

# Chapitre 3

## État de l'art

### 3.1 Introduction

Dans ce chapitre, nous avons réalisé une étude approfondie sur les avancées récentes de la recherche concernant les techniques de Machine Learning utilisées pour l'analyse du Big Data. Ces avancées nous ont permis de proposer une nouvelle technique d'analyse du Big Data, qui sera présentée en détail dans le chapitre suivant.

### 3.2 IA au service de l'analyse du Big Data : Classification des techniques et paramètres d'évaluation

L'analyse du Big Data est une entreprise complexe qui requiert des outils et des techniques appropriés. L'intelligence artificielle (IA) est devenue un élément essentiel de l'analyse du Big Data, offrant des méthodes efficaces pour découvrir des modèles, prédire des tendances et prendre des décisions automatisées. Le Machine Learning, en particulier le Deep Learning, a révolutionné notre capacité à traiter les Big Data.

La **Figure 3.1** présente une **taxonomie des techniques d'analyse de Big Data basée sur les sous-domaines de l'intelligence artificielle**. La classification proposée comprend quatre grandes catégories, à savoir le Machine Learning, les méthodes basées sur les connaissances et les inférences, les algorithmes de prise de décision, les méthodes de recherche et la théorie de l'optimisation (**Russell et Norvig, 2020**)[18]. De plus, les paramètres qualitatifs les plus importants pour évaluer chaque méthode d'analyse du

Big Data et comprendre ses avantages et ses inconvénients sont identifiés comme suit [19] :

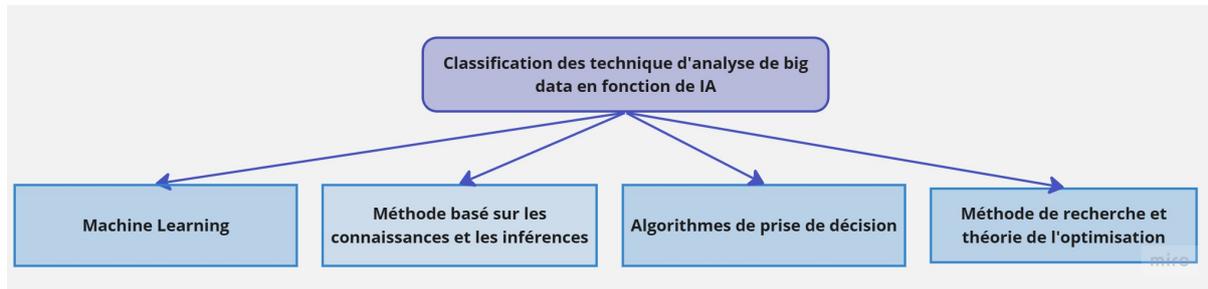


FIGURE 3.1 – Classification des techniques d’analyse big data en utilisant l’IA.

- **Scalabilité** : Il s’agit de la capacité d’un mécanisme ou d’un système à s’adapter rapidement aux changements, tels qu’une augmentation de la taille des données ou du nombre d’utilisateurs, sans compromettre la qualité de l’analyse.
- **Efficacité** : Cela fait référence au rapport entre la méthode utilisée et les besoins totaux en termes de temps et de coûts.
- **Précision** : Elle est évaluée en tenant compte de divers paramètres, tels que les erreurs de données et la capacité prédictive des algorithmes, pour déterminer la fiabilité et l’exactitude des résultats.
- **Confidentialité** : Cela concerne les pratiques mises en place pour garantir que les données sont utilisées uniquement à des fins autorisées et qu’elles sont protégées contre l’accès non autorisé.
- **Complexité** : La complexité d’un algorithme mesure la quantité de ressources nécessaires pour l’exécuter en fonction de la taille de l’entrée. Elle est généralement exprimée en notation Big O et permet de choisir les algorithmes les plus efficaces pour résoudre un problème donné.

Il est important que ces concepts sont cruciaux lors du développement et de l’application de nouvelles techniques d’analyse de Big Data basées sur l’IA. Les chercheurs et les praticiens doivent prendre en compte ces aspects pour assurer des résultats précis, une efficacité opérationnelle, une évolutivité et une confidentialité appropriées dans leurs solutions d’analyse des Big Data.

### 3.3 Mécanismes du Machine Learning

Il existe plusieurs méthodes de Machine Learning utilisées pour l'analyse du Big Data, chacune ayant ses avantages et ses limites. La **Figure 3.2** présente certains des algorithmes les plus couramment utilisés pour chaque type de Machine Learning :

