

REPUBLICQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj

Faculté des Sciences et de la technologie

Département d'Electronique

Mémoire

Présenté pour obtenir

LE DIPLOME DE MASTER

FILIERE : **Electronique**

Spécialité : industries électroniques

Par

➤ **BENSEFIA ADEM**

Intitulé

Identification des appareils électriques par combinaison des classifieurs

Évalué le : 12/09/2021

Par la commission d'évaluation composée de :*

<i>Nom & Prénom</i>	<i>Grade</i>	<i>Qualité</i>	<i>Etablissement</i>
<i>M.Yahia Belhaddad</i>	<i>MCB</i>	<i>Président</i>	<i>Univ-BBA</i>
<i>Mme. Laouamri Asma</i>	<i>MAA</i>	<i>Encadreur</i>	<i>Univ-BBA</i>
<i>M. Salah Eddine Mezaache</i>	<i>MCB</i>	<i>Examineur</i>	<i>Univ-BBA</i>

Année Universitaire 2020/2021

Résumé

Les systèmes d'identification des appareils électriques devront fournir les informations sur l'identité de l'appareil en état de fonctionnement, l'intervalle de consommation et la quantité de consommation. La conception de tels systèmes est basée sur l'implémentation d'une méthode de classification. Les méthodes de classification peuvent se distinguer aux approches intrusive et non intrusive, la plupart d'entre eux reposent le choix d'une structure de classifieur (discriminateur linéaire, KNN, SVM...) et le critère d'évaluation des performances. Notre travail s'inscrit dans le cadre de conception des systèmes multi-classifieurs (MCS), qui s'attachent à trouver comment plusieurs classifieurs peuvent être combinés pour obtenir de meilleurs systèmes de classification.

Deux systèmes d'identification des appareils électriques sont implémentés, le premier système est basé sur l'analyse harmonique et le deuxième est basé sur les paramètres statistiques du courant électrique. Un système multi-classifieurs qui combine les sorties des trois classificateurs (KNN, ADI et ANN) est conçu et le taux d'identification obtenu est satisfaisant.

Abstract :

The electrical system identification should provide information on the identity of the device in working order, the consumption interval and the amount of consumption. The design of such systems is based on the implementation of a classification method. Classification methods can be distinguished from intrusive and non-intrusive approaches, most of them are based on the choice of a classifier structure (linear discriminator, KNN, SVM ...) and the performance evaluation criterion. Our work falls within the framework of Multi-Classifier Systems (MCS) design, which is concerned with finding how multiple classifiers can be combined to achieve better classification systems.

Two systems for identifying electrical devices are implemented, the first system is based on harmonic analysis and the second is based on statistical parameters of electric current. A multi-classifier system which combines the outputs of the three classifiers (KNN, ADI and ANN) is designed and the identification rate obtained is satisfactory.

ملخص

يجب أن توفر أنظمة تحديد الأجهزة الكهربائية معلومات عن هوية الجهاز في ترتيب العمل ، وفترة الاستهلاك ومقدار الاستهلاك. يعتمد تصميم هذه الأنظمة على تنفيذ طريقة التصنيف. يمكن تمييز طرق التصنيف عن الأساليب المتطفلة وغير التدخلية ، حيث يعتمد معظمها على اختيار هيكل المصنف (المُمَيِّز الخطي ، KNN ، SVM ...) ومعيار تقييم الأداء. يندرج عملنا في إطار تصميم أنظمة التصنيف المتعددة (MCS) ، والذي يهتم بإيجاد كيفية دمج المصنفات المتعددة لتحقيق أنظمة تصنيف أفضل. تم تنفيذ نظامين لتحديد الأجهزة الكهربائية ، النظام الأول يعتمد على التحليل التوافقي والثاني يعتمد على المعايير الإحصائية للتيار الكهربائي. تم تصميم نظام متعدد المصنفات يجمع بين مخرجات المصنفات الثلاثة (KNN و ADI و ANN) ومعدل التقييم الذي تم الحصول عليه مرضٍ.

REMERCIEMENTS

Je remercie Dieu le tout puissant de m'avoir donné la force, la volonté et le courage à mettre en lumière ce modeste travail.

