République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Mohamed El Bachir El Ibrahimi de Borj Bou Arréridj
Faculté des Mathématiques et d'Informatique
Département d'Informatique



MEMOIRE
Présenté en vue de l'obtention du diplôme
Master en Informatique
Spécialité : Réseaux et Multimédia

THEME

Étude des Techniques d'Interprétabilité Appliquées aux Modèles CNN pour la Détection Précoce de la Maladie d'Alzheimer à Partir d'Images IRM.

Présenté par :
Chihab Imane Khadidja
Zouache Hanane

Soutenu publiquement le : 30 / 06 / 2026
Devant le jury composé de :
Président : Dre. Badaoui Atika
Examinatice : Dre. Boutouhami Sara
Encadrant : Dr. Nouioua Mourad

Année Universitaire : 2025/2026

Dédicace

À mes chers parents, pour leurs sacrifices, leurs prières, leur amour inconditionnel et leur soutien indéfectible.

À mes quatre sœurs : mon aînée Ghofrane, ainsi que Khouloud, Sabrine et Takwa,pour leur affection, leur soutien moral et leurs encouragements constants.

À ma tante, pour son soutien constant et son aide tout au long de mon parcours universitaire.

À ma belle-famille, pour leurs encouragements.

À toutes les personnes qui m'ont accompagnée, de près ou de loin, dans ce travail.

Imane Khadidja

Dédicace

Louange à Allah, le Premier et le Dernier, l'Apparent et le Caché, qui m'a accordé la force, la patience et la détermination nécessaires pour mener à bien ce modeste travail. Que la paix et les bénédictions soient sur notre Prophète Mohammed (paix et salut sur lui).

***À mon cher père**, En reconnaissance de ton soutien constant, de tes grands sacrifices et de tes précieux conseils qui ont éclairé mon chemin. Je te dédie ce travail en témoignage de mon amour et de ma fidélité, en espérant toujours être à la hauteur de tes attentes.*

***À ma chère mère**, Source de tendresse et de générosité, dont les prières ont été mon soutien à chaque étape. Les mots ne sauraient exprimer toute ma gratitude et mon amour pour toi. Que Dieu te protège et te comble de santé et de bonheur.*

***À mes chers frères : ElHachemi, Othmane et Khaled**, Vous êtes une bénédiction de Dieu et un soutien dont je suis fière tout au long de ma vie. Merci pour votre amour sincère, vos encouragements constants et votre soutien qui ont toujours été pour moi une source de force et de détermination. Que Dieu vous protège et préserve les liens d'amour et de fraternité qui nous unissent.*

***À ma chère sœur Ahlam**, Tu es bien plus qu'une sœur ; tu es une amie et une compagne de route. Merci pour ton affection, ton attention constante et tes paroles encourageantes qui m'ont toujours donné espoir et confiance. Que Dieu te protège et remplisse ta vie de bonheur et de réussite.*

***À tous les membres de ma chère famille**, Petits et grands, merci pour votre amour, votre confiance et votre soutien permanent, ainsi que pour tous les encouragements et l'aide que vous m'avez apportés.*

***À mes honorables enseignants**, Qui ont contribué à ma formation scientifique et académique et m'ont transmis leur savoir et leurs précieux conseils tout au long de mon parcours universitaire. Je leur adresse mes sincères remerciements et ma profonde gratitude.*

***À toutes les personnes qui ont contribué**, de près ou de loin, à la réalisation de ce travail, J'adresse mes plus sincères remerciements et ma reconnaissance, et je vous dédie ce modeste effort en témoignage de mon estime et de ma gratitude.*

Hanane

Remerciement

C’est avec un immense plaisir et une profonde gratitude que nous dédions ces quelques lignes à tous ceux qui ont contribué, de près ou de loin, à la réalisation et à l’aboutissement de ce travail. En premier lieu, nous remercions Dieu le Tout-Puissant de nous avoir donné la volonté, le courage et la santé nécessaires pour mener à bien ce projet, et de nous avoir guidés sur le chemin du savoir.

Nos remerciements s’adressent également aux membres du jury pour l’intérêt qu’ils ont porté à notre travail de fin d’études et pour avoir accepté de l’évaluer et de l’enrichir par leurs remarques avisées.

Une mention spéciale et des remerciements chaleureux sont adressés à l’équipe de l’incubateur d’entreprises de l’université Enfin, nous tenons à remercier chaleureusement tous les enseignants et le personnel administratif du département d’Informatique de l’Université de Bordj Bou Arréridj, ainsi que nos familles et amis qui nous ont toujours encouragés.

Résumé

Cette thése visait à développer un système intelligent et innovant pour faciliter le diagnostic précoce de la maladie d'Alzheimer, en s'appuyant sur l'imagerie par résonance magnétique (**IRM**) et des techniques d'apprentissage profond. Pour ce faire, plusieurs modèles de réseaux neuronaux convolutifs (**CNN**), tels que **Xception**, **MobileNet** et **EfficientNet**, ont été utilisés afin d'extraire des caractéristiques riches et subtiles des images. Ces données ont ensuite été intégrées à des algorithmes d'apprentissage automatique classiques, tels que **SVM**, **KNN** et **Random Forest**, pour classifier les stades d'évolution de la maladie.

Les résultats expérimentaux ont démontré que la combinaison de réseaux de neurones convolutifs et d'algorithmes d'apprentissage automatique permet d'obtenir une précision optimale de la classification des maladies. Le modèle hybride (**Xception + SVM**) a atteint la plus grande généralisabilité et la meilleure précision diagnostique. Afin de garantir la fiabilité des résultats médicaux et d'instaurer un climat de confiance avec les cliniciens, Méthodes d'explicabilité de l'IA, telles que **Grad-CAM** et **SHAP**, ont été intégrées pour fournir une analyse visuelle et systématique qui clarifie le processus de décision du modèle.

De plus, la recherche a abouti à la conception d'une plateforme médicale intégrée utilisant le framework **Flask**. Cette plateforme permet aux utilisateurs de gérer intelligemment les données des patients, de télécharger des images **IRM**, d'exécuter des algorithmes de prédiction, de consulter les résultats de classification et d'interagir avec les interprétations visuelles du modèle.

Mots-clés : Maladie d'Alzheimer, imagerie par résonance magnétique (IRM), Apprentissage Profond, Réseaux de Neurones Profonds, Apprentissage Automatique, classification médicale, Intelligence Artificielle Explicable, Grad-CAM, SHAP.

Abstract

This thesis aimed to develop an intelligent and innovative system to facilitate the early diagnosis of Alzheimer's disease, using magnetic resonance imaging (**MRI**) and deep learning techniques. To achieve this, several convolutional neural network (**CNN**) models, such as **Xception**, **MobileNet**, and **EfficientNet**, were used to extract rich and subtle features from the images. This data was then integrated with classical machine learning algorithms, such as **SVM**, **KNN**, and **Random Forest**, to classify the stages of disease progression.

The experimental results demonstrated that combining convolutional neural networks and machine learning algorithms provides optimal accuracy in disease classification. The hybrid model (**Xception + SVM**) achieved the greatest generalizability and the best diagnostic accuracy. To ensure the reliability of medical results and foster trust with clinicians, "Explainable AI methods, such as **Grad-CAM** and **SHAP**, have been integrated to provide a visual and systematic analysis that clarifies the model's decision-making process.

Furthermore, the research has led to the design of an integrated medical platform using the **Flask** framework. This platform allows users to intelligently manage patient data, upload **MRI** images, run prediction algorithms, view classification results, and interact with the model's visual interpretations.

Keywords : Alzheimer's disease, Magnetic Resonance Imaging (MRI), Deep Learning, Deep Neural Networks, Machine Learning, Medical Classification, Explainable Artificial Intelligence (XAI), Grad-CAM, SHAP.

ملخص

هدفت هذه الأطروحة إلى تطوير نظام ذكي ومبتكر لتسهيل التشخيص المبكر لمرض الزهايمر، باستخدام التصوير بالرنين المغناطيسي (MRI) وتقنيات التعلم العميق. ولتحقيق ذلك، استُخدمت عدة نماذج من الشبكات العصبية الالتفافية (CNN)، مثل Xception و MobileNet و EfficientNet، لاستخلاص خصائص دقيقة وغنية من الصور. ثم دُمجت هذه البيانات مع خوارزميات التعلم الآلي التقليدية، مثل SVM و KNN و Random Forest، لتصنيف مراحل تطور المرض.

أظهرت النتائج التجريبية أن الجمع بين الشبكات العصبية الالتفافية وخوارزميات التعلم الآلي يوفر دقة مثلى في تصنيف المرض. وقد حقق النموذج الهجين (Xception + SVM) أعلى مستوى من التعميم وأفضل دقة تشخيصية. ولضمان موثوقية النتائج الطبية وتعزيز ثقة الأطباء، دُمجت تقنيات تفسير الذكاء الاصطناعي القابل للتفسير مثل Grad – CAM و SHAP، لتوفير تحليل بصري ومنهجي يوضح عملية اتخاذ القرار في النموذج.

علاوة على ذلك، أدى البحث إلى تصميم منصة طبية متكاملة باستخدام إطار عمل Flask. تتيح هذه المنصة للمستخدمين إدارة بيانات المرضى بكفاءة، وتحميل صور الرنين المغناطيسي، وتشغيل خوارزميات التنبؤ، وعرض نتائج التصنيف، والتفاعل مع التفسيرات المرئية للنموذج.

الكلمات المفتاحية: مرض الزهايمر، التصوير بالرنين المغناطيسي (MRI)، التعلم العميق، الشبكات العصبية العميقة، التعلم الآلي، التصنيف الطبي، الذكاء الاصطناعي القابل للتفسير، Grad – CAM، SHAP.

Table des Matières

Liste des Figures **xiii**

Liste des Tableaux **xvi**

Introduction Générale **1**

1 Contexte Général et Etat de l'Art **4**

1.1 Introduction 4

1.2 Définition de la maladie d'Alzheimer 5

1.3 Causes biologiques et facteurs de risque de la maladie d'Alzheimer 5

1.3.1 Causes biologiques 5

1.3.2 Facteurs de risque 6

1.4 Symptômes et stades de la maladie d'Alzheimer 6

1.4.1 Symptômes précoces (Stades 1 à 3) 6

1.4.2 Symptômes avancés (Stades 4 à 7) 6

1.5 Méthodes de diagnostic de la maladie d'Alzheimer 7

1.6 Imagerie par Résonance Magnétique (IRM) 7

1.6.1 Principe de fonctionnement de l'IRM 8

1.6.2 L'IRM dans le diagnostic de la maladie d'Alzheimer 9

1.7 Intelligence artificielle dans le domaine médical 11

1.8 Conclusion 13

2 Conception du Modèle CNN pour la Classification de la Maladie d'Alzheimer et Intégration des Techniques d'Interprétabilité **14**

2.1 Introduction 14

2.2 Machine learning et deep learning 15
2.2.1 L'apprentissage automatique (Machine learning) 15
2.2.2 L'apprentissage profond (Deep learning) 16
2.2.3 L'architecture du système d'aide au diagnostique de la maladie d'Alzheimer 17
2.3 Les réseaux de neurones convolutifs (CNNs) 18
2.3.1 Rôle des CNN comme extracteurs 18
2.3.2 L'architecture d'un Réseaux de Neurones Convolutif (CNN) 18
2.3.3 Le bloc d'extraction des caractéristiques 19
2.3.4 Le bloc de classification 21
2.3.5 Les paramètres d'un CNN 22
2.3.6 Les fonctions d'activation 23
2.3.7 Le transfert d'apprentissage (Transfer Learning) 25
2.4 Les CNNs pour l'analyse des images médicales 26
2.5 Modèles utilisés dans notre projet 27
2.5.1 EfficientNet 27
2.5.2 MobileNet 28
2.5.3 Xception 30
2.6 Algorithmes de classification 31
2.6.1 SVM (Support Vector Machine) 32
2.6.2 KNN (K-Nearest Neighbors) 32
2.6.3 Random Forest 33
2.7 Évaluation des modèles de classification 33
2.8 Manque d'explicabilité des modèles de Deep Learning 34
2.9 L'intelligence artificielle explicable (XAI) 34
2.10 Concept de l'intelligence artificielle explicable (XAI) 35
2.11 Les techniques d'explicabilité utilisées 35
2.11.1 La méthode SHAP 35
2.11.2 La méthode Grad-CAM 39
2.11.3 Synthèse et comparaison des méthodes d'explicabilité 40
2.12 Conclusion 41

3 Méthodologie Proposée **42**

3.1 Introduction 42
3.2 Architecture générale de notre travail 43
3.3 Modélisation Conceptuelle UML 43
3.3.1 Diagramme de cas d'utilisation 43
3.3.2 Diagramme de classes 45
3.3.3 Diagramme de séquence 46
3.4 Présentation du jeu de données 48
3.5 Prétraitement et ingénierie des données 49
3.5.1 Analyse de la distribution et gestion du déséquilibre 49
3.5.2 Redimensionnement et Normalisation 50
3.5.3 Organisation et partitionnement du jeu de données 50
3.6 Paramétrage des CNNs 50
3.6.1 Normalisation et conversion en tenseurs 51
3.7 Intégration des techniques d'explicabilité 51
3.7.1 Grad-CAM 51
3.7.2 SHAP 52
3.8 Conclusion 54

4 Implémentation, Résultats et Discussion **55**
4.1 Introduction 55
4.2 Résultats expérimentaux des modèles CNN 55
4.2.1 EfficientNet +SVM 55
4.2.2 EfficientNet + KNN 57
4.2.3 EfficientNet + Random Forest 57
4.2.4 MobileNet + SVM 58
4.2.5 MobileNet + KNN 58
4.2.6 MobileNet + Random Forest 59
4.2.7 Xception + SVM 60
4.2.8 Xception + KNN 61
4.2.9 Xception + Random Forest 61
4.3 Comparaison des performances 62
4.4 Analyse et discussion 63
4.5 Choix du modèle final 64

4.6 Interprétation des résultats obtenus par les méthodes d'explicabilité 64
4.6.1 Analyse des Resultats de la méthode Grad-CAM 64
4.6.2 Analyse des Resultats de la méthode SHAP 66
4.7 Environnement de travail et technologies utilisées 67
4.7.1 Langage de programmation 67
4.7.2 Environnements de développement 67
4.7.3 Technologies Web 67
4.7.4 Bibliothèques Python 68
4.7.5 Matériel utilisé 69
4.8 Présentation de la plateforme 69
4.8.1 Interfaces du médecin 69
4.8.2 Admin login 75
4.8.3 Admin dashboard 75
4.9 Contributions du projet 77
4.9.1 Limites du système 77
4.10 Conclusion 78

Conclusion générale **79**

Références **80**

Table des figures

1.1 : Les trois axes de coupe pour la visualisation du cerveau par IRM A gauche : coupe axiale, au milieu : coupe sagittale, à droite : coupe frontale 9
1.2 Évolution de la maladie d'Alzheimer à travers les différents stades de démence . 10
1.3 Illustration des quatre principaux phénomènes observés dans la maladie d'Alzheimer : atrophie hippocampique, diminution du volume cérébral, altérations de la substance blanche et hypométabolisme cérébral. 11

2.1 Relation entre Intelligence Artificielle, Machine Learning et Deep Learning . . 16
2.2 Architecture d'un système d'aide au diagnostique de la maladie d'Alzheimer . . 17
2.3 Architecture standard d'un réseau de neurones convolutif (CNN) 18
2.4 Mécanisme calculatoire d'une opération de convolution 19
2.5 Comparaison des opérations de Max-Pooling et d'Average-Pooling 20
2.6 Processus d'aplatissement tensoriel (Flatten Layer) 21
2.7 Structure générale et flux décisionnel d'un réseau CNN 21
2.8 Effet du paramètre Stride sur le balayage spatial 22
2.9 Visualisation des cartes de caractéristiques intermédiaires 23
2.10 Représentation graphique de la fonction d'activation ReLU 24
2.11 Application du mécanisme de Dropout pendant l'entraînement 24
2.12 Architecture de EfficientNet 27
2.13 Architecture de MobileNet 29
2.14 Les convolutions séparables en profondeur 30
2.15 Principe de fonctionnement de SVM 32
2.16 Principe de fonctionnement de KNN 32
2.17 Principe de fonctionnement de Random Forest 33
2.18 L'IA explicable et l'IA classique 35

2.19 Principe théorique de l'explicabilité par SHAP Image sur une IRM cérébrale (Rouge : contribution positive, Bleu : contribution négative)[1]. 38

3.1 Architecture générale du système de diagnostic et d'explicabilité 43
3.2 Diagramme de cas d'utilisation du Médecin Spécialiste 44
3.3 Diagramme de cas d'utilisation de l'Administrateur Système 45
3.4 Diagramme de classes global de la plateforme médicale 46
3.5 Diagramme de séquence du processus d'analyse et d'explicabilité 47
3.6 Diagramme de séquence du processus d'analyse et d'explicabilité 48

4.1 Matrices de confusion et ROC du modèle EfficientNet avec SVM 56
4.2 Matrice de confusion et ROC du modèle EfficientNet + KNN 57
4.3 Matrice de confusion et ROC du modèle EfficientNet + Random Forest 57
4.4 Matrice de confusion et courbes ROC du modèle MobileNet avec SVM 58
4.5 Matrice de confusion et courbes ROC du modèle MobileNet + KNN 58
4.6 Matrice de confusion et courbes ROC du modèle MobileNet + Random Forest . 59
4.7 Matrice de confusion et courbes ROC du modèle Xception avec SVM 60
4.8 Matrice de confusion et courbes ROC du modèle Xception + KNN 61
4.9 Matrice de confusion et courbes ROC du modèle Xception + Random Forest . . 61
4.10 Visualisation explicable par l'approche Grad-CAM. 65
4.11 Analyse multi-classe par les valeurs SHAP. 66
4.12 Interface d'authentification du médecin 69
4.13 signup doctor de la plateforme 70
4.14 Dashboard médecin 71
4.15 Page d'aide et d'exploration du système médical 72
4.16 Ajouter patient 72
4.17 Upload IRM 73
4.18 Run AI Pipeline 73
4.19 Résultat classification Moderate Dementeed 74
4.20 Génération automatique du rapport PDF 74
4.21 Admin login 75
4.22 Dashboard Admin 75
4.23 Gestion des médecins par l'administrateur 76

4.24 Paramètres de l'administrateur 76
4.25 Interface administrateur pour l'envoi des messages aux médecins 77

Liste des tableaux

2.1 Architecture de EfficientNet-B0 28
2.2 Architecture de MobileNet 29
2.3 Architecture du modèle Xception 31
2.4 Correspondance terminologique entre théorie des jeux et apprentissage automatique 36
2.5 Visualisations SHAP et leurs propriétés interprétatives 38
2.6 Comparaison conceptuelle entre Grad-CAM et SHAP dans le système proposé . 41

3.1 Distribution quantitative des examens IRM par classe 49
3.2 Répartition volumétrique des sous-ensembles de données 50

4.1 Comparaison globale des performances des différents modèles 62

Introduction Générale

Contexte général et motivation

La maladie d'Alzheimer est l'une des maladies neurodégénératives les plus fréquentes chez les personnes âgées. Elle entraîne une détérioration progressive des fonctions cognitives, notamment la mémoire, le raisonnement et les capacités d'apprentissage, ce qui affecte considérablement la qualité de vie des patients ainsi que celle de leurs proches. Face à cette évolution progressive, un diagnostic précoce constitue un enjeu majeur afin de ralentir l'évolution de la maladie et d'améliorer la prise en charge médicale.

Parmi les techniques d'imagerie médicale, l'Imagerie par Résonance Magnétique (IRM) occupe une place essentielle, car elle permet de visualiser les modifications structurelles du cerveau associées à la maladie d'Alzheimer. Toutefois, l'analyse manuelle des images IRM reste une tâche complexe, longue et dépend fortement de l'expérience du spécialiste.

Au cours des dernières années, l'intelligence artificielle, et plus particulièrement le Deep Learning, a connu un développement remarquable dans le domaine de l'analyse d'images médicales. Les réseaux de neurones convolutifs (CNN) ont démontré leur capacité à extraire automatiquement des caractéristiques pertinentes et à améliorer les performances des systèmes de classification.

Dans ce mémoire, nous proposons une approche hybride combinant des modèles CNN et des algorithmes de Machine Learning afin d'améliorer la classification des différents stades de la maladie d'Alzheimer à partir d'images IRM. Afin de renforcer la confiance dans les résultats obtenus, nous intégrons également des techniques d'intelligence artificielle explicable (XAI), notamment Grad-CAM et SHAP, qui permettent d'interpréter les décisions du modèle et d'identifier les régions cérébrales ayant influencé la prédiction.

Problématique

Malgré les performances remarquables des modèles de Deep Learning dans le domaine de l'analyse d'images médicales, plusieurs défis limitent encore leur utilisation en pratique clinique.

Parmi ces défis figurent le manque de données médicales annotées, le déséquilibre entre les classes, le risque de surapprentissage (*Overfitting*) ainsi que le caractère peu interprétable des modèles de Deep Learning, souvent considérés comme des *boîtes noires*. Cette absence d'explication réduit la confiance des professionnels de santé dans les décisions produites par ces systèmes.

Par ailleurs, l'utilisation d'approches hybrides combinant les modèles CNN avec des algorithmes de Machine Learning constitue une piste prometteuse, mais leur efficacité dans la classification des stades de la maladie d'Alzheimer reste à évaluer.

Dans ce contexte, la problématique de ce mémoire est la suivante :

> **Comment améliorer la classification des différents stades de la maladie d'Alzheimer à partir d'images IRM en combinant des modèles CNN et des algorithmes de Machine Learning, tout en garantissant une meilleure interprétabilité des résultats grâce aux techniques d'intelligence artificielle explicable ?**

Objectifs du travail

L'objectif principal de ce mémoire est de développer une approche intelligente, performante et interprétable permettant la classification automatique des stades de la maladie d'Alzheimer à partir d'images IRM.

Plus précisément, ce travail vise à :

- Étudier les performances des architectures CNN **EfficientNetB0**, **MobileNetV2** et **Xception**.
- Utiliser ces architectures comme extracteurs de caractéristiques (*Feature Extractors*).
- Combiner les caractéristiques extraites avec plusieurs algorithmes de Machine Learning (**SVM**, **KNN** et **Random Forest**).

- Comparer les performances des différentes approches hybrides.
- Évaluer les modèles à l'aide de plusieurs métriques (*Accuracy*, *Precision*, *Recall*, *F1-score* et *ROC-AUC*).
- Intégrer les techniques d'explicabilité **Grad-CAM** et **SHAP** afin d'expliquer les décisions des modèles.
- Développer une plateforme web médicale intelligente permettant la prédiction, la visualisation des explications et la génération automatique d'un rapport PDF.

Organisation du mémoire

Ce mémoire est organisé en quatre chapitres :

- **Chapitre 1 :** présente les notions générales relatives à la maladie d'Alzheimer, les techniques de diagnostic ainsi que les principales applications de l'intelligence artificielle dans le domaine médical.
- **Chapitre 2 :** décrit les fondements du Deep Learning, les réseaux de neurones convolutifs, les architectures **EfficientNetB0**, **MobileNetV2** et **Xception**, ainsi que les techniques d'intelligence artificielle explicable **Grad-CAM** et **SHAP**.
- **Chapitre 3 :** présente la méthodologie proposée, incluant la préparation des données, l'entraînement des modèles, les approches hybrides basées sur le Machine Learning, les métriques d'évaluation ainsi que la conception UML de la plateforme.
- **Chapitre 4 :** décrit l'implémentation de la plateforme web médicale, présente les résultats expérimentaux obtenus, leur interprétation à l'aide de **Grad-CAM** et **SHAP**, ainsi que les principales interfaces destinées aux médecins et à l'administrateur.

Chapitre 1

Contexte Général et Etat de l'Art

1.1 Introduction

La maladie d'Alzheimer est un trouble neurodégénératif courant, provoquant une détérioration graduelle et irrémédiable des fonctions cognitives, y compris la mémoire et les compétences intellectuelles avancées. Cette pathologie représente un enjeu majeur pour la santé publique à l'échelle mondiale [2].

Ce chapitre est destiné à exposer les concepts fondamentaux concernant la maladie, en commençant par sa définition et son caractère neurodégénératif, avant de traiter les causes et facteurs de risque, qu'ils soient d'origine génétique, environnementale ou associés au mode de vie. On décrit aussi les symptômes et les divers niveaux de la maladie, en séparant les manifestations initiales des formes plus avancées qui conduisent à une diminution de l'autonomie. Ceci contribue à une meilleure compréhension du progrès de la maladie ainsi que de ses impacts sur les aptitudes cognitives et fonctionnelles du patient.

Le chapitre traite également des principales modalités diagnostiques, notamment le diagnostic clinique et celui reposant sur les techniques d'imagerie médicale, en soulignant l'importance de l'imagerie par résonance magnétique (IRM) dans la détection des altérations structurelles associées à la pathologie. En dernier lieu, ce chapitre expose la fonction de l'intelligence artificielle dans le secteur médical, en explicitant ses apports au diagnostic des pathologies neurologiques, notamment par l'intermédiaire de l'analyse des images médicales visant à perfectionner la justesse diagnostique.

1.2 Définition de la maladie d'Alzheimer

La maladie d'Alzheimer est une maladie neurodégénérative chronique caractérisée par une atrophie cérébrale, notamment au niveau du cortex cérébral et de l'hippocampe, entraînant une détérioration progressive de la mémoire et des fonctions cognitives [3].

1.3 Causes biologiques et facteurs de risque de la maladie d'Alzheimer

La maladie d'Alzheimer est le produit de mécanismes complexes engendrant à la fois des causes biologiques et des facteurs de risque. Les causes biologiques se rapportent aux modifications neuropathologiques constatées dans le cerveau. Divers facteurs de risque, qu'ils soient associés à l'âge, à la génétique ou au mode de vie, peuvent faciliter l'émergence et l'évolution de la maladie.[4]

Il est établi que la maladie ne découle pas naturellement de la mort des neurones liée au vieillissement. Effectivement, cette pathologie résulte d'un amassement de protéines atypiques au sein du cerveau.

1.3.1 Causes biologiques

L'accumulation de deux protéines majeures, la bêta-amyloïde et la protéine tau, caractérise la maladie d'Alzheimer, et ces dépôts sont à l'origine des lésions cérébrales propres à cette pathologie. Ces altérations se manifestent principalement sous deux formes neuropathologiques majeures :[5]

- **Les plaques amyloïdes :** elles sont principalement composées de l'accumulation de la protéine bêta-amyloïde à l'extérieur du neurone.
- **Les dégénérescences neurofibrillaires (DNF) :** elles se constituent de l'accumulation de la protéine tau chimiquement altérée (phosphorylée de manière anormale) à l'intérieur du neurone.

Ces mécanismes entraînent une altération progressive de la mémoire, des fonctions cognitives et du comportement.

1.3.2 Facteurs de risque

Plusieurs facteurs de risque ont été identifiés dans le développement de la maladie d'Alzheimer : L'âge représente le facteur principal, car la prévalence augmente fortement après 65 ans. Des facteurs génétiques, tels que la présence de l'allèle APOE-04, sont également associés à un risque accru [6].

En outre, des facteurs modifiables liés au mode de vie, comme l'hypertension, le diabète, l'obésité, la sédentarité, le tabagisme, ainsi que l'isolement social, peuvent contribuer au déclin cognitif. La prise en compte de ces facteurs est essentielle pour mettre en place des stratégies de prévention visant à retarder l'apparition de la maladie.

1.4 Symptômes et stades de la maladie d'Alzheimer

Après avoir présenté la maladie, il est important de connaître ses symptômes et son évolution en stades. La maladie se déroule en **7 stades**, du léger au sévère[7].

