots) est directement connectée à la couche correspondante du décodeur (D_1, D_2, \dots) via une connexion résiduelle (skip connection). Cette méthode permet de transférer les caractéristiques extraites à chaque niveau de l'encodeur vers les couches correspondantes du décodeur sans modification significative.

La principale limite de l'approche avec les connexions résiduelles est qu'elles n'établissent pas de relation complexe entre les caractéristiques multi-échelles.

La figure (b) représente l'attention croisée dans laquelle, les couches de l'enco-

deur et du décodeur sont connectées via un module d'attention croisée. Dans cette configuration les caractéristiques de l'encodeur (E_1, E_2, \dots) sont utilisées comme mémoire globales (keys et values), les caractéristiques du décodeur (D_1, D_2, \dots) servent de requêtes (Queries). Le module d'attention croisée apprend à pondérer dynamiquement les relations entre les caractéristiques de l'encodeur et les caractéristiques du décodeur à chaque niveau. Les principaux avantages de cette approche :

- Facilite une interaction approfondie entre les niveaux de l'encodeur et du décodeur, optimisant ainsi le flux d'informations.
- Renforce la précision et la clarté des cartes de saillance, même dans des scénarios complexes.

Ces avantages de l'attention croisée motivent notre choix d'expérimenter cette approche.

En plus des Transformers, nous avons également exploré l'architecture des réseaux de neurones convolutifs (CNN) dans le cadre de notre recherche. Bien que les CNN présentent certaines limitations, notamment en ce qui concerne la prise en compte du contexte à long terme, leur capacité à capturer efficacement les relations locales dans les données en font une approche pertinente dans certaines situations.

Une expérimentation de cette approche consiste, par exemple, à faire varier la taille du noyau de convolution et la valeur du rembourrage afin d'évaluer leur impact sur l'extraction des caractéristiques spectrales. L'objectif est de comparer les performances des CNN à celles des différentes architectures de Transformers. En explorant ces différentes approches, nous visons à identifier les avantages et les limites de chacune d'elles dans le traitement des données de réflectance.

Notre architecture de réseau de neurones convolutifs est similaire à celle présentée dans la Figure 3.1, à la seule différence que la sortie de notre modèle est une régression.

4.3 Méthodes d'apprentissage

Dans cette section, nous nous appuyons principalement sur les travaux de Loshchilov Ilya et al. [55] pour expliquer les méthodes d'apprentissage de nos modèles. Nous y abordons les algorithmes les plus couramment utilisés, notamment l'algorithme de descente de gradient et l'optimiseur AdamW.

4.3.1 Algorithme de descente de gradient

Notre modèle utilise l'algorithme de descente de gradient pour optimiser les paramètres θ en minimisant la fonction de coût $J(\theta)$. Cette méthode itérative ajuste les paramètres du modèle afin de réduire l'erreur entre les prédictions et les valeurs réelles.

Principe mathématique :

Soit $J(\theta)$ une fonction différentiable que l'on souhaite minimiser, où θ représente le vecteur des paramètres. Le gradient de J en θ , noté $\nabla J(\theta)$, indique la direction de la plus forte augmentation de J . Pour minimiser J , il convient donc de se déplacer dans la direction opposée au gradient.

Formule de mise à jour :

À chaque itération t , les paramètres sont mis à jour selon la formule :

$$\theta_{t+1} = \theta_t - \eta \nabla J(\theta_t) \tag{4.2}$$

Où :

- θ est le vecteur des paramètres à l'itération t ;
- η est le taux d'apprentissage, un hyperparamètre déterminant la taille du pas de mise à jour, le taux d'apprentissage utilisé dans notre modèle est de 0.00001 ;
- $\nabla J(\theta)$ est le gradient de J évalué en θ .

Étapes de l'algorithme :

1. **Initialisation** : Choisir un point de départ θ_0 et fixer le taux d'apprentissage η .
2. **Calcul du gradient** : Évaluer $\nabla J(\theta_t)$.
3. **Mise à jour des paramètres** : Appliquer la formule de mise à jour pour obtenir θ_{t+1} .
4. **Convergence** : Répéter les étapes 2 et 3 jusqu'à ce que $\|\nabla J(\theta_t)\|$ soit inférieur à un seuil prédéfini, indiquant que l'on a atteint un minimum local ou global.

4.3.2 Optimiseur AdamW

En complément de la descente de gradient classique, nous utilisons l'optimiseur AdamW pour améliorer la convergence et la généralisation du modèle. AdamW est une variante de l'algorithme Adam qui décorrèle la régularisation par décroissance du poids des mises à jour des gradients, ce qui conduit à de meilleures performances sur divers ensembles de données, il est souvent choisi dans les modèles d'apprentissage profond, car il combine la rapidité et la stabilité d'Adam avec une régularisation plus précise [55], ce qui se traduit généralement par des performances accrues dans les modèles.

4.3.3 Fonction de Perte

Dans le contexte de la régression, la fonction de perte utilisée est la racine carrée de l'erreur quadratique moyen, définie comme suit :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.3)$$

Où y_i représente les valeurs réelles et \hat{y}_i les prédictions du modèle. Cette fonction mesure la racine carrée de l'écart quadratique moyen entre les prédictions et les valeurs réelles, fournissant une mesure intuitive de la performance du modèle.

4.3.4 Partition des jeux de données

Nos jeux de données sont divisés en deux parties :

- **Ensemble d'entraînement** : Constitue 80 % des données totales et est utilisé pour ajuster les paramètres du modèle.
- **Ensemble de test** : Constitue 20 % des données totales.

Le jeu de données Challenge inclut déjà une colonne spécifiquement dédiée à la séparation des données d'entraînement et de test, cette division est estimée à une division 80-20. Cette conception vise à garantir une comparaison cohérente des résultats entre tous les participants utilisant ces données pour entraîner et évaluer leurs modèles.

4.4 Conclusion

Ce chapitre a été consacré à la présentation de la méthodologie adoptée dans le cadre de notre étude. Nous avons commencé par introduire les deux jeux de données utilisés, Challenge et celui issu des sciences forensique(Criminalistique), en effectuant une analyse descriptive pour mieux comprendre leurs caractéristiques. Ensuite, nous avons détaillé l'ensemble des architectures de modèles utilisées, en expliquant leur fonctionnement et leurs spécificités. Enfin, nous avons présenté la méthode d'apprentissage adoptée ainsi que la stratégie de division des jeux de données. Ce chapitre a permis de poser les bases méthodologiques nécessaires pour mieux interpréter les résultats des expérimentations, qui seront abordés dans le chapitre suivant.

Chapitre 5

Résultats et Discussions

5.1 Estimation des concentrations et des durées

Dans toutes nos expérimentations, lorsque nous ne précisons pas les paramètres, nous utiliserons un taux d'apprentissage de 10^{-4} , 2500 itérations, une taille de mini-lots de 32. Pour les jeux de données, lorsqu'on ne précise pas une transformation spécifique cela suppose que les jeux de données n'ont subi aucune transformation. Il convient de noter que, dans le cadre de nos expérimentations avec des architectures complètes de Transformers (comprenant un encodeur et un décodeur), le nombre de têtes d'attention est identique pour les deux modules, sauf indication contraire.

- **Expérimentation :1**

La première expérimentation a été réalisée avec l'architecture des réseaux de neurones convolutifs. Nous avons fait varier la taille du noyau ainsi que le rembourrage pour nos expérimentations. Ces différentes variations sont présentées par le tableau 5.1.

TABLE 5.1 – Tableau des valeurs expérimentales CNN

Taille du noyau	Rembourrage	Challenge	Criminalistique
		RMSE	RMSE
3	1	0.7902	0.3308
4	2	0.8950	0.3841
5	3	0.7986	0.3029
7	5	0.7611	0.3161
8	6	0.7842	0.2755
9	7	0.8130	0.2953

Le résultat pour les CNN, qui est de **0.7611** pour Challenge, a été obtenu avec une taille de noyau de 7 et un rembourrage de 5. De même, avec Criminalistique, nous avons obtenu une erreur de **0.2755**, avec un noyau de taille 8 et un rembourrage de 6.

Les figures 5.1 et 5.2 montrent, respectivement, l'évolution des erreurs d'entraînement et de validation sur les jeux de données Challenge et Criminalistique. La première correspond à une configuration avec un noyau de taille 7 et un rembourrage de 5, tandis que la seconde utilise un noyau de taille 8 et un rembourrage de 6.

