

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

UNIVERSITÉ BADJI MOKHTAR-ANNABA
BADJI MOKHTAR- ANNABAUNIVERSITY



جامعة باجي مختار – عنابة

Faculté : Sciences de l'Ingénierat
Département : Electronique
Domaine : Sciences et Techniques
Filière : Automatique
Spécialité : Automatique et Système

Mémoire

Présenté en vue de l'obtention du Diplôme de Master

Thème :

Surveillance automatique de la qualité d'eau basée sur
les techniques d'apprentissage approfondi

Présenté par: *NEDJAI AIA ABDELOUHAB*

Encadrant: *SAADI MED NACER MCB UBM ANNABA*

Jury de Soutenance :

Bekaik Mounir	MCB	UBM Annaba	Président
SAADI MED NACER	MCB	UBM Annaba	Encadrant
Ramdani Messaoud	MCA	UBM Annaba	Examineur

Année Universitaire : 2020/2021



REMERCIEMENTS

Avant toute chose, on remercie Dieu le tout puissant de m'avoir donné le courage, la patience et la force de reprendre mes études après une longue rupture (depuis 2005).

Tous mes remerciements à Monsieur SAADI MOHAMED NACER De m'avoir encadré et en particulier pour son soutien, sa gentillesse, ses encouragements, ses conseils pendant la période de la réalisation de ce travail. Nos sincères remerciements aux Messieurs les membres de jury qui nous ont fait l'honneur en acceptant d'examiner ce travail.

Mes salutations aux enseignants du département d'électronique qui ont contribué à notre formation.

Mes remerciements aussi à mes collègues au travail de m'avoir encouragé de continuer mes études.

Enfin, un remerciement spécial aux étudiants de la 2eme année Master 20/21 du département d'électronique plus spécialement de l'option automatique et système.



Dédicaces

Je voudrais dédier cet humble travail à ma chère maman et mon cher père A mes frères à mon épouse mon premier soutien sur ma retour a l'AFAC et de m'avoir encouragé et soutenir tout au long de mes

Études

A mes enfants « Ghadir et Ayhem Taki » Que Dieu les protège et leurs offre la chance et le bonheur.

A toute ma famille.

A mes collègues au travail

Merci !

TABLE DES MATIERES.....I
LISTE DES FIGURES.....IV
LISTE DES TABLEAUX..... V
LISTE D'ABREVIATIONS.....VI
INTRODUCTION GENERALE.....1

Chapitre I

Systèmes de production d'eau potable et surveillance industrielle

I.1	Introduction.....	5
I.2	Traitement physique	6
I.2.1	Prétraitement	6
I.2.2	Le dégrillage	6
I.3	Traitement physico-chimique	7
I.3.1	Pré-chloration	7
I.3.2	Clarification	7
I.3.2.1	Coagulation-floculation.....	7
I.3.3	Décantation	11
I.3.4	Filtration sur sable.....	11
I.4	Traitement chimique.....	11
I.4.1	Post-chloration (désinfection).....	11
I.4.2	Stockage de l'eau.....	12
I.5	Propriétés générales d'une eau de boisson et celle destinée à la consommation humaine.....	12
I.5.1	Les différents paramètres spécifiques qui commandent la qualité des eaux de consommation.....	13
I.5.1.1	Les paramètres organoleptiques.....	13
I.5.1.2	Les paramètres physico-chimiques.....	14
I.5.1.3	Les matières organiques.....	14
I.6	Surveillance des procédés de production d'eau potable.....	15
I.6.1	Surveillance industrielle : définitions et état de l'art	15
I.6.2	Définitions des termes de base utilisés en surveillance	16
I.6.3	Méthodes de surveillance	18
I.7	Conclusion.....	20

Chapitre II

Deep learning et CNN

II.1	Introduction.....	21
II .2	L’IA, Intelligence artificielle	22
II.3	La machine Learning	23
II.3.1	Apprentissage supervisé : apprendre par l’exemple	24
II.3.2	Apprentissage non supervisé : apprendre en explorant	24
II .4	Le Deep Learning	25
II.5	La méthode de réseaux neurones convolutionnels (CNN).....	27
II.5.1	Architecture globale de CNN	28
II.5.1.1	Couche convolutive	28
II.5.1.2	Couche de pooling	29
II.5.1.3	Couche totalement connectée	29
II.5.1.4	Couche de correction (ReLU)	30
II.5.1.5	Nombre de filtres	32
II.5.1.6	Forme de filtres	32
II.6	Conclusion.....	33

Chapitre III

Simulation par CNN

III.1	Introduction.....	34
III.2	Classification des images.....	38
III.3	Architecture du réseau.....	40
III.3.1	Hyper-paramètres du modèle de référence	41
III.4	Résultats obtenus et discussion.....	43
III.5	Conclusion.....	47
	Conclusion générale.....	48

Liste de Figures

Figure 1 Répartition de l'eau sur la terre (site web memento de Degérmont).....1

Figure I.1 Un modèle typique de potabilisation appliquée à une eau de surface.....6

Figure I.2 Coagulation-floculation.....8

Figure I.3 Essai « Jar-Test ».....15

Figure I.4 Composantes de la surveillance industrielle.....18

Figure I.5 Classification globale des méthodes de surveillance 19

Figure II.1 Intelligence artificielle, machine learning e, deep learning et convolutional network.....21

Figure II.2 Réseau de neurones simple, Réseau de neurones profond 25

Figure II.3 Les types de couches d'un réseau de neurones26

Figure II.4 Réseau neuronal convolutif28

Figure II.5 Architecture simple de CNN..... 28

Figure II.6 Max pooling..... 29

Figure II.7 courbe la fonction ReLU..... 30

Figure II.8 courbe la fonction sigmoid..... 31

Figure II.9 courbe la fonction de tangente hyperbolique..... 32

Figure III.1 Manque de données au mois de mars et juillet 2015 sur le tableau d'évolution des paramètres de la qualité d'eau de la station de CHAIBA année 2015.....35

Figure III.2 Echantillon forme d'image avec les niveaux de gris 38

Figure III.3 Classification de la qualité de l'eau38

Figure III.4 Exemple - Eau de bonne qualité.....39

Figure III.5 Exemple – Eau de moyenne qualité..... 39

Figure III.6 Exemple – Eau de mauvaise qualité..... 39

Figure III.7 Exemple – Eau de très mauvaise qualité.....40

Figure III.8 Architecture du modèle..... 41

Figure III.9 Répartition de données d'apprentissage (51 image).....43

Figure III.10 Courbe de précision du modèle d'apprentissage..... 43

Figure III.11 Courbe d'erreur du modèle d'apprentissage.....44

Figure III.12 Courbe de précision du modèle test44

Figure III.13 Courbe d'erreur du modèle test.....45

Figure III.14 Matrice de confusion.....46

Liste des Tableaux

Tableau I. 1	Les différents coagulants.....	8
Tableau I.2	Les différentes flocculant.	10
Tableau III.1	Grille normative pour estimer la qualité de l'eau en Algérie.....	36
Tableau III.2	Exemple de classement manuel des paramètres d'analyse de la qualité d'eau de la station de CHAIBA année 2015.....	37
Tableaux III.3	Hyper-paramètres du modèle de référence adopté	42
Tableau III.4	Répartition de 30 images de test.....	45

Liste d'abréviations

PH: potentiel Hydrogène

T°:Température

TU: Turbidité

UV: Ultra violets

Ca²⁺: peroxide de calcium

Mg²⁺ : Magnesium

Na⁺: Sodium

DBO : Demande Biochimique en oxygène

DCO : Demande chimique en oxygène

Fe : Fer

Mn : Maganése

Cu : cuivre

Zn : Zinc

Cd: cadmuim

Pb : plomb

OMS : l'Organisation mondiale de la santé

NFU (NTU): (Nephelometric Turbidity Unit)

MES : Matières en suspension

RNA : Réseaux de neurons

C : Conductivité

OD : Oxygène dissous

CO₂ : Dioxyde de carbone

Caco₃ : Carbonate de calcium

CaO₂ : peroxide de calcium

Hco₃ : Bicarbonates

CL : Chloures

CNN : Convolutional Neural Network

ReLu : Rectified Linear Units

MLP : Multi-layer perceptron

ANN :Artificial neural network



Figure 1 Répartition de l'eau sur la terre

INTRODUCTION GENERALE

Les rêves les plus fous de l'humanité est de trouver la vie ailleurs que sur notre planète. Dans tout l'Univers, l'eau est une molécule que l'homme recherche avidement, car sa découverte à la surface ou dans l'atmosphère d'une planète lointaine libèrerait aussitôt la réalisation de ces rêves.

L'eau constitue un élément essentiel dans la vie et l'activité humaine. C'est une composante majeure des mondes minéral et organique. Dans le monde présent, l'eau participe à toutes les activités quotidiennes notamment, domestiques, industrielles et agricoles ce qui la rend un élément récepteur exposé à tous les genres de pollution. Le phénomène de la pollution contribue de façon considérable à la limitation des ressources en eau potable.

Sa formule peut s'écrire de façon très simple : H_2O (ce qui signifie que cette molécule résulte de l'union d'un atome d'oxygène avec 2 atomes d'hydrogène), **c'est l'eau.**

Derrière ce mot, s'impose une image, née des voyages spatiaux, celle de la planète bleue : la Terre.

L'eau est en effet la substance minérale la plus répandue à la surface du globe. Elle en constitue l'hydrosphère. Son volume est estimé à $1\,385 \cdot 10^6 \text{ km}^3$, dont environ 97,4 % dans les océans (couvrant 71 % de la surface terrestre), mais malheureusement cette eau est salée et non potable et ne convient pas à l'irrigation, 2 % sous forme de glace et 0,6 % seulement (de l'ordre de $8 \cdot 10^6 \text{ km}^3$) constituant les eaux continentales (y compris les nappes souterraines et l'humidité des sols). À la surface du sol, on évalue à environ $350\,000 \text{ km}^3$ le volume des eaux continentales disponibles (fleuves, lacs) ; les glaces des calottes polaires représentent un volume de $25 \cdot 10^6 \text{ km}^3$. Enfin, il y a $13\,000 \text{ km}^3$ d'eau dans l'atmosphère sous forme de vapeur et de nuages. L'évaporation annuelle est estimée à $500\,000 \text{ km}^3$ et les précipitations sur les continents à $110\,000 \text{ km}^3$ par an.

L'eau est le constituant majeur de la matière vivante (synonyme de vie biologique). Il entre en moyenne pour 80 % dans sa composition.

Élément majeur du monde minéral et biologique, l'eau est aussi le vecteur privilégié de la vie et de l'activité humaine. À l'heure actuelle, l'utilisation globale de l'eau, en additionnant les usages domestiques, industriels et agricoles, représente le chiffre impressionnant de 700 m^3 par an et par habitant en moyenne, sachant que. Et encore, les disparités sont énormes selon les régions du monde : de moins de 100 m^3 pour les pays en voie de développement à plus de $2\,000 \text{ m}^3$ pour les États-Unis. Il est donc certain que les besoins en eau de l'humanité ne cesseront de croître. La croissance de cette consommation d'eau a été très forte ces dernières décennies. Environ 10% de cette consommation sert à l'usage domestique [1]. Ceci implique la nécessité impérieuse de protéger l'eau et de la traiter, que ce soit pour produire une eau propre à la consommation humaine ou à des usages spécifiques industriels, ou pour limiter les rejets de pollution dans le milieu naturel.

L'exigence d'une réglementation très stricte des pouvoirs publics est alors bien justifiée. Durant ces dernières décennies, des efforts importants ont été réalisés dans le développement des méthodes de contrôle et de surveillance automatique des eaux potables. Ces méthodes dites de haut niveau disposent d'outils qui sont plutôt orientés vers la communication avec un opérateur expert. Celles-ci représentent les techniques de l'intelligence artificielle (IA) qui peuvent être utilisées en raison de leur robustesse et de leur capacité à tenir compte de la nature dynamique et

complexe du procédé. Ce type de technique est de plus en plus accepté dans l'industrie de production d'eau potable en tant qu'outil de modélisation et de contrôle des procédés qui servent comme outil de base pour l'aide à la décision. Leur réponse est plus élaborée et peut être obtenue soit à partir de données brutes venant directement des variables de surveillance ; soit à partir de données traitées venant des sorties de traitements de bas niveau [2]. Il est tout à fait logique de supposer que le problème de contrôle et de surveillance de l'eau brute peut être vu comme un problème de reconnaissance de formes, représentant l'ensemble des observations ou mesures liées aux caractéristiques de celle-ci.

Parmi les techniques d'IA connues dans ce domaine, on trouve les réseaux de neurones artificiels et plus particulièrement les réseaux de neurones convolutifs (CNN) qui représentent la technique approfondie la plus pertinente qu'on peut utiliser dans la surveillance et contrôle des systèmes de la qualité d'eau. Ces techniques se démarquent des autres outils par leur capacité d'apprentissage rapide et leur généralisation.

Les multi données de paramètres physico-chimiques exploités dans la surveillance et de contrôle de la qualité d'eau, tels que la conductivité, pH, la température, l'oxygène dissous, et les MES (matières en suspension), sont représentés sous forme d'images à partir du CNN et transmise vers une station de contrôle qui assure l'acquisition et le traitement des données. La technique devant être utilisée au niveau du système de décision doit pouvoir effectuer un contrôle quasi permanent de cette ressource précieuse. De plus, on peut utiliser les sorties de ce premier contrôle comme des variables d'entrée dans un autre système de surveillance [3].

Ce travail a été réalisé au niveau de la station de production d'eau potable « CHAIBA ANNABA » de la commune de Sidi Amar ANNABA. Cette station fournit de l'eau à plus de 700 000 habitants (2012) et a une capacité nominale de traitement de 1000 l/s-3600M³/h à partir de l'eau brute pompée à partir du barrage de « CHAFIA ELTARF » de capacité 160 million de M³ environ.

Le projet de fin d'étude est composé de trois chapitres finalisés par une conclusion générale.

Le premier chapitre est consacré aux étapes générales d'une usine de production d'eau potable, la plus complète et la plus récente, on aborde également le sujet de propriétés générales d'une eau de boisson ainsi que les différents paramètres spécifiques qui commande la qualité de l'eau de consommation, en particulier ceux relatifs au procédé de coagulation. La dernière partie de ce chapitre est consacrée à la surveillance des procédés de production d'eau potable et les différentes méthodes de surveillance.

Le deuxième chapitre est dédié à la clarification des différents passages de l'apprentissage approfondi arrivant à la représentation de la méthode utilisée pour la réalisation de la surveillance de la qualité d'eau potable celle de CNN.

Le troisième chapitre représente plus spécialement la partie simulation et les résultats obtenus.

On finalise ce travail de mémoire par une conclusion générale qui résume l'étude réalisée dans ce travail et les recommandations envisagés adéquate.

Chapitre I Systèmes de production d'eau potable et surveillance industrielle

I.1 Introduction :

L'eau qui est fournie par le réseau de distribution doit être conforme aux normes de potabilités (limites), de qualité fixée par la réglementation. Lorsque la limite de qualité est dépassée, l'eau est déclarée non potable. Le principal objectif d'une station de production d'eau potable est de fournir un produit qui satisfait à un ensemble de normes de qualité à un prix raisonnable pour le consommateur. L'OMS dans ses directives établies en 1983 impose à chaque pays d'établir sa propre législation, en fonction des critères locaux, en y incluant en particulier les données économiques.

L'eau est le composé le plus abondant sur la surface du globe ; sans elle, la vie comme nous le savons cesserait d'exister. On note que 5 millions de personnes meurent chaque année dans le monde du fait de la mauvaise qualité de l'eau ! L'assurance de la qualité et de l'hygiène de l'eau est l'une des exigences prioritaires de la santé publique. Les eaux à visée de potabilisation pour la consommation humaine sont de différentes natures. Les eaux souterraines constituent 1/5 des réserves d'eau douce soit environ 1000 milliards de m³ [4]. Elles sont généralement d'excellente qualité physico-chimique et bactériologique. Néanmoins, les terrains traversés en influent fortement la minéralisation. Les eaux de surface se répartissent en eaux courantes ou stockées. Elles sont généralement riches en gaz dissous, en matières en suspension et organiques, ainsi qu'en plancton. Elles sont très sensibles à la pollution minérale et organique de type nitrate et pesticide d'origine agricole. Dans le but d'avoir une eau de qualité, les contrôles sont de plus en plus rigoureux.