1.4.1 Symptômes précoces (Stades 1 à 3)

Au stade précoce de la maladie d'Alzheimer, les symptômes sont souvent discrets et peuvent être confondus avec un vieillissement normal. Les premiers signes apparaissent généralement sous forme de troubles de la mémoire récente, notamment une difficulté à retenir de nouvelles informations et des oublis fréquents dans la vie quotidienne. Le patient peut également présenter une baisse de l'attention, une légère désorientation temporelle, ainsi que des troubles du langage tels que la recherche de mots.

À ce stade, la personne conserve en général une autonomie partielle, bien que les troubles cognitifs commencent progressivement à impacter sa qualité de vie, correspondant aux premiers niveaux allant de : (1) l'absence de déficience cognitive au (2) déclin cognitif très léger, puis au (3) déficit cognitif léger.[8]

1.4.2 Symptômes avancés (Stades 4 à 7)

Aux stades avancés, les troubles cognitifs deviennent sévères et entraînent une perte importante d'autonomie. Le patient peut présenter une désorientation spatio-temporelle marquée, des

difficultés majeures de communication et une incapacité à reconnaître les proches.

Des troubles comportementaux peuvent également apparaître, tels que l'agitation, l'anxiété ou l'apathie. Dans les phases les plus sévères, la personne dépend entièrement d'une assistance pour les activités quotidiennes, ce qui nécessite une prise en charge médicale et sociale adaptée, correspondant aux niveaux allant de : (4) déclin cognitif modéré, (5) déclin cognitif relativement sévère, (6) déclin cognitif sévère, jusqu'au (7) déclin cognitif très sévère.[8]

1.5 Méthodes de diagnostic de la maladie d'Alzheimer

Le diagnostic de la maladie d'Alzheimer repose sur une approche multidimensionnelle combinant l'évaluation clinique, les tests neuropsychologiques et les techniques d'imagerie médicale. À ce jour, il n'existe pas de méthode unique permettant de confirmer la maladie de manière définitive en pratique courante. Cependant, l'association de plusieurs examens permet d'augmenter significativement la précision diagnostique, notamment dans les stades précoces.

- **Une évaluation clinique :** examen des capacités cognitives et du comportement à l'aide de tests spécifiques.
- **L'IRM (Imagerie par Résonance Magnétique) :** observation d'une diminution du volume de l'hippocampe et d'autres régions cérébrales.
- **La tomographie par émission de positrons (TEP) :** utilisant du glucose pour observer le métabolisme du cerveau.
- **L'analyse du LCR (Liquide Céphalo-Rachidien) :** dosage de la protéine amyloïde et de la protéine tau comme marqueurs de la pathologie.

Dans la section suivante, nous allons présenter la méthode de diagnostic la plus utilisée pour la détection de la maladie d'Alzheimer, qui est l'IRM. [9]

1.6 Imagerie par Résonance Magnétique (IRM)

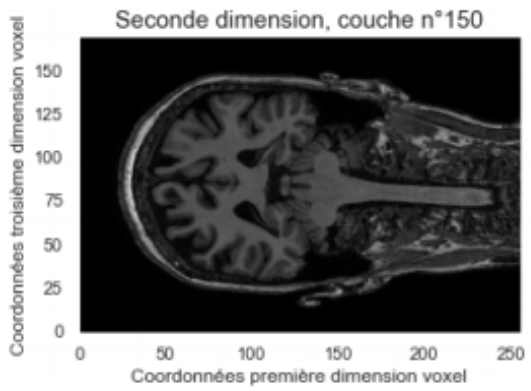
L'Imagerie par Résonance Magnétique (IRM) est une technique d'imagerie médicale non invasive permettant d'obtenir des images détaillées des structures internes du cerveau. Elle est largement utilisée dans le diagnostic des maladies neurologiques, notamment la maladie d'Alzheimer , car elle permet d'observer l'atrophie cérébrale et les anomalies structurelles associées à cette pathologie.[10]

1.6.1 Principe de fonctionnement de l'IRM

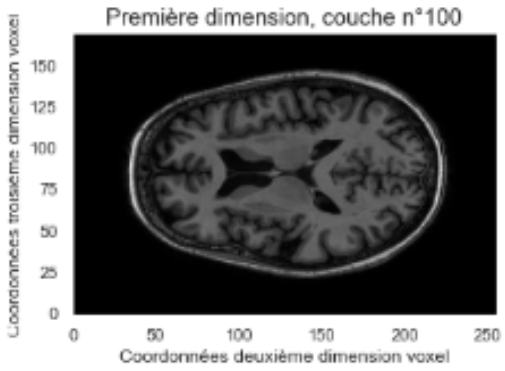
L'Imagerie par Résonance Magnétique (IRM) s'appuie sur les caractéristiques des noyaux d'hydrogène contenus dans les tissus biologiques. Quand un patient est exposé à un champ magnétique fort, les noyaux se positionnent en accord avec la direction du champ. Par la suite, une impulsion de radiofréquence est appliquée afin de troubler cet agencement. Quand l'impulsion cesse, les noyaux regagnent leur état d'équilibre et émettent un signal électromagnétique qui est capté par l'appareil de scanner.

Grâce à l'utilisation de gradients magnétiques, le scanner est capable de déterminer avec précision la source de chaque signal. Cela permet une reconstruction d'images détaillées en 2D et 3D des structures internes du corps humain et du cerveau. L'IRM est donc capable de différencier divers types de tissus, d'évaluer la connectivité neuronale (DTI) et de quantifier l'activité cérébrale (IRM fonctionnelle).

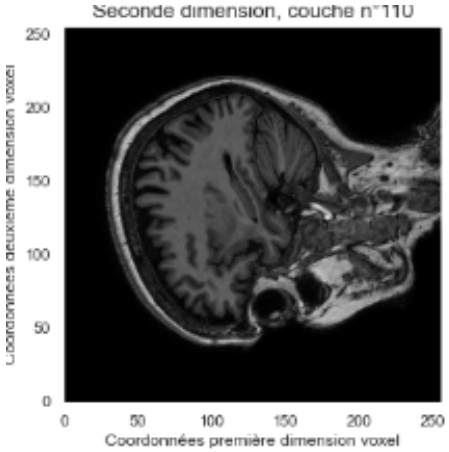
La Figure 1.1 illustre les principaux plans de coupe utilisés en imagerie par résonance magnétique pour la visualisation du cerveau. Elle présente les trois axes fondamentaux : axial, sagittal et frontal.



(a) Schéma IRM 1



(b) Schéma IRM 2



(c) Schéma IRM 3

FIGURE 1.1 – : Les trois axes de coupe pour la visualisation du cerveau par IRM A gauche : coupe axiale, au milieu : coupe sagittale, à droite : coupe frontale

1.6.2 L'IRM dans le diagnostic de la maladie d'Alzheimer

L'outil principal pour diagnostiquer la maladie d'Alzheimer est l'imagerie par résonance magnétique (IRM). L'IRM est un outil efficace pour identifier diverses anomalies typiques de la maladie d'Alzheimer. La Figure 1.2 présente des images d'IRM illustrant l'évolution de la maladie d'Alzheimer à travers différents stades de démence.[2]

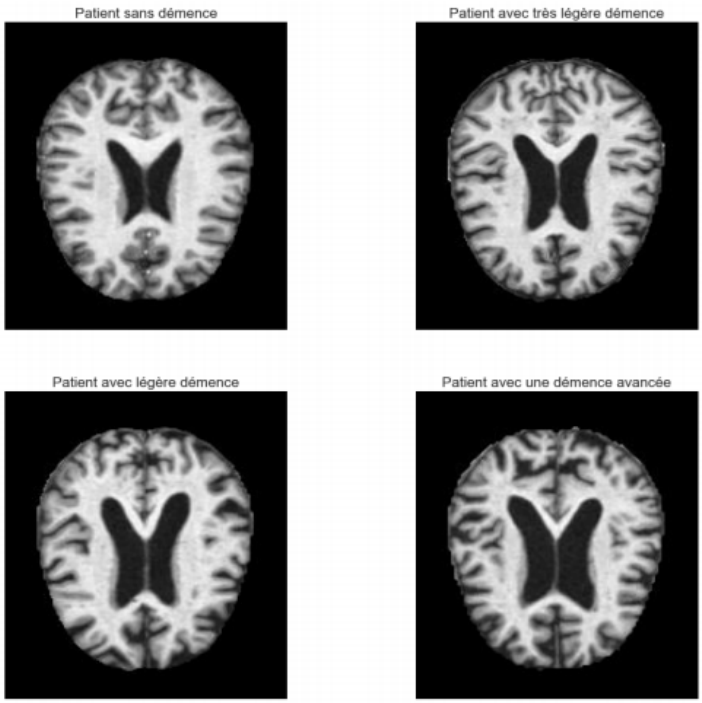
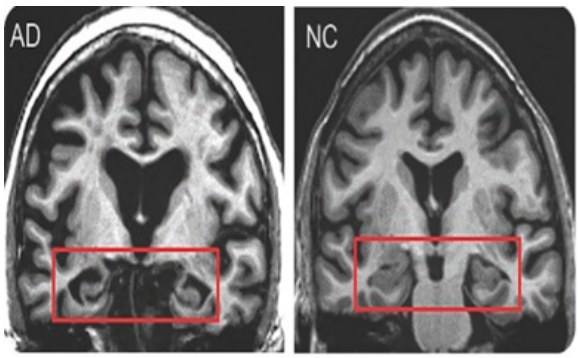


FIGURE 1.2 – Évolution de la maladie d'Alzheimer à travers les différents stades de démence

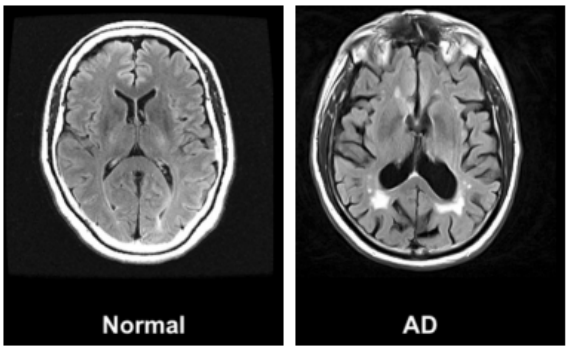
Comme le montre la Figure 1.2, l'évolution de la maladie d'Alzheimer est associée à de nombreuses altérations structurales et fonctionnelles du cerveau. Ces changements, visibles sur l'IRM, permettent d'identifier des biomarqueurs caractéristiques de la maladie et de suivre son évolution à travers les différents stades de démence. Voici les principales anomalies mises en évidence :

1. **Atrophie hippocampique :** diminution du volume, marqueur sûr de la maladie.
2. **Diminution du volume cérébral :** réduction générale du cortex, principalement dans les lobes temporaux et pariétaux.
3. **Altérations de la substance blanche :** détérioration des connexions neuronales, évaluée par IRM en tenseur de diffusion (DTI).
4. **Hypométabolisme cérébral :** baisse de l'activité dans certaines zones, identifiée grâce à l'IRM fonctionnelle (IRMf).

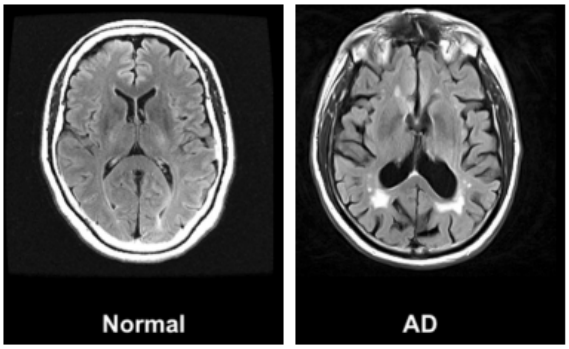
La figure suivante présente une illustration des principales anomalies précédemment décrites associées à la maladie d'Alzheimer.



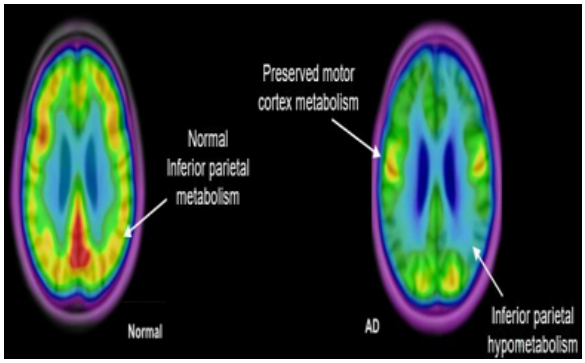
(a) Atrophie hippocampique. Source : Figure 1 de McKeith *et al.*, *Neurology*, 2017 [11].



(b) Diminution du volume cérébral. Source : Figure 1 de Chandra *et al.*, *Journal of Neurology*, 2019 [12].



(c) Altérations de la substance blanche. Source : Figure 2 de Chandra *et al.*, *Journal of Neurology*, 2019 [12].



(d) Hypométabolisme cérébral. Source : image du National Institute on Aging (NIH), diffusée via Wikimedia Commons [13].

FIGURE 1.3 – Illustration des quatre principaux phénomènes observés dans la maladie d'Alzheimer : atrophie hippocampique, diminution du volume cérébral, altérations de la substance blanche et hypométabolisme cérébral.

1.7 Intelligence artificielle dans le domaine médical

L'intelligence artificielle (IA) occupe une place croissante dans le domaine médical, en particulier grâce aux avancées récentes du machine learning et du deep learning. Ces approches permettent d'analyser de grandes quantités de données cliniques et d'imagerie afin d'améliorer la détection, le diagnostic et le suivi de nombreuses pathologies [14].

L'intelligence artificielle a considérablement contribué à l'amélioration du diagnostic des maladies neurologiques, en particulier grâce à sa capacité à analyser des données complexes et volumineuses telles que les images IRM, les scanners (CT), les signaux EEG et les données cliniques [15].

Pour le diagnostic de la maladie d'Alzheimer, une variété de données biologiques et cliniques peut être exploitée, notamment :

1. **Images médicales :** analyse approfondie des IRM et des TEP pour identifier précocement les altérations neurodégénératives, en particulier la diminution volumique de certaines structures cérébrales.
2. **Les données fonctionnelles et comportementales :**les changements dans les activités quotidiennes, la perte d'autonomie et des troubles comme l'anxiété ou les perturbations du sommeil, permettant de détecter précocement un déclin cognitif.

Les approches basées sur le machine learning et le deep learning permettent de détecter des anomalies parfois difficiles à identifier par une interprétation humaine, notamment dans les stades précoces de certaines pathologies. Ces approches sont exploités pour des différents objectifs tels que [2] :

1. **Classification des patients :** distinction efficace entre sujets sains, patients avec troubles cognitifs légers (MCI) et patients atteints de d'un maladie tel que l'Alzheimer.
2. **Détection automatique des anomalies :** segmentation précise des structures cérébrales afin d'identifier la dégénérescence neuronale et les zones atteintes.
3. **Quantification volumétrique :** estimation fiable des volumes de l'hippocampe, du cortex et d'autres régions clés touchées par la maladie.

Afin de réaliser ces objectifs, il existe une variété de méthodes de machine learning et de deep learning, telles que [16] :

1. **les réseaux de neurones convolutifs (CNN) :** Les réseaux de neurones convolutifs (CNN) sont devenus un élément central des études de neuroimagerie, notamment dans le domaine de la maladie d'Alzheimer, en raison de leur capacité à apprendre directement des représentations spatiales à partir des données d'IRM.
2. **les réseaux de neurones récurrents (RNN) :** Les réseaux de neurones récurrents (RNN) étendent les capacités de l'apprentissage profond en modélisant les dynamiques temporelles dans les données longitudinales [41]. Dans le cas de la maladie d'Alzheimer, ils sont particulièrement utiles pour l'étude de la progression de la maladie et la prédiction de la conversion du trouble cognitif léger (MCI) vers la maladie d'Alzheimer (AD), car ils permettent de capturer les dépendances séquentielles dans les données d'imagerie ou les mesures cliniques.

1.8 Conclusion

Dans ce chapitre, nous avons présenté des informations complètes sur la maladie d'Alzheimer, depuis sa définition, ses causes et facteurs de risque, jusqu'aux symptômes et aux stades d'évolution, en passant par les méthodes de diagnostic, avec un accent particulier sur l'imagerie par résonance magnétique (IRM) comme outil précis et efficace pour surveiller le cerveau et évaluer les altérations liées à la maladie.

Nous avons également exploré le rôle de l'intelligence artificielle dans le domaine médical, notamment pour l'analyse des images médicales et le soutien à la prise de décision clinique, en insistant sur l'importance des techniques d'interprétabilité pour rendre les résultats des modèles plus transparents et fiables, en identifiant les régions cérébrales les plus influentes dans le processus de diagnostic.

Ainsi, la compréhension de la maladie, de ses symptômes, des techniques d'imagerie médicale, combinée à l'utilisation de l'apprentissage automatique et profond avec des méthodes d'interprétation des modèles, constitue une base solide pour les chercheurs et les médecins afin d'améliorer la précision du diagnostic et de soutenir les décisions cliniques.

Dans le chapitre suivant, nous présenterons la méthodologie adoptée pour la classification des images IRM, avec un focus sur l'intégration des techniques d'interprétabilité pour renforcer la compréhension et la confiance dans les décisions des modèles proposés.

Chapitre 2

Conception du Modèle CNN pour la Classification de la Maladie d'Alzheimer et Intégration des Techniques d'Interprétabilité

2.1 Introduction

Dans le contexte actuel, la classification automatique des images IRM à l'aide de modèles d'apprentissage profond représente une approche prometteuse pour améliorer la détection et le suivi de la maladie d'Alzheimer. Les réseaux de neurones convolutifs (CNN) se distinguent par leur capacité à extraire automatiquement des caractéristiques pertinentes à partir des images cérébrales, ce qui explique leur utilisation croissante dans le domaine de l'imagerie médicale.[2]

Cependant, malgré leurs performances élevées, l'exploitation de ces modèles dans un cadre médical reste limitée par le manque de transparence dans leur processus de décision. En effet, une bonne précision de classification ne suffit pas toujours ; il est également nécessaire de comprendre sur quelles informations le modèle s'appuie afin d'évaluer la cohérence des résultats et de renforcer la confiance des professionnels de santé.

Dans ce contexte, les techniques d'interprétabilité jouent un rôle essentiel, car elles per-

mettent de rendre les prédictions des CNN plus compréhensibles. Elles fournissent des explications visuelles et quantitatives facilitant l'analyse des décisions du modèle et contribuent à améliorer la fiabilité globale du système.

Ainsi, ce chapitre présente les notions fondamentales du Deep Learning, les architectures CNN utilisées, ainsi que les principales techniques d'interprétabilité adoptées dans ce travail, notamment Grad-CAM et SHAP. L'objectif est de proposer une approche de classification à la fois performante et interprétable, adaptée aux exigences du domaine médical.

2.2 Machine learning et deep learning

Le domaine de la traitement d'images en intelligence artificielle (AI) comprend plusieurs branches. Parmi les plus courantes, on trouve la classification des images IRM à l'aide des approches de deep learning.

2.2.1 L'apprentissage automatique (Machine learning)

L'apprentissage automatique, couramment mentionné en anglais « Machine Learning », est une discipline de l'informatique, des sciences des données et de l'intelligence artificielle qui permet aux systèmes d'apprendre et d'évoluer automatiquement à partir des données, sans nécessiter de programmation explicite. Il utilise des algorithmes et des modèles statistiques pour analyser les informations, identifier des motifs (patterns), prédire des résultats et s'améliorer continuellement grâce à l'intégration de nouvelles données [17].

L'apprentissage automatique se divise principalement en : l'apprentissage supervisé (Supervised Learning), l'apprentissage non supervisé (Unsupervised Learning) et l'apprentissage par renforcement (Reinforcement Learning). La classification supervisée est une tâche d'apprentissage supervisé dans laquelle un modèle apprend à attribuer une classe (étiquette) à une donnée en se basant sur des exemples déjà étiquetés.

Chaque exemple de la base d'apprentissage contient : (1) les caractéristiques (features) qui représentent les données d'entrée et (2) l'étiquette (label) qui correpond à la classe correcte associée à ces données. L'objectif de la classification supervisée est d'apprendre une fonction capable de prédire la classe inconnues de nouvelles données à partir des ses attributs.

2.2.2 L'apprentissage profond (Deep learning)

Au sein du machine learning, l'apprentissage profond est un sous-domaine qui utilise des réseaux neuronaux multicouches inspirés de la structure du cerveau humain. Ces réseaux se distinguent par leur capacité à traiter les données de manière flexible et ajustable. Ils apprennent à partir des données et améliorent continuellement leurs performances [17]. Cette hiérarchie et les relations d'inclusion entre ces concepts sont illustrées par la Figure 2.1 [2].

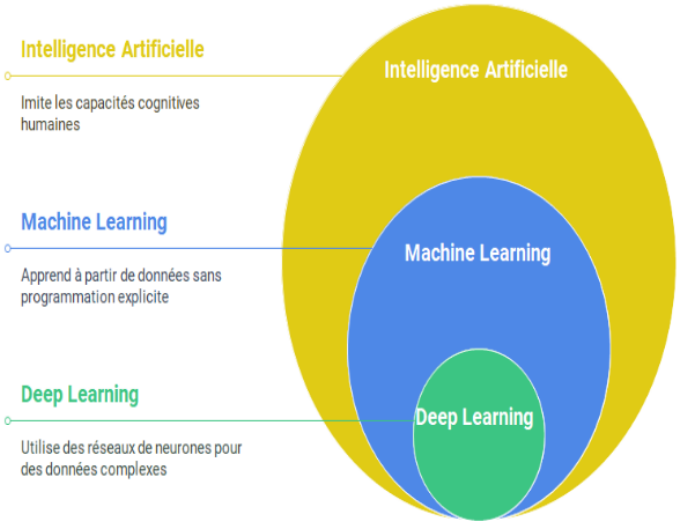


FIGURE 2.1 – Relation entre Intelligence Artificielle, Machine Learning et Deep Learning

L'avantage des méthodes de deep learning par rapport aux méthodes de machine learning réside dans la façon d'extraire les caractéristiques à partir des données. Dans les approches de machine learning, les caractéristiques pertinentes doivent souvent être obtenues manuellement avec des méthodes ou des algorithmes externes. Cela nécessite souvent l'intervention d'un expert du domaine. La qualité de ces caractéristiques influence directement les performances du modèle. En revanche, l'apprentissage profond réalise une extraction automatique et implicite des caractéristiques à partir des données brutes. Dans le cas des réseaux de neurones convolutifs (CNNs), les couches convolutionnelles apprennent progressivement des représentations de plus en plus abstraites et discriminantes. Cela élimine la nécessité d'une phase explicite d'extraction de caractéristiques, ce qui est particulièrement utile pour traiter des données non structurées comme les images ou les signaux [2].

2.2.3 L'architecture du système d'aide au diagnostique de la maladie d'Alzheimer

Dans le cadre de ce mémoire, nous mettons en œuvre une approche de classification supervisée pour la prédiction précoce de la maladie d'Alzheimer à partir d'images d'imagerie par résonance magnétique (IRM). L'approche proposée repose sur une architecture hybride associant le deep learning et le machine learning. Plus précisément, les modèles de deep learning sont utilisés pour l'extraction automatique des caractéristiques à partir des images IRM, tandis que les algorithmes de machine learning assurent la phase de classification supervisée. L'approche proposée intègre également des méthodes d'explicabilité, appelées en anglais explainable artificial intelligence methods, afin de fournir des interprétations des prédictions réalisées et d'améliorer la transparence du processus de classification. Les méthodes d'explicabilité nous permettent d'interpréter les décisions du modèle et d'identifier les régions de l'image ayant le plus influencé la prédiction. L'architecture du système d'aide au diagnostic de la maladie d'Alzheimer est résumée par la figure suivante.

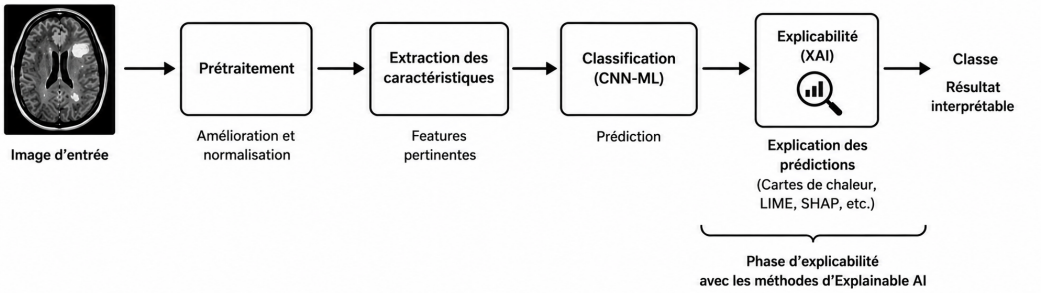


FIGURE 2.2 – Architecture d'un système d'aide au diagnostique de la maladie d'Alzheimer

Dans la suite de ce chapitre, nous allons présenter les différentes phases du processus de diagnostique, en commençant par la présentation des réseaux de neurones convolutifs en général, puis des réseaux de neurones convolutifs préentraînés utilisés dans cette étude pour l'extraction efficace des caractéristiques, ensuite les méthodes de classification et finalement, les méthodes de l'IA explicable utilisées pour l'interprétation des résultats obtenus.

2.3 Les réseaux de neurones convolutifs (CNNs)

Les réseaux de neurones convolutifs (CNNs) ont récemment connu une évolution notable, avec des performances particulièrement remarquables dans le domaine de la vision par ordinateur et du traitement d'images. Ces modèles sont utilisés dans des situations fréquentes de classification ou de détection d'objet, d'image, de typage, etc., parce qu'ils permettent d'extraire eux-mêmes les informations pertinentes des données, sans l'exigence d'une intervention manuelle.

2.3.1 Rôle des CNN comme extracteurs

Les réseaux de neurones convolutifs (CNN) jouent un rôle fondamental en tant qu'extracteurs de caractéristiques dans le traitement des images. Ils permettent de détecter automatiquement des motifs importants tels que les contours, les textures et les structures complexes.

Dans ce projet, nous avons utilisé EfficientNet, MobileNet et Xception comme extracteurs de caractéristiques à partir des images IRM cérébrales. Ces caractéristiques sont ensuite utilisées comme entrée pour des algorithmes de Machine Learning (SVM, KNN, Random Forest) afin de réaliser la classification des stades de la maladie d'Alzheimer.

2.3.2 L'architecture d'un Réseaux de Neurones Convolutif (CNN)

Un réseau de neurones convolutif (CNN) représente une architecture d'apprentissage profond spécifiquement structurée pour traiter des données organisées en grille bidimensionnelle, telles que les images numériques. Sa spécificité réside dans son aptitude à apprendre de façon adaptative et automatisée des hiérarchies de descripteurs spatiaux, allant des textures et motifs géométriques élémentaires jusqu'à des représentations sémantiques hautement complexes [18].

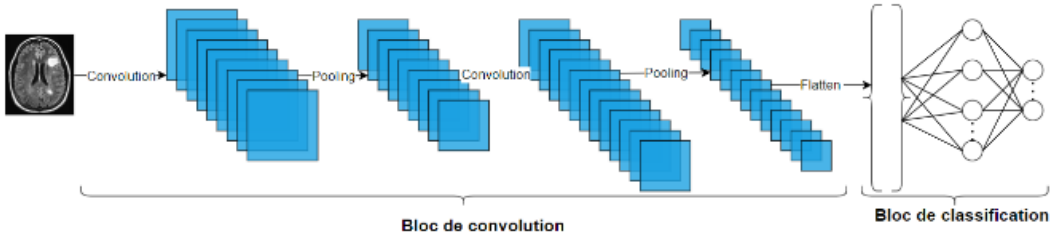


FIGURE 2.3 – Architecture standard d'un réseau de neurones convolutif (CNN)

2.3.3 Le bloc d'extraction des caractéristiques

Dans un réseau de neurones convolutifs (CNN), le bloc d'extraction des caractéristiques est composé de plusieurs types de couches qui travaillent en synergie pour transformer l'image d'entrée en représentations numériques exploitables. Les principales couches de ce bloc sont :

2.3.3.1 Les couches convolutionnelles

La couche de convolution constitue l'une des briques fondamentales des CNN et permet d'extraire des caractéristiques à partir de différentes régions locales des données. Elle est composée d'un ensemble de noyaux tridimensionnels (kernels), chaque neurone agissant comme un noyau. La taille du noyau (longueur × largeur) est conçue artificiellement, par exemple 3×3 ou 5×5, tandis que le nombre de canaux (channels), correspondant au nombre de cartes de caractéristiques produites, influence à la fois la capacité d'extraction des caractéristiques et la complexité calculatoire ; augmenter le nombre de canaux améliore l'extraction des caractéristiques mais augmente également la charge computationnelle.