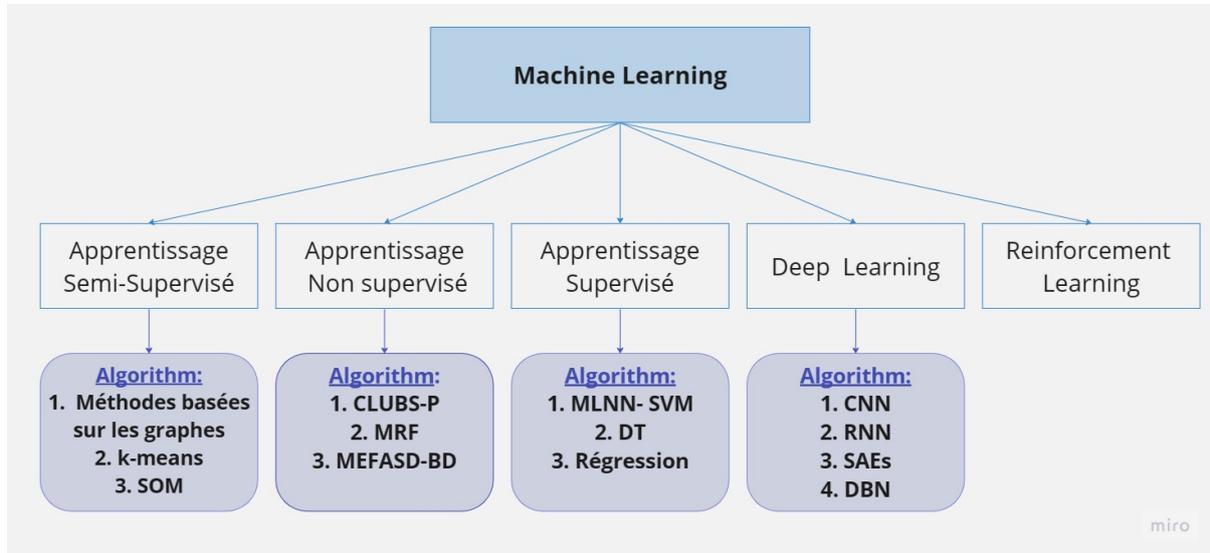


FIGURE 3.2 – Taxonomie des algorithmes les plus courants pour chaque type de Machine Learning utilisé pour l'analyse du Big Data.

#### 3.3.1 Apprentissage supervisé

La première catégorie nécessite des efforts manuels considérables pour mettre les données dans un format adapté aux algorithmes d'apprentissage. Il existe plusieurs algorithmes de Machine Learning supervisés couramment utilisés pour l'analyse de gros volumes de données (Big Data). Ces algorithmes sont souvent utilisés pour résoudre des problèmes de classification, où le modèle apprend à prédire des étiquettes discrètes ou des catégories prédéfinies pour de nouvelles données, ou de régression, où le modèle apprend à prédire des valeurs continues ou numériques pour de nouvelles données (voir la **Figure 3.3**), qui sont des tâches courantes dans l'analyse de données. Voici quelques-uns des algorithmes les plus couramment utilisés :

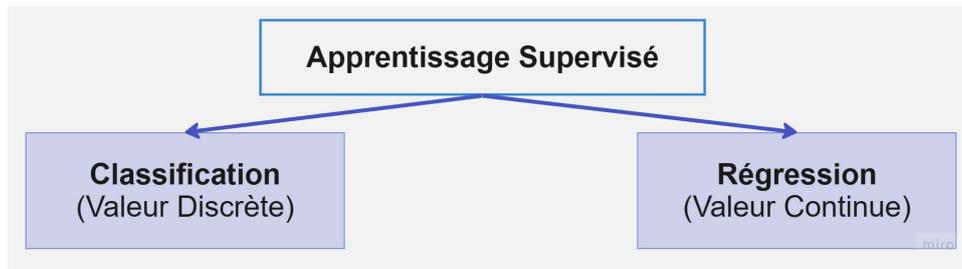


FIGURE 3.3 – L'apprentissage supervisé

### 3.3.1.1 Réseau neuronal multicouche entraîné avec Support Vector Machine (SVM)

Dans leur article, **AIZubi (2020) [20]**, les auteurs ont développé une nouvelle technique de classification de données massives pour la reconnaissance du diabète en utilisant des techniques de MapReduce et de Machine Learning. Ils collectent d'abord les données à partir du Big Data et utilisent le modèle MapReduce pour diviser efficacement les ensembles de données en sous-ensembles plus petits. Ensuite, une procédure de nettoyage des données est utilisée pour éliminer le bruit généré par les données collectées. Les caractéristiques sélectionnées sont ensuite formées à l'aide d'un réseau neuronal multicouche entraîné avec SVM. Le réseau neuronal est appliqué pour classer les nouveaux échantillons. Les résultats expérimentaux montrent que cette méthode est efficace pour la reconnaissance du diabète, avec des taux d'erreur, de sensibilité, de spécificité et d'exactitude acceptables.

### 3.3.1.2 Modèle d'arbre de décision (DT)

Dans leur article, **Nair et Shetty (2018) [2]**, les auteurs ont développé un système de prédiction en temps réel de l'état de santé à distance en appliquant des modèles de Machine Learning à d'énormes flux de données. Le système est conçu en utilisant Apache Spark dans un environnement cloud. Un modèle d'arbre de décision est créé à partir des données de santé actuelles et appliqué au flux de données pour prédire l'état de santé. Le modèle d'arbre de décision joue un rôle très important dans le Machine Learning. Il est capable de gérer à la fois les variables continues et confidentielles, et fournit une indication claire de prédiction ou de classification sans nécessiter beaucoup de calculs, permettant ainsi une meilleure approximation, quelle que soit la complexité des données. Les algorithmes les plus couramment utilisés pour cela sont "Classification and Regression Trees" (CART) et "Iterative Dichotomiser 3" (ID3).

La structure présentée conduit à une efficacité optimale en termes de temps et de système. La confidentialité des données est protégée en utilisant un compte secondaire Twitter.

La **Figure 3.4** suivante représente ce système évolutif : l'utilisateur tweete ses attributs de santé, et l'application les reçoit en temps réel, extrait les attributs et applique le modèle de Machine Learning pour prédire son état de santé. Les résultats sont ensuite envoyés directement à l'utilisateur via un message instantané pour prendre les mesures appropriées.

