

Master of Science conjoint HES-SO - UNIL en Sciences de la santé Orientation Technique en radiologie médicale

L'intelligence artificielle au service de la détection de la maladie de Parkinson à l'Iode-123 Ioflupane en Médecine Nucléaire

Adrien Arrigo
Guillaume Favre-Gillioz

Sous la direction de
Prof. Jérôme Schmid
De la Haute Ecole de Santé de Genève

Sous la co-direction de
Dr. Valentina Garibotto
Du Service de Médecine Nucléaire
des Hôpitaux Universitaires de Genève

Expert : Dr. Jonas Richiardi
Du département de radiologie
du Centre Hospitalier Universitaire Vaudois

Lausanne, HES-SO Master, 2021

Remerciements

Nous tenons tout d'abord à chaleureusement remercier notre directeur, Docteur Schmid Jérôme, Professeur ordinaire à la Haute Ecole de Santé de Genève, qui nous a guidé et nous a fournis de précieux conseils tout au long de cette étude. Nous remercions ensuite notre co-directrice Doctoresse Valentina Garibotto pour son expertise clinique et ces corrections.

Notre reconnaissance s'adresse également au Docteur Nicolas Nicastro pour ces conseils sur la base de données PPMI et son apprentissage à ce sujet, ainsi que le partage de données.

Nous remercions aussi M. Denis Brault, TRM en Service de Médecine nucléaire des HUG, pour le partage d'informations sur les protocoles ainsi que pour la collecte de données.

Enfin, nous remercions l'Expert du jury, Docteur Richiardi Jonas, responsable de recherche clinique dans le Département de Radiologie du CHUV, pour son expertise en Apprentissage Machine ainsi que ses corrections.

Table des matières

1	L'intelligence artificielle.....	15
1.1	Bref historique.....	15
1.2	IA et Apprentissage machine.....	15
1.3	Apprentissage profond (Deep Learning).....	17
1.3.1	Principe et fonctionnement.....	17
1.3.2	Application du DL en radiologie.....	21
1.3.3	Type de réseaux neuronaux.....	22
1.4	Données.....	24
1.4.1	Données d'entraînement et de validation.....	25
1.4.2	Données de test.....	26
1.5	Préparation des données.....	27
1.5.1	Pre-processing.....	27
1.5.2	Augmentation.....	30
1.5.3	Balancing.....	32
1.6	Evaluation du réseau.....	32
1.6.1	Mesure quantitative.....	33
1.6.2	Evaluation visuelle.....	36
2	Maladie de Parkinson.....	39
2.1	Anatomo-physiopathologie.....	39
2.2	Méthode diagnostique.....	41
2.2.1	Pré-diagnostic.....	41
2.2.2	Diagnostic Clinique.....	42
2.2.3	Examen au DaTSCAN.....	43
3	Diagnostic assisté par ordinateur de la maladie de Parkinson.....	47
3.1	Etudes sur PPMI.....	48
3.1.1	Machine Learning classique.....	48
3.1.2	Deep Learning.....	48
3.2	Normalisation des images.....	49
3.2.1	Normalisation géométrique.....	49
3.2.2	Normalisation d'intensité.....	50
3.2.3	Masquage.....	51
3.3	Traitement des images en ML.....	51
3.4	Traitement des images en DL.....	52
3.5	Augmentation des données.....	52

3.6	Transfer Learning.....	52
4	Rôles et compétences du TRM en relation avec l'IA.....	53
4.1	Un environnement de travail en plein évolution.....	53
4.2	Les prérequis techniques nécessaires.....	54
4.2.1	Base de mathématiques et statistiques.....	54
4.2.2	Langage de Programmation.....	54
4.2.3	Gestion des données	55
4.2.4	Evaluation de l'IA	56
4.2.5	Entrainement des utilisateurs.....	56
4.2.6	Assurance qualité	56
5	Méthode.....	57
5.1	Question de recherche et objectifs	57
5.2	Collecte des bases de données	58
5.3	Choix du réseau et paramétrage	61
5.4	Evaluation visuelle des bases de données	62
5.5	Evaluation quantitatives des bases de données	63
5.6	Standardisation des données	64
5.7	Scénarios des tests	65
6	Résultats.....	67
6.1	Evaluation quantitative des prédictions	67
6.1.1	Evaluations de données issues de la même base de données	68
6.1.2	Evaluation de données issues d'une autre base d'entraînement	70
6.1.3	Evaluation de données partiellement représentées dans les données d'entraînement	70
7	Discussion.....	73
7.1	Performances du réseau par rapport à l'état de l'art	73
7.2	Evaluation intra-base de données	73
7.2.1	Préparation des données.....	74
7.2.2	Evaluation interbase de données	77
7.3	Evaluation interbase mélangée	78
7.3.1	Mélange de deux bases de données	79
7.3.2	Mélange de trois bases de données	80
7.3.3	Transfer Learning.....	82
7.3.4	Evaluation des CAM.....	82
7.3.5	Évaluation du temps d'opération du réseau.....	84
7.4	Rôles du TRM et recommandations	84

7.5	Limites et perspectives de recherches.....	87
7.5.1	Volume de données à prendre en compte	87
7.5.2	Extraction striatale.....	87
7.5.3	Normalisation d'intensité	88
7.5.4	Lissage des images.....	88
7.5.5	Augmentation des données.....	88
7.5.6	Balancing de classe	89
7.5.7	Balancing des bases de données.....	89
7.5.8	Utilisation de données cliniques	89
7.5.9	Transfer Learning.....	90
8	Conclusion.....	91
9	Références.....	93
10	Annexes.....	103
10.1	Tableau récapitulatif des études antérieures	103
10.2	CAD et contrôles qualité.....	106
10.3	Architecture du réseau utilisé.....	108
10.4	Tableau récapitulatif des différents entraînements et tests.....	111

Liste des tableaux

Tableau 1 : Symptômes pré-diagnostics	41
Tableau 2 : Procédure au DaTSCAN	44
Tableau 3 : Paramètres d'acquisition d'image.....	45
Tableau 4 : Paramètres de traitement d'images.....	45
Tableau 5 : Représentation du volume de données de chaque base en : Nombre de patients contrôles (N contrôle), en nombre de patients atteints de la Maladie de Parkinson (N Parkinson), ainsi que leur nombre additionné.	60
Tableau 6: Récapitulation du nombre de données de la base Symbia après ajout des 40 images.	60
Tableau 7 : score de la qualité d'image des différentes bases de données, 5 = excellent, 4 = très bon, 3 = bon, 2 = mauvais, 1 = très mauvais.	62
Tableau 10: Code instauré aux tableaux de résultats.	67
Tableau 11 : Exemples de codes pour les résultats avec leur signification.	68
Tableau 12: Résultats des données PPMI.....	68
Tableau 13: Résultats des données Toshiba.	69
Tableau 14: Résultats des données Symbia.	69
Tableau 15: Résultats des évaluations de données issues d'une autre base d'entraînement.	70
Tableau 16: Résultats des évaluations PPMI et Symbia.	70
Tableau 17: Résultats des évaluations de PPMI + Symbia augmentée géométriquement. ...	71
Tableau 18: Résultats des données Symbia augmentées géométriquement + Toshiba.	71
Tableau 19: Résultats des données PPMI + Symbia + Toshiba.....	72
Tableau 20: Résultats des données PPMI + Symbia augmentée géométriquement + Toshiba.	72
Tableau 21 : Résultats des données de Transfert Learning PPMI augmentées géométriquement vers Symbia.....	72
Tableau 22 : Récapitulation du nombre d'erreurs systématiques insérées par la nouvelle base de données de test	74
Tableau 23 : Liste d'études CAD au DaTSCAN.....	104
Tableau 24 : Différents entraînements et tests effectués avec : Sp (normalisation des données sous SPM), Soo (lissage des images), GrBal (balancing de groupe), AugGéo (augmentation géométrique des données) et TL (Transfert Learning).	112

Liste des figures

Figure 1: Différentes méthode d'apprentissage, (Sharma, 2017)	16
Figure 2 De l'IA au Deep Learning (Fujita, 2020).....	18
Figure 3 : Perceptron (Sharma, 2021).....	19
Figure 4: Les couches d'un Réseau Neuronal (Formation personnalisée: procédure pas à pas TensorFlow Core, 2020)	20
Figure 5 : Evolution du CAD (Fujita, 2020)	22
Figure 6 : Exemple de structure d'un CNN (Sharma, 2017).....	23
Figure 7 : Exemple de l'application du Max Pooling (Sharma, 2017).....	24
Figure 8 : Performance du CAD et nombre de données (Fujita, 2020).....	26
Figure 9: Normalisation du ventricule gauche sur l'axe court (Wong et al., 2016)	29
Figure 10: Distribution de courbes d'intensité (Zhao et al., 2018)	30
Figure 11 : Chat augmenté (Despois, 2018).....	31

Figure 12 : Balance de blanc sur des images (Galdran et al., 2017)	31
Figure 13 : La courbe ROC, (Delacour et al., 2020)	33
Figure 14 : Changement du seuil (Calmant, 2008)	34
Figure 15 : AUC d'un modèle idéal (Narkhede, 2019)	35
Figure 16 : AUC d'un modèle avec un AUC suboptimal (Narkhede, 2019)	35
Figure 17 : Formule du taux d'exactitude (Accuracy (error rate), 2019)	36
Figure 18 : Processus la carte d'activation de classe (Zhou et al., 2016)	37
Figure 19 : Exemple de carte de saillance (Simonyan et al., 2014)	38
Figure 20 : Différentes parties d'un CNN examiné par la technique de visualisation des caractéristiques (Olah et al., 2017)	38
Figure 21 : Constructions... ou le ciel ? (Olah et al., 2017)	39
Figure 22: Anatomie du cerveau (Dubois-Mazeyrie, 2017)	40
Figure 23 : SPECT-CT (« GE Healthcare Unveils Discovery NM 630 SPECT Medgadget », 2011)	44
Figure 24 : Analyse Semi-quantitative (Shimizu et al., 2017)	47
Figure 25 : Template MNI (Garcia-Gomez et al., 2013)	50
Figure 26 : Normalisation d'intensité cérébrale (Castillo-Barnes et al., 2018)	51
Figure 27: L'image « 1 » provient de la base de données Symbia, la « 2 » de la base de données Toshiba et la « 3 » de la base de données PPMI. Ces images sont non normalisées, hormis la « 3 », la base PPMI ayant été récoltée avec une pré-normalisation..	59
Figure 28 : Images comparatives du DaTSCAN d'un patient sain (1) et atteint de la Maladie de Parkinson de grade 3 (2). Les signes classiques observés sont : une asymétrie gauche/droite des noyaux gris ainsi que leur forme. En rouge la différence de symétrie des noyaux est observable, ainsi que la forme hétérogène de ceux-ci dans l'image (2). A noter que le reste de l'encéphale apparaît plus sombre sur l'image « saine » par la haute concentration du radiotracer dans les noyaux gris, pour un patient pathologique, la concentration faible du radiotracer dans les noyaux gris se manifeste par une gamme dynamique plus faible dans l'image.	59
Figure 29: Critères de recherche PPMI (PPMI data base, 2020). Les cases cochées en bleu représentent les catégories sélectionnées afin de n'avoir que les images de SPECT avec le bon radiotracer, les patients pathologiques, contrôles et SWEDD, ainsi que celles déjà normalisées (« Parkinson's Progression Markers Initiative Search Criteria », 2020)	60
Figure 30: Réseau neuronal avec des raccourcis de connexion (He et al., 2016a)	61
Figure 31: Histogrammes des intensités des lots d'images de chaque base de données, respectivement Toshiba (1), Symbia (2), Symbia normalisée (3) et PPMI (4). Le bruit de fond de l'image est retiré afin d'avoir une répartition claire des courbes	63
Figure 32 : Images comparatives de la même imagerie DatSCAN provenant de la base de données Symbia respectivement sans et avec normalisation spatiale utilisant SPM12	64
Figure 33: Représentation des différents types d'entraînements et tests présents dans les résultats, 1) Evaluation des données issues de la même base d'entraînement, 2) Evaluation de données issues d'une autre base d'entraînement et 3) Evaluation de données partiellement représentées dans les données d'entraînement.	67
Figure 34 : Evaluation des données issues de la même base d'entraînement	74
Figure 35 : Evaluation de données issues d'une autre base d'entraînement	77
Figure 36 : Evaluation de données partiellement représentées dans les données d'entraînement	78
Figure 37 : Comparatif des images Symbia normalisées sous SPM12 (1) et Toshiba (2)	80
Figure 38: Superposition d'une CAM et d'une imagerie DaTSCAN d'un patient sain type	83

Figure 39 : Superposition d'une CAM et d'une imagerie DaTSCAN d'un patient pathologique type.....	83
Figure 40: Superposition d'une CAM et d'une imagerie DaTSCAN avec répartition irrationnelle.....	84
Figure 41: Superposition d'une CAM et d'une imagerie DaTSCAN avec répartition irrationnelle.....	84
Figure 42 Diagramme illustrant notre architecture ResNet.....	110

Liste des abréviations

Sigle	Acronyme anglais	Équivalent en français
ACC	Accuracy	Exactitude
AUC	Area Under the Curve	Aire sous la courbe
ANN	Artificial Neural Network	Réseau Neuronal Artificiel
CAD	Computed Assisted Diagnosis	Aide au Diagnostic Assisté
CADSC	Computer Aided Image Analysis Subcommittee	Comité de Détection Assistée par Ordinateur en Imagerie Diagnostique
CAM	Class Activation Mapping	Cartographie par Activation de Classe
CNN	Convolutional Neural Network	Réseau de neurones convolutionnels
CT	Computerized Tomography	Tomographie computerisée
Dataset	Dataset	Ensemble de données
DAT	Dopaminergic Active Transporter	Transporteur Dopaminergique Actif
DICOM	Digital imaging and Communications in Medicine	Imagerie numérique et Communications en Médecine
DL	Deep Learning	Apprentissage en profondeur
DLTK	Deep Learning Toolkit	Boîte à outils d'apprentissage profond
DNN	Deep Neural Network	Réseau Neuronal Profond
FB	Fan Beam	Faisceau en éventail
FDA	Food and Drug Administration	Agence américaine des produits alimentaires et médicamenteux
FN	False Negative	Faux Négatif
FP	False Positive	Faux Positif
GAP	Global Average Pooling Layer	Couche Globale Moyenne de Mise en Commun
LEHR	Low Energy and High Resolution	Faible Energie et Haute Résolution
LEUR	Low Energy and Ultra High Resolution	Faible Energie et Ultra Haute Résolution
LPHR	Low Penetration High Resolution	Faible Pénétration et haute Résolution
ML	Machine Learning	Apprentissage Machine
PPMI	Parkinson's Progression Marker Initiative	Initiative des Marqueurs de la Progression de Parkinson
RELU	Rectified Linear Units	Unités linéaires rectifiées
ResNet	Residual Network	Réseau Résiduel
ROI	Region of Interest	Région d'intérêt
RSNA	Radiological Society of North America	Société Nord-Américaine de Radiologie

ROC	Receiver Operating System	Fonction d'Efficacité du Récepteur
SPECT	Single Photon Emission Computed Tomography	Tomographie par Emission Monophotonique
SPM	Statistical Parametric Mapping	Cartographie Statistique Paramétrique
SVM	Support Vector Machine	Machine à vecteurs de support
SWEDD	Scan Without Evidence of Dopaminergic Deficit	Scanner sans Preuve de Déficit Dopaminérgique
TL	Transfer Learning	Transfert d'Apprentissage
TN	True Negative	Vrai Négatif
TP	True Positive	Vrai Positif

Sigle	Acronyme français
ACC	Association Canadienne des Radiologues
BDD	Base de Données
HAS	Haute Autorité de Santé
HUG	Hôpitaux Universitaire de Genève
IA	Intelligence Artificielle
IC	Intervalle de Confiance
IRM	Imagerie par Résonance Magnétique
kV	Kilovolt
MN	Médecine Nucléaire
mAs	Milliampère seconde
Se	Sensibilité
SNR	Rapport Signal sur Bruit
Sp	Spécificité
TDM	Tomodensitométrie
TRM	Technicien en radiologie médicale
US	Ultrason

Résumé

Objectif : La scintigraphie cérébrale du transporteur de la dopamine (DAT) est utilisée dans le diagnostic de la maladie de Parkinson. Cependant, l'inspection visuelle des images DAT dépend du niveau d'expertise du lecteur et possède une spécificité relativement faible dans le stade précoce de la maladie. Dans ce contexte, l'intelligence artificielle (IA) pourrait aider les cliniciens en améliorant la précision du diagnostic tout en réduisant la variabilité de l'observateur. L'objectif principal de cette étude est d'évaluer les performances d'une IA sur différentes bases de données et des techniques de préparation de données. Un second objectif est d'investiguer des possibles rôles du TRM en IA et de proposer des recommandations dans la pratique.

Méthodologie : Trois bases de données ont été récoltées, deux bases locales provenant d'appareillages de constructeur différents des Hôpitaux Universitaires de Genève et une autre, plus volumineuse, provenant de la base internationale PPMI. Différentes stratégies d'entraînement du système d'IA basé sur l'apprentissage profond ont été étudiées et évaluées, telles que des techniques de normalisation et augmentation des données, ainsi que des approches de mélange des bases de données.

Résultat : Avec une augmentation géométrique, l'évaluation de la base de données PPMI a fourni une exactitude de 95.6%. En évaluant la base de données clinique, le meilleur résultat donnait une exactitude de 82.1% avec une normalisation et une augmentation des données. En faisant un transfert d'apprentissage avec la base de données locale pré-entraînée à partir de la base PPMI, une exactitude de 87.5% a été obtenue en évaluant les images locales.

Discussion : La normalisation des données est nécessaire au bon traitement des images par le réseau. Cependant, ces techniques doivent être préparées et appliquées correctement au risque de réduire les performances. Le mélange de bases a montré d'aussi bons résultats que l'utilisation de l'augmentation géométrique seule. Toutefois, le mélange de bases via le transfert d'apprentissage de la base PPMI vers les bases locales est la technique ayant montré les résultats les plus élevés.

Conclusion : Un mélange de données peut être bénéfique au travers du transfert d'apprentissage. Avec les bonnes compétences, le TRM peut apporter, grâce à son expertise, une plus-value sur l'implémentation d'une IA diagnostique.

Abstract

Purpose: Brain dopamine transporter (DAT) SPECT imaging is used in the diagnosis of Parkinson's disease (PD). However, visual inspection of DAT images depends on the reader's level of expertise and has a relatively low specificity in early PD. In this context, Artificial Intelligence (AI) could support clinicians by improving diagnostic accuracy while reducing observer variability. The primary objective of this study is to evaluate the diagnosis performance of an artificial intelligence (AI) on different databases with various data preparation techniques. A secondary objective is to investigate possible roles of radiographers in AI by giving recommendations in such practices.

Methodology: Three datasets were collected, two local datasets from different manufacturers acquired at the Geneva University Hospital, and the last larger dataset was obtained from the international database PPMI. Different training strategies of the IA based on Deep Learning were studied and evaluated, including data normalization and augmentation, as well as data fusion approaches.

Results: With a geometric augmentation, the PPMI dataset led to an accuracy of 95.6%. By evaluating the main clinical comparative dataset, the best result was obtained by using the normalization and augmentation approaches – resulting in an accuracy of 82.1%. By doing a Transfer Learning on the local dataset from the PPMI database, an accuracy of 87.5% was achieved.

Discussion: Data normalization is necessary for the proper processing of images by the network. However, the techniques should be assessed and used correctly to avoid potential drops in performances. The data fusion produced results as good as the data augmentation alone. Nonetheless, the transfer learning technique from the PPMI database to the local dataset produced the best results.

Conclusion: Data fusion can be beneficial using Transfer Learning. With the proper skills, the radiographer can bring, due to its expertise, added value on the implementation of a diagnostic AI.

1 L'intelligence artificielle

1.1 Bref historique

Selon Bini et al. (2018), l'intelligence artificielle (IA) est une notion qui peut être définie comme l'intelligence humaine démontrée par les machines. En science informatique, on la définit comme l'étude d'agents intelligents qui sont des appareils qui perçoivent leur environnement et prennent des actions pour maximiser leurs chances d'atteindre des objectifs. Bien que ce soit un phénomène qui prend de plus en plus d'ampleur, l'intelligence artificielle existe depuis plusieurs décennies. Une des premières formes d'IA ayant vu le jour est un système créé par Arthur Lee Samuel en 1952. Après avoir rassemblé les données et regardé toutes les options disponibles avant de prendre une décision, ce système était capable de jouer au jeu de dame de manière autonome. Pour que cela fonctionne, chaque mouvement possible devait être programmé dans un algorithme en y fournissant des instructions au préalable. Ces systèmes s'appuyaient sur la modélisation d'éléments correspondant au monde réel, ainsi que sur des relations logiques avec lesquelles il est possible de performer des opérations arithmétiques. Cela permettait de connaître les différentes étapes par lesquelles ce système procédait, entraînant ainsi une certaine reproductibilité. Un gros désavantage était l'étape de modélisation nécessitant un investissement très important de temps, ce qui limitait l'accessibilité à l'IA. De plus, la précision de ces modèles dépendait directement de la compréhension de l'homme sur le monde qui l'entoure ainsi que sur sa complexité.

L'IA de nos jours comporte une place grandissante dans différents domaines tels que la santé (Jiang et al., 2017). Contrairement à une IA basée sur des jeux d'échecs où il est possible d'avoir une intégration de règles et stratégies très bien délimitées, en médecine, cela est plus délicat. Il va de soit de s'accorder que le fonctionnement du corps humain est d'une grande complexité, dans laquelle les classifications les plus détaillées et approfondies de pathologies ne demeurent que des simplifications plus ou moins fortes de processus biologiques sous-jacents. Cela limite les systèmes d'IA qui se basent ainsi sur des modèles qui ne sont pas encore pleinement compris par l'homme.

1.2 IA et Apprentissage machine

Autour de l'an 1980, on commence à voir l'émergence d'une nouvelle génération de systèmes d'IA. Considéré comme une sous branche de l'IA, l'apprentissage machine, ou Machine Learning (ML) en anglais, est capable d'améliorer ses performances en apprenant par l'expérience grâce à des algorithmes. Traitant les données de manière statistique, le ML est capable d'apprendre à faire des tâches sans la nécessité de fournir des instructions. Cela leur permet de jouer sur des jeux bien plus complexes comprenant une infinité de combinaisons possibles telles que le jeu chinois Go (Stoutamire et al., 1991). C'est grâce à l'observation d'un

jeu important de données que le système peut faire une abstraction interne, permettant de s'affranchir de la phase contraignante de la modélisation.

Une des applications les plus courantes du ML est l'exploration de données (Data Mining) qui consiste à extraire des modèles et des connaissances à partir de grandes bases de données (Maglogiannis, 2007). En effet, chaque cas, chaque base de données utilisée par du ML peut être décrite avec un même ensemble de caractéristiques. Ces caractéristiques peuvent être des variables de nature continue, catégorielle ou binaire. Prenons l'exemple d'un algorithme utilisé dans la reconnaissance d'images d'espèces de fleurs d'Iris. L'algorithme doit donc reconnaître ce qui caractérise les iris grâce à des caractéristiques telles que la longueur du pétale, sa largeur et sa couleur. A la suite d'un entraînement, en prenant en compte l'ensemble de ces caractéristiques, l'algorithme va donc pouvoir les classer selon leur espèce.

Adéquatement choisi, le nombre de données nécessaires à un bon entraînement n'a pas forcément besoin d'être élevé. Cela signifie que ces caractéristiques sont stables et qu'elles devraient donc se retrouver dans la variété de données que le système est susceptible de rencontrer. Toutefois, si la caractéristique n'est pas optimalement sélectionnée, il va falloir la rechoisir, la réentraîner et la re-tester.

Il existe différentes catégories d'apprentissages qui sont l'apprentissage supervisé, l'apprentissage non-supervisé et l'apprentissage par renforcement (Lison, 2012), comme illustré dans la Figure 1.

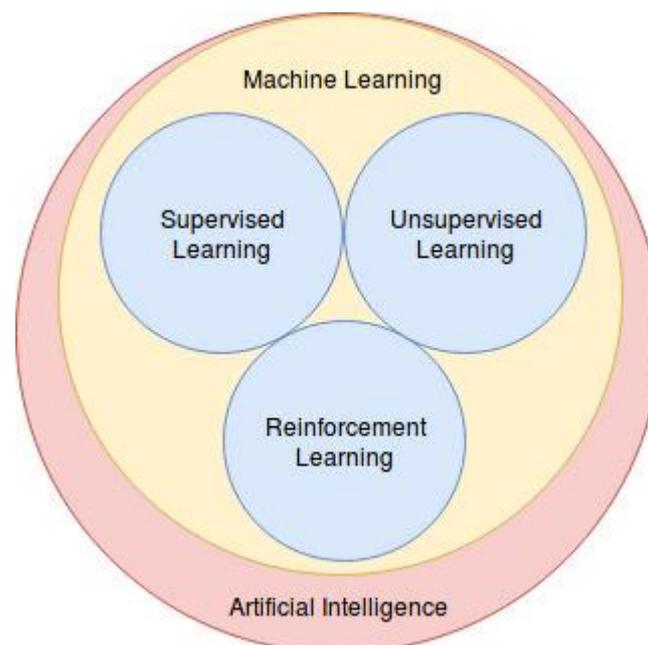


Figure 1: Différentes méthode d'apprentissage, (Sharma, 2017)

Durant l'entraînement, si les iris sont étiquetées (i.e. associées avec une information à un type de donnée) selon leur espèce et que nous indiquons quelles caractéristiques permettent de les discriminer entre elles, alors on parle d'apprentissage supervisé. C'est une technique typique du ML traditionnel. A l'opposé, l'apprentissage non supervisé consiste à ne fournir aucune information sur les classes ou les caractéristiques des données. L'algorithme va par lui-même classer les données dans des catégories qui peuvent s'avérer correspondre aux espèces respectives pour reprendre l'exemple des Iris (Yadav & Jadhav, 2019). Souvent, les ingénieurs ne savent pas nécessairement quels sont les attributs les plus appropriés pour construire leur modèle avant d'avoir fait au préalable des tests. C'est pourquoi les développeurs peuvent rentrer dans de longs cycles de développement. En effet, ils doivent identifier et développer de nouveaux attributs, reconstruire des modèles, mesurer les résultats, et répéter le processus jusqu'à ce que les résultats deviennent satisfaisants. La troisième catégorie s'agit de l'apprentissage par renforcement où le système acquiert un certain comportement dans un environnement en prenant des actions et en y quantifiant les résultats (Sharma, 2017). Si l'agent obtient un résultat correct, il obtient un point de récompense qui va amplifier l'agent à agir avec ce même type d'action. Il s'agit-là d'une boucle de validation rétrospective consistant à réajuster les coefficients de décision dans le réseau.

Dans le domaine de la médecine, les systèmes semi-supervisés sont les plus adaptés en raison du coût élevé pour étiqueter les données. Seulement une partie des données sont étiquetées, alors qu'une autre partie, bien plus abondante, demeure non étiquetée. Par exemple on va spécifier uniquement le nom d'espèce de la plante ou bien distinguer un malade d'un non malade. Il existe différents algorithmes se basant sur le ML traditionnel qui sont les arbres décisionnels (Pal et al., 2003), les machines à vecteurs de support (SVM) (Vishwanathan et al., 2002), les k plus proches voisins (Keller et al., 1985), les modèles Gaussien mixtes (Wolfinger et al., 1993) et la classification naïve bayésienne (Zhang et al., 2009).

1.3 Apprentissage profond (Deep Learning)

1.3.1 Principe et fonctionnement

Selon Zhang et al. (2017), grâce à l'augmentation de la puissance de calcul des ordinateurs et l'avènement du phénomène « Big Data » nous avons pu voir émerger durant ces vingt dernières années l'apparition de réseaux neuronaux. Le Big Data désigne une quantité de données volumineuses dont l'usage d'outils classiques de gestion de base de données ne peuvent vraiment traiter. Dérivée du ML, cette approche s'inspire du fonctionnement du cortex cérébral visuel et de son réseau neuronal. En effet, il s'agit donc de réseaux structurés en couches de neurones artificiels (Artificial Neural Network ou ANN) qui peuvent avoir un

nombre de couches profondes (Deep Neural Network ou DNN). Une couche est un bloc de neurones ou de nœuds opérant ensemble à une certaine profondeur dans un réseau neuronal. Pouvant atteindre une profondeur de plus de 100 couches, on appelle cette nouvelle approche d'apprentissage le « Deep Learning » et appartient au paradigme du ML (Figure 2).

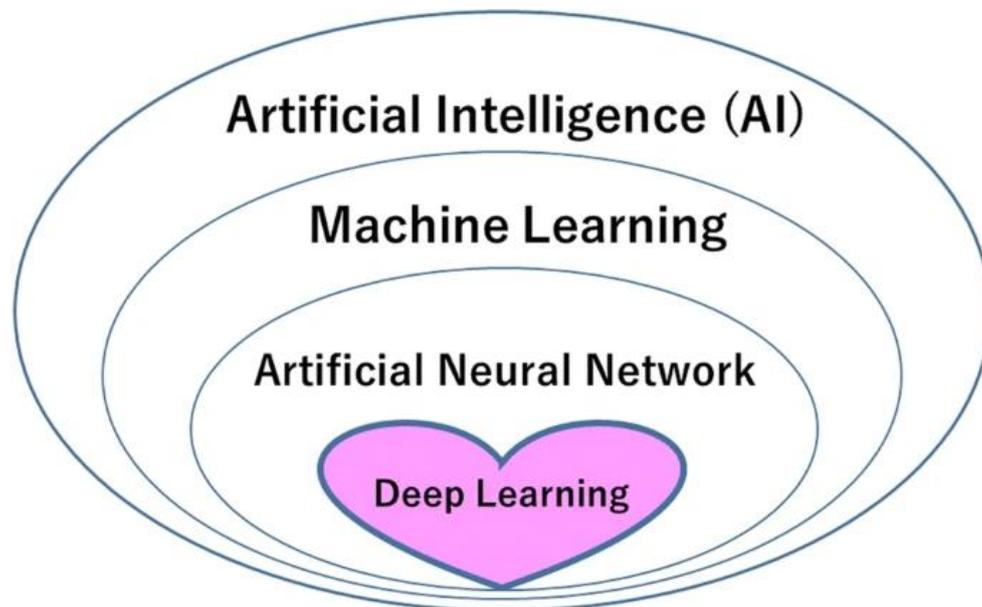


Figure 2 De l'IA au Deep Learning (Fujita, 2020)

Il se distingue du ML traditionnel par sa capacité à traiter des bases de données plus grandes, plus complexes est aussi moins structurées. L'apprentissage profond permet de s'affranchir du risque de mal choisir des caractéristiques en laissant le système les trouver lui-même. Cette nouvelle méthode se base sur l'apprentissage semi-supervisé (Tapia, 2017). Son intérêt est le suivant : durant l'entraînement, la seule information qui est fournie est l'étiquette (dans notre exemple d'Iris, l'espèce). L'algorithme va extraire par lui-même quelles sont les caractéristiques les plus prédictives lui permettant de classer les Iris en différentes espèces. Le choix final des caractéristiques résultant du DL est parfois surprenant ou difficilement compréhensible du fait que leur nombre est souvent très élevé, d'où l'effet « boîte noire ». Afin de mieux appréhender en quoi consiste un réseau de neurone, prenons l'exemple d'un perceptron qui est un réseau de neurones à une seule couche, illustré dans la Figure 3.

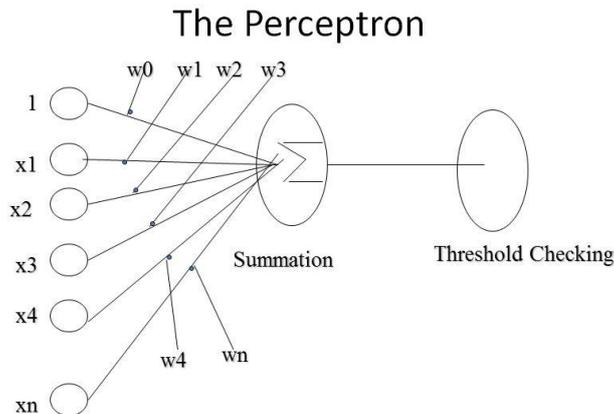


Figure 3 : Perceptron (Sharma, 2021)

Le neurone est l'unité fondamentale d'un réseau et selon Hansen (2019), chaque neurone correspond à une variable d'entrée et son niveau d'activation est défini par un coefficient d'activation allant de 0 à 1. Connectés entre eux, les neurones ont leur propre poids de connexion. A cela s'ajoute le biais qui permet de définir à partir de quel niveau le neurone a une valeur significative. Le biais peut s'ajouter ou se soustraire au niveau de connexion et d'activation des neurones connectés. Toutes ces unités sont reliées à un neurone de sortie qui reçoit la somme des niveaux d'activation pondérés par leurs poids de connexion respectifs. En bout de chaîne se trouve alors une fonction d'activation, qui peut être exprimée de plusieurs façons (ReLU, Sigmoid, Softmax, etc.) (Yamashita et al., 2018). L'acheminement de l'information se fait dans une seule direction, des variables d'entrées aux variables de sorties, d'où le terme « feedforward neural networks ». Si le but de ce réseau est d'effectuer une classification binaire, alors on va attribuer à cette fonction un seuil qui va lui permettre de classer les données selon leur niveau d'activation respectif. Tous ces paramètres de configuration sont internes au modèle. Afin d'entraîner un perceptron, nous allons chercher à minimiser l'erreur de prédiction à l'aide d'un jeu de données d'entraînement. La fonction de perte, également référée comme la fonction de « coût », mesure la compatibilité entre les prédictions de classification faites par le réseau et la classification réelles des données (Yamashita et al., 2018). Plus le nombre d'erreurs est petit, plus la fonction de perte sera faible. Pour se faire, nous allons utiliser un algorithme de descente de gradient ayant pour rôle de minimiser la perte en faisant varier les poids des neurones et les biais. C'est pourquoi les pondérations de connexion et les biais ont des valeurs initiales choisies aléatoirement qui vont être remises à jour entre chaque observation (Svozil et al., 1997). Il s'agit donc d'un processus d'optimisation itératif ayant recours à un algorithme d'entraînement par rétropropagation.

Il existe aussi des paramètres externes au modèle, indépendant des données et réglables qui influent sur les paramètres de l'algorithme d'optimisation (Romero et al., 2015). Bien réglés,

ils peuvent réduire le temps d'entraînement du modèle et améliorer ces performances (Smith et al., 2018). On retrouve par exemple la taille du lot (batch size), le nombre d'époch, la fonction de perte à optimiser ainsi que la méthode d'optimisation associée (ex. descente du gradient) avec ces propres paramètres comme le taux d'apprentissage ou le coefficient d'inertie (momentum).

