Université de Poitiers

Institut de Recherche XLIM :
UMR CNRS 7252, 86073 Poitiers Cedex 9,
France



UFR des sciences fondamentales et appliquées

École Doctorale :
Mathématiques, Informatique, Matériaux, Mécanique, Énergétique (MIMME)

Thèse

en vue de l'obtention du

Grade de Docteur en Informatique

Spécialité doctorale :

Traitement du Signal et des Images

Directrice de thèse :

Christine FERNANDEZ MALOIGNE

Encadrants :

Thierry URRUTY
Pascal BOURDON

par

Lobna FEZAI

Représentation uniforme de l'imagerie médicale

Soutenu le jj/mm/aaaa, devant la commission d'examen :

Guy CARRAULT	Examinateur
Alain MIRRANVILLE	Examinateur
Muriel VISANI	Rapporteur
Christophe CHARRIER	Rapporteur
Christine FERNANDEZ MALOIGNE	Directeur
Thierry URRUTY	Encadrant
Pascal BOURDON	Encadrant, invité

TABLE DES MATIÈRES

LISTE DES FIGURES viii

LISTE DES ABRÉVIATIONS ix

INTRODUCTION GÉNÉRALE xi

1 État de l'art : Intelligence artificielle 2

1.1 Histoire rapide de l'IA . . . 3

1.2 Apprentissage automatique . . . 6

1.2.1 Apprentissage supervisé . . . 7

1.2.2 Apprentissage non supervisé . . . 8

1.3 Apprentissage profond . . . 8

1.3.1 Le réseau de neurones . . . 9

1.3.2 Types de réseau de neurones . . . 10

1.3.3 Fonctions d'erreur . . . 16

1.3.3.1 Erreur absolue moyenne . . . 16

1.3.3.2 Erreur quadratique moyenne . . . 17

1.3.3.3 Entropie croisée . . . 17

1.3.4 La descente de gradient . . . 21

1.3.5 Fonctions d'activation . . . 22

1.3.6 Divers hyperparamètres, performance et généralisation du modèle . . . 22

1.3.6.1 Hyperparamètres du modèle . . . 24

1.3.6.2 Performance et généralisation du modèle . . . 25

1.3.7 Métriques . . . 29

1.4 L'apprentissage profond pour l'image . . . 32

1.4.1 L'image aux yeux de la machine . . . 33

1.4.2 L'imagerie médicale . . . 35

1.4.3 L'évolution des réseaux convolutifs . . . 37

1.4.4 Apprentissage profond pour l'imagerie médicale . . . 40

1.4.5 Défis et perspectives 45

2 État de l'art : Meta-apprentissage et attention 48

2.1 Problématique et solution au manque de données 52
2.2 Few-shot learning 54
2.3 Meta-apprentissage 56
2.4 Solutions de l'état de l'art 59
2.4.1 Apprentissage par transfert 59
2.4.2 Apprentissage de l'espace métrique 63
2.4.2.1 Réseau siamois 64
2.4.2.2 Réseau de correspondance (Matching network) 65
2.4.2.3 Réseau prototypique (Prototypical Networks) 67
2.4.2.4 Réseau de relation (Relation network) 68
2.4.3 Des solutions d'apprentissage liées à l'algorithme d'optimisation 70
2.4.3.1 Mécanisme de mémoire externe 70
2.4.3.2 Model Agnostic Meta learning (MAML) 74
2.4.3.3 SNAIL 76
2.5 Mécanisme d'attention 77
2.5.1 Carte d'attention 77
2.5.1.1 Objectif 77
2.5.1.2 Contribution 78
2.5.1.3 Approche et résultat 79
2.5.2 Réseau d'attention spatiale pour la classification en few-shot learning . 80
2.5.2.1 Objectif 80
2.5.2.2 Contribution 80
2.5.2.3 Approche et résultat 81
2.5.3 Apprentissage d'une représentation discriminante profonde basée sur la carte d'attention pour la classification de scènes 82
2.5.3.1 Objectif 82
2.5.3.2 Contribution 83
2.5.3.3 Approche et résultat 83
2.5.4 Apprentissage visuel dynamique en quelques coups sans oublier 84
2.5.4.1 Objectif 84
2.5.4.2 Contribution 85
2.5.4.3 Approche et résultat 85
2.5.5 Branche d'attention : Apprentissage du mécanisme d'attention pour une interprétation visuelle 86

2.5.5.1 Objectif . . . 86
2.5.5.2 Contribution . . . 86
2.5.5.3 Approche et résultats . . . 87
2.6 Conclusion . . . 88

3 Anonymisation profonde 89
3.1 Motivation . . . 90
3.2 L'apprentissage profond et la confidentialité . . . 94
3.2.1 Les données : le pétrole d'aujourd'hui . . . 95
3.2.2 La confidentialité en jeu . . . 99
3.2.3 Empreintes digitales . . . 99
3.3 Base de données . . . 100
3.4 L'anonymisation liée aux équipements d'acquisition de l'IRM . . . 102
3.4.1 Approches et architectures . . . 103
3.4.1.1 Jeu de données . . . 103
3.4.1.2 La classification en fonction de différents équipements . . . 104
3.4.1.3 La reconstruction de l'IRM . . . 104
3.4.2 Reformulation mathématique . . . 106
3.4.3 Expérimentations et résultats . . . 107
3.5 L'anonymisation liée à l'identité du patient . . . 111
3.5.1 Approches et architectures . . . 113
3.5.1.1 Jeu de données . . . 113
3.5.1.2 La classification en fonction de l'identité de patients . . . 114
3.5.1.3 La reconstruction de l'IRM . . . 114
3.5.2 Expérimentations et résultats . . . 114
3.6 Conclusion . . . 117

4 Meta-apprentissage attentif 119
4.1 Motivation . . . 120
4.2 État de l'art . . . 123
4.3 Approches et reformulations . . . 125
4.3.1 Réseau de neurones attentif . . . 125
4.3.2 Réseau de neurones attentif amélioré . . . 129
4.4 Expérimentations et résulats . . . 130
4.4.1 Bases de données . . . 130
4.4.1.1 Base de données ISIC 2018 . . . 131

4.4.1.2 Base de données des radiographies du thorax 131
4.4.2 Expérimentations 132
4.4.3 Résultats 133
4.4.4 Conclusion 134

CONCLUSION GÉNÉRALE **135**

BIBLIOGRAPHIE **137**

LISTE DES FIGURES

1 La différence entre l'apprentissage profond et les méthodes d'apprentissage automatique classiques [115] xii

2 Détection des objets en utilisant l'apprentissage profond [249] xiii

3 Reconnaissance faciale [231] xiv

4 Reconnaissance des sentiments[249] xiv

5 Illustration du problème de l'oubli catastrophique [126] xvi

6 Synthèse des défis de l'apprentissage profond xix

1.1 Le diagramme hiérarchique des sous domaines de l'intelligence artificielle . . 6

1.2 Structure de la base de données 6

1.3 Comparaisons entre un neurone biologique et un neurone artificiel [70][135] . . 9

1.4 Exemple d'un réseau de neurones [70] 10

1.5 Exemple de la structure de CNN [229] 11

1.6 Exemple d'une convolution par un seul filtre [260] 12

1.7 Exemples de cartes de caractéristiques résultantes de la convolution par trois filtres [260] 13

1.8 Les cartes caractéristiques résultantes suite à une ReLU [260] 14

1.9 Les cartes caractéristiques résultantes suite à une opération de max-pooling de taille 2x2 [260] 15

1.10 Valeur de l'erreur entropie croisée lorsque la probabilité cible est égale à 1 inspiré de [35] 18

1.11 Processus d'apprentissage pour une classification binaire 19

1.12 Processus d'apprentissage pour une classification multiclasse 20

1.13 Influence du nombre de couches et du nombre d'époques d'entraînement sur la valeur de l'erreur du modèle, inspirée de [106]. (a) représente les valeurs de la MAE en fonction du nombre d'époques pour des profondeurs différentes du réseau de neurones. (b) montre l'influence directe du nombre de couches sur la MAE du modèle. 24

1.14 Influence de nombre de neurones sur l'entraînement [44] 25

1.15 Le sous-ajustement et le sur-ajustement [1] 26

1.16 Courbes illustrant le dilemme biais-variance [1] 27

1.17 Illustration du problème de sur-ajustement et d'interruption prématurée, inspirée de [236] 28
1.18 Illustration de la stratégie de la validation croisée [236] 28
1.19 Influence de la valeur du terme de régularisation sur un réseau de neurones entrainé pour une tâche binaire de classification [44] 29
1.20 Illustration des éléments de la matrice de confusion [31] 30
1.21 Matrice de confusion [31] 31
1.22 Représentation de l'image [154] 33
1.23 Une partie de la matrice représentative d'une image en niveaux de gris [114] . 34
1.24 Illustration de la normalisation de Nyul et Udupa [124] 34
1.25 La chronologie des architectures de l'apprentissage profond 37
1.26 Précision de la classification et les propriétés des architectures des réseaux de neurones profonds appliquées sur la base de données de validation de ImageNet 40

2.1 Importance de la quantité de données sur la performance de l'apprentissage profond [5] 50
2.2 La MAE en fonction de la taille de la base de données [106] 50
2.3 Les implications du manque de données 51
2.4 Few-shot learning, meta-learning et mécanisme d'attention 54
2.5 La structure de la base de données de N-way K-shot learning 55
2.6 Jeu de données pour une approche d'apprentissage classique 57
2.7 Jeu de données pour une approche de meta-apprentissage : L'ensemble de données est en fait un meta-ensemble divisé entre les épisodes qui composent les époques. Nous distinguons deux processus d'apprentissage et deux apprenants : le meta-apprenant, un modèle qui apprend à travers les épisodes, et un deuxième modèle appelé base-apprenant, incorporé et entraîné à l'intérieur d'un épisode par le meta-apprenant. Considérons une époque de meta-apprentissage composée de plusieurs tâches de classification (plusieurs épisodes). Au cours d'une tâche de classification T définie par un support set de $N * K$ images et un query set de Q images, base-apprenant est initialisé et entraîné sur le support set. Ensuite, il est appliqué sur les images de query set pour prédire leurs classes. À la fin de chaque épisode, les poids du meta-apprenant sont mises à jour en se basant sur la valeur de la fonction d'erreur de classification de query set. La différence entre les techniques mentionnées dans ce chapitre est l'approche de fonctionnement du meta-apprenant. 58
2.8 Solutions de l'état de l'art 60
2.9 Combinaison de meta-apprentissage et d'apprentissage par transfert : FT désigne le fine tuning. SS désigne "Scaling and Shifting", une opération de transfert simple définit par Sun et al. [240]. Le Meta-Transfer Learning représente l'approche principale. Le Hard Task Meta-Batch est une stratégie d'entraînement qui favorise les tâches difficiles. 63

2.10 Illustration de l'approche du réseau siamois [125] 64
2.11 Architecture du réseau siamois [117] 65
2.12 Architecture du réseau siamois avec une fonction d'erreur triplets [117] 66
2.13 Architecture du réseau de correspondance [264] : l'extraction des caractéristiques des images de support set et query set est effectuée respectivement par les modèles g_θ et f_θ. Les vecteurs caractéristiques alimentent une couche softmax suite à laquelle et en se basant sur la distance cosinus entre le vecteur caractéristique de l'image requête et les vecteurs caractéristiques des images du support set., l'image requête est classée. 66
2.14 Architecture du réseau prototypique [232] 67
2.15 Classification few-shot learning de Omniglot [242] 68
2.16 Classification few-shot learning de miniImagenet [242] 68
2.17 Architecture du réseau de relation [242] 69
2.18 Illustration de la comparaison des réseaux de neurones métriques [268] 69
2.19 Architecture LSTM 71
2.20 Illustration de l'approche des réseaux de neurones à mémoire augmentée[268] . 73
2.21 MAML pipeline 75
2.22 Une étape de meta -entraînement d'un processus MAML [17] 75
2.23 Gauche : images de ILSVRC-2013. Milieu-gauche : Carte d'attention prédites. Milieu-droite : cartes d'attention seuillées avec le bleu qui montre les régions d'intérêt les plus importantes. Droite : segmentation résultante en se basant sur les masques d'attention produits [226] 78
2.24 Images d'origine, cartes d'attention et masques d'attention résultants [143] . . 82
2.25 Aperçu de l'architecture finale : 1) Un réseau attentionnel-génératif : Réseau CNN ResNet-18 pré-entrainé sur la base de données Imagenet et ajusté sur la nouvelle base et l'architecture Grad-CAM (Gradient-poids et algorithme basé sur les cartes d'activation de classe) pour générer les carte d'attention. 2) Architecture CNN à deux flux pour la fusion des images originales avec les cartes d'attention en combinant les deux fonctions d'erreur [143]. 84
2.26 Aperçu de l'architecture : 1) Une première partie d'extractions de caractéristiques de classification basée sur ConvNet 2) Un classificateur à quelques coups comme générateur de poids. [72] 85
2.27 Différene entre la structure classique de carte d'activation de classe (CAM) et la structure basée sur la branche d'attention (ABN) 87

3.1 Les domaines d'application de l'exploitation de données [244] 94
3.2 Quantité de données produite par minute en 2014 [27] 96
3.3 Quantité de données produite par minute en 2021 [27] 97
3.4 Processus d'exploitation des données inspiré de [244] 98
3.5 Accessibilité et l'exploitation des données dans les différents secteurs [244] . . 98

3.6 La distribution de la base de données ADNI en fonction d'âge et de sexe en 2022 101
3.7 Architecture proposée pour la classification des équipements d'IRM 105
3.8 Architecture proposée pour la reconstruction de l'IRM 106
3.9 Courbe de la précision de classification de l'entraînement et de la validation de la classification 107
3.10 Courbe de la fonction d'erreur de l'entraînement et de la validation de la classification 108
3.11 Matrice de confusion des données test de la classification 109
3.12 Des échantillons des images reconstruites 110
3.13 La courbe de la fonction d'erreur d'auto-encodeur 110
3.14 La courbe de la PSNR en fonction de Lambda 111
3.15 La courbe de la PSNR en fonction de la précision de classification du test . . . 112
3.16 Matrice de confusion des données de test de notre approche finale 112
3.17 Courbe de la précision de classification et de la perte de la méthode de la classification classique 115
3.18 Courbe de la précision de classification et de la perte de la méthode de réseau seamois 115
3.19 Matrice de confusion des données test de la classification 116
3.20 Exemple d'images reconstruites par l'auto-encodeur avec la fonction d'erreur adaptée. À gauche, l'image originale et l'image reconstruite d'un patient souffrant de la maladie d'Alzheimer, à droite, les images d'un patient sain. . . 116
3.21 Matrice de confusion des données de test de notre approche finale 117

4.1 L'architecture du réseau de neurones attentif proposé 126
4.2 Base de données ISIC 2018 131
4.3 Quelques échantillons des différentes classes des scanners thoraciques [269] . . 132
4.4 Des échantillons des cartes d'attention et des masques d'attention résultants de la base de données ISIC 2018 133

LISTE DES ABRÉVIATIONS

BAC Balanced Accuracy

BCE Binary Cross-Entropy

CA Classification Accuracy

CCE Categorical Cross-Entropy

CAM Carte d'Activation de Classe

DL Deep Learning

DNN Deep Neural Network

DT Decision Tree

FN False Negative

FP False Postive

IA Intelligence artificielle

KNN K-Nearest Neighbors

MAML Model-Agnostic Meta-Learning

ML Machine Learning

MSE Mean Squared Error

MAE Mean Absolute Error

NLP Natural Language Processing

NLU Natural Language Understanding

ReLU Rectified Linear Unit

RF Random Forest

LR Logistic Regression

SCCE Sparse Categorical Cross-Entropy

SVM Support-Vector Machine

TN True Negative

TP True Positive

Tanh Tangente hyperbolique

INTRODUCTION GÉNÉRALE

Le domaine médical est un domaine aux enjeux socio-economiques primordiaux. La recherche dans ce domaine est essentielle pour aboutir rapidement au meilleur diagnostic, prévenir le développement des maladies et découvrir les traitements les plus efficaces pour guérir les patients ou au moins ralentir la progression des maladies et améliorer la qualité de vie des malades. Les données revêtent une grande importance dans le domaine médical, y compris l'imagerie médicale. L'exploitation optimale de l'imagerie médicale est un domaine de recherche très actif pour améliorer le diagnostic et les soins proposés. Indispensable dans ce domaine, l'intelligence artificielle apporte une aide au diagnostic, pronostic et soins personnalisés.

L'intelligence artificielle est une branche de l'informatique incorporant des algorithmes permettant de rendre la machine plus performante et plus efficace dans des multiples tâches, notamment de segmentation ou de classification, en minimisant la marge d'intervention de l'être humain. L'apprentissage automatique est une sous-section de l'intelligence artificielle dont le principe est de se baser sur un ensemble d'exemples et des expériences pour apprendre à effectuer une tâche spécifique sans que la machine soit explicitement codée dans ce sens. Par conséquent, la machine est capable d'effectuer des tâches réservées précédemment à l'humain d'une façon plus rapide, plus optimale et moins coûteuse. Plusieurs techniques sont proposées dans le cadre de l'apprentissage automatique, notamment la machine à vecteurs de support (SVM), la forêt aléatoire (RF) et l'apprentissage profond (DL). Ce dernier offre un potentiel exceptionnel dans la recherche dans le domaine médical et introduit une piste encore plus riche et plus variée des solutions technologiques médicales permettant de traiter les données à des vitesses énormes, d'une façon efficace et avec une bonne précision.

L'apprentissage profond a suscité beaucoup d'intérêt dans la dernière décennie. Les applications récentes dans plusieurs domaines sont davantage axées sur le déploiement de l'apprentissage profond plutôt que d'autres techniques d'apprentissage automatique classiques pour des multiples raisons. En effet, afin de simuler de plus en plus l'approche d'apprentissage de l'être humain, l'apprentissage profond se sert des modèles mathématiques conçus pour simuler l'apprentissage du cerveau humain. Le principe se base sur un réseau de neurones artificiel composé de couches de neurones ayant des connexions favorisant l'échange des informations entre eux. Ce réseau permet d'exploiter les dépendances statistiques pour découvrir des relations sémantiques latentes significatives et d'extraire des caractéristiques pertinentes. L'extraction de caractéristiques est l'une des étapes primordiales pour effectuer une tâche. Ces réseaux d'apprentissage profond sont capables de résoudre des problèmes complexes et de dégager des éléments de compréhension à partir d'une grande quantité de données qui auraient auparavant été perdues, oubliées ou négligées, notamment dans le secteur de l'imagerie médicale. La sélection automatique de caractéristiques en déployant l'apprentissage profond a surpassé les méthodes heuristiques ou manuelles traditionnelles et a permis d'aboutir à des compromis intéressants entre la précision, le coût et la vitesse des algorithmes (fig. 1)[263].

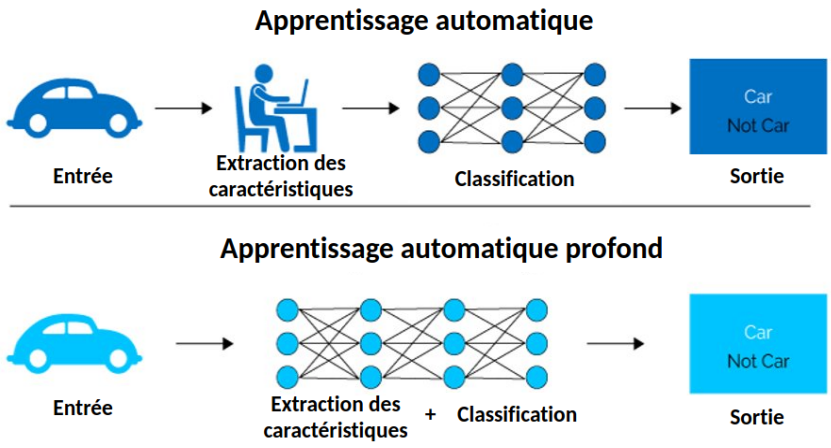


FIGURE 1 – La différence entre l'apprentissage profond et les méthodes d'apprentissage automatique classiques [115]

Ainsi, les algorithmes d'apprentissage profond sont mieux adaptés à différents cas d'utilisation tels que la vision par ordinateur (détection des objets (voir fig. 2), reconnaissance

faciale (voir fig. 3), reconnaissance des sentiments (voir fig. 4)), traitement de signal ou le traitement du langage naturel, dépassant même les capacités humaines pour certaines tâches. Par exemple, en 2016, une solution basée sur un réseau de neurones profond a été proposé avec une meilleure capacité de classification et de reconnaissance des objets que l'homme [247]. Au cours de la même année, une autre méthode d'apprentissage automatique appelée AlphaGo a battu les champions du monde au jeu de GO [225]. Des marques d'automobile se sont mises en compétition pour produire les meilleures voitures autonomes en déployant l'apprentissage profond [82]. L'apprentissage profond est donc devenu omniprésent dans notre quotidien, parfois sans prise de conscience, il est incorporé dans nos voitures, nos smartphones, nos appareils connectés, nos maisons et nos établissements. Il est utilisé dans des applications aux domaines variés, notamment la sécurité [262], l'éducation, l'analyse du climat, les prédictions financières et le diagnostic médical auquel nous nous intéressons particulièrement dans cette thèse.

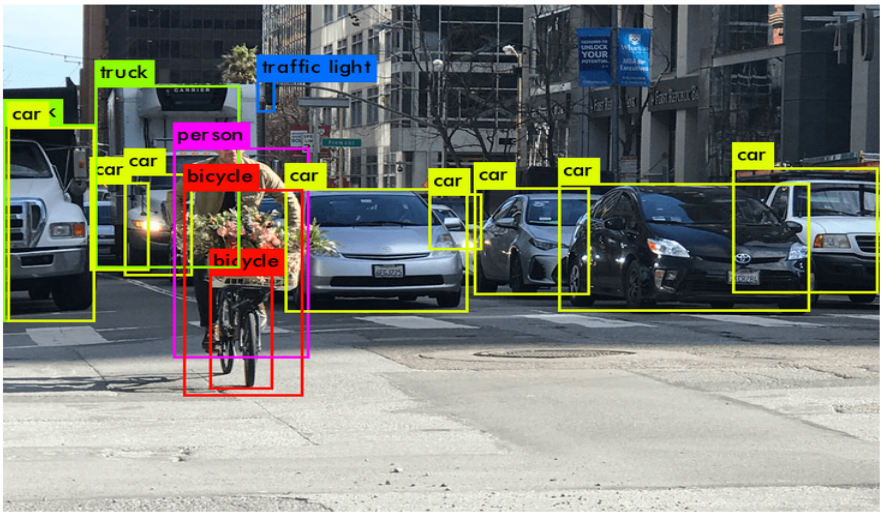


FIGURE 2 – Détection des objets en utilisant l'apprentissage profond [249]

L'apprentissage profond appliqué au domaine médical s'est énormément développé dans cette dernière décennie : des chatbots capables de réaliser un diagnostic symptomatique des patients, des analyses cliniques automatisées et des algorithmes d'apprentissage profond appliqués sur l'imagerie médicale pour identifier des maladies rares ou des types spécifiques d'une certaine pathologie. Les méthodes à base d'apprentissage profond sont conçues comme

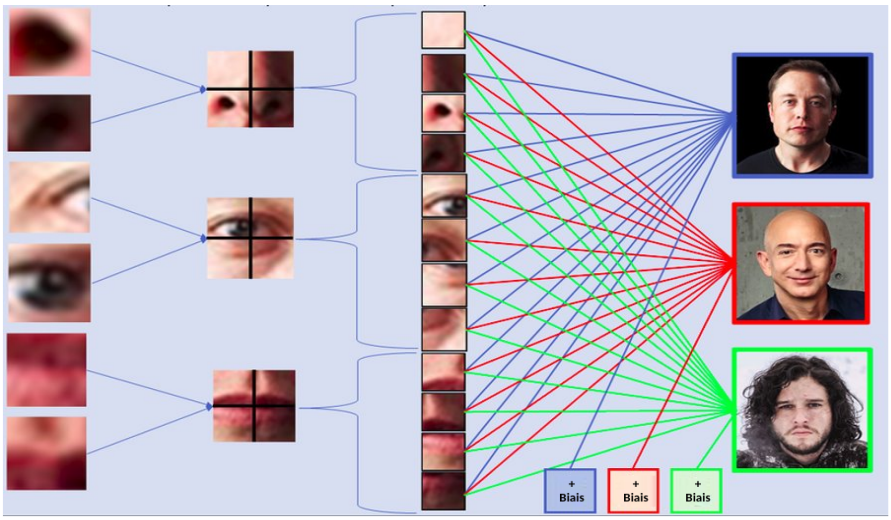


FIGURE 3 – Reconnaissance faciale [231]



FIGURE 4 – Reconnaissance des sentiments[249]

un outil d'aide au diagnostic et de soutien des experts médicaux et non pas comme un outil visant à supplanter le médecin.

Par exemple, Google s'est intéressé à l'apprentissage profond afin d'aider les cliniciens à gérer les données et les résultats des patients. Dans ce sens, l'article "apprentissage profond pour les dossiers médicaux électroniques" [195] montre la possibilité de réduire la charge administrative en améliorant la compréhension des traitements et des besoins des patients. La boîte à outils d'apprentissage profond peut également fournir un soutien indispensable aux professionnels de santé eux-mêmes. Le prestataire de soins de santé NHS au Royaume-Uni a reconnu la valeur de cette technologie et s'est engagé à devenir un leader dans le domaine des soins de santé grâce à l'apprentissage profond. Un investissement dans des solutions d'apprentissage profond, notamment AWS d'Amazon et Aidoc, permet de contourner certains des défis du domaine médical et en particulier les exigences rigoureuses de l'imagerie médicale.

Un rapport de comparaison des performances de l'apprentissage profond avec celles des professionnels de santé dans la détection des pathologies à partir de l'imagerie médicale, publié par le Lancet Digital Health Journal [149], a prouvé la crédibilité et la précision des algorithmes profonds dans le diagnostic médical à l'aide de l'imagerie médicale. Selon ce rapport, toujours sous quelques réserves, les "performances des modèles d'apprentissage profond sont équivalentes à celles des professionnels de santé".

Si l'apprentissage profond dans les soins de santé n'en est encore qu'aux premiers stades de son potentiel, il a déjà donné des résultats significatifs. Ses bénéfices ont été reconnus par des institutions et des organisations médicales de premier plan. L'avenir est toujours entre les mains des professionnels de santé, mais ils sont désormais soutenus par une technique qui comprend leurs besoins et qui diminue les stress qu'ils subissent au quotidien.

Eric Topol, cardiologue, généticien et auteur du livre "Deep Medicine" [255], a souligné, dans une interview sur l'application de l'apprentissage profond pour restaurer les soins de santé, l'immense importance de cette technologie dans le domaine de la santé [47]. Il a exprimé aussi que l'application médicale la plus prometteuse est le diagnostic basé sur l'imagerie médicale représentant un premier dépistage sous la supervision du professionnel de santé. Il a mis en relief aussi les principaux défis techniques et pratiques de l'incorporation l'IA dans la médecine, notamment la sécurité, les biais des algorithmes et les inégalités expliquées par le fait que l'IA n'est accessible qu'à ceux qui peuvent se l'offrir à cause des coûts élevés de calcul et de stockage. Ceci nous amène à la dernière partie de cette introduction. En effet, bien que l'apprentissage profond soit une solution idéale pour de multiples problématiques, la mise en œuvre de l'apprentissage profond au sein du domaine médical fait l'objet de multiples critiques et plusieurs défis techniques, scientifiques, éthiques et sociaux sont rencontrés.

a) Défis techniques

L'apprentissage profond exige une grande quantité de ressources, notamment du calcul pour exécuter les algorithmes complexes rapidement et de la mémoire pour stocker les modèles et les données, provoquant un besoin énorme d'énergie. Ces exigences sont considérées souvent

comme une entrave qui réduit significativement les domaines d'application des réseaux de neurones, notamment des cas d'utilisation en temps réel ou des applications dans des systèmes embarqués avec des ressources limitées en mémoire, en calcul ou en batterie.

Pour minimiser le besoin en ressources mémoire, les réseaux de neurones classiques essaient de trouver la meilleure représentation des données présentes et finissent par détruire les connaissances précédemment acquises. Cet automatisme engendre un phénomène appelé l'oubli catastrophique [66] (fig. 5).

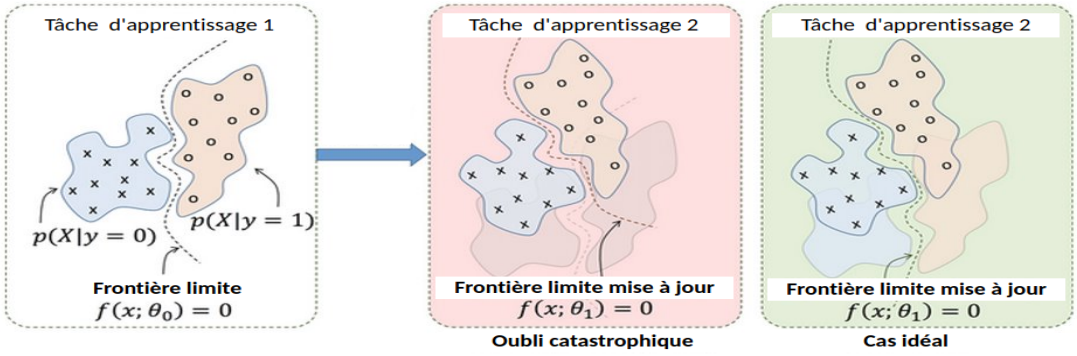


FIGURE 5 – Illustration du problème de l'oubli catastrophique [126]

En outre, l'apprentissage profond se base simplement sur un algorithme d'optimisation qui représente ses problèmes classiques de convergence (sous-apprentissage) [21], de manque de généralisation (sur-apprentissage) [89], de gradient évanescent (une diminution très rapide des valeurs des gradients entraînant l'arrêt prématuré de l'apprentissage) [95] et de l'explosion de gradients [241] (voir chap. 1).

Enfin, après avoir fourni toutes les ressources nécessaires, un dernier défi, récurrent et très critique pour l'apprentissage profond, est la non-disponibilité et la qualité des données. En effet, l'apprentissage profond est un outil très gourmand en termes de données et la collecte de données peut s'avérer difficile pour la simple raison de l'absence d'une quantité suffisante de données de bonne qualité ou pour d'autres raisons liées aux défis éthiques sociaux cités au-dessus.

Pour conclure, les données, la mémoire, le calcul et l'énergie représentent des ressources clés pour réussir dans ce domaine. En revanche, ils peuvent être des verrous lors du déploiement des modèles profonds. Par conséquent, plusieurs techniques sont proposées dans cette direction

pour minimiser le besoin de données et réduire la complexité de l'apprentissage (voir chap. 2) notamment le meta-apprentissage sur lequel se base notre deuxième contribution.

b) Défis scientifiques

L'apprentissage profond est un domaine de recherche expérimental. Les réseaux de neurones font intervenir plusieurs hyper-paramètres dont le réglage n'est toujours pas évident et un très grand nombre de paramètres mis à jour itérativement à travers l'algorithme d'optimisation. L'initiation aléatoire de ces paramètres ainsi que le tâtonnement nécessaire pour fixer les hyper-paramètres peuvent engendrer encore plus de complexité et de besoins en termes de calcul et de mémoire. Les protocoles expérimentaux de recherche peuvent être exigeants afin d'obtenir les meilleures performances. Par conséquent, nous pouvons nous trouver dans l'obligation de tester plusieurs architectures possibles et d'effectuer plusieurs tâtonnements pour répondre aux questions scientifiques liées au choix des hyper-paramètres et à la structure du modèle.

En outre, un réseau de neurones artificiel est une structure complexe qui se compose d'un grand nombre de neurones et de connexions difficiles à comprendre ou à décrire mathématiquement. Ainsi, les démarches logiques derrière les décisions prises ne sont toujours pas évidentes. En revanche, obtenir une performance maximale en utilisant un modèle moins complexe est un véritable défi. Dans ce sens, la compréhension, l'interprétation et la robustesse de l'apprentissage profond soulèvent beaucoup de préoccupations scientifiques [212].

Des études dans cette direction ont donné naissance à une nouvelle piste importante de recherche connue sous le nom d'intelligence artificielle explicable ou "explainable AI". Parmi les techniques proposées pour expliciter les décisions des modèles de l'apprentissage profond, nous nous intéressons dans notre travail au phénomène de l'attention (voir chap. 2).

c) Défis éthiques et sociaux

La confidentialité et le respect de la vie privée sontdevenusdevenues deux termes liés étroitement à l'intelligence artificielle et en particulier à l'apprentissage profond. Des préoccupations légitimes expliquées par l'omniprésence et la puissance de ces outils font

émerger un discours éthique et social de nos jours (voir chap. 3.6). Ce problème éthique provoque des restrictions à l'accès aux données, limitant ainsi le problème de confidentialité, mais aussi le progrès favorable des recherches dans le secteur de l'apprentissage profond dans de multiples domaines.

Par ailleurs, les solutions proposées par l'apprentissage profond visent à aider tous les êtres humains équitablement, et donc à les soulager d'un travail épuisant et à faciliter leurs tâches quotidiennes. En revanche, comme mentionné ci-dessus, l'apprentissage profond est une solution gourmande en termes de ressources de données, de calcul et de mémoire nécessitant des dispositifs coûteux pour traiter les données. Si les données sont un facteur limitant essentiel pour les institutions publiques de recherche, les calculs le sont également. En réduisant les ressources nécessaires pour trouver les bons hyperparamètres pour une tâche donnée et le besoin de données, nous faisons un pas en avant vers la démocratisation et l'accessibilité de l'apprentissage profond pour tous [85].

Pour terminer, un dernier défi de taille est à noter : le problème écologique. Le coût énergétique nécessaire pour effectuer les calculs complexes exigés par les algorithmes de l'apprentissage profond peut rapidement devenir énorme. Une telle consommation d'énergie fait de l'apprentissage profond une solution qui ne respecte pas l'environnement et le développement durable.Trouver des méthodes pour réduire la complexité et donc la consommation d'énergie nécessaire pour l'apprentissage pourrait être une des clés pour limiter l'impact écologique néfaste du déploiement des réseaux neurones.

Tout au long de cette thèse, nous proposons des solutions à ces défis de l'utilisation de l'apprentissage profond, en particulier dans le domaine de l'imagerie médicale avec ses différentes particularités (fig. 6). Nous espérons pouvoir résoudre ces difficultés afin de permettre l'exploitation optimale des réseaux de neurones pour le meilleur diagnostic médical possible. Les solutions que nous abordons peuvent être généralisées dans de multiples domaines et avec des adaptations variées.

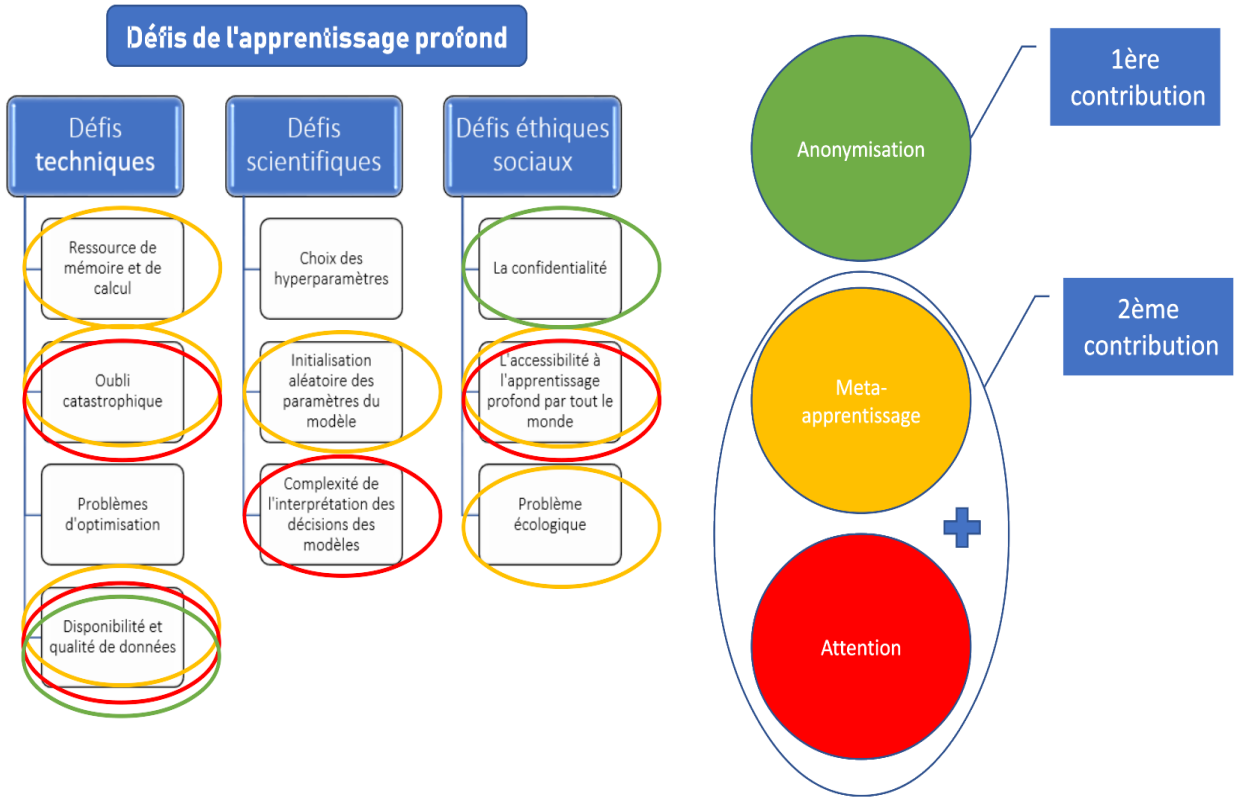


FIGURE 6 – Synthèse des défis de l'apprentissage profond

Dans la figure 6, nous résumons les défis mentionnés au-dessus et nous présentons nos deux majeures contributions permettant de résoudre plusieurs de ces défis. Les défis sont entourés par des cercles de couleur en lien avec celle de nos contributions sur la droite.

Les principales contributions de cette thèse sont en lien avec les défis rencontrés lors du déploiement de l'apprentissage profond dans le secteur de l'imagerie médicale. La première contribution majeure est la reconstruction des images médicales en vue de l'anonymisation en éliminant toutes les traces sur l'identité des patients et de garder toutes les informations significatives pour le diagnostic médical (voir chap. 3.6). Cette contribution nous permet d'accéder à une plus grande quantité de données tout en préservant la vie privée des individus. La revue des approches de meta-apprentissage et du mécanisme d'attention (voir chap. 2) constitue une des contributions dérivées de ce travail de thèse indiqué par l'ovale en orange dans la figure 6. Les techniques de meta-apprentissage permettent de résoudre la problématique de l'apprentissage avec peu de données, de favoriser une meilleure initialisation des paramètres et

d'exploiter les anciennes expériences pour résoudre les nouvelles tâches. Ainsi, nous réduisons non seulement le besoin en termes de données, mais aussi en termes de ressources de mémoire et de calcul, nous résolvons le problème de l'oubli catastrophique, nous limitons le problème écologique et nous permettons une meilleure accessibilité à l'apprentissage profond par tout le monde afin de relever tous les défis discutés ci-dessus. D'autre part, le mécanisme d'attention permet de mettre en relief des régions d'intérêt dans les données et donc minimiser la marge d'erreur, focaliser l'attention sur les informations les plus pertinentes, réduire le temps et la complexité de calcul, permettant ainsi de limiter encore les différents défis de l'apprentissage profond. La deuxième contribution majeure est l'élaboration d'une approche à l'intersection du meta-apprentissage et de l'attention pour résoudre le problème du manque de données, en particulier dans le contexte de maladies rares.

Ainsi, après ce chapitre introductif, le reste du manuscrit est organisé comme suit : le chapitre 1 présente l'état de l'art dédié à l'introduction de l'intelligence artificielle, de l'apprentissage automatique et de l'apprentissage profond. Nous nous intéressons surtout à ce dernier, ses formulations mathématiques, les notions en lien ainsi que les travaux les plus reconnus. Nous commençons par une brève histoire de l'intelligence artificielle et ses affiliations dans la section 1.1. Nous nous penchons sur l'apprentissage automatique et ses deux catégories principales dans la section 1.2. Par la suite, nous introduisons dans la section 1.3 l'apprentissage profond que nous utilisons tout au long de cette thèse, nous définissons les notions en lien et les métriques déployées pour l'évaluation. Dans la section 1.4, nous présentons la notion de l'image numérique et l'imagerie médicale en particulier. Nous détaillons ensuite l'évolution des modèles d'apprentissage profond pour les images et en particulier ceux dédiés aux images médicales.

Le chapitre 2 est un autre chapitre de revue de l'état de l'art qui aborde une problématique à laquelle nous sommes confrontés lors de l'utilisation de l'apprentissage profond : le manque de données. Nous commençons dans la section 2.1 par la présentation de ce problème et ses implications, nous élaborons une vue d'ensemble des approches permettent de lutter contre ce problème du manque de données. Nous définissons la problématique reconnue sous le nom de few-shot learning dans la section 2.2 et la notion de l'approche de meta-apprentissage dans

la section 2.3. Nous présentons ensuite un résumé des techniques en lien avec le few-shot learning y compris quelques techniques de meta-apprentissage dans la section 2.4. Nous organisons les techniques, en fonction de la stratégie d'apprentissage, en trois catégories principales : apprentissage par transfert, apprentissage de l'espace métrique et des solutions liées à l'algorithme d'optimisation. Nous détaillons l'état de l'art des méthodes, en examinant et en discutant leurs architectures et leurs résultats. Finalement, nous proposons dans la section 2.5 une étude de la littérature des techniques du mécanisme d'apprentissage. Par la suite, dans chaque chapitre, nous rappelons brièvement les travaux les plus récents relatifs au sujet traité et nous situons l'approche proposée parmi celles proposées dans l'état de l'art.

Le chapitre 3 s'intéresse à la première principale contribution. Nous commençons la première section 3.1 en présentant la motivation qui nous a menée à travailler dans cette direction. Ensuite, dans la section 3.2 nous annonçons le contexte général et les préoccupations en lien avec la confidentialité lors de l'utilisation de l'apprentissage profond. La section 3.3 s'intéresse à la description des données utilisées tout au long de ce chapitre. Le reste du chapitre contient deux sections principales : la première, 3.4, présente notre approche de l'anonymisation liée aux équipements d'acquisition de l'IRM et les expérimentations : le jeu de données, la classification en fonction des équipements d'acquisition de données, la reconstruction des images en vue de l'anonymisation, la reformulation mathématique de l'approche et la conclusion. La deuxième section, 3.5, a une structuration similaire de la section 3.4 avec un objectif d'anonymisation lié à l'identité du patient.

Le chapitre 4 s'organise comme suit : tout d'abord, nous commençons par une introduction générale du contexte et de la motivation dans la section 4.1. Ensuite, nous explorons dans une deuxième section 4.2 une revue de la littérature des algorithmes de meta-learning utilisée dans les travaux précédents dans le domaine de l'imagerie médicale. Puis, nous expliquons les deux approches proposées dans la section suivante 4.3. Dans la section 4.4, nous décrivons la base de données, nous mettons en œuvre le pipeline présenté dans la section précédente et nous présentons les expérimentations, les résultats et la conclusion.

Finalement, le chapitre suivant conclut cette thèse et fournit une liste de recommandations pour les orientations futures possibles et les améliorations du travail existant.

Chapitre

1 État de l'art : Intelligence artificielle

Sommaire

1.1 Histoire rapide de l'IA 3
1.2 Apprentissage automatique 6
1.2.1 Apprentissage supervisé 7
1.2.2 Apprentissage non supervisé 8
1.3 Apprentissage profond 8
1.3.1 Le réseau de neurones 9
1.3.2 Types de réseau de neurones 10
1.3.3 Fonctions d'erreur 16
1.3.4 La descente de gradient 21
1.3.5 Fonctions d'activation 22
1.3.6 Divers hyperparamètres, performance et généralisation du modèle 22
1.3.7 Métriques 29
1.4 L'apprentissage profond pour l'image 32
1.4.1 L'image aux yeux de la machine 33
1.4.2 L'imagerie médicale 35
1.4.3 L'évolution des réseaux convolutifs 37
1.4.4 Apprentissage profond pour l'imagerie médicale 40
1.4.5 Défis et perspectives 45

1.1 Histoire rapide de l'IA

En jetant un regard rétrospectif sur la longue histoire de l'avancement de l'humanité, nous pouvons considérer que l'intelligence artificielle (IA) est née en 1936 avec les premières bases mathématiques théoriques de la science informatique "computing machines" [259]. La théorie de la calculabilité représente la logique mathématique derrière les algorithmes informatiques. Cette théorie permet de concevoir une classe de problèmes informatiques pouvant être résolus à l'aide d'un nombre fini d'instructions et un espace défini de mémoire. Alan Turing, un mathématicien et cryptologue britannique, a établi le premier ordinateur imaginaire, également connu sous le nom de la machine de Turing. Cette machine fictive, selon la thèse de Church, est un concept mathématique abstrait qui permet de résoudre tous les algorithmes [42]. Elle permet aussi d'émettre des réponses fondamentales à la théorie de la décidabilité. Cette contribution était la théorie de base derrière le fonctionnement des premiers ordinateurs ENIAC (ELectonic Numerical Integrator And Computer) [159] [75]. En effet, Alan a aussi établi les premières notions liées aux algorithmes de l'intelligence artificielle [258] [257].

Dans la même période, John von Neumann, un mathématicien et physicien américain, s'inspire du fonctionnement du cerveau de l'être humain pour donner naissance à la théorie des jeux et du comportement économique [266] [265]. Son concept mathématique a permis d'optimiser l'espace mémoire des algorithmes et a amélioré le fonctionnement de ENIAC [46]. Ce fut aussi un prototype important pour les systèmes de l'intelligence artificielle.

En 1947, dans son rapport "Intelligent Machinery", Alan aborde l'exemple du jeu d'échecs pour juger si la machine peut se passer pour un être humain dans un jeu [257]. En 1950, Alan a conçu un test pour pouvoir évaluer l'intelligence de la machine dans le référentiel de l'être humain. Le test consiste à affecter une tâche habituellement effectuée par l'être humain à la machine [258]. Si l'être humain, en communiquant avec la machine dans le contexte de cette tâche, ne peut pas distinguer qu'il s'agit d'une machine, la machine est considérée intelligente. Pendant les 10 années qui suivaient, la communauté s'intéresse de plus en plus à ce concept d'intelligence artificielle définie par McCarthy comme la science et l'ingénierie de la fabrication

des machines intelligentes [297] [177]. Au fil des années, d'autres versions du test de Turing et d'autres éléments sont apparus importants pour la définition de l'intelligence de la machine.

En 1956, une conférence sur l'intelligence artificielle a eu lieu au Dartmouth College aux États-Unis [43][51]. Des personnages importants qui allaient influencer le domaine par la suite ont participé et assisté à cette conférence. Un projet a été initié par des chercheurs dans ce sens, notamment John McCarthy de l'université Stanford [267] [189] et Marvin Minsky de MIT [158][164]. Cette conférence était l'acte de naissance de l'IA comme un secteur de recherche autonome. Depuis, elle était incorporée dans tous les domaines et a prouvé sa performance dans des multiples tâches, notamment l'étude de climat [136], la robotique [84], la littérature [23], le trading [49], la médecine [87] et l'analyse de texte [48].

En revanche, toujours en analogie avec l'être humain, certaines tâches nécessitent un apprentissage. La machine intelligente doit développer et améliorer les performances à partir de ses expériences. Ainsi, l'apprentissage automatique émerge. Une machine capable d'apprendre est une machine plus intelligente, capable d'exécuter des tâches d'une façon plus efficace sans le guidage direct et l'intervention de l'être humain. En 1959, Arthur Samuel a défini l'apprentissage automatique en tant que la science qui permet à la machine d'apprendre, sans explicitement la programmer [214] [213].

Une nouvelle définition par Tom Mitchell en 1997 établie que l'apprentissage automatique est la science qui permet à une machine de s'améliorer automatiquement à chaque expérience [111] [167].

La définition moderne de Mitchell :

Une machine est dite capable d'apprendre d'une expérience E, respective à une classe de tâche T et une mesure de performance P, si P en T s'améliore après E.

Exemple : Jouer à l'échec

E = L'expérience de jouer à plusieurs parties d'échec

T = La tâche de jouer à l'échec

P = La probabilité que la machine gagne à la prochaine jeu

En 2003, Peter Norvig [1] définit l'intelligence artificielle comme la science de la création des machines intelligentes capables d'exécuter des tâches qui peuvent être réalisées par l'être humain d'une façon plus optimale et plus rapide [217] [207] [86]. Les tâches peuvent être un calcul, une tâche de contrôle, une tâche répétitive, une tâche gourmande en temps ou une tâche complexe [157][276].

Les scientifiques essayent de plus en plus de simuler la façon d'apprentissage de l'être humain. L'apprentissage profond est une branche de l'apprentissage automatique où nous utilisons des réseaux de neurones artificiels pour faire apprendre à la machine. Effectivement, le terme réseau de neurones est inspiré de la composition et du fonctionnement du cerveau humain. L'objectif est de s'approcher au mieux à la façon avec laquelle l'être humain apprend pour pouvoir reproduire des tâches réservées précédemment à l'humain. Les avancements dans le domaine informatique et l'augmentation des capacités du calcul aboutissent à une énorme progression des techniques d'apprentissage profond (fig.1.1).

En conclusion, si la science de l'intelligence artificielle (IA) est considérée la branche fille de la science informatique [134], nous pouvons considérer que le "machine learning" ou l'apprentissage automatique (ML) est une sous-branche et le "deep learning" ou l'apprentissage profond (DL) son arrière-petit-fils (fig.1.1).

Plusieurs études dans des multiples domaines ont réussi à faire naître l'un des outils les plus influents sur l'avancement de l'humanité. L'apprentissage automatique a procuré des nouvelles pistes importantes d'exploitations et d'explorations des données dans des multiples tâches complexes d'ingénieries dans des domaines variés qui ne pouvaient pas être effectuées manuellement.

1. directeur de recherche à Google et ancien responsable de la division de programmation de satellites et robots d'exploration chez NASA

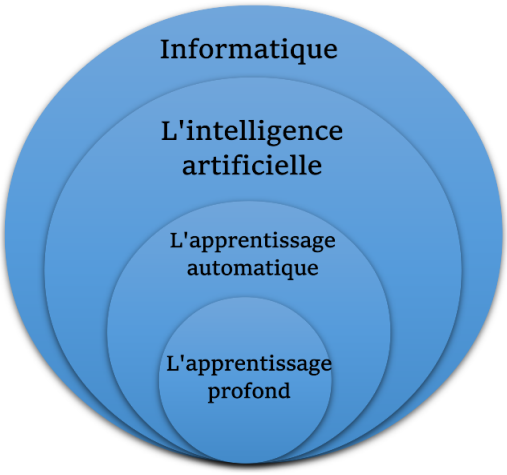


FIGURE 1.1 – Le diagramme hiérarchique des sous domaines de l'intelligence artificielle

1.2 Apprentissage automatique

La première étape d'un projet basé sur l'apprentissage automatique est l'acquisition et la préparation des données. L'étape de la récolte des données en quantité et qualité suffisantes est cruciale pour un bon apprentissage. Cela peut éviter des biais de représentativité et améliorer le modèle et les prédictions. Une autre étape primordiale est la préparation et le nettoyage des données recueillies avant le déploiement. En effet, des données peuvent être inutiles, incomplètes, endommagées ou à modifier afin d'être comprises par l'algorithme. Plusieurs approches sont proposées dans ce but, telles que la visualisation de données, le prétraitement de données ou la normalisation avec ses multiples variantes. Ensuite, la base de données se divise principalement en deux parties utilisées dans deux phases principales de l'apprentissage automatique (fig.1.2).

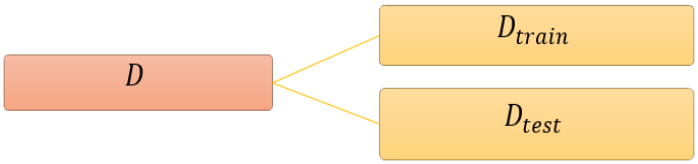


FIGURE 1.2 – Structure de la base de données

a) Une phase d'apprentissage

Cette phase, dite aussi phase d'entraînement, consiste en une étape de modélisation d'un problème et l'estimation de ses paramètres en utilisant la partie de la base de données réservée à l'entraînement. Cette phase peut inclure aussi une étape de validation pour laquelle une partie des données est réservée (voir section 1.3.6).

b) Une phase de test

Pendant cette phase, le modèle appris est testé sur la partie de la base de données non utilisée lors de l'apprentissage dite base de test et des métriques de performances sont mesurées (voir section 1.3.7) .