En conclusion, le modèle d'arbre de décision permet d'économiser considérablement de temps et d'argent en utilisant pleinement les technologies existantes plutôt que de développer de nouvelles technologies pour effectuer la même tâche. Avec quelques modifications mineures, cette application peut être utilisée pour prédire la présence de diverses maladies.

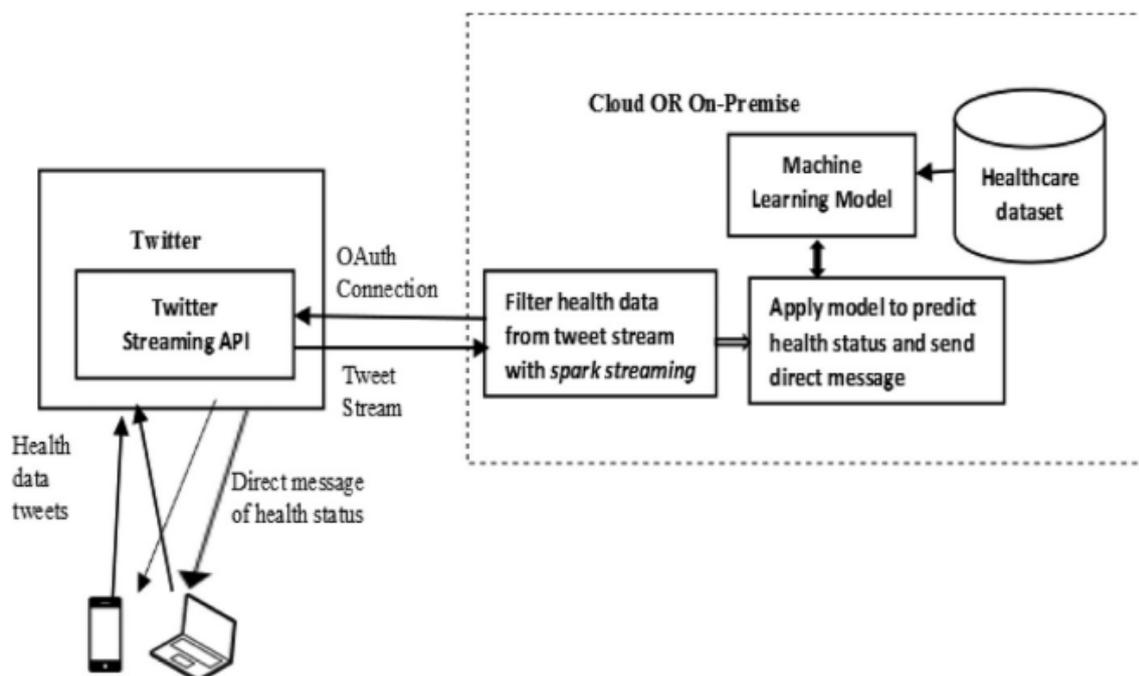


FIGURE 3.4 – Modèle de système de surveillance de la santé en temps réel [2].

### 3.3.1.3 Algorithmes de régression avec apprentissage d'ensemble

Dans une étude menée par **Asencio-Cortés et al. (2018)** [3], l'objectif était de prédire les événements sismiques en Californie en utilisant des algorithmes de régression combinés à un apprentissage d'ensemble, dans le contexte du Big Data. Les chercheurs ont

analysé les données sur les tremblements de terre en Californie de 1970 à 2017, totalisant 1 Go d'informations réparties en 27 ensembles de données. L'objectif était de prédire la magnitude des tremblements de terre pour les sept prochains jours.

Leur approche utilisait le framework Apache Spark ainsi que des modèles d'apprentissage automatique de la bibliothèque H2O, tels que la régression linéaire, les machines à renforcement de gradient (GBM), l'apprentissage profond et les forêts aléatoires (voir **Figure 3.5**). Ils ont également exploité l'infrastructure cloud d'Amazon. Les résultats obtenus étaient très prometteurs, avec des erreurs relatives avoisinant les 10

Les principaux avantages de la méthode proposée étaient des niveaux élevés de parallélisme et de mise à l'échelle. Cependant, elle présentait une efficacité relativement faible pour le traitement de grands ensembles de données.

En résumé, l'analyse de données massives pour la prédiction de la magnitude des séismes ouvre une nouvelle voie de recherche très prometteuse. Cette approche pourrait être utile pour traiter simultanément de grandes quantités de données contenant de nombreuses variables.

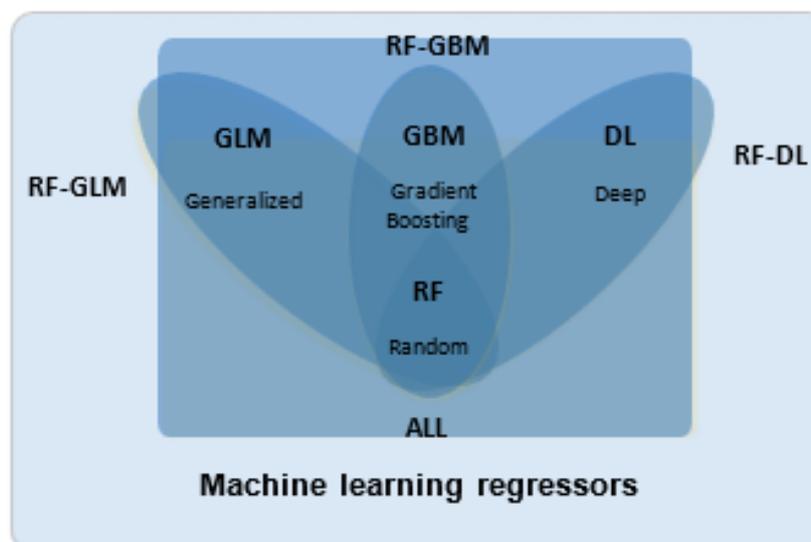


FIGURE 3.5 – Exemple d'algorithmes de régression avec apprentissage d'ensemble [3].