Afin d'avoir un exemple plus concret, nous pouvons observer un perceptron multicouche (Multilayer perceptron) entraîné dans la classification d'Iris dans la Figure 4.

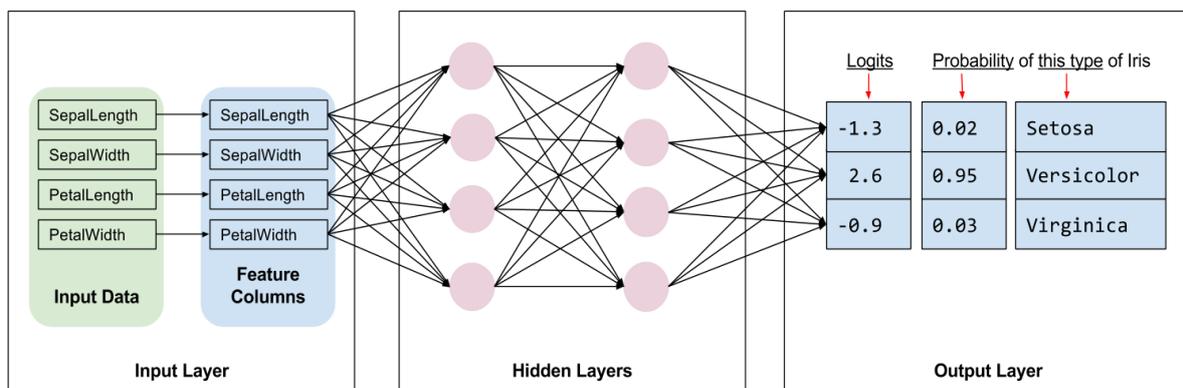


Figure 4: Les couches d'un Réseau Neuronal (Formation personnalisée: procédure pas à pas | TensorFlow Core, 2020)

On peut voir la couche d'entrée (input) correspondant aux variables d'entrées des caractéristiques de la fleur d'iris telles que la longueur, la largeur et la couleur d'une pétale. Les couches suivantes (couches cachées) correspondent à un niveau différent de représentation d'une même caractéristique qui va se convertir en une représentation de plus en plus complexe et abstraite. Chaque neurone d'une couche particulière est connecté à tous les neurones de la couche suivante (Svozil et al., 1997). Enfin, il y a la couche de sortie (output) aboutissant à une fonction d'activation classant les espèces prédites avec une certaine probabilités de se tromper. Il existe différents types de couches qui sont les couches de détection, de convolution, d'agrégation, de connexion complètes (Fully-Connected) et de terminaison correspondant à la couche de classification (Baker et al., 2017). S'agissant d'une classification multi-classe dans le cas des Iris, on a généralement recours à une fonction d'activation de type Softmax (Ren et al., 2017).

Il existe différentes architectures et approches de Deep Learning qui dépendent du type de données traité, du type d'apprentissage utilisé ainsi que de la nature de la tâche à résoudre. On retrouve son utilisation dans de nombreux domaines tels que la reconnaissance d'images, la traduction automatique, la prédiction financière (Fischer et al., 2018), la pharmaceutique (Ekins, 2016), la robotique (Levine et al., 2018), les systèmes de modération de réseaux

sociaux (Badjatiya et al., 2017) et aussi dans l'imagerie médicale (Shen et al., 2017). En effet, on voit l'apparition de logiciels d'assistance au diagnostic ou à la détection de maladies (Chartrand et al., 2017). La suite de ce travail va se concentrer sur les systèmes d'assistance au diagnostic dans le secteur de la radiologie.

1.3.2 Application du DL en radiologie

Selon la Société nord-américaine de Radiologie (RSNA), une crise est en train d'apparaître en Radiologie (van Ginneken et al., 2011). Avec la progression technologique des systèmes d'acquisition d'images, le nombre et la complexité des images produites a rapidement augmenté. La cadence de travail chez les radiologues a considérablement augmenté et leur rôle est devenu plus difficile selon Taylor et al. (2016). De plus, les coûts de la santé en lien avec l'imagerie médicale s'accroissent rapidement. En raison d'une augmentation du volume d'informations lors de scans individuels, il devient de plus en plus nécessaire de standardiser l'interprétation des images par l'humain et de trouver de nouvelles approches dans la manière de travailler.

Nous voyons aujourd'hui l'émergence de systèmes basés sur le Deep Learning dans le but d'assister le médecin lors de l'interprétation d'images radiologiques, plus connus sous le nom de Computer Assisted Diagnosis (CAD). D'après la RSNA, ces systèmes pourraient en partie répondre à cette problématique. La RSNA définit les systèmes CAD comme des algorithmes de machine dans l'aide au processus d'interprétation d'images. En cas de succès, cela peut accélérer le processus de diagnostic, réduire les erreurs de diagnostic et améliorer l'évaluation quantitative. Malgré le besoin grandissant et évident de logiciels d'assistance de type CAD, ce terrain de recherche existe depuis plusieurs décennies. Nous pouvons voir dans la Figure 5 : Evolution du CAD que le développement du CAD a connu plusieurs phases alternées d'intérêts et de désintérêts dans la recherche. La progression de la recherche dans ce domaine est fortement liée à la progression d'IA (Fujita, 2020).

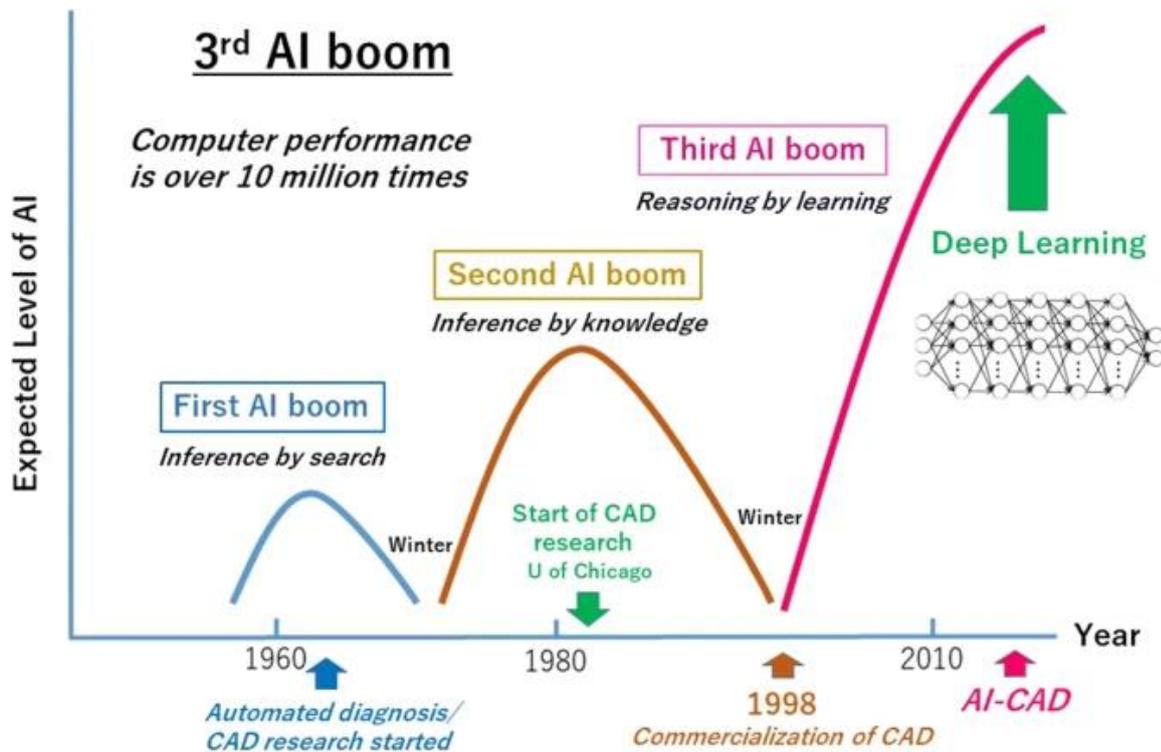


Figure 5 : Evolution du CAD (Fujita, 2020)

Toutefois, son utilisation en milieu clinique reste faible à l'exception de la mammographie aux USA (Taylor et al., 2016). En effet, tout le processus allant du développement de l'algorithme à son implémentation en clinique fait face à de nombreux défis. Selon la RSNA, pour y répondre, sa mise en place exige la maîtrise de divers champs d'expertise en traitement d'images, en ML, en physique médicale et en radiologie. C'est pourquoi le CAD est un terrain de recherche interdisciplinaire.

1.3.3 Type de réseaux neuronaux

1.3.3.1.1 CNN

Selon Krizhevsky et al. (2012), les réseaux neuronaux convolutionnels sont spécialisés dans la reconnaissance d'objets. Inspiré par la partie visuelle du cortex, c'est un modèle ayant une grande capacité d'apprentissage. Ce sont des réseaux s'appuyant sur le Deep Learning et utilisant la propagation de l'information vers l'avant. Sharma (2016) présente comment lorsqu'on donne un lot de données images étiquetées au CNN, une région locale ou un champ réceptif associé à un groupe de neurones est détecté et va être traité comme illustré dans la Figure 6. A partir d'une matrice ayant un certain fenêtrage, une valeur sera calculée à chaque position de cette dernière.

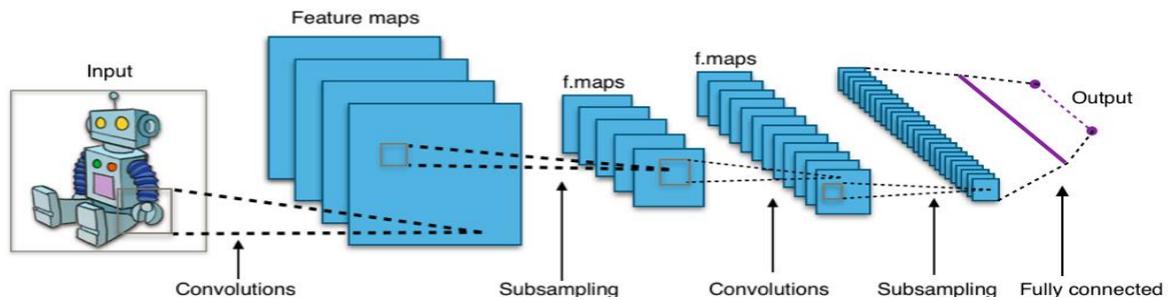


Figure 6 : Exemple de structure d'un CNN (Sharma, 2017)

Ayant une architecture hiérarchique, le champ réceptif illustrant par exemple une main va suivre une série de transformations en passant par des couches de convolution, des couches de Pooling et une couche Fully-Connected. Ceux sont les mêmes types de couches que l'on peut retrouver dans des réseaux de neurones standards. Contrairement aux réseaux standards, les neurones dans une couche sont seulement connectés à une petite région de la couche précédente. Ayant également moins de paramètres, cela les rend plus faciles et plus rapides à entraîner.

Bien qu'on retrouve aussi des hyperparamètres, comme le taux d'apprentissage, il existe des hyperparamètres supplémentaires telles que le remplissage, le pas ainsi que le nombre et la taille des kernels. Les Kernels sont un ensemble de paramètres apprenables qui sont appliqués lors de l'opération de Convolution.

Nous allons décrire ces différents types de couches plus en détail. Les couches les plus souvent utilisées dans l'extraction des caractéristiques sont les couches de convolution, de pooling et de ReLU (*Introduction au Deep Learning avec MATLAB*, s. d.)

- Couches de convolution : Ce type de couche est la composante clé d'un CNN. Elle permet de faire émerger les caractéristiques d'une image à travers un jeu de filtres convolutionnels, dont chacun va activer certaines caractéristiques des images. Selon Sharma (2017), chaque neurone est connecté à une région locale de l'image. La taille de cette région correspond à la taille du filtre. Chaque neurone va produire un point relié entre le poids et une caractéristique d'entrée se situant dans le champ réceptif. Chaque calcul conduit à l'extraction d'une carte de caractéristiques.
- La couche de « pooling » : Souvent placée entre deux couches de convolutions, elle réduit la dimension des données de sorties en performant des sous-échantillonnages non-linéaires, chose qui permet la réduction du nombre de paramètres que le réseau doit apprendre. Une des techniques est le « max pooling » qui est processus de discrétisation. Cela consiste à choisir la valeur de pixel la plus haute d'une région selon sa taille. En d'autres mots, le max pooling sélectionne la plus haute valeur incluse dans la fenêtre de

l'image couverte par une matrice plus petite. On peut voir un exemple avec une matrice de pooling lors du sous-échantillonnage d'une image dans la Figure 7.

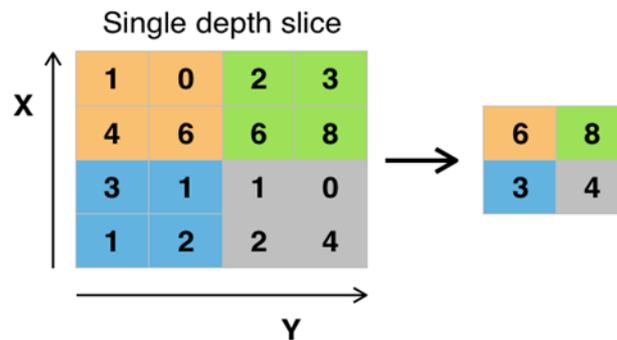


Figure 7 : Exemple de l'application du Max Pooling (Sharma, 2017)

- La couche ReLU : La couche de correction ReLU est une fonction d'activation qui permet un entraînement plus rapide et efficace en transformant les valeurs en entrée négatives à 0 tout en conservant les valeurs positives.

Ces trois opérations sont répétées de nombreuses fois comme illustré dans la Figure 6.

Après l'extraction des caractéristiques, l'architecture du CNN procède à la classification.

- Couche Fully-connected (FC) : L'avant dernière couche utilisée est la couche FC qui va produire un vecteur avec un nombre de dimensions égales au nombre de classes que le réseau doit prédire. Ce vecteur contient la probabilité d'appartenance pour chaque classe selon l'image traitée. L'objectif de cette couche est d'aplatir les caractéristiques apprises par les couches convolutionnelles et de toutes les combiner.
- Couche de fonction d'activation : La toute dernière couche du classificateur va utiliser une fonction Softmax ou Sigmoid pour calculer une probabilité de prédiction pour chaque classe correspondante à l'ensemble des caractéristiques mesurées sur l'objet en question.

On retrouve l'utilisation des réseaux CNN dans différentes applications telles que la reconnaissance de visage (Lawrence et al., 1997), la modélisation de phrases (Kalchbrenner et al., 2014), la segmentation anatomique (Wachinger et al., 2018), le pilotage automatique de voiture (Bojarski et al., 2016), et aussi en clinique dans le cadre de la classification de nodules pulmonaires en radiologie (Yamashita et al., 2018).

1.4 Données

Comme le dit Currie et al. (2019), si l'intelligence artificielle est l'engin qui conduira vers l'avant le futur de la technologie, les données sont l'essence sur lequel il se repose. En effet, les algorithmes utilisés requièrent un vaste nombre de données pour fonctionner correctement.

Ces données doivent être accessibles, utilisables et validées. Selon l'Association Canadienne des Radiologues (ACR) (Tang et al., 2018), l'absence de données représentatives constitue un des obstacles majeurs dans le développement d'applications IA le domaine de la radiologie. Il s'agit de systèmes de diagnostic assistés par ordinateur (CAD) basés sur du Machine Learning. Il existe différentes bases de données remplissant des objectifs distincts. On retrouve :

- Les données d'entraînement et de validation
- Les données de test

1.4.1 Données d'entraînement et de validation

Selon Petrick et al. (2013), la base de données utilisée pour le design d'un système CAD s'appelle la base de données d'entraînement. Selon l'ACR (Tang et al., 2018), sa disponibilité est une des trois composantes clés les plus importantes dans le développement d'un CAD. En effet, un jeu de données d'entraînement adéquatement sélectionné et suffisamment large permet au développeur de choisir et affiner les caractéristiques correspondant aux propriétés inhérentes à la maladie au sein de la population cible (Petrick et al., 2013). Elle doit contenir des images radiologiques de patients avec et sans présence de la maladie cible puisque le CAD doit être capable de distinguer ses caractéristiques pathologiques et non-pathologiques. Ce jeu de données devrait inclure la gamme d'anormalités possibles qui sont vues dans la pratique et qui correspondent à la pathologie cible. Toutefois, l'utilisation exclusive d'images n'est pas suffisante. Afin de bénéficier de l'apport du Big Data dans le domaine de la radiologie, la mise en corrélation d'informations cliniques pertinente est un point critique (Currie et al., 2019a). C'est pourquoi, la collection d'images combinées avec d'autres types de données telles que des données de diagnostic et des informations patient-spécifiques font partie intégrante d'une base de données d'entraînement (Petrick et al., 2013). Il peut s'agir d'informations telles que des biomarqueurs à partir de tests sanguins, mais aussi l'âge, la démographie, des antécédents familiaux, des maladies concomitantes et des antécédents d'expositions environnementaux. Additionnellement, il peut être nécessaire d'avoir des données spécifiques à la maladie cible telles que son type, sa taille, ou son stade pour définir des sous-groupes de personnes dans l'échantillon de données.

L'ACR estime qu'une mauvaise sélection de base de données de référence peut fortement impacter les performances du CAD (Tang et al., 2018). C'est pourquoi il devrait y avoir une description claire des caractéristiques telles que les paramètres d'acquisition et de reconstruction des images.

Bien que les algorithmes de Machine Learning fonctionnent correctement avec une base de données faible, c'est moins le cas pour du Deep Learning. En effet, le DL est « DATA

HUNGRY » (Mikolajczyk, 2018). Cela signifie que plus il y aura de données, plus l'algorithme DL utilisé par le CAD sera performant. Les algorithmes traditionnels vont plus rapidement atteindre un plateau de performance malgré l'augmentation du nombre de données selon la Figure 8.

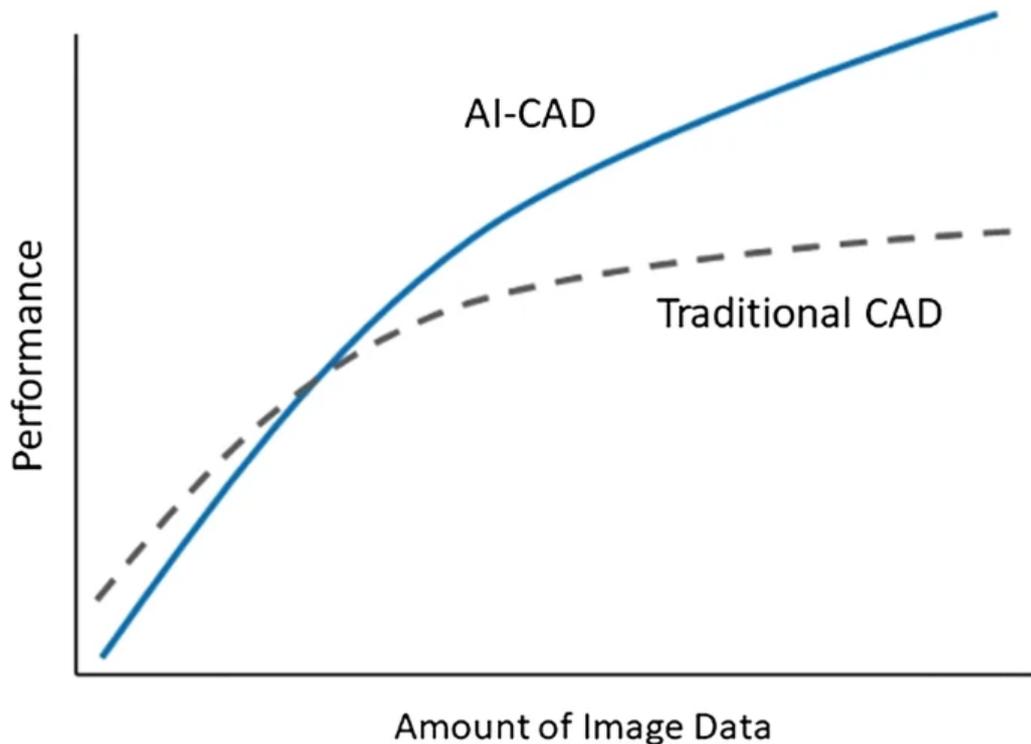


Figure 8 : Performance du CAD et nombre de données (Fujita, 2020)

Il existe une autre base de données également utilisée dans le développement et l'évaluation intermédiaire du CAD qui est la base de données de validation. Les résultats obtenus sont utilisés comme un guide d'optimisation du processus d'apprentissage de l'algorithme. Indépendant de la base de données d'entraînement, ce jeu de données permet de spécifier des stratégies d'entraînement en ajustant les réglages des méta paramètres (Brownlee, 2017). On retrouve par exemple le nombre d'époch, la taille de lot (« batch ») ou le nombre d'itérations (Mikko, 2019).

1.4.2 Données de test

Selon (Petrick et al., 2013), les données utilisées pour l'évaluation des performances du système CAD s'appellent les données de test. En principe elles devraient correspondre à la population cible sur laquelle l'IA va être appliquée. De plus, les paramètres d'acquisition d'image et la préparation du patient devraient y être aussi représentatifs. Un jeu de données de test adéquatement sélectionné permet d'estimer les performances du système CAD et facilite les comparaisons avec d'autres systèmes CAD avec la même intention d'utilisation.

Afin d'avoir une interprétation adéquate des résultats, des critères d'inclusion et d'exclusion des données devraient être clairement spécifiés et justifiés.

Lors de l'évaluation des performances, concernant la prévalence de la maladie dans les bases de données du CAD, il est commun qu'elle soit supérieure à la prévalence de la population cible, particulièrement lorsque la prévalence de la maladie est faible. Lorsque les performances d'évaluation sont comparées entre deux systèmes CAD en compétition, l'augmentation de la prévalence des cas dans la base de données de test peut accentuer la différence de performance entre ces deux derniers.

La présence de biais peut influencer les performances du CAD telles que les biais de vérification lorsque le diagnostic de la maladie n'est pas effectué systématiquement avec les mêmes standards. C'est pourquoi la détermination du statut du patient à propos de la maladie cible devrait être effectué avec un gold standard. Il peut aussi y avoir des biais de sélection lorsque les cas inclus dans la base de données de correspondent pas à la population cible. Il est dans tous les cas important de rendre connaissance et discuter de tout biais potentiels susceptibles d'influencer les résultats du CAD.

Lorsqu'un réseau est entraîné, il peut parfois y avoir un risque de sur-apprentissage. C'est un problème courant en ML. Cela se produit lorsque les performances du CAD sont excellentes avec des données sur lequel il s'est entraîné et médiocres lors de l'évaluation de nouvelles données (*Overfitting in Machine Learning*, 2017). En d'autres termes, le modèle ne parvient pas à se généraliser correctement. Normalement, les éléments pertinents à retenir par le réseau sont les caractéristiques intrinsèques d'un objet d'étude et sont considérées comme du signal qui est bénéfique. A l'inverse, il peut aussi retenir des informations non pertinentes considérées comme du bruit. Un algorithme performant devrait pouvoir mémoriser uniquement le signal et le distinguer du bruit. Si le modèle utilisé est trop flexible ou complexe, il risque de mémoriser le bruit au détriment du signal. Pour le prévenir, il faut par exemple augmenter le nombre de données d'entraînements, réduire le nombre d'itérations et supprimer les caractéristiques non pertinentes.

1.5 Préparation des données

1.5.1 Pre-processing

Une fois que la base de données est constituée, les images radiologiques incluses sont souvent traitées dans une phase de pré-traitement (Pre-processing). C'est une technique d'exploration de données qui transforme les images brutes dans un format compréhensible par le système CAD. Selon la Société nord-américaine de Radiologie (Ginneken et al., 2011), elle est effectuée par de nombreuses études scientifiques traitant du CAD. Cela permet de comparer des images provenant de patients différents (Towey et al., 2011). Mais encore, cela réduit la

variabilité entre différentes marques d'appareils et différents protocoles utilisés. Par exemple, lors de la quantification d'un emphysème détecté au scanner avec des doses, des épaisseurs de coupe et des algorithmes de reconstruction différents ont un fort impact sur le score d'emphysème de patients provenant d'un centre à un autre ce qui peut biaiser les performances du CAD. En effet, les ordinateurs sont considérés comme des « blind number cruncher » et peuvent facilement être trompés par des différences facilement ignorables par l'humain (van Ginneken et al., 2011). C'est pourquoi les paramètres de scan (ex. pitch, mAs et kV) et les algorithmes de construction devraient, dans des conditions idéales, être constants, chose qui est impossible en pratique. Afin de limiter cette variabilité, une normalisation doit être appliquée aux images.

Il est important de relever que le recours au prétraitement dépend du contexte d'utilisation et qu'il n'existe pas à ce jour de gold standard. C'est pourquoi il y aura toujours une variante d'une étude à une autre. Selon Ortiz et al. (2016), la phase de pré-traitement est composée de deux phases : la normalisation spatiale et la normalisation d'intensité.

1.5.1.1 *Normalisation spatiale*

Tout d'abord, il faut commencer par une normalisation spatiale pour réduire les différences de positionnement (Taylor et al., 2016; Towey et al., 2011). Par exemple, cela permet de s'assurer que les pixels (ou voxel pour les images 3D) d'images de chaque sujet soit à la même position anatomique dans une région comme le cerveau (Illán et al., 2012; Martinez-Murcia et al., 2013). Il peut être nécessaire d'appliquer une transformation linéaire de type rigide telles que des translations et des rotations. Il est aussi possible d'effectuer des déformations d'images grâce à une transformation élastique dans le but d'harmoniser les variabilités des morphologies. En effet, une étude s'intéressant au CAD dans l'imagerie cardiaque au scanner a utilisé une normalisation spatiale pour réduire la variabilité inter-sujet (Wong et al., 2016). En effet, le cœur n'est pas situé à la même position dans le thorax, et la taille et la forme sont également variables. Ils ont notamment utilisé une normalisation du rayon des ventricules gauches (VG) afin qu'ils aient tous la même taille selon la Figure 9.

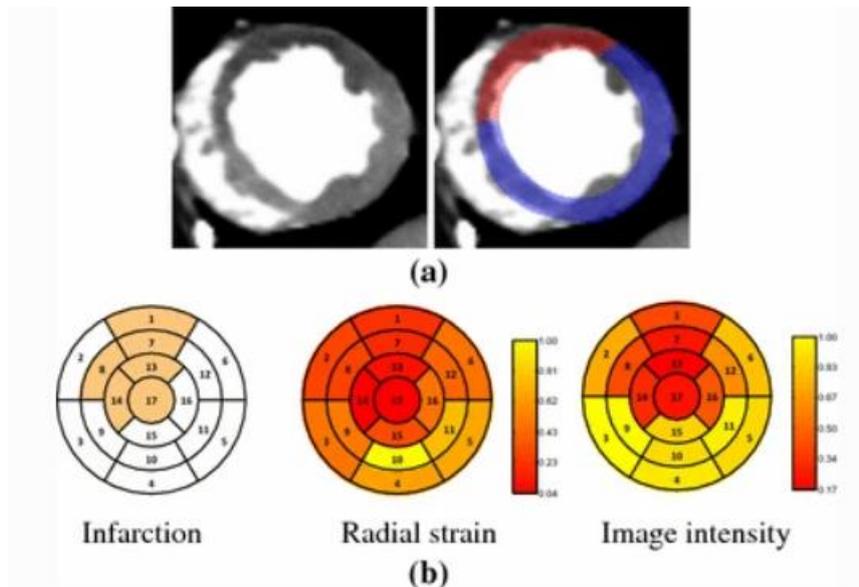


Figure 9: Normalisation du ventricule gauche sur l'axe court (Wong et al., 2016)

Dans la Figure 9, La partie en rouge correspond à une région du VG identifiée comme infarctée par un expert. Nous pouvons y voir la représentation des régions du VG en sous parties. Plus une région est rouge, plus elle est infarctée. Une normalisation radiale est effectuée entre l'image de gauche et l'image du centre. Nous pouvons voir que le niveau d'atteinte des régions se trouvent légèrement modifiées après la normalisation radiale.

1.5.1.2 Normalisation d'intensité

Selon Brahim et al., (2019), la normalisation d'intensité est requise lorsqu'on souhaite réduire la variabilité inter-sujet. Elle devient appropriée dans toute analyse d'image avec plus d'un sujet pour assurer la comparabilité au sein et au travers des images des différents sujets. C'est tout particulièrement la variabilité des intensités de pixels que l'on souhaite réduire.

Selon (Zhao et al., 2018), la visualisation des échelles d'intensité peut se faire à l'aide d'un histogramme comme illustré dans la Figure 10. Nous pouvons voir deux graphiques représentant les courbes de distribution d'intensité des valeurs de pixels d'images cérébrales selon leur niveau d'intensité (Zhao et al., 2018). Cela provient de 30 images cérébrales acquises par IRM dans le contexte d'une mise en place d'un système DL pour la segmentation de tumeurs cérébrales. Sur l'image de gauche nous pouvons voir une distribution d'échelles d'intensité très variable d'une image à une autre avant la normalisation. Sur l'image de droite nous pouvons voir une distribution d'échelles d'intensité qui sont bien plus homogènes après la normalisation.

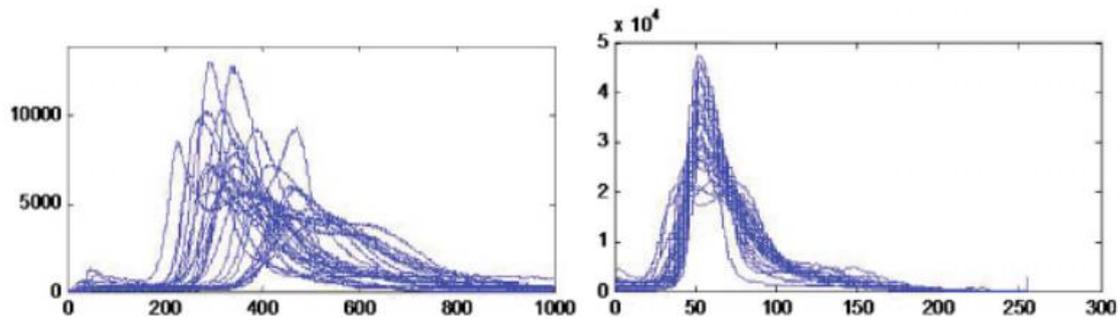


Figure 10: Distribution de courbes d'intensité (Zhao et al., 2018)

1.5.2 Augmentation

Selon Fawzi et al. (2016), un problème courant de la classification est celui du manque de données disponibles pour entraîner des classificateurs précis et robustes. Pour réduire la rareté des données, une approche est basée sur l'augmentation artificielle de ces dernières. Cela consiste en la transformation de données disponibles en de nouvelles données tout en préservant les mêmes étiquetages. En effet, des transformations suffisamment faibles conservent les propriétés des images radiologiques. Cette pratique est courante lorsqu'on souhaite entraîner de larges réseaux d'apprentissages profonds. On retrouve :

- L'augmentation géométrique
- L'augmentation d'intensité

Lorsqu'on évalue la performance d'un modèle algorithmique, on souhaite déterminer s'il se généralise bien. Il doit être capable de faire de bonnes prédictions sur des données qui n'ont jamais été traitées en reconnaissant leurs caractéristiques importantes et en ignorant les détails insignifiants. L'intérêt est de réduire le risque de surapprentissage (« Data face », 2020). Il existe d'autres techniques d'augmentations ayant recours à de la saturation, l'ajout de bruit, du floutage ou bien des changements de couleurs etc. La Figure 11 illustre une série d'augmentations géométriques, d'intensités ou autre à partir d'une image originale.

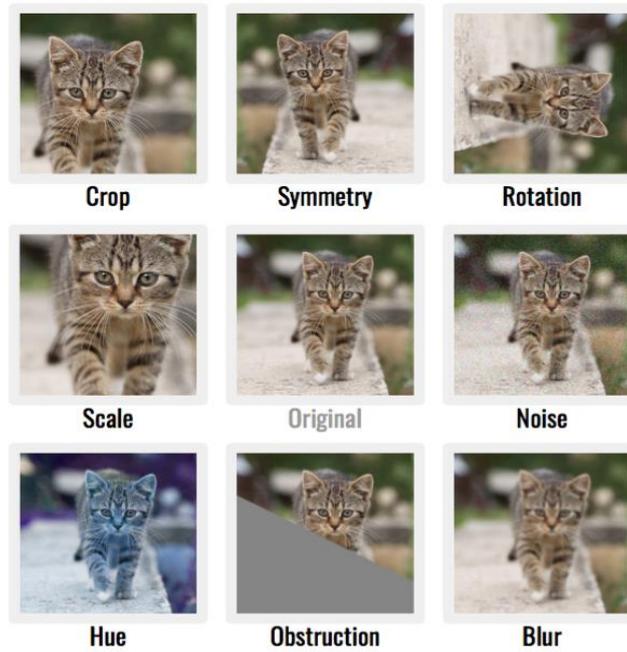


Figure 11 : Chat augmenté (Despois, 2018)

1.5.2.1 Augmentation Géométrique

L'augmentation géométrique des images facilite l'apprentissage de caractéristiques sans être impacté par une différence de rotation ou d'échelle entre elles (Yadav et al., 2019). On peut par exemple trouver des transformations de type géométriques telles que des rotations, des déformations, des recadrages, des étirements, des inversions ou des translations.

1.5.2.2 Augmentation d'Intensité

Il existe d'autres techniques d'augmentation telles que l'augmentation du contraste, de la brillance ou de la balance en blancs, appliquées par ex. durant l'analyse d'images de lésions cutanées (Galdran et al., 2017). Ces techniques entraînent l'algorithme de DL à une variabilité étendue lors de l'apprentissage de caractéristiques tout en devenant moins sensibles aux facteurs de variation. Nous pouvons voir un exemple d'augmentation de photos de lésions de la peau avec une balance de blanc dans la Figure 12.