FIGURE 2.4 – Mécanisme calculatoire d'une opération de convolution

L'opération de convolution consiste à faire glisser le noyau sur l'image d'entrée, à multiplier ses valeurs par celles des pixels correspondants et à sommer les résultats pour obtenir une carte de caractéristiques (feature map). Chaque pixel de cette carte représente ainsi la combinaison des pixels de l'image originale couverts par le noyau à la position correspondante, permettant une extraction hiérarchique des caractéristiques visuelles. [18].

2.3.3.2 Les couches de Pooling

Après avoir extrait les cartes de caractéristiques de la couche de convolution, une couche de pooling est ajoutée pour réduire les dimensions des données et extraire les caractéristiques les plus importantes. Les neurones sont connectés aux domaines réceptifs locaux de la couche précédente sans chevauchement, et chaque région des cartes de caractéristiques est agrégée pour produire une valeur représentative. Les fonctions de pooling les plus utilisées sont le max pooling et le average pooling, le max pooling sélectionnant la valeur maximale de chaque région pour représenter cette zone et extraire les caractéristiques essentielles, ce qui est particulièrement utile pour l'extraction des caractéristiques de bas niveau



FIGURE 2.5 – Comparaison des opérations de Max-Pooling et d'Average-Pooling

2.3.3.3 Couche d'aplatissement (Flatten layer)

Cette couche permet de convertir les cartes de caractéristiques tridimensionnelles en un vecteur unidimensionnel, afin qu'elles puissent être utilisées par une couche entièrement connectée. Par exemple, une carte de taille $5 \times 5 \times 2$ sera transformée en un vecteur de 50 éléments. Cette étape est indispensable, car les couches denses ne peuvent traiter que des données sous forme de vecteur unidimensionnel

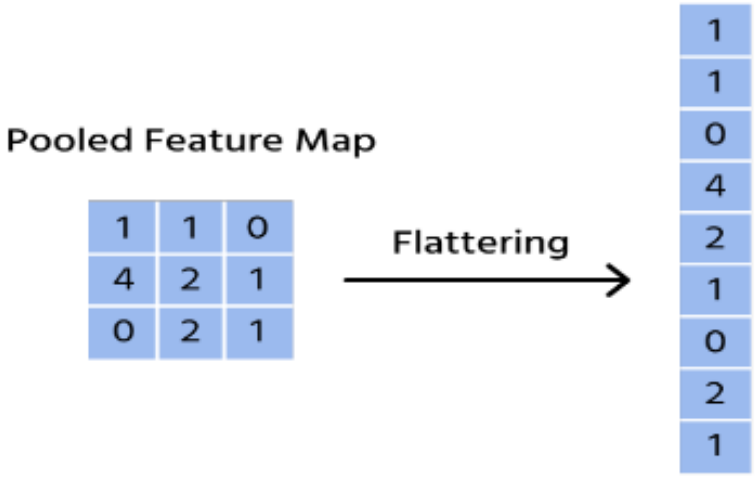


FIGURE 2.6 – Processus d'aplatissement tensoriel (Flatten Layer)

2.3.4 Le bloc de classification

Le bloc de classification reçoit en entrée le vecteur aplati provenant du bloc d'extraction des caractéristiques.

La principale fonction des couches entièrement connectées (EC) est d'apprendre à associer ce vecteur à la classe correcte de l'image. Elles permettent également d'améliorer l'apprentissage et la qualité de la décision en empilant plusieurs couches EC successives [2].

La couche de sortie génère un vecteur dont la taille est égale au nombre de classes. Elle utilise la fonction d'activation Softmax afin d'estimer la probabilité d'appartenance de l'image à chacune des classes [2] .

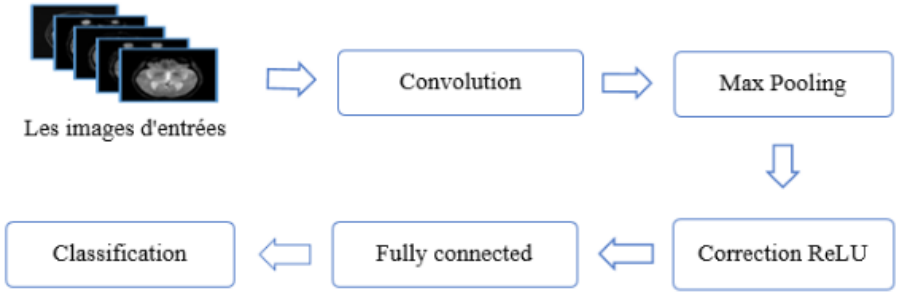


FIGURE 2.7 – Structure générale et flux décisionnel d'un réseau CNN

La figure 2.3 peut être detaillée par la figure 2.7, les images d'entrée constituent la base de données essentielle pour la classification. La couche de convolution (Convolution) permet

d'extraire les caractéristiques importantes des images à l'aide de filtres, tandis que la couche de Max Pooling réduit la taille des images et limite le risque de sur-apprentissage. La fonction d'activation ReLU améliore l'efficacité du traitement, et la couche entièrement connectée (Fully Connected) relie tous les neurones de la couche précédente à ceux de sortie. Enfin, le processus de classification attribue à chaque image la classe qui lui correspond.[3]

2.3.5 Les paramètres d'un CNN

2.3.5.1 Stride

Le paramètre *stride* est le nombre de pixels par lequel le noyau (kernel) se déplace lorsqu'on applique la convolution sur une image. Autrement dit, c'est la "taille du pas" que fait le noyau en parcourant l'image.

- **S**i stride = 1 $\rightarrow$ le noyau avance d'un pixel à chaque fois (déplacement précis, sortie plus grande).
- **S**i stride = 2 $\rightarrow$ le noyau avance de deux pixels à chaque fois (déplacement plus rapide, sortie plus petite).

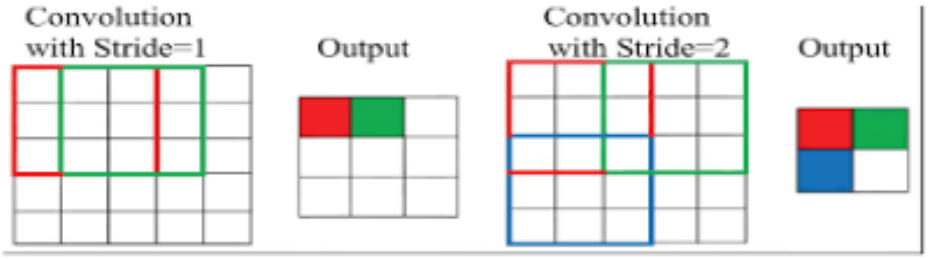


FIGURE 2.8 – Effet du paramètre Stride sur le balayage spatial

On peut en déduire que la taille de la stride influence directement les dimensions de sortie : une stride plus grande réduit les dimensions de sortie en parcourant l'image plus rapidement, tandis qu'une stride plus petite conserve davantage de détails, produisant ainsi des dimensions de sortie plus larges.

2.3.5.2 Les filtres convolutionnels

Le processus de convolution (Convolution) consiste à faire glisser un ensemble de filtres sur l'image d'entrée pour extraire les caractéristiques importantes. La profondeur de la couche

convolutionnelle correspond au nombre de ces filtres, et chaque filtre génère sa propre carte de caractéristiques. Le nombre de filtres q signifie que l'image sera convoluée q fois, produisant q cartes de caractéristiques qui mettent en évidence différents motifs dans l'image. Le choix de la taille et du nombre de filtres dépend de l'application et est souvent prédéfini dans les architectures CNN existantes

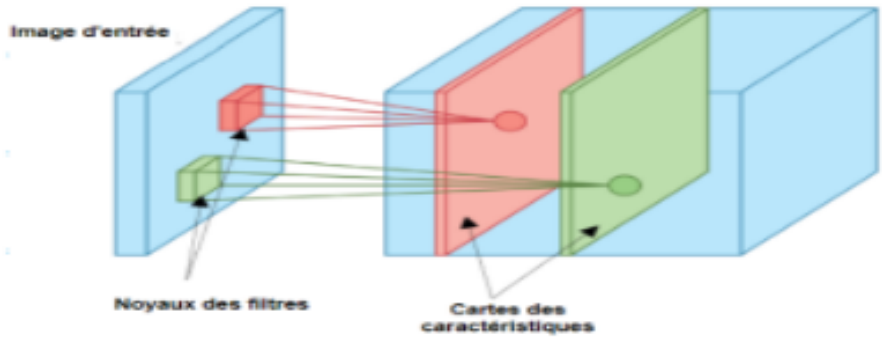


FIGURE 2.9 – Visualisation des cartes de caractéristiques intermédiaires

2.3.5.3 Zero-Padding dans les réseaux convolutionnels

Le *zero-padding* est utilisé pour ajouter des zéros autour de l'image lorsque la taille des filtres ne correspond pas, ce qui évite la réduction progressive de la taille des cartes de caractéristiques (*Feature Maps*) ou permet de conserver la taille de sortie [19].

2.3.6 Les fonctions d'activation

Les fonctions d'activation sont des fonctions mathématiques utilisées pour contrôler la sortie d'un réseau neuronal. Elles décident si un neurone doit être activé ou non, en fonction de son influence sur la précision de la prédiction. Les fonctions d'activation jouent un rôle essentiel dans l'amélioration des performances du modèle, et il en existe plusieurs types, parmi lesquels nous pouvons citer ceux présentés par Nennouche et Atchi [20].

2.3.6.1 ReLU

La fonction *ReLU* transforme chaque sortie d'un neurone en zéro si elle est négative, et la laisse inchangée si elle est positive, ce qui permet au réseau d'apprendre plus efficacement.

$$G(E) = \max(0, E) = \begin{cases} E & \text{si } E \geq 0 \\ 0 & \text{sinon} \end{cases}$$

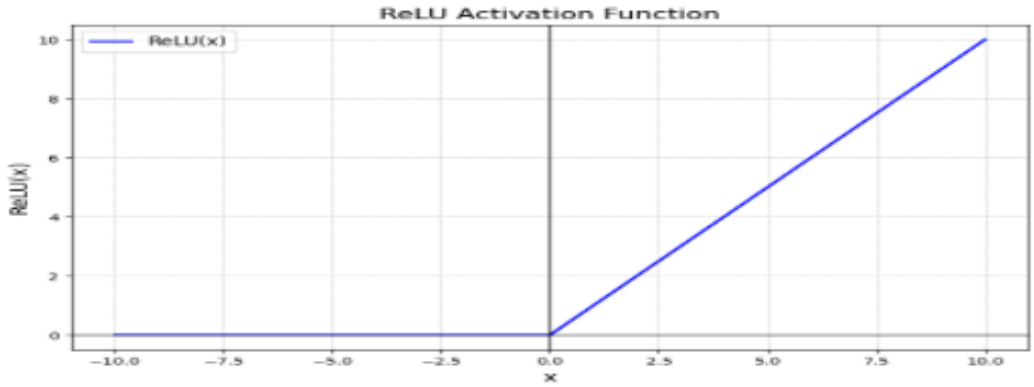


FIGURE 2.10 – Représentation graphique de la fonction d'activation ReLU

2.3.6.2 Softmax

La fonction d'activation *Softmax* est courrament utilisée en sortie, la fonction *Softmax* calcule la probabilité de chaque classe en normalisant les valeurs du vecteur d'entrée.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{2.1}$$

Avec z_i le i-ème élément du vecteur d'entrée et K le nombre de classes.

2.3.6.3 Dropout

Le dropout est une méthode pour renforcer les réseaux neuronaux et éviter le surapprentissage. Pendant l'entraînement, certaines neurones sont désactivées au hasard (leurs sorties deviennent zéro). Cela oblige le réseau à utiliser les autres neurones et à mieux généraliser.(Définition de CNN (avec la partie sur Dropout) [21].

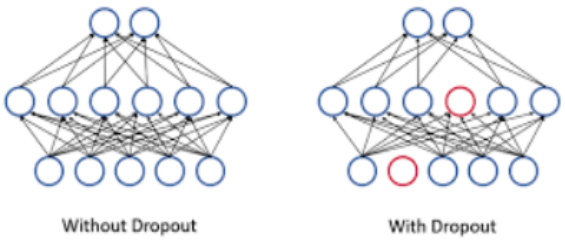


FIGURE 2.11 – Application du mécanisme de Dropout pendant l'entraînement

Après avoir présenté l'architecture générale des CNN, ainsi que leurs couches et paramètres principaux, nous introduirons ensuite des modèles préentraînés avancés, tels qu'EfficientNet et MobileNet, qui constituent le cœur de notre étude expérimentale[2].

2.3.7 Le transfert d'apprentissage (Transfer Learning)

En Deep Learning, le transfert d'apprentissage est une méthode largement répandue qui consiste à exploiter un modèle déjà entraîné sur une vaste base de données (telle qu'ImageNet) comme base pour entreprendre une nouvelle tâche. Plutôt que de former un CNN depuis le début, ce qui exige souvent des volumes considérables de données et un temps de calcul conséquent, il est possible de réemployer les couches de convolution d'un modèle déjà formé, ayant déjà acquis la capacité d'extraire des caractéristiques communes des images. On se limite alors à ré-entraîner uniquement les couches entièrement connectées les plus récentes afin de les ajuster à la nouvelle tâche. Cette méthode offre d'excellentes performances même en utilisant des ensembles de données de taille plutôt réduite[22].

2.3.7.1 Types de transfert d'apprentissage

Il existe trois pratiques adjacentes ou sous-paramètres du transfert d'apprentissage. La distinction entre elles – ainsi que pour le transfert d'apprentissage en général – résulte en grande partie des changements dans la relation entre le domaine source, le domaine cible et les tâches à accomplir :

- **Le transfert inductif :** Ce type se produit lorsque les tâches source et tâches cible sont différentes, peu importe les différences ou similitudes entre le domaine source et le domaine cible (c'est-à-dire les ensembles de données) [23].
- **L'apprentissage non supervisé :** ressemble au transfert inductif, mais aucune donnée n'est étiquetée. Contrairement au transfert inductif, qui est souvent supervisé, il permet d'apprendre à partir de données non annotées. Un exemple courant est la détection de fraude, où le modèle identifie les comportements anormaux à partir de motifs communs dans les transactions [23].
- **Le transfert transductif :** se produit lorsque les tâches source et cible sont les mêmes, mais que les ensembles de données diffèrent. Les données source sont souvent étiquetées tandis que les données cible ne le sont pas. Un exemple courant est l'utilisation d'un

modèle de classification de textes entraîné sur des avis de restaurants pour classer des avis de films [23].

2.4 Les CNNs pour l'analyse des images médicales

Les réseaux de neurones convolutifs (CNN) jouent un rôle essentiel dans le domaine de l'imagerie médicale grâce à leur capacité à analyser automatiquement les images et à extraire des caractéristiques pertinentes.Ils sont utilisés dans plusieurs applications :

- Détection des maladies à partir des images médicales
- Classification des images (IRM, scanner, radiographie)
- Aide au diagnostic médical
- Segmentation des organes et des lésions

Les CNN sont largement utilisés pour la détection et la classification de nombreuses maladies à partir d'images médicales, telles que les tumeurs cérébrales, le cancer du poumon et les maladies cardiovasculaires. Ils permettent d'analyser différents types d'images comme l'IRM, les scanners (CT) et les radiographies, offrant ainsi une aide précieuse pour le diagnostic automatisé et le suivi des patients [24].

Dans le cas de la maladie d'Alzheimer, les CNN sont utilisés pour analyser les images IRM du cerveau afin de détecter les changements structurels liés à la maladie, tels que l'atrophie de l'hippocampe. Ils permettent de classifier les patients selon différents stades (Non Demented, Very Mild, Mild, Moderate), facilitant ainsi un diagnostic précoce et une meilleure prise en charge [25].

Dans ce projet, les CNN (Xception, MobileNet, EfficientNet) sont utilisés pour extraire des caractéristiques à partir des images IRM du cerveau. Ces caractéristiques sont ensuite utilisées comme entrée pour des algorithmes de Machine Learning (SVM, KNN, Random Forest) afin de réaliser la classification des stades de la maladie d'Alzheimer.

La suite de ce chapitre est dstiné à la présentation des ces mzthodes anisi que les technique d'interprétabilité.

2.5 Modèles utilisés dans notre projet

2.5.1 EfficientNet

EfficientNet est un groupe de réseaux de neurones qui apprend à reconnaître des images. Il a été créé par Tan et Le en 2019 [26]. Ce qui rend EfficientNet spécial, c'est qu'il peut apprendre mieux tout en utilisant moins de ressources computationnelles.[27]

2.5.1.1 Architecture

EfficientNet est une famille de réseaux de neurones convolutifs proposée pour améliorer les performances des modèles et réduire le nombre de paramètres. Son architecture repose sur une méthode appelée *"compound scaling"*, qui permet d'augmenter de manière équilibrée la profondeur, la largeur et la résolution du réseau.

Contrairement aux approches traditionnelles, EfficientNet optimise l'utilisation des ressources et maintient une grande précision. Cette architecture est particulièrement adaptée aux tâches de classification d'images complexes [28].

L'architecture géneral de efficientnet est présentée par Figure 2.12 et Table 2.1.

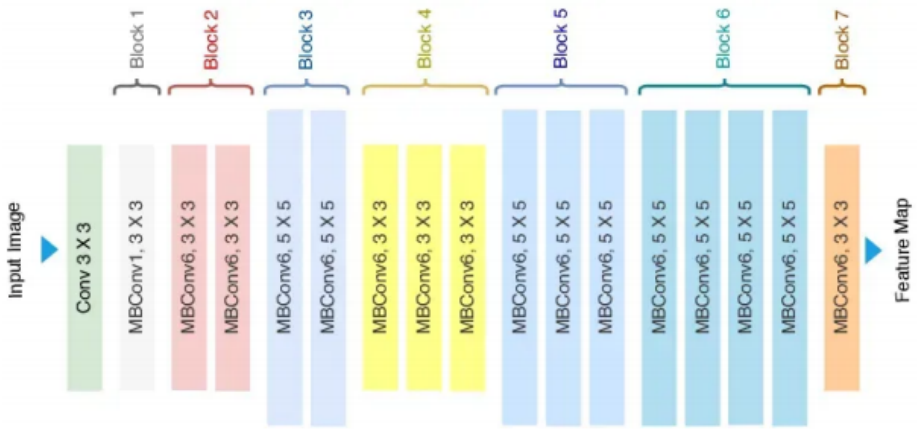


FIGURE 2.12 – Architecture de EfficientNet

Stage	Operator	Resolution	Channels	Layers
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 + Pooling + FC	7×7	1280	1

TABLE 2.1 – Architecture de EfficientNet-B0

2.5.1.2 Avantages

EfficientNet est une série d'architectures de réseaux neuronaux convolutionnels qui ont gagné en popularité grâce à leur efficacité en matière de ressources informatiques et des performances compétitives sur diverses tâches de vision par ordinateur. L'idée principale derrière EfficientNet est de faire évoluer l'architecture du réseau de manière méthodique, la profondeur, la largeur et la résolution, afin d'améliorer les performances sans augmentation importante des coûts computationnels [29]. De plus, sa forte capacité de généralisation réduit les risques de surapprentissage lors de l'utilisation de jeux de données généralement limités et déséquilibrés, caractéristiques de la plupart des contextes de recherche médicale.

2.5.2 MobileNet

MobileNet est une architectures de réseaux de neurones convolutifs (CNN) conçue par Google pour être légère, rapide et consommer peu d'énergie.

2.5.2.1 Architecture

L'architecture de MobileNet repose sur l'utilisation des convolutions séparables en profondeur "*Depthwise Separable Convolution*", qui permettent de réduire considérablement le nombre de paramètres et le coût computationnel par rapport aux CNN classiques.

Cette architecture décompose la convolution classique en deux étapes : une convolution depthwise suivie d'une convolution pointwise. Cette approche accélère le modèle et le rend plus adapté aux environnements à ressources limitées. [30].

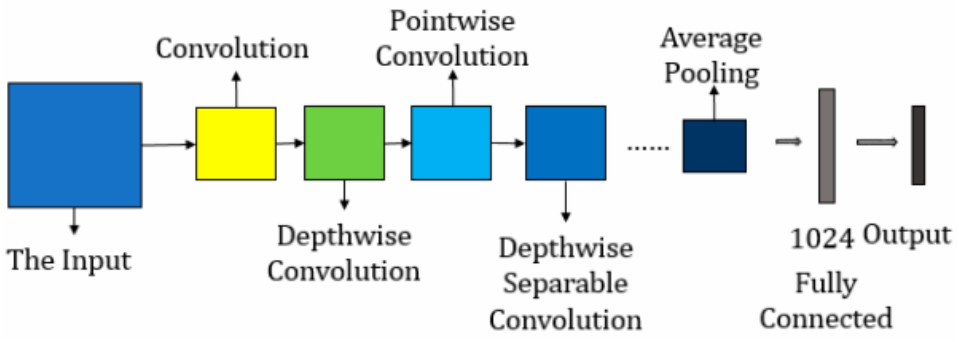


FIGURE 2.13 – Architecture de MobileNet

Stage	Operator	Resolution	Channels	Layers
1	Conv3x3	224×224	32	1
2	Depthwise + Pointwise	112×112	64	1
3	Depthwise + Pointwise	112×112	128	2
4	Depthwise + Pointwise	56×56	256	2
5	Depthwise + Pointwise	28×28	512	6
6	Depthwise + Pointwise	14×14	1024	2
7	Avg Pool + FC	7×7	1024	1

TABLE 2.2 – Architecture de MobileNet

2.5.2.2 Avantages

MobileNet offre pluieurs avantagess pour les applications de données médicales, notamment dans les situations qui requièrent une analyse en temps réel, une faible consommation de ressources informatiques et un déploiement dans des environnements cliniques aux ressources limitées. Malgré sa légèreté, MobileNet offre des performances compétitives en termes de précision, dans les tâches de classification d'images. Il représente un bon compromis entre vitesse d'exécution et efficacité.

Cependant, comparé à des modèles plus complexes comme EfficientNet, il peut présenter une légère baisse de précision dans certains cas, mais reste très utile lorsque les ressources sont

limitées [30].

2.5.3 Xception

Le modèle Xception, proposé par François Chollet, est une architecture de réseau de neurones convolutifs conçue comme une amélioration des réseaux Inception [31].

2.5.3.1 Architecture

L'architecture de Xcepetioon est illusté dans la Figure 2.14 et Table 2.3. Xception (Extreme Inception) repose entièrement sur l'utilisation des convolutions séparables en profondeur (depthwise separable convolutions).

Cette approche consiste à décomposer la convolution classique en deux opérations : une convolution extitdepthwise, appliquée indépendamment à chaque canal, suivie d'une convolution extitpointwise (1×1) permettant de combiner les informations. Elle permet ainsi de réduire le nombre de paramètres et d'améliorer l'efficacité du modèle. Grâce à cette architecture, Xception est capable de capturer des caractéristiques complexes avec une meilleure efficacité computationnelle [32].

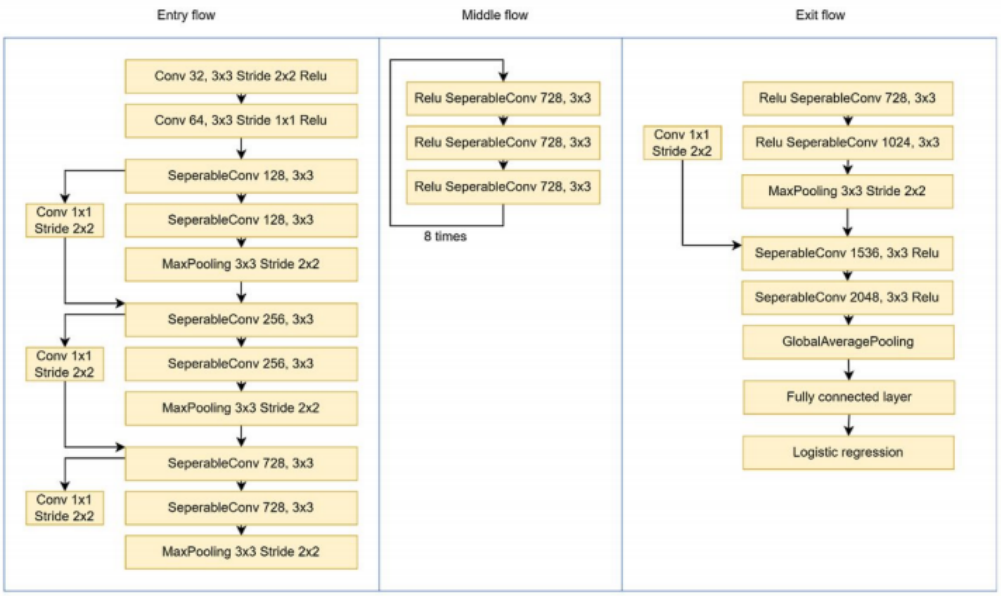


FIGURE 2.14 – Les convolutions séparables en profondeur

Stage	Operator	Resolution	Channels	Layers
1	Conv3x3	299×299	32	1
2	Conv3x3	299×299	64	1
3	Entry Flow (SepConv)	149×149	128	3
4	Entry Flow (SepConv)	75×75	256	3
5	Entry Flow (SepConv)	38×38	728	3
6	Middle Flow (SepConv)	19×19	728	8
7	Exit Flow (SepConv)	10×10	1024	3
8	Global Avg Pool + FC	1×1	2048	1

TABLE 2.3 – Architecture du modèle Xception

2.5.3.2 Avantages

Xception (Extreme Inception) offre des bénéfices significatifs dans l'imagerie médicale et l'analyse de données grâce à son architecture de convolution séparée en profondeur,permettant une extraction de fonctionnalités hautement efficace et précise à partir de données médicalse complexes. La capacité du modèle à dissocier les corrélations spatiales et par canal lui permet de saisir des motifs subtils et complexes dans les images médicales, comme les anomalies tissulaires, les limites des lésions et les variations morphologiques. Ces éléments sont fréquemment déterminants pour un diagnostic médicale précis.[32]

2.6 Algorithmes de classification

Les algorithmes de Machine Learning dédiés à la classification des données permettent d'apprendre des modèles à partir des caractéristiques extraites afin de prédire correctement les classes associées.

Dans ce projet, ces algorithmes sont utilisés après l'extraction des caractéristiques par les réseaux de neurones convolutifs (CNN) pour classer les différents stades de la maladie d'Alzheimer. Dans cette section, nous présentons les algorithmes de classification utilisés dans ce projet.

2.6.1 SVM (Support Vector Machine)

Le Support Vector Machine (SVM) est un algorithme de classification supervisée qui consiste à déterminer un hyperplan optimal séparant les différentes classes [33].

