

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE
UNIVERSITÉ MOHAMED BOUDIAF - M'SILA

FACULTÉ : TECHNOLOGIE

DEPARTEMENT : ELECTRONIQUE



DOMAINE : Sciences et Technologies

FILIÈRE : Électronique

OPTION : Électronique des Systèmes Embarqués

Mémoire présenté pour l'obtention du diplôme de Master Académique

Par: Tawfik Beghriche

Ben taybe farida

Intitulé

**Prédiction des maladies les plus fréquentes
par les techniques de deep learning**

Soutenu publiquement le ** october 2020 devant un jury composé de:

Mr. DJERIOUI Mohamed

Mr. BRIK Youcef

Université de M'sila

Université de M'sila

Université de M'sila

Université de M'sila

Co-Encadreur

Encadreur

Année universitaire : 2019 /2020

*Remerciements

Tout d'abord, Nous remercions **ALLAH** miséricorde de nous avoir donné la volonté, le courage et la patience qui nous ont permis de réaliser ce travail.

Nous tenons à exprimer nos vifs remerciements a nos encadreurs **Mr. Djrioi Mohamed** et **Mr. BRIK Youcef** d'avoir proposé ce sujet et pour avoir dirigé notre travail durant la période d'élaboration de ce mémoire.

Nous exprimons notre profonde gratitude pour nous avoir fait profiter de leurs connaissances, mais aussi de leurs méthodes de travail, et surtout de leur rigueur scientifique.

Nos remerciements s'adressent également à tous les membres de Jury, qui ont accepté de nous honorer

Nos remerciements vont aussi à tous les enseignants du département d'électronique sans oublier toutes les personnes de ce même département.

Enfin, nous remercions A tous ceux qui ont contribué de près ou de loin à la réalisation de ce projet de fin d'études.

Dédicace

A mes chers parents.

A mes chers frères et sœurs.

Toute la famille.

Tous mes amis.

*Tous les collègues de notre promotion
(2020).*

Tawfik Beghriche

Dédicace

A mes chers frères et sœurs.

Toute la famille.

Tous mes amis.

*Tous les collègues de notre promotion
(2020).*

Farida Ben taybe

Table des matières

LISTE DES FIGURES	iv
LISTE DES TABLEAUX	v
ABREVIATION ET NOTATION	vi
INTRODUCTION GENERALE	1

CHAPITRE I La biomédicale et l'apprentissage des machines

I.1.INTRODUCTION	03
I.2.SI QUOI LE DEPISTAGE ET POURQUOI?	03
I.3. BIOMEDICAL ET L'ELECTRONIQUE	04
I.3.1 Electromyogramme	04
I.3.2 Electrocardiogramme	05
I.3.3.Maladies cardiaques (Heart Disease)	05
I.3.4.Maladie d'Alzheimer	06
I.3.5.La maladie de diabète	07
I.3.6.Dépistage du cancer	07
I.4.INTELLIGENCE ARTIFICIELLE	08
I.4.1 Historique	08
I.4.2.Multiples applications de l'intelligence artificielle	09
I.4.3.Future de l'intelligence artificielle	12
I.5.QU'EST CE QUE L'APPRENTISSAGE AUTOMATIQUE OU MACHINE LEARNING ?	12
I.5.1.L'Apprentissage	13
I.5.2. Apprentissage des machines et ses applications a la biomedicale	14
I.6. CATEGORIES DES MACHINES D'APPRENTISSAGE	14
I.6.1. Machines par séparation	14
I.6.1.1. Machines à Vecteurs de Support (SVMs)	14
I.6.1.2 Les Réseaux de Neurones	17
I.6.1.2.1. Présentation de quelques Types de réseaux neuronaux	17
I.6.1.2.2.Types de problèmes résolubles par réseau de neurones	20
I.6.2. Machines par modélisation	21
I.7.LES DIFFÉRENTES MÉTHODES DE LA CLASSIFICATION ET L'APPRENTISSAGE	22

I.8. CONCLUSION	26
-----------------------	----

CHAPITRE II Système proposé

II.1. INTRODUCTION	27
II.2. MACHINE LEARNING VS DEEP LEARNING	27
II.2.1. Quelles sont leurs différences ?	28
II.2.2. Les champs d'application	29
II.2.2.1. Les champs d'application d'un Machine learning	30
II.2.2.2. Les champs d'application d'un Deep learning	30
II.3. POURQUOI LE DEEP LEARNING?	31
II.4. DEFINITIONS DE DEEP LEARNING	32
II.4.1. Comment fonctionne Le Deep Learning	33
II.4.2. Quels sont les réseaux de neurones pour l'apprentissage en profondeur ?	34
.....	34
II.4.3. Méthodes d'apprentissage approfondi	35
II.4.4. Les applications de l'apprentissage profond	37
II.5. DIFFERENTS TYPES DE MODELES D'APPRENTISSAGE EN PROFONDEUR EXPLIQUES	37
II.6. ARCHITECTURES D'APPRENTISSAGE EN PROFONDEUR	38
II.6.1. RÉSEAU DE NEURONES PROFONDS __DNN__	39
II.6.2. RESEAU DE NEURONES CONVOLUTIONNELS	41
II.6.2.1. Historique	41
II.6.2.2. Principe générale	41
II.6.2.3. L'opération de convolution	42
II.6.2.4. La couche ReLU (rectification linéaire)	44
II.6.2.5. Le pooling	44
II.6.2.6. Le flattening (ou mise à plat)	46
II.6.2.7. fully connected	46
II.6.2.8. le padding	47
II.6.3. RESEAUX DE NEURONES A ACTION DIRECTE (DFFnn)	48
II.6.3.1. Historique	48
II.6.3.2. Principe générale	48

II.6.3.3.Le perceptron (neurone)	49
II.6.3.4. Réseaux de neurones artificiels	50
II.6.3.5. Softmax activation function	51
II.6.3.6. loss function	51
II.6.4. Réseaux de neurones récurrents RNN	52
II.6.5. Réseaux LSTM	53
II.7. LES FONCTIONS D'ACTIVATION	54
II.7.1 Sigmoidé	55
II.7.2 Tangente hyperbolique	55
II.7.3 Softmax	55
II.7.4 Unité linéaire rectifiée ("Rectified Linear Unit", ReLU)	55
II.7.5 Variantes de ReLU	56
II.8.6 Maxout	56
II.8.CONCLUSION	57
CHAPITRE III Résultats expérimentaux	
III.1. INTRODUCTION	58
III.2. BASE DE DONNÉES UTILISÉE	58
III.2.1. Heart Disease	58
III.2.2. Brest cancer	59
III.2. 3. Diabète	61
III.2. 4. L'Alzheimer	62
III.3. PROTOCOLE EXPÉRIMENTAL	63
III.3.1. Taux de classification	64
III.3.2. Taux d'erreur	64
III.4. RÉSULTATS DE SIMULATION	65
III.4.1. Avec la base Heart Disease	65
III.4.2. Avec la base Breast cancer	67
III.4.3. Avec la base diabetes	68
III.4.4. Avec la base Alzheimer	70
III.5. CONCLUSION	71

CONCLUSION GENERALE.....	72
REFERENCES	73

LISTE DES FIGURES

Fig. I.1. la forme d'un signal EMG	04
Fig. I.2. Signal ECG	05
Fig. I.3. le processus typique du machine learning	12
Fig I.4. une modèle de ML qui essaye d'obtenir la meilleure approximation possible de F.....	13
Fig I.5. Création de B bootstrap à partir des exemples d'apprentissage, chacun servira à entraîner un arbre de décision. Si c'est un cas de classification, la classe finale est attribuer par vote majoritaire. Si c'est un cas de régression, on fait la moyenne des prédictions.....	15
Fig I.6. On cherche un hyperplan qui divise les observations en deux catégories. Considérons un exemple x qu'on veut classer, si $f(x) > 0$, il appartient à la classe des cercles, sinon il appartient à la classe des carrés. Dans cette figure on peut voir qu'il existe une infinité d'hyperplans séparateurs possibles	16
Fig I.7. L'hyperplan optimal (en vert) avec la marge maximale. Les échantillons remplis sont des vecteurs supports	16
Fig I.8. Exemple d'un problème non linéairement séparable. La courbe devient une bande linéaire après avoir appliqué la transformation ϕ non linéaire.....	17
Fig I.9. La figure représente le fonctionnement d'un neurone formel	18
Fig I.10. Perceptron multicouche	19
Fig I.11. Exemple classification	20
Fig I.12. Exemple Régression	20
Fig. I.13. Le principe de MMG	21
Fig. I.14. Le principe de MMC	22
Fig I.15. l'aprontissage supervisees	23
Fig I.16. Extrait de la classification taxinomique de Linné.....	24
Fig. I.17. Schéma d'un modèle semi-supervisé	26
Fig. II.1. le relation entre AI , ML et DL	27
Fig. II.2. La différence de performance entre le Deep Learning et la plupart des algorithmes de ML en fonction de la quantité de données	32
Fig. II.3. Le procède du ML classique comparé à celui du Deep Learning	32
Fig. II.4. Système de reconnaissance de formes	33

Fig.II.5.Chronologie graphique montrant le développement des cinq principales architectures d'apprentissage en profondeur par date, de 1990 à 2015	38
Fig.II.6.le fonctionnement d'un réseau de neurones profonds DNN	40
Fig.II.7.La différence entre une vision par ordinateur et une vision humaine	41
Fig.II.8.l'architecture général de la CNN	42
Fig.II.9.comment fonctionne la convolution	43
Fig.II.10.les fonctions d'activation ReLU et Leaky ReLU et la différence entre elles.....	44
Fig.II.11.la différence entre Max pooling et Average pooling	45
Fig.II.12.comment fonctionne le flattening	46
Fig.II.13.la distribution des couches fully_conncted	47
Fig.II.14.un padding de 2 sur une matrice de taille 32×32	47
Fig.II.15. la distribution d'un Réseaux de neurones à action direct	49
Fig.II.16. comment fonctionne une neurone	50
Fig.II.17. Image avec des cercles et des flèches démontrant l'interrelation entre les couches réseau d'entrée, de sortie, cachées et contextuelles	52
Fig.II.18.Image avec cercles et flèches montrant la cellule de mémoire LSTM et le flux d'informations à travers les différentes portes	53
Fig.II.19.Diagramme d'une cellule GRU typique	54
Fig. II.20.representation d'un différent types de activation functions	56
Fig.III.1. la forme des données des patients de cardiaques	58
Fig.III.2. la forme des données des patients de cancer du sein	60
Fig.III.3. la forme des données des patients de diabète.....	61
Fig.III.4. la forme des données des patients d'Alzheimer	62
Fig.III.5. Taux de classification illustré dans la matrice de confusion	64
Fig. III.6. Taux d'erreur illustré dans la matrice de confusion.	64
Fig. III.7. la différence entre accuracy et loss d'apprentissage et de test avec l'ensemble de données des patients des cardiaques avec le model DNN.	66
Fig. III.8. la matrix de confusion des patients des cardiaques avec le model DNN.....	66
Fig. III.9. la différence entre accuracy et loss d'apprentissage et de test avec l'ensemble de données des patients de cancer du sein avec le model DNN	68
Fig. III.10. la matrix de confusion des patients de cancer du sein avec le model DNN.....	68

Fig. III.11. la différence entre accuracy et loss d'apprentissage et de test avec l'ensemble de données des patients du diabète avec le model DNN	69
Fig. III.12. la matrix de confusion des patients du diabète avec le model DNN.....	70
Fig. III.13. la différence entre accuracy et loss d'apprentissage et de test avec l'ensemble de données des patients d'Alzheimer avec le model DNN.....	71

LISTE DES TABLEAUX

Tableau II.1. Synthèse les différences entre ML et DL.....	29
Tableau II.2. quelques applications pour un groupe d'architecteurs.....	39
Tableau III.1. Informations d'attributs de la base de données des patients de cardiaques.....	59
Tableau III.2. Informations d'attributs de la base de données des patients de cancer du sein.....	60
Tableau III.3. Informations d'attributs de la base de données des patients de diabète.....	62
Tableau III.4. Informations d'attributs de la base de données des patients d'Alzheimer.....	63
Tableau III.5. Résultats obtenus avec la base de données des patients des cardiaques.....	65
Tableau III.6. Résultats obtenus avec la base de données des patients de cancer du sein.....	67
Tableau III.7. Résultats obtenus avec la base de données des patients du diabète.....	69
Tableau III.8. Résultats obtenus avec la base de données des patients d'Alzheimer.....	70

LISTE DES ABREVIATIONS

EMG : Electromyogramme

ECG : Electrocardiogramme

NHLBI : National Heart, Lung, and Blood Institute

GPS : genaral problem solver

IA : Intelligence Artificielle

FAIR : Facebook Artificial Intelligence Research

SVM : Support Vector Machines

RNA : Réseaux de Neurones Artificiels

MLP : Multi Layer Perceptron

MMG : Modèle de Mélange Gaussien

MMC : Modèles de Markov Cachée

ML : Machine Learning

DL : Deep Learning

DNN : deep neural network

DFFnn : deep feed forward neural network

CNN : convlutional neural network

RNN : Réseau neuronal recurrent

LSTM : Long Short-Term Memory

Introduction générale

INTRODUCTION GENERALE

En médecine, l'un des principaux défis est de prélever un échantillon ou un signal vital vers un patient via un capteur spécifique ou par analyse physico-chimique pour l'étudier et le traiter et extraire des informations pertinentes pour identifier les maladies liées. En raison du coût très élevé de ces traitements et du nombre de patients qui augmentent de jour en jour, la nécessité d'automatiser ce traitement ou ce comportement est devenue une nécessité et une solution à ce problème, et avec le développement de l'intelligence artificielle ces derniers temps années où l'on peut remplacer le professionnel de santé par une machine où l'on peut Ces derniers prennent des décisions critiques et un taux d'erreur très faible par rapport aux facilités fournies par l'inclusion des techniques d'intelligence artificielle, notamment de la rapidité de mise en œuvre, il est capable de diagnostiquer des centaines, mais des milliers de patients en très peu de temps par rapport aux médecins.

Nous pouvons supposer que le problème de l'identification des maladies est considéré comme le problème de la reconnaissance des formes, où les modèles représentent un groupe d'observations ou d'échantillons de patients. Actuellement, la reconnaissance des formes a été largement étudiée ces dernières années dans la littérature scientifique. Cependant, malgré toutes les recherches approfondies, les techniques de reconnaissance disponibles dans ce domaine sont encore inférieures aux capacités visuelles humaines, car dans ces dernières ce sont des opérations mathématiques qui contiennent un taux d'erreur, même s'il est faible.

Pour activer la reconnaissance des formes, l'apprentissage automatique (ML) peut apprendre à séparer les formes de l'une des autres formes où les données doivent être linéaires. C'est pourquoi le soi-disant apprentissage en profondeur a émergé qui est même capable de traiter des données non linéaires, car de nombreuses techniques ont été développées pour améliorer les performances des systèmes biomédicaux tels que la reconnaissance des tumeurs, la segmentation des images médicales, la classification des cellules et l'identification des maladies. Malheureusement, dans les systèmes de la littérature, nous trouvons des techniques efficaces mais leur mise en œuvre peut être longue ou rapide mais ne donnent pas suffisamment de résultats satisfaisants.

Dance ce mémoire, nous présentons un système de reconnaissance des maladies à l'aide d'un ensemble de données biomédicales pour plusieurs maladies différentes (diabète, Alzheimer ...) en incluant des techniques d'apprentissage en profondeur où la différence entre ces techniques est mise en évidence en fonction des moyens de relier les couches et les méthodes des «neurones», il existe également un apprentissage non censuré, supervisé et amélioré dans les algorithmes d'apprentissage en profondeur, car ils indiquent la façon dont l'algorithme est alimenté en données par les chercheurs. Au-delà de cela, les réseaux de neurones convolutifs (CNN) sont typiques de la reconnaissance de données d'image, vidéo et audio, en raison de leur capacité à travailler avec des données denses. Les réseaux de neurones récurrents (RNN) sont utilisés pour traiter les langages naturels, tandis que les LSTM sont des différences dans les RNN qui visent à conserver des informations structurées basées sur les données.

Afin de compléter ce travail, notre thèse est organisée en trois chapitres comme suit:

- Le premier chapitre donne un aperçu général de la biomédecine et de sa relation avec l'apprentissage automatique.
- Dans le chapitre deux, nous mettrons en évidence les différences les plus importantes entre l'apprentissage automatique et l'apprentissage en profondeur, puis nous expliquerons comment fonctionnent les techniques d'apprentissage en profondeur.
- Le troisième chapitre présente et discute les résultats obtenus avec quatre bases de données réelles différentes et une comparaison des résultats pour déterminer la technique qui donne de meilleures performances que d'autres.

Chapitre I
**La biomédicale et
l'apprentissage des machines**

I.1.INTRODUCTION

Les termes « biomédecine » et « biomédical » ont des origines et des trajectoires différentes en anglais, en allemand «Biomedizin» et en français « biomédecine », mais leur sens est très similaire aujourd'hui. «Biomédical» est apparu pour la première fois dans les écrits d'auteurs américains et britanniques dans les années 1920, suivi dix ans plus tard par la «biomédecine». Le dictionnaire médical américain (1923) le définit comme «une médecine clinique fondée sur les principes de la physiologie et de la biochimie». Ainsi, dès le début, la biomédecine et la recherche biomédicale ont été comprises comme une sorte de médicament étroitement associé à l'expérimentation et au laboratoire plutôt qu'au savoir du médecin et de la clinique.

I.2.SI QUOI LE DEPISTAGE ET POURQUOI?

Un dépistage est une action de santé publique, c'est -à-dire une liste d'opérations entreprendre de caractère collectif en vue de l'amélioration de la santé d'une population.

Le dépistage d'une maladie a des retombées favorables chez les sujets correctement identifiés comme positifs ou négatifs.

Pour les vrais positifs, les avantages peuvent être une meilleure efficacité de l'intervention(traitement ou mesure préventive) instaurée plus précocement, de meilleures chances de survie des patients, ou une amélioration de la qualité de vie des patients par réduction des traitements , et une économie de ressources générée par une réduction du coût total de la prise en charge de la maladie (traitements initiaux moins radicaux et diminution de la morbi-mortalité).

Pour les vrais négatifs, les avantages peuvent être un sentiment de tranquillité éprouvé par les patients et un éventuel allègement des mesures de surveillance. Par ailleurs, le dépistage peut contribuer à l'équité en permettant l'accès aux soins à une partie de la population dépistée [1].

Après de l'importance de dépistage médicale sur la sante humaine on peut utiliser des nouvelles méthode plus développée et précisé Pour améliorer l'efficacité et la résultat de notre dépistage comme la dépistage numérique et ci sa Quesque on va voir dans notre mémoire

I.3. BIOMEDICAL ET L'ELECTRONIQUE

Dans nos jours, l'étude des signaux biomédicaux et l'extraction de l'information utile d'une façon automatique sont très importantes et aident les médecins à bien comprendre le comportement de différents tissus cellulaires afin de faire un bon diagnostic et à construire les bases de données de patients. Dans la suite, nous citons les signaux biomédicaux les plus utilisés dans le traitement automatiques des patients à l'aide de l'apprentissage des machines:

I.3.1. Electromyogramme

L'électromyogramme (EMG) est enregistré en utilisant un appareil d'électromyographie, qui mesure le potentiel électrique du muscle. Le système nerveux central (le cerveau, la moelle épinière et les nerfs périphériques) contrôle l'action des fibres musculaires qui se traduit en général aux mouvements. Le muscle se compose de cellules spéciales qui sont capables de faire des dilatations et des relaxations et qui sont contrôlées par des neurones.

Un EMG peut être enregistré par deux méthodes, EMG de surface (en mettant des électrodes sur la surface de la peau) qui est plus populaire que la deuxième appelée EMG intramusculaire (où une électrode à aiguille est insérée dans la peau) parce qu'elle est non invasive.

Un EMG de surface mesure un seul ou plusieurs unités motrices, il est appelé potentiel d'action de l'unité motrice (PAUM). Le potentiel réel voisine les 100 mV mais à cause des couches de tissus conjonctives et de la peau, l'EMG est un signal complexe avec une amplitude beaucoup plus petite (environ 5 mV) (Figure I.1).

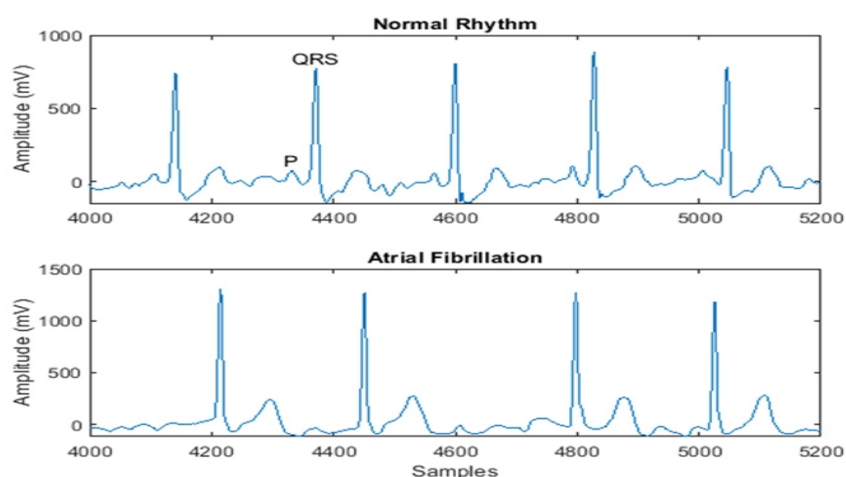


Fig. I.1. la forme d'un signal EMG.

I.3.2. Electrocardiogramme

L'électrocardiogramme (ECG) est la représentation de l'activité électrique du cœur, ce rythme cardiaque est contrôlé par la cellule stimulatrice connu sous le nom du nœud sinusal (SA).

Les ondes PQRST et parfois U constituent un cycle ECG (Figure I.2) : l'onde P est aperçue lorsque le nœud SA déclenche l'impulsion qui se propage dans les oreillettes et crée la contraction auriculaire ; l'intervalle PQ (où segment isométrique) est la durée de propagation de l'impulsion de l'oreillette vers les ventricules, permettant au flux sanguin de s'écouler dans la même direction; le complexe QRS se produit quand l'impulsion arrive aux ventricules et traduit la contraction ventriculaire.

Le segment ST est la période de la dépolarisation des ventricules, dans le cas normal, il doit être isoélectrique, sinon, le niveau d'amplitude et la pente de ce segment sont des indicateurs de l'état ischémique du myocarde et enfin l'onde T qui exprime la repolarisation ventriculaire, c'est l'état de repos des ventricules (relaxation).

Les impulsions électriques sont enregistrées par le biais des électrodes placées sur des surfaces particulières du corps, en utilisant un ECG 12-dérivations ou 12-canal (dans les hôpitaux) ou ECG 3-canal.