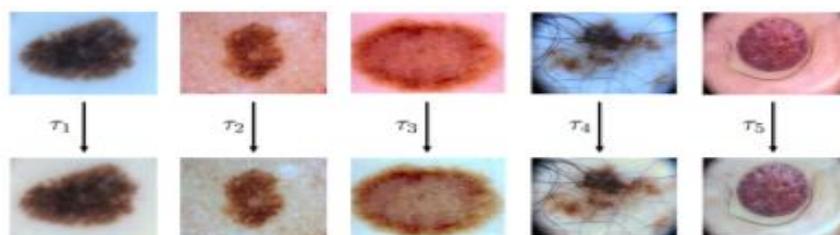


Figure 12 : Balance de blanc sur des images (Galdran et al., 2017)

1.5.3 Balancing

Des déséquilibres de données ont lieu lorsqu'il y'a une différence de représentation de classes (ex. entre les sujets pathologiques et les sujets les sujets non-pathologiques). En effet, lorsqu'il y a un déséquilibre dans les classes, le classificateur aura tendance à plus favoriser une classe au détriment d'une autre. Par conséquent, on risque de constater une détérioration des performances du classificateur par rapport aux classes les moins représentée (Batista et al., 2004).

Il existe différentes méthodes telles que le sur-échantillonnage randomisé des données sous-représentées pour contourner ce problème qui est une technique relativement simple. Cette technique de duplication est toutefois connue pour créer du sur-apprentissage puisque les données ajoutées ne sont que des pures copies. Inversement, on peut aussi faire du sous-échantillonnage randomisé en réduisant le nombre d'image de la classe la plus représentée.

Il existe également des méthodes d'échantillonnages plus complexes et moins sensibles au sur-apprentissage telles que :

- Les paires de Tomek : Cette technique consiste à créer des paires de classes en combinant une image pathologique et une image non-pathologique ayant la distance la plus proche autour du seuil de décision. On appelle ces paires liées « Tomek ». Etant une technique de sous-échantillonnage, l'image de la classe majoritaire associé à une image de la classe minoritaire va être supprimée (Batista et al., 2004).
- L'approche « SMOTE » : Cette technique consiste à créer des nouvelles images appartenant à la classe minoritaire en interpolant des images ayant des caractéristiques proches (Jiang et al., 2016).
- La règle du plus proche voisin : Cette technique consiste à éliminer les images de la classe majoritaire ayant une distance de la limite de décision (patho / pas patho) qui soit éloignée du seuil de décision. En effet, ces cas qui sont très saillants sont moins pertinents pour l'apprentissage. (Kriminger et al., 2012)

Ces techniques peuvent même être combinées entre elles. Selon Batista et al. (2004), qui ont testé ces méthodes, la combinaison de la technique SMOTE et de Tomek serait la plus recommandée lorsqu'il y a peu de cas positifs. Lorsqu'il y a beaucoup de cas positifs, la méthode simple du suréchantillonnage randomisé serait la plus appropriée.

1.6 Evaluation du réseau

On peut retrouver différents types d'évaluations. On peut procéder à une évaluation quantitative basée sur des mesures ainsi qu'à une évaluation visuelle basée sur la représentation et l'utilisation des caractéristiques par le système.

1.6.1 Mesure quantitative

Selon Petrick et al. (2013), l'évaluation des performances d'un CAD dépend de la nature de l'application clinique du logiciel. Son évaluation est importante car elle permet :

- L'estimation de la performance de l'algorithme
- L'établissement de son efficacité
- La facilitation des comparaisons entre différents systèmes avec applications similaires
- L'optimisation de l'algorithme durant le développement CAD

Le comité de détection assistée par ordinateur en imagerie diagnostique (CADSC) propose différentes mesures d'évaluation des performances du système CAD de diagnostic qui sont (Huo et al., 2013) :

- La sensibilité : pourcentage correct de cas réellement positifs (vrais positifs)
- La spécificité : pourcentage correct de cas réellement négatifs (vrais négatifs)
- L'aire sous la courbe ROC : sera expliqué à la page suivante.

Ces trois mesures de test diagnostic sont indépendantes de la prévalence de la maladie d'un jeu de données. Afin de rendre plus intuitif la notion de spécificité, on préfère parler de fraction de faux positifs qui est définie comme :

$$\text{Fraction de faux positifs} = 1 - \text{spécificité}$$

Une autre mesure couramment utilisée est la Courbe ROC (Receiver Operating Characteristics). Cette courbe trace le taux de vrais positifs en fonction du taux de faux positifs pour différents seuils de classification. De ce fait, cette mesure permet également la sélection d'un seuil adéquat selon la Figure 13.

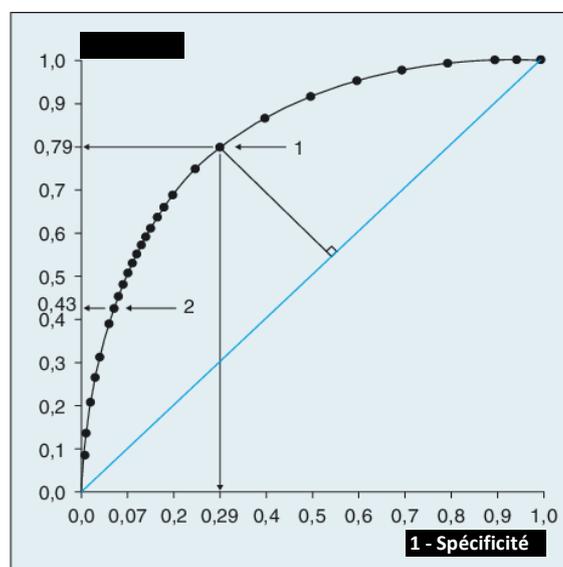
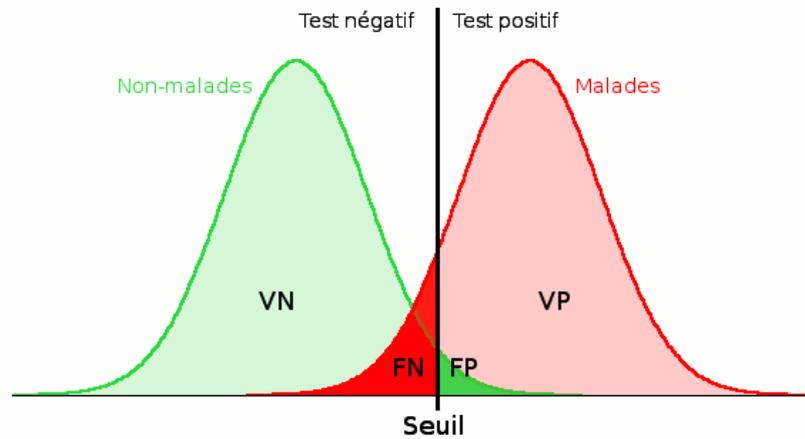


Figure 13 : La courbe ROC, (Delacour et al., 2020)

Par exemple, diminuer le seuil de classification augmente le nombre d'éléments classés comme positifs et donc le nombre de vrais positifs, ce qui par conséquent augmente le nombre de faux positifs. La sensibilité et la spécificité vont également varier (Figure 14). Le choix du seuil dépend du contexte d'utilisation clinique.



Néanmoins, la courbe ROC ne permet pas un aperçu global des performances d'un test diagnostique puisqu'elle donne une information pour un seuil choisi. Afin d'avoir une estimation incluant toutes les valeurs de seuils de classification, il est possible de mesurer par calcul de son intégrale qui est :

- *L'aire sous la courbe ROC (AUC) :*

Cette valeur est normalisée entre 0 et 1. Un modèle dont la totalité des réponses sont fausses aura un AUC de 0. Inversement, un modèle dont la totalité de réponses sont justes aura un AUC de 1. Cette mesure permet de comparer les performances globales de plusieurs test quantitatifs indépendamment de leur seuil de classification choisi. La Figure 15 représente un modèle ayant classé correctement la totalité des réponses et qui a donc un AUC de 1 (Narkhede, 2019). Les courbes correspondant aux distributions des vrais négatifs et aux vrais positifs ne se chevauchent pas.

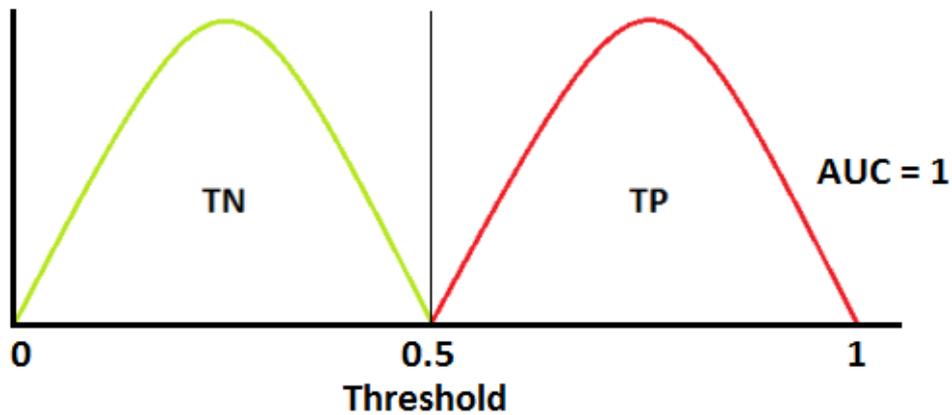


Figure 15 : AUC d'un modèle idéal (Narkhede, 2019)

Le deuxième exemple correspond cette fois-ci à un modèle ayant un AUC de 0.7 selon la Figure 16. Le modèle a donc 70% de chance de distinguer correctement un vrai positif d'un vrai négatif. Cela se traduit par un chevauchement des deux courbes, introduisant ainsi les erreurs de type 1 et 2.

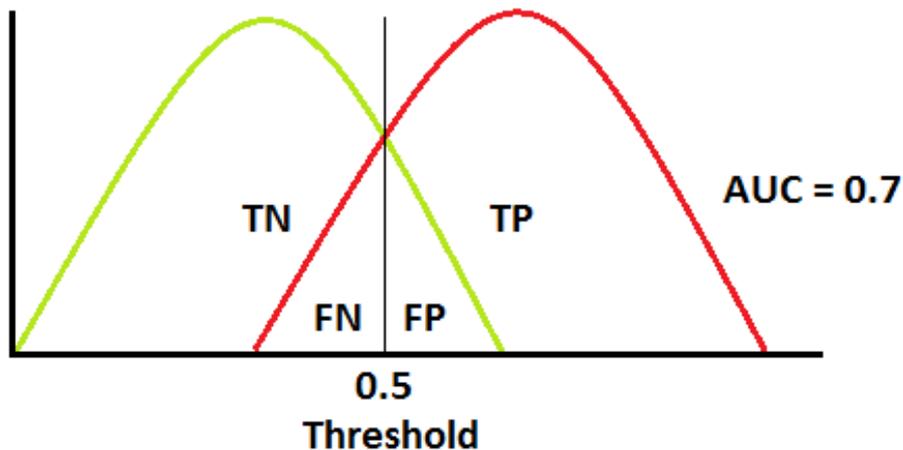


Figure 16 : AUC d'un modèle avec un AUC suboptimal (Narkhede, 2019)

Très souvent utilisé, on utilise aussi le taux d'exactitude qui permet d'évaluer un algorithme de ML (*Accuracy (error rate)*, 2019). Il permet de savoir à quelle fréquence l'algorithme classe une donnée correctement. Il s'agit du rapport entre le nombre de données correctement évaluées sur le nombre total de données évaluées. Dans le cadre d'un test diagnostique binaire, la formule utilisée est représentée selon la Figure 17. Les termes « TP », « TN », « FP », « FN », correspondent aux vrais positifs, aux vrais négatifs, aux faux positifs et enfin aux faux négatifs.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Figure 17 : Formule du taux d'exactitude (Accuracy (error rate), 2019)

Moins couramment utilisée, on peut mesurer la probabilité post-test qui est employée lorsque le test diagnostique a effectué sa prédiction aboutissant à deux mesures :

- En cas de test positif : La *valeur prédictive positive* qui est la probabilité d'être réellement malade
- En cas de test négatif : La *valeur prédictive négative* qui est la probabilité d'être réellement en bonne santé

Etant des mesures dépendantes de la prévalence de la maladie dans les données, leur interprétation implique une certaine vigilance lors de l'interprétation des résultats.

1.6.2 Evaluation visuelle

En apprentissage semi-supervisé, le CNN apprend lui-même les caractéristiques les plus prédictives lors de la classification. Cependant, le phénomène de boîte noire limite la capacité de l'Homme à comprendre quelles sont les informations qui sont retenues par l'algorithme. Il est également difficile de savoir comment elles sont conceptualisées. On ne sait pas non plus s'il utilise des informations pertinentes. Il existe différentes méthodes afin de rendre le comportement d'un réseau neuronal plus transparent lors de l'évaluation d'images à classer (Pratt et al., 2019). Lors de son développement, ces techniques permettent par exemple de mieux cibler les modifications à effectuer telles que :

- L'ajustement des hyperparamètres durant la phase d'optimisation
- Le pré-traitement des images

1.6.2.1 CAM

Selon (Zhou et al., 2016), les cartes de classes d'activation (Class Activation Mapping) indiquent les régions discriminatoires utilisées par le réseau neuronal pour effectuer des prédictions de classes. Cette approche a beaucoup contribué dans le domaine de la classification. Elle repose sur la représentation de cartes de chaleur mettant en avant les pixels d'une image conduisant le modèle à l'associer à une certaine classe selon ces caractéristiques. Cette technique requiert une modification de l'architecture du réseau et permet la visualisation des dernières étapes de la classification des images ou des dernières couches de convolution en le réentraînant.

Juste avant la couche finale d'activation (Softmax), on connecte les dernières couches de convolution d'un réseau CNN à une couche GAP (Global Average Pooling layer). Cette couche

clé engendre un regroupement global des cartes de convolution des caractéristiques en produisant en sortie une moyenne pour chaque carte de caractéristiques (Gagana, 2019). Pour rappel, une carte de caractéristiques est la donnée de sortie d'un filtre de convolution de la couche précédente. Selon la Figure 18, nous avons n cartes de caractéristiques différentes. La couche GAP crée n scores qui sont la moyenne de chaque carte de caractéristiques. Ces scores sont ensuite connectés à la prédiction par classe selon n poids (W_1, W_2, \dots, W_n) dépendant de chaque classe. La CAM pour une classe donnée est alors une somme des cartes d'activations pondérée par ces poids W_i .

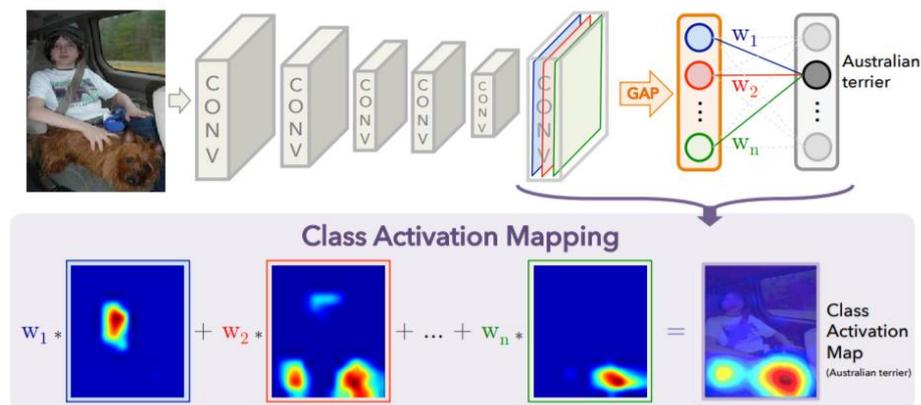


Figure 18 : Processus la carte d'activation de classe (Zhou et al., 2016)

En suréchantillonnant les cartes d'activation à la taille des images d'entrée et en les projetant sur elles, on peut identifier les régions d'activation les plus pertinentes selon la classe. Enfin, on peut identifier les régions d'activation importantes des images traitées en faisant une projection arrière des poids des connexions neuronales de la couche FC sur la carte de convolution des caractéristiques. Un exemple d'un CAM est donné dans la Figure 18.

1.6.2.2 Saliency map

Selon (Simonyan et al., 2014), les cartes de saillance (saliency maps) peuvent être utilisées pour les réseaux peu supervisés. Considérons une image donnée appartenant à une certaine classe définie par la fonction du score de classification. On souhaite mettre dans un certain ordre les pixels de l'image qui la composent selon leur niveau d'importance lors de la classification. C'est pourquoi on appelle aussi cette technique, la carte d'attribution des pixels (Smilkov et al., 2017). Comme nous l'avons vu, cette fonction de classification dépend des valeurs de poids et de biais du modèle. Plus la magnitude des éléments de poids est élevée, plus l'importance des pixels correspondant sera grande pour attribuer l'image à une certaine classe. Sachant que la fonction de classification utilise un modèle non-linéaire, on considère que la magnitude du poids est un vecteur dérivé de cette fonction qui est déterminée par propagation inverse. Ensuite, on met dans l'ordre les pixels du vecteur de poids pour obtenir

une carte de saillance. Les pixels blancs sont les pixels ayant le plus d'importance alors que les pixels noirs ont en le moins. La Figure 19 montre des exemples de cartes de saillance. Comme nous pouvons le voir, ces cartes sont assez bruitées.



Figure 19 : Exemple de carte de saillance (Simonyan et al., 2014)

1.6.2.3 Feature Visualisation

Selon (Olah et al., 2017), il est possible de visualiser à différents niveaux la représentation des caractéristiques d'un réseau. Communément appelé « la visualisation des caractéristiques », cette technique est considérée comme un outil très puissant et simple d'utilisation. Elle nous permet de voir les caractéristiques apprises par le réseau ainsi que les concepts qu'il en retire. Dans la Figure 20, il est possible de visualiser les caractéristiques à différentes étapes et à différentes échelles de la structure du réseau.

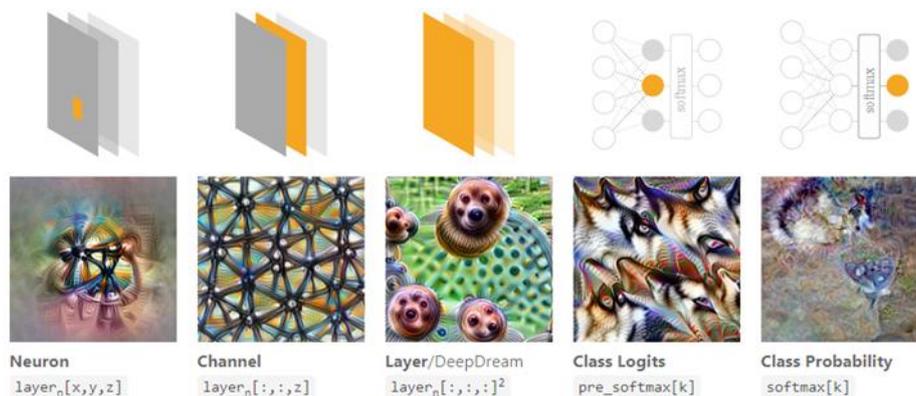


Figure 20 : Différentes parties d'un CNN examiné par la technique de visualisation des caractéristiques (Olah et al., 2017)

On peut s'intéresser à un groupe de neurones sur une des premières couches de convolution ou bien s'intéresser à une des dernières couches impliquées dans la classification (Class Logits, Class Probability). Comme nous pouvons le constater, la représentation des caractéristiques varie considérablement selon la position et l'échelle de visualisation choisie.

Nous pouvons voir dans la Figure 21 (gauche), un lot d'images représentant des constructions. L'image à droite représente une carte de visualisation des caractéristiques retenues par le réseau pour effectuer sa classification. Il est intéressant de constater que le ciel, qui n'est pas un élément inhérent à une construction soit pourtant utilisé pour faire une prédiction. Cette information n'aurait pas pu être fournie sans l'aide de ces outils d'évaluations visuelles. C'est d'ailleurs un exemple assez clair de sur-apprentissage du modèle qui a intégré une caractéristique (ciel) qui ne devrait pas être utilisée. En effet, le ciel ne devrait pas être pris en compte pour décider du type de construction que le modèle doit classer.



Figure 21 : Constructions... ou le ciel ? (Olah et al., 2017)

2 Maladie de Parkinson

La maladie de Parkinson est la deuxième maladie neurodégénérative la plus commune dans le monde après la maladie d'Alzheimer (Schrug et al., 2015). Selon la Haute Autorité de Santé, c'est aussi la seconde cause de handicap moteur d'origine neurologique chez les personnes âgées (« Guide parcours de soins maladie de Parkinson », 2016). Son incidence, en augmentation, suit le changement démographique de la population en vieillissement. En 2011 sa cause exacte n'est pas connue. Elle serait d'origine multifactorielle et combine des facteurs génétiques et environnementaux.

2.1 Anatomico-physiopathologie

La maladie de Parkinson se caractérise par un dépôt d'agrégats protéiques à l'intérieur des neurones (Eger et al., 2009). Ces protéines agrégées sont appelées corps de Lewy et sont responsables de la mort des cellules neuronales. Ces dépôts sont retrouvés dans la maladie de Parkinson ainsi que dans la démence à corps de Lewy. Une région particulièrement touchée est la substance noire qui est le siège de production des neurones dopaminergiques au niveau du mésencéphale selon la Figure 22. Le degré de sévérité des

symptômes moteurs corrélatent avec l'importance de la perte cellulaire de la substance noire. Néanmoins, il existe également une corrélation au niveau du noyau dorsal vague et le locus ceruleus qui produisent également des neurones. Ces neurones produisent principalement de la dopamine, et leur disparition provoque par conséquent une chute progressive des taux de dopamine dans le striatum.

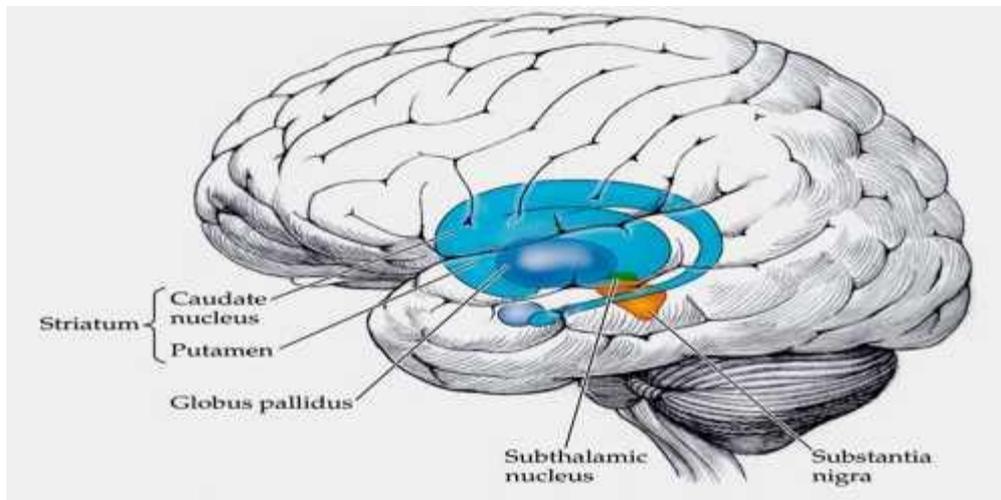


Figure 22: Anatomie du cerveau (Dubois-Mazeyrie, 2017)

Le striatum est composé du noyau caudé et du putamen. Il fait partie des noyaux gris centraux, présents dans chaque hémisphère, qui sont à leur tour composés par le pallidum, le noyau sous thalamique et la substance noire. Les noyaux gris centraux jouent un rôle primordial dans l'élaboration des mouvements. Pour en revenir au striatum, les neurones qui la composent, reçoivent un grand nombre d'afférences provenant du thalamus, du tronc cérébral ainsi que de différentes régions corticales telles que :

- Le cortex moteur
- Le cortex somato-sensoriel
- Le cortex visuel

Le but de ces neurones est d'intégrer ces afférences et sont ensuite projetées sur les neurones du pallidum. En passant par divers relais neuronaux, les neurones du pallidum contrôlent l'activation des noyaux thalamiques qui sont impliqués dans la génération des mouvements. En effet, ces noyaux ont à leur tour des neurones projetant sur les aires motrices corticales et permettent l'activation d'un acte moteur.

Les neurones dopaminergiques de la substance noire sont reliés au striatum via le putamen. Cette voie de liaison s'appelle le faisceau nigro-strié qui se termine par des synapses au niveau du putamen. Le putamen est composé de récepteurs à la dopamine D1 et D2 qui ont respectivement un effet activateur et un effet inhibiteur sur les neurones du striatum.

Dans la maladie de Parkinson, la perte neuronale au niveau de la substance noire y provoque une diminution de la concentration en dopamine ainsi que dans le striatum. On peut aussi voir une diminution des terminaisons dopaminergiques striatales ainsi qu'une dégénération du faisceau de communication nigro-strié. Cette carence en dopamine engendre des perturbations de l'équilibre des ganglions gris centraux, ce qui joue un rôle déterminant dans la pathogénèse de la maladie.

2.2 Méthode diagnostique

2.2.1 Pré-diagnostic

Le diagnostic de la maladie de Parkinson repose sur l'identification des symptômes moteurs classiques qui sont incorporés dans les critères de diagnostic (Schrage et al., 2015). Néanmoins, au moins 50% des neurones nigrostriataux sont déjà perdus au moment du diagnostic. De plus, d'autres anomalies pathologiques sont susceptibles d'apparaître avant même le début de la dégénération nigrostriatale. C'est pourquoi les patients avec des diagnostics ultérieurs de la maladie de Parkinson se présentent à leur médecin généraliste avec des symptômes pouvant présenter des signes avant-coureurs de la maladie de Parkinson. On retrouve donc les symptômes moteurs liés à la dégradation neuronale mais aussi des symptômes non-moteurs n'y étant pas liés. On peut retrouver des symptômes pouvant arriver bien avant le diagnostic de la maladie qui sont représentés dans le Tableau 1 par ordre d'incidence décroissant.

5 ans pré-diagnostic	2 ans pré-diagnostic
Tremblements	Tremblements
Troubles de l'équilibre	Rigidité
Constipation	Raideurs à l'épaule
Hypotension	Problèmes de mémorisation
Troubles de l'érection	Insomnies
Vertiges	
Troubles du sommeil paradoxal (REM)	
Dépression	
Anxiété	

Tableau 1 : Symptômes pré-diagnostic

On pourrait même retrouver jusqu'à dix ans avant le diagnostic de Parkinson l'apparition de tremblements et de problèmes de constipation.

2.2.2 Diagnostic Clinique

La maladie de Parkinson est une maladie lente et progressive (Eger et al., 2009). Selon l'HAS, en dépit de ces signes précurseurs, la spécificité diagnostic de la maladie de Parkinson à un stade précoce est relativement faible. Le médecin généraliste a un rôle important dans la détection de la maladie de Parkinson. Etant généralement le premier consulté, il peut être le premier à soupçonner une maladie de Parkinson. Il est donc habilité à évoquer une présence éventuelle d'une maladie neurologique après avoir évalué des fameux troubles moteurs comme :

- Une difficulté de l'initiation des mouvements
- Une lenteur des mouvements
- Une pauvreté du mouvement

Ainsi qu'au moins un des signes suivants qui sont :

- La rigidité
- Le tremblement au repos
- L'instabilité posturale

Le patient est par la suite redirigé vers un médecin neurologue afin de procéder à une série de tests neurologiques plus approfondis. Il va notamment regarder le patient marcher, examiner leur hypertonie au niveau du poignet et des coudes, leur écriture, il va poser des questions sur l'odorat ou encore le sommeil. C'est à lui que revient la confirmation ou non du diagnostic de de Parkinson. Bien que la certitude du diagnostic clinique puisse parfois être élevée, cela n'est souvent pas le cas. En effet, des processus de suivi et de réévaluation de l'état de la progression de la maladie et des symptômes peuvent être nécessaires. C'est pourquoi la durée entre les premières consultations et le diagnostic final peut durer plusieurs mois voire même des années (« Guide parcours de soins maladie de Parkinson », 2016).

Lorsque le médecin neurologue a un tableau clinique manquant, ou lorsqu'il y'a des signes évocateurs possible d'une maladie autre que celle de Parkinson, on peut être amené à faire différents tests complémentaires comme des tests génétiques, des tests sanguins ou de l'imagerie médicale (« Guide parcours de soins maladie de Parkinson », 2016). En imagerie, on peut retrouver :

- L'IRM ou le scanner fournit une information sur les structures morphologiques
 - Permet le diagnostic différentiel de maladies cérébro-vasculaires ou d'atrophies corticales
- Le DaTSCAN fournit une information sur les processus métaboliques au niveau striatum
 - Mesure la perte dopaminergique neuronale fonctionnelle

2.2.3 Examen au DaTSCAN

La scintigraphie cérébrale au DaTSCAN¹ est un examen pratiqué en Médecine Nucléaire. Il permet l'observation de la densité des terminaisons des neurones dopaminergiques fonctionnelles au niveau du striatum lors de la maladie de Parkinson.

2.2.3.1 Indication

Elle est destinée chez des patients atteints de symptômes parkinsoniens cliniquement vagues dont le but est d'effectuer un diagnostic différentiel (*compendium.ch*, 2020). En effet, elle est requise lorsqu'on veut faire un diagnostic différentiel entre un tremblement essentiel et des syndromes parkinsoniens liée à la maladie parkinson idiopathique, à l'atrophie multi systématisée ou à la paralysie supra nucléaire progressive. Néanmoins, il ne permet pas le diagnostic différentiel des trois syndromes parkinsoniens entre eux.

Elle également indiquée lors du diagnostic différentiel entre la démence à corps de Lewy et la maladie d'Alzheimer. Elle permet aussi la détection précoce de parkinsonismes neurodégénératifs ou bien la détection de syndromes parkinsoniens d'origine médicamenteuse.

2.2.3.2 Principe

L'examen consiste à évaluer la bio-cinétique d'un produit radiopharmaceutique injecté au patient qui va notamment se fixer au niveau du striatum. Le principe actif du DaTSCAN est l'ioflupane [¹²³I] étant un radiotraceur composé d'une molécule vectrice (loflupane) qui est analogue de la cocaïne. Elle présente une forte affinité pour la protéine de transport présynaptique de la dopamine (DAT=Dopaminergic Active Transporter). On peut examiner l'intégrité des neurones dopaminergiques nigrostriataux au niveau du striatum grâce au marqueur radioactif à l'iode-123. Avec une demi-vie de 13,2h cet isotope émet notamment un rayonnement gamma de 159 KeV (Pigrée et al., 2000).

La réception de ce rayon sur les détecteurs de l'appareil d'imagerie SPECT nous permet d'imager la bio-cinétique du marqueur au niveau des régions striatales (Single Photon Emission Computed Tomography). Nous pouvons voir un exemple d'appareil SPECT dans la Figure 23.

¹ DaTSCAN est un nom commercial du radiotraceur composé de l'ioflupane. (*compendium.ch*, 2020)



Figure 23 : SPECT-CT (« GE Healthcare Unveils Discovery NM 630 SPECT | Medgadget », 2011)

2.2.3.3 Procédure

La procédure de l'examen inclut la préparation du patient, le positionnement, l'acquisition et le traitement des images, selon l'Association Européenne de Médecine Nucléaire (Darcourt et al., 2010). Il est important d'éviter, avant l'examen, la consommation de médicaments ou l'abus de drogues susceptibles d'influencer l'analyse quantitative de la liaison du produit injecté. La suite de la procédure est présentée dans le Tableau 2.

Avant l'injection	Prise orale du perchlorate de sodium protégeant la thyroïde de l'accumulation d'iode libre radioactif
Injection	Injection intra-veineuse de 185Mbcq d'Ioflupane
Après l'injection	Attendre entre 3h et 6h après l'injection pour commencer l'acquisition des images Encourager le patient à uriner juste avant de commencer l'examen
Positionnement sur la table d'examen	Placer le patient en décubitus dorsal Eviter une contention trop ferme pour immobiliser la tête du patient

Tableau 2 : Procédure au DaTSCAN

Il est important de privilégier le confort du patient. En effet, cela permet une diminution du risque d'artéfacts de mouvements. De plus, l'obliquité du crane peut-être corrigée durant le traitement des images.

2.2.3.3.1 Acquisition des images

Une fois que le patient est placé sur la table, il faut procéder à l'acquisition des images qui peut durer environ 30 minutes (Darcourt et al., 2010). Il est possible d'utiliser entre deux et trois têtes détectrices. Les collimateurs à privilégier sont ceux qui sont paramétrés à l'iode-123 ou les collimateurs en éventail (fan-beam). L'utilisation de ce dernier permet un meilleur

compromis entre la résolution spatiale et taux de capacité de détection. Néanmoins, les collimateurs LEHR ou LEUHR (Low Energy-(Ultra-)High Resolution) sont aussi tolérés. Nous pouvons voir dans Tableau 3 des paramètres d'acquisition supplémentaires.

Matrice	128x128
Angle d'incrémentation (360 degrés)	Tous les 3 degrés
Mode d'acquisition	Step and Shoot
Rayon de rotation	Le plus faible possible
Nombre de particules à détecter	1 millions (inférieur si correction de dispersion)

Tableau 3 : Paramètres d'acquisition d'image

2.2.3.3.2 Traitement des images

Après que les images sont acquises, il faut procéder à un traitement des images qui sont initialement dans un état brut sous forme de sinogrammes (Darcourt et al., 2010). Tout d'abord, il faut évaluer les projections en mode cinématique pour évaluer le degré de mouvement ou d'artéfacts. La méthode de reconstruction est itérative et se fait par rétroprojection filtrée. Des paramètres de traitement d'images supplémentaires sont illustrés dans le Tableau 4.

Reconstruction du cerveau	Cerveau en entier Résolution la plus haute
Filtration	Filtre passe-bas (ex : Butterworth) Préfiltration des projections avec un filtre 3D / Filtre 2D sur les images reconstruire
Correction d'atténuation	Correction d'homogénéité selon la technique de Chang ($\mu=0.10-0.12 \text{ cm}^{-1}$) / Mesure de la carte d'atténuation acquise en séquentiel ou simultané avec un balayage de transmission ou un scanner

Tableau 4 : Paramètres de traitement d'images

2.2.3.4 Interprétation des images

Lorsque le traitement des images est terminé, le TRM envoie les images au médecin nucléaire qui va les interpréter. Il est possible de faire une interprétation visuelle ainsi qu'une évaluation semi-quantitative des régions striatales.

2.2.3.4.1 Interprétation visuelle

L'évaluation visuelle permet d'examiner le niveau de normalité de la fixation spécifique du DAT. Lorsqu'il est anormal, on peut évaluer la magnitude de la fixation du DAT. Cela fournit particulièrement une information sur l'asymétrie entre le striatum de gauche et de droite, ainsi

que sur leur forme. Il faut prendre en compte l'âge ainsi que des informations morphologiques issues d'éventuels CT ou IRM (Darcourt et al., 2010).