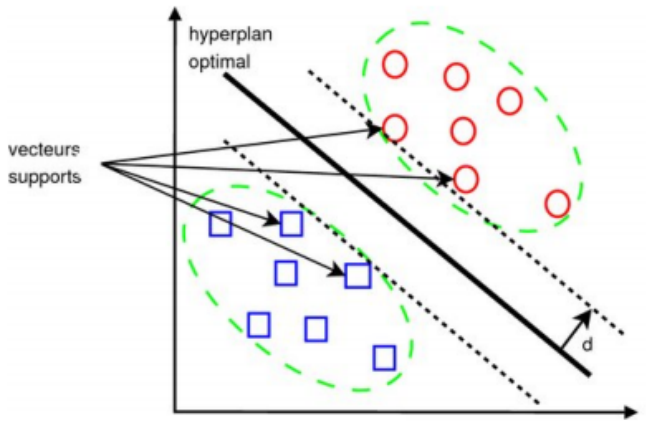


FIGURE 2.15 – Principe de fonctionnement de SVM

Comme illustré dans la Figure 2.15, SVM maximise la marge entre les classes, ce qui lui permet d'offrir une bonne capacité de généralisation.

2.6.2 KNN (K-Nearest Neighbors)

Le K-Plus proche voisin, K-Nearest Neighbors (KNN) est un algorithme simple basé sur la distance entre les données. Il classe un nouvel échantillon en fonction des classes de ses plus proches voisins [34].

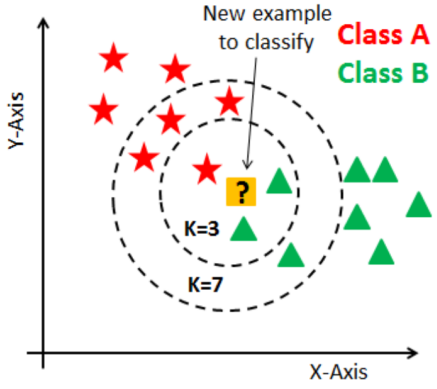


FIGURE 2.16 – Principe de fonctionnement de KNN

KNN présente plusieurs avantages, notamment sa simplicité d'implémentation et l'absence

de phase d'entraînement complexe. Cependant, il comporte aussi certaines limites, car il est sensible au bruit et peut engendrer un coût computationnel élevé lorsqu'il est appliqué à de grandes bases de données.

2.6.3 Random Forest

Comme l'illustre la Figure 2.17, l'algorithme Random Forest repose sur un ensemble d'arbres de décision. Il construit plusieurs arbres et combine leurs résultats afin d'améliorer la précision et la robustesse du modèle.

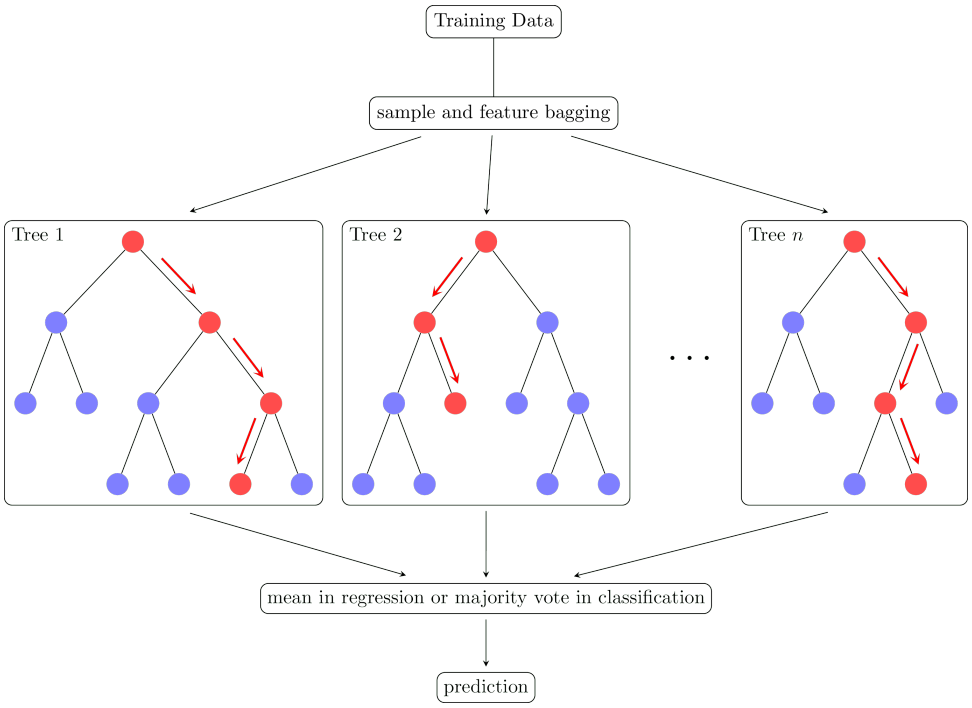


FIGURE 2.17 – Principe de fonctionnement de Random Forest

Random Forest fonctionne en construisant plusieurs arbres de décision à partir de sous-ensembles aléatoires de données, puis en agrégeant leurs résultats à l'aide d'un vote majoritaire afin d'obtenir la prédiction finale. Cet algorithme est robuste face au surapprentissage et efficace pour les problèmes de classification complexes [35].

2.7 Évaluation des modèles de classification

L'évaluation des performances des modèles est une étape essentielle pour mesurer leur efficacité. Plusieurs métriques sont utilisées :

- **Accuracy** : proportion des prédictions correctes
- **Precision** : capacité du modèle à éviter les faux positifs
- **Recall** : capacité à détecter correctement les vrais positifs
- **F1-score** : moyenne harmonique entre précision et rappel

La matrice de confusion est également utilisée pour visualiser les performances du modèle en comparant les prédictions aux valeurs réelles [36].

2.8 Manque d'explicabilité des modèles de Deep Learning

Malgré les performances élevées des modèles de Deep Learning, ces modèles présentent certaines limites, notamment le risque de surapprentissage, la nécessité de disposer de grandes quantités de données et leur complexité computationnelle. Parmi ces limites, le problème de la « boîte noire » constitue l'un des enjeux les plus critiques, car il rend difficile la compréhension des décisions prises par le modèle. Cette absence de transparence pose un problème majeur dans des domaines sensibles comme la médecine, où l'interprétation des résultats est cruciale pour garantir la confiance et la fiabilité des systèmes proposés.

Les modèles de Deep Learning sont souvent considérés comme des "boîtes noires", car il est difficile de comprendre comment ils prennent leurs décisions. Cette absence d'interprétabilité limite leur utilisation dans des domaines sensibles comme la médecine, où la transparence est essentielle [37].

Dans ce projet, nous avons cherché à traiter ce problème en intégrant des techniques d'interprétabilité aux modèles proposés afin de mieux expliquer les prédictions obtenues.

2.9 L'intelligence artificielle explicable (XAI)

L'évolution de l'intelligence artificielle a conduit à la création de certains modèles qui agissent comme des « boîtes noires », ce qui complique la compréhension de leur production de résultats, même du côté des développeurs. Toutefois, la capacité d'expliquer le fonctionnement de ces systèmes est cruciale, car elle assure leur bon déroulement, le respect des réglementations et la possibilité pour les utilisateurs de comprendre et remettre en question les décisions prises.

2.10 Concept de l'intelligence artificielle explicable (XAI)

L'intelligence Artificielle Explicable, en alglais Explainable Artificial Intelligence (XAI),regroupe un ensemble de méthodes et de processus permettant aux utilisateurs de comprendre et de faire confiance aux résultats des modèles d'apprentissage automatique. Elle offre la possibilité de caractériser les modèles d'IA, leurs conséquences prévues et leurs préjugés possibles, tout en estimant leur exactitude, leur justesse, leur transparence et leurs résolutions.

L'explicabilité est indispensable pour les organisations, car elle consolide la confiance lors de l'implémentation des systèmes d'IA et encourage une démarche responsable dans l'évolution de ces technologies [38]. La différence netre un approche classique et une approche explicable est illustrée dans la Figure 2.18.

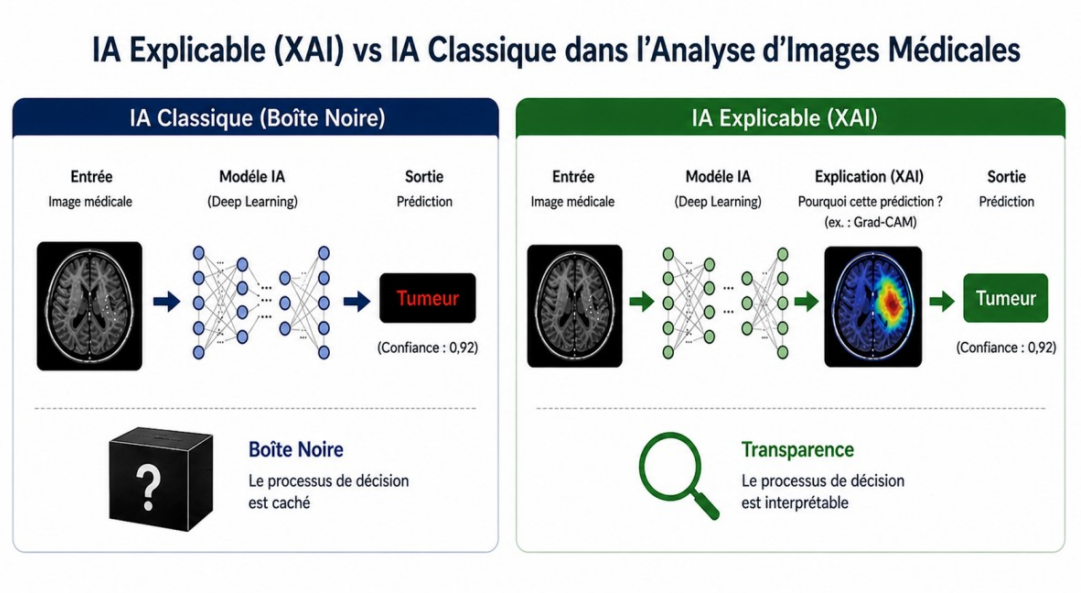


FIGURE 2.18 – L'IA explicable et l'IA classique

2.11 Les techniques d'explicabilité utilisées

2.11.1 La méthode SHAP

La méthode SHAP (SHapley Additive exPlanations) trouve ses racines dans la théorie des jeux coopératifs, plus précisément dans les valeurs de Shapley proposées par Lloyd Shapley en 1953. L'analogie centrale consiste à considérer chaque caractéristique (*feature*) d'un modèle comme un joueur participant à un jeu coopératif dont le paiement correspond à la prédiction

du modèle pour un échantillon donné [39]. Les valeurs de Shapley satisfont quatre propriétés fondamentales :

- **Efficacité :** La somme des contributions de tous les joueurs égale le paiement total.
- **Symétrie :** Deux joueurs apportant la même contribution à toutes les coalitions reçoivent un partage égal.
- **Additivité :** Dans un jeu composé de plusieurs sous-jeux, la contribution d'un joueur au jeu combiné est la somme de ses contributions à chaque sous-jeu.
- **Joueur nul :** Un joueur ne contribuant à aucune coalition reçoit une part de zéro.

La formule mathématique classique de la valeur de Shapley ϕ_i pour un joueur (ou une caractéristique) i est donnée par la formule suivante :

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left(v(S \cup \{i\}) - v(S)\right) \tag{2.2}$$

où :

- $\phi_i(v)$: valeur de Shapley associée à la caractéristique i.
- N : ensemble de toutes les caractéristiques.
- S : sous-ensemble des caractéristiques.
- $v(S)$: valeur de la coalition.
- $v(S \cup \{i\}) - v(S)$: contribution marginale de la caractéristique i.

Cette formule calcule la contribution moyenne de chaque caractéristique à la prédiction finale.

2.11.1.1 Connexion au machine learning

La transition des valeurs de Shapley vers l'analyse SHAP a été formalisée par Lundberg et Lee en 2017. L'équivalence entre les terminologies de la théorie des jeux et du Machine Learning est résumée dans le tableau suivant :

TABLE 2.4 – Correspondance terminologique entre théorie des jeux et apprentissage automatique

Théorie des jeux	**Machine Learning**
Joueurs	Caractéristiques (*features*)
Paiement	Prédiction du modèle pour un échantillon
Contribution marginale	Valeur de la caractéristique pour l'échantillon
Valeur de Shapley du joueur j	Valeur SHAP de la caractéristique j

Une propriété cruciale héritée des valeurs de Shapley est l'additivité : pour un échantillon x, la prédiction du modèle $f(x)$ s'exprime comme la somme de la prédiction moyenne et des valeurs SHAP de toutes ses caractéristiques :

$$f(x) = \phi_0 + \sum_{j=1}^{M} \phi_j \tag{2.3}$$

où ϕ_0 représente la prédiction moyenne du modèle sur l'ensemble des données. Cette décomposition permet d'interpréter chaque valeur SHAP ϕ_j comme la contribution d'une caractéristique à l'écart entre la prédiction individuelle et la prédiction moyenne.[39].

Dans certains contextes de modèles linéaires ou d'approximations spécifiques, la contribution peut être formulée sous la forme :

$$\phi_j = \beta_j (x_j - \mathbb{E}[x_j]) \tag{2.4}$$

2.11.1.2 Calcul des valeurs SHAP

Le calcul exact des valeurs SHAP à partir de la formule théorique est généralement trop coûteux en termes de calcul. Plusieurs méthodes d'approximation ont été développées [39] :

- **TreeSHAP :** Pour les modèles basés sur des arbres (*XGBoost, Random Forest, LightGBM, CatBoost*), permet de calculer exactement les valeurs SHAP en exploitant la structure arborescente du modèle.
- **DeepExplainer / GradientExplainer :** Pour les réseaux de neurones, utilisent respectivement le graphe de calcul ou la différentiabilité du modèle.
- **KernelSHAP :** Méthode *model-agnostic* (indépendante du modèle) applicable à tout modèle supervisé, basée sur un échantillonnage pondéré à partir d'un *background dataset* représentatif (comme pour le *SVM* et le *KNN*).

2.11.1.3 Visualisations et interprétation

Les auteurs décrivent quatre visualisations principales, chacune offrant des perspectives locales (échantillon individuel) ou globales (ensemble du modèle) [39].

TABLE 2.5 – Visualisations SHAP et leurs propriétés interprétatives

Type de graphique	Niveau	Avantage principal	Limitation
Bar plot	Global	Classement des caractéristiques par importance moyenne (unités interprétables)	Aucune indication sur la direction ou la monotonicité de l'effet
Beeswarm plot	Global	Vue d'ensemble de la distribution des valeurs SHAP avec directionnalité via la couleur	Difficulté à extraire précisément la forme des relations
Scatter plot	Global / Local	Détail sur la direction, l'amplitude et la variabilité de l'impact d'une caractéristique	Une seule caractéristique par figure
Waterfall plot	Local	Décomposition exhaustive de la prédiction individuelle	Focus unique sur un seul échantillon

2.11.1.4 Un exemple d'interprétation des prédictions à l'aide de la méthode SHAP

À côté de l'image originale (à gauche), le modèle projette l'explication locale (à droite) sous forme de pixels colorés :

- **Les zones en rouge :** Représentent une contribution positive. Ce sont les régions anatomiques cérébrales qui augmentent la probabilité du modèle à attribuer un stade spécifique de la maladie d'Alzheimer.
- **Les zones en bleu :** Représentent une contribution negative. Ce sont les caractéristiques visuelles qui diminuent cette probabilité et éloignent le diagnostic de ce stade[39].

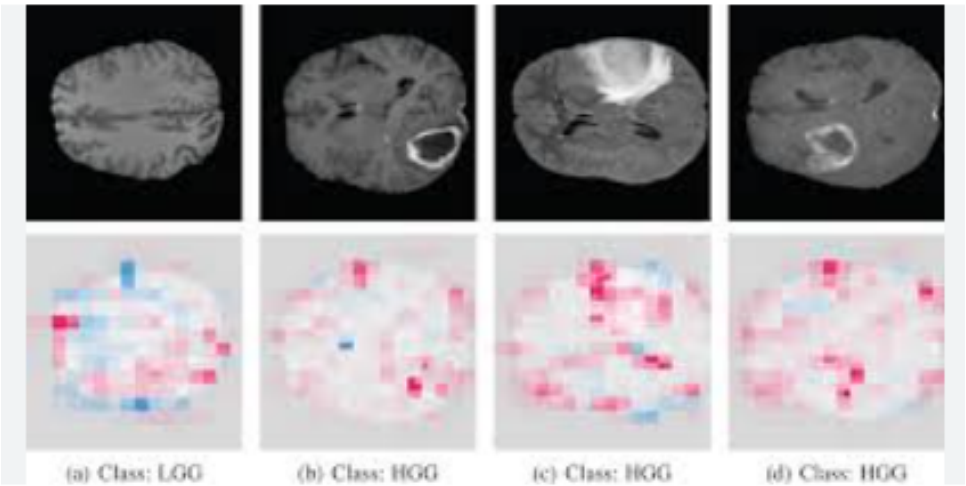


FIGURE 2.19 – Principe théorique de l'explicabilité par SHAP Image sur une IRM cérébrale (Rouge : contribution positive, Bleu : contribution négative)[1].

2.11.2 La méthode Grad-CAM

Grad-CAM (*Gradient-weighted Class Activation Mapping*) est une méthode d'explicabilité qui permet de visualiser les régions importantes d'une image utilisées par les réseaux de neurones convolutifs (CNN) pour prendre une décision. Cette technique exploite les gradients de la classe cible au niveau de la dernière couche convolutionnelle afin de générer une carte de localisation mettant en évidence les zones les plus influentes dans la prédiction [40].

L'un des principaux avantages de Grad-CAM est sa compatibilité avec différents types d'architectures CNN, notamment les modèles de classification d'images, les modèles à sorties structurées, ainsi que ceux utilisés pour les tâches multimodales et l'apprentissage par renforcement, sans nécessiter de modification ou de réentraînement du modèle [40].

2.11.2.1 Fonctionnement général de Grad-CAM

Le fonctionnement de Grad-CAM repose sur un ensemble d'étapes simples qui permettent de transformer les informations internes du réseau de neurones en une carte visuelle explicative :

- **Calcul du score cible :** Le modèle analyse l'image et produit un score pour chaque classe. On choisit le score de la classe spécifique à expliquer.
- **Calcul des gradients :** On calcule les gradients de ce score par rapport aux cartes de caractéristiques (*feature maps*) de la dernière couche convolutionnelle. Cela permet de voir quelles régions influencent la décision.
- **Pondération des canaux :** Les gradients sont moyennés pour chaque canal afin d'obtenir un poids d'importance (α_k^c). Plus le poids est grand, plus le canal est important pour la prédiction.
- **Génération de la heatmap :** On combine toutes les *feature maps* en les multipliant par leurs poids respectifs, puis on les additionne. On applique ensuite une fonction ReLU pour garder seulement les influences positives. Enfin, on redimensionne la carte pour l'adapter à la taille de l'image originale.

2.11.2.2 Formulation mathématique de Grad-CAM

Le calcul de la carte Grad-CAM s'articule autour de deux étapes fondamentales :

1. **Importance des caractéristiques (α_k^c) :** Pour évaluer la sensibilité de chaque carte de caractéristiques A^k de la dernière couche convolutive par rapport à la classe prédite c (ex. Alzheimer), nous calculons le gradient du score brut y^c par rapport à ces activations, moyenné spatialement par *Global Average Pooling* :

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{i,j}^k} \tag{2.5}$$

   Où Z est le nombre total de pixels de la carte (par exemple 14×14). En Python, cela correspond à l'utilisation de `tf.GradientTape()` et `tf.reduce_mean()`.

2. **Combinaison linéaire et filtrage ReLU** : Les poids α_k^c sont ensuite combinés linéairement avec les cartes A^k. L'application de la fonction ReLU permet d'éliminer les gradients négatifs (bruits) pour ne conserver que les éléments visuels qui ont une influence positive sur le diagnostic :

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \tag{2.6}$$

   En code, cette opération est matérialisée par la ligne : `heatmap = np.maximum(heatmap, 0)`.

L'un des principaux avantages de la méthode Grad-CAM réside dans son application directe aux réseaux de neurones convolutifs (CNN) performants, sans nécessiter de modification de leur architecture ni de phase de réentraînement. De plus, Grad-CAM fournit des explications visuelles claires, facilitant l'interprétation des prédictions du modèle et contribuant à la validation du diagnostic automatique.

La figure

2.11.3 Synthèse et comparaison des méthodes d'explicabilité

Afin de synthétiser les différences fondamentales entre les deux approches d'explicabilité (*XAI*) adopté dans cette étude, le Tableau 2.6 présente une comparaison conceptuelle entre Grad-CAM et SHAP selon leur niveau de vision, de calcul et la nature de leurs résultats.

TABLE 2.6 – Comparaison conceptuelle entre Grad-CAM et SHAP dans le système proposé

Caractéristique	**Grad-CAM (Global/Régional)**	**SHAP (Local/Pixel-level)**
Niveau de Vision	Vue macroscopique basée sur les régions d'activation (*Region-level*).	Vue microscopique basée sur la contribution individuelle de chaque pixel (*Pixel-level*).
Niveau de Calcul	Calculé au niveau des couches profondes du CNN (*Last Conv Layer*).	Calculé au niveau des sorties probabilités finales du classifieur (SVM).
Nature du Résultat	Identifie **"Où"** le modèle a concentré son attention visuelle globale.	Identifie **"Quels"** pixels spécifiques soutiennent (Rouge) ou défavorisent (Bleu) la classe.

2.12 Conclusion

Dans ce chapitre, nous avons présenté les différentes étapes de la méthodologie adoptée pour la classification de la maladie d'Alzheimer à partir des images IRM. Nous avons d'abord introduit les notions de base liées au Machine Learning, au Deep Learning ainsi qu'aux réseaux de neurones convolutifs, qui constituent le fondement de notre approche.

Par la suite, nous avons décrit les principaux modèles CNN utilisés, à savoir Xception, MobileNet et EfficientNet, en mettant en évidence leurs caractéristiques et leurs performances dans le traitement des images. Une approche hybride a également été proposée, combinant l'extraction de caractéristiques par les CNN avec des classifieurs traditionnels tels que SVM, kNN et Random Forest, afin d'améliorer la précision de la classification.

Nous avons également intégré des techniques d'interprétabilité, notamment Grad-CAM et SHAP, permettant d'expliquer les décisions des modèles et de renforcer la confiance dans les résultats obtenus. Enfin, une interface graphique a été développée pour faciliter l'utilisation du système et offrir une visualisation claire des prédictions.

Ainsi, ce chapitre a permis de définir une méthodologie complète, combinant performance et interprétabilité, qui sera évaluée et analysée dans le chapitre suivant à travers les résultats expérimentaux.

Chapitre 3

Méthodologie Proposée

3.1 Introduction

Ce chapitre expose de manière détaillée la démarche méthodologique conçue pour la détection précoce et la classification multi-classe des différents stades de la maladie d'Alzheimer à partir d'images de résonance magnétique (IRM).

Pour répondre aux exigences de précision diagnostique et de transparence du secteur médical, notre approche s'appuie sur un pipeline hybride innovant. Ce dernier combine la puissance d'extraction de caractéristiques des réseaux de neurones convolutifs (CNN) avec la robustesse statistique de classifieurs avancés, à savoir les machines à vecteurs de support (SVM), les k plus proches voisins (KNN) et les forêts aléatoires (Random Forest). Un autre apport de ce travail est l'intégration des techniques d'explicabilité, visant à améliorer la transparence et la compréhension des décisions du modèle.

Nous détaillerons dans un premier temps les protocoles de prétraitement, incluant la normalisation et le redimensionnement rigoureux des données issues de la base quantitative exploitée. Dans un second temps, nous présenterons la formalisation conceptuelle et architecturale du système à travers une modélisation UML (Unified Modeling Language) unifiée, structurant les interactions cliniques au sein de la plateforme.

Enfin, une attention particulière sera dédiée à l'intégration des modules d'intelligence artificielle explicable (XAI), notamment Grad-CAM et SHAP. Ces techniques sont indispensables

pour transcrire les prédictions algorithmiques complexes en indicateurs visuels et quantitatifs directement exploitables et vérifiables par les praticiens.

3.2 Architecture générale de notre travail

L'architecture globale de notre solution logicielle est structurée en plusieurs blocs fonctionnels séquentiels, allant de l'acquisition de l'image IRM brute jusqu'à la génération d'un rapport diagnostique transparent. La Figure 3.1 schématise les interconnexions directes entre le pipeline de classification hybride et les moteurs d'explicabilité locale.

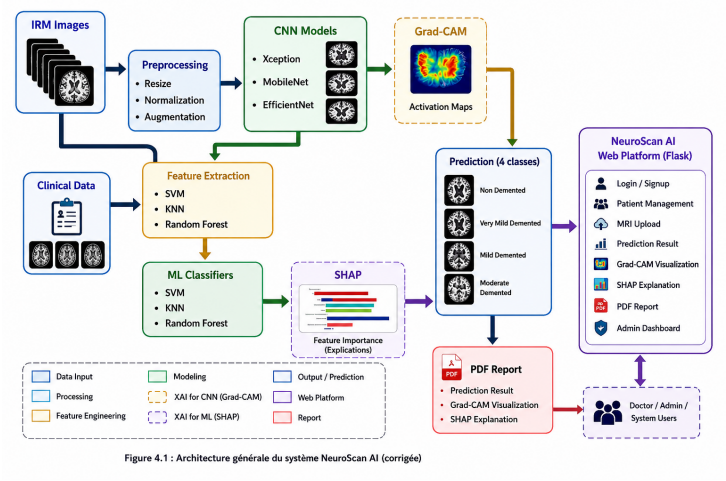


FIGURE 3.1 – Architecture générale du système de diagnostic et d'explicabilité

3.3 Modélisation Conceptuelle UML

La spécification des besoins fonctionnels et l'organisation structurelle de notre plateforme médicale sont formalisées à l'aide du langage de modélisation unifié UML. Cette étape indispensable sécurise la logique logicielle avant son déploiement sur l'infrastructure Cloud.

3.3.1 Diagramme de cas d'utilisation

Le diagramme de cas d'utilisation cadre les frontières du système en identifiant clairement les rôles et privilèges des deux acteurs clés : le Praticien Médical (Neurologue / Radiologue) et l'Administrateur du système.

Le médecin spécialiste interagit avec la plateforme pour charger l'examen IRM d'un patient,

sélectionner le modèle computationnel cible, exécuter l'analyse diagnostique automatisée, puis explorer de manière interactive les cartes thermiques issues de Grad-CAM et les analyses de SHAP. Enfin, il peut valider les résultats cliniques et exporter un rapport officiel au format PDF. L'administrateur, quant à lui, assure la gestion des utilisateurs, la maintenance des modèles et la sécurité des accès.

Les Figures 3.2 et 3.3 illustrent respectivement ces interactions fonctionnelles.