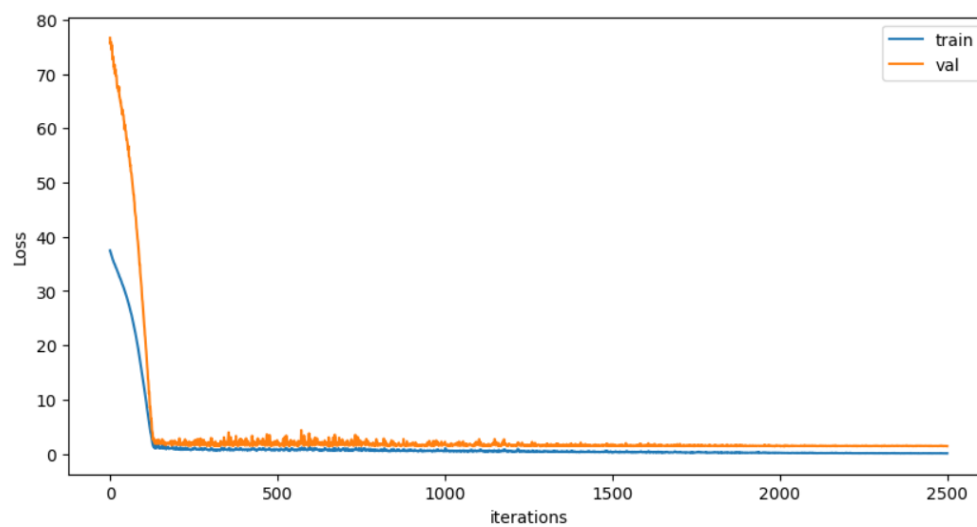


FIGURE 5.1 – Évolution de l'erreur avec Challenge

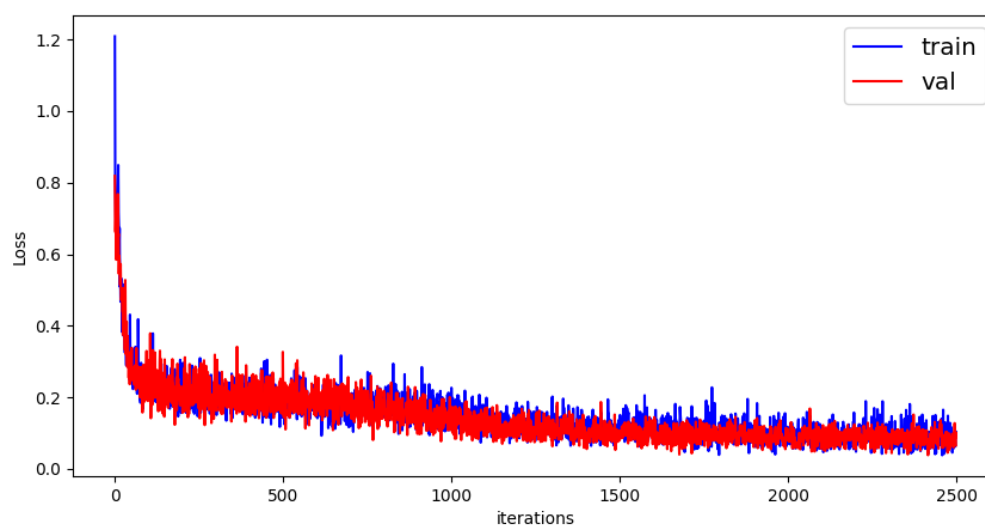


FIGURE 5.2 – Évolution de l'erreur du jeu de données Criminalistique

- **Interprétation Expérimentation 1**

Les valeurs plus élevées du RMSE pour Challenge observées dans toutes les configurations, indiquent que ces données sont plus difficiles à modéliser. Cela est probablement dû à leur haute dimensionnalité avec 682 longueurs d'onde et au bruit qu'elles contiennent. Un noyau plus large (de taille 7) semble mieux adapté pour capturer des relations pertinentes dans ces données complexes.

En revanche, les valeurs plus faibles du RMSE obtenues avec Criminalistique suggèrent que ce jeu de données est plus facile à modéliser. Cela pourrait s'expliquer par des relations plus évidentes entre les variables, possiblement dues à un nombre réduit de longueurs d'onde (130 exactement) ou à l'ajout de données contextuelles.

L'analyse des courbes des figures 5.1 et 5.2 révèle que, pour les données Challenge, la perte de validation est initialement instable, mais finit par décroître en suivant une tendance similaire à celle de la perte d'entraînement. Pour le jeu de données Criminalistique, l'évolution de la courbe montre une bonne généralisation du modèle, sans surapprentissage. Le modèle semble donc bien adapté aux données, avec une performance stable et efficace au fil des itérations.

- **Expérimentation :2**

La deuxième expérimentation a été réalisée en utilisant l'architecture des Transformers, en se focalisant uniquement sur la partie encodeur. Lors de cette phase, nous avons fait varier le nombre de têtes d'attention ainsi que le nombre de couches. Ces variations et les différents résultats sont présentés par le tableau 5.2

TABLE 5.2 – Les différentes valeurs du RMSE avec les Transformers

Nombre de tête	Nombre de couches	Challenge	Criminalistique
		RMSE	RMSE
2	4	0.8003	0.3399
2	6	0.8590	0.3156
3	3	0.8483	0.3320
4	3	0.8451	0.2782
4	4	0.8786	0.3145
5	6	0.8756	0.3382
6	4	0.8098	0.3026
8	4	0.7930	0.4222

En testant différentes configurations des Transformers, nous avons constaté que la meilleure configuration est obtenue avec 8 têtes d’attention et 4 couches, ce qui donne un RMSE de **0.7930** pour le jeu de données Challenge. Pour le jeu de données Criminalistique, un RMSE de **0.2782** a été obtenu avec une configuration de 4 têtes d’attention et 3 couches.

Les figures 5.3 et 5.4 présentent l’évolution des pertes d’entraînement et de validation correspondant aux meilleures valeurs de RMSE obtenues pour chaque jeu de données. La figure 5.3 illustre les résultats pour une architecture composée de 8 têtes d’attention et 4 couches avec le jeu de données Challenge, tandis que la figure 5.4 correspond à une configuration avec 4 têtes d’attention et 3 couches avec le jeu de données Criminalistique.

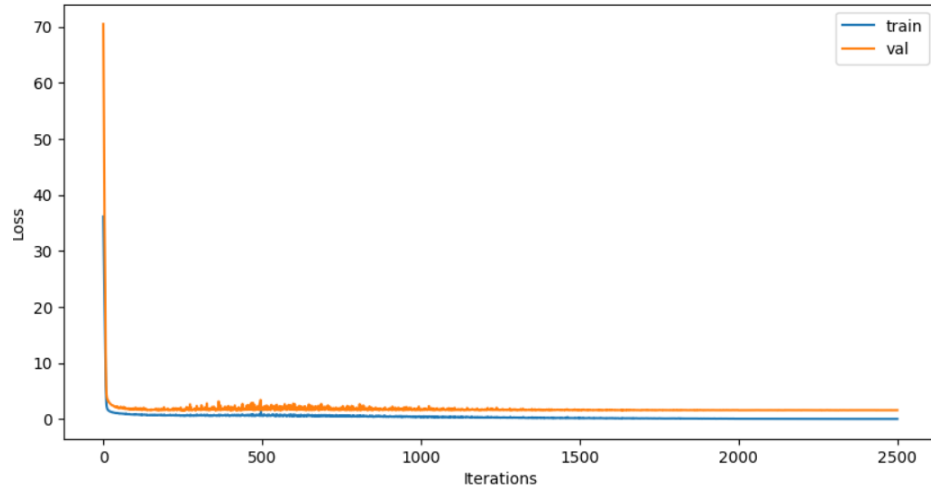


FIGURE 5.3 – Évolution de l'erreur du jeu de données Challenge

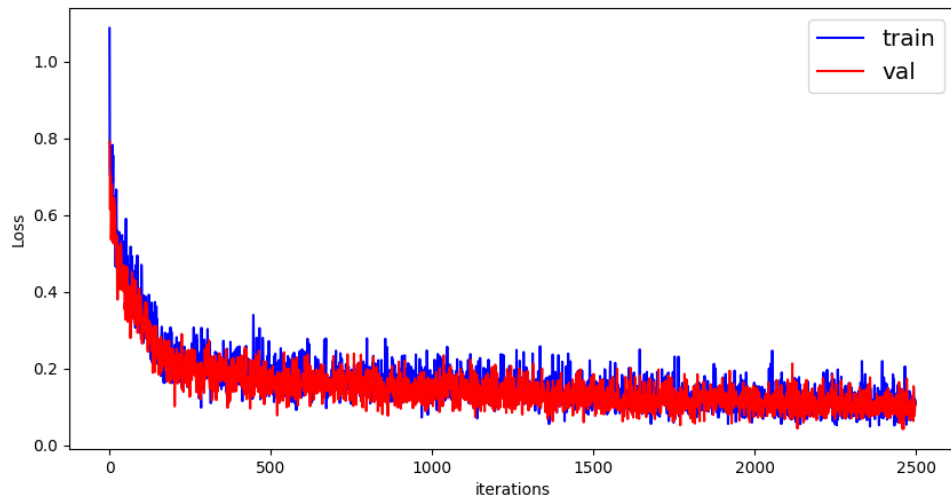


FIGURE 5.4 – Évolution de l'erreur du jeu de données Criminalistique

• Interprétation Expérimentation 2

Les résultats des différents RMSE indiquent que les meilleures performances sont obtenues avec le jeu de données Criminalistique, dont la faible dimensionnalité permet de capturer efficacement les interactions spectrales avec seulement 4 têtes d'attention, atteignant un RMSE de **0.2782**. À l'inverse, le jeu de données Challenge, plus complexe en raison de sa forte dimensionnalité, a obtenu un RMSE de **0.7930**. L'ensemble de ces performances reste toutefois légèrement moins bon que celles obtenues lors des

expérimentations avec les CNN.