En outre, il faut tenir compte :

- Du niveau de contraste
- De la suppression du fond
- Des artéfacts techniques tels que :
 - Mouvements de la tête
 - Artéfact d'atténuation
 - Centre de rotation
 - Manque d'homogénéité
- De la médication

2.2.3.4.2 Evaluation semi-quantitative :

En plus de l'interprétation visuelle, l'analyse semi-quantitative est recommandée afin d'évaluer objectivement les niveaux de fixations striatales (Darcourt et al., 2010). Cette étape est faite avec des logiciels de quantification tels que QuantiSPECT, DaTView ou DaTQUANT comme illustré dans la Figure 24. Généralement, on définit des Régions d'intérêts (ROI) sur les coupes axiales / obliques. Il existe deux manières possibles qui sont la sélection de la coupe avec le plus de haut niveau de fixation spécifique ou la sélection de tout le volume striatal.

Puis, il faut :

- Définir un ROI pour évaluer la fixation spécifique du DAT dans le striatum ainsi que les régions striatales (noyau caudé, putamen)
- Définir un ROI avec une absence ou une présence faible pour évaluer la fixation non spécifique (ex : cervelet)
- Recourir à une taille de ROI avec une forme et une taille standardisée
 - Idéalement, faire un ROI adapté à la morphologie à partir d'images de fusion IRM
- Comparer les valeurs de fixation spécifiques à un sujet normal en ajustant avec l'âge.

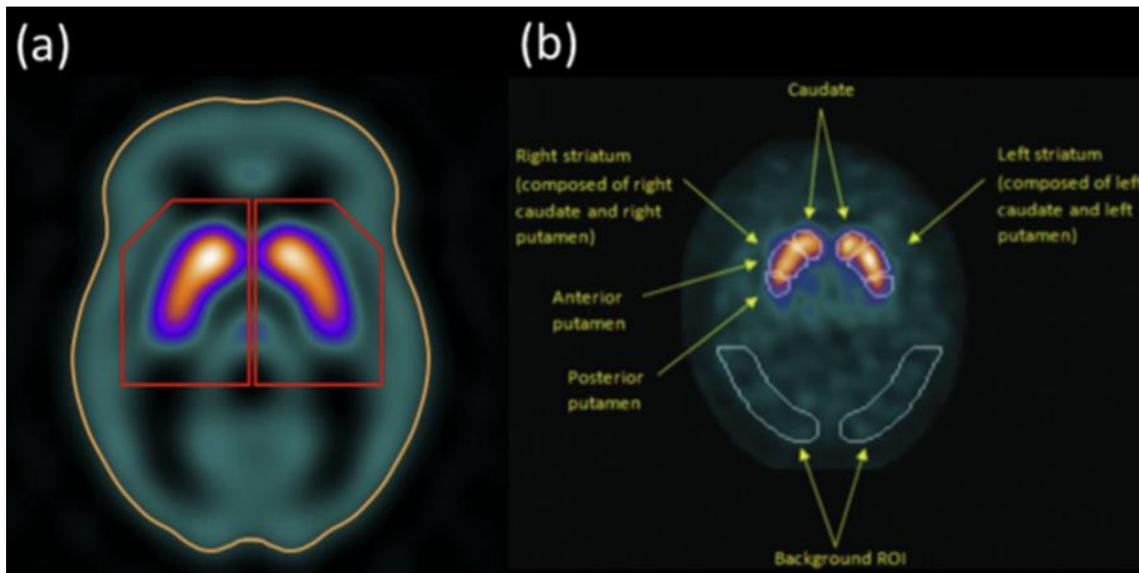


Figure 24 : Analyse Semi-quantitative (Shimizu et al., 2017)

La formule de la fixation spécifique du produit radiopharmaceutique sur la région striatale est la suivante :

$$\frac{(\text{Nombre de comptage d'événements moyens du ROI striatal} - \text{nombre de comptage d'événement du ROI non spécifique})}{\text{nombre de comptage d'événement du ROI non spécifique}}$$

En outre, l'analyse semi-quantitative doit prendre en compte les mesures de corrections et est comparée avec des fixations du DAT spécifiques de sujets sains obtenues par des techniques de ROI ajustés avec l'âge. Généralement, on analyse le striatum en entier composé de la tête du noyau caudé et le putamen. La détermination du ratio entre le putamen et le noyau caudé peut également être une aide (binding ratio).

3 Diagnostic assisté par ordinateur de la maladie de Parkinson

Depuis environ dix ans, on retrouve dans la littérature l'apparition d'études basées sur les systèmes CAD dans le diagnostic de la maladie de Parkinson au DaTSCAN en Médecine Nucléaire. Elles se sont notamment intéressées aux performances du CAD.

Le Tableau 21 de l'annexe Tableau récapitulatif des études antérieures 10.1 recense une vingtaine d'études en décrivant les bases de données utilisées (BDD), les méthodes d'entraînement, le pré-traitement d'image, le type d'algorithme de classification, le nombre d'images ayant été utilisés ainsi que les performances diagnostiques.

3.1 Etudes sur PPMI

Selon notre recherche de la littérature, vingt-neuf études ont été recensées concernant l'utilisation du CAD à l'examen DATSCAN et l'évaluation de ces performances de classification. Dans cette section, seules les études ayant utilisé la base de données PPMI (Parkinson's Progression Markers Initiative) sont traitées. Il s'agit d'une base de données accessible au public et qui regroupe de nombreuses données de patients composées d'images radiologiques de différentes sources (IRM, CT, MN) ainsi que des informations sur des compositions sanguines tels que des biomarqueurs. Selon les bases de données d'entraînement utilisées, 17 études ont utilisé la base de données PPMI à partir de l'année 2014 jusqu'à ce jour. Elles sont toutes basées sur du ML.

3.1.1 Machine Learning classique

Dix de ces études ont utilisé du ML classique. Toutes les études utilisant du ML traditionnel ont un entraînement supervisé et ont utilisé la méthode de classification SVM à l'exception de Rumman et al. (2018), qui ont utilisé une technique plus récente basée sur le MLP. De plus, quatre d'entre elles la compare à d'autres techniques comme le BT, le K-NN, le LR, le NB, le RF, le RM, ou le MLP (Castillo-Barnes et al., 2020; Mabrouk et al., 2019; Oliveira et al., 2018; Prashanth et al., 2017). Oliveira et al., auteurs de deux études en 2015 et 2018 ont eu recours à la méthode SVM et ont obtenu un taux d'exactitude de 97.9%. Ce sont les meilleures performances que l'on puisse retrouver dans la littérature parmi les études ayant utilisé du ML classique.

3.1.2 Deep Learning

Six études ont utilisé du DL. Elles sont basées majoritairement sur des réseaux de type CNN et ont toutes bénéficié d'un entraînement semi-supervisé. Sa première utilisation remonte en 2016 par (Ortiz et al.) qui obtiennent une exactitude de 95%. On peut aussi citer le réseau PD-net qui a été élaboré par Choi et al. (2017). C'est un réseau de type 3D qui a été spécialement conçu dans le cadre du diagnostic de la maladie de Parkinson au SPET. Il a également été utilisé et amélioré par Pianpanit et al. (2017). Ces deux études ont respectivement eu une exactitude de 96% et de 96.9%. Tagaris et al. (2017) ont utilisé le ResNet-50 qui est un réseau de type récurrent. Les réseaux résiduels ont des capacités de représentation puissantes et ont de très bons résultats en détection d'image (Fung, 2017). De plus, ils ont aussi montré des résultats importants dans le domaine de la détection d'objets et la reconnaissance de visage. Composé d'un réseau dont les connexions pouvant directement sauter à une couche suivante, cela lui permet d'éviter certaines couches lorsque le réseau subit une baisse de performances à cause de certaines d'entre elles (« Residual Networks (ResNet) - Deep Learning », 2020). La spécificité de ce réseau est de réduire le problème de la « disparition du gradient »

(vanishing gradients) souvent observé avec des réseaux à de nombreuses couches (réseau profond). Certaines couches, pouvant être responsables d'une dégradation lors de la classification d'images, vont être évitées grâce à un système de régulation récursif. C'est pourquoi il s'agit d'une technique de saut de connexions (« Residual Networks (ResNet) - Deep Learning », 2020). Les meilleures performances basées sur du DL sont détenues par (Quan et al., 2019) qui obtiennent une exactitude de 98.5%. Ce sont également les meilleures performances que l'on puisse retrouver dans toute la littérature que ce soit du ML classique ou du DL.

3.2 Normalisation des images

3.2.1 Normalisation géométrique

La normalisation géométrique peut être effectuée à l'aide de SPM (Statistical Parametric Mapping, qui est un logiciel couramment utilisé dans des études (Castillo-Barnes et al., 2018, 2020; Choi et al., 2017; Martinez-Murcia et al., 2013a). Il permet l'automatisation du procédé de normalisation géométrique. On retrouve aussi le logiciel moins connu de Shit (Sheffield Image Registration Toolkit) utilisé par plusieurs études (Illán et al., 2012; Martinez-Murcia et al., 2013a; Taylor & al., 2016).

On peut également retrouver une déformation élastique. Cette seconde étape permet de déformer l'image qui se base sur une image de référence que l'on appelle un Template. Des études ont créé leur propre Template (Illán et al., 2012; Martinez-Murcia et al., 2013a; Oliveira et al., 2015; Taylor & al., 2016a). Par exemple, Illán et al. (2012), ont créé leur propre Template en sélectionnant à partir d'un jeu de données de patients, une centaine d'images de patients sains et les ont sommées pour obtenir une seule image de référence (Template). On peut voir dans la Figure 25, un exemple de template « MNI » disponible sur internet et représenté dans les 3 plans de l'espace. Il a été construit à partir de 30 images de sujets sains en médecine nucléaire et a notamment été utilisé par plusieurs études (Garcia-Gomez et al., 2013; Mabrouk et al., 2019; Magesh et al., 2020; Prashanth et al., 2014). La région hypercaptante correspond typiquement à la région striatale.

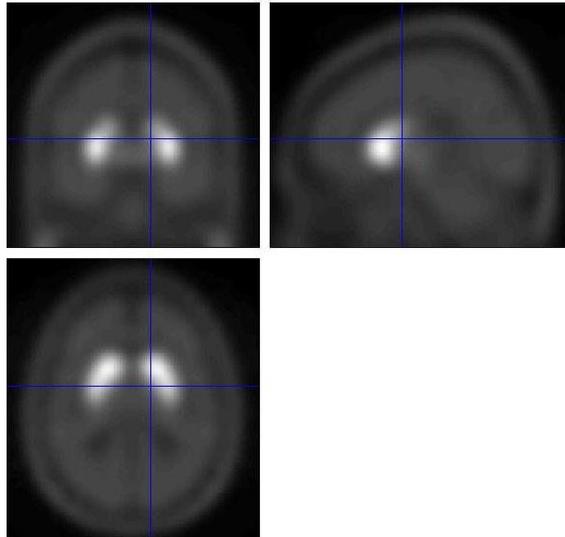


Figure 25 : Template MNI (Garcia-Gomez et al., 2013)

3.2.2 Normalisation d'intensité

La normalisation d'intensité des images permet d'homogénéiser les échelles d'intensité au travers de la base de données (Ortiz et al., 2016). En effet, il y a une différence entre chaque sujet, à propos des valeurs de captation dans des régions d'activité spécifiques comme la région striatale (transporteurs DAT) et les régions d'activités non spécifiques comme le cortex, le cervelet ou le réseau vasculaire dans le cadre du DatSCAN (Illán et al., 2012; Martinez-Murcia et al., 2013a; Rojas et al., 2013). Le niveau de captation dans les régions d'activité spécifiques doit correspondre aux valeurs d'intensité d'image pour chaque patient. Il existe différentes méthodes de normalisation d'intensité telles que :

- La normalisation d'intensité à l'intégrale (Illán et al., 2012; Ortiz et al., 2016a; Rojas et al., 2013).
- La normalisation d'intensité au maximum (Illán et al., 2012a; Martinez-Murcia et al., 2013b; Oliveira & Castelo-Branco, 2015) : Selon les premiers auteurs, cette technique serait plus stable que la normalisation à l'intégrale. Ils se réfèrent au 85^{ème} ou au 95^{ème} percentiles de l'histogramme d'intensité pour éviter des phénomènes de saturation à cause de valeurs aberrantes. Cette technique est la plus utilisée.
- La normalisation d'intensité à la moyenne (Taylor et al., 2016).
- La normalisation d'intensité selon la distribution alpha-stable (Augimeri et al., 2016a; Castillo-Barnes et al., 2018) : Selon les derniers auteurs, cette dernière technique serait plus efficace que les techniques utilisées jusqu'à maintenant.

Par exemple, la Figure 26 représente des images cérébrales acquises au DaTSCAN avant la normalisation d'intensité (deuxième ligne) et après normalisation d'intensité (première ligne) basée sur la distribution d'alpha-stable (Castillo-Barnes et al., 2018).

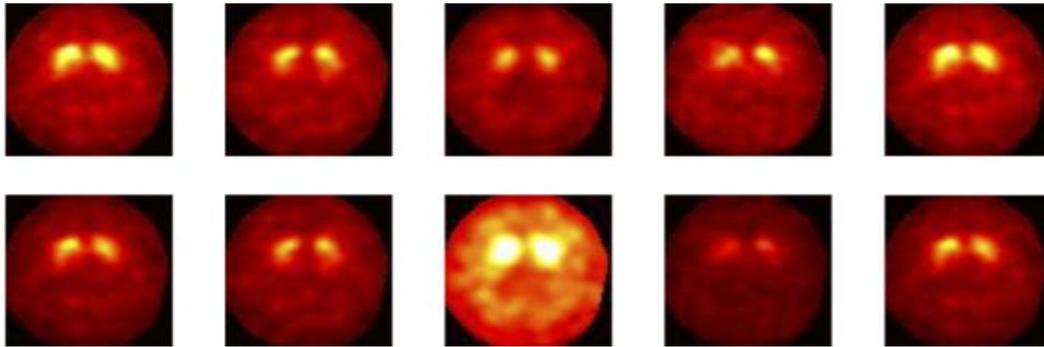


Figure 26 : Normalisation d'intensité cérébrale (Castillo-Barnes et al., 2018)

3.2.3 Masquage

Une fois que les images ont été normalisées, un ROI des images peut être sélectionné pour extraire les régions pertinentes. Cette procédure de masquage a un double intérêt (Ortiz et al., 2016) :

- Evite de conduire les différentes étapes comme l'extraction de caractéristiques avec des voxels non pertinents sachant que le DaTSCAN est seulement focalisé sur le striatum.
- La réduction de la dimension des données permet une moins grande charge de calcul de l'algorithme DL et potentiellement une amélioration des performances.

De la même façon que la normalisation géométrique, Il est possible d'utiliser un Template qui va attribuer à la région extérieure à la zone d'intérêt une valeur nulle aux voxels externes. Quatre autres études ont procédé à une suppression des régions extra-striatales (Klyuzhin et al., 2018; Oliveira et al., 2018b; Pianpanit et al., 2017; Rumman et al., 2018). Une autre étude a uniquement supprimé la région extra-cérébrale avec un Template (Castillo-Barnes et al., 2020).

3.3 Traitement des images en ML

Concernant le pré-traitement des images, 8 études sur 11 ont utilisé une normalisation spatiale. Les trois meilleures études ont utilisé le Template MNI (Martinez-Murcia et al., 2013a; Oliveira et al., 2018a; Oliveira et al., 2015). La plupart des études n'ayant pas utilisé de la normalisation spatiale ont eu les moins bons résultats (Brahim et al., 2017; Mabrouk et al., 2019; Rumman et al., 2018). Concernant la normalisation d'intensité, 7 études sur 11 l'ont utilisée. Les 4 meilleurs études ont utilisé la normalisation d'intensité (Martinez-Murcia et al., 2013a; Oliveira et al., 2018a; Oliveira et al., 2015; Prashanth et al., 2017), celle basée sur le

maximum étant la plus utilisée. Quant à l'étude de Oliveira et al. (2018), ils ont notamment utilisé une normalisation spatiale, une normalisation d'intensité et un filtre de lissage gaussien, qui, rappelons-le, détiennent le meilleur résultat dans la littérature avec du ML. Il semble que la combinaison de la normalisation d'intensité et la normalisation spatiale permettent d'avoir de très bonnes performances dans le cadre du ML.

3.4 Traitement des images en DL

Concernant le pré-traitement des images, bien que le concept de normalisation spatiale soit souvent évoqué, il n'est pas appliqué. En effet, les images de la base de données PPMI ont déjà eu un processus de normalisation spatiale avec une transformation affine à l'aide du Template MNI (Marek et al., 2018). De plus, un filtre gaussien de 6 mm a été appliqué sur toutes les images. Seulement une étude sur six a utilisé une normalisation spatiale (Magesh et al., 2020). A propos de la normalisation d'intensité, deux études ont utilisé cette technique et ont eu une exactitude respective de 95.2% et 95.% (Magesh et al., 2020; Ortiz et al., 2016). La moitié des études, ont procédé à une suppression des régions extra striatales (Klyuzhin et al., 2018; Ortiz et al., 2016; Pianpanit et al., 2017).

3.5 Augmentation des données

D'après nos recherches, seules les études basées sur du DL ont utilisé l'augmentation des données. La moitié d'entre elles ont utilisé des techniques d'augmentation géométrique et/ou d'intensité. Quan et al. (2019), avec les meilleures performances, ont notamment augmenté leur base de données d'entraînement avec des techniques de translation et de variations d'intensité. Choi et al. (2017), ont uniquement utilisé des techniques d'inversion entre le côté gauche et droite. Magesh et al. (2020), ont également utilisé des techniques d'inversion et y ont ajouté des translations. Enfin, Klyuzhin et al. (2018), dont seulement l'aire sous la courbe ROC a été mesurée (AUC=0.69), ont utilisé des augmentations géométriques telles que des translations, des rotations et une autre technique d'augmentation basée sur le floutage.

3.6 Transfer Learning

Deux études ont eu recours au principe de transfert d'apprentissage (Transfer Learning) afin de préparer leur réseau (Magesh et al., 2020; Quan et al., 2019;). Cette technique permet de pré-entraîner un réseau neuronal à partir de bases de données relativement grandes (Bengio et al., 2012). Son objectif est de procéder à un ajustement des poids du réseau neuronal. Ensuite, il faut fournir la base de données d'intérêt sur lequel le réseau va devoir s'entraîner à nouveau. Grâce à un entraînement antérieur, il aura gardé en mémoire les poids qui va donc bénéficier d'un transfert de connaissance. Magesh et al. (2020) et Quan et al. (2019), ont les

deux fait un TL à partir de la base de données en ligne ImageNet. Avec de l'augmentation de données, le TL sont les seules techniques qui ont été utilisées Quan et al. (2019).

4 Rôles et compétences du TRM en relation avec l'IA

4.1 Un environnement de travail en plein évolution

La radiologie est un secteur ayant vécu de grands changements. Comme le dit Sloane et al. (2017), la profession du TRM est dans un état de fluctuation. En Anglais, ce terme signifie un état d'incertitude à propos de ce qui devrait être fait (à la suite d'un événement important) avant l'établissement d'une nouvelle direction d'action (*state of flux - definition*, 2021). En l'occurrence, la série d'évènements importants à laquelle la profession fait face est l'évolution fulgurante de la technologie et de l'automatisation.

Selon le rapport de Veille Métier de l'Association Suisse des Techniciens en Radiologie Médicale (ASTRM), le métier TRM est fortement impacté par la digitalisation de l'imagerie médicale (RVM, 2019). Un des principaux rôles du TRM en progression concerne le traitement des données à la suite de leur acquisition. Avec l'arrivée de l'IA, cette profession sera même amenée à interpréter les résultats diagnostics produits par l'IA à la suite d'une prestation radiologique à un faible niveau de complexité. Le choix du logiciel d'IA approprié à utiliser selon la nature de l'examen et son indication pourra aussi lui appartenir. Ses nouveaux rôles et responsabilité impliqueront le TRM à un nombre accru de contact directs avec des professionnels de la santé dans un contexte interdisciplinaire. Elle ne se limitera donc plus aux contacts avec le médecin radiologue / nucléariste. Il y aura donc l'émergence du rôle de partenaire clinique-gestionnaire de l'information. Il y a par exemple les rapports au médecin, des retours sur la validité des réseaux neuronaux et de son utilisation, mais aussi une contribution à la recherche et au développement de la qualité des soins.

En 2007, l'association américaine des médecins (AAPM) a créé le sous-comité en imagerie diagnostique en CAD (CADSC), dans un souci de pouvoir créer des standards dans ce domaine et tenir les membres informés des nouveaux développements (Huo et al., 2013). Selon eux, en plus des radiologues, il pourrait être judicieux d'entraîner les TRM à l'utilisation du CAD, surtout si la préparation ou le positionnement du patient peut affecter le système. Comprendre l'application visée du système CAD ainsi que ses limites est important car une mauvaise utilisation peut créer des résultats mauvais et cela peut même devenir dangereux. Une évaluation selon la modalité et l'application du CAD devrait être faite par la société américaine des TRM (American Society of Radiologic Technologist ou ASRT) pour déterminer s'il est pertinent de les former. Bien qu'il soit conseillé que le TRM puisse bénéficier d'un entraînement au CAD, ce dernier ne semble pas inclus comme le radiologue durant tout le

processus de développement et de mise en place du système. Pourtant, le TRM est un professionnel de la santé avec un vaste champ de compétences, fortement impliqué dans le secteur de la radiologie selon le rapport sur le champ d'activité du TRM en (Karlegger et al., 2013). Il a un rôle indispensable dans la chaîne de production d'images sans nommer la relation avec le patient. C'est pourquoi il est fort probable qu'il possède une expertise lui offrant une place légitime dans la participation au développement de systèmes CAD.

4.2 Les prérequis techniques nécessaires

Concernant les radiologues utilisant le CAD, ils doivent comprendre comment un système CAD fonctionne (Ginneken et al., 2011). En effet, cette dernière traite le problème en le découpant en différentes composantes. C'est pourquoi il faut avoir une compréhension de base sur ces composantes afin de comprendre pourquoi le système CAD peut parfois faire des erreurs alors que, pour l'homme, ces erreurs paraissent évidentes. Il est donc primordial d'avoir un certain nombre de prérequis que le TRM doit maîtriser afin de pouvoir comprendre, améliorer et surveiller des systèmes d'IA.

4.2.1 Base de mathématiques et statistiques

Le Machine Learning (ML) est une technique basée sur des algorithmes et des mathématiques appliquées. Cela signifie qu'il faut avoir de bonnes connaissances concernant l'algorithmie (Savaram, 2019), mais aussi sur les modèles statistiques et probabilistes. En effet, la majorité des algorithmes effectuent des prédictions se basant sur des conditions pouvant être incertaines et doivent prendre des décisions fiables. C'est pourquoi les décisions prises dépendent d'un modèle probabiliste, d'où l'importance d'avoir des bases en probabilité (Savaram, 2019). Avoir des bases en mathématiques est aussi fondamentale pour développer une expertise en IA (Eremenko, 2020).

4.2.2 Langage de Programmation

Travailler sur du ML est différent de devoir travailler sur des choses plus petites comme la création d'application simples ou bien de faire du script de pages web. En effet cela requiert des connaissances supplémentaires. Afin de se spécialiser en ML, il faut avoir des connaissances solides telles que l'architecture d'un ordinateur, les algorithmes, la structures des données et leur complexité. Cela implique donc des connaissances en programmation et en science computationnelle. Le machine Learning ne se limite pas à un langage de programmation spécifique (Savaram, 2019) et dépend surtout de l'objet d'étude. On peut par exemple retrouver le langage Python qui supporte par exemple la bibliothèque de ML de TensorFlow. Il existe aussi le langage R ou le langage C/C++.

4.2.3 Gestion des données

4.2.3.1 *Modélisation des données*

La tâche critique des algorithmes de ML est d'analyser des modèles de données pas toujours structurées, ce qui requiert la science de la modélisation des données (Savaram, 2019). La modélisation des données permet d'identifier la structure des données sous-jacentes, découvrir des tendances et compléter des écarts dans des parties où les données n'existent pas. Connaître des concepts de modélisation des données aide à créer des algorithmes efficaces.

4.2.3.2 *Curation des données*

Avoir des connaissances en exploration des données est aussi un prérequis pouvant être d'une grande aide lors de l'élaboration de bases de données (Eremenko, 2020). Souvent, les sources d'informations proviennent de différents systèmes séparés et elles sont parfois bien intégrées entre-elles. Mais la plupart du temps, elles sont très mal agencées. La curation des données est une solution à ce problème. Selon Thirumuruganathan et al. (2020), la curation de données est le processus de découvrir, intégrer et nettoyer des données provenant de sources différentes pour des tâches d'analyse. Bien que cela paraisse chronophage, c'est une étape cruciale car elle joue un rôle important pour libérer leur valeur. C'est par conséquent une étape que le TRM doit pouvoir maîtriser. L'intégration des données consiste à les organiser et les structurer dans un support unique. En plus de les authentifier, il faut aussi vérifier qu'il n'y ait pas d'erreur ou de valeurs manquantes qui pourraient être corrigées. Dans le cas contraire il peut être judicieux de supprimer celles qui sont non pertinentes. C'est durant l'étape de curation que l'on peut être amené à exclure de la base de données des images qui ont été acquises selon un protocole non conforme. Il faut donc qu'un processus de curation des données soit mis en place. Le format DICOM est un standard de format permettant de structurer les images radiologiques digitalisées en incluant par exemple les paramètres d'acquisition d'image ou de reconstruction (Currie et al., 2019). Cette curation est souvent faite manuellement par la main d'experts en annotant les images DICOM afin d'entraîner les algorithmes et les faire réagir de manière appropriée. Toutefois, ce format ne suffit pas puisqu'il n'intègre pas les données cliniques telles que le diagnostic ou toutes les caractéristiques personnelles du patient.

4.2.3.3 *Prétraitement des données*

En milieu pratique, les systèmes CAD peuvent être amenés à travailler sur des bases de données d'images provenant de protocoles ou de sources pouvant être différents (Ginneken et al., 2011). Afin d'homogénéiser ces données, il est crucial de standardiser les images comme vu à la section 1.5 sur la préparation des données. Bien que le TRM ait une expertise en traitement d'image (section 4), il existe de nombreuses techniques dont il n'est peut-être

pas accoutumé. Il doit donc apprendre des techniques de traitement d'images propres aux systèmes d'interprétation d'images telles que la normalisation ou l'augmentation des données.

4.2.4 Evaluation de l'IA

Afin pouvoir évaluer les performances de l'IA, il est nécessaire de maîtriser les métriques d'interprétation de ce dernier selon la nature d'utilisation (RVM, 2019). Dans le cadre d'un système CAD à visée diagnostique, ceux sont les mesures de performances diagnostics qui doivent être comprises par le TRM. Il faut par exemple citer les notions de sensibilité, de spécificité, les courbes ROC, telles qu'explicitées dans la section 1.6 du cadre théorique concernant l'interprétation d'un réseau neuronal. Savoir interpréter un graphe illustrant l'aire sous la courbe ROC et identifier par exemple les faux positifs sont donc des prérequis.

4.2.5 Entraînement des utilisateurs

Comme dit au début de la section 4, il est important d'entraîner le TRM pour bénéficier du plein potentiel des outils de CAD (Huo et al., 2013). Il n'existe néanmoins pas de guide de bonne pratique pour savoir comment entraîner les TRM, sachant que ceux-ci doivent entre autres recevoir des remises à niveau.

4.2.6 Assurance qualité

Selon Huo et al. (2013), il est important de s'assurer qu'un système CAD installé en milieu clinique fonctionne correctement à son installation et qu'il continue de fonctionner durant son utilisation. Bien que ces systèmes soient fiables, il peut arriver que le système soit corrompu par inadvertance ou bien qu'il y'ai des changements significatifs des systèmes d'acquisitions d'images. Des variations concernant la population étudiée peut aussi avoir lieu. Ces différents points peuvent provoquer une chute de performances des systèmes CAD. Cela peut par conséquent modifier le profil de risque-bénéfice initialement établi. Afin de préserver la sécurité du patient, des systèmes de contrôles qualités doivent être développés et implémentés pour les CAD sur les sites cliniques. Ainsi, les utilisateurs finaux peuvent régulièrement contrôler les performances de l'appareil et relever toute différence majeure de performance pouvant être critique. Le TRM pouvant être amené à effectuer des assurances qualité doit donc comprendre l'objectif de ces contrôles et y être entraîné.

La manière dont sont utilisés ces systèmes CAD par les utilisateurs finaux doit aussi être surveillée. Bien que ce genre d'outil soit habituellement utilisé en tant que second lecteur, des variations dans sa manière d'être utilisée peut potentiellement réduire son efficacité d'utilisation. Le profil risque-bénéfice du système peut aussi en être impacté. Des procédures d'utilisation doivent donc être mises en place.

Le CADSC propose des recommandations pour y remédier (Huo et al., 2013). Une liste des points importants à prendre en compte est disponible dans l'Annexe 10.2.

En outre, les tests d'acceptation doivent être documentés et doivent faire office de référence pour les futurs contrôles afin de vérifier la consistance des résultats. De plus, il est important d'encourager et d'entraîner le site à faire des tests d'acceptance. Cela requiert donc de suivre les instructions de la procédure fournie par le fabricant ou celle qui est élaborée par le site.

Puisque la base de données de test fournie pour les contrôles qualités par le fabricant est souvent petite et non représentative de la population étudiée sur le site, il est encouragé de récolter et utiliser sa propre base de données de test de manière rétrospective ou prospective. Néanmoins, l'utilisateur est le seul responsable en cas de mauvaise utilisation de cette dernière.

5 Méthode

Sur la base de notre cadre théorique nous avons pu mettre en évidence le fort potentiel du DL pour le CAD de la maladie de Parkinson, ainsi que l'importance des données et des traitements relatifs. Néanmoins, il n'est pas possible de considérer comme optimale la meilleure approche publiée. En effet, les bases de données utilisées par les études sont différentes. De plus, en considérant uniquement celles ayant utilisées la base de données PPMI, il réside tout-de-même une variabilité inter-étude concernant sa taille, ainsi que des échantillons qui y sont inclus. En outre, lorsque des études utilisent des basent de données locales qui ont toujours un échantillon très faible à disposition, les performances sont en conséquence réduites. Pourtant, aucune d'entre elles ont essayé de s'en affranchir en profitant de l'apport de plusieurs bases de données afin d'augmenter leur taille d'échantillon.

Dans un contexte clinique, cette problématique mérite d'être étudiée concernant la mise en place d'une IA dédiée au diagnostic de la maladie de Parkinson à travers la question et les objectifs de recherche suivants.

5.1 Question de recherche et objectifs

Notre question de recherche est : « Comment l'association de plusieurs bases de données issues de protocoles différents influe sur les performances de prédiction d'une IA sur la maladie de parkinson au DaTSCAN ? ». Les objectifs sont les suivants :

1. Comprendre l'apport de plusieurs bases de données à l'entraînement sur les performances de l'IA et établir la transférabilité de l'IA d'un protocole d'acquisition vers un autre.
2. Evaluer l'impact des méthodes d'augmentation et de balancing.

3. Evaluer l'impact des méthodes de normalisation spatiale et de lissage des images.
4. Evaluer l'impact du mélange de bases de données et du Transfer Learning.
5. Un autre objectif est de proposer des recommandations professionnelles aux TRM afin de contribuer, améliorer et créer un système CAD

5.2 Collecte des bases de données

Afin d'augmenter le volume de données et d'évaluer les variations de performances dues à la fusion de bases de données, 3 bases différentes ont été collectées.

1. Base « Symbia » : fournie par le service de médecine nucléaire des Hôpitaux Universitaires de Genève (HUG)
 - a. Images de qualité conventionnelle due à l'utilisation d'un SPECT-CT standard 2 têtes
 - b. Données cliniques utilisées comme base de référence car ce SPECT-CT est encore utilisé en routine clinique
2. Base « Toshiba » : fournie par le service de médecine nucléaire des HUG
 - a. Images de haute qualité grâce à un SPECT-CT 3 têtes spécialisé pour l'imagerie cérébrale
 - b. Base collectée afin d'affiner la prédiction de l'IA
3. Base « PPMI » : provient de la PPMI (Parkinson Progression Marker Initiative, 2011) étant un programme de recherche regroupant entre autres de l'imagerie radiologique concernant la Maladie de Parkinson provenant d'hôpitaux du monde entier
 - a. Images de qualité variable selon le protocole utilisé en clinique par les différentes institutions, néanmoins pré-normalisées par le groupe de recherche de la PPMI sur la base d'un protocole de post-traitement commun
 - b. Base de données volumineuse augmentant le nombre d'images
 - c. Utilisée à des fins de comparatif avec d'autres études ayant utilisé cette base avec des réseaux de neurones (= « benchmark »)

Des exemples d'images des différentes bases de données sont présentées dans la Figure 27 :

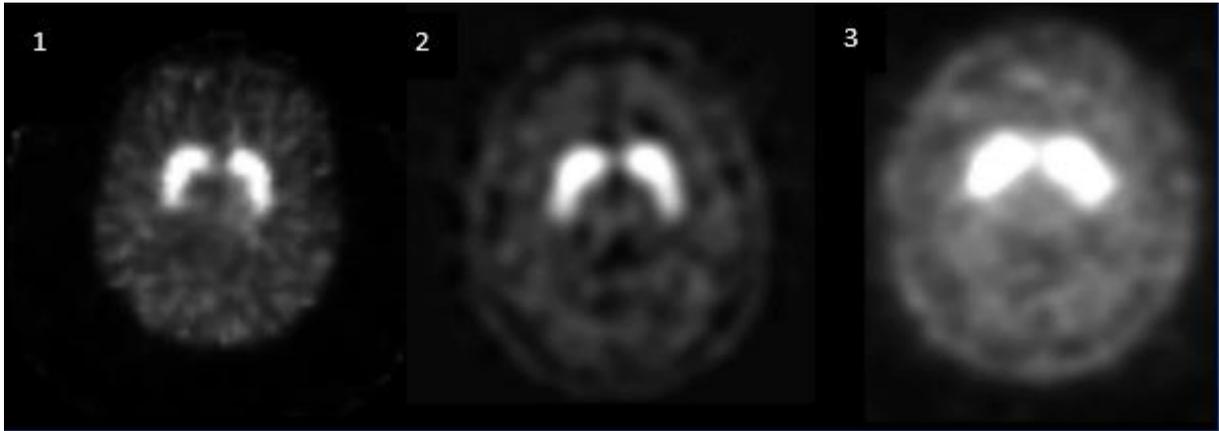


Figure 27: L'image « 1 » provient de la base de données Symbia, la « 2 » de la base de données Toshiba et la « 3 » de la base de données PPMI. Ces images sont non normalisées, hormis la « 3 », la base PPMI ayant été récoltée avec une pré-normalisation.