Nous allons donner un aperçu, en premier lieu dans ce chapitre, des étapes générales d'une usine de production d'eau potable, la plus complète et la plus récente (figure I.1), en deuxième lieu, on aborde le sujet de propriétés générales d'une eau de boisson ainsi que les différents paramètres spécifiques qui commandent la qualité de l'eau de consommation, en particulier ceux relatifs au procédé de coagulation.

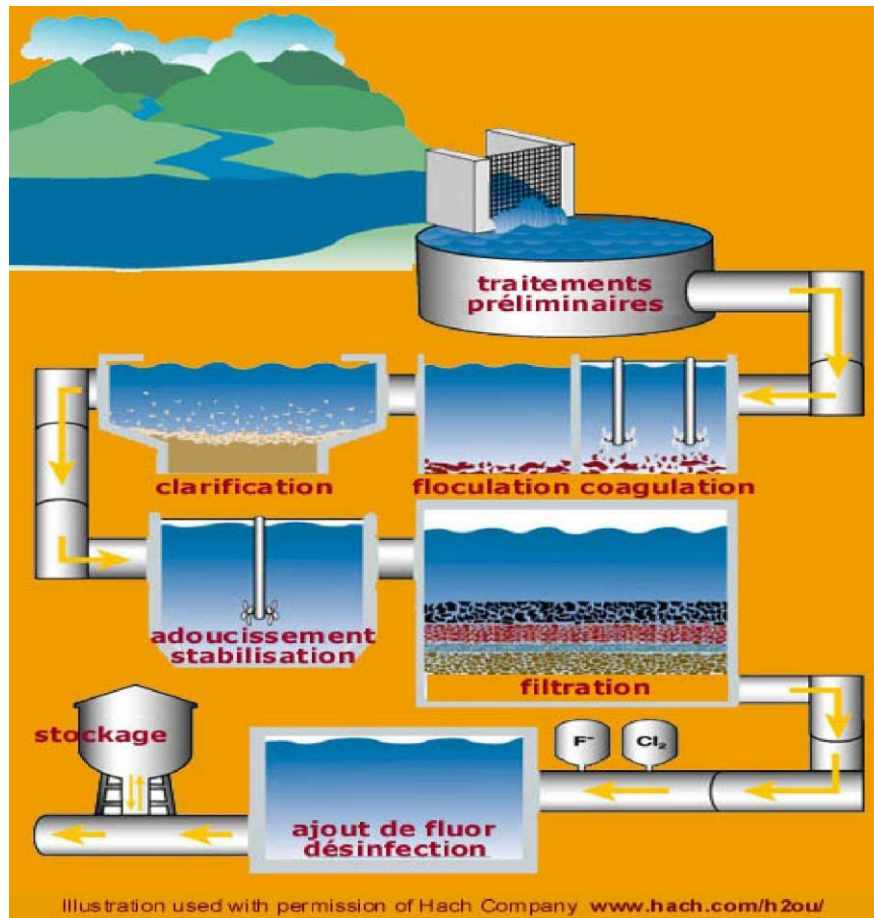


Figure I.1 Un modèle typique de potabilisation appliquée à une eau de surface.

I.2 Traitement physique

I.2.1 Prétraitement

Les eaux subissent une étape de prétraitement sur le site de captage avant d'être conduites vers l'unité de traitement. Le prétraitement consiste à faire un dégrillage afin d'éliminer les particules de grosse taille, les branches, le sable... [5].

I.2.2 Le dégrillage

Le dégrillage permet d'enlever et arrêter les corps flottants, les gros déchets et les débris de dimensions intermédiaires (passant à travers la grille, afin d'éviter qu'ils interfèrent avec le fonctionnement des équipements aval). Il sert également à empêcher l'accès des poissons aux ouvrages avals (conduite et installation de traitement). Après un dégrillage (traitement primaire où les particules de diamètres importants sont retenues), les eaux brutes sont orientées vers la station du traitement.

I.3 Traitement physico-chimique

I.3.1. Pré-chloration

Si les eaux à traiter contiennent beaucoup de matières organiques, ou encore de l'ammoniaque, du fer ou du manganèse, une étape de pré-chloration (cette étape s'effectue dans un bassin de mélange) préalable est nécessaire comme c'est le cas ici à la station de CHAIBA SIDIAMAR ANNABA. Elle permet d'éliminer plus facilement ces substances au cours de l'étape suivante dite de clarification. L'oxydation est une opération essentielle à tout traitement des eaux. La station utilisée pour l'oxydation des matières organiques, le fer, le manganèse et l'ammoniaque un oxydant comme le chlore gazeux ou Hypochlorite de sodium.

Actuellement, la station utilise dans l'eau des doses entre 2,5 à 3 g/m³ de chlore.

I.3.2. Clarification

La clarification consiste à agréger sous forme de floccs, les matières en suspension(MES) organique et minérales et les substances colloïdales. Ces matières indésirables s'étant agglomérées, elle se décantent au fond du bassin de décantation ou elles sont régulièrement extraites. Par contre, l'eau surnageant est filtrée sur sable [6].

La clarification comprend les opérations suivantes :

- Coagulation
- Flocculation
- Filtration

I.3.2.1 Coagulation-flocculation:

La coagulation-flocculation est un procédé physico-chimique de clarification des eaux. Il réside dans la formation, par l'addition de coagulant, trames floconneuses appelées " Floccs". La coagulation flocculation, traditionnellement considérée comme un procédé lié à l'élimination des états dispersés (matières en suspension ou colloïdes), peut devenir efficace vis -à vis de l'abattement de la charge organique dissoute à condition de se placer dans des conditions optimales souvent différentes de celles exigées pour l'élimination de la turbidité [7]

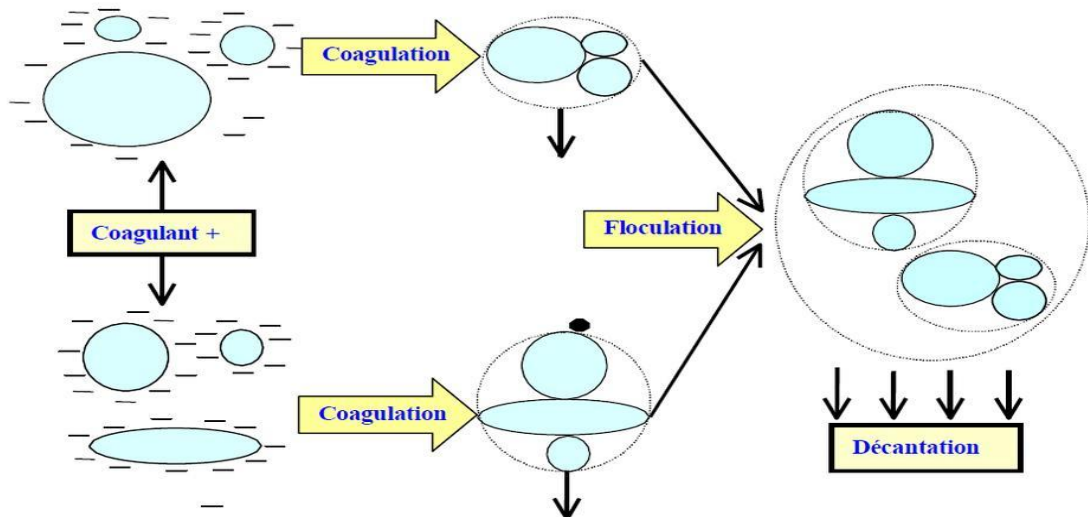


Figure I.2 Coagulation-floculation

➤ **Coagulation:**

La coagulation est un processus qui consiste à neutraliser les charges portées par les substances colloïdales ou dissoutes indésirables à l'aide d'un produit chimique de charge opposée, appelé coagulant avec une agitation rapide (le bassin de coagulation doit être équipé d'une unité mécanique de mélange rapide), afin de faciliter leur agglomération en flocons décantables ou filtrables.

Les coagulants sont des produits capables de neutraliser les charges des colloïdes présents dans l'eau. La difficulté principale est de déterminer la quantité optimale de réactif à injecter en fonction des caractéristiques de l'eau brute. Un mauvais contrôle de ce procédé peut entraîner une augmentation importante des coûts de fonctionnement et le non-respect des objectifs de qualité en sortie. Le choix du coagulant pour le traitement de l'eau de consommation doit tenir compte de l'innocuité du produit, de son efficacité et de son coût.

Nom de coagulant	Formes disponibles	Aspect	Densité	Le taux de traitement usuel	Gamme de pH d'utilisation
Sulfate d'alumine	Concassé Noisettes Poudre Liquide	Blanchâtre	1	10 à 150 g/m ³	5,7 < pH < 7,5
Polychlorures basiques d'alumine (PCBA)	Liquide	Blanchâtre	1,45	10 à 100 g/m ³	étendue
chlorure ferrique	crystallisé liquide	Brun. Déliquescent Brun	1,45	05 à 100 g/m ³	5 < pH < 8,5
chlorosulfate ferrique	solution	Brun rouge	1,5	05 à 100 g/m ³	5 < pH < 8,5
Sulfate ferrique	Poudre Liquide	Blanche Rouge brun	1	10 à 100 g/m ³	5 < pH < 8,5
Sulfate ferreux	Poudre	Cristallin Couleur vert clair	0,9	10 à 100 g/m ³	pH > 7,8

Tableau I. 1 Les différents coagulants. [8]

Le type de coagulant et la dose ont une influence sure :

- La bonne ou la mauvaise qualité de l'eau clarifiée,
- Le bon ou le mauvais fonctionnement de la floculation et de la filtration,
- Le coût d'exploitation.

Les deux principaux types de coagulant sont :

- Les sels de fer (chlorure ferrique) CLF3
- Les sels d'aluminium (sulfate d'aluminium), produit utilisé dans la station de traitement de CHAIBA

Le choix du coagulant peut varier avec la température et la saison. Le sulfate d'aluminium, par exemple, est un coagulant utilisé pour une température d'eau supérieure à 10-12°C. On peut rappeler également que plus un coagulant a de charges positives, plus son efficacité est grande. Par la suite, nous allons énumérer l'ensemble des paramètres influençant le bon fonctionnement du procédé de coagulation. [9]

▪ L'influence du paramètre pH

Le pH a une influence primordiale sur la coagulation. Il est d'ailleurs important de remarquer que l'ajout d'un coagulant modifie souvent le pH de l'eau. Cette variation est à prendre en compte afin de ne pas sortir de la plage optimale de précipitation du coagulant. La plage du pH optimal est la plage à l'intérieur de laquelle la coagulation a lieu suffisamment rapidement. En effet, une coagulation réalisée à un pH non optimal peut entraîner une augmentation significative du temps de coagulation. En général, le temps de coagulation est compris entre 15 secondes et 3 minutes. Le pH a également une influence sur l'élimination des matières organiques.

▪ L'influence de la dose de coagulant

La dose de réactif est un paramètre à prendre en compte. Le coagulant qui est habituellement fortement acide a tendance à abaisser le pH de l'eau. Pour se placer au pH optimal, il est possible d'ajouter un acide ou une base. Une dose de coagulant excessive entraîne une augmentation du coût d'exploitation, tandis qu'un dosage insuffisant conduit à une qualité de l'eau traitée insuffisante.

La difficulté principale est de déterminer la quantité optimale de réactif à injecter en fonction des caractéristiques de l'eau brute. A l'heure actuelle, il n'existe pas de modèle de connaissance simple qui permet de déterminer le taux de coagulant en fonction des différentes variables affectant le procédé. La détermination du taux de coagulant est effectuée par analyse hors ligne au laboratoire tous les jours, à l'aide d'un essai expérimental appelé « Jar-test » (Figure I.3). Cet essai consiste à mettre des doses croissantes de coagulant dans des récipients contenant la même eau brute. Après quelques instants, on procède sur l'eau décantée à toutes les mesures utiles de

qualité de l'eau. La dose optimale est donc déterminée en fonction de la qualité des différentes eaux comparées. L'inconvénient de cette méthode est de nécessiter l'intervention d'un opérateur. On voit ici tout l'intérêt de disposer d'un moyen automatique pour effectuer cette détermination.

▪ L'influence de la température

La température joue un rôle important. En effet, une température basse, entraînant une augmentation de la viscosité de l'eau, crée une série de difficultés dans le déroulement du processus : la coagulation et la décantation du floc sont ralenties et la plage du pH optimal diminue. Pour éviter ces difficultés, une solution consiste à changer de coagulant en fonction des saisons.

▪ L'influence de la turbidité

La turbidité est, elle aussi, un paramètre influant sur le bon fonctionnement du procédé de coagulation. Dans une certaine plage de turbidité, l'augmentation de la concentration en particules doit être suivie d'une augmentation de la dose de coagulant.

Quand la turbidité de l'eau est trop faible, on peut augmenter la concentration en particules par addition d'argiles. Dans le cas de fortes pluies, l'augmentation des MES (matière en suspension) favorise une meilleure décantation. Enfin, pour grossir et alourdir le floc, on ajoute un adjuvant de floculation

➤ Floculation

La floculation est l'étape de traitement qui suit la coagulation. Elle vise à favoriser la croissance de floccs par une agitation lente et prolongée de l'eau provenant des bassins de coagulation. Elle est réalisée dans un bassin pourvu d'une unité mécanique d'agitation et implique habituellement l'ajout d'un flocculant (le bassin de floculation doit être équipé d'une unité mécanique de mélange lente). Elle complète la phase de la coagulation et vise à assurer une plus grande cohésion du floc et une meilleure vitesse de sédimentation. L'adjuvant ou le flocculant peut être introduit dans un bassin de floculation est poly-électrolyte.

Nom de flocculant	Formes disponibles	Origine	Taux de traitement usuel
Flocculant minéral la silice activée	Liquide: se prépare sur l'installation	Neutralisation d'une solution de silicate de soude par un acide sulfurique, sulfate d'aluminium chlore, dioxyde de carbone	0,2 à 1 g/m ³
flocculant organique naturels: l'alginate	Poudre	Extrait d'algue	0,2 à 1 g/m ³
L'aqualgine	Poudre	Alginate purifié	0,2 à 1 g/m ³
l'amidon	Poudre	Pomme de terre, tapioca extraits de graisse végétale	0,2 à 1 g/m ³
Flocculant organique de synthèse Non anionique anionique	solution Poudre émulsion	fabriqués en laboratoire à partir de polyacrylamides	0,02 à 0,5 g/m ³

Tableau I.2 Les différents flocculants. [8]

I.3.3 Décantation

La décantation a pour but d'éliminer les floccs issus de la coagulation et floculation, elle se fait grâce au bassin de décantation, le volume de ce dernier est 3400 m³. Le temps nécessaire pour la décantation des floccs est deux heures.

I.3.4 Filtration sur sable

La filtration est la barrière ultime et obligatoire de la filière de traitement des eaux dans la majeure partie des cas. Elle vise à réaliser ou à compléter, à travers un lit filtrant, la réduction des particules en suspension, des coliformes, des virus, des parasites ainsi que la turbidité. Sans elle, plusieurs filières de traitement ne pourraient obtenir de crédits pour l'enlèvement des virus et des kystes de protozoaires.