Les courbes de pertes illustrent l'évolution des pertes d'entraînement et de validation pour chaque jeu de données. Pour les données Challenge, la convergence des pertes est progressive, ce qui est cohérent avec leur forte dimensionnalité et leur complexité élevée. À l'inverse, pour le jeu de données Criminalistiques, la convergence est plus rapide et stable, reflétant leur faible dimensionnalité et leur nature moins complexe. Ces observations confirment l'importance d'adapter l'architecture du modèle en fonction de la dimensionnalité et de la complexité des données.

- **Expérimentation :3**

Pour la troisième expérimentation, nous avons utilisé l'architecture des Visions Transformers avec encodeur uniquement et en modulant le nombre de patch ainsi que le nombre de tête d'attention.

Dans cette expérimentation, chaque observation est divisée en patch, auxquels sont associées leurs positions dans le modèle. L'objectif est d'évaluer si le regroupement de plusieurs réflectances peut révéler des associations significatives et améliorer la précision des estimations. Les résultats des expérimentations sont présentés sur le tableau 5.3 .

TABLE 5.3 – RMSE avec les différentes configurations des ViT

Nombre de têtes	Taille des patches	Challenge	Criminalistique
		RMSE1	RMSE2
4	15	0.7793	0.4434
4	10	0.9257	0.3341
6	15	0.7497	0.4378
6	20	0.8096	0.3468
6	10	0.7677	0.3827
6	5	0.7587	0.3525
6	4	0.8098	0.4174
8	4	0.7930	0.4307

Dans cette configuration, nous avons obtenu un RMSE de **0.7497**, soit le meilleur résultat pour les données du Challenge, en regroupant les données en patches de taille 15 et en utilisant 4 têtes d’attention. En revanche, avec le jeu de données Criminalistique, bien que le modèle conserve une certaine efficacité (RMSE de **0.3341**), ses performances restent inférieures à celles des configurations précédentes. Cela suggère que ce découpage en patches ne permet pas de capturer de manière optimale les relations entre les différentes réflectances. À l’inverse, le jeu de données Challenge semble mieux s’adapter à cette architecture, avec un RMSE plus faible que celui obtenu avec les autres configurations.

Les figures 5.5 et 5.6 présentent l’évolution des pertes d’entraînement et de validation obtenues pour les jeux de données Challenge et Criminalistique, respectivement. La première correspond à une architecture avec 4 têtes d’attention et une taille de patch de 10, tandis que la seconde utilise 6 têtes d’attention et une taille de patch de 15.

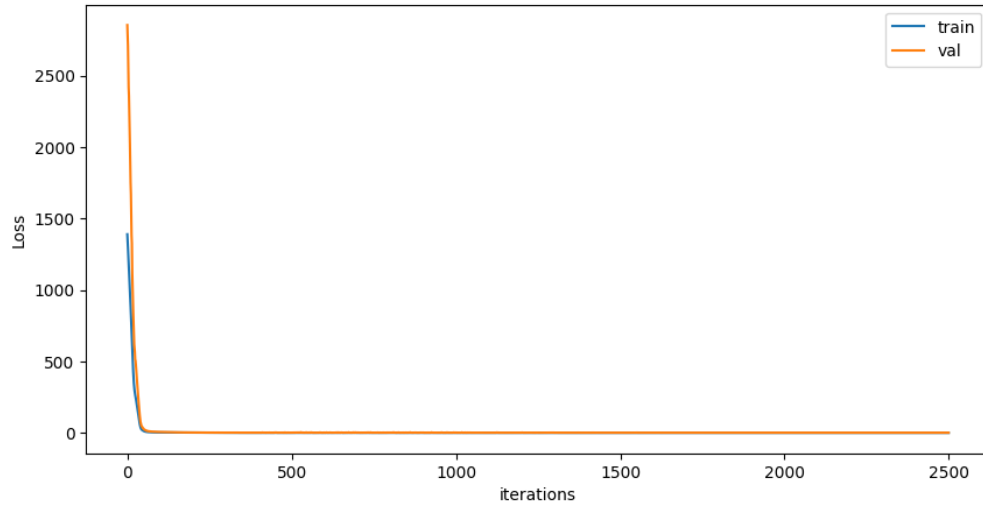


FIGURE 5.5 – Évolution de l'erreur du jeu de donnée Challenge

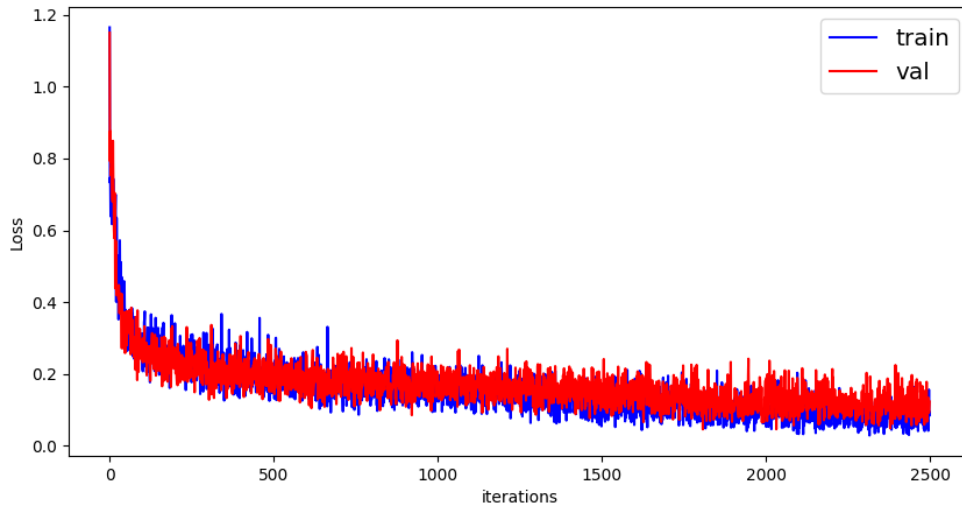


FIGURE 5.6 – Évolution de l'erreur du jeu de donnée Criminalistique

• Interprétation Expérimentation 3

Dans le tableau 5.3, nous constatons que le RMSE du jeu de données Challenge affiche de meilleurs résultats que ceux obtenus avec les architectures précédentes. Cela indique que le regroupement des spectres sous forme de patches a permis de capturer davantage de relations pertinentes. En revanche, pour le jeu de données Criminalistique, les résultats sont moins bons comparés aux architectures précédentes. Cela suggère qu'il existe moins de relations significatives dans ce jeu de données lorsqu'on

regroupe les spectres sous forme de patches, ce qui introduit davantage de bruit.

En observant les courbes de pertes des deux jeux de données dans la figure 5.5 et 5.6, nous remarquons que, pour les graphiques du jeu de données Challenge, la perte diminue rapidement au cours des premières époques avant de se stabiliser autour d'une valeur constante de 0.7497. L'absence de divergence entre les courbes d'entraînement et de validation suggère un faible surapprentissage.

En revanche, pour les courbes associées au jeu de données Criminalistique, les deux configurations montrent un modèle bien entraîné, avec une perte de validation qui suit la même tendance que celle de l'entraînement, Cela indique qu'un patch de taille 15 capture mieux les relations entre ces données complexes.

• Expérimentation : 4

La quatrième expérimentation porte sur une architecture complète des Transformers. Nous avons utilisé à la fois l'encodeur et le décodeur, tout en faisant varier le nombre de têtes d'attention et le nombre de couches dans chaque composant.

Plusieurs tests ont été réalisés :

Premier test : Nous avons divisé notre jeu de données en deux parties égales. La première moitié a été passée à travers l'encodeur, tandis que la seconde moitié a été transmise au décodeur. Ensuite, l'attention croisée a été appliquée, tout en gardant le nombre de couches constant.

Le tableau 5.6 présente un récapitulatif des différentes expérimentations effectuées ayant conduit au résultat d'un RMSE égal à **0.317** pour le jeu de données Criminalistique et de **0.8801** pour le jeu de données Challenge.

TABLE 5.4 – RMSE avec les différentes configurations de l’attention croisée

Nombre de têtes	Nombre de couches	Challenge	Criminalistique
		RMSE	RMSE
2	3	0.9226	0.3393
4	3	0.9023	0.3765
5	3	0.8801	0.3888
10	3	1.0275	0.3341
17	3	0.9137	0.3241
20	3	0.9267	0.3170

Les figures 5.7 et 5.8 illustrent l’évolution des pertes d’entraînement et de validation pour les jeux de données Challenge et Criminalistique, respectivement. La figure 5.7 correspond à une architecture comportant 5 têtes d’attention et 3 couches, tandis que la figure 5.8 présente les résultats obtenus avec 20 têtes d’attention et 3 couches.