*Mes profonds remerciements à mon encadreur **Dr. A .LAOUAMRI** qui m'a aidé avec ses précieux conseils et ses orientations tout au long de mon travail.*

*J'exprime un profond remerciement à Monsieur **Hacine Gharbi Abdenour** et Mr. **Ghazali Fateh** Pour la disponibilité, la compréhension, les conseils et les encouragements avec lesquelles ils ont bien voulu diriger ce travail.*

Je remercie également mes professeurs pour la qualité de enseignement qui 'ils m'ont prodigué au cours de ces années passées à l'université de Bordj Bou Arreridj

Enfin, j'adresse mes sincères remerciements aux membres du jury. Je les remercie pour leur patience ainsi que pour leur lecture attentive de ce travail.

Dédicaces

Je dédie ce projet :

A mes parents, pour l'amour qu'ils m'ont toujours donné, leurs encouragements et toute l'aide qu'ils m'ont apportée durant mes études. Aucun mot, aucune dédicace ne pourrait exprimer mon respect, ma considération, et mon amour pour les sacrifices qu'ils ont consentis pour mon instruction et mon bien-être.

A mon encadreur M. Asma Laouamri pour avoir m'encadré

A tous les membres de ma famille, petits et grands, qui m'ont toujours aidé et encouragé, qui étaient toujours à mes côtés.

A Mes Frères et ma Sœur Je leur dédie ce travail pour tous les sacrifices qu'ils n'ont cessé de m'apporter tout au long de mes années d'études.

À tous mes chers amis et mes collègues

A Les enfants de ma sœur : HIDAYA et MAZEN

Sommaire

Chapitre I : Théorie des systèmes d'identification des appareils électriques

I.1	INTRODUCTION	4
I.2	CONTEXTE GENERAL SUR L'IDENTIFICATION DES APPAREILS ELECTRIQUES.....	4
I.2.1	Approches intrusives:.....	4
I.2.2	Approches semi-intrusives (Semi-Intrusive Load Monitoring SILM) :.....	5
I.2.3	Approches non intrusive (Non-Intrusive Load Monitoring NILM) :.....	5
I.3	CONCEPTION D'UN SYSTEME D'IDENTIFICATION DES APPAREILS ELECTRIQUES....	6
I.3.1	L'ANALYSE HARMONIQUE	6
I.3.2	Paramètres statistiques	7
I.4	TRAVAUX PRECEDENTS SUR L'IDENTIFICATION DES APPAREILS ELECTRIQUES.....	7
I.5	Bases de données des courants électriques	8
I.5.1	La base PLAID :.....	8
I.5.2	La base COOLL :.....	10
I.6	Conclusion	11

Chapitre II : Méthodes de classification et Approches de Combinaison des classifieurs

II.1	Introduction.....	13
II.2	Classification	13
II.2.1	Classification supervisée :	13
II.2.2	Classification non supervisée :	13
II.3	Classifieurs.....	13
II.4	Méthodes de classification.....	14
II.4.1	Classification KNN :	14
A.	Définition	14
B.	Principe de fonctionnement	14
C.	Étapes de la classification KNN :.....	15
D.	Différentes distances utilisées.....	15
II.4.2	Classification par les Réseaux de neurones :.....	16
A.	Définition :	16
B.	Les réseaux de neurones artificiels	16

C. Classification neuronale :	17
D. Le perceptron multicouche.	18
II.4.3 Classification discriminante :	18
A. Définition :	18
B. Analyse Discriminante Linéaire (LDA) :	18
B.1 Apprentissage du modèle LDA	18
B.2 Test du modèle	19
II.4.4 Autres méthodes de classification :	19
II.5 Taux de classification	19
II.6 Combinaison des classifieurs	20
II.6.1 Pourquoi combiner les classifieurs ?	20
II.6.2 Types de sortie du classifieur :	20
II.6.3 Mesures de performances d'un classifieur	21
II.6.4 Stratégies de combinaison des classifieurs	21
II.6.4.1 Combinaison séquentielle :	21
II.6.4.2 Combinaison parallèle :	22
A. Approches de combinaison parallèle	23
A.1 Fonctions de combinaison sans apprentissage si la sortie est de type « classe ».....	23
□ Méthodes de vote	23
A.2 Fonctions de combinaison avec apprentissage si la sortie est de type « classe.....	24
• Méthode de Bayes.....	24
II.6.4.3 Combinaison hybride :	25
II.7 Conclusion	25

Chapitre III: Implémentation et Résultats de classification et de combinaison des classifieurs