Ce procédé de filtration est celui qui est le plus utilisé à la station du CHAIBA SIDI AMAR ANNABA. Il est caractérisé par un lavage intermittent de son média filtrant. Ce lavage de filtre se fait comme suit:

- Fermeture de la vanne de sortie d'eau filtrée.
- Ouverture de la vanne d'entrée d'eau de lavage.
- Mise en service de la pompe de dé colmatage pendant 30 à 50 secondes.
- Mise en marche du surpresseur de lavage.
- Ouverture de la vanne d'entrée d'air de lavage pendant 8 à 30 minutes.
Puis fermeture de la vanne d'entrée d'air.
- Arrêt du surpresseur de lavage.
- Mise en service des pompes de rinçage pendant 8 à 10

I.4 Traitement chimique

I.4.1 Post-chloration (désinfection)

En raison de la présence occasionnelle de germes (Entérocoques, Escherichia Colis) l'injection d'hypochlorite de sodium existante sera conservée pour assurer ainsi une désinfection de l'eau distribuée dans le réseau. La désinfection vise à tuer ou inactiver les germes pathogènes, qui peuvent se trouver dans l'eau, susceptibles de causer des maladies infectieuses chez l'homme. Cette désinfection à l'eau de javel sera asservie au débit entrant.

La stérilisation est obtenue par un traitement chimique. Les réactifs utilisés ont été le chlore et certains composés chlorés, le brome, l'iode, l'ozone et le permanganate de potassium.

I.4.2 Stockage de l'eau

Le stockage de l'eau s'effectue dans des réservoirs situés généralement en hauteur : bassins enterrés au sommet des collines ou châteaux d'eau. Ils fonctionnent selon le principe des vases communicants pour assurer une pression régulière et suffisante au sein du réseau en fonction du rythme de consommation. Ils constituent aussi une réserve de sécurité en cas d'incident sur le réseau ou de hausse anormale de la consommation. Pour pouvoir satisfaire à tout moment la demande en eau potable des abonnés, un réservoir de stockage d'une capacité de 3600 m³ a été créé sur le lieu de traitement en forme de bache aérienne; une réserve qui permet de gérer les pointes de consommation en différents points du réseau parmi la quantité réservée à SIDER ELHADJAR.

I.5 Propriétés générales d'une eau de boisson et celle destinée à la consommation humaine

L'eau de boisson produite et celle destinée à la consommation humaine on passant par les différentes étapes citées ci-dessus soumis à une certaine norme d'analyse et contrôle des différents paramètres spécifiques plus précisément ceux relatifs au procédé de coagulation.

L'eau de boisson peut être définie en se référant à l'OMS comme une eau ne contenant en quantité dangereuse ni substances chimiques; ni germes nocifs à la santé humaine. Elle doit être agréable à boire que les circonstances le permettent, pour être distribué, Une eau destinée à la consommation humaine est considérée comme potable si elle répond à des exigences de qualité bien définies sur le plan de ses caractéristiques physico- chimique et bactériologiques qui ne doivent pas porter atteinte à la santé du consommateur pour la qualité des eaux distribuées, on se réfère essentiellement à deux aspects :

- ✓ La satisfaction de l'utilisateur, subjective car fondée essentiellement sur la quantité organoleptique et visuelle : « L'eau doit être aussi agréable à boire que les circonstances le permettent. »;
- ✓ Et, surtout, la composition et sa compatibilité avec l'hygiène et la protection de la santé publique (qualité sanitaire).

L'eau de distribution publique est définie par des normes nationales découlant des travaux de l'Organisation mondiale de la santé (OMS) et de directives européennes. L'OMS dans ses directives établies en 1983 a regroupé les paramètres concernant la qualité de l'eau en cinq catégories :

- La qualité microbiologique;
- Les composés inorganiques ayant une conséquence sur la santé ;
- Les composés organiques ayant une conséquence sur la santé ;
- La qualité esthétique (transparence, absence de coloration) ;
- Les constituants radioactifs,

I.5.1 Les différents paramètres spécifiques qui commandent la qualité des eaux de consommation.

I.5.1.1 Les paramètres organoleptiques

➤ La couleur

La couleur de l'eau est dite vraie ou réelle lorsqu'elle est due aux seules substances en solution, elle est dite apparente quand les substances en suspension y ajoutent leur propre coloration. L'eau de boisson ne doit pas présenter une coloration qui dépasse 15 mg/l à l'échelle platine.

➤ La turbidité

La turbidité correspond, après filtration à 0.45 µm, aux particules fines présentes dans le filtrat, appelées colloïdes et qui rendent l'eau trouble. La turbidité exprime l'atténuation d'un faisceau de lumière blanche au contact de particules colloïdes. Elle est d'autant plus élevée que la densité des particules contenue dans l'eau est forte. Il existe plusieurs échelles pour la mesurer. Une des plus utilisées est l'unité NFU (Nephelometric Turbidity Unit). La turbidité des eaux alimentaires doit être inférieure à **1 NFU** au point de mise en distribution et inférieure à **2 NFU** au robinet du consommateur. Sa valeur considérée comme très forte pour des eaux de surface présentant une turbidité supérieure à 3 000 NFU. Les valeurs peuvent atteindre 10 000 NFU dans des bassins versants fortement érodés ou lors d'épisodes pluvieux. Dans ce cas le paramètre n'est plus significatif et il faut revenir aux MES.

La turbidité est également considérée comme un indicateur indirect du risque microbiologique, car certains microorganismes tels que les virus peuvent y être associés (par adsorption). La turbidité d'une eau est l'inverse de sa transparence ; elle est due à la présence des matières en suspension dans l'eau. L'eau de boisson ne doit présenter une turbidité **>5NTU**.

➤ L'odeur

Une eau destinée à l'alimentation doit être inodore. La présence d'une odeur est un signe de pollution ou de la présence de la matière organique en décomposition.

I.5.1.2 Les paramètres physico-chimiques

➤ La température

La température influe considérablement sur la vitesse des réactions chimiques ou biochimiques dans l'eau, la température optimale de l'eau de consommation est entre 9°C et 12°C. Elle ne doit pas dépasser 15°C.

➤ PH

Le PH influe sur les phénomènes de corrosion ou d'entartrage des conduites. On admit que le PH des eaux de consommation doit être compris entre 6.5 et 8.5

➤ La dureté

La dureté d'une eau est associée à la présence en solution et à l'état dissous des ions bivalents calcium (Ca^{2+}) et magnésium (Mg^{2+}).

I.5.1.3 Les matières organiques

➤ Le carbone organique total(COT)

Le COT est défini comme la concentration en mg de carbone/l d'échantillon, c'est une méthode directe pour mesurer la teneur en matières organiques d'une eau. Il consiste à brûler l'échantillon d'eau à 950 °C sous l'action de l'oxygène gazeux et de mesurer le CO_2 dégagé, et par suite de mesurer le carbone total (organique et minéral). Le COT est une méthode rapide et présente plusieurs avantages tels que : on peut analyser l'eau quelle que soit sa qualité (de forte salinité, acide, basique, ou contient des composés toxiques). Mais malgré ses avantages, il exige un appareillage couteux [10.11].

La détermination du carbone organique total (COT) ou dissous (COD) passera donc par l'élimination du carbone minéral avant la mesure, ou par sa soustraction du carbone total. Le principe de la mesure du carbone organique dans les eaux consiste en une minéralisation par oxydation des matières organiques (M.O), suivi de la mesure du CO_2 libéré.

➤ Alcalinité

L'alcalinité d'une correspond à la présence d'ions hydroxydes, carbonates et hydrogéné carbonates.

✓ Conclusion :

Cette première partie du chapitre un a fait l'objet de la présentation des différentes étapes de production de l'eau potable dans les stations de traitement de l'eau brute acheminée à partir des barrages. Les différentes chaînes de production ont été décrites d'une façon détaillée.

I.6 Surveillance des procédés de production d'eau potable

I.6.1 surveillance industrielle : définitions et état de l'art

Les procédés de production d'eau potable ont un fonctionnement complexe qui ne peut pas être mesuré, modélisé et interprété que d'une façon partielle à cause du fait de la complexité des phénomènes mis en jeu mais aussi par leur nature non stationnaire et aléatoire ; ils peuvent donc avoir des fonctionnements différents d'une expérience à l'autre, pour les mêmes conditions expérimentales.

Quand on parle de surveillance des eaux potables, il s'agit en fait de connaître l'état de l'eau en continu (à chaque instant) à partir des différents paramètres ayant trait à sa qualité. Traditionnellement, on procède généralement à un certain nombre de mesures utiles des différents paramètres de l'eau brute pour le test de sa qualité tels que : le contrôle bactériologique, le contrôle de désinfection, et le contrôle physico-chimique (pH, température, turbidité, conductivité, oxygène dissous), et pour décider après sur son état propre, et par la suite chercher les techniques et méthodes pour la rendre potable. Ces méthodes sont déterminées à l'aide d'un essai expérimental appelé Jar-test.



Figure I.3 Essai « Jar-Test »

En général dans les usines importantes, un seul essai est effectué par jour [12]. L'opérateur fera un nouvel essai entre temps pour changer la dose du coagulant uniquement si la qualité de l'eau traitée se dégrade. L'inconvénient de cette technique est qu'elle nécessite de façon non stop des interventions et des déplacements sur site de l'opérateur. Cette technique ne

permet pas de suivre finement l'évolution de la qualité de l'eau brute et aura le désavantage d'avoir un temps de réponse relativement long. En voici tout l'intérêt de disposer d'un contrôle automatique de ce procédé pour une meilleure efficacité de traitement et une réduction des coûts d'exploitation.

Le contrôle de la qualité de l'eau au niveau des usines de production doit se faire de façon immédiate en se basant sur une surveillance continue des paramètres descripteurs de la qualité à travers toute une chaîne de traitement qui assure à la fois la surveillance et le contrôle. L'automatisation de ces procédés pour la surveillance a permis de s'affranchir de certaines faiblesses humaines, d'accroître la fiabilité et d'améliorer les conditions d'exploitation et des performances de l'installation par la suppression d'un maximum de tâches répétitives et pénibles pour l'opérateur. Automatiser les procédés de traitement peut avoir des objectifs divers, les plus fréquents sont d'augmenter les performances du système de production, de garantir la qualité du produit, de la surveillance des installations et des équipements de traitement, de diminuer les coûts de production et d'améliorer la sécurité de l'installation industrielle et de son environnement. De plus, l'automatisation couplée à un stockage d'informations, permet d'envisager des études statistiques de données recueillies ouvrant la voie à des études de modélisation et donc d'optimisation des procédés de traitement. Plus récemment, l'automatisation et la supervision par ordinateur ont conduit à une augmentation du confort en permettant une meilleure maîtrise même à distance d'un grand nombre d'informations, ce qui permet de simplifier les tâches d'exploitation, de surveillance, de maintenance et de gestion [13]. L'utilisation des techniques issues du domaine de l'intelligence artificielle apparaît comme la principale alternative pour aborder ces problèmes lorsqu'il est nécessaire de prendre en compte l'intervention des experts du domaine ou de traiter l'information de nature qualitative.

I.6.2 Définitions des termes de base utilisés en surveillance

Les différentes définitions trouvées dans plusieurs travaux [14, 15, 16] laissent imaginer d'établir un lexique sur les termes et les terminologies qui seront utiles et intéressantes pour la compréhension de certaines notions sur la surveillance industrielle.

- **Fonctionnement normal d'un système** : Un système est dit dans un état de fonctionnement normal lorsque les variables le caractérisant (variables d'état, variables de sortie, variables d'entrée, paramètres du système) demeurent au voisinage de leurs valeurs nominales. Le système est dit défaillant dans le cas contraire [13].
- **Dégradation** : une dégradation représente une perte de performances d'une des fonctions assurées par un équipement [14]. Si les performances sont au-dessous du seuil d'arrêt

défini dans les spécifications fonctionnelles de cet équipement, il n'y a plus dégradation mais défaillance.

- **Défaillance** : une défaillance est l'altération ou la cessation de l'aptitude d'un ensemble à accomplir sa ou ses fonctions requises avec les performances définies dans les spécifications techniques. En général, une défaillance est la cause d'une anomalie [14,16].

- **Un défaut** : se définit comme une anomalie du comportement d'un système sans forcément remettre en cause sa fonction.

- **Panne** : une panne est l'inaptitude d'une entité (composant ou système) à assurer une fonction requise [17].

- **Surveillance**

La surveillance est un dispositif passif, informationnel qui analyse l'état du système et fournit des indicateurs. La surveillance consiste notamment à détecter et classer les défaillances en observant l'évolution du système puis à les diagnostiquer en localisant les éléments défaillants et en identifiant les causes premières.

La surveillance se compose donc de deux fonctions principales qui sont la détection et le diagnostic (figure I.4).

- **La détection**

C'est l'étape qui décide si le système est soumis à un défaut ou pas.

- **Le diagnostic**

Le diagnostic se décompose en deux fonctions la localisation et l'identification. La localisation permet de déterminer le sous-ensemble fonctionnel défaillant tandis que l'identification de la cause consiste à déterminer les causes qui ont mené à une situation anormale.

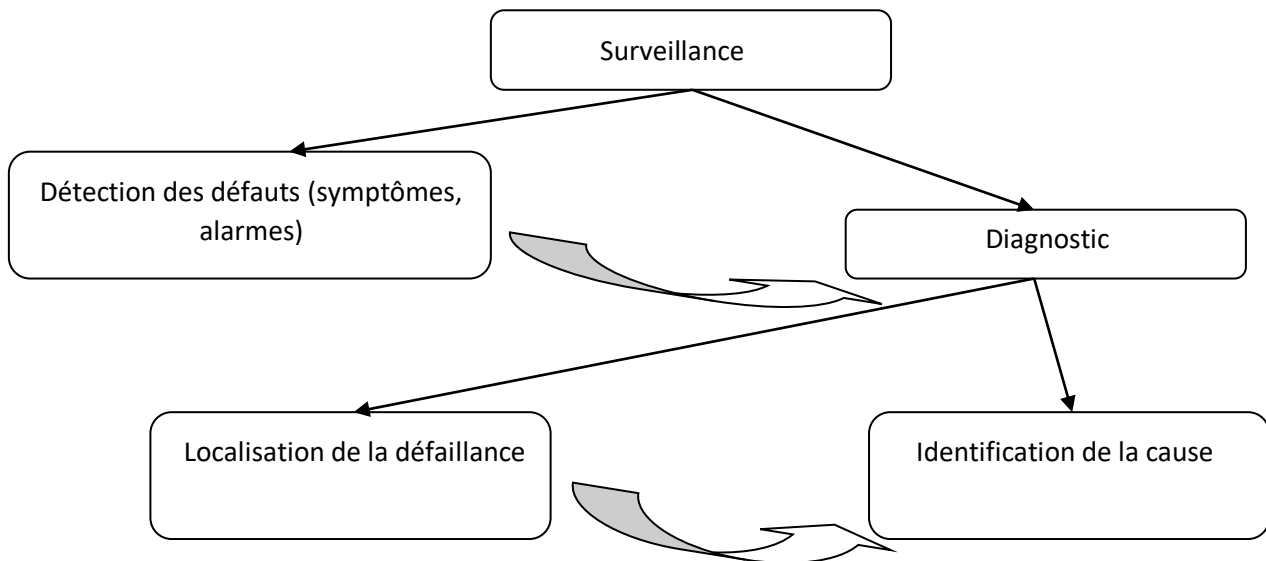


Figure I.4 Composantes de la surveillance industrielle

I.6.3 Méthodes de surveillance

Les méthodes de surveillance sont généralement divisées en deux groupes (Dash *et al.*, 2000) :

Méthodes de surveillance *avec modèle*

Méthodes de surveillance *sans modèle*

- Les premières se basent sur l'existence d'un modèle formel de l'équipement et utilisent généralement des techniques de l'Automatique (Combacau, 1991).

- Les techniques de *surveillance sans modèle* sont divisées en deux parties.

- ✓ La première partie correspond aux outils statistiques de *Traitement du Signal* qui sont généralement qualifiés d'outils de traitement de bas niveau, parce qu'ils sont en contact direct avec le signal capteur, et ne servent généralement que pour la génération d'alarmes brutes, sans aucune information concernant leur signification. Les outils statistiques établissent des tests sur les signaux d'acquisition, tests qui ne sont capables d'assurer que la fonction détection de défaillances.
- ✓ La deuxième partie est celle des techniques dites de haut niveau et qui sont plutôt orientées vers la communication avec l'expert. Celles-ci représentent les techniques de *l'Intelligence Artificielle* (IA) et servent comme outil de base pour l'aide à la décision. Leur réponse est donc plus élaborée que celle des techniques de bas niveau et elles sont capables de détecter, interpréter (association à un mode) et diagnostiquer les défaillances.