Le **Tableau 3.1** fourni présente une comparaison entre différents algorithmes d'apprentissage supervisé :

TABLE 3.1 – Comparaison entre les algorithmes supervised learning [6].

Références	Algorithme	Évolutivité	Efficacité	Précision	Confidentialité
(BDA-MC)[20]	MLNN entraîné avec SVM	X	X	oui	X
ML-SBD [2]	Le modèle d'arbre de décision	oui	oui	oui	oui
RAC-BD [3]	régression	oui	X	oui	X

### 3.3.2 Apprentissage non supervisé

Les algorithmes d'apprentissage non supervisé sont utilisés pour découvrir des structures ou des modèles cachés dans les données sans avoir besoin d'une étiquette ou d'une réponse prédéfinie pour chaque observation. Un exemple courant d'application d'apprentissage non supervisé est l'algorithme de recommandation de produits d'Amazon [21].

Les algorithmes non supervisés peuvent être classés en deux catégories principales : le clustering (ou regroupement), qui regroupe les données similaires en clusters, et la réduction de dimensionnalité, qui permet de visualiser les données pour faciliter leur compréhension.

#### 3.3.2.1 CLUBS-P

Dans leur article, **Ianni et al. (2020) [22]** soulignent que les méthodes traditionnelles d'analyse de données, telles que l'analyse de régression et les tests d'hypothèses, ne sont pas adaptées à la gestion des big data générées par les applications et les dispositifs modernes. Ils proposent d'utiliser des algorithmes de clustering qui peuvent identifier des modèles et des relations dans les données sans nécessiter de connaissances ou d'hypothèses préalables.

Dans leur étude, **Ianni et al. (2020) [22]** ont présenté un algorithme de clustering basé sur le centroïde pour les big data, appelé CLUBS-P. Ils ont évalué deux implémenta-

tions différentes de CLUBS-P pour déterminer sa faisabilité dans le contexte des big data. Les résultats obtenus ont montré une très bonne évolutivité de l'algorithme et une amélioration par rapport à l'algorithme de clustering parallèle de référence, à savoir K-means. De plus, l'implémentation de CLUBS-P basée sur le passage de messages a permis une utilisation optimale des ressources des nœuds disponibles lorsque toutes les fonctionnalités offertes par Spark n'étaient pas nécessaires.

### 3.3.2.2 Systèmes flous évolutifs multi-objectifs

L'article **Pulgar-Rubio et al. (2017) [23]** présente une nouvelle méthode appelée MEFASD-BD (Multi-objective Evolutionary Fuzzy Systems) qui est la première approche d'évolution des systèmes flous pour la découverte de sous-groupes dans le contexte du Big Data. Cette méthode utilise le modèle MapReduce avec Apache Spark pour traiter efficacement de grandes quantités de données. L'objectif est d'analyser la qualité des sous-groupes obtenus pour chaque bloc de données afin d'améliorer leur pertinence. MEFASD-BD est spécialement conçu pour gérer des ensembles de données de haute dimension, ce qui en fait une solution adaptée aux défis du Big Data.

Les résultats des tests réalisés ont démontré une réduction significative du temps d'exécution tout en maintenant des niveaux élevés de qualité. Cette méthode ouvre de nouvelles perspectives pour la découverte de sous-groupes dans les environnements de Big Data.

### 3.3.2.3 Champs aléatoires de Markov (MRF)

Dans l'article **Ip et al. (2018) [24]**, un aperçu du Big Data et des techniques d'apprentissage automatique dans la protection des cultures est présenté. De plus, la possibilité d'utiliser des champs aléatoires de Markov (Markov random fields) qui prennent en compte la composante spatiale entre les sites voisins pour modéliser la résistance aux herbicides est examinée. Les résultats de l'expérience ont démontré la performance de l'approche proposée.

Le Tableau 3.2 présente une comparaison entre trois algorithmes d'apprentissage non supervisé :

TABLE 3.2 – Comparaison entre les algorithmes Unsupervised learning [6]

Références	Algorithme	Évolutivité	Efficacité	Précision	Confidentialité
EBD-C [22]	CLUBS-P	oui	oui	oui	X
MEFASD-BD [23]	Systèmes flous évolutifs multi-objectifs	oui	oui	oui	X
BDML-CP [24]	MarKov random fields	oui	X	oui	X

### 3.3.3 Apprentissage semi-supervisé

L'analyse de Big Data peut être réalisée à l'aide d'algorithmes d'apprentissage semi-supervisé pour traiter de grandes quantités de données de manière plus efficace. L'idée sous-jacente est que les données non étiquetées contiennent des informations utiles qui peuvent être exploitées pour améliorer la généralisation des modèles. Voici quelques exemples d'algorithmes semi-supervisés qui peuvent être utilisés pour l'analyse de Big Data :

#### 3.3.3.1 Méthodes basées sur les graphes

Les méthodes de classification semi-supervisées basées sur les graphes, présentées par **Zhu (2005)** [25], définissent un graphe où les nœuds représentent des exemples étiquetés et non étiquetés de l'ensemble de données, et les arêtes reflètent la similarité entre les exemples (elles peuvent être pondérées). Ces méthodes utilisent généralement des techniques de propagation d'étiquettes sur le graphe. Les méthodes basées sur les graphes sont généralement non paramétriques, discriminantes et transformationnelles, ce qui signifie qu'elles ne peuvent pas être facilement étendues à de nouveaux points qui ne font pas partie de l'ensemble d'apprentissage.