Une comparaison d'image Contrôle vs. Pathologique est présentée dans la Figure 28. Une explication du contexte anatomo-pathologique de la Maladie de Parkinson est présente dans la section 2.2.3.4.1 ainsi que de l'interprétation semi-quantitative utilisée Médecine nucléaire avec le DaTSCAN à la section 2.2.3.4.2.

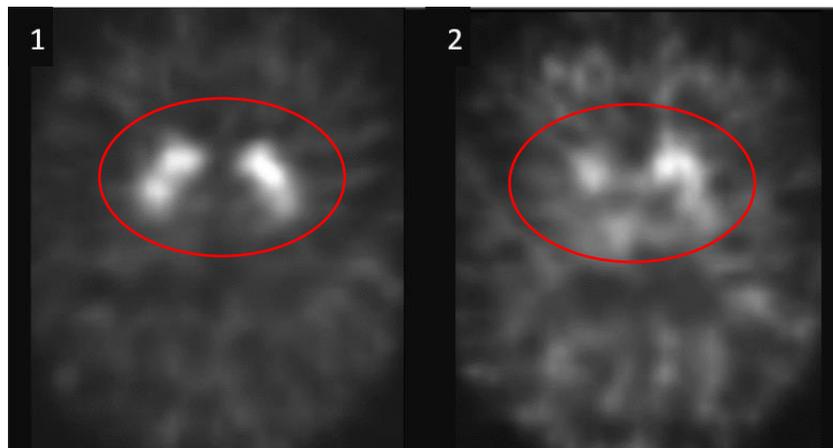


Figure 28 : Images comparatives du DaTSCAN d'un patient sain (1) et atteint de la Maladie de Parkinson de grade 3 (2). Les signes classiques observés sont : une asymétrie gauche/droite des noyaux gris ainsi que leur forme. En rouge la différence de symétrie des noyaux est observable, ainsi que la forme hétérogène de ceux-ci dans l'image (2). A noter que le reste de l'encéphale apparaît plus sombre sur l'image « saine » par la haute concentration du radiotracer dans les noyaux gris, pour un patient pathologique, la concentration faible du radiotracer dans les noyaux gris se manifeste par une gamme dynamique plus faible dans l'image.

Afin d'avoir une base de données comportant que des images SPECT avec le bon radiotracer, les données de la PPMI ont été sélectionnées selon une méthode de recherche avancée présentée dans la Figure 29.

Figure 29: Critères de recherche PPMI (PPMI data base, 2020). Les cases cochées en bleu représentent les catégories sélectionnées afin de n'avoir que les images de SPECT avec le bon radiotracteur, les patients pathologiques, contrôles et SWEDD, ainsi que celles déjà normalisées (« Parkinson's Progression Markers Initiative | Search Criteria », 2020)

Une liste d'identifiants de patients ainsi que de leurs images respectives sont finalement sélectionnées selon le protocole recherché. Les images intitulées "Cobalt" ou "Dual Isotope" sont donc systématiquement écartées. Les images conservées sont donc celles utilisant comme traceur l'ioflupane marquée à l'Iode-123.

Le Tableau 5 représente le nombre total d'images récoltées par catégories :

Base de données	N contrôles	N parkinson	Total
Toshiba	50	49	99
Symbia	55	43	98
PPMI	198	466	664

Tableau 5 : Représentation du volume de données de chaque base en : Nombre de patients contrôles (N contrôle), en nombre de patients atteints de la Maladie de Parkinson (N Parkinson), ainsi que leur nombre additionné.

Le volume de données présenté dans le Tableau 5 présente la base de données Symbia telle que récoltée au départ, cependant, 40 images ont été ajoutées en fin de recherche et entièrement utilisées pour la phase de test (Tableau 6). Les 40 images supplémentaires de la base de données Symbia récoltées a posteriori sont composées de 17 cas contrôle et 23 pathologiques.

Base de données	N entrainement	N test	Total
Symbia Priori	74	16	90
Symbia Posteriori	74	56	130

Tableau 6: Récapitulation du nombre de données de la base Symbia après ajout des 40 images.

5.3 Choix du réseau et paramétrage

Dans le cadre de cette étude, un réseau de type ResNet (He et al., 2016a, 2016b) a été utilisé et brièvement décrit dans la section 3.1.2 sur le DP. Il s'agit donc d'un réseau neuronal 3D avec un acheminement de l'information vers l'avant ayant la spécificité d'avoir des raccourcis de connexion entre certaines couches selon la Figure 30.

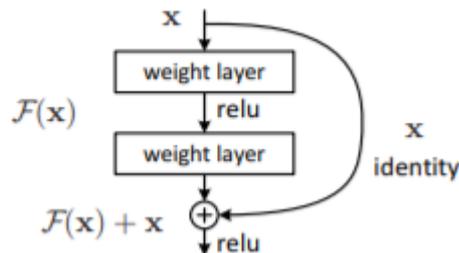


Figure 30: Réseau neuronal avec des raccourcis de connexion (He et al., 2016a)

En pratique, le réseau est bâti à partir d'une version modifiée de la librairie DLTK (Rajchl, 2021), laquelle est basée sur l'architecture TensorFlow (*TensorFlow*, s. d.) de Google. Notre réseau ResNet est décrit plus en détails dans l'annexe 10.3.

Le volume de l'image est découpé en n blocs de 96x96x48 pixels lors de l'entraînement ($n = 2$) et de l'inférence ($n = 15$). Ces blocs sont invariables en taille et en nombre, mais ont une position aléatoire dans le volume à chaque itération. Ces blocs sont fournis en entrée du réseau lors de l'entraînement et la prédiction. Lors de la prédiction, le réseau estime pour chacun de ces blocs les probabilités des classes « pathologique » ou « sain ». La moyenne des probabilités de tous les blocs par classe la plus élevée donne la prédiction finale ainsi qu'un certain degré de certitude du réseau concernant ce choix.

De manière générale, les images sont divisées de la sorte : sur le nombre d'images total, 80% sont utilisés pour l'entraînement et 20% pour le test, et sur ces 80% d'entraînement, 80% sont conservés pour de l'entraînement pur et 20% pour la phase de validation du réseau. Les données utilisées restent identiques pour chaque entraînement et tests par la suite, entre 2 tests à N=56 images par exemple, les patients tests sont toujours les mêmes et ne sont jamais représentés dans les données d'entraînement et vice-versa. Pour les tests sur la base de données Symbia, tel qu'expliqué dans le Tableau 6 de la section 5.2, plus d'images pour la phase de test peuvent être utilisés.

Les spécificités du serveur de calcul utilisé sont un Intel Xeon Gold 5122 à 4 CPUs de 3.6 GHz, 64 Go de RAM, et 4 GPUs Nvidia GTX 1080 avec 8 Go de RAM. Avec cette installation, un temps d'entraînement de 3h10 est nécessaire pour 25'000 itérations, indépendamment de la taille de la base d'entraînement

5.4 Evaluation visuelle des bases de données

Une évaluation visuelle de la qualité des images est effectuée par deux Techniciens en Radiologie Médicale (TRM), auteurs de ce mémoire, de façon concertée (Tableau 7).

	SNR	Résolution	Contraste	Taille de Matrice
Toshiba	5	5	4	128x128
Symbia	1	4	3	128x128
PPMI	3	3	3	128x128

Tableau 7 : score de la qualité d'image des différentes bases de données, 5 = excellent, 4 = très bon, 3 = bon, 2 = mauvais, 1 = très mauvais.

Les images de la base de données Symbia sont acquises avec un SPECT 2-têtes T-16 équipé d'un collimateur LPHR (Low Penetration-High Resolution). Ce collimateur permet de conserver une résolution élevée au détriment du SNR (Rapport signal sur bruit). La tête du patient n'est pas fixée.

Les images de la base de données Toshiba sont acquises avec un SPECT 3-têtes (GCA-9300A/UI Toshiba Medical Systems AG, Oetwil am See, Switzerland) équipé d'un fan-beam et d'un collimateur LEHR. La tête du patient est systématiquement fixée dans une têtère afin de réduire les artefacts liés au mouvement. Une méthode « Step-and-shoot » de 30 minutes avec 60 projections sur 360°. La reconstruction est effectuée en retro projection filtrée utilisant un filtre « Shepp and Logan ». Une correction d'atténuation uniforme Chang de 0.15 cm² de coefficient est ensuite utilisée afin de compenser l'atténuation des photons.

Les images de la base de données PPMI sont évaluées avec leur pré-normalisation, les images sont donc de qualité acceptable en tous points pour le diagnostic visuel. Une reconstruction itérative de type HOSEM (Hermes ordered-subset expectation maximization) est d'abord appliquée, suivie d'une correction d'atténuation hélicoïdale et d'une correction d'atténuation « Chang 0 ».

Une comparaison visuelle des histogrammes des 3 bases est effectuée afin d'apprécier leurs différences en répartition des intensités illustrée dans la Figure 31:

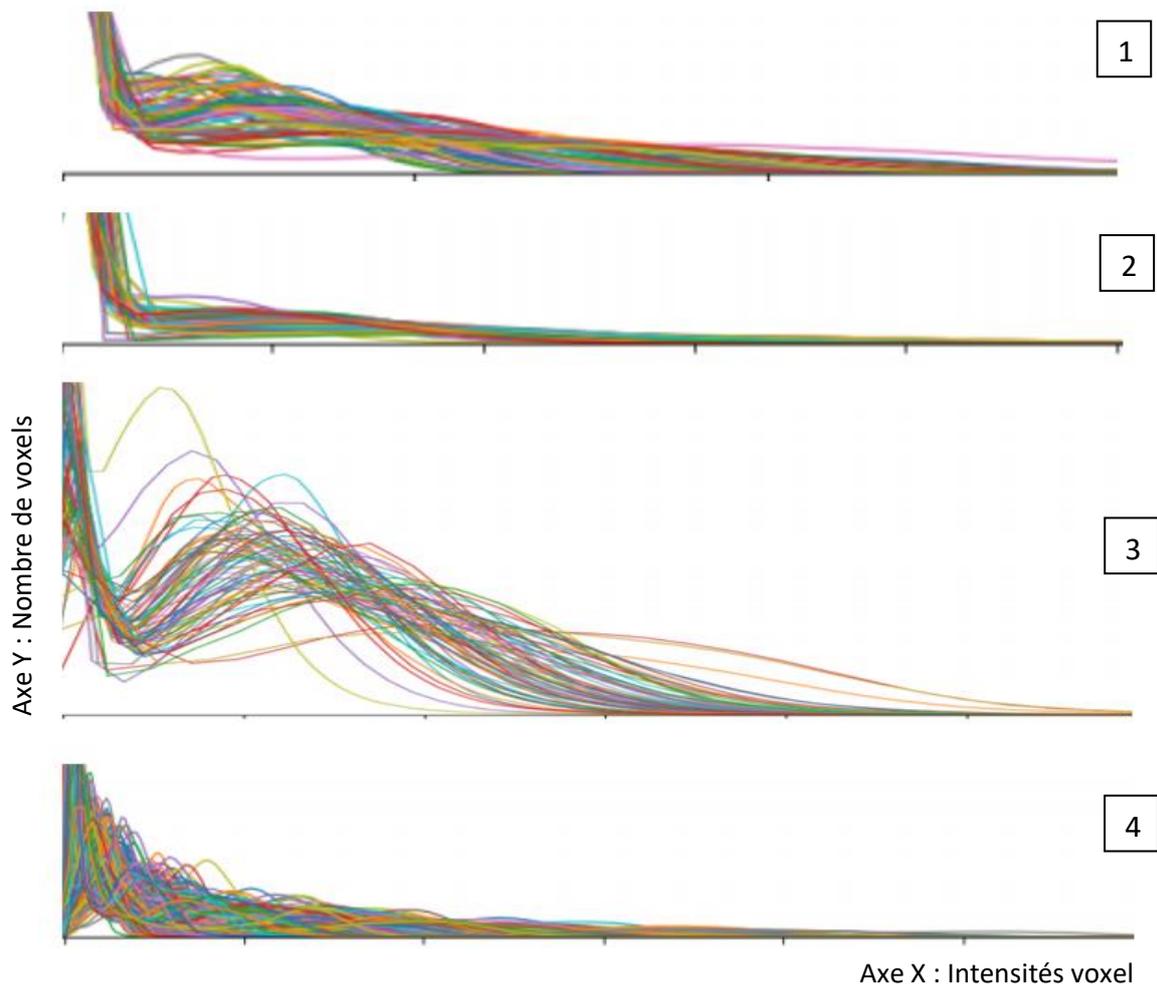


Figure 31: Histogrammes des intensités des lots d'images de chaque base de données, respectivement Toshiba (1), Symbia (2), Symbia normalisée (3) et PPMI (4). Le bruit de fond de l'image est retiré afin d'avoir une répartition claire des courbes

Les graphiques de la Figure 31 montrent que les images des bases de données Toshiba (1) et Symbia (2), ne comportant aucunes normalisations, sont plutôt « plates ». À la suite de la normalisation de la base de données Symbia (2 à 3), les courbes d'intensité se sont « caractérisées » en formant une cloche et en se distinguant plus aisément les unes des autres et ressemblent davantage à celles que l'on voit sur les courbes de la base PPMI (4). Le premier « pic » dans ces bases est plus marqué et pourrait permettre à l'IA de mieux catégoriser ces images.

5.5 Evaluation quantitatives des bases de données

Afin de comparer les performances de l'IA, nous évaluons de façon systématique les différentes mesures de performances diagnostic : exactitude, sensibilité, spécificité et AUC (cf. section Mesure quantitative 1.6.1).

Les différents tests et calculs sont effectués par le programme Stata 16 (*Stata | New in Stata*, 2021). Les résultats de l'exactitude sont comparés statistiquement avec le test non

paramétrique de Mc-Nemar tel qu'utilisé par Choi et al. (2017) pour les comparaisons intrabase (Symbia sur Symbia par exemple), étant des données dépendantes. Les comparaisons statistiques des tests interbase (PPMI sur Symbia par exemple), étant des données indépendantes, sont testées avec un test de proportion. La valeur P de ce test doit être inférieure à 0.05 ($p < 0.05$) afin de valider de la significativité statistique. De plus, un intervalle de confiance selon Wald est systématiquement calculé.

5.6 Standardisation des données

Afin de s'assurer que l'IA concentre son attention sur la pathologie et non les différences dues à l'acquisition de l'imagerie, une normalisation des données est appliquée.

Base de données « Symbia » : Les données non normalisées sont redimensionnées afin d'être à la même taille (1). Ensuite, une normalisation spatiale a été appliquée aux données Symbia (2) afin d'en étudier l'impact sur les performances de l'IA.

Cette normalisation, présentée dans la Figure 32, est effectuée avec le logiciel « Statistical Parametric Mapping » (SPM) version 12 selon l'approche « Old Normalise » qui consiste à recaler nos images SPECT à une image de référence (« 123I-FP-CIT ») de façon affine puis non-rigide (*SPM12 Software - Statistical Parametric Mapping 2020*). Ce template est défini dans l'espace MNI (Montreal Neurologic Institute) et est basé sur le moyennage de 30 cerveaux sans signes de déficit dopaminergiques nigrostriatal obtenu avec une autre base de données (cf. Figure 25). Dans la suite, cette approche sera nommée « SPM ».

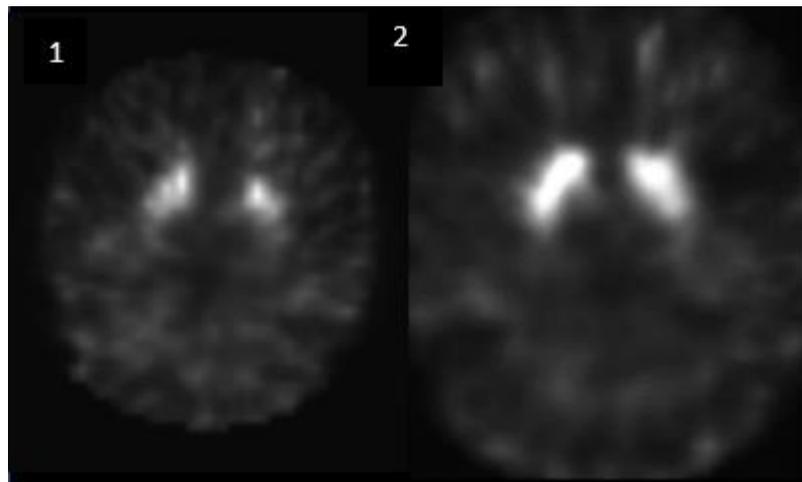


Figure 32 : Images comparatives de la même imagerie DatSCAN provenant de la base de données Symbia respectivement sans et avec normalisation spatiale utilisant SPM12.

Base de données « Toshiba » : Les images provenant du SPECT Toshiba étant déjà normalisées par l'acquisition particulière de cette Gamma caméra SPECT à 3 têtes, aucune standardisation de pré-traitement n'a été jugée nécessaire.

Une dernière normalisation de lissage des images est expérimentée sur les données Symbia et Toshiba (PPMI étant déjà lissées). Un filtre gaussien de 6.0mm relatif à la largeur à mi-hauteur de la Gaussienne (Full Width at Half Maximum (FWHM)) est appliqué afin de réduire le bruit de fond des images comme décrit par Wisniewski, et al. (2013) dans le protocole de processing d'image pour le calcul du « Striatal Binding Ratio » (SBR).

Base de données « PPMI » : Les images étant issues de PPMI ont déjà été standardisées selon le protocole mentionné. Un filtre Gaussien 3D de 6.0mm, puis une normalisation spatiale ont été appliqués aux images afin de les normaliser en intensité et dans l'espace.

Afin d'augmenter le volume des données des bases les plus petites et de conserver une certaine variabilité dans les données, une augmentation géométrique est testée :

Augmentation Géométrique : création de nouvelles données artificielles sur la base d'une déformation rigide des données originales (translations 3D jusqu'à 10mm, inversion gauche/droite et rotation jusqu'à 2° appliquées de façon aléatoire).

Afin de comparer les performances de l'IA dans des cas équivalents de volumes de données, un balancing est utilisé :

Balancing de Groupe : méthode de duplication de données afin d'avoir le même nombre de données entre 2 bases de données par une approche d'« Upsampling ».

Le balancing, en français « rééquilibrage » expliqué à la section 1.5.3 Balancing, décrit généralement une méthode entre deux « classes » telles que Contrôle vs. Pathologique. Le balancing de « groupe » utilisé dans notre méthode rééquilibre entre deux bases de données le nombre d'images de celles-ci afin qu'il n'y ait pas de différences entre elles. L'approche par Upsampling augmente le nombre d'images de la base ayant le moins d'images.

5.7 Scénarios des tests

Différents tests ont été conduits afin d'évaluer les performances de l'IA par rapport aux stratégies adoptées. Ces tests se répartissent dans différentes catégories.

Tests « X/X » :

Une IA entraînée sur une base des données « X » est appliquée sur des données issues de la même base de données X (les images utilisées pour l'entraînement étant différentes de celles utilisées pour les tests).

La première série de tests basées sur la base de données PPMI servira de référence aux autres études, étant donné la taille conséquente de cette base. Ensuite des tests sur la base de données clinique Symbia exploreront l'utilisation de la normalisation des données et de

l'application d'augmentation géométrique. Finalement, les données Toshiba seront testées afin de valider leur pertinence.

Tests « X/Y » :

L'IA entraînée sur la base de données « Y » est testée sur des données de la base « X ». Ce test permet d'évaluer les performances de l'IA à faire des prédictions sur une base de données différente de celle utilisée pour entraîner son modèle.

Tests « X/XY » et tests « X/XYZ » :

La fusion de différentes bases de données (« X » avec « Y », ou « X » avec « Y » et « Z ») est appliquée au niveau de l'entraînement en regroupant les images d'entraînement des bases respectives. La méthode d'équilibrage entre groupes est alors considérée, chaque groupe désignant une base de données différente. L'IA entraînée sur ce mélange de bases est alors testée sur une base de données « X », l'idée étant d'investiguer l'apport de la combinaison de bases dans les performances de l'IA, par exemple dans un but de transférabilité de l'IA d'un centre à un autre.

Tests « X=>Y/Y » :

L'approche de transfert d'apprentissage (transfer learning) est appliquée en utilisant les paramètres du réseau obtenus après un entraînement sur la base « X » comme **initialisation** des paramètres du réseau entraîné sur la base de données « Y ». L'initialisation aléatoire décrite en Section 5.3 n'est donc pas utilisée. Dans nos tests, les poids obtenus à la suite de l'entraînement de la base PPMI augmentée sont transférés.

La majorité des entraînements proposés sont testés avec la base de données Symbia, celle-ci étant la base de données clinique utilisée actuellement et comportant le plus d'intérêt. D'autres tests sont cependant proposés afin de valider et conserver les meilleures stratégies.

Chaque test est présenté dans le Tableau 22 de l'Annexe 10.3.

Une analyse des erreurs au sein de chaque test permettra d'identifier de potentielles erreurs systématiques. Une évaluation qualitative à l'aide de l'approche CAM expliquée dans la section 1.6.2.1 CAM effectuée sur chacune de ces images afin d'identifier si les erreurs proviennent de la détection de l'IA, ou de l'ambivalence de diagnostic des images elles-mêmes. Les images seront donc également analysées par la Dresse Garibotto Cheffe de Service de Médecine Nucléaire du HUG afin d'avoir une notion concernant le diagnostic clinique de ces images. Afin de conserver une méthode rigoureuse, l'IA n'étant pas entraînée avec les bindings ratio, le médecin non plus et devra donner son diagnostic par l'analyse de l'image seule.

6 Résultats

Les résultats quantitatifs sont organisés de la façon suivante comme présentés dans la Figure 33 :

1. Evaluation des données issues de la même base d'entraînement (Section 6.1.1) : dans cette section, les 3 bases de données sont entraînées individuellement et testées sur leurs propres données. C'est là que les normalisations et augmentations diverses sont évaluées et comparées.
2. Evaluation de données issues d'une autre base d'entraînement (Section 6.1.2) : dans cette section, des tests sont effectués sur des images issues de bases de données différentes de celles utilisées pour l'entraînement. Les résultats de ces tests amèneront un éclaircissement concernant la capacité de l'IA à détecter la Maladie de Parkinson entraînée sur une base et testée sur une autre.
3. Evaluation de données partiellement représentées dans les données d'entraînement (Section 6.1.3) : dans cette section, les bases de données sont fusionnées et entraînées ensemble, puis testées sur les bases individuellement.

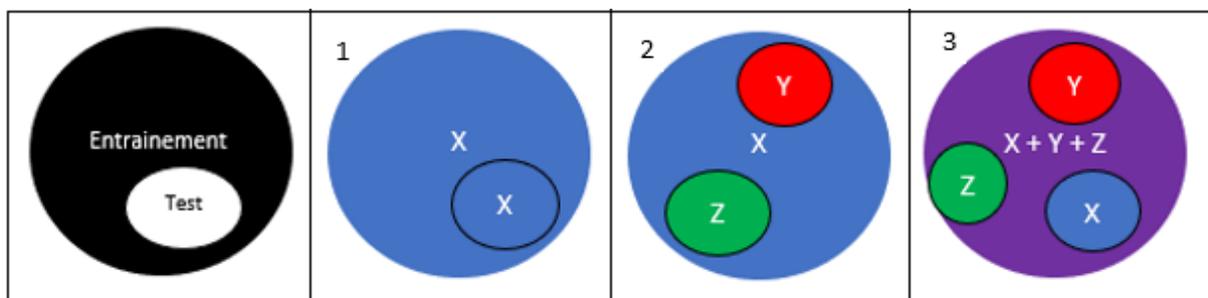


Figure 33: Représentation des différents types d'entraînements et tests présents dans les résultats, 1) Evaluation des données issues de la même base d'entraînement, 2) Evaluation de données issues d'une autre base d'entraînement et 3) Evaluation de données partiellement représentées dans les données d'entraînement.

6.1 Evaluation quantitative des prédictions

Afin d'assurer une meilleure compréhension des résultats, un code est instauré aux tableaux de résultats et présentés dans le Tableau 8 :

Bases de données			Caractéristiques complémentaires			
Symbia	Toshiba	PPMI	SPM12	Lissage	Entraînement avec augment. Géométrique	Transfer Learning
Sy	T	P2	(S12)	(Soo)	(Gé)	(TL)

Tableau 8: Code instauré aux tableaux de résultats.

Les caractéristiques complémentaires présentées entre parenthèses sont liées uniquement à la base de données dont le nom précède la parenthèse. Si aucunes précisions ne sont données, la base de données est donc non altérée : sans normalisation complémentaire à celles évoquées, augmentation ou lissage.

Le Tableau 9 montre quelques exemples de codes avec leur signification :

Données entraînement	Données Test	Description
Sy (S12, Soo)	Sy (S12, Soo)	Entraînement de la base de données Symbia normalisée sous SPM12 et lissées, testée sur la même base
P2 (Gé) + Sy (Gé)	Sy	Entraînement de la base de données PPMI augmentée géométriquement fusionnée avec la base de données Symbia augmentée géométriquement, testée sur la base de données Symbia
T + Sy (S12, Gé)	Sy (S12)	Entraînement de la base de données Toshiba sans normalisation, lissage ou augmentation fusionnée avec les données Symbia normalisées sous SPM et augmentées géométriquement, testé sur les données Symbia normalisées sous SPM12
P2 (Gé)=>Sy (TL, S12)	Sy (S12)	Pré-entraînement de la base de données PPMI augmentée géométriquement transférée sur un entraînement de la base de données Symbia normalisée sous SPM12

Tableau 9 : Exemples de codes pour les résultats avec leur signification.

Pour chaque résultat, le nombre d'images utilisé pour l'entraînement (« N train ») et le test (« N test ») sont rappelés.

6.1.1 Evaluations de données issues de la même base de données

6.1.1.1 Données PPMI

Données entraînement	Données Test	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
P2	P2	94.7 (90.6-98.8)	97.3	90	99.2	552	113
P2 (Gé)	P2	95.6 (91.8-99.4)	98.6	90	99	2208	113

Tableau 10: Résultats des données PPMI.

Les résultats des tests de la base de données PPMI sont présentés dans le Tableau 10. Le meilleur résultat est le test de PPMI entraîné sur PPMI avec augmentation géométrique montrant une exactitude = 95.6% pour une sensibilité de 98.6%, une spécificité de 90% et une

AUC = 99%. Cependant, la différence entre le test avec et sans augmentation géométrique n'est statistiquement pas significatif.

6.1.1.2 Données Toshiba

Données entraînement	Données Test	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
T	T	89.5 (75.7-100)	77.7	100	95.5	79	19
T (Gé)	T	89.5 (75.7-100)	88.8	90	97.8	316	19
T (Soo)	T (Soo)	84.2 (67.8-100)	88.8	80	95.5	79	19

Tableau 11: Résultats des données Toshiba.

Les résultats de la base de données Toshiba sont présentés dans le Tableau 11. Le test présentant les meilleurs résultats est celui de l'entraînement Toshiba testé sur Toshiba avec une exactitude = 89.5% pour une sensibilité de 77.7%, une spécificité de 100% et une AUC = 95.5%. L'augmentation améliore la sensibilité au détriment de la spécificité et n'est donc pas conservée pour les tests suivants sur cette base. Les résultats avec lissage présentent une exactitude moins bonne (sans significativité statistique), une sensibilité acceptable, mais une spécificité réduite et la méthode de lissage ne sera donc également pas conservée pour les données Toshiba.

Le nombre de données de test faible (n=16-19) pour les données Symbia et les données Toshiba dans les prochains tableaux ne permet pas de démontrer une différence statistiquement significative en raison d'une faible puissance de ces derniers et ce point sera développé dans la discussion.

6.1.1.3 Données Symbia

Données entraînement	Données Test	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
Sy	Sy	83.3 (66.1-100)	75	90	91	74	16
Sy (S12)	Sy (S12)	71.4 (59.5-83.2)	62.1	81.5	78.3	74	56
Sy (S12, Gé)	Sy (S12)	82.1 (72.1-92.2)	82.8	81.5	84.5	296	56
Sy (Soo)	Sy (Soo)	72.2 (51.5-92.9)	62.5	80	78.7	74	16
Sy (S12, Soo)	Sy (S12, Soo)	62.5 (49.8-75.2)	100	22.2	82.7	74	56
Sy (S12, Soo, Gé)	Sy (S12, Soo)	51.8 (38.7-64.9)	100	0	71.1	296	56

Tableau 12: Résultats des données Symbia.

Les résultats des tests sur la normalisation de la base de données Symbia sont présentés dans le Tableau 12. Le test montrant le meilleur résultat est celui de Symbia normalisé sous SPM12 géométriquement augmenté avec une exactitude = 82.8% pour une sensibilité de 82.8%, une spécificité de 81.5% et une AUC = 84.5%. L'augmentation de la base de données Symbia (N=74 à N=296) présente une amélioration de l'exactitude ainsi que de la sensibilité. Les tests avec lissage montrent une décroissance de l'exactitude significative ($p < 0.05$) ainsi que de la spécificité allant de 10-30% la spécificité chutant jusqu'à 0%. De ce fait, le lissage n'est pas conservé pour les prochains tests du fait de son inconsistance. Pour rappel, les tests avec N=16 images font partie de l'ancienne série de données Symbia (Tableau 6) qui n'a pas été complétée par les 40 images de la série 2. Ces résultats sont conservés à titre informatifs.

6.1.2 Evaluation de données issues d'une autre base d'entraînement

Données entraînement	Données Test	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
P2	Sy (S12)	76.8 (65.7-87.8)	65.5	88.9	83	552	56
P2 (Gé)	Sy (S12)	73.2 (61.6-84.8)	48.3	100	83.1	2208	56
Sy (S12, Gé)	P2	64.6 (55.8-73.4)	100	0	48.2	296	113

Tableau 13: Résultats des évaluations de données issues d'une autre base d'entraînement.

Les résultats des évaluations de données issues d'une autre base d'entraînement sont présentés dans le Tableau 13. Les entraînements de la base de données PPMI testée sur les données Symbia montrent une exactitude de 76.8% sans augmentation géométrique contre 73.2% avec (non significative), ainsi qu'une sensibilité réduite pour une spécificité améliorée avec ladite augmentation. L'entraînement de Symbia augmentée géométriquement testée sur PPMI montre une exactitude de 64.6% pour une sensibilité à 100%, mais une spécificité de 0%.

6.1.3 Evaluation de données partiellement représentées dans les données d'entraînement

6.1.3.1 Entraînement interbase 2 bases (PPMI et Symbia)

Données entraînement	Données Test	Groupe Balancing	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
P2 + Sy (S12)	Sy (S12)	Non	71.4 (59.6-83.3)	48.3	96.3	84.4	625	56
P2 + Sy (S12)	Sy (S12)	Oui	75 (63.7-86.3)	55.2	96.3	85.6	1102	56
P2 + Sy (S12)	P2	Oui	93.8 (89.4-98.2)	98.6	85	99.3	1102	113

Tableau 14: Résultats des évaluations PPMI et Symbia.

Les résultats des tests sur la base de données PPMI et Symbia sont présentés dans le Tableau 14. Le balancing de groupe montre une amélioration non significative de l'exactitude de 71.4% à 75% ainsi que de la sensibilité (48.3% vs. 55.2%) en passant d'un entraînement de N=625 à 1102. Le balancing de groupe est donc conservé pour les tests interbase suivants. Le test sur PPMI montre une exactitude de 93.8% pour une sensibilité de 98.6% et une spécificité de 85%.

6.1.3.2 Entraînements 2 bases (PPMI et Symbia augmentée)

Données entraînement	Données Test	Groupe Balancing	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
P2 + Sy (S12, Gé)	Sy (S12)	Oui	71.4 (59.6-83.3)	62.1	81.5	80	1102	56
P2 + Sy (S12, Gé)	P2	Oui	94.7 (90.6-98.8)	97.2	90	99.3	1102	113

Tableau 15: Résultats des évaluations de PPMI + Symbia augmentée géométriquement.

Les résultats des tests entraînés sur les bases de données PPMI et Symbia normalisées sous SPM12 augmentée géométriquement avec un balancing de groupe (N=1102) sont présentés dans le Tableau 15. Les résultats des tests sur les bases Symbia normalisés sous SPM (N=56) et PPMI (N=113) montrent, respectivement, une exactitude = 71.4% et 94.7% pour une sensibilité de 62.1% et 97.2% ainsi qu'une spécificité de 81.5% et 90%. A noter que bien que l'augmentation géométrique ait été appliquée à la base Symbia (74 → 296, Tableau 12), le nombre final d'images d'entraînement reste à 1102 sous l'effet du groupe balancing (i.e. Symbia passe de 296 images à 1102/2=552 images pour l'entraînement).

6.1.3.3 Entraînement interbase 2 bases (Symbia augmentée et Toshiba)

Données entraînement	Données Test	Groupe Balancing	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
T + Sy (S12, Gé)	Sy (S12)	Non	73.2 (61.6-84.8)	79.3	66.7	85.9	375	56

Tableau 16: Résultats des données Symbia augmentées géométriquement + Toshiba.

Les résultats des tests entraînés sur les bases de données fusionnées Symbia augmentée géométriquement et Toshiba (N=375) sont présentés dans le Tableau 16. Le test sur la base de données Symbia normalisée sous SPM12 (N=56) montre une exactitude de 73.2% pour une sensibilité de 79.3% et une spécificité de 66.7%.

6.1.3.4 Entraînement interbase 3 bases (PPMI + Symbia + Toshiba)

Données entraînement	Données Test	Groupe Balancing	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
P2 + T + Sy (S12)	Sy (S12)	Oui	82.1 (72.1-92.2)	65.5	100	85.3	1653	56
P2 + T + Sy (S12)	P2	Oui	92.9 (88.2-97.6)	100	80	98.3	1653	113

Tableau 17: Résultats des données PPMI + Symbia + Toshiba.

Les résultats des tests entraînés sur les bases de données fusionnées PPMI, Symbia et Toshiba avec balancing de groupe (n=1653) sont présentés dans le Tableau 17. Les tests Symbia normalisé sous SPM12 ainsi que PPMI montrent respectivement une exactitude de 82.1% et 92.9% pour une sensibilité de 65.5% et 100% et une spécificité de 100% et 80%.

6.1.3.5 Entraînement interbase 3 bases (PPMI + Symbia Augmentée + Toshiba)

Données entraînement	Données Test	Groupe Balancing	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
P2 + T + Sy (S12, Gé)	Sy (S12)	Oui	67.9 (55.6-80.1)	48.3	88.9	80.5	1653	56
P2 + T + Sy (S12, Gé)	P2	Oui	95.6 (91.8-99.4)	97.2	92.5	99	1653	113

Tableau 18: Résultats des données PPMI + Symbia augmentée géométriquement + Toshiba.