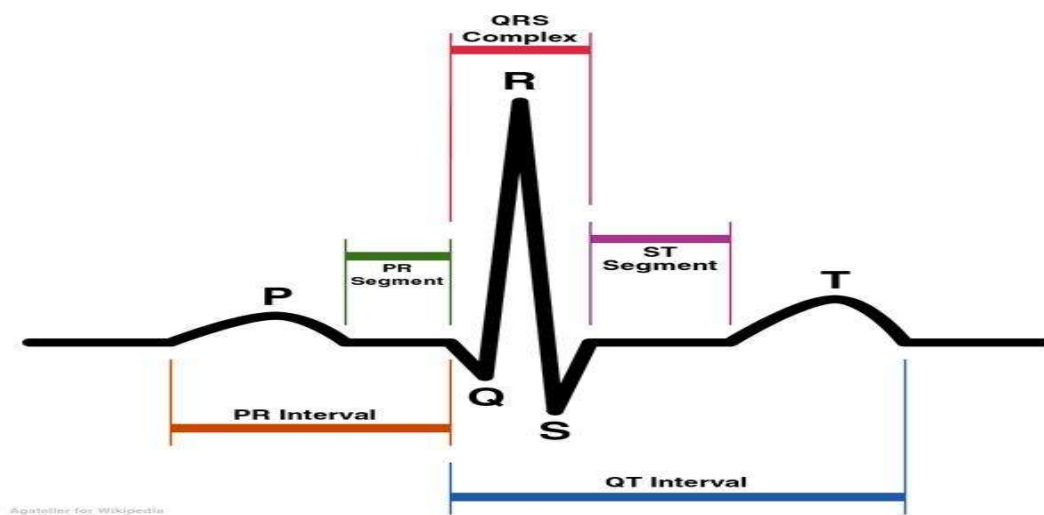


Fig. I.2. Signal ECG.

I.3.3.Maladies cardiaques (Heart Disease)

Le nombre de personnes souffrant de maladies cardiaques est en augmentation. Un diagnostic précis à un stade précoce suivi d'un traitement ultérieur approprié peut permettre de sauver un nombre des vies considérable. De nouvelles données publiées par l'Institut national du cœur, des poumons et du sang (NHLBI) montrent que les femmes plus âgées, en particulier, sont plus à risque de contracter une maladie cardiaque.

Une étude récente sur le terrain d'une maladie cardiaque peut être contrôlée efficacement si elle est diagnostiquée à un stade précoce. Mais ce n'est pas facile de faire un diagnostic précis en raison de nombreux facteurs complexes de maladies cardiaques. Par exemple, de nombreux symptômes cliniques sont associés à de nombreux organes humains autres que le cœur et très souvent, les maladies cardiaques peuvent présenter divers syndromes.

En raison de cette complexité, il est nécessaire d'automatiser le processus de diagnostic médical qui peut aider les médecins dans le processus de diagnostic.

I.3.4.Maladie d'Alzheimer

900.000 personnes sont touchées par la maladie d'Alzheimer en France et chaque année [23] . trouble à la mémoire, difficulté à se repérer dans le temps et l'espace, perte progressive du langage et de l'autonomie...sont quelques-uns des symptômes caractéristiques des maladies neuro de génératives comme la maladie d'Alzheimer.

Dans notre cerveau, les neurones sont organisés en réseau. ces cellules transmettent l'influx nerveux et les informations. Dans la maladie d'Alzheimer, ce tissu cérébral est lésé .dans le cerveau des malades, on observe trois choses caractéristiques de la maladie :

- Une perte de neurones
- La formation de plaques dites amyloïdes, dues à l'accumulation de certaines protéines(les protéines Tau)qui composent la structure interne des neurones.

Toutes ces lésions progressent dans le cerveau année après année. Elles commencent par toucher les structures les plus internes du cerveau, dont l'hippocampe qui est dédiée à la mémoire. C'est la raison pour laquelle la maladie se manifeste généralement au début par des troubles de la mémoire.

Puis, les lésions atteignent les zones postérieures et externes du cerveau qui traitent les informations plus globales, liées à la gestuelle et au langage. D'où des troubles du langage (nommés aphasie).les patients ne parviennent plus à nommer les objets (c'est l'agnosie), ils ne reconnaissent plus les choses, ont des difficultés à coordonner leurs gestes et effectuer des

gestes concrets (ce qui est appelé apraxie)...A terme, le cerveau des malades finit par s'atrophier totalement.

Quand les premiers signes de la maladie d'Alzheimer apparaissent, le processus est déjà à l'œuvre depuis plusieurs années.

Ces signes comportent souvent des troubles de la mémoire, de la pensée du jugement, déraisonnement, du langage et des troubles du comportement (agitation, agressivité, insomnie, dépression, anxiété, paranoïa, hallucinations visuelles).A un stade plus avancé, la personne devient dépendante, ne reconnaît plus ses proches et court des risques de chutes et d'accidents.

Le diagnostic repose avant tout sur l'observation des signes et sur l'entretien avec la famille.

Un examen neurologique et des tests évaluant la mémoire ainsi que les capacités à faire face aux activités quotidiennes sont également réalisés .Prise de sang, électroencéphalogramme, ponction lombaire, scanner du crâne, doppler des vaisseaux du cou sont parfois nécessaires pour écarter une cause curable de démence.

L'Avancement pour dépister la maladie d'Alzheimer, on utilise des tests neuropsychologique, l'imagerie médicale pour visualiser la zone responsable de la mémoire dans le cerveau, ou encore la ponction lombaire pour détecter les marqueurs biologiques de la maladie[2].

I.3.5.La maladie de diabète

Le diabète correspond à une élévation anormale de la glycémie. Définie par le taux de sucre dans le sang .Cette augmentation de la glycémie peut provoquer à plus ou moins long terme des lésions de différents organes. Comme les yeux, les reins, les nerfs et les vaisseaux sanguins. Près de 90%des diabétiques vivent pendant des années avec cette maladie sans le savoir car le diabète(ou diabète sucre même si le terme est un peu dépassé)est évoqué lorsque la glycémie à jeun est supérieure ou égale à 1,20g/l.il est conseillé de vérifier ce chiffre une seconde fois d'avoir deux dosages de la glycémie[3].

I.3.6.Dépistage du cancer

Le dépistage consiste en une série d'examens médicaux, répétés à intervalles réguliers, qui permettent de découvrir la présence éventuelle d'une maladie bien avant qu'elle ne commence à causer des anomalies perceptibles (symptômes). Seuls certains cancers sont actuellement concernés par le dépistage. Le dépistage permet de détecter certains cancers (ou certaines anomalies précancéreuses) longtemps avant l'apparition des premiers symptômes. Son objectif est clair :

- éviter un cancer grâce au traitement de lésions précancéreuses.
- augmenter les chances de guérison (ou permettre un traitement nettement moins agressif) grâce à une prise en charge précoce de la maladie.

Le dépistage n'est actuellement possible que pour une liste limitée de cancers. D'autres dépistages ne font pas l'objet d'un consensus scientifique, même si certains patients y ont recours après discussion avec leur médecin traitant.

I.4.INTELLIGENCE ARTIFICIELLE

C'est une discipline scientifique relative au traitement des connaissances et au raisonnement dans le but de permettre à une machine d'exécuter des fonctions normalement associées à l'être humain. L'intelligence artificielle tente de reproduire les processus cognitifs humains dans le but de réaliser des actions « intelligente ». Elle est comme « la construction des programmes informatique qui s'adonnent à des tâches qui sont pour l'instant accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que :

- L'apprentissage perceptuel.
- L'organisation de la mémoire et le raisonnement critique.

La norme ISO 2382-28 définit l'intelligence artificielle comme la « capacité d'une unité fonctionnelle à exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement et l'apprentissage ». Qualifiée de prochaine révolution informatique, l'intelligence artificielle est au cœur de tous les sujets d'actualités, il semble indispensable de définir cette technologie de rupture et de clarifier son régime juridique, mais aussi d'identifier les applications en cours ou en développement dans les entreprises et les bénéfices qu'elles en tirent.

I.4.1.Historique

Il est intéressant de revenir sur les origines et l'historique de l'intelligence artificielle afin de bien comprendre ses premières orientations et ses perspectives pour l'avenir:

- 1950 : Alan M. Turing, mathématicien et théoricien précurseur de l'informatique, lance le concept d'intelligence artificielle.
- 1955-1956: Lancement du premier programme d'intelligence artificielle par Allen Newell, John C. Shaw et Herbert A. Simon, Logic Theorist.
- 1957: Modélisation des jeux d'échec.

- 1958: John McCarthy invente le Lisp (list processing), langage de programmation interactif (développement au MIT).
- 1958: Construction du premier réseau neuronal, le Perceptron, de Frank Rosenblatt, machine dite connexionniste.
- 1959: Élaboration du premier GPS (général problem solver) -fin de la première période de l'intelligence artificielle.
- 1970: Néoconnexionnisme.
- 1989: DeepThought, supercalculateur d'IBM, deux millions de coups par seconde.
- 1990 - 1997: Développement de Deep Blue rebaptisé DeeperBlue: conception d'un système de 256 processeurs fonctionnant en parallèle, chaque processeur peut calculer environ trois millions de coups par seconde.
- 2009: Le MIT a lancé un projet visant à repenser la recherche en intelligence artificielle.
- 2011: Watson, le superordinateur d'IBM remporte deux des trois manches du jeu télévisé Jeopardy! La performance a consisté pour cette intelligence artificielle à répondre à des questions de culture générale.
- 2013: Humain Brain Project. Google ouvre un laboratoire de recherches dans les locaux de la NASA.
- 2014: DeepKnowledge Ventures : nommé à son conseil d'administration VITAL, un algorithme capable d'élaborer ses décisions en analysant les bilans comptables des entreprises potentiellement intéressantes, le test clinique, la propriété intellectuelle et les précédents investissements.
- 2015: Facebook Artificial Intelligence Research (FAIR). Google rend sa technologie d'intelligence artificielle TensorFlow accessible à tous. Développement d'une crainte que l'intelligence artificielle dépasse à terme les performances de l'intelligence humaine.
- 2016: Amelia d'IPSoft un agent virtuel. Aussi, AlphaGo bat trois fois consécutives le champion du monde de jeu de go, Lee Se-Dol en cinq manches.

I.4.2. Multiples applications de l'intelligence artificielle

L'intelligence artificielle a déjà trouvé de multiples usages au sein de la société :

- **Les moteurs de recherche** : Tous les moteurs de recherche (y compris des agences de voyage) sont basés sur des systèmes intelligents d'extraction, d'analyse, et de classification de données pour produire le plus rapidement possible un résultat pertinent à la requête de l'utilisateur. C'est ainsi que Google a implémenté un système utilisant les techniques de machine learning pour son moteur de recherche en octobre 2015, intitulé RankBrain9. Ce système convertit de grandes quantités de texte en vecteurs mathématiques pour aider le système à deviner le sens des mots ou des phrases qu'il ne connaît pas et ainsi traiter les 15 % de requêtes jamais effectuées auparavant qu'il reçoit chaque jour.
- **Les moteurs de recommandation** : En s'appuyant sur les données issues de la navigation et des achats d'un utilisateur, des sites comme Amazon ou Netflix sont capables de lui proposer d'autres produits similaires qui pourraient l'intéresser. Ces technologies prédictives sont aussi utilisées pour les plateformes de publicité en ligne (Google, Criteo) pour proposer aux visiteurs des contenus d'annonceurs en rapport avec les pages qu'ils ont visité.
- **La traduction automatique** : Elle s'appuie sur des algorithmes de modélisation statistique du langage naturel. Ils intègrent les règles de construction de chaque langue.
- **Les assistants personnels (Siri, Cortana, Google Now...)** : Ils sont déployés sur les Smartphones qui s'appuient sur plusieurs briques technologiques: la reconnaissance vocale pour convertir le son en texte, le langage naturel pour comprendre le sens des mots, un moteur de recherche pour trouver réponse à la question et la synthèse vocale pour communiquer la réponse à l'utilisateur, la planification pour la gestion d'événements, etc.
- **Les agents conversationnels** : Ils sont utilisés dans les domaines du support client et du télémarketing et consistent en des fenêtres de chat qui s'ouvrent toutes seules sur un site web, ou de serveur vocal qui répond aux questions 24h sur 24. Ils utilisent le langage naturel et leur accès à de vastes bases de données leur permet de répondre aux questions les plus simples.
- **Les véhicules autonomes** : Si certains prototypes roulent déjà sur les routes au contact des autres véhicules, les voitures qui se garent toutes seules ou qui freinent par anticipation sont déjà une réalité. Le pilotage automatique des avions ou la gestion de

trajectoire des véhicules spatiaux, ou encore les drones se basent aussi sur l'intelligence artificielle.

- **Les systèmes de navigation GPS :** Développé en 1968 par l'Institut de recherche de Stanford, cet algorithme permet d'optimiser le cheminement entre plusieurs points dans un réseau en se basant sur le coût du trajet ou la distance parcourue.
- **Les finances :** Elles sont gérées par des systèmes intelligents pour organiser leurs opérations, investir en bourse et gérer leurs biens, mais aussi pour repérer des transactions qui sortent de l'ordinaire. Les banques possèdent aussi des systèmes experts d'évaluation de risques liés à l'octroi de crédits (credit-scoring).
- **Le cyber sécurité :** Les acteurs du cyber sécurité ont adopté les techniques du machine learning afin de détecter des comportements anormaux dans les systèmes d'information, et de déceler les menaces persistantes pour éviter des opérations d'espionnage ou d'extraction de données¹⁰. Près de 300 paramètres (heures et IP des connexions et des machines, téléchargements, etc.) sont pris en compte pour établir le modèle d'analyse comportementale dont la première phase d'apprentissage dure environ une semaine. Citons la jeune pousse lyonnaise Sentryo qui intègre des algorithmes de machine learning pour sécuriser les sites industriels critiques.
- **Les jeux vidéo :** Ils emploient des techniques d'intelligence artificielle pour donner vie aux personnages non joueurs ou encore pour créer des univers entiers à partir d'algorithmes. En 1997, DeepBlue, le superordinateur d'IBM avait battu Garry Kasparov, champion du monde d'échecs en titre. En 2016, c'est DeepMind, le programme d'intelligence artificielle de Google qui a annoncé la victoire de son programme AlphaGo contre le champion d'Europe de go en titre, Fan Hui¹³. Ce résultat prend appui sur la technologie des réseaux neuronaux, que nous décrivons précédemment. Deepmind avait déjà développé un système d'intelligence artificielle capable de déterminer l'action la plus judicieuse pour battre l'homme dans une vingtaine de jeux d'arcade.
- **La médecine :** L'ordinateur Watson d'IBM vise à compiler la plus large base de données dédiée à la santé dans le monde, portant sur 300 millions de patients. La branche santé d'IBM a acquis quatre sociétés médicales depuis sa création : Phytel (santé & population), Explorys (fichiers santé de cliniques), Merge Healthcare (imagerie médicale) et Truven (analytics médicales). Ce supercalculateur a notamment fait ses preuves dans un service d'oncologie pour des services personnalisés :

analysant ADN, dossier du patient et autres publications et essais cliniques. L'ordinateur propose un protocole adapté au patient. L'entreprise américaine Enlitic propose des technologies d'intelligence artificielle pour analyser l'imagerie médicale, qui permettraient de déceler des fractures de façon plus performante que les radiologues d'après les dires de la société.

I.4.3.Future de l'intelligence artificielle

Nous venons de voir que l'intelligence artificielle est déjà une réalité pour de multiples applications. Mais au vu des recherches qui sont menées dans le monde entier, elle a encore de beaux jours devant elle. Les applications de l'intelligence artificielle pour le futur peuvent aller encore plus loin notamment dans le domaine de : voitures autonomes, robotique, bâtiments connectés et la médecine.

I.5.QU'EST CE QUE L'APPRENTISSAGE AUTOMATIQUE OU MACHINE LEARNING ?

L'apprentissage automatique est une discipline de l'IA qui offre aux ordinateurs la possibilité d'apprendre à partir d'un ensemble d'observations que l'on appelle ensemble d'apprentissage[4].

Chaque observation, comme par exemple « j'ai mangé tels et tels aliments à tel moment de la journée pendant telle période ce qui a causé telle maladie » est décrite au moyen de deux types de variables:

- Les premières sont appelées les variables prédictives (ou attributs ou caractéristiques), dans notre exemple mon âge, mon dossier médical, mes antécédents médicaux. Ce sont les variables à partir desquelles on espère pouvoir faire des prédictions. Les n variables prédictives associées à une observation seront notées comme un vecteur $x=(x_1... x_n)$ à n composantes. Un ensemble de M observations sera constitué de M tels vecteurs $x(1)... x(M)$ [4].

- Une variable cible dont on souhaite prédire la valeur pour des événements non encore observés. Dans notre exemple, il s'agirait de la maladie contractée. on notera y cette variable cible [4].

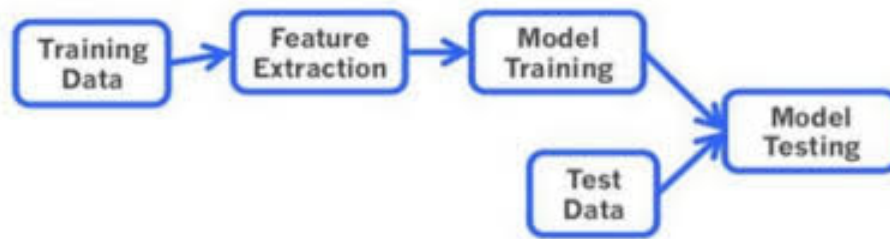


Fig. I.3. le processus typique du machine learning

En résumé, la valeur de la variable y dépend de :

- Une fonction $F(x)$ déterminée par les variables prédictives.
- Un bruit $\varepsilon(x)$ qui est le résultat d'un nombre de paramètres dont on ne peut pas tenir compte.

Aussi bien F que ε ne seront jamais connues mais l'objectif d'un modèle de ML est d'obtenir la meilleure approximation possible de F à partir des observations disponibles. Cette approximation sera notée f , on l'appelle la fonction de prédiction[4].

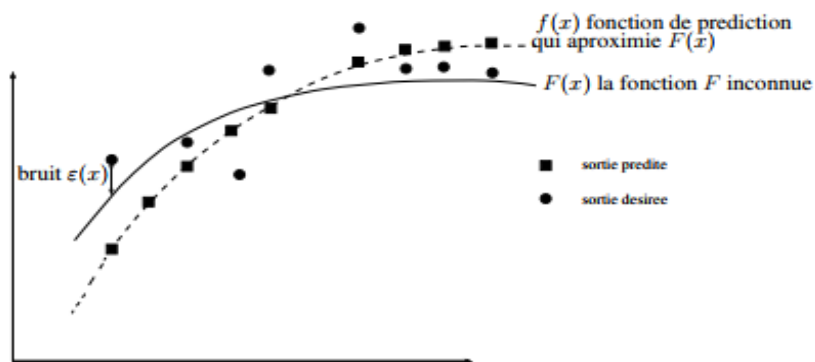


Fig I.4. une modèle de ML qui essaye d'obtenir la meilleure approximation possible de F

Voici quelques exemples d'utilisation du ML[4]:

- Vision par ordinateur
- Détection de fraude
- Classification (image, texte, vidéo, son ...)
- Les publicités ciblées
- Diagnostic médical

I.5.1.L'Apprentissage

L'apprentissage est le mécanisme par lequel le système nerveux s'adapte au monde extérieur. Les mécanismes précis de l'apprentissage par le cerveau restent très largement incompris et impliquent à la fois des modifications génomiques en partie sous le contrôle d'OTX2 (méthylations des citosines, rétrotransposons) , des modifications neuronales (augmentation de la surface d'échange des synapses, du nombre de récepteurs en post synaptique ou du nombre de neuromédiateurs en pré-synaptiques) mais aussi du microenvironnement où les pieds astrocytaires jouent également un rôle primordial.

L'apprentissage en machine learning est très probablement fondamentalement différent de celui du cerveau. En effet, aujourd'hui les principales techniques d'apprentissage sont dites supervisées, à savoir que l'on fournit à l'algorithme des vérités terrains .Notre cerveau lui semble apprendre de manière largement non supervisé. Une des définitions historiques de Tom Michels de l'apprentissage machine disait qu'un système était intelligent si pour une tâche T mesurée par une performance P, P s'améliore à la tâche T avec l'expérience E. Les récentes avancées en calcul parallèle ont permis dans de nombreuses applications visuelles d'égaliser voire de dépasser les capacités humaines pour la tâche qui leur est affectée. Néanmoins l'enjeu actuel est la

transférabilité de ces connaissances et de passer de réseaux dits discriminatifs à des réseaux génératifs en les intégrant dans des représentations plus larges que de simples tâches.

L'apprentissage machine se fait essentiellement par itérations successives en cherchant à minimiser l'erreur globale d'un modèle mesurée par sa fonction de coût. Ce type de problème se réfère à l'ensemble des techniques d'optimisation et plus particulièrement de minimisation. Une des techniques de minimisation utilisées en apprentissage machine dans notre travail c'est Réseaux Neurones Convolutionnels CNN.

I.5.2. Apprentissage des machines et ses applications a la biomédicale

Les chercheurs en intelligence artificielle visent toujours à programmer des machines capables d'effectuer des tâches qui requièrent de l'intelligence. Cependant, programmer des machines capables de s'adapter à toutes les situations et éventuellement d'évoluer en fonction de nouvelles contraintes est difficile. L'enjeu est de contourner cette difficulté en dotant la machine de capacités d'apprentissage lui permettant de tirer profit de son expérience. C'est pourquoi parallèlement aux recherches sur le raisonnement automatique se sont développées des recherches sur l'apprentissage par les machines en anglais « machine Learning ». Le principal objectif de ses recherches est la résolution automatique des problèmes complexes par la prise de décision sur la base d'observations de ces problèmes.

L'utilisation de l'apprentissage automatique pour les applications biomédicales connaît une augmentation considérable. Ce regain d'intérêt a plusieurs causes. D'une part, l'application réussie des techniques d'apprentissage automatique dans différents domaines D'autre part, le développement le plus récent est l'avènement des dossiers médicaux électroniques.

I.6. CATEGORIES DES MACHINES D'APPRENTISSAGE

I.6.1. Machines par séparation

I.6.1.1. Machines à Vecteurs de Support (SVMs)

Les Support Vector Machine (SVM) (machines à vecteur de support) sont des algorithmes de classification binaire non linéaire très puissant.