#### 3.3.3.2 Méthode des k-means

La méthode des k-means, proposée par **MacQueen (1967)** [26], est une méthode de clustering utilisée pour diviser automatiquement un ensemble de données en k groupes. Elle fonctionne en sélectionnant k centres de groupe initiaux, puis en les affinant itérativement de la manière suivante : chaque instance est assignée au centre de groupe le

plus proche, puis chaque centre de groupe est mis à jour pour être la moyenne des instances qui lui sont assignées. L'algorithme converge lorsque l'affectation des instances aux clusters ne change plus. Les clusters sont pré-initialisés avec des instances choisies au hasard dans l'ensemble de données.

### 3.3.3.3 Self Organizing Map (SOM)

Le Self Organizing Map (SOM), présenté par **Haykin (2003) [27]**, est un algorithme populaire de réseau de neurones artificiels basé sur l'apprentissage non supervisé. Le SOM peut projeter des données de grande dimension dans un espace de dimension inférieure, ce qui peut être utile pour analyser des modèles dans l'espace de recherche ayant une structure complexe. Les résultats sont visuels et faciles à analyser. Son rôle principal est de réaliser une projection non linéaire des données de haute dimension sur un espace de faible dimension. Les cartes auto-organisatrices sont largement utilisées dans la classification de données.

Le Tableau suivant présente une comparaison entre trois algorithmes d'apprentissage semi-supervisé :

TABLE 3.3 – Comparaison entre les algorithmes semi-supervised learning.

Références	Algorithme	Évolutivité	Efficacité	Précision	Confidentialité
MCA-MO[26]	Méthode des k-means	X	oui	oui	X
SSL-G [25]	Méthodes basées sur les graphes	X	oui	oui	X
NN-ACF[27]	Self Organizing Map	oui	oui	oui	oui

### 3.3.4 Deep Learning

Les techniques de Deep Learning sont un domaine de l'informatique qui se concentre sur le développement de réseaux de neurones artificiels capables d'analyser et d'apprendre à partir de données pour prendre des décisions et faire des prédictions. Les données massives sont l'un des principaux domaines où les techniques de Deep Learning sont appliquées, car la quantité de données générées aujourd'hui est trop importante

pour que les techniques traditionnelles puissent les traiter efficacement.

Dans cette partie, nous présenterons quelques-uns des algorithmes de Deep Learning les plus couramment utilisés pour l'analyse de Big Data :

#### **3.3.4.1 Convolutional Neural Networks**

Li Shen et al. (2019) [4] présentent un algorithme de Deep Learning, en particulier les Convolutional Neural Networks (CNN), pour la détection précise du cancer du sein à partir d'images mammographiques de dépistage. L'approche "end-to-end" utilisée permet d'obtenir d'excellentes performances, avec des AUC élevées et une réduction des faux positifs et des faux négatifs.

Cette approche permet de réduire significativement le besoin d'annotations de ROI et a de nombreuses applications en imagerie médicale, en plus de la détection du cancer du sein sur les mammographies de dépistage.

L'algorithme a été testé sur deux ensembles de données indépendants : des mammographies en film numérisé de la base de données numérique pour la mammographie de dépistage (CBIS-DDSM) et des images de mammographie numérique plein champ (FFDM) de la base de données INbreast. Les résultats montrent une performance remarquable avec une AUC (Area Under the Curve) par image de 0,88 pour le meilleur modèle individuel sur l'ensemble CBIS-DDSM, et une amélioration à 0,91 grâce à une moyenne de quatre modèles. Sur l'ensemble INbreast, l'AUC par image atteint 0,95 pour le meilleur modèle individuel et 0,98 avec la moyenne de quatre modèles. Les sensibilité et spécificité des modèles sont également rapportées.

Une découverte importante de l'étude est la possibilité de transférer le classificateur d'image entière entraîné sur les mammographies CBIS-DDSM vers les images FFDM INbreast, en utilisant uniquement un sous-ensemble des données INbreast pour l'affinage, sans nécessiter d'annotations de lésions supplémentaires.

La Figure 3.6 représente l'approche end-to-end des Convolutional Neural Networks pour la détection du cancer du sein.