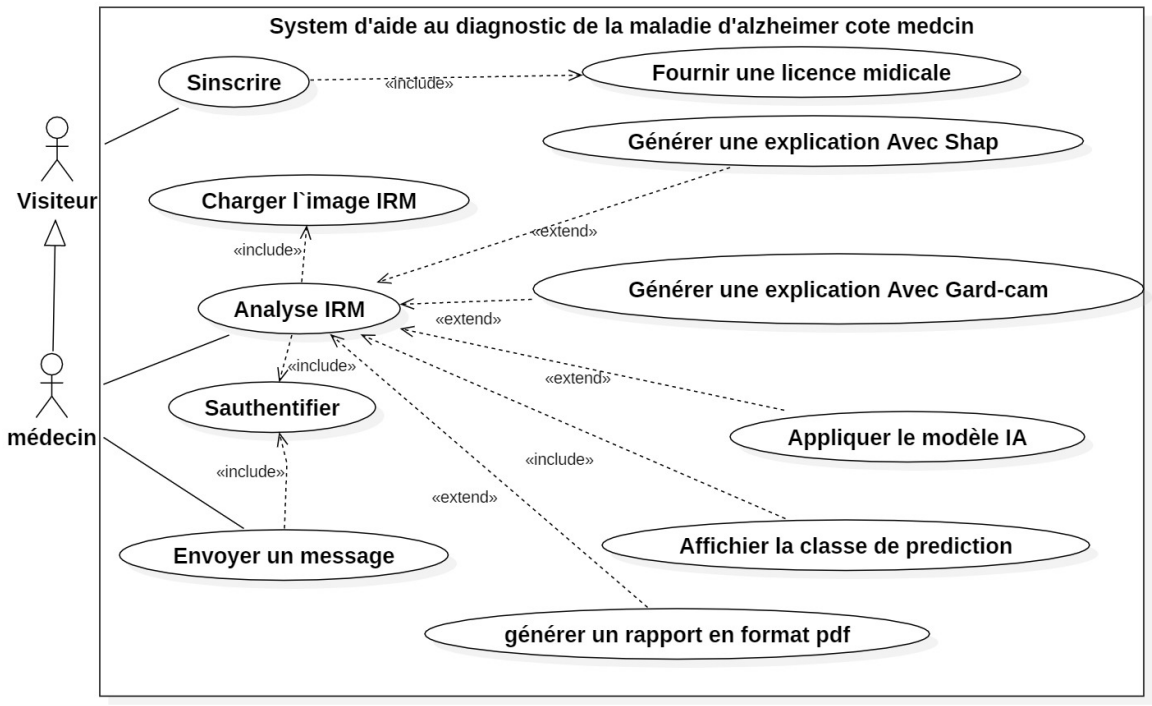


FIGURE 3.2 – Diagramme de cas d'utilisation du Médecin Spécialiste

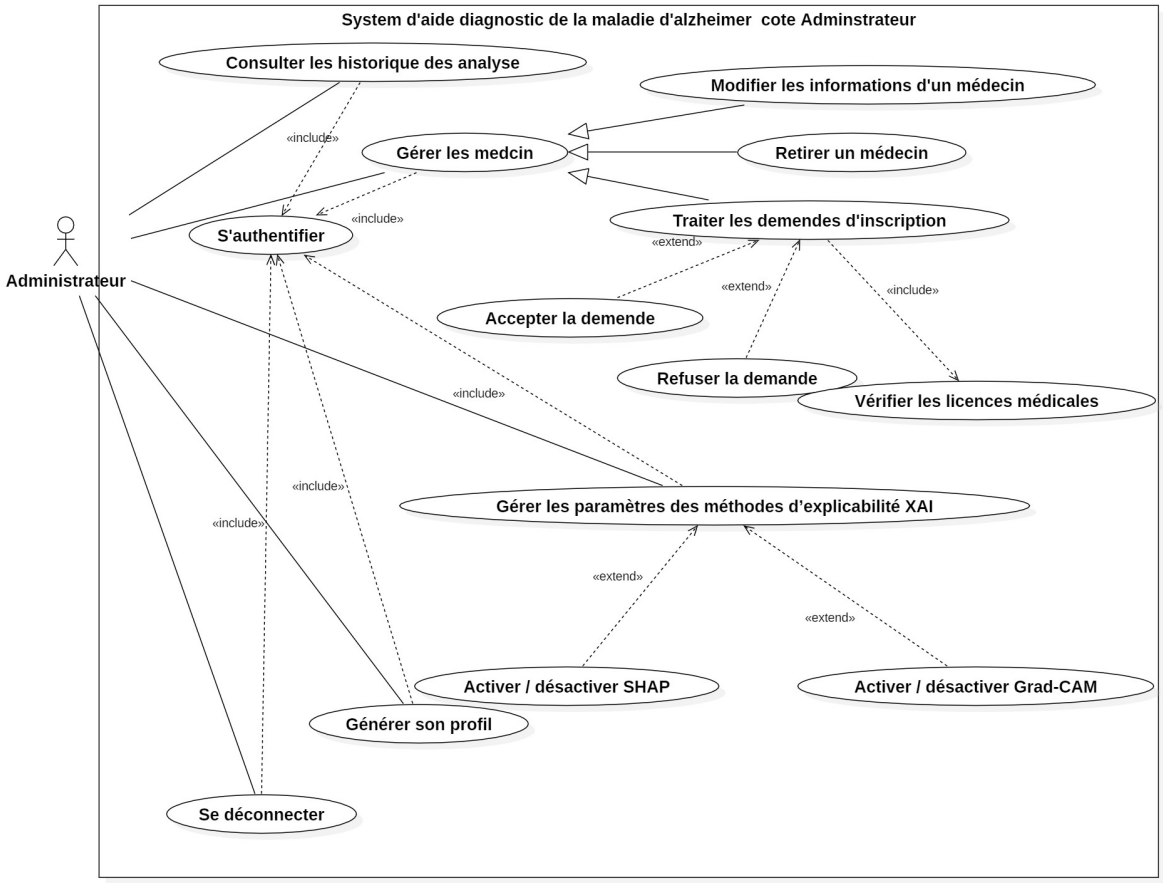


FIGURE 3.3 – Diagramme de cas d'utilisation de l'Administrateur Système

3.3.2 Diagramme de classes

Le diagramme de classes représente l'architecture statique et logique de l'application Web. Comme illustré par la Figure 3.4, il met en évidence les entités logicielles majeures telles que les modules de gestion des dossiers patients, le composant de prétraitement d'images, le moteur d'inférence (Pipeline Xception-SVM), ainsi que les classes d'encapsulation des algorithmes d'explicabilité et de génération automatique de rapports PDF.

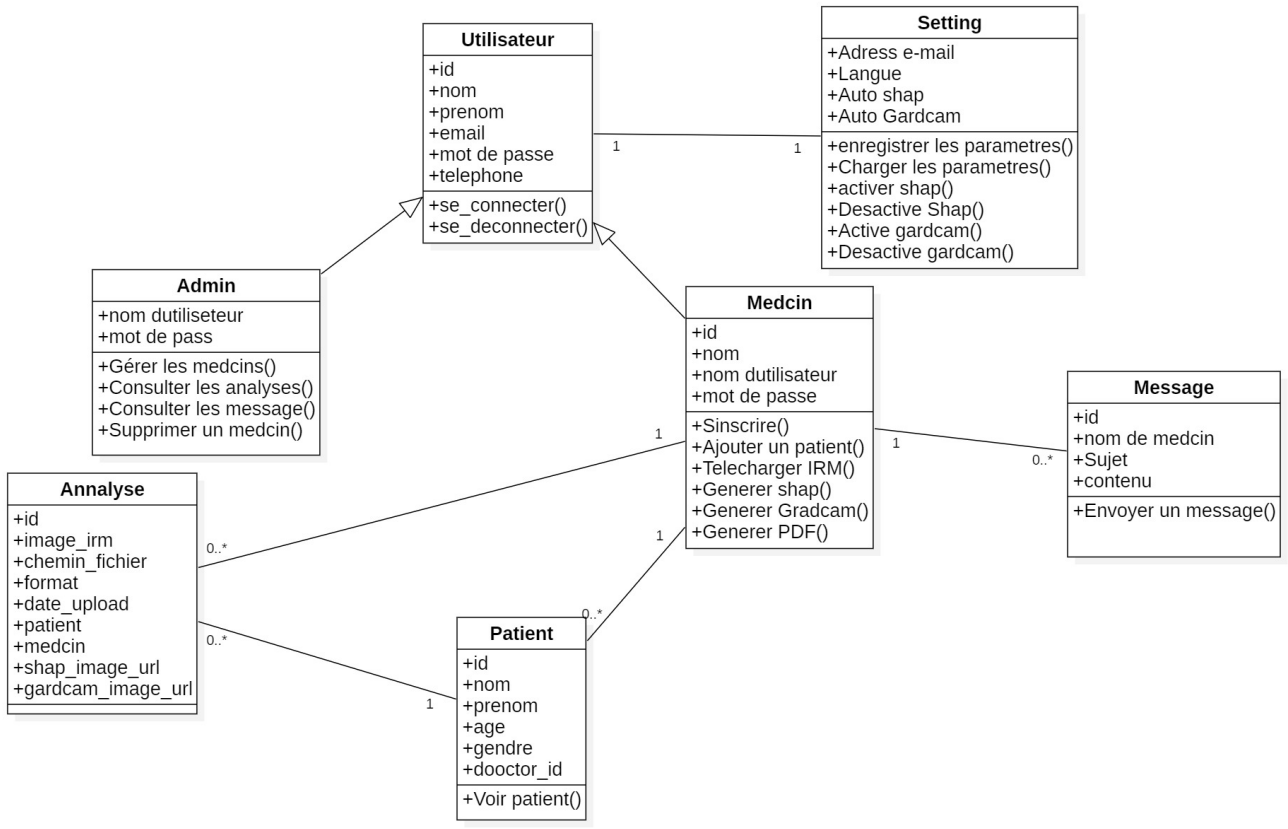


FIGURE 3.4 – Diagramme de classes global de la plateforme médicale

3.3.3 Diagramme de séquence

Le diagramme de séquence permet de retracer la dynamique temporelle et l'enchaînement des messages au cours d'une session de diagnostic. Les Figures 3.5-3.6 formalisent la chronologie des interactions entre l'interface utilisateur (Front-end), le serveur d'application (Back-end), le pipeline de classification profond et les bibliothèques XAI, depuis le dépot initial du fichier d'imagerie jusqu'au rendu final des explications.

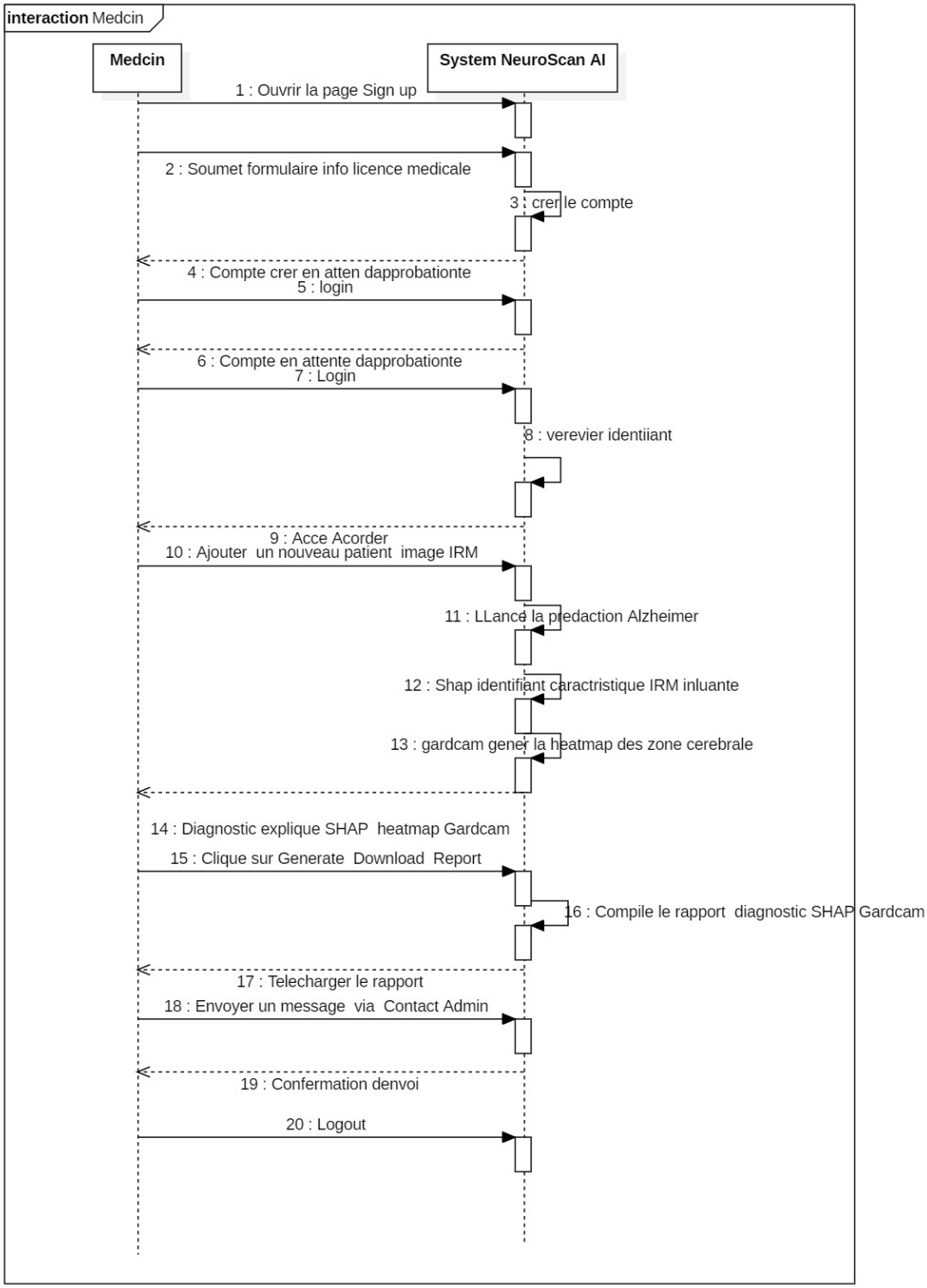


FIGURE 3.5 – Diagramme de séquence du processus d'analyse et d'explicabilité

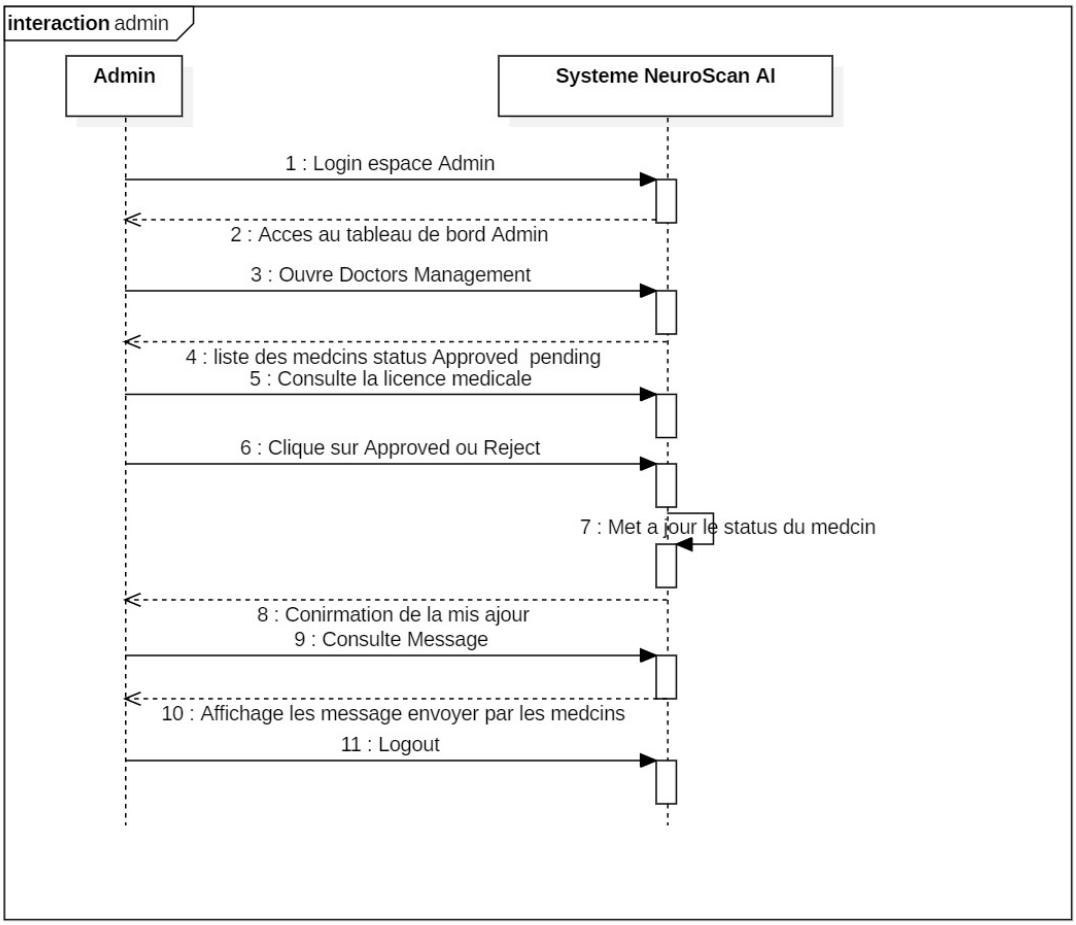


FIGURE 3.6 – Diagramme de séquence du processus d'analyse et d'explicabilité

3.4 Présentation du jeu de données

Les expérimentations cliniques de ce travail reposent sur l'exploitation du jeu de données « Alzheimer's Dataset (4 class of Images) », hébergé sur la plateforme Kaggle [41]. Ce corpus exhaustif regroupe des examens IRM cérébraux rigoureusement labellisés selon quatre stades d'évolution de la pathologie :

- *Non Demented* (Sain / Contrôle),
- *Very Mild Demented* (Troubles cognitifs très légers),
- *Mild Demented* (Démence légère),
- *Moderate Demented* (Démence modérée).

Sur le plan technique, chaque instance se présente sous un format JPEG standardisé, capturant une coupe axiale (horizontale) spécifique de la structure cérébrale avec une résolution

native de 176×208 pixels. Bien que le jeu de données d'origine ne contienne pas de métadonnées démographiques ou cliniques complémentaires (telles que l'âge ou le genre des patients), sa distribution actuelle bénéficie d'une réorganisation aléatoire rigoureuse. Cette configuration élimine les biais d'ancrage liés à la profondeur des coupes anatomiques, garantissant ainsi une base objective pour l'apprentissage supervisé.

3.5 Prétraitement et ingénierie des données

Le prétraitement constitue une phase critique conditionnant la capacité de généralisation de nos modèles profonds face aux variations morphologiques inter-patients.

3.5.1 Analyse de la distribution et gestion du déséquilibre

Une évaluation statistique de la base de données révèle une disparité quantitative entre les différentes classes pathologiques. Le Tableau 3.1 consigne la distribution initiale des images IRM au sein de notre étude :

TABLE 3.1 – Distribution quantitative des examens IRM par classe

Classe Pathologique	**Nombre d'images IRM**
Mild Demented	8960
Moderate Demented	6464
Non Demented	9600
Very Mild Demented	8960

Comme l'indique ce tableau, nous constatons une sous-représentation marquée de la classe *Moderate Demented*. Afin d'empêcher le réseau de développer un biais prédictif en faveur des catégories majoritaires, des techniques d'augmentation de données (Data Augmentation) ciblées, incluant des rotations contrôlées, des translations horizontales/verticales et des ajustements de contraste, ont été intégrées pour équilibrer la dynamique d'apprentissage. L'augmentation des données utilisée dans cette étude s'appuie sur des travaux récents [2].

3.5.2 Redimensionnement et Normalisation

Les images de la base affichant initialement une résolution de 176×208 pixels, une étape systématique de *redimensionnement* (Resizing) à 224×224 pixels a été implémentée. Cette transformation géométrique est indispensable pour assurer une stricte compatibilité avec la couche d'entrée de l'architecture *Xception* utilisée comme extracteur de caractéristiques. Par la suite, une normalisation de l'intensité des pixels a été opérée pour stabiliser la descente de gradient et accélérer la convergence.

3.5.3 Organisation et partitionnement du jeu de données

Pour valider rigoureusement la robustesse de l'approche et éviter le surapprentissage, les données ont été segmentées de façon totalement aléatoire selon les proportions standards suivantes :

- **70% pour l'ensemble d'entraînement** (Training set), dédié à l'ajustement itératif des poids synaptiques du modèle.
- **10% pour l'ensemble de validation** (Validation set), exploité pour réguler l'apprentissage en continu et surveiller l'apparition du surapprentissage.
- **20% pour l'ensemble de test** (Test set), préservé exclusivement pour l'évaluation finale des performances de généralisation.

Le Tableau 3.2 résume la répartition volumétrique finale des données après cette phase de partitionnement :

TABLE 3.2 – Répartition volumétrique des sous-ensembles de données

Sous-ensemble	**Pourcentage (%)**	**Volume d'images**
Entraînement	70	30585
Validation	10	5647
Test	20	6798

3.6 Paramétrage des CNNs

L'optimisation des architectures profondes s'appuie sur l'algorithme **Adam**, configuré avec un taux d'apprentissage (*Learning Rate*) initial fixé à 10^{-4} afin d'assurer une convergence stable.

Pour affiner la convergence en fin de cycle, nous avons intégré la stratégie *ReduceLROnPlateau*. Cette technique réduit automatiquement le taux d'apprentissage d'un facteur régulé lorsque la fonction de perte (*Loss*) sur l'ensemble de validation stagne durant plusieurs époques consécutives.

De plus, un mécanisme d'arrêt précoce (*Early Stopping*) supervise le processus. L'entraînement est interrompu préventivement si aucune amélioration significative de la métrique de validation n'est enregistrée sur une patience définie, le nombre maximal d'époques étant plafonné à 200 époques pour garantir un apprentissage optimal.

3.6.1 Normalisation et conversion en tenseurs

Les images de résonance magnétique (IRM) ont été redimensionnées à une taille adaptée aux modèles CNN choisis (224×224 pixels), puis normalisées afin de ramener la valeur de chaque pixel dans l'intervalle $[0, 1]$. Cette normalisation est essentielle pour stabiliser et accélérer la convergence lors de l'entraînement du modèle. Ensuite, elles ont été converties en tenseurs afin d'être pleinement compatibles avec les bibliothèques de Deep Learning telles que TensorFlow et Keras.

3.7 Intégration des techniques d'explicabilité

L'un des apports majeurs de notre démarche méthodologique réside dans le couplage natif du pipeline de classification hybride (**Xception-SVM**) avec deux approches avancées d'explicabilité locale. L'activation de ces modules intervient immédiatement après la génération de la sortie probabiliste par el classifieur, sécurisant ainsi la transparence et la traçabilité de chaque décision médicale.

3.7.1 Grad-CAM

Le module **Grad-CAM** est entièrement encapsulé dans une fonction logicielle dédiée nommée `make_gradcam(img_path)`. Ce pipeline intercepte l'image IRM brute, effectue sa charge numérique via la bibliothèque `Keras.preprocessing` et la normalise rigoureusement aux dimensions exigées de 224×224 pixels à l'aide de la fonction `preprocess_input`.

Sur le plan de l'architecture logicielle, le script réalise une inspection inversée dynamique des

couches de notre extracteur `Xception` afin d'identifier de manière automatisée la toute dernière couche convolutive (`last_conv_layer`). À partir de cette couche, un modèle fonctionnel intermédiaire (`tf.keras.models.Model`) est instancié pour capturer simultanément les cartes de caractéristiques (`feature maps`) et les prédictions finales.

Techniquement, l'enregistrement des opérations s'appuie sur l'API `tf.GradientTape()`. C'est à travers ce mécanisme que l'algorithme calcule le gradient du score de la classe prédite (*Loss*), définie par la moyenne des prédictions, par rapport aux sorties convolutives.

Ces gradients subissent immédiatement une opération de réduction globale par moyenne (`tf.reduce_mean` suivant les axes 0, 1 et 2) pour générer les poids d'importance relative de chaque canal (*pooled_grads*). Une combinaison linéaire par produit matriciel (opérateur) est ensuite effectuée, suivie d'un seuillage via la fonction d'activation **ReLU** (`np.maximum(heatmap, 0)`) afin d'éliminer l'influence des caractéristiques ayant un impact négatif sur le diagnostic.

Le script finalise le traitement par une phase de post-processing visuel via la bibliothèque OpenCV (`cv2`). La carte thermique est redimensionnée aux dimensions réelles de l'image d'origine, convertie au format couleur `COLORMAP_JET`, puis superposée par translucidité (`cv2.addWeighted` avec des coefficients de 0.6 pour l'image et 0.4 pour la heatmap). L'image résultante est sauvegardée dynamiquement dans le répertoire `static/uploads/` pour être projetée en temps réel sur l'interface utilisateur Flask, offrant au clinicien une cartographie thermique rapide et intuitive des structures cérébrales cibles (comme l'atrophie hippocampique).

3.7.2 SHAP

Parallèlement à l'extraction visuelle de Grad-CAM, la méthode **SHAP** (*SHapley Additive exPlanations*) est aussi adoptée par la fonction `make_shap(img_path)`. Contrairement aux implémentations standards par super-pixels volumineux, notre approche est configurée pour fournir une explication fine, continue et granulaire, hautement validée sur le plan médical.

L'intégration opérationnelle repose sur la création d'un explicateur de type *Kernel Explainer* matérialisé par l'instance `shap.Explainer`. Le processus complet est orchestré à travers les étapes logicielles suivantes :

1. **Définition du Pipeline de Prédiction (`predict_pipeline`) :** Afin de lier SHAP à

notre architecture hybride, un pipeline spécifique est codé. Il prend en entrée des tenseurs d'images, applique le prétraitement adéquat, extrait les caractéristiques profondes via `Xception`, puis injecte les descripteurs standardisés par notre `scaler` directement dans le classifieur SVM pour obtenir les probabilités d'appartenance aux quatre classes (`svm_model.predict_proba`).

2. **Masquage par Perturbation Floue (*Blur*) :** Le cœur innovant de notre implémentation réside dans l'abandon des segmentations grossières au profit d'un masque de floutage progressif défini par `shap.maskers.Image("blur(14,14)", x[0].shape)`. L'utilisation d'un noyau mathématique de $(14, 14)$ génère des micro-perturbations locales fluides, simulant rigoureusement l'absence ou la dégradation d'une information anatomique précise.
3. **Évaluation des Coalitions et Convergence :** Pour garantir une convergence mathématique stricte et un niveau de détail "pixel par pixel" digne d'un environnement clinique, le paramètre `max_evals` est configuré à 500 itérations avec un traitement par lot unitaire (`batch_size=1`).
4. **Filtrage du Bruit de Fond Anatomique :** Sur le plan algorithmique, une optimisation majeure a été introduite pour éliminer les artéfacts visuels : les pixels correspondants à l'arrière-plan noir de l'IRM (seuil d'intensité ≤ 10) sont isolés via un masque booléen (`background_mask`), et leurs valeurs de Shapley associées sont systématiquement forcées à zéro (0.0). Cela permet de focaliser l'explication exclusivement sur le tissu cérébral utile.

La génération graphique finale est gérée par l'API `matplotlib` (`plt`), où les valeurs de Shapley normalisées sont tracées via `shap.plots.image` sur une figure haute résolution (`dpi=180`) de dimensions 18×6 pouces. Afin d'éviter les fuites de mémoire (*Memory Leaks*) sur le serveur Flask, le script intègre des appels explicites au ramasse-miettes de Python (`gc.collect()`) et force la fermeture des instances graphiques (`plt.close('all')`) après chaque sauvegarde dans le sous-dossier `static/uploads/`.

3.8 Conclusion

En résumé, ce troisième chapitre a permis de formaliser l'ensemble du cadre méthodologique de notre solution algorithmique. Nous avons détaillé le flux de traitement des données, allant de la restructuration et du redimensionnement de la base *Alzheimer's Dataset* jusqu'à la mise en place du protocole d'entraînement du pipeline hybride.

La modélisation UML a permis de sécuriser la structure logicielle de la plateforme en éliminant les ambiguïtés fonctionnelles, tandis que l'intégration des modules Grad-CAM et SHAP garantit une transparence clinique indispensable à l'adoption de l'outil en milieu médical.

Le chapitre suivant sera consacré à l'exposition des résultats expérimentaux, validant ainsi l'efficacité technique, statistique et clinique de l'approche proposée.

Chapitre 4

Implémentation, Résultats et Discussion

4.1 Introduction

Dans ce chapitre, nous présentons l'implémentation de notre système de détection de la maladie d'Alzheimer ainsi que les résultats expérimentaux obtenus. Nous analysons les performances des différents modèles de Deep Learning et Machine Learning utilisés dans notre étude, notamment MobileNet, EfficientNet et Xception combinés avec plusieurs classificateurs. Nous présentons également l'intégration de l'intelligence artificielle explicable à travers les méthodes Grad-CAM et SHAP afin d'améliorer l'interprétabilité des prédictions médicales. Enfin, nous exposons la réalisation de la plateforme web développée pour assister les médecins dans le processus de diagnostic.