Les résultats des tests entraînés sur les bases de données fusionnées PPMI, Symbia augmentée géométriquement et Toshiba avec balancing de groupe (n=1653) sont présentés dans le Tableau 18. Les tests Symbia normalisé sous SPM12 ainsi que PPMI montrent respectivement une exactitude de 67.9% et 95.5% pour une sensibilité de 48.3% et 97.2% et une spécificité de 88.9% et 92.5%.

6.1.3.6 Entraînement par Transfert Learning (PPMI augmenté géométriquement vers Symbia)

Données entraînement	Données Test	Exactitude (IC) (%)	Sensibilité (%)	Spécificité (%)	AUC (%)	N train	N test
P2 (Gé)=>Sy (TL, S12)	Sy (S12)	87.5 (78.8-96.2)	82.8	92.6	94	296	56
P2 (Gé)=>Sy (TL, S12, Gé)	Sy (S12)	78.6 (67.8-89.3)	72.4	85.2	86.2	296	56

Tableau 19 : Résultats des données de Transfert Learning PPMI augmentées géométriquement vers Symbia.

Les résultats des tests entraînés sur les bases de données Symbia avec un « Transfert Learning » à partir de la base PPMI augmentée (n=296) sont présents dans le Tableau 19. Le meilleur résultat est le test sur Symbia sans augmentation géométrique (n=56) avec une

exactitude de 87.5% pour une sensibilité de 82.8% et une spécificité de 92.6%. L'augmentation géométrique de la base Symbia à l'entraînement montre des résultats généraux moins bons.

7 Discussion

À noter que les numéros des tests sont utilisés ici à des fins de compréhension pour le lecteur. Les différents numéros des tests correspondants sont trouvés dans le Tableau 22 de l'Annexe 10.4.

7.1 Performances du réseau par rapport à l'état de l'art

Afin d'évaluer l'efficacité du réseau ResNet utilisé et comparer ses performances aux études préexistantes recensées dans le Tableau 21, la base de données PPMI a été utilisée comme base de référence. Avec un entraînement de 443 images de la PPMI, une exactitude de 94.7%, une sensibilité de 97.3% et une spécificité de 90% (N test=113) ont été obtenues selon le Tableau 10. Aucune normalisation supplémentaire à celle utilisée par Wisniewski et al. (2013) dans la Tableau 21 n'a été ajoutée. En y appliquant de l'augmentation géométrique à l'entraînement, le modèle a obtenu une exactitude de 95.6% et une légère hausse de la sensibilité (98.6%), toujours selon le Tableau 10.

En comparant les résultats obtenus avec d'autres ayant utilisé la base de données PPMI avec du DL, Quan et al. (2019) ont obtenu une exactitude supérieure de 98.5% avec un entraînement de 427 images (Tableau 21). À savoir que les données tirées de la base de données PPMI ne sont pas toutes identiques d'une étude à une autre et ne prennent pas forcément les mêmes critères de sélection en compte. Il est donc important de savoir que leur performance respective ne sont pas aussi comparables qu'elles pourraient l'être. Ayant aussi utilisé de l'augmentation géométrique, ces derniers ont effectué un TL à partir de la base de données ImageNet. Choi et al. (2017) présente une exactitude de 96% sur la base de données PPMI avec de l'augmentation géométrique (Tableau 21), proche des 95.6% présentés ici. Une fourchette entre 95-98.5% d'exactitude est retrouvée dans différentes études utilisant du DL (Ortiz et al., 2016; Quan et al., 2019) et entre 82.2-97.9% pour le ML (Brahim et al., 2017; Oliveira et al., 2018) dans laquelle le résultat de l'exactitude du test de PPMI se trouve.

Take-home message : Le réseau utilisé dans le cadre de cette étude a des performances similaires à l'état de l'art.

7.2 Evaluation intra-base de données

L'évaluation intra-base des données (Figure 34) a permis d'étudier l'impact de certaines techniques de préparation de données de manière séparée en vue de les conserver pour les phases suivantes relatives à l'évaluation interbase de données et le mélange des données

d'entraînement (Section 5.7). Selon la spécificité des trois bases de données, différentes techniques de préparation ont été investiguées.

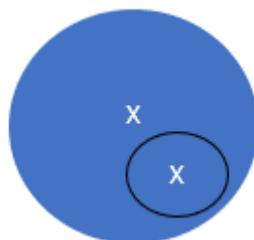


Figure 34 : Evaluation des données issues de la même base d'entraînement

7.2.1 Préparation des données

7.2.1.1 Normalisation spatiale

En analysant les images Symbia, il a été décidé de procéder à une normalisation spatiale. En effet, les images de cette base étaient souvent désaxées et avaient subi des rotations. Sans préparation des images, avec un entraînement de 74 images et une évaluation sur 16 images, une exactitude de 83.3% (IC=66.1-100%) a été obtenue selon le Tableau 12. Après normalisation des données par le logiciel SPM12 expliquée dans la Section 5.6, avec une évaluation de 16 images, une performance de 87.5% (IC=71.3-100%) a été obtenue due à une sensibilité améliorée de 83.3% et une spécificité de 92%. En raison d'un faible échantillon, il n'y a pas eu cependant de différence statistiquement significative. Néanmoins, ces résultats montrent une amélioration de la prédiction du réseau.

Dans un 2^{ème} temps, 40 images de la base de données Symbia normalisée sous SPM12 ont été ajoutées, donnant la nouvelle base de test à 56 images. Les résultats avec cette nouvelle base de test chutent fortement, allant d'une exactitude de 87.5% à 71.4%. Cette chute peut être expliquée par le nombre d'erreurs systématiques introduites avec la nouvelle base de données, étant particulièrement élevées, comme présenté dans le Tableau 20. Ces « erreurs systématiques » sont généralement des images catégorisées par le médecin nucléariste comme étant « peu évidentes » et graduées 1/3 sur une échelle allant de « 1 = peu pathologique » à « 3 = très pathologique ».

	N images	N erreurs systématiques	Taux d'erreurs systématiques
Premier batch	16	2	12.5%
Deuxième Batch	40	9	22.5%
Base finale	56	11	20%

Tableau 20 : Récapitulation du nombre d'erreurs systématiques insérées par la nouvelle base de données de test

Afin d'avoir des tests statistiques plus robustes et d'avoir une puissance plus élevée, ainsi que mieux refléter la réalité clinique, la base de données Symbia avec 56 images, bien que

comportant de nombreuses erreurs systématiques est conservée pour la suite des comparaisons.

Avant d'opter pour la normalisation spatiale de SPM12, une autre approche de normalisation spatiale appelée « Brain Align » avait été testée. Celle-ci consiste à centrer et à réaligner les images de façon rigide sur la base de segmentation de cerveaux effectuées sur les scanner d'atténuation du protocole SPECT. La segmentation est automatisée par une méthode de Deep Learning (Akkus et al., 2018). En entraînant le réseau avec 69 images pour 17 images de test sous Brain Align, une exactitude de 82.3% (IC=64.2-100%) est obtenue. Le nombre d'images est légèrement inférieur due à l'indisponibilité des CT d'atténuation de 4 images.

En raison de l'absence de plusieurs scanners d'atténuation dans l'enrichissement de l'échantillon de 40 images, ainsi que de moins bons résultats obtenus par cette technique, cette normalisation a été finalement écartée.

Concernant la base de données PPMI, celle-ci est déjà normalisée selon le protocole expliqué dans la section 5.6.

Concernant la base de données Toshiba, il a été considéré que ces images étaient de bonne qualité et que la pré-normalisation du mode d'acquisition permettait déjà une normalisation spatiale suffisante. Néanmoins, cette base de données aurait pu tout-de-même bénéficier d'une transformation élastique avec le Template MNI en utilisant tout comme les autres bases de données, le logiciel SPM12. Cela aurait réellement permis de normaliser les trois bases de données entre elles.

Take-home message : La normalisation spatiale est une technique de préparation des images incontournable à utiliser.

7.2.1.2 Filtration Gaussienne

Afin de conserver une ressemblance avec le protocole de normalisation de la base de données PPMI pour les tests interbase futurs, l'application d'un filtre Gaussien de 6.0 mm a été testé sur les images Toshiba et Symbia. Néanmoins, les résultats obtenus dans les deux cas ont démontré une réduction des performances selon le Tableau 11 et le Tableau 12. En effet, l'application d'un lissage sur les images Symbia normalisées sous SPM a provoqué une réduction non significative de l'exactitude d'environ 9% (sensibilité = 100%, spécificité =30%). Le lissage testé sur la base de données Symbia non normalisée a provoqué une baisse non significative de l'exactitude d'environ 11% (Tableau 12), pouvant expliquer qu'un lissage sur des données normalisées et variables en intensité n'est pas bénéfique. Cependant, pour valider cela, il aurait fallu le graphique des intensités de la base de données Symbia normalisée avec lissage.

Concernant la base de données Toshiba, l'application du filtre a produit une réduction non significative de l'exactitude de 89.5% à 84.2% avec un entraînement de 79 images et une évaluation sur 19 images (Tableau 11). Néanmoins, tout comme la base Symbia, une amélioration de la sensibilité a été constatée, pour une réduction de la spécificité. Pour la suite de l'étude, l'utilisation de cette technique de lissage a été exclue.

Malgré les résultats décevants lors de l'utilisation d'un lissage gaussien, différentes études ont elles reporté une amélioration de leurs résultats (Oliveira et al., 2018b; Oliveira & Castelo-Branco, 2015; Rojas et al., 2013; Rumman et al., 2018). Par exemple, Rojas et al. (2013) ont utilisé une filtration avec une variance de 0.9 et un noyau de convolution de taille 9x9 dans le but de diminuer le bruit des images. De même, Oliveira et al. (2018) ont utilisé un filtre gaussien de 16mm à mi-hauteur, alors que Rumman et al. (2018) ont utilisé une filtration adaptative consistant à ne pas éroder les bords ou tout autres structures afin de conserver les petits détails.

7.2.1.3 *Augmentation géométrique*

Cette technique utilisée par Choi et al. (2017), Klyuzhin et al. (2018), Magesh et al. (2020), et Quan et al. (2019) est souvent associée à une amélioration des performances. Cette pratique permettrait de réduire le risque d'overfitting (Magesh et al., 2020; Quan et al., 2019). L'augmentation géométrique a donc été évaluée sur les trois bases de données à disposition comme expliqué dans la section 5.7. En raison d'un volume faible de données de la base de données Symbia et Toshiba, l'apposition d'une augmentation géométrique était particulièrement pertinente. Sur la base de données Toshiba, l'augmentation géométrique n'a pas eu d'impact sur l'exactitude (Tableau 11). Concernant la base de données Symbia, l'augmentation géométrique a permis une amélioration non significative de l'exactitude de 71.4% (IC= 59.5-83.2%) à 82.1% (IC=72.1-92.2%) (Tableau 12). L'augmentation géométrique de la base de données PPMI, largement plus volumineuse a vu son exactitude s'améliorer de 94.7% (552 images) à 95.6% (2208 images) sans différence statistiquement significative (Tableau 10).

Il est intéressant de remarquer qu'à chaque fois que l'exactitude augmentait, elle était reliée à une amélioration de la sensibilité (de 83.3% à 100% avec Symbia et de 97.3% à 98.6% pour PPMI). L'augmentation géométrique semble donc en effet bénéfique de manière générale pour les performances de l'IA lors de l'évaluation intra-base. Ce constat rejoint Quan et al. (2019) qui estiment qu'une augmentation géométrique est efficace pour entraîner des modèles à partir de petite base de données. Néanmoins, la plus-value de l'augmentation géométrique semble diminuer avec la taille de la base de données. Le principe de « state of the art » demande que les installations radiologiques datent d'au maximum 5 ans afin de garder une qualité d'image

correcte et d'éviter les pannes récurrentes (European Society of Radiology, 2014). Ceci implique un changement relativement fréquent des installations et des protocoles d'imageries, compliquant la réutilisation des données à des fins d'entraînement d'IA. Les volumes de données faibles découlant de ce principe rendent l'augmentation géométrique intéressante.

Take-home message : L'évaluation visuelle des données est nécessaire afin d'apprécier les caractéristiques des données mais aussi le niveau de variabilité entre données afin d'appliquer des éventuelles techniques de normalisation.

7.2.2 Evaluation interbase de données

L'objectif de cette seconde phase est d'évaluer les performances du réseau entraîné sur base de données lorsqu'il fait des prédictions sur une autre base de données (Figure 35) selon la section 5.7. Dans le contexte du CAD au DaTSCAN, des connaissances scientifiques à ce sujet manquent. C'est pourquoi, cette méthode a été explorée en utilisant deux bases de données (Symbia et PPMI).

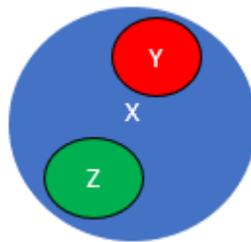


Figure 35 : Evaluation de données issues d'une autre base d'entraînement

Premièrement, la base de données PPMI a été évaluée avec un modèle entraîné sur une version augmentée et normalisée de la base de données Symbia obtenant une performance de 64.6% (IC=55.78-73.42%). En comparaison avec de l'intra-base, cela a engendré une baisse significative de 20%, ce qui signifie que le modèle n'est pas parvenu à généraliser ses performances (Section 6.1.1.1). Pourtant, des techniques de normalisation étaient similaires entre les 2 bases. En effet, ces deux bases ont subi une normalisation spatiale à l'aide de templates définis dans l'espace MNI. Il est possible que le réseau ne soit pas parvenu à extraire les caractéristiques les plus discriminantes. Une augmentation ne permet pas de compenser un manque de variabilité/richeesse des données inhérente à la maladie de Parkinson.

Ensuite, l'opération inverse a été effectuée toujours selon le Tableau 14, la base de données Symbia a été évaluée à partir d'un modèle entraîné sur une version augmentée de PPMI, obtenant une exactitude de 73.2% (IC=61.6-84.8%). En refaisant le même test sans augmenter la base de données PPMI, l'exactitude est restée similaire.

Le test d'une base de données, sans représentation de celle-ci dans l'entraînement (Section 6.1.1) engendre une baisse des performances. Cependant, cette détérioration est plus importante quand l'entraînement en question provient d'une base de données peu volumineuse. Ceci démontre bien que la taille de la base de données, mais également la nature de celle-ci à un effet important sur les performances d'un réseau.

Il est donc possible d'extraire des caractéristiques, bien que pas optimal, d'une base de données et de les transposer dans une autre. Cela a justifié l'intérêt d'explorer cette caractéristique en combinant les bases de données en faisant du mélange de données d'entraînement ou en recourant à l'approche de transfert d'apprentissage (cf. Transfer dans Section 7.3.2).

Ces résultats suggèrent qu'une augmentation géométrique serait bénéfique en évaluation intra-base afin de rendre son modèle plus robuste à la variabilité liée aux conditions géométriques d'acquisitions mais ne permet pas de compenser un manque de variabilité dans une petite base de données. Au contraire, elle pourrait même être contre-productive (Tableau 13).

Take-home message : Evaluer un réseau sur une base de données absente à l'entraînement ne semble pas être bénéfique, voire délétère sur les performances.

7.3 Evaluation interbase mélangée

Cette troisième étape consistait à évaluer les performances du réseau lors d'entraînement bénéficiant d'un mélange de plusieurs bases de données d'entraînement (Figure 36) expliqué à la Section 5.7.

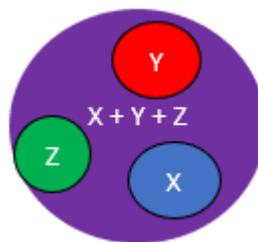


Figure 36 : Evaluation de données partiellement représentées dans les données d'entraînement

Castillo-Barnes et al. (2018), ont notamment suggéré d'utiliser un mélange de bases de données afin d'obtenir une meilleure performance du système de classification. Cette étape a également permis de vérifier l'efficacité du rééquilibrage (5.6) des données parmi les bases de données ayant un échantillon plus faible. Puisque la base de données PPMI est la plus grande, la duplication des images a été effectuée uniquement sur base de données Symbia et Toshiba. L'augmentation géométrique n'a pas été utilisée sur la base PPMI durant l'entraînement combinant plusieurs bases en raison d'un échantillon beaucoup plus grand que

les deux autres bases de données. En effet, l'augmentation géométrique de cette base crée environ 2200 images. En considérant que les bases Symbia et Toshiba ont 70 images, chacune d'entre elle devraient être dupliquées 30 fois pour arriver à une taille égale à la base de données PPMI.

Etant la base de données clinique sur laquelle cette étude se concentre particulièrement, la base de données Symbia a été plus fréquemment évaluée.

7.3.1 Mélange de deux bases de données

Le premier mélange contenait la base de données PPMI et la base de données Symbia normalisée sous SPM12 (Section 6.1.3.1). Le test sur la base de données Symbia normalisée obtient une exactitude de 71.4% (IC=59.5-83.2%) selon le Tableau 14. Le même résultat est trouvé en évaluation intra-base de Symbia (Section 6.1.1.3). Toutefois, la spécificité a été améliorée au détriment de la sensibilité. L'ajout de la base de données PPMI ne semble donc pas avoir perturbé le réseau (test 15), mais n'a toutefois pas pu être profitable. En rééquilibrant la base de données Symbia avec la base de données PPMI, le modèle a non significativement augmenté son exactitude d'environ 4% en évaluant la base de données Symbia (test 16).

En évaluant cette fois-ci la base de données PPMI, le modèle a obtenu une exactitude de 93.8% (IC=89.4-98.2%) sans augmentation géométrique des bases (test 18), en comparaison, avec les résultats intra-base de PPMI augmentée géométriquement, ayant obtenu une exactitude supérieure mais non significative de 94.7% (IC=90.6-98.8%) (Section 6.1.1.1). Le rééquilibrage a donc été bénéfique à la base de données la plus petite lors de son évaluation, mais a réduit légèrement les performances de la base la plus volumineuse. Cela peut venir d'un surentrainement de la base rééquilibrée.

Ayant pu confirmer que le rééquilibrage des données Symbia était bénéfique en évaluant sur des données de même nature, une augmentation géométrique y a été combinée avant l'étape de rééquilibrage (Section 6.1.3.2). En conséquence, le modèle a eu une réduction non significative d'environ 3.5%. La combinaison de ces deux techniques n'a pas permis une amélioration des performances, comparée avec l'utilisation d'une duplication seule (Test 19 vs. 16). Néanmoins, cela a provoqué une augmentation de spécificité au détriment de la sensibilité. En effectuant la même combinaison en évaluant la base de données PPMI, le modèle a légèrement augmenté son exactitude de 93.8 (IC=89.4-98.2%) à 94.7% (IC=90.6-98.8%) (test 21). Il aurait fallu comparer l'augmentation géométrique à la duplication des données lors de l'évaluation de PPMI pour vérifier que cette amélioration des performances soit expliquée par un impact délétère causé par la duplication des images Symbia. En effet, plus l'augmentation géométrique des images est importante, moins la duplication des images sera importante.

Le dernier test de cette phase était de combiner la base de données Symbia avec la base de données Toshiba et d'évaluer uniquement la base de données Symbia (Section 6.1.3.3). Etant deux bases de données à peu près équivalentes en taille, un rééquilibrage par duplication des données n'était pas nécessaire. Comme l'augmentation géométrique des images symbia en intra-base avait un impact positif (test 5), elle a été à nouveau utilisée. En évaluant les images de la base de données Symbia, le modèle a obtenu une exactitude de 73.2% (IC=61.6-84.8%) (test 22). Bien que ce résultat soit meilleur qu'en évaluation intra-base sans augmentation (test 4), il est inférieur à Symbia en intra-base avec une augmentation géométrique (test 5). Aucune des comparaisons montrées sont statistiquement significatives dans ce paragraphe, ce qui signifie probablement que la base de données Toshiba n'ait pas permis d'affiner le modèle. Il est important de noter que la base de données Toshiba est composée uniquement d'images facilement classifiables par le médecin nucléaire. L'ajout de cas « faciles » a peut-être joué en défaveur du modèle. Une autre possibilité serait la nature des images Toshiba qui proviennent d'un système d'acquisition différent de Symbia. Les caractéristiques potentiellement extraites par l'IA des images Toshiba qui sont plutôt de haute résolution en apparence n'ont peut-être pas pu être utilisées lors de l'évaluation des images Symbia qui sont de moins bonne résolution comme expliqué dans la Section 5.4. De plus, les images Toshiba n'ont pas subi une normalisation spatiale avec l'outil SPM12. Par conséquent, ces deux bases de données, bien que visuellement proches, hormis un zoom, comme présenté dans la Figure 37, suites à la normalisation de Symbia, n'étaient en fait pas si proches.

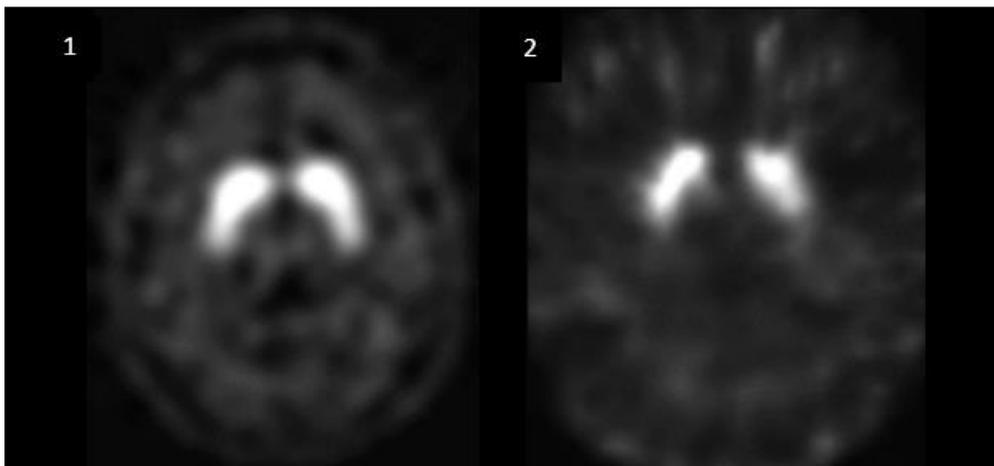


Figure 37 : Comparatif des images Symbia normalisées sous SPM12 (1) et Toshiba (2)

7.3.2 Mélange de trois bases de données

Afin de complexifier le mélange des bases de données, le mélange de trois bases a été testé à l'entraînement dont les résultats se trouvent aux sections 6.1.3.4 et 6.1.3.5. Etant généralement bénéfique, le rééquilibrage des données par duplication a été utilisé pour la totalité des tests qui suivent. Sans augmentation géométrique, selon le Tableau 17 (test 24),

l'entraînement des trois bases de données a augmenté, de manière non significative, l'exactitude du modèle lors de l'évaluation des images Symbia de 75% (IC=63.76-86.3%) à 82.1% (IC=72.1-92.2%) par rapport à un entraînement sans Toshiba (Tableau 14, test 15). De plus, la spécificité est passée de 96.3% à 100%. Il est cependant difficile d'affirmer que ce soient les images Toshiba en-elle-même qui ont amélioré ces performances, ou plutôt la taille de l'échantillon artificiel d'entraînement.

De plus, en évaluant cette-fois ci la base de données PPMI (test 26), l'ajout de la base de données Toshiba n'a pas eu d'impact par rapport au test 18 du Tableau 14. Dans ces conditions, il est possible que combiner d'autres bases de données d'échantillons de petite taille et dupliquées ne soit pas bénéfique.

En reproduisant cette fois-ci le même test et en ajoutant une augmentation géométrique à la base de données Symbia (section 6.1.3.5), le modèle a subi une diminution non significative de l'exactitude d'environ 15% (test 24 vs. test 27). Il est donc possible qu'une simple duplication de la base de données Symbia soit plus efficace qu'une augmentation géométrique lorsque plusieurs bases de données sont mélangées à l'entraînement. Néanmoins, en évaluant cette fois-ci la base de données PPMI, cela a permis d'augmenter de manière non significative l'exactitude du modèle de 2.5% (test 26, Tableau 17 vs. test 29, Tableau 18). Ces résultats peuvent venir du fait que l'augmentation géométrique de la base de données Symbia, ajoute de la variabilité comparée au rééquilibrage et réduit le facteur de duplication de celui-ci dû à la taille de la base augmentée.

Il est important de noter que l'étude des mélanges de plusieurs bases de données d'entraînement, combinées avec du rééquilibrage et/ou de l'augmentation, peut donner des résultats parfois contre-intuitifs. Il est difficile de pouvoir affirmer si les méthodes qui ont été utilisées ont réellement eu impact positif ou si ceci est lié à l'aléa de l'entraînement. Telle qu'utilisée, le mélange de plusieurs bases de données n'a pas permis d'obtenir des performances supérieures à l'entraînement en intra-base.

Les résultats globaux des différents mélanges de bases sont peu clairs et semblent réduire la performance du réseau dans sa tâche. Certains tests semblent bénéfiques alors que tout portait à croire que non et inversement. Il est possible que la complexité d'un tel réseau et des paramètres qu'il faut évaluer rendent l'utilisation d'une fusion de bases compliquée à mettre en place. Même pour une base de données clinique de petite taille, une gestion des données est plus percutante que l'augmentation de son volume par une méthode telle que testée ici.

7.3.3 Transfer Learning

La dernière étape consistait à « mélanger » indirectement les bases de données en les pré-entraînant. Peu de tests ont été effectués, car cette méthode a été développée très tardivement et n'était pas prévue dans la méthodologie initiale. Par déduction, à la suite des différents résultats obtenus a priori, le pré-entraînement de la base de données PPMI augmentée géométriquement semblait être la meilleure solution, suivie d'un entraînement de la base de données Symbia normalisée sous SPM12 et initialisée avec l'entraînement de PPMI, testée sur elle-même. Ce cheminement, présenté dans la section 6.1.3.6, a permis d'obtenir une exactitude de 87.5% (IC=78.8-96.2%) (Tableau 19), étant le meilleur résultat pour les tests de la base de données Symbia toutes normalisations/augmentation confondue. Bien que le TL soit la technique montrant les meilleurs résultats, aucune différence statistiquement significative a été obtenue.

Le Transfer Learning semble donc la meilleure technique à utiliser lors de la combinaison de plusieurs bases de données. Puisqu'une seule base de données est entraînée, il n'y a pas d'équilibrage de bases de données requises. Sachant que l'augmentation géométrique était une technique efficace en intra-base, la réutiliser en augmentant les images Symbia à l'entraînement combinée avec du TL a cette fois-ci joué un impact négatif, mais non significatif selon le Tableau 19.

Take-home message : Le mélange de différentes bases de données à l'entraînement semble donner des résultats prometteurs lorsque le réseau évalue une base de données locale, notamment lorsque le TL est utilisé.

7.3.4 Evaluation des CAM

Une évaluation des CAM de l'entraînement de Symbia augmentée a montré les particularités suivantes : pour une prédiction « contrôle », le réseau a systématiquement proposé une répartition tel que présentée dans la Figure 38, centrée sur l'encéphale, de forme circulaire, comprenant les deux noyaux gris. Pour une prédiction « pathologique » le réseau a systématiquement proposé une répartition telle que présentée dans la Figure 39, large, comprenant une grande partie de l'encéphale, presque rectangulaire.

Les résultats des CAM présentés ici sont uniquement pour des prédictions typiques « contrôles » ou « pathologiques », mais, dans certains cas, quand le réseau donnait une prédiction incorrecte, les images pathologiques avaient une CAM ressemblant à la Figure 38 (contrôle) et inversement. Il semblerait que lorsque le réseau décrit le centre de l'image comme

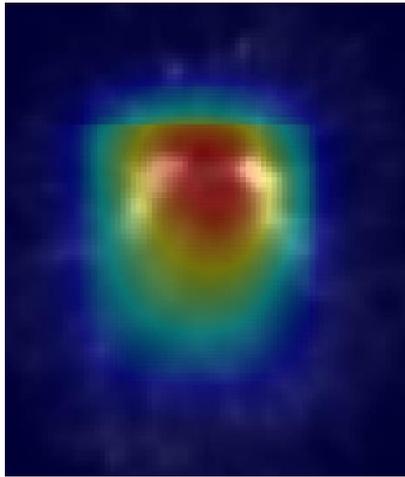


Figure 38: Superposition d'une CAM et d'une imagerie DaTSCAN d'un patient sain type

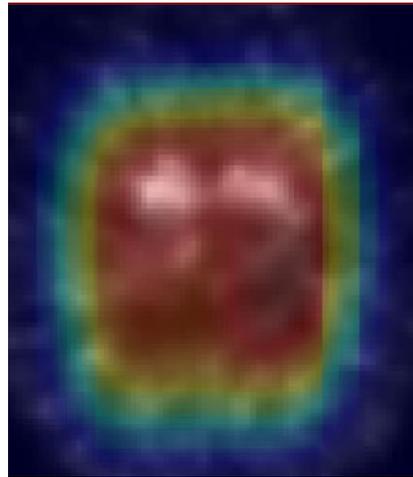


Figure 39: Superposition d'une CAM et d'une imagerie DaTSCAN d'un patient pathologique type

étant saine, il s'arrête là, mais que s'il n'est pas satisfait du résultat, il va élargir le champ de recherche. Une possibilité serait que le réseau ne prédit pas uniquement sur des caractéristiques liées à l'image, mais également à la taille de la zone qu'il doit investiguer. Si le réseau a rapidement trouvé que la taille des noyaux gris ou leur symétrie était suffisante, il prédit le patient comme sain, alors que s'il doit élargir son champ de vision, c'est qu'il prédit automatiquement qu'il est pathologique. À savoir que l'analyse de CAM seule n'est pas suffisante afin de comprendre le comportement du réseau et ne propose pas une vérité absolue. Cependant, une analyse des cartes de saillance expliquée dans la section 1.6.2.2 ou la visualisation des caractéristiques expliquée à la section 1.6.2.3 peuvent encore permettre de complémentariser ces observations sur les CAM.

Afin de pallier le problème de cartographie du réseau, une extraction striatale peut être, à nouveau, une solution afin de rendre celui-ci plus précis lors des prédictions et d'éviter des potentielles erreurs.

Dans des cas particuliers, certaines prédictions pathologiques avaient des répartitions telles que celles de la Figure 41 et de la Figure 40, unilatérales ou antilogiques, pas concentrée sur les noyaux gris. Ces images montrent une investigation singulière du réseau. Les noyaux gris sont presque entièrement écartés de la zone de recherche et seul la matière blanche est conservée dans l'analyse.

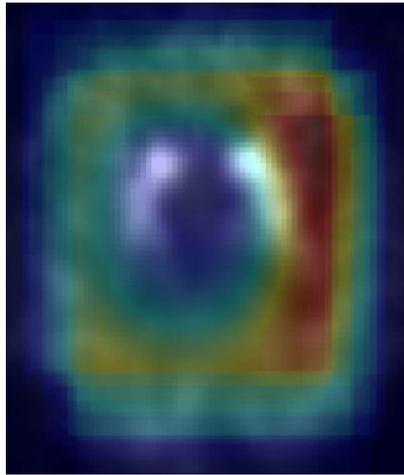


Figure 41: Superposition d'une CAM et d'une imagerie DaTSCAN avec répartition singulière

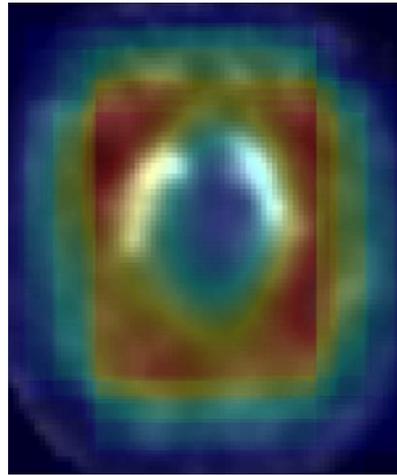


Figure 40: Superposition d'une CAM et d'une imagerie DaTSCAN avec répartition singulière

7.3.5 Évaluation du temps d'opération du réseau

Durant l'entraînement, les images de la base d'entraînement sont utilisées à plusieurs reprises, donc pour des plus petites bases de données que PPMI, il serait donc possible de réduire ce temps en diminuant le nombre d'itérations. Ces temps d'entraînements ne varient pas aussi du fait que les paramètres du réseau tel que la taille des blocs par exemple ne changent pas. L'utilisation d'une augmentation géométrique n'augmente pas le temps de calcul car celle-ci est faite avant l'entraînement. De plus, le pré-entraînement de la base de données PPMI en TL reste enregistré et n'as pas besoin d'être systématiquement entraîné à nouveau. Le temps d'exécution d'une prédiction d'une image par le réseau nécessite environ 160 ms, ce temps particulièrement court facilite donc une implémentation en clinique du réseau.

7.4 Rôles du TRM et recommandations

Comme présenté dans la Section 4, le TRM a un rôle important à jouer dans l'amélioration, l'implémentation et le suivi d'une IA de diagnostic. Ses connaissances en protocoles d'acquisition en imagerie médicale et leur analyse permettent notamment au TRM d'apporter une notion de modification des données afin de faciliter le travail et l'entraînement d'une IA basée sur ces données.

Afin de compléter son panel de connaissances dans le domaine, un développement des bases en mathématiques, statistiques et langages informatiques sont néanmoins nécessaires. Ces points ne sont pas suffisamment approfondis durant les études du Bachelor, mais pourtant, comme explicité par le journal de l'Académie Suisse des Sciences Techniques (Curioni et al., 2019), tous curriculums scientifiques devraient avoir accès à une éducation sur la pensée computationnelle probabiliste. Ils ajoutent que : « La digitalisation ainsi que le traitement d'informations intelligentes affecte presque la totalité des terrains de réflexions humains. » ceci

devant, et plus particulièrement dans le domaine de la radiologie, mener à repenser, plus tôt dans le cursus, l'implication de notre profession dans l'apprentissage d'un CAD ou de façon général de systèmes basés sur l'IA aussi présents aujourd'hui dans l'acquisition.

A travers ce travail, nous avons relevé que la notion ayant le plus péjoré notre compréhension du réseau est celle du langage informatique. Le réseau étant codé en Python et les scripts de lancement en langage shell linux, nous avons dû apprendre à nous y accommoder avec à notre Directeur de mémoire ainsi qu'en autodidacte. Notre implication dans les modifications de l'algorithme était de ce fait faible et ne nous a pas permis de modifier les paramètres tels que nous l'aurions imaginé. Cela nous a aussi freiné dans la confection de la méthodologie, ne connaissant pas les limites de l'algorithme et de ses particularités, beaucoup de notions ne pouvaient être testées et validées.