III.1 Préliminaire	27
III.2 Implémentation des systèmes d'identification	27
III.3 Configuration matérielle et logicielle.....	30
III.4 Résultats d'implémentation et discussions.....	30
III.4.1 Expériences sur le système d'identification (SI) en utilisant l'analyse harmonique.....	30
III.4.1.1 Identification avec le classifieur KNN	30

III.4.1.2 Identification avec le classifieur ADI.....	31
III.4.1.3 Identification avec le classifieur ANN.....	31
III.4.2 Expérience sur le système d'identification en utilisant les paramètres statistiques.....	33
III.4.2.1 Identification avec le classifieur KNN.....	33
III.4.2.2 Identification avec le classifieur ADI.....	33
III.4.2.3 Identification avec le classifieur ANN.....	33
III.4.2.4 Comparaison entre les résultats des deux systèmes d'identification	34
III.4.3 Conclusion	34
III.5 Implémentation d'un système multi-classifieur (SMC).....	34
III.5.1 Architecture du système proposé.....	34
III.5.2 Fonction de combinaison ou 'méthode de Bayes'	35
III.5.3 Résultats de combinaison.....	36
III.5.4 Conclusion.....	38
Conclusion générale	39
Bibliographe	40

Liste des figures :

Figure I.1 : Principe de mesure de l'approche non intrusive.....	5
Figure I.2 : Structure de système NILM.....	5
Figure I.3 : Schéma synoptique de conception d'un système d'identification.....	6
Figure I.4 – Deux exemples de signaux de la base PLAID.....	9
Figure II.1 : Principe de fonctionnement de l'algorithme K-ppv.....	14
Figure II.2 : Fonctionnement de KNN.....	15
Figure II.3 : Structure d'un neurone formel.....	17
Figure II.4 : Un réseau perceptron multicouche.	18
Figure II.5 : Schéma de combinaison séquentielle de classifieurs.....	22
Figure II.6 : Combinaison parallèle de classifieurs.....	23
Figure II.7 : Taxonomie des méthodes de Combinaison.....	24
Figure II.8 : Combinaison hybride de classifieurs.....	25
Figure III-1 : Visualisation du réseau réalisé sur la boîte à outils	32
Figure III-2 : Architecture de la combinaison proposée.....	35

Liste des tableaux

Tableau I.1 : Quelques travaux dans le domaine de l'identification des appareils électriques.....	8
Tableau I.2 : Résumé des appareils trouvés dans la base PLAID.....	9
Tableau I.3 : Résumé de la base de données COOLL	10
Tableau III-1 : Sélection de la configuration optimal de classificateur KNN.....	30
Tableau III-2 : Taux de classification de KNN en fonction de nombre d'harmoniques.....	31
Tableau III-3 : Sélection de la configuration optimal de classificateur ADI, Avec un nombre d'harmonique = 25.....	31
Tableau III-4 : Taux de classification de l'ADI en fonction de nombre d'harmoniques.....	31
Tableau III-5 : Taux de classification de l'ANN en fonction de nombre d'harmoniques.....	32
Tableau III-6 : Taux de classification de KNN en fonction de nombre de vecteurs statistiques.....	33
Tableau III-7 : Taux de classification d'ADI en fonction de nombre de vecteurs statistiques.....	33

Tableau III-8 : Taux de classification d'ANN en fonction de nombre de vecteurs statistiques.....	33
Tableau III.9 : Comparaison des résultats de classification des deux systèmes.....	34
Tableau III.10 : Taux de classification des vecteurs après combinaison des classifieurs en fonction des paramètres statistiques.....	37
Tableau III.11 : Taux de classification des signaux après combinaison des classifieurs en fonction des paramètres statistiques.....	37

Introduction Générale

A cause de l'émergence demande de l'énergie et l'apparition des sources d'énergie renouvelables, plusieurs travaux de recherche ont été développés pour la gestion intelligente des réseaux électriques, d'une part et pour la supervision de la consommation de l'énergie, d'autre part. Une partie de ces recherches a été menée sur la surveillance de la charge et sur la conception des systèmes d'identification des appareils électriques, dans l'objectif de mieux contrôler et limiter la consommation d'énergie.