Après ces deux phases, le modèle est prêt à être déployé avec un nouveau jeu de données. Dans la pratique, la base de données est souvent divisée en trois parties : une partie pour l'entraînement, une partie pour la validation et une partie pour le test. Le jeu de données de validation, similairement au jeu de données du test, n'est pas observé par le modèle. Les données de validation proviennent de la même distribution que le jeu de donnée de l'entraînement et elles sont utilisé dans la phase d'apprentissage. Elles sont utilisées pour estimer l'erreur de généralisation du modèle et à valider la progression du processus de l'apprentissage.

L'apprentissage automatique peut être classé en deux grandes branches :

1.2.1 Apprentissage supervisé

Dans cette branche d'apprentissage, nous possédons des données d'entrée et nous connaissons la nature/les classes de la sortie. Ainsi, pour ce type d'apprentissage, les données doivent être labellisées. Les tâches d'apprentissage supervisé peuvent être divisées comme des problèmes de "régression" ou de "classification". Dans le cadre d'une régression, l'objectif est de prédire une sortie à valeur continue, ce qui signifie faire correspondre des variables d'entrée à une fonction définie continue. Dans le cadre d'une classification, les résultats sont plutôt sous la forme d'une sortie discrète (par exemple : 0 ou 1). En d'autres termes, l'objectif est de

faire correspondre les variables d'entrée à des catégories discrètes. Nous pouvons formaliser le problème d'apprentissage automatique supervisé mathématiquement en tant qu'un problème d'estimation ou de modélisation. L'algorithme cherche à trouver une fonction f telle que $Y = f(X)$, où X représente l'entrée et Y représente la sortie. Il existe plusieurs techniques classiques connues d'apprentissage supervisé telles que la régression logistique (LR) [40], l'arbre de décision (DT) [25], la forêt aléatoire (RF) [13], les K-voisins les plus proches (KNN) [3] et les machines à vecteurs de support (SVM) [110].

1.2.2 Apprentissage non supervisé

Dans le cadre de l'apprentissage non supervisé, nous utilisons des données d'entrée non labellisées. La sortie est déduite sur la base des relations et des corrélations entre les variables des données d'entrée. Les tâches d'apprentissage non supervisé peuvent être divisées principalement comme des problèmes de partitionnement de données ou des problèmes de regroupement (clustering).

1.3 Apprentissage profond

Certes, l'apprentissage automatique était déjà un point d'inflexion dans l'histoire de la science et la technologie. Cependant, l'avancement le plus important a eu lieu avec l'apparition de l'apprentissage profond et l'extraction automatique des caractéristiques des données. En effet, l'étape classique d'extraction manuelle des caractéristiques est une opération cruciale pour permettre aux algorithmes à prendre les meilleures décisions d'une façon optimale. Les résultats dépendaient principalement de la qualité des caractéristiques conçues [139]. Cependant, accomplir la tâche manuellement consomme beaucoup du temps et d'efforts. À ceci s'ajoute la subjectivité et la fatigue humaines. L'apprentissage profond permet d'apprendre automatiquement des caractéristiques de plus en plus élevées d'une couche à une autre et a montré des performances records dans l'extraction des caractéristiques de segmentation et de classification.

1.3.1 Le réseau de neurones

Le réseau de neurones sert à extraire automatiquement des caractéristiques connexes et à découvrir des relations sémantiques latentes significatives entre les caractéristiques des données [54]. L'objectif est de déduire un modèle ou une fonction qui permet de prendre les bonnes décisions dans une certaine tâche.

La structure d'un réseau de neurones artificiel par analogie au réseau de neurones biologique se base sur les neurones. Pour comprendre le concept de l'apprentissage profond, nous détaillons le fonctionnement des neurones. En effet, le neurone artificiel reçoit des données d'entrée assimilées à des pulsations électrochimiques reçues par les dendrites d'un neurone biologique. Une opération, prenant en considération la pondération de l'entrée, est effectuée à l'aide d'une fonction d'activation (fig. 1.3). Les étapes sont les suivantes :

1. Additionner la multiplication de toutes les entrées x_i et leurs poids respectifs w_i : $\sum_i w_i x_i$.
2. Additionner le total obtenu avec le biais : $z = \sum_i w_i x_i + b$.
3. Appliquer une fonction d'activation (voir section 1.3.5).

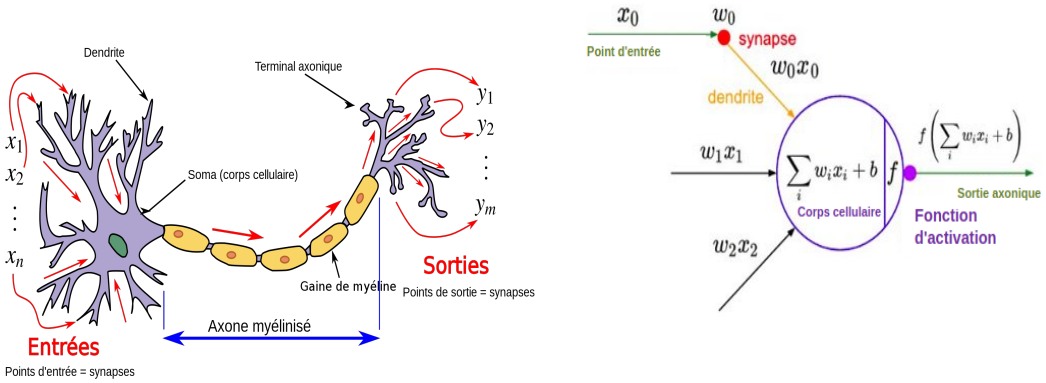


FIGURE 1.3 – Comparaisons entre un neurone biologique et un neurone artificiel [70][135]

Le réseau de neurones de l'apprentissage profond est constitué typiquement de neurones, présents comme des nœuds, organisés en couches. La structure de base d'un réseau neuronal se compose d'une couche d'entrée, une couche de sortie et des couches cachées (figure 1.4).

Les paramètres comprenant les poids et les biais représentés respectivement par w et b dans la figure 1.3 sont ajustés pendant le processus d'apprentissage de manière à minimiser une

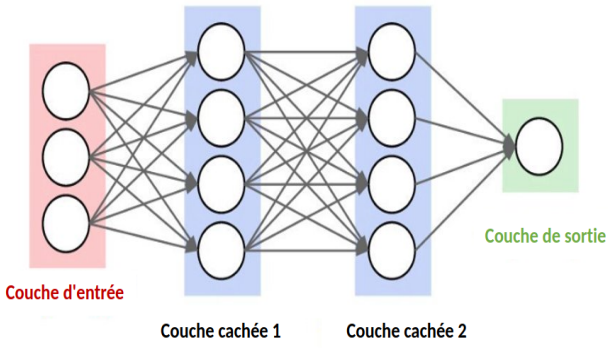


FIGURE 1.4 – Exemple d'un réseau de neurones [70]

fonction de coût permettant de mesurer l'erreur de la prédiction commise par le réseau (voir section 1.3.3). Les étapes incluses dans l'apprentissage sont les suivantes :

1. Introduire les données et lancer le modèle pour obtenir les prédictions.
2. Calculer la fonction de coût.
3. Déployer l'algorithme d'optimisation pour ajuster les poids du modèle afin d'améliorer les prédictions. Cette étape de mise à jour des poids est appelée rétro-propagation ou "back-propagation".
4. Répéter les étapes 1 à 4, représentant un seul cycle ou une époque d'apprentissage, jusqu'à atteindre un certain nombre d'époques ou une certaine condition d'arrêt liée à la valeur de la fonction de coût.

Le nombre d'époques est l'un des hyperparamètres à définir pour l'apprentissage. À chaque cycle, la fonction de coût devrait théoriquement se réduire et le réseau commence ainsi à produire des prédictions plus proches des valeurs cibles.

1.3.2 Types de réseau de neurones

Le réseau de neurone décrit dans la section précédente n'est adapté pour les tâches complexes. Des améliorations ont donné naissance à plusieurs architectures. Les deux les plus connus sont : les réseaux de neurone convolutifs (CNN) et les réseaux de neurones récurrents (RNN). Le tableau 1.1 résume les différences entre ces deux types de réseaux.

	CNN	RNN
Architecture	Réseau de neurones à direction direct (feedforward), fait intervenir des champs réceptifs, des filtres et des opérations de pooling.	Réseau récurant qui s'alimente par la sortie
Type de données	Données spatiales telles que des images	Données temporelles ou séquentielles telles que de la vidéo ou du texte
Sortie	La taille de la sortie est fixée à l'avance	La taille de la sortie peut varier
Cas d'utilisation	Reconnaissance faciale, diagnostic de l'imagerie médicale, analyse des images, classification des images, segmentation et détection des objets	Traduction de texte, Traitement du langage naturel (NLP), compréhension du langage naturel (NLU), reconnaissance vocale, analyse de la parole, analyse sentimentale

TABLE 1.1 – Résumé de différences entre CNN et RNN inspiré de [187]

Dans notre travail, nous nous intéressons à l'imagerie en particulier. L'architecture la plus utilisée pour l'image est les CNN [18]. Les CNN se sont montrés performants dans de nombreuses tâches de vision par ordinateur telles que la reconnaissance, la détection et la segmentation d'objets. Donc, dans la suite de cette sous-section, nous allons détailler le fonctionnement des CNN (fig. 1.5).

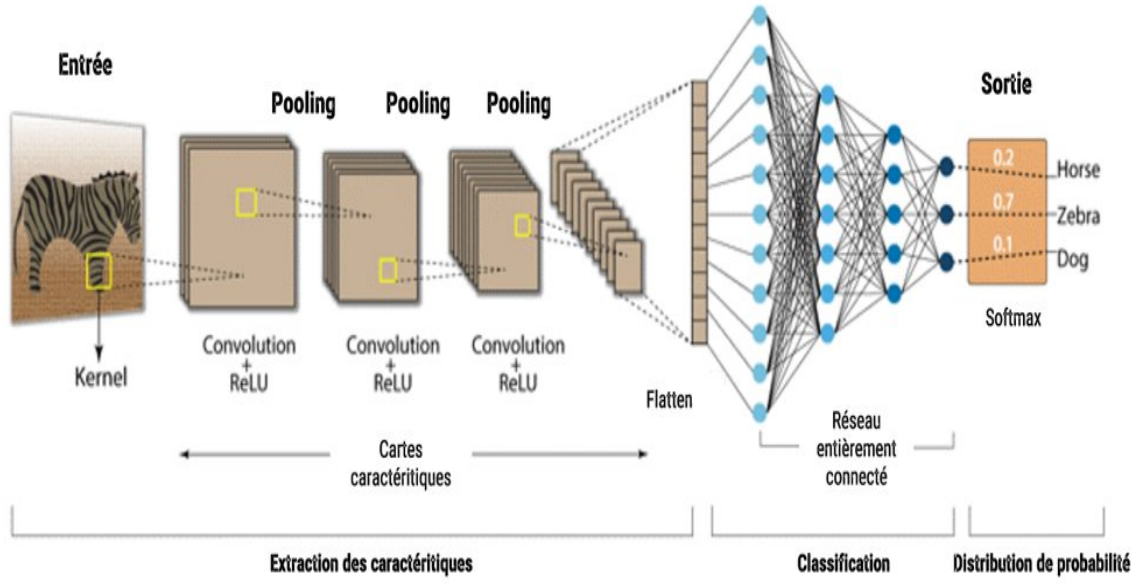


FIGURE 1.5 – Exemple de la structure de CNN [229]

Un bloc CNN effectue principalement 3 types d'opérations à une image :

— Convolution

— Pooling

— ReLU

Nous définissons chacune de ses opérations.

a) La convolution

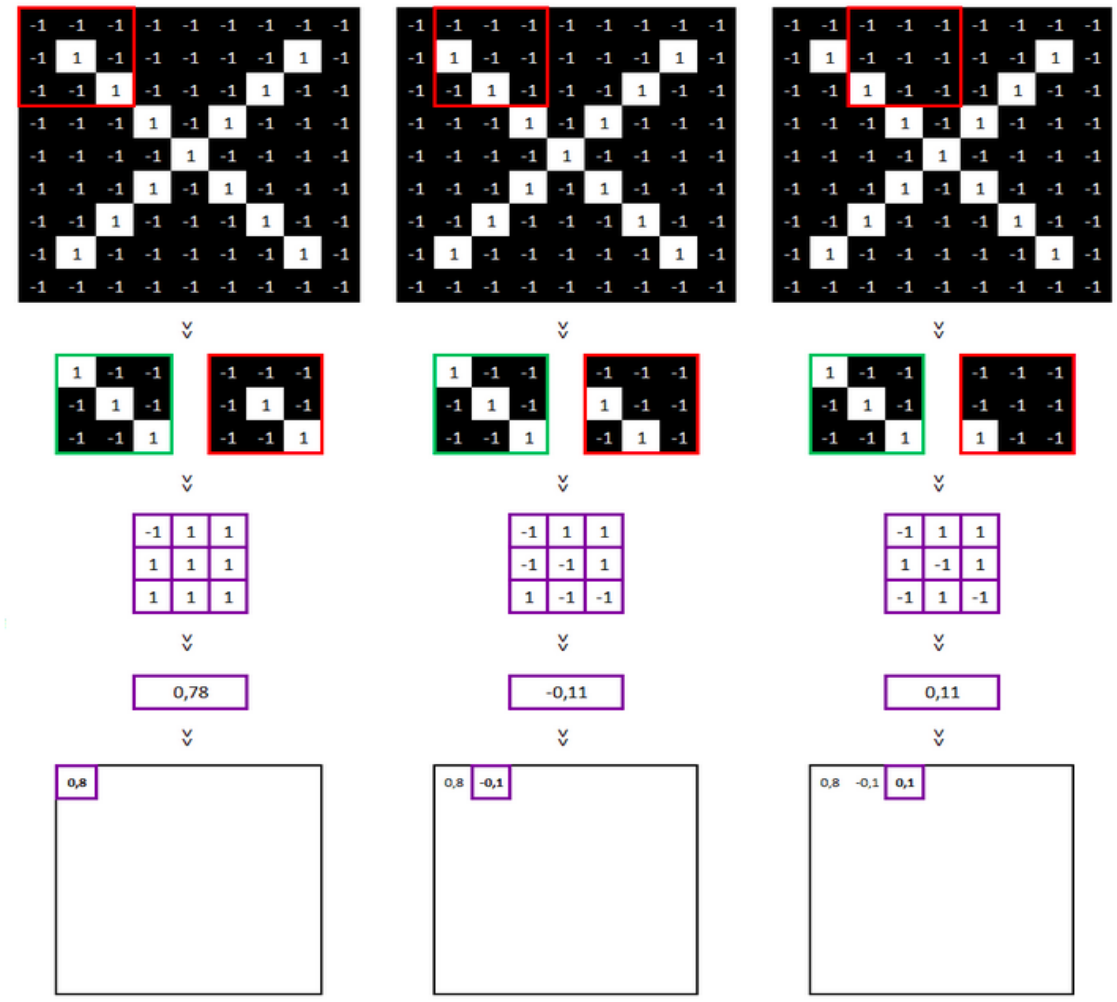


FIGURE 1.6 – Exemple d'une convolution par un seul filtre [260]

La convolution est une opération mathématique simple. Elle est effectuée en multipliant le filtre par une partie de l'image et en sommant le résultat. La partie de l'image cible est

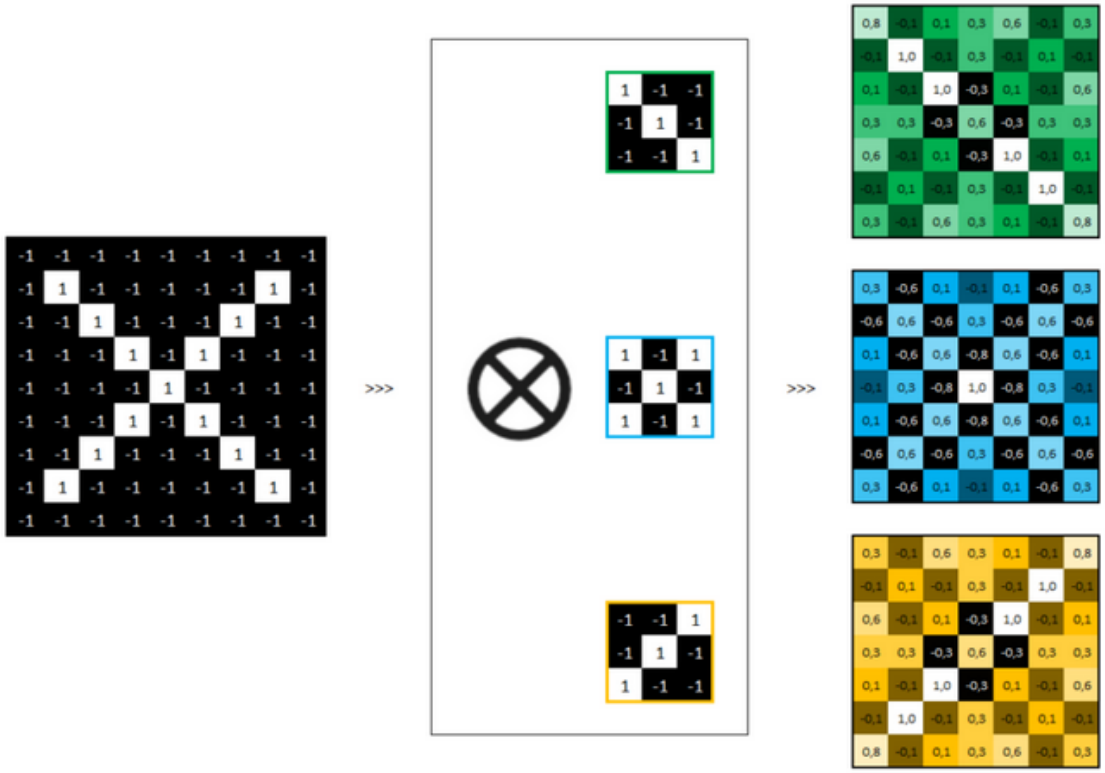


FIGURE 1.7 – Exemples de cartes de caractéristiques résultantes de la convolution par trois filtres [260]

appelée le champ réceptif. L'avantage de cette approche est que les neurones d'une couche s'intéressent particulièrement à l'extraction des caractéristiques visuelles dans une petite portion de l'image. L'objectif est de repérer des caractéristiques appelées aussi un feature, un filtre ou un noyau convolutif (kernel) dans les images d'entrée. Le principe est de faire "glisser" le filtre, de calculer le produit de convolution entre le filtre et la région de l'image balayée (fig. 1.6) et produire les cartes caractéristiques connues aussi sous le nom des cartes d'activations. Les cartes caractéristiques résultantes nous indique l'emplacement des features dans l'image : plus la valeur est élevée en une partie de l'image, plus cette partie ressemble au feature (fig. 1.7). Contrairement aux techniques classiques, les features ne sont pas prédéfinies en avance, mais apprises par le réseau lors la phase d'entraînement.

Le choix de la taille de filtre est un paramètre à préciser. La convolution inclut d'autres paramètres tels que le stride et le padding avec le filtre. Le stride correspond à la taille du pas de convolution pour déplacer les filtres. L'augmentation des strides réduit donc la taille de la sortie. Le padding fait référence à des zéros ajoutés aux données d'entrée afin que la taille de sortie de

la couche soit la même que celle de l'entrée. Il permet au filtre d'atteindre le bord de l'image en s'adaptant au pas de la convolution.

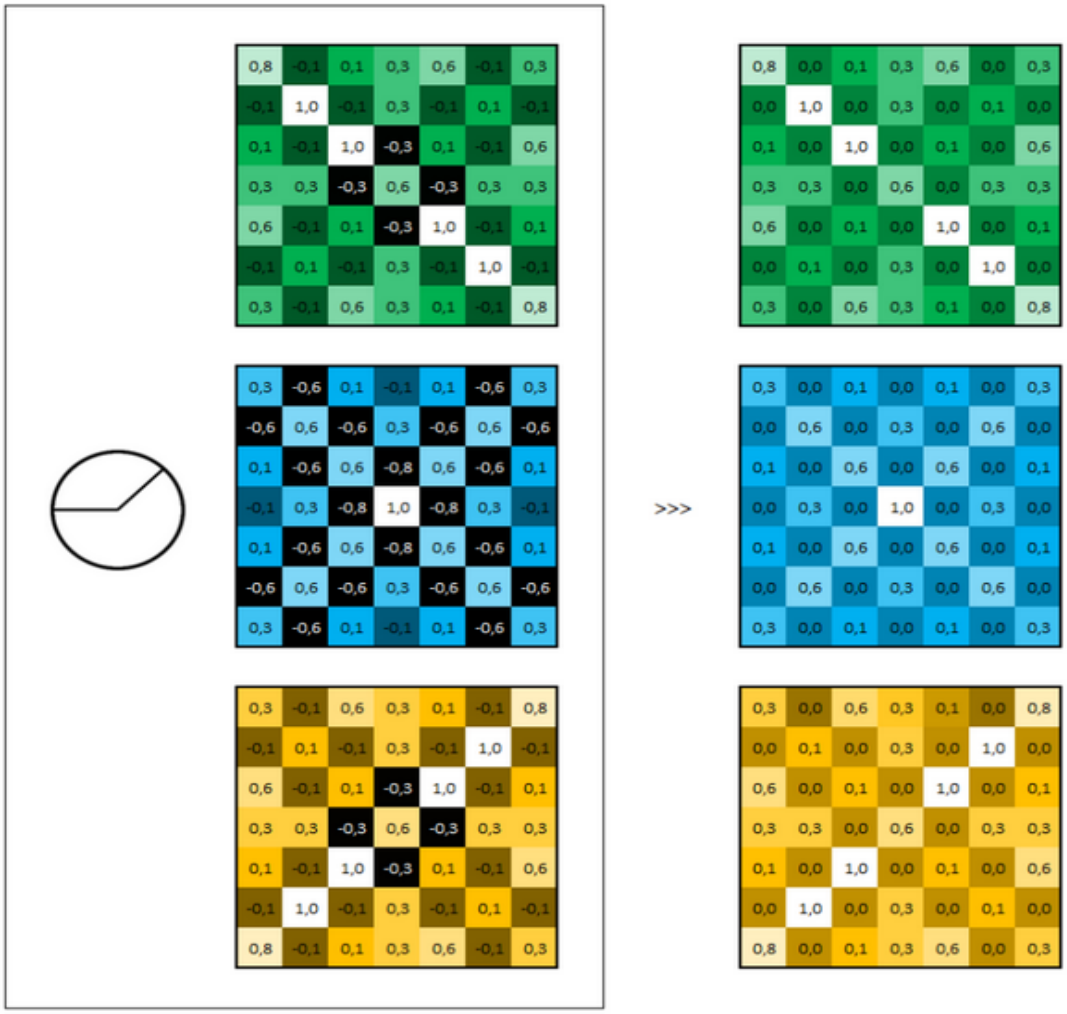


FIGURE 1.8 – Les cartes caractéristiques résultantes suite à une ReLU [260]

b) Unité linéaire rectifiée (ReLU)

La fonction ReLU définie dans la section 1.3.5 permet de transformer les valeurs négatives en 0 (fig. 1.8).

c) Le pooling

Ils existent deux techniques principales de pooling : le max-pooling (prend la valeur maximale de chaque portion de l'image en entrée) ou le average-pooling (prend la médiane

de chaque portion). Le max-pooling est le plus répandu dans les cas d'utilisation classique des CNN (fig. 1.9). La taille de la portion de l'image est à préciser.

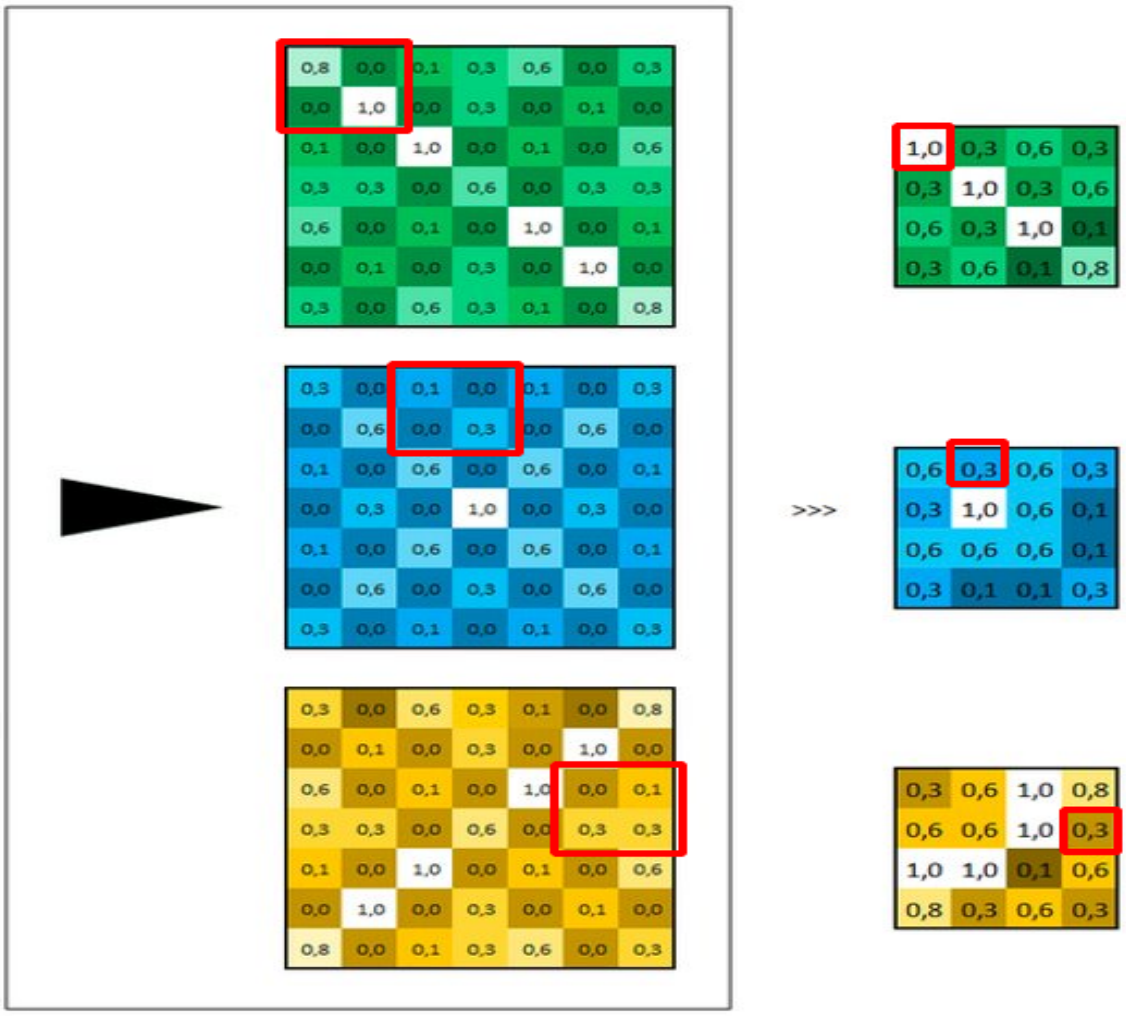


FIGURE 1.9 – Les cartes caractéristiques résultantes suite à une opération de max-pooling de taille 2x2 [260]

d) Couche d'aplatissement

La couche d'aplatissement appelée Flatten prend la totalité des valeurs de nos matrices précédemment calculées, et à les empiler, en vue de les exploiter dans la couche d'entrée d'un réseau de neurones multicouche entièrement connecté de structure classique. Le reste de réseau permet d'aboutir à des probabilités d'appartenance pour prédire la classe à laquelle l'image d'entrée appartient.

L'objectif est d'apprendre à extraire les caractéristiques de l'image les plus utiles pour minimiser une certaine fonction d'erreur (voir la section suivante) afin d'effectuer une tâche cible. Les réseaux de neurones convolutifs apprennent plusieurs filtres et donc plusieurs caractéristiques en parallèle pour une image en entrée. Classiquement, le nombre de filtres par couche de convolution est entre 32 à 1024 filtres en parallèle. Les premiers filtres permettent d'extraire des caractéristiques de bas niveau comme les lignes, les bords et les contours. Les filtres dans les couches les plus profondes permettent d'extraire des caractéristiques de plus haut niveau comme les formes.

1.3.3 Fonctions d'erreur

L'objectif ultime de la phase d'apprentissage est de résoudre un problème d'optimisation de la fonction de perte, appelée aussi la fonction de coût ou la fonction d'erreur. Cette fonction est utilisée pour arbitrer la performance d'un modèle lors de l'apprentissage. Elle mesure la ressemblance entre les prédictions du modèle et les valeurs cibles et décide de l'ajustement à effectuer sur les poids par l'algorithme d'apprentissage du réseau de neurones à chaque itération. Plusieurs fonctions de perte sont utilisées communément dans les réseaux de neurones. Ces fonctions incorporent toujours la notion de mesure des distances entre les sorties du modèle correspondant aux prédictions du réseau et les valeurs cibles données en entrée [198]. Dans le reste de cette partie, nous présentons les fonctions d'erreur les plus utilisées dans les réseaux de neurones.

1.3.3.1 Erreur absolue moyenne

L'erreur absolue moyenne (Mean Absolute Error (MAE)) est appelée aussi fonction de perte L1. Elle consiste à calculer la somme moyenne des valeurs absolues des différences entre les valeurs cibles et les prédictions du réseau de neurones.

$$E_{MAE} = 1/N \sum_{i=1}^{N} |y_{p_i} - y_i| \tag{1.1}$$

avec

$$\begin{cases} y_i & \text{la ième valeur cible} \\ y_{p_i} & \text{la ième valeur prédite} \\ N & \text{le nombre de données de l'apprentissage} \end{cases}$$

L'erreur absolue moyenne n'est pas sensible aux valeurs extrêmes et peut engendrer des problèmes de convergence [101]. Elle est utilisée souvent pour les problématiques de régression.

1.3.3.2 Erreur quadratique moyenne

L'erreur quadratique moyenne (Mean Squared Error (MSE)) est appelée aussi fonction de perte L2. Elle consiste à calculer la somme moyenne des carrées des différences entre les valeurs cibles et les prédictions du réseau de neurones :

$$E_{MSE} = 1/N \sum_{i=1}^{N} (y_{p_i} - y_i)^2 \tag{1.2}$$

L'erreur quadratique moyenne est sensible aux valeurs extrêmes. Elle est souvent utilisée comme fonction d'erreur dans les problématiques de régression, surtout lorsqu'on observe une distribution normale des données.

1.3.3.3 Entropie croisée

La combinaison de cette fonction de perte avec la fonction d'activation notamment softmax permet l'interprétation des sorties d'un modèle comme probabilités d'appartenance à une classe [78].

$$E_{sc} = -\sum_{i=1}^{N} y_i log(y_{p_i}) \tag{1.3}$$

Les valeurs de probabilités de cette problématique sont dans l'intervalle [0,1] et la valeur de log sur cet intervalle est négatives. Par conséquent, pour avoir une valeur positive de la fonction d'erreur, un signe moins est ajouté à l'équation.

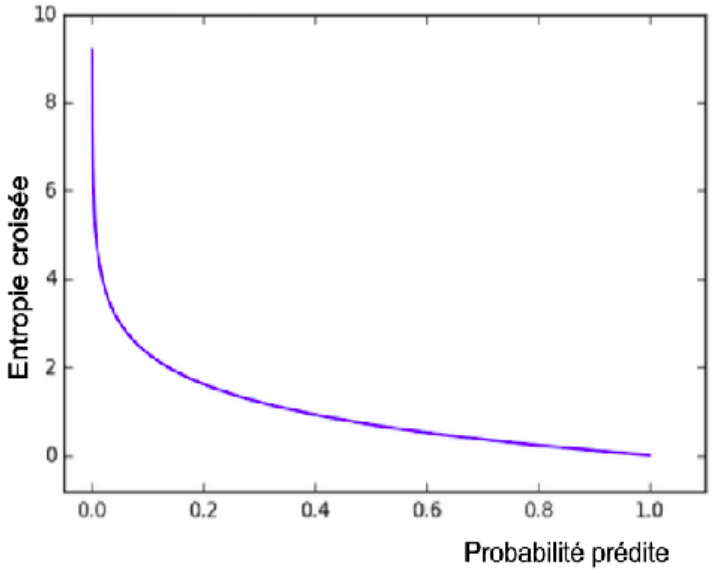


FIGURE 1.10 – Valeur de l'erreur entropie croisée lorsque la probabilité cible est égale à 1 inspiré de [35]

Cette fonction d'erreur est utilisée souvent pour les problématiques de classification. Elle a amélioré nettement la performance des modèles avec les fonctions d'activation sigmoïde and softmax en comparaison avec la MSE. L'entropie croisée permet une convergence stable et rapide.

a) Entropie croisée binaire

L'entropie croisée binaire (Binary Cross-Entropy (BCE)) est une fonction d'erreur utilisée pour des tâches de classification binaire (des tâches avec deux classes seulement).

$$E_{BCE} = -\sum_{i=1}^{N}(y_i log(y_{p_i}) + (1 - y_i) log(1 - y_{p_i})) \tag{1.4}$$

avec

$$\begin{cases} y_i & \text{la ième valeur cible} \\ y_{p_i} & \text{la ième valeur prédite} \\ N & \text{le nombre de données de l'apprentissage} \end{cases}$$

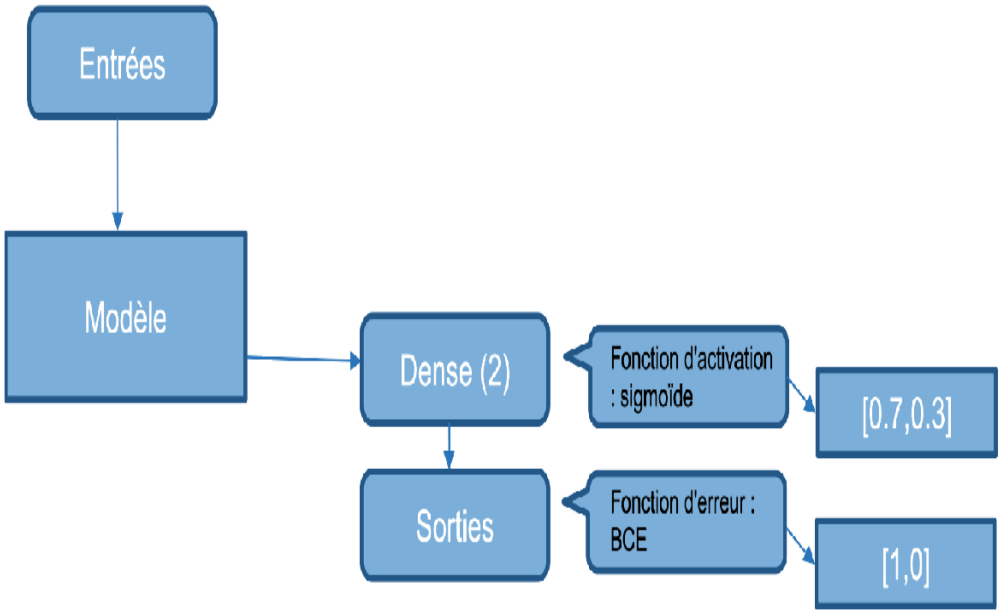


FIGURE 1.11 – Processus d'apprentissage pour une classification binaire

b) Entropie croisée catégorique

L'entropie croisée catégorique (Categorical Cross-Entropy (CCE)) est une fonction d'erreur utilisée pour des tâches de classification multiclasse. Ce sont des tâches avec plusieurs classes, mais un individu peut appartenir seulement à une seule classe.

$$E_{CCE} = -\sum_{i=1}^{N}\sum_{c=1}^{M} y_{i,c} log(y_{p_{i,c}}) \tag{1.5}$$

avec

$\begin{cases} y_{i,c} & \text{la valeur de la probabilité cible d'appartenance de l'individu i à la classe c} \\ y_{p_{i,c}} & \text{la valeur de la probabilité prédite d'appartenance de l'individu i à la classe c} \\ M & \text{le nombre de classe} \\ N & \text{le nombre de données de l'apprentissage} \end{cases}$

c) Entropie croisée catégorique éparse

L'entropie croisée catégorique éparse (Sparse Categorical Cross-Entropy (SCCE)) suit le mème principe que l'entropie croisée catégorique. La différence est la forme/type de donnée. Par exemple pour une classification à 3 classes :

— Dans le cadre de l'entropie croisée catégorique à 3 classes, y_i aurait l'une des valeurs suivantes : [1,0,0], [0,1,0], [0,0,1].

— Dans le cadre de l'entropie croisée catégorique éparse à 3 classes, y_i aurait l'une des valeurs suivantes : 1, 2, 3.

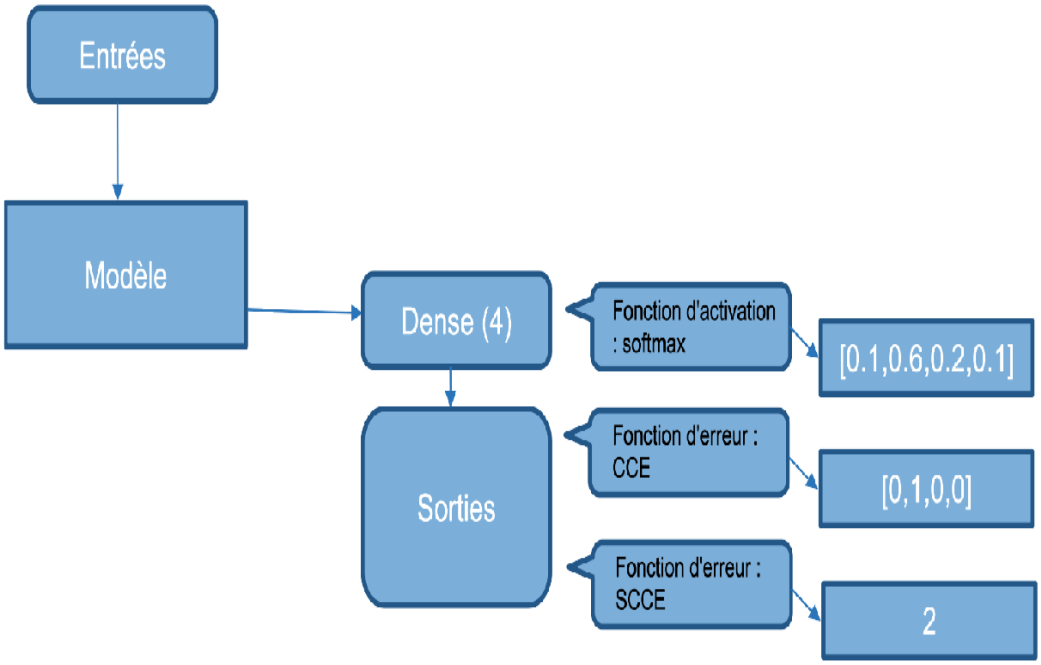


FIGURE 1.12 – Processus d'apprentissage pour une classification multiclasse

L'usage dépend seulement des données et de la façon avec laquelle nous voulons les introduire à notre réseau de neurones. En revanche, nous devons noter que l'entropie croisée catégorique éparse est moins gourmande en termes de temps et de mémoire de calcul puisqu'elle utilise des valeurs entières comme entrées contrairement à l'entropie croisée catégorique qui utilise des vecteurs en entrée.

1.3.4 La descente de gradient

L'objectif principal de l'apprentissage profond peut se formuler comme un problème de minimisation d'une fonction de coût. En revanche, résoudre un tel problème avec des millions de paramètres n'est pas simple. La plupart des algorithmes d'apprentissage utilise la descente de gradient qui représente une approximation utilisant une approche itérative.

Trouver le minimum de la fonction de coût nécessite des opérations différentielles. L'algorithme de la descente de gradient est l'algorithme d'optimisation le plus fréquent dans ce cadre. Il s'agit d'un algorithme qui minimise une fonction en prenant à chaque itération la direction de la plus grande descente.

La fonction d'optimisation $J(\Theta)$ est paramétrée par les poids qui renseignent sur l'importance de la connexion entre les neurones. L'algorithme consiste à calculer le gradient de cette fonction d'erreur respectivement à chaque poids et mettre à jour ces poids de façon à minimiser au mieux la fonction d'erreur.

L'équation classique de l'algorithme est la suivante :

$$\Theta = \Theta - \mu \bigtriangledown_{\Theta} J(\Theta) \tag{1.6}$$

où $\mu \bigtriangledown_{\Theta} J(\Theta)$ est le terme incrémental dans la direction du gradient descendant. $\bigtriangledown_{\Theta}$ définit le gradient relatif aux poids et μ le pas d'adaptation, un petit nombre positif appelé taux d'apprentissage, déduit des valeurs propres de la matrice composée par les données [102].

L'algorithme de descente de gradient a progressé au fil du temps et plusieurs variantes sont désormais disponibles dans des outils informatiques et des plateformes en ligne, y compris TensorFlow [2]. Le choix entre ces variantes dépend de plusieurs facteurs. Il s'agit d'un algorithme itératif et adaptatif pour atteindre le minimum de la fonction cible, dans notre cas la fonction d'erreur. Le vrai challenge est de définir la trajectoire optimale qui nous conduit à la solution. La mise à jour des paramètres, d'une itération à l'autre, renseigne sur l'exactitude et la rapidité de la convergence.

2. https://www.tensorflow.org/?hl=fr

L'approche classique de la descente de gradient est lente puisqu'il s'agit d'une unique mise à jour des paramètres, d'où l'introduction de la descente de gradient stochastique SGD. SGD est toujours un algorithme itératif où la différentiation se fait par rapport à chaque entrée x^i et classe y^i :

$$\Theta = \Theta - \mu \bigtriangledown_\Theta J(\Theta; x^i; y^i) \tag{1.7}$$

Bien que cet algorithme soit amélioré, il peut ne pas converger et rester dans un état d'oscillation autour de la solution. La nouvelle approche proposée consiste à passer à un mini-lot SGD ((mini-batch SGD) :

$$\Theta = \Theta - \mu \bigtriangledown_\Theta J(\Theta; x^{i:i+n}; y^{i:i+n}) \tag{1.8}$$

1.3.5 Fonctions d'activation

Les neurones du réseau fonctionnent à travers des fonctions d'activation qui permettent d'effectuer une opération sur les données provenant des neurones de la couche précédente à l'aide d'une fonction mathématique. Cette dernière peut influencer nettement la performance du réseau. Ainsi, le choix de la fonction d'activation des neurones est important.

Les neurones de la couche d'entrée intègrent généralement la fonction identité comme fonction d'entrée. En effet, seulement une sommation pondérée est effectuée au niveau de chaque neurone avant de transmettre les données aux neurones de la couche suivante. Le choix de la fonction d'activation des neurones de la couche de sortie dépend souvent de la tâche finale du réseau de neurones (voir Tab.1.2). Notamment, pour des problématiques de classification, comme explicité dans la section 1.3.3, la fonction softmax est la plus adaptée [22].

1.3.6 Divers hyperparamètres, performance et généralisation du modèle

Bien que les réseaux de neurones ne puissent pas être considérés récents (voir section 1.1), ils ont connu récemment des développements dynamiques et leur utilisation est devenue

Fonction d'activation	Formule	Cas d'utilisation	Intervalle
Identité	x	La sortie du neurone activé est la même que l'entrée. Pas d'opération effectuée sur l'entrée.	$[-\infty, +\infty]$
Sigmoïde	$1/(1+e^{-x})$	La sigmoïde logistique prend en entrée la valeur réelle et la normalise entre [0,1] pour produire une sortie probabiliste.	$[0, 1]$
Tanh	$2/(1+e^{-2x})-1$	La courbe de la tangente hyperbolique est similaire à celle de la sigmoïde logistique. Elle est parfois plus performante grâce à sa symétrie, utilisée surtout pour les réseaux complexes multicouches.	$[-1, 1]$
ReLU	$max(0, x)$	Unité de rectification linéaire peut accélérer la convergence du gradient stochastique (voir section 1.3.4) et réduit la complexité par rapport à la fonction sigmoïde ou tanh grâce à sa formule simple [282]. En revanche, son utilisation peut rencontrer un problème d'activation, surtout si le taux d'apprentissage est élevé [102].	$[0, +\infty]$
Softmax	$e^{x_i} / \sum_j e^{x_j}$	x_i est un élément du vecteur d'entrée et $\sum_j e^{x_j}$ est la somme de toutes les fonctions exponentielles de toutes les entrées. Cela garantit que la valeur de la sortie doit dans $[0, 1]$ et que la somme soit égale à 1, constituant ainsi une distribution de probabilité bien définie (voir section 2.4.1). Utilisée surtout pour les tâches de classification en normalisant le résultat et produisant des sorties probabilistes.	$[0, 1]$

TABLE 1.2 – Fonctions d'activation

beaucoup plus répandue de nos jours grâce à la présence des outils de calcul parallèle, notamment les cartes GPU, permettant des calculs plus rapides et moins chers. En parallèle, la capacité de stockage s'est améliorée, permettant d'avoir une plus grande quantité de données pour améliorer l'apprentissage. En effet, plusieurs paramètres entrent en jeu pour aboutir à un

apprentissage optimal. Le choix de l'architecture, comprenant le nombre de couches, le nombre de neurones par couches, le type du réseau, l'initialisation des poids, la fonction d'activation et la fonction d'erreur, est critique pour permettre le meilleur apprentissage [14]. L'algorithme d'optimisation et ses hyperparamètres, notamment le taux d'apprentissage, interviennent majoritairement dans le processus d'apprentissage [102][243]. Cependant, la performance d'un apprentissage ne peut pas réellement être mesurée pendant la phase d'entraînement, mais plutôt pendant la phase de test sur un nouveau jeu de données. Cette section est conçue pour aborder ces problématiques en détaillant l'influence de quelques hypermètres d'un réseau de neurones sur l'apprentissage et en présentant l'état de l'art des techniques favorisant la généralisation des modèles.

1.3.6.1 Hyperparamètres du modèle

Différentes expérimentations sont effectuées dans ce sens pour étudier l'influence des multiples hyperparamètres sur la performance de l'apprentissage [106].

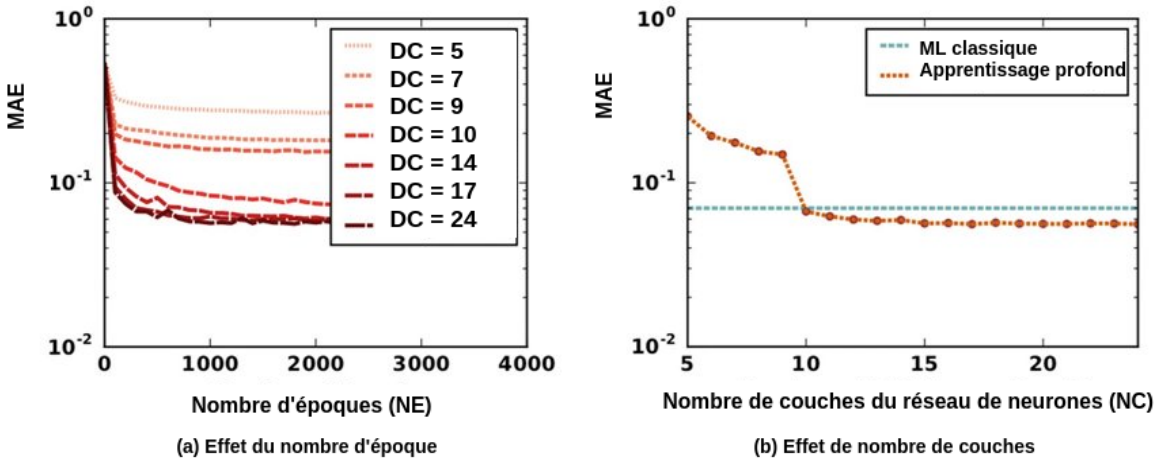


FIGURE 1.13 – Influence du nombre de couches et du nombre d'époques d'entraînement sur la valeur de l'erreur du modèle, inspirée de [106]. (a) représente les valeurs de la MAE en fonction du nombre d'époques pour des profondeurs différentes du réseau de neurones. (b) montre l'influence directe du nombre de couches sur la MAE du modèle.

Jha et al. [106] mettent en évidence l'importance du choix du nombre d'époques d'entraînement et du nombre de couches du réseau de neurones (fig. 1.13). En effet, plus le nombre de couches et le nombre d'époques sont élevés, plus la MAE diminue. Notamment,

l'apprentissage profond, dans cette expérimentation, a dépassé les performances des méthodes classiques de l'apprentissage automatique à partir de l'utilisation de 10 couches. En revanche, nous remarquons qu'à partir d'une certaine époque et à partir d'un certain nombre de couches, la MAE est plutôt stable.

Nous pouvons aussi observer clairement l'effet du choix du nombre de neurones dans la fig. 1.14. Il s'agit d'un réseau de neurones à une seule couche cachée entrainé pour une tâche de classification binaire de couleurs rouge et vert. Nous pouvons observer que plus le nombre de neurones est important, plus la classification est meilleure. Cependant, cet étroit ajustement engendre une fonction de modélisation complexe et peut provoquer un sur-ajustement ou un sur-apprentissage (voir section 1.3.6.2).

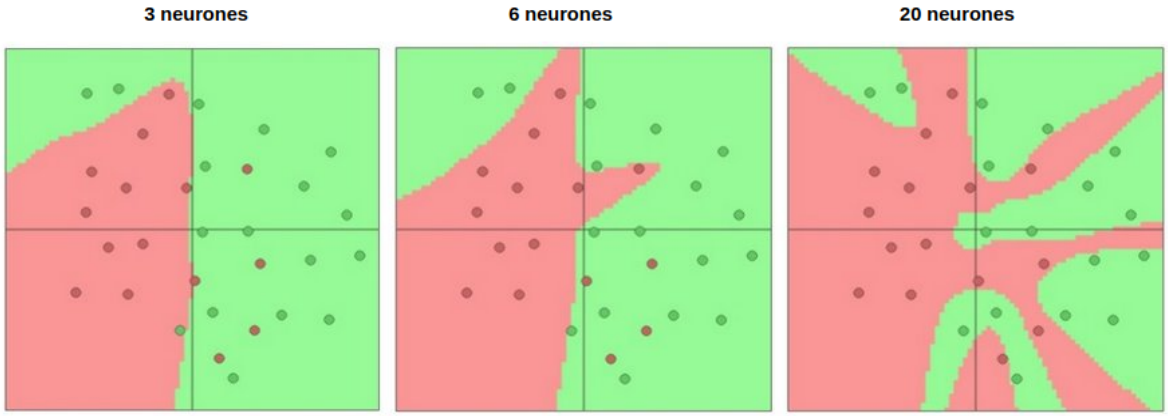


FIGURE 1.14 – Influence de nombre de neurones sur l'entraînement [44]

Pour conclure, le bon choix des hyperparamètres est nécessaire pour optimiser l'apprentissage. Cependant, fixer ces hyperparamètres n'est pas toujours évident et de nos connaissances, il n'y a toujours pas des règles prédéfinies dans ces sens [146][65]. En effet, c'est souvent par tâtonnement et en se basant sur les expérimentations de l'état de l'art qu'on peut décider des paramètres à choisir.

1.3.6.2 Performance et généralisation du modèle

La performance d'un modèle se mesure par des métriques calculées en utilisant un jeu de données inconnu, n'ayant pas été introduite au réseau lors de la phase d'apprentissage (voir section 1.3.7). Ces métriques mesurent principalement la généralisation du modèle sur

ces données de test. Deux problématiques sont souvent rencontrées : le sous-ajustement et le sur-ajustement aux données d'apprentissage, appelé aussi le sur-apprentissage. Pour mieux comprendre la fig 1.15, nous devons définir le risque empirique qui désigne la moyenne de la fonction d'erreur sur l'ensemble de données. L'objectif ultime est de minimiser ce risque empirique. La fonction du risque empirique fait intervenir deux types d'erreur : une erreur d'approximation (biais) et une erreur d'estimation (variance). Il est généralement impossible minimiser les deux erreurs simultanément. Donc, nous cherchons un compromis entre les deux dans un point de complexité optimale (fig. 1.16). Ces phénomènes d'apprentissage peut engendrer une difficulté à faire les bonnes prédictions pour de nouvelles données. En effet, minimiser la fonction d'erreur d'apprentissage sur les données d'entraînement n'est pas l'objectif ultime de l'apprentissage profond. Le véritable objectif consiste à obtenir un modèle permettant la meilleure performance de prédiction sur des nouveaux jeux de données. Par conséquent, lors de la phase d'apprentissage, nous devons s'intéresser non seulement à améliorer la performance, mais aussi à généraliser le modèle [89].