La récolte des données est particulièrement importante. Une sélection de données trop « faciles » à prédire pour l'IA inclura des biais dans son développement. Généralement, dans des conditions expérimentales, l'échantillon d'images contient toutes les données nécessaires, ce qui facilite le travail de l'IA. En pratique, il arrive que certaines informations ne soient pas disponibles et l'IA doit donc pouvoir être capable de poser un diagnostic avec une quantité d'information pouvant parfois manquer. Il faut donc qu'il ait une certaine capacité d'adaptation et ne pas entrainer une IA avec des informations souvent indisponibles dans des conditions réelles. De plus, selon Ginneken et al. (2011), les constructeurs doivent ouvrir la porte du PACS aux concepteurs de CAD afin de faciliter la récolte de données et y appliquer différentes formes de manipulations. Le TRM ayant accès à ces informations ainsi que les données génétiques, sanguines et autre du patient peut améliorer la performance de son réseau et la rendre plus fine. Il est donc important d'avoir un professionnel capable de faire l'analyse approfondie des données et de leur contenu, ainsi que des données non radiologiques, lui permettant une meilleure compréhension du développement d'un CAD et des possibilités qu'il apporte dans le diagnostic. Sachant que le Médecin Nucléaire utilise le « binding ratio » pour rendre son diagnostic (Section 2.2.3.4.2), l'IA peut certainement bénéficier de cette information telles que Prashanth et al. (2014).

La normalisation des images, quel qu'elle soit, doit être précise et réfléchie si l'on veut que le réseau se concentre sur la pathologie donnée (Section 7.2.1). Dans une optique de standardisations des procédures, le TRM pourrait se charger de la création de protocoles de pré-traitement d'image. Le TRM donc pourrait créer un protocole contenant les différents processus de traitement d'images qu'il estime nécessaire au bon fonctionnement du CAD dans le contexte donné. Le TRM devra se concerter avec le médecin radiologue, expert en physiopathologie, afin de ne pas dénaturer les propriétés distinctives de la pathologie et les

ingénieurs développeurs du système d'IA, connaissant les limites de celle-ci. Une relation collaborative permettra de s'assurer de l'intégralité du réseau et des données qui lui sont fournies. Dans un souci de passage de la théorie à la pratique, le TRM devra donc savoir analyser les points problématiques dans les protocoles cliniques afin de les améliorer et les adapter aux besoins de l'IA. À cela, une compréhension approfondie des caractéristiques de l'image (ex. distribution des intensités) et de ses principes de formation (ex. physique et méthodes de reconstruction) doivent être apprivoisées afin d'améliorer encore ces méthodes de normalisation et de les valider. Le TRM pourra, à la suite d'un approfondissement de ses connaissances en mathématiques et en langage informatique, automatiser les processus importants de traitement d'image, offrant un gain de temps, ainsi qu'une meilleure reproductibilité. De ce fait, l'implémentation en clinique de ces réseaux sera accompagnée par un professionnel avisé et expert du domaine, récoltant lui-même les données nécessaires. Dans une pratique clinique où le TRM sait déjà récolter des informations via l'anamnèse du patient, faire du post-traitement à des fins d'amélioration des données et apprendre à reconnaître les anomalies existantes sur l'image, le TRM a déjà une partie des éléments en main afin de développer et de suivre les IA diagnostiques.

De plus, la participation active du TRM dans le développement des IA de diagnostic, entre autres, pour des examens à faible niveau de complexité selon le Rapport Veille Métier sont l'avenir de la digitalisation (RVM, 2019). Le TRM verra son rôle basculer dans la surveillance des appareils et la résolution de problèmes normalement réservées aux opérateurs des productions industrielles. Le TRM sera un pont entre la pratique clinique et l'industrie, améliorant les algorithmes, participant à la simplification des appareils et développant les informations nécessaires lors des examens. Le TRM pourra de plus apporter ses connaissances et son expertise à l'industrie afin d'améliorer l'affichage des appareils radiologiques, les données enregistrées ou encore communiquer les résultats de l'IA ainsi que son interprétation (RVM, 2019). En addition, le rapport propose une intégration du TRM dans des laboratoires d'essai et de certification d'IA, ainsi qu'une nomination de TRM référents dans les grandes institutions sanitaires. Cette position permettra au TRM d'analyser, entre autres, les pratiques d'utilisation du CAD et de la valeur ajoutée des soins spécifiques à l'imagerie médicale.

Comme exprimé par Petrick et al. (2013), la capacité analytique d'images médicales du TRM lui fournit également la capacité d'entreprendre des actions de contrôle qualité. Lorsque le réseau semble donner des résultats contre-intuitifs, l'analyse visuelle des prédictions telles que les CAM (Section 2.2.3.4.1) est un outil qu'il est intéressant d'utiliser afin d'avoir une information visuelle sur la performance de l'IA. Petrick et al. (2013) proposent entre autres de

valider que le CAD regarde bien la zone d'intérêt et ne produit pas sa prédiction sur des analyses aléatoires au sein de l'image.

Concernant les indicateurs cliniques, certains d'entre eux peuvent primer sur d'autres (sensibilité, spécificité, aire sous la courbe ROC). Dans notre situation, selon Dr. Garibotto, la spécificité était plus importante que la sensibilité. Il était selon elle, tout-à-fait acceptable de valider le CAD pour la Maladie de Parkinson avec une très bonne spécificité et une sensibilité moins bonne. Cela dépend des décisions thérapeutiques pouvant être prises en fonction de la pathologie, de ses tenants et aboutissants. Le TRM doit donc pouvoir faire le lien avec des connaissances approfondies en statistiques pour savoir quels sont les indicateurs cliniques les plus importants selon la nature de l'examen.

Afin de s'assurer de la généralisation de ces recommandations, une étude qualitative portant sur les connaissances manquantes des TRM autour des professionnels travaillant avec les CAD apporterait une vision complémentaire plus juste du sujet.

Take-home message : Afin de faire bénéficier de son expertise à une équipe de recherche en tant que TRM, ce dernier requiert un certain nombre de prérequis sous-jacents à l'Intelligence Artificielle.

7.5 Limites et perspectives de recherches

7.5.1 Volume de données à prendre en compte

La première limite concerne le nombre des différents paramètres, mélanges et méthodes utilisées. Plusieurs tests supplémentaires auraient été nécessaires afin de compléter les résultats et valider certains paramètres ou méthodes utilisées. Cependant, il est concevable que tester toutes les possibilités systématiquement aurait mené à un nombre d'évaluations trop significatif vu le temps disponible. Par exemple, le rééquilibrage a rapidement été validé sans approfondissement particulier de la méthode, ceci ayant pu amener des biais dans l'analyse des méthodes suivant.

7.5.2 Extraction striatale

Dans notre étude, la normalisation avec SPM12 de la base de données clinique Symbia a réduit le volume extra-cérébral sans pour autant totalement le supprimer. On peut imaginer qu'il aurait fallu supprimer cette région qui comporte des voxels cliniquement inexploitable. En allant plus loin, des études antérieures ont notamment procédé à une suppression des régions extra-striatales (Ortiz et al., 2016; Pianpanit et al., 2017; Rumman et al., 2018). Avec des exactitudes respectives de 95%, 96.9% et 94% lors de l'évaluation de PPMI, cette technique aurait pu être utilisée afin de permettre au modèle de se focaliser uniquement sur cette région d'intérêt. D'autres études ont utilisé une technique consistant à distinguer les

régions striatales de chaque hémisphère des patients dans le processus de classification (Augimeri et al., 2016; Segovia et al., 2012; Towey et al., 2011). Si seulement un côté était considéré comme pathologique, alors le cas était considéré comme pathologique. Etant utilisée uniquement sur des bases de données locales (Tableau 21), cette technique mérite d'être étudiée sur la base de données PPMI dans le futur. L'avis d'un médecin nucléaire à ce sujet est également crucial.

7.5.3 Normalisation d'intensité

Comme présenté dans le Tableau 21, certaines études ont procédé à des techniques de normalisation d'intensité. Il a été constaté que la base de données Symbia, avait une variabilité importante d'intensité d'une image à une autre, et plus particulièrement après normalisation sous SPM12 comme présenté à la Figure 31. Une technique de normalisation d'intensité basée sur la distribution alpha-stable a été envisagée comme suggéré par deux études (Augimeri et al., 2016a; Castillo-Barnes et al., 2020). Néanmoins, cette approche nécessitait la segmentation des régions striatales pour être appliquée correctement, ce qui n'a pas pu être réalisé aux vues du temps imparti et de sa complexité. Cette approche a donc été abandonnée.

7.5.4 Lissage des images

Les résultats obtenus à l'aide d'une filtration gaussienne ont démontré des résultats mitigés pouvant s'expliquer par une technique de lissage trop simple et un paramétrage inadéquat du niveau de lissage. Il serait intéressant, pour des études futures, de tester différentes techniques de lissage. Il serait ainsi possible d'apprécier la variation de performances de classification en fonction du degré de filtration des images ainsi que du degré de complexité. Par exemple, la technique de filtration adaptative utilisée par Rumman et al. (2018), brièvement décrite dans la section 7.2.1.2 pourrait être comparée.

7.5.5 Augmentation des données

L'augmentation géométrique utilisée est basée sur une méthode relativement simple, composée uniquement de translations, de rotations et de symétries. Toutefois, il existe d'autres techniques d'augmentation comme la variation des échelles d'intensité ou la variation du niveau du floutage selon la section 1.5.2. Dans une étude future, il serait intéressant d'utiliser d'autres techniques d'augmentation. Lors du mélange de base de données issues de systèmes d'acquisition différents, faire varier la résolution des images et le niveau des échelles d'intensité permettrait de s'affranchir de cette variabilité dans les données pouvant être néfaste pour un modèle. Ceci pourrait expliquer le manque d'efficacité de l'augmentation géométrique lors du mélange de plusieurs bases de données durant cette étude.

7.5.6 Balancing de classe

Parmi les bases de données à disposition, en intra-base, bien qu'il y eût une surreprésentation dans la parité des images pathologique par rapport aux images contrôles, aucune technique n'a été utilisée pour s'en affranchir. Demeurant une question ouverte dans le monde scientifique sur la meilleure méthode à utiliser, notre étude ne s'est pas penchée sur cette problématique. Selon les différentes méthodes qui ont été proposées dans la section 1.5.3, il serait intéressant d'évaluer ces techniques comme celle de Tomek (Batista et al., 2004) sur la variation des performances en termes de sensibilité et de spécificité.

7.5.7 Balancing des bases de données

Pour s'affranchir des différences de tailles entre les bases de données mélangées à l'entraînement, nous avons décidé de procéder à un équilibrage par duplication des bases de données de taille d'échantillon inférieures. Cependant, l'augmentation de la taille d'échantillon en dupliquant des données peut s'avérer être critiquable. En effet, durant l'entraînement, le réseau ne va pas découvrir de nouvelles images lui permettant d'affiner son modèle, mais uniquement le pousser à augmenter sa concentration sur la plus petite base de données. Des techniques d'équilibrage plus avancées devraient être étudiées dans le futur. Par exemple, un équilibrage des données grâce à des méthodes d'augmentation ne se limitant pas à de simple duplication pourraient être testées. Bien qu'il soit avéré qu'un réseau entraîné fonctionne mieux avec une base d'entraînement élevée, peut-être qu'une solution serait de faire un nivellement par le bas (i.e. du « downsampling ») en réduisant la taille des bases de données élevées pour correspondre à celles des bases les plus petites. Afin de conserver une richesse d'information dans la base de données, les cas jugés comme trop faciles à classer pourraient être supprimés. Néanmoins, ce genre de décision dépend du contexte d'utilisation du système CAD.

En outre, bien que de nombreuses techniques de normalisation des données pourraient être utilisées, il faut garder à l'esprit que c'est un processus chronophage s'il ne peut pas être entièrement automatisé. L'idéal serait de trouver un compromis entre la charge de préparation des données et des performances optimales du modèle. Un champ d'étude à explorer serait de rendre un modèle plus robuste à la variabilité dans les données grâce à des techniques d'augmentation plus complexes.

7.5.8 Utilisation de données cliniques

En plus d'informations cliniques, le médecin nucléariste utilise aussi le Binding Ratio en tant qu'analyse semi-quantitative (Section 2.2.3.4.2). Il aurait donc pu être envisageable d'utiliser aussi cette information dans l'entraînement du réseau. Néanmoins, comme cette information se calcule sur la base des images et que l'entraînement en DL est souvent semi-supervisé,

l'IA devrait par elle-même découvrir cette caractéristique. En effet, le principe du DL est de laisser le système choisir quelles sont les caractéristiques les plus pertinentes.

Dans le but d'améliorer les performances de diagnostic du système CAD de cette étude, des informations cliniques indépendantes des données images auraient pu être ajoutées telles que l'âge. Klyuzhin et al. (2018), ont notamment utilisé l'âge en tant que covariable et ils ont pu observer une augmentation de l'AUC de leur système CAD d'environ 5%. En effet, selon Correira et al. (2016), le niveau de captation spécifique du DaTSCAN tend à diminuer naturellement avec l'âge. Une autre piste à étudier serait d'intégrer au système CAD des données cliniques spécifiant les types de symptômes relevés chez le patient. Castillo-Barnes et al. (2018) ont d'ailleurs utilisé des biomarqueurs disponibles sur la base de données PPMI dans le système de classification. Ils ont par exemple utilisé des tests de plasma sanguin, de liquides céphalo-rachidiens ou des échantillons d'acides ribonucléique (ARN). En cas de disponibilité d'autres systèmes d'imagerie comme du CT ou de l'IRM, ces données auraient pu être utilisées par le réseau afin d'obtenir des informations morphologiques et fonctionnelles (Darcourt et al., 2010). Néanmoins, l'intégration de toutes ces informations dans un système d'IA peut être une tâche difficile. De plus, une variabilité dans le parcours du patient dans le processus de diagnostic de la maladie peut parfois limiter la disponibilité de ces informations. Cela encourage le développement de bases de données telles que PPMI mais également des bases de données locales, qui devraient inclure des informations ne se limitant pas qu'à la radiologie. Il faudrait aussi augmenter l'accessibilité de ces bases de données à des fins de recherches, tout en encourageant la standardisation des processus de prise en charge des patients au niveau régional et international.

7.5.9 Transfer Learning

L'utilisation du TL a permis d'obtenir de très bons résultats dans cette étude. A notre connaissance, cette étude est la toute première à avoir utilisé la base de données PPMI en pré-entraînement afin de faire du Transfer Learning dans le cadre d'un CAD pour la maladie de Parkinson. Un des avantages de cette technique, lors de l'utilisation de plusieurs bases de données, est de s'affranchir du problème de différence de taille entre elles. En effet, nous avons pu utiliser cette technique sans nous préoccuper de la grande différence de taille entre Symbia et PPMI, tout en ayant d'excellents résultats. De plus, la base de données PPMI a été augmentée géométriquement ce qui a accentué cette différence de taille. Ces performances justifient que des études s'intéressant au DaTSCAN utilisent cette technique lorsqu'ils souhaitent recourir à plus d'une base de données d'entraînement. Selon la section 3.6, du TL a été effectuée avec des bases de données n'ayant pas de lien avec la nature des données d'entraînement. Par exemple, la base de données ImageNet est composée d'images de la vie de tous les jours. Pourtant, les résultats qui ont été obtenus sont positifs. Il serait

intéressant d'utiliser des images pas forcément basées sur du DaTSCAN en guise d'en pré-entraînement, ayant des caractéristiques pouvant rendre plus attentif le réseau à la nature des structures à analyser et à classer. Il serait par exemple possible de pré-entraîner le réseau afin qu'il puisse comprendre que la région à analyser est composée de deux structures distinctes qui sont le striatum de l'hémisphère gauche et le striatum de l'hémisphère droite.

8 Conclusion

La tendance actuelle dans le diagnostic clinique de la Maladie de Parkinson au DaTSCAN présente des études avec des bases de données petites à l'exception de celles utilisant la base de données PPMI. La préparation des données est connue pour être une étape importante au bon fonctionnement d'un réseau et de ses performances. Avec de petites bases, une augmentation géométrique aide à augmenter la variabilité des données sans avoir la limitation d'acquérir plus d'images de patients. Cependant, le mélange de plusieurs bases de données, provenant d'anciens protocoles d'imagerie de l'hôpital ou de protocoles de différents lieux peuvent permettre de créer de plus grandes bases de données en s'affranchissant d'un temps élevé nécessaire pour une collecte d'un nombre d'images suffisant.

Une normalisation de la base de données clinique Symbia sous SPM12 partant d'un recalage élastique sur un template dans l'espace MNI a montré des résultats prometteurs bien que non significatifs. Cependant, le lissage des images n'a pas été bénéfique, alors que la grande majorité des études qui en ont fait usage montraient de bons résultats. Il est donc important de bien évaluer ses données et leurs caractéristiques afin de faciliter la compréhension du réseau à leur analyse. La normalisation des données est donc une étape nécessaire au développement d'un CAD diagnostic.

Après avoir normalisé les données, il faut s'assurer d'un volume suffisant de celles-ci, possible par l'acquisition d'une base de données externe, comme PPMI. Avec la base PPMI et la base Toshiba issue d'un protocole antérieur de l'hôpital, plusieurs mélanges ont été possibles afin de tester la validité de ces fusions et leur pertinence. Cependant, les différents mélanges effectués n'ont pas montré d'amélioration des performances du réseau. Même après avoir normalisé les données afin qu'elles se ressemblent dans l'espace, le réseau n'a pas réussi à généraliser les caractéristiques des différents protocoles. Il est donc clair qu'une normalisation spatiale des données, bien qu'élastique, ne suffit pas à créer une ressemblance suffisante aux yeux du CAD. Dans notre étude, le mélange de données, autre qu'avec du Transfer Learning, n'est donc pas recommandée dans le cadre du diagnostic de la Maladie de Parkinson au DaTSCAN.

Le réseau évalué sur la base de données PPMI seule a cependant démontré des performances élevées, montrant la pertinence de celui-ci et de son efficacité. Pourtant, le « mélange » de données par Transfer Learning à partir de la base PPMI a montré d'excellents résultats. La base de données PPMI, étant accessible au public et d'un volume conséquent, permet un pré-entraînement efficace afin d'améliorer les performances de classification du réseau entraîné sur la base de données locale.

Les machines et protocoles changeant souvent, la réutilisation des données est compliquée et demande du temps à leurs préparation. C'est pour cela que la place du TRM est cruciale par sa connaissance en imagerie médicale et à la qualité de celles-ci. Cependant, afin d'avoir un bagage suffisant, le TRM doit développer ses connaissances en mathématiques, statistiques et informatique pour participer activement à la mise en place de systèmes de CAD et d'IA en général, tout en approfondissant ses compétences en traitement de données et contrôles qualité.

9 Références

- Accuracy (error rate)*. (2019, mai 17). DeepAI. <https://deepai.org/machine-learning-glossary-and-terms/accuracy-error-rate>
- Akkus, Z., Kostandy, P., Philbrick, K. A., & Erickson, B. J. (2018). Extraction of brain tissue from CT head images using fully convolutional neural networks. In E. D. Angelini & B. A. Landman (Éds.), *Medical Imaging 2018 : Image Processing* (p. 71). SPIE. <https://doi.org/10.1117/12.2293423>
- Augimeri, A., Cherubini, A., Cascini, G. L., Galea, D., Caligiuri, M. E., Barbagallo, G., Arabia, G., & Quattrone, A. (2016a). CADA—computer-aided DaTSCAN analysis. *EJNMMI Physics*, 3(1), 4. <https://doi.org/10.1186/s40658-016-0140-9>
- Badjatiya, P., Gupta, S., Gupta, M., & Varma, V. (2017). Deep Learning for Hate Speech Detection in Tweets. *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*, 759-760. <https://doi.org/10.1145/3041021.3054223>
- Baker, B., Gupta, O., Naik, N., & Raskar, R. (2017). Designing Neural Network Architectures using Reinforcement Learning. *arXiv:1611.02167 [cs]*. <http://arxiv.org/abs/1611.02167>
- Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, 6(1), 20-29. <https://doi.org/10.1145/1007730.1007735>
- Bengio, Y. (2012). Deep Learning of Representations for Unsupervised and Transfer Learning. *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 17-36. <http://proceedings.mlr.press/v27/bengio12a.html>
- Bini, S. A. (2018). Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing : What Do These Terms Mean and How Will They Impact Health Care? *The Journal of Arthroplasty*, 33(8), 2358-2361. <https://doi.org/10.1016/j.arth.2018.02.067>
- Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J., & Zieba, K. (2016). End to End Learning for Self-Driving Cars. *arXiv:1604.07316 [cs]*. <http://arxiv.org/abs/1604.07316>
- Brahim, A., Khedher, L., Górriz, J. M., Ramírez, J., Toumi, H., Lespessailles, E., Jennane, R., & Hassouni, M. (2017). A proposed computer-aided diagnosis system for Parkinson's disease classification using 123I-FP-CIT imaging. *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. <https://doi.org/10.1109/ATSIP.2017.8075510>
- Brahim, Abdelbasset, Jennane, R., Riad, R., Janvier, T., Khedher, L., Toumi, H., & Lespessailles, E. (2019). A decision support tool for early detection of knee OsteoArthritis using X-ray imaging and machine learning : Data from the OsteoArthritis Initiative. *Computerized Medical Imaging and Graphics*, 73, 11-18. <https://doi.org/10.1016/j.compmedimag.2019.01.007>
- Brownlee, J. (2017, juillet 13). What is the Difference Between Test and Validation Datasets? *Machine Learning Mastery*. <https://machinelearningmastery.com/difference-test-validation-datasets/>
- Calmant. (2008). *Module 137 : Etudes épidémiologiques* [Webbapps]. <https://webapps.fundp.ac.be/umdb/biostats2017/biostat/modules/module137/module.php>

- Castillo-Barnes, D., Martínez-Murcia, F. J., Ortiz, A., Salas-Gonzalez, D., Ramírez, J., & Górriz, J. M. (2020). Morphological Characterization of Functional Brain Imaging by Isosurface Analysis in Parkinson's Disease. *International Journal of Neural Systems*, 30(09), 2050044. <https://doi.org/10.1142/S0129065720500446>
- Castillo-Barnes, D., Ramírez, J., Segovia, F., Martínez-Murcia, F. J., Salas-Gonzalez, D., & Górriz, J. M. (2018). Robust Ensemble Classification Methodology for I123-Ioflupane SPECT Images and Multiple Heterogeneous Biomarkers in the Diagnosis of Parkinson's Disease. *Frontiers in Neuroinformatics*, 12, 53. <https://doi.org/10.3389/fninf.2018.00053>
- Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdal, M., Turcotte, S., Pal, C. J., Kadoury, S., & Tang, A. (2017). Deep Learning : A Primer for Radiologists. *RadioGraphics*, 37(7), 2113-2131. <https://doi.org/10.1148/rg.2017170077>
- Choi, H., Ha, S., Im, H. J., Paek, S. H., & Lee, D. S. (2017). Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *NeuroImage: Clinical*, 16, 586-594. <https://doi.org/10.1016/j.nicl.2017.09.010>
- Compendium.ch. (s. d.). Consulté 5 janvier 2021, à l'adresse <https://compendium.ch/fr/product/1050185-datscan-123-i-ioflupane-sol-inj>
- Compendium.ch. (2020, janvier 1). Compendium. <https://compendium.ch/fr/product/1050185-datscan-123-i-ioflupane-inj-los/mpro>
- Correia, & Pereira. (2016). *DaTSCAN quantification : Comparative analysis of different processing algorithms*. https://lenitudesmedicalcenter.pt/wp-content/uploads/2016/10/Poster-EANM-2016_DaTSCAN.pdf
- Curioni, A., & Lukas, C. (2019). *Recommandation for an AI Strategy in Switzerland*. satw.
- Currie, G., Hawk, K. E., Rohren, E., Vial, A., & Klein, R. (2019a). Machine Learning and Deep Learning in Medical Imaging : Intelligent Imaging. *Journal of Medical Imaging and Radiation Sciences*, 50(4), 477-487. <https://doi.org/10.1016/j.jmir.2019.09.005>
- Darcourt, J., Booij, J., Tatsch, K., Varrone, A., Vander Borght, T., Kapucu, Ö. L., Någren, K., Nobili, F., Walker, Z., & Van Laere, K. (2010). EANM procedure guidelines for brain neurotransmission SPECT using 123I-labelled dopamine transporter ligands, version 2. *European Journal of Nuclear Medicine and Molecular Imaging*, 37(2), 443-450. <https://doi.org/10.1007/s00259-009-1267-x>
- Data face : Data augmentation. (2020, janvier 6). *Calyps DATA Intelligence*. <https://www.calyps.ch/analyse/data-face-data-augmentation/>
- Delacour, H., Servonnet, A., Perrot, A., Vigezzi, J., & Ramirez, J. (2020). *La courbe ROC (receiver operating characteristic) : Principes et principales applications en biologie clinique*.
- Despois. (2018). *Stop Feeding Garbage To Your Model ! —The 6 biggest mistakes with datasets and how to avoid them*. | Hacker Noon. Hackernoon. <https://hackernoon.com/stop-feeding-garbage-to-your-model-the-6-biggest-mistakes-with-datasets-and-how-to-avoid-them-3cb7532ad3b7>
- Dubois-Mazeyrie. (2017). *Vers un nouveau traitement pour Parkinson ? - Collège de France*. Fondation Collège de France. <https://www.fondation-cdf.fr/2017/09/27/vers-un-nouveau-traitement-pour-parkinson/>

Eger, A., Gaudet-Blavignac, C., & Hammer, A. (2009, juin 26). *La Maladie de Parkinson*. Université de Genève.
http://www.medecine.unige.ch/enseignement/apprentissage/module4/immersion/archives/2008_2009/travaux/09_r_parkinson.pdf

Ekins, S. (2016). The Next Era : Deep Learning in Pharmaceutical Research. *Pharmaceutical Research*, 33(11), 2594-2603. <https://doi.org/10.1007/s11095-016-2029-7>

Eremenko. (2020, février 16). *Neural Network and AI Skills : What Your Business Needs to Know | Udemty for Business*. Udemty Blog. <https://blog.udemy.com/neural-network-and-ai-skills-what-your-business-needs-to-know/>

European Society of Radiology. (2014). Renewal of radiological equipment. *Insights into Imaging*, 5(5), 543-546. <https://doi.org/10.1007/s13244-014-0345-1>

Fawzi, A., Samulowitz, H., Turaga, D., & Frossard, P. (2016). Adaptive data augmentation for image classification. *2016 IEEE International Conference on Image Processing (ICIP)*, 3688-3692. <https://doi.org/10.1109/ICIP.2016.7533048>

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669. <https://doi.org/10.1016/j.ejor.2017.11.054>

Formation personnalisée : Procédure pas à pas | TensorFlow Core. (2020). TensorFlow. https://www.tensorflow.org/tutorials/customization/custom_training_walkthrough

Fujita, H. (2020). AI-based computer-aided diagnosis (AI-CAD) : The latest review to read first. *Radiological Physics and Technology*, 13(1), 6-19. <https://doi.org/10.1007/s12194-019-00552-4>

Fung, V. (2017). *An Overview of ResNet and its Variants | by Vincent Fung | Towards Data Science*. <https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035>

Gagana, B. (2019, octobre 2). *Class Activation maps*. Medium. <https://medium.com/@GaganaB/class-activation-maps-551477720679>

Galdran, A., Alvarez-Gila, A., Meyer, M. I., Saratxaga, C. L., Araújo, T., Garrote, E., Aresta, G., Costa, P., Mendonça, A. M., & Campilho, A. (2017). Data-Driven Color Augmentation Techniques for Deep Skin Image Analysis. *arXiv:1703.03702 [cs]*. <http://arxiv.org/abs/1703.03702>

Garcia-Gomez, Huertas, & Lojo. (2013). *NITRC: 18F-DOPA PET and 123I-FP-CIT (Ioflupane—DaTSCAN) SPECT templates for SPM normalisation. : Tool/Resource Info*. <https://www.nitrc.org/projects/spmtemplates/>

GE Healthcare Unveils Discovery NM 630 SPECT | Medgadget. (2011). *Medgadget*. <https://www.medgadget.com/2011/06/ge-healthcare-unveils-discovery-nm-630-spect.html>

Guide parcours de soins maladie de Parkinson. (2016). *Haute Autorité de Santé*, 86. https://www.has-sante.fr/jcms/c_1242645/fr/guide-parcours-de-soins-maladie-de-parkinson

Hansen, C. (2019, août 5). *Neural Networks : Feedforward and Backpropagation Explained*. Machine Learning From Scratch. <https://mlfromscratch.com/neural-networks-explained/>

He, K., Zhang, X., Ren, S., & Sun, J. (2016a). *Deep Residual Learning for Image Recognition*. 770-778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

- He, K., Zhang, X., Ren, S., & Sun, J. (2016b). Identity Mappings in Deep Residual Networks. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Éds.), *Computer Vision – ECCV 2016* (p. 630-645). Springer International Publishing. https://doi.org/10.1007/978-3-319-46493-0_38
- Huo, Z., Summers, R. M., Paquerault, S., Lo, J., Hoffmeister, J., Armato, S. G., Freedman, M. T., Lin, J., Ben Lo, S.-C., Petrick, N., Sahiner, B., Fryd, D., Yoshida, H., & Chan, H.-P. (2013). Quality assurance and training procedures for computer-aided detection and diagnosis systems in clinical use. *Medical Physics*, *40*(7), 077001. <https://doi.org/10.1118/1.4807642>
- Illán, I. A., Górriz, J. M., Ramírez, J., Segovia, F., Jiménez-Hoyuela, J. M., & Ortega Lozano, S. J. (2012a). Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging : Automatic assistance to PD diagnosis in DaTSCAN imaging. *Medical Physics*, *39*(10), 5971-5980. <https://doi.org/10.1118/1.4742055>
- Introduction au Deep Learning avec MATLAB*. (s. d.). Consulté 4 octobre 2020, à l'adresse <https://ch.mathworks.com/fr/campaigns/offers/deep-learning-with-matlab.html>
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare : Past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230-243. <https://doi.org/10.1136/svn-2017-000101>
- Jiang, K., Lu, J., & Xia, K. (2016). A Novel Algorithm for Imbalance Data Classification Based on Genetic Algorithm Improved SMOTE. *Arabian Journal for Science and Engineering*, *41*(8), 3255-3266. <https://doi.org/10.1007/s13369-016-2179-2>
- Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences. *arXiv:1404.2188 [cs]*. <http://arxiv.org/abs/1404.2188>
- Karlegger. (2013). *Technique en Radiologie Médicale : Champ d'activité et besoins du marché du travail*. Secrétariat d'Etat à la formation, à la recherche et à l'innovation (SEFRI). https://www.astrm.ch/files/Dokumente/Verband/Projekte/rapport_econcept.pdf
- Keller, J. M., Gray, M. R., & Givens, J. A. (1985). A fuzzy K-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-15*(4), 580-585. <https://doi.org/10.1109/TSMC.1985.6313426>
- Klyuzhin, I., Shenkov, N., Rahmim, A., & Sossi, V. (2018). Use of deep convolutional neural networks to predict Parkinson's disease progression from DaTscan SPECT images. *Journal of Nuclear Medicine*, *59*(supplement 1), 29-29.
- Koushik, J. (2016). Understanding Convolutional Neural Networks. *arXiv:1605.09081 [stat]*. <http://arxiv.org/abs/1605.09081>
- Kriminger, E., Príncipe, J. C., & Lakshminarayan, C. (2012). Nearest Neighbor Distributions for imbalanced classification. *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 1-5. <https://doi.org/10.1109/IJCNN.2012.6252718>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25*, 1097-1105.
- Lawrence, S., Giles, C. L., Ah Chung Tsoi, & Back, A. D. (1997). Face recognition : A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, *8*(1), 98-113. <https://doi.org/10.1109/72.554195>

Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., & Quillen, D. (2018). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *The International Journal of Robotics Research*, 37(4-5), 421-436. <https://doi.org/10.1177/0278364917710318>

Lison, P. (2012). *An introduction to machine learning*. <https://docplayer.net/22095982-An-introduction-to-machine-learning.html>

Mabrouk, R., Chikhaoui, B., & Bentabet, L. (2019). Machine Learning Based Classification Using Clinical and DaTSCAN SPECT Imaging Features : A Study on Parkinson's Disease and SWEDD. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2), 170-177. <https://doi.org/10.1109/TRPMS.2018.2877754>

Magesh, P. R., Myloth, R. D., & Tom, R. J. (2020). An Explainable Machine Learning Model for Early Detection of Parkinson's Disease using LIME on DaTSCAN Imagery. *Computers in Biology and Medicine*, 126, 104041. <https://doi.org/10.1016/j.compbimed.2020.104041>

Maglogiannis, I. G. (2007). *Emerging Artificial Intelligence Applications in Computer Engineering : Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*. IOS Press.

Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., Coffey, C. S., Caspell-Garcia, C., Simuni, T., Jennings, D., Tanner, C. M., Trojanowski, J. Q., Shaw, L. M., Seibyl, J., Schuff, N., Singleton, A., Kieburtz, K., Toga, A. W., Mollenhauer, B., Galasko, D., Chahine, L. M., ... Sherer, T. (2018). The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*, 5(12), 1460-1477. <https://doi.org/10.1002/acn3.644>

Martinez-Murcia, F. J., Górriz, J. M., Ramírez, J., Moreno-Caballero, M., Gómez-Río, M., & The Parkinson's Progression Markers Initiative. (2013a). Parametrization of textural patterns in I123-ioflupane imaging for the automatic detection of Parkinsonism : Parametrization of textural patterns in 123I-ioflupane imaging. *Medical Physics*, 41(1), 012502. <https://doi.org/10.1118/1.4845115>

Mikko. (2019, mai 21). Hyperparameter Optimization with Keras. *Medium*. <https://towardsdatascience.com/hyperparameter-optimization-with-keras-b82e6364ca53>

Mikolajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. *2018 International Interdisciplinary PhD Workshop (IIPHDW)*, 117-122. <https://doi.org/10.1109/IIPHDW.2018.8388338>

Narkhede, S. (2019, mai 26). *Understanding AUC - ROC Curve*. *Medium*. <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>

Olah, C., Mordvintsev, A., & Schubert, L. (2017, novembre 7). *Feature Visualization*. *Distill*. <https://distill.pub/2017/feature-visualization>

Oliveira, F. P. M., & Castelo-Branco, M. (2015). Computer-aided diagnosis of Parkinson's disease based on [¹²³ I]FP-CIT SPECT binding potential images, using the voxels-as-features approach and support vector machines. *Journal of Neural Engineering*, 12(2), 026008. <https://doi.org/10.1088/1741-2560/12/2/026008>

Oliveira, F. P. M., Faria, D. B., Costa, D. C., Castelo-Branco, M., & Tavares, J. M. R. S. (2018a). Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [123I]FP-CIT SPECT images. *European Journal of Nuclear Medicine and Molecular Imaging*, 45(6), 1052-1062. <https://doi.org/10.1007/s00259-017-3918-7>

- Ortiz, A., Martínez-Murcia, F. J., García-Tarifa, M. J., Lozano, F., Górriz, J. M., & Ramírez, J. (2016a). Automated Diagnosis of Parkinsonian Syndromes by Deep Sparse Filtering-Based Features. In Y.-W. Chen, S. Tanaka, R. J. Howlett, & L. C. Jain (Éds.), *Innovation in Medicine and Healthcare 2016* (Vol. 60, p. 249-258). Springer International Publishing. https://doi.org/10.1007/978-3-319-39687-3_24
- Overfitting in Machine Learning : What It Is and How to Prevent It.* (2017, septembre 7). EliteDataScience. <https://elitedatascience.com/overfitting-in-machine-learning>
- Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, *86*(4), 554-565. [https://doi.org/10.1016/S0034-4257\(03\)00132-9](https://doi.org/10.1016/S0034-4257(03)00132-9)
- Parkinson Progression Marker Initiative. (2011). The Parkinson Progression Marker Initiative (PPMI). *Progress in Neurobiology*, *95*(4), 629-635. <https://doi.org/10.1016/j.pneurobio.2011.09.005>
- Parkinson's Progression Markers Initiative | Search Criteria. (2020). PPMI. <https://www.ppmi-info.org/access-data-specimens/request-specimens/>
- Petrick, N., Sahiner, B., Armato, S. G., Bert, A., Correale, L., Delsanto, S., Freedman, M. T., Fryd, D., Gur, D., Hadjiiski, L., Huo, Z., Jiang, Y., Morra, L., Paquerault, S., Raykar, V., Samuelson, F., Summers, R. M., Tourassi, G., Yoshida, H., ... Chan, H.-P. (2013). Evaluation of computer-aided detection and diagnosis systems. *Medical Physics*, *40*(8), 087001. <https://doi.org/10.1118/1.4816310>
- Pianpanit, T., Lolak, S., Sawangjai, P., Ditthapron, A., Leelaarporn, P., Marukatat, S., Chuangsuwanich, E., & Wilaiprasitporn, T. (2017). Neural network interpretation of the Parkinson's disease diagnosis from SPECT imaging. *arXiv:1908.11199 [cs, eess, stat]*. <http://arxiv.org/abs/1908.11199>
- Pigrée, G., & Josset, M. (2000). *Etude de la répartition de radionucléides artificiels émetteurs gamma entre une usine de dépollution et son milieu environnant; Cas de la station d'épuration du District du Grand Caen*. Association pour le controle de la radioactivité dans l'OUEST. https://inis.iaea.org/collection/NCLCollectionStore/_Public/40/007/40007584.pdf?r=1&r=1
- Prashanth, R., Dutta Roy, S., Mandal, P. K., & Ghosh, S. (2014). Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging. *Expert Systems with Applications*, *41*(7), 3333-3342. <https://doi.org/10.1016/j.eswa.2013.11.031>
- Prashanth, R., Dutta Roy, S., Mandal, P. K., & Ghosh, S. (2017). High-Accuracy Classification of Parkinson's Disease Through Shape Analysis and Surface Fitting in 123I-Ioflupane SPECT Imaging. *IEEE Journal of Biomedical and Health Informatics*, *21*(3), 794-802. <https://doi.org/10.1109/JBHI.2016.2547901>
- Pratt, H., Coenen, F., & Zheng, Y. (2019). Feature Visualisation of Classification of Diabetic Retinopathy Using a Convolutional Neural Network. *KHD@IJCAI*.
- Quan, J., Xu, L., Xu, R., Tong, T., & Su, J. (2019). DaTscan SPECT Image Classification for Parkinson's Disease. *arXiv:1909.04142 [cs, eess, stat]*. <http://arxiv.org/abs/1909.04142>
- Rajchl. (2021). *DLTK* [Python]. DLTK. <https://github.com/DLTK/DLTK> (Original work published 2017)
- Ren, Y., Zhao, P., Sheng, Y., Yao, D., & Xu, Z. (2017). Robust Softmax Regression for Multi-class Classification with Self-Paced Learning. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2641-2647. <https://doi.org/10.24963/ijcai.2017/368>

- Residual Networks (ResNet)—Deep Learning. (2020, juin 3). *GeeksforGeeks*.
<https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/>
- Rojas, A., Górriz, J. M., Ramírez, J., Illán, I. A., Martínez-Murcia, F. J., Ortiz, A., Gómez Río, M., & Moreno-Caballero, M. (2013). Application of Empirical Mode Decomposition (EMD) on DaTSCAN SPECT images to explore Parkinson Disease. *Expert Systems with Applications*, *40*(7), 2756-2766.
<https://doi.org/10.1016/j.eswa.2012.11.017>
- Romero, A., Radeva, P., & Gatta, C. (2015). Meta-Parameter Free Unsupervised Sparse Feature Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(8), 1716-1722.
<https://doi.org/10.1109/TPAMI.2014.2366129>
- Rumman, M., Tasneem, A. N., Farzana, S., Pavel, M. I., & Alam, M. A. (2018). Early detection of Parkinson's disease using image processing and artificial neural network. *2018 Joint 7th International Conference on Informatics, Electronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, 256-261. <https://doi.org/10.1109/ICIEV.2018.8641081>
- RVM. (2019). Face à la digitalisation de la santé : Quel avenir pour le métier de TRM ? *Ergone*.
- Savaram. (2019). *Skills Need To Become A Machine Learning Engineer*. <https://mindmajix.com/what-skills-needed-for-machine-learning-jobs>
- Schrag, A., Horsfall, L., Walters, K., Noyce, A., & Petersen, I. (2015). Prediagnostic presentations of Parkinson's disease in primary care : A case-control study. *The Lancet Neurology*, *14*(1), 57-64.
[https://doi.org/10.1016/S1474-4422\(14\)70287-X](https://doi.org/10.1016/S1474-4422(14)70287-X)
- Segovia, F., Górriz, J. M., Ramírez, J., Álvarez, I., Jiménez-Hoyuela, J. M., & Ortega, S. J. (2012). Improved Parkinsonism diagnosis using a partial least squares based approach : Improved Parkinsonism diagnosis using a partial least squares based approach. *Medical Physics*, *39*(7Part1), 4395-4403. <https://doi.org/10.1118/1.4730289>
- Sharma. (2017). *Convolutional Neural Networks in Python*.
<https://www.datacamp.com/community/tutorials/convolutional-neural-networks-python>
- Sharma. (2021). *Perceptron in Neural Networks—Artificial Intelligence Examples and Tutorials*.
<http://www.dailyfreecode.com/code/perceptron-neural-networks-3331.aspx>
- Shen, D., Wu, G., & Suk, H.-I. (2017). Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, *19*(1), 221-248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>
- Shimizu, S., Namioka, N., Hirose, D., Kanetaka, H., Hirao, K., Hatanaka, H., Takenoshita, N., Kaneko, Y., Ogawa, Y., Tsugawa, A., Umahara, T., Sakurai, H., & Hanyu, H. (2017). Comparison of diagnostic utility of semi-quantitative analysis for DAT-SPECT for distinguishing DLB from AD. *Journal of the Neurological Sciences*, *377*, 50-54. <https://doi.org/10.1016/j.jns.2017.03.040>
- Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Deep Inside Convolutional Networks : Visualising Image Classification Models and Saliency Maps. *arXiv:1312.6034 [cs]*. <http://arxiv.org/abs/1312.6034>
- Sloane, C., & al., f. (2017). Informing radiography curriculum development : The views of UK radiology service managers concerning the 'fitness for purpose' of recent diagnostic radiography graduates. *Radiography*, *23*, S16-S22. <https://doi.org/10.1016/j.radi.2017.05.013>
- Smilkov, D., Thorat, N., Kim, B., Viégas, F., & Wattenberg, M. (2017). SmoothGrad : Removing noise by adding noise. *arXiv:1706.03825 [cs, stat]*. <http://arxiv.org/abs/1706.03825>

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters : Part 1 -- learning rate, batch size, momentum, and weight decay. *arXiv:1803.09820 [cs, stat]*.
<http://arxiv.org/abs/1803.09820>

SPM12 Software—Statistical Parametric Mapping. (2020).
<https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>

Staff, R. T., Ahearn, T. S., Wilson, K., Counsell, C. E., Taylor, K., Caslake, R., Davidson, J. E., Gemmell, H. G., & Murray, A. D. (2009). Shape analysis of 123I-N- ω -fluoropropyl-2- β -carbomethoxy-3 β -(4-iodophenyl) nortropane single-photon emission computed tomography images in the assessment of patients with parkinsonian syndromes: *Nuclear Medicine Communications*, 30(3), 194-201.
<https://doi.org/10.1097/MNM.0b013e328314b863>

Stata | New in Stata. (2021). <https://www.stata.com/new-in-stata/>

state of flux—Definition. (2021). WorNet-Online. http://www.wordnet-online.com/state_of_flux.shtml

Stoutamire, D. (1991). *Machine Learning, Game Play, and Go*.
<http://citeseerx.ist.psu.edu/viewdoc/versions?doi=10.1.1.44.5254>

Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, 39(1), 43-62.
[https://doi.org/10.1016/S0169-7439\(97\)00061-0](https://doi.org/10.1016/S0169-7439(97)00061-0)

Tagaris, A., Kollias, D., & Stafylopatis, A. (2017). Assessment of Parkinson's Disease Based on Deep Neural Networks. In G. Boracchi, L. Iliadis, C. Jayne, & A. Likas (Éds.), *Engineering Applications of Neural Networks* (Vol. 744, p. 391-403). Springer International Publishing.
https://doi.org/10.1007/978-3-319-65172-9_33

Tang, A., Tam, R., Cadrin-Chênevert, A., Guest, W., Chong, J., Barfett, J., Chepelev, L., Cairns, R., Mitchell, J. R., Cicero, M. D., Poudrette, M. G., Jaremko, J. L., Reinhold, C., Gallix, B., Gray, B., & Geis, R. (2018). Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Canadian Association of Radiologists Journal*, 69(2), 120-135.
<https://doi.org/10.1016/j.carj.2018.02.002>

Tapia, J., & Aravena, C. (2017). Gender Classification from NIR Iris Images Using Deep Learning. In B. Bhanu & A. Kumar (Éds.), *Deep Learning for Biometrics* (p. 219-239). Springer International Publishing. https://doi.org/10.1007/978-3-319-61657-5_9

Taylor, J., & al. (2016a). Clinical Adoption of CAD : Exploration of the Barriers to Translation through an Example Application. *Procedia Computer Science*, 90, 93-98.
<https://doi.org/10.1016/j.procs.2016.07.029>

TensorFlow. (s. d.). TensorFlow. Consulté 5 janvier 2021, à l'adresse
<https://www.tensorflow.org/?hl=fr>

Thirumuruganathan, S., Tang, N., Ouzzani, M., & Doan, A. (2020). *Data Curation with Deep Learning* (Version 1) [Data set]. OpenProceedings.org. <https://doi.org/10.5441/002/EDBT.2020.25>

Thurston, M. (2018). *Artificial intelligence in nuclear medicine : Automated interpretation of Ioflupane-123 DaTScan for Parkinson's disease using deep learning* [Text/html]. 943 words.
<https://doi.org/10.1594/ECR2018/C-0668>

- Towey, D. J., Bain, P. G., & Nijran, K. S. (2011). Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images: *Nuclear Medicine Communications*, 32(8), 699-707.
<https://doi.org/10.1097/MNM.0b013e328347cd09>
- van Ginneken, B., Schaefer-Prokop, C. M., & Prokop, M. (2011a). Computer-aided diagnosis : How to move from the laboratory to the clinic. *Radiology*, 261(3), 719-732.
<https://doi.org/10.1148/radiol.11091710>
- Vishwanathan, S. V. M., & Narasimha Murty, M. (2002). SSVM : A simple SVM algorithm. *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, 3, 2393-2398 vol.3. <https://doi.org/10.1109/IJCNN.2002.1007516>
- Wachinger, C., Reuter, M., & Klein, T. (2018). DeepNAT : Deep convolutional neural network for segmenting neuroanatomy. *NeuroImage*, 170, 434-445.
<https://doi.org/10.1016/j.neuroimage.2017.02.035>
- Wisniewski, G., Seibyl, J., & Marek, K. (2013). *DatScan SPECT Image Processing Methods for Calculation of Striatal Binding Ratio (SBR)*. 1.
- Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in Statistics - Simulation and Computation*, 22(4), 1079-1106.
<https://doi.org/10.1080/03610919308813143>
- Wong, K. C. L., Tee, M., Chen, M., Bluemke, D. A., Summers, R. M., & Yao, J. (2016). Regional infarction identification from cardiac CT images : A computer-aided biomechanical approach. *International Journal of Computer Assisted Radiology and Surgery*, 11(9), 1573-1583.
<https://doi.org/10.1007/s11548-016-1404-5>
- Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, 6(1), 113. <https://doi.org/10.1186/s40537-019-0276-2>
- Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks : An overview and application in radiology. *Insights into Imaging*, 9(4), 611-629.
<https://doi.org/10.1007/s13244-018-0639-9>
- Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning : Progress in machine intelligence for rational drug discovery. *Drug Discovery Today*, 22(11), 1680-1685.
<https://doi.org/10.1016/j.drudis.2017.08.010>
- Zhang, M.-L., Peña, J. M., & Robles, V. (2009). Feature selection for multi-label naive Bayes classification. *Information Sciences*, 179(19), 3218-3229. <https://doi.org/10.1016/j.ins.2009.06.010>
- Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., & Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Medical Image Analysis*, 43, 98-111.
<https://doi.org/10.1016/j.media.2017.10.002>
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). *Learning Deep Features for Discriminative Localization*. 2921-2929.
https://openaccess.thecvf.com/content_cvpr_2016/html/Zhou_Learning_Deep_Features_CVPR_2016_paper.html

10 Annexes

10.1 Tableau récapitulatif des études antérieures

Auteurs	BDD	Méthode d'entraînement	Pre-processing	Classification	N. d'image	Se (%)	Sp (%)	AUC	Acc (%)
Staff et al. (2009)	BL	#	-	SQ	25 C / 27 PD	79.4	90	0.94	#
Towey et al. (2011)	BL	AS+	NS* SRES	ML (NB)	37 C / 79 PD	93.7	97.3	#	94.8
Segovia et al. (2012)	VV	AS+	NS* NIm	ML (SVM)	94 C / 95 PD	93.7	95.7	0.97	94.7
Illán et al. (2012)	BL	AS+	NS** NIM	ML (SVM)	108 C / 100 PD	89	93.2	0.97	#
Rojas et al. (2013)	VN	As+	NS** NIM FG SRES	ML (SVM)	39 C / 41 PD	95.1	94.9	0.95	95
Martinez-Murcia et al. (2013)	PPMI VN VV	As+	NS** NIM	ML (SVM)	111 C / 158 PD	99.1	96.2	#	97.4
Prashanth et al. (2014)	PPMI	AS-	NS*	ML (SVM)	179 C / 369 PD	96.5	95	#	96.1
Oliveira et al. (2015)	PPMI	AS+	NS* NIM FG	ML (SVM)	209 C / 445 PD	97.7	98.1	#	97.9
Taylor & al. (2016)	PPMI	#	NS*(ShIRT) NIm SERS	#	200 C / 100 PD	#	#	0.98	94
Prashanth et al. (2017)	PPMI	AS+	NS NIm SERS	ML (BT, NB, RF, SVM*)	208 C / 427 PD	97.4	97.2	0.99	97.3
Augimeri et al. (2016)	BL	AS+	NS*(FSL- FLIRT) NI- α SERS	ML (SVM)	12 C / 31 PD	100	100	1	100
Ortiz et al. (2016)	PPMI	AS-	Nii SERS	DL	#	96	94	0.95	95
Choi et al. (2017)	PPMI SNUH	AS-	DA (G/D)	DL (CNN- 3D (PD net))	170 C / 379 PD / 77 SWEDD	94.2	100	0.99	96

Tagaris et al. (2017)	BL	AS-	-	DL (CNN (ResNet-50), CNN-RNN)	40 C / 100 PD	#	#	#	74
Brahim et al. (2017)	PPMI	AS+	NIm	ML (SVM)	111 C / 158 PD	91.2	93.1	#	92.6
Pianpanit et al. (2017)	PPMI	AS-	SERS	ML (SVM) DL (CNN (PD NET, <u>Deep PD NET*</u> modifié))	159 C / 448 PD	96.4	97.9	0.99	96.9
Klyuzhin et al. (2018)	PPMI	AS-	SERS DA1	DL (CNN)	305 C / 65 PD	#	#	0.69	#
Thurston et al. (2018)	BL	AS-	DA2	DL (CNN (Inception v3))	54 C / 54 PD	96.3	66.7	#	87
Castillo-Barnes et al. (2018)	PPMI(+BM)	AS-	NS** NI- α	ML (SVM)	194 C / 194 PD (SWEDD inclus)	96.1	97.4	0.98	96.7
Oliveira et al. (2018)	PPMI	AS-	NS*(ITK) NIM FG SERS	ML (K-NN, LR, <u>SVM*</u>)	209 C / 443	98	97.6	#	97.9
Rumman et al. (2018)	PPMI	AS+	SERS FG	ML(MLP)	100 C / 100 PD	100	88	#	94
Mabrouk et al. (2019)	PPMI	AS+	SERS	ML (<u>K-NN*</u> , LR, MLP, RM, SVM)	157 C / 342 PD	#	#	0.88	82.2
Quan et al. (2019)	PPMI	AS-	TL DA4	DL (CNN (Inception v3))	168 C / 259 PD	99	97.7	0.99	98.5
Magesh et al. (2020)	PPMI	AS-	NS- NIM DA3 TL	DL(CNN(VG G16))	170 C / 346 PD	97.6	90.9	0.94	95.2
Castillo-Barnes et al. (2020)	PPMI	AS+	NS**	ML (<u>SVM</u> , NB, MLP,	193 C / 193 PD	#	#	0.98	97

Tableau 21 : Liste d'études CAD au DaTSCAN

Remarque : Lorsque plusieurs réseaux sont évalués, ceux qui sont soulignés sont les plus performants.

Liste d'abréviations :

Base de données :

BM : Biomarqueurs de la PPMI
BL : Images d'une Base locale
PPMI : Parkinson's Progression Markers Initiative
SNUH: Seoul National University Hospital
VN : Images de l'hôpital Virgen de las Nieves
VV : Images de l'hôpital Virgen de las Victoria

Entraînement :

AS+ : Apprentissage supervisé
AS- : Apprentissage semi-supervisé

Nombre d'images :

C : Cas Contrôle
PD : Cas Parkinsonien
SWEDD: Scan Without Evidence of Dopaminergic Deficit

Pre-processing:

NS : Normalisation spatiale
NS* : Normalisation spatiale avec SPM
NS** : Normalisation spatiale avec SPM avec Template
NS* : Normalisation spatiale sans SPM avec Template
NI_m : Normalisation d'intensité à la moyenne
NI_M : Normalisation d'intensité au maximum
NI_I : Normalisation d'intensité à L'intégrale
NI_α : Normalisation d'intensité avec distribution α-stable
FG : Filtre Gaussien
SERS : Suppression des région extra striatales
DA1 : Data augmentation avec rotation, translation, flou
DA2 : Data Augmentation avec flip horizontal, rotation et changement d'échelle
DA3 : Data Augmentation avec translation, inversion
DA4 : Data Augmentation avec translation, inversion, brillance
TL : partir de la base de données ImageNet

Classification :

SQ : Analyse semi-quantitative basée sur le binding ratio
ML: Machine Learning Traditionnel
 K-NN: k-nearest neighbor
 LR: Linear Regression
 RF: Random Forest
 SVM: Support Vector Machine
 MLP: MultiLayer Perceptron
 NB: Naive Bayes
DL: Deep Learning

: pas d'information

- : rien

10.2 CAD et contrôles qualité

Selon le CADSC, il est recommandé de procéder à des contrôles qualités d'un système CAD utilisé en milieu clinique dont les spécificités sont les suivantes (Huo et al., 2013) :

Quand ?

- Installation d'un système CAD ou à la suite de nouvelles mise-à-jours du logiciel incluant celles du système d'exploitation. Cela peut impliquer des vérifications des performances/des conformités spécifiées par le fabricant (via des publications scientifiques, ou manuels d'utilisateur)
- Remplacement/Mise à niveau du système d'acquisition d'image ou de ses logiciels pouvant affecter la qualité de l'image (Tube RX, source d'Energie, détecteur, algorithmes de reconstruction d'image)
- Vérification d'éventuelles interactions avec des mise-à-jours ou l'installation de nouveaux logiciels sur l'ordinateur où se trouve le système CAD
 - Ne devrait pas faire varier les performances !
- Déviation de performances importantes des spécifications originales
- Procédures spécifiques considérées comme nécessaire sur une base régulière pour s'assurer de la consistance des performances (hebdomadaire, mensuel)
- Changement des paramètres d'acquisition (mAs, Kv, algorithme de reconstruction, etc.)

Par qui ?

- La procédure d'assurance qualité doit être effectuée par un personnel qui a été désigné et bénéficiant d'un entraînement adéquat (TRM, physicien médical, radiologue, fabricant)

Quel matériel/outils utiliser ?

- Utilisation d'un jeu de données images cliniques avec les normes de références connues qui est fourni par le fabricant ou qui est collecté sur le site clinique même
- Utilisation de logiciels fournis et validés par le fabricant ou d'un tiers fournisseur pour :
 - Enregistrement et récupération du jeu de données images du QA sélectionné sur un répertoire distinct des données images des patients
 - Marquage, validation et stockage des normes de références appropriées dans un format spécifié par le fabricant pour une notation automatisée des résultats des performances du CAD (VP, FP, exactitude, etc.)
 - Enregistrement des résultats sous format DICOM pour chaque image utilisée dans le QA.

- Affichage des résultats pour des objectifs de QA
- Calcul des résultats
- Utilisation éventuelle d'un fantôme destiné au QA des systèmes CAD avec une procédure validée.

Quelles mesures utiliser ?

- Dépend de l'application du système CAD et des objectifs du QA. Dans la situation d'un CAD diagnostic il est possible de se référer à la section 1.6. Il est important de relever que ces résultats permettent uniquement d'apprécier la variation des performances avec une même base de données images. Ces résultats ne doivent pas être comparés avec des résultats provenant de la population reportée par le fabricant du CAD ou d'autres études à moins que la base de données soit représentative de la population et que la taille d'échantillon soit suffisante.

Quelles informations fournir ?

- Contenu du journal du QA
 - Date et raison du QA
 - ID de l'utilisateur
 - Type de procédure
 - Localisation des données de test et des résultats
- Mesures de performance (section 1.6.1)
- Echec de communication entre le CAD et des systèmes nécessaires (système d'acquisition, écran d'affichage de l'ordinateur, autre dispositif d'imagerie)
- Enregistrement d'erreurs qui sont critiques

Quels critères de réussite ? Les limites de tolérances dépendent de l'application, de la taille d'échantillon et de la base de données de test utilisée

- Concernant les tests périodiques, il faut définir les niveaux de changement des performances comme acceptable, passable, inquiétant ou inacceptable
- Lors de mise-à-jours, uniquement mesurer la limite de tolérance maximale lorsque les performances diminuent

En cas d'échec ?

- Intervenir en localisant la panne pour savoir si un changement du software ou d'un hardware est en lien ou non avec le système CAD et qui est responsable des pertes de performances

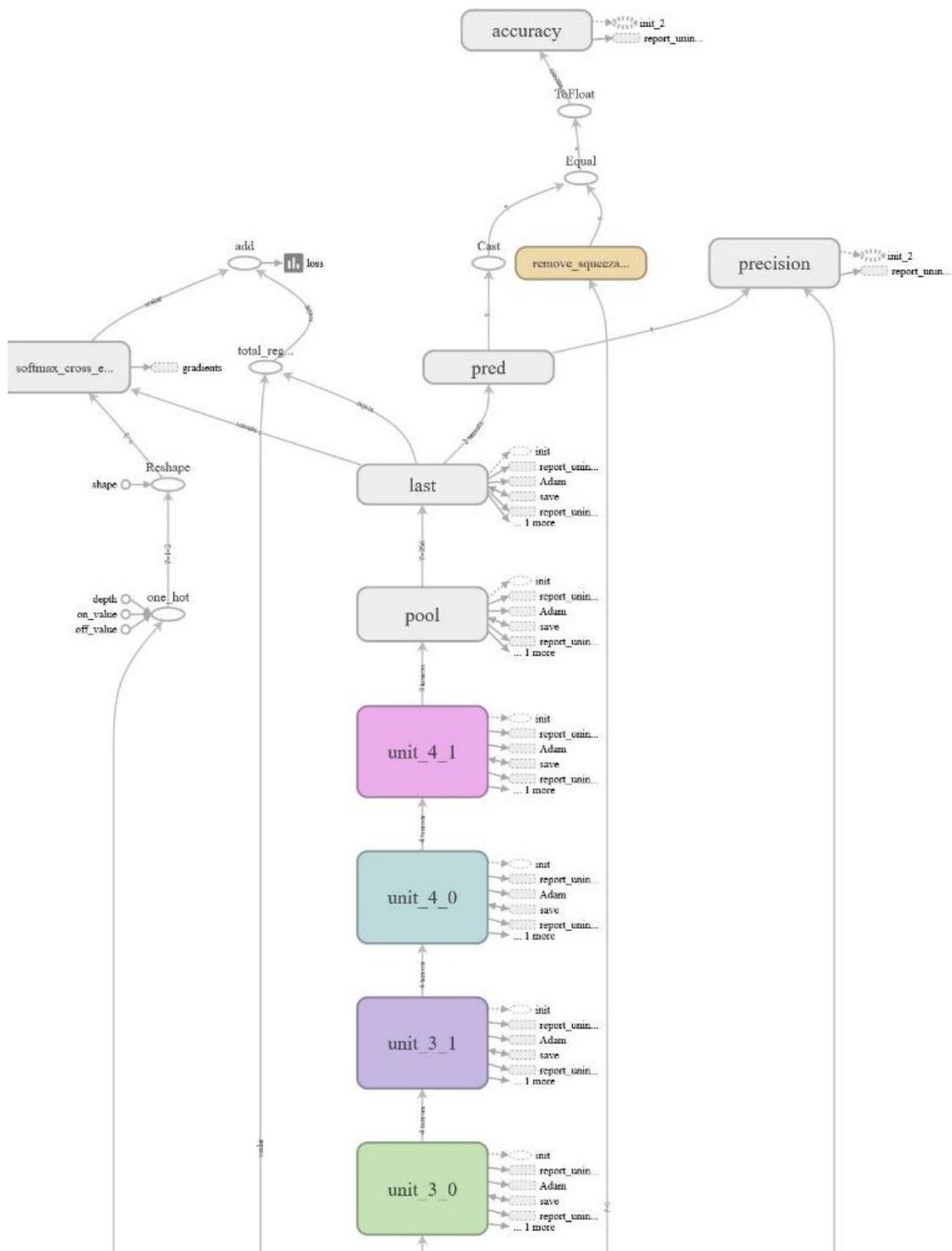
- Si l'échec est le résultat d'une contradiction entre la base de données test du QA et une autre base de données, il faut investiguer pour déterminer si la base de données de test est représentative de la population de patients sur le site

10.3 Architecture du réseau utilisé

Le diagramme de la Figure 42 généré par TensorFlow illustre l'architecture globale du modèle ResNet utilisé dans ce travail. On peut notamment observer en couleur 8 unités, chaque unité implémentant le concept d'unité résiduelle de l'approche ResNet. Ces unités intitulées « *unit_{i,j}* » sont distribuées par paires (*unit_{i,1}*, *unit_{i,2}*) dans un niveau de résolution *i*. Le réseau comporte 4 niveaux de résolution, $i \in [1,4]$. Les dernières couches du réseau comportent une couche de « batch normalisation » (BN, nœud « pool ») suivie d'une couche entièrement connectée (nœud « last »).

Notre réseau ResNet se compose donc de 4 niveaux d'échelles qui réduisent la dimensionalité des caractéristiques avec des convolutions. A chaque niveau, 2 unités « résiduelles » sont utilisées. Les paramètres du réseau ResNet sont les valeurs des noyaux de convolution, les paramètres de la couche BN ainsi que les poids de la dernière couche entièrement connectée. Hors situation de transfer learning, les paramètres de la couche BN sont initialisés à zéro tandis que celles des autres couches sont initialisées de façon aléatoire en suivant une distribution uniforme dans l'intervalle $[-\sqrt{3/n}, \sqrt{3/n}]$ où *n* correspond au nombre d'éléments des noyaux de convolution ou au nombre de poids.

La fonction de perte utilisée est une forme régularisée de la fonction « softmax cross entropy » qui est minimisée avec la méthode « Adam ». Un taux d'apprentissage de 0.001 est maintenu constant durant l'optimisation. L'optimisation est effectuée sur 25000 itérations avec un test de validation toutes les 100 itérations. En fin d'entraînement le modèle ayant l'exactitude la plus élevée à un test de validation est celui conservé.



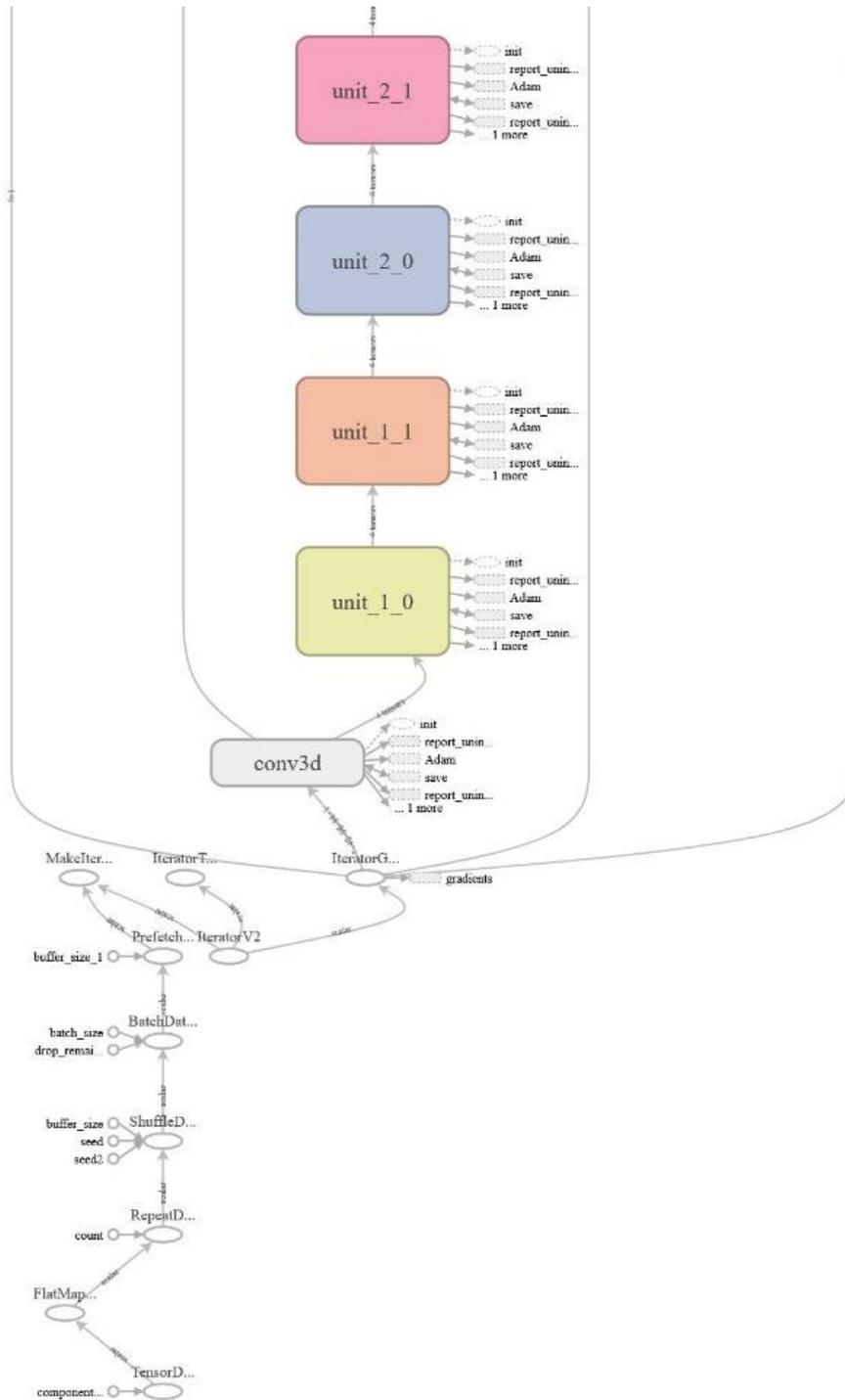


Figure 42 Diagramme illustrant notre architecture ResNet

10.4 Tableau récapitulatif des différents entrainements et tests

No.	Training			Evaluation			SPM	Soo	GrBal	AugGéo
	Symbia	Toshiba	PPMI	Symbia	Toshiba	PPMI				
X/X										
Entrainement PPMI benchmark										
1			X			X				
2			X			X				X
Normalisation sur Symbia										
3	X			X						
4	X			X			X			
5	X			X			X			X
Lissage sur Symbia										
6	X			X				X		
7	X			X			X	X		
8	X			X			X	X		X
Toshiba										
9		X			X					
10		X			X					X
11		X			X			X		
X/Y										
Interbase entrainement 1 base										
12			X	X			X			
13			X	X			X			X
14	X					X	X			X
X/XY										
Interbase entrainement 2 bases										
15	X		X	X						
16	X		X	X					X	
17	X		X		X				X	
18	X		X			X			X	
19	X		X	X					X	Sy
20	X		X		X				X	Sy
21	X		X			X			X	Sy
22	X	X		X			X			Sy
23	X	X			NN		X			Sy
X/XYZ										
Interbase entrainement 3 bases										
24	X	X	X	X					X	
25	X	X	X		X				X	
26	X	X	X			X			X	
27	X	X	X	X					X	Sy

28	X	X	X		X				X	Sy
29	X	X	X			X			X	Sy
X=>Y/Y										
Transfert Learning PPMI augmenté géométriquement vers Symbia										
30	X			X		TL	X			
31	X			X		TL	X			X

Tableau 22 : Différents entraînements et tests effectués avec : Sp (normalisation des données sous SPM), Soo (lissage des images), GrBal (balancing de groupe), AugGéo (augmentation géométrique des données) et TL (Transfert Learning).