Les systèmes d'identification des appareils électriques devront fournir les informations sur l'identité de l'appareil en état de fonctionnement, l'intervalle de consommation et la quantité de consommation. La conception de tels systèmes est basée sur l'implémentation d'une méthode de classification, dont le rôle est de permettre d'associer des classes aux ensembles des mesures qui appartiennent à une base de données contenant les signaux des appareils électriques. Aujourd'hui, une grande variété de méthodes de classification est proposée dans la littérature. Ces méthodes peuvent se distinguer aux approches intrusive et non intrusive, la plupart d'entre eux reposent sur un schéma général en trois points essentiels :

1. Le choix d'une structure de classifieur (discriminateur linéaire, KNN, SVM...).
2. Le critère d'évaluation des performances du classifieur. Le critère le plus classique est le taux de bonne décision ou de classification.
3. La sélection du modèle ou sélection d'une forme de représentation.

Mais récemment, dans le domaine de la reconnaissance, le concept de combinaison de plusieurs résultats issus de différents classifieurs est proposé comme une nouvelle direction de développement des systèmes d'identification. Deux raisons majeures sont derrière la combinaison:

- **La Précision:** une décision plus fiable peut être obtenue en combinant l'avis (les sorties) de plusieurs classifieurs.
- **L'Efficacité:** un problème complexe peut être décomposé en plusieurs sous-problèmes qui sont plus faciles à comprendre et à résoudre.

Notre travail s'inscrit dans le cadre de conception des systèmes multi-classifieurs (MCS), qui s'attachent à trouver comment plusieurs classifieurs peuvent être combinés pour

obtenir de meilleurs systèmes de classification. Il y a deux voies principales de recherches sur les systèmes MCS. La première s'intéresse à la façon avec laquelle sont conçus les classifieurs individuels, et la seconde s'oriente vers la recherche des différents types de combinaisons de ces classifieurs.

Dans l'objectif d'implémenter une structure de combinaison pour les systèmes d'identification des appareils électrique et d'améliorer les performances de classification ; Notre contribution portera, tout d'abord, sur l'étude, l'implémentation et la validation de trois types de classifieurs individuellement. Par la suite, un système multi classifieur sera conçu en implémentant une approche de combinaison.

Ce travail est divisé en trois chapitres, qui sont :

- Le premier chapitre du manuscrit évoque l'état de l'art dans le domaine de l'identification des charges. Il a permis de nous éclairer sur les différentes méthodes utilisées dans ce domaine, d'identifier leurs avantages et inconvénients.
- Le deuxième chapitre sera consacré à l'étude de la problématique associée à la combinaison de classifieurs. Nous rappelons tout d'abord ce qu'on entend par classifieur dans le cadre de la combinaison. Nous passons en revue les principales stratégies de combinaison (approche séquentielle, parallèle et hybride) qui diffèrent principalement par l'agencement des classifieurs. Ce chapitre va nous tracer le chemin qui nous mène à choisir la méthode de combinaison pour l'identification des appareils électriques.
- Le troisième chapitre a pour but d'implémenter notre système et discuter les résultats obtenus en effectuant différentes expériences.

**Chapitre I : Théorie des systèmes
d'identification des appareils électriques**

I.1 INTRODUCTION

Les travaux de recherche relatifs au suivi de la consommation d'énergie électrique ont vu plusieurs développements et avancements, ces dernières années. Et vu que la façon dont le réseau électrique procure de l'énergie au consommateur a changé, de l'état classique qui était la circulation de l'électricité en un seul sens (de la station de production vers le consommateur) vers un nouveau état dont le flux d'énergie à deux sens, à cause de l'exploitation des sources d'énergies renouvelables ; le besoin de la gestion intelligente du réseau électrique était une nécessité afin d'améliorer la supervision et la mesure de la consommation de l'énergie.

Cependant, la surveillance intelligente de la charge et l'identification des appareils électriques ont vu plusieurs développements et avancements surtout avec l'arrivée de l'intelligence artificielle, dans l'objectif de mieux contrôler et limiter la consommation d'énergie et aussi pour superviser le réseau électrique domestique pour une alimentation en énergie plus efficace et plus sécurisée.

I.2 CONTEXTE GENERAL SUR L'IDENTIFICATION DES APPAREILS ELECTRIQUES

L'identification des appareils domestiques est un domaine très vaste où les études datent depuis les années 80, c'est un domaine qui consiste à reconnaître les appareils dans un réseau et surveiller leur fonctionnement et leur consommation d'énergie individuellement. Trois approches peuvent être considérées en ce qui concerne l'identification électrique : intrusive, semi intrusive, et non intrusive [1] [2].