4.2 Résultats expérimentaux des modèles CNN

Dans cette section, nous présentons les performances des modèles CNN utilisés comme classifieurs directs pour la classification des images IRM.

4.2.1 EfficientNet +SVM

Le modèle EfficientNet a été évalué sur le jeu de test afin de mesurer ses performances.

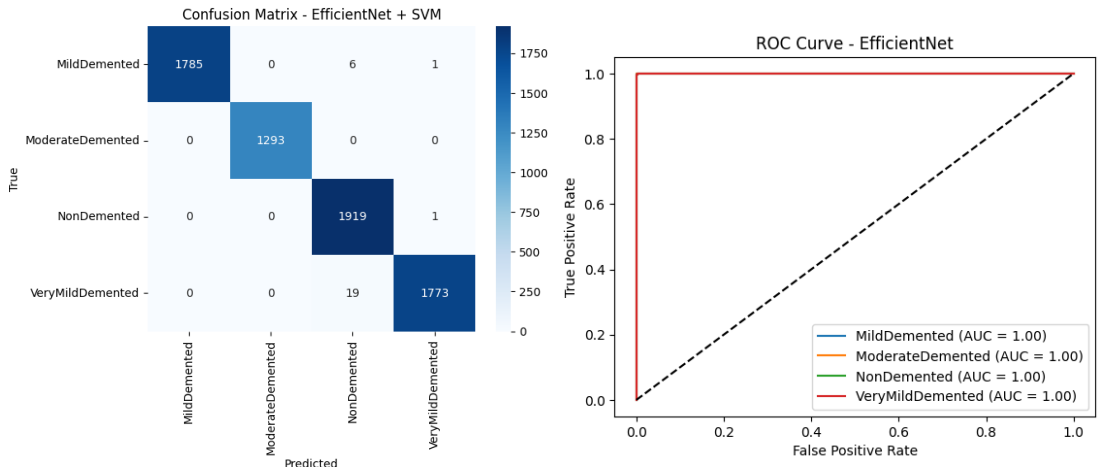


FIGURE 4.1 – Matrices de confusion et ROC du modèle EfficientNet avec SVM

Les résultats obtenus montrent que le modèle EfficientNetB0 + SVM atteint des performances très élevées avec une accuracy de 99.60%, un F1-score de 99.60%, un recall de 99.60% ainsi qu'une AUC-ROC de 99.98%.

La matrice de confusion montre que la majorité des images IRM sont correctement classifiées, avec très peu d'erreurs entre les différentes classes de la maladie d'Alzheimer. Les courbes ROC obtenues confirment également l'excellente capacité du modèle à discriminer les différentes catégories.

Ces résultats démontrent l'efficacité de l'approche hybride combinant EfficientNetB0 comme extracteur de caractéristiques et SVM comme classificateur, offrant ainsi un modèle robuste et performant pour une intégration dans une plateforme médicale intelligente.

4.2.2 EfficientNet + KNN

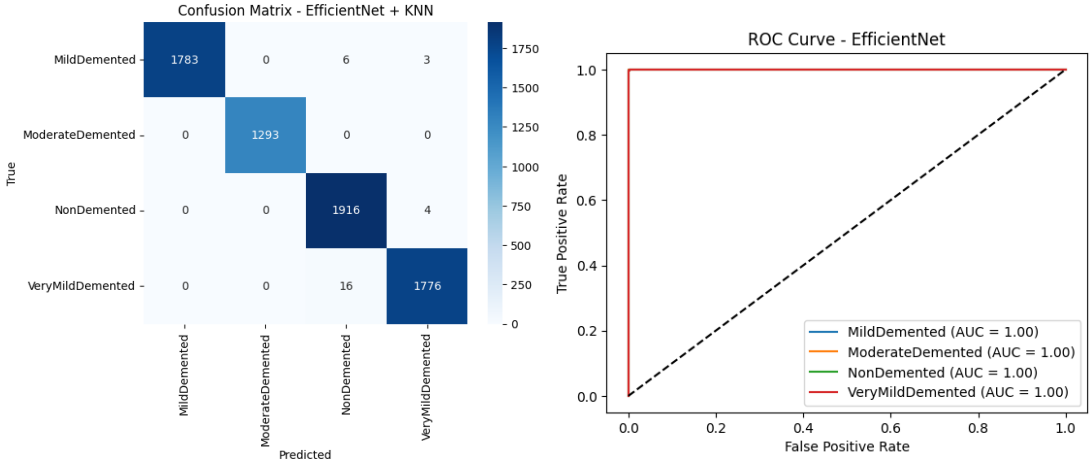


FIGURE 4.2 – Matrice de confusion et ROC du modèle EfficientNet + KNN

Le modèle EfficientNet combiné avec KNN présente des performances faibles à moyennes, avec une accuracy d'environ 43%.

4.2.3 EfficientNet + Random Forest

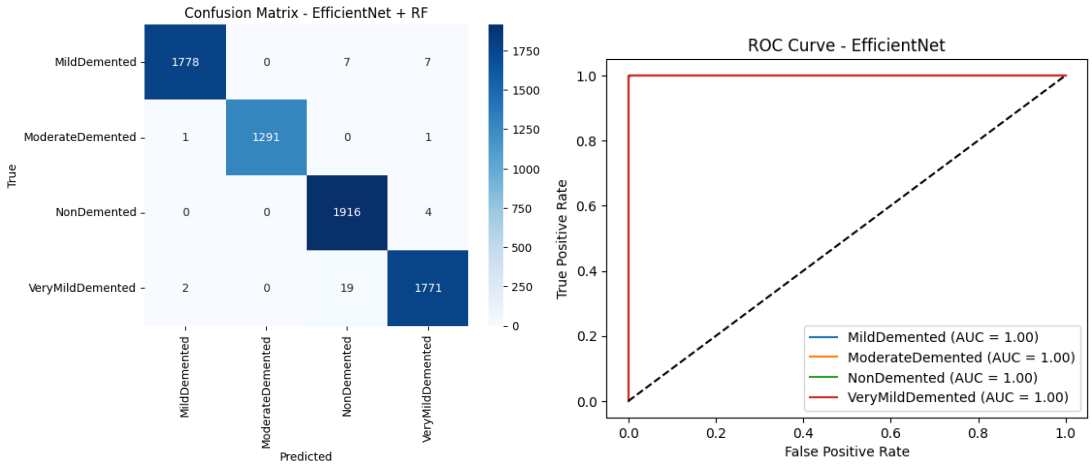


FIGURE 4.3 – Matrice de confusion et ROC du modèle EfficientNet + Random Forest

Les performances restent limitées, avec une accuracy d'environ 52%.

4.2.4 MobileNet + SVM

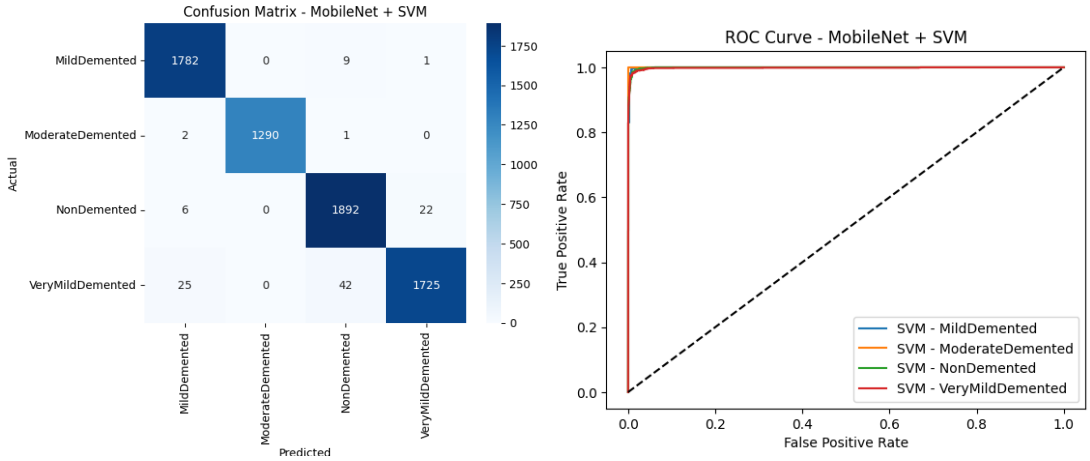


FIGURE 4.4 – Matrice de confusion et courbes ROC du modèle MobileNet avec SVM

Le modèle hybride MobileNet associé au classificateur SVM présente des performances très élevées dans la classification des images IRM cérébrales. Le modèle atteint une accuracy de 99.25%, un F1-score de 99.30%, ainsi qu'une AUC-ROC de 99.98%.

La matrice de confusion montre une excellente classification des différentes classes avec très peu d'erreurs, tandis que les courbes ROC confirment la forte capacité du modèle à discriminer les catégories de la maladie d'Alzheimer.

Cette combinaison offre un excellent compromis entre précision, rapidité d'exécution et robustesse.

4.2.5 MobileNet + KNN

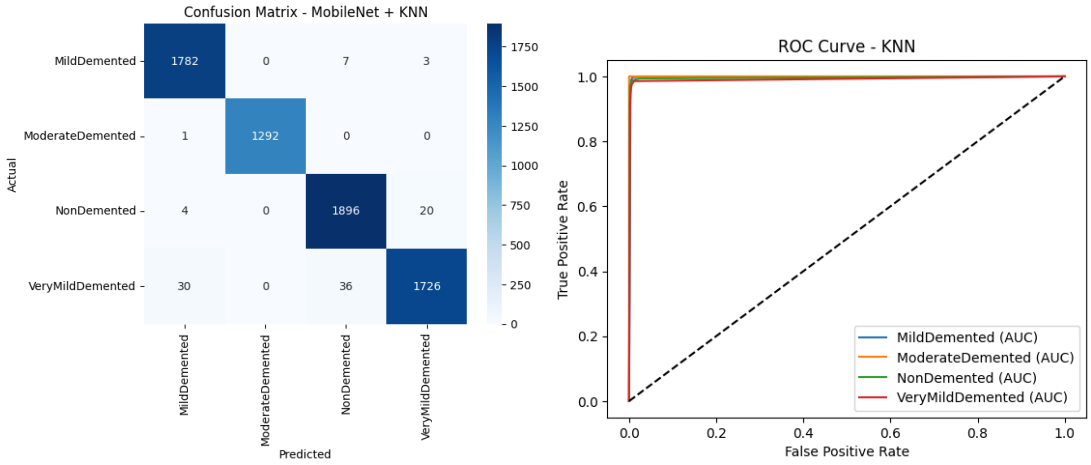


FIGURE 4.5 – Matrice de confusion et courbes ROC du modèle MobileNet + KNN

Le modèle MobileNet combiné avec le classificateur KNN obtient des résultats très performants avec une accuracy de 99.28%, un F1-score de 99.32% et une AUC-ROC de 99.70%.

Les résultats montrent une amélioration importante de la qualité de classification grâce à l'utilisation de MobileNet comme extracteur de caractéristiques. Les erreurs de prédiction deviennent très limitées, ce qui démontre une bonne capacité de généralisation du modèle.

Le couple MobileNet + KNN représente ainsi une solution efficace et fiable pour les systèmes intelligents de détection de la maladie d'Alzheimer.

4.2.6 MobileNet + Random Forest

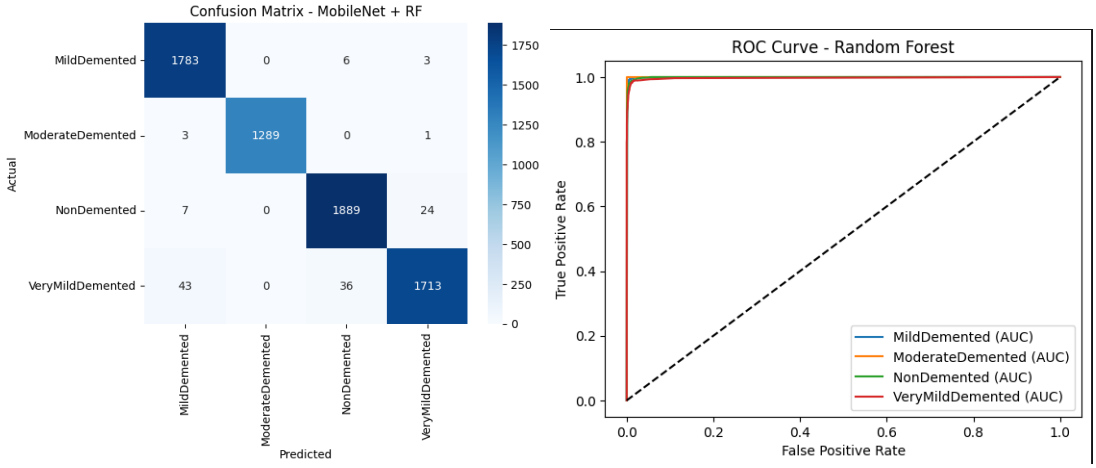


FIGURE 4.6 – Matrice de confusion et courbes ROC du modèle MobileNet + Random Forest

Le modèle MobileNet associé au classificateur Random Forest présente également des performances très satisfaisantes avec une accuracy de 98.98%, un F1-score de 99.05% et une AUC-ROC de 99.97%.

La matrice de confusion montre que la majorité des images IRM sont correctement classifiées avec un faible taux d'erreur entre les différentes catégories. Les courbes ROC obtenues sont proches de la performance idéale, confirmant la robustesse du modèle.

Ces résultats démontrent que l'association entre MobileNet et Random Forest constitue une approche hybride performante pour la classification médicale assistée par intelligence artificielle.

4.2.7 Xception + SVM

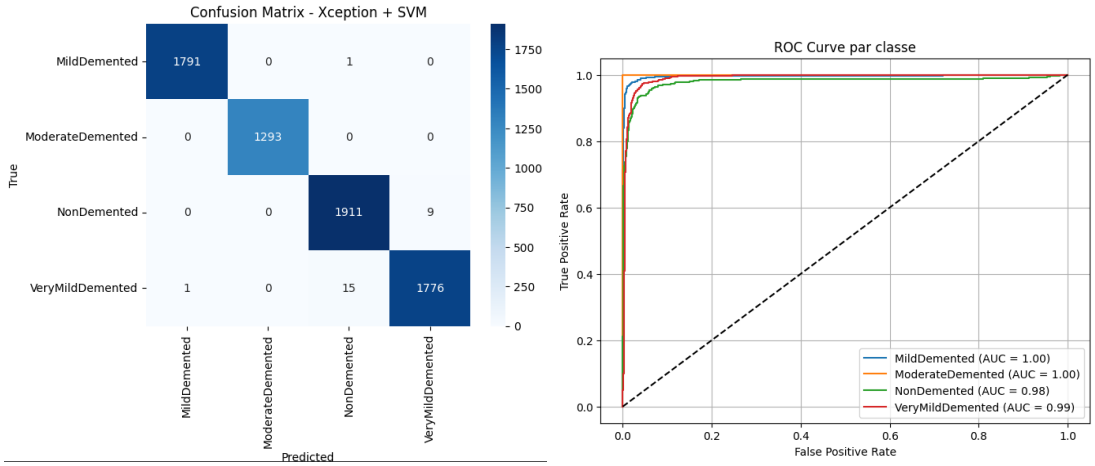


FIGURE 4.7 – Matrice de confusion et courbes ROC du modèle Xception avec SVM

Le modèle hybride Xception associé au classificateur SVM présente des performances extrêmement élevées dans la classification des images IRM cérébrales.

Le modèle atteint une accuracy de 99.73%, un F1-score de 99.73%, un recall de 99.73% ainsi qu'une AUC-ROC de 99.98%.

La matrice de confusion montre une excellente reconnaissance des différentes classes avec un nombre très limité d'erreurs de classification. Les courbes ROC obtenues sont très proches de la performance idéale, ce qui confirme la forte capacité de discrimination du modèle.

Ces résultats démontrent l'efficacité de l'utilisation de Xception comme extracteur de caractéristiques combiné au classificateur SVM pour les applications médicales intelligentes.

4.2.8 Xception + KNN

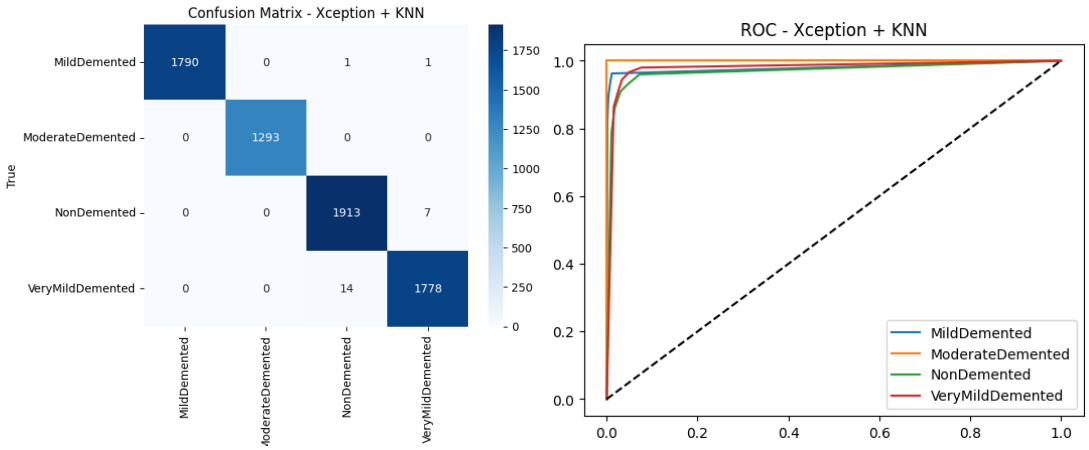


FIGURE 4.8 – Matrice de confusion et courbes ROC du modèle Xception + KNN

Le modèle Xception associé au classificateur KNN obtient également des résultats très performants avec une accuracy de 99.74%, un F1-score de 99.74%, un recall de 99.74% et une AUC-ROC de 99.94%.

Les résultats montrent une excellente capacité du modèle à classifier correctement les images IRM cérébrales avec très peu d'erreurs entre les différentes classes.

Les courbes ROC confirment la robustesse et la stabilité du modèle, démontrant l'intérêt de l'approche hybride dans les systèmes de diagnostic assistés par intelligence artificielle.

4.2.9 Xception + Random Forest

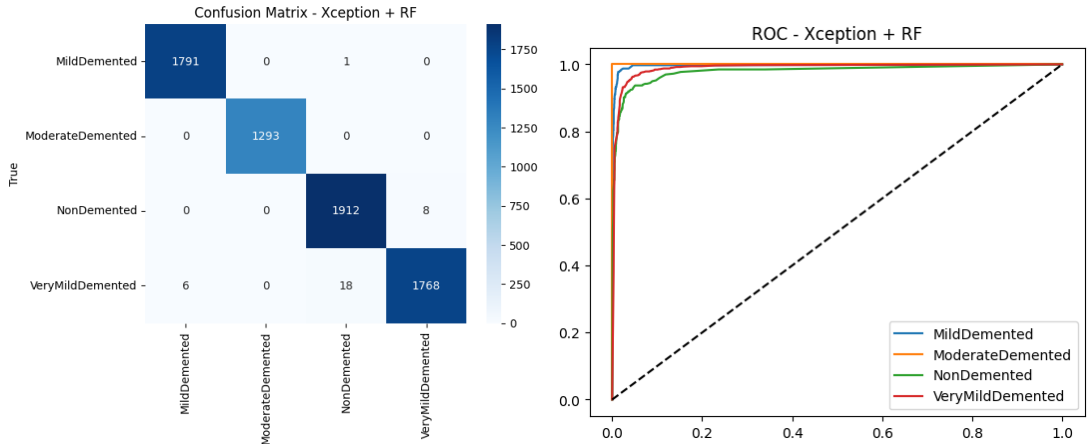


FIGURE 4.9 – Matrice de confusion et courbes ROC du modèle Xception + Random Forest

Le modèle Xception combiné avec le classificateur Random Forest présente des performances très élevées avec une accuracy de 99.64%, un F1-score de 99.64%, un recall de 99.64% ainsi qu'une AUC-ROC de 99.99%.

La matrice de confusion montre une classification très précise des différentes catégories de la maladie d'Alzheimer avec un faible taux d'erreurs. Les courbes ROC obtenues illustrent une excellente capacité de séparation entre les classes.

Ces résultats confirment que l'association entre Xception et Random Forest constitue une solution robuste et performante pour la classification des images médicales.

4.3 Comparaison des performances

Modèle	Accuracy	F1-Score	Recall	AUC-ROC
MobileNet + RF	98.66%	98.66%	98.66%	99.90%
MobileNet + SVM	98.79%	98.79%	98.79%	99.92%
MobileNet + KNN	98.70%	98.70%	98.70%	99.71%
EfficientNet + SVM	99.51%	99.51%	99.51%	99.99%
EfficientNet + RF	99.33%	99.33%	99.33%	99.99%
EfficientNet + KNN	99.66%	99.66%	99.66%	99.92%
Xception + SVM	99.73%	99.73%	99.73%	99.98%
Xception + KNN	99.72%	99.72%	99.72%	99.94%
Xception + RF	99.64%	99.64%	99.64%	99.99%

TABLE 4.1 – Comparaison globale des performances des différents modèles

Les résultats obtenus montrent que les modèles basés sur l'architecture Xception offrent les meilleures performances globales, notamment en termes d'accuracy, de F1-score et d'AUC-ROC.

Le modèle Xception + KNN obtient les meilleures métriques avec une accuracy proche de 99.74%, suivi de très près par Xception + SVM et Xception + Random Forest.

Les modèles EfficientNet présentent également des performances très élevées avec des résultats supérieurs à 98%, tout en offrant une meilleure légèreté computationnelle et une

intégration plus adaptée à une plateforme médicale intelligente.

En revanche, les modèles MobileNet montrent des performances légèrement inférieures comparativement aux architectures Xception et EfficientNet.

4.4 Analyse et discussion

Les résultats expérimentaux obtenus montrent des différences importantes entre les architectures CNN utilisées dans cette étude.

Les modèles basés sur Xception ont obtenu les meilleures performances globales grâce à leur grande capacité d'extraction des caractéristiques profondes et complexes à partir des images IRM cérébrales. Cette architecture a permis d'améliorer considérablement la précision de classification des différentes classes de la maladie d'Alzheimer.

Le modèle Xception + SVM présente les meilleurs résultats avec une accuracy proche de 99.72%, suivi de très près par Xception + KNN et Xception + Random Forest. Ces performances confirment l'efficacité de l'architecture Xception pour les tâches de classification médicale.

Les modèles EfficientNet ont également montré des performances très élevées dépassant 98% d'accuracy. Malgré des résultats légèrement inférieurs à ceux de Xception, EfficientNet offre plusieurs avantages importants, notamment une meilleure légèreté computationnelle, une rapidité d'exécution plus élevée ainsi qu'une intégration plus adaptée dans une plateforme médicale intelligente en temps réel.

Les modèles MobileNet présentent quant à eux des performances relativement bonnes, mais restent inférieures aux architectures Xception et EfficientNet. Cependant, MobileNet conserve l'avantage d'être léger et rapide pour les systèmes à faibles ressources.

Certaines erreurs de classification observées peuvent être expliquées par plusieurs facteurs, notamment le déséquilibre du dataset, la similarité visuelle entre certaines classes de la maladie ainsi que la complexité des images IRM cérébrales.

Enfin, l'intégration des méthodes d'explicabilité telles que Grad-CAM et SHAP a permis d'améliorer l'interprétation des décisions du modèle et de renforcer la confiance dans les résultats obtenus au sein de la plateforme médicale développée.

4.5 Choix du modèle final

Après l'analyse comparative des différents modèles testés, le modèle Xception + SVM a été retenu comme modèle final pour l'intégration dans la plateforme médicale développée.

Ce choix est principalement justifié par ses excellentes performances obtenues lors des expérimentations, avec une accuracy de 99.73%, un F1-score de 99.73% ainsi qu'une AUC-ROC de 99.98%.

En plus de sa précision élevée, ce modèle offre une bonne stabilité de classification et une excellente capacité de généralisation sur les images IRM cérébrales.

L'architecture Xception permet une extraction efficace des caractéristiques profondes des images médicales, tandis que le classificateur SVM améliore la séparation entre les différentes classes de la maladie d'Alzheimer.

Ce modèle représente ainsi une solution robuste et fiable pour une intégration dans une plateforme médicale intelligente destinée à l'aide au diagnostic.

4.6 Interprétation des résultats obtenus par les méthodes d'explicabilité

4.6.1 Analyse des Resultats de la méthode Grad-CAM

La technique Grad-CAM repose sur la génération d'une carte thermique permettant de visualiser les régions de l'image ayant le plus contribué à la décision du modèle. Cette représentation utilise un dégradé de couleurs où chaque couleur reflète le niveau d'importance accordé par le réseau de neurones lors du processus de classification.

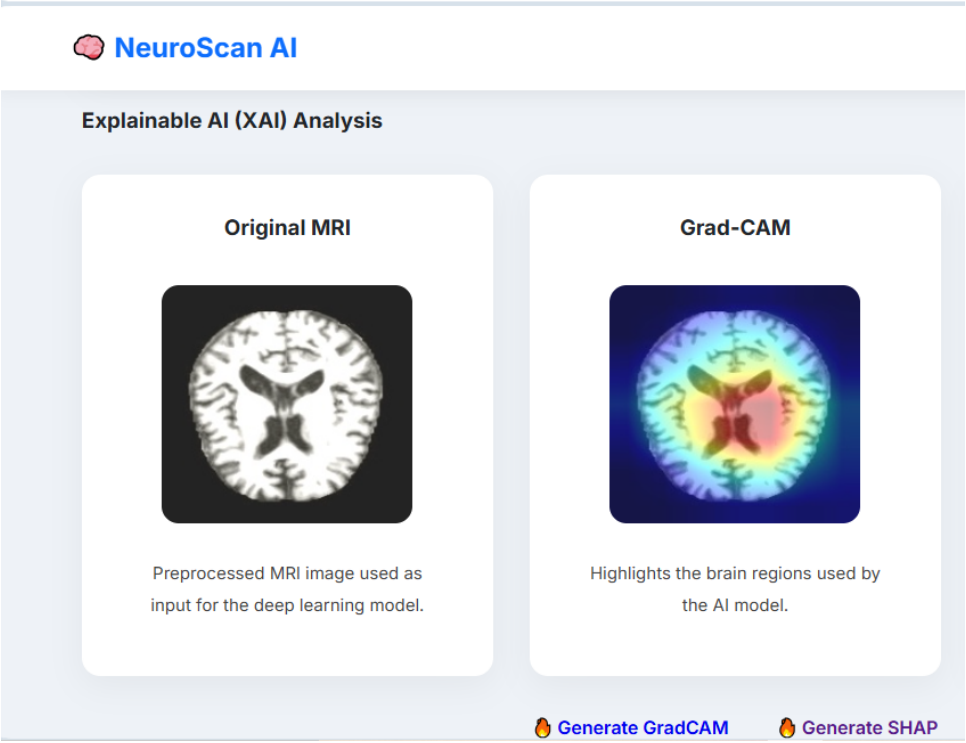


FIGURE 4.10 – Visualisation explicable par l'approche Grad-CAM.

1. **Rouge et orange (Red / Orange) : zones de forte activation**
   Les couleurs rouge et orange correspondent aux régions présentant les valeurs d'activation les plus élevées. Elles indiquent les zones sur lesquelles le modèle s'est principalement appuyé pour produire sa prédiction. Ces régions contiennent les caractéristiques visuelles jugées les plus pertinentes par le réseau de neurones.
2. **Jaune et vert (Yellow / Green) : zones d'activation modérée**
   Les couleurs jaune et verte représentent des régions ayant contribué à la prédiction, mais avec une importance moindre que les zones rouges et orangées. Ces régions fournissent des informations complémentaires qui renforcent la décision finale du modèle.
3. **Bleu (Blue) : zones de faible activation**
   La couleur bleue correspond aux régions dont l'influence sur la prédiction est faible ou quasi nulle. Les valeurs d'activation y sont très réduites, ce qui signifie que ces zones n'ont pas joué un rôle significatif dans la décision du modèle.