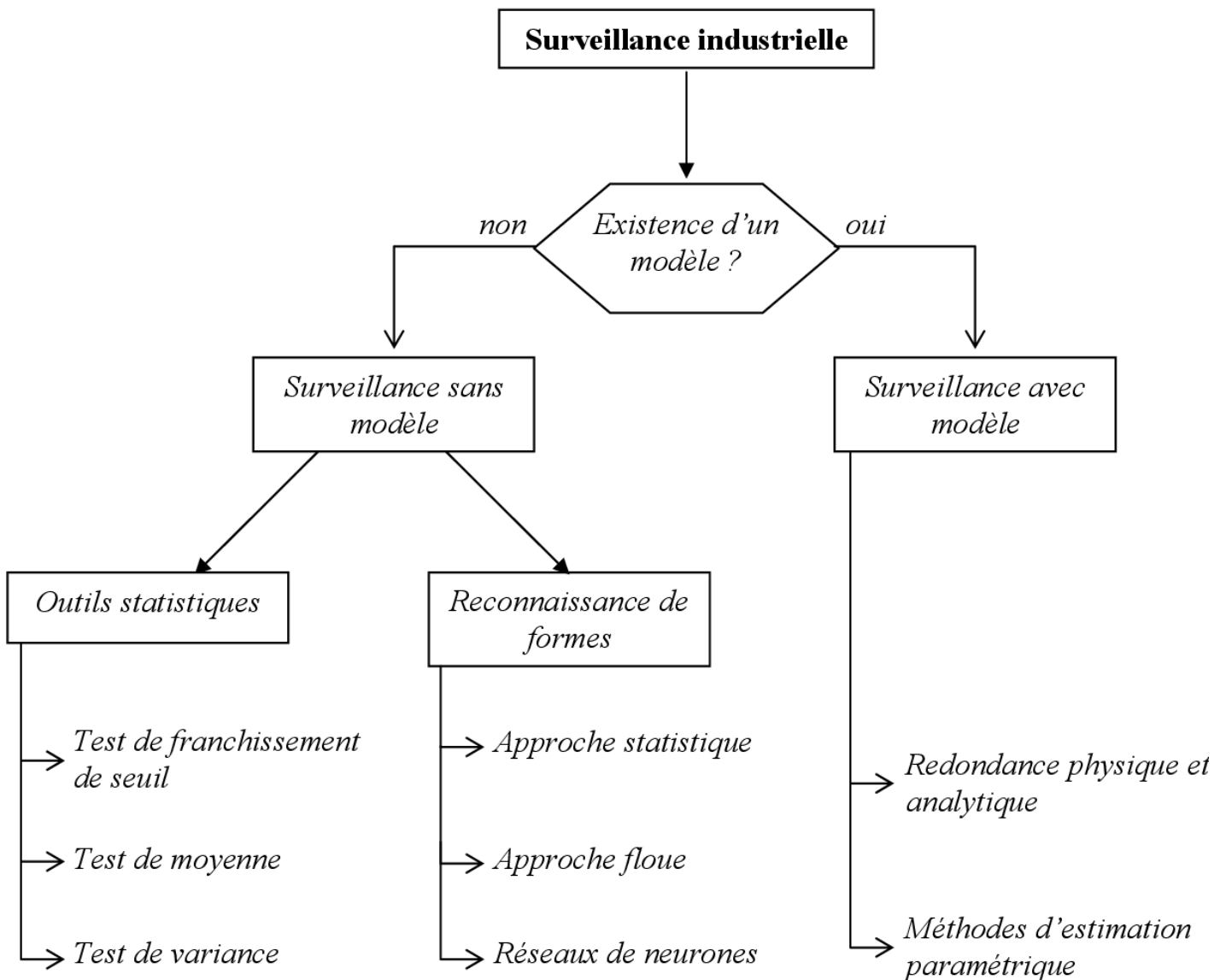


Figure I.5 Classification globale des méthodes de surveillance

I.7 Conclusion

Ce premier chapitre a servi à présenter les différentes étapes d'un modèle typique de potabilisation appliquée à une eau de surface, la plus complète et la plus innovée. Nous avons décrit les différents paramètres spécifiques qui commandent la qualité des eaux de consommation, nous avons également énumérer l'ensemble des paramètres influençant le bon fonctionnement du procédé de coagulation, la dernière partie de ce chapitre est consacrée à la surveillance des procédés de production d'eau potable et les différentes méthodes de surveillance.

Le chapitre suivant est consacré à la partie de l'apprentissage profond et les réseaux de neurones convolutifs (CNN).

Chapitre II Deep learning et CNN

II .1 Introduction

L'intelligence artificielle c'est bien le domaine qui est en plein croissance récemment. De la découverte de visages, aux assistants conversationnels, en passant par les véhicules autonomes et les systèmes de conseils d'achats sur internet et le domaine de traitement et surveillance de la qualité d'eau potable , ces nouvelles technologies approprient notre quotidien. Et dans ce large domaine, un type de méthodes fait plus particulièrement parler de lui : le Deep Learning.

Le Deep Learning, ou apprentissage profond, est par exemple à l'origine du très médiatique succès, Le potentiel de ces technologies semble étonnant. D'ailleurs, dans les dernières années, l'Institut Gartner les a placées tout en haut, au pic des souhaits qu'elles peuvent générer.

Mais entre le potentiel et ce que l'on doit réellement en attendre, entre Deep Learning et autres domaines de l'intelligence artificielle, il n'est pas forcément facile de s'y retrouver...

En effet, le Deep Learning est déjà exploré depuis quelque temps en domaine de traitement et surveillance de la qualité d'eau potable. D'abord par le monde de la Recherche puis par celui de la Recherche et du Développement. A l'heure où les premières applications commerciales vont arriver sur le marché, il nous semble important de pouvoir instruit : comprendre ce dont il s'agit, quelles en sont les applications.

C'est donc ce que nous souhaite proposer comme application concernant notre sujet de surveillance de la qualité d'eau basés sur des techniques d'apprentissage profond cas de CNN.

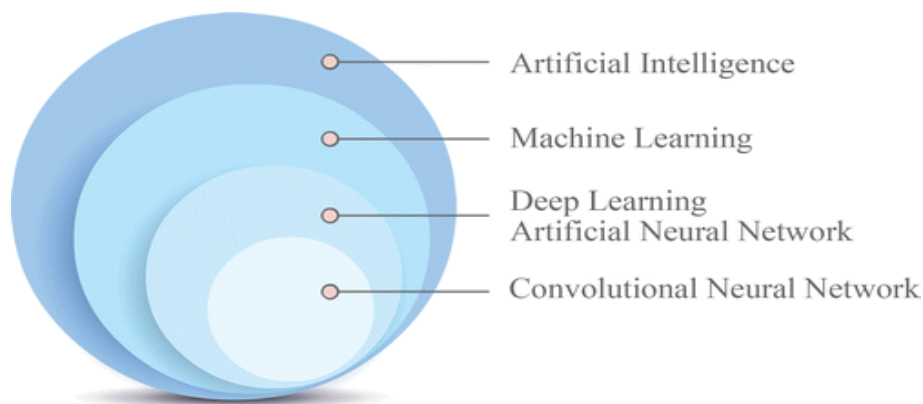


Figure II.1 Intelligence artificielle, machine learning e, deep learning et convolutional network

Le schéma ci-dessus (Figure II .1) évoque les différentes zones de l'évolution l'intelligence artificielle on arrive au système le plus développé dans ce domaine (convolutional Neural Network) cas de l'application utilisé de notre travaille de mémoire, ce chapitre est dédié à la clarification des différents passages de l'apprentissage approfondi arrivant à la représentation de la méthode utilisée pour la réalisation de la surveillance de la qualité d'eau potable celle de CNN.

II .2 L'IA, Intelligence artificielle

Un vaste domaine scientifique, notion apparue fin des années cinquante début des années soixante, l'intelligence artificielle peut-être définie comme l' « ensemble de techniques permettant à des machines d'accomplir des tâches et de résoudre des problèmes normalement réservés aux humains et à certains animaux » [18].

Il s'agit donc de techniques qui visent à reproduire, imiter, simuler l'intelligence, ou en tout cas les capacités que l'on peut associer à ce terme : capacité à percevoir son environnement, à interagir avec lui, à communiquer au moyen d'un langage, à mettre en œuvre un raisonnement et à résoudre des problèmes, à planifier des actions afin d'atteindre un but, à représenter la connaissance, à appliquer cette connaissance au moment opportun...

Comme dit précédemment, le domaine est donc très vaste et s'organise en de multiples branches qui associent sciences cognitives, mathématiques, électronique, informatique...

Les évolutions en matière d'IA ont suivi différents courants. Dans les années quatre vingt, par exemple, les « systèmes experts » correspondent à une approche basée sur la capacité à reproduire un raisonnement logique. Ces programmes permettent de répondre à des questions ou problèmes dans un domaine très spécifique, où la machine sert à mettre en œuvre des règles logiques décrites par des experts humains et implémentées dans des bases de connaissance. Ce type d'approche est encore assez proche de l'informatique classique où il faut identifier à l'avance toutes les informations à fournir au système. Elle peut vite être limitée pour des cas complexes où de multiples paramètres doivent être pris en compte.

Une autre approche consiste plutôt à s'inspirer du fonctionnement du cerveau, sa façon de modéliser et de s'adapter à l'environnement. La **Machine Learning**, et donc le **Deep Learning**, relèvent plutôt de cette approche.

II.3 La machine Learning :

Apparu à la fin des années cinquante, la machine Learning correspond au fait de donner la capacité d'apprendre à la machine, par elle-même, quelle que soit la situation, sans que l'on ait formellement à écrire (ni même à connaître) toutes les règles.

La machine learning s'inscrit dans un processus de traitement d'un problème complexe, qui peut se présenter comme suit :

- On a identifié et clairement posé un problème à résoudre.
- On a identifié un résultat à atteindre en sortie afin d'accomplir une tâche (une prédiction, une identification, une recommandation, la réponse à une question, une traduction...).
- On a recensé et préparé des données à exploiter (des images, du texte, des données numériques, des signaux sonores...) : les données d'entrée.
- On veut produire automatiquement le résultat attendu à partir des données d'entrée. Pour cela, un modèle de traitement automatique doit être trouvé et le but est que ce soit la machine qui trouve par elle-même les paramètres de ce modèle c'est-à-dire les bons réglages. Pour cela, elle a besoin d'un apprentissage, permis par la mise en place d'un algorithme d'apprentissage, sur la base de données d'entraînement.
- Le modèle est testé, évalué et optimisé.
- Le modèle est mis en production : la machine généralise ce qu'elle a appris sur des données nouvelles. Les tâches préalablement identifiées peuvent être accomplies.
- Le modèle est mis à jour régulièrement.

La phase d'apprentissage repose sur un algorithme d'apprentissage c'est-à-dire la mise en place d'une architecture et d'un programme informatiques qui permettent à une machine de recevoir des données d'entrée, d'effectuer une série de traitements utilisant ces données d'entrée, de produire un résultat en sortie et, surtout, de s'améliorer pour produire ce résultat. Ce dernier point suppose qu'un objectif de résultat (et éventuellement une tolérance par rapport à l'atteinte de cet objectif) soit communiqué au système et que des données d'entraînement et les moyens de mesurer ses performances lui soient fournis. C'est en cherchant à améliorer ses performances sur les données d'entraînement que la machine va apprendre. Une fois que l'apprentissage sera terminé, la « machine » pourra produire le résultat de façon autonome sur des données d'entrée qu'elle n'aura encore jamais rencontrées.

Mais, selon le résultat visé, les données dont on dispose au départ et le contexte, il existe plusieurs façons de faire apprendre à une machine. On distingue notamment **l'apprentissage supervisé, l'apprentissage non supervisé** (d'autres méthodes existent mais nous ne les développerons pas ici)

II.3.1 Apprentissage supervisé : apprendre par l'exemple

Dans le cas de l'apprentissage supervisé, le système est avisé dans son apprentissage. On lui indique le type de résultat à aboutir en l'alimentant d'exemples. Pour cela, on lui fournit des données d'entrée pour lesquelles le résultat est connu et communiqué au système. Le but est qu'il puisse ensuite généraliser ce qu'il a appris pour des données non connues. Par exemple, si le système doit apprendre à reconnaître des qualité des paramètres d'eau c'est bonne où autre dans une image, on lui fournit des images où la est la qualité bonne ,moyenne ,mauvaise, très mauvaise , sont signalées et où le label « bonne qualité » est associé. On parle ainsi de données étiquetées ou annotées. Le jeu de données d'entraînement annotées permet au système de calculer ses erreurs en comparant ses résultats avec les résultats connus et ainsi d'ajuster le modèle pour progresser. Une partie des données annotées (non utilisées pendant l'entraînement) pourra également servir à vérifier l'efficacité du modèle, une fois l'apprentissage terminé.

II.3.2 Apprentissage non supervisé : apprendre en explorant

Dans le cas de l'apprentissage non supervisé, on ne donne pas d'exemples de résultats attendus au système. Seules les données d'entrée sont fournies et le système doit apprendre, de façon autonome, la meilleure façon d'explorer les données. Il doit chercher à identifier dans le jeu de données une façon de les structurer (trouver des modèles ou « patterns ») ou encore à extraire des caractéristiques particulières. La performance permettant l'ajustement du modèle est alors appréciée grâce à des indicateurs objectifs, comme, par exemple, des calculs de variabilité intra ou interclasses.

Notons que, si l'apprentissage non supervisé est une sorte de « graal pour les chercheurs en intelligence artificielle » [19] et si les autre apprentissage ouvre de belles perspectives, l'apprentissage supervisé reste de loin le plus exploré et pratiqué à l'heure actuelle, notamment pour des applications de régression ou de classification pour lesquelles il donne de meilleurs résultats.

II .4 Le Deep Learning : un sous-ensemble de méthodes d'apprentissage machine

En fonction des différentes problématiques posées, les développeurs et chercheurs en IA ont exploré et mis au point une plate-forme d'ensemble d'algorithmes. Parmi les plus connues, on peut citer les réseaux de neurones convolutifs , en anglais Convolutional Neural Network (CNN).

Le Deep Learning peut être vu comme une sous-catégorie des « réseaux de neurones ». En effet, il s'agit d'un ensemble d'outils et de méthodes d'apprentissage automatique basés sur l'utilisation « avancée » de réseaux de neurones. On utilise les mots « deep » ou « profond » en référence au nombre de couches de neurones qui constituent ces réseaux : plus le nombre de couches est grand plus le réseau est profond (Figure II .2) et plus il permet de traiter des problèmes complexes. Les réseaux de neurones « simples », c'est-à-dire de profondeur plus délicat, sont eux nommés «Shallow Networks ». [20]

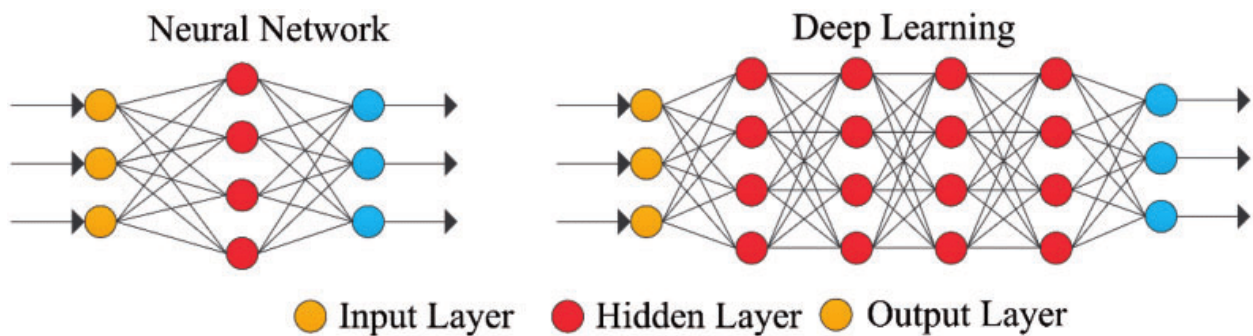


Figure II.2 Réseau de neurones simple

Réseau de neurones profond

L'utilisation des réseaux de neurones en apprentissage automatique n'a rien de nouveau. Le tout premier algorithme, le Perceptron, a été proposé dès 1957 par Franck Rosenblatt.