I.2.1 Approches intrusives: Les méthodes de suivi intrusives ILM (Intrusive Load Monitoring), consistent à installer des semi compteurs attachés à chaque appareil, qui envoient l'information à une unité centrale de traitement pour l'inclure dans le rapport de consommation générale [3]. L'ILM peut être trouvé dans :

- Sous-comptage électrique : un calculateur d'énergie est utilisé pour chaque appareil.
- Appareils intelligents : les appareils ont des circuits communicants qui leur permettent de reporter leur consommation à une unité centrale.
- Sondage électrique : un signal électrique est injecté au réseau domestique et le signal retourné est analysé. L'intrusivité vient ici du fait d'injecter au réseau un signal extérieur qui pourrait affecter la qualité du réseau. [2]

L'approche intrusive n'est pas une solution fiable, car elle est complexe et coûteuse.

I.2.2 Approches semi-intrusives (Semi-Intrusive Load Monitoring SILM) : La surveillance semi-intrusive de la charge est en quelque sorte une version détendue de l'ILM où au lieu d'utiliser des compteurs d'énergie un par appareil, nous utilisons un compteur d'énergie pour mesurer un groupe d'appareils. L'approche SILM partage les mêmes inconvénients que l'ILM mais avec moins d'intrusion.

I.2.3 Approches non intrusive (Non-Intrusive Load Monitoring NILM) : La surveillance de la charge non intrusive (NILM) est un domaine où la principale préoccupation est de décomposer une consommation d'énergie agrégée en ses différents éléments constitutifs de manière non intrusive en utilisant un seul compteur d'énergie. Par exemple, au lieu d'introduire différents compteurs d'énergie placés tout autour du foyer nous plaçons un seul compteur d'énergie au niveau du panneau principal [4]. La **figure I.1** donne un schéma sur le principe de mesure de cette approche. Le système basé sur l'approche NILM est divisé en trois parties (**figure I.2**):

- L'acquisition des données : c'est le prétraitement et le stockage des données globales.
- L'extraction des caractéristiques de chaque appareil à travers la modélisation des charges.
- Les algorithmes pour faire la désagrégation ou l'identification de chaque charge.

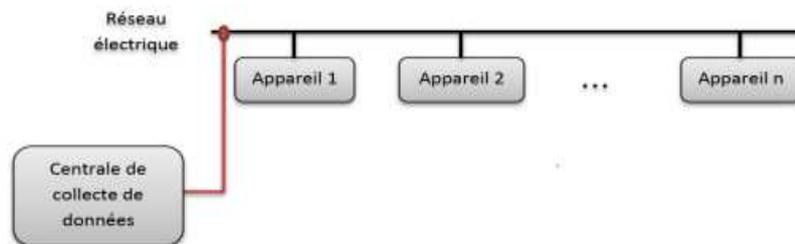


Figure I.1 : Principe de mesure de l'approche non intrusive [3]

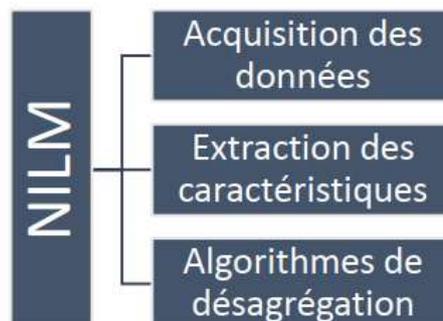


Figure I.2 : Structure de système NILM

I.3 CONCEPTION D'UN SYSTEME D'IDENTIFICATION DES APPAREILS ELECTRIQUES

Les appareils électriques peuvent être identifiés à partir de la signature de leurs courants électriques. Cependant, le système d'identification de ces appareils indiqué à la figure I.3, doit contenir :

- une phase d'apprentissage pour apprendre et modéliser les différentes classes d'appareils.
- une phase de test pour classifier un ensemble de signaux de test et évaluer les performances du système d'identification. Ces deux phases exigent la préparation d'une base de données d'apprentissage et une base de données de test.
- une étape d'extraction des caractéristiques du signal qui sont considéré comme empreinte du signal électrique.
- une phase de classification qui est un acte cognitif ou une procédure qui permet d'assigner à l'appareil une famille à laquelle il doit appartenir (une classe). Cette phase utilise une des classificateurs standards tels que : ADI, KNN, GMM, etc.