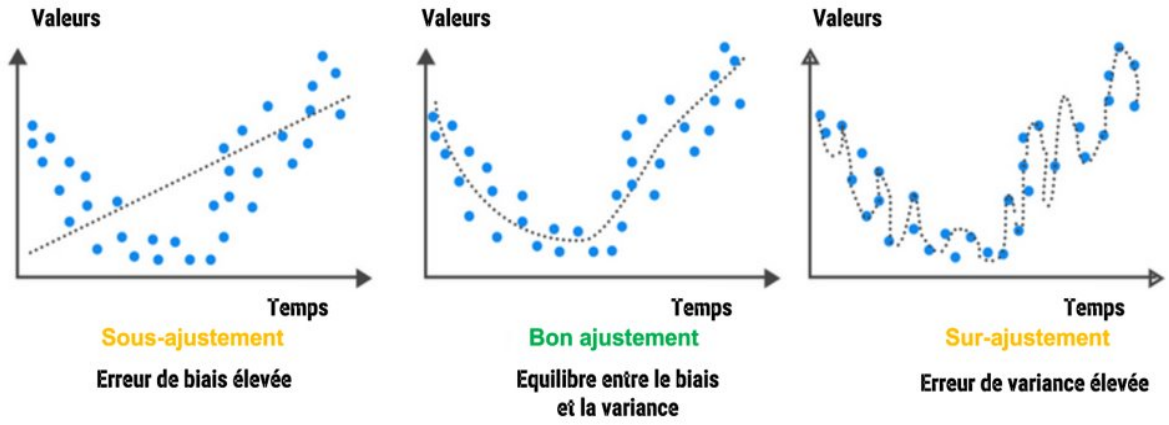


FIGURE 1.15 – Le sous-ajustement et le sur-ajustement [1]

Plusieurs approches sont proposées pour assurer la généralisation et limiter le sur-ajustement et le sous-ajustement.

a) Interruption prématurée :

Les données sont souvent divisées en trois : entraînement, validation et test (voir section 1.2). À chaque époque, le modèle est testé sur les données de validation (fig. 1.17). L'idée de

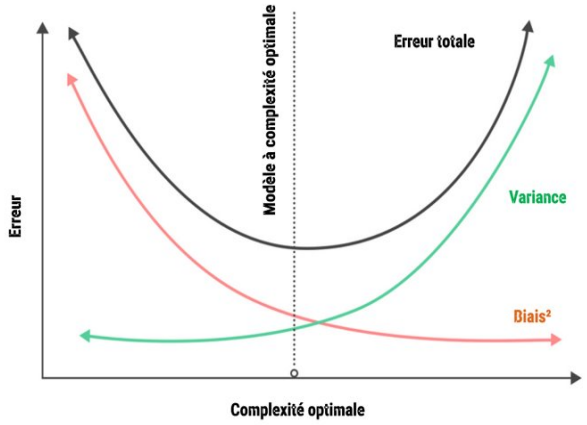


FIGURE 1.16 – Courbes illustrant le dilemme biais-variance [1]

l'interruption prématurée de l'apprentissage au point de la complexité optimale du modèle sert à éviter le phénomène de sur-ajustement et donc à améliorer la généralisation du modèle (early stopping). Le processus d'apprentissage est donc modifié comme suit :

1. Introduire les données et lancer le modèle pour obtenir les prédictions.
2. Calculer la fonction de coût.
3. Déployer l'algorithme d'optimisation pour ajuster les poids du modèle afin d'améliorer les prédictions.
4. Lancer le modèle appris sur le jeu de validation et calculer la fonction d'erreur.
5. Répéter les étapes 1 à 5 jusqu'à ce que le nombre d'époques défini soit atteint ou que la valeur de l'erreur de validation augmente.

b) La validation croisée :

La validation croisée consiste à diviser un ensemble de données donné en k sous-ensembles, chacun utilisé comme ensemble de validation à un moment donné. Par exemple, dans la fig 1.18, $k = 5$, l'ensemble de données est divisé en 5 sous-ensembles. Dans la première itération, la première section en rouge est utilisée pour valider le modèle, et les autres en vert sont utilisées pour l'entraînement du modèle, etc. Finalement, la moyenne des k erreurs de validation est alors considérée comme une estimation de l'erreur de validation totale. La validation croisée est une technique efficace pour surmonter le problème de généralisation du modèle.

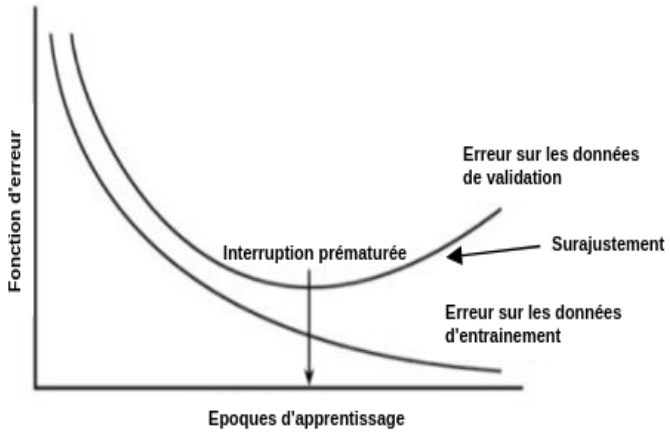


FIGURE 1.17 – Illustration du problème de sur-ajustement et d'interruption prématurée, inspirée de [236]

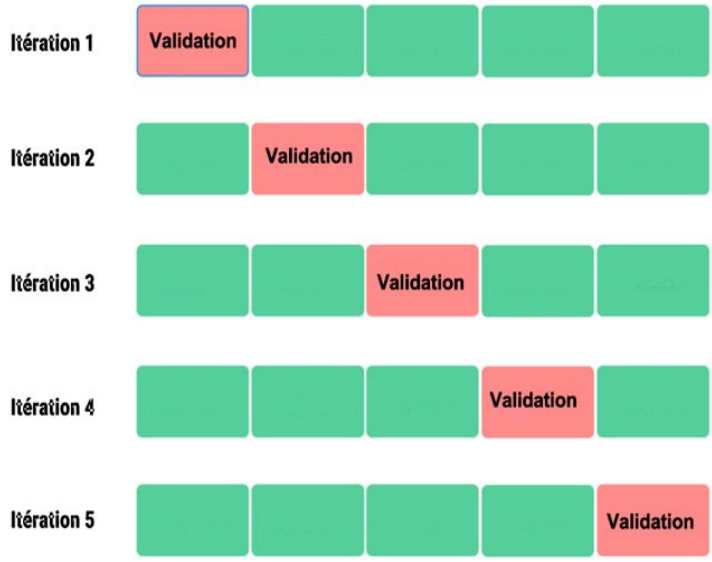


FIGURE 1.18 – Illustration de la stratégie de la validation croisée [236]

c) Dropout :

Le dropout est l'une des techniques pour lutter contre le sur-ajustement. Le concept de cette technique consiste à perturber l'apprentissage du modèle en désactivant temporairement un pourcentage de l'ensemble des neurones du réseau. Ce nombre de neurones à désactiver est un autre hyperparamètre à choisir. D'une époque à une autre, une sélection aléatoire des neurones à

désactiver est effectuée. Ainsi, le modèle apprend avec une nouvelle configuration et augmente son pouvoir de généralisation [234] [254] [89].

d) La régularisation :

La régularisation est une autre technique utilisée pour limiter le phénomène de sur-ajustement [129]. Elle consiste à ajouter un terme de pénalisation à la fonction d'erreur afin de modérer les valeurs des poids. Le choix de la valeur de ce terme est important pour éviter un sous-ajustement ou un problème de convergence [21][153]. Nous observons l'influence de la valeur du terme de régularisation dans la fig. 1.19. Plus le terme de régularisation est élevé, plus on évite le sur-ajustement. En parallèle, nous devons éviter également un sous-apprentissage.

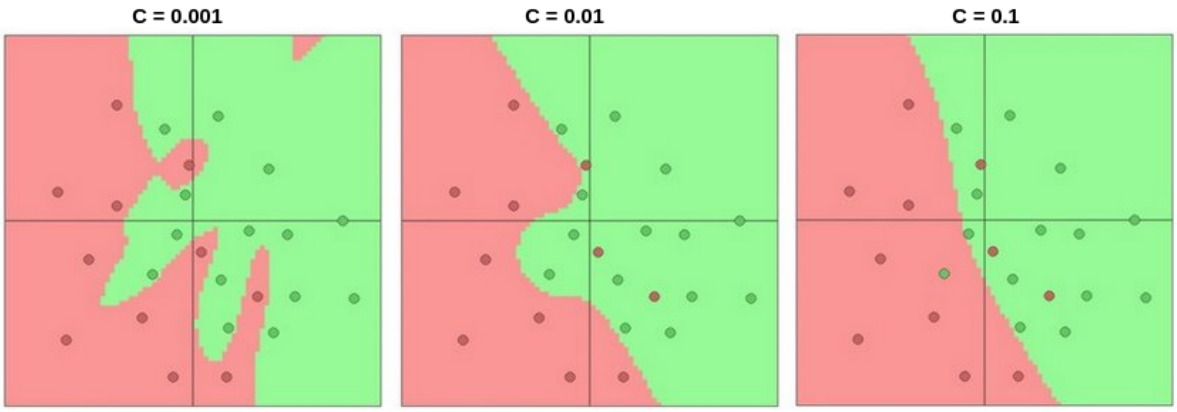


FIGURE 1.19 – Influence de la valeur du terme de régularisation sur un réseau de neurones entrainé pour une tâche binaire de classification [44]

1.3.7 Métriques

La classification est l'une des tâches principales des techniques de l'apprentissage automatique. Afin de légitimer l'utilisation de ces techniques dans des applications variées et importantes et en raison de leur impact direct sur des décisions critiques, notamment dans le domaine médical, une évaluation correcte et robuste de ces algorithmes est primordiale. La performance de ces méthodes est généralement mesurée à l'aide de métriques comparant les prédictions des modèles avec une vérité de terrain, qui est souvent manuellement annotée par les experts et donnent des indications sur le type des erreurs commises lors des prédictions. Pour

commencer, nous basons le calcul des métriques sur les valeurs présentées sur un tableau connu sous le nom de la matrice de confusion (voir fig 1.21). Une matrice de confusion renseigne généralement sur quatre indicateurs :

Les bonnes prédictions :

— Vrais positifs (True Positive TP) : Des données bien prédites dans la classe cible

— Vrais négatifs (True Negative TN) : Des données bien prédites en dehors de la classe cible

Les fausses prédictions :

— Faux négatifs (False Positive FN) : Des données mauvaisement prédites en dehors de la classe cible

— Faux positifs (False Positive FP) : Des données mauvaisement prédites dans la classe cible

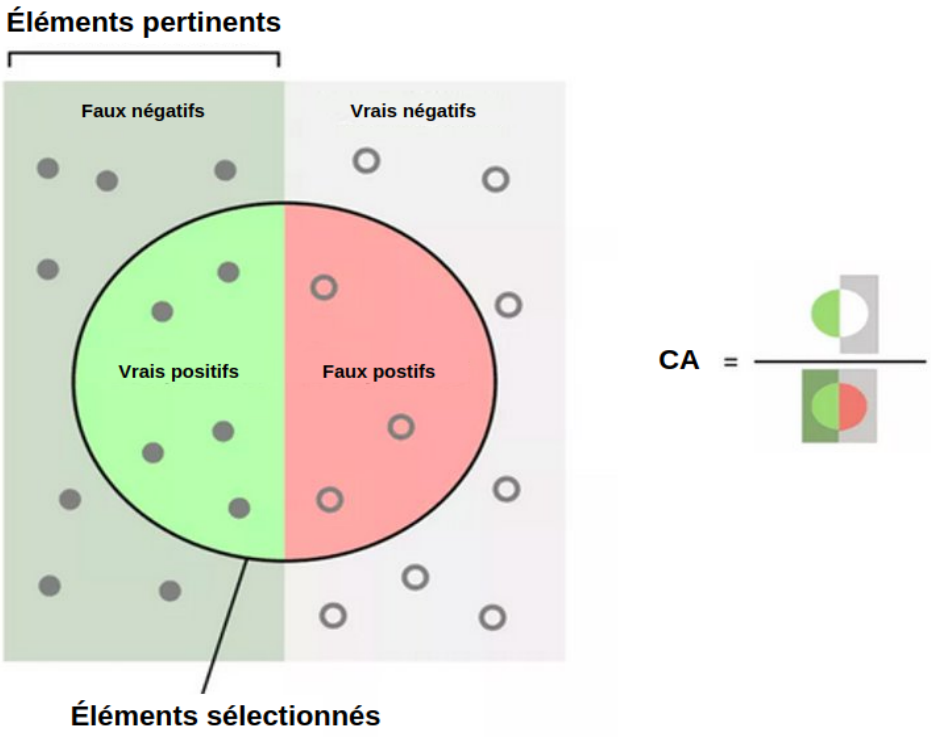


FIGURE 1.20 – Illustration des éléments de la matrice de confusion [31]

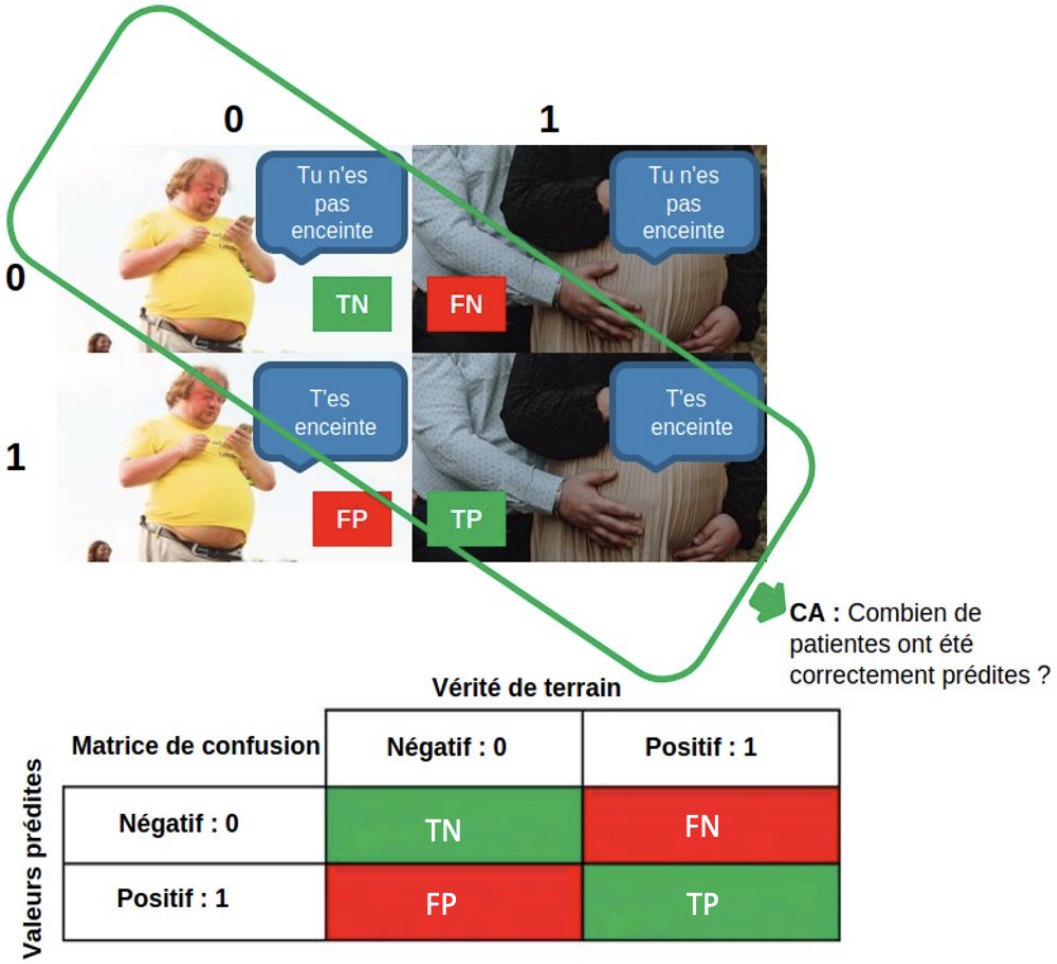


FIGURE 1.21 – Matrice de confusion [31]

a) Précision de classification (CA) :

CA indique le pourcentage de bonnes prédictions

$$CA = \frac{TP + TN}{TP + TN + FP + FN}$$

CA présente des limites dans le cadre de données déséquilibrées, il vaut mieux utiliser la précision de classification balancée.

b) Précision de classification balancée (BAC) :

Elle s'agit d'une métrique de classification introduisant une pondération arbitraire ou selon le nombre de classes et leurs proportions dans la totalité de la base de données [32].

$$\begin{aligned}\text{Balanced Accuracy Weighted} &= \frac{w_1\,\text{sensitivity} + w_2\,\text{specificity}}{w_1 + w_2} \\ &= \frac{w_1}{w_1 + w_2}\frac{TP}{TP + FN} + \frac{w_2}{w_1 + w_2}\frac{TN}{TN + FP}\end{aligned}$$

$$\text{Binary balanced Accuracy} = \frac{1}{2} \times \frac{TP}{TP + FN} + \frac{1}{2} \times \frac{TN}{TN + FP}$$

c) Spécificité :

Elle renseigne sur la capacité d'un modèle à identifier les vraies positives

$$\text{Spécificité} = \frac{TP}{TP + FP}$$

d) Sensibilité :

Elle renseigne sur la proportion des bonnes prédictions parmi les données prédites dans la classe cible.

$$\text{Sensibilité} = \frac{TP}{TP + TN}$$

1.4 L'apprentissage profond pour l'image

Récemment, l'apprentissage profond a réussi à résoudre plusieurs problèmes dans une variété de domaines. Les données d'entrée peuvent être de différents types, notamment des données numériques, des images, des vidéos et des signaux. Dans notre travail, nous nous intéressons en particulier à l'image dans le domaine médical où l'apprentissage profond est utilisé pour accomplir plusieurs tâches, comme la détection, la reconnaissance et la

segmentation des objets et la classification des images. Dans ce chapitre, nous expliquons tout d'abord la notion d'image. Ensuite, nous abordons l'historique des techniques d'apprentissage profond appliquées aux images du domaine médical en soulignant l'évolution des architectures déployées et leurs différents avantages et inconvénients. Enfin, nous nous intéressons en particulier à la tâche de classification et les travaux dans l'état de l'art la concernant.

1.4.1 L'image aux yeux de la machine

Une image est définie comme étant une fonction à deux variables x et y [12]. En effet, pour la machine, une image digitale de 8 bits consiste en une grille de pixels de valeur dans l'intervalle [0,255] (fig.1.22) [76]. Nous pouvons distinguer deux types d'image.

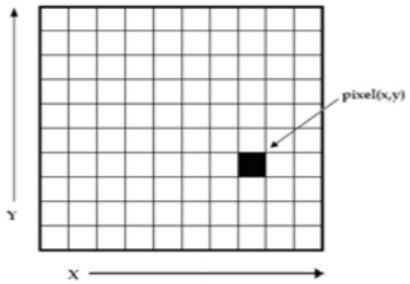


FIGURE 1.22 – Représentation de l'image [154]

a) Image en niveaux de gris

L'image est représentée par une matrice à deux dimensions, et la valeur du pixel représente l'intensité des niveaux de gris. La valeur 0 représente les pixels les plus sombres (noir) et la valeur 255 représente les pixels les plus clairs (blanc).

b) Image couleur

Les images couleurs (RVB) sont représentées sur trois canaux dans une matrice à trois dimensions. Chaque canal correspond à une des couleurs primaires (rouge, vert, bleu). Ainsi, chaque pixel est représenté par trois valeurs respectivement à l'intensité de chaque couleur dans ce pixel [104].

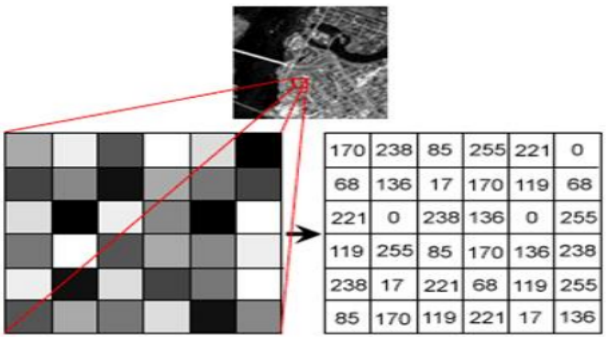


FIGURE 1.23 – Une partie de la matrice représentative d'une image en niveaux de gris [114]

Plusieurs techniques de traitement sont effectuées pour préparer les images d'entrée. Le choix dépend de la tâche, le domaine et la nature du jeu de données. Nous citons quelques exemples de ces techniques, mais la liste n'est pas exhaustive.

— Normalisation de la taille : il s'agit de redimensionner les images pour avoir la même taille à l'entrée au réseau de neurones ;

— Interpolation des données : cette approche sert à homogénéiser la résolution des données d'entrée ;

— Normalisation de l'intensité : il s'agit d'une étape de prétraitement d'image, utilisée pour modifier l'intensité des pixels d'une image donnée. Plusieurs solutions sont proposées dans ce sens [124][179] [2] (voir Tab. 1.3) ;

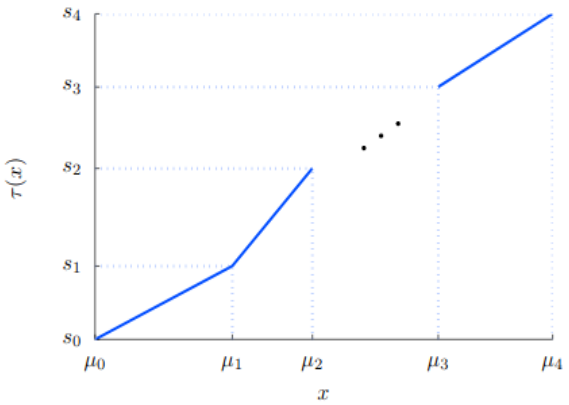


FIGURE 1.24 – Illustration de la normalisation de Nyul et Udupa [124]

— Filtrage : nous pouvons appliquer des filtres sur les images ou sur les histogrammes des images pour éliminer des informations non pertinentes. Nous pouvons aussi

Technique de normalisation	**Formule**	**Cas d'utilisation**
Normalisation linéaire Min-Max	$x' = (x - x_{min})/(x_{max} - x_{min})$	Une méthode classique de normalisation de la plage d'intensité utilisée surtout avec une distribution de données plus au moins uniforme.
Mise à l'échelle logarithmique	$x' = log(x)$	Une méthode utilisée lorsque la distribution suit la loi de puissance.
Clipping	si $x > max$, alors $x' = max$ si $x < min$, alors $x' = min$	Une méthode utilisée lorsque la distribution contient des valeurs aberrantes.
Z-score	$x' = (x - \mu)/\sigma$ où σ écart type et μ est la valeur moyenne	Une méthode utilisée lorsque la distribution ne contient pas des valeurs aberrantes.
Nyul et Udupa	$x' = s_i + (x - \mu_i)(s_{i+1} - s_i)/(\mu_{i+1} - \mu_i)$ avec $x \in [\mu_i, \mu_{i+1}]$ (fig.1.24)	Une méthode de normalisation plus raffinée basée sur la mise en correspondance des histogrammes pour assurer l'uniformité des échelles d'intensité [178].

TABLE 1.3 – Techniques de normalisation

redimensionner ou recadrer les données ou prendre juste une partie des images prenant en considération une information préliminaire sur la région la plus importante de la donnée ;

— Autres : nous pouvons diviser l'image en petites parties ou échantillonner des petits patchs de l'image pour diminuer la complexité du calcul. Il y a également plusieurs autres techniques d'augmentation de données, notamment les rotations, les transformations affines, la mise à l'échelle, la modification de la texture ou les couleurs [224]. Cependant, dans plusieurs cas et domaines d'utilisation, ces modifications ne sont pas tolérées et peuvent engendrer un biais important à la prise de la décision.

1.4.2 L'imagerie médicale

De nos jours, l'apprentissage profond est très populaire dans différents domaines avec des données d'une grande variété de propriétés. Dans le domaine médical, l'imagerie médicale est l'une des composantes principales et omniprésentes dans le diagnostic, le suivi et compréhension des maladies. Les pixels qui composent l'image représentent ses

caractéristiques, notamment l'atténuation des rayons X, la densité de l'eau, l'impédance acoustique, l'activité électrique, etc. Les images peuvent être analysées par l'homme ou automatiquement par le biais de techniques informatiques intelligentes. Plusieurs modalités d'imagerie médicale sont utilisées. Dans cette section, nous souhaitons d'abord présenter un bref résumé de ces modalités pour mettre en évidence les particularités du domaine médical et les défis à relever. Ensuite, dans la prochaine section, nous présentons les architectures de réseaux de neurones profonds les plus populaires, adaptées et déployées pour le diagnostic et la détection des anomalies dans le domaine médical.

Le concept d'acquisition des images radiographiques (rayons X) est l'exposition du corps à un type de rayonnement appelé rayon X pour obtenir les parties du corps dans différentes nuances de noir et de blanc. Ceci est dû au fait que les différents tissus absorbent différentes quantités de rayonnement. La tomographie assistée par ordinateur (CT) est une combinaison de rayons X et d'un ordinateur. Cette modalité présente des images plus détaillées que les images à rayons X, déployant un faisceau étroit de rayons X qui tourne autour d'une partie spécifique du corps pour produire une série d'images sous différents angles. Ainsi, l'ordinateur utilise ces angles pour créer un balayage bidimensionnel (2D) en coupe transversale exposant l'intérieur de la partie du corps. L'ordinateur peut empiler plusieurs de ces scans les uns sur les autres pour créer une image détaillée de l'organe dans toutes les directions. Dans la tomographie par émission de positons (TEP), on fait introduire un agent de contrast contenant des traceurs radioactifs.

Contrairement aux modalités précédentes, l'imagerie médicale par résonance magnétique (IRM) ne fait pas intervenir les radiations ionisantes nocives. De puissants aimants, des ondes radio et un ordinateur sont utilisés pour obtenir des images détaillées des tissus et structures de l'intérieur du corps. L'IRM est considérée une modalité très importante puisqu'elle met en évidence les différences entre les tissus et les structures sains et malades, mieux que toute autre technologie existante. L'IRM est utilisée notamment pour l'imagerie cérébrale. Elle permet de détecter les changements dans le flux sanguin dans les différentes parties du cerveau, aidant ainsi à déterminer les fonctions associées à chaque partie du cerveau. Cette technologie a favorisé un grand progrès dans le diagnostic de maladies telles que la maladie

d'Alzheimer, la sclérose en plaques et les tumeurs. L'imagerie par tenseur de diffusion (DTI) est une technique de neuro-imagerie basée sur l'IRM permettant d'estimer l'emplacement, l'orientation et l'anisotropie des trajets de la matière blanche du cerveau (https://www.imagilys.com/diffusion-tensor-imaging-dti/).

Toutes ces modalités sont importantes pour le diagnostic médical. Plusieurs examens se basent sur l'imagerie médicale pour prendre les bonnes décisions. Les images médicales se caractérisent par des informations quantitatives et des objets sans orientation canonique. La différence entre les différentes modalités et les différentes tâches (la détection des maladies, la quantification des anomalies et la segmentation des organes) sont énormes. La prise en compte de ces différences et l'adaptation des algorithmes en conséquence permettent d'améliorer considérablement les performances de ces modèles.

La performance de l'apprentissage profond s'est beaucoup améliorée dans les tâches de classification. Cette tâche est très importante dans le domaine de l'imagerie médicale. L'un des principaux avantages de l'apprentissage profond est la capacité à extraire automatiquement les caractéristiques en détectant les relations latentes entre les différentes composantes des données d'entrée. Cette extraction automatique est un facteur clé dans différentes tâches médicales et en particulier dans les tâches de classification.

1.4.3 L'évolution des réseaux convolutifs

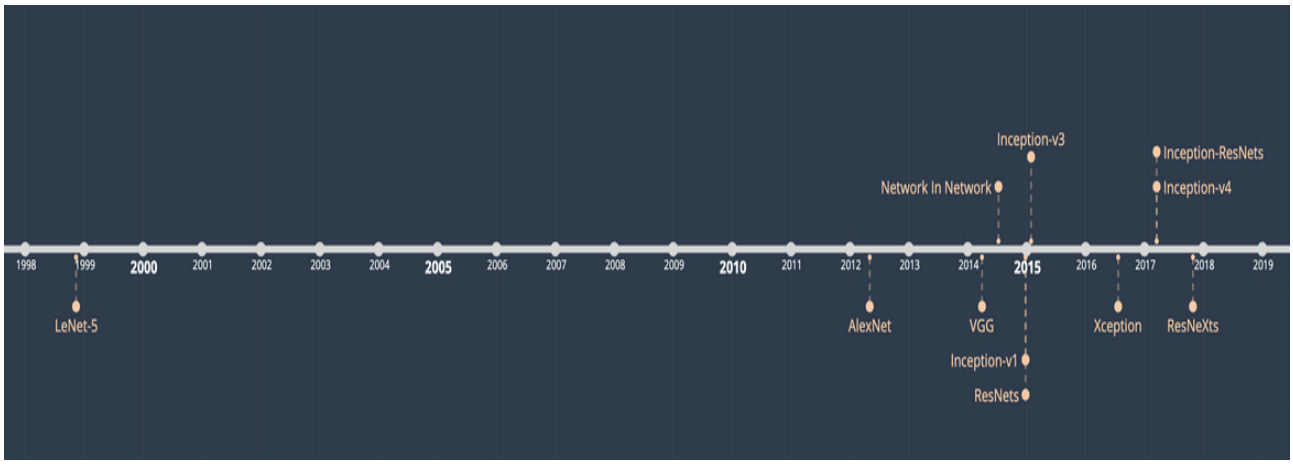


FIGURE 1.25 – La chronologie des architectures de l'apprentissage profond

Les CNN ont été introduits par Kunihiko Fukushima en 1980 [68]. La première application dans l'analyse d'images médicales était proposée par Lo et al. [152] en 1995. LeCun et al. [138] ont connu le premier véritable succès avec des CNN en 1990 sur la base de données MNIST pour la reconnaissance de chiffres manuscrits avec l'architecture LeNet comprenant 3 couches de convolutions et 2 couches entièrement connectées avec environ 60000 paramètres [137]. L'architecture proposée par LeCun et al. a été capable d'obtenir des résultats de précision très proches de l'état de l'art (un taux d'erreur de 1% et un taux de rejet de 9% à partir d'un jeu de données d'environ 1000 échantillons. L'utilisation d'un grand nombre de couches cachées et d'images plus grandes peut améliorer les résultats. Cependant, le nombre de paramètres d'entraînement peut augmenter rapidement proportionnellement au nombre de couches et à la taille de données d'entrée, d'où la notion de filtre ou noyau (kernel) liée aux couches de convolution. En effet, des filtres sont appliqués sur différentes positions de l'entrée pour extraire des cartes de caractéristiques locales. Ainsi, un neurone ne reçoit plus la totalité de l'image, mais seulement la région définie par le filtre. Une autre couche, classiquement utilisée pour minimiser la complexité des CNN, est la couche de pooling qui est considérée comme une couche de résumé de la couche en entrée.

Les défis principaux rencontrés déployant cette architecture sont le gradient évanescent [16] et le calcul complexe avec les ressources informatiques limitées [15]. Plusieurs solutions ont été proposées pour faire face à ces limites, notamment des solutions liées à l'algorithme d'optimisation (Momentum, AdaGrad, RMSProp, Adam) [205] ou des solutions liées à la façon d'apprentissage, notamment l'apprentissage semi-supervisé ou l'entrainement à deux phases [93]. D'autres solutions liées à l'architecture sont déployées telles que la mémoire à long terme (LSTM) pour gérer les difficultés des dépendances à long terme de la descente de gradient [96][18] ou l'intégration de la fonction d'activation ReLU ayant une bonne performance dans différentes tâches d'imagerie [128][74].

Cela nous amène à la prochaine étape importante de l'histoire de l'apprentissage profond dans l'imagerie, à savoir la contribution de Krizhevsky et al. en 2012 au défi ImageNet, une base référence pour la classification et la détection d'objets sur des centaines de catégories d'objets et des millions d'images, avec l'architecture AlexNet [128][206]. AlexNet est une

architecture CNN similaire à LeNet, employant des noyaux avec de grands champs réceptifs dans les couches proches de l'entrée et des noyaux avec des plus petits champs réceptifs en s'approchant de la sortie. L'architecture consiste en 8 couches dont 5 couches de convolution et 3 couches entièrement connectées. La principale différence entre AlexNet et LeNet réside dans la fonction d'activation où AlexNet utilise l'unité linéaire rectifiée (ReLu) à la place de la tangente hyperbolique (tanh). En outre, AlexNet gagne en termes du temps de calcul, surtout avec l'évolution des GPU permettant d'obtenir de meilleurs résultats et d'être le modèle gagnant de l'ImageNet Large Scale Visual Recognition Challenge (ILSVRC) en 2012.

Suite à ImageNet, les architectures plus profondes avec des noyaux plus petits ont été proposées comme la façon la plus directe pour améliorer les performances d'un réseau de neurones. Visual Geometry Group (VGG) est l'une des architectures reconnues utilisant toujours le ReLU se composant de 13 couches convolutives et 3 couches entièrement connectées et comportant environ 138 millions de paramètres [228]. Par la suite, plusieurs variantes plus profondes ont été proposées, notamment le modèle à 19 couches VGG19 ou OxfordNet, l'architecture gagnante du défi ImageNet de 2014. Nous pouvons évoquer également le réseau à 22 couches reconnu sous le nom de Inception-v1 ou GoogLeNet avec une architecture plus complexe dans le but d'améliorer l'efficacité de l'apprentissage et réduire également le nombre de paramètres à environ 5 millions [147][246].

L'architecture Inception a connu différents successeurs développés tels que Inception-v3 proposé par Szegedy et al. [247] en 2015 incluant 24M paramètres et visant à avoir des calculs plus efficaces déployant des méthodes de factorisation et de normalisation par lot (batch normalisation) [99]. Cependant, augmenter le nombre de paramètres de l'architecture pour obtenir de meilleures performances affaiblit la généralisation et sature la précision. Dans ce sens, l'équipe de recherches de Microsoft, He et al. [91] ont proposé une nouvelle architecture ResNet en 2015 pour résoudre ce problème. Cette architecture inclut des sauts de connexions (skip connections) et également une normalisation par lots (batch normalisation) tout en construisant des modèles plus profonds incluant jusqu'à 152 couches sans compromettre la généralisation du modèle [99]. De nos jours, la mission de battre les scores de l'état de l'art sur la base référence ImageNet avec des architectures classiques devient de plus en plus difficile.

Cependant, Google n'arrête pas d'innover avec des nouvelles architectures telles que la nouvelle version de Inception nommée Inception-v3 ou Xception proposée par François Chollet [39] ou Inception-v4 la version améliorée proposée par Szegedy et al. [245] avec quelques changements simples incluant plus de modules de convolutions et l'uniformisation du nombre de filtres pour chaque module. La même équipe de chercheurs a proposé une autre architecture nommée Inception-ResNet-V2 intégrant des blocs résiduels [245]. ResNeXt-50, la version améliorée de ResNet, est également une des architectures largement répandues. Elle intègre plus de tours parallèles dans un module. Cette approche a été incorporée aussi dans quelques travaux dans l'architecture Inception. Cette approche permet de résoudre le problème de l'empreinte mémoire. Nous pouvons citer aussi les architectures U-Net et ses variantes [296], SE-Net avec le bloc "queeze and excitation", EfficientNets et les variantes de B0 à B9 basées sur une approche de redimensionnement de l'architecture en vue de balancer entre sa profondeur, sa résolution et sa largeur [251]. Nous pouvons admettre que la concurrence a saturé et ces architectures sont largement utilisées dans plusieurs domaines, notamment l'analyse d'imagerie médicale.

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
VGG16	528 MB	0.713	0.901	138,357,544	23
InceptionV3	92 MB	0.779	0.937	23,851,784	159
ResNet50	98 MB	0.749	0.921	25,636,712	-
Xception	88 MB	0.790	0.945	22,910,480	126
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
ResNeXt50	96 MB	0.777	0.938	25,097,128	-

FIGURE 1.26 – Précision de la classification et les propriétés des architectures des réseaux de neurones profonds appliquées sur la base de données de validation de ImageNet

1.4.4 Apprentissage profond pour l'imagerie médicale

Plusieurs types de données sont exploités dans le domaine médical [28][182]. En revanche, l'imagerie est la composante principale dans la perspective de l'aide au diagnostic médical [55]. Geert Litjens et al. [148] évoquent 300 contributions d'apprentissage profond dans les

applications d'images médicales telles que la classification d'images, la détection d'objets, la segmentation et dans de multiples domaines médicaux, à savoir la neurologie, l'ophtalmologie, l'oncologie, la cardiologie, la pneumologie, l'abdominal, la musculo-squelettique…

Les tâches médicales font l'objet de recherche en intelligence artificielle depuis les années 1970. La solution consistait principalement en des algorithmes mathématiques simples. Entre-temps, la recherche en IA a débuté avec GOFAI (good old-fashioned artificial intelligence) associée à Haugeland [261]. Cependant, ces premières intuitions d'intelligence artificielle et les simples algorithmes étaient loin moins performants que les solutions basées sur l'apprentissage profond, surtout avec l'évolution des architectures CNN.

Plusieurs revues essayent de couvrir les travaux de recherche déployant l'apprentissage profond en analyse d'images médicales [148][223][220][81]. En effet, l'apprentissage profond attire l'attention de la communauté scientifique dans le domaine médical au cours des dernières années et ces architectures de base constituent aujourd'hui l'état de l'art pour multiples tâches liées à l'imagerie médicale et sont disponibles dans différents outils de programmation tels que Keras [3] (fig. 1.26). Par conséquent, il s'agit d'un sujet de recherche émergent dans le domaine médical grâce à ses nombreuses perspectives innovantes, comme le débruitage d'images médicales en utilisant un réseau de neurones résiduel [109], la détection de pathologies thoraciques [11], la segmentation avec un apprentissage faiblement supervisée [53][116], la fusion d'images médicales [279] et différentes autres applications.

Les modèles multi-flux sont des architectures ayant comme entrée des données de multiples sources sous forme de canaux présentés à la couche d'entrée et pouvant être fusionnés à n'importe quelle couche du réseau. Ces architectures ont une application importante pour les tâches du traitement d'images médicales [113]. Par exemple, dans le cadre de la détection d'anomalies, le contexte est important. Afin d'augmenter le contexte, des patchs plus grands sont donnés en entrée. Cependant, cette approche directe augmente significativement le nombre de paramètres et les besoins en capacité et mémoire de calcul. Dans ce sens, des nouvelles architectures ont été élaborées intégrant le contexte dans une représentation connexe avec des informations locales à haute définition [113][169] [284]. D'autres techniques de vision

3. https://keras.io/

artificielle peuvent également inspirer les futures perspectives de l'imagerie médicale [174] [235] [290].

D'autres architectures ont déjà été déployées dans ce domaine. Nous pouvons citer les auto-encodeurs empilés (SAE) et les réseaux de croyance profonds (DBN), dans lesquels les réseaux de neurones profonds sont entraînés couche par couche d'une manière non supervisée (pré-entraînement). Ensuite, un ajustement du réseau empilé est effectué. Cependant, ces techniques sont gourmandes et complexes en termes de temps et capacité de calcul [148].

La classification d'images est l'une des principales applications de l'apprentissage profond dans le domaine médical ayant en entrée une ou plusieurs images médicales et en sortie une seule variable informative correspondant au diagnostic. Les architectures les plus populaires pour la tâche de classification sont toujours AlexNet et VGG [148]. Néanmoins, plusieurs travaux dans le domaine médical ont fait intervenir des architectures différentes, comme l'Inception-v3 pour la détection automatique de la rétinopathie diabétique et de l'œdème maculaire diabétique sur la rétine [83] ou la classification du cancer de la peau en utilisant l'apprentissage profond [56]. Cependant, c'est aussi l'un des secteurs qui manque énormément de données expliquant la popularité de l'apprentissage par transfert pour ce type de tâche [148][250].

Certains travaux ont intégré un réseau pré-entraîné uniquement comme extracteur de caractéristiques, d'autres ajustent ou affinent l'entraînement d'une partie ou de la totalité du modèle aux données du domaine médical. Deux articles références étudient les différents résultats de ces deux approches et ont donné des résultats contradictoires [8][120][148]. Antony et al. [8] soulignent, dans leur travail lié à l'arthrose radiographique du genou, qu'affiner le modèle a clairement surpassé l'utilisation du réseau pré-entraîné uniquement comme extracteur de caractéristiques (précision de $57,6\%$ contre $53,4\%$). D'un autre côté, l'application de Kim et al. [120] de l'apprentissage profond pour la cytopathologie de la thyroïde met en évidence qu'intégrer le CNN pré-entraînés uniquement comme extracteur de caractéristiques est plus performant que l'ajustement (précision de $70,5\%$ contre $69,1\%$). Nous pouvons observer que même dans ce cas où l'ajustement est nuisible, les résultats en termes de précision sont étroits. Cependant, en conclusion et considérant les résultats des deux approches mentionnées,

l'intégration d'un réseau pré-entraînée rend l'entraînement plus rapide et plus efficace. Par exemple, pour la détection automatique de la rétinopathie diabétique et de l'œdème maculaire diabétique [83] et la classification du cancer de la peau [56], l'Inception-v3 a été affiné sur des données médicales pour obtenir une précision proche de celle de l'expert humain. Manegola et al. [160] ont expérimenté certaines tâches pour comparer la performance de l'entraînement à partir de zéro et l'ajustement d'un réseau pré-entraîné. Les résultats obtenus mettent en évidence que l'ajustement est plus performant, surtout avec un petit jeu de données (environ 1000 images de lésions cutanées).

Les méthodes classiques de l'apprentissage automatique représentaient la base des méthodes utilisées dans différentes applications jusqu'à l'apparition de l'apprentissage profond. L'apprentissage profond ne se limite pas aux modèles standards CNN. Il comprend d'autres architectures qui peuvent intégrer un pré-entraînement non supervisé. Nous pouvons citer les auto-encodeurs (AE), les auto-encodeurs automatiques empilés (SAE), les machines de Boltzmann restreintes (RBM), les réseaux de croyance profonds (DBN), qui sont en effet des SAE, mais avec des blocs de RBM au lieu de AE, les auto-encodeurs variationnels (VAE), les réseaux adversariaux génératifs (GAN)... Le pré-entraînement non supervisé utilise des échantillons non liés à la tâche cible pour initialiser les paramètres du réseau, ce qui permet de trouver les paramètres optimaux pour un ajustement et donc d'améliorer les performances de classification. À notre connaissance, les premiers articles incorporant ces techniques dans la classification d'images médicales étaient liées neuro-imagerie en 2013, notamment l'utilisation des DBN pour la classification de la maladie d'Alzheimer (MA) basée sur l'IRM [26][190]. Suk et al. [237] ont également déployé des DBN, mais aussi des DBN multimodaux (MM-DBN). L'objectif, par analogie avec l'architecture multi-flux du CNN, est la combinaison des informations complémentaires apportées par les différentes modalités de l'imagerie médicale IRM et TEP en entrées, obtenant ainsi les meilleurs résultats dans trois problèmes de classification. Suk et al. ont utilisé les AE et les SAE pour la classification d'images de la maladie d'Alzheimer (MA) et le déficit cognitif léger (DCL), toujours en combinant l'IRM et la TEP dans quatre tâches de classification binaire [238][239]. Ces travaux ont abordé certaines difficultés, mais ils ont mis en évidence la robustesse et la performance

de ces techniques d'apprentissage profond dans les tâches de neuro-imagerie en particulier et de l'imagerie médicale en général. Ils ont également confirmé que la profondeur du réseau neuronal et la quantité de données sont effectivement des facteurs clés pour aboutir à la meilleure classification. En outre, il y a toujours des limitations et des inconvénients tels que la difficulté de l'interprétation des résultats et la limite maximale de deux modalités bimodales d'IRM et de TEP, alors qu'il est généralement bénéfique de fusionner autant de modalités et de types de données que possible pour utiliser le maximum d'informations. L'opération de combinaison des données était une simple concaténation des caractéristiques des modalités dans un vecteur, résultant une performance basse comparée à celle du SVM multimodal. Dans ce sens, Suk et al. proposent une approche pour résoudre le problème des données insuffisantes ou incomplètes avec l'apprentissage profond multimodal [290]. D'autres travaux s'inspirent également des réseaux multimodaux pour exploiter le maximum de données disponibles [174][235]. Des nombreux défis liés à ces architectures demeurent au cœur de la recherche.

Les champs d'applications de l'apprentissage profond ont évolué au cours de ces dernières années. En effet, on ne peut pas se permettre de ne pas mentionner le secteur de recherche d'images basée sur le contenu (CBIR) où l'utilisation de l'apprentissage profond est une solution directe, voire nécessaire. CBIR présente une technique de recherche dans une base de données massive permettant de détecter des cas similaires pour une meilleure compréhension de l'entrée. En effet, la taille et l'hétérogénéité des bases de données d'images médicales exigent l'utilisation de systèmes de recherche basés sur le contenu pour une organisation efficace et une meilleure aide aux experts cliniques. Le principal défi de cette tâche est l'extraction des caractéristiques et des relations pertinentes entre ces caractéristiques. Par conséquent, grâce à sa capacité d'extraire automatiquement les caractéristiques et les relations latentes, l'apprentissage profond est une orientation immédiate pour une telle tâche d'imagerie médicale.

La plupart des articles relatifs au CBIR intègrent des CNN pré-entraînés pour l'extraction de caractéristiques. Anavi et al. [4] et Liu et al. [150] ont utilisé 5 couches de CNN et des couches entièrement connectées pour l'extraction de caractéristiques des radiographies. Anavi et al. ont utilisé un réseau pré-entraîné dans la phase d'extraction des caractéristiques. Liu et al. ont utilisé une différente technique en utilisant des images à codes à barres de Radon. Liu et al. expliquent

les résultats inférieurs à l'état de l'art par l'utilisation des petits patches de 96 pixels. Chung et al. [41] utilisent des CNN siamois (SCNN) pour apprendre automatiquement les représentations d'images radiographiques. Le SCNN proposé par Chung et al. exigent des données en paires d'images en entrée. Pour valider les représentations d'images apprises pat leur réseau profond, ces représentations sont utilisées pour une tâche de recherche d'images médicales basée sur le contenu d'une base de données d'images de fond de rétinopathie diabétique. Le SCNN a prouvé son efficacité en surpassant les performances de l'architecture classique CNN. Shah et al. [222] ont manipulé des images MR de la prostate déployant la combinaison des CNN avec des forêts de hachage (forêts aléatoires non supervisées). 1000 caractéristiques résumées dans une grande matrice ont été extraites. Ensuite, les forêts de hachage sont déployées pour compresser cette matrice en descripteurs spécifique à chaque volume de l'IRM.

En effet, les CNN dominent les travaux dans le secteur de l'imagerie médicale, en particulier ces dernières années. Geert Litjens et al. [148] ont mentionné que parmi 47 articles liés à la classification d'images médicales entre 2015 et 2017, 36 articles utilisaient des CNN, 5 articles utilisaient des AE et 6 articles utilisaient des RBM. Cela peut être expliqué par la présence des modèles CNN pré-entraînés dans les outils de programmations disponibles, leur flexibilité et leur haute performance par rapport aux autres techniques.

Ainsi, de nombreuses combinaisons de différentes techniques existantes ont montré un potentiel important dans ce domaine [181]. Cependant, d'autres combinaisons et d'autres techniques utilisées sur l'imagerie naturelle peuvent avoir des perspectives prometteuses pour les images médicales. Ainsi, l'apprentissage profond dans le domaine de l'imagerie médicale n'en est qu'à ses débuts et de nombreux défis liés à la nature de ce domaine et aux spécificités de ses données restent non résolus.

1.4.5 Défis et perspectives

Pour conclure, l'apprentissage profond est une technique introduisant un groupe de méthodes informatiques et mathématiques permettant à un algorithme d'apprendre automatiquement les caractéristiques à partir d'un ensemble de données d'entrée et les règles

aboutissant aux meilleures décisions pour une certaine tâche. L'application de cette technique dans le domaine de l'imagerie médicale nécessite une évaluation, une validation et une précision plus poussées.

Bien que l'apprentissage profond réussisse à extraire des informations pertinentes dans des multiples domaines, le défi reste d'adapter ces architectures et techniques d'apprentissage profond existantes aux particularités et aux modalités du domaine médical. Dans cette section, nous fournissons une vue d'ensemble de certains défis et limitations de l'apprentissage profond spécifiques a ce domaine.

Le principal défi dans le domaine médical est la quantité de données réduite en raison de nombreux obstacles tels que l'accès aux données, la confidentialité, la protection des données et de multiples autres défis dans cette perspective. Esteva et al. [56] prouvent qu'il ne s'agit pas du premier défi réel. Ils utilisent 18 ensembles de données publiques et plus de 100000 d'images d'entraînement pour la classification du cancer de la peau au niveau de l'épiderme avec des réseaux neuronaux profonds. Mais, dans la plupart des cas, problème principal est plutôt la disponibilité de données labellisées de bonne qualité. Par conséquent, des multiples techniques sont testés pour atténuer ces défis, comme l'apprentissage par transfert expliqué ci-dessus pour fournir un meilleur point de départ que l'initialisation aléatoire du poids. En outre, la plupart des images sont associées à des rapports cliniques. Dans ce sens, l'extraction des informations de ces rapports et la possibilité de les incorporer dans notre réseau neuronal est bénéfique. Cependant, pour les maladies rares, un nombre suffisant de données labellisées avec un rapport clair et correct peut être indisponible.

Un autre défi lié aux données est le déséquilibre entre les classes. Dans cette perspective, des nombreuses techniques telles que le GAN, le sur-échantillonnage ou l'augmentation des données sont proposées, généralement pour les images naturelles. Cependant, l'efficacité de ces techniques pour les images médicales reste un défi à relever.

Les particularités des modalités de l'image médicale, telles que la propriété niveaux de gris et le défi du traitement de l'image médicale en 3D, constituent des contraintes supplémentaires

toujours liées aux données et plus spécifiques au domaine. Parmi les solutions proposées pour surmonter ce dernier défi est de traiter l'image 3D comme une pile de 2D et la patcher.

L'un des défis de l'apprentissage profond est le problème de la visualisation ou de l'interprétation efficace de la représentation des caractéristiques latentes. Certaines approches ont été testées pour la visualisation des poids formés et pour l'interprétation des représentations des caractéristiques latentes telles que la carte d'activation de classe (CAM). D'autres pistes doivent être abordés par les communautés scientifiques de recherche et les experts cliniques, en collaboration.

Plusieurs hyperparamètres sont toujours définis manuellement dans l'apprentissage profond. Le choix, parfois aléatoire, fait de l'apprentissage profond, pour de nombreux acteurs liés au domaine, une boîte noire et un problème de fiabilité. En effet, à notre connaissance, à l'heure actuelle, aucune technique d'apprentissage profond ne peut être généralisée comme étant la meilleure solution pour toutes les tâches et pour toutes les modalités. Par conséquent, ce domaine reste au centre de la recherche malgré son évolution et développement rapides.

Chapitre

2 État de l'art : Meta-apprentissage et attention

Sommaire

2.1 Problématique et solution au manque de données 52
2.2 Few-shot learning 54
2.3 Meta-apprentissage 56
2.4 Solutions de l'état de l'art 59
2.4.1 Apprentissage par transfert 59
2.4.2 Apprentissage de l'espace métrique 63
2.4.3 Des solutions d'apprentissage liées à l'algorithme d'optimisation 70
2.5 Mécanisme d'attention 77
2.5.1 Carte d'attention 77
2.5.2 Réseau d'attention spatiale pour la classification en few-shot learning 80
2.5.3 Apprentissage d'une représentation discriminante profonde basée sur la carte d'attention pour la classification de scènes 82
2.5.4 Apprentissage visuel dynamique en quelques coups sans oublier 84
2.5.5 Branche d'attention : Apprentissage du mécanisme d'attention pour une interprétation visuelle 86
2.6 Conclusion 88

De nos jours, l'apprentissage profond est utilisé dans des nombreuses tâches dans des domaines différents avec la même performance, voire meilleure que l'être humain. En 1980, Fukushima a créé les premiers CNN (section 1.4.3). Depuis lors, grâce l'évolution des capacités de calcul et de stockage et l'avancement réalisé par la communauté de l'apprentissage automatique, la performance des algorithmes d'apprentissage profond n'a pas cessé de s'améliorer sur des multiples tâches liées à l'imagerie. En 2015, He et al. [91] ont confirmé que leur modèle a dépassé les performances des humains dans la classification de la base référence ImageNet. Les modèles d'apprentissage profond ont connu un grand succès dans des applications variées, généralement pour lesquelles les bases étaient relativement abondantes.