Les réseaux de neurones artificiels s'inspirent (sans en être la copie) du fonctionnement des réseaux de neurones des cerveaux humains ou animaux. Le cerveau est constitué d'un très grand nombre de cellules spécialisées, les neurones, étroitement liées entre elles par des milliers d'interconnexions. Chaque neurone est capable de traiter de multiples informations en provenance de son environnement et de réagir en créant des messages électrochimiques. Un neurone seul ne réalise que des opérations élémentaires, mais associés en réseau, les neurones permettent aux hommes et aux animaux d'accomplir des tâches cognitives

extrêmement complexes. Les réseaux de neurones artificiels reprennent ce principe en interconnectant de multiples unités élémentaires, les neurones formels, dans le but de résoudre des problèmes. Ils sont organisés en couches où chaque couche répond à un niveau d'abstraction d'un problème.

Dans le cas des réseaux multicouches (de type « Perceptrons multicouches ») :

- La première couche correspond aux données d'entrée (Figure II.3)
- La dernière couche correspond aux sorties attendues, par exemple les différentes classes dans un problème de classification comme notre cas (les quatre classes de qualité d'eau potable).
- Les couches intermédiaires sont appelées « couches cachées ». Les sorties d'une couche deviennent les entrées de la couche suivante (on va détailler les couches sur la présentation de CNN ci-après).

Mais, de leur invention jusque dans les années 2000, les réseaux de neurones utilisés étaient des réseaux peu profonds : au maximum deux couches cachées. Ils étaient utilisés parmi la gamme des outils de machine learning mais pour des usages limités du fait de la nécessité de disposer de grandes puissances de calculs. Aussi, d'autres algorithmes, tels les CNN ou les Random Forest, jugés plus performants, leur étaient souvent préférés.

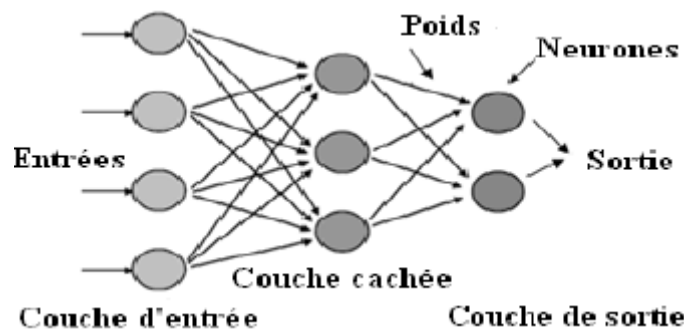


Figure II.3 Les types de couches d'un réseau de neurones

En 2012, les résultats extraordinaires acquis lors d'une rivalité d'analyse d'images agitent directement le paysage scientifique.

ILSVRC est une compétition annuelle où des groupes de recherche apprécient leurs algorithmes de traitement d'images sur le jeu de données ImageNet (une des plus grande base mondiale d'images annotées). En 2011, les plus faibles taux d'erreur de classification de la compétition

ILSVRC étaient d'environ 25%. En 2012, un algorithme permet de faire baisser ce record à 16%. Cet algorithme, **SuperVision**, est basé sur un réseau de neurones « profond ».

Il était impossible d'ajouter un nombre suffisant de couches cachées pour obtenir d'excellents résultats d'apprentissage et classification à l'époque mais avec les progrès matériels (des processeurs graphiques) et l'internet et la constitution de gigantesques bases de données annotées il est devenu possible de minimiser au maximum l'erreur et le temps à consacrer à l'apprentissage devenait aussi acceptable. Evidemment, les algorithmes ont également continué à progresser. On ne peut pas ajouter de plus en plus de couches sans mettre au point de nouvelles techniques : nouvelles architectures, nouveaux types de couches, nouvelles méthodes de calculs... C'est cet ensemble de nouvelles méthodes (et celles à l'état de recherche) qui, aujourd'hui, constituent le Deep Learning.

II.5 La méthode de réseaux neurones convolutionnels (CNN).

Avec l'approche conventionnelle de l'apprentissage automatique, ces propriétés ont été identifiées manuellement, en s'appuyant sur l'enregistrement en mémoire des caractéristiques des motifs d'entrée : en substance, les entrées connues ont été introduites dans le réseau neuronal et pour chacune d'entre elles, ce qui en fait un très long travail.

Les réseaux convolutifs de l'anglais Convolutional Neural Network ont été pilotés initialement par Fukushima [20], il a résulté une structure du réseau nerveux hiérarchique inspirée par le travail de recherche de Hubel [21]. Lecun [22], les a généralisés pour classer les chiffres avec gain optimal et pour reconnaître les numéros de contrôle manuscrit par LeNet-5 qui est montré à la (FigureII.4). Ciresan [23] a utilisé les réseaux convolutifs et réalisé les meilleures prouesses dans la littérature pour la découverte d'objets variés pour des bases de données d'images : MNIST, NORB, CIFAR10 et l'ensemble de données ImageNet.

Les réseaux CNN se concentrent principalement sur le fait que l'entrée sera composée d'images sous la forme d'une matrice de pixels. Elle a deux dimensions pour une image aux niveaux de gris

Cela permet d'ajuster l'architecture à mettre en place pour répondre au mieux à la nécessité de traiter un type de données spécifiques. Les réseaux neuronaux convolutifs diffèrent des autres formes de réseaux neuronaux artificiels ou les idées sur le type propre d'entrées sont utilisées, cela permet de mettre une architecture réseau beaucoup plus performante. [24] [25] [26]

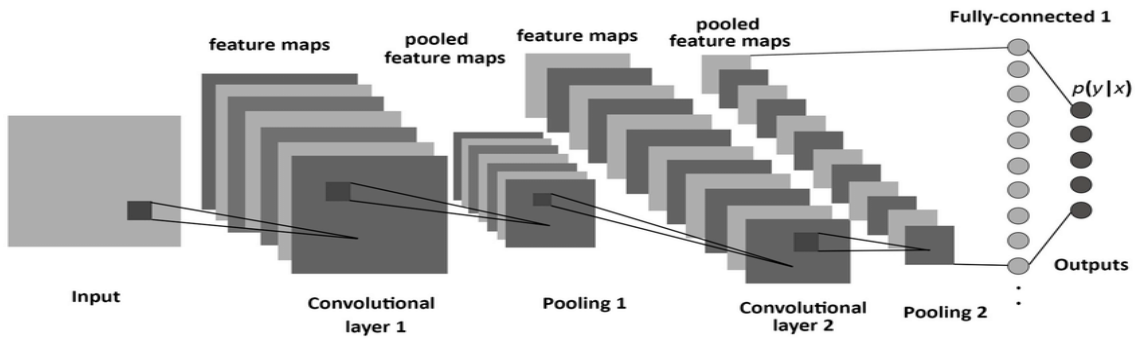


Figure II. 4 Réseau neuronal convolutif

II.5.1 Architecture globale de CNN

Les CNN sont composés de trois types de couches : des couches convolutives, des couches de meeting et des couches entièrement connectées. L'architecture CNN est formée par une accumulation de couches de traitement indépendantes. Une architecture CNN réduite pour la classification des images est illustrée à la (figure II .5). [24] [25]

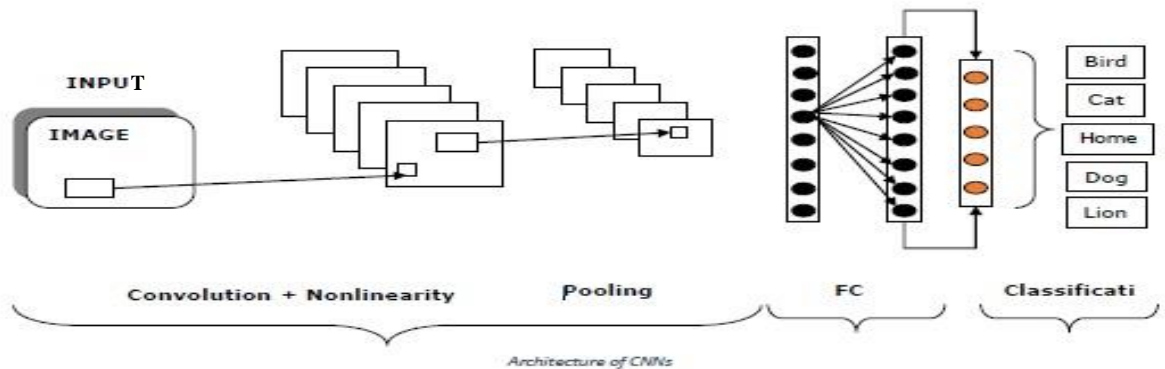


Figure II.5 Architecture simple de CNN

II.5.1.1 Couche convolutive

Les couches convolutives forment le noyau du réseau convolutif. Ces couches se composent d'une grille rectangulaire de neurones qui ont un petit champ réceptif étendu à travers toute la profondeur du volume d'entrée. Ainsi, la couche convolutive est juste une convolution d'image de la couche précédente, où les poids spécifient le filtre de convolution.

La couche convolutive déterminera la sortie des neurones qui sont connectés aux régions locales de l'entrée par le calcul du produit scalaire entre leurs poids et la région connectée au volume

d'entrée. ReLu vise à appliquer une fonction d'activation « élémentaire » telle qu'une fonction sigmoïde à la sortie de l'activation produite par la couche précédente. [24] [25] [26]

II.5.1.2 Couche de pooling

Après chaque couche convolutive, il peut y avoir une couche de pooling. Cette couche sous échantillonne le long de la dimensionnalité spatiale de l'entrée donnée, L'image d'entrée est découpée en une série de rectangles de n pixels de côté ne se chevauchant pas (pooling). Chaque rectangle peut être vu comme une tuile. Le signal en sortie de tuile est défini en fonction des valeurs prises par les différents pixels de la tuile, ce qui réduira davantage le nombre de paramètres au sein de cette activation. Il y a plusieurs façons de faire cette mise en commun, comme prendre la moyenne (average) ou le maximum, ou une combinaison linéaire prise par des neurones dans le bloc. Par exemple, la (figure. II.6) montre le max pooling sur une fenêtre 2×2 . [24] [25] [26]

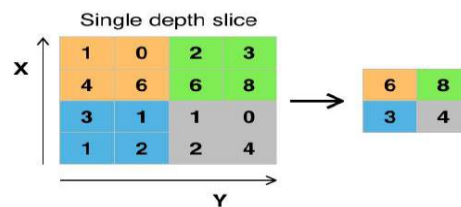


Figure II.6 Max pooling

II.5.1.3 Couche totalement connectée

Il est plus juste que le raisonnement de haut niveau dans le réseau neuronal se fait via des couches entièrement connectées, suite à plusieurs opérations de couches de convolution et de max-pooling,

Dans les réseaux de neurones convolutifs, chaque couche agit comme un filtre de détection pour la présence de caractéristiques spécifiques ou de motifs présents dans les données d'origine. Les premières couches d'un réseau convolutif détectent des caractéristiques qui peuvent être reconnues et traduites plus facilement. Les couches suivantes détectent de plus en plus des caractéristiques plus profondes. La dernière couche du réseau convolutif est capable de faire une classification ultra-spécifique en combinant toutes les caractéristiques spécifiques détectées par les couches précédentes dans les données d'entrée.

Les couches totalement connectées font les mêmes tâches que celles des ANN standard et tenteront de produire des notes de classe à partir des activations, pour les utiliser pour la classification. Il est également conseillé d'utiliser ReLu entre ces couches pour améliorer les performances. [24] [25] [26]

II.5.1.4 Couche de correction (ReLU)

A base d'une couche qui va exécuter une fonction mathématique (fonction d'activation) sur les signaux de sortie, on peut améliorer l'efficacité du traitement en insérant cette dernière entre les couches de traitement.

La fonction d'activation : Non Linearity ReLU (Rectified Linear Units) ReLU :

$$F(x) = \max(0, x)$$

Cette fonction force les neurones à retourner des valeurs positives. La couche de correction ReLU remplace donc toutes les valeurs négatives reçues en entrées par des zéros. Elle joue le rôle de fonction d'activation. [24] [25]

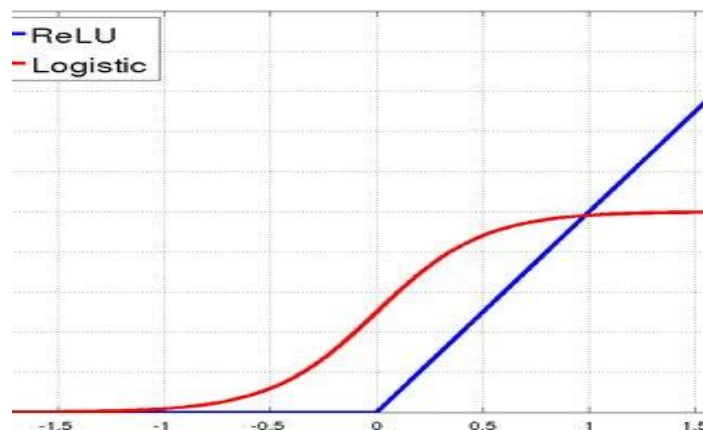


Figure II.7 courbe la fonction ReLU

Souvent, la correction Relu est préférable et le plus utilisée suite aux résultats performantes, néanmoins il existe d'autre forme fonction de correction, on évoque en ce qui suit quelques unes :

- ◆ La correction par la fonction sigmoïde, résolue par la formule (1), elle associe à tout réel x une valeur appartenant à l'intervalle] 0,1[(figure II.8).

$$f(x) = \frac{1}{(1+e^{-x})} \quad \dots\dots\dots (1)$$

C'est l'une des premières fonctions utilisées par les experts du domaine qui offre en plus l'avantage de normaliser les entrées qu'elle reçoit mais qui depuis a beaucoup perdu en popularité due aux problèmes suivants :

- L'incapacité de la fonction Sigmoid d'associer la valeur 0 à une entrée empêchant ainsi d'éliminer des inputs lors du calcul des sorties des couches inférieures de neurones ce qui alourdit les calculs.

- L'inefficacité de l'algorithme de descente de gradient lorsqu'il est associé à cette fonction notamment à cause de la saturation qu'il finit par provoquer (inexistence de la valeur 0).

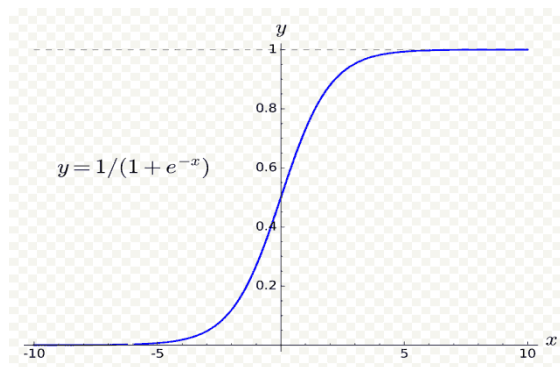


Figure II.8 courbe la fonction sigmoid

- ◆ La correction par tangente hyperbolique, elle est résolue par la formule (2), elle associe à tout réel x une valeur entre]-1, 1[, représenté ci-dessous (figure II.9)

$$f(x)=\tanh(x) = \frac{1-e^{-2x}}{1+e^{-2x}} \dots\dots\dots (2)$$

Cette fonction corrige la faille la plus grave de la fonction sigmoid à savoir sa faible compatibilité avec l'algorithme de la descente de gradient à la base du processus d'apprentissage cependant l'emploi de cette fonction d'activation ne suffit toujours à se prémunir du problème du *Vanishing gradient problème* (affaiblissement de gradient) qui consiste en l'affaiblissement des changements apportées aux poids par l'algorithme de descente de gradient lors de l'entraînement qui se traduit par le fait que le réseau de neurone n'apprenne plus.