Le principe des SVM consiste à construire une bande séparatrice non linéaire de largeur maximale qui sépare deux ensembles d'observations et à l'utiliser pour faire des prédictions. l'astuce des SVM pour y parvenir consiste à utiliser une transformation ϕ non linéaire qui envoie les points $x(1), \dots, x(M)$ de l'espace original à n dimensions (n est le nombre de variable prédictives) vers des nouveaux points $\phi(x(1)), \dots, \phi(x(M))$ dans un espace de dimension plus grand que n ou ils seront plus facile à séparer

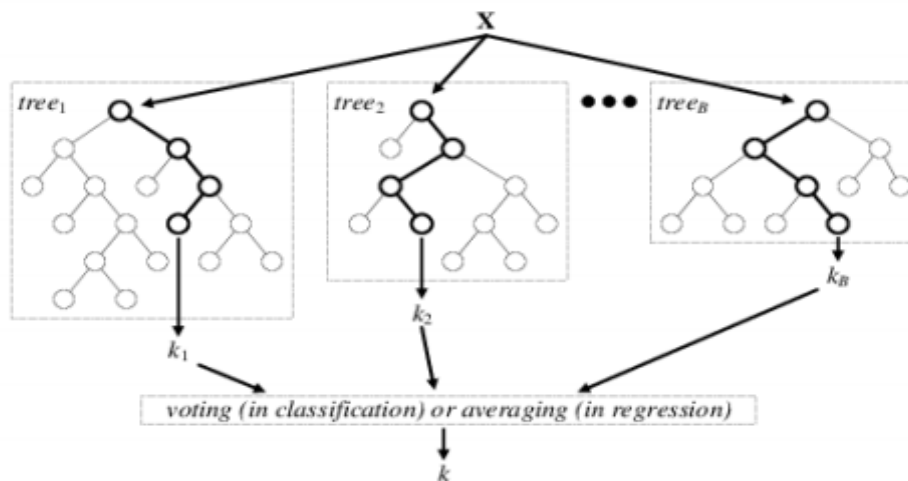


Fig I.5. Création de B bootstrap à partir des exemples d'apprentissage, chacun servira à entraîner un arbre de décision. Si c'est un cas de classification, la classe finale est attribuer par vote majoritaire. Si c'est un cas de régression, on fait la moyenne des prédictions

Les SVM sont des classificateurs qui reposent sur deux idées clés :

La première idée consiste à trouver un séparateur linéaire de largeur maximale, c'est la notion de marge maximale. La marge est la distance entre la frontière de séparation et les échantillons les plus proches. Ces derniers sont appelés vecteurs supports. Le problème est de trouver cette frontière séparatrice optimale.

Dans le cas où le problème est linéairement séparable, le choix de l'hyperplan séparateur n'est pas évident. Il existe en effet une infinité d'hyperplans séparateurs, dont les performances en phase d'apprentissage sont identiques, mais dont les performances en phase de test peuvent être très différentes.

Pour résoudre ce problème, il a été montré, qu'il existe un unique hyperplan optimal, défini comme l'hyperplan qui maximise la marge entre les échantillons et l'hyperplan séparateur.

Il existe des raisons théoriques à ce choix. Vapnik a montré que la capacité des classes d'hyperplans séparateurs diminue lorsque leur marge augmente.

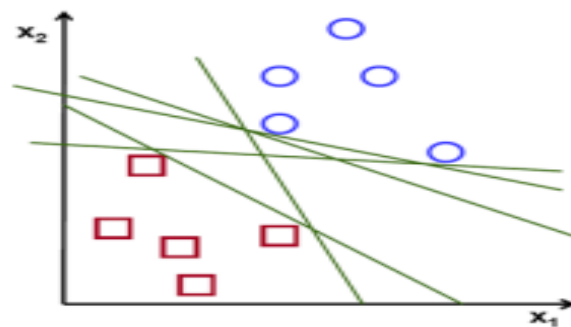


Fig I.6. On cherche un hyperplan qui divise les observations en deux catégories. Considérons un exemple x qu'on veut classer, si $f(x) > 0$, il appartient à la classe des cercles, sinon il appartient à la classe des carrés. Dans cette figure on peut voir qu'il existe une infinité d'hyperplans séparateurs possibles.

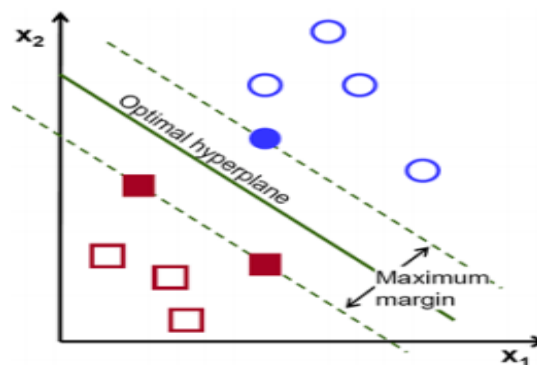


Fig I.7. L'hyperplan optimal (en vert) avec la marge maximale. Les échantillons remplis sont des vecteurs supports

Afin de pouvoir traiter des cas où les données ne sont pas linéairement séparables, la deuxième idée clé des SVM est de transformer l'espace de représentation des données d'entrées en un espace de plus grande dimension, dans lequel il est probable qu'il existe une séparation linéaire. Ceci est réalisé grâce à une fonction noyau, qui doit respecter les conditions du théorème de Mercer, et qui a l'avantage de ne pas nécessiter la connaissance explicite de la transformation à appliquer pour le changement d'espace.

Les fonctions noyaux permettent de transformer un produit scalaire dans un espace de grande dimension, ce qui est coûteux, en une simple évaluation ponctuelle d'une fonction. Cette technique est connue sous le nom de kernel trick.

Les deux fonctions noyaux les plus utilisées sont le noyau polynomial et le noyau gaussien

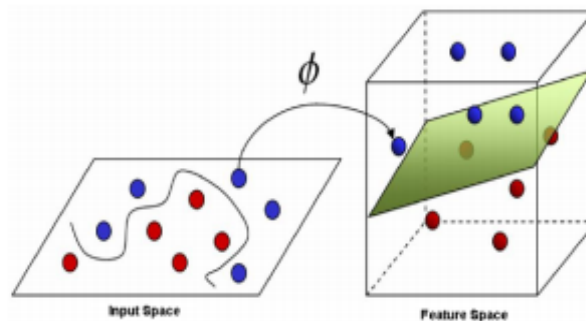


Fig I.8. Exemple d'un problème non linéairement séparable. La courbe devient une bande linéaire après avoir appliqué la transformation ϕ non linéaire

le SVM permet de traiter des problèmes de classification non linéaire complexe et aussi ils constituent une alternative aux réseaux de neurones car plus faciles à entraîner mais Les SVM sont souvent moins performants que les forêts aléatoires[4].

I.6.1.2 Les Réseaux de Neurones :

Les réseaux de neurones proposent une simulation du fonctionnement de la cellule nerveuse à

L'aide d'un automate.

Les réseaux neuronaux sont constitués d'un ensemble de neurones (nœuds) connectés entre eux par des liens qui permettent de propager les signaux de neurone à neurone.

Grâce à leur capacité d'apprentissage, les réseaux neuronaux permettent de découvrir des relations complexes non-linéaires entre un grand nombre de variables, sans intervention externe.

De ce fait, ils sont largement utilisés dans de nombreux problèmes de classification (ciblage marketing, reconnaissance de formes, traitement de signal,...) d'estimation (modélisation de phénomènes complexes,...) et prévision (bourse, ventes,...).

Il existe un compromis entre clarté du modèle et pouvoir prédictif.

Plus un modèle est simple, plus il sera facile à comprendre, mais moins il sera capable de prendre en compte des dépendances trop variées.

I.6.1.2.1. Présentation de quelques Types de réseaux neuronaux :

Il existe beaucoup de types de réseaux neurones, chaque type étant développé pour un objectif particulier[5].

- **Neurone Formel :**

Un neurone formel est une représentation mathématique et informatique d'un neurone biologique. Le neurone formel possède généralement plusieurs entrées et une sortie qui correspondent respectivement aux dendrites et au cône d'émergence du neurone biologique (point de départ de l'axone)[5].

Les actions excitatrices et inhibitrices des synapses sont représentées, la plupart du temps, par des coefficients numériques (les poids synaptiques) associés aux entrées[5].

Les valeurs numériques de ces coefficients sont ajustées dans une phase d'apprentissage. Dans sa version la plus simple, un neurone formel calcule la somme pondérée des entrées reçues, puis applique à cette valeur une fonction d'activation, généralement non linéaire[5].

La valeur finale obtenue est la sortie du neurone[5].

Le neurone formel est l'unité élémentaire des réseaux de neurones artificiels dans lesquels il est associé à ses semblables pour calculer des fonctions arbitrairement complexes, utilisées pour diverses applications en intelligence artificielle[5].

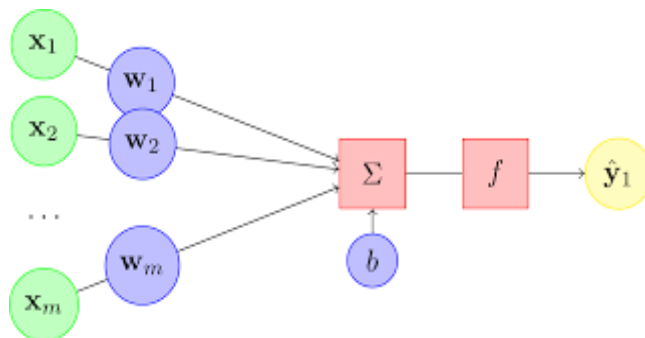


Fig I.9. La figure représente le fonctionnement d'un neurone formel.

- **Neurones multicouche :**

Le perceptron multicouche (multi layer perceptron MLP) est un classifieur linéaire de type réseau neuronal formel organisé en plusieurs couches (Figure I.9) au sein desquelles une information circule de la couche d'entrée vers la couche de sortie uniquement ; il s'agit donc d'un réseau de type feedforward (en). Chaque couche est constituée d'un

nombre variable de neurones, les neurones de la couche de sortie correspondant toujours aux sorties du système[5].

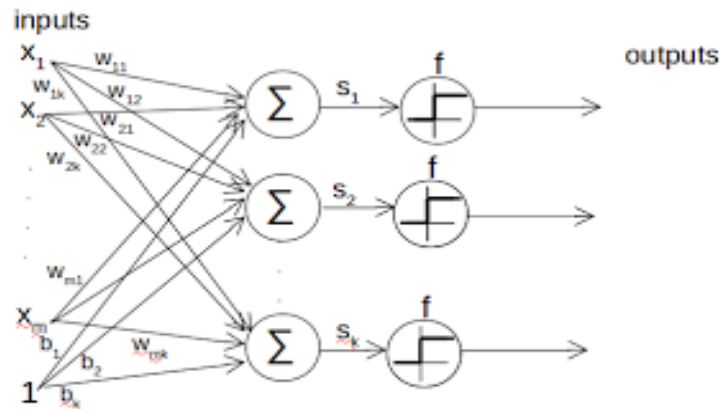


Fig I.10.Perceptron multicouche

- **Neurones récurrents :**

Les réseaux de Neurones récurrents (RNNs) permettent d'analyser les séquences de vecteurs tout comme les modèles de Markov cachés .Le temps entre ici en ligne de compte car les sorties (de la coche de sortie et/ou de la coche cachées) calculées à l'instant t sont ré-injectées en entrée du réseau et/ou en entrée de la couche cachée[5].

- **Réseaux de Hopfield :**

Le réseau de neurones d'Hopfield est un modèle de réseau de neurones récurrents à temps discret dont la matrice des connexions est symétrique et nulle sur la diagonale et où la dynamique est asynchrone (un seul neurone est mis à jour à chaque unité de temps). Il a été découvert par le physicien John Hopfield en 1982. Sa découverte a permis de relancer l'intérêt dans les réseaux de neurones qui s'était essoufflé durant les années 1970 à la suite d'un article de Marvin Minsky et Seymour Papert[5].

Un réseau de Hopfield est une mémoire adressable par son contenu : une forme mémorisée est retrouvée par une stabilisation du réseau, s'il a été stimulé par une partie adéquate de cette forme[5].

- **Réseaux Neurones Convolutionnels :**

En apprentissage automatique, un réseau de neurone convolutifs (ou réseau de neurones

à convolution, ou CNN ou ConvNet) est un type de réseau de neurones artificiels acycliques

dans lequel le motif de connexion entre les neurones est inspiré par le cortex visuel des

animaux. Les neurones de cette région du cerveau sont arrangés de sorte à ce qu'ils

correspondent à des régions qui se chevauchent lors du pavage du champ visuel. Leur

fonctionnement est inspiré par les processus biologiques, ils consistent en un empilage

multicouche de perceptrons, dont le but est de prétraiter de petites quantités d'informations[5].

Les réseaux neuronaux convolutifs ont de larges applications dans la reconnaissance d'un vecteur d'image et vidéo (uni, double et triple dimension), les systèmes de recommandation et le traitement du langage naturel. Ces caractéristiques font en sorte qu'elle soit une des techniques dont l'application peut être très adéquate pour la classification des données biomédicales, comme nous verrons dans le chapitre 2 où on va expliquer les CNN en détails[5].

I.6.1.2.2. Types de problèmes résolubles par réseau de neurones

Les réseaux de neurones permettent de résoudre deux grands types de problèmes :

- Les problèmes de classifications, où l'on cherche à prédire une classe, une catégorie. L'on distingue classiquement les classifications binaires (deux classes à prédire), des classifications "multi-class" où plusieurs classes sont à prédire[6].



Fig I.11. Exemple classification.

- Les problèmes de régressions, où l'on cherche à prédire une fonction continue[6].

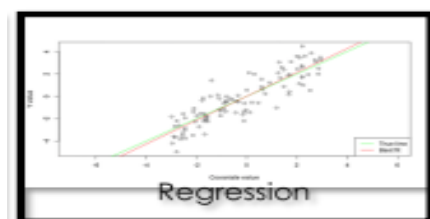


Fig I.12. Exemple Régression.

Cette distinction est importante car pour chaque type de problème, l'évaluation des performances de ce réseau se fera de manière différente. Cette mesure de performance, également appelée fonction de coût "cost-function", reflète de combien se trompe le réseau de neurone pour la tâche qui lui est affectée. Par exemple dans le cadre de problèmes de classifications la log-loss 1.3 ou l'entropie-croisée est classiquement utilisée.

(1-3)

J est la fonction de coût, M représente le nombre de classes à prédire, $y_{o,c}$, un indicateur binaire valant 0 ou 1 en fonction de si la classe c est correctement prédite pour l'observation o et $p_{o,c}$ la probabilité prédite pour chacune des classes de l'observation o . Ce genre de fonction de coût pénalise donc particulièrement les résultats qui se trompent et sont sûrs d'eux[6].

I.6.2. Machines par modélisation

I.6.2.1. Modèle de mélange gaussiens

Un modèle de mélange gaussien (MMG) est une catégorie de modèle probabiliste qui stipule que tous les points de données générés sont dérivés d'un mélange de distributions gaussiennes finies sans paramètres connus. Les paramètres des modèles de mélange gaussien sont dérivés d'une estimation maximum a posteriori ou d'un algorithme itératif d'anticipation-maximisation à partir d'un modèle antérieur bien entraîné. Les modèles de mélange gaussien sont très utiles pour modéliser des données, en particulier des données provenant de plusieurs groupes.

Les modèles de mélange gaussien sont utilisés dans les systèmes biométriques, où le modèle paramétrique aide à comprendre les caractéristiques ou les mesures associées à des caractéristiques telles que les caractéristiques spectrales du tractus vocal.

Les modèles de mélange gaussien sont également utilisés pour l'estimation de la densité et sont considérés comme les techniques les plus matures sur le plan statistique pour la classification.

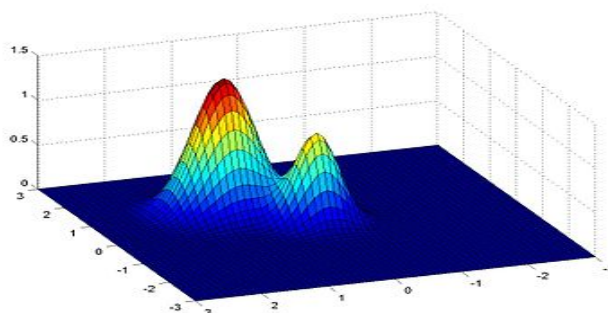


Fig. I.13. Le principe de MMG.

I.6.2.2. Chaines de Markov cachées

Les modèles de Markov cachées MMCs sont des modèles statistiques, riches et largement utilisés en traitement du signal. Ils sont développés par Andrew Markov (étudiant de Tchebychev), et ils sont premièrement, orientés vers des objectifs linguistiques dans des travaux de littérature Russe. Ceux sont des outils efficaces en modélisation des données séquentielles ou 'time-series Data'. Utilisés par la suite dans des problèmes de reconnaissance de la parole par Baker, leur théorie de base est introduite par Baum et ses collègues à la fin des années soixante.

Actuellement, ces modèles sont, de plus en plus adoptés en reconnaissance automatique de la parole, pour l'analyse des séquences ADN, et dans des problèmes liés à l'écriture et le traitement de texte. Aussi, leur utilisation en visionique est vaste. Ils sont implémentés en segmentation d'images, reconnaissance des visages, interprétation des gestes, ainsi que la modélisation de l'arrière-plan et récemment en traitement vidéo.

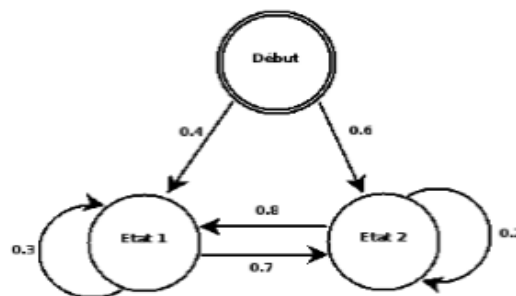


Fig. I.14. Le principe de MMC.

I.7.LES DIFFÉRENTES MÉTHODES DE LA CLASSIFICATION ET L'APPRENTISSAGE :

Méthodes supervisées :

Dans le cas de l'apprentissage supervisé, on dispose d'un ensemble de données étiquetées, ou d'exemples qui se sont vus associés une classe par un professeur ou un expert. Cet ensemble d'exemples constitue la base d'apprentissage[5].

Les méthodes d'apprentissage supervisé se donnent alors comme objectif général de construire à partir de la base d'apprentissage, ou fonctions de classement. Une telle fonction permet, à partir de la description d'un objet, de reconnaître un attribut particulier, la classe (Figure I.15.)[5].

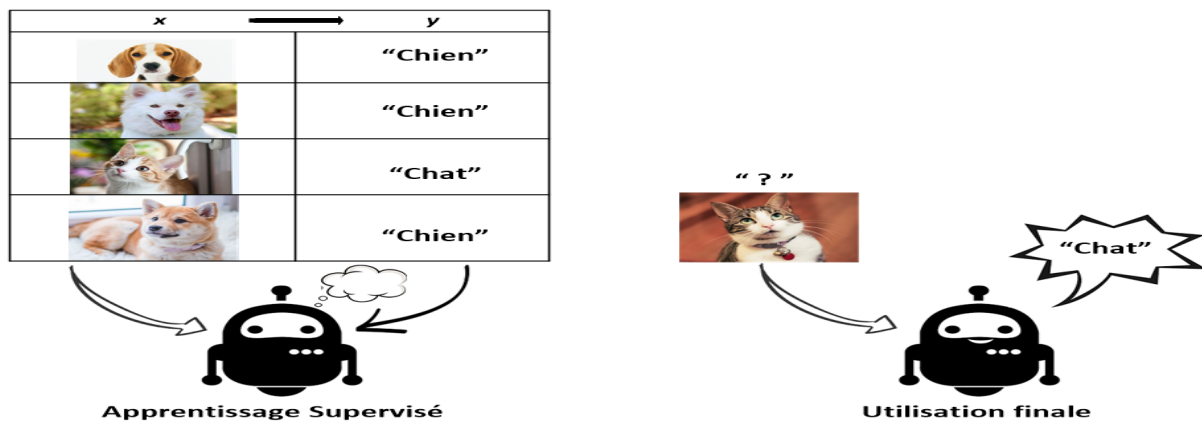


Fig I.15. l'aprontissage supervisees

Dans [(3), l'inférence inductive est définie comme un processus qui à partir d'une connaissance spécifique observée sur certains objets et d'une hypothèse inductive initiale, permet d'obtenir une assertion inductive impliquant ou rendant compte fortement ou faiblement des observations. Dans le cas de l'apprentissage inductif supervisé, qui est un sous domaine de l'inférence inductive, la connaissance spécifique consiste en un ensemble d'objets appartenant à des classes connues. L'assertion inductive est exprimée par une règle de classification qui assigne une classe à chaque objet. L'implication forte est satisfaite si la règle classe correctement tous les objets connus[5].

Méthodes non-supervisées :

L'apprentissage non-supervisé, encore appelé apprentissage à partir d'observations ou découverte, consiste à déterminer une classification« sensée » à partir d'un ensemble d'objets ou de situations données (des exemples non étiquetés)[5].

On dispose d'une masse de données indifférenciées, et l'on désire savoir si elles possèdent une quelconque structure de groupes. Il s'agit d'identifier une éventuelle tendance des données à être regroupées en classes. Ce type d'apprentissage, encore appelé Cluster ING ou Cluster Analysais, se trouve en classification automatique et en taxinomie numérique. Cette forme de classification existe depuis des temps immémoriaux. Elle concerne notamment les

sciences de la nature (**FigI.15**), les classifications des documents et des livres mais également la classification des sciences élaborées au cours des siècles par les philosophes [4] .

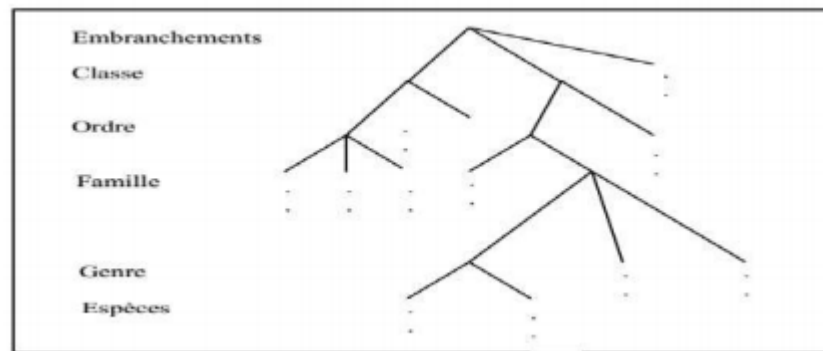


Fig I.16. Extrait de la classification taxinomique de Linné.

L'automatisation de la construction de classification constitue aujourd'hui un véritable domaine de recherche. La notion clé utilisée pour créer des classes d'objets est une mesure de la similarité entre les objets. Les classes ou concepts sont construits de façon à maximiser la similarité intra-classes et à minimiser la similarité interclasses.

L'apprentissage non supervisé correspond également à la classification conceptuelle, où une collection d'objets forme une classe si cette classe peut être décrite par un concept, compte tenu d'un ensemble de concepts prédéfinis [5] .

Méthodes hiérarchiques :

En classification hiérarchique ascendante. Le procédé consiste à grouper les observations individuelles en classes par a partie de la même classe.

Les méthodes se distinguent par le choix de la distance entre les observations et la définition de la stratégie d'agrégation.