On pourrait dire que la machine réussit mieux que les humains à l'exploitation des milliards d'images. Par contre, les expérimentations prouvent que les meilleurs résultats sont obtenus avec des énormes bases de données. Effectivement, la quantité et la qualité des données sont des critères primordiaux pour extraire les meilleurs traits et caractéristiques de données, en particulier en utilisant les techniques de l'apprentissage profond (fig. 2.1). Ce besoin observé de données a découragé de nombreux acteurs de la communauté scientifique de la recherche n'ayant accès qu'à de petits échantillons de données de profiter de la puissance de l'apprentissage profond. L'influence de la taille de la base de données est beaucoup plus importante dans le cadre de l'apprentissage profond que pour les techniques classiques de l'apprentissage automatique (ML). Dans ce sens, une étude comparative équitable est réalisée sur l'effet de la taille de la base de données sur la fonction d'erreur MAE d'un modèle d'apprentissage profond ElemNet et deux modèles ML conventionnels déjà prouvés plus performants que la forêt aléatoire [106]. Comme illustré dans la figure 2.2, la taille de la base de données d'entraînement impacte davantage ElemNet que les modèles la forêt aléatoire. Nous observons que la courbe d'erreur présente une diminution plus forte de l'erreur avec l'augmentation de la taille de la base pour l'apprentissage profond par rapport aux modèles de la forêt aléatoire. Nos résultats soulignent que le modèle DNN peut non seulement bénéficier davantage de la taille de la base de données par rapport aux modèles ML classiques, mais aussi qu'il peut les dépasser même avec une base de données d'environ 4 000 échantillons.

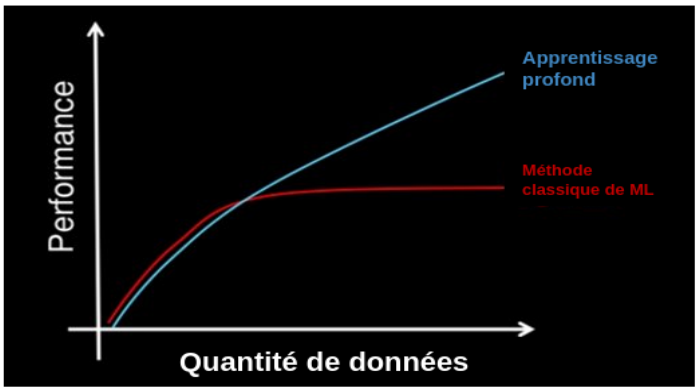


FIGURE 2.1 – Importance de la quantité de données sur la performance de l'apprentissage profond [5]

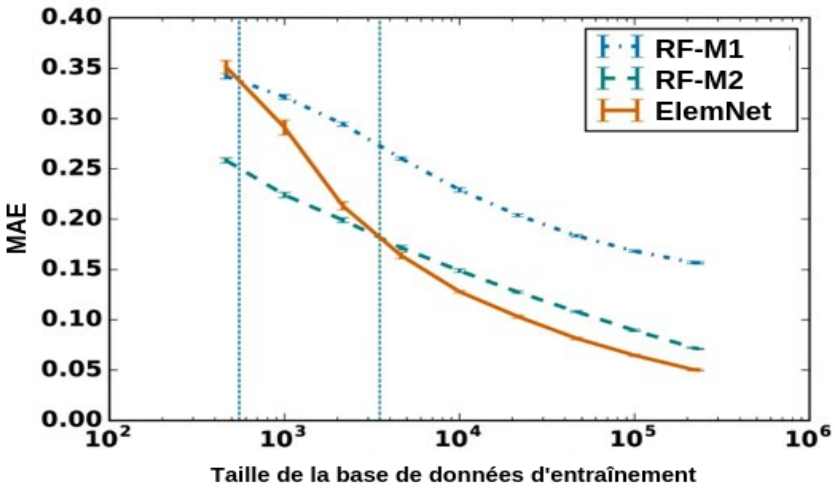


FIGURE 2.2 – La MAE en fonction de la taille de la base de données [106]

Cependant, dans plusieurs cas et à causes de nombreux obstacles, la collecte d'une grande base de données équilibrée et bien libellée avec une bonne qualité est une mission difficile. Dans des applications du monde réel, il n'est pas toujours possible de posséder une large base suffisante pour atteindre les mêmes performances mentionnées au-dessus. Parfois, la tâche consiste à classer des images avec seulement quelques échantillons par classe. Pour ce type de tâches, les algorithmes classiques d'apprentissage profond, étant des techniques très gourmandes en termes de données, échouent et sont toujours loin de la performance humaine.

À ceci s'ajoute, l'entraînement d'un modèle avec une grande quantité de données est gourmand et couteux en termes de temps et de capacité de calcul. En outre, un modèle DNN entraîné pour une tâche doit être entraîné de zéro pour une nouvelle tâche, en dépit de la

corrélation qui peut exister entre les deux tâches. Ensuite, après avoir relevé ces chalenges, ces données peuvent expirer rapidement, ce qui nécessiterait une nouvelle collecte de données et un nouvel entraînement. Ainsi, il faut trouver de nouvelles approches permettant l'utilisation de l'apprentissage profond avec peu de données en visant d'autres critères d'un réseau de neurones tels que l'algorithme d'optimisation, la fonction d'erreur [77] et l'initialisation des hyper-paramètres [270].

Dans la section suivante, nous abordons quelques implications directes du manque de données et les solutions de l'état de l'art. Ensuite, nous représentons la problématique de l'apprentissage en quelques coups (ou sur peu d'échantillons), dit "few-shot learning" et le principe de meta-apprentissage. Le reste du chapitre est consacré aux techniques fondamentales utilisées pour résoudre la problématique de few-shot learning en présentant l'état de l'art des méthodes, en détaillant l'évolution des architectures et en examinant et en discutant les résultats.

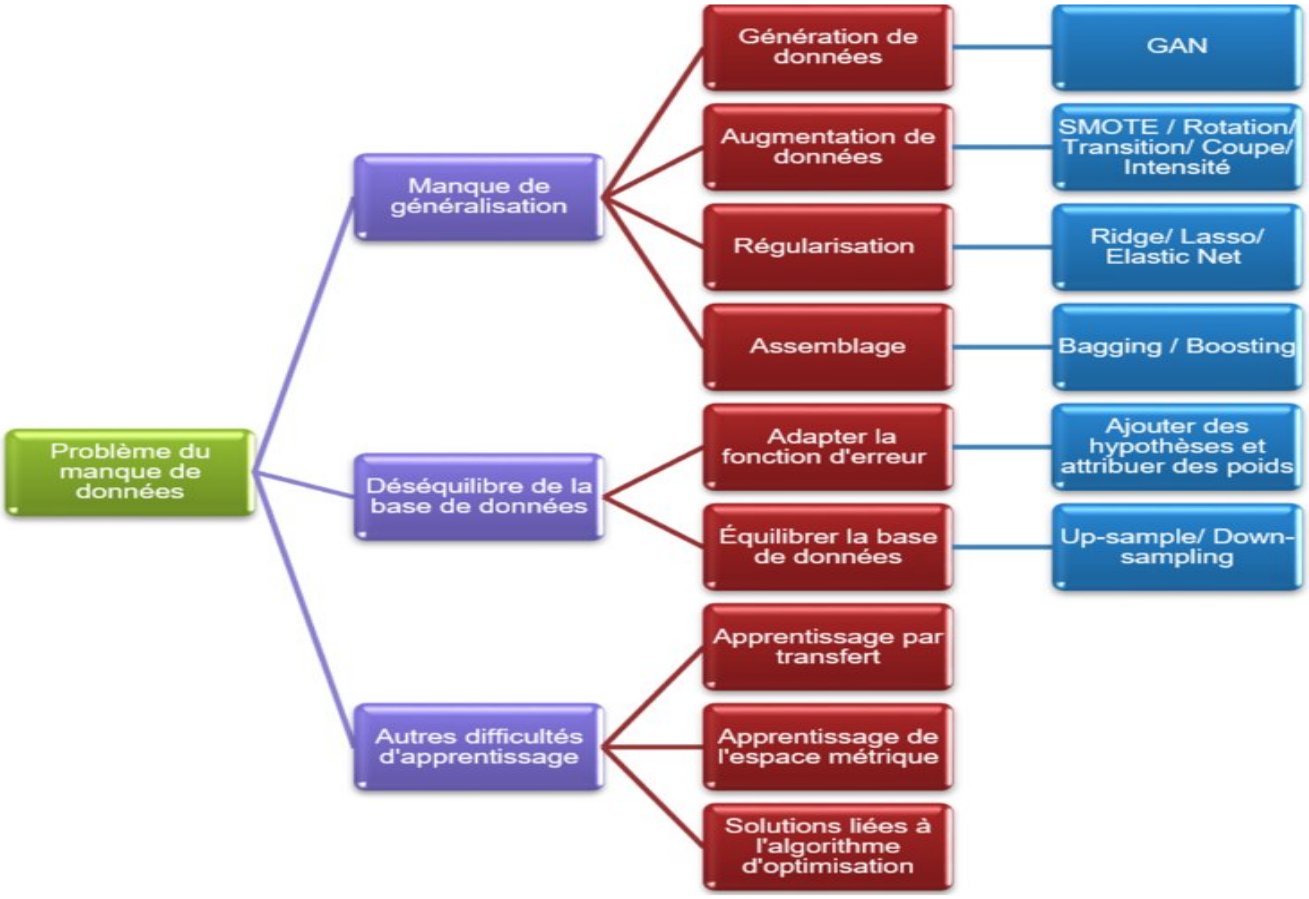


FIGURE 2.3 – Les implications du manque de données

2.1 Problématique et solution au manque de données

Le manque de données peut engendrer plusieurs problèmes d'apprentissage comme le manque de généralisation, le déséquilibre de la base de données et les défis d'optimisation (fig. 2.3).

Pour résoudre les difficultés de généralisation liées à la quantité de données, plusieurs approches sont proposées [71]. Nous avons déjà évoqué quelques solutions pour favoriser la généralisation des modèles en général, notamment la régularisation, la validation croisée et les dropouts (voir section 1.3.6.2). Dans cette section, nous évoquons trois autres solutions répandues liées surtout à la quantité de données.

Réseau Adverse Génératif (GAN) : La génération des données en utilisant les réseaux de neurones est une nouvelle tendance pour lutter contre l'instabilité des modèles et permettre la généralisation [211][194]. Goodfellow et al. [79] ont introduit en 2014 les réseaux de neurones adverses génératifs. Il s'agit d'un algorithme qui fait intervenir deux réseaux de neurones :

— Un générateur (G) : Générer des images à partir, par exemple, d'un vecteur de bruit avec l'objectif de devenir de plus en plus performant pour la génération des images proches des images réelles.

— Un discriminateur (D) : Distinguer les images réelles des images générées par le générateur avec l'objectif de devenir de plus en plus performant à trouver les différences qui sont de plus en plus légères.

Les deux réseaux sont entraînés simultanément en parallèle. Lors de l'entraînement, (D) est alimenté par des images réelles et des images générées par (G) et espère distinguer leurs provenances. En parallèle, le générateur apprend à générer des images qui permettent de tromper le classificateur (D) [123] [210].

Les modèles génératifs offrent des perspectives variées dans plusieurs secteurs, notamment dans le domaine médical [285] et dans le multimédia [69]. Cependant, comme le cas de toutes les technologies aussi puissantes, ils pourraient engendrer des multiples problèmes éthiques telles que les fake news et la manipulation criminelle. En revanche, ils s'avèrent aussi que les

GAN présentent des solutions efficaces pour détecter les manipulations des images [88][166] ou le deep fake [275]. Ainsi, les dangers qui viennent avec cette technologie, ne doivent pas être un obstacle à son progrès. Néanmoins, il est important d'établir une régulation et des solutions de traçabilité des données.

Augmentation de données : L'augmentation de données consiste à effectuer des modifications sur les images sans toucher à l'essentiel de leurs contenus, notamment la rotation, introduire un bruit ou un flou, zoomer, modifier le contraste et la luminosité. L'objectif n'est donc pas de créer de nouvelles données, mais d'apporter une certaine variabilité dans les données en favorisant la généralisation des modèles.

Méthodes d'assemblage : Le bagging et le boosting représentent des solutions au manque de données. En effet, par analogie à la façon de prise de décision de l'être humain, ces deux méthodes ensemblistes font intervenir plusieurs modèles spécialisés sur des parties de la base de données pour recueillir plus avis sur la bonne prédiction [193].

Le bagging consiste à sous-échantillonner l'entraînement sur des parties et de générer plusieurs modèles pour chaque partie de données. Ainsi, il faut moyenner dans le cadre d'une régression ou organiser un vote des différentes prédictions dans le cadre d'une classification. Le bagging est déployé surtout aux modèles à forte variance pour favoriser leur stabilité. Le principe du boosting se différencie du bagging par le fait que les différents classificateurs sont pondérés à chaque prédiction avec un poids plus fort quand la prédiction est incorrecte.

Le bagging et le boosting réduisent la variance de l'estimation en combinant plusieurs estimateurs provenant de différents modèles, ainsi le modèle résultant est plus stable. Les deux méthodes produit N modèles à partir d'un seul, mais le boosting fait en sorte d'ajouter de nouveaux modèles qui réussissent là où les modèles précédents échouent. Les deux techniques produisent plusieurs minibases d'entraînement par échantillonnage aléatoire, mais seul le boosting alloue des pondérations aux données afin de favoriser les données les plus difficiles. Si le problème est une faible performance, le boosting est le meilleur candidat puisqu'il pourrait générer des erreurs plus faibles. Si le problème est le sur-ajustement, le bagging est le meilleur

candidat puisque le boosting peut augmenter ce problème de sur-ajustement aux données avec son approche de pondération.

Les solutions mentionnées au-dessus sont quelques exemples liés au manque de données. D'autres solutions sont proposées pour résoudre le problème des bases de données déséquilibrées (fig. 2.3), notamment la manipulation de la fonction d'erreur ou l'échantillonnage. Dans le reste de chapitre, nous abordons d'autres solutions pour résoudre les problèmes d'apprentissage sur de petits jeux de données. Pour commencer, il faut bien définir deux termes largement répandus : Few-shot learning (la problématique de manque de données) et le meta-apprentissage (une approche d'apprentissage qui propose des solutions permettant de résoudre la problématique de few-shot learning) (voir fig. 2.4).

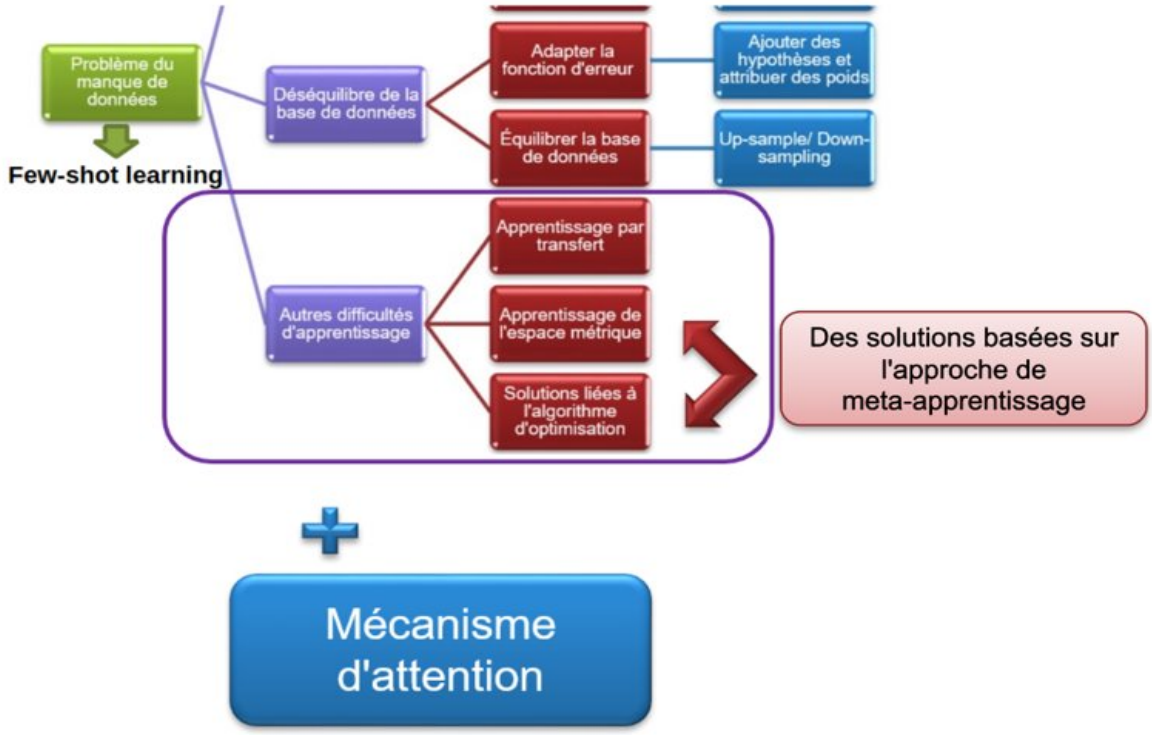


FIGURE 2.4 – Few-shot learning, meta-learning et mécanisme d'attention

2.2 Few-shot learning

La problématique d'apprentissage à partir de peu d'échantillons est appelée few-shot learning. Résoudre le problème d'optimisation de la minimisation de la fonction d'erreur

d'un réseau profond avec que quelques échantillons pour l'entraînement est une mission difficile puisque cette fonction peut dévier facilement de la solution optimale. La communauté scientifique de la recherche s'est intéressée à ce problème depuis quelques années et plusieurs techniques sont désormais disponibles [271][29]. Parmi les solutions développées les plus répandues, nous pouvons citer l'approche meta-apprentissage qu'on va définir dans la section suivante avec ses multiples techniques soulignées parmi les solutions de l'état de l'art dans la section 2.4.

La structure classique d'une base de données pour l'apprentissage profond se divise principalement en deux parties d'entraînement et de test (fig. 2.6). Le few-shot learning respecte la même structure. Nous pouvons référer au "few-shot learning" par "N-way K-shot learning"(fig. 2.5). N désigne le nombre de classes ou "ways" et K désigne le nombre des échantillons ou "shots" par classe dans le cadre d'une classification [196].

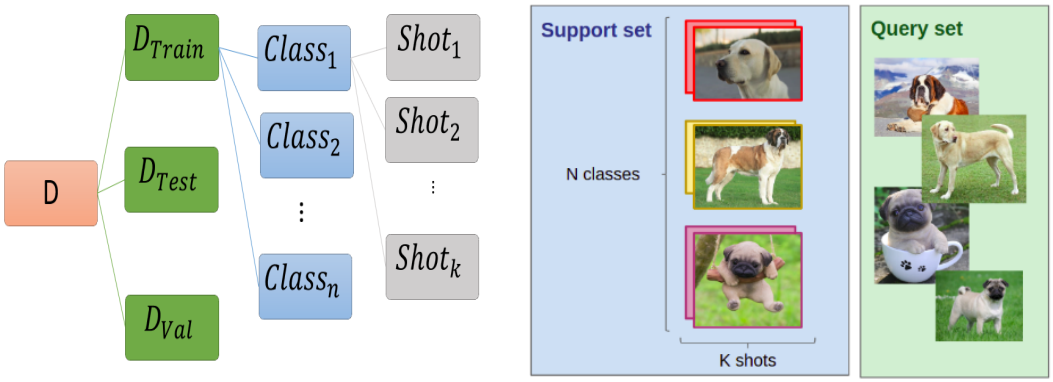


FIGURE 2.5 – La structure de la base de données de N-way K-shot learning

Comme illustré dans la figure 2.5, nous définissons le problème de classification d'images N-way K-shot comme suit : étant donné un ensemble de données pour l'entraînement (Support set), nous espérons classer les images de l'ensemble de test ou de requête (Query set) parmi les N classes. En effet, lorsque K est petit, nous parlons de few-shot learning. Le one-shot learning et le zero-shot learning désignent des algorithmes ayant le même principe avec l'hypothèse d'avoir respectivement une seule paire (K=1) ou zéro paire (K=0) de données d'apprentissage en entrée [132][292][280]. Le défi ici n'est pas le manque de données étiquetées, mais plutôt la rareté des données elles même.

Le concept de zero-shot learning se base sur la possession d'une idée générale sur les caractéristiques fondamentales de l'image telles que l'apparence ou les propriétés d'un certain objet [202][199]. Ainsi, nous pouvons conclure que le one-shot learning ou le few-shot learning sont justement une extension de l'hypothèse de zero-shot learning[57][64][289].

Il s'agit de penser à gagner de l'expérience à partir d'autres tâches, d'où la considération de few-shot learning comme un problème de meta-apprentissage.

2.3 Meta-apprentissage

Le meta-apprentissage est une révolution dans le domaine de recherche en intelligence artificielle. Le principe de l'intelligence artificielle et en particulier l'apprentissage profond est d'essayer d'imiter l'apprentissage de l'être humain [45] [171]. Cependant, un modèle qui ne maitrise qu'une seule tâche ne peut pas être considéré comme vraiment intelligent. Dans cette perspective, le meta-apprentissage entre en jeu. Le meta-apprentissage est désormais proposé pour rendre l'apprentissage plus généralisé, rapide, flexible aux changements et adaptable à différentes tâches avec un jeu de données réduit et quelques itérations d'entraînement [97][62].

Le terme meta-apprentissage fait référence à l'approche "apprendre à apprendre". En 1998, Thrun et al. [252] ont déclaré qu'étant donné une tâche à effectuer, un algorithme apprend "si sa performance à la tâche s'améliore avec l'expérience" et étant donné un ensemble de tâches à effectuer, un algorithme apprend si "sa performance à chaque tâche s'améliore avec l'expérience et avec le nombre de tâches effectuées". Il s'agit exactement du concept du meta-apprentissage. En effet, un modèle n'apprend pas à effectuer une certaine tâche spécifique, il apprend au fur et à mesure à effectuer de nombreuses tâches. Ainsi, d'une tâche à une autre, il améliore sa façon d'apprendre : il apprend à apprendre [253].

Au lieu de concevoir les modèles pour extraire les caractéristiques, nous espérons concevoir une architecture qui apprend la meilleure façon d'apprendre pour aboutir aux meilleurs résultats. Par conséquent, la mission à court terme est de définir la fonction de prédiction pour la tâche actuelle, mais également d'une tâche à une autre, le modèle conclut la manière la

plus optimale pour apprendre. Le défi de l'approche de meta-apprentissage est d'entraîner un modèle sur différentes tâches d'apprentissage et d'utiliser les connaissances acquises lors de ces expériences pour résoudre de nouvelles tâches d'apprentissage avec une base limitée de données et moins d'itérations de descente de gradient nécessaires lors de l'entraînement [6][65].

Ravi et al. [196] se sont intéressés à résoudre la problématique de Few-shot learning en utilisant le meta-apprentissage. Ils ont détaillé dans leur travail la structure du jeu de données utilisée dans l'approche de meta-apprentissage. En effet, dans une approche classique d'apprentissage profond, la base de données est divisée en un jeu de données d'entraînement et un jeu de données de test (fig. 2.6).

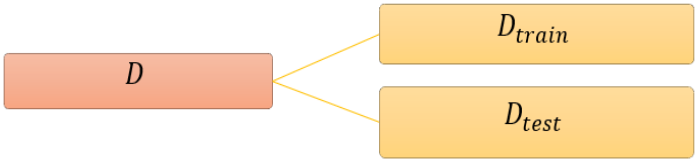


FIGURE 2.6 – Jeu de données pour une approche d'apprentissage classique

Dans l'approche de meta-apprentissage, l'entraînement est divisé en époques, les époques sont composées des épisodes et l'ensemble de données est en fait un meta-ensemble divisé entre les épisodes. Chaque épisode consiste à une tâche de classification avec une phase d'entraînement à n classes de k échantillons (support set) et une phase de test sur une minibase de test (query set)(fig. 2.7).

Ainsi, une tâche de classification de N-way K-shot est intégrée dans l'approche de meta-apprentissage comme un meta-test. Pour le meta-entraînement, nous utilisons une grande base de données disponible. Le processus de meta-entraînement comporte un certain nombre d'épisodes. Les épisodes sont composés en respectant la même structure de notre tâche de classification de meta-test. Par conséquent, pour chaque épisode, nous échantillonnons N classes et K images de la base de meta-entraînement pour le support set et quelques images pour le query set. À la fin de chaque épisode, les poids du modèle sont mises à jour d'une façon à minimiser l'erreur sur le query set. C'est ainsi que le modèle gagne en généralisation et

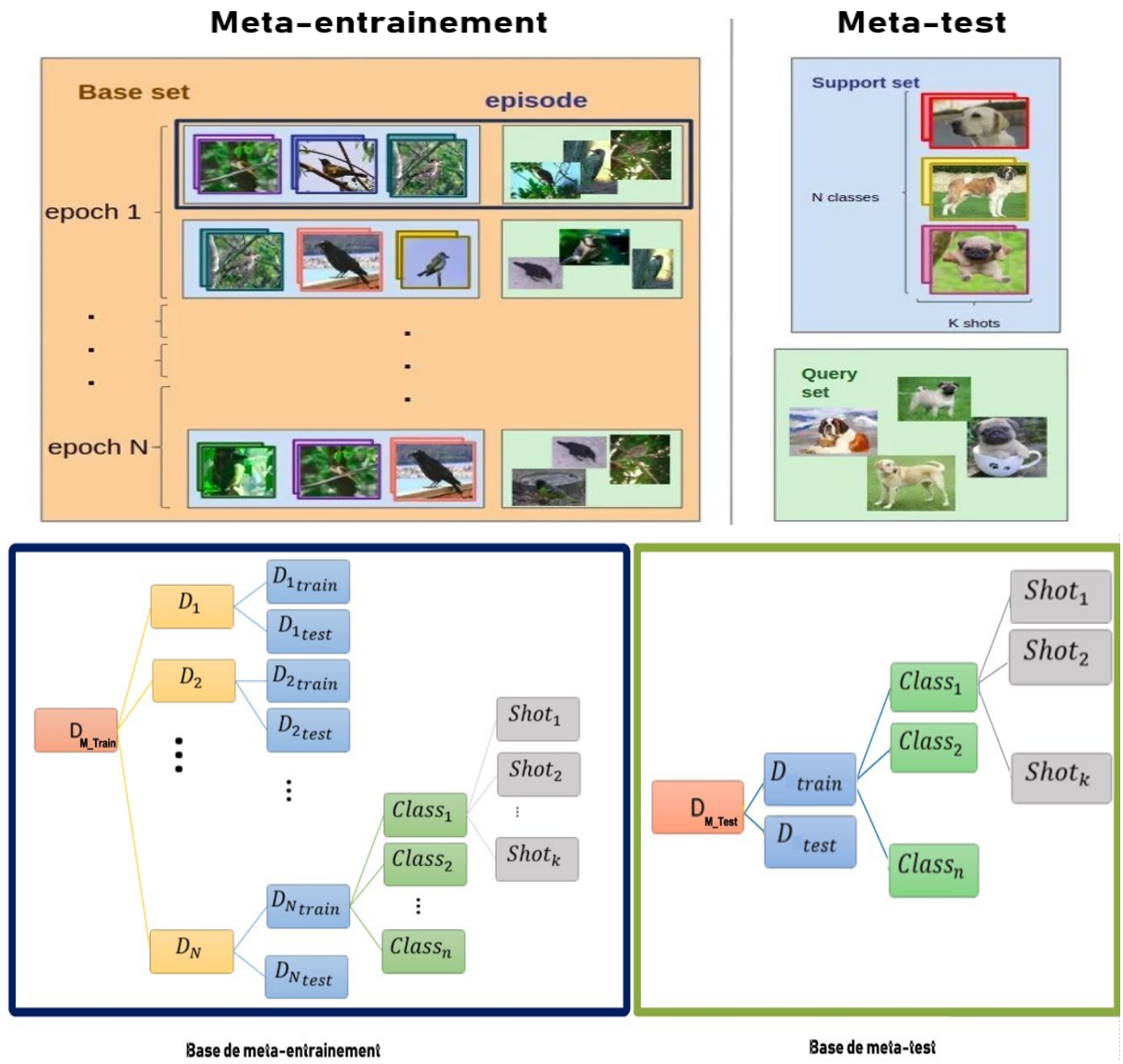


FIGURE 2.7 – Jeu de données pour une approche de meta-apprentissage : L'ensemble de données est en fait un meta-ensemble divisé entre les épisodes qui composent les époques. Nous distinguons deux processus d'apprentissage et deux apprenants : le meta-apprenant, un modèle qui apprend à travers les épisodes, et un deuxième modèle appelé base-apprenant, incorporé et entraîné à l'intérieur d'un épisode par le meta-apprenant. Considérons une époque de meta-apprentissage composée de plusieurs tâches de classification (plusieurs épisodes). Au cours d'une tâche de classification T définie par un support set de $N * K$ images et un query set de Q images, base-apprenant est initialisé et entraîné sur le support set. Ensuite, il est appliqué sur les images de query set pour prédire leurs classes. À la fin de chaque épisode, les poids du meta-apprenant sont mises à jour en se basant sur la valeur de la fonction d'erreur de classification de query set. La différence entre les techniques mentionnées dans ce chapitre est l'approche de fonctionnement du meta-apprenant.

flexibilité d'un épisode à une autre. Finalement, la performance du modèle est mesurée sur la tâche de classification du meta-test.

a) La démarche du meta-apprentissage :

1. Acquérir des connaissances à travers plusieurs tâches.
2. Utiliser les connaissances acquises dans des nouvelles tâches.

L'un des défis de meta-apprentissage est la similarité entre les tâches précédentes et la nouvelle tâche.

Plusieurs techniques sont désormais proposées dans le cadre du meta-apprentissage pour résoudre le problème de manque de données. Ces techniques peuvent être divisées en deux catégories : les solutions liées à l'espace métrique et les solutions liées à l'algorithme d'optimisation.

Maintenant, qu'ont été définies la notion de la problématique de Few-shot learning et l'approche de meta-apprentissage, les deux sections suivantes sont dédiées à expliquer les techniques les plus utilisées dans l'état de l'art pour résoudre la problématique du manque de données ou de few-shot learning. La section 2.4 représente des solutions en lien avec l'apprentissage, parmi lesquelles il y a des solutions basées sur l'approche de meta-apprentissage (voir fig.2.8). La section 2.5 représente une revue de littérature des travaux incorporant le mécanisme d'attention permettant aussi d'améliorer l'apprentissage, surtout avec le défi de manque de données.

2.4 Solutions de l'état de l'art

2.4.1 Apprentissage par transfert

L'apprentissage par transfert est une technique très largement utilisée où les informations collectées à partir d'une abondance de données dans une étape à laquelle nous nous référons par un pré-entraînement sont utilisées pour fournir un meilleur point de départ que l'initialisation aléatoire des poids du réseau de neurones. En effet, l'initialisation arbitraire de paramètres n'est pas considérée la façon la plus optimale. Dans cette perspective, plusieurs techniques cherchent à trouver les valeurs optimales d'initialisation de paramètres, notamment dans l'apprentissage

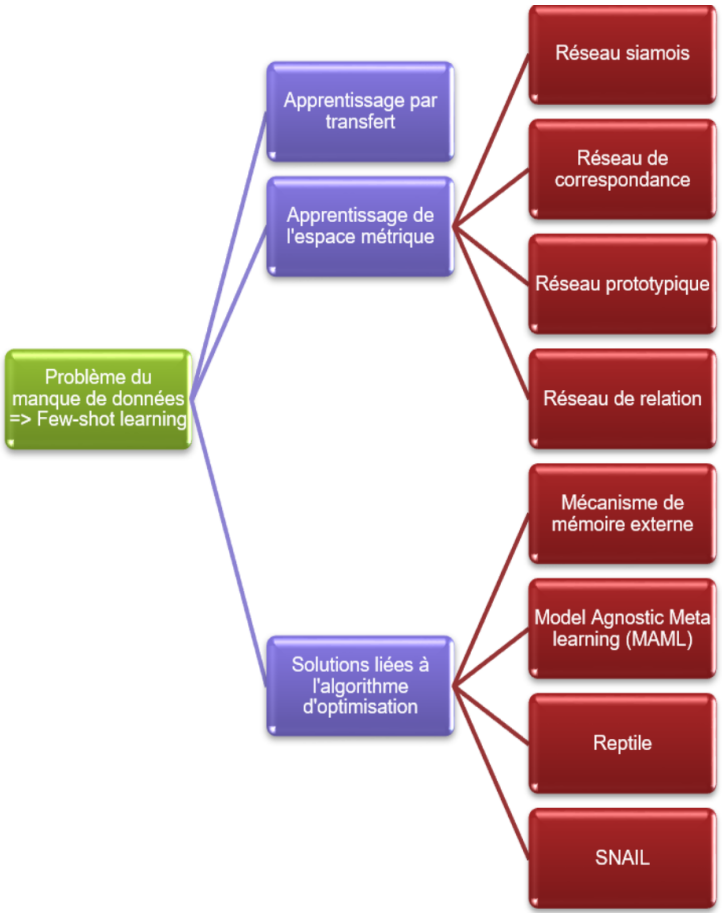


FIGURE 2.8 – Solutions de l'état de l'art

par transfert [184][185]. Au lieu de partir d'une initialisation arbitraire pour apprendre une tâche, un transfert de connaissances d'un ancien apprentissage d'une autre tâche est réalisé. L'apprentissage profond par transfert est une approche dans laquelle nous utilisons les premières couches d'un modèle déjà entraînées pour une autre tâche comme premières couches d'un nouveau modèle [250]. L'apprentissage par transfert permet un apprentissage plus rapide, plus efficace et plus optimale avec un besoin réduit de données pour effectuer une nouvelle tâche [273][20].

Afin d'expliquer l'apprentissage par transfert, deux termes doivent être définis : le domaine et la tâche.

On définit un espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$ [45] :

$$\begin{cases} \Omega & \text{Espace fondamental : } \mathcal{P}(\Omega), \text{ ensemble des évenements possibles} \\ \mathcal{F} & \text{Sigma-algebra est un ensemble de sous-ensembles des évenements } \Omega \Rightarrow \mathcal{F} \subseteq \mathcal{P}(\Omega) \\ \mathcal{P} & \text{Mesure de probabilité tel que } \mathcal{P}(\Omega) = 1 \end{cases}$$

Un domaine est noté $D = (\Omega, \mathcal{P}(X))$ [185]

$$\begin{cases} \Omega & \text{L'espace fondamentale} \\ \mathcal{P}(X) & \text{Distribution de probabilité telle que } X = x_1, ..., x_n \in \Omega \end{cases}$$

Un vecteur $\mathcal{P}$ de N éléments est appelé un vecteur de distribution de probabilité si

$$\begin{cases} \mathcal{P}(i) & \text{est une valeur positive } \forall\, i \in [|\, 1, N\, |]. \\ \sum_{i\,\in\,\Omega} \mathcal{P}(i) = 1 & \text{(Loi des probabilités totales)} \end{cases}$$

Une tâche est noté $T = (y, f(x))$ [185]

$$\begin{cases} y & \text{Espace des classes de sortie} \\ f(x) & \text{Fonction de prédiction notée en tant qu'une fonction de probabilité conditionnelle } \mathcal{P}(y \mid x) \end{cases}$$

Ainsi, **l'apprentissage par transfert** peut être défini comme suit [250] : Ayant une tâche d'apprentissage T_t dans un domaine D_t et un modèle pré-entraîné dans un autre domaine D_s pour effectuer une tâche T_s. L'apprentissage par transfert tente d'améliorer la performance de la fonction de prédiction $f_T()$ pour une tâche T_t en utilisant les connaissances apprises en D_s pour la tâche T_s, avec $D_s \neq D_t$ et/ou $T_s \neq T_t$. En outre, la taille de la base de données de D_s est beaucoup plus importante D_t, $N_s \gg N_t$ favorisant l'apprentissage pour la tâche T_s. Ainsi, c'est important de pouvoir utiliser le savoir acquis dans cette expérience d'apprentissage pour les prochains apprentissages (autres domaines ou autres tâches). Nous parlons d'un apprentissage par transfert profond si $f_T()$ est fonction non linéaire représentant un réseau de neurones.

Démarche : L'apprentissage profond a deux variantes. La première utilise l'apprentissage par transfert comme technique d'extraction de caractéristiques et utilise ensuite un autre classificateur tel que la machine à vecteurs de support (SVM), la forêt aléatoire (RF), K-Nearest Neighbors (KNN) et d'autres techniques différentes. Cependant, de nombreuses expériences ont montré que ces techniques ont des performances limitées avec des données limitées [144]. La deuxième variante fait intervenir un ajustement ou un réglage fin. Pour cette dernière variante, le processus de l'apprentissage par transfert peut se diviser en deux étapes :

a) Extraction des caractéristiques ou "feature extraction"

1. Sélectionner un domaine source dans lequel un modèle est pré-entraîné avec une large base de données de haute qualité.
2. Éliminer quelques couches et en particulier la dernière couche Softmax.
3. Rattacher des couches adaptées au domaine et au jeu de données cibles.
4. Entrainer la totalité du modèle pour la nouvelle tâche en gardant les premières couches pré-entraînées figées.

b) Ajustement ou "Fine Tuning"

On entraine une partie ou la totalité du modèle, y compris les premières couches pré-entraînées, avec un taux d'apprentissage très faible.

Limites et perspectives : Cette approche nécessite quand même un pré-entraînement et l'ajustement à la nouvelle tâche peut aussi être un vrai défi. En effet, dans certains cas, une large quantité de données peut être nécessaire pour l'ajustement. Ainsi, l'adaptation au domaine et l'apprentissage par transfert de multiples domaines constituent l'une des préoccupations des chercheurs de nos jours.

Plusieurs études sont menées pour améliorer les résultats de l'apprentissage par transfert, notamment une combinaison entre le meta-apprentissage et l'apprentissage par transfert, illustrée dans la figure 2.9, proposée par Sun et al. [240] pour résoudre un problème de few-learning. Cette figure nous souligne déjà la différence entre les différentes approches.

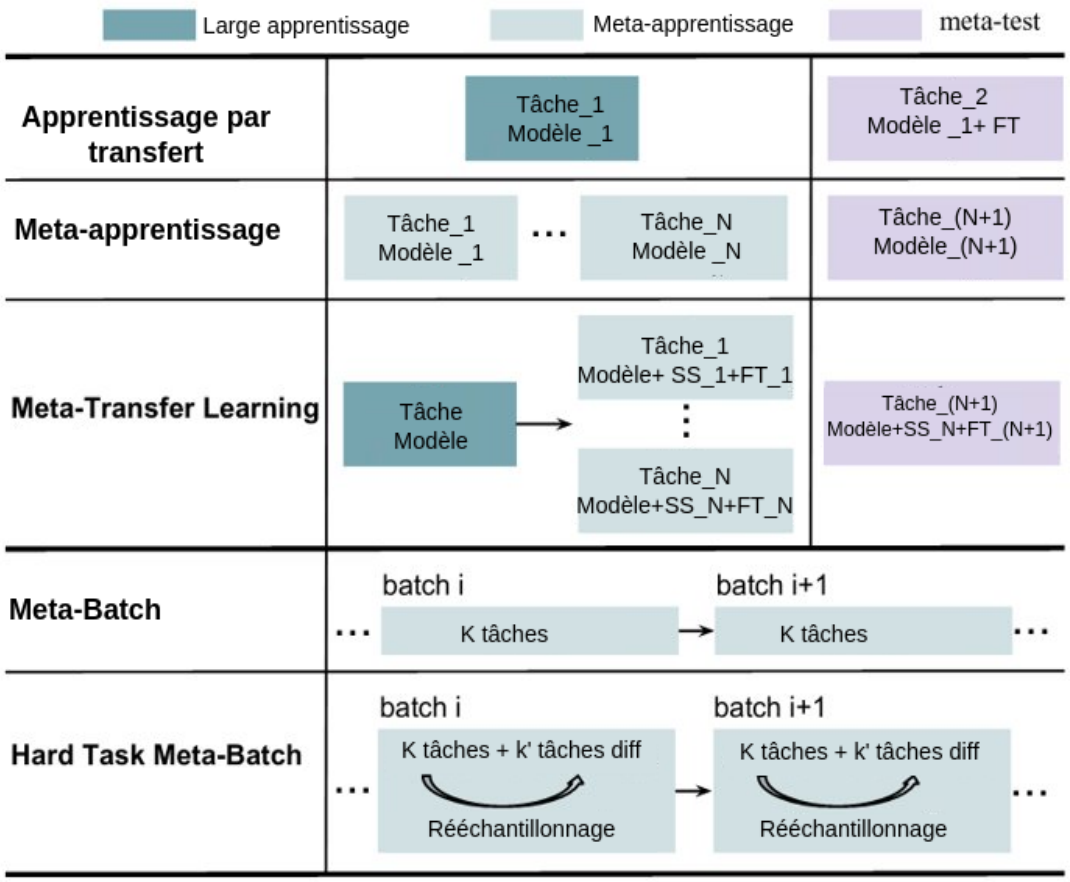


FIGURE 2.9 – Combinaison de meta-apprentissage et d'apprentissage par transfert : FT désigne le fine tuning. SS désigne "Scaling and Shifting", une opération de transfert simple définit par Sun et al. [240]. Le Meta-Transfer Learning représente l'approche principale. Le Hard Task Meta-Batch est une stratégie d'entraînement qui favorise les tâches difficiles.

2.4.2 Apprentissage de l'espace métrique

L'approche consiste à créer un espace métrique dans lequel les caractéristiques de nos données sont repérées. Ainsi, les nouvelles données d'entrée sont comparées dans cet espace. Cet espace métrique est considéré comme un espace discriminant créé à l'aide d'une métrique de distance afin de calculer la similarité et la dissimilarité des caractéristiques entre différentes images de différentes classes pour positionner les images de la même classe aussi proches que possible les unes des autres et aussi éloignées que possible des images de classes différentes. Cette approche est utilisée dans plusieurs algorithmes pour résoudre la problématique de few-shot learning grâce à la possibilité de la combinaison de plusieurs méthodes d'extractions de caractéristiques représentatives et des techniques variées de la comparaison des représentations. En 2009, Palatucci et al. [183] ont introduit la notion de

classificateur basée sur les relations sémantiques (SOC). Ils proposent explicitement un espace de caractéristiques sémantiques conçu manuellement et un algorithme permettant de faire correspondre chaque nouvelle sortie à un point de cet espace pour conclure la classe. Socher et al. [233] ont étendu cette idée de l'espace de caractéristiques sémantiques par une approche basée sue le réseau de neurones. Dans la suite de cette section, nous citons les principaux derniers réseaux de neurones qui utilisent des approches métriques.

2.4.2.1 Réseau siamois

Le principe d'un réseau siamois est de donner un ensemble de données d'entrée sous la forme de paires [273]. La seule labellisation nécessaire est de dire si la paire contient des données appartenant à la même classe ou à deux classes différentes. En 2015, Kochet et al.[125] ont proposé les réseaux siamois pour résoudre le problème de few-shot learning. Le modèle est formé de deux réseaux neuronaux convolutifs jumeaux permettant l'extraction des caractéristiques des images d'entrée. Au début de l'entraînement, le modèle reçoit donc une paire d'images. Les représentations résultantes alimentent à un autre réseau pour les comparer et décider si elles appartiennent à la même classe ou à des classes différentes. Ensuite, le modèle est déployé pour une tâche de classification de few-shot learning, où chaque image de requête est comparée à toutes les images du support set et assignée à la classe la plus proche.

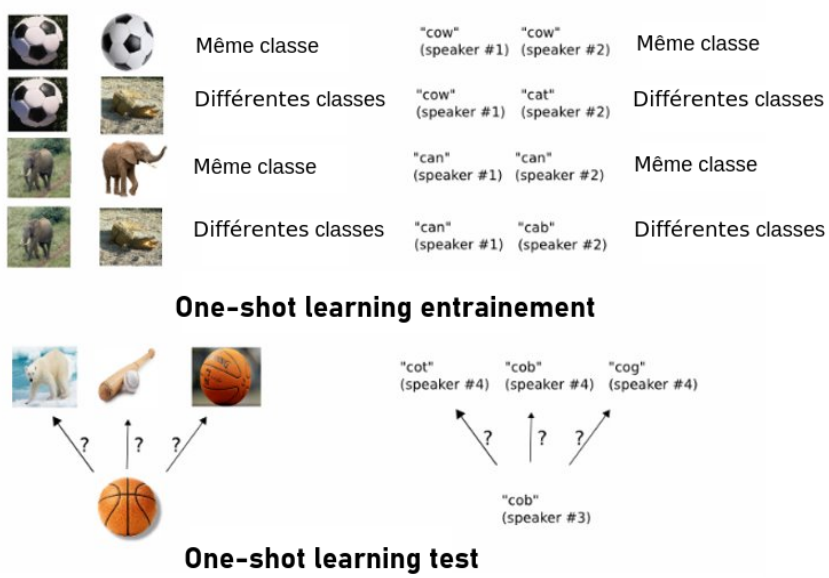


FIGURE 2.10 – Illustration de l'approche du réseau siamois [125]

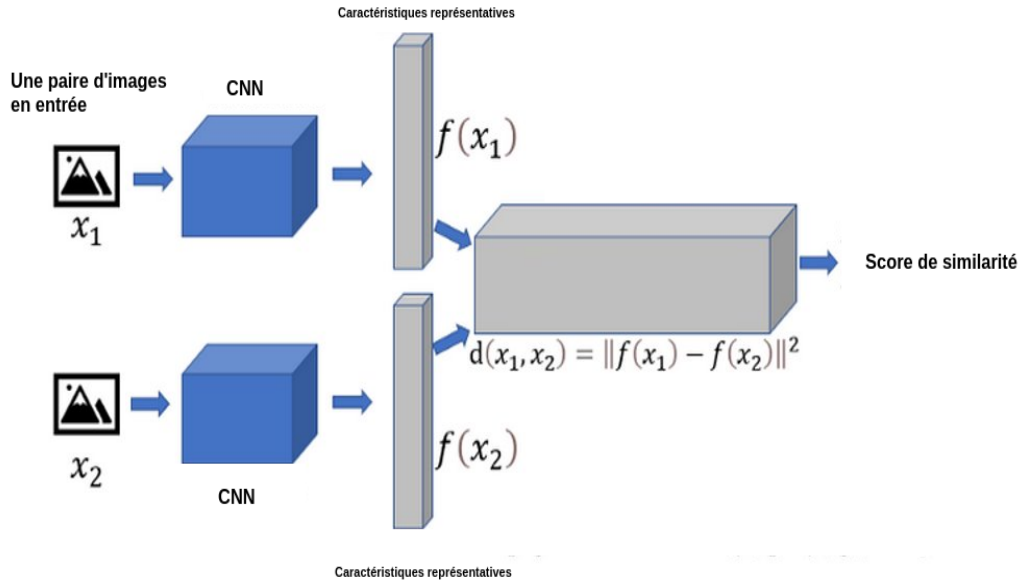


FIGURE 2.11 – Architecture du réseau siamois [117]

Les réseaux siamois sont largement utilisés pour résoudre le problème de few-shot learning. Néanmoins, il est à noter que la tâche sur laquelle ils sont entraînés (comparaison d'une paire d'images) diffère de la tâche cible finale (classification) et l'algorithme exige une structure de paire de données.

Plusieurs adaptations sont proposées, notamment l'introduction d'un programme bayésien (BPL) [133]. L'idée de cet algorithme est de faire apprendre à notre réseau une fonction de calcul de distance avec un seuil de degré de similarité comme hyperparamètre [248]. Une autre adaptation consiste à l'utilisation de la fonction d'erreur de triplets qui prend en entrée des triplets d'entrées [219] (voir fig. 2.12).

2.4.2.2 Réseau de correspondance (Matching network)

En 2016, Vinyals et al. [264] propose une amélioration au réseau siamois. Leurs réseaux de correspondance classifient les images requêtes en comparant leurs vecteurs de caractéristiques aux vecteurs des images du support set. Les résultats prouvent que leur approche surclasse les réseaux siamois et les réseaux de neurones à mémoire augmentée.

L'algorithme utilise une grande base de données pour entrainer le modèle à trouver les meilleurs vecteurs représentatifs des images comme souligné dans l'approche de

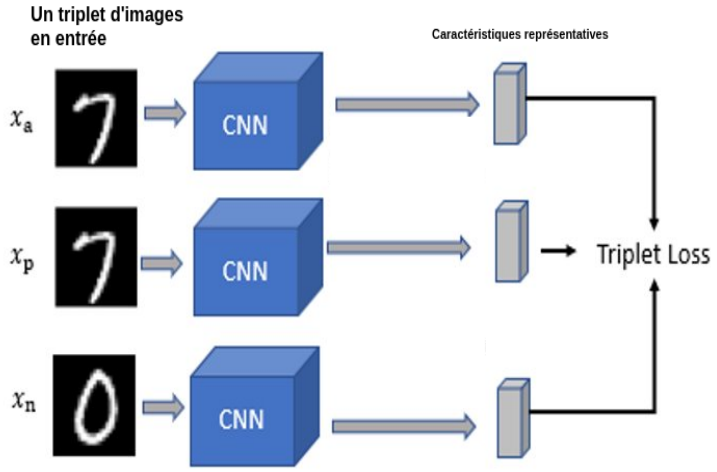


FIGURE 2.12 – Architecture du réseau siamois avec une fonction d'erreur triplets [117]

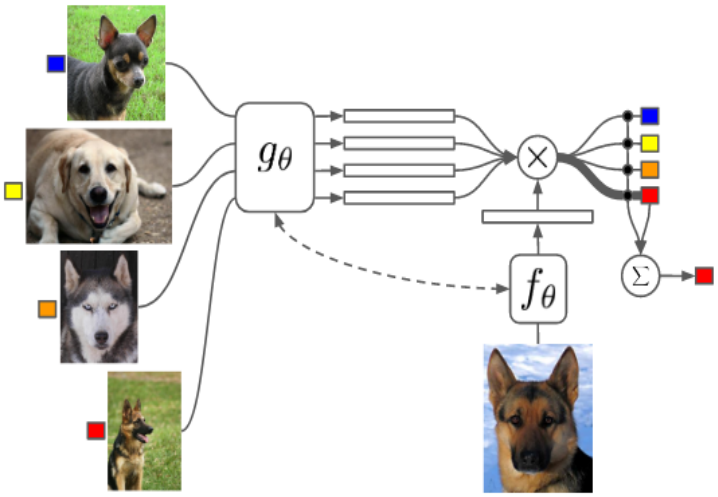


FIGURE 2.13 – Architecture du réseau de correspondance [264] : l'extraction des caractéristiques des images de support set et query set est effectuée respectivement par les modèles g_θ et f_θ. Les vecteurs caractéristiques alimentent une couche softmax suite à laquelle et en se basant sur la distance cosinus entre le vecteur caractéristique de l'image requête et les vecteurs caractéristiques des images du support set., l'image requête est classée.

meta-apprentissage (voir section 2.3). Dans chaque épisode, les images de support set et query set sont données en entrée à un réseau de convolutions qui sert à l'extraction des caractéristiques. Ensuite, chaque image de query set est classée à l'aide d'une couche softmax en se basant sur la distance cosinus entre le vecteur caractéristique de l'image requête et les vecteurs caractéristiques des images du support set. Finalement, une fonction d'erreur d'entropie croisée est calculée et les poids sont mis à jour (voir section 1.3.3). Par conséquent, les réseaux de correspondance apprennent à calculer des représentations des images et donc à

les classifier en les comparant aux autres représentations des images de différentes classes (fig. 2.13).

Contrairement à l'apprentissage classique, où les réseaux de neurones apprennent à extraire les caractéristiques pertinentes qui décrivent les classes, les réseaux de correspondance apprennent plutôt les caractéristiques pertinentes pour discriminer entre les différentes classes. Vinyals et al. font intervenir également un mécanisme d'attention (voir section 2.5) et un mécanisme de mémoire externe LSTM bidirectionnel (voir section 2.4.3.1) pour gagner en expérience. Cependant, il faut noter que ces améliorations pourrait être couteuses en termes du temps de calcul. Ce travail est proposé comme une solution au problème de one-shot learning.

2.4.2.3 Réseau prototypique (Prototypical Networks)

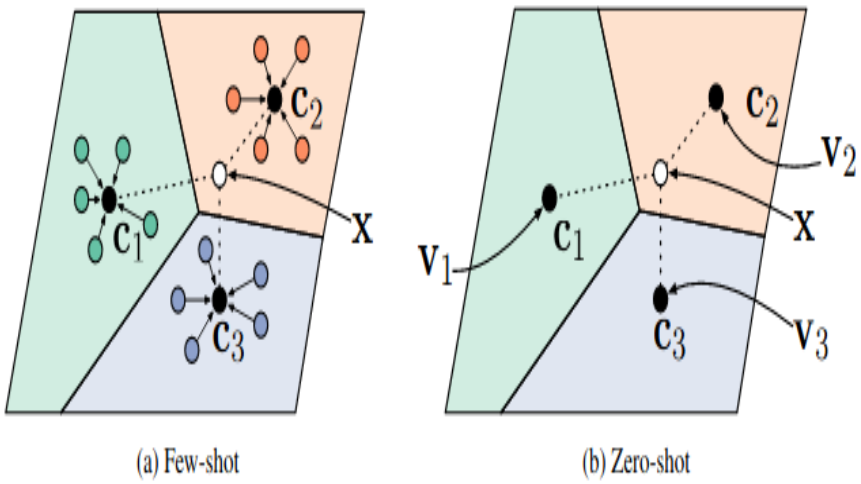


FIGURE 2.14 – Architecture du réseau prototypique [232]

Contrairement au réseau de correspondance où les images du query set sont comparées à toutes les images de support set, dans les réseaux prototypiques, une représentation moyennée des K échantillons de chaque classe du support set, appelée prototype, est positionnée dans l'espace métrique. Ainsi, les vecteurs représentatifs des images du query set sont ensuite comparés uniquement à ces représentations moyennées des différentes classes (fig. 2.14). Une autre particularité des réseaux prototypiques consiste à utiliser la distance euclidienne au lieu de la distance cosinus. Dans le cadre d'un problème de one-shot learning, l'algorithme revient donc à celui des réseaux de correspondance. Les résultats de tâches de classification few-shot

learning et one-shot learning des réseaux prototypiques prouvent une amélioration par rapport aux réseaux de correspondance [232].