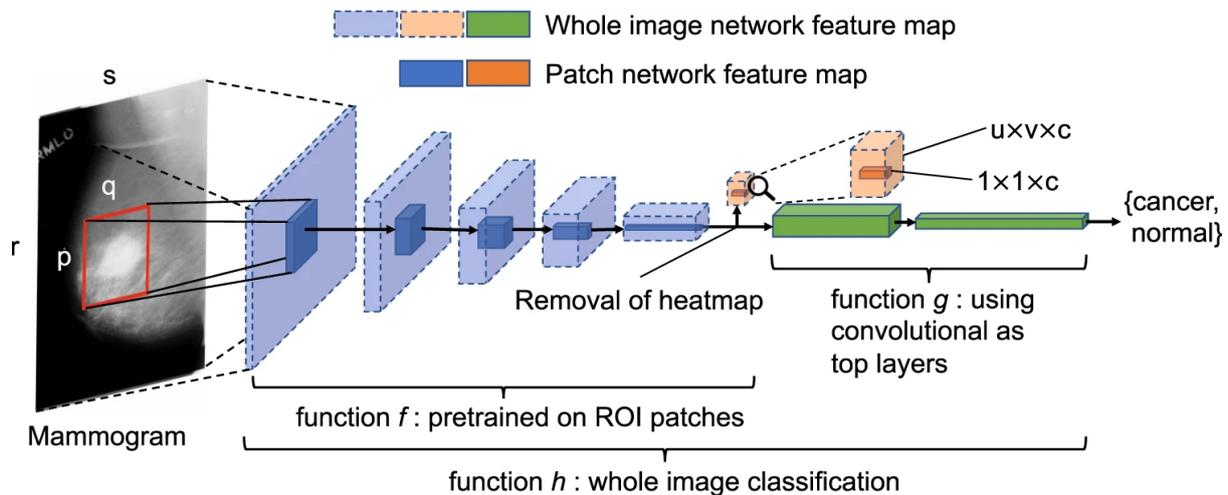


FIGURE 3.6 – Approche end-to-end des Convolutional Neural Networks pour la détection du cancer du sein [4]

### 3.3.4.2 Recurrent Neural Networks

Dans leur article, Bilal Jan et al. (2017) [28] expliquent que les modèles traditionnels de Deep Learning tels que les autoencodeurs empilés, les Deep Belief Networks (DBN) et les Convolutional Neural Networks (CNN) ne sont pas adaptés à l'apprentissage de caractéristiques pour les données de séries temporelles. Les Recurrent Neural Networks (RNN) sont présentés comme une solution efficace pour traiter les données de séries temporelles, en particulier dans les applications de traitement du langage naturel. Les RNN apprennent les caractéristiques des données de séries temporelles en stockant l'information des entrées précédentes dans leur état interne, ce qui leur permet de prendre en compte les relations entre les données successives. Cependant, les RNN peuvent rencontrer des problèmes de disparition du gradient, ce qui limite leur capacité à capturer des dépendances à long terme. Des variantes des RNN ont été développées pour résoudre ce problème, telles que les Long Short-Term Memory (LSTM) et les Gated Recurrent Units (GRU). Les RNN ont obtenu des performances remarquables dans de nombreuses applications de traitement du langage naturel, de reconnaissance vocale et de traduction automatique.

La Figure 3.7 représente l'architecture des Recurrent Neural Networks.

Hammou, Lahcen et Mouline (2020) [29] ont introduit une nouvelle technologie efficace pour l'analyse émotionnelle en utilisant des Réseaux de Neurones Récurrents. Les auteurs ont adopté un texte rapide avec des variables de réseau de neurones récurrents

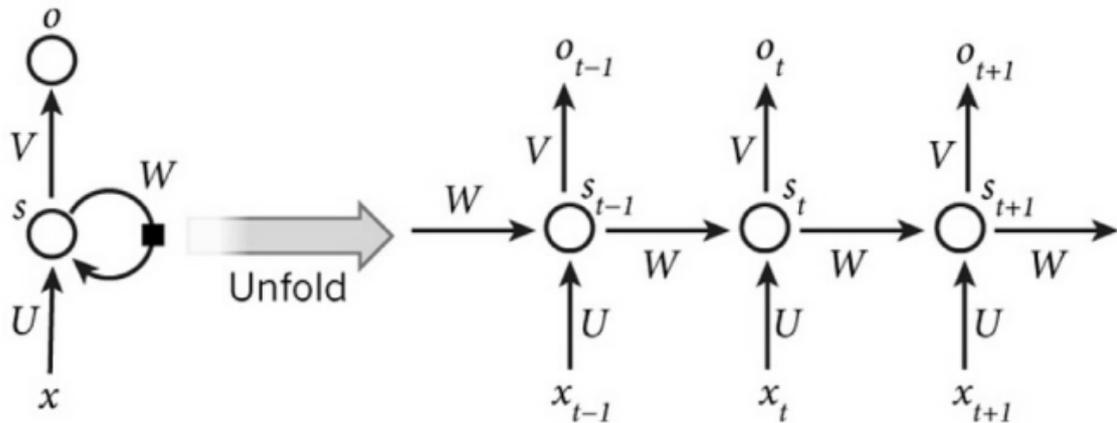


FIGURE 3.7 – Architecture des Recurrent Neural Networks [? ]

pour représenter et classer les données de texte. De plus, un système distribué basé sur l'apprentissage automatique distribué pour des analyses en temps réel a été suggéré. Les tests effectués démontrent que la méthode fournie dépasse les méthodes LSTM, BTSTM et GRU en termes de précision de la classification, et peut gérer d'excellentes données pour l'analyse émotionnelle.

Les RNN sont particulièrement efficaces lorsqu'il s'agit de traiter de gros volumes de données, car leur capacité à traiter des données séquentielles leur permet d'extraire des caractéristiques et des modèles à partir de vastes quantités de données de manière efficace. De plus, comme les RNN apprennent à partir de données passées, ils peuvent s'adapter aux ensembles de données changeants et aux nouvelles informations, ce qui les rend hautement adaptables et utiles dans diverses applications telles que la reconnaissance de la parole, la classification d'images, la prédiction de séquences et la compréhension du langage naturel. Dans l'ensemble, les RNN sont une technique puissante et polyvalente d'apprentissage en profondeur qui est précieuse pour la manipulation de gros volumes de données.