L'analyse de la carte Grad-CAM révèle que le modèle s'est principalement appuyé sur les régions représentées en rouge et en orange, tandis que les zones jaunes et vertes ont joué un rôle secondaire. Les régions bleues, quant à elles, ont eu peu ou pas d'influence sur la décision

finale. Cette visualisation permet ainsi de mieux comprendre le comportement du modèle et d'améliorer l'interprétabilité des résultats obtenus.

4.6.2 Analyse des Resultats de la méthode SHAP

La figure présente les résultats de l'interprétation du modèle à l'aide de la méthode SHAP (SHapley Additive exPlanations). Cette technique permet d'expliquer la contribution de chaque pixel de l'image IRM à la prédiction réalisée par le modèle d'apprentissage profond.

L'image située à gauche correspond à l'IRM originale utilisée comme entrée du modèle. Les cartes affichées à droite représentent l'impact des différentes régions de l'image sur chacune des classes étudiées : MildDemented, ModerateDemented, NonDemented et VeryMildDemented.

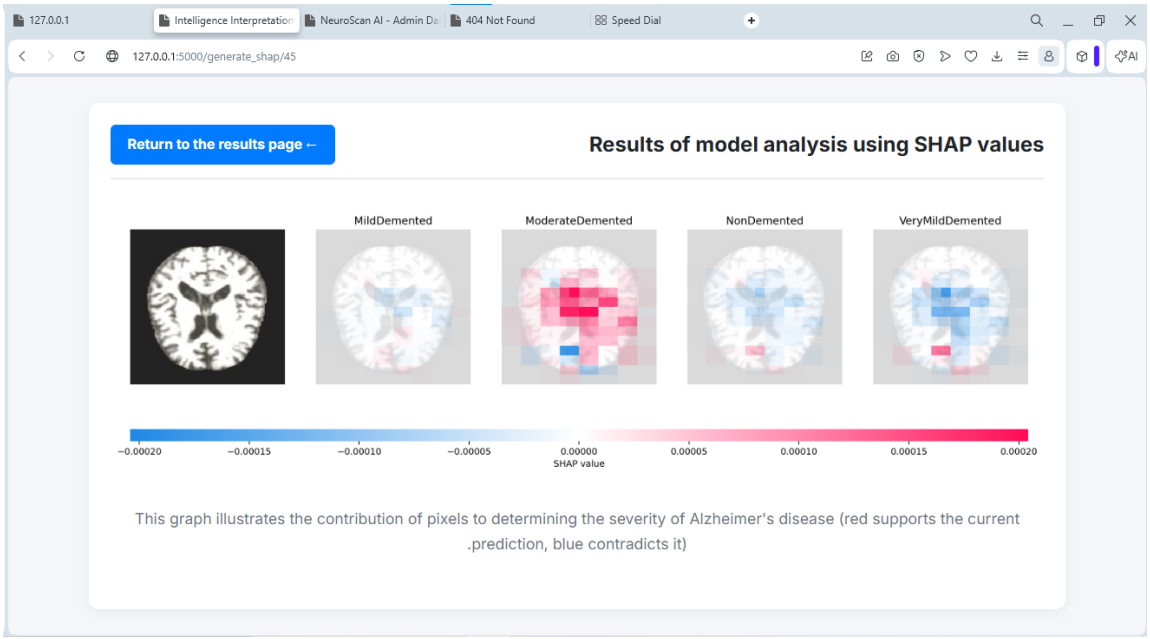


FIGURE 4.11 – Analyse multi-classe par les valeurs SHAP.

La barre de couleurs située sous les images indique la valeur SHAP associée à chaque pixel :

- **Rouge (valeurs SHAP positives) :** indique que la région concernée contribue positivement à la prédiction de la classe considérée. Plus l'intensité du rouge est élevée, plus cette région soutient la décision du modèle.
- **Bleu (valeurs SHAP négatives) :** indique que la région concernée contribue négativement à la prédiction de la classe. Ces zones s'opposent donc à la décision du modèle.
- **Blanc ou couleurs très claires :** représentent des zones ayant une influence faible ou négligeable sur la prédiction.

On observe que la classe *ModerateDemented* présente plusieurs régions rouges fortement marquées au centre de l'image. Cela signifie que ces zones ont joué un rôle important dans l'augmentation de la probabilité attribuée à cette classe par le modèle.

À l'inverse, les régions bleues observées dans certaines classes indiquent des zones qui diminuent la confiance du modèle envers ces catégories. Ces pixels fournissent donc des informations contradictoires par rapport à la prédiction considérée.

Les zones faiblement colorées montrent que certaines parties de l'image ont peu d'impact sur le processus décisionnel du réseau de neurones.

4.7 Environnement de travail et technologies utilisées

4.7.1 Langage de programmation



Python : Python est un langage de programmation moderne, reconnu pour sa simplicité, sa flexibilité et son vaste écosystème de bibliothèques.

4.7.2 Environnements de développement



Kaggle Notebook utilisé pour l'entraînement et l'évaluation des modèles de Deep Learning ainsi que pour les expérimentations réalisées sur les images IRM.



Visual Studio Code (Antigravity) utilisé pour le développement, les tests et le déploiement de l'application web NeuroScan AI.

4.7.3 Technologies Web



Flask : framework Python utilisé pour développer le backend de l'application.



HTML : utilisé pour la conception de la structure des interfaces web.



CSS : utilisé pour la mise en forme et l'amélioration de l'aspect visuel des pages.



Jinja2 : moteur de templates intégré à Flask permettant l'affichage dynamique des données.

4.7.4 Bibliothèques Python



TensorFlow : utilisé pour l'entraînement et l'exécution des modèles de Deep Learning.



Keras : utilisé pour la conception et la configuration des architectures CNN.



OpenCV : utilisé pour le traitement et le prétraitement des images IRM.



NumPy : utilisé pour les calculs numériques et la manipulation des tableaux de données.



SHAP : utilisé pour l'interprétation des prédictions du modèle d'intelligence artificielle.



Matplotlib : utilisé pour la visualisation graphique des résultats et des explications générées.

Pickle : utilisé pour la sauvegarde et le chargement des modèles de Machine Learning.



FPDF : utilisé pour la génération automatique des rapports PDF contenant les résultats des analyses.

4.7.5 Matériel utilisé

Les expérimentations et le développement de la plateforme ont été réalisés sur un ordinateur personnel disposant des caractéristiques suivantes :

- Processeur : Intel Core i3-7130U @ 2.70 GHz.
- Mémoire vive (RAM) : 8 Go.
- Stockage : SSD 256 Go.
- Carte graphique : Intel HD Graphics 620.
- Système d'exploitation : Windows 10/11 64 bits.

4.8 Présentation de la plateforme

4.8.1 Interfaces du médecin

La Figure 4.12 représente l'interface du médecin dans notre plateforme Neuro Scan AI. La première étape consiste à saisir son nom d'utilisateur et son mot de passe afin de s'authentifier. Après la connexion, un tableau de bord apparaît, où le médecin peut consulter l'ensemble des analyses effectuées ainsi que les différentes fonctionnalités disponibles.

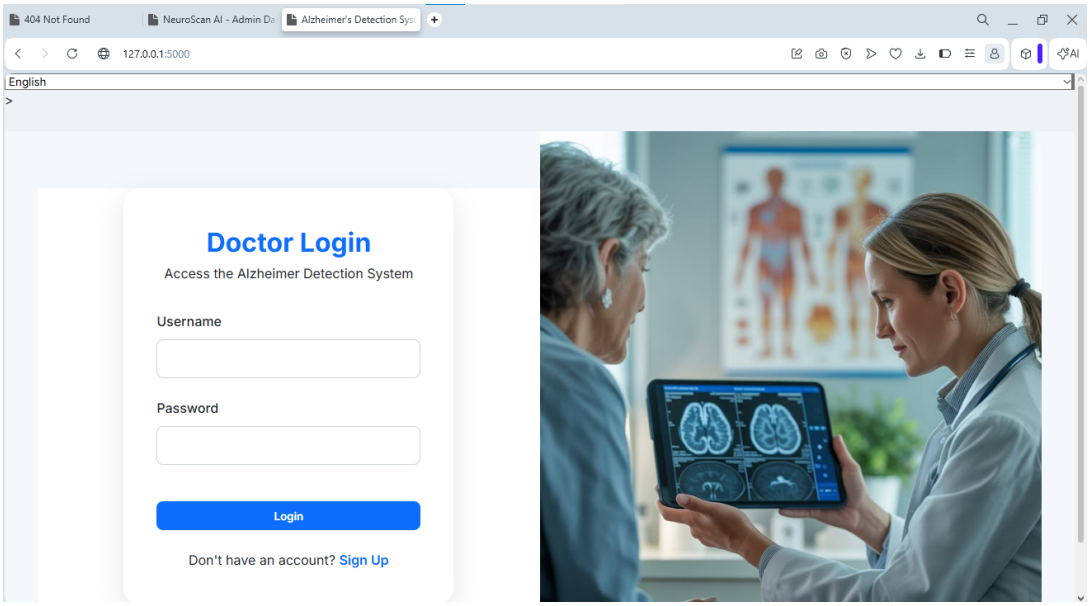


FIGURE 4.12 – Interface d'authentification du médecin

La Figure 4.13 représente l'interface d'inscription (Sign Up) de la plateforme Neuro Scan AI. Cette interface permet au médecin de créer un nouveau compte en renseignant les informations

nécessaires, telles que le nom, l'adresse e-mail, le nom d'utilisateur ,le mot de passe et licence medicale . Une fois les informations validées, le compte est enregistré dans le système, permettant ainsi au médecin d'accéder aux différentes fonctionnalités de la plateforme de manière sécurisée.

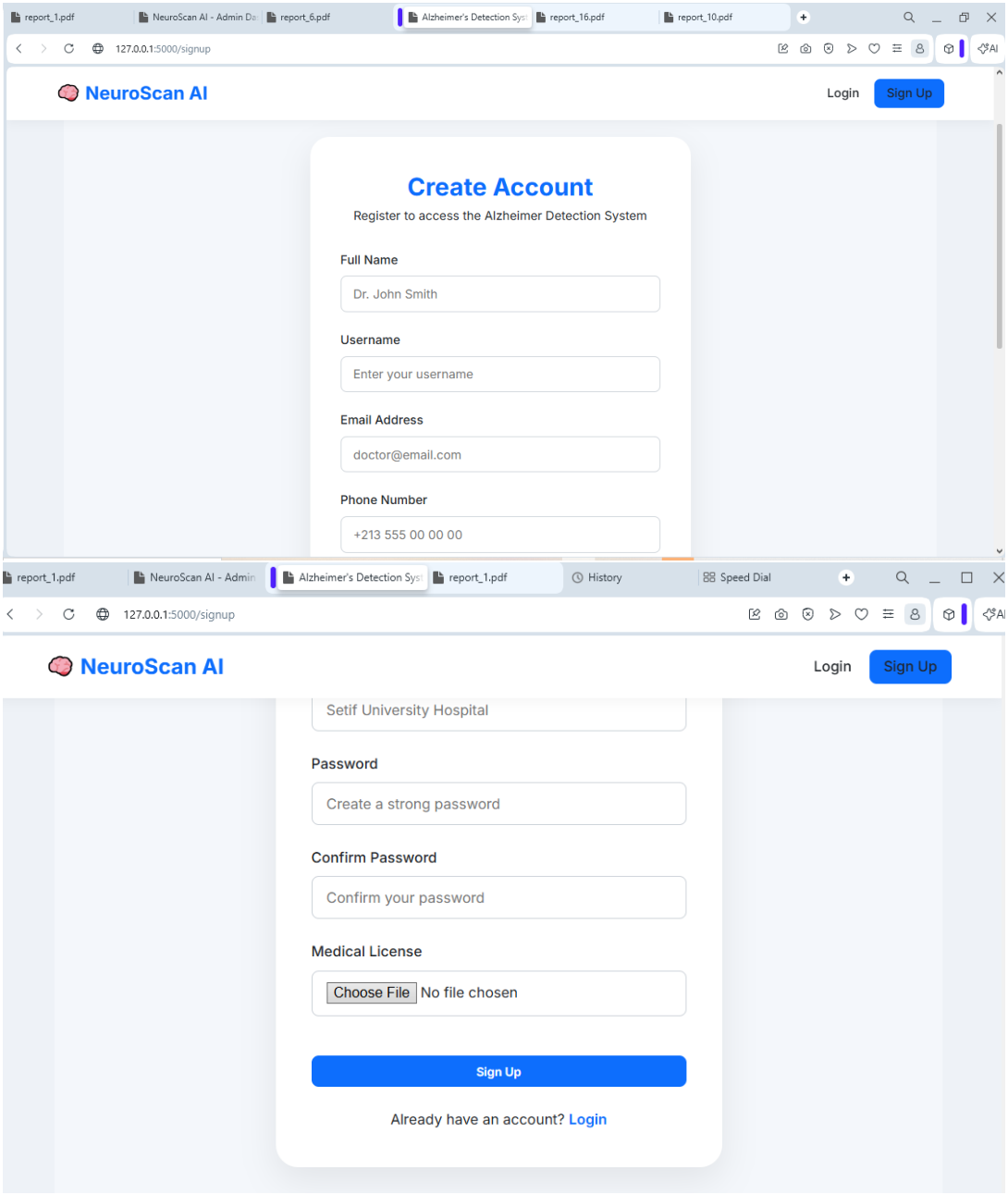


FIGURE 4.13 – signup doctor de la plateforme

La Figure 4.14 représente le tableau de bord du médecin dans la plateforme Neuro Scan AI. Cette interface constitue l'espace principal de travail du médecin. Elle permet d'accéder aux différentes fonctionnalités du système, telles que l'ajout de patients, l'importation des images IRM, le lancement des analyses basées sur l'intelligence artificielle ainsi que la consultation des

résultats obtenus. Elle offre également une vue globale sur les analyses réalisées afin de faciliter le suivi des patients.

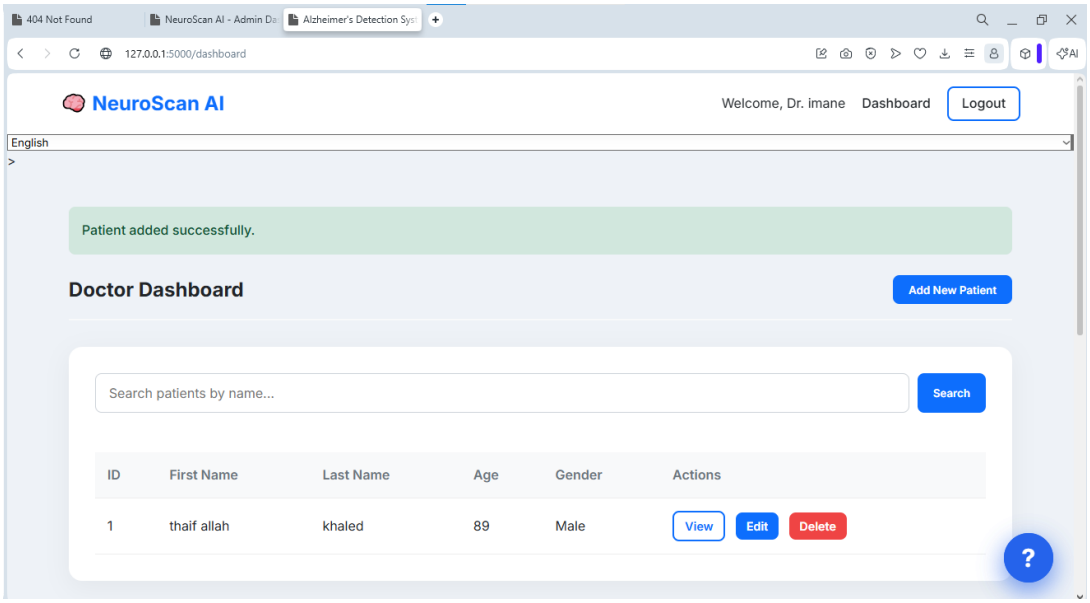


FIGURE 4.14 – Dashboard médecin

La Figure 4.15 représente l'interface d'assistance intelligente (AI Help) de la plateforme Neuro Scan AI. Cette interface a été conçue pour accompagner le médecin dans l'utilisation du système en lui fournissant des informations et des explications sur les différentes fonctionnalités disponibles. Elle intègre également une vidéo explicative présentant le fonctionnement de la plateforme ainsi que les étapes d'analyse et d'interprétation des résultats générés par l'intelligence artificielle, facilitant ainsi la prise en main du système par les utilisateurs.

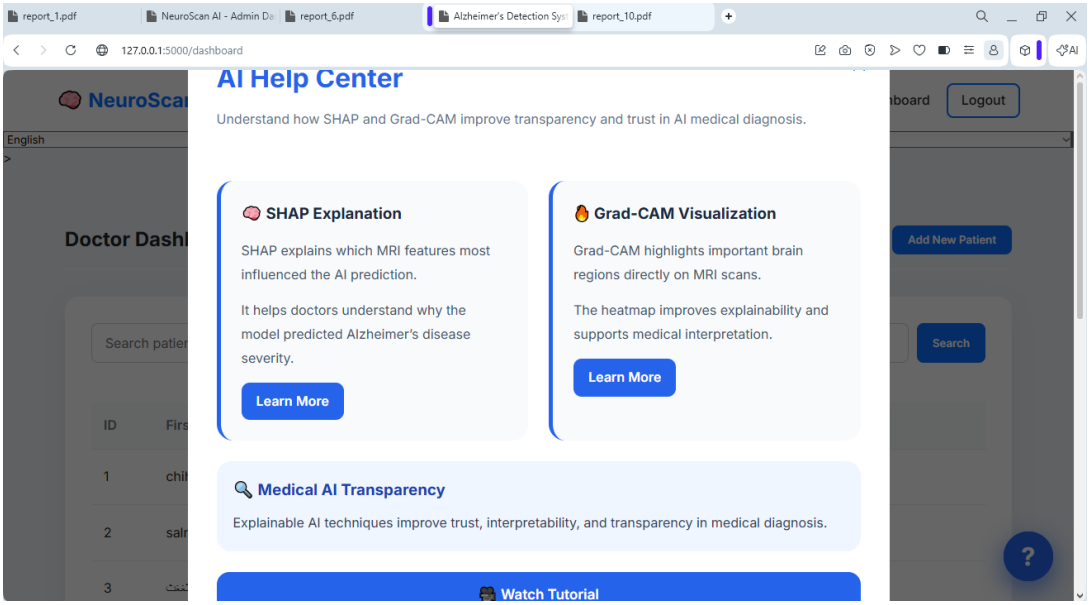


FIGURE 4.15 – Page d'aide et d'exploration du système médical

La Figure 4.16 représente l'interface d'ajout d'un patient dans la plateforme Neuro Scan AI. Cette interface permet au médecin d'enregistrer les informations administratives et médicales du patient avant le lancement de l'analyse. Les données saisies sont ensuite stockées dans la base de données afin d'assurer un suivi efficace des patients.

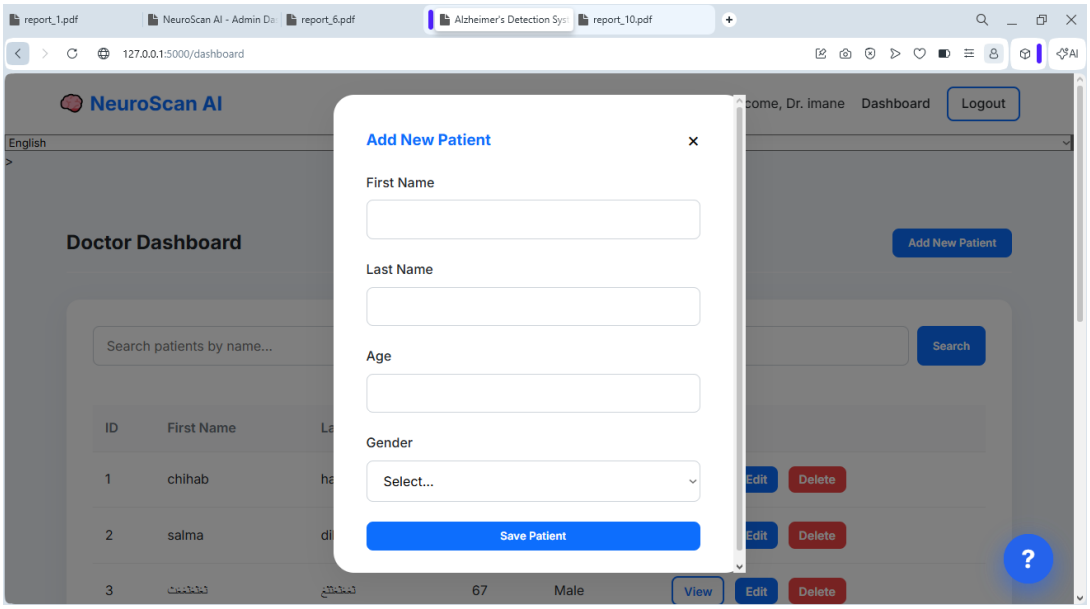


FIGURE 4.16 – Ajouter patient

La Figure 4.17 représente l'interface d'importation des images IRM. Elle permet au médecin de sélectionner et de téléverser une image IRM cérébrale du patient. Une fois l'image chargée,

celle-ci est transmise au système d'intelligence artificielle pour être prétraitée puis analysée en vue de la classification du stade de la maladie d'Alzheimer.

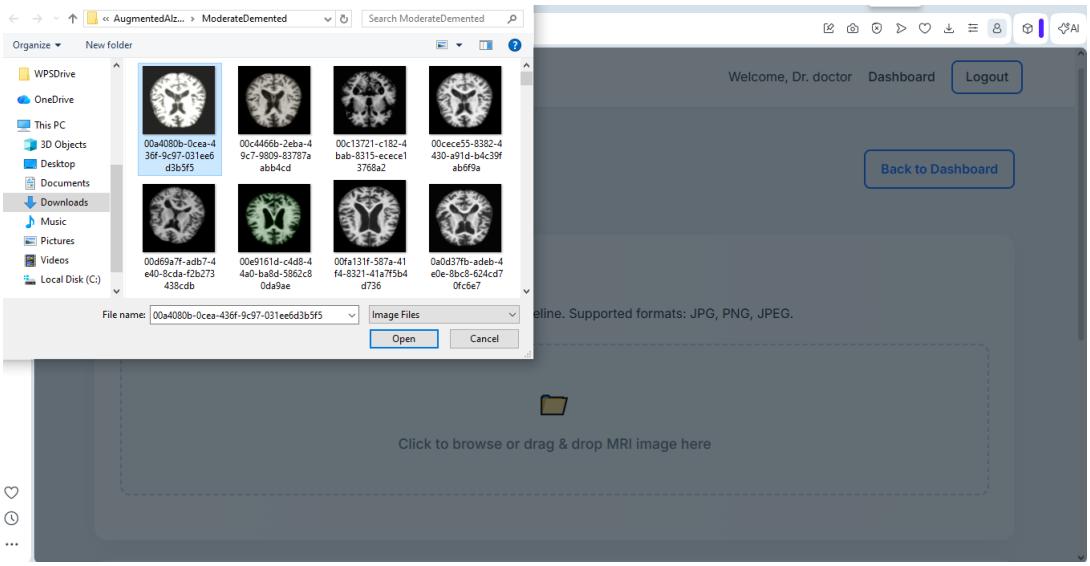


FIGURE 4.17 – Upload IRM

La Figure 4.18 représente l'interface de lancement du pipeline d'intelligence artificielle (AI Pipeline) de la plateforme Neuro Scan AI. Après l'importation de l'image IRM, le médecin peut démarrer le processus d'analyse en cliquant sur le bouton « Run AI Pipeline ». Cette action permet de lancer automatiquement les différentes étapes de traitement et de classification de l'image.

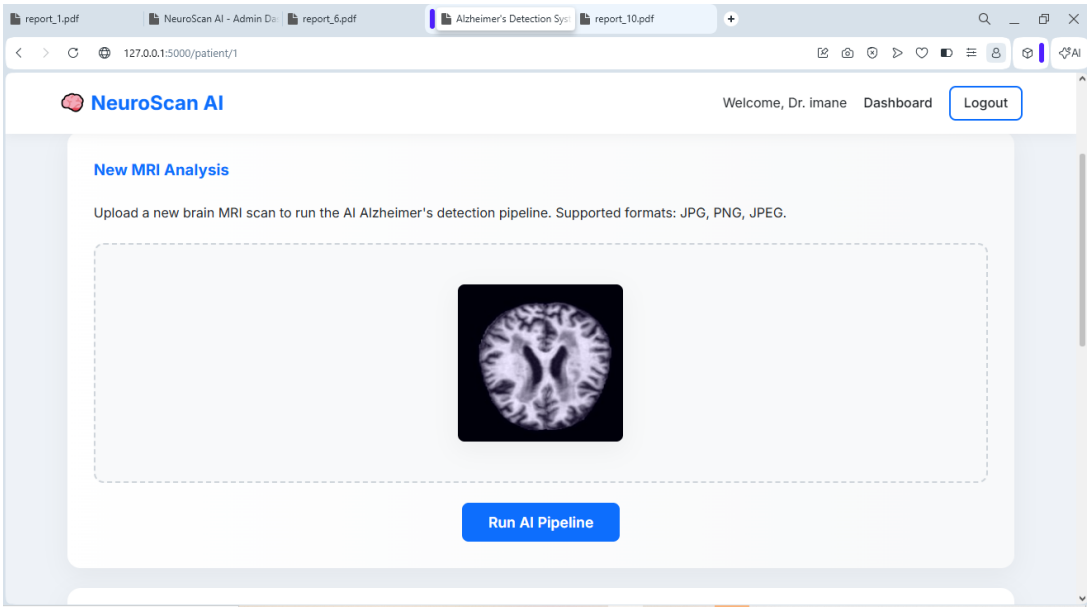


FIGURE 4.18 – Run AI Pipeline

La Figure 4.19 représente l'interface d'affichage des résultats de classification. Après l'exécution du pipeline d'intelligence artificielle, le système affiche la classe prédite du patient. Dans cet exemple, le modèle a classifié l'image IRM dans la catégorie « Moderate Demented », indiquant un stade modéré de la maladie d'Alzheimer.

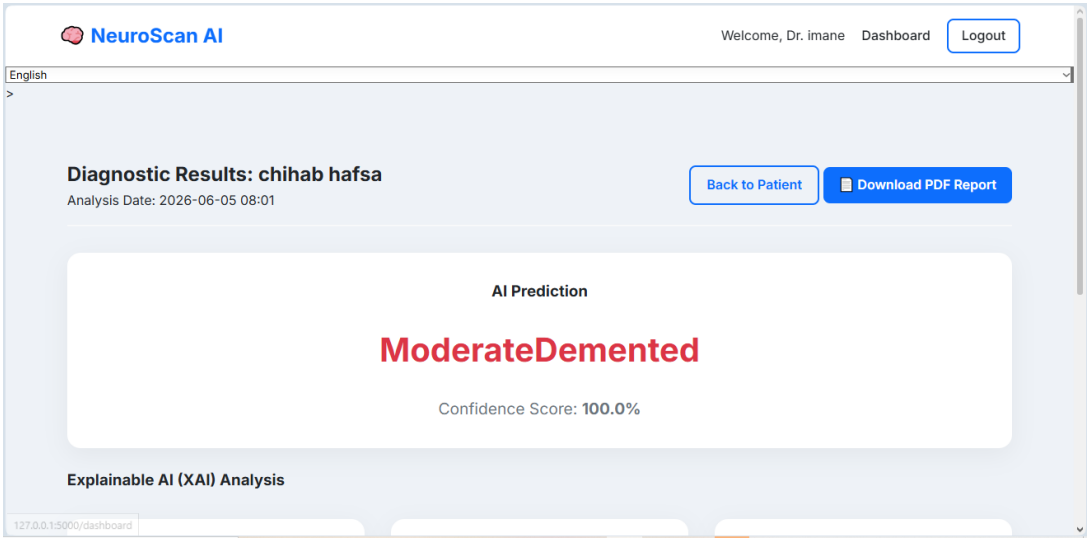


FIGURE 4.19 – Résultat classification Moderate Dementeed

La Figure 4.20 représente le rapport PDF généré automatiquement après l'analyse de l'image IRM. Ce document contient les informations du patient, le résultat de la classification, ainsi que les visualisations Grad-CAM et SHAP permettant d'interpréter la décision du modèle. Il facilite la consultation, l'archivage et le partage des résultats médicaux.