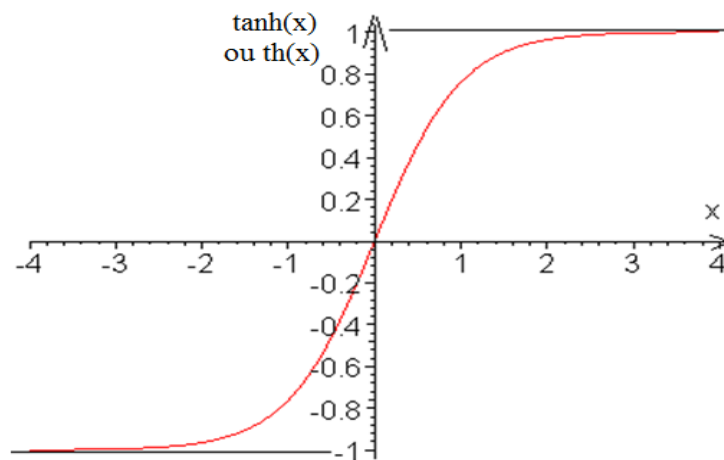


Figure II.9 courbe la fonction de tangente hyperbolique

Reste, la correction Relu est la meilleure comme nous avons indiqué auparavant, car il en résulte la formation de réseau neuronal plusieurs fois plus rapide, sans faire une différence significative à la généralisation de précision.

Les CNNs utilisent plus d'hyper paramètres qu'un MLP standard. Même si les règles habituelles pour les taux d'apprentissage et des constantes de régularisation s'appliquent toujours, il faut prendre en considération les notions de nombre de filtres, leur forme

II.5.1.5 Nombre de filtres :

Comme la taille des images intermédiaires diminue avec la profondeur du traitement, les couches proches de l'entrée ont tendance à avoir moins de filtres tandis que les couches plus proches de la sortie peuvent en avoir davantage. Pour égaliser le calcul à chaque couche, le produit du nombre de caractéristiques et le nombre de pixels traités est généralement choisi pour être à peu près constant à travers les couches. Pour préserver l'information en entrée, il faudrait maintenir le nombre de sorties intermédiaires (nombre d'images intermédiaire multiplié par le nombre de positions de pixel) pour être croissante (au sens large) d'une couche à l'autre.

Le nombre d'images intermédiaires contrôle directement la puissance du système, dépend du nombre d'exemples disponibles et la complexité du traitement.

II.5.1.6 Forme de filtres :

Les formes de filtre varient grandement dans la littérature. Ils sont généralement choisis en fonction de l'ensemble de données. Les meilleurs résultats sur les images de MNIST (28x28) sont habituellement dans la gamme de 5x5 sur la première couche, tandis que les ensembles de données d'images naturelles (souvent avec des centaines de pixels dans chaque dimension) ont tendance à utiliser de plus grands filtres de première couche de 12x12, voire 15x15.

II.6 Conclusion

Dans ce chapitre, nous avons présenté, en premier lieu, les méthodes d'intelligence artificielle, puis la machine learning, le deep learning et les réseaux de neurone convolutifs. Les notions importantes qui sont en relation avec l'apprentissage profond (définition, architectures) ont été abordées ainsi qu'une vision générale sur l'apprentissage profond, toute on donnant en détail la méthode choisie dans notre travail de recherche qui est le CNNs (Convolutional Neural Network). Ces réseaux sont capables d'extraire des caractéristiques d'images présentées en

entrée et de classifier ces caractéristiques. Ils sont fondés sur la notion de « champs récepteurs » (receptive fields), ils implémentent aussi l'idée de partage des poids qui permettant de réduire beaucoup le nombre de paramètres libres de l'architecture. Ce partage des poids permet en outre de réduire les temps de calcul, l'espace mémoire nécessaire et également d'améliorer les capacités de généralisation du réseau.

A ce stade, nous disposons donc une solution dans le domaine de surveillance de la qualité d'eau potable à base des techniques d'apprentissage approfondi en choisissant les CNNs. Nous allons donc aborder dans le dernier chapitre la classification automatique de la qualité d'eau potable de la station de CHAIBA en utilisant les CNNs alimentés en entrée par des données prélevées pendant huit ans (période 2013-2020) et fournies par laboratoire centrale de l'ADE ANNABA et biens sûr notre repère de comparaison utilisé est la grille normative pour l'estimation la qualité de l'eau en Algérie.

Chapitre III Simulation par CNN

III.1 Introduction :

Dans cette partie, nous considérons la surveillance de la qualité de l'eau potable comme un problème de classification, où les classes correspondent aux différents états de l'eau. Le système de surveillance de la qualité de l'eau que nous proposons est basé sur traitement des données (apprentissage, classification et décision). Une base de données (figure III.1) sur les analyses d'eau de production de la station de CHAIBA ainsi que l'eau de source du barrage de CHAFIA est utilisée (plusieurs visites ont été effectués au niveau de la station de CHAIBA et au niveau de laboratoire centrale de l'ADE d'Annaba et même au niveau du barrage de CHAFIA). Les étapes de traitement (apprentissage, classification, décision) conçues dans le but de la surveillance de la qualité de l'eau du processus constituent la part de notre contribution dans ce travail.

Lors de la collecte des données, nous signalons les problèmes rencontrés tels que le manque de données et des données anormales pendant le fonctionnement de la station en raison de pannes de capteurs et de panne de réseau de distribution auprès du barrage de CHAFIA (endommagement de la conduite principale pendant une longue durée). Les tableaux d'analyse de qualité de l'eau de la station et du barrage des années écoulées (2013-2020) montrent ce manque de paramètres d'analyse pendant les mois d'arrêt. A titre d'exemple dans la figure III.1, les données manquantes sont représentées par une étoile sur le tableau d'évolution des paramètres de la qualité d'eau de la station de CHAIBA de l'année 2015.

La surveillance permanente de la qualité de l'eau implique la mesure en continu d'un certain nombre de paramètres. Ceux-ci peuvent se classer en deux grandes familles : les paramètres usuels et spécifiques. Un exemple de données collectées dans les différents processus cités précédemment est montré dans la figure III.1.

Dans le système de surveillance de la qualité de l'eau, les données sont souvent manquantes. Les paramètres de qualité de l'eau surveillés sont surveillés qu'une seule fois à un certain moment. Parce que le temps est irréversible, les données manquantes ne peuvent plus être acquises, elles ne peuvent donc être remplacées. A cet effet, nous avons éliminés les données manquantes dans la figure III.1. De ce fait, nous avons redimensionné ce tableau en réduisons le nombre des paramètres, initialement égale à 18 à 12 paramètres (au lieu de 18 se trouvant sur le tableau d'évolution de la qualité d'eau de la station ou du barrage et au lieu de 15 paramètres sur la grille normative (tableau III.1).

indis/param	UNITE	NORME	janvet	fevrier	mars	avril	mai	juin	juillet
PH	*	6,5-8,5	7,5	7,38	*	7,16	7,18	7,13	*
Température	*	*	11,2	11,7	*	12,3	12,1	11,4	*
conductivite	µs/cm	3125	459	391	*	409	361	418	*
TDS	mg/l	*	229,5	195,5	*	204,5	180,5	209	*
Salinite	g/l	*	0,2	0,2	*	0,2	0,2	0,2	*
Turbidite	NTU	5	1,5	2,6	*	4,8	6,8	3	*
Chlore résidu	g/l	*	0,8	0,8	*	0,8	0,1	0,8	*
Couleur	mg/Pt	25	*	*	*	*	*	*	*
residus secs	mg/l	2000	*	*	*	*	*	*	*
TA	mg/lCaCO ₃	*	0	0	*	0	0	0	*
TAC	mg/lCaCO ₃	*	93,84	84	*	79,04	86,7	92,13	*
Durete total	mg/l	500	186	164	*	154	152	160	*
Ammonium	mg/l	0,5	*	<0,06	*	<0,06	<0,06	<0,06	*
Nitrites	mg/l	0,1	*	<0,02	*	<0,02	<0,02	<0,02	*
Nitrates	mg/l	50	*	<1,1	*	<1,1	<1,1	<1,1	*
Orthosphosph	mg/l	0,5	*	<0,07	*	<0,07	<0,07	<0,07	*
indice perma	mg/l	3	<0,5	1,66	*	1,6	1,63	<0,5	*
Calcium	mg/l	200	53,7	46,49	*	44,88	44,08	48,09	*
Magnesium	mg/l	150	11,17	11,66	*	10,2	10,2	9,72	*
Fer total	mg/l	0,3	<0,01	<0,01	*	<0,01	<0,01	0,06	*
Aluminium	mg/l	0,2	0,18	0,2	*	0,32	0,21	0,31	*
CO ₂ total	mg/l	*	107,63	100,6	*	97,54	112,75	128,57	*
CO ₂ libre	mg/l	*	6,94	10,42	*	12,73	19,68	16,21	*
Bicarbonates	mg/l	*	114,43	102,48	*	96,38	105,77	112,36	*
Carbonates	mg/l	*	0	0	*	0	0	0	*
Chlorure	mg/l	500	33,68	35,07	*	31,65	30,98	33,68	*
Sulfates	mg/l	400	*	*	*	*	*	*	*

Figure III.1 Manque de données au mois de mars et juillet 2015 sur le tableau d'évolution des paramètres de la qualité d'eau de la station de CHAIBA année 2015

Pour estimer la qualité de l'eau, on se base sur la grille normative de l'Algérie (tableau III.1).

Dans cette grille, on distingue quatre (04) types de qualité d'eau :

- Bonne
- Moyenne
- Mauvaise
- Très mauvaise

Paramètres Physico-chimiques	Bonne	Moyenne	Mauvaise	Très mauvaise
PH	6.5-8.5	6.5-8.5	>6, <9	>5, <9
T°C	25	25-30	30-35	>35
Minéralisation mg/l	300-1000	1000-1200	1200-1600	>1600
Ca ²⁺ mg/l	40-100	100-200	200-300	>300
Mg ²⁺ mg/l	30	30-100	100-150	>150
Na ⁺ mg/l	10-100	100-200	200-500	>500
Chlorures mg/l	10-150	150-300	300-500	>500
Sulfates mg/l	50-200	200-300	300-400	>400
Organiques :				
O ₂ dissous %	>100	100-50	50-30	>30
DBO ₅ mg/l	5	5-10	10-15	>15
DCO mg/l	20	20-40	40-50	>50
Matières organiques	5	5-10	10-15	>15

Tableau III.1 Grille normative à douze paramètres pour l'estimation la qualité de l'eau en Algérie.

Une autre contrainte enregistrée est que les données collectées nous ont été fournies sous un format de papier (27 tableaux de l'évolution de la qualité d'eau de la station de CHAIBA et du barrage de CAHAFIA). Nous avons converti ces tableaux en des tableaux de format Excel suivant la forme ci-dessous :

- 0 : manque de données,
- 1 : lorsque le paramètre figure dans la colonne « bonne » de la grille normative,
- 2 : lorsque le paramètre figure dans la colonne « moyenne » de la grille normative,
- 3 : lorsque le paramètre figure dans la colonne « mauvaise » de la grille normative,
- 4 : lorsque le paramètre figure dans la colonne « très mauvaise » de la grille normative.

m/p	janvier	février	mars	avril	mai	juin	juillet	aout	sept	oct	nov	dec
ph	2	2	2	2	2	1	2	1	2	2	0	0
Température	1	1	1	1	1	1	1	1	1	2	0	0
Oxygène dissous	0	0	0	0	0	0	0	0	0	0	0	0
Calcium	1	1	1	0	0	0	1	1	1	1	0	0
Magnésium	1	1	0	0	0	0	0	0	0	0	0	0
Ammonium	2	2	3	2	2	2	2	2	2	2	0	0
Nitrites	2	2	2	2	2	2	2	2	2	2	0	0
Nitrates	0	0	0	0	0	0	0	1	1	1	0	0
Sodium	3	2	2	2	2	2	2	2	3	3	0	0
Fer total	1	1	1	2	1	1	4	1	1	1	0	0
Chlorure	1	1	1	1	1	1	1	1	1	1	0	0
Sulfate	1	1	0	0	0	0	0	0	0	0	0	0

TableauIII.2 Exemple de classement manuel des paramètres d'analyse de la qualité d'eau de la station de CHAIBA année 2015.

Dans la présente étude, on a utilisé la méthode de duplication de données pour améliorer l'efficacité de l'application CNN. Tant que la méthode de duplication est appropriée, les résultats obtenus seront satisfaisants. Nous avons multiplié par 3 les tableaux existants, dont le nombre s'élève à 27, nous obtenons donc $27 \times 3 = 81$ tableaux. Ceux-ci seront transformés en 81 images avec des niveaux de gris dans les images décrit ci-dessous :

- 1= le niveau de gris est égal à 0, il représente la couleur noire
- 2= le niveau de gris est égal à 75, il représente la couleur grise foncée
- 3= le niveau de gris est égal à 150, il représente la couleur grise claire
- 4= le niveau de gris est égal à 200, il représente la couleur grise très claire
- 0= le niveau de gris est égal à 255, il représente la couleur fond blanc

La figure III.2 ci-dessous représente un échantillon d'une forme d'image avec les niveaux de gris après transformation du tableau de données des paramètres des analyses.



Figure III.2 Echantillon forme d'image avec les niveaux de gris

III.2 Classification des images

Le système de classification des images qu'on a conçu fait le classement selon le choix cité ci-dessous :

- On sélectionne la classe très mauvaise si le nombre 0 est supérieur à 4.
- On sélectionne la classe mauvaise si le nombre 3 est supérieur à 13.
- On sélectionne la classe moyenne si le nombre 3 est inférieur à 13 et le nombre 2 est supérieur à 23.
- Sinon la classe est sélectionnée bonne
- Si le nombre $3 < 13$ et le nombre $2 > = 23$ la classe est moyenne sinon la classe est bonne

La figure III.3 représente la classification de la qualité de l'eau suivant le système de classification décrit précédemment.