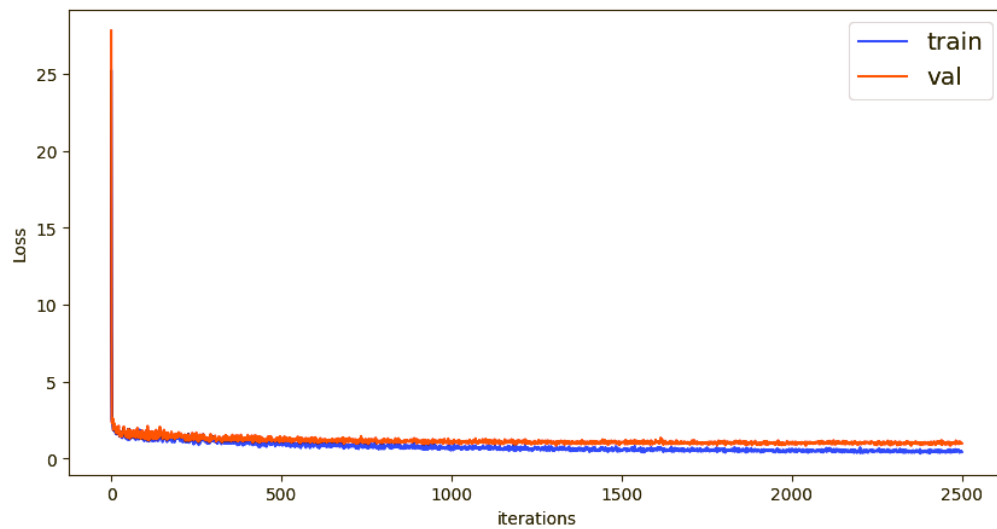


FIGURE 5.7 – Courbe de l’attention croisée jeu de donnée Challenge

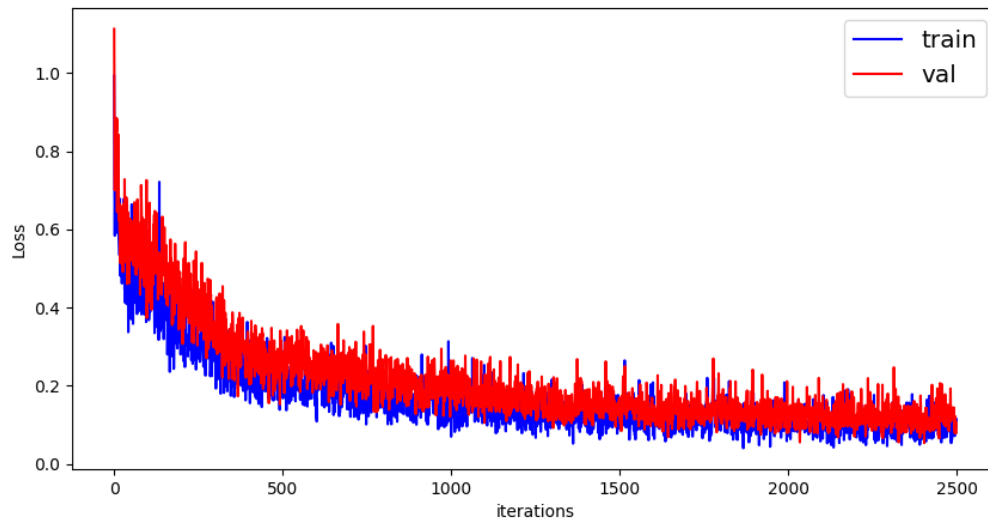


FIGURE 5.8 – Courbe de l’attention croisée jeu de donnée Criminalistique

Deuxième test : Le deuxième test que nous avons réalisé portait exclusivement sur les données Criminalistique. Disposant à la fois de données spectrales et environnementales, nous avons cherché à évaluer l’intérêt de leur combinaison à travers un mécanisme d’attention croisée. Pour ce faire, les données spectrales ont été introduites dans l’encodeur, tandis que les données environnementales ont été injectées dans le décodeur. Dans le cadre de ce test, seuls les types de substrat ont été utilisés comme données contextuelles.

Avec cette attention croisée en faisant varier le nombre de tête et en laissant invariant le nombre de couches du décodeur et de l’encodeur à 3, nous obtenons un RMSE de **0.1977** comme étant le meilleur pour cette expérimentation.

Le tableau 5.5 présente un récapitulatif des différentes expérimentations effectuées ayant conduit au résultat d’un RMSE égal à **0.1977**.

TABLE 5.5 – Résultats des RMSE avec l’attention croisée en utilisant seulement substrat

Nombre de têtes	Nombre de couches	Criminalistique
2	3	0.2335
4	3	0.1977
5	3	0.2455
10	3	0.2744
17	3	0.2059
20	3	0.2370

La courbe 5.9 présente les résultats de l’évolution des pertes d’entraînement et de validation pour la meilleure configuration sur le jeu de données Criminalistique.

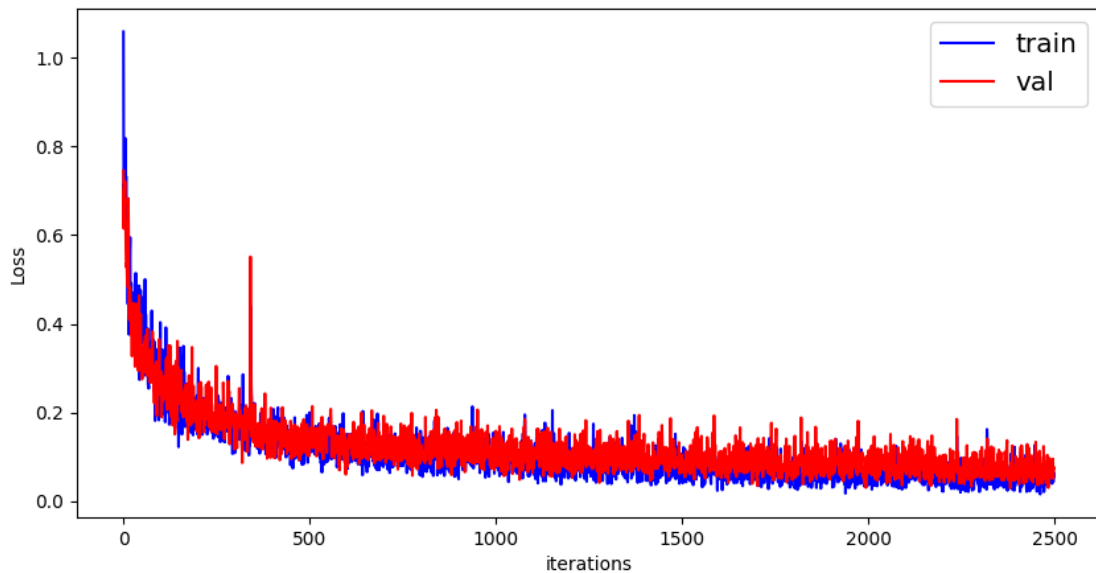


FIGURE 5.9 – Résultats attention croisée avec seulement le Substrat

Troisième test : Le troisième test que nous avons réalisé portait sur les données Criminalistiques ; nous voulions évaluer l’impact de l’intégration de toutes les données environnementales dans le modèle.

Le tableau 5.6 montre que le meilleur RMSE, soit **0.2112**, est obtenu avec une configuration de 20 têtes et 3 couches. Toutefois, ce résultat reste moins performant

que celui du "Deuxième test", qui ne prenait en compte que le substrat comme donnée environnementale.

TABLE 5.6 – Résultats des RMSE avec toutes les caractéristiques environnementales

Nombre de têtes	Nombre de couches	Criminalistique
2	3	0.2329
4	3	0.2170
5	3	0.2455
10	3	0.2115
17	3	0.2504
20	3	0.2112

La courbe 5.10 illustre l'évolution du RMSE pour les phases d'entraînement et de validation, ayant convergé vers une valeur finale de **0.2112**.