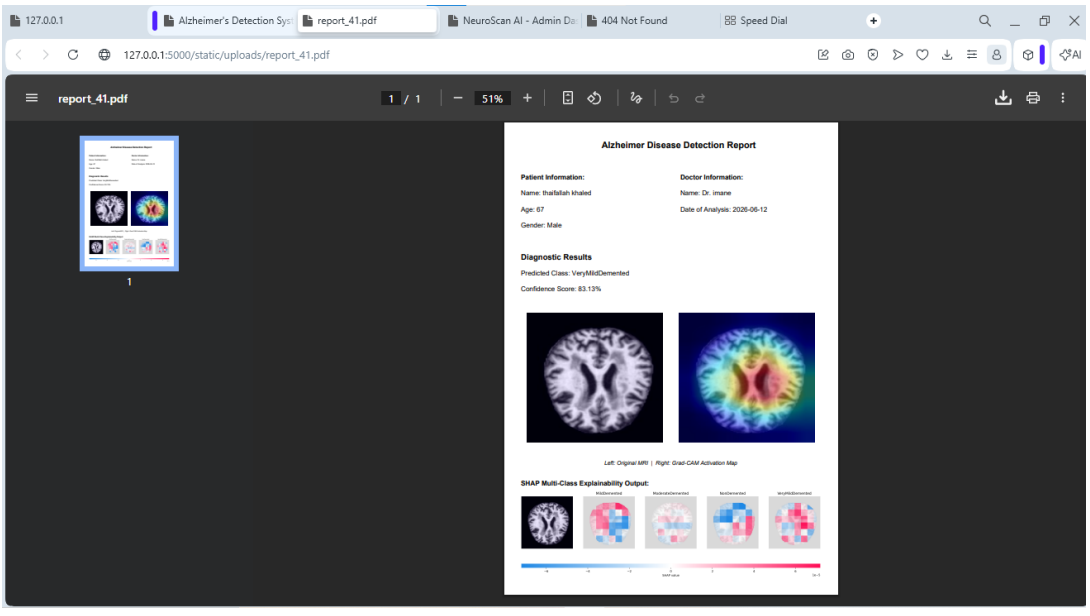


FIGURE 4.20 – Génération automatique du rapport PDF

4.8.2 Admin login

La Figure 4.21 représente l'interface d'authentification de l'administrateur. Elle permet d'accéder de manière sécurisée aux fonctionnalités de gestion de la plateforme à l'aide d'un identifiant et d'un mot de passe.

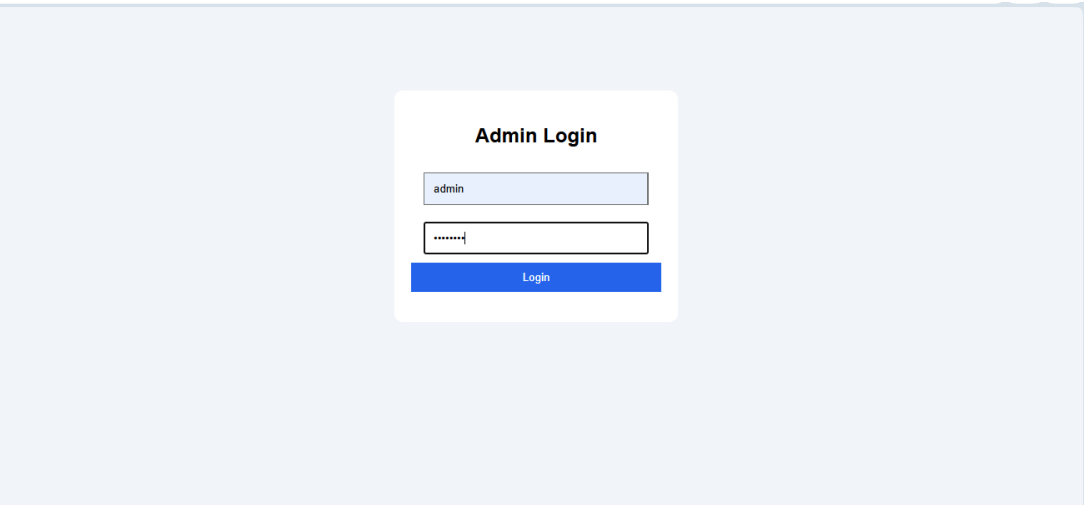


FIGURE 4.21 – Admin login

4.8.3 Admin dashboard

La Figure 4.22 représente le tableau de bord de l'administrateur. Cette interface offre une vue globale sur les activités de la plateforme et permet d'accéder aux différentes fonctionnalités de gestion.

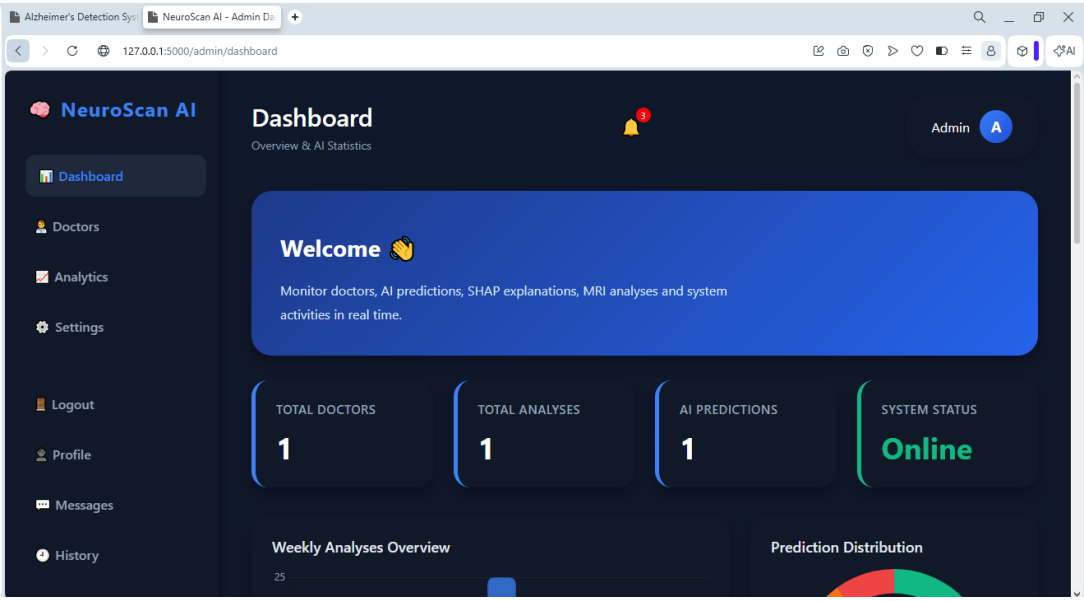


FIGURE 4.22 – Dashboard Admin

La Figure 4.23 représente l'interface de gestion des médecins. Elle permet à l'administrateur de consulter, ajouter, modifier ou supprimer les comptes des médecins enregistrés dans la plateforme.

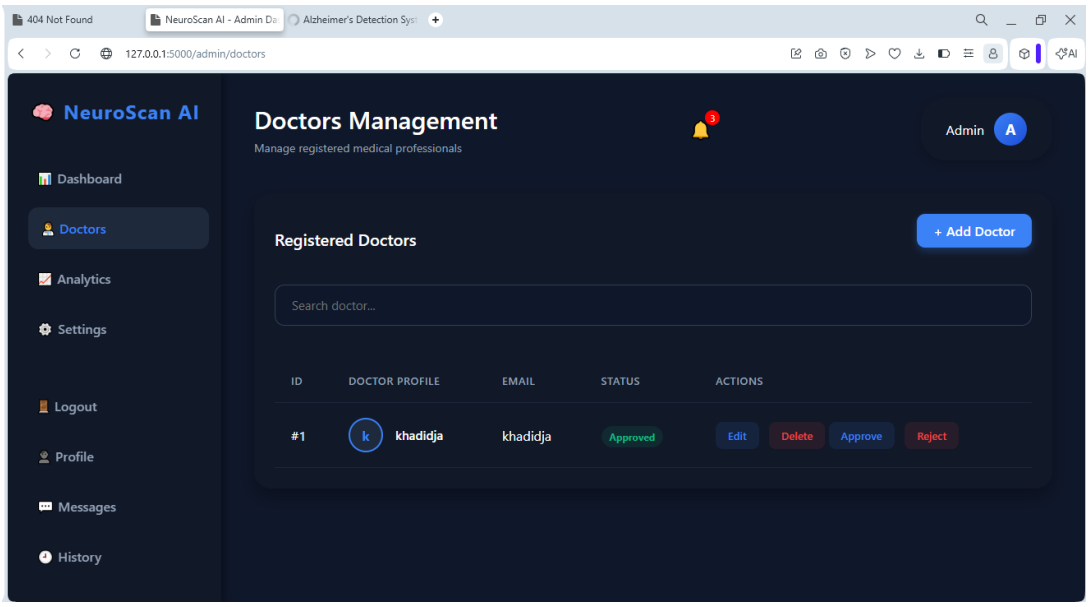


FIGURE 4.23 – Gestion des médecins par l'administrateur

La Figure 4.24 représente l'interface des paramètres de l'administrateur. Elle permet de gérer les informations du compte ainsi que certaines configurations de la plateforme.

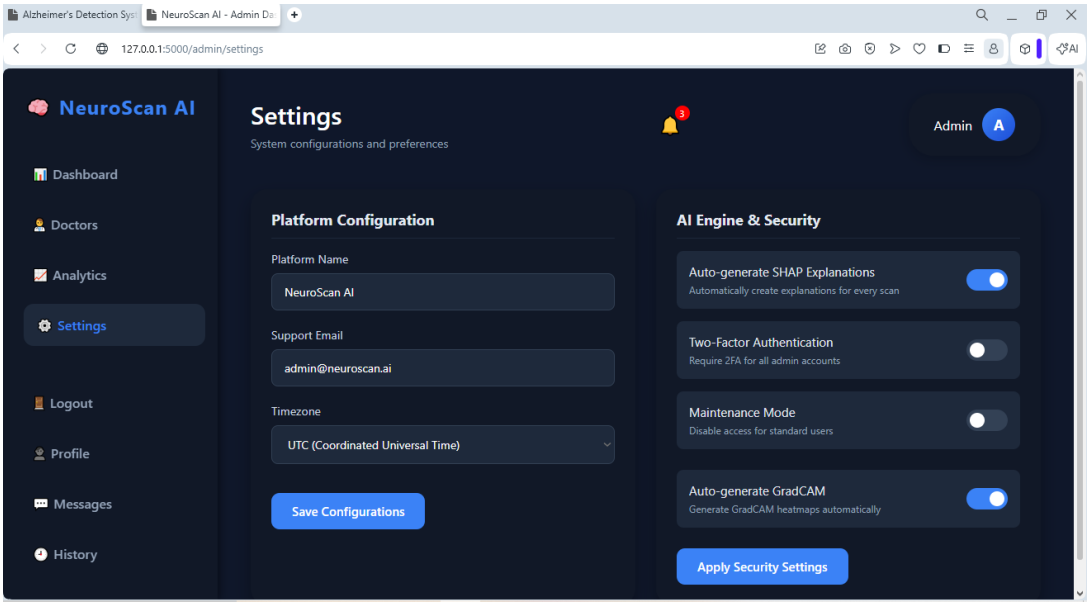


FIGURE 4.24 – Paramètres de l'administrateur

La Figure 4.25 représente l'interface d'envoi des messages aux médecins. Cette fonction-

nalité permet à l'administrateur de communiquer avec les utilisateurs et de transmettre des notifications ou des informations importantes.

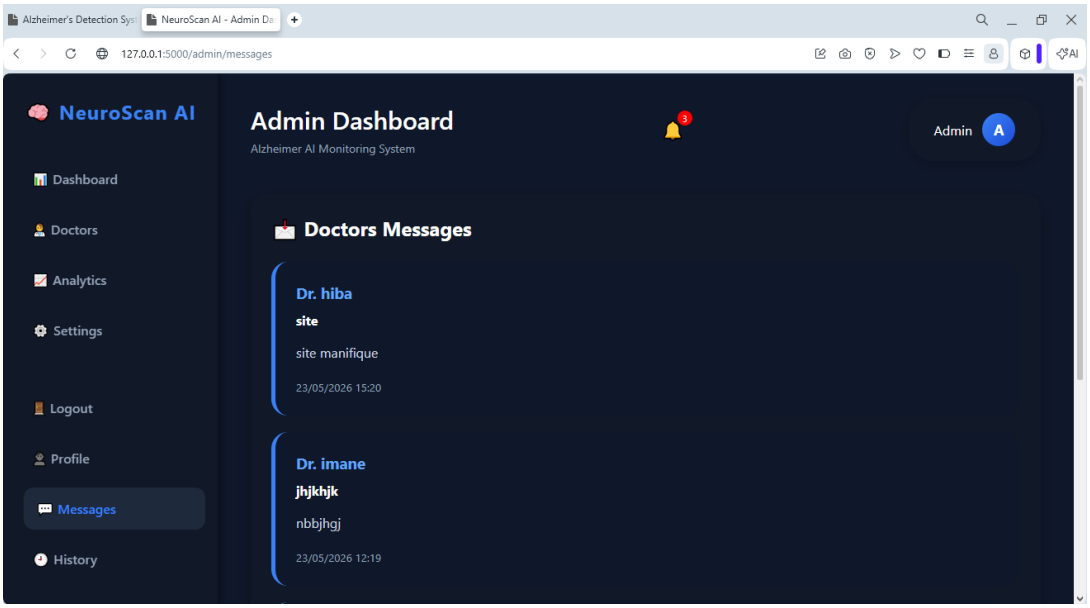


FIGURE 4.25 – Interface administrateur pour l'envoi des messages aux médecins

4.9 Contributions du projet

Notre projet présente plusieurs contributions importantes :

• Développement d'une plateforme médicale intelligente • Intégration de modèles CNN pré-entraînés • Classification automatique de la maladie d'Alzheimer • Intégration des techniques Explainable AI (Grad-CAM et SHAP) • Développement d'interfaces dédiées aux médecins et administrateurs • Génération d'analyses médicales interprétables

4.9.1 Limites du système

Malgré les excellents résultats obtenus, notre système présente certaines limites. Le dataset utilisé reste relativement limité comparé aux grandes bases de données médicales internationales. De plus, les performances peuvent varier selon la qualité des images IRM et les équipements médicaux utilisés.

Par ailleurs, l'application actuelle fonctionne principalement comme un système d'assistance au diagnostic et ne remplace pas l'expertise médicale professionnelle.

Dans les travaux futurs, nous prévoyons d'intégrer davantage de données cliniques, d'améliorer les performances temps réel et de développer une version mobile intelligente destinée aux médecins.

4.10 Conclusion

Dans ce chapitre, nous avons présenté l'implémentation complète de notre système intelligent de détection de la maladie d'Alzheimer basé sur les images IRM cérébrales.

Les différents modèles expérimentés ont montré des performances très élevées, notamment l'approche Xception + KNN qui a obtenu les meilleurs résultats avec une accuracy de 99.74

Nous avons également intégré des méthodes d'intelligence artificielle explicable telles que Grad-CAM et SHAP afin d'améliorer la transparence et l'interprétation des prédictions médicales.

Enfin, nous avons développé une plateforme web intelligente permettant aux médecins de charger des images IRM, obtenir des prédictions automatiques et visualiser les explications associées.

Ces résultats démontrent le potentiel important de l'intelligence artificielle dans l'assistance au diagnostic précoce de la maladie d'Alzheimer.

Conclusion Générale

Dans ce travail, nous avons proposé une approche intelligente basée sur l'utilisation des réseaux de neurones convolutifs (CNN) combinés avec des algorithmes de Machine Learning pour la classification des stades de la maladie d'Alzheimer à partir des images IRM cérébrales.

Plusieurs architectures de Deep Learning ont été étudiées, notamment Xception, MobileNet et EfficientNet, utilisées comme extracteurs de caractéristiques profondes. Les caractéristiques extraites ont ensuite été exploitées par différents classifieurs de Machine Learning tels que SVM, KNN et Random Forest afin d'améliorer les performances de classification.

Les résultats expérimentaux obtenus ont montré que les approches hybrides CNN + Machine Learning permettent d'obtenir des performances très élevées. Le modèle Xception combiné avec SVM a présenté les meilleurs résultats avec une accuracy supérieure à 99%, une excellente capacité de généralisation ainsi qu'une forte robustesse dans la classification des différentes catégories de la maladie d'Alzheimer.

L'intégration des techniques d'intelligence artificielle explicable, notamment Grad-CAM et SHAP, a permis de mieux interpréter les décisions des modèles et de visualiser les régions importantes des images IRM ayant influencé les prédictions. Ces approches améliorent considérablement la transparence du système et renforcent la confiance dans les résultats obtenus, ce qui représente un aspect essentiel dans le domaine médical.

Dans le cadre de ce projet, une plateforme médicale intelligente basée sur Flask a également été développée afin de permettre la gestion des patients, le chargement des images IRM, l'exécution des modèles de classification ainsi que la visualisation des résultats et des interprétations explicables.

Malgré les résultats très satisfaisants obtenus, certaines limites restent présentes, notamment

le déséquilibre du dataset, la taille limitée des données disponibles ainsi que les contraintes liées aux ressources matérielles utilisées lors des expérimentations.

Comme perspectives futures, il serait intéressant d'utiliser des datasets médicaux plus larges et plus diversifiés, d'intégrer d'autres architectures de Deep Learning plus récentes, ainsi que de déployer la plateforme dans un environnement cloud afin d'améliorer son accessibilité et ses performances en temps réel.

Comme perspectives futures, nous envisageons :

• Développer une application mobile • Intégrer davantage de modèles d'intelligence artificielle • Ajouter des fonctionnalités de suivi médical • Déployer la plateforme dans des établissements de santé • Améliorer les performances du système avec de nouvelles données

Enfin, ce travail démontre l'intérêt et l'efficacité des approches hybrides combinant Deep Learning, Machine Learning et intelligence artificielle explicable dans le développement de systèmes médicaux intelligents destinés à l'aide au diagnostic précoce de la maladie d'Alzheimer.

Références

[1] ResearchGate, "SHAP output of correctly classified images," Computer Modeling in Engineering & Sciences, 2024, consulté le 12 juin 2026. [Online]. Available : https://www.researchgate.net/figure/SHAP-output-of-correctly-classified-images_fig15_381249045

[2] Y. S. Rahal, "Une approche hybride pour la classification des irm de la maladie d'alzheimer : Combinaison de modèles cnn et de classifieurs machine learning avec augmentation de données," Mémoire de Master, Université de Bordj Bou Arréridj (BBA), 2025.

[3] K. Benali Khodja and A. Guennaz, "Classification des stades de la rétinopathie diabétique à l'aide des techniques d'apprentissage profond," Mémoire de master, Universitè de Tlemcen, 2023, mention : Génie Biomédical.

[4] Fondation pour la Recherche Médicale. (2020) Maladie d'alzheimer : facteurs de risque et facteurs protecteurs. [Online]. Available : https://www.frm.org/

[5] P. Krolak-Salmon, "Physiopathologie de la maladie d'alzheimer : le rôle central de la plaque amyloïde et de la protéine tau," NPG Neurologie - Psychiatrie - Gériatrie, vol. 20, no. 120, Supplement, pp. 120S2–120S6, 2020.

[6] A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta, and A. Upadhya, "A cnn model : Earlier diagnosis and classification of alzheimer disease using mri," in 2020 International Conference on Smart Electronics and Communication (ICOSEC). Trichy, India : IEEE, Sep. 2020, pp. 1–6.

[7] Alzheimer's Association. (2022) Les 7 stades de la maladie d'alzheimer. [Online]. Available : https://www.alz.org/

[8] A. Pouchon, E. Fakra, F. Haesebaert, G. Legrand, M. Rigon, E. Schmitt, P. Conus, T. Bou-

gerol, M. Polosan, and C. Dondé, “Intervenir précocement dans les stades débutants du trouble bipolaire : pourquoi, quand et comment,” L’Encéphale, vol. 48, no. 1, pp. 60–69, February 2022.

[9] K. Sesse, “Système d’aide au diagnostic pour la détection précoce de la maladie d’alzheimer,” Master’s thesis, Université Ibn Khaldoun – Tiaret, June 2025.

[10] G. B. Frisoni and et al., “The clinical use of structural mri in alzheimer disease,” Nature Reviews Neurology, vol. 6, no. 2, pp. 67–77, 2010.

[11] I. G. McKeith, B. F. Boeve, D. W. Dickson, G. Halliday, J.-P. Taylor, D. Weintraub, D. Aarsland, J. Galvin, J. Attems, C. G. Ballard, A. Bayston, T. G. Beach, F. Blanc, N. Bohnen, L. Bonanni, J. Bras, P. Brundin, D. Burn, A. Chen-Plotkin, J. E. Duda, O. El-Agnaf, H. Feldman, T. J. Ferman, D. Ffytche, H. Fujishiro, D. Galasko, J. G. Goldman, S. N. Gomperts, N. R. Graff-Radford, L. S. Honig, A. Iranzo, K. Kantarci, J. B. Leverenz, S. Lewis, C. Lippa, A. Lunde, M. Masellis, E. Masliah, P. McLean, B. Mollenhauer, T. J. Montine, E. Moreno, E. Mori, M. Murray, J. T. O’Brien, S. Orimo, R. B. Postuma, S. Ramaswamy, O. A. Ross, D. P. Salmon, A. Singleton, A. Taylor, A. Thomas, P. Tiraboschi, J. B. Toledo, J. Q. Trojanowski, D. Tsuang, Z. Walker, M. Yamada, and K. Kosaka, “Diagnosis and management of dementia with lewy bodies : Fourth consensus report of the dlb consortium,” Neurology, vol. 89, no. 1, pp. 88–100, 2017.

[12] A. Chandra, G. Dervenoulas, and M. Politis, “Magnetic resonance imaging in alzheimer’s disease and mild cognitive impairment,” Journal of Neurology, vol. 266, no. 6, pp. 1293–1302, 2019.

[13] National Institute on Aging, “Pet scan-normal brain-alzheimers disease brain,” 2026, image diffus’ee via Wikimedia Commons, domaine public.

[14] R. Sun, E. Deutsch, and L. Fournier, “Intelligence artificielle et imagerie médicale,” Bulletin du Cancer, vol. 109, no. 1, pp. 83–88, January 2022, received 26 May 2021, accepted 2 September 2021.

[15] E. Villain, “Utilisation de l’intelligence artificielle pour l’aide au diagnostic des patients atteints de pathologies neuro dégénératives,” Theses, Université Paul Sabatier - Toulouse III, Dec. 2021. [Online]. Available : https://theses.hal.science/tel-03684132

[16] M. R. Shaikh, A. Jeyabose, and R. V. Arjunan, “Deep learning for alzheimer’s disease : advances in classification, segmentation, subtyping, and explainability,” BioMedical Engineering OnLine, vol. 24, no. 1, p. 150, 2025.

[17] IBM, “What is machine learning ?” https://www.ibm.com/topics/machine-learning, 2026, [En ligne ; consulté le 30-mars-2026].

[18] J. Rimmer, P. Harrison, and ital., “An introduction to convolutional neural networks for image processing,” Journal of Deep Learning Applications, vol. 14, no. 2, pp. 112–128, 2018.

[19] IBM, “Que sont les réseaux de neurones convolutifs (cnn) ?” https://www.ibm.com/fr-fr/topics/convolutional-neural-networks, 2024, [En ligne ; accédé le 16-juin-2026].

[20] M. Nennouche and A. M. Atchi, “Fusion de caractéristiques pour la classification des différents niveaux de démence de la maladie d’alzheimer,” Mémoire de Master, Université, 2023.

[21] Encord AI Glossary, “Understanding dropout layers in deep learning architecture,” https://encord.com/glossary/dropout-definition/, 2023, consult’e le 10 juin 2026.

[22] M. A. Morid, A. Borjali, and G. Del Fiol, “A scoping review of transfer learning applications in medical image analysis,” BMC Medical Informatics and Decision Making, vol. 21, pp. 1–18, 2021.

[23] IBM Cloud Education, “A taxonomy of transfer learning and its applications,” https://www.ibm.com/topics/transfer-learning, 2023, consult’e le 11 juin 2026.

[24] A. Esteva and A. Robicquet, “Applications of deep learning in medical imaging,” Nature Medicine, vol. 27, pp. 123–134, 2021.

[25] S. Sarraf and G. Tofighi, “Alzheimer’s disease classification using deep learning on mri images,” IEEE Access, vol. 10, pp. 45 678–45 690, 2022.

[26] M. Tan and Q. Le, “Efficientnet : Rethinking model scaling for convolutional neural networks,” in International conference on machine learning. PMLR, 2019, pp. 6105–6114.

[27] B. Koonce, EfficientNet. Berkeley, CA : Apress, 2021.

[28] M. Tan and Q. V. Le, "Efficientnet : Rethinking model scaling for convolutional neural networks," Proceedings of the International Conference on Machine Learning (ICML), 2021.

[29] M. Raghu and C. Zhang, "Deep learning applications in medical image analysis using efficientnet," IEEE Access, vol. 10, pp. 112 345–112 356, 2022.

[30] A. G. Howard, M. Zhu, B. Chen, and D. Kalenichenko, "Mobilenets : Efficient convolutional neural networks for mobile vision applications," IEEE, 2021.

[31] S. Siddharth and N. Sujatha, "Multiple grapes leaf disease identification using an optimal deep learning model : Xception," in 2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN). Villupuram, India : IEEE, April 2023, pp. 1–6.

[32] F. Chollet, "Xception : Deep learning with depthwise separable convolutions," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[33] W. S. Noble, "Support vector machines in medical applications," Nature Biotechnology, vol. 38, pp. 123–130, 2022.

[34] M. Odusami, R. Maskeliūnas, and R. Damaševičius, "An intelligent system for early detection of alzheimer's disease using machine learning on mri images," Diagnostics, vol. 12, no. 2, p. 328, 2022.

[35] R. Jack and E. Smith, "Application of random forests algorithm in medical data classification : A contemporary review," Biomedical Signal Processing and Control, vol. 82, p. 104520, 2023.

[36] D. Chicco and G. Jurman, "Guidelines for evaluation metrics selection in medical image classification tasks," IEEE Access, vol. 12, pp. 15 432–15 445, 2024.

[37] F. Doshi-Velez and B. Kim, "Explainability in artificial intelligence," Harvard University, 2022.

[38] IBM. (2024) What is Explainable AI ? Présentation claire des concepts et des

avantages de l'XAI. Think Topics. Consulté le 12 juin 2026. [Online]. Available : https://www.ibm.com/think/topics/explainable-ai

[39] A. V. Ponce-Bobadilla, V. Schmitt et al., "Practical guide to SHAP analysis : Explaining supervised machine learning model predictions in drug development," Clinical and Translational Science, 2024, cité 720 fois.

[40] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam : Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626.

[41] "alzheimers-dataset-4-class-of-images." [Online]. Available : https://www.kaggle.com/datasets/psraju2024/alzheimers-dataset-4-class-of-images