2.4.2.4 Réseau de relation (Relation network)

Le module de relation reçoit la concaténation de la sortie d'un module d'extraction des caractéristiques d'une image de query set et la sortie d'un modèle qui calcule un prototype pour chaque classe. Il produit en sortie un score de relation pour chaque couple (fig. 2.17). En appliquant un softmax à ces scores de relation, le modèle décide de la classe [242]. En outre, la fonction de distance n'est pas définie à l'avance, mais apprise par l'algorithme. Les réseaux de relation ont dépassé les performances des réseaux profonds traditionnels, siamois, prototypiques, de correspondance, à mémoire augmentée et MAML (Model-Agnostic Meta-Learning) dans des tâches de classification de zero-shot learning et few-shot learning (fig. 2.16).

Model	Fine Tune	5-way Acc.		20-way Acc.	
		1-shot	5-shot	1-shot	5-shot
MANN	N	82.8%	94.9%	-	-
CONVOLUTIONAL SIAMESE NETS	N	96.7%	98.4%	88.0%	96.5%
CONVOLUTIONAL SIAMESE NETS	Y	97.3%	98.4%	88.1%	97.0%
MATCHING NETS	N	98.1%	98.9%	93.8%	98.5%
MATCHING NETS	Y	97.9%	98.7%	93.5%	98.7%
SIAMESE NETS WITH MEMORY	N	98.4%	99.6%	95.0%	98.6%
NEURAL STATISTICIAN	N	98.1%	99.5%	93.2%	98.1%
META NETS	N	99.0%	-	97.0%	-
PROTOTYPICAL NETS	N	98.8%	99.7%	96.0%	98.9%
MAML	Y	98.7 ± 0.4%	**99.9 ± 0.1%**	95.8 ± 0.3%	98.9 ± 0.2%
RELATION NET	N	**99.6 ± 0.2%**	**99.8± 0.1%**	**97.6 ± 0.2%**	**99.1± 0.1%**

FIGURE 2.15 – Classification few-shot learning de Omniglot [242]

Model	FT	5-way Acc.	
		1-shot	5-shot
MATCHING NETS	N	43.56 ± 0.84%	55.31 ± 0.73%
META NETS	N	49.21 ± 0.96%	-
META-LEARN LSTM	N	43.44 ± 0.77%	60.60 ± 0.71%
MAML	Y	48.70 ± 1.84%	63.11 ± 0.92%
PROTOTYPICAL NETS	N	49.42 ± 0.78%	**68.20 ± 0.66%**
RELATION NET	N	**50.44 ± 0.82%**	65.32 ± 0.70%

FIGURE 2.16 – Classification few-shot learning de miniImagenet [242]

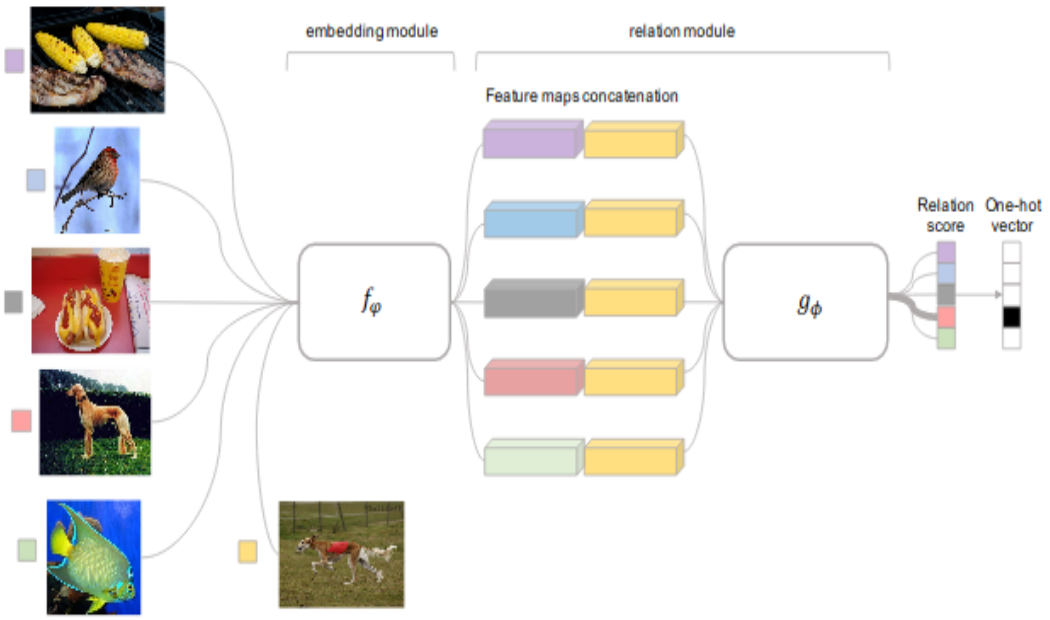


FIGURE 2.17 – Architecture du réseau de relation [242]

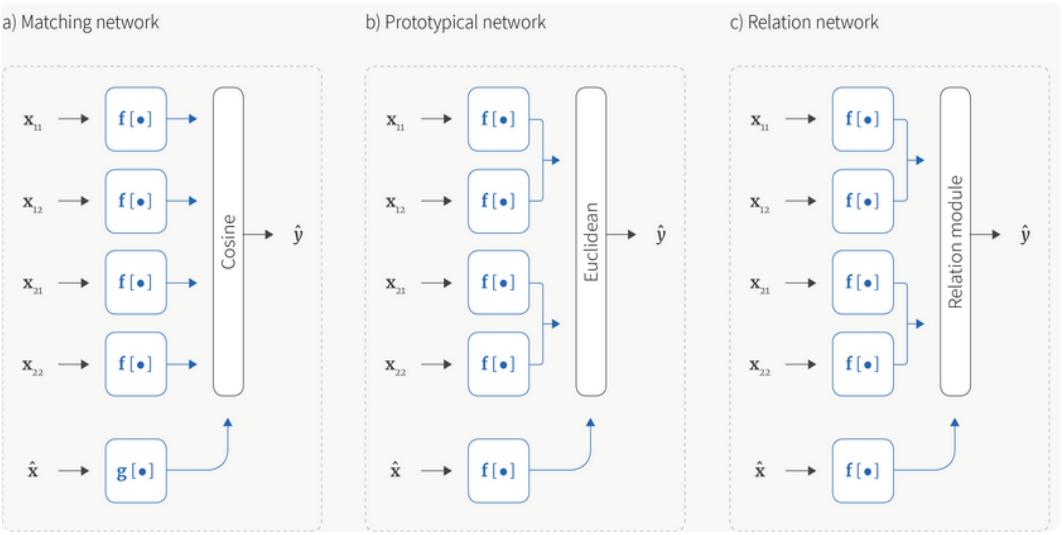


FIGURE 2.18 – Illustration de la comparaison des réseaux de neurones métriques [268]

La figure 2.18 résume la différence entre les trois derniers réseaux de neurones. Plusieurs travaux ont été proposés ces dernières années en se basant sur le calcul de la distance métrique. Les approches peuvent largement varier en fonction de la structure de l'architecture et de la variation des méthodes d'extraction des vecteurs caractéristiques et des méthodes de calcul de la distance entre ces vecteurs [52][9].

L'hypothèse contraignante de ces approches est la similarité et la distribution des données entre le meta-entraînement et le meta-test. Une autre limite est liée aux particularités du domaine d'application et le niveau de complexité de la classification à effectuer.

2.4.3 Des solutions d'apprentissage liées à l'algorithme d'optimisation

2.4.3.1 Mécanisme de mémoire externe

Par analogie à l'apprentissage humain, l'incorporation d'un mécanisme de mémoire est une perspective importante pour profiter davantage des expériences précédentes. Dans ce but, plusieurs travaux présentent des multiples approches basées sur des mécanismes de mémoire. Nous commençons tout d'abord par la définition d'un mécanisme important de mémoire appelé LSTM.

Réseaux à longue mémoire à court terme (LSTM) :

L'une des principales limites de l'architecture RNN simple est la disparition du gradient, qui rend difficile l'apprentissage des corrélations à long terme. Pour résoudre ce problème, une architecture RNN plus sophistiquée est proposée : les réseaux à longue mémoire à court terme ou "Long Short Term Memory networks" (LSTM). L'architecture traditionnelle des RNN est utile lorsque des informations un peu plus anciennes ne sont pas pertinentes pour la tâche actuelle. Cependant, lorsque l'écart entre les informations requises et la position dans laquelle elles sont utilisées est important, un mécanisme de mémoire spécifique doit être inclus. Ceci est lié principalement au fait qu'il est difficile d'apprendre les dépendances à long terme avec la descente de gradient [16][96]. Les LSTM sont toujours des RNN. La mémoire dans les LSTM est appelée cellules C_t utilisant l'état précédent H_{t-1} et l'entrée courante X_t. En effet, ces cellules décident des informations pertinentes à conserver et des informations à ignorer (fig. 2.19).

Le module répété dans une architecture RNN simple contient une seule couche. Cependant, le module répété des LSTM a une architecture différente composée de quatre couches en interaction (figure 2.19). Un LSTM utilise les portes (opération ponctuelle d'addition ou de

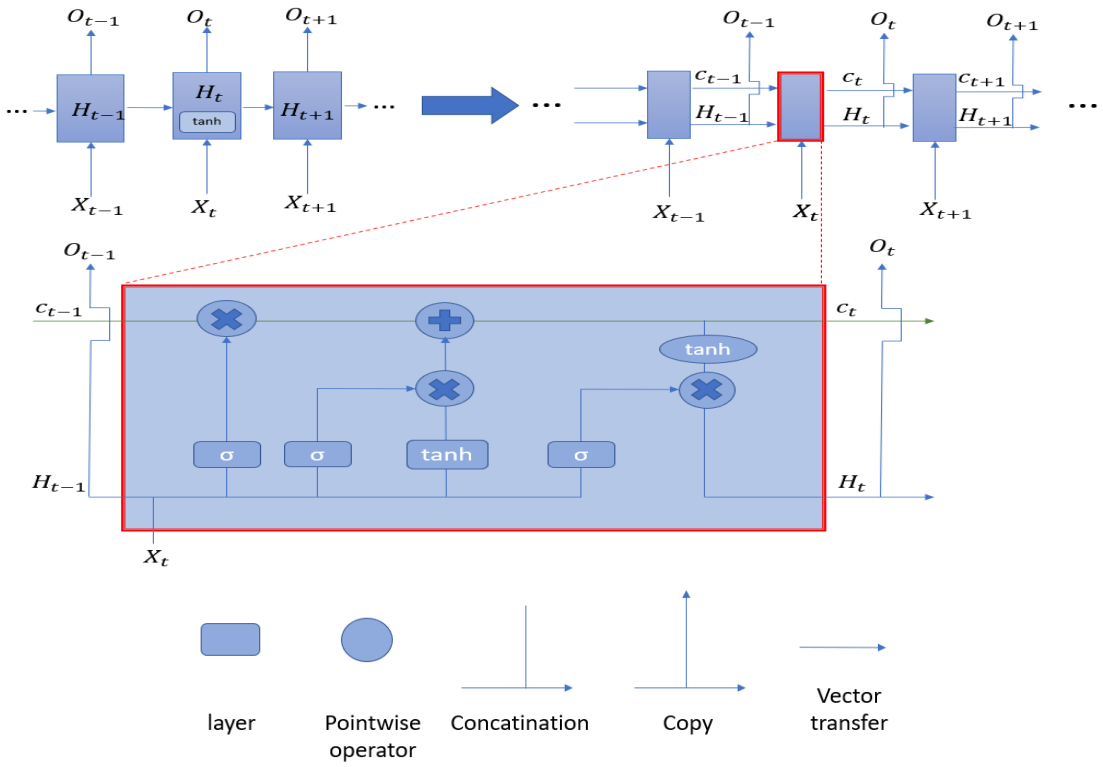


FIGURE 2.19 – Architecture LSTM

multiplication) pour contrôler les informations à conserver ou à ignorer dans l'état de la cellule C_t.

La couche sigmoïde ou tanh donne des nombres entre -1 et 1. Elle décide de la quantité d'information à conserver.

a) Porte d'oubli (Forget gate layer)

La porte d'oubli est la première étape du LSTM. Elle décide de l'information à ignorer ou à conserver dans l'état de la cellule en utilisant l'état caché H_{t-1} et l'entrée actuelle X_t et résulte une valeur entre 0 et 1 pour chaque nombre dans l'état de la cellule C_{t-1}. La valeur 1 est utilisée pour conserver entièrement l'information tandis que la valeur 0 est utilisée pour la rejeter complètement. La fonction de la première couche dénommée f_t est décrite comme suit

$$f_t = \sigma(v_f X_t + a_f H_{t-1} + b_f) \tag{2.1}$$

b) Mise à jour de l'état de la cellule

Cette étape décide des nouvelles informations pertinentes à stocker dans l'état de la cellule. Elle est composée de deux parties.

1. La porte d'entrée (Input gate layer) est une couche sigmoïde qui décide des valeurs à mettre à jour. Sa fonction qu'on désigne par e_t est la suivante :

$$e_t = \sigma(v_i X_t + a_i H_{t-1} + b_i) \tag{2.2}$$

2. Une couche de tanh créant un vecteur de nouvelles valeurs à ajouter à l'état de la cellule C_t. Sa fonction, qu'on désigne par N_t, est la suivante :

$$N_t = tanh(v_N X_t + a_N H_{t-1} + b_N) \tag{2.3}$$

On utilise les deux dernières étapes pour mettre à jour l'ancien état de la cellule C_{t-1} et créer un nouvel état de la cellule C_t. Tout d'abord, l'ancien état C_{t-1} est multiplié par f_t pour oublier les informations non pertinentes et ajouter $e_t N_t$ les nouvelles valeurs candidates.

$$C_t = f_t C_{t-1} + e_t N_t \tag{2.4}$$

c) Sortie

La sortie dépend de l'état de la cellule, du dernier état caché et de l'entrée actuelle. Tout d'abord, nous utilisons une couche sigmoïde sur l'entrée actuelle et le dernier état caché pour décider de l'information à produire de l'état de la cellule. Ensuite, on inclut l'état de la cellule par tanh en donnant des valeurs entre -1 et 1. Enfin, on le multiplie par la sortie de la porte sigmoïde.

H_t est calculé comme suit :

$$H_t = \sigma(v X_t + a H_{t-1} + b) tanh(C_t) \tag{2.5}$$

Meta-LSTM : Ravi et al. [196] ont proposé un réseau d'optimisation basé sur le principe des LSTM. Les poids θ de base-apprenant sont représentés par l'état des cellules du LSTM, ce qui conduit à une mise à jour classique de LSTM. Nous pouvons faire une analogie directe entre une étape de mise à jour LSTM et une rétro propagation du gradient descendant avec $f_t = 1$ et e_t le taux d'apprentissage. Par conséquent, ce modèle apprend et mémorise son apprentissage.

Les réseaux de neurones à mémoire augmentée : Santoro et al. [215] ont introduit les réseaux de neurones à mémoire augmentée ou "memory augmented neural network" (MANN) avec l'idée de base que de nouvelles images provenant de nouvelles classes pourraient être classées en utilisant des informations stockées sur la classification d'images précédentes. Il utilise une architecture permettant d'utiliser l'apprentissage antérieur à chaque étape, composée d'un RNN (contrôleur) et d'une mémoire augmentée. Cette solution peut résoudre des problèmes de few-shot learning [216]. Leur modèle utilise un réseau neuronal récurrent qui apprend à la fois comment stocker et comment récupérer des informations pertinentes à partir de données antérieures. Le réseau est constitué d'un contrôleur (un LSTM ou un réseau de type feed-forward) qui stocke les mémoires dans un réseau et les récupère pour les utiliser dans la classification (fig. 2.20). La mémoire récupérée est une somme pondérée de toutes les mémoires stockées, pondérée par les similarités cosinusoïdales transformées en soft-max. Au fil du meta-entraînement, le contenu de la mémoire s'adapte à la tâche actuelle et la classification s'améliore. Ce travail est similaire aux travaux de Hochreiter et al. [95] qui utilisent aussi les architectures RNN avec mémoire [94] [96]).

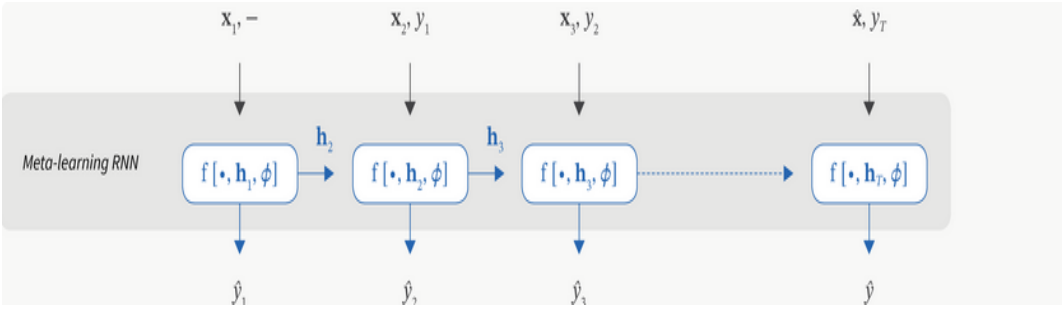


FIGURE 2.20 – Illustration de l'approche des réseaux de neurones à mémoire augmentée[268]

Neural Turning Machine (NTM) : Un réseau neuronal à mémoire augmentée inspiré du réseau neuronal LSTM utilisant un mécanisme d'adressage de la mémoire à deux niveaux : un premier basé sur le contenu et un second basé sur l'emplacement, contrairement au MANN

qui intègre uniquement un mécanisme d'adressage de la mémoire basé sur le contenu sans le mécanisme d'adressage de la mémoire basé sur l'emplacement.

Dans la même optique, d'autres méthodes exploitent l'idée d'étendre les réseaux neuronaux avec une mémoire externe [173][172][79]. La principale limite de ces algorithmes est toujours la capacité du calcul et du stockage.

2.4.3.2 Model Agnostic Meta learning (MAML)

Dans cette sous-section, nous optons à détailler davantage la technique puisqu'elle sera à la base de notre approche proposée dans le chapitre 4. En 2017, Finn et al. [61] ont écrit un papier de référence sur cette technique de meta-apprentissage. Ils ont défini l'idée d'un algorithme "agnostique" fonctionnant sur n'importe quel réseau entraîné avec la descente de gradient et sur de multiples tâches d'apprentissage. Les paramètres du modèle sont explicitement construits de telle sorte qu'un petit nombre d'étapes de gradient avec un petit ensemble de données d'apprentissage d'une nouvelle tâche produira une bonne performance sur cette tâche [7]. MAML se base sur le principe de généralisation et entraîne le modèle pour qu'il soit facile à ajuster. L'idée principale est de former les paramètres initiaux du modèle de manière à maximiser la performance du modèle sur une nouvelle tâche après juste quelques itérations avec un petit jeu de données. MAML relâche l'hypothèse d'une architecture précise du réseau de neurones (RNN [216], réseau siamois [125]). Il fonctionne sur différentes architectures avec différentes fonctions d'erreur. Une autre contrainte résolue en déployant le MAML est la structure de l'ensemble de données, notamment en paires requises dans différents autres travaux [133], [248], [219].

L'algorithme de MAML, illustré dans la figure 2.22, se résume comme suit :

1. Échantillonner un support set et un query set de la base de données de meta-entraînement.
2. Initialiser le base-apprenant f_θ par les poids du meta-apprenant M_θ (voir section 2.3).
3. Ajuster le base-apprenant f_θ par quelques itérations sur le support set.
4. Évaluer f_θ sur le query set et calculer la fonction d'erreur.
5. Utiliser la fonction d'erreur calculée dans l'étape (4) tout le long de l'épisode pour mettre à jour les poids du meta-apprenant M_θ.

Algorithm 1 Model-Agnostic Meta-Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters
1: randomly initialize θ
2: **while** not done **do**
3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
4: **for all** $\mathcal{T}_i$ **do**
5: Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ with respect to K examples
6: Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
7: **end for**
8: Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$
9: **end while**

Algorithm 2 MAML for Few-Shot Supervised Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters
1: randomly initialize θ
2: **while** not done **do**
3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
4: **for all** $\mathcal{T}_i$ **do**
5: Sample K datapoints $\mathcal{D} = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$
6: Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ using $\mathcal{D}$ and $\mathcal{L}_{\mathcal{T}_i}$ in Equation (2) or (3)
7: Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
8: Sample datapoints $\mathcal{D}_i' = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$ for the meta-update
9: **end for**
10: Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$ using each $\mathcal{D}_i'$ and $\mathcal{L}_{\mathcal{T}_i}$ in Equation 2 or 3
11: **end while**

FIGURE 2.21 – MAML pipeline

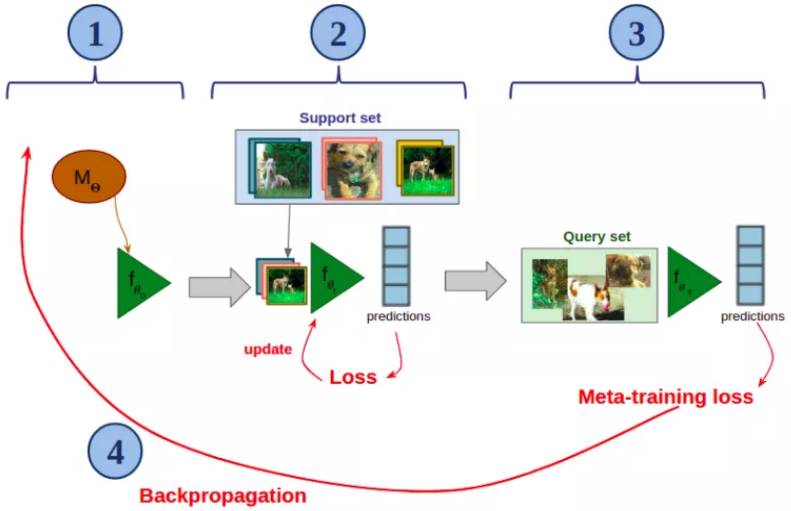


FIGURE 2.22 – Une étape de meta -entraînement d'un processus MAML [17]

Au cours du meta-entraînement, le MAML apprend des paramètres d'initialisation qui permettent au modèle de s'adapter rapidement et efficacement à une nouvelle tâche. Un simple ajustement dans la phase d'apprentissage de la nouvelle tâche avec quelques itérations, voire une seule, suffit pour que le modèle soit performant. Techniquement, MAML maximise la sensibilité des fonctions d'erreur des nouvelles tâches par rapport aux paramètres.

Soit f_θ notre meta-modèle f paramétré par θ initialisé aléatoirement, $p(t)$ la distribution des tâches et J_θ la fonction d'erreur. Soit T un lot de N tâches tel que $T_i \sim p(T)$. Pour chaque tâche T_i, nous entraînons le modèle en utilisant K entrés et calculons la fonction d'erreur.

Pour un apprentissage classique, la phase d'apprentissage est définie comme suit :

$$J_\theta = \sum_{D_{entranement}} c(\hat{y}_i \mid y_i, s_i, \theta) \tag{2.6}$$

$$\theta' \leftarrow \theta - \alpha \Delta_\theta J_\theta \tag{2.7}$$

Pour le meta-apprentissage avec MAML, la phase d'apprentissage est définie comme suit :

$$J'_\theta = \sum_{\mathcal{D}_{metaentranement}} J_\theta \tag{2.8}$$

$$\theta \leftarrow \theta - \beta \Delta_\theta J'_\theta \tag{2.9}$$

Plusieurs améliorations de MAML sont proposées, notamment Reptile, une adaptation de premier ordre de l'algorithme du MAML [176][175] ou le Meta-SGD développé par Li et al.(2017) [145] pour apprendre non seulement l'initialisation des paramètres, mais aussi le taux d'apprentissage convenable et la direction de la mise à jour. Il n'y a pas de gagnant absolu de toutes les techniques, l'approche à utiliser dépend des données, de l'architecture et des paramètres [63].

2.4.3.3 SNAIL

Mishra et al. [165] ont introduit une architecture d'un meta-apprenant attentif (SNAIL). Le réseau prédit à partir d'un ensemble de tuples (données, étiquette) l'étiquette manquante pour le dernier exemple. Leur système n'est pas récurrent, et prend toute la séquence de données de support set en une seule fois. L'architecture est basée sur l'alternance de convolutions et de couches d'attention. Plusieurs autres travaux se basent sur un mécanisme d'attention [36]. La prochaine section est dédiée pour mieux expliquer le principe d'attention et les principaux travaux connexes.

2.5 Mécanisme d'attention

Le défi principal de l'intelligence artificielle est toujours la recherche des meilleures singularités qui caractérisent les données d'entrée. Ces caractéristiques sont importantes pour établir la tache de classification avec moins de données. Le meta-apprentissage utilise les connaissances acquises d'une expérience à une autre pour optimiser l'apprentissage. En revanche, pour décider d'une façon efficace et rapide, l'être humain ne donne pas le même intérêt à toutes les parties d'une image. Effectivement, il fait attention à certaines régions plus qu'à d'autres. Récemment, inspiré toujours par l'apprentissage de l'humain, le mécanisme d'attention attire de plus en plus la communauté scientifique de recherche. Cette technique a pour objectif d'établir des relations entre les différentes caractéristiques pour mettre en avant les régions les plus importantes pour l'identification de chaque classe. Elle sert aussi à expliquer les décisions prises par le réseau de neurones dans la perspective de pouvoir comprendre et expliquer les mécanismes de l'IA (explainable AI) [84] [80].

Le mécanisme d'attention peut être utilisé d'autres types de données, notamment le texte, mais il est très répandu pour l'image [108].

2.5.1 Carte d'attention

L'un des principaux mots clés en lien avec le mécanisme d'attention est la carte d'attention (attention map or saliency map) explorée par Simonyan et al. [226] pour visualiser le modèle de classification des images (fig. 2.24).

2.5.1.1 Objectif

Simonyan et al. [226] ont pour objectif de visualiser le processus de classification d'images par un réseau profond et à mettre en avant les caractéristiques influençant le plus la décision du CNN.

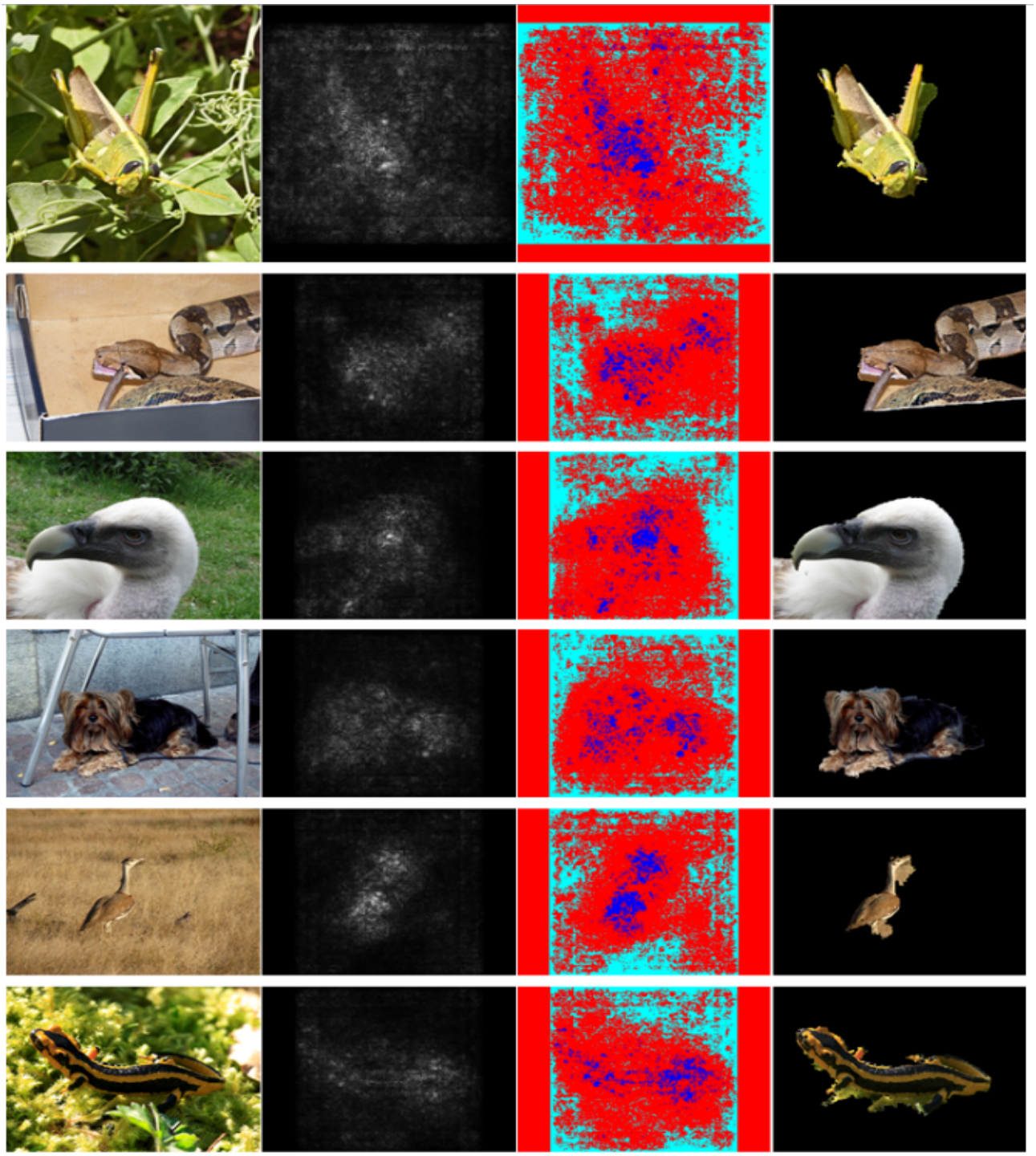


FIGURE 2.23 – Gauche : images de ILSVRC-2013. Milieu-gauche : Carte d'attention prédites. Milieu-droite : cartes d'attention seuillées avec le bleu qui montre les régions d'intérêt les plus importantes. Droite : segmentation résultante en se basant sur les masques d'attention produits [226]

2.5.1.2 Contribution

Ils ont utilisé deux techniques de visualisation : La visualisation d'une représentation par classe (Class Model Visualisation) et la visualisation de la région d'intérêt spécifique à l'image

(image-Specific Class Saliency Visualisation). La première vise à créer une image artificielle représentant la classe d'intérêt. La deuxième est basée sur la génération d'une carte d'attention par classe.

2.5.1.3 Approche et résultat

Étant donné une image I, une classe c et une classification CNN, la fonction de score d'une classe est définie comme suit :

$$S_c(I) = w_c^T + b_c \tag{2.10}$$

avec

$$\begin{cases} I & \text{Représentation vectorielle de l'image} \\ w_c & \text{Vecteur de poids} \\ b_c & \text{Le biais} \end{cases}$$

Dans le cadre d'une représentation linéaire, nous pouvons détecter facilement le pixel le plus influençant dans la prise de décision. Cependant, la fonction de score de classe n'est certainement pas linéaire dans un réseau de neurones profond. En effet, étant donné I_O l'image échantillon pour laquelle nous visons à noter les pixels respectivement à leur influence sur la classe c. Nous pouvons penser à approximer la fonction d'évaluation de la classe par une fonction linéaire dans le voisinage de I_O grâce à une expansion de Taylor du premier ordre [226] :

$$S_c(I) \approx w^T + b \tag{2.11}$$

avec w le dérivé de S_c respectivement à l'image I en I_O comme suit :

$$w = \frac{\partial S_c}{\partial I} \mid_{I_O} \tag{2.12}$$

Cette approche a été utilisée précédemment dans le cadre de la classification Bayésienne [10].

Génération de the class saliency maps : Soit I_O (n lignes et m colonnes) l'image échantillon sur laquelle on espère évaluer les pixels en fonction de leur influence sur la classe c. La carte d'attention résultante $M \in \mathbf{R}^{m \times n}$ est produite en calculant la dérivée (eq. 2.12) avec une seule rétro-propagation comme suit :

— Si l'image est en échelle de gris : $M_{ij} =\mid w_h(i,j) \mid$ où $h(i,j)$ est l'indice de l'élément de w respectivement au pixel de I dans la i-ème ligne et la j-ème colonne.

— Si l'image est multicanale (RVB) : $M_{ij} = max_z \mid w_h(i,j,z) \mid$ où $h(i,j,z)$ est l'indice de l'élément de w respectivement au pixel de I sur la i-ème ligne, la j-ème colonne et la z-ème couleur du canal.

Les cartes d'attention établies mettent en évidence les régions d'intérêt de l'image en lien avec la classe d'intérêt et leur calcul n'est pas couteux puisqu'il ne nécessite qu'une seule rétro-propagation.

2.5.2 Réseau d'attention spatiale pour la classification en few-shot learning

2.5.2.1 Objectif

On espère à travers cette approche améliorer les algorithmes de meta-apprentissage et utiliser davantage les connaissances acquises des caractéristiques extraites d'une expérience à une autre.

2.5.2.2 Contribution

Cette approche utilise l'attention canal en parallèle avec le module d'attention spatiale (C-SAM) afin d'extraire les informations les plus pertinentes en utilisant des échantillons de plusieurs classes à partir de différentes tâches de classification avec peu de données

labellisées [291]. Zhang et al. révèlent que contrairement aux travaux précédents [196] [141] où le meta-apprentissage consiste à apprendre de plusieurs tâches et utiliser les connaissances préalables pour une nouvelle tâche, cette approche vise à apprendre une représentation des classes.

2.5.2.3 Approche et résultat

Dans cette approche, le meta-apprentissage est déployé en tant que réseau de relation pour calculer la similarité entre des données libellées et non libellées en utilisant le C-SAM. Deux techniques du mécanisme d'attention sont utilisées : l'attention en canal liée aux caractéristiques globales [277] et l'attention spatiale liée aux caractéristiques locales [186]. À partir d'une carte de caractéristiques (feature map), le mécanisme d'attention en canal extrait un tenseur 1D activé par la fonction Sigmoïde et le mécanisme d'attention spatiale produit un masque de caractéristiques de la même taille que la carte de caractéristiques. Les deux techniques d'attention ont pour objectif d'obtenir des valeurs d'activation élevées sur les cartes de caractéristiques pertinentes et faibles sur les cartes de caractéristiques non pertinentes redondantes.

Zhang et al. [291] conjuguent les deux techniques d'attention citées ci-dessus pour aboutir aux meilleurs résultats et extraire les caractéristiques les plus pertinentes relativement aux tâches. Ils ont inclus un module de métrique de relation déployant un paramètre d'apprentissage indiquant la pertinence des caractéristiques au lieu de la métrique de distance [242] [92]. Les deux techniques sont complémentaires. L'attention de canal perd de l'information en multipliant la carte de caractéristiques par les valeurs d'activation inférieures. Placée dans différentes couches de convolution, lorsque l'attention du canal affaiblit l'information dans certaines cartes de caractéristiques, l'attention spatiale, en utilisant le masque d'attention, met en évidence de nombreuses régions pertinentes de chaque carte de caractéristiques. Les cartes de caractéristiques de sortie résultantes des deux techniques d'attention sont additionnées, permettant la mise en relief de la région d'intérêt la plus discriminante. Ce travail inclut également des différentes fonctions de perte personnalisées évaluées sur différentes parties du réseau de neurones afin d'améliorer ses performances.

2.5.3 Apprentissage d'une représentation discriminante profonde basée sur la carte d'attention pour la classification de scènes

2.5.3.1 Objectif

Dans le cadre de la classification de scènes d'images de télédétection, l'apprentissage profond est largement utilisé. Cependant, plusieurs défis sont présents, notamment les "grandes variations intra-classe" ((intra-class inconsistency), la "petite dissimilarité interclasse" (inter-class indistinction), la taille réduite des objets et leur dispersion dans les images de scènes par rapport à l'image naturelle. En revanche, les modèles CNN utilisent uniquement l'entrée RVB originale sans faire attention à la région la plus pertinente de ces images RVB. Par conséquent, le cout de calcul est élevé et les parties les plus précieuses des données sont parfois négligées. Afin d'améliorer la cohérence intra-classe et la discrimination interclasse, un mécanisme d'attention est incorporé.

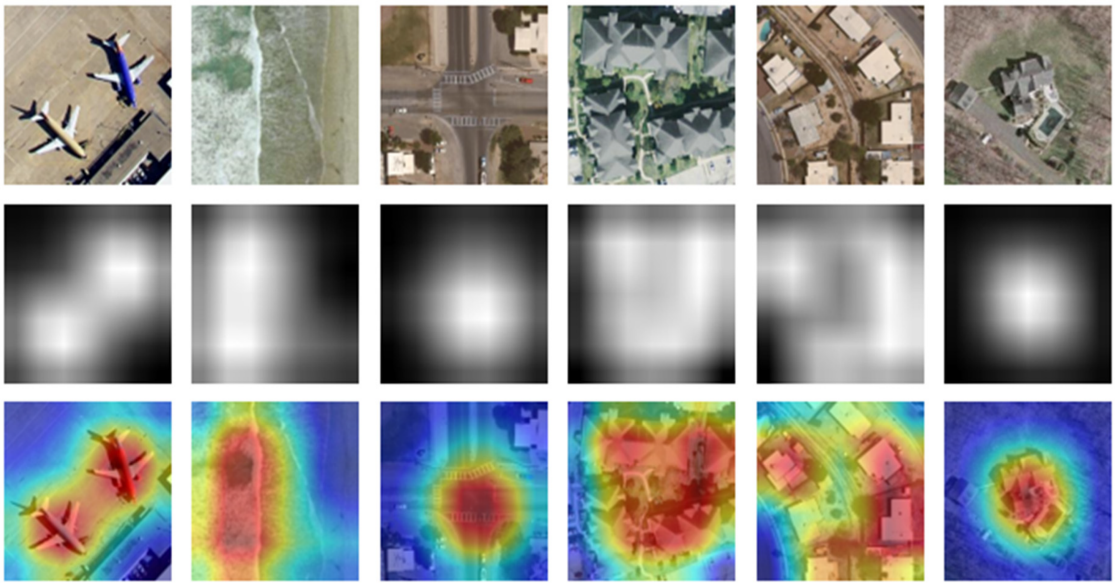


FIGURE 2.24 – Images d'origine, cartes d'attention et masques d'attention résultants [143]

2.5.3.2 Contribution

Cet article s'intéressé la tâche de classification de scènes d'images de télédétection. Afin d'apprendre une représentation de classe robuste et discriminative à partir de ce type d'images, Li et al. [143] font intervenir le mécanisme d'attention dans leurs CNN. Leur approche inclut une carte d'activation de classe (CAM) fournissant une information supplémentaire sur les régions pertinentes par rapport à l'architecture des modèles profonds RVB classiques. La partie importante de ce travail est la fusion des caractéristiques des cartes d'attention et des caractéristiques de l'image originale. Ce travail utilise une fonction d'erreur reconnue sous le nom fonction d'erreur centrale récemment développée qui a bien fonctionné dans la tâche de reconnaissance des visages par apprentissage profond afin d'améliorer la performance et la discrimination [274].

2.5.3.3 Approche et résultat

De nombreuses techniques sont proposées avec différentes métriques de distance pour créer l'espace discriminant pour la tâche de classification de scènes d'images de télédétection [272] [38]. Cependant, les architectures classiques ne font pas attention à la région discriminante la plus pertinente de l'image d'entrée représentant les caractéristiques spécifiques d'une classe particulière. Dans ce sens, Li et al proposent une approche d'apprentissage de représentation discriminante avec la carte d'attention (DDRL-AM). L'approche commence par la génération des cartes d'attention (AM) pour toutes les images. Chaque pixel est noté selon sa pertinence dans l'image originale. Ainsi, dans un second temps, ces cartes d'attention et les images originales sont données en entrée au réseau de neurones. Ce réseau inclut une opération de fusion assurant une meilleure utilisation de cette connaissance d'attention acquises dans les cartes d'attention avec l'image originale. Cette fusion est inspirée par une architecture à double-flux pour la fusion de caractéristiques proposée par Simonyan et al. [227] comprenant une couche de sortie de classification Softmax finale fusionnant les deux réseaux. Différents autres travaux sont faits dans cette direction [58] [30]. Li et al. abordent une architecture

comprenant un réseau attentionnel-génératif pour générer les cartes d'attention et un modèle de fusion de caractéristiques profondes (fig.2.25).

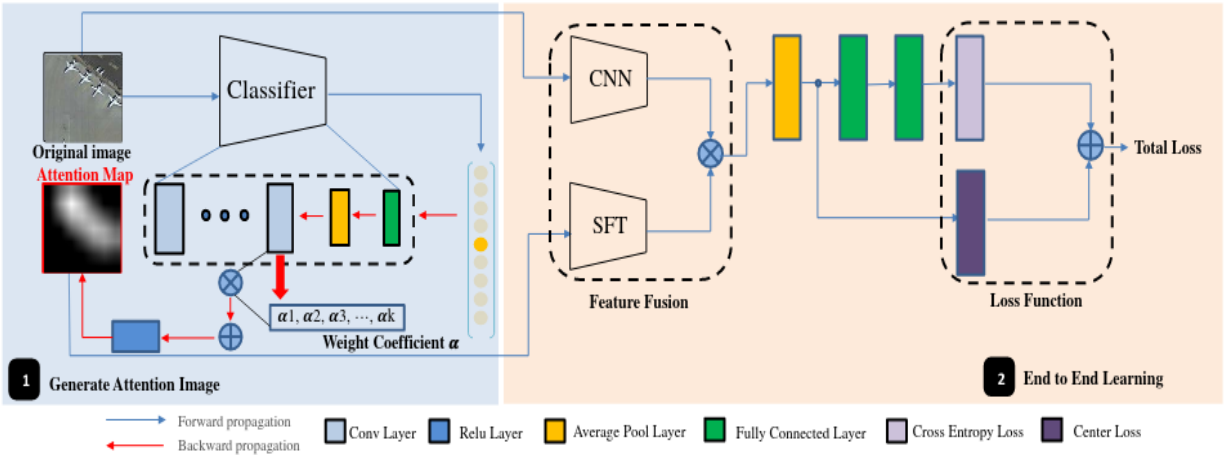


FIGURE 2.25 – Aperçu de l'architecture finale : 1) Un réseau attentionnel-génératif : Réseau CNN ResNet-18 pré-entrainé sur la base de données Imagenet et ajusté sur la nouvelle base et l'architecture Grad-CAM (Gradient-poids et algorithme basé sur les cartes d'activation de classe) pour générer les carte d'attention. 2) Architecture CNN à deux flux pour la fusion des images originales avec les cartes d'attention en combinant les deux fonctions d'erreur [143].

Pour améliorer la capacité de discrimination du modèle, outre l'optimisation de la fonction de perte d'entropie croisée classique, Li et al. intègrent la fonction de perte de centre décidant du meilleur centre pour les caractéristiques de la classe et influençant la décision par rapport aux différentes classes [274].

2.5.4 Apprentissage visuel dynamique en quelques coups sans oublier

2.5.4.1 Objectif

Gidaris et al. [72] estiment que l'utilisation du meta-apprentissage pour few-shot learning doit être plus efficace et plus rapide sans affaiblir et sacrifier la précision sur les classes initiales sur lesquelles le réseau neuronal s'est pré-entrainé.

2.5.4.2 Contribution

L'approche intègre deux techniques permettant d'apprendre de nouvelles classes à partir de quelques échantillons, sans oublier les classes initiales sur lesquelles le modèle a été pré-entraîné. En premier lieu, le réseau de neurones inclut une pondération basée sur un mécanisme d'attention. Dans ce but, des vecteurs de poids de classification des classes initiales sont générés. Ensuite, le classificateur de CNN est implémenté en tant qu'une fonction de similarité cosinusoïdale entre les représentations des caractéristiques et les vecteurs de classification. Ainsi, le réseau neuronal inclut à la fois les classes initiales et les nouvelles classes. Cette fusion permet une meilleure généralisation des représentations de caractéristiques sur les nouvelles classes.

2.5.4.3 Approche et résultat

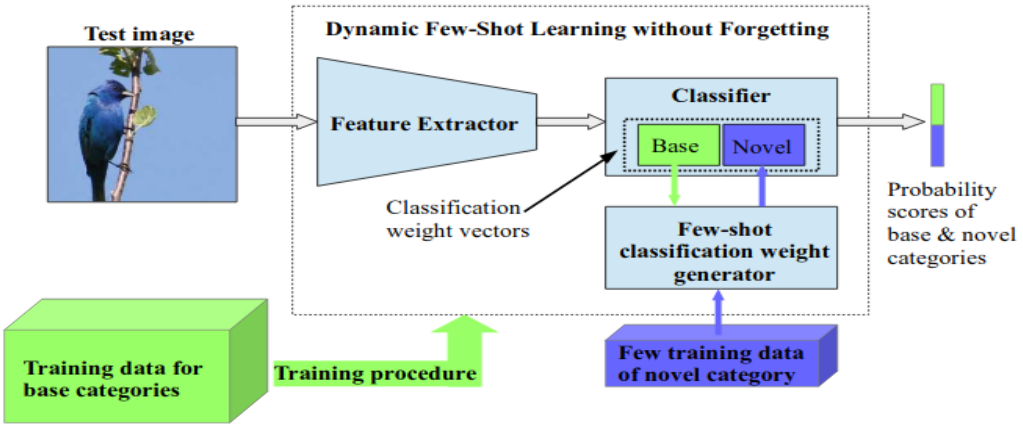


FIGURE 2.26 – Aperçu de l'architecture : 1) Une première partie d'extractions de caractéristiques de classification basée sur ConvNet 2) Un classificateur à quelques coups comme générateur de poids. [72]

Les deux parties de la fig 2.26 sont entrainées sur les classes initiales avec une large base de données. Pendant l'étape de test (meta-test), le générateur de poids intègre le mécanisme d'attention qui se base sur les poids pour décider. En effet, ce mécanisme prend en entrée les quelques échantillons de la nouvelle classe et le vecteur de poids de classification des classes initiales préalablement créé (rectangle vert dans la boîte du classificateur) et génère le vecteur

de poids de la nouvelle classe (rectangle bleu dans la boite de classificateur) permettant au CNN d'être performant sur les anciennes classes et les nouvelles classes aussi.

2.5.5 Branche d'attention : Apprentissage du mécanisme d'attention pour une interprétation visuelle

2.5.5.1 Objectif

Grace aux cartes d'attention, nous espérons expliquer visuellement les décisions prises par le réseau de neurones. Les cartes d'attention mettent en évidence les régions pertinentes qui influencent le plus la décision. Introduire une branche d'attention permet de payer plus d'attention aux caractéristiques qui permettent de distinguer les classes et de produire deux fonctions d'erreur séparées.

2.5.5.2 Contribution

Fukui et al. divisent l'explication visuelle en deux catégories. La première est basée sur la perturbation de gradient en introduisant un bruit [230] ou une autre perturbation sur les données d'entrée (données auxiliaires) [221] [34]. Suite à ces perturbations, les cartes d'attention sont générées. La deuxième catégorie génère ces derniers en se basant sur la décision de classification. La technique classique reconnue sous le nom de carte d'activation de classe (CAM) nécessite le remplacement de la couche entièrement connectée par une couche de convolution et une couche de global average pooling GAP qui dégradent les performances du CNN [168] [294] (fig.2.27).

Afin d'éviter cette dégradation, Fukui et al. [67] proposent l'incorporation de la branche d'attention (ABN). Cette technique d'attention permet de générer les cartes d'attention en une propagation vers l'avant (forward propagation). Cette structure permet en même temps d'expliciter la décision du réseau de neurones et d'améliorer sa performance. Cette structure de

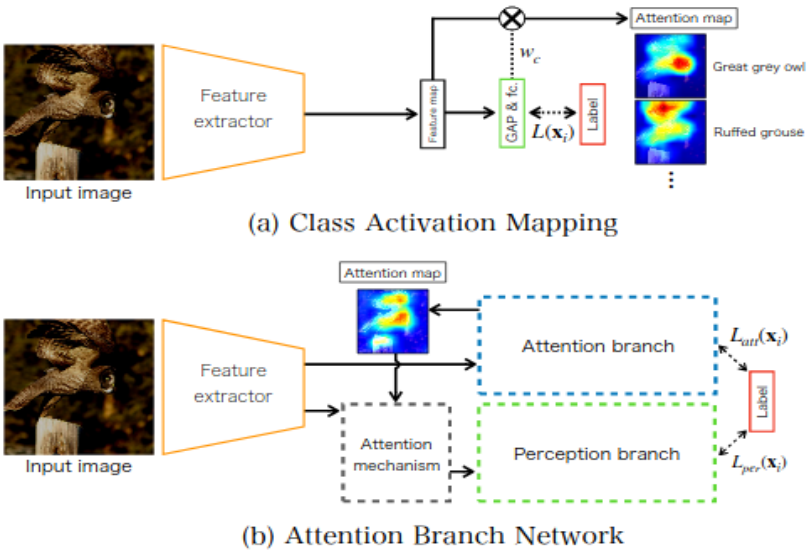


FIGURE 2.27 – Différene entre la structure classique de carte d'activation de classe (CAM) et la structure basée sur la branche d'attention (ABN)

mécanisme d'attention permet de mieux comprendre la prise de décision du CNN et d'améliorer ses performances par la même occasion.

2.5.5.3 Approche et résultats

ABN comprend trois parties :

— Extraction de caractéristique : plusieurs couches de convolution pour extraire les caractéristiques de données d'entrée.

— Branche d'attention : un mécanisme d'attention basé sur la réponse de la décision de classification pour produire les cartes d'attention.

— Branche de perception : plusieurs couches de convolution ayant comme entrée les caractéristiques et les cartes d'attention et donnant en sortie les probabilités de classes.

L'ABN est entraîné en calculant les fonctions de perte dans les branches d'attention et de perception. La branche d'extraction de caractéristiques et la branche perception sont conçues en ajustant des modèles VGGNet et ResNet. La branche attention est introduite après la branche d'extraction de caractéristiques. Cependant, contrairement au CAM classique où la sortie est les cartes d'attention, les cartes d'attention sont ici une sortie intermédiaire et sont ensuite l'entrée de la branche perception.

2.6 Conclusion

Dans ce chapitre, nous avons défini la problématique de manque de donnée connue sous le nom de Few-shot learning et l'approche tendance de meta-apprentissage. Nous avons présenté un état d'art intensif divisé en trois catégories en se basant sur la stratégie d'apprentissage et e, détaillant l'évolution des solutions proposées. Ensuite, toujours dans l'optique de l'amélioration de l'efficacité de l'apprentissage, nous nous sommes intéressés au mécanisme d'attention. Nous avons présenté une revue de littérature des différents travaux en examinant et en discutant leurs architectures et leurs résultats. Cette étude d'état de l'art est très enrichissante pour pouvoir améliorer tous les algorithmes d'apprentissage. Dans le cadre de notre thèse, ces notions et ces approches nous seront utile dans le chapitre 3 et 4 pour expliquer nos approches et motiver partiellement les choix qu'on va effectuer.

Chapitre

3 Anonymisation profonde

Sommaire

3.1 Motivation 90

3.2 L'apprentissage profond et la confidentialité 94

3.2.1 Les données : le pétrole d'aujourd'hui 95

3.2.2 La confidentialité en jeu 99

3.2.3 Empreintes digitales 99

3.3 Base de données 100

3.4 L'anonymisation liée aux équipements d'acquisition de l'IRM 102

3.4.1 Approches et architectures 103

3.4.2 Reformulation mathématique 106

3.4.3 Expérimentations et résultats 107

3.5 L'anonymisation liée à l'identité du patient 111

3.5.1 Approches et architectures 113

3.5.2 Expérimentations et résultats 114

3.6 Conclusion 117

3.1 Motivation

La confidentialité ! Un mot qui nous tient au cœur en tant qu'individu, un mot qui fait référence à de nombreux sujets et domaines dans notre quotidien [218]. La confidentialité est un pacte de confiance, de discrétion, de réassurance et de sécurité établi avec une seconde partie. La notion de confidentialité dans le domaine de la sécurité de l'information numérique a été définie par l'organisation internationale de normalisation (ISO) comme « le fait de s'assurer que l'information n'est accessible qu'à ceux dont l'accès est autorisé ».

a) Le secret professionnel

C'est un terme lié étroitement à la confidentialité. Le professionnel a le devoir de préserver l'intimité et la vie privée de son client. Cette exigence consiste principalement à protéger les informations personnelles des individus, mais également de leur permettre de partager librement les données nécessaires pour que le professionnel agisse d'une façon efficace et optimale.