Dans l'algorithme de base, le calcul de la distance (il s'agit plus exactement d'une quantité critère que l'on appelle distance par abus de langage) fait par récurrence à partir de la matrice des distances entre observations.

Méthodes non hiérarchiques :

La classification non hiérarchique ou partitionnement, aboutissant à la décomposition de l'ensemble de tous les individus en m ensemble disjoints ou classes d'équivalence, le nombre m de classes est fixé. Le résultat obtenu est alors une partition de l'ensemble des individus, un ensemble de parties, ou classes de l'ensemble I des individus telles que :

- Toute classe soit non vide
- Deux classes distinctes sont disjointes.

· Tout individu appartient à une classe. Cet algorithme porte le nom de "agrégation autour de centres variables". Une version légèrement différente, connue sous le nom de "nuées dynamiques" consiste à représenter chaque groupe non pas par son centre, mais par un ensemble de points (noyau) choisis aléatoirement à l'intérieur de chaque groupe.

On calcule alors une distance "moyenne" entre chaque observation et ces noyaux et l'on procède à l'affectation.

Méthodes non paramétriques :

Un classifié est dit non paramétrique si aucune distribution statistique paramétrique n'est exploitée, seule la distance spectrale sera alors prise en compte. Cette catégorie comprend notamment les méthodes fondées sur la minimisation de distance (hyper boîte ou parallélépipédique, la distance minimale et la distance de Mahalanobis, K plus proches voisins, K-means, ISODATA, etc.), de nouvelles méthodes apparues récemment s'ajoutent à cette catégorie comme les réseaux neuronaux et les Machines à Support Vecteurs (SVM)[5].

Méthodes paramétriques :

Un classifieur est dit paramétrique s'il associe à la signature spectrale (ou profil) une distribution statistique connue, le plus fréquemment pour le traitement d'images, la loi normale ou multi normale. Cette association offre la possibilité d'affecter à chaque pixel une probabilité d'appartenance à une classe donnée[5].

Methods structurelles : Ce type de méthodes exploite des informations structurelles et contextuelle d'un objet, elles analysent l'objet en termes de ses composantes (primitives) et de leurs propriétés, on trouve par exemple l'analyse syntaxique d'une forme ou un objet à partir d'une grammaire, la distance d'arbres, la distance de graphes (isomorphismes de graphes, de sous-graphes, avec correction d'erreurs, etc.). Dans la méthode structurelle la classe se présente principalement sous la forme de petites régions rondes[5].

Méthodes Renforcement : Les données en entrée sont les mêmes que pour l'apprentissage supervisé, cependant l'apprentissage est guidé par l'environnement sous la forme de récompenses ou de pénalités données en fonction de l'erreur commise lors de l'apprentissage[5].

Méthode Semi-Supervisé : Les données d'entrée sont constituées d'exemples étiquetés et non étiquetés. Ce qui peut être très utile quand on a deux types de données, car cela permet de ne pas en laisser de côté et d'utiliser toute l'information[5].

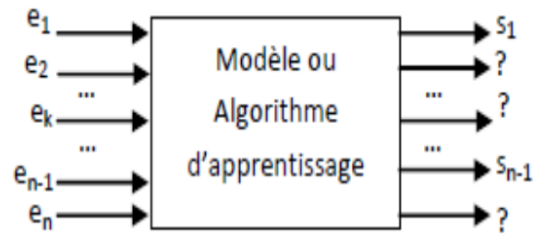


Fig. I.17. Schéma d'un modèle semi-supervisé.

I.8.CONCLUSION

Dans ce chapitre, nous avons présenté la relation de notre travail avec le dépistage médical et son importance sur la santé humaine, nous avons vu aussi des notions sur la biomédecine et sa relation avec les machines d'apprentissage (Machine Learning). Nous avons vu ce qu'est une machine d'apprentissage (ML) et nous avons montré sa relation avec l'amélioration des traitements et des décisions médicales (dépistage numérique) par les professionnels de la santé.

Nous avons ainsi présenté les différentes catégories des machines d'apprentissage et également les différents types d'apprentissage. Dans le prochain chapitre, nous allons voir le principe de l'apprentissage profond (Deep learning) et les réseaux de neurones convolutifs (CNNs)

Chapitre II
Systeme propose

II.1. INTRODUCTION :

Dans le cadre de notre travail, la solution proposée est de reconnaître la pathologie à travers des techniques de reconnaissance de formes en utilisant des techniques d'apprentissage en profondeur. Dans ce chapitre, nous allons d'abord introduire des concepts sur l'apprentissage en profondeur et ensuite nous commencerons à clarifier les différences entre l'apprentissage en profondeur et l'apprentissage automatique. Ensuite, nous verrons ce qu'est le Deep Learning (DL) et quelle est son utilisation et comment cela fonctionne, nous couvrirons également la partie programmation et son essence générale de l'application de la classification. Enfin, nous mentionnons en détail la plupart des types de DL c'est DNN, CNN, DFFnn et LSTM.

II.2. MACHINE LEARNING VS DEEP LEARNING

Beaucoup de gens s'inquiètent de l'IA, car ils ne pouvaient pas croire que les ordinateurs d'aujourd'hui puissent apprendre et prendre des décisions intelligentes. Cependant, les bases de l'intelligence artificielle sont aujourd'hui à la portée de tous .

L'apprentissage automatique ML et l'apprentissage profond DL sont les deux concepts les plus importants qui rendent l'intelligence artificielle possible. On confond bien souvent ces deux termes, alors qu'ils désignent deux méthodes bien distinctes employées dans des champs d'application différents .

Le ML et DL font partie de l'intelligence artificielle que les deux approches conduisent les ordinateurs à prendre des décisions intelligentes. et l'apprentissage en profondeur est une sous-catégorie de l'apprentissage automatique comme le montre l'image ci-dessous, l'utilisation du ML et du DL est devenue dans de nombreux domaines et conditions, dans lesquels la machine peut prendre des décisions intelligentes similaires à celles prises par les humains, les deux technologies nécessitent définitivement quantités de données, qu'il utilise comme base d'apprentissage, la similitude s'arrête là [7] .

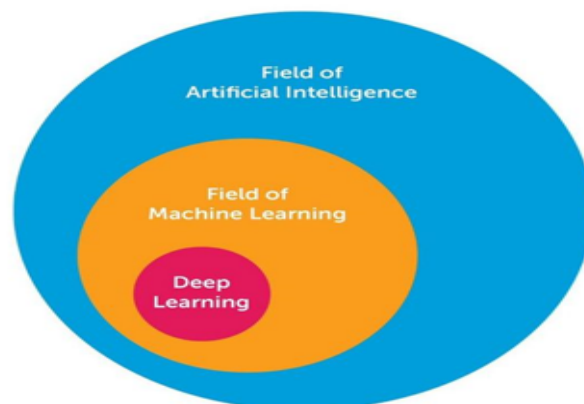


Fig.II.1.le relation entre AI, ML et DL

II.2.1. Quelles sont leurs différences ?

L'apprentissage automatique ou ML est la technique la plus ancienne et la plus simple que DL, elle dépend d'un algorithme qui adapte le système lui-même basé sur le réseau du cerveau humain , L'application de cette technologie implique la présence de données structurées et classifiées car le système est alimenté avec ces données qui permet lui de comprendre comment classer de nouvelles données similaires, Sur la base de cette classification, le système effectue alors les actions programmées, Il peut par exemple, déterminer si l'image montre un chien ou un chat , après la première étape d'utilisation, l'algorithme est amélioré sur la base des avis des développeurs, qui informent le système des notes incorrectes et indiquent les catégories correctes [7] .

Alors que l'apprentissage en profondeur ou DL ne nécessite pas de données structurées, car le système fonctionne sur plusieurs couches de réseaux de neurones qui combinent différents algorithmes qui dépendent du cerveau humain. Ainsi, le système peut fonctionner à partir de données non structurées , cette approche est particulièrement adaptée aux tâches complexes, lorsque tous les aspects des objets à traiter ne peuvent pas être classés dans le sens du courant. Le système d'apprentissage en profondeur définit lui-même les caractéristiques discriminatoires. Sur chaque couche, vous recherchez une nouvelle norme spécifique pour l'objet, qui sert de base pour décider du classement de la conservation de l'objet à la fin du processus .

De plus, la technologie requise pour l'apprentissage en profondeur est plus complexe. Il nécessite plus de ressources informatiques et est beaucoup plus cher que le machine learning, il n'est donc pas attractif, du moins pour le moment, pour une utilisation en groupe par les entreprises .

De plus, l'apprentissage en profondeur nécessite beaucoup plus de données, le système doit contenir plus de 100 millions d'entrées pour obtenir des résultats fiables, mais le ML fonctionne lorsqu'il y a une petite quantité de données pour obtenir des résultats fiables[7].

Dans le tableau ci-dessous **Tableau II.1**[7].

Nous avons mis les différences les plus importantes entre les techniques ML et DL :

<i>En terme de</i>	<i>Machine learning</i>	<i>Deep Learning</i>
Organisation des données	Données structurées	Données non structurées
quantité de données	hautes performances sur un petit et moyen ensemble de données	fonctionne très bien sur un big data
Ingénierie des fonctionnalités	comprendre soigneusement les caractéristiques de la façon dont il représente les données	nécessaire pour comprendre la meilleure fonctionnalité spécifique qui représente les données
entrainement	Entrainement par l'humain nécessaire	Système d'apprentissage autonome
algorithme	Algorithme modifiable	Réseauxdeneuronal d'algorithmes
matériel nécessaire	fonctionne sur une machine bas de gamme	de préférence nécessite une machine avec GPU
temps d'exécution	de quelques minutes à quelques heures	cela nécessite un temps allant jusqu'à 2-3 semaines
interopérabilité	certain algorithmes sont faciles à interpréter comme, la logistique et l'arbre de décision, et rares où comme certains sont presque difficiles comme SVM	peut être difficile ou impossible dans la plupart des cas
Champ d'application	Action simples de routine	Tache complexes

Tableau II.1.Synthèse les différences entre ML et DL

II.2.2.Les champs d'application

On peut considérer le Machine learning comme une technologie précurseur du Deep learning. Concrètement, toutes les tâches accomplies à l'aide du Machine learning peuvent être résolues

avec le Deep learning. Il ne faut donc pas nécessairement opposer apprentissage profond et automatique .

Le Deep learning mobilise beaucoup plus de ressources et n'est donc pas un processus efficace. Les champs d'application des deux technologies sont donc en principe bien délimités : toute tâche que le Machine learning peut exécuter doit être traitée par cette même technique .

Pour les entreprises, utiliser ces technologies représente un avantage concurrentiel énorme, car les deux apprentissages, automatique ou profond, ne sont pas encore la norme dans le quotidien professionnel[7].

II.2.2.1. Les champs d'application d'un Machine learning

Marketing en ligne : quelles mesures marketing apportent des résultats ? Les humains sont généralement mauvais pour passer en revue de grandes quantités de données et fournir des estimations fiables. Dans ce cas, il vaut mieux utiliser des outils d'analyse marketing qui s'appuient sur le Machine learning. Ils évaluent des données définies et peuvent fournir des diagnostics fiables à propos du type de contenu capable d'aboutir à une conversion, des contenus que les clients veulent lire et des canaux marketing les plus efficaces pour conclure une vente[7].

Support client : les chatbots peuvent s'appuyer sur le Machine learning. Ils s'orientent en fonction des mots-clés trouvés dans la question de l'utilisateur et, par des questions pour obtenir plus d'informations ou prendre des décisions, dialoguent avec l'utilisateur jusqu'à lui apporter la réponse désirée .

Vente : ce qui fonctionne pour Netflix et Amazon est aussi idéal pour la vente. Grâce au Machine learning, les systèmes peuvent anticiper avec précision les produits et services qui pourraient intéresser les clients sur leur site. Ils peuvent ainsi faire des recommandations détaillées, ce qui facilite la vente avec des gammes de produits très large ou des produits hautement personnalisables .

Informatique décisionnelle : le Machine learning peut aussi servir à visualiser les données importantes de l'entreprise et à rendre différentes prévisions compréhensibles pour les décideurs humains .

II.2.2.2. Les champs d'application d'un Deep learning

Sécurité informatique : contrairement aux solutions basées sur le Machine learning, les systèmes IT et de cyber-sécurité qui s'appuient sur le Deep learning peuvent identifier aussi

bien les dangers documentés que les risques jusqu'alors inconnus grâce à leur capacité à détecter les anomalies dans les patterns connus du réseau neuronal. Le Deep learning démultiplie l'efficacité des mesures de sécurité .

Support client : les chatbots basés sur le Deep learning comprennent l'expression naturelle des personnes et ne sont pas limités à l'utilisation de mots-clés précis. Le dialogue est clairement plus efficace et les solutions proposées plus susceptibles de répondre à la demande .

Création de contenu : le Deep learning peut servir à automatiser la création de contenu. À partir d'une base de données de contenus suffisamment fournie, le système peut créer un nouveau contenu ou effectuer des traductions en autonomie .

Assistant vocal : les assistants numériques, comme Siri, Alexa ou Google, reposent sur le Deep learning. Les premiers assistants numériques commencent à faire leur apparition en entreprise. Les utilisateurs peuvent s'exprimer naturellement pour demander, par exemple, d'abandonner une commande, d'envoyer un email, de créer un rapport ou de lancer une recherche .

Outre les champs d'application cités, les deux technologies s'utilisent dans de nombreux domaines du quotidien, comme la médecine, les sciences ou la mobilité[7].

II.3. POURQUOI LE DEEP LEARNING?

Les algorithmes de ML décrits dans la première partie fonctionnent bien pour une grande variété de problèmes. Cependant ils ont échoués à résoudre quelques problèmes majeurs de l'IA telle que la reconnaissance vocale et la reconnaissance d'objets.

Le développement du deep learning fut motivé en partie par l'échec des algorithmes traditionnels dans de telle tâche de l'IA .

Mais ce n'est qu'après que de plus grandes quantités de données ne soit disponibles grâce notamment au Big Data et aux objets connectés et que les machines de calcul soient devenues plus puissantes qu'on a pu comprendre le potentiel réel du Deep Learning .

Une des grandes différences entre le Deep Learning et les algorithmes de ML traditionnelles c'est qu'ils s'adaptent bien, plus la quantité de données fournie est grande plus les performances d'un algorithme de Deep Learning sont meilleurs. Contrairement à plusieurs algorithmes de ML classiques qui possèdent une borne supérieure à la quantité de données qu'ils peuvent recevoir des fois appelée "plateau de performance", les modèles de Deep Learning n'ont pas de telles limitations (théoriquement) et ils sont même allés jusqu'à dépasser la performance humaine dans des domaines comme l'image processing[8].

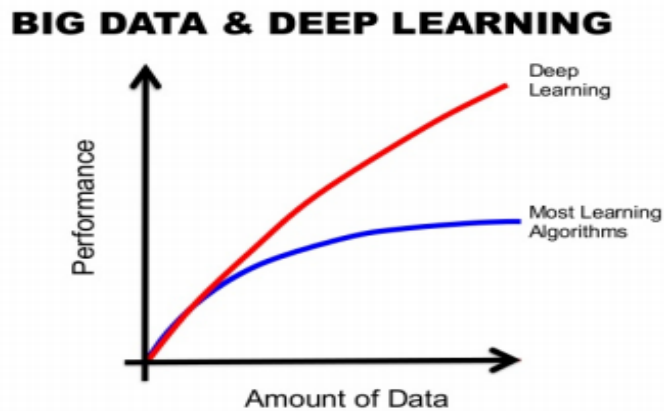


Fig. II.2. La différence de performance entre le Deep Learning et la plupart des algorithmes de ML en fonction de la quantité de données

Autre différence entre les algorithmes de ML traditionnelles et les algorithmes de Deep Learning c'est l'étape de l'extraction de caractéristiques. Dans les algorithmes de ML traditionnelles l'extraction de caractéristiques est faite manuellement, c'est une étape difficile et coûteuse en temps et requiert un spécialiste en la matière alors qu'en Deep Learning cette étape est exécutée automatiquement par l'algorithme [8].

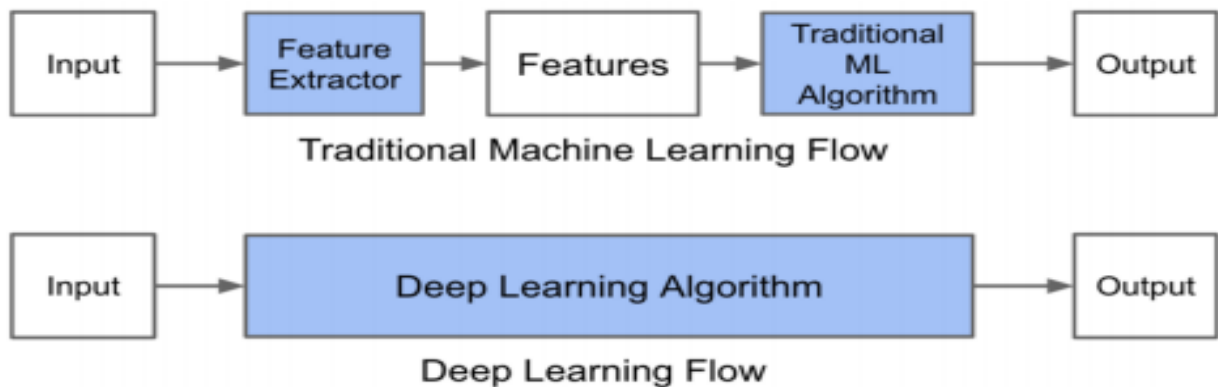


Fig. II.3. Le procédé du ML classique comparé à celui du Deep Learning

L'apprentissage en profondeur fournit un cadre très souple, universel (presque ?) Et pouvant être appris pour représenter des informations mondiales, visuelles et linguistiques, par le deep learning peut apprendre à la fois sans surveillance et sous surveillance.

II.4. DEFINITIONS DE DEEP LEARNING

L'apprentissage profond est un type d'apprentissage automatique (ML) et d'intelligence artificielle (AI) qui imite la façon dont les humains acquièrent certains types de connaissances, et c'est un élément important de la science des données, qui comprend des

statistiques prédictives et de la modélisation, et est également très utile pour les scientifiques chargés de collecter, d'analyser et d'interpréter de grandes quantités de données, plus rapidement et plus facilement .

Dans sa forme la plus simple, l'apprentissage en profondeur peut être considéré comme un moyen d'automatiser les analyses prédictives. Alors que les algorithmes d'apprentissage automatique traditionnels sont linéaires, les algorithmes d'apprentissage profond sont empilés dans une hiérarchie pour augmenter la complexité et l'abstraction, ce qui leur permet d'accomplir des tâches non linéaires et plus complexes [9].

II.4.1. Comment fonctionne Le Deep Learning

Les programmes informatiques qui utilisent Le Deep Learning passent par le même processus que l'apprentissage du jeune enfant pour identifier le chien. Chaque algorithme de la hiérarchie applique une transformation non linéaire à son entrée et utilise ce qu'il apprend pour créer un modèle statistique en sortie. Les itérations se poursuivent jusqu'à ce que la sortie ait atteint un niveau de précision acceptable. Le nombre de couches de traitement par lesquelles les données doivent passer est ce qui a inspiré le <<deep>> [9].

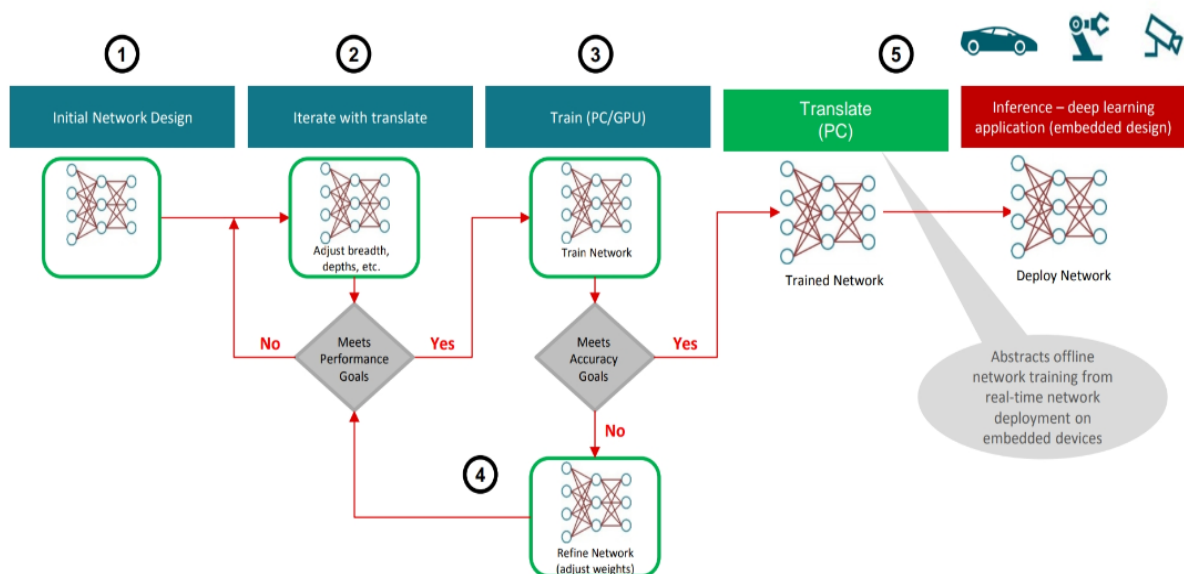


Fig. II.4. Système de reconnaissance de formes.

Dans l'apprentissage machine traditionnel, le processus d'apprentissage est supervisé, et le programmeur doit être extrêmement précis lorsqu'il indique à l'ordinateur le type de choses qu'il doit rechercher pour décider si une image contient un chien ou non. Il s'agit d'un processus laborieux appelé extraction de caractéristiques, et le taux de réussite de l'ordinateur dépend entièrement de la capacité du programmeur à définir avec précision un ensemble de

caractéristiques pour « chien ». L'avantage de Le Deep Learning est que le programme construit lui-même l'ensemble de caractéristiques sans supervision. L'apprentissage non supervisé est non seulement plus rapide, mais il est aussi généralement plus précis .

Au départ, le programme informatique peut recevoir des données d'entraînement – un ensemble d'images pour lesquelles un humain a étiqueté chaque image « chien » ou « pas chien » avec des méta-tags. Le programme utilise les informations qu'il reçoit des données d'entraînement pour créer un ensemble de caractéristiques pour « chien » et construire un modèle prédictif. Dans ce cas, le modèle que l'ordinateur crée en premier lieu peut prédire que tout ce qui se trouve dans une image qui a quatre pattes et une queue doit être étiqueté « chien ». Bien entendu, le programme n'est pas au courant des étiquettes « quatre pattes » ou « queue ». Il se contentera de rechercher des modèles de pixels dans les données numériques. À chaque itération, le modèle prédictif devient plus complexe et plus précis .

Contrairement au jeune enfant, qui mettra des semaines, voire des mois, à comprendre le concept de « chien », un programme informatique qui utilise des algorithmes d'apprentissage profond peut se voir montrer un ensemble de formation et trier des millions d'images, identifiant avec précision en quelques minutes les images qui contiennent des chiens .