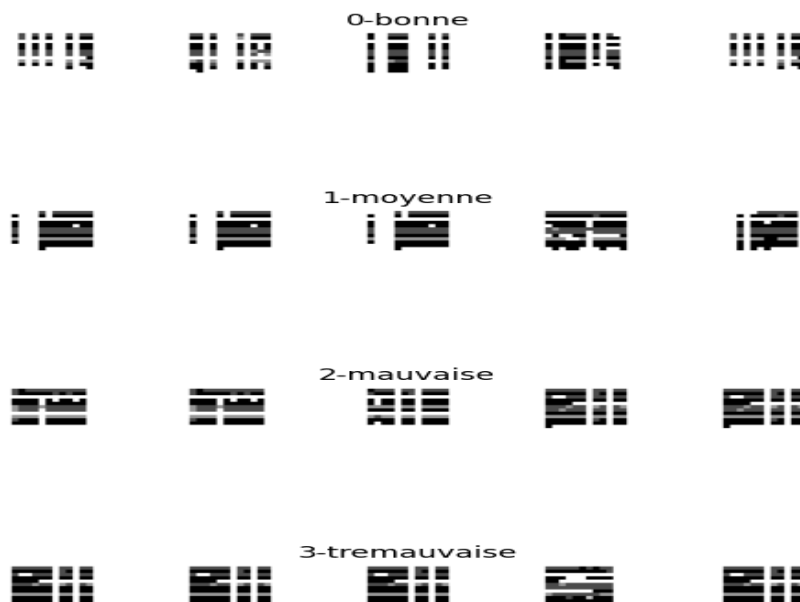


Figure III.3 Classification de la qualité de l'eau



Figure III.4 Exemple - Eau de bonne qualité

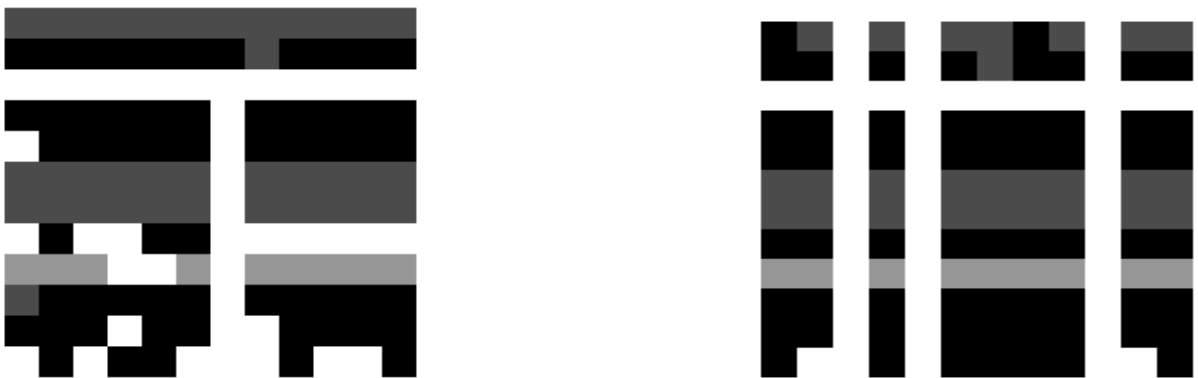


Figure III.5 Exemple - Eau de moyenne qualité

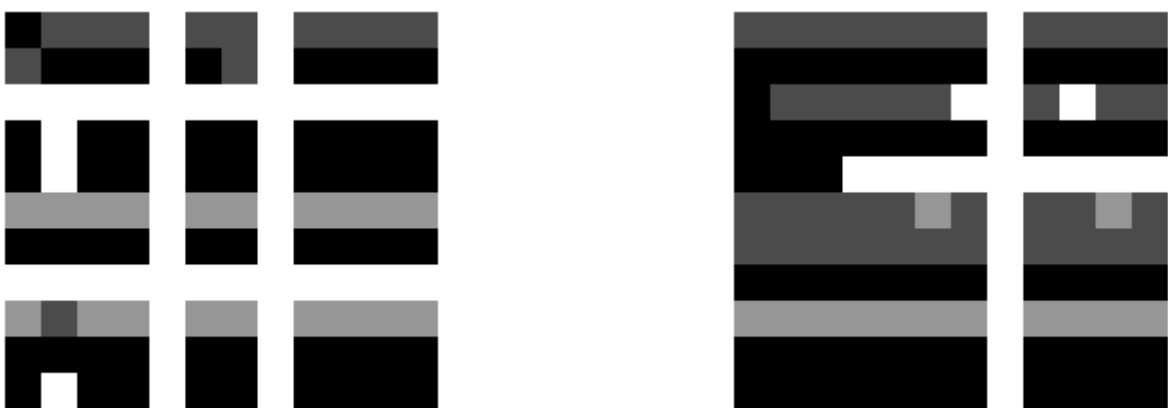


Figure III.6 Exemple - Eau de mauvaise qualité

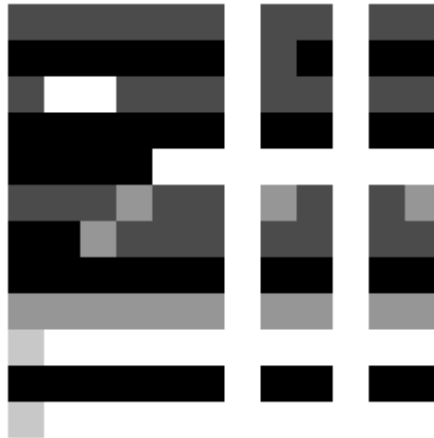


Figure III.7 Exemple – Eau de très mauvaise qualité

Figure III .3 Échantillons de 04 classes de la qualité d'eau typiques de la base de données

III.3 Architecture du réseau

Dans ce paragraphe, nous présentons l'architecture du réseau CNN que nous avons conçu afin de représenter les classes de la qualité de l'eau suivant le système de classification décrit précédemment.

Dans le processus de construction du modèle CNN, la sélection des hyper-paramètres adéquats du CNN peut améliorer efficacement la précision du diagnostic de défaut ainsi que les vitesses d'apprentissage et de test du modèle CNN-Softmax conçu. Généralement, les principaux hyper-paramètres incluent le nombre de couche, la taille des noyaux de convolution, les noyaux de sous-échantillonnage, les optimiseurs, les fonctions d'activation et le coefficient d'apprentissage, etc. Ces configurations ont un impact direct sur les performances du modèle CNN adopté. Une architecture CNN est développée après plusieurs essais répétés et qu'on va la considérer comme un modèle de référence avec lequel on poursuivra notre évaluation.

Le modèle de référence CNN de référence, qui donne un taux de précision le plus optimal, se compose d'une couche de pooling (SL), deux couches de convolution (CL), trois couches entièrement connectées (FC) et est doté aussi d'un classifieur Softmax dans la sortie. Dans l'étape de sélection de la taille des noyaux, le CL ultérieur doit avoir une taille de noyau plus petite que celle du CL antérieur, car après chaque opération CL et SL, la taille des données de sortie sera diminuée, ce qui réduit significativement le coût des calculs et améliore la vitesse d'apprentissage.

Le modèle que nous présentons dans la figure III.8 comporte l'architecture suivante :

- Une (01) couche d'average,
- Deux (02) couches de convolution,
- Une (01) couche de maxpooling,
- Trois (03) couches de fullyconnected

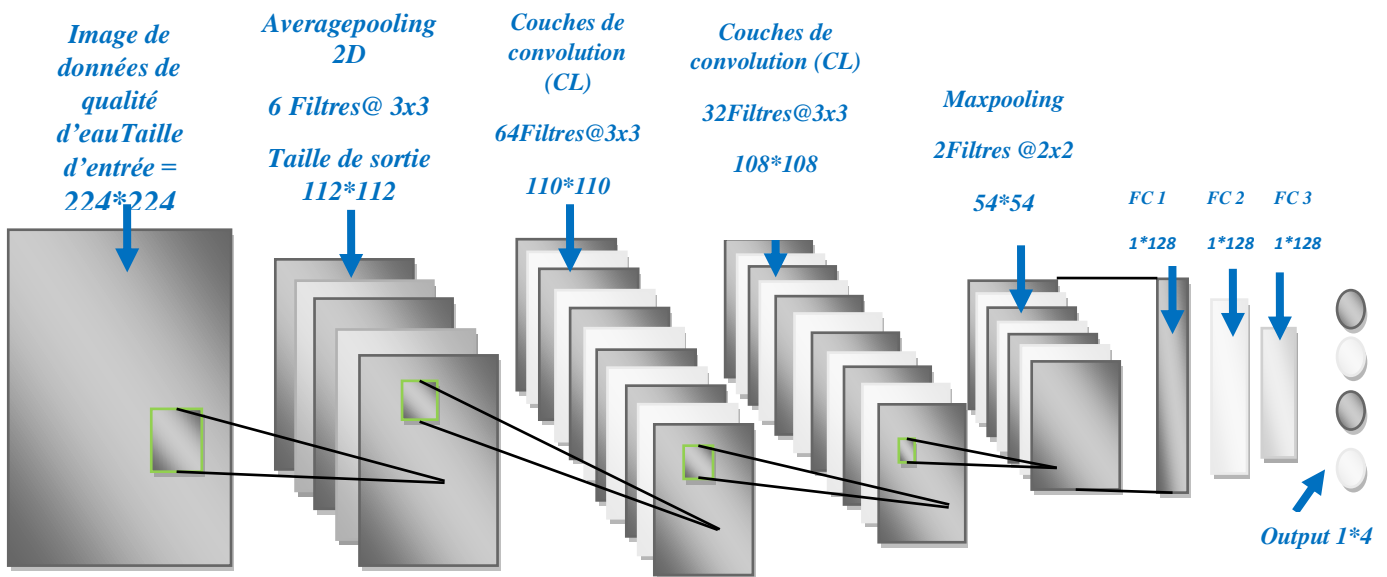


Figure III.8 Architecture du modèle

III.3.1 Hyper-paramètres du modèle de référence

L'image en entrée est de taille 224x224, l'image passe d'abord par la première couche d'averagepooling 2D composée de 6 filtres de tailles 3x3, la taille de sortie sera de 112x112 suivi d'une couche convolution composée de 64 filtres de tailles 3x3, cette couche de convolution est suivie fonction d'activation ReLU. Cette fonction force les neurones à retourner des valeurs positives, après cette convolution 64 features maps de taille 110x110 seront créés, une deuxième couche de convolution est appliquée, elle composée de 32 filtres de taille 3x3 et après une opération d'activation ReLU, la sortie sera de taille 108x108.

On applique ensuite Maxpooling de 2 filtres de taille 2x2 pour réduire la taille de l'image ainsi que la quantité de paramètres et de calcul. À la sortie de cette couche, nous aurons la taille 54x54.

Après ces deux couches de convolution, nous utilisons un réseau de neurones composé de trois couches fullyconnected. Ces couches ont respectivement 128, 64, 32 neurones où la fonction d'activation utilisée est ReLU, la quatrième couche est un softmax qui permet de calculer la distribution de probabilité des 4 classes.

Le tableau III.3 présente une description détaillée concernant les hyper-paramètres du modèle de référence adopté.

Nombre de couche	Type	Spécification de la couche	Taille de sortie
Couche 01	Entrée	Taille d'entrée = 224*224	224*224
Couche 02	Pooling2D moyenne	Nombre de filtre =6 Taille de kernel=3x3	112*112
Couche 03	convolution	Nombre de filtre =64 Taille de kernel=3x3 Activation='relu' Stride=1	110*110
Couche 04	convolution	Nombre de filtre =32 Taille de kernel=3x3 Activation='relu' Stride=1	108*108
Couche 05	MaxPool2D	Nombre de filtre =2 Taille de kernel=2x2	54*54
Couche 06	Dropout	128 couches cachées Dropout=0.5	1*128
Couche 07	Flatten		1*128
Couche 08	Dense	activation='relu'	1*128
Couche 09	Dense	activation='softmax'	1*4

Tableaux III 3 : Hyper-paramètres du modèle de référence adopté.

Après le choix des hyper-paramètres, 51 images ont été utilisées pour l'apprentissage et 30 images pour le test.

L'apprentissage a donné une répartition des données représenté sur la figure III.9

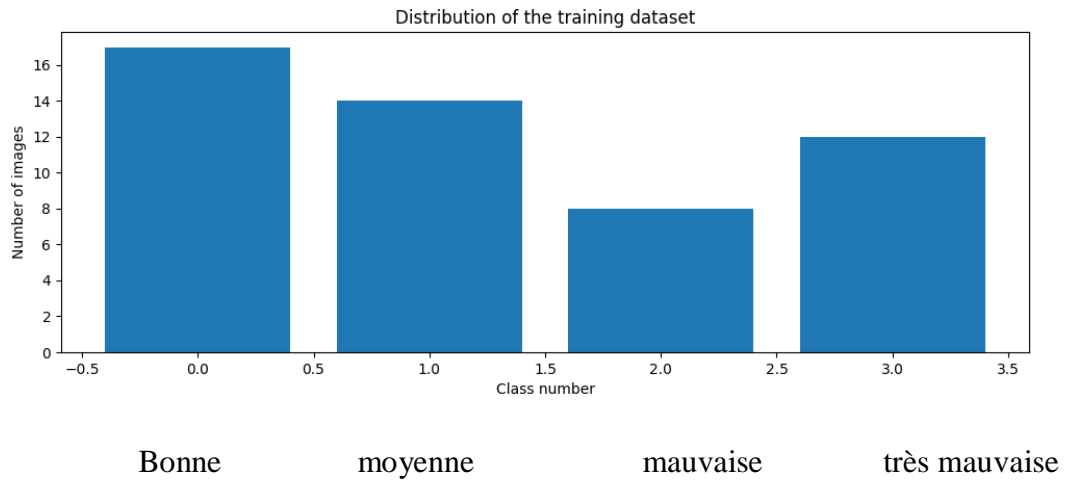


Figure III.9 Répartition de données d'apprentissage (51 images).

III.4 Résultats obtenus et discussion

Les résultats en termes de précision et d'erreur ainsi que la matrice de confusion pour le modèle de référence adopté sont illustrés dans les figures.