Un professionnel qui ne respecte pas la confidentialité de ses clients risque d'être sanctionné. Les sanctions varient de l'avertissement à l'interdiction de la poursuite de l'activité professionnelle ou même l'emprisonnement. L'individu retient le droit de la poursuite juridique et de la demande des dédommagements.

b) Confidentialité et anonymisation

Les informations confidentielles peuvent prendre plusieurs formes et englobent plusieurs domaines : social, politique, juridique, médical et économique. Elles peuvent être des données partagées volontairement ou involontairement sur le cloud [156][24], des paroles confiées à un professionnel ou un dossier émis à une institution, etc.

L'anonymisation implique un processus supplémentaire pour protéger l'accès à l'identité. L'anonymat est un statut qui rend la mission d'identification de l'individu absolument

impossible. Ce processus inclut l'élimination de tous les rapports et les repères permettant de tracer l'identité de la personne.

c) Pseudo-anonymisation

Par ignorance ou par faute d'inattention, on a l'impression parfois d'être anonyme alors qu'on ne l'est pas. Notamment, sur les réseaux sociaux, on pense souvent qu'on est anonyme en utilisant un faux profil ou un pseudo. Cependant, c'est loin d'être le cas. En effet, on est bien identifié sur internet à cause de plusieurs techniques avancées. Ainsi, de nos jours, on doit être bien vigilant pour s'assurer de la protection de la confidentialité. Ces avancées technologiques provoquant une pseudo-anonymisation font peur aux individus aussi bien qu'aux institutions. Un autre domaine d'application auquel nous allons nous intéresser le plus dans notre travail est le domaine médical. On a tendance à croire que dès que les coordonnées personnelles (nom, prénom, âge...) sont éliminées, l'identité du patient est bien préservée. En réalité, des approches avancées proposées par exemple par l'apprentissage profond permettent de dévoiler l'identité du patient sans avoir besoin de ces méta-données. Dans le cadre de ce travail, nous abordons la problématique de l'anonymisation en imagerie médicale. Effectivement, des traces invisibles à l'œil nu dans l'imagerie médicale permettent à nos réseaux de neurones de divulguer l'identité du patient.

L'apprentissage profond est un outil puissant qui permet d'effectuer de multiples tâches complexes, parfois inaccessibles d'une façon classique à l'être humain (voir chapitre 2). Cependant, il s'agit d'une arme à double tranchant. Son pouvoir, notamment dans le diagnostic médical, augmente simultanément avec sa capacité à violer facilement la vie privée des personnes. Par conséquent, il y a toujours des questions sur la confidentialité des échanges effectuées et les institutions se méfient souvent du partage des données. Ainsi, la problématique de la pseudo-anonymisation est de plus en plus urgente.

Les données sont importantes pour la communauté scientifique, car elles sont la source même des connaissances d'un domaine. Elles apportent des éléments concrets et servent à faire avancer la science et la technologie. La science de données permet d'exploiter ces données

brutes pour étudier et analyser un système complexe comme une entreprise ou encore un organe ! La fiabilité des données scientifiques et techniques, la confiance de leur exploitation, représentent donc des enjeux socio-économiques primordiaux. La masse de données à gérer est également un enjeu important, résumé dans le contexte de Big data, ou données massives, que des outils d'IA permettent de traiter.

Les données massives ne signifient pas simplement un changement dans l'échelle de la preuve apportée à une hypothèse scientifique. Le changement dans la volumétrie modifie aussi les formes de gestion de cette information, avec notamment l'introduction des algorithmes et de l'intelligence artificielle dans le raisonnement avec des mises en corrélation des données. La prédiction par IA à partir des données n'est donc plus une simple méthode basée sur la causalité (une prédiction basée sur une hypothèse), mais un "raisonnement" s'appuyant sur des probabilités et des croisements de données hétérogènes, en grand nombre, et dont il peut être complexe d'analyser les causes et les fondements.

L'apprentissage profond, en particulier, est une technique gourmande en termes de données. Donc, il est important de pouvoir instaurer un climat de confiance mutuelle avec les individus et les institutions afin de les rassurer par rapport à la véritable anonymisation des données. Ainsi, d'une part, les institutions peuvent s'engager auprès de leurs clients pour respecter leur zone d'intimité. D'autre part, elles peuvent se permettre de fournir avec sérénité et en toute sécurité les données avec les chercheurs sans risquer leur obligation envers leurs clients et en étant certaines que leurs identités resteront confidentielles.

Dans ce chapitre, nous abordons la problématique de pseudo-anonymisation en imagerie médicale, nous prouvons qu'il s'agit d'un véritable risque et nous proposons des solutions pour l'éviter. En effet, les réseaux de neurones ont montré des performances record dans de multiples tâches médicales. Cependant, la quantité et la qualité des données sont des exigences cruciales. Les données sont l'un des problèmes les plus difficiles à résoudre lors du déploiement de modèles d'apprentissage profond. L'un des principaux défis réside dans les protocoles de confidentialité des institutions, en particulier dans le domaine médical. Bien que les métadonnées soient généralement déjà exclues de la base de données fournie, des

nombreuses caractéristiques invisibles dans les images peuvent aider à tracer les données anonymes.

Nous espérons combattre le feu par le feu. Dans un premier lieu, nous utilisons l'apprentissage profond pour trouver ces traces invisibles. Ensuite, nous proposons d'utiliser également l'apprentissage profond pour les exclure. Pour ce travail, nous nous concentrons sur l'imagerie par résonance magnétique (IRM). Cependant, les solutions proposées peuvent être généralisées pour des multiples types d'imagerie dans le domaine médical, mais aussi dans d'autres domaines. Dans cette perspective d'anonymisation, ce chapitre inclut deux contributions majeures. Pour la première contribution, nous nous intéressons à l'une des caractéristiques les plus importantes de l'IRM : l'équipement utilisé pour l'acquisition. En effet, spécialement pour les maladies rares, connaitre un détail lié à la machine ou l'établissement d'acquisition peut permettre d'identifier directement le patient. Tout d'abord, nous cherchons à produire un algorithme capable de bien distinguer plusieurs équipements IRM de différents constructeurs. À cette fin, nous utilisons une architecture de réseau de neurones convolutif pour travailler sur cette tâche de classification d'images médicales. Par la suite, nous allons reconstruire l'IRM d'entrée à l'aide d'un auto-encodeur. La dernière étape consiste à utiliser l'auto-encodeur afin d'induire en erreur le classifieur qui classifie l'équipement IRM. Les données résultantes représentent des IRM toujours valides pour le diagnostic médical, mais sans information sur l'équipement d'acquisition. La deuxième contribution de chapitre consiste à préserver directement l'identité du patient. Nous prouvons tout d'abord la possibilité de distinguer les IRM des différents patients, ce qui est problématique, surtout dans le cadre des maladies rares. Ensuite, le même processus est adapté pour aboutir à des IRM anonymisées utilisables pour le diagnostic avec une protection réelle de l'identité du patient.

Ce travail est une initiation d'un processus rassurant les institutions médicales quant à la confidentialité de leurs patients et permettant aux chercheurs d'obtenir des données d'une haute qualité visuelle nécessaires pour les modèles d'aide au diagnostic médical. Dans ce but, nous proposons la première étape vers une véritable anonymisation des données d'imagerie médicale basée sur les caractéristiques de classification extraites à l'aide de réseaux de neurones profonds. À notre connaissance, ce travail est le premier à utiliser l'auto-encodeur en conjonction avec

l'approche d'identification dans le domaine médical à des fins d'anonymisation. Le reste du chapitre est organisé comme suit : la section 2 est réservée à l'introduction du contexte, la section 3 présente la base de données utilisée, la section 4 et la section 5 détaillent respectivement la première et la deuxième contribution : l'approche proposée, les expériences, les résultats et la discussion. Enfin, la section 6 conclut le travail et ses perspectives.

3.2 L'apprentissage profond et la confidentialité

La collecte des bases de données de haute qualité est souvent une tâche difficile. De nombreux défis sont à relever dans cette perspective, notamment l'accès aux données pour des considérations juridiques et éthiques [19]. Ces bases de données sont importantes pour l'avancement de la recherche. Cependant, de nombreuses entreprises, qui collectent massivement des données, ont réussi à pénétrer la vie privée des individus. Ces entreprises étaient les premières bénéficiaires des avancées de l'intelligence artificielle et en particulier de l'apprentissage profond, ce qui a soulevé des véritables problèmes de confidentialité.

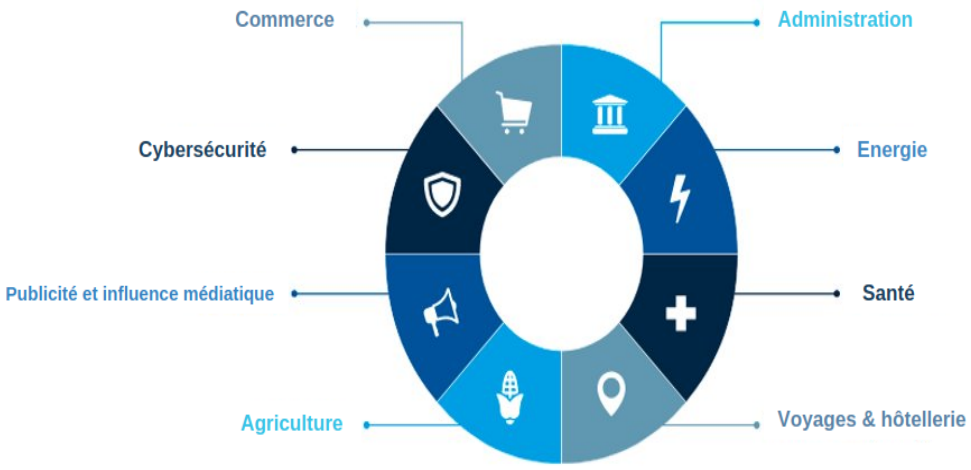


FIGURE 3.1 – Les domaines d'application de l'exploitation de données [244]

Au début, on ne se rendait pas compte du danger que représentent ces avancements. La technologie moderne offre un style de vie et un bien-être inestimables à l'humanité. Les nouvelles technologies ont envahi notre quotidien. Les objets intelligents occupent nos maisons, nos établissements et nos rues (voir fig. 3.1). L'intelligence artificielle a déjà réussi

à changer notre mode de vie. Mais avec le progrès des capacités de calcul et de stockage de données, l'apprentissage profond a son mot à dire. De nos jours, une grande quantité de données est produite et l'apprentissage profond est incorporé dans toutes nos machines. Cette combinaison est une arme à double tranchant. En effet, son bénéfice est énorme puisqu'avec une grande quantité de données, les réseaux de neurones omniprésents sont de plus en plus performants. Par conséquent, ils sont capables d'effectuer de multiples tâches avec une grande efficacité et des records de précision. Cependant, ce bénéfice s'est accompagné d'un risque de confidentialité. Effectivement, cette dominance sur les données et la progression de la puissance de l'apprentissage profond a permis à des entreprises une utilisation controversée à des fins de commercialisation, mais aussi de manipulation sociale, politique, éthique et économique (voir section 3.2.2). Ainsi, les individus aussi bien que les institutions ont commencé à s'interroger "enfin" par rapport à la confidentialité de leurs données [283][288]. Nous allons voir dans les sections suivantes que les résultats donnés par les réseaux de neurones sont souvent bénéfiques s'ils sont bien interprétés et utilisés dans un cadre surveillé. Cependant, ce n'est toujours pas le cas [208][122]. Les entreprises de nos jours sont en forte compétition de collection de données. Les données sont considérées comme le pétrole d'aujourd'hui. En effet, cette tendance récente a fait de ce sujet une préoccupation très pressante pour la plupart des institutions de différents domaines et des gouvernements [142][151]. Il s'agit d'un défi très intéressant : d'une part, on veut fournir des modèles performants avec des données accessibles, mais d'autre part, les données doivent être protégées contre les manipulations intentionnelles et accidentelles de la confidentialité.

3.2.1 Les données : le pétrole d'aujourd'hui

À travers les figures 3.2 et 3.3, nous observons quelques sources expliquant l'explosion de la quantité de données. Involontairement et parfois volontairement, à travers un partage, un j'aime, un clic, on est en train de créer une énorme base de données qui permet à des acteurs de décider par rapport à notre avenir.

La pandémie de 2020, par exemple, a affecté de nombreux secteurs qui ont tourné au ralenti, voire se sont totalement arrêtés, mais les données, elles, n'ont cessé d'être produites [27] ! En effet, la variété des sources d'entrées permettent d'accroître considérablement la collecte de données, tandis que les coûts d'infrastructure de la gestion et du stockage diminuent. La puissance ascendante des données donne naissance à des entreprises spécialistes dans l'exploitation des données. Ces entreprises se nourrissent des données et grâce à l'apprentissage par transfert (voir section 2.4.1), elles sont capables de performer d'une tâche à une autre et d'un secteur à un autre sans plus avoir besoin d'une grande quantité de données. Elles possèdent des modèles de plus en plus puissants et consistants qui peuvent à la fois faire avancer l'humanité et mettre en danger la sphère de confidentialité (section 3.2.2).

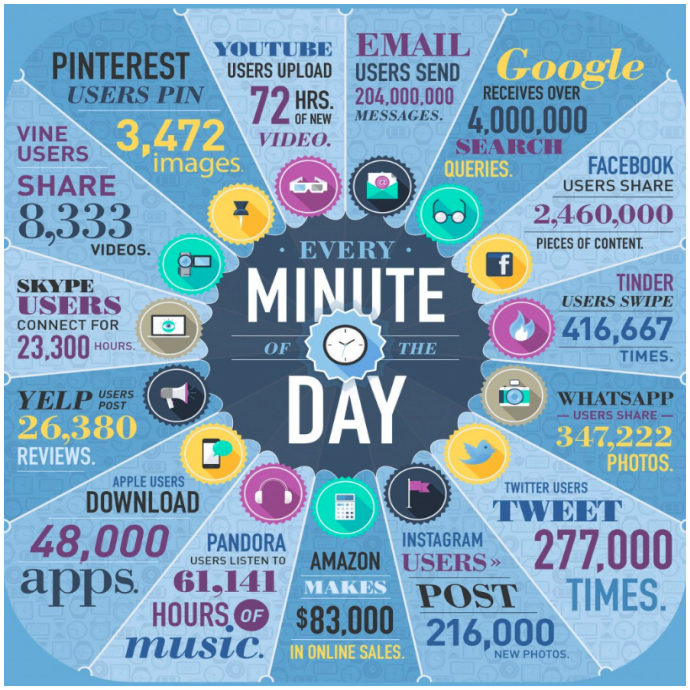


FIGURE 3.2 – Quantité de données produite par minute en 2014 [27]

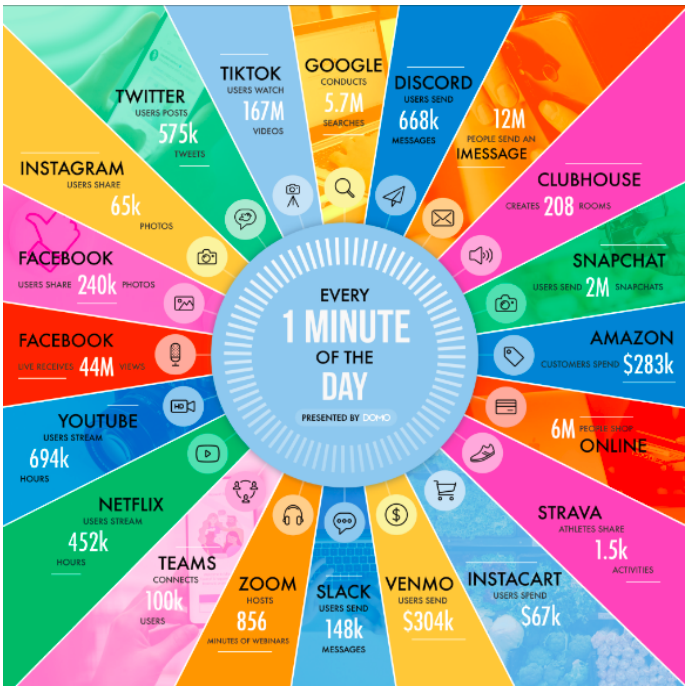


FIGURE 3.3 – Quantité de données produite par minute en 2021 [27]

La figure 3.4 présente l'exploitation de données comme un processus de recyclage. Dans le "big data 1.0", les sources de données et les outils de stockage et de calcul réduits, freinés par des coûts élevés, résultent des capacités analytiques limitées et restreintes à des acteurs spécifiques. L'omniprésence des objets connectés et de l'intelligence artificielle dans tous les secteurs, l'évolution des capacités du calcul et du stockage et l'avancement des recherches favorisant la puissance et la performance de l'apprentissage profond a donné naissance au "big data 2.0". Cette version se caractérise par une analyse de données avancée qui peut mettre en danger l'identité des individus et leur confidentialité. Dans "Big Data Next", il est toujours possible de concevoir des réseaux profonds produisant des modèles rentables et puissants dans des multiples domaines, mais aussi tout en respectant la confidentialité de tous les acteurs.

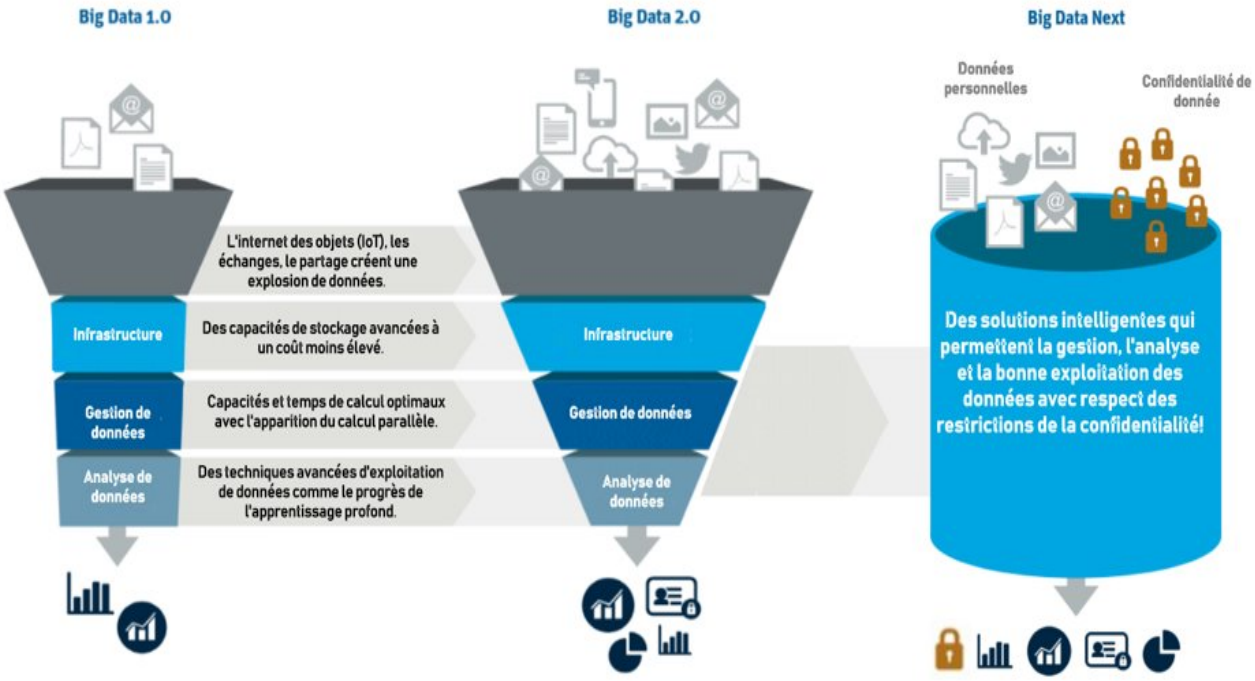


FIGURE 3.4 – Processus d'exploitation des données inspiré de [244]

Industrie	Niveau de surveillance réglementaire	Facilité de collecte de données	Omniprésence de la technologie	Index de maturité
Publicité et influence médiatique	3	3	3	3.0
Voyage et hôtellerie	3	2	3	2.7
Cybersécurité	2	2	3	2.3
Commerce	3	2	2	2.3
Energie	2	2	1	1.7
Santé	1	1	2	1.3
Administration	1	2	1	1.3
Agriculture	2	1	1	1.3

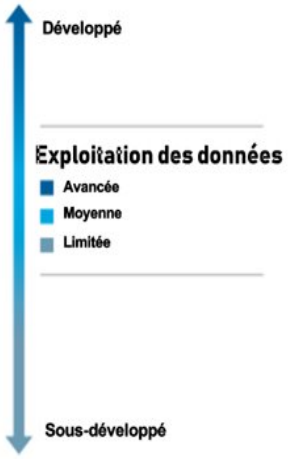


FIGURE 3.5 – Accessibilité et l'exploitation des données dans les différents secteurs [244]

SVB Analytics a établi un index de maturité (fig. 3.5) qui permet d'examiner le niveau d'incorporation de l'exploitation de données dans des différents secteurs en se basant sur trois caractéristiques primordiales : restrictions règlementaires, facilité de la collecte des données et le niveau d'intégration technologique. Plus le score total est élevé, plus l'exploitation est

développée. Par conséquent, nous pouvons voir que notamment dans le secteur de la santé, le score est faible et la marge de croissance est énorme. Cependant, nous pouvons observer aussi que les restrictions réglementaires et l'inaccessibilité aux données sont des défis importants dans ce domaine.

3.2.2 La confidentialité en jeu

Les innovations proposées par l'apprentissage profond ont permis de changer les règles du jeu de tous les secteurs. Parfois, l'accès à des données confidentielles et spécifiques à l'aide des partenaires ou par le biais des accords engagés peut être un besoin vital pour les acteurs du marché afin de maintenir un avantage concurrentiel ou pour manipuler la population dans une direction ou une autre. Par conséquent, et afin préserver la confidentialité, des nombreuses réglementations sont apparues. Cependant, plusieurs tâches importantes n'ont pas pu être traitées efficacement par l'intelligence artificielle en raison de ces restrictions sur les données, notamment dans le domaine médical où les protocoles de confidentialité sont stricts et ne permettent pas de faciliter l'accès aux données aux chercheurs. En effet, l'éventualité d'une ré-identification du patient est une préoccupation émergente [105][191][60][281][130].La manipulation des données des patients profiterait financièrement aux compagnies d'assurance entre autres. Par conséquent, les institutions médicales disposent de protocoles stricts et complexes pour ouvrir les données à la recherche publique, ce qui empêche l'accélération de la recherche pour les applications cliniques du monde réel [112][287][73]. Ainsi, en utilisant l'apprentissage automatique pour protéger la vie privée, nous pouvons faire une énorme différence dans différentes tâches du domaine médical telles que le diagnostic, pronostic et la guérison des maladies.

3.2.3 Empreintes digitales

Le domaine des empreintes digitales a également bénéficié de l'apprentissage profond. Plusieurs travaux de recherche ont montré des performances élevées pour l'identification de dispositifs dans différents domaines d'application, notamment pour les appareils médicaux

[118][278][201][161]. Howard et al. se sont intéressés à l'identification des dispositifs du rythme cardiaque à l'aide de réseaux neuronaux. À ce jour, aucune investigation n'a été menée en utilisant l'apprentissage profond pour l'identification des dispositifs d'IRM et surtout dans un but d'anonymisation. Néanmoins, dans le domaine médical, l'assistance médicale intelligente et l'aide au diagnostic sont des directions de recherche émergentes pour aider les médecins à exploiter les données médicales [37] et la confidentialité est l'un des problèmes les plus critiques rencontrés lors de la collecte de données. Par conséquent, l'utilisation de l'apprentissage profond pour préserver la vie privée pourrait changer la donne dans le domaine médical[127][103].

3.3 Base de données

Le big data est souvent associé à ce qu'on pourrait appeler la malédiction des "5V " : volume, vitesse, variété, variabilité et véracité. Cette notion est particulièrement vraie dans le domaine de la santé : Les données sur un patient sont très nombreuses, elles peuvent être compilées très rapidement et doivent être analysées vite, en temps réel, elles sont des sources très diverses, de nature hétérogène et n'ont pas toutes le même poids et on ne peut pas toujours leur apporter la même confiance. Mais cette malédiction potentielle des 5V, si elle est maîtrisée, peut permettre d'aller vers une médecine « 5P » : préventive, prédictive, participative, personnalisée, pertinente. L'imagerie médicale est un outil précieux qui contribue largement à cette médecine 5P mais aussi à la problématique des 5V. Ainsi, un scanner corps entier génère plusieurs milliers d'images comprenant chacune plusieurs millions de pixels. L'imagerie « multimodale » faisant appel à plusieurs modes d'acquisition en IRM amplifie encore la quantité d'informations.

De nos jours, l'IRM est largement utilisée, étant une imagerie non ionisante et non invasive offrant une grande résolution spatiale et temporelle et de riches informations anatomiques et physiologiques. Habituellement, différentes séquences d'IRM sont fournies, telles que Flair, T1, T2, T1 post-contraste, correspondant à différentes techniques d'excitation des spins magnétiques à l'intérieur du corps humain, ce qui donne lieu à différents contrastes d'échelle de gris (voir section 1.4.2). Nous limiterons notre expérience à des séquences T2. Utiliser la

totalité de l'IRM n'est pas toujours l'approche la plus efficace, car toutes les régions ne sont pas pertinentes pour notre classifieur. En revanche, utiliser l'ensemble de l'image rend la complexité de l'approche élevée, ce qui rendrait la méthode coûteuse en termes de calcul. Par conséquent, l'utilisation d'une zone plus petite de notre IRM peut s'avérer plus efficace et plus précis que l'utilisation de l'IRM entière. Ainsi, nous avons choisi d'utiliser les 10 coupes centrales de notre IRM d'entrée, quelles que soient les dimensions initiales.

Les données utilisées pour la préparation de cet article proviennent de la base de données d'Alzheimer's Disease Neuroimaging Initiative (ADNI) [4]. L'ADNI a été lancée en 2003 sous la forme d'un partenariat public-privé, dirigé par le chercheur principal Michael W. Weiner, MD. L'objectif principal de l'ADNI est de vérifier si l'imagerie par résonance magnétique (IRM), la tomographie par émission de positons (TEP), d'autres marqueurs biologiques et l'évaluation clinique et neuropsychologique peuvent être combinés pour mesurer l'évolution de la déficience cognitive légère (MCI) et de la maladie d'Alzheimer précoce (AD).

L'ADNI est l'une des plus grandes bases de données cliniques disponibles et de qualité contrôlée, avec une large distribution d'âge et de sexe. Elle vise à développer une compréhension claire de la maladie d'Alzheimer. L'utilisation d'un tel ensemble de données peut améliorer les techniques du diagnostic pour la détection précoce de la maladie d'Alzheimer et donc fournir une meilleure aide aux cliniciens.

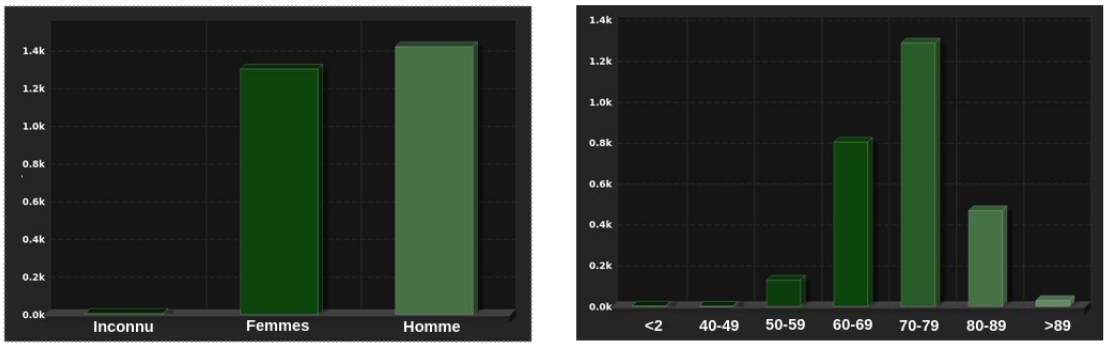


FIGURE 3.6 – La distribution de la base de données ADNI en fonction d'âge et de sexe en 2022

4. http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

3.4 L'anonymisation liée aux équipements d'acquisition de l'IRM

Les multiples étapes des pipelines de dispositifs d'acquisition de l'imagerie médicale laissent des empreintes révélatrices qui peuvent être utilisées comme des signatures d'identification. Généralement, bien que les détails personnels du patient soient exclus des données d'entrée, appliquer des architectures simples d'apprentissage profond peut contribuer à l'identification de l'équipement d'acquisition, et donc conduire à quelques institutions médicales possible représentant déjà des premiers indices sur l'identité le patient. Et lorsqu'il s'agit de maladies rares, cela peut permettre d'identifier directement le patient. De nouvelles techniques sont nécessaires de toute urgence pour limiter ce risque de confidentialité.

Notre première contribution a pour but de surmonter la problématique des traces liées au dispositif d'acquisition dans l'IRM afin de protéger la vie privée des patients. La première partie sert à fournir un modèle qui a la capacité de classifier l'IRM en fonction du constructeur de l'équipement. Ces empreintes laissées sur l'image par l'équipement IRM lors de l'acquisition sont invisibles à l'œil nu. Il est donc difficile de se fier aux méthodes classiques d'extraction de caractéristiques. Par conséquent, nous proposons d'utiliser l'apprentissage profond pour la classification. Le réseau de neurones prend en entrée l'IRM et donne en sortie le constructeur du dispositif de d'acquisition. Nous testons notre approche sur un jeu de données de la base ADNI. Notre classifieur a montré de bonnes performances pour distinguer les différents constructeurs de machines IRM.

La génération et la reconstruction de données est également l'un des défis les plus émergents dans le domaine médical (voir section 3.2.2). De nombreux travaux de recherche récents se concentrent sur cette tâche tout en abordant les questions d'anonymisation [287][112]. L'anonymisation représente le processus d'élimination irréversible des données permettant d'identifier la personne par tous les moyens possibles. Cependant, nous avons déjà prouvé que la suppression des méta-données de l'image IRM n'est clairement pas suffisante pour une véritable anonymisation. Dans cette perspective, la deuxième partie de notre contribution scientifique

est de proposer une approche basée sur un encodeur-décodeur pour reconstruire les données IRM sans les empreintes liées à l'équipement. Enfin, nous combinons notre classifieur avec l'auto-encodeur afin de générer des données synthétiques de haute qualité et résistantes aux attaques de confidentialité.

Pour valider notre proposition, nous mesurons la métrique Peak Signal-to-Noise Ratio (PSNR) sur l'image reconstruite afin de nous assurer de la qualité de l'image et des faibles performances de la classification. Nous collaborons avec des spécialistes du domaine pour approuver la qualité de ces données pour le diagnostic médical. Enfin, nous observons que la reconstruction automatique de l'IRM diminue la précision du classifieur de 95 % à 86 %. Mais, en utilisant notre fonction d'erreur personnalisée pour entraîner notre modèle, la précision du classifieur est réduite à 58 %.

3.4.1 Approches et architectures

Dans cette section, nous détaillons les architectures déployées, en présentant leurs hyperparamètres, en examinant et en discutant leurs résultats.

3.4.1.1 Jeu de données

Nous avons choisi de travailler sur les trois fabricants dominants d'IRM : SIEMENS, Philips Medical Systems et GE MEDICAL SYSTEMS. Le protocole d'IRM de la base de données que nous avons choisi pour nos expériences se concentre sur l'imagerie du plan d'acquisition axial cohérent sur des scanners 3T en utilisant des séquences pondérées en T2. Nous avons choisi d'entraîner notre modèle d'abord sur les scanners de dépistage de 500 patients pour chaque fabricant en utilisant 1/5 de cet ensemble total de données d'entraînement comme ensemble de données de validation. Ensuite, pour évaluer la performance de notre modèle, nous avons utilisé le scanner de dépistage de 100 patients différents comme ensemble de données du test.

Deux autres phases sont importantes avant l'étape de classification : le prétraitement des IRM et l'extraction des coupes (voir section 1.4.2). L'étape de prétraitement effectue une étape de correspondance d'histogramme pour normaliser les distributions des valeurs de pixel

à travers les volumes d'image dans l'ensemble de données. Elle comprend également une étape de normalisation de la carte des voxels. Les images d'entrée, pré-traitées et échantillonnées à la dimension $228 \times 228 \times 1$, sont les entrées de notre classifieur.

3.4.1.2 La classification en fonction de différents équipements

La première partie de notre travail vise à fournir un classifieur capable de distinguer les différents équipements IRM. Pour ce faire, notre modèle d'apprentissage profond est basé sur un CNN. Afin de trouver le modèle optimal pour la meilleure prédiction, nous avons effectué une recherche approfondie de la meilleure architecture de réseau de neurones et de ses hyperparamètres. Nous avons effectué des essais systématiques en commençant par une architecture à deux blocs et en augmentant progressivement la profondeur pour améliorer la performance d'apprentissage de notre modèle jusqu'à s'approcher d'un point de saturation. L'architecture retenue du réseau, illustrée sur la figure 4.1, se compose de quatre blocs de convolution. Elle est similaire à l'architecture de base de classifieur utilisé par Vinyals et al. [264] précédés d'une couche de mise en commun de la moyenne globale. Chaque bloc de convolution est constitué d'une convolution $(3 \times 3 \times 1)$, suivie d'une fonction d'activation non linéaire ReLU et d'une couche de max-pooling $(2 \times 2 \times 1)$. Les deux premiers blocs contiennent respectivement 32 et 64 filtres, suivis de deux blocs de 128 filtres. La sortie des quatre blocs convolutifs est transmise à une couche d'aplatissement (Flatten) suivie de deux blocs denses de respectivement de 32 et 3 unités, d'une fonction d'activation ReLU et d'une fonction d'activation Softmax (voir section 1.3.5). La sortie finale est le nombre de classes représentant le nombre de fabricants d'IRM.

3.4.1.3 La reconstruction de l'IRM

Une contribution scientifique importante de notre travail consiste à la reconstruction de l'IRM après l'avoir encodé dans un espace dimensionnel latent afin d'effacer les empreintes digitales liées à l'équipement. Cette étape de reconstruction est la partie la plus importante dans la perspective de l'anonymisation de l'IRM. Par conséquent, nous utilisons un auto-encodeur

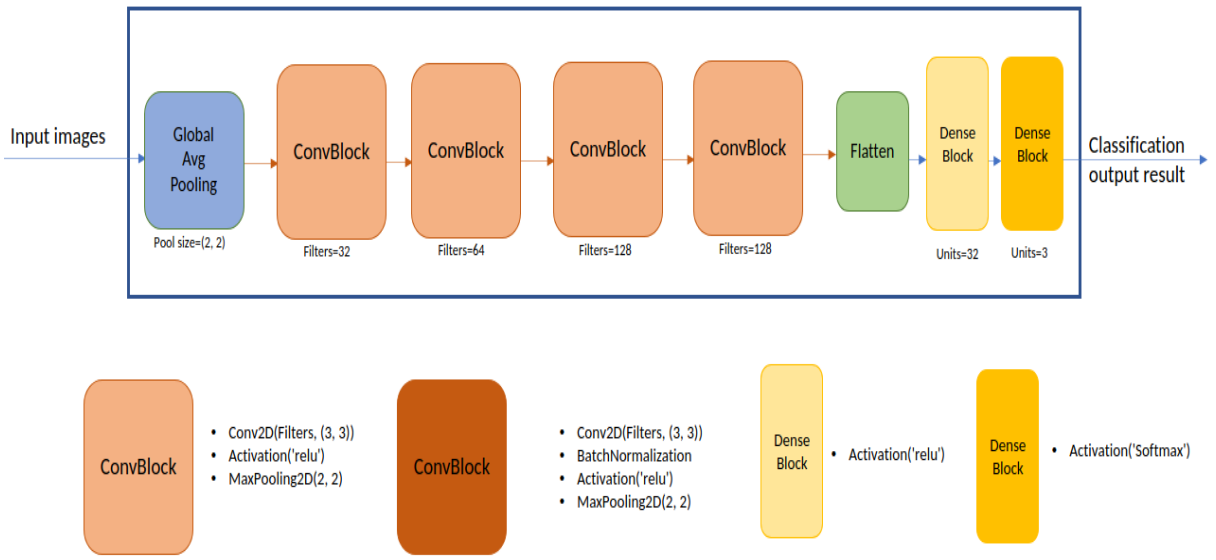


FIGURE 3.7 – Architecture proposée pour la classification des équipements d'IRM

illustré dans la figure 3.8 comme la branche d'encodeur-décodeur pour reconstruire les coupes d'IRM de $228 \times 228 \times 1$ et on observe les résultats. L'encodeur consiste en 3 blocs de convolution, chaque bloc incorporant une couche de convolution et une couche de batch normalization. Une couche de downsampling max-pooling suit les deux premiers blocs de convolution. Le premier bloc contient 32 filtres, suivi de deux blocs de 64 filtres. De même, le décodeur se compose de 3 blocs convolutifs, chaque bloc comprenant une couche de convolution et une couche de batch normalization. Une couche d'upsamling suit les deux premiers blocs de convolution. Le premier et le deuxième bloc contiennent respectivement 64 et 32 filtres, suivis d'un bloc d'un seul filtre reconstruisant l'IRM d'entrée.

L'étape suivante est la combinaison du classifieur et l'auto-encodeur afin de pouvoir minimiser au mieux le risque de pseudo-anonymisation (fig. 3.8). La section suivante détaille la fonction d'erreur que nous avons adaptée pour entraîner notre auto-encodeur. L'entraînement séquentiel du classifieur et de l'auto-encodeur diminue la précision du classifieur tout en maintenant la capacité de l'auto-encodeur à reconstruire une IRM de qualité.

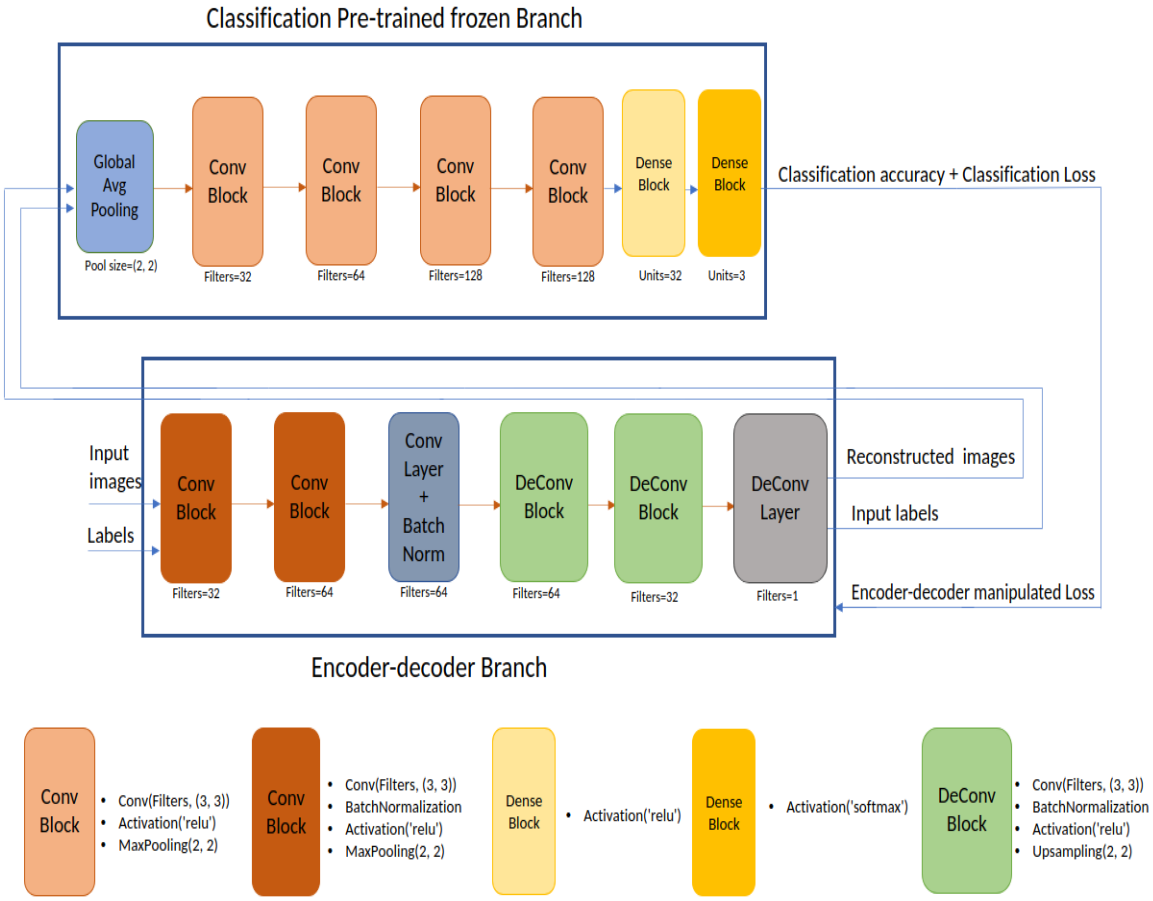


FIGURE 3.8 – Architecture proposée pour la reconstruction de l'IRM

3.4.2 Reformulation mathématique

L'étape finale de notre pipeline consiste à manipuler la fonction de perte de notre auto-encodeur pour qu'il soit contraint d'avoir le minimum de précision de classification possible, ce qui peut également se traduire par une valeur de fonction d'erreur élevée. On peut reformuler mathématiquement cette étape de notre approche comme un problème d'optimisation représenté par le système de Karush–Kuhn–Tucker (KKT).

Soit X l'entrée et $L_\theta(X)$ la fonction d'erreur MSE de notre auto-encodeur.

Objectif : $\hat{\theta} = argmin L_\theta(X)$

$X' = f_\theta(X)$: La reconstruction de X par l'auto-encodeur

où

$$\begin{cases} \theta & \text{poids} \\ f_\theta & \text{encodeur-decodeur} \end{cases}$$

Alors : θ représente la solution du système KKT suivant :

$$(A) : \begin{cases} min_\theta L_\theta(X) \\ max_\theta L_c(X') \end{cases}$$

où

$$\begin{cases} L_c & \text{l'erreur de classification} \\ L_c(X') = L_c(f_\theta(X)) & \\ max_\theta L_c(X') & \text{la contrainte du kkt} \end{cases}$$

La solution θ de (A) vérifie $min_\theta(L_\theta(X) - \lambda L'(X)$ ou $min_\theta(L_\theta(X) + \lambda exp(-L'(X))$.

Cette nouvelle fonction d'erreur prend en compte la contrainte d'avoir la valeur d'erreur de classification la plus élevée possible tout en conservant la performance de la reconstruction d'IRM.

3.4.3 Expérimentations et résultats

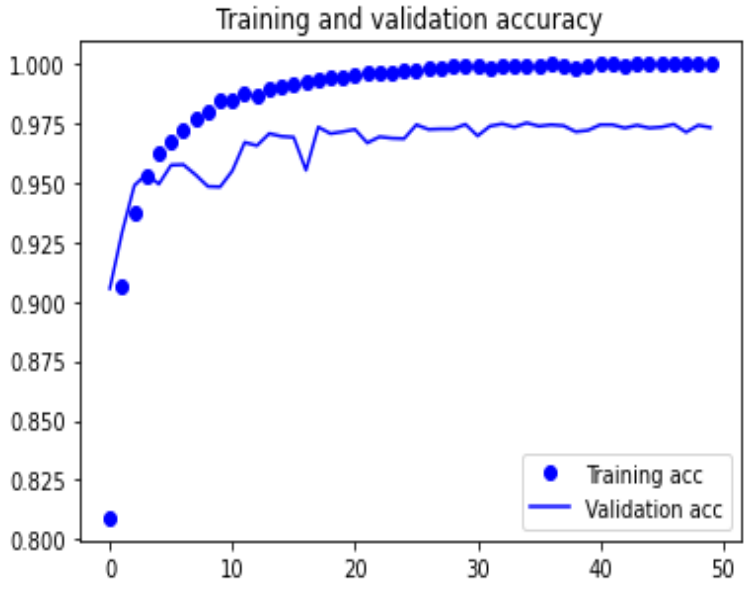


FIGURE 3.9 – Courbe de la précision de classification de l'entraînement et de la validation de la classification

Nous procédons à un entrainement par cycle. Un cycle est composé de plusieurs époques. Le passage d'un cycle à un autre suit simplement les principes d'un apprentissage par transfert.

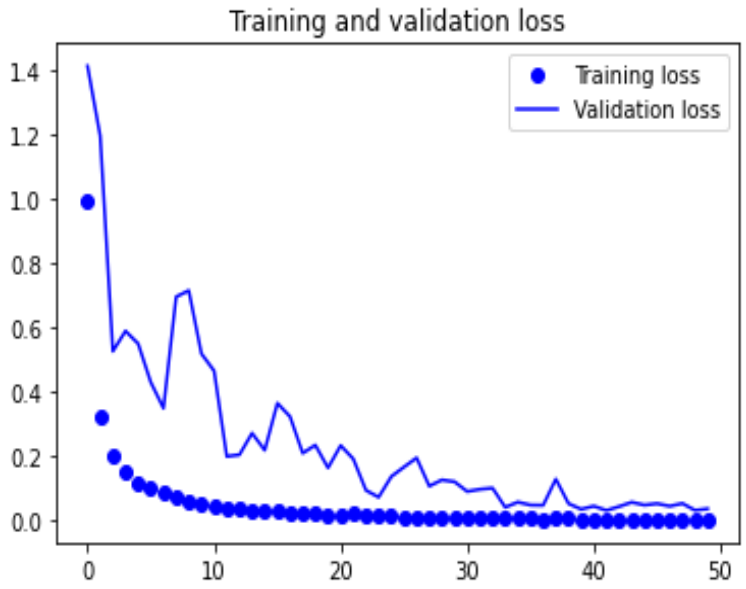


FIGURE 3.10 – Courbe de la fonction d'erreur de l'entraînement et de la validation de la classification

Notre entraînement pour la classification ne comprend que 5 cycles de 50 époques. Les courbes de la précision de classification de l'entraînement et de la validation du premier cycle, illustrées dans la Figure 3.9 montrent que notre apprentissage atteint une précision de classification très élevé après seulement 20 époques. Nous avons validé notre modèle pour la classification des constructeurs d'IRM sur l'ensemble de données du test. Comme le montrent le tableau 3.1 et la matrice de confusion (fig. 3.11), notre approche atteint une précision de classification de 95 % et une perte très réduite (fig.3.10) démontrant la performance de notre méthode et sa possible utilisation pour des applications cliniques réelles.

MRI	**GE**	**Philips**	**Siemens**
Precision	0.92	0.97	0.98
Recall	0.99	0.93	0.95
CA	0.97	0.97	0.98
F1-score	0.95	0.95	0.97

TABLE 3.1 – Résultats sur les données test de la classification des constructeurs d'IRM

Notre auto-encodeur a également prouvé une bonne performance. Plusieurs méthodes dans l'état de l'art sont proposées pour évaluer la qualité des images après des manipulations de compression ou de déformation [50][33][209]. Pour ce travail d'initiation, on a choisi d'utiliser la PSNR, un rapport utilisé comme mesure de la qualité visuelle entre l'image originale et une

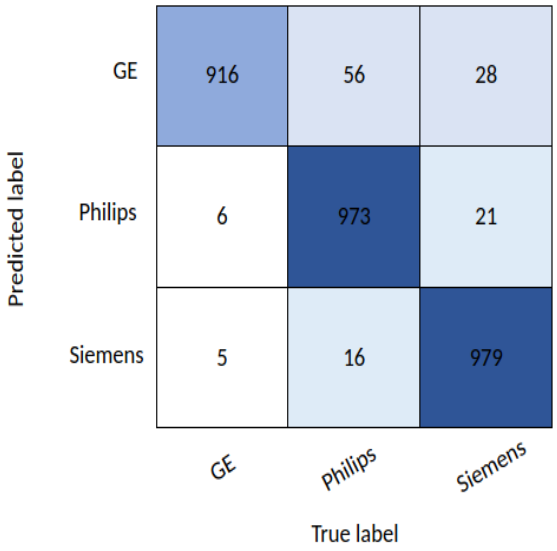


FIGURE 3.11 – Matrice de confusion des données test de la classification

image reconstruite. Cette métrique est utilisée principalement pour observer la qualité visuelle de l'IRM reconstruite. Bien que la PSNR a atteint seulement 21,98 dB après un seul cycle, visuellement, la qualité de la reconstruction était déjà élevée. Pour confirmer cette qualité, nous avons collaboré avec des spécialistes dans le domaine de l'imagerie médicale. Ces derniers ont confirmé que ces images résultantes peuvent être utilisées dans les applications cliniques pour lesquelles elles sont destinées, à savoir les tâches de diagnostic médical. Des échantillons d'images d'entrée de l'auto-encodeur comparés aux images reconstruites sont illustrés dans la figure 3.12. Nous montrons également la courbe de fonction d'erreur décroissante de notre auto-encodeur (Fig 3.13) prouvant sa performance.

La dernière étape de notre approche est la combinaison adaptée de notre classifieur et notre auto-encodeur. Tout d'abord, nous testons les images reconstruites de l'ensemble de données de test dans notre classifieur et la précision de classification diminue à 86 %. Ensuite, nous intégrons la fonction d'erreur améliorée liant le classifieur à l'auto-encodeur. Nous avons testé plusieurs valeurs de λ et comparé l'efficacité de notre algorithme sur le jeu de données (Figure 3.14). L'objectif de cette étape est de trouver le meilleur compromis entre la précision de classification de la classification et les performances de l'auto-encodeur (Figure 3.15). Nous avons diminué la précision à 58 % en maintenant un PSNR de 18,48 dB en utilisant $\lambda = -0,008$.

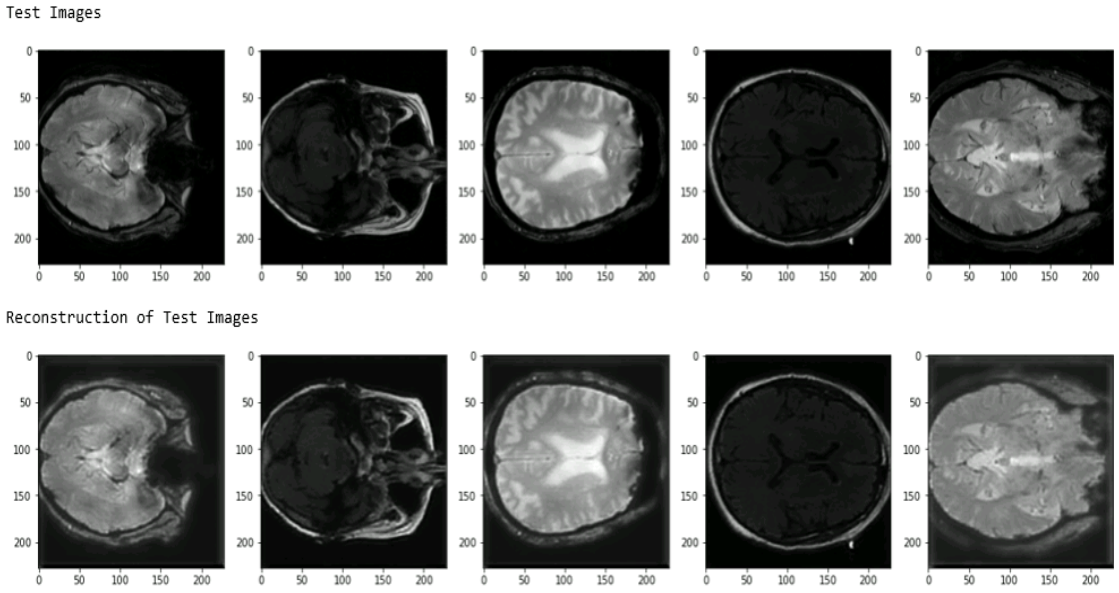


FIGURE 3.12 – Des échantillons des images reconstruites

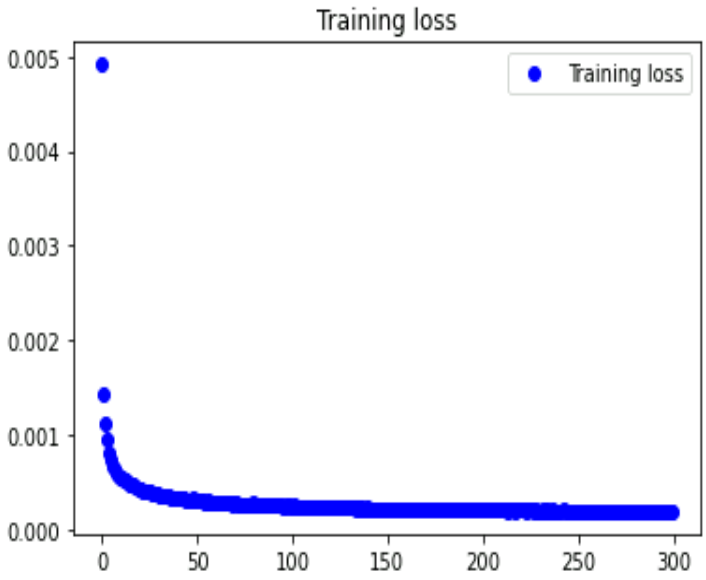


FIGURE 3.13 – La courbe de la fonction d'erreur d'auto-encodeur

Par conséquent, et comme le montre la matrice de confusion (fig. 3.16), notre reconstruction est capable de diminuer l'efficacité du classifieur et de l'induire en erreur. D'autre part, la valeur PSNR proche de PSNR à l'origine prouve que notre approche préserve une qualité visuelle de l'IRM permettant de l'utiliser pour de multiples tâches du diagnostic médical. Nous avons, également à cette étape, montré quelques échantillons d'images reconstruites à des experts médicaux avec lesquels nous collaborons étroitement. Ils ont pu évaluer visuellement

les changements dans les caractéristiques des tissus (matières blanche et grise) et l'atrophie cérébrale progressive des images reconstruites de patients souffrant d'Alzheimer, validant ainsi la qualité des images reconstruites et le diagnostic possible sur cette tâche spécifique.
Nous avons cherché la source qui explique la valeur de la PSNR limitée en dépit de la bonne qualité de la reconstruction confirmée par nos spécialistes. L'investigation a donné qu'il y avait des IRM de phase dans la base de données avec des spécifications différentes. Leurs éliminations n'a pas influé sur les résultats de classification, mais a amélioré de loin la valeur de PSNR à une moyenne totale de 42.37 dB.