Pour atteindre un niveau de précision acceptable, les programmes de Deep Learning nécessitent l'accès à d'immenses quantités de données d'entraînement et de puissance de traitement, qui n'étaient pas facilement accessibles aux programmeurs avant l'ère du big data et de l'informatique dématérialisée. Comme les programmes de deeplearning peuvent créer des modèles statistiques complexes directement à partir de leurs propres résultats itératifs, ils sont capables de créer des modèles prédictifs précis à partir de grandes quantités de données non étiquetées et non structurées. C'est important car l'internet des objets (IoT) continue à devenir plus omniprésent, car la plupart des données que les humains et les machines créent sont non structurées et ne sont pas étiquetées[9].

II.4.2. Quels sont les réseaux de neurones pour l'apprentissage en profondeur?

Un type d'algorithme d'apprentissage automatique avancé, connu sous le nom de réseaux de neurones artificiels, est à la base de la plupart des paradigmes d'apprentissage en profondeur, de sorte que l'apprentissage en profondeur peut parfois être appelé apprentissage neuronal profond ou réseau neuronal profond[9].

Les réseaux de neurones se présentent sous de nombreuses formes différentes, y compris les réseaux de neurones récurrents, les réseaux de neurones à convolution, les réseaux de

neurones artificiels et les réseaux de neurones pour une action précoce, chacun avec des cas d'utilisation spécifiques (nous verrons l'explication de chacun plus tard). Cependant, ils fonctionnent tous d'une manière quelque peu similaire, remplissent les données et laissent le formulaire déterminer par lui-même s'il a fourni la bonne interprétation ou pris la bonne décision pour une donnée particulière.

- Dans le cas de l'application (classification), nous avons généralement le chemin de notre programme informatique écrit par tous les algorithmes (DL) du début à la fin comme suit:
- Liste des bibliothèques utilisées
- Afficher nos données
- Division des données (formation et tests)
- Formation de notre modèle (création d'un modèle formé)
- tests
- Résultat (précision et perte avec courbe tracée pour chacun)

Les réseaux de neurones incluent un processus pour tester la quantité d'erreur résultant de la décision du réseau en tant que tel, il a donc besoin d'énormes quantités de données pour s'y entraîner.

Ce n'est pas un hasard si les réseaux de neurones ne sont devenus populaires qu'après que la plupart des entreprises ont adopté des analyses de données approfondies et l'accumulation de grandes quantités de données depuis les premières itérations du modèle incluent des hypothèses quelque peu formées sur le contenu d'une image ou de parties de discours, les données utilisées doivent être classées. Pendant la phase de formation afin que le modèle puisse voir si son hypothèse est correcte. Cela signifie que si de nombreuses entreprises qui utilisent de grandes quantités de données ont une grande quantité de données, les données non organisées sont moins utiles car les données non organisées ne peuvent être analysées via un modèle d'apprentissage en profondeur qu'après la formation et un niveau de précision acceptable, Alors que les modèles ML ne peuvent pas être formés aux données non structurées.

II.4.3. Méthodes d'apprentissage approfondi

Diverses méthodes peuvent être utilisées pour créer de puissants modèles d'apprentissage en profondeur. Ces techniques comprennent la réduction du taux d'apprentissage, le transfert de l'apprentissage, la formation à gratter et l'abandon .

- **Une diminution du taux d'apprentissage:** le taux d'apprentissage est un paramètre excessif - un facteur qui détermine le système ou corrige les conditions de son travail avant le

processus d'apprentissage, où il contrôle la taille du changement auquel le modèle est soumis en réponse à l'erreur estimée chaque fois que les poids du modèle sont modifiés. Des taux d'apprentissage trop élevés peuvent conduire à des processus d'apprentissage instables ou à un groupe de poids sous-optimaux, et des taux d'apprentissage très faibles peuvent conduire à un long processus de formation qui peut se boucher.

- **La méthode d'abaissement du taux d'apprentissage** : également appelée adoucissement du taux d'apprentissage ou taux d'apprentissage adaptatif - consiste à adapter le taux d'apprentissage pour augmenter les performances et réduire le temps de formation. Les ajustements les plus simples et les plus courants du taux d'apprentissage pendant la formation comprennent des techniques pour réduire le taux d'apprentissage au fil du temps.

- **Transfert d'apprentissage** : Ce processus implique la maîtrise d'un modèle déjà créé; Nécessite une interface d'un réseau existant. Tout d'abord, les utilisateurs alimentent le réseau existant avec de nouvelles données contenant des étiquettes qui étaient auparavant inconnues. Une fois les modifications apportées à la grille, de nouvelles tâches peuvent être mises en œuvre avec des capacités de classification plus spécifiques. Cette méthode a une fonctionnalité qui nécessite beaucoup moins de données que les autres, réduisant ainsi le temps de calcul à quelques minutes ou heures.

- **Entraînement à partir de zéro** : Cette méthode oblige le développeur à collecter une grande collection de données désagrégées et à construire une architecture réseau capable d'apprendre les fonctionnalités et le modèle. Cette technologie est particulièrement utile pour les nouvelles applications, ainsi que pour les applications qui ont un grand nombre de catégories de sortie. En général, c'est une approche moins courante, car elle nécessite des quantités excessives de données, ce qui fait que la formation prend des jours ou des semaines.

Pour améliorer le résultat (précision et perte), nous avons généralement essayé d'ajuster les facteurs suivants:

- Quantité de données pour la conduite de la formation (modification de la division entre formation et test)
- Le nombre de couches dans notre modèle
- Le nombre de neurones dans chaque couche
- Époques (nombre de sessions de formation)
- la valeur de Batch_size (afficher la vitesse de rotation)

II.4.4. Les applications de l'apprentissage profond

Comme les modèles d'apprentissage profond traitent l'information de manière similaire au cerveau humain, ils peuvent être appliqués à de nombreuses tâches que les gens accomplissent. Le Deep Learning est actuellement utilisé dans la plupart des outils de reconnaissance d'images, de traitement du langage naturel et des logiciels de reconnaissance vocale. Ainsi, en entraînant par apprentissage profond des algorithmes de reconnaissance faciale sur 200 millions d'images de visages, le système FaceNet de la société Google obtient un taux d'identification correcte de 99,63 p. 100. Le nombre d'applications potentielles de l'apprentissage profond est immense. C'est la raison pour laquelle cette méthode d'apprentissage s'est imposée ces dernières années. Ces techniques permettent d'améliorer la reconnaissance d'images en général et de créer des applications pour la biométrie (reconnaissance d'empreintes digitales ou d'iris), la médecine (avec, par exemple, le diagnostic de mélanomes à partir d'images de grains de beauté et l'analyse de radiographies), la voiture autonome (reconnaissance d'obstacles, de véhicules, de panneau de signalisation, etc.), par exemple. Elles permettent aussi d'améliorer la reconnaissance de la parole, avec des systèmes comme Siri, ou le profilage des individus, pour la recommandation et la publicité ciblée, ou encore les logiciels de jeux, comme on l'a vu en mars 2016 lorsque le programme informatique AlphaGo l'a emporté sur Lee Sedol, l'un des meilleurs joueurs mondiaux de go, en ayant fait appel à de l'apprentissage profond et à de l'apprentissage par renforcement. Enfin, et surtout, les techniques d'apprentissage supervisé aident à anticiper le futur sur la base du passé, ce qui permet d'évaluer, avec une précision inconnue auparavant, les risques potentiels d'investissements, d'accidents, de maladies, etc. Or, la prédiction aide à prendre des décisions en calculant les conséquences les plus probables de chaque action. De ce fait, les systèmes prédictifs recourant à de l'apprentissage profond jouent un rôle de plus en plus important dans le monde contemporain où on les utilise pour trancher dans les situations délicates à la place des hommes.

II.5. DIFFERENTS TYPES DE MODELES D'APPRENTISSAGE EN PROFONDEUR EXPLIQUES

Le Deep Learning est une classe de techniques d'apprentissage automatique qui exploitent de nombreuses couches de traitement non linéaire de l'information pour l'extraction et la transformation d'entités supervisées ou non, pour l'analyse et la classification de modèles. Il se

compose de nombreuses couches hiérarchiques pour traiter les informations de manière non linéaire, où un concept de niveau inférieur aide à définir les concepts de niveau supérieur. Les réseaux de neurones artificiels peu profonds ne sont pas capables de gérer une quantité importante de données complexes, qui sont évidentes dans de nombreuses applications de routine telles que la parole naturelle, les images, la récupération d'informations et d'autres applications de traitement de l'information de type humain. L'apprentissage en profondeur convient à de telles applications. Avec l'apprentissage en profondeur, il est possible de reconnaître, de classer et de classer les modèles de données pour une machine avec relativement moins d'efforts. Google est un pionnier de l'expérimentation du deeplearning initié par Andrew Ng.

Le Deep Learning offre un traitement multicouche de type humain par rapport à l'architecture peu profonde. L'idée de base de l'apprentissage en profondeur est d'utiliser un traitement hiérarchique utilisant de nombreuses couches d'architecture. Les couches d'architecture sont organisées de manière hiérarchique. Après plusieurs pré-entraînements, l'entrée de chaque couche va à sa couche adjacente. Le plus souvent, une telle pré-formation d'une couche sélectionnée exécutée de manière non supervisée.

Le Deep Learning suit une approche distribuée de la gestion des mégadonnées. La méthode suppose que les données sont générées en tenant compte de nombreux facteurs, de différents moments et de différents niveaux. Le deeplearning facilite l'agencement et le traitement des données en différentes couches selon leur durée (occurrence), leur échelle ou leur nature.

II.6. ARCHITECTURES D'APPRENTISSAGE EN PROFONDEUR

Le nombre d'architectures et d'algorithmes utilisés dans l'apprentissage en profondeur est important et varié, dans cette partie nous examinerons les cinq plus importantes des architectures du deeplearning, qui ont été l'une des méthodes les plus utilisées dans différentes applications au cours des vingt dernières années et sont encore utilisés encore aujourd'hui et fortement dans divers domaines[10].

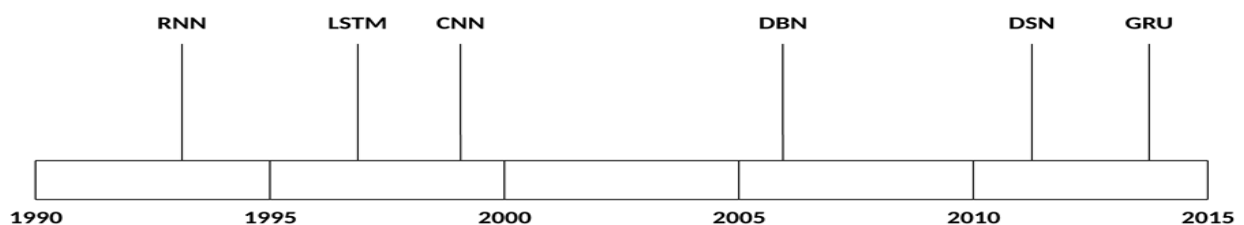


Fig.II.5.Chronologie graphique montrant le développement des cinq principales architectures d'apprentissage en profondeur par date, de 1990 à 2015

Dans le tableau suivant, nous mettons quelques applications pour chaque algorithme :

<i>Architecture</i>	<i>Application</i>
<i>DNN</i>	Usage général
<i>DFFnn</i>	vision par ordinateur et reconnaissance faciale
<i>CNN</i>	Reconnaissance d'images, analyse vidéo, traitement du langage naturel
<i>RNN</i>	en langage naturel, reconnaissance de
<i>LSTM</i>	l'écriture manuscrite, reconnaissance vocale, reconnaissance des gestes, sous-titrage des images

Tableau II.2.quelques applications pour un groupe d'architecteurs

Explorons maintenant ces architectures et méthodes utilisées pour les former.

II.6.1. RÉSEAU DE NEURONES PROFONDS __DNN__ :

Ces réseaux de neurones multicouches peuvent comporter des millions de neurones, répartis en plusieurs dizaines de couches. Ils sont utilisés en deeplearning pour concevoir des mécanismes d'apprentissage supervisés et non supervisés.

Dans ces architectures mathématiques, chaque neurone effectue des calculs simples mais les données d'entrées passent à travers plusieurs couches de calcul avant de produire une sortie. Les résultats de la première couche de neurones servent d'entrée au calcul de la couche suivante et ainsi de suite. Il est possible de jouer sur les différents paramètres de l'architecture du réseau : le nombre de couches, le type de chaque couche, le nombre de neurones qui composent chaque couche...etc.

Le perceptron multicouche, capable de traiter des phénomènes non linéaires, est un exemple de ce type de réseau. Concrètement, les premières couches permettront d'extraire des caractéristiques simples (comme des contours) que les couches suivantes combineront pour former des concepts de plus en plus complexes et abstraits : assemblages de contours en motifs, de motifs en parties d'objets, de parties d'objets en objets, etc.

L'espoir est que plus on augmente le nombre de couches, plus les réseaux de neurones apprennent des choses compliquées, abstraites, correspondant de plus en plus à la manière dont un humain raisonne. Il est néanmoins difficile de mettre au point des mécanismes d'apprentissage efficaces pour chacune des couches intermédiaires (appelées couches profondes ou couches cachées). La solution consiste à allier apprentissage supervisé (comme cela est fait pour un réseau à deux couches) et un apprentissage non-supervisé, utilisant des exemples dont la valeur de sortie est inconnue ,(inspiré d'un article d'Interstices) [11].

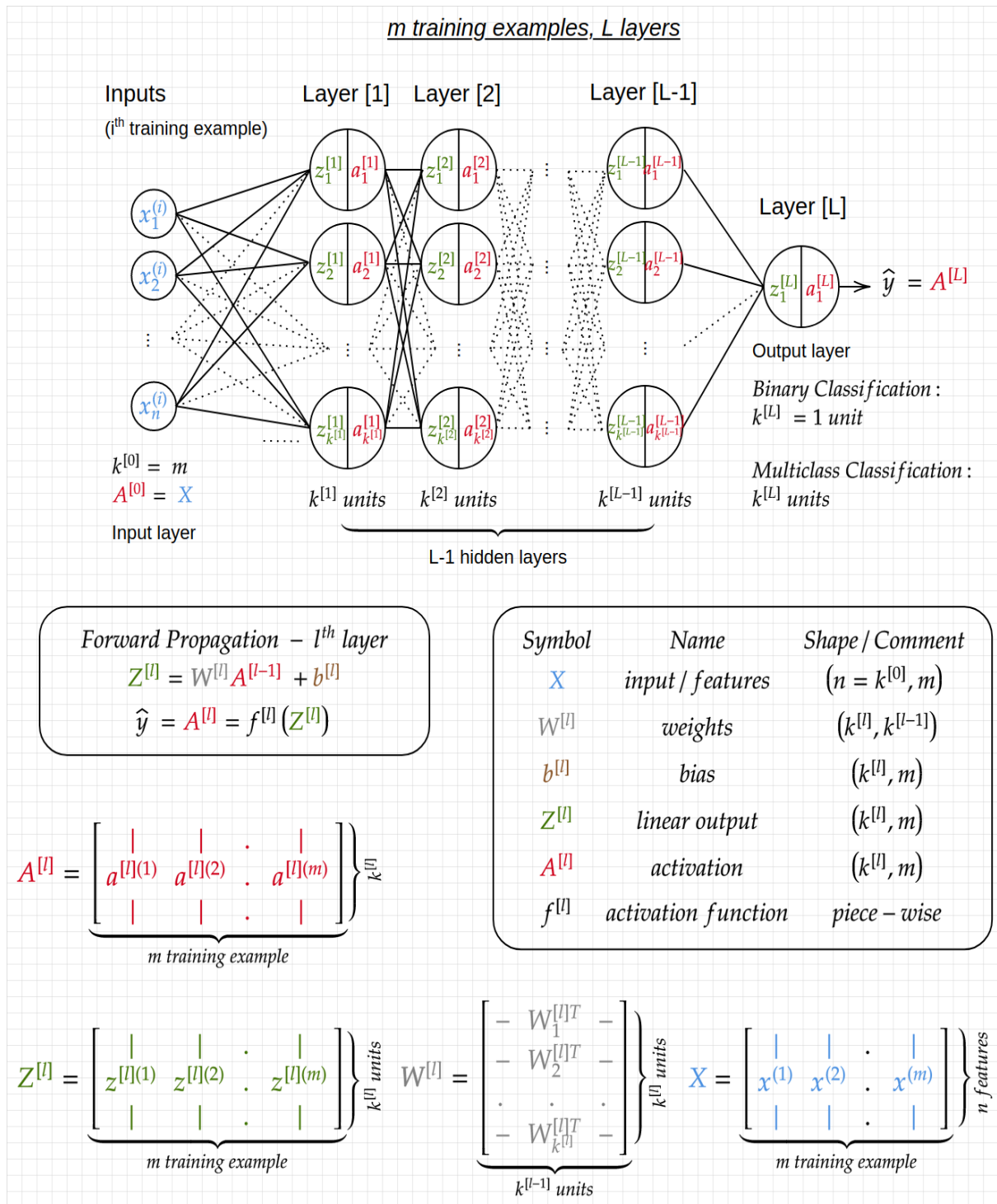


Fig.II.6.le fonctionnement d'un réseau de neurones profonds DNN .

II.6.2.RESEAU DE NEURONES CONVOLUTIONNELLS :

II.6.2.1.Historique :

Ces dernières années, les réseaux de neurones convolutifs (CNN) ont été largement utilisés dans diverses tâches de reconnaissance visuelle par ordinateur et ont obtenu de bons résultats

par rapport aux méthodes traditionnelles. La classification des images est l'une des tâches fondamentales et importantes de la reconnaissance visuelle, la méthode CNN étant également appliquée à d'autres tâches de reconnaissance visuelle (telles que la détection d'objets, la localisation d'objets et la segmentation sémantique ..etc), son architecture généralement dérivée de l'architecture de réseau humain dans la classification d'images[12] .

II.6.2.2.Principe générale

Nous allons voir ensemble comment fonctionne le réseau convolutif, en abordant brièvement le principal outil mathématique caché derrière, ainsi que toutes les étapes de l'analyse faite par l'algorithme. En tant qu'êtres humains, nous pouvons distinguer des images de formes et de couleurs différentes, mais l'ordinateur n'a pas cette capacité, Au contraire il a la capacité d'effectuer des centaines de calculs en une fraction de seconde, ce que l'esprit humain est incapable de faire, L'ordinateur ne voit l'univers que par des nombres, si nous les endommageons, L'image ci-dessous en tant qu'êtres humains nous la voyons comme une image contenant trois chiens, Quant à l'ordinateur il le voit comme une matrice tridimensionnelle $nw \times nh \times nc$ (nh : hauteur, nw : Largeur, nc : profondeur) Quant à la profondeur (nc) Il se compose de valeurs allant de 0 à 255 Ce sont simplement des pixels répartis sur la largeur et la largeur Quant à la profondeur elle nous renseigne sur le nombre de couleurs dans l'image (si $nc = 1$: l'image est grayscale ou $nc = 3$ l'image est RGB), Dans notre cas cette image est RGB[12].



Fig.II.7.La différence entre une vision par ordinateur et une vision humaine

Ainsi, les réseaux de neurones convolutions recevra cette image (figure 1), et un ensemble de processus lui sera appliqué pour identifier les formes qu'il contient aux fins d'utilisation dans une application particulière (figure 2), tel que ces processus sont divisés en deux catégories principales : Les filtres (l'extraction de l'information) : pour faire ressortir, par exemple, les arêtes verticales, les variations de coloration, etc..., sachant que plus les filtres interviennent

tard dans l'enchaînement, plus ils permettent de détecter des formes complexes (et plus abstraites) Les simplifications (l'analyse de l'information) : pour alléger les calculs et dégager les informations principales Ensuite, il donnera la probabilité que cette image représente un robot, un chien , un humain...etc ,grâce à un réseau de neurones artificiels ANN[12].

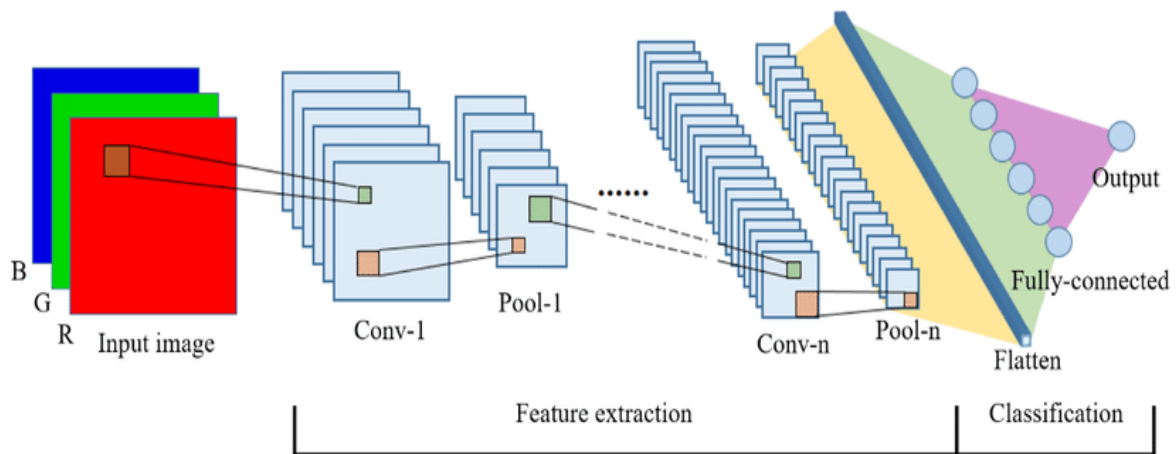


Fig.II.8.l'architecture général de la CNN

II.6.2.3.L'opération de convolution :

La convolution est le cœur du réseau de neurones convolutif et à l'origine c'est un **opération** mathématique basé sur la multiplication de la matrice des nombres avec un filtre (également connu sous le nom de kernel), qui déterminerait la présence de certaines caractéristiques ou motifs dans l'image d'origine (revenu), puis il peut être utilisé plusieurs filtres afin d'extraire Différentes fonctionnalités et le filtre est de petite taille pour effacer toute l'image et applique les calculs appropriés entre les valeurs du filtre et les points de couleur (pixels) dans l'image ,afin d'extraire des fonctionnalités Ce processus se traduit par une nouvelle matrice avec des dimensions inférieures par rapport à l'image d'origine [12].