Afin de pouvoir extraire les caractéristiques des signaux nous devons décomposer le signal du courant en utilisant l'analyse harmonique et/ou les paramètres statistiques.

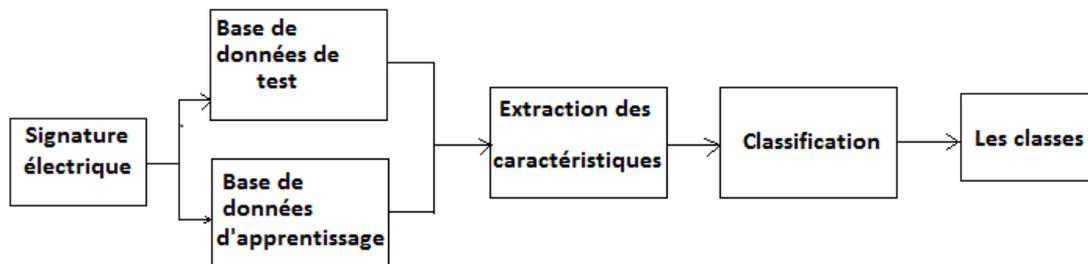


Figure I.3 : Schéma synoptique de conception d'un système d'identification

I.3.1 L'ANALYSE HARMONIQUE [5]

La méthode universelle d'analyse harmonique est celle de la transformée de Fourier (FT). Dans le cas discret, le vecteur du signal courant $s[n]$ de longueur L peut être analysé en utilisant la transformée de Fourier discrète (TFD) :

$$S[K] = \sum_{N=0}^{L-1} S[n] \text{Exp}\left(\frac{-j2\pi kn}{L}\right) ; K = 0, \dots, K - 1$$

Notre approche pour l'extraction de la chaîne de vecteurs est principalement basée sur la division de chaque signal de courant électrique en fenêtres superposées par une demi-période, puis sur le calcul de la série de Fourier discrète (SDF) dans chaque fenêtre :

$$S[n] = \sum_{N=0}^{L-1} C_K \text{Exp}\left(\frac{-j2\pi kn}{N}\right)$$

où N est la période de s [n] en échantillons et Ck sont les coefficients de la série de Fourier exprimés comme suit :

$$C_K = \frac{1}{N} \sum_{N=0}^{L-1} S[n] \text{Exp}\left(\frac{-j2\pi kn}{N}\right)$$

Les coefficients obtenus sont appelés coefficients STFS. Ils peuvent être empilés dans un vecteur de coefficients STFS (calculé sur une fenêtre), qui est appelé une observation. Ainsi, chaque signal de courant électrique est représenté par une séquence vectorielle, qui peut être mise dans une matrice harmonique MH (P, R) représentant le signal complet, où P est le nombre d'harmoniques (sa valeur optimale sera discutée plus tard) et R est le nombre de fenêtres lié à la durée du signal de courant d'entrée.

I.3.2 Paramètres statistiques [5]

Le calcul des paramètres statistiques est réalisé sur les lignes de la matrice harmonique. On obtient des vecteurs statistiques de longueur égale à P (nombre d'harmoniques). Nous avons sélectionné des mesures statistiques jusqu'au deuxième ordre calculé sur les valeurs absolues de la matrice :

- La moyenne : $\mu = E[M H]$
est calculée par la formule : $\mu(p) = \frac{1}{R} \sum_{r=1}^R M H(p, R), p = 1, \dots, P;$
- L'écart type : $\sigma = \sqrt{E[(M H - \mu)^2]}$
est calculée par la formule : $\sigma(p) = \sqrt{\frac{1}{R} \sum_{r=1}^R [(M H(p, R) - \mu(p))^2]}$

Enfin, chaque courant électrique d'appareil est représenté par un vecteur composé de deux valeurs statistiques temporelles caractérisant l'harmonique d'ordre p :

$$V(p) = [\mu(p) \quad \sigma(p)]$$

I.4 TRAVAUX PRECEDENTS SUR L'IDENTIFICATION DES APPAREILS ELECTRIQUES

Bien que plusieurs travaux soient menés dans le domaine de l'identification des appareils électriques et la conception d'un système de surveillance de charge, peu de ces travaux ont utilisé une base de données échantillonnée à haute fréquence. Ainsi, le **tableau I.1** énumère la liste de travaux et d'études menées en utilisant des bases de données

échantillonnées à **30hz** ou plus. Ces travaux sont classés selon leurs meilleurs résultats et les méthodes de classification utilisées pour les obtenir.