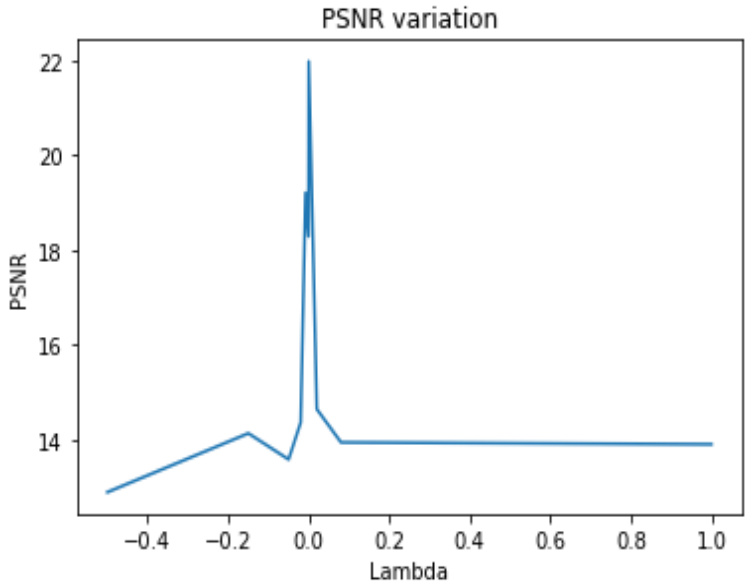


FIGURE 3.14 – La courbe de la PSNR en fonction de Lambda

3.5 L'anonymisation liée à l'identité du patient

Notre travail était à notre connaissance le premier travail qui traite la problématique des empreintes des équipements d'acquisition d'imagerie médicale ayant comme objectif l'anonymisation profonde. Suite aux résultats satisfaisants sur les fabricants IRM, nous proposons d'aller plus loin dans cette perspective.

La deuxième contribution de ce chapitre est consacrée aux traces liées à l'identité du patient présentes dans l'IRM. En effet, nous espérons que l'approche s'adapte à cette

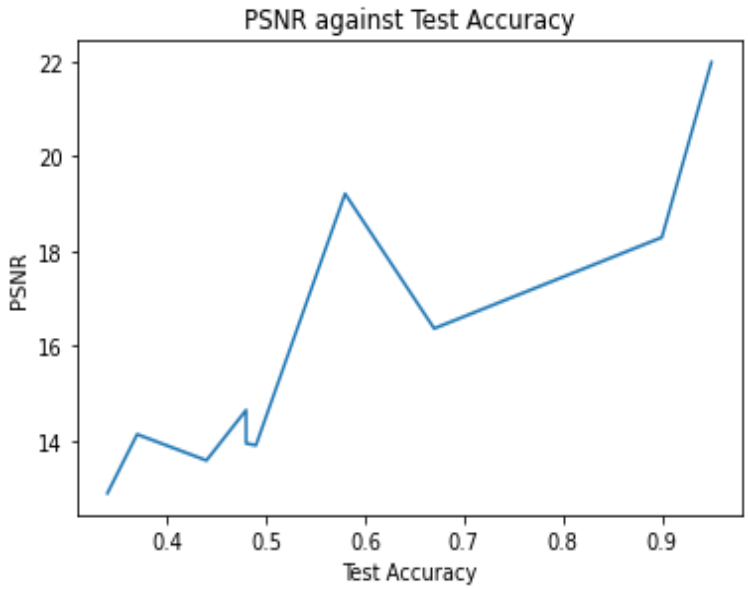


FIGURE 3.15 – La courbe de la PSNR en fonction de la précision de classification du test

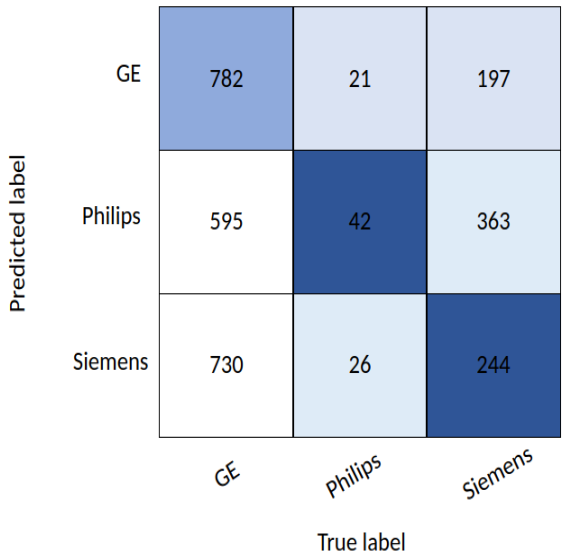


FIGURE 3.16 – Matrice de confusion des données de test de notre approche finale

nouvelle problématique et pour protéger la vie privée des patients. Comme précédemment, les méta-données qui permettent d'identifier le patient sont exclues. En utilisant l'apprentissage profond, nous prouvons encore un problème de pseudo-anonymisation en retrouvant des traces distinguant les patients. La première partie de cette section a pour but de fournir un modèle qui a la capacité de classifier l'IRM en fonction de l'identité du patient. Cependant, contrairement aux dispositifs d'acquisition, la quantité de données n'était pas suffisante pour effectuer un

entraînement classique. Dans ce but, nous proposons d'utiliser l'une des techniques présentées dans le chapitre 2 permettant d'utiliser les réseaux de neurones avec une petite quantité de données. Ensuite, similairement à la section précédente, nous procédons à une reconstruction des données d'une façon à produire des images d'une qualité suffisante pour le diagnostic médical et à préserver l'identité des patients.

Notre dernière contribution combine les deux réseaux : celui du classifieur et celui de la reconstruction. L'objectif est de tromper le classifieur tout en conservant la meilleure qualité d'image reconstruite. Nos expériences ont prouvé que cette combinaison de réseaux améliore la performance globale. Cette piste pourrait être très utile dans la direction de l'anonymisation dans le domaine de l'imagerie médicale.

3.5.1 Approches et architectures

L'objectif de cette partie est d'utiliser l'apprentissage profond pour résoudre le problème de l'identification des patients à travers l'IRM lié afin de protéger leur vie privée. Dans cette sous-section, nous allons présenter notre pipeline avec les architectures et les approches proposées.

3.5.1.1 Jeu de données

Dans cette partie, nous utilisons toujours des séquences des IRM pondérées en T2 avec un plan d'acquisition axial cohérent sur des scanners 3T. Nous avons choisi d'entraîner notre modèle d'abord sur les scanners de dépistage de 20 patients, 9 IRM par patient, en utilisant 2 IRM de cet ensemble total de données d'entraînement comme ensemble de données de validation. Ensuite, pour évaluer la performance de notre modèle, nous avons utilisé 3 scanners de dépistage des 20 patients comme ensemble de données du test.

On effectue une extraction de 20 coupes centrales (voir section 1.4.2). L'étape de prétraitement comporte une normalisation de Nyul and Udupa et également une étape de normalisation de la carte des voxels. Les images d'entrée, pré-traitées et échantillonnées à la dimension $228 \times 228 \times 1$, sont les entrées de notre classifieur.

3.5.1.2 La classification en fonction de l'identité de patients

Dans la première partie de notre travail, nous envisageons fournir un classifieur capable de distinguer l'identité des patients. Pour ce faire, nous avons procédé au déploiement de la même architecture utilisée pour l'identification des équipements d'acquisition. Nous avons également effectué les essais en nous basant sur le nombre de blocs et en augmentant progressivement la profondeur pour améliorer la performance d'apprentissage de notre modèle. L'architecture retenue du réseau est la même architecture illustrée sur la figure 4.1 et décrite dans la section 3.4.1.2. La sortie finale est le nombre de classes représentant le nombre de patients. Nous avons jugé les résultats moyennement satisfaisants à cause de la quantité de données limitée. Ainsi, nous avons implémenté une architecture siamoise (voir sec 2.4.2.1). Cette approche a amélioré la performance de classification.

3.5.1.3 La reconstruction de l'IRM

Le pipeline de cette partie est similaire à la reconstruction de l'IRM liée aux équipements 3.8. Effectivement, cette étape de reconstruction, éliminant les dernières traces en lien avec l'identité des patients, est la partie la plus importante dans la perspective de l'anonymisation de l'IRM. Identiquement à la section 3.4.1.2, une fonction d'erreur est adaptée pour combiner le classifieur et l'auto-encodeur. L'entraînement séquentiel du nouvel auto-encodeur diminue la précision du classifieur tout en maintenant la performance de la reconstruction de l'auto-encodeur.

3.5.2 Expérimentations et résultats

Similairement à l'approche de classification adaptée au-dessus, nous avons procédé par à un entrainement par 5 cycles de 50 époques. Les courbes de la précision de classification de l'entraînement et de la validation du dernier cycle, illustrées dans la figure 3.17 montrent que notre apprentissage atteint une précision de classification limitée. Nous avons validé notre modèle pour la classification des patients sur l'ensemble de données du test. Notre approche atteint seulement une précision de classification de 67 %.

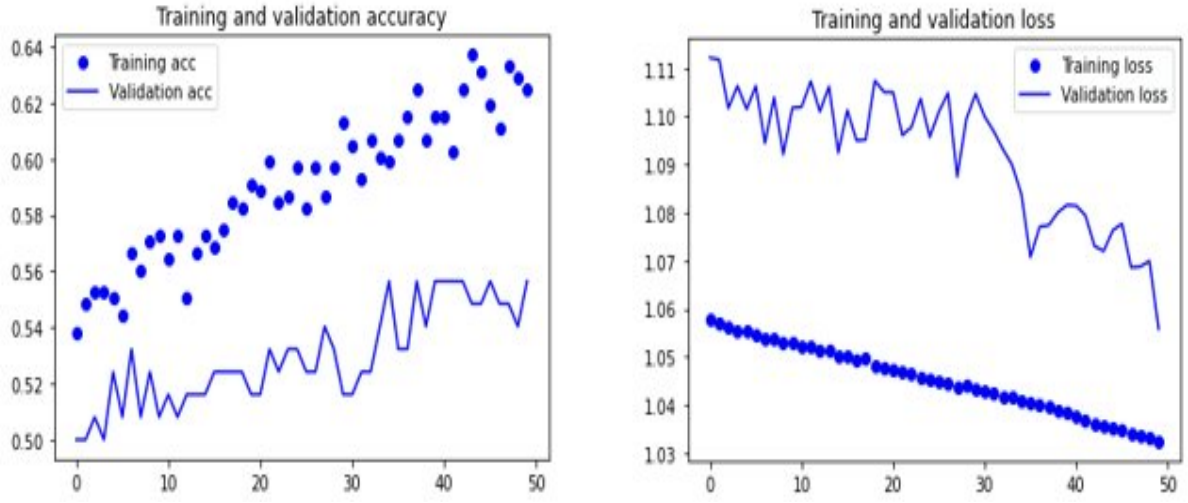


FIGURE 3.17 – Courbe de la précision de classification et de la perte de la méthode de la classification classique

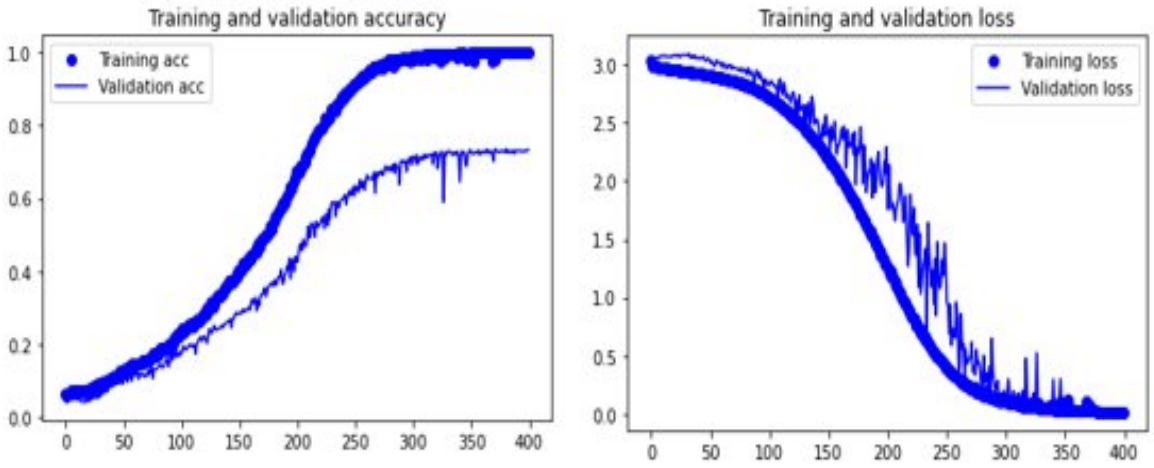


FIGURE 3.18 – Courbe de la précision de classification et de la perte de la méthode de réseau seamois

Afin d'améliorer ces résultats de classification et en raison de la quantité de données réduite, nous avons proposé une architecture siamoise. Nous avons entrainé notre modèle sur 400 époques (fig. 3.18). Nous avons déjà constaté une nette amélioration sur les courbes de précision, de classification et de perte. Nous avons validé notre modèle pour la classification des patients sur l'ensemble de données du test. Comme le montre la matrice de confusion (fig. 3.19), notre approche atteint une précision de classification de 92 %, démontrant la performance de notre méthode et sa possible utilisation pour des applications cliniques réelles.

Confusion Matrix

```
[[19  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  3  0  0  1  0  2  0  0  0  0  1  0  0  2  0  0  0  0  0]
 [ 0  0  8  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0 11  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  4  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  4  0  0  1  0  0  0  0  0  0  0  1  0  0  0]
 [ 0  0  0  0  0  0 10  0  0  0  0  0  1  0  0  1  0  0  0  0]
 [ 0  0  0  0  1  0  0  5  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  1  1  0  0  0  0  7  3  0  0  0  1  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  1 15  0  0  0  0  0  0  0  0  0  0]
 [ 2  0  0  0  0  0  0  0  0  1 12  0  1  0  0  0  0  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  1  0 12  0  2  0  0  0  0  0  0]
 [ 1  0  0  0  0  0  0  0  0  0  0  0 10  0  0  0  1  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  0  0  0  0  9  0  0  3  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  1  0 12  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  2  0  0 12  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0  2  0  0  9  0  0  0]
 [ 0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  4  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 11  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  5]]
```

FIGURE 3.19 – Matrice de confusion des données test de la classification

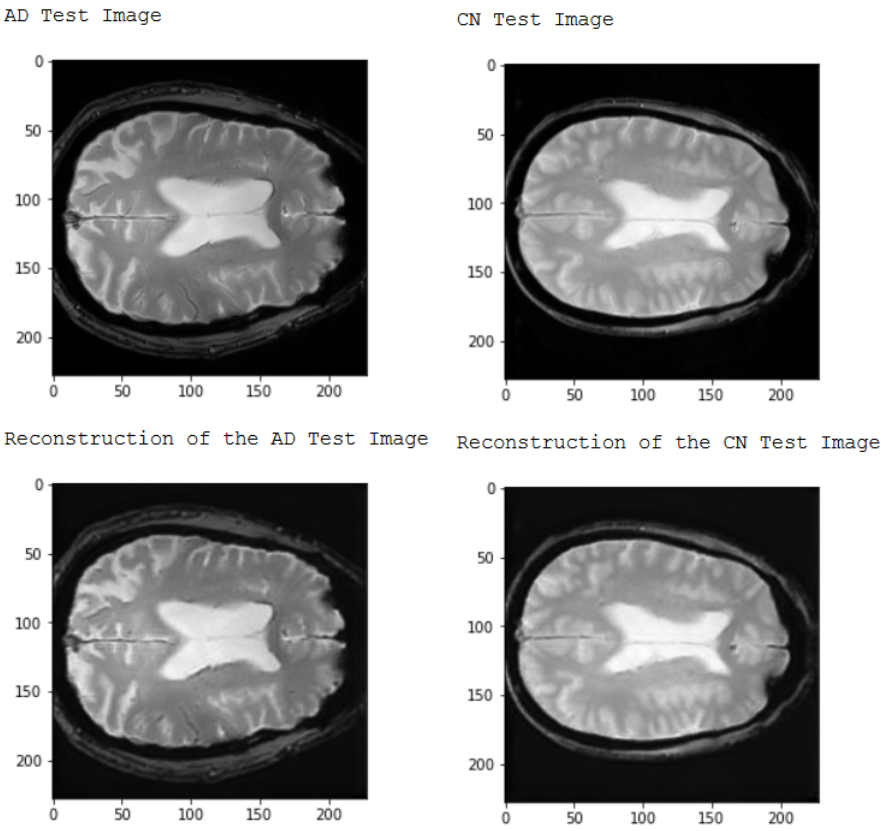


FIGURE 3.20 – Exemple d'images reconstruites par l'auto-encodeur avec la fonction d'erreur adaptée. À gauche, l'image originale et l'image reconstruite d'un patient souffrant de la maladie d'Alzheimer, à droite, les images d'un patient sain.

La dernière étape de notre approche est la combinaison adaptée de notre classifieur et notre auto-encodeur, en intégrant la fonction d'erreur améliorée liant le classifieur à l'auto-encodeur. L'objectif de cette étape est toujours le meilleur compromis entre la précision de classification de la classification et les performances de l'auto-encodeur. Notre auto-encodeur adapté a prouvé une bonne performance. La PSNR a atteint 31,98 dB. La qualité des images reconstruites était également confirmée auprès de nos spécialistes (fig. 3.20). Nous avons réussi à diminuer la précision de classification à 63 % comme le montre la matrice de confusion (fig. 3.21), notre reconstruction est capable de diminuer l'efficacité du classifieur et de l'induire en erreur. D'autre part, la valeur PSNR prouve que notre approche préserve une qualité visuelle de l'IRM nécessaire aux multiples tâches du diagnostic médical.

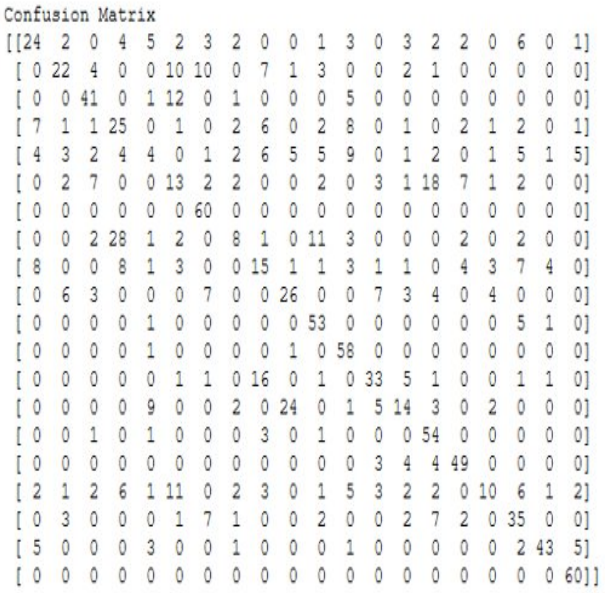


FIGURE 3.21 – Matrice de confusion des données de test de notre approche finale

3.6 Conclusion

Dans la première partie de chapitre, nous nous sommes concentrés sur les traces de l'équipement d'acquisition IRM. Nous avons proposé deux architectures différentes pour deux tâches différentes. La première architecture vise à classifier les appareils d'IRM par rapport aux fabricants. Cette étape a légitimé notre préoccupation quant à la capacité d'identifier les patients par l'identification de l'équipement utilisé pour l'acquisition de l'IRM, en particulier pour les

maladies rares. La deuxième architecture vise à reconstruire une IRM avec une approche basée sur un auto-encodeur. Elle génère des images reconstruites qui permettent de lutter efficacement contre les attaques de pseudo-anonymisation. La combinaison adaptée de ces deux architectures constitue une première tentative d'anonymisation. Cette dernière étape a prouvé le potentiel de notre approche pour diminuer la capacité de la distinction de l'équipement d'acquisition IRM. Les résultats obtenus montrent que les empreintes digitales de l'appareil sont importantes, mais que leur élimination est possible. Ce cadre est le premier pas vers une véritable anonymisation de l'imagerie médicale.

Pour répondre au mieux au besoin urgent créé par les préoccupations liées à la confidentialité, nous avons continué dans la même perspective en utilisant l'apprentissage profond pour éliminer les traces invisibles qui peuvent aider à identifier un patient. Par conséquent, nous avons utilisé principalement le même pipeline. En revanche, pour cette partie, la quantité de données était relativement limitée. C'est pourquoi, nous avons eu recours à l'une des méthodes les plus répandues pour la problématique de few-shot learning : le réseau siamois. Ainsi, nous avons établi et prouvé un processus significatif qui offre à l'institution médicale une garantie de confidentialité et permet aux chercheurs d'obtenir des données anonymes de haute qualité visuelle pour des futures applications cliniques. Notre approche peut également être testée sur différentes séquences IRM et d'autres types d'imagerie médicale. Une perspective claire est d'évaluer nos images anonymes dans différentes tâches de diagnostic.

Chapitre

4 Meta-apprentissage attentif

Sommaire

4.1 Motivation 120
4.2 État de l'art 123
4.3 Approches et reformulations 125
4.3.1 Réseau de neurones attentif 125
4.3.2 Réseau de neurones attentif amélioré 129
4.4 Expérimentations et résulats 130
4.4.1 Bases de données 130
4.4.2 Expérimentations 132
4.4.3 Résultats 133
4.4.4 Conclusion 134

4.1 Motivation

Le secteur médical est un secteur de service critique, important et polémique. Dans ce domaine, on déploie différents examens cliniques résultants de la recherche et des multiples technologies biomédicales pour diagnostiquer et traiter les pathologies, les traumatismes et les blessures, classiquement en ayant recours à la prescription de médicaments, l'intervention chirurgicale ou d'autres multiples formes de thérapies des soins de santé. Tous les secteurs sont importants. Néanmoins, le secteur médical est un secteur qui est très sacré parce qu'il est en lien direct avec l'espérance et la qualité de vie garantie par la bonne santé. Cela explique l'intérêt de la communauté de recherche à développer des nouvelles solutions et à proposer des nouvelles techniques pour améliorer l'efficacité du diagnostic et du soin médical. Le diagnostic précoce et la médecine prédictive permettent la décélération et peut-être le traitement définitif des maladies considérées encore incurables il y a quelques années. Dans cette direction, les travaux de recherche ont donné naissance à de multiples techniques qui se basent sur l'intelligence artificielle. Ces techniques ont abouti à des performances record dans de nombreuses tâches médicales [148][87][140]. La performance des techniques utilisant l'apprentissage profond a surpassé même les performances humaines dans certaines applications [90]. Cependant, comme expliqué dans le chapitre 2, la quantité et la qualité des données sont des exigences cruciales permettant d'extraire les bonnes caractéristiques, d'entrainer des modèles les plus complexes et d'obtenir les meilleurs résultats. En revanche, dans plusieurs cas d'utilisation, la collecte d'une grande quantité de données équilibrées, étiquetées, et de haute qualité est une tâche difficile. De nombreux défis sont à prendre en compte dans cette perspective tels que l'accès aux données, la confidentialité des données, l'indisponibilité des données, la qualité des données, l'étiquetage [223]. En outre, l'entraînement du modèle avec une telle quantité de données est coûteux en termes de temps de calcul. De plus, après avoir relevé ces défis, ces données peuvent devenir rapidement obsolètes ou peuvent nécessiter une mise à jour, ce qui implique une nouvelle collecte de données et un nouvel entraînement. En particulier, dans le domaine de l'imagerie médicale, les données représentent un obstacle significatif lors du déploiement de l'apprentissage profond pour différentes tâches, notamment la détection, la reconnaissance

et la segmentation des objets et la classification des images. Cette dernière tâche est l'un des premiers et des plus importants cas d'utilisation de l'apprentissage profond dans le domaine médical, prenant en entrée une ou plusieurs images médicales et donnant en sortie une seule variable informative pour aider l'expert médical à réaliser son diagnostic. Néanmoins, c'est aussi une des tâches qui a un besoin avide de données et qui en manque énormément, ce qui explique la popularité de l'apprentissage par transfert pour une telle tâche [148]. L'apprentissage par transfert est une technique permettant l'utilisation d'un modèle pré-entraîné, généralement sur des millions d'images naturelles, pour compenser le manque de données dans le domaine médical. En effet, cette approche utilise les hyperparamètres du modèle pré-entraîné sur une tâche de classification sur une autre sans avoir à recommencer l'apprentissage de zéro. La disponibilité de ces modèles pré-entraînés et le fait qu'ils puissent simplement être appliqués directement à n'importe quelle image médicale facilitent leur utilisation. Cependant, à cause des limites mentionnées dans la section 2.4.1, cette technique présente des performances limitées lorsqu'il s'agit de maladies rares. Une étude directe de l'impact du volume de données et de la similarité du domaine a prouvé que l'efficacité de l'apprentissage par transfert dépend de la similarité du domaine et du volume de données disponibles pour l'ajustement [20]. Spécifiquement pour les maladies rares avec régime de données faibles, la simple disponibilité d'un ensemble de données suffisant est difficile, ainsi l'ajustement du modèle conduira directement à un sur-apprentissage. Par conséquent, afin d'atténuer les difficultés liées à la classification des maladies rares, des percées ont été réalisées pour trouver de nouvelles façons d'utiliser l'apprentissage profond avec peu de données en se basant sur d'autres critères critiques différents pour former un réseau de neurones, tels que l'algorithme d'optimisation, la fonction de perte et l'initialisation des hyperparamètres. Les défis liés aux maladies rares retiennent peu l'attention malgré son importance et son rythme de développement incroyablement élevé. En effet, 7 000 maladies sont considérées comme des maladies rares connues [107] touchant 400 millions de personnes dans le monde [119]. Il s'agit donc d'un sujet de recherche émergent abordé par la communauté de l'imagerie médicale.

Du point de vue de l'apprentissage automatique, les maladies rares dans une population de patients représentent peu de données, ce qui représente des classes peu représentées. De

nos jours, plusieurs solutions ont été expérimentées pour faire face au manque de données étiquetées et de bases de données déséquilibrées, principalement divisées en deux catégories : les techniques basées sur les données et les techniques basées sur les connaissances préalables (voir section 2.1). L'approche de méta-apprentissage propose d'entraîner un modèle flexible sur diverses tâches connexes avec des petites bases de données afin d'offrir la possibilité d'apprendre à effectuer plusieurs tâches et d'utiliser les expériences acquises pour résoudre de nouvelles tâches d'apprentissage sans avoir à entraîner à partir de zéro. En d'autres termes, au lieu de concevoir des modèles pour extraire des caractéristiques, on conçoit une architecture qui apprend quelle est la meilleure manière d'apprendre pour obtenir les meilleurs résultats. Le méta-apprentissage est utilisé pour rendre l'apprentissage plus rapide, flexible aux changements et adaptable à différentes tâches en utilisant seulement un petit ensemble de données et quelques itérations de descente de gradient d'entraînement. Il est basé sur l'architecture du modèle, les hyperparamètres et les jeux de données : le modèle est entraîné par le méta-apprenant pour pouvoir apprendre sur différentes tâches. De nombreux algorithmes ont été expérimentés tels que Model-Agnostic Meta-Learning (MAML), adversarial Meta Learning (ADML), fast Context Adaptation with Meta-Learning (CAML) et Reptile [286][180]. Dans notre travail, nous avons choisi une version adaptée de MAML comme approche de meta-apprentissage.

Il est toujours difficile de trouver la meilleure représentation des caractéristiques d'une classe afin de pouvoir effectuer une bonne classification avec une quantité limitée de données étiquetées. Le méta-apprentissage vise à utiliser les connaissances antérieures et les concepts appris pour pouvoir apprendre à apprendre. Récemment, le mécanisme d'attention a attiré l'attention de la communauté scientifique de chercheurs (voir section 2.5). Inspiré par la façon dont l'homme prête attention à des régions saillantes précises d'une image pour prendre une décision, le mécanisme d'attention vise à imiter le comportement humain pour extraire les caractéristiques les plus pertinentes. En effet, l'idée principale est d'utiliser la relation entre les caractéristiques de différentes classes dans différentes tâches pour pouvoir se concentrer sur les parties saillantes de l'entrée. Le mécanisme d'attention est apparu d'abord dans le traitement du langage naturel (NLP) et la compréhension du langage naturel (NLU). Ensuite, ses applications ont commencé à couvrir d'autres domaines, notamment la vision par ordinateur.

Dans cet article, nous proposons une nouvelle approche pour la classification des maladies rares. La méthode proposée combine une technique de méta-apprentissage avec un mécanisme d'attention afin de proposer un modèle qui améliore sa capacité à apprendre plus rapidement d'une tâche à l'autre en utilisant les expériences acquises précédemment et les connaissances antérieures importantes, en généralisant les multiples concepts appris et en combinant plusieurs compétences déjà apprises.

4.2 État de l'art

Les tâches d'analyse d'imagerie médicale sont toujours assez délicates et complexes, même pour des professionnels de santé expérimentés, étant donné qu'elles nécessitent une compréhension et une attention approfondies aux détails des images. Des années d'expérience sont nécessaires pour devenir qualifié dans une tâche telle notamment la classification ou la segmentation [59] et l'annotation des tumeurs [100], des lésions [163] et des organes humains. Pour ces raisons, l'introduction de l'intelligence artificielle pourrait s'avérer être un progrès significatif et d'une grande aide au diagnostic médical.

Bien que l'apprentissage profond ait abouti à plusieurs avancées dans des multiples tâches du domaine médical, il y a toujours des limitations liées à la disponibilité des données [197]. En effet, les réseaux de neurones convolutifs réussissent à identifier les caractéristiques clés avec une grande quantité de données, mais ils ont tendance à être fortement dépendants des données sur lesquelles ils s'entraînent. Dans le cadre de l'imagerie médicale, dû à la rareté de certaines maladies, les données sont souvent en nombre limité. Pour faire face à cette limitation, des travaux de recherche sont en cours afin d'optimiser l'exploitation d'une quantité de données très réduite. Une réflexion directe dans cette direction consiste à réduire la dépendance des modèles vis-à-vis des données en déployant une approche similaire par analogie à la manière dont les humains peuvent inférer à partir de moins de données. Ainsi sont apparus les approches de méta learning qui tente de diminuer la dépendance aux données en se basant sur une représentation générale, plutôt que spécifique.

Les approches de meta-learning sont utilisées sur des multiples types de données, notamment les données textes [200], mais dans notre travail, nous nous intéressons principalement à l'imagerie médicale [155]. Dans cette section, nous abordons brièvement quelques travaux qui visent la résolution du problème de few-shot learning dans le domaine de l'imagerie médicale en utilisant des techniques de meta-learning. Notamment, pour la segmentation, Zhao et al. [293] utilisent l'augmentation de données dans le cadre du one-shot learning pour la segmentation des IRM cérébrales. Ronneberger et al. [203] ont utilisé le réseau U-Net pour la segmentation des structures neuronales dans les piles de microscopie électronique en few-shot learning. Lahiani et al. [131] ont utilisé des réseaux de neurones profonds à déconvolution de couleur pour la problématique de one-shot learning de la segmentation des tissus cancéreux. Dans [204] et [162], les auteurs ont utilisé des réseaux de neurones profonds respectivement de type "squeeze & excite" et V-net pour segmenter des images volumétriques. En incorporant l'utilisation de réseaux génératifs [79] et en s'inspirant des modèles multimodaux [295], Mondal et al. [170] effectuent la segmentation d'IRM cérébrale 3D multimodale.

Pour la classification d'images médicales, Puch et al. [192] ont souligné encore le besoin de réussir l'implémentation de l'apprentissage profond avec une petite quantité de données suite à une étude du problème de la rareté de certaines maladies. Ainsi, ils ont déployé des réseaux de triplets pour reconnaître les maladies du cerveau. Les auteurs ont fait recourt également à la recherche sur grille pour identifier les hyperparamètres idéaux pour l'architecture. Orting et al. [181] ont également utilisé les réseaux de triplets pour l'analyse des images médicales de l'emphysème dans les scans thoraciques. Kim et al. [121] ont expérimenté le réseau de neurones d'appariement comme approche de meta-learning pour le diagnostic du glaucome à partir d'images du fond de l'œil. Les auteurs ont effectué l'augmentation des données afin d'éviter un sur-ajustement. En effet, une mission clé lors de la manipulation d'imagerie médicale est le maintien d'une image d'haute définition pour l'utilisation dans le cadre clinique réel. Dans ce sens, les auteurs ont recadré, simplement, les images au centre dans le but de se concentrer davantage sur les caractéristiques distinctives de la macula et du disque optique.Similairement, Tri-Cong et al. [188] ont incorporé l'augmentation de

données dans une application importante dans le domaine d'imagerie médicale en lien avec lésion cutanée, en particulier le mélanome. Dans ce dernier travail, trois types d'augmentation étaient soulignés : augmentation géométrique, normalisation des couleurs entre les bases de données et les transformations basées sur les spécificités de l'application pour le mélanome. Un pipeline de deux phases a donné des bonnes performances de classification sur des bases de données reconnues. Les auteurs ont mis en relief l'importance de la préservation de la signification sémantique des classes visées et étaient effectivement conscients que, notamment, les déformations ne sont pas possibles dans toutes les applications. Cependant, l'augmentation de données reste toujours une pratique éthiquement controversée pour le diagnostic médical et encore, elle augmente la quantité de donnée et donc le temps et le cout de calcul. Hu et al. [98] ont utilisé Reptile comme approche de méta-apprentissage pour le diagnostic de la rétinopathie diabétique. Mini-Imagenet a été utilisé pour initialiser les poids lors de meta-entraînement dans le cadre 5-shot 5-way few-shot learning. Le réseau de neurones siamois a été exploré par Chyng et al. [41] pour la recherche d'imagerie médicale basée sur le contenu (CBIR) dans le cadre de la rétinopathie diabétique.

Tous ces travaux aboutissent à des résultats prometteurs pour résoudre le problème du nombre de données limité dans le domaine médical, ce qui nous a encouragé à proposer une approche générique adaptée pour toutes les modalités d'imagerie médicale en se basant sur l'approche de meta-learning et en incorporant un mécanisme d'attention.

4.3 Approches et reformulations

4.3.1 Réseau de neurones attentif

L'objectif de notre travail est d'utiliser le méta-apprentissage pour résoudre le problème de la classification de quelques images médicales. L'attention est incorporée pour améliorer l'efficacité de notre méta-apprentissage. Nous cherchons à trouver l'initialisation optimale des paramètres à partir d'une distribution de tâches similaires, en déplaçant les poids initiaux du modèle vers une position optimale, en utilisant l'approche MAML pour utiliser les

connaissances préalables et une branche d'attention pour apprendre à quelle région le réseau doit faire attention d'une tâche à l'autre. Notre modèle pré-entrainé nécessite alors qu'un simple ajustement des paramètres lors d'une seconde phase d'apprentissage sur la nouvelle tâche. Peu, voir qu'une seule itération de descente de gradient, suffit pour que le modèle soit performant et atteigne la convergence. Techniquement, notre approche maximise la sensibilité des fonctions de perte des nouvelles tâches par rapport aux paramètres et incorpore une branche d'attention.

On formule la classification d'images de maladies rares comme un problème de few-shot learning. Nous visons à entraîner un réseau de neurones en utilisant des données de méta-entraînement (maladies communes), de telle sorte qu'en présence de nouvelles tâches associées à quelques échantillons de données du méta-test (les maladies rares), le modèle s'adapte rapidement avec quelques itérations de descente de gradient pour traiter les nouvelles tâches.

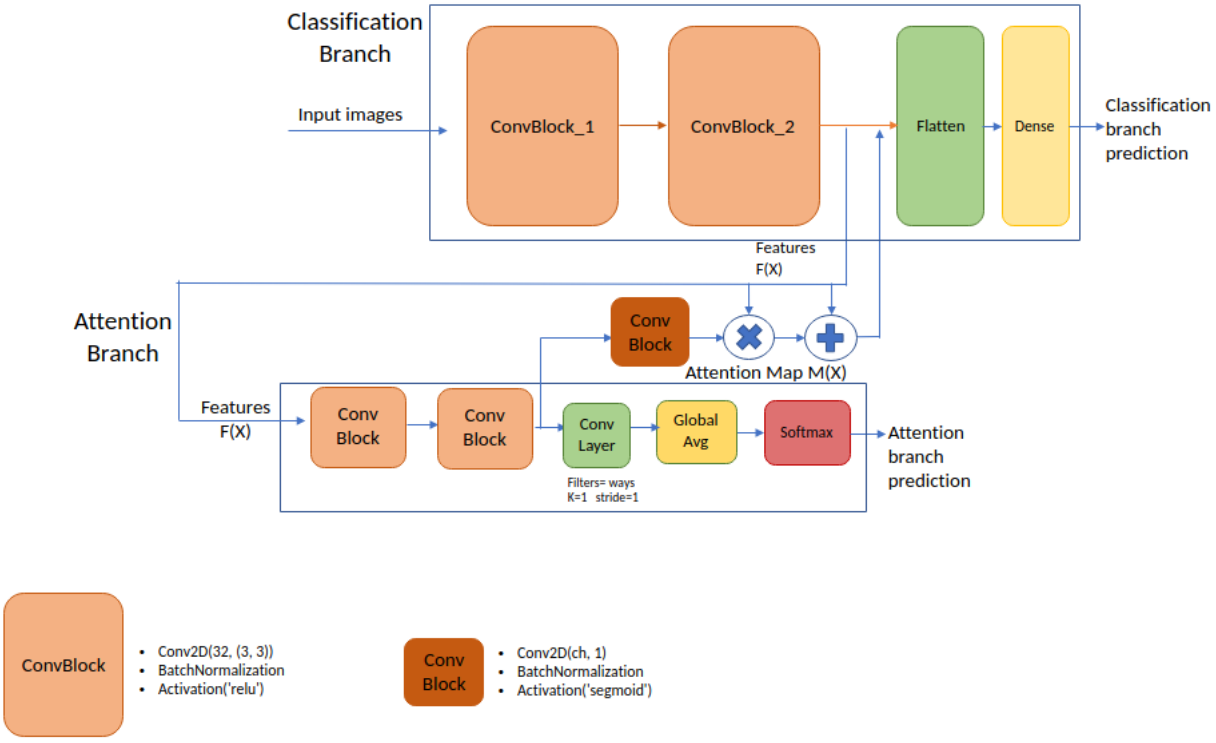


FIGURE 4.1 – L'architecture du réseau de neurones attentif proposé

L'architecture du réseau proposé de notre apprenant, illustrée dans la figure 4.1, est basée sur trois branches : la branche d'extraction de caractéristiques, la branche de classification et la branche d'attention. Les images ISIC 2018, échantillonnées à 90×120, ont servi à l'apprentissage de la branche d'extraction de caractéristiques [256].

La branche d'extraction de caractéristiques est constituée de 4 blocs de convolution comme l'architecture de base du classificateur utilisé par Vinyals et al. (2016). Chacun de ces blocs consiste en une convolution (3×3) de 32 filtres, suivie d'une normalisation par lot, d'une fonction ReLU et d'un max-pooling (2×2). La sortie des quatre blocs de convolution est envoyée à notre branche d'attention. Elle se compose de quatre blocs de convolution. Chaque bloc de convolution de la branche d'attention consiste en une convolution de (3×3). Les deux premiers et deuxièmes blocs contiennent respectivement 32 et 16 filtres, suivis d'une normalisation par lots et d'une ReLU. Ces quatre blocs sont suivis d'une couche global average pooling aidant à l'extraction du CAM. La sortie de la branche d'attention est multipliée par les caractéristiques de sortie de la branche de classification aplaties et avant d'être envoyée dans un softmax suivi d'une couche dense.

Comme nous pouvons le voir, les caractéristiques profondes extraites de la branche de classification servent d'entrée à la branche d'attention. Une fois convoluées par la branche d'attention, les caractéristiques sont renvoyées à la branche de classification.

Soit $F(X)$ la sortie de notre bloc d'extraction de caractéristiques et soit $M(X)$ la carte d'attention produite par la branche d'attention. Le mécanisme d'attention résultant C(X) est expliqué dans l'équation suivante :

$$C(X) = F(X) * M(X) + F(X) \tag{4.1}$$

L'équation 4.1 peut également être écrite comme suit :

$$C(X) = F(X) * (M(X) + 1) \tag{4.2}$$

Le résultat $C(X)$ alimente la branche de classification composée d'un bloc de base suivi d'une couche global average pooling, de deux couches denses de respectivement 256 et du nombre de classes, et d'une couche d'activation softmax.

La fonction d'erreur des branches d'attention et de classification est la fonction d'entropie croisée calculant l'erreur entre la classe prédite et la classe réelle.

Soit f_θ notre méta-apprenant f paramétré par θ initialisé aléatoirement, $p(t)$ la distribution des tâches et J_θ la fonction de perte. Les θ sont constitués de θ_C et θ_A respectivement les paramètres de la branche du classificateur et les paramètres de la branche d'attention. Soit T un lot de N tâches tel que $T_i \sim p(T)$. Pour chaque tâche T_i, nous entraînons le modèle en utilisant k échantillons de n classes et calculons la fonction de perte.

$$J_{T_i}(f_{\{\theta_C,\theta_A\}_i}) = -\sum_{x_j,y_j\sim T_i} y_i log(f_{\{\theta_C,\theta_A\}_i}(x_j) + (1-y_j)log(1-f_{\{\theta_C,\theta_A\}_i}(x_j))$$

$$\{\theta'_C,\theta'_A\} \leftarrow \{\theta_C,\theta_A\} - \alpha\nabla_{\{\theta_C,\theta_A\}} J_{T_i}(f_{\{\theta_C,\theta_A\}})$$

où

$$\begin{cases} \theta_C & \text{Paramètres initiaux de la branche de classification} \\ \theta_A & \text{Paramètres initiaux de la branche d'attention} \\ \theta' & \text{Paramètres optimaux de la tâche } T_i \\ \alpha & \text{hyperparamèteres} \\ \nabla_{\{\theta_C,\theta_A\}} J_{T_i}(f_{\{\theta_C,\theta_A\}}) & \text{gradient de la tâche } T_i \end{cases}$$

Par conséquent, on obtient N paramètres optimaux θ'_i. Avant de passer au lot de tâches suivant, une meta mise à jour est effectuée sur les paramètres initiaux afin de faire passer le modèle à la position optimale.

$$J'_{\{\theta_C,\theta_A\}} = \sum_{T_i\sim p(T)} J_{T_i}(f_{\{\theta'_C,\theta'_A\}})$$

$$\{\theta'_C,\theta'_A\} \leftarrow \{\theta'_C,\theta'_A\} - \beta\nabla_\theta J'_{\{\theta_C,\theta_A\}}$$

où

$$\begin{cases} \theta_C & \text{Paramètres initiaux de la branche de classification} \\ \theta_A & \text{Paramètres initiaux de la branche d'attention} \\ \beta & \text{hyperparamèteres} \\ \nabla_\theta J'_\theta & \text{gradient de chaque tâche } T_i \text{ par rapport à } \theta'_i \end{cases}$$

Cette étape de meta mise à jour représente le méta-entraînement. On met principalement à jour θ en prenant la moyenne des gradients de chaque tâche T_i par rapport au paramètre θ'_i.

On formule la tâche de classification d'images de maladies rares comme un problème d'apprentissage à N voies, K-shots few-shot learning. Le jeu de données est divisé en deux parties : $D_{Meta-train}$ et $D_{Meta-test}$. Notre objectif est de meta-entraîner le réseau de neurones en utilisant le jeu de données de méta-entraînement $D_{Meta-train}$ contenant les classes de maladies ayant les plus grands jeux de données disponibles (fig.4.2) de sorte que, étant donné les nouvelles tâches dans la phase de meta-test utilisant le jeu de données de meta-test $D_{Meta-test}$ pour les maladies les plus rares, le modèle peut rapidement adapter le modèle de réseau via quelques itérations de descente de gradient.

Le pipeline de notre approche incluant la branche d'attention est représenté dans l'algorithme suivant.

4.3.2 Réseau de neurones attentif amélioré

Cependant, l'algorithme ci-dessus incluant le mécanisme de la branche d'attention n'est pas façon optimale d'exploitation des connaissances préalables. En fait, d'une étape à l'autre de notre algorithme de meta-apprentissage, on met à jour les poids du mécanisme d'attention sans tenir compte de la fonction d'erreur de la branche d'attention. L'idée est donc de pénaliser la mise à jour en utilisant une combinaison de l'erreur de la prédiction de la branche de classification avec l'erreur de la branche d'attention. La modification est incluse dans l'étape de la meta mise à jour effectuée sur les paramètres initiaux pour le déplacer vers la position optimale.

$$J'_{\{\theta_C,\theta_A\}} = \sum_{T_i \sim p(T)} (1 - log(J_{T_{i_{att}}}))J_{T_i}(f_{\{\theta'_C,\theta'_A\}}) \tag{4.3}$$

Algorithm 5.1 Paying-attention MAML algorithm

Require: p(T): distribution over tasks of the $D_{Meta-train}$ Dataset
Require: α & β Learner and meta-learning learning rates
randomly initialize $\{\theta'_C, \theta'_A\}$
while number of meta-epochs is not done **do**
  Sample batch of tasks $T_i \sim p(T)$ of the $D_{Meta-train}$ Dataset
  for all T_i in the batch **do**
    Sample k data points $D = (x(j), y(j))$ of the n classes from T_i sampled dataset
    while number of learner-updates is not done **do**
      Compute $\nabla_{\{\theta_C,\theta_A\}} J_{T_i}(f_{\{\theta_C,\theta_A\}})$ using $J_{T_i}(f_{\{\theta_C,\theta_A\}})$ in eq (4)
      Compute gradient descent using eq (5)
    end while
  end for
  Compute the meta-learner loss function $J'_{\{\theta_C,\theta_A\}}$ using eq (6)
  Update the meta-learner parameters $\{\theta_C, \theta_A\}$ using the eq (7)
end while

Une étude expérimentale des différentes fonctions a été incorporée pour choisir la meilleure représentation du mécanisme d'attention dans notre algorithme. L'amélioration suivante consisterait à construire l'architecture pour la tâche suivante de classification des mêmes classes, en tenant compte des performances de la branche d'attention sur la tâche précédente.

$$C(X) = F(X) * ((1 - log(L_{att}) * M(X) + 1) \tag{4.4}$$

4.4 Expérimentations et résulats

4.4.1 Bases de données

Afin d'évaluer les performances de la méthode proposée, nous avons sélectionné deux jeux de données médicales :

4.4.1.1 Base de données ISIC 2018

. On utilise le jeu de données ISIC 2018 the International Skin Imaging Collaboration dataset for Skin Lesion Analysis Towards Melanoma Detection Dataset [163](fig. 4.2). ISIC représente la plus grande collection publiquement disponible d'images dermoscopiques de lésions cutanées de qualité. Nous utilisons des images de lésions cutanées de sept maladies de la peau, dont le nævus mélanocytaire (6705), le mélanome (1113), la kératose bénigne (1099), le carcinome basocellulaire (514), la kératose actinique (327), la lésion vasculaire (142) et le dermatofibrome (115). Pour reproduire l'environnement clinique, nous considèrons les quatre classes avec le plus grand nombre de cas comme des maladies courantes (base de meta-entraînement) et les trois classes restantes comme des maladies rares (base de meta-test).

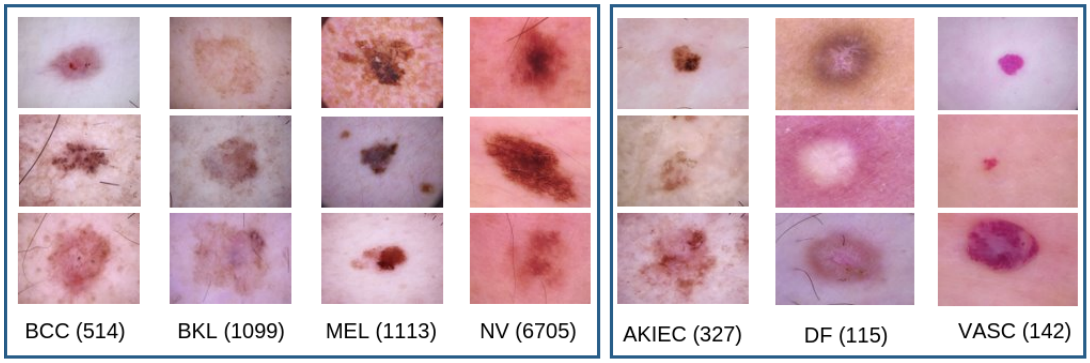


FIGURE 4.2 – Base de données ISIC 2018

4.4.1.2 Base de données des radiographies du thorax

Notre base de radiographie du thorax est un ensemble de données publiques ouvertes de 112 120 radiographies du thorax en vue frontale de 30 805 patients uniques avec 14 pathologies thoraciques étiquetées, chaque image pouvant avoir plusieurs étiquettes, notamment Atelectasis (11535), Consolidation (2772), Infiltration (13307), Pneumothorax (19871), Œdème (5746), Emphysème (6323), Fibrose (1353), Effusion (5298), Pneumonie (4667), Épaississement pleural (2303), Cardiomégalie (2516), Nodule (1686), Masse (3385) et Hernie (227) (fig.4.3). Nous avons élaboré un jeu de données contenant uniquement les images à étiquette unique d'un total de 30 974, dont Atelectasis (4212), Consolidation (1094), Infiltration (3959), Pneumothorax (9552), Œdème (2138), Emphysème (2706), Fibrose (307), Effusion (2199),

Pneumonie (1314), Épaississement pleural (634), Cardiomégalie (895), Nodule (727), Masse (1127) et Hernie (110). Nous simulons le problème des maladies rares à un problème de few-shot learning en utilisant les classes avec le plus grand nombre de cas comme maladies communes (base de meta-entraînement) et les trois classes restantes comme maladies rares (base de meta-test).