La relation de la taille de l'image de sortie :

nh' : taille de l'image de sortie (largeur).

nh : taille de l'image d'entrée (largeur).

k : taille du kernel utilisé (largeur)

p : le padding appliqué

s : le stride appliqué

Le calcul est relativement simple et peut être effectué manuellement, mais plus la taille de l'image est grande, plus le filtre(kernel) est grand, plus il devient complexe, ce qui nécessite une intervention informatique. En supposant que nous ayons une matrice dimensionnelle 3×4 et un kernel 2×2 et le stride=1, le calcul sera comme indiqué dans l'image ci-dessous (figure 03). Ensuite, pour calculer la deuxième valeur de l'image de sortie, on va utiliser le paramètre stride de la convolution qui représente le nombre de pixels de l'image d'entrée que nous allons convertir pour réappliquer la chaîne. Mais il peut arriver que la convolution dépasse de l'image si on choisit mal le kernel et le stride, Par exemple si nous prenons le même exemple précédent, mais que nous modifions la valeur de stride à 3, nous rencontrerons un problème après le premier déformation. En outre vous remarquerez que le résultat de l'habillage est une image plus petite que l'image d'entrée. Pour éliminer ces deux problèmes, on a un paramètre « padding » qui est utilisé dans les convolutions et qui, simplement, ajoute des 0 autour de l'image d'entrée, pour augmenter la taille de l'image de sortie ou éviter les dépassements [12].

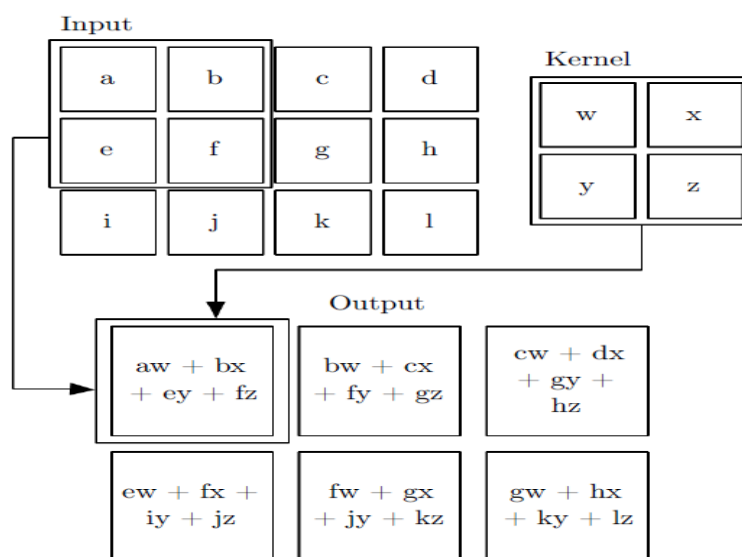


Fig.II.9.comment fonctionne la convolution

II.6.2.4. La couche ReLU (rectification linéaire) :

La fonction d'activation ReLU est la fonction dite de « correction » largement utilisée dans le deep learning. C'est plutôt l'une des fonctions les plus utilisées de nos jours en raison de ses caractéristiques la plus importante:

- 1- Le fait de ne pas activer tous les neurones en même temps permet également d'accélérer les calculs car les valeurs négatives sont mises à zéro, ce qui signifie que

toutes les valeurs égales ou inférieures à zéro deviennent nulles , tandis que les valeurs positives restent les mêmes .

- 2- Comme on l'a vu, une convolution va réaliser des opérations d'additions/multiplications : les valeurs en sorties sont donc linéaires par rapport à celles en entrée.
- 3- dans une image, la linéarité n'est pas très présente ni importante (par exemple, les variations entre valeurs de pixels peuvent être importantes dans une région, l'image a des coins...).
- 4- En ne modifiant pas les données positives, ReLU n'impacte pas les caractéristiques mises en évidence par la convolution, au contraire : elle les met davantage en évidence en creusant l'écart (valeurs négatives) « entre » deux caractéristiques (par exemple le nez et les yeux).

A noter qu'il existe d'autres ReLU comme par exemple LeakyReLU qui cassent moins la linéarité des données, même si ReLU est très largement préférée et utilisée [12].

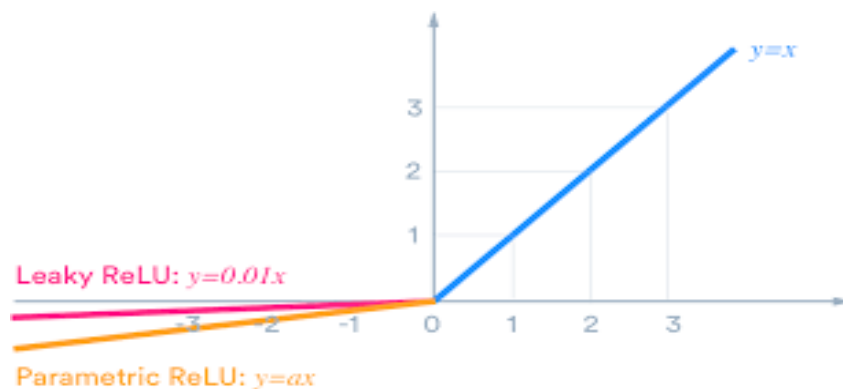


Fig.II.10.les fonctions d'activation ReLU et LeakyReLU et la différence entre elles

II.6.2.5.Le pooling :

Le pooling est une opération simple qui consiste à remplacer un carré de pixels (généralement 2×2 ou 3×3) par une valeur unique. De cette manière, l'image diminue en taille et se retrouve simplifiée (lissée) , Non seulement cela réduit le nombre de calculs nécessaires, mais cela empêche également le sur-ajustement[12].

Pour appliquer le pooling, on commence par sélectionner un carré de pixels de taille 2×2 puis on calcule la valeur qui va venir remplacer ce carré , Ensuite on décale ce carré vers la droite de 1 cases si le stride (= pas) vaut 1 par exemple (généralement, il vaut 1 ou 2) , Une fois arrivé au bout à droite, on recommence tout à gauche en décalant une fois vers le bas (d'un pas égal au stride) et en allant de nouveau vers la droite [13].

Il existe plusieurs types de pooling :

Le *maxpooling* : qui revient à prendre la valeur maximale de la sélection. C'est le type le plus utilisé car il est rapide à calculer (immédiat), et permet de simplifier efficacement l'image [12].

Le *meanpooling* (ou *averagepooling*) : soit la moyenne des pixels de la sélection : on calcule la somme de toutes les valeurs et on divise par le nombre de valeurs. On obtient ainsi une valeur intermédiaire pour représenter ce lot de pixels [12].

Le *sumpooling* : c'est la moyenne sans avoir divisé par le nombre de valeurs (on ne calcule que leur somme) [12].

La différence entre eux :

le **Mean** et le **sum** sont quasi-identiques , On peut simplement noter que le **sum** fait augmenter la valeur des pixels de l'image à creuser les écarts entre régions de l'image ce qui peut mettre en avant certaines caractéristiques , tandis que le **mean** permet de rester borné , tandis que le **max-pooling** va avoir tendance à retenir les caractéristiques (features) les plus marquées et simples de la sélection de pixels, comme par exemple une arête verticale. A l'inverse, mean étant une moyenne, seules les features moins marquées ressortiront . De manière générale, il est recommandé d'utiliser **max-pooling**, car il se distingue de mean-pooling sur les cas extrêmes et est quasiment équivalent à mean-pooling dans les autres cas [12].

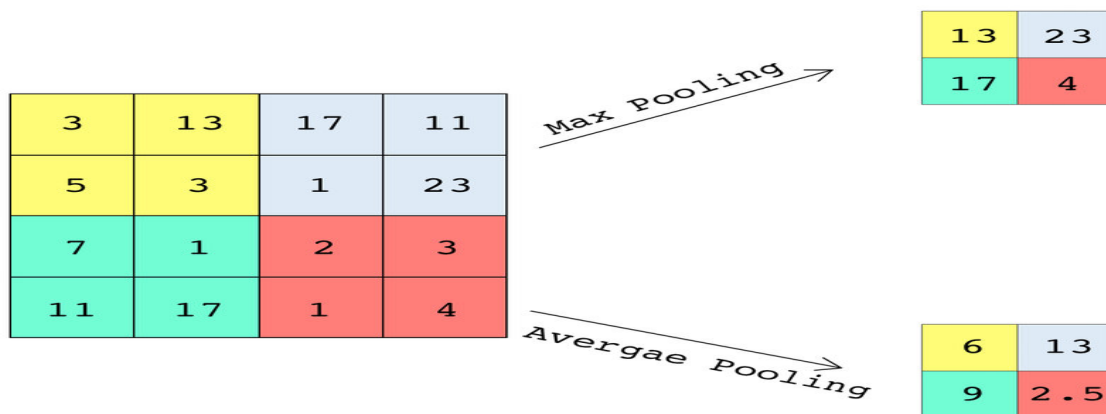


Fig.II.11. la différence entre Max pooling et Averagepooling

II.6.2.6. Le flattening (ou mise à plat) :

C'est la dernière étape de la partie « extraction des informations » L'entrée de cette couche est un matrices qui comprend des informations sur les emplacements des formes complexes spécifiques (motifs) dans l'image , Une opération d'aplatissement (flattening) fait de l'entrée (la matrice) que nous avons pour en faire un (long) vecteur de valeurs , chacune de ces valeurs représentant la probabilité d'un class de l'ensemble de classifications sur lesquelles le réseau est formé puis transmis à un réseau de neurones pour prédire la probabilité de la sortie [12].

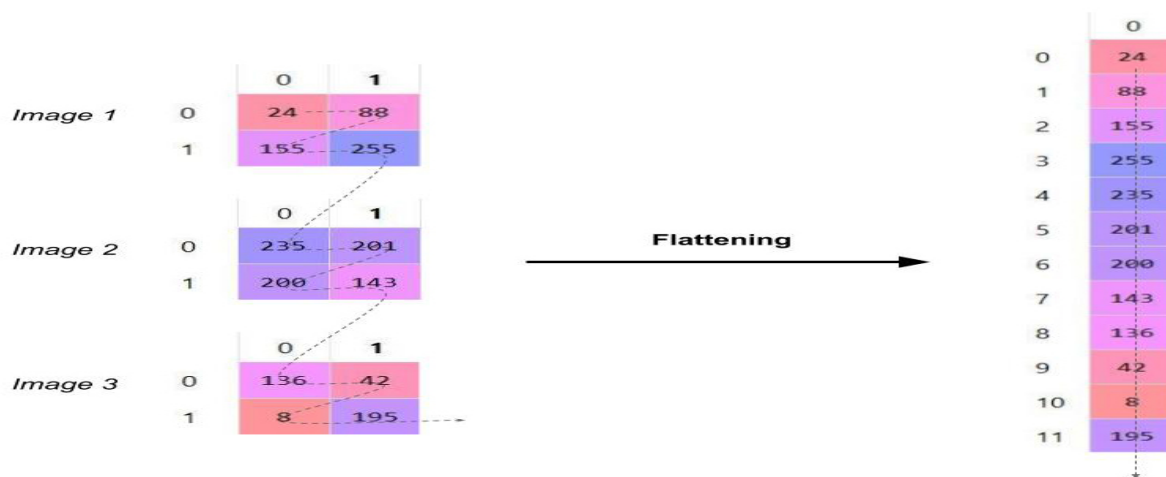


Fig.II.12.comment fonctionne le flattening

II.6.2.7.fully connected :

Les réseau de neurones artificiels est un réseau qui contient des couches ça a appelé fullyconnectedlayers or dense , ces couches qui reçoit un vecteur de valeurs et renvoie des probabilités pour chaque classe de prédiction , cette couches contient des neurones , Il existe différents types de ces neurones [12]:

les neurones d'entrée : qui reçoivent la somme des signaux pondérés de la couche de flattening qui envoient leur valeur à tous les neurones de la couche suivante , Dans le cas du CNN ce sera la valeur d'un pixel précis pour chaque neurone .

les neurones cachés : organisés en couches, qui vont envoyer la somme des signaux qu'ils reçoivent aux neurones de la couche suivante .

les neurones de sortie : qui reçoivent la somme des signaux pondérés de la dernière couche cachée qui prédire la probabilité de la sortie , et Chaque neurone de sortie représente alors une prédiction spécifique. Par exemple, le premier neurone représente la prédiction « c'est un chat », le deuxième « c'est un chien »... etc.

II.6.3. RESEAUX DE NEURONES A ACTION DIRECTE (DFFnn):

II.6.3.1. Historique :

En 1943, McCulloch et Pitts (1943) ont proposé un modèle de calcul inspiré du cerveau humain, qui a lancé la recherche sur le réseau de neurones artificiels (ANN), ceux-ci sont capables d'apprendre et de reconnaître et peuvent résoudre un large éventail de problèmes complexes, les réseaux de neurones Feedforward (FFNN) sont le type spécial de modèles ANN, et sa représentation structurelle le rend attrayant car il permet de percevoir un modèle de calcul (une fonction) dans une structure / réseau, forme De plus cette structure qui en fait un approximateur de fonction universel, qui a la capacité d'approximation de toute fonction continue (Hornik, 1991).

Par conséquent, les réseaux de neurones Feedforward (FFNN) peuvent résoudre un large éventail de problèmes, certains d'entre eux ci-dessous :

- comme la reconnaissance des formes (Jain et al., 2000).
- regroupement et classification (Zhang, 2000).
- approximation de la fonction (Selmic et Lewis, 2002).
- contrôle (Lamand Leung, 2006).
- bioinformatique (Mitra et Hayashi, 2006).
- traitement du signal (Niranjan et Principe, 1997).
- traitement de la parole (Gorin et Mammone, 1994).

La structure d'un FFNN se compose de plusieurs neurones (unités de traitement) disposés couche par couche et les neurones d'une couche ont des connexions (poids) à partir des neurones de sa couche précédente [15].

II.6.3.2. Principe générale :

Les réseaux de feedforward profonds, également souvent appelés réseaux neuronaux de feedforward, ou perceptrons multicouches (MLP), sont les modèles d'apprentissage profond par excellence, et cela s'appelle comme ça parce que les informations circulent uniquement à travers le réseau neuronal via les signaux d'entrée, puis vers les couches cachées, (une ou plusieurs couches) et enfin à travers la couche de sortie, et il n'y a pas de connexions de rétroaction dans lesquelles les sorties du modèle sont réinjectées en lui-même, lorsque les réseaux de neurones à action directe sont étendus pour inclure des connexions de rétroaction,

mais avant de parler de réseaux de neurones à action directe et comment cela fonctionne, comprenons pourquoi nous avons besoin de ce type de réseaux de neurones .

Le but d'un réseau à action directe est d'approximer une fonction \mathbf{f}^* , par exemple, pour un classificateur $\mathbf{y} = \mathbf{f}^* (\mathbf{x})$ mappe une entrée x à une catégorie y , un réseau à action directe définit un mappage $\mathbf{y} = \mathbf{f} (\mathbf{x}; \boldsymbol{\theta})$ et apprend la valeur des paramètres $\boldsymbol{\theta}$ qui donnent la meilleure approximation de fonction [16].

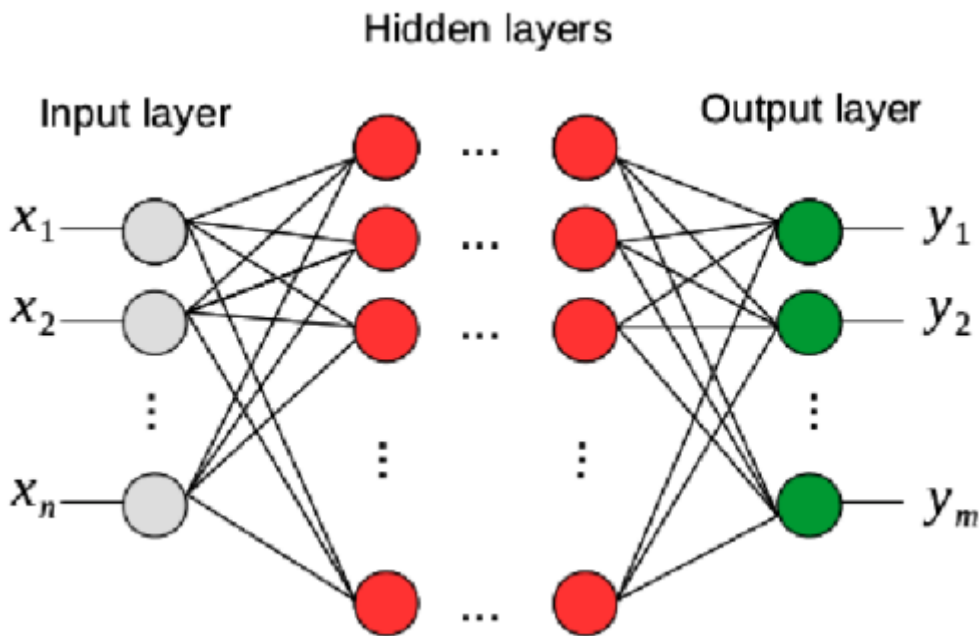


Fig.II.15. la distribution d'un Réseaux de neurones à action direct

II.6.3.3.Le perceptron (neurone) :

Le réseau neuronal est né d'un algorithme d'apprentissage machine très populaire nommé perceptron, ils ont été développés dans les années 1950 et 1960 par le scientifique Frank Rosenblatt, inspiré des travaux antérieurs de Warren McCulloch et Walter Pitts , les modèles traditionnels comme le perceptron (le perceptron est une unité fondamentale du réseau neuronal) qui prend des entrées réelles et donne une sortie booléenne ne fonctionne que si les données sont linéairement séparables.

Ensuite, nous avons le modèle de neurone sigmoïde qui est similaire au perceptron, mais le modèle sigmoïde est légèrement modifié de sorte que la sortie du neurone sigmoïde est beaucoup plus lisse que la sortie fonctionnelle par étape du perceptron. Bien que nous ayons introduit la fonction de neurone sigmoïde non linéaire, elle n'est toujours pas en mesure de séparer efficacement les données non linéaires, mais si nous connectons plusieurs neurones

sigmoïdes de manière efficace, nous pouvons approximer la combinaison de neurones à toute relation complexe entre l'entrée et la sortie, nécessaire pour traiter les données non linéaires [17].

La relation entre l'entrée et la sortie dans le cas d'un neurone :

x_i : Signaux d'entrée .

w_i : Les poids entre les signaux d'entrée et neurone .

b : le biais d'un neurone .

$f()$: la fonction d'activation d'un neurone .

y : la sortie (décision) .

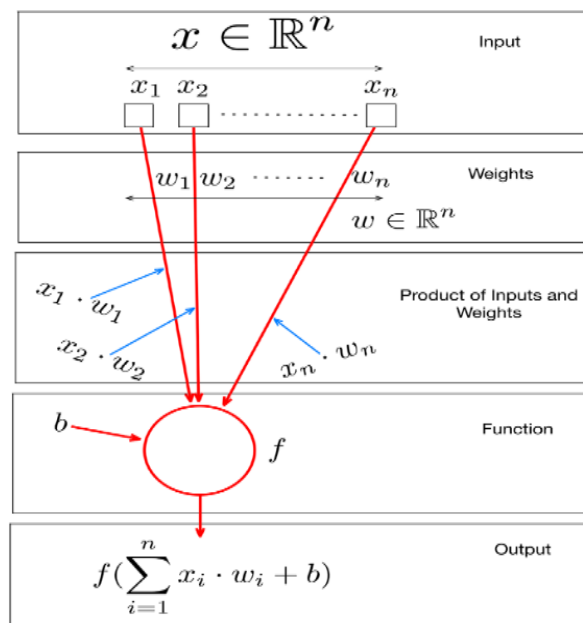


Fig.II.16.comment fonctionne une neurone

II.6.3.4. Réseaux de neurones artificiels :

le réseau neuronal multicouche est composé de nombreux neurones sigmoïdes et il est capable de gérer des données séparables non linéairement, les couches présentes entre les couches d'entrée et sortie sont appelées couches cachées et utilisées pour gérer les relations complexes non linéairement séparables entre entrée et sortie[14].

La relation entre l'entrée et la sortie dans le cas réseaux de neurones avec un seul couche cachée (hiddenlayer) :

xi :Signaux d'entrée dirigés vers la couche invisible .

xj :Signaux du couche invisible dirigés vers le couche de sortie .

wi: Les poids entre la couche d'entrée et la couche cachée .

wj: Les poids entre la couche cachée et la couche de sortie .

bj : le biais de la couche cachée .

bj : le biais de la couche de sortie .

g() : la fonction d'activation de la couche cachée .

f() : la fonction d'activation de la couche de sortie .

y : la sortie (décision) .

II.6.3.5. Softmax activation function :

la fonction Softmax est la fonction d'activation la plus fréquemment utilisée en apprentissage profond pour les problèmes de classification comme une fonction d'activation de sortie, ce qui la rend meilleure que d'autres dans ce cas, c'est que la sortie est toujours positive, quelle que soit l'entrée, en appliquant la fonction softmax nous obtiendrions une distribution de probabilité prédite et notre sortie réelle est également une distribution de probabilité, nous pouvons comparer ces deux distributions pour calculer la perte du réseau[17].

Si : $j = 1$ à k

II.6.3.6. loss_function :

Dans cette section, nous parlerons de la fonction de perte pour la classification binaire et multi-classes. Le but de cette fonction est d'indiquer au modèle qu'une correction doit être effectuée dans le processus d'apprentissage.

En général, le nombre de neurones dans la couche de sortie serait égal au nombre de classes, et dans le cas d'une classification binaire, nous ne pouvons utiliser qu'un seul neurone sigmoïde qui génère la probabilité $P(Y = 1)$ donc nous pouvons obtenir $P(Y = 0) = 1 - P(Y = 1)$, mais dans le cas de la classification, nous utiliserons la perte d'entropie croisée pour comparer la distribution de probabilité prédite et la distribution de probabilité réelle[17].

La perte d'entropie croisée pour la classification binaire est donnée par:

La perte d'entropie croisée pour la classification multi-classes est donnée par:

II.6.4. Réseaux de neurones récurrents RNN :

Le RNN est l'une des architectures de réseau de base à partir desquelles d'autres architectures d'apprentissage en profondeur sont construites. La principale différence entre un réseau multicouche typique et un réseau récurrent est que, plutôt que des connexions à rétroaction complète, un réseau récurrent peut avoir des connexions qui reviennent dans les couches précédentes (ou dans la même couche). Cette rétroaction permet aux RNN de conserver la mémoire des entrées passées et de modéliser les problèmes dans le temps.

Les RNN se composent d'un riche ensemble d'architectures (nous examinerons ensuite une topologie populaire appelée LSTM). Le différenciateur clé est la rétroaction au sein du réseau, qui pourrait se manifester à partir d'une couche cachée, de la couche de sortie ou d'une combinaison de celles-ci[10].