### 3.3.4.3 Autoencoders et Stacked Autoencoders (SAEs)

Akrufa Hajirahimova et Aybeniz Aliyeva (2020) [5] présentent les Stacked Autoencoders (SAEs) comme l'une des techniques de Deep Learning les plus largement utilisées. Les SAEs sont construits en empilant plusieurs autoencodeurs, qui sont les réseaux de neurones feed-forward les plus typiques. Un autoencodeur est une structure d'apprentissage non supervisée qui possède trois couches : la couche d'entrée,

la couche cachée et la couche de sortie, comme représenté dans la Figure 3.8. Le processus d'entraînement d'un autoencodeur consiste en deux étapes : l'étape d'encodage et l'étape de décodage. L'encodeur est utilisé pour mapper les données d'entrée dans une représentation cachée, et le décodeur est utilisé pour reconstruire les données d'entrée à partir de la représentation cachée.

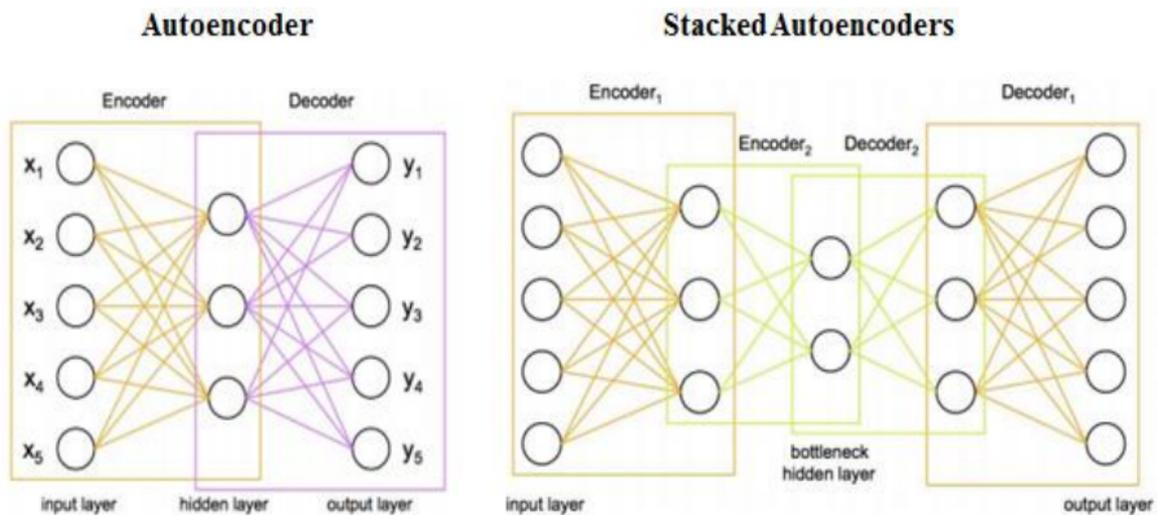


FIGURE 3.8 – Architecture des autoencodeurs [5]

Les SAEs sont généralement entraînés en deux étapes : la préformation et le réglage fin. Dans l'étape de préformation, chaque modèle d'autoencodeur est entraîné de manière non supervisée couche par couche, de bas en haut. Cette opération est répétée jusqu'à ce que les paramètres de toutes les couches cachées soient entraînés. Une fois que toutes les couches cachées sont entraînées, l'algorithme de rétropropagation est utilisé pour minimiser la fonction de coût et mettre à jour les poids avec un ensemble d'entraînement étiqueté afin d'effectuer un réglage fin.

#### 3.3.4.4 Deep Belief Network (DBN)

D'après l'article de Bilal Jan et al. (2017) [28], le réseau de croyance profonde (Deep Belief Network, DBN) est un modèle qui peut apprendre la représentation des caractéristiques à partir de données étiquetées et non étiquetées en utilisant des techniques non supervisées et supervisées. L'architecture du DBN, comme illustré dans la Figure 3.9, comprend une couche d'entrée, des couches cachées et une couche de sortie, avec

deux couches connectées directement formant une machine de Boltzmann restreinte (RBM)<sup>1</sup>.