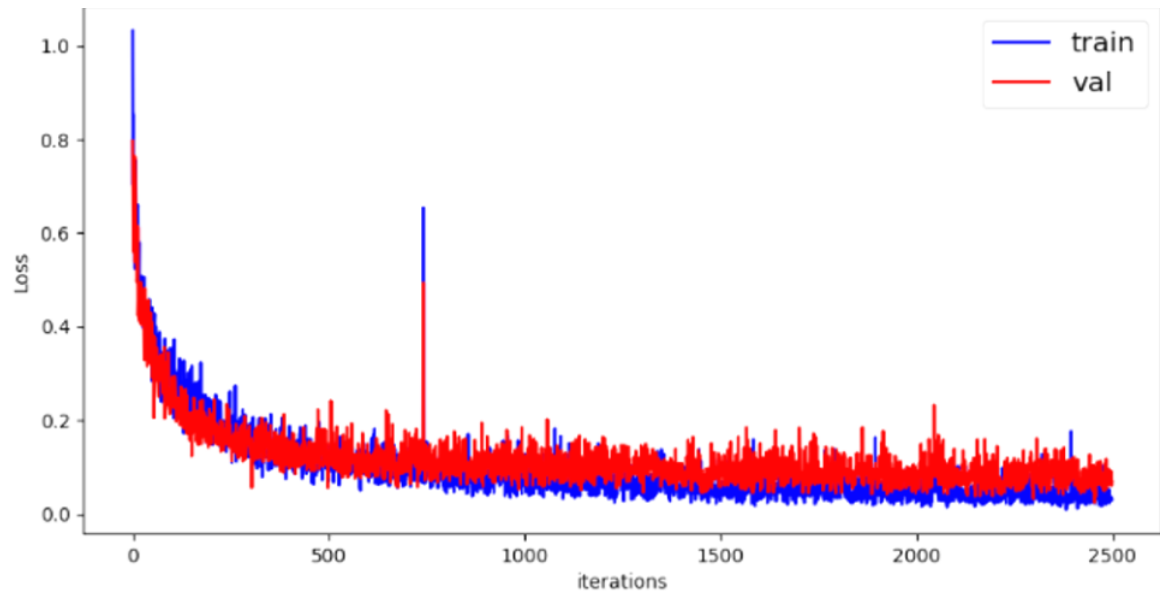


FIGURE 5.10 – Résultats attention croisée avec toutes les données contextuelles

• Interprétation Expérimentation 4

Dans le **premier test**, les données ont été divisées en deux parties égales : une moitié a été traitée par l'encodeur et l'autre par le décodeur. L'attention croisée a

ensuite été appliquée pour capturer les interactions entre les deux modules. Les résultats des RMSE montrent qu'en utilisant l'attention croisée, un nombre élevé de têtes d'attention, soit 20 têtes, est nécessaire pour capturer efficacement les interactions entre les entrées des deux modules (encodeur et décodeur) des Transformers.

La figure 5.7 présente la courbe d'évolution de l'erreur pour le jeu de données Challenge, un léger surapprentissage est observé, probablement dû au faible nombre de têtes d'attention (5 têtes) et à la complexité élevée des données, caractérisée par un grand nombre de colonnes. Cette configuration pousse le modèle à explorer davantage de relations, amplifiant ce phénomène. En revanche, la figure 5.8 du jeu de données Criminalistique montre de bonnes performances par rapport aux autres architectures, indiquant que l'attention croisée permet de mieux modéliser les interactions entre ces données.

Le deuxième test s'est concentré sur les données Criminalistique, composées à la fois de données spectrales et contextuelles. En croisant ces deux types de données, nous avons observé une meilleure interaction entre les variables, tout en nécessitant un faible nombre de têtes d'attention. Avec 4 têtes, le modèle atteint un RMSE de **0.1977**, surpassant toutes les autres architectures déjà explorées. Cela suggère qu'en combinant les données contextuelles et spectrales, nous parvenons à capturer davantage de relations entre les données, ce qui améliore les performances globales.

Pour le Troisième test, l'objectif était d'évaluer l'impact de l'intégration de l'ensemble des données environnementales. Nous avons observé que le modèle atteint son meilleur RMSE de **0.2112** avec une configuration de 20 têtes. Toutefois, cette performance reste inférieure à celle obtenue lors du deuxième test. Ce résultat peut s'expliquer par le fait que certaines variables contextuelles introduisent du bruit, ce qui complique l'apprentissage pour le modèle. À l'inverse, le deuxième test, qui ne prenait en compte que le substrat comme donnée environnementale, a permis d'obtenir de meilleures performances avec seulement 4 têtes, probablement en raison de la

simplicité et de la pertinence de cette variable.

5.2 Etude d'explicabilité

Les Transformers reposent sur le mécanisme d'attention, qui calcule des poids attribuant une importance relative à chaque élément d'entrée par rapport aux autres. Ces poids d'attention permettent d'identifier les caractéristiques les plus influentes dans la prise de décision du modèle, offrant ainsi un premier niveau d'explicabilité. Cette étude d'explicabilité a été effectuée sur le jeu de données Criminalistique, compte tenu de la sensibilité de ce domaine quant à l'obtention de résultats pertinents.

Dans le cadre de l'analyse spectrale, il est crucial de comprendre quelles longueurs d'onde le modèle considère comme pertinentes pour effectuer ses estimations. Cela permet non seulement de valider les résultats obtenus, mais également d'offrir des informations exploitables pour les experts du domaine.

- **Fixation des Graines et Sélection des Données** : elle garantit que l'échantillon X utilisé est le même à chaque exécution. Soit $X \in \mathbb{R}^{n \times d}$, où n est le nombre d'observations dans le batch, et d est la dimension des caractéristiques (longueurs d'onde, spectres, etc.).
- **Extraction des Poids d'Attention** : Les poids d'attention, notés A , sont obtenus du modèle. Le mécanisme d'attention est calculé via :

$$A = \text{softmax} \left(\frac{QK^\top}{\sqrt{d_k}} \right) \quad (5.1)$$

avec :

- Q : Représentation des requêtes (queries) de dimension d_k ,
- K : Représentation des clés (keys) de dimension d_k ,

- $\sqrt{d_k}$: Facteur de normalisation,
- softmax : Fonction pour convertir les scores en probabilités.

La sortie $A \in \mathbb{R}^{n \times n}$ une matrice de relation entre chaque paire d'éléments ou une version réduite, comme une moyenne sur les têtes d'attention.

- **Moyenne des Poids d'Attention** : Pour simplifier, une moyenne \bar{A} des poids d'attention est calculée sur les dimensions non nécessaires :

$$\bar{A} = \frac{1}{h} \sum_{i=1}^h A_i \quad (5.2)$$

Où h est le nombre de têtes d'attention. Le résultat $\bar{A} \in \mathbb{R}^n$ représente l'importance pour chaque observation dans X .

- **Filtrage des Poids d'Attention Importants** : On identifie les poids \bar{A}_i supérieurs à un seuil τ :

$$I = \{i \mid \bar{A}_i > \tau\} \quad (5.3)$$

Le seuil τ ($\tau = 0.08$) a été fixé manuellement à la suite de plusieurs essais. En dessous de cette valeur, presque toutes les longueurs d'onde étaient retenues, rendant l'analyse peu informative. Au-dessus, seules deux longueurs d'onde étaient sélectionnées 945 nm et 1019 nm, ce qui limitait l'explicabilité. Ce seuil représente donc un compromis entre exhaustivité et sélectivité.

- **Normalisation des Poids d'Attention** : Les poids \bar{A} des indices importants I sont normalisés entre 0 et 1 pour une visualisation cohérente :

$$A_{\text{norm}}(i) = \frac{\bar{A}_i - \min(\bar{A}_I)}{\max(\bar{A}_I) - \min(\bar{A}_I)} \quad \text{pour } i \in I \quad (5.4)$$

Cette normalisation permet d'attribuer une couleur ou une intensité visuelle à chaque point.

- **Analyse des Longueurs d'Onde** : Les longueurs d'onde associées aux indices I sont extraites. Si λ représente les longueurs d'onde des caractéristiques dans

X , les longueurs d'onde importantes sont données par :

$$\Lambda_{\text{important}} = \{\lambda_i \mid i \in I\} \quad (5.5)$$

- **Visualisation et Résumé** : Pour chaque longueur d'onde λ_i , on affiche :
 - Sa valeur dans X : x_i ,
 - Son importance (poids normalisé) : $A_{\text{norm}}(i)$.

Graphiquement :

- λ est représenté sur l'axe des abscisses,
- x_i est représenté sur l'axe des ordonnées,
- Les couleurs et tailles des points sont proportionnelles à $A_{\text{norm}}(i)$.

La figure 5.11 illustre les poids d'attention attribués aux réflectances les plus significatives par notre modèle dans le cadre de **l'expérimentation 4**, réalisée sur les données Criminalistique. Ces poids proviennent d'un mécanisme d'attention croisée basé sur les longueurs d'onde. Plus une zone tend vers le jaune, plus sa contribution à la prédiction est importante. Il apparaît que les longueurs d'onde les plus influentes sont 945 nm, 1019 nm, 1100.125 nm et 1316 nm. Un tableau récapitulatif de ces longueurs d'onde est disponible dans l'annexe A.