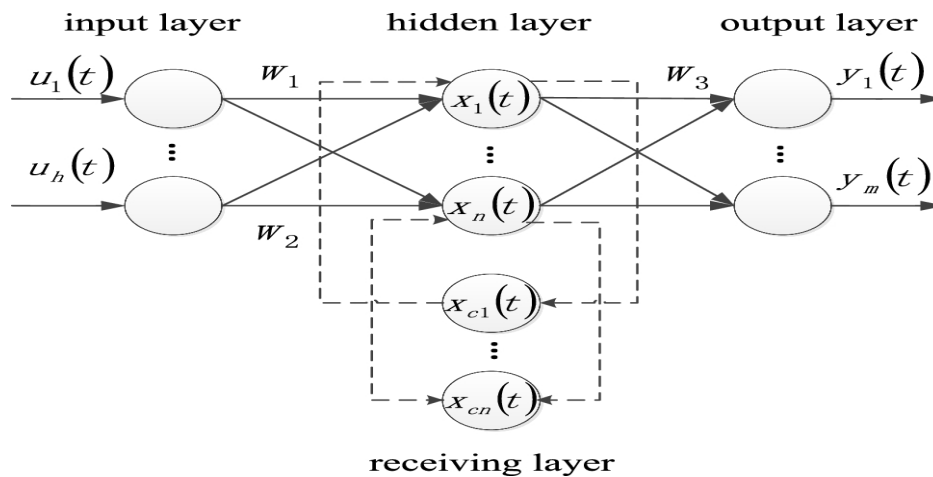


Fig.II.17.Image avec des cercles et des flèches démontrant l'interrelation entre les couches réseau d'entrée, de sortie, cachées et contextuelles

Les RNN peuvent être dépliés dans le temps et entraînés avec une rétro-propagation standard ou en utilisant une variante de rétro-propagation appelée rétro-propagation dans le temps (BPTT)[10].

II.6.5. Réseaux LSTM :

Le LSTM a été créé en 1997 par Hochreiter et Schmidhuber, mais il a gagné en popularité ces dernières années en tant qu'architecture RNN pour diverses applications. Vous trouverez des LSTM dans des produits que vous utilisez tous les jours, tels que les smartphones. IBM a

appliqué des LSTM dans IBM Watson® pour la reconnaissance vocale conversationnelle fixant des jalons.

Le LSTM s'est éloigné des architectures de réseaux neuronaux typiques basées sur les neurones et a plutôt introduit le concept d'une cellule de mémoire. La cellule mémoire peut conserver sa valeur pendant une courte ou longue durée en fonction de ses entrées, ce qui permet à la cellule de se souvenir de ce qui est important et pas seulement de sa dernière valeur calculée.

La cellule de mémoire LSTM contient trois portes qui contrôlent la façon dont les informations entrent ou sortent de la cellule. La porte d'entrée contrôle quand de nouvelles informations peuvent circuler dans la mémoire. La porte d'oubli contrôle quand une information existante est oubliée, permettant à la cellule de se souvenir de nouvelles données. Enfin, la porte de sortie contrôle le moment où les informations contenues dans la cellule sont utilisées dans la sortie de la cellule. La cellule contient également des poids qui contrôlent chaque porte. L'algorithme de formation, généralement BPTT, optimise ces poids en fonction de l'erreur de sortie réseau résultante[10].

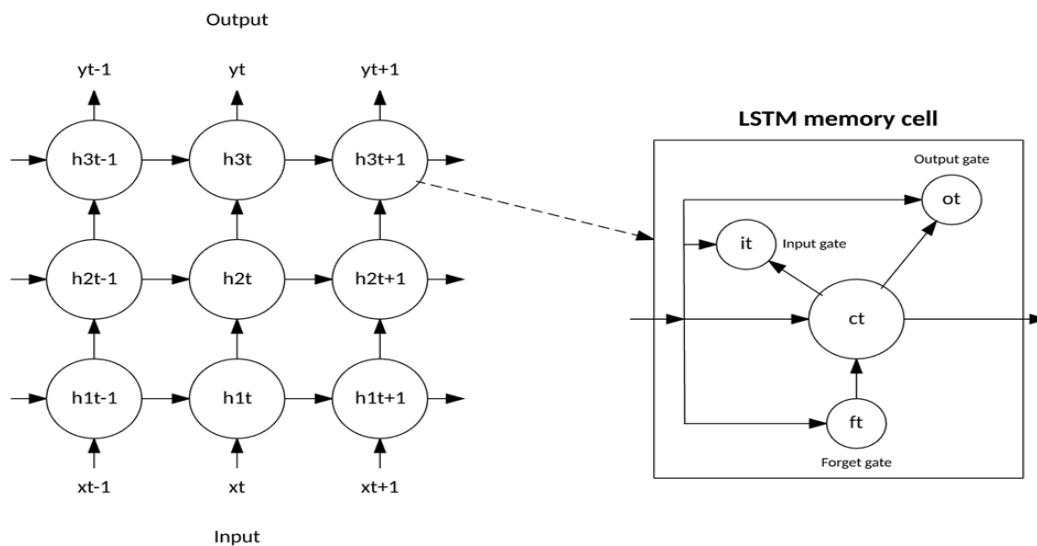


Fig.II.18.Image avec cercles et flèches montrant la cellule de mémoire LSTM et le flux d'informations à travers les différentes portes

En 2014, une simplification du LSTM a été introduite, appelée l'unité récurrente fermée. Ce modèle a deux portes, se débarrassant de la porte de sortie présente dans le modèle LSTM. Pour de nombreuses applications, le GRU a des performances similaires à celles du LSTM, mais sa simplicité signifie moins de poids et une exécution plus rapide.

Le GRU comprend deux portes: une porte de mise à jour et une porte de réinitialisation. La porte de mise à jour indique la quantité de contenu de cellule précédente à conserver. La porte

de réinitialisation définit comment incorporer la nouvelle entrée avec le contenu de cellule précédent. Un GRU peut modéliser un RNN standard simplement en définissant la porte de réinitialisation à 1 et la porte de mise à jour à 0.

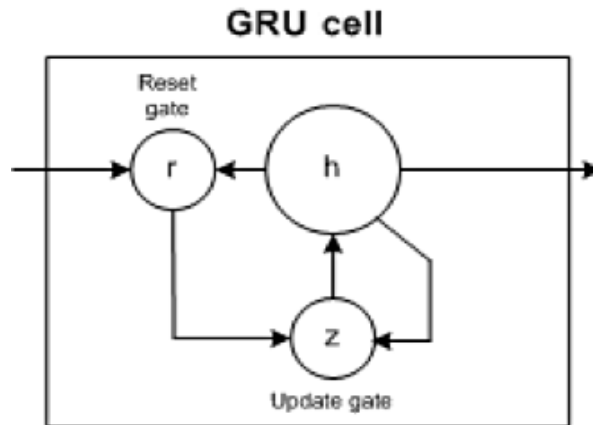


Fig.II.19.Diagramme d'une cellule GRU typique

Le GRU est plus simple que le LSTM, peut être formé plus rapidement et peut être plus efficace dans son exécution. Cependant, le LSTM peut être plus expressif et avec plus de données, peut conduire à de meilleurs résultats[10].

II.7. LES FONCTIONS D'ACTIVATION

- Elles permettent que chaque couche d'un réseau de neurones fonctionne de manière non linéaire.
- Il en existe un très grand nombre et elles peuvent être combinées.
- Une combinaison judicieuse de fonctions d'activation appropriées des couches de l'entrée vers la sortie améliore l'extraction des caractéristiques.
- Les fonction active sigmoïde et relu c'est le plus utilisees dans notre travail(pour l'application de classification) et la fonction softmax quand il y a plusieurs classe (chez nous le cas d'alzheimer) [22].

Exemples de fonctions d'activation (représentées par $g(x)$ avec $x =$ valeur de la couche d'entrée) :

II.7.1 Sigmoïde

- Une fonction sigmoïde transforme les variables en valeurs allant de 0 à 1 et elle est couramment utilisée pour générer une distribution de Bernoulli.

II.7.2 Tangente hyperbolique

- La dérivée $g' = (1 - g^2)$ facilite la procédure avec les algorithmes de rétro-propagation.

Elle est couramment utilisée dans la couche de sortie.

II.7.3 Softmax

- La sortie Softmax peut être considérée comme une distribution de probabilité sur les catégories.
- Elle est couramment utilisée dans la couche de sortie.

II.7.4 Unité linéaire rectifiée ("RectifiedLinear Unit", ReLU)

- Cette fonction d'activation (et ses variantes) est fréquemment utilisée en apprentissage profond.
- Fonction utilisée pour les couches cachées ("hiddenlayers") des réseaux de neurones.
- **Avantages et inconvénients :**
 - Coût de calcul faible et fonction facile à optimiser.
 - Convergence plus rapide que sigmoïde ou tanh.
 - Améliore la propagation du gradient.
 - Si $x < 0 \Rightarrow$ poids non actualisés \Rightarrow pas d'apprentissage.
 - Non différentiable à zéro.
 - Pas centrée en zéro et non bornée.

II.7.5 Variantes de ReLU

Softplus ou SmoothReLU :

Autres variantes de ReLU: LeakyReLUs, Noisy ReLUs, PReLUs ("ParametricReLUs"), ELUs ("ExponentialLinear Unit")

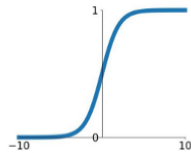
II.8.6 Maxout

La matrice de pondération de cette fonction est un tableau à 3 dimensions dans lequel la 3ème dimension correspond à la connexion entre couches voisines[22].

Activation Functions

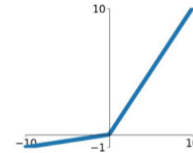
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



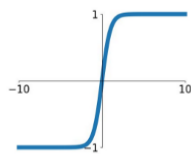
Leaky ReLU

$$\max(0.1x, x)$$



tanh

$$\tanh(x)$$

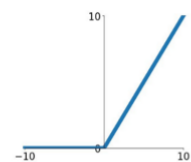


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ReLU

$$\max(0, x)$$



ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

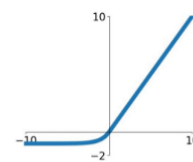


Fig. II.20. représentation d'un différent types de activation fonctions

II.8.CONCLUTION :

Dans ce chapitre, différentes techniques d'apprentissage en profondeur appliquées pour la classification qui constitue notre système de reconnaissance des maladies sont mises en évidence ,Nous avons fait une comparaison entre les techniques d'apprentissage en profondeur et les techniques d'apprentissage automatique que nous avons expliquées dans le chapitre précédent. Ensuite, nous avons mentionné ce qui rend l'apprentissage en profondeur unique à l'utilisation au lieu des techniques d'apprentissage automatique, et nous avons expliqué en détail quels sont les algorithmes d'apprentissage en profondeur les plus importants et comment ils fonctionnent.

Le chapitre suivant est entièrement consacré à la simulation, et est dédié à l'évaluation des performances des techniques de deep learning que nous avons étudiées dans ce chapitre, et l'objectif de cette étude est l'apport de ces méthodes en vue de leur application dans le domaine de l'identification des maladies, où de nombreuses expériences seront menées pour maximiser les performances du système et choisir la technique qui a la haute performance.

Chapitre III

Résultats expérimentaux

III.1. INTRODUCTION

Ce dernier chapitre est spécifiquement dédié à la mise en place et à l'évaluation de notre système en utilisant de nombreuses techniques d'apprentissage en profondeur (DNN, DFF, LSTM) et cela en les appliquant à différentes bases de données biomédicales réelles (cardiopathie, cancer du sein, diabète, Alzheimer) .

III.2. BASE DE DONNÉES UTILISÉE

Afin d'évaluer correctement notre système proposé, il existe quatre bases de données différentes utilisées pour extraire les performances globales, Ces données proviennent de sources biomédicales réelles, voici une description des propriétés de chaque base de données.

III.2.1. Heart Disease

Cet ensemble de données date de 1988 et se compose de quatre bases de données: Cleveland, Hongrie, Suisse et Long Beach V, Il contient 76 attributs, y compris l'attribut prédit, mais toutes les expériences publiées se réfèrent à l'utilisation d'un sous-ensemble de 14 d'entre eux. Le champ "cible" fait référence à la présence d'une maladie cardiaque chez le patient. Il est de valeur entière : **0 : pas de maladie et 1 : maladie**.

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
...
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

1025 rows x 14 columns

Fig.III.1. la forme des données des patients de cardiaques

N° de paramètre	Abréviation	signification
1	Age	âge en années
2	sexe	1 = male; 0 = femelle
3	cp	type de douleur thoracique
4	trestbps	pression artérielle au repos (en mm Hg à l'admission à l'hôpital)

5	chol	cholestérol sérique en mg/dl
6	fsb	glycémie à jeun > 120 mg / dl) (1 = vrai; 0 = faux
7	restecg	Résultats électro cardiographiques au repos
8	thalach	fréquence cardiaque maximale atteinte
9	exang	angine de poitrine induite par l'exercice (1 = oui; 0 = non)
10	oldpeak	Dépression ST induite par l'exercice relatif au repos
11	Slope	la pente du segment ST d'exercice de pointe
12	ca	nombre de vaisseaux majeurs (0-3)
13	thal	1 = normal; 2 = défaut fixe; 3 = défaut réversible
14	target	0 = pas de maladie et 1 = maladie

Tableau III.1. Informations d'attributs de la base de données des patients de cardiaques .

III.2.2. Breast cancer

Dans le monde, le cancer du sein est le type de cancer le plus répandu chez les femmes et le deuxième en termes de taux de mortalité , Le diagnostic de cancer du sein est effectué lorsqu'une bosse anormale est détectée (par auto-examen ou radiographie) ou un petit grain de le calcium est vu (sur une radiographie) , Après la découverte d'une bosse suspecte, le médecin procédera à un diagnostic pour déterminer si elle est cancéreuse et, dans l'affirmative, si elle s'est propagée à d'autres parties du corps.Cet ensemble de données sur le cancer du sein a été obtenu auprès des hôpitaux de l'Université du Wisconsin, Madison de Dr William H. Wolberg .

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
0	842302	M	17.99	10.38	122.80	1001.0	0.11840
1	842517	M	20.57	17.77	132.90	1326.0	0.08474
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960
3	84348301	M	11.42	20.38	77.58	386.1	0.14250
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030
...
564	926424	M	21.56	22.39	142.00	1479.0	0.11100
565	926682	M	20.13	28.25	131.20	1261.0	0.09780
566	926954	M	16.60	28.08	108.30	858.1	0.08455
567	927241	M	20.60	29.33	140.10	1265.0	0.11780
568	92751	B	7.76	24.54	47.92	181.0	0.05263

569 rows x 7 columns

Fig.III.2. la forme des données des patients de cancer du sein .

N° de paramètre	Signification	signification
0	id	ID number
1	mean_radius	moyenne des distances du centre aux points du périmètre
2	mean_texture	écart type des valeurs d'échelle de gris
3	mean_perimeter	taille moyenne de la tumeur centrale
4	mean_area	
5	mean_smoothness	erreur standard pour la variation locale des longueurs de rayon
6	diagnosis	Le diagnostic des tissus mammaires (M = malin, B = bénin)

Tableau III.2. Informations d'attributs de la base de données des patients de cancer du sein .

III.2. 3. diabètes

Cet ensemble de données provient à l'origine de l'Institut national du diabète et des maladies digestives et rénales , L'objectif de l'ensemble de données est de prédire par diagnostic si un patient souffre ou non de diabète, sur la base de certaines mesures diagnostiques incluses dans l'ensemble de données ,Plusieurs contraintes ont été placées sur la sélection de ces instances à partir d'une base de données plus importante. En particulier, tous les patients ici sont des femmes d'au moins 21 ans d'origine indienne Pima , cette données se composent de plusieurs variables prédictives médicales et d'une variable cible, le résultat. Les variables prédictives incluent le nombre de grossesses que la patiente a eues, son IMC, son niveau d'insuline, son âge, etc.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
338	9	152	78	34	171	34.2	0.893	33	1
334	1	95	60	18	58	23.9	0.260	22	0
316	3	99	80	11	64	19.3	0.284	30	0
105	1	126	56	29	152	28.7	0.801	21	0
36	11	138	76	0	0	33.2	0.420	35	0
...
257	2	114	68	22	0	28.7	0.092	25	0
108	3	83	58	31	18	34.3	0.336	25	0
25	10	125	70	26	115	31.1	0.205	41	1
371	0	118	64	23	89	0.0	1.731	21	0
481	0	123	88	37	0	35.2	0.197	29	0

768 rows x 9 columns

Fig.III.3. la forme des données des patients de diabète

N° de paramètre	Abréviation	signification
1	Pregnancies	Nombre de fois enceintes

2	Glucose	Concentration de glucose plasmatique à 2 heures dans un test de tolérance au glucose par voie orale
4	BloodPressure	Pression artérielle diastolique (mm Hg)
5	SkinThickness	Épaisseur du pli cutané du triceps (mm)
6	Insulin	Insuline sérique 2 heures (mu U / ml)
7	BMI	Indice de masse corporelle (poids en kg / (hauteur en m) ^ 2)
8	DiabetesPedigreeFunction	Fonction pedigree du diabète
9	Age	âge en années
10	Outcome	Variable de classe (0 ou 1) 268 sur 768 sont 1, les autres sont 0

Tableau III.3. Informations d'attributs de la base de données des patients de diabète.

III.2. 4. L'Alzheimer

L'Alzheimer est la cause la plus courante de démence car elle nuit aux compétences mentales et sociales, ce qui conduit à une perturbation plus précise du fonctionnement quotidien dans la vie normale, c'est aussi une dégénérescence des cellules cérébrales saines qui conduit à une baisse continue de la mémoire ainsi que des capacités intellectuelles , nous utiliserons les données fournies par le projet ADNI pour développer un modèle informatique qui aide au diagnostic de la maladie en utilisant un modèle de réseau de neurones profonds .

	DX.bl	AGE	PTGENDER	PTEDUCAT	PTETHCAT	PTRACCAT	APOE4	MMSE	imputed_genotype	APOE	Genotype
0	AD	81.3	Male	18	Not Hisp/Latino	White	1	20	True		3,4
1	LMCI	67.5	Male	10	Hisp/Latino	White	0	27	False		3,3
2	CN	73.7	Male	16	Not Hisp/Latino	White	0	29	True		3,3
3	LMCI	80.4	Female	13	Not Hisp/Latino	White	0	25	True		3,3
4	AD	73.9	Female	12	Not Hisp/Latino	White	1	24	True		3,4
...
623	LMCI	74.4	Female	12	Not Hisp/Latino	White	1	29	True		3,4
624	LMCI	69.4	Male	19	Not Hisp/Latino	White	1	29	True		3,4
625	LMCI	75.6	Female	16	Not Hisp/Latino	White	0	28	True		3,3
626	LMCI	83.4	Male	18	Not Hisp/Latino	White	0	26	True		3,3
627	LMCI	69.6	Female	12	Not Hisp/Latino	White	0	27	False		3,3

Fig.III.4. la forme des données des patients d'Alzheimer

N° de paramètre	Abréviation	signification
1	Age	Âge au départ
2	PTGENDER	Sexe
4	PTEDUCAT	Années d'études
5	PTETHCAT	Origine ethnique
6	PTRACCAT	Course
7	APOE4	Génotype APOE4
8	MMSE	Score MMSE au test cognitif
9	imputed_genotype	Désignation spécifique au défi , TRUE = a des génotypes imputés
10	DX.bl	Diagnostic à l'inclusion

Tableau III.4. Informations d'attributs de la base de données des patients d'Alzheimer.

III.3. PROTOCOLE EXPÉRIMENTAL

Après avoir décrit les caractéristiques de chaque base de données. Maintenant, l'opération suivante consiste à implémenter un modèle et obtenir des résultats sous forme de probabilité ou de classe. L'étape suivante consiste à déterminer l'efficacité du modèle basé sur une métrique utilisant des jeux de données de test. Différentes mesures de performance sont utilisées pour évaluer différents algorithmes d'apprentissage automatique, métrique pour l'évaluation d'algorithmes d'apprentissage automatique est la précision (en Anglais : *Accuracy*), erreurs, intervenons pour voir ce que sont ces métriques.

III.3.1. Taux de classification

Le taux de classification est le nombre de prédictions correctes faites par le modèle sur toutes les sortes de prédictions effectuées.

		Actual	
		Positives(1)	Negatives(0)
Predicted	Positives(1)	TP	FP
	Negatives(0)	FN	TN

Fig.III.5. Taux de classification illustré dans la matrice de confusion.

$$(III.1)$$

Dans le numérateur, nos prédictions correctes (vrais positifs et vrais négatifs) (marquées en rouge dans la figure ci-dessus) et dans le dénominateur, sont le genre de toutes les prédictions faites par l’algorithme (correctes ou non).

III.3.2. Taux d’erreur

L’erreur dans les problèmes de classification est le nombre de prédictions incorrectes faites par le modèle sur tous les types de prédictions effectuées.

		Actual	
		Positives(1)	Negatives(0)
Predicted	Positives(1)	TP	FP
	Negatives(0)	FN	TN

Fig. III.6. Taux d’erreur illustré dans la matrice de confusion.

$$(III.2)$$

Dans le numérateur, nos prédictions incorrectes (faux positifs et faux négatifs) (marquées en rouge dans la figure ci-dessus) et dans le dénominateur, sont le genre de toutes les prédictions faites par l’algorithme (correctes ou non).

III.4. RÉSULTATS DE SIMULATION

III.4.1. Avec la base Heart Disease

Dans cette partie, nous montrons les résultats obtenus avec la base de données des patients des cardiaques avec de nombreux modèles d'apprentissage en profondeur et également les résultats des années dernière avec de nombreux classificateurs d'apprentissage automatique (svm , knn , ...etc), le **Tableau III.5.** résume les performances de tous ces classificateurs .

Model	epochs	Accuracy	Loss	Time	Accuracy	Loss	Time
		Train %	Train %	train(us)	Test %	Test %	test(us)
<i>DNN</i>	200	90.98	26	31	98.53	10	39
<i>DFFnn</i>	200	81	47	4ms	82.9	36	1ms
<i>LSTM</i>	100	99.8	2	1ms	96.5	22	196
<i>SVM</i>					85		
<i>Decision Tree</i>					71		
<i>Logistic Regression</i>					81		
<i>Random Forest</i>					80		
<i>KNN</i>					87		

Tableau III.5. Résultats obtenus avec la base de données des patients des cardiaques.

Après avoir obtenu les résultats (accuracy et loss) avec tous les modèles, nous avons comparé ces résultats et comme nous le notons dans le tableau ci-dessus **Tableau III.5.** , les résultats que nous avons obtenus avec ces données en utilisant le modèle DNN (**accuracy_testing = 98.53** et **loss_testing =10**) C'est mieux que tous les résultats que nous avons obtenus avec les autres modèles et classificateurs, et les résultats étaient excellents comme le montre dans **Fig. III.7.** Et **Fig. III.8.** .