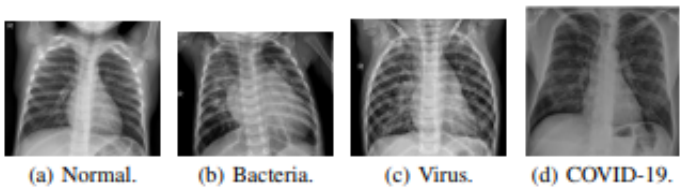


FIGURE 4.3 – Quelques échantillons des différentes classes des scanners thoraciques [269]

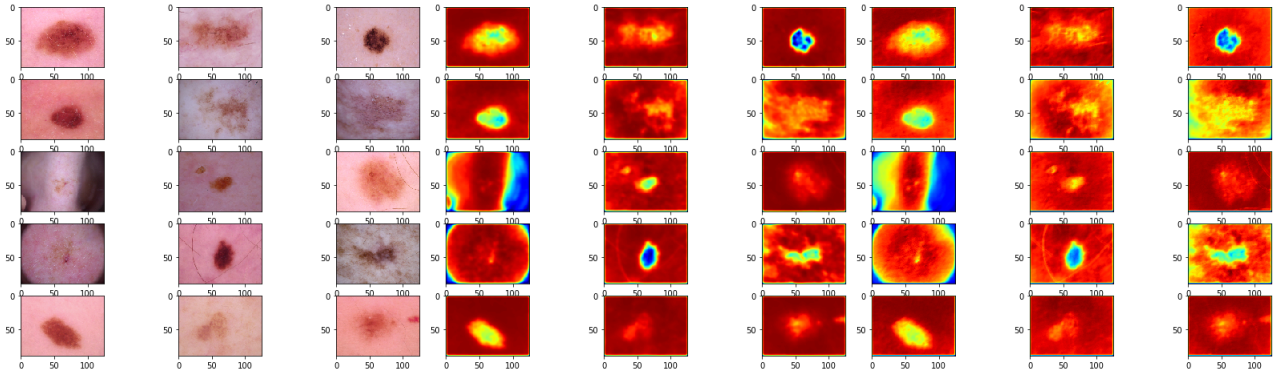
4.4.2 Expérimentations

Notre modèle d'apprentissage profond est basé sur un réseau neuronal profond (DNN) composé de plusieurs couches consécutives de neurones. Afin de trouver le meilleur modèle pour la prédiction de l'enthalpie de formation, on effectue une recherche approfondie de la meilleure architecture de modèle DNN ainsi que dans l'espace des hyperparamètres. Nous avons effectué une recherche systématique dans un large espace d'architecture de réseau neuronal, en commençant par une architecture à deux couches et en augmentant progressivement la profondeur pour améliorer la capacité d'apprentissage de notre modèle jusqu'à atteindre un point de saturation. Nous avons exploré différentes combinaisons du nombre d'unités de neurones par couche. Une couche d'exclusion a été ajoutée à chaque fois que le nombre de neurones entre des couches consécutives changeait pour éviter un sur-ajustement. L'erreur de test a commencé à osciller dans de petites limites au-delà de l'architecture à 17 couches. La recherche d'architecture a été poursuivie jusqu'à un modèle DNN à 24 couches où l'erreur de test est restée la même que pour le réseau à 17 couches. Nous pensons que le modèle d'apprentissage profond a déjà appris les caractéristiques nécessaires qu'il pouvait trouver dans l'ensemble de données d'entraînement à ce stade, car l'augmentation de la profondeur n'a pas amélioré davantage les performances du modèle. La meilleure architecture de modèle

est présentée dans le tableau 3. Nous avons également expérimenté différents types de fonctions d'activation, et nous avons constaté que ReLU était le plus performant.

4.4.3 Résultats

La visualisation des cartes d'attention nous a confirmé la pertinence de notre proposition (fig. 4.4). Un meilleur seuillage adapte selon le cas de l'utilisation et en prenant en considération l'avis des spécialistes dans le domaine médical permettra d'obtenir des masques plus raffinés et donc des résultats, de classification et dans un autre niveau de segmentation, plus pertinents.



(a) Images d'origine **(b) Cartes d'attention** **(c) Masques d'attention**

FIGURE 4.4 – Des échantillons des cartes d'attention et des masques d'attention résultants de la base de données ISIC 2018

Les résultats ont approuvé l'efficacité de notre pipeline. En combinant le MAML avec l'attention et en n'utilisant que 5 images par classes pour l'ajustement, on arrive à obtenir une précision de classification de 65% pour la base de données de radiographie thoracique et de 68% pour la base de données ISIC2018. Notre réseau de neurones attentif a encore amélioré nos résultats en atteignant respectivement une précision de classification de 69% pour la radiographie thoracique et une amélioration plus importante de la précision de classification pour ISIC2018 s'élevant de 76%.

	Radiographie Thoracique	**ISIC 2018**
MAML	53%	62%
MAML+Attention	65%	68%
MAML+ Attention améliorée	69%	76%

TABLE 4.1 – Accuracy de meta-test sur les deux différentes bases de données

4.4.4 Conclusion

Dans ce chapitre, nous avons proposé une méthode appelée MAML attentif, basée sur le meta-apprentissage et le mécanisme d'attention, permettant de déployer l'apprentissage profond avec peu de données. Les résultats obtenus sont moyennement satisfaisants dans un secteur critique qui nécessite beaucoup de précision. Cependant, ce chapitre n'est que l'initialisation d'une perspective très importante dans le domaine de l'imagerie médicale. Les hyper-paramètres testés sont choisis arbitrairement et souvent par tâtonnement. Un autre paramétrage, notamment le nombre de données par classe et l'architecture de base, permettrait d'améliorer les résultats obtenus. Notre solution peut être utilisée en combinaison avec notre pipeline proposé dans le chapitre précédent pour souligner les régions d'intérêt pour notre anonymisation et afin d'utiliser moins de données.

CONCLUSION GÉNÉRALE

L'imagerie médicale est une composante primordiale du domaine de la santé sur laquelle se base le diagnostic et le traitement de certaines maladies. Elle est essentielle pour suivre et évaluer l'évolution d'une maladie particulière, en vue de sa guérison éventuelle ou également afin de guider les médecins lors des interventions chirurgicales. L'imagerie médicale est un secteur qui couvre un large éventail d'applications, principalement la classification et la segmentation. Le principe général de l'analyse d'images médicales est de mesurer et d'analyser les caractéristiques les plus distinctives de l'image. L'analyse automatique des images est l'une des applications les plus connues de l'apprentissage profond.

En effet, l'intelligence artificielle est sans doute l'une des étapes charnières dans l'histoire de l'humanité. La quantité de données disponibles, l'avancement des algorithmes et l'amélioration des capacités du matériel informatique, donnent à l'apprentissage profond, en particulier une marge de développement importante. Le déploiement de l'apprentissage profond dans le secteur de santé favorise des solutions plus adaptées et plus préventives, apportant une amélioration au diagnostic, au suivi et traitement médical des patients et une assistance intéressante aux professionnels de santé.

Dans ce contexte, cette thèse s'est intéressée à deux défis principaux liés à l'utilisation de l'apprentissage profond dans l'imagerie médicale : la confidentialité et la disponibilité de données.

Dans le troisième chapitre, nous avons étudié le problème lié à la confidentialité : la pseudo-anonymisation. Nous avons réussi à classer des IRM en premier lieu selon la marque de l'équipement d'acquisition et en second lieu selon l'identité des patients. Nous avons proposé également un pipeline complet qui traite cette problématique en générant des nouvelles images.

Les IRM résultantes permettent toujours d'effectuer les analyses et le diagnostic nécessaires sans détenir les détails permettant l'identification du fabricateur de l'équipement d'acquisition de l'IRM ou du patient.

Par ailleurs, nous nous sommes penchés sur le meta-learning qui est une approche tendance de l'apprentissage profond. Les techniques du meta-learning ont pour objectif de résoudre la problématique de few-shot learning. Le few-shot learning désigne un contexte d'application dans lequel la quantité de données disponible pour l'apprentissage est très limitée, notamment dans le cadre des maladies rares. Ainsi, l'introduction de meta-apprentissage dans le secteur médical peut radicalement changer le cours de l'aide au diagnostic médical. Le meta-apprentissage et ses variantes présentent plusieurs avantages par rapport aux méthodes de l'apprentissage profond conventionnelles. D'une part, ils résolvent la problématique de few-shot learning. D'autre part, ils favorisent un apprentissage optimal et un ajustement rapide d'une tâche à une autre.

Une revue de l'état de l'art des méthodes de meta-apprentissage et des techniques dans le deuxième chapitre nous a permis d'utiliser les techniques du meta-apprentissage adaptées pour profiter de l'apprentissage profond dans plusieurs applications. Nous avons aussi souligné la notion du mécanisme d'attention utilise dans des multiples travaux. Nous avons eu recours aux réseaux siamois pour faire face au problème du manque de données lors de la classification des patients dans le troisième chapitre. Les méthodes de classification en vue d'aide au diagnostic médical dans le quatrième chapitre sont également basées sur des techniques de meta-apprentissage. Ces dernières, en combinaison avec le mécanisme d'attention, ont prouvé la validité d'une perspective importante de l'utilisation de l'apprentissage profond dans le secteur de l'imagerie médicale avec peu de données.

Dans cette thèse, nous avons fait face à plusieurs défis scientifiques et techniques spécifiques au domaine de l'imagerie médicale, notamment la qualité et la quantité de données, et l'arbitrage entre les différentes techniques de prétraitement, d'apprentissage et de meta-apprentissage. Toutes les solutions proposées font intervenir une combinaison de techniques classiques de traitement d'image et de méthodes d'apprentissage profond.

L'une des perspectives les plus importantes de ce travail est l'introduction du mécanisme d'attention dans le processus d'anonymisation pour comparer les régions présentant un risque de confidentialité aux régions nécessaires pour le diagnostic. Ensuite, nous espérons proposer un outil complet incorporant la combinaison des solutions des deux chapitres permettant l'anonymisation et le diagnostic adapté. Cet outil pourrait être utilisé automatiquement par les institutions médicales sur les données brutes acquises, permettant simultanément une adaptation rapide aux nouvelles maladies et ses spécificités, et une véritable anonymisation.

L'apprentissage profond joue ainsi un rôle fondamental dans le domaine de l'imagerie médicale en fournissant aux professionnels de la santé des informations qui leur permettent d'identifier les problèmes à un stade précoce, offrant ainsi des soins aux patients beaucoup plus personnalisés et pertinents. L'avenir des soins de santé n'a jamais été aussi passionnant. Non seulement l'IA offre la possibilité de développer des solutions qui répondent à des besoins très spécifiques du secteur, mais l'apprentissage profond dans le domaine des soins de santé peut devenir incroyablement puissant pour soutenir les cliniciens et transformer radicalement les soins aux patients.

BIBLIOGRAPHIE

[1] Overfitting and methods of addressing it, 2021.

[2] Normalization, 2022.

[3] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3) :175–185, 1992.

[4] Yaron Anavi, Ilya Kogan, Elad Gelbart, Ofer Geva, and Hayit Greenspan. Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval. In *Medical Imaging 2016 : Computer-Aided Diagnosis*, volume 9785, page 978510. International Society for Optics and Photonics, 2016.

[5] Chief Scientist at Baidu Andrew Ng. Extract data conference.

[6] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in neural information processing systems*, pages 3981–3989, 2016.

[7] Antreas Antoniou, Harrison Edwards, and Amos Storkey. How to train your maml. *arXiv preprint arXiv :1810.09502*, 2018.

[8] Joseph Antony, Kevin McGuinness, Noel E O'Connor, and Kieran Moran. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 1195–1200. IEEE, 2016.

[9] Mikel Artetxe and Holger Schwenk. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7 :597–610, 2019.

[10] David Baehrens, Timon Schroeter, Stefan Harmeling, Motoaki Kawanabe, Katja Hansen, and Klaus-Robert MÃžller. How to explain individual classification decisions. *Journal of Machine Learning Research*, 11(Jun) :1803–1831, 2010.

[11] Yaniv Bar, Idit Diamant, Lior Wolf, Sivan Lieberman, Eli Konen, and Hayit Greenspan. Chest pathology detection using deep learning with non-medical training. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 294–297. IEEE, 2015.

[12] Bruno Barbosa, António JR Neves, Sandra C Soares, and Isabel D Dimas. Analysis of emotions from body postures based on digital imaging. *SIGNAL 2018 Editors*, page 81, 2018.

[13] Mariana Belgiu and Lucian Drăguţ. Random forest in remote sensing : A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114 :24–31, 2016.

[14] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In *Neural networks : Tricks of the trade*, pages 437–478. Springer, 2012.

[15] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning : A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8) :1798–1828, 2013.

[16] Yoshua Bengio, Patrice Simard, Paolo Frasconi, et al. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2) :157–166, 1994.

[17] Etienne Bennequin. Few-shot image classification with meta-learning, 2020.

[18] Jose Bernal, Kaisar Kushibar, Daniel S Asfaw, Sergi Valverde, Arnau Oliver, Robert Martí, and Xavier Lladó. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging : a review. *Artificial intelligence in medicine*, 95 :64–81, 2019.

[19] Daniel Bernau, Philip-William Grassal, Jonas Robl, and Florian Kerschbaum. Assessing differentially private deep learning with membership inference. *arXiv preprint arXiv :1912.11328*, 2019.

[20] Michael Bernico, Yuntao Li, and Dingchao Zhang. Investigating the impact of data volume and domain similarity on transfer learning applications. In *Proceedings of the Future Technologies Conference*, pages 53–62. Springer, 2018.

[21] Chris M Bishop. Training with noise is equivalent to tikhonov regularization. *Neural computation*, 7(1) :108–116, 1995.

[22] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.

[23] Aline FS Borges, Fernando JB Laurindo, Mauro M Spínola, Rodrigo F Gonçalves, and Claudia A Mattos. The strategic use of artificial intelligence in the digital era : Systematic literature review and future research directions. *International Journal of Information Management*, 57 :102225, 2021.

[24] Maged N Kamel Boulos, Ann C Brewer, Chante Karimkhani, David B Buller, and Robert P Dellavalle. Mobile medical and health apps : state of the art, concerns, regulatory control and certification. *Online journal of public health informatics*, 5(3) :229, 2014.

[25] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Routledge, 2017.

[26] Tom Brosch, Roger Tam, Alzheimer's Disease Neuroimaging Initiative, et al. Manifold learning of brain mris by deep learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 633–640. Springer, 2013.

[27] Domo business cloud. Data never sleeps, 2022.

[28] Sandie Cabon, Fabienne Porée, Antoine Simon, Olivier Rosec, Patrick Pladys, and Guy Carrault. Video and audio processing in paediatrics : A review. *Physiological measurement*, 40(2) :02TR02, 2019.

[29] Leslie Casas, Gustavo Carneiro, Nassir Navab, and Vasileios Belagiannis. Few-shot meta-denoising. *arXiv preprint arXiv :1908.00111*, 2019.

[30] Souleyman Chaib, Huan Liu, Yanfeng Gu, and Hongxun Yao. Deep feature fusion for vhr remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8) :4775–4784, 2017.

[31] Clément Côme Charles Tremblay. Accuracy : définition, calcul et limites, 2021.

[32] Clément Côme Charles Tremblay. La balanced accuracy weighted, pour aller plus loin que l'accuracy, 2021.

[33] Christophe Charrier, Laurence T Maloney, Hocine Cherifi, and Kenneth Knoblauch. Maximum likelihood difference scaling of image quality in compression-degraded images. *JOSA A*, 24(11) :3418–3426, 2007.

[34] Aditya Chattopadhyay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++ : Improved visual explanations for deep convolutional networks. *arXiv preprint arXiv :1710.11063*, 2017.

[35] Nagesh Singh Chauhan. Loss functions in neural networks, 2021.

[36] Kan Chen, Trung Bui, Chen Fang, Zhaowen Wang, and Ram Nevatia. Amc : Attention guided multi-modal correlation learning for image search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2644–2652, 2017.

[37] Min Chen, Xiaobo Shi, Yin Zhang, Di Wu, and Mohsen Guizani. Deep features learning for medical image analysis with convolutional autoencoder neural network. *IEEE Transactions on Big Data*, 2017.

[38] Gong Cheng, Ceyuan Yang, Xiwen Yao, Lei Guo, and Junwei Han. When deep learning meets metric learning : Remote sensing image scene classification via learning discriminative cnns. *IEEE transactions on geoscience and remote sensing*, 56(5) :2811–2821, 2018.

[39] François Chollet. Xception : Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[40] Evangelia Christodoulou, Jie Ma, Gary S Collins, Ewout W Steyerberg, Jan Y Verbakel, and Ben Van Calster. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, 110 :12–22, 2019.

[41] Yu-An Chung and Wei-Hung Weng. Learning deep representations of medical images using siamese cnns with application to content-based image retrieval. *arXiv preprint arXiv :1711.08490*, 2017.

[42] B Jack Copeland. The church-turing thesis. 1997.

[43] Daniel Crevier. *AI : the tumultuous history of the search for artificial intelligence*. Basic Books, Inc., 1993.

[44] cs231n. Cs231n : Convolutional neural networks for visual recognition.

[45] Guy Davidson and Michael C Mozer. Sequential mastery of multiple tasks : Networks naturally learn to learn. *arXiv preprint arXiv :1905.10837*, 2019.

[46] Martin Davis. Mathematical logic and the origin of modern computers. In *Studies in the History of Mathematics*, pages 137–167. 1987.

[47] Nicola Davis. Interview : Cardiologist eric topol : 'ai can restore the care in healthcare', 2019.

[48] Min-Yuh Day. Introduction to artificial intelligence for text analytics. 2022.

[49] Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. Deep direct reinforcement learning for financial signal representation and trading. *IEEE transactions on neural networks and learning systems*, 28(3) :653–664, 2016.

[50] Boubacar Diallo, Thierry Urruty, Pascal Bourdon, and Christine Fernandez-Maloigne. Robust forgery detection for compressed images using cnn supervision. *Forensic Science International : Reports*, 2 :100112, 2020.

[51] Stephanie Dick. Artificial intelligence. 2019.

[52] Georgiana Dinu, Angeliki Lazaridou, and Marco Baroni. Improving zero-shot learning by mitigating the hubness problem. *arXiv preprint arXiv :1412.6568*, 2014.

[53] Jose Dolz, Christian Desrosiers, Li Wang, Jing Yuan, Dinggang Shen, and Ismail Ben Ayed. Deep cnn ensembles and suggestive annotations for infant brain mri segmentation. *Computerized Medical Imaging and Graphics*, page 101660, 2019.

[54] OS Eluyode and Dipo Theophilus Akomolafe. Comparative study of biological and artificial neural networks. *European Journal of Applied Engineering and Scientific Research*, 2(1) :36–46, 2013.

[55] Andre Esteva, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher. Deep learning-enabled medical computer vision. *NPJ digital medicine*, 4(1) :1–9, 2021.

[56] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639) :115, 2017.

[57] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4) :594–611, 2006.

[58] Christoph Feichtenhofer, Axel Pinz, and Andrew Zisserman. Convolutional two-stream network fusion for video action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1933–1941, 2016.

[59] Alexandre Fenneteau, Pascal Bourdon, David Helbert, Christine Fernandez-Maloigne, Christophe Habas, and Remy Guillevin. Learning a cnn on multiple sclerosis lesion segmentation with self-supervision. In *3D Measurement and Data Processing, IS&T Electronic Imaging 2020 Symposium*, 2020.

[60] Samuel G Finlayson, John D Bowers, Joichi Ito, Jonathan L Zittrain, Andrew L Beam, and Isaac S Kohane. Adversarial attacks on medical machine learning. *Science*, 363(6433) :1287–1289, 2019.

[61] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1126–1135. JMLR. org, 2017.

[62] Chelsea Finn and Sergey Levine. Meta-learning and universality : Deep representations and gradient descent can approximate any learning algorithm. *arXiv preprint arXiv :1710.11622*, 2017.

[63] Chelsea Finn, Kelvin Xu, and Sergey Levine. Probabilistic model-agnostic meta-learning. In *Advances in Neural Information Processing Systems*, pages 9516–9527, 2018.

[64] Chelsea Finn, Tianhe Yu, Tianhao Zhang, Pieter Abbeel, and Sergey Levine. One-shot visual imitation learning via meta-learning. *arXiv preprint arXiv :1709.04905*, 2017.

[65] Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimilano Pontil. Bilevel programming for hyperparameter optimization and meta-learning. *arXiv preprint arXiv :1806.04910*, 2018.

[66] Robert M French. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4) :128–135, 1999.

[67] Hiroshi Fukui, Tsubasa Hirakawa, Takayoshi Yamashita, and Hironobu Fujiyoshi. Attention branch network : Learning of attention mechanism for visual explanation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10705–10714, 2019.

[68] Kunihiko Fukushima. Neocognitron : A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4) :193–202, 1980.

[69] Chuang Gan, Naiyan Wang, Yi Yang, Dit-Yan Yeung, and Alex G Hauptmann. Devnet : A deep event network for multimedia event detection and evidence recounting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2568–2577, 2015.

[70] Maria Teresa Gaudio, Gerardo Coppola, Lorenzo Zangari, Stefano Curcio, Sergio Greco, and Sudip Chakraborty. Artificial intelligence-based optimization of industrial membrane processes. *Earth Systems and Environment*, 5(2) :385–398, 2021.

[71] Benyamin Ghojogh and Mark Crowley. The theory behind overfitting, cross validation, regularization, bagging, and boosting : tutorial. *arXiv preprint arXiv :1905.12787*, 2019.

[72] Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4367–4375, 2018.

[73] Christopher E Gillies, Daniel F Taylor, Brandon C Cummings, Sardar Ansari, Fadi Islim, Steven L Kronick, Richard P Medlin Jr, and Kevin R Ward. Demonstrating the consequences of learning missingness patterns in early warning systems for preventative health care : A novel simulation and solution. *Journal of Biomedical Informatics*, 110 :103528, 2020.

[74] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.

[75] Herman H Goldstine and Adele Goldstine. The electronic numerical integrator and computer (eniac). *Mathematical Tables and Other Aids to Computation*, 2(15) :97–110, 1946.

[76] Rafael C Gonzalez. *Digital image processing*. Pearson education india, 2009.

[77] Santiago Gonzalez and Risto Miikkulainen. Improved training speed, accuracy, and data utilization through loss function optimization. *arXiv preprint arXiv :1905.11528*, 2019.

[78] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning (adaptive computation and machine learning series). *Cambridge Massachusetts*, pages 321–359, 2017.

[79] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[80] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. iclr'15. *arXiv preprint arXiv :1412.6572*, 2015.

[81] Hayit Greenspan, Bram Van Ginneken, and Ronald M Summers. Guest editorial deep learning in medical imaging : Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5) :1153–1159, 2016.

[82] Sorin Grigorescu, Bogdan Trasnea, Tiberiu Cocias, and Gigel Macesanu. A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, 37(3) :362–386, 2020.

[83] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22) :2402–2410, 2016.

[84] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. Xai—explainable artificial intelligence. *Science Robotics*, 4(37) :eaay7120, 2019.

[85] Ghouthi Boukli Hacene. Processing and learning deep neural networks on chip. machine learning. *Ecole nationale supérieure Mines-Télécom Atlantique*, 2019.

[86] Alon Halevy, Peter Norvig, and Fernando Pereira. The unreasonable effectiveness of data. *IEEE intelligent systems*, 24(2) :8–12, 2009.

[87] Pavel Hamet and Johanne Tremblay. Artificial intelligence in medicine. *Metabolism*, 69 :S36–S40, 2017.

[88] Othman A Hanshal, Osman N Ucan, and Yousef K Sanjalawe. Hybrid deep learning model for automatic fake news detection. *Applied Nanoscience*, pages 1–11, 2022.

[89] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1) :1–12, 2004.

[90] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers : Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[91] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[92] Nathan Hilliard, Lawrence Phillips, Scott Howland, Artëm Yankov, Courtney D Corley, and Nathan O Hodas. Few-shot learning with metric-agnostic conditional embeddings. *arXiv preprint arXiv :1802.04376*, 2018.

[93] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7) :1527–1554, 2006.

[94] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02) :107–116, 1998.

[95] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmidhuber, et al. Gradient flow in recurrent nets : the difficulty of learning long-term dependencies, 2001.

[96] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8) :1735–1780, 1997.

[97] Kyle Hsu, Sergey Levine, and Chelsea Finn. Unsupervised learning via meta-learning. *arXiv preprint arXiv :1810.02334*, 2018.

[98] Shi Hu, Jakub Tomczak, and Max Welling. Meta-learning for medical image classification. 2018.

[99] Sergey Ioffe and Christian Szegedy. Batch normalization : Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv :1502.03167*, 2015.

[100] Ali Işın, Cem Direkoğlu, and Melike Şah. Review of mri-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science*, 102 :317–324, 2016.

[101] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[102] Robert A Jacobs. Increased rates of convergence through learning rate adaptation. *Neural networks*, 1(4) :295–307, 1988.

[103] Hossein Jafari, Oluwaseyi Omotere, Damilola Adesina, Hsiang-Huang Wu, and Lijun Qian. Iot devices fingerprinting using deep learning. In *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*, pages 1–9. IEEE, 2018.

[104] Anil K Jain. *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.

[105] Yeon Uk Jeong, Soyoung Yoo, Young-Hak Kim, and Woo Hyun Shim. De-identification of facial features in magnetic resonance images : software development using deep learning technology. *Journal of medical Internet research*, 22(12) :e22739, 2020.

[106] Dipendra Jha, Logan Ward, Arindam Paul, Wei-keng Liao, Alok Choudhary, Chris Wolverton, and Ankit Agrawal. Elemnet : Deep learning the chemistry of materials from only elemental composition. *Scientific reports*, 8(1) :1–13, 2018.

[107] Jinmeng Jia, Ruiyuan Wang, Zhongxin An, Yongli Guo, Xin Ni, and Tieliu Shi. Rdad : A machine learning system to support phenotype-based rare disease diagnosis. *Frontiers in genetics*, 9 :587, 2018.

[108] Xiang Jiang, Mohammad Havaei, Gabriel Chartrand, Hassan Chouaib, Thomas Vincent, Andrew Jesson, Nicolas Chapados, and Stan Matwin. On the importance of attention in meta-learning for few-shot text classification. *arXiv preprint arXiv :1806.00852*, 2018.

[109] Worku Jifara, Feng Jiang, Seungmin Rho, Maowei Cheng, and Shaohui Liu. Medical image denoising using convolutional neural network : a residual learning approach. *The Journal of Supercomputing*, 75(2) :704–718, 2019.

[110] Thorsten Joachims. 11 making large-scale support vector machine learning practical. In *Advances in kernel methods : support vector learning*, page 169. MIT press, 1999.

[111] Michael I Jordan and Tom M Mitchell. Machine learning : Trends, perspectives, and prospects. *Science*, 349(6245) :255–260, 2015.

[112] James Jordon, Daniel Jarrett, Jinsung Yoon, Paul Elbers, Patrick Thoral, Ari Ercole, Cheng Zhang, Danielle Belgrave, and Mihaela van der Schaar. Hide-and-seek privacy challenge synthetic data generation vs. patient re-identification with clinical time-series data. 2020.

[113] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36 :61–78, 2017.

[114] Naushad Karim. Image processing and computer vision in java, 2014.

[115] Jagreet Kaur. Automatic log analysis using deep learning and ai, 2020.

[116] Hoel Kervadec, Jose Dolz, Meng Tang, Eric Granger, Yuri Boykov, and Ismail Ben Ayed. Constrained-cnn losses for weakly supervised segmentation. *Medical image analysis*, 54 :88–99, 2019.

[117] Renu Khandelwal. One-shot learning with siamese network, 2021.

[118] Mehdi Kharrazi, Husrev T Sencar, and Nasir Memon. Blind source camera identification. In *2004 International Conference on Image Processing, 2004. ICIP'04.*, volume 1, pages 709–712. IEEE, 2004.

[119] M Khoury and R Valdez. Rare diseases, genomics and public health : an expanding intersection. *Genomics and Health Impact Blog*, 2016.

[120] Edward Kim, Miguel Corte-Real, and Zubair Baloch. A deep semantic mobile application for thyroid cytopathology. In *Medical Imaging 2016 : PACS and Imaging Informatics : Next Generation and Innovations*, volume 9789, page 97890A. International Society for Optics and Photonics, 2016.

[121] Mijung Kim, Jasper Zuallaert, and Wesley De Neve. Few-shot learning using a small-sized dataset of high-resolution fundus images for glaucoma diagnosis. In *Proceedings of the 2nd international workshop on multimedia for personal health and health care*, pages 89–92, 2017.

[122] Taehoon Kim and Jihoon Yang. Latent-space-level image anonymization with adversarial protector networks. *IEEE Access*, 7 :84992–84999, 2019.

[123] Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. *Advances in neural information processing systems*, 27, 2014.

[124] Jesse Knight, Graham W Taylor, and April Khademi. Equivalence of histogram equalization, histogram matching and the nyul algorithm for intensity standardization in mri. *Journal of Computational Vision and Imaging Systems*, 3(1), 2017.

[125] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, page 0. Lille, 2015.

[126] Soheil Kolouri, Nicholas Ketz, Xinyun Zou, Jeffrey Krichmar, and Praveen Pilly. Attention-based structural-plasticity. *arXiv preprint arXiv :1903.06070*, 2019.

[127] Jaidip Kotak and Yuval Elovici. Iot device identification using deep learning. In *Conference on Complex, Intelligent, and Software Intensive Systems*, pages 76–86. Springer, 2020.

[128] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[129] Anders Krogh and John Hertz. A simple weight decay can improve generalization. *Advances in neural information processing systems*, 4, 1991.

[130] Aditya Kuppa, Lamine Aouad, and Nhien-An Le-Khac. Towards improving privacy of synthetic datasets. In *Annual Privacy Forum*, pages 106–119. Springer, 2021.

[131] Amal Lahiani, Jacob Gildenblat, Irina Klaman, Nassir Navab, and Eldad Klaiman. Generalizing multistain immunohistochemistry tissue segmentation using one-shot color deconvolution deep neural networks. *arXiv preprint arXiv :1805.06958*, 2018.

[132] Brenden Lake, Ruslan Salakhutdinov, Jason Gross, and Joshua Tenenbaum. One shot learning of simple visual concepts. In *Proceedings of the annual meeting of the cognitive science society*, volume 33, 2011.

[133] Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266) :1332–1338, 2015.

[134] Jean-Louis Laurière. *Intelligence artificielle : résolution de problèmes par l'homme et la machine*. Eyrolles Paris, 1987.

[135] D Cat laz. Neuron3, 2018. CC BY-SA 4.0.

[136] Walter Leal Filho, Tony Wall, Serafino Afonso Rui Mucova, Gustavo J Nagy, Abdul-Lateef Balogun, Johannes M Luetz, Artie W Ng, Marina Kovaleva, Fardous Mohammad Safiul Azam, Fátima Alves, et al. Deploying artificial intelligence for climate change adaptation. *Technological Forecasting and Social Change*, 180 :121662, 2022.

[137] Yann LeCun. The mnist database of handwritten digits. *http ://yann. lecun. com/exdb/mnist/*, 1998.

[138] Yann LeCun, Bernhard E Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne E Hubbard, and Lawrence D Jackel. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*, pages 396–404, 1990.

[139] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, 1998.

[140] June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo, and Namkug Kim. Deep learning in medical imaging : general overview. *Korean journal of radiology*, 18(4) :570–584, 2017.

[141] Yoonho Lee and Seungjin Choi. Gradient-based meta-learning with learned layerwise metric and subspace. *arXiv preprint arXiv :1801.05558*, 2018.

[142] Ang Li, Yixiao Duan, Huanrui Yang, Yiran Chen, and Jianlei Yang. Tiprdc : Task-independent privacy-respecting data crowdsourcing framework for deep learning with anonymized intermediate representations. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 824–832, 2020.

[143] Jun Li, Daoyu Lin, Yang Wang, Guangluan Xu, and Chibiao Ding. Deep discriminative representation learning with attention map for scene classification. *arXiv preprint arXiv :1902.07967*, 2019.

[144] Xiaomeng Li, Lequan Yu, Chi-Wing Fu, and Pheng-Ann Heng. Difficulty-aware meta-learning for rare disease diagnosis. *arXiv preprint arXiv :1907.00354*, 2019.

[145] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd : Learning to learn quickly for few-shot learning. *arXiv preprint arXiv :1707.09835*, 2017.

[146] Jason Liang, Elliot Meyerson, Babak Hodjat, Dan Fink, Karl Mutch, and Risto Miikkulainen. Evolutionary neural automl for deep learning. *arXiv preprint arXiv :1902.06827*, 2019.

[147] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv :1312.4400*, 2013.

[148] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42 :60–88, 2017.

[149] Xiaoxuan Liu, Livia Faes, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith Shamdas, Christoph Kern, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging : a systematic review and meta-analysis. *The lancet digital health*, 1(6) :e271–e297, 2019.

[150] Xinran Liu, Hamid R Tizhoosh, and Jonathan Kofman. Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 2872–2878. IEEE, 2016.

[151] Giovanni Livraga and Stefano Paraboschi. First report on privacy metrics and data sanitisation.

[152] S-CB Lo, S-LA Lou, Jyh-Shyan Lin, Matthew T Freedman, Minze V Chien, and Seong Ki Mun. Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Transactions on Medical Imaging*, 14(4) :711–718, 1995.

[153] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv :1711.05101*, 2017.

[154] Maria Lyra, Agapi Ploussi, Antonios Georgantzoglou, and PC Ionescu. Matlab as a tool in nuclear medicine image processing. *MATLAB-A Ubiquitous tool for the practical engineer*, pages 477–500, 2011.

[155] Gabriel Maicas, Andrew P Bradley, Jacinto C Nascimento, Ian Reid, and Gustavo Carneiro. Training medical image analysis systems like radiologists. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 546–554. Springer, 2018.

[156] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies*, 2019(4), 2019.

[157] John McCarthy. What is artificial intelligence ? 2007.

[158] John McCarthy, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, 27(4) :12–12, 2006.

[159] Scott McCartney. Eniac : The triumphs and tragedies of the world's first computer. 1999.

[160] Afonso Menegola, Michel Fornaciali, Ramon Pires, Sandra Avila, and Eduardo Valle. Towards automated melanoma screening : Exploring transfer learning schemes. *arXiv preprint arXiv :1609.01228*, 2016.

[161] Kevin Merchant, Shauna Revay, George Stantchev, and Bryan Nousain. Deep learning for rf device fingerprinting in cognitive communication networks. *IEEE Journal of Selected Topics in Signal Processing*, 12(1) :160–167, 2018.

[162] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net : Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.

[163] Md Ashraful Alam Milton. Automated skin lesion classification using ensemble of deep neural networks in isic 2018 : Skin lesion analysis towards melanoma detection challenge. *arXiv preprint arXiv :1901.10802*, 2019.

[164] Marvin Minsky. Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1) :8–30, 1961.

[165] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. A simple neural attentive meta-learner. *arXiv preprint arXiv :1707.03141*, 2017.

[166] Shubha Mishra, Piyush Shukla, and Ratish Agarwal. Analyzing machine learning enabled fake news detection techniques for diversified datasets. *Wireless Communications and Mobile Computing*, 2022, 2022.

[167] Tom M Mitchell, Svetlana V Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L Malave, Robert A Mason, and Marcel Adam Just. Predicting human brain activity associated with the meanings of nouns. *science*, 320(5880) :1191–1195, 2008.

[168] Volodymyr Mnih, Nicolas Heess, Alex Graves, et al. Recurrent models of visual attention. In *Advances in neural information processing systems*, pages 2204–2212, 2014.

[169] Pim Moeskops, Max A Viergever, Adriënne M Mendrik, Linda S de Vries, Manon JNL Benders, and Ivana Išgum. Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging*, 35(5) :1252–1261, 2016.

[170] Arnab Kumar Mondal, Jose Dolz, and Christian Desrosiers. Few-shot 3d multi-modal medical image segmentation using generative adversarial learning. *arXiv preprint arXiv :1810.12241*, 2018.

[171] Kevin R Moon, Alfred O Hero, and B Véronique Delouille. Meta learning of bounds on the bayes classifier error. In *2015 IEEE Signal Processing and Signal Processing Education Workshop (SP/SPE)*, pages 13–18. IEEE, 2015.

[172] Tsendsuren Munkhdalai, Alessandro Sordoni, Tong Wang, and Adam Trischler. Metalearned neural memory. *arXiv preprint arXiv :1907.09720*, 2019.

[173] Tsendsuren Munkhdalai and Hong Yu. Meta networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2554–2563. JMLR. org, 2017.

[174] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696, 2011.

[175] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv :1803.02999*, 2018.

[176] Alex Nichol and John Schulman. Reptile : a scalable metalearning algorithm. *arXiv preprint arXiv :1803.02999*, 2, 2018.

[177] Peter Norvig. Artificial intelligence : Early ambitions. *New Scientist*, 216(2889) :ii–iii, 2012.

[178] László G Nyúl and Jayaram K Udupa. On standardizing the mr image intensity scale. *Magnetic Resonance in Medicine : An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(6) :1072–1081, 1999.

[179] László G Nyúl, Jayaram K Udupa, and Xuan Zhang. New variants of a method of mri scale standardization. *IEEE transactions on medical imaging*, 19(2) :143–150, 2000.

[180] Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. Tadam : Task dependent adaptive metric for improved few-shot learning. In *Advances in Neural Information Processing Systems*, pages 721–731, 2018.

[181] Silas Nyboe Ørting, Jens Petersen, Veronika Cheplygina, Laura H Thomsen, Mathilde MW Wille, and Marleen De Bruijne. Feature learning based on visual similarity triplets in medical image analysis : A case study of emphysema in chest ct scans. In *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 140–149. Springer, 2018.

[182] David Ouyang, Bryan He, Amirata Ghorbani, Neal Yuan, Joseph Ebinger, Curtis P Langlotz, Paul A Heidenreich, Robert A Harrington, David H Liang, Euan A Ashley,

et al. Video-based ai for beat-to-beat assessment of cardiac function. *Nature*, 580(7802) :252–256, 2020.

[183] Mark Palatucci, Dean Pomerleau, Geoffrey E Hinton, and Tom M Mitchell. Zero-shot learning with semantic output codes. In *Advances in neural information processing systems*, pages 1410–1418, 2009.

[184] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE transactions on neural networks*, 22(2) :199–210, 2010.

[185] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10) :1345–1359, 2009.

[186] Jongchan Park, Sanghyun Woo, Joon-Young Lee, and In So Kweon. Bam : Bottleneck attention module. *arXiv preprint arXiv :1807.06514*, 2018.

[187] David Petersson. Deep learning : réseaux neuronaux rnn et cnn quelles différences ?, 2020.

[188] Tri-Cong Pham, Chi-Mai Luong, Muriel Visani, and Van-Dung Hoang. Deep cnn and data augmentation for skin lesion classification. In *Asian Conference on Intelligent Information and Database Systems*, pages 573–582. Springer, 2018.

[189] FATHIMA ANJILA PK. What is artificial intelligence ? *"Success is no accident. It is hard work, perseverance, learning, studying, sacrifice and most of all, love of what you are doing or learning to do"*., page 65, 1984.

[190] Sergey M Plis, Devon R Hjelm, Ruslan Salakhutdinov, Elena A Allen, Henry J Bockholt, Jeffrey D Long, Hans J Johnson, Jane S Paulsen, Jessica A Turner, and Vince D Calhoun. Deep learning for neuroimaging : a validation study. *Frontiers in neuroscience*, 8 :229, 2014.

[191] W Nicholson Price and I Glenn Cohen. Privacy in the age of medical big data. *Nature medicine*, 25(1) :37–43, 2019.

[192] Santi Puch, Irina Sánchez, and Matt Rowe. Few-shot learning with deep triplet networks for brain imaging modality recognition. In *Domain Adaptation and Representation*

Transfer and Medical Image Learning with Less Labels and Imperfect Data, pages 181–189. Springer, 2019.

[193] J Ross Quinlan et al. Bagging, boosting, and c4. 5. In *Aaai/Iaai, vol. 1*, pages 725–730, 1996.

[194] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv :1511.06434*, 2015.

[195] Alvin Rajkomar and Eyal Oren. Deep learning for electronic health records, 2018.

[196] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. 2016.

[197] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing : Overview, challenges and the future. In *Classification in BioApps*, pages 323–350. Springer, 2018.

[198] Russell Reed and Robert J MarksII. *Neural smithing : supervised learning in feedforward artificial neural networks*. Mit Press, 1999.

[199] Scott Reed, Zeynep Akata, Honglak Lee, and Bernt Schiele. Learning deep representations of fine-grained visual descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 49–58, 2016.

[200] Anthony Rios and Ramakanth Kavuluru. Few-shot and zero-shot multi-label learning for structured label spaces. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, volume 2018, page 3132. NIH Public Access, 2018.

[201] Shamnaz Riyaz, Kunal Sankhe, Stratis Ioannidis, and Kaushik Chowdhury. Deep learning convolutional neural networks for radio identification. *IEEE Communications Magazine*, 56(9) :146–152, 2018.

[202] Bernardino Romera-Paredes and Philip Torr. An embarrassingly simple approach to zero-shot learning. In *International Conference on Machine Learning*, pages 2152–2161, 2015.

[203] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[204] Abhijit Guha Roy, Shayan Siddiqui, Sebastian Pölsterl, Nassir Navab, and Christian Wachinger. 'squeeze & excite'guided few-shot segmentation of volumetric images. *Medical image analysis*, 59 :101587, 2020.

[205] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv :1609.04747*, 2016.

[206] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3) :211–252, 2015.

[207] Stuart J Russell. *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010.

[208] Jihyeon Ryu, Yifeng Zheng, Yansong Gao, Sharif Abuadbba, Junyaup Kim, Dongho Won, Surya Nepal, Hyoungshick Kim, and Cong Wang. Can differential privacy practically protect collaborative deep learning inference for the internet of things ? *arXiv e-prints*, pages arXiv–2104, 2021.

[209] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment : A natural scene statistics approach in the dct domain. *IEEE transactions on Image Processing*, 21(8) :3339–3352, 2012.

[210] Ruslan Salakhutdinov. Learning deep generative models. *Annual Review of Statistics and Its Application*, 2 :361–385, 2015.

[211] Pouya Samangouei, Maya Kabkab, and Rama Chellappa. Defense-gan : Protecting classifiers against adversarial attacks using generative models. *arXiv preprint arXiv :1805.06605*, 2018.

[212] Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller. *Explainable AI : interpreting, explaining and visualizing deep learning*, volume 11700. Springer Nature, 2019.

[213] Arthur G Samuel. Phonemic restoration : insights from a new methodology. *Journal of Experimental Psychology : General*, 110(4) :474, 1981.

[214] Arthur L Samuel. Some studies in machine learning using the game of checkers. ii—recent progress. *Computer Games I*, pages 366–400, 1988.

[215] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850. PMLR, 2016.

[216] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. One-shot learning with memory-augmented neural networks. *arXiv preprint arXiv :1605.06065*, 2016.

[217] Afşar Saranli, Stuart Russel, and Peter Norvig. Artificial intelligence : a modern approach. 2003.

[218] Tanja Schroeder, Maximilian Haug, Heiko Gewald, et al. Data privacy concerns using mhealth apps and smart speakers : Comparative interview study among mature adults. *JMIR Formative Research*, 6(6) :e28025, 2022.

[219] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet : A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[220] P Seebock. Deep learning in medical image analysis. *Master's thesis, Vienna University of Technology, Faculty of Informatics*, 2015.

[221] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam : Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[222] Amit Shah, Sailesh Conjeti, Nassir Navab, and Amin Katouzian. Deeply learnt hashing forests for content based image retrieval in prostate mr images. In *Medical Imaging 2016 : Image Processing*, volume 9784, page 978414. International Society for Optics and Photonics, 2016.

[223] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19 :221–248, 2017.

[224] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1) :1–48, 2019.

[225] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587) :484–489, 2016.

[226] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks : Visualising image classification models and saliency maps. *arXiv preprint arXiv :1312.6034*, 2013.

[227] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568–576, 2014.

[228] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*, 2014.

[229] Vishrut Singhal. What do you mean by convolutional neural network ?, 2021.

[230] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. Smoothgrad : removing noise by adding noise. *arXiv preprint arXiv :1706.03825*, 2017.

[231] Dave Smith. Cutting-edge face recognition is complicated. these spreadsheets make it easier, 2018.

[232] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pages 4077–4087, 2017.

[233] Richard Socher, Milind Ganjoo, Christopher D Manning, and Andrew Ng. Zero-shot learning through cross-modal transfer. In *Advances in neural information processing systems*, pages 935–943, 2013.

[234] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout : a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1) :1929–1958, 2014.

[235] Nitish Srivastava and Ruslan R Salakhutdinov. Multimodal learning with deep boltzmann machines. In *Advances in neural information processing systems*, pages 2222–2230, 2012.

[236] StatSoft. Réseaux de neurones, 1984-2013.

[237] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 101 :569–582, 2014.

[238] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Latent feature representation with stacked auto-encoder for ad/mci diagnosis. *Brain Structure and Function*, 220(2) :841–859, 2015.

[239] Heung-Il Suk and Dinggang Shen. Deep learning-based feature representation for ad/mci classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 583–590. Springer, 2013.

[240] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 403–412, 2019.

[241] Ruo-Yu Sun. Optimization for deep learning : An overview. *Journal of the Operations Research Society of China*, 8(2) :249–294, 2020.

[242] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare : Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018.

[243] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR, 2013.

[244] Silicon Valley Bank (SVB). Big data next : Capturing the promise of big data. big data report 2015, 2015.

[245] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[246] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[247] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[248] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface : Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.

[249] Ilyes Talbi. Quelles sont les applications de la computer vision ?

[250] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. pages 270–279, 2018.

[251] Mingxing Tan and Quoc Le. Efficientnet : Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[252] Sebastian Thrun and Lorien Pratt. Learning to learn : Introduction and overview. In *Learning to learn*, pages 3–17. Springer, 1998.

[253] Sebastian Thrun and Lorien Pratt. *Learning to learn*. Springer Science & Business Media, 2012.

[254] Vincent Tinto. Dropout from higher education : A theoretical synthesis of recent research. *Review of educational research*, 45(1) :89–125, 1975.

[255] Eric Topol. *Deep medicine : how artificial intelligence can make healthcare human again*. Hachette UK, 2019.

[256] Kittler.H Tschandl.P, Rosendahl.C. H. the ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. sci. data 5, 180161 doi :10.1038/sdata.2018.161. 2018.

[257] Alan Turing. Intelligent machinery. 1948. *The Essential Turing*, pages 395–432, 1969.

[258] Alan M Turing. Computing machinery and intelligence. In *Parsing the turing test*, pages 23–65. Springer, 2009.

[259] Alan Mathison Turing et al. On computable numbers, with an application to the entscheidungsproblem. *J. of Math*, 58(345-363) :5, 1936.

[260] Kévin Vancappel. Tutoriel | deep learning : le réseau neuronal convolutif (cnn), 2021.

[261] Andre Vellino. Artificial intelligence : The very idea : J. haugeland, (mit press, cambridge, ma, 1985) ; 287 pp. *Artificial Intelligence*, 29 :349–353, 09 1986.

[262] Benoit Vibert, Jean-Marie Le Bars, Christophe Charrier, and Christophe Rosenberger. Logical attacks and countermeasures for fingerprint on-card-comparison systems. *Sensors*, 20(18) :5410, 2020.

[263] Sandra Vieira, Walter HL Pinaya, and Andrea Mechelli. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders : Methods and applications. *Neuroscience & Biobehavioral Reviews*, 74 :58–75, 2017.

[264] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016.

[265] John Von Neumann. The computer and the brain. In *The Computer and the Brain*. Yale university press, 2012.

[266] John Von Neumann and Oskar Morgenstern. Theory of games and economic behavior. In *Theory of games and economic behavior*. Princeton university press, 2007.

[267] Demetris Vrontis, Michael Christofi, Vijay Pereira, Shlomo Tarba, Anna Makrides, and Eleni Trichina. Artificial intelligence, robotics, advanced technologies and human resource management : a systematic review. *The International Journal of Human Resource Management*, 33(6) :1237–1266, 2022.

[268] S. Prince W. Zi, L. S. Ghoraie. Few-shot learning and meta-learning i, 2019.

[269] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8 : Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.

[270] YAQING Wang, J Kwok, LM Ni, and Q Yao. Generalizing from a few examples : A survey on few-shot learning. *arXiv preprint arXiv :1904.05046*, 2019.

[271] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples : A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3) :1–34, 2020.

[272] Yuebin Wang, Liqiang Zhang, Hao Deng, Jiwen Lu, Haiyang Huang, Liang Zhang, Jun Liu, Hong Tang, and Xiaoyue Xing. Learning a discriminative distance metric with label consistency for scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8) :4427–4440, 2017.

[273] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1) :9, 2016.

[274] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515. Springer, 2016.

[275] Mika Westerlund. The emergence of deepfake technology : A review. *Technology Innovation Management Review*, 9(11), 2019.

[276] Patrick Henry Winston. *Artificial intelligence*. Addison-Wesley Longman Publishing Co., Inc., 1992.

[277] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam : Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–19, 2018.

[278] Qingyang Wu, Carlos Feres, Daniel Kuzmenko, Ding Zhi, Zhou Yu, Xin Liu, et al. Deep learning based rf fingerprinting for device identification and wireless security. *Electronics Letters*, 54(24) :1405–1407, 2018.

[279] Kai-jian Xia, Hong-sheng Yin, and Jiang-qiang Wang. A novel improved deep convolutional neural network model for medical image fusion. *Cluster Computing*, 22(1) :1515–1527, 2019.

[280] Yongqin Xian, Bernt Schiele, and Zeynep Akata. Zero-shot learning-the good, the bad and the ugly. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4582–4591, 2017.

[281] Liyang Xie, Kaixiang Lin, Shu Wang, Fei Wang, and Jiayu Zhou. Differentially private generative adversarial network. 2018.

[282] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv :1505.00853*, 2015.

[283] Runhua Xu. *Functional encryption based approaches for practical privacy-preserving machine learning*. PhD thesis, University of Pittsburgh, 2020.

[284] Wei Yang, Yingyin Chen, Yunbi Liu, Liming Zhong, Genggeng Qin, Zhentai Lu, Qianjin Feng, and Wufan Chen. Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. *Medical image analysis*, 35 :421–433, 2017.

[285] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging : A review. *Medical image analysis*, 58 :101552, 2019.

[286] Wei Ying, Yu Zhang, Junzhou Huang, and Qiang Yang. Transfer learning via learning to transfer. In *International Conference on Machine Learning*, pages 5085–5094, 2018.

[287] Jinsung Yoon, Lydia N Drumright, and Mihaela Van Der Schaar. Anonymization through data synthesis using generative adversarial networks (ads-gan). *IEEE journal of biomedical and health informatics*, 24(8) :2378–2388, 2020.

[288] Da Yu, Huishuai Zhang, Wei Chen, and Tie-Yan Liu. Do not let privacy overbill utility : Gradient embedding perturbation for private learning. *arXiv e-prints*, pages arXiv–2102, 2021.

[289] Tianhe Yu, Chelsea Finn, Annie Xie, Sudeep Dasari, Tianhao Zhang, Pieter Abbeel, and Sergey Levine. One-shot imitation from observing humans via domain-adaptive meta-learning. *arXiv preprint arXiv :1802.01557*, 2018.

[290] Lei Yuan, Yalin Wang, Paul M Thompson, Vaibhav A Narayan, Jieping Ye, Alzheimer's Disease Neuroimaging Initiative, et al. Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. *NeuroImage*, 61(3) :622–632, 2012.

[291] Yan Zhang, Min Fang, and Nian Wang. Channel-spatial attention network for fewshot classification. *PloS one*, 14(12), 2019.

[292] Ziming Zhang and Venkatesh Saligrama. Zero-shot learning via semantic similarity embedding. In *Proceedings of the IEEE international conference on computer vision*, pages 4166–4174, 2015.

[293] Amy Zhao, Guha Balakrishnan, Fredo Durand, John V Guttag, and Adrian V Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8543–8553, 2019.

[294] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.

[295] Tongxue Zhou, Su Ruan, and Stéphane Canu. A review : Deep learning for medical image segmentation using multi-modality fusion. *Array*, 3 :100004, 2019.

[296] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++ : A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

[297] Alexander Zimmer. Intelligence artificielle for future ? *Bulletin des médecins suisses*, 102(42) :1352–1352, 2021.

Représentation uniforme de l'imagerie médicale

Résumé :

Le domaine médical est à la fois critique et vaste. C'est un terrain avec une large marge d'innovation et d'amélioration face aux enjeux souvent très importants voir vitaux. L'apprentissage profond de son côté représente une perspective importante dans de multiples domaines et en particulier dans le domaine de l'imagerie médicale. La restriction souvent rencontrée lors du déploiement de cette technique dans ce domaine est les données : la disponibilité et la confidentialité. Dans ce travail de thèse, nous proposons d'offrir aux experts médicaux des multiples nouvelles pratiques pour utiliser l'apprentissage profond avec une quantité limitée de données. Nous soulignons également le danger de la pseudo-anonymisation et nous proposons un pipeline permettant une véritable anonymisation liée à l'identité du patient et à l'équipement d'acquisition.

Mots clés : Intelligence artificielle, apprentissage profond, anonymisation, attention, meta-apprentissage, few-shot-learning.

Abstract :

The area of medicine is critical and enormous. This is a field with great potential for innovation and improvement in the face of challenges often very important, if not vital. Deep learning, on the other hand, represents an important perspective in several fields, particularly in the field of medical imaging. The limitation often encountered when deploying this technique in this area is the data : availability and privacy. In this thesis work, we propose to offer medical experts multiple new practices to use deep learning with a limited amount of data. We also highlight the danger of pseudo-anonymization and provide a pipeline for true anonymization related to patient identity and acquisition equipment.

Key-words : Artificial Intelligence, deep-learning, anonymisation, attention, meta-learning, few-shot-learning.

Publications :

- DeepMRS : An End-to-End Deep Neural Network for Dementia Disease Detection using MRS Data : IEEE ISBI 2020, Iowa, USA
- Deep anonymization of medical imaging : Multimedia Tools and Applications