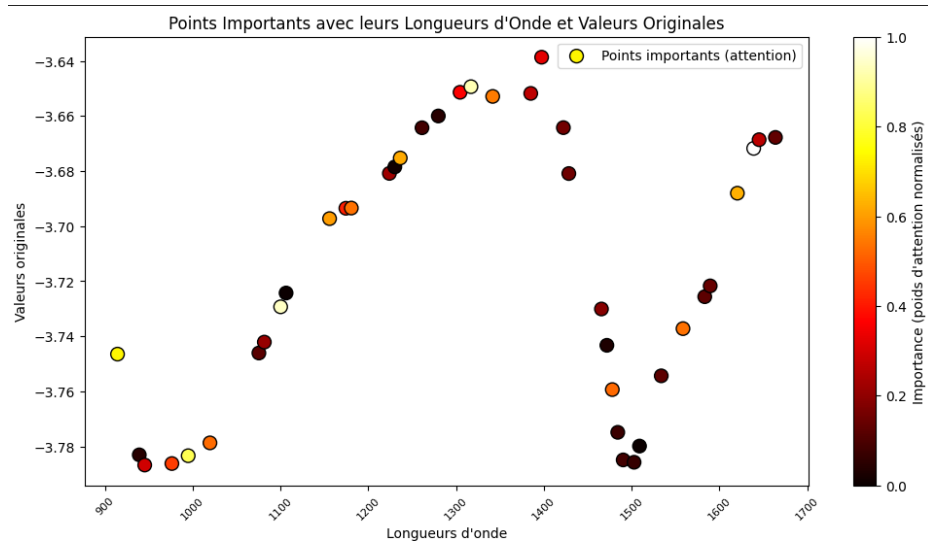


FIGURE 5.11 – *Explicabilité Transformer*

5.3 Synthèse des résultats des expérimentations

Cette synthèse des résultats consiste en une comparaison de nos différents résultats avec l'approche PLSR utilisée dans les travaux de Matthieu Lesnoff et al. [1], ayant obtenu un RMSE de 0.64 sur le même ensemble de données Challenge.

Les figures 5.12 et 5.13 présentent respectivement un récapitulatif des performances de nos différentes architectures comparées à celle du modèle PLSR, ainsi qu'une comparaison de ces architectures appliquées aux différents jeux de données.

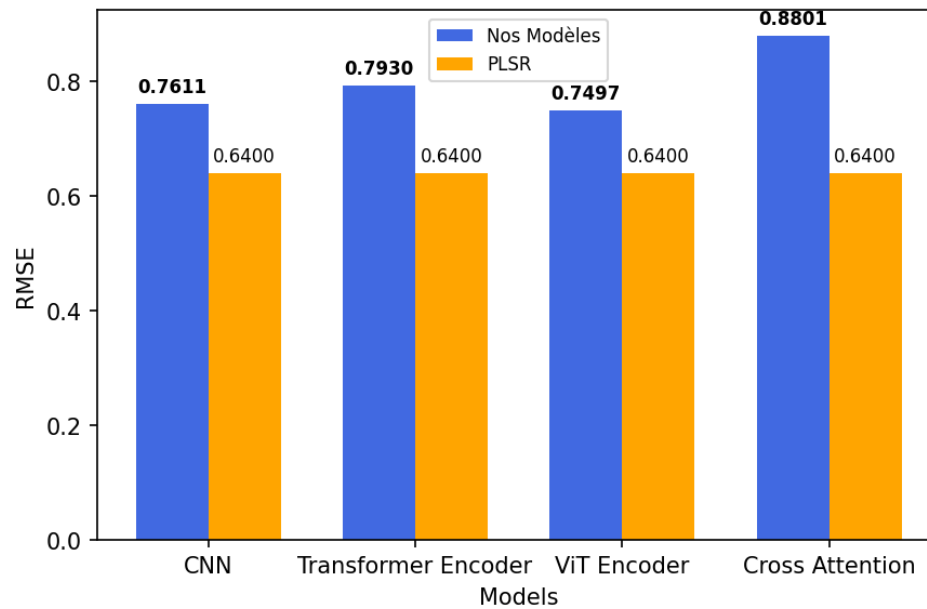


FIGURE 5.12 – comparaison architecture avec le PLSR

La figure 5.12 montre que les performances de nos différentes architectures sont inférieures à celles des PLSR. Cependant, le modèle ViT montre des résultats prometteurs, suggérant qu'avec une configuration optimisée des patches, ces performances pourraient être encore améliorées.

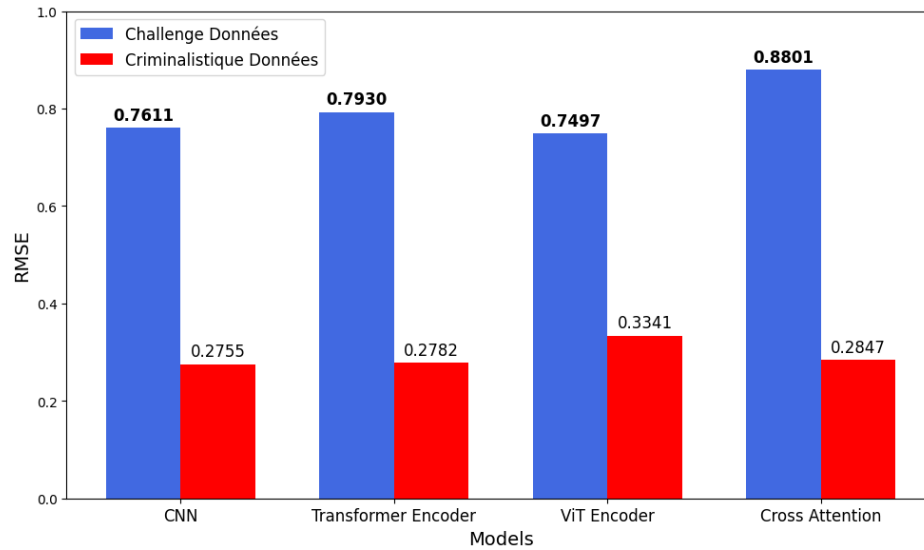


FIGURE 5.13 – Comparaison des architectures avec les deux jeux de données

Nous observons à la figure 5.13 que de manière générale, les modèles s'adaptent mieux au jeu de données Criminalistique qu'à Challenge. Cela s'explique par la complexité accrue du jeu de données Challenge, qui contient un plus grand nombre de caractéristiques.

5.4 Conclusion

Cette section dédiée aux résultats et analyses des expérimentations nous a permis de mieux comprendre le comportement de nos modèles sur les deux jeux de données étudiés, Challenge et Criminalistique. Nous avons constaté que, de manière générale, les modèles offrent de meilleures performances sur le jeu de données Criminalistique par rapport à Challenge. Cette différence peut s'expliquer par la complexité accrue du jeu de donnée Challenge, qui contient bien plus de caractéristiques, rendant les relations intra-caractéristiques plus difficiles à capturer. Après avoir testé différentes architectures, nous avons identifié que la meilleure configuration pour le jeu de données Challenge est l'utilisation des Vision Transformers. Avec des patches de taille 15, cette

architecture semble mieux modéliser les relations entre les différentes caractéristiques. Cela nous a permis d’obtenir une erreur globale de **0.7497**, qui est la plus faible parmi toutes les architectures testées sur Challenge. Pour le jeu de données Criminalistique, l’architecture de l’attention croisée a donné les meilleurs résultats. Cette performance est due à la séparation des données, en données spectrales et environnementales. En croisant ces deux types de données, le modèle a pu capturer des relations plus riches entre les caractéristiques avec un RMSE de **0.1977**. Ces observations confirment l’importance d’adapter l’architecture des modèles en fonction des spécificités des jeux de données pour optimiser les performances.

Enfin, l’analyse des poids d’attention du modèle Transformer a mis en évidence les longueurs d’onde 945.266 nm, 1019.598 nm, 1100.125 nm et 1316.927 nm comme particulièrement influentes pour le jeu de données Criminalistique. Ces résultats offrent une piste d’interprétation précieuse pour la compréhension des phénomènes physico-chimiques sous-jacents, tout en illustrant la capacité des Transformers à capturer des relations spectrales pertinentes.

Chapitre 6

Conclusion Générale

Ce mémoire a exploré différentes architectures pour le traitement de données spectroscopiques à partir de deux jeux de données : Criminalistique et Challenge. L'objectif principal était d'évaluer la pertinence d'approches d'apprentissage profond, telles que les CNN et les Transformers, pour estimer des mesures de concentration ou de datation relative des taches de sang à partir de réflectances présentant des niveaux de complexité variés. L'ensemble des expérimentations visait à identifier des configurations optimisées permettant d'améliorer la précision des estimations tout en tenant compte des spécificités propres à chaque jeu de données.

Nous avons réalisé plusieurs expérimentations, notamment l'estimation des concentrations à partir du jeu de données Challenge, ainsi que l'estimation de la durée d'exposition des taches de sang à partir du jeu de données Criminalistique. Ces expérimentations ont été effectuées à l'aide des CNN, des Vision Transformers utilisant uniquement l'encodeur, ainsi que de l'architecture complète des Transformers exploitant le mécanisme d'attention croisée entre l'encodeur et le décodeur. Ces expérimentations ont montré que certaines architectures étaient mieux adaptées à un jeu de données qu'à un autre. Ainsi, les Vision Transformers ont obtenu les meilleures per-

performances sur le jeu Challenge, avec une erreur globale de 0.7449. Alors que, pour le jeu de donnée Criminalistique, l'architecture basée sur l'attention croisée a été la plus performante, atteignant une erreur globale de 0.19. Ces résultats soulignent l'importance d'adapter les modèles aux spécificités des données pour maximiser leur efficacité. Sur le jeu Challenge, bien que les Transformers aient obtenu des résultats inférieurs à ceux de la régression PLSR (RMSE=0.64), ils se démarquent par leur capacité à offrir une meilleure interprétabilité à l'analyse des poids d'attention. Dans le domaine des sciences forensiques les résultats obtenus avec les Transformers pour la datation relative des taches de sang ouvrent des perspectives prometteuses, notamment dans des contextes où l'explicabilité des modèles est essentielle.