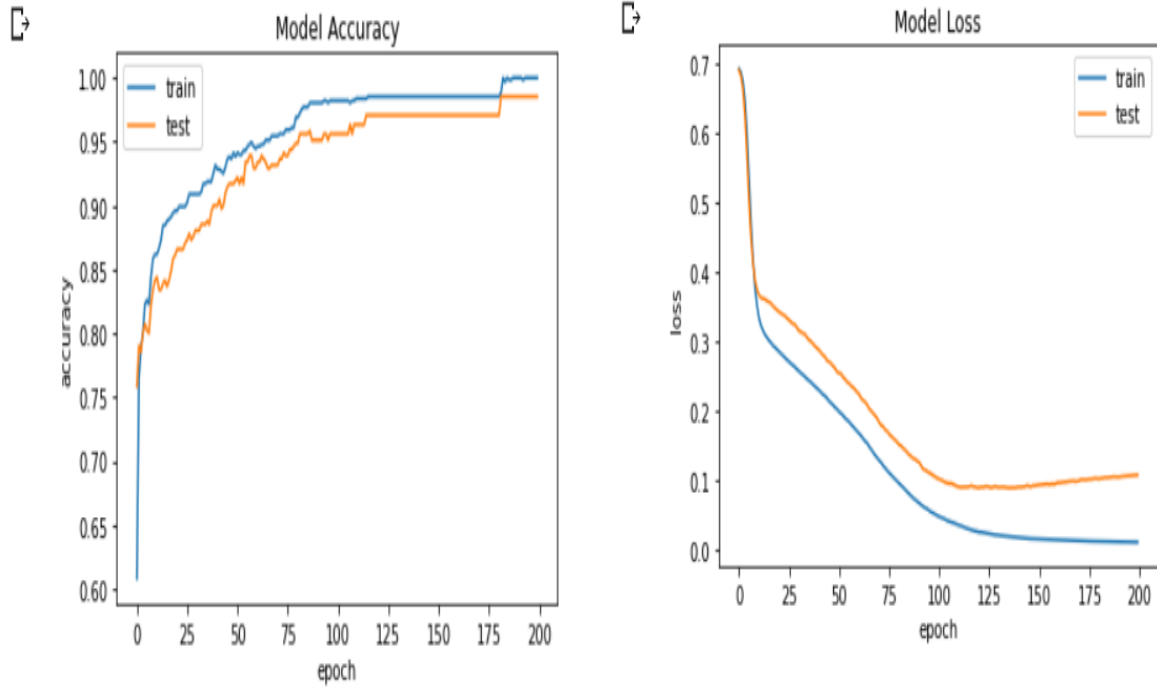


Fig. III.7. la différence entre accuracy et loss d'apprentissage et de test avec l'ensemble de données des patients des cardiaques avec le model DNN .

Nous avons également la matrice de confusion qui montre également combien notre modèle a été formé :

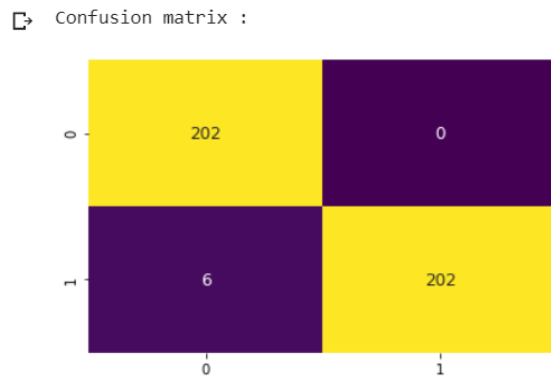


Fig. III.8. la matrice de confusion des patients des cardiaques avec le model DNN .

III.4.2. Avec la base Breast Cancer

Dans cette partie, nous montrons les résultats obtenus avec la base de données des patients de cancer du sein avec de nombreux modèles d'apprentissage en profondeur et également les résultats des années dernière avec de nombreux classificateurs d'apprentissage automatique (svm , knn , ...etc), le **Tableau III.6.** résume les performances de tous ces classificateurs .

Model	epochs	Accuracy	Loss	Time	Accuracy	Loss	Time
		Train %	Train %	train(us)	Test %	Test %	test(us)
<i>DNN</i>	200	99.6	0.7	33	100	0.28	91
<i>DFFnn</i>	200	95.5	12	3ms	92.9	14	2ms
<i>LSTM</i>	50	97	13	4ms	91.6	31	387
<i>SVM</i>					63		
<i>Decision Tree</i>					99		
<i>Logistic Regression</i>					64		
<i>Random Forest</i>					99		
<i>KNN</i>					72		

Tableau III.6. Résultats obtenus avec la base de données des patients de cancer du sein.

Après avoir obtenu les résultats (accuracy et loss) avec tous les modèles, nous avons comparé ces résultats et comme nous le notons dans le tableau ci-dessus **Tableau III.6.** , les résultats que nous avons obtenus avec ces données en utilisant le modèle DNN (**accuracy_testing = 100** et **loss_testing = 0,28**) C'est mieux que tous les résultats que nous avons obtenus avec les autres modèles et classificateurs, et les résultats étaient excellents comme le montre dans **Fig. III.9.** Et **Fig. III.10.** .

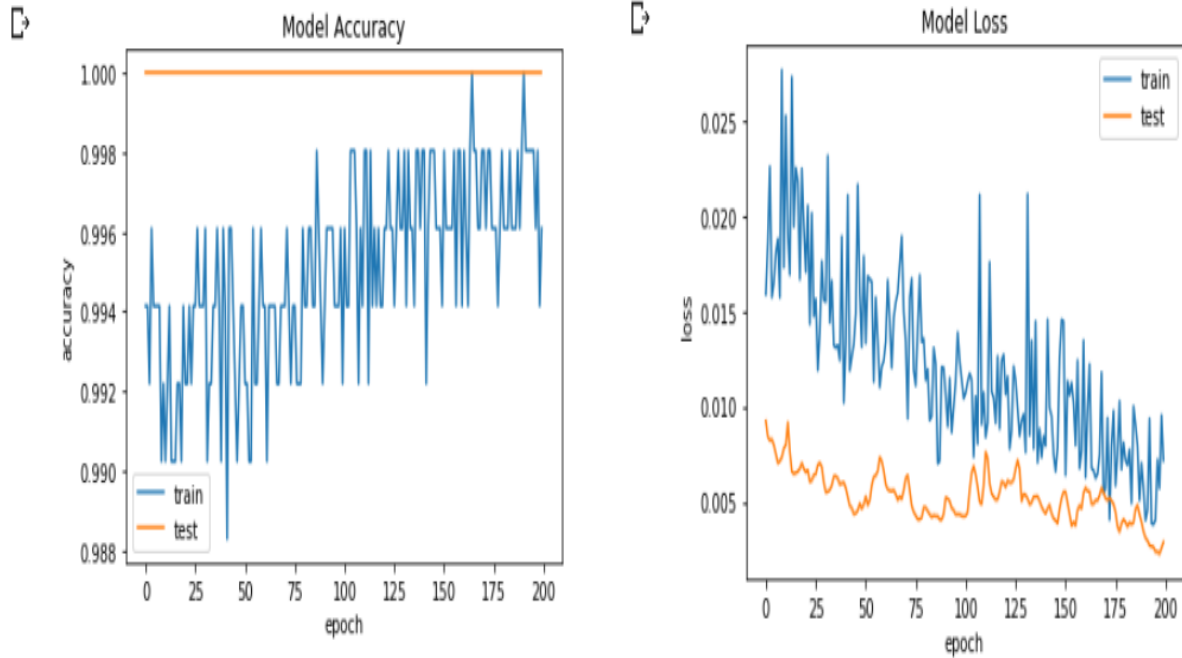


Fig. III.9. la différence entre accuracy et loss d'apprentissage et de test avec l'ensemble de données des patients de cancer du sein avec le model DNN .

Nous avons également la matrice de confusion qui montre également combien notre modèle a été formé :

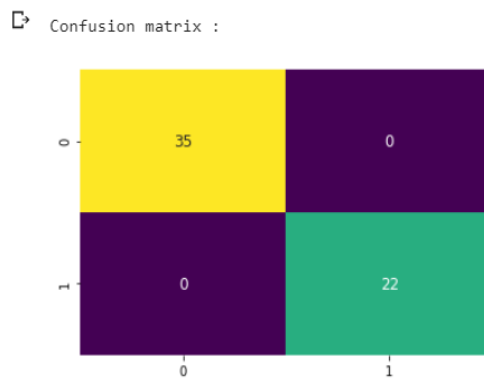


Fig. III.10. la matrice de confusion des patients de cancer du sein avec le model DNN .

III.4.3. Avec la base diabetes

Dans cette partie, nous montrons les résultats obtenus avec la base de données des patients du diabète avec de nombreux modèles d'apprentissage en profondeur et également les résultats des années dernière avec de nombreux classificateurs d'apprentissage automatique (svm , knn , ...etc), le **Tableau III.7.** résume les performances de tous ces classificateurs .

Model	epochs	Accuracy	Loss	Time	Accuracy	Loss	Time
		Train %	Train %	train(us)	Test %	Test %	test(us)
<i>DNN</i>	500	87	27	180	92	20	137
<i>DFFnn</i>	200	80.9	42	4ms	74	61	2ms
<i>LSTM</i>	50	85	31	307	67	59	796
<i>SVM</i>					77.6		
<i>Decision Tree</i>					100		
<i>Logistic Regression</i>					100		
<i>Random Forest</i>					100		
<i>KNN</i>					100		

Tableau III.7. Résultats obtenus avec la base de données des patients du diabète.

Après avoir obtenu les résultats (accuracy et loss) avec tous les modèles , nous avons comparé ces résultats et comme nous le notons dans le tableau ci-dessus **Tableau III.7.** , Les résultats que nous avons obtenus avec ces données avec le modèle DNN (**accuracy_testing = 92** et **loss_testing = 20**) C'était meilleur que les résultats que nous avons obtenus des modèles LSTM et DFF , alors qu'il était si proche des résultats avec les classificateurs d'apprentissage automatique (M_L) , les résultats sont bons comme le montre dans **Fig. III.11.** Et **Fig. III.12.**

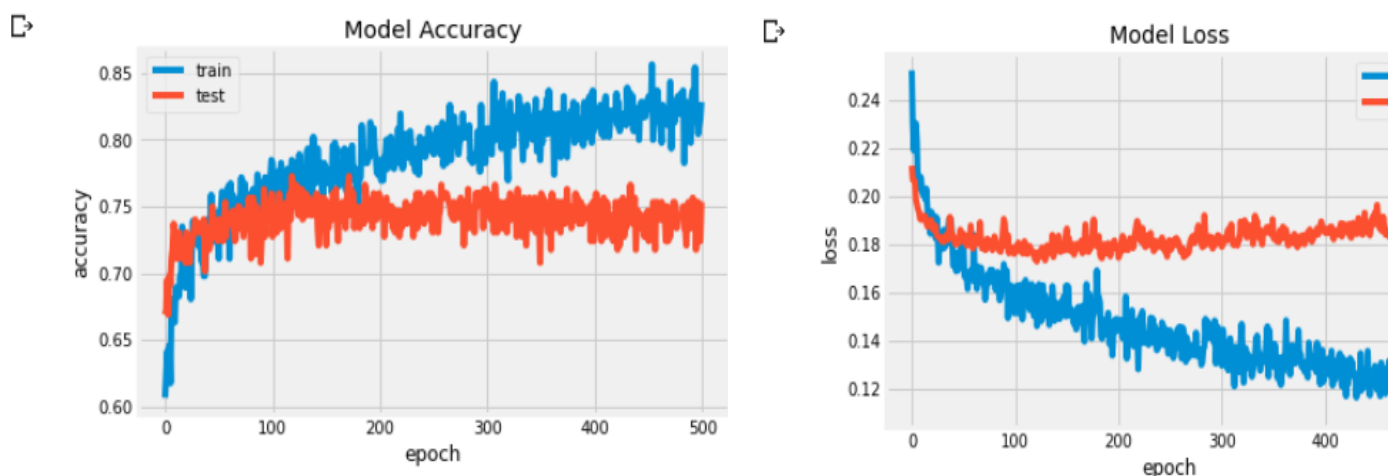


Fig. III.11. la différence entre accuracy et loss d'apprentissage et de test avec l'ensemble de données des patients du diabète avec le model DNN .

Nous avons également la matrice de confusion qui montre également combien notre modèle a été formé :

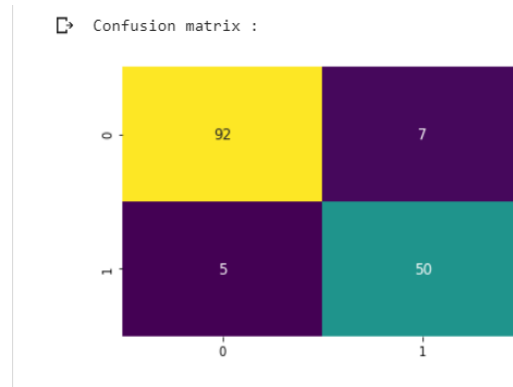


Fig. III.12. la matrice de confusion des patients du diabète avec le model DNN .

III.4.4. Avec la base Alzheimer

Dans cette partie, nous montrons les résultats obtenus avec la base de données des patients d'Alzheimer avec de nombreux modèles d'apprentissage en profondeur et également les résultats des années dernière avec de nombreux classificateurs d'apprentissage automatique (svm , knn , ...etc) , le **Tableau III.8.** résume les performances de tous ces classificateurs .

Model	epochs	Accuracy	Loss	Time	Accuracy	Loss	Time
		Train %	Train %	train(us)	Test %	Test %	test(us)
<i>DNN</i>	200	78.6	48	80	74.6	56	81
<i>DFFnn</i>	500	68	63	9ms	73	66	2ms
<i>LSTM</i>	100	81	48	563	70	64	191
<i>SVM</i>					55		
<i>Decision Tree</i>					63		
<i>Logistic Regression</i>					75		
<i>Random Forest</i>					66		
<i>KNN</i>					73		

Tableau III.8. Résultats obtenus avec la base de données des patients d'Alzheimer.

Après avoir obtenu les résultats (accuracy et loss) avec tous les modèles , nous avons comparé ces résultats et comme nous le notons dans le tableau ci-dessus **Tableau III.7.** , Les résultats que nous avons obtenus avec ces données avec le modèle DNN (**accuracy_testing = 92** et **loss_testing = 20**) C'était meilleur que les résultats que nous avons obtenus des modèles LSTM et DFF , alors qu'il était si proche des résultats avec les classificateurs d'apprentissage automatique (M_L) , les résultats sont bons comme le montre dans **Fig. III.13.**

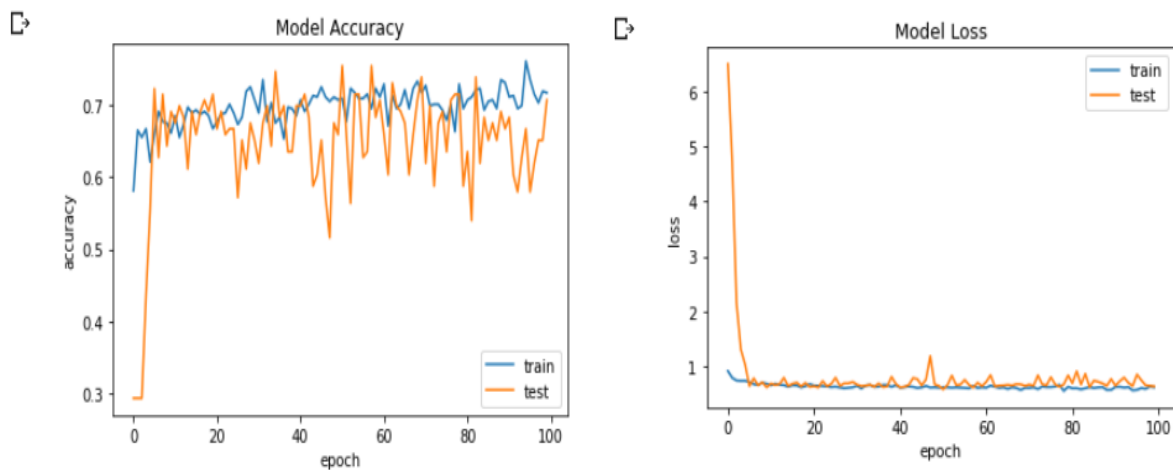


Fig. III.13. la différence entre accuracy et loss d'apprentissage et de test avec l'ensemble de données des patients d'Alzheimer avec le model DNN .

III.5. CONCLUSION

Ce dernier chapitre a fait l'objet d'une étude de simulation liée à la mise en œuvre de techniques d'apprentissage en profondeur appliquées à la science de l'identification des maladies, cette étude ayant permis de jeter un regard sur la plupart des techniques d'apprentissage en profondeur (DNN, LSTM, DFF, ... etc) et qu'en les étudiant, en savoir plus sur eux et évaluer séparément les performances de chacun d'eux .

Cette étude détaillée a été menée dans le but de choisir de manière décisive l'approche proposée, les paramètres associés au taux de reconnaissance (précision et perte) étant les facteurs pertinents qui ont permis d'évaluer les méthodes étudiées et de rendre possible la comparaison entre leurs performances, où après cette comparaison, nous avons constaté qu'en général, ce modèle de réseau de neurones profonds donne de meilleures performances que d'autres modèles et classificateurs, en particulier en termes d'identification des maladies et lorsque nous utilisons ce type de données spécialement .

Conclusion générale

CONCLUSION GENERALE

Notre objectif de travail s'est concentré sur la mise en œuvre d'un système de reconnaissance des maladies à l'aide de techniques d'apprentissage en profondeur.

Au stade de la sélection des technologies que nous utiliserons dans notre système, le critère que nous avons pris en considération est la forme des données que nous allons utiliser, car les données ont plusieurs formes (1D, 2D, 3D), et les techniques d'apprentissage en profondeur diffèrent en fonction de leur utilisation de tout type de ces données, par exemple, l'une d'entre elles n'utilise qu'une seule forme de ces formes ou nous donne les meilleures performances avec celle-ci, mais une autre peut utiliser avec toutes ces formes de données, dans notre cas, l'ensemble de données que nous avons utilisé est one_dimension.

Nous avons ensuite testé des techniques d'apprentissage en profondeur à l'aide de quatre vraies bases de données biomédicales avec une dimension de différentes maladies pour vérifier la validité du système proposé, car il nous permet de connaître les avantages et les inconvénients de chaque technique et les différences entre une technique et une autre.

Si ce système s'applique dans les hôpitaux, le nombre de traitements sera réduit ce qui aura lieu contribuera à réduire l'intervention du médecin, surtout dans certains cas qui ne nécessitent pas du tout son intervention, ce qui réduira le matériel médical et les personnes concernées.

Dans ce cas, DNN a prouvé son efficacité grâce à ses performances élevées par rapport à d'autres techniques d'apprentissage en profondeur et à des techniques d'apprentissage automatique, comme illustré par l'ensemble de tableaux du chapitre précédent.

Références

REFERENCES

- [1] Guide méthodologique : Comment évaluer a priori un Programme de dépistage ? ,Service évaluation technologique , Service évaluation économique Mai 2004 années .
- [2]<https://livre.fnac.com/a1854678/Marina-Carrere-d-Encausse-La-maladie-d-Alzheimer>
- [3]<https://sante.journaldesfemmes.fr/fiches-maladies/2499922-diabete-definition-causes-symptomes-traitement/>
- [4]Mémoire de fin d'études, pour l'obtention du diplôme de Master en Informatique, Option: Modèle Intelligent et Décision(M.I.D),Thème : Deep Learning pour la classification des images, Réalisé par : Moualek Djaloul YoucefAnnée universitaire : 2016-2017 .
- [5] Mémoire Master Professionnel , Domaine : Informatique et Technologie de l'Information,Filière : Informatique,Spécialité : fondamentale,Présenté par : Boughaba Mohammed et Boukhris Brahim,Thème:L'apprentissage profond (Deep Learning) pour la classification et la recherche d'images par le contenu Année universitaire 2016/2017 .
- [6] Mémoire de fin d'étude , en vue de l'obtention du diplôme de Master en Informatique Option Systèmes intelligents , Thème : Développement d'une Architecture Basée sur l'Apprentissage Profond (Deep Learning) pour la Détection d'Intrusion dans les Réseaux , Préparé par : Mr. MIMOUNE Zakarya. Année Universitaire 2018/2019.
- [7] <https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/deep-learning-vs-machine-learning/>
- [8] Mémoire de fin d'étude , en vue de l'obtention du diplôme de Master en informatique , Option :Systèmes Intelligents , Thème :La classification des images satellitaires par l'apprentissage profonde (deep learning) , Préparé par : Abdelaziz HABBA et Omar ISHAK
- [9] <https://actualiteinformatique.fr/intelligence-artificielle/definition-deep-learning>
- [10] <https://developer.ibm.com/technologies/artificial-intelligence/articles/cc-machine-learning-deep-learning-architectures/>
- [11] <https://dataanalyticspost.com/Lexique/reseau-de-neurones-profonds/>
- [12] <https://penseeartificielle.fr/focus-reseau-neurones-convolutifs/>

- [13] <https://www.aliens-sci.com/convolutional-neural-network-cnn/>
- [14] <https://penseeartificielle.fr/focus-reseau-neurones-artificiels-perceptron-multicouche/>
- [15] <https://www.sciencedirect.com/science/article/abs/pii/S0952197617300234>
- [16] <https://towardsdatascience.com/deep-learning-feedforward-neural-network-26a6705dbdc7>
- [17] <https://hackernoon.com/deep-learning-feedforward-neural-networks-explained-c34ae3f084f1>
- [18] <https://hackernoon.com/building-a-feedforward-neural-network-from-scratch-in-python-d3526457156b>
- [19] <https://medium.com/swlh/neural-networks-4b6f719f9d75>
- [20] <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>, Consulté Mai 2019.
- [21] <https://www.kaggle.com/ronitf/heart-disease-uci>, Consulté Mai 2019.
- [22] <http://biochimej.univangers.fr/Page2/COURS/Zsuite/6BiochMetabSUITE/5IntelligenceArtificielle/1IntelligenceArtificielle.htm>
- [23] <https://alzheimer-recherche.org/la-maladie-alzheimer/quest-maladie-dalzheimer/definition-et-chiffres/>

ملخص:

هذا العمل هو جزء من مجال الذكاء الاصطناعي لحل المشاكل المتعلقة بالصحة. قمنا بتطبيق نظام التعرف على علم الأمراض باستخدام تقنيات التعلم العميق . تتمثل مساهمتنا الرئيسية في هذا العمل ، من ناحية ، في استخدام مجموعة من الميزات لتقليل عدد العلاجات التي تقلل من المعدات الطبية والموظفين اللازمين. من ناحية أخرى ، قدمت DNN كفاءتها من حيث حساب الوقت والأداء لاتخاذ قرار مهم بشأن علم الأمراض.

كلمات مفتاحية : أمراض القلب، سرطان الثدي، مرض السكري ، التصنيف ، التعلم العميق ، تعلم الآلة

، CNN ، DNN ، DFF .

Summary :

This work is part of the field of artificial intelligence to solve health-related problems. We implemented a pathology recognition system using deep learning techniques . Our main contribution to this work is, on the one hand, to use a range of features to reduce the number of treatments that reduce the medical equipment and personnel needed. On the other hand, DNN provided its efficiency in terms of calculating time and performance for important pathology decision making.

Mots clés : heart disease, breast cancer, diabetes, classification, Machine learning , Deep learning , DNN, CNN, DFF .