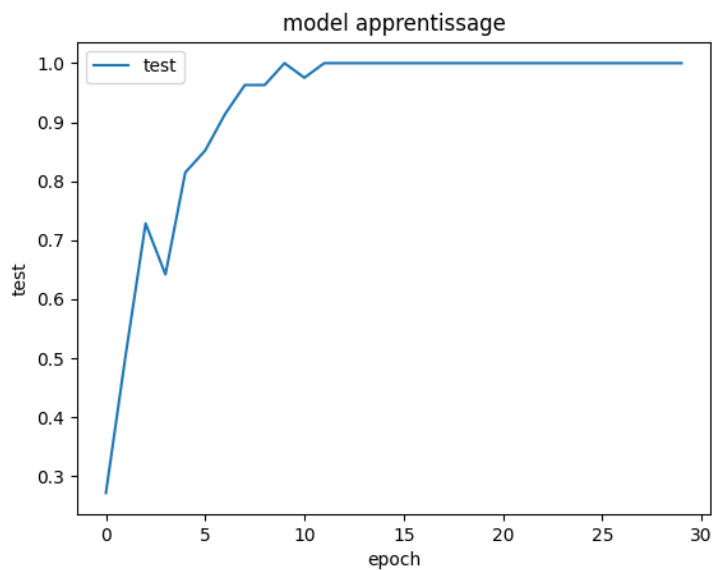


Figure III.10 Courbe de précision du modèle d'apprentissage

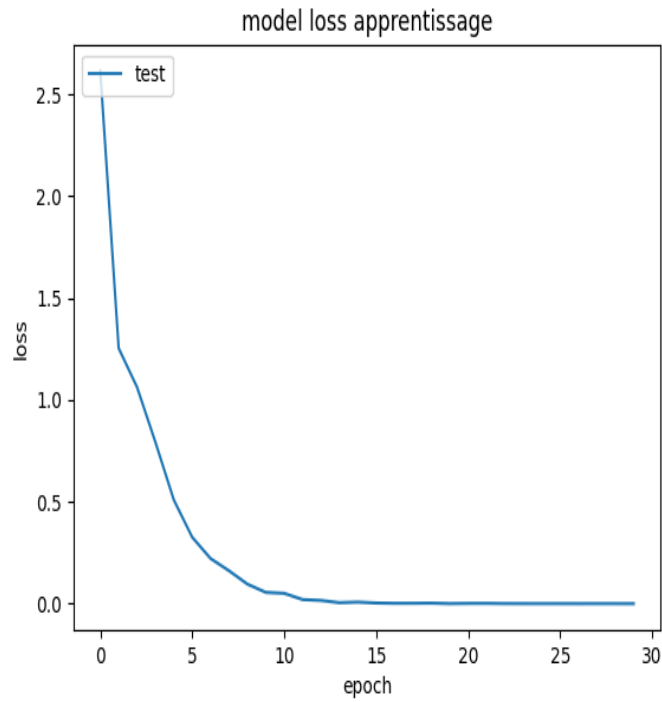


Figure III.11 Courbe d'erreur du modèle d'apprentissage

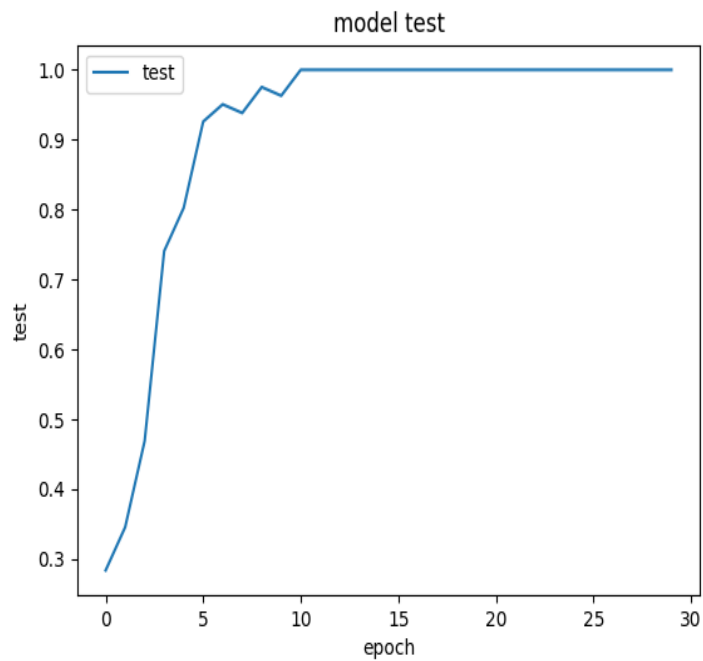


Figure III.12 Courbe de précision du modèle test

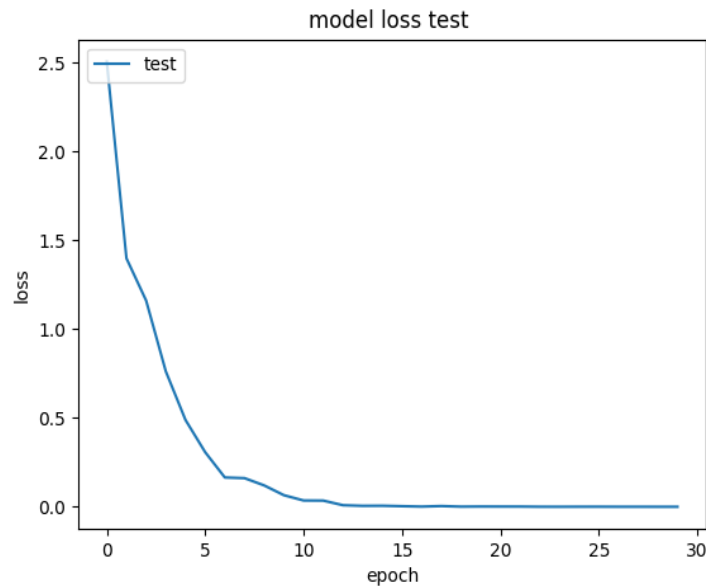


Figure III.13 Courbe d'erreur du modèle test

En analysant des résultats obtenus, On constate les remarques suivantes :

- La précision de l'apprentissage et de test augmente avec le nombre d'époque, ceci reflète qu'à chaque époque le modèle apprend plus d'informations. Si la précision est diminuée alors on aura besoin de plus d'information pour faire apprendre notre modèle et par conséquent on doit augmenter le nombre d'époque et vice versa.
- De même, l'erreur d'apprentissage et de la validation diminue avec le nombre d'époque.

- Matrice de confusion

Pour le test des 30 images, nous obtenons les résultats suivants :

8 : états « Bonne qualité ».

8 : états « Moyenne qualité ».

7 : états « Mauvaise qualité ».

7 : états « Très mauvaise ».

Les classe	bonne	moyenne	Mauvaise	Très mauvaise
Bonne	8	0	0	0
Moyenne	0	8	0	0
Mauvaise	0	0	7	1
Très mauvaise	0	0	0	6

Tableau III.4 Répartition de 30 images de test

Nous allons maintenant calculer le taux de précision suivant la formule cette formule ci-dessous

- Précision = la somme de bon classification des 4 classes /(la somme de mauvaise classification plus la somme de bon classification des 4 classes).

Le taux de précision sera égal à :

$$\text{Taux de précision} = 29 / (1 + 29) = 0.966$$

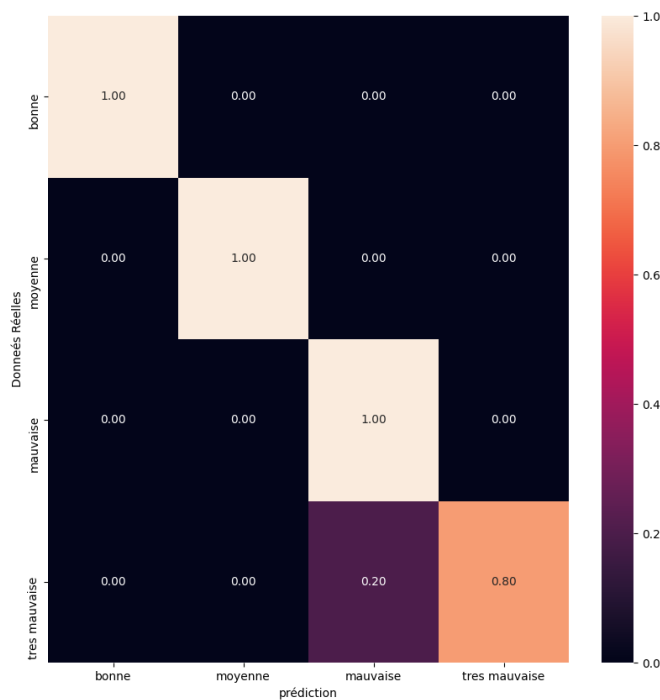


Figure III.14 Matrice de confusion

Sur la figure III.14, l'axe des abscisses de la matrice de confusion représente l'étiquette de catégorie prédite et l'axe des ordonnées représente l'étiquette de catégorie réelle. Le nombre d'échantillons d'ensembles de données de test est de 30, et il existe 4 classes, comme indiqué précédemment. Les éléments sur la diagonale principale de la matrice de confusion représentent le numéro d'échantillon de la classification correcte de chaque état. Dans la figure III.14, chaque colonne quantifie les informations de taux de précision. Chaque ligne quantifie les informations de taux de rappel. On peut évidemment voir sur la figure III.14 que le résultat de diagnostic le plus bas se produit dans la condition 2.

Un seul des 30 échantillons de test sur l'ensemble de test est mal classé par la prédiction. L'étiquette réelle de l'échantillon mal classé est l'état « très mauvaise ». A partir de la figure III.14, nous constatons que la précision des autres états est de 100 %. Le modèle a bien classé les images des autres états « Bonne, Moyenne et Mauvais » et il a mal classé les images de l'état « très mauvaise ». Le taux de précision totale est égal à 96,66%.

III.5 Conclusion

Nous avons présenté dans ce chapitre une approche de classification de la qualité de l'eau, basée sur les réseaux de neurones convolutifs. Le modèle CNN de référence adopté a donné de bons résultats obtenus en termes de précision et d'erreur que ce soit dans la phase d'apprentissage ou dans phase de test.

Conclusion générale

Le travail présenté dans ce mémoire a porté sur l'utilisation des réseaux de neurones convolutifs appliqués à la surveillance de la qualité des eaux potables. Ce travail a été structurée autour de trois chapitres essentiels.

Le premier a été consacré à une introduction au domaine de l'eau, il a permis de présenter les systèmes de production d'eau potables ainsi que les différentes méthodes de surveillance des systèmes industriels.

Le second chapitre a été consacré à la présentation des méthodes de l'intelligence artificielle tel que l'apprentissage approfondi et plus particulièrement les réseaux de neurones convolutifs.

Enfin le troisième et le dernier chapitre, a fait l'objet d'une étude en simulation concernant la mise en œuvre d'un modèle d'apprentissage, à base de réseaux de neurones convolutifs, appliqué dans le domaine de la surveillance de la qualité des eaux potables de la station de CHAIBA de la wilaya d'Annaba. Nous avons rencontré quelques problèmes dans la phase d'implémentation à cause du manque de données, mais ce problème a été surmonter grâce à l'application de la méthode de duplication des données.

A partir des bons résultats obtenus, nous espérons que l'application qu'on a développée peut être alternative prometteuse dans l'avenir pour une surveillance intelligente et automatique de la qualité des eaux potables.

Bibliographie

- [2]. **SLIM. A** « contribution de l'analyse multivariée a l'étude de régression et la classification supervisée des données environnementales » mémoire de master, université de m'sila 2012
- [3]. **LADJAL .M** « Contribution au développement de systèmes de surveillance innovants dédiés au contrôle de la qualité des eaux potables » thèse de doctorat, université de M'sila 2013.
- [4]. **CARDOT, 1999 CLAUDE CARDOT**, agrégé de génie biologique-biochimie « panorama complet et synthétique des procédés physico-chimiques et biologiques de traitement des eaux potables et eaux usées ».
- [5]. **DAJOZ R.** (2000). Précis d'écologie : « Cours Et Exercices Résolus. 7 ième édition. Dunod, Paris. 613p ».
- [6]. **MOUASSA S.** (2006). « Impact Du Périmètre D'irrigation Sur La Qualité Des Eaux Souterraines De La Nappe Alluviale De Guelma Et Sur Les Sols ». Mémoire de Magister, Université Badji Mokhtar. Annaba. 120p.
- [7]. (**Degrémont, 2005**). « Tome 1 présente les bases des différents procédés et technologies de traitement de l'eau et son contexte (définition de l'eau, du traitement...) et les différentes analyses concernant la qualité de l'eau ».
- [8]. **A. Charafeddine HAMMOUDI, Mérouane BOUCHEMAL** « Analyse de la qualité des eaux de la station de traitement de Hammam Debegh » Université Larbi Ben M'hidi– Oum El Bouaghi –Faculté des Sciences et Sciences appliquées –Département-d'Hydraulique, Master II en Hydraulique.
- [9]. (**LIND2,1994 ; LIND3,1994 ;CIDF-Ldes Eaux,2000**) L'influence De HRH de Leon · 2006 · Cité 51 fois — « *des* paramètres influençant le bon fonctionnement du procédé de coagulation ».
- [11]. **C.R.A.A.G.** (Centre de Recherches Astronomiques et Géophysiques Algérien) (2004). « Etude Géophysique Par Gravimétrie De La Région D'Oued Zénati, Guelma ». Rapport Interne. 42 p.
- [12]. **H. Hernandez**, « Développement d'un capteur logiciel pour la prédiction de la dose de coagulant dans une station de traitement d'eau potable en vue de son diagnostic », Rapport LAAS N° 05175, 6ème Congrès des Doctorats de l'Ecole Doctorat Systèmes (EDSYS), 17-20 Mai 2005, Toulouse, France.
- [13]. **HECTOR RICARDO HERNANDEZ DE LEON**, « Supervision et diagnostic des procédés de production d'eau potable », Thèse de Doctorat, Institut National des Sciences Appliquées de Toulouse, CNRS, N° 06584, France, 2006.

- [14]. **Daniel Racoceanu**, « Contribution à la surveillance des systèmes de production en utilisant les techniques de l'intelligence artificielle », Thèse d'HDR, Université de Franche-Comté de Besançon, France, N° d'ordre : 140, 2006.
- [15]. **Nicolas Pallaut**, « Méthodologie de surveillance dynamique à l'aide des réseaux neuro-flous temporels », Thèse de Doctorat, Université de Franche-Comté de Besançon, France, N° d'ordre : 1135, 2006.
- [16]. **Nelly Olivier-Maget**, « Surveillance des systèmes dynamiques hybrides », Thèse de Doctorat, Université de Toulouse, France, N° d'ordre : 913, 2007.
- [17]. **R. Zemouri, D. Racoceanu, N. Zerhouni**, « Réseaux de neurones récurrents à fonctions de base radiales RRFR : Application à la surveillance dynamique », Revue Systèmes /JESA, 2003, Vol. 37, N°1, pp. 49-81.
- [19]. **FRANCE INTE**. « La quête du Graal des chercheurs en intelligence artificielle ». France Inter [en ligne]. [Consulté le 31 août 2018]. Disponible à l'adresse : <https://www.franceinter.fr/emissions/c-est-deja-demain/c-est-deja-demain-24-janvier-2018>
- [20]. **FUKUSHIMA Kunihiko, MAYAKE Sei**. «Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. 1982 ».
- [21]. **HUBEL David hunter, WIESEL Torsten nils**. « Ferrier lecture-functional architecture of macaquemonkey visual cortex. 1977 Hongwei. Semantic clustering and convolutional neural networks for short text categorization. 2015 ».
- [22]. **LECUN Yann, BOTTOU Leon, BENGIO Yoshua, HAFFNER Patrick**. Gradient-based « learning applied to document recognition. 1998 ».
- [23]. **CIRESAN Dan C, MEIER Ueli, MASCI Jonathan, MARIA GAMBARDELLA Luca, SCHIDHUBER Jurden**. «Flexible, high performance convolutional neural networks for image classification. 2011 ».
- [24]. **O'SHEA Keiron, NASH Ryan**. « An introduction to convolutional neural networks. 2015 ».
- [25]. **WANG Peng, XU Jiaming, XU Bo, LIU Chenglin, ZHANG Heng, WANG Fangyuan HAO Hongwei**. «Semantic clustering and convolutional neural networks for short text categorization. 2015 ».

Référence cite web

[1].<https://www.suezwaterhandbook.fr/eau-et-generalites/l-eau-ses-proprietes.degrémont> 2005

[10]. **BOEGLIN Jean-claude**. « Propriétés des eaux naturelles. 10 janv. 2001 » — Article de référence | Réf : *G1110* v1 ... <https://www.techniques-ingenieur.fr/base-documentaire/environnement-securite-th5/eau-proprietes->

[18]. **LECUN, Yann**. « Qu'est-ce que l'intelligence artificielle ? » [en ligne]. Disponible à l'adresse :https://www.college-de-france.fr/media/yann_lecun/UPL4485925235409209505_Intelligence_ArtificielleY_LeCun.pdf

[26].<http://adventuresinmachinelearning.com/convolutional-neural-networks-tutorial-tensorflow/>Consultez le 2/aout /2018.

RESUME

L'eau est un élément indispensable à la Vie. Elle est indispensable à l'Homme. Sans eau, aucun organisme qu'il soit végétal ou animal, ne peut vivre. A cet effet, il est donc très important de la contrôler et de la surveiller pour préserver une bonne qualité. Des efforts ont été engagés ces dernières années par l'état algérien en traitant prioritairement l'approvisionnement en quantité et en qualité de l'eau potable au peuple algérien. Pour cela, il a fourni tous les efforts nécessaires à la mise en œuvre des différents moyens pour développer des méthodes de contrôle et de surveillance automatique de l'eau. Ce travail s'inscrit dans ce titre dans ce cadre, en proposant une nouvelle approche pour le contrôle et la surveillance de la qualité des eaux potables. Ce mémoire aborde le problème de la surveillance des eaux potables en fonction des caractéristiques physico-chimiques de l'eau brute à l'aide d'un modèle basé sur les techniques de l'apprentissage approfondi basé sur les réseaux de neurones convolutifs(CNN).

ملخص

الماء عنصر أساسي في الحياة. إنه ضروري للإنسان. بدون الماء ، لا يمكن لأي كائن حي ، سواء كان نباتاً أو حيوانياً ، أن يعيش لهذا الغرض ، من المهم للغاية التحكم فيه ومراقبته للحفاظ على توفير جودة جيدة ، وقد بذلت الدولة الجزائرية جهوداً في السنوات الأخيرة من خلال التعامل معه كأولوية لتوفير الكمية والنوعية من المياه الصالحة للشرب لفائدة الشعب الجزائري من أجل ذلك ، فقد بذلت كل الجهود اللازمة الهادفة إلى توفير مختلف الوسائل لتطوير طرق التحكم الآلي والمراقبة الجيدة للمياه. وهذا العمل جزء من عنوان هذا المطلب، من خلال اقتراح نهج جديد لضبط ومراقبة جودة المياه الصالحة للشرب. تتناول هذه الرسالة مشكلة مراقبة مياه الشرب للخصائص الفيزيائية والكيميائية للمياه الخام باستخدام نموذج يعتمد على تقنيات التعلم المتعمقة مثل اقتراحنا لتقنية CNN.