Parmi les principales contributions de ce mémoire, nous pouvons citer l'adaptation des architectures des Transformers et des CNN aux données spectroscopiques, caractérisées par des niveaux de complexité distincts. L'utilisation des architectures des Transformers pour l'estimation de la datation relative des taches de sang et de la concentration en protéines dans les fourrages constitue une approche novatrice dans le domaine spectral. Par ailleurs, la comparaison des différentes architectures avec celles reconnues pour leurs performances dans la littérature met en lumière notre contribution originale à la recherche dans le domaine.

Une autre contribution majeure réside dans l'analyse de l'explicabilité des modèles, rendue possible par l'exploitation des mécanismes d'attention. L'identification des longueurs d'onde les plus influentes, notamment 945.266 nm, 1019.598 nm, 1100.125 nm et 1316.927 nm, fournit une piste d'interprétation précieuse pour la compréhension des phénomènes physico-chimiques. Cette capacité à interpréter les décisions du modèle représente un avantage significatif, en particulier dans des contextes sensibles comme la criminalistique, où la transparence des modèles est essentielle.

Certaines limites ont été identifiées. La forte dimensionnalité et le bruit potentiel du jeu de données Challenge ont posé des défis à certaines architectures, notamment

aux Transformers avec attention croisée. En revanche, le jeu de données Criminalistique a produit de bons résultats grâce à la combinaison des données environnementales et spectrales.

Plusieurs pistes d'amélioration peuvent être envisagées. L'utilisation d'architectures hybrides combinant des CNN, pour capturer les relations locales, et des Transformers, pour modéliser les relations globales, pourrait améliorer les performances. De plus, augmenter la taille des jeux de données permettrait d'accroître la robustesse des modèles même si l'accès à de grandes quantités de données spectroscopie proche infra-rouge reste un défi.

Ces résultats ouvrent des perspectives intéressantes pour de futures applications et avancées méthodologiques non destructives dans l'analyse et le traitement des données spectrales.

Bibliographie

- [1] M. LESNOFF, M. MAXIME et R. JEAN-MICHEL, « Comparison of locally weighted pls strategies for regression and discrimination on agronomic nir data », *Journal of Chemometrics*, vol. 34, 2020.
- [2] Y. KEQIANG, F. SHIYAN et Z. YANRU, « Heavy metal hg stress detection in tobacco plant using hyperspectral sensing and data-driven machine learning methods », *Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy*, vol. 245, 2021.
- [3] Z. HONG, L. DONGXIAO, H. XINDAN et M. XIAOJING, « Infrared metamaterial for surface-enhanced infrared absorption spectroscopy : pushing the frontier of ultrasensitive on-chip sensing », *International Journal of Optomechatronics*, vol. 15, 2021.
- [4] A. FRANTISHEK, A. HADIS, S. MANAF et A. DHANANJAY, « Explainable predictive modeling for limited spectral data », *Chemometrics and Intelligent Laboratory Systems*, vol. 225, 2022.
- [5] A. TESFAYE, X. WENBO et F. JINLONG, « Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images », *BMC Bioinformatics*, vol. 9, 2008.
- [6] S. ALEXANDER, W. LILY et A. C. F, « Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images », *Remote Sensing*, vol. 14, 2022.
- [7] G. FENG et L. JIANFEI, « Simple dilated convolutional neural network for quanti-

- tative modeling based on near infrared spectroscopy techniques », *Chemometrics and Intelligent Laboratory Systems*, vol. 232, 2023.
- [8] R. CORDEIRO, R. ANTÓNIO, P. OCTAVIAN et S. PEDRO, « Neural Architecture Search for 1D CNNs—Different Approaches Tests and Measurements », *Sensors*, vol. 21, 2021.
- [9] B. E. JANN, G. MADS et S. THOMAS, « Data Augmentation of Spectral Data for Convolutional Neural Network (CNN) Based Deep Chemometrics », *arXiv preprint*, vol. arXiv, 2017.
- [10] J. FRAUNHOFER et W. WOLLASTON, « Prismatic and Diffraction Spectra », *Harper and Brothers*, vol. 92, 1898.
- [11] W. SVANTE, S. MICHAEL et E. LENNART, « PLS-regression : a basic tool of chemometrics », *Chemometrics and Intelligent Laboratory Systems*, vol. 58, 2001.
- [12] S. EMAN, A. EL-KHAIR et S. D. A., « NIR spectroscopy-multivariate analysis for rapid authentication, detection and quantification of common plant adulterants in saffron (*Crocus sativus* L.) stigmas », *LWT*, vol. 122, 2020.
- [13] L. WEIJIE, B. SYLVIE, C. SONGCHAO, L. ALEXANDRE, R. C. M. G. C. et J. BENOIT, « Visible, near- and mid-infrared spectroscopy coupled with an innovative chemometric strategy to control apple puree quality », *Food Control*, vol. 120, 2021.
- [14] M. G. ARAMDAS, D. BAPPA, M. DAYESH, H. ITTAI, B. KATJA, S. R. N., P. KIRAN, D. ASHWINI, M. SHAIESH et K. R. M., « Monitoring the Foliar Nutrients Status of Mango Using Spectroscopy-Based Spectral Indices and PLSR-Combined Machine Learning Models », *Remote Sensing*, vol. 13, 2021.
- [15] HU et CHUANMIN, « Remote detection of marine debris using satellite observations in the visible and near infrared spectral range : Challenges and potentials », *Remote Sensing of Environment*, vol. 259, 2021.
- [16] C. SECHUL, « Portable analytics as a contemporary environmental microplastic research tool : The Pitches and Hitches », *Trends in Environmental Analytical Chemistry*, vol. 10, 2024.

- [17] L. XUE, Z. Y. ZHANG LIANGXIAO, W. DU, W. XUEFANG, Y. LI, Z. WEN et L. PEIWU, « Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils », *Trends in Food Science and Technology*, vol. 101, 2020.
- [18] G. ZHIMING, B. A. OSEI, Y. LIMEI, C. QUANSHENG, S. JIYONG, E.-S. H. R. et Z. XIAOBO, « Intelligent evaluation of taste constituents and polyphenols-to-amino acids ratio in matcha tea powder using near infrared spectroscopy », *Food Chemistry*, vol. 353, 2021.
- [19] A. H. Z., J. RAHUL, M. R. EVI, C. MYOUNG-GUN, K. KYUNG-HWAN, K. GEON-WOO et C. BYOUNG-KWAN, « Nondestructive measurement of anthocyanin in intact soybean seed using Fourier Transform Near-Infrared (FT-NIR) and Fourier Transform Infrared (FT-IR) spectroscopy », *Infrared Physics and Technology*, vol. 111, 2020.
- [20] K.-A. KAMILA et K. JOANNA, « The application of spectroscopic techniques in combination with chemometrics for detection adulteration of some herbs and spices », *Microchemical Journal*, vol. 153, 2020.
- [21] N. SAID, M. ELSAYED, E.-E. S. SAFA, M. W. S., R. N. Y. et H. A. A., « Estimation of key potentially toxic elements in arid agricultural soils using Vis-NIR spectroscopy with variable selection and PLSR algorithms », *Front. Environ. Sci.*, vol. 11, 2023.
- [22] M. I. S., B.-C. FERNANDO et S. DAVID, « Accurate non-destructive prediction of peach fruit internal quality and physiological maturity with a single scan using near infrared spectroscopy », *Food Chemistry*, vol. 335, 2021.
- [23] H. XUEQIN, X. DANPING, S. HAN, M. JINFANG, Z. YONGXIN, C. JIAZE, Y. YANYAN et H. FURONG, « Estimation of chemical oxygen demand in different water systems by near-infrared spectroscopy », *Ecotoxicology and Environmental Safety*, vol. 243, 2022.
- [24] L. YONGZONG, A. ERIC, D. HONGWEI et H. YONGGUANG, « Quantitative Assessment of Cold Injury in Tea Plants by Terahertz Spectroscopy Method », *Agronomy*, vol. 13, 2023.