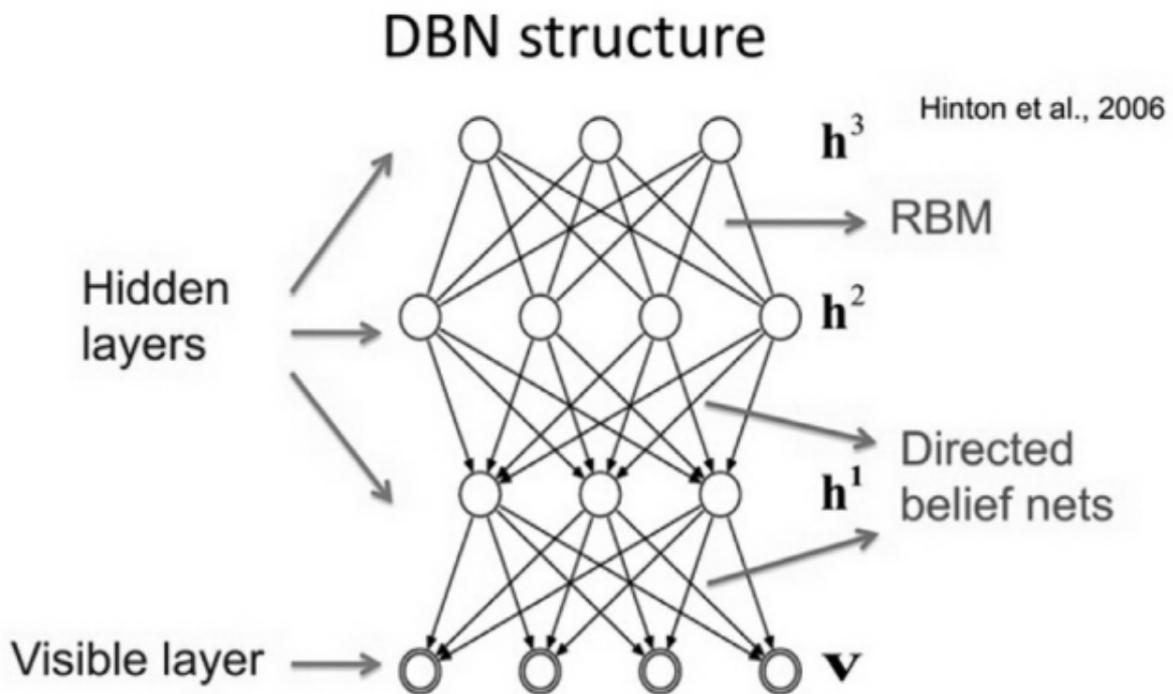


FIGURE 3.9 – Deep Belief Network [6].

Les chercheurs exploitent la puissance de la discrimination de DBN pour traiter efficacement les Big Data. Afin de gérer des quantités massives de données en un temps de traitement réduit, un modèle d'architecture basé sur les unités de traitement graphique (GPU) a été proposé. Cependant, l'implémentation de ce modèle pour les données à grande échelle présente divers défis, notamment le transfert de données entre la mémoire client et la mémoire globale. Une solution efficace consiste à stocker tous les paramètres et la solution d'entraînement dans la mémoire globale pendant la phase d'entraînement, tandis qu'un traitement parallèle des données permet des mises à jour simultanées sur chaque bloc d'informations. L'implémentation sur GPU démontre son efficacité lors de l'incorporation de plusieurs millions de paramètres dans RBM [28].

Le tableau suivant (**Tableau ??**) présente une comparaison entre quatre algorithmes de Deep Learning :

1. RBM comporte deux couches, les nœuds des deux couches étant entièrement connectés entre eux et aucune connectivité entre les nœuds au sein de la même couche

TABLE 3.4 – Comparaison entre les algorithmes Deep learning.

Références	Algorithme	Évolutivité	Efficacité	Précision	Confidentialité
DL-BCDSM [4]	CNN	oui	oui	oui	X
DL-BDA[28]	DBN	oui	X	oui	X
SDL-BDA [5]	SAEs	X	oui	X	X
PDRN-BDA)[29] DL-BDA[28]	RNN	oui	X	oui	X

En conclusion, le volume de données massives générées à une vitesse impressionnante dépasse les capacités des techniques d'apprentissage automatique traditionnelles pour fournir des solutions optimales aux ensembles de données complexes. Les techniques de deep learning ont réalisé des avancées significatives en abordant les problèmes de données en permettant l'identification de prédictions plus précises à partir de vastes quantités de données, en dévoilant des motifs et des structures complexes. Le deep learning utilise des réseaux de neurones artificiels pour effectuer un traitement multicouches des données, permettant aux algorithmes d'extraire des caractéristiques de haut niveau à partir des données brutes. Cette approche simplifie non seulement le traitement des données, mais améliore également la précision et l'efficacité des modèles d'apprentissage automatique.

### 3.4 Comparaison entre Les algorithmes de chaque type du Machine Learning

Une comparaison entre les différents types d'algorithmes de Machine Learning est présentée dans le Tableau ci-dessous :

TABLE 3.5 – Comparaison Entre Les algorithmes de chaque type du Machine Learning.

Type d'apprentissage	Nom de l'algorithme	Année	Auteur(s)	Complexité	Domaines
Supervised Learning	SVM trained multi layer neural network	2020	AlZubi	$O(N^3)$	Classification, Régression
Supervised Learning	Decision tree model	2018	Nair, et Shetty	$O(n)$	Classification, Régression
Supervised Learning	Regression algorithms with ensemble learning	2018	Asencio-cortés et al	$O(N^2)$	Régression
Unsupervised Learning	CLUBS-P	2020	Anni et al	$O(n \log(n))$	Clustering
Unsupervised Learning	Systèmes flous évolutifs multi-objectifs	2017	Pulgar-Rubio et al	$O(n)$	Clustering
Unsupervised Learning	Markov random fields	2018	IP et al	Non spécifier	Traitement et segmentation d'image
DL	RNN	2020	Hammou, Lahcen et Mouline	$O(N^2)$	Traitement du langage naturel
DL	DBN	2017	Bilal Jana et al	Non spécifier	Classification, Régression
DL	$CNN_{EndtoEnd}$	2019	Li Shen et al et al	$O(N^2)$	Vision par ordinateur, Traitement d'image
SSL	Méthodes basées sur graphes	2005	Zhu	$O(N^3)$	Classification, Régression, Clustering
SSL	Méthode des k-means	1967	MacQueen	$O(n)$	Clustering
SSL	Self Organizing Map (SOM)	2003	Haykin	Non spécifier	Clustering, Réduction de dimensionnalité