- [25] STANFORD GRADUATE SCHOOL OF BUSINESS, « Explainer : What is machine learning? ». <https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf>, 2024. Stanford Human-Centered AI Institute, consulté le 2 octobre 2024.
- [26] Q. TING, Y. YUANZHEN, sun HAOJIE, H. TINGTING, W. XUECHENG, W. YAGI, W. ZHENFENG, Z. LINGYUN, Z. WEIFENG et Y. MING, « Rapid discrimination and quantification of kudzu root with its adulterant part using FT-NIR and a machine learning algorithm », *Vibrational Spectroscopy*, vol. 116, 2021.
- [27] M. AMBRA, F. NUNZIA et F. DAVIDE, « Use of machine learning tools and NIR spectra to estimate residual moisture in freeze-dried products », *Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy*, vol. 293, 2023.
- [28] M. I. S. M. H. TAN, M. F. JAMLOS, A. F. OMAR, K. KAMARUDIN et M. A. JAMLOS, « Ganoderma boninense classification based on near-infrared spectral data using machine learning techniques », *Chemometrics and Intelligent Laboratory Systems*, vol. 232, 2023.
- [29] W. MEGAN, O. D. AL-JUMEILY, A. ISMAIL, K. IFTIKHAR, B. JASON et TANG, « Palm-sized Near-Infrared Spectroscopy and Machine Learning Analytics for the Detection of Endogenous Constituents and Drugs in Human Fingernails », *IEEE*, vol. 232, 2023.
- [30] Y. KEQIANG, F. SHIYAN et Z. YANRU, « Heavy metal Hg stress detection in tobacco plant using hyperspectral sensing and data-driven machine learning methods », *Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy*, vol. 245, 2021.
- [31] A. NUNIK, S. EDO et S. AGUSTAMI, « An automatic generation of pre-processing strategy combined with machine learning multivariate analysis for NIR spectral data », *Journal of Agriculture and Food Research*, vol. 13, 2023.
- [32] R. MATHEUS, L. FLAVIO, D. C. P. H., Z. M. POLTRONIERI, D. L. L. M., N. I. TAVARES, R. ISABELLA, C. T. R.P., D. P. V. LUCIANA, R. RENAN, S. M. C.S., F. P. LYRA, R. WANDERSON, F. P. R. et K. R. A., « Skin cancer diagnosis using

- NIR spectroscopy data of skin lesions in vivo using machine learning algorithms », *Biocybernetics and Biomedical Engineering*, vol. 44, 2024.
- [33] N. J. F. ISINGIZWE, H. E. A., V. MATTIA, T. CHRISTOPHER, P. W. JACOBUS et O. U. LINUS, « Feature Reduction for the Classification of Bruise Damage to Apple Fruit Using a Contactless FT-NIR Spectroscopy with Machine Learning », *Foods*, vol. 12, 2023.
- [34] G. STANISŁAW et G. WOJCIECH, « Supporting soil and land assessment with machine learning models using the Vis-NIR spectral response », *Geoderma*, vol. 405, 2022.
- [35] C.-T. J.P., L. D. S. M. MARIA et B. D. FERNANDES, « On-line monitoring of egg freshness using a portable NIR spectrometer in tandem with machine learning », *Journal of Food Engineering*, vol. 306, 2021.
- [36] C. DELUZARCHE, « Définition | deep learning - apprentissage profond | futura tech ». <https://www.futura-sciences.com/tech/definitions/intelligence-artificielle-deep-learning-17262/>, 2022. Futura Sciences, consulté le 2 juin 2024.
- [37] Z. S. V., P. E. V., V. V. E., V. M. A., S. E. V., Z. A. D., S. A. V., Z. T. S. et G. A. P., « Hyperspectral Non-Imaging Measurements and Perceptron Neural Network for Pre-Harvesting Assessment of Damage Degree Caused by Septoria/Stagonospora Blotch Diseases of Wheat », *Agronomy*, vol. 13, 2023.
- [38] Z. HONG, Z. CHU, T. M. FARAG, W. XINHUA, H. YONG, Q. ZHENGJUN et L. YUFEI, « Wheat Kernel Variety Identification Based on a Large Near-Infrared Spectral Dataset and a Novel Deep Learning-Based Feature Selection Method », *Frontiers in Plant Science*, vol. 11, 2021.
- [39] S. SERGEY, S. NIR, B. MIKHAIL, B. NADIA, M. DAVID et K. IFTACH, « ER-R : Improving regression by deep learning and prior knowledge utilization for fluorescence analysis », *Chemometrics and Intelligent Laboratory Systems*, vol. 236, 2021.
- [40] P. DÁRIO et M. PUNEET, « An automated deep learning pipeline based on advanced optimisations for leveraging spectral classification modelling », *Analy-*

- tica Chimica Acta* *Chemometrics and Intelligent Laboratory Systems* 192, vol. 215, 2021.
- [41] W. SZYMON et J. MAŁGORZATA, « Deep neural networks in profiling of apple juice adulteration based on voltammetric signal of the iridium quadruple-disk electrode », *Chemometrics and Intelligent Laboratory Systems*, vol. 209, 2021.
 - [42] S. HUI, S. LINWEI, W. JINJIN, X. ZHIBING, Z. SUWEI, W. ZIHAN, W. HUIJIE et Y. JIANHUA, « NIR spectroscopy combined with 1D-convolutional neural network for breast cancerization analysis and diagnosis », *Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy*, vol. 287, 2023.
 - [43] L. HUIYU, S. CHEN, W. GONGJI, S. QINRU, Y. KAI et L. ZEFENG, « BloodNet : An attention-based deep network for accurate, efficient, and costless bloodstain time since deposition inference », *Briefings in Bioinformatics*, vol. 24, 2023.
 - [44] Y. A. A. HOSSEIN, H. G. ALI et P. VAHID, « Deep Neural Networks Meet CSI-Based Authentication », *arXiv*, vol. 215, 2018.
 - [45] V. ASHISH, S. NOAM, P. NIKI, U. JAKOB, J. LLION, G. A. N, K. ŁUKASZ et P. ILLIA, « Attention is All you Need », *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
 - [46] B. L. DOSOVITSKIY ALEXEY, K. ALEXANDER, W. DIRK, Z. XIAOHUA, U. THOMAS, D. MOSTAFA, M. MATTHIAS, H. GEORG, G. SYLVAIN, U. JAKOB et H. NEIL, « An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale », *arXiv*, vol. 30, 2021.
 - [47] D. R. R. et C. T., « Root mean square error (RMSE) or mean absolute error (MAE) ? – Arguments against avoiding RMSE in the literature », *Geoscientific Model Development*, vol. 7, 2014.
 - [48] R. P. J., D. MICHIEL, E. SANNE et H. MIA, « Sensitivity and generalized analytical sensitivity expressions for quantitative analysis using convolutional neural networks », *Critical Reviews in Analytical Chemistry*, vol. 36, 2006.
 - [49] C. JOHN, Z. HUASHA, J. BOBBY, C. YE et M. JIANGCHANG, « Machine learning at the limit », *2015 IEEE International Conference on Big Data (Big Data)*, vol. 242, 2015.

- [50] K. MOEZ, « Convolutional Neural Networks : A Survey », *Computers*, vol. 12, 2023.
- [51] P. ADAM, G. SAM, M. FRANCISCO, L. ADAM, B. JAMES, C. GREGORY, K. TREVOR, L. ZEMING, G. NATALIA, A. LUCA, D. ALBAN, K. ANDREAS, Y. EDWARD, D. ZACH, R. MARTIN, T. ALYKHAN, C. SASANK, S. BENOIT, F. LU, B. JUNJIE et C. SOUMITH, « PyTorch : An Imperative Style, High-Performance Deep Learning Library », *arXiv.org*, vol. 16, 2019.
- [52] SCIENCESCONF, « Chemometrics 2018 - page officielle ». <https://chemom2018.sciencesconf.org/resource/page/id/5.html>, 2018. SciencesConf, consulté le 19 novembre 2024.
- [53] Z. YINGZHU, N. CHONGJIA, L. CHEUNG-CHI, joty SHAFIQ, C. E. SIONG et M. BIN, « Cross Attention with Monotonic Alignment for Speech Transformer », *ISCA*, vol. 16, 2020.
- [54] L. QIWU, su JIAOJIAO, Y. CHUNHUA, G. WEIHUA, S. OLLI et L. LI, « CAT-EDNet : Cross-Attention Transformer-Based Encoder–Decoder Network for Saliency Defect Detection of Strip Steel Surface », *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022.
- [55] L. ILYA et H. FRANK, « Decoupled Weight Decay Regularization », *arXiv.org*, vol. 16, 2019.

Annexe A

Longueurs d'onde importantes

TABLE A.1 – Tableau des longueurs d'onde et des valeurs originales

Longueur d'onde (nm)	Valeur originale
914.294	-3.7465
939.072	-3.7830
945.266	-3.7867
976.238	-3.7862
994.821	-3.7833
1019.598	-3.7787
1075.348	-3.7461
1081.542	-3.7421
1100.125	-3.7293
1106.319	-3.7243
1155.874	-3.6973
1174.457	-3.6935
1180.652	-3.6934
1224.012	-3.6808
1230.206	-3.6785
1236.401	-3.6753
1261.178	-3.6643
1279.761	-3.6600
1304.539	-3.6514
1316.927	-3.6494
1341.705	-3.6529
1385.065	-3.6518
1397.454	-3.6387