Algorithme D'apprentissage	Caractéristique utilisée (Régime transitoire/permanent)	Précision %	Apprentissage (Supervisé/Non supervisé/les deux)
SVM	Permanent/Transitoire	78%-85.8%	Supervisé
	Permanent/Transitoire	85%-90%	Supervisé
Naïve Bayes	Permanent/Transitoire	97.09%-99.21% (Permanent) 97.69%-99.17% (Transitoire)	Supervisé
	Permanent	81.41%-81.42%	Supervisé
	Permanent	66.6%-85.9%	Supervisé
Neural Networks	Permanent/Transitoire	84%	Supervisé
	Permanent/Transitoire	89.59%	Supervisé
	Permanent/Transitoire	48%-100% (selon l'appareil)	Supervisé
KNN	Permanent/Transitoire	68%-76%	Supervisé
	Permanent/Transitoire	43.6%-87.3%	Supervisé
HMM	Permanent	81.5%	Non supervisé
	Permanent	64.1%-99.8%	Non supervisé
	Permanent	60.3%-87.2%	Non supervisé

Tableau I.1 : Quelques travaux dans le domaine de l'identification des appareils électriques [6]

I.5 Bases de données des courants électriques

I.5.1 La base PLAID :

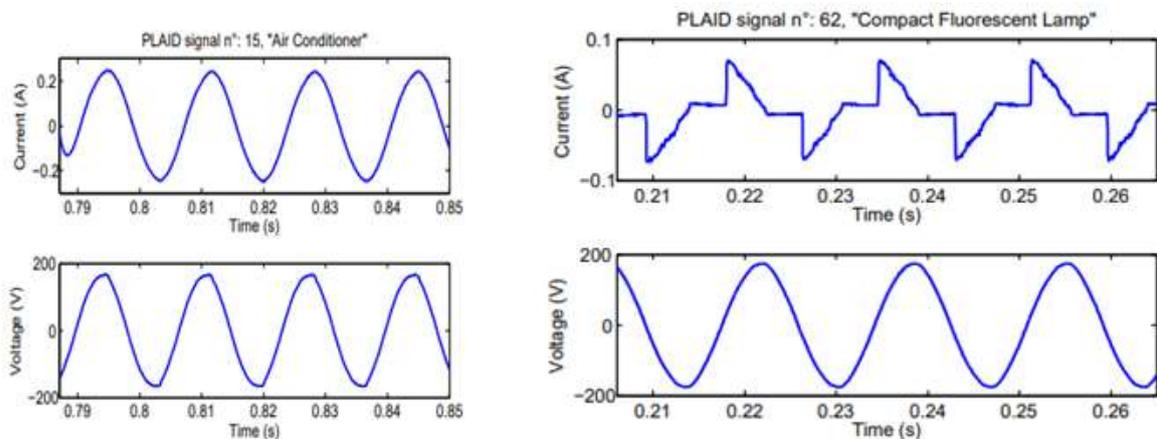
La base 'Plug Load Appliance Identification Dataset' (PLAID) est une base de données publique de signatures électriques. Ces signatures sont des mesures de courant et de tension prises au cours de l'été 2013 à Pittsburgh, Pennsylvanie, États-Unis, à partir de 55 ménages. Cette base de données contient 11 types d'appareils et pour chaque appareil de trois à six

exemples (l'ensemble de données contient un total de 1074 signatures). Chaque signature de la base de données fait quelques secondes de durée, contenant le régime transitoire de mise en marche et une partie du régime établi (correspondant à la phase de la consommation régulière). Ces signaux sont échantillonnés à une fréquence de 30 kHz. **Le tableau I.2** résume les appareils constituant la base de données : les différents types d'appareils et le nombre d'exemples existant dans la base pour chaque type.

Type d'appareil	Nombre d'exemples
Climatiseur	66
Lampe Fluo-compact	175
Ventilateur	115
Réfrigérateur	38
Séchoir	156
Radiateur	35
Lampe à incandescence	114
Ordinateur portable	172
Micro-ondes	139
Aspirateur	38
Machine à laver	26
Total	1074

Tableau I.2 : Résumé des appareils trouvés dans la base PLAID [2]

Deux exemples de signaux de PLAID sont donnés dans la **figure I.4**. [2]



a) Courant et tension d'un climatiseur –zoom b) Courant et tension d'une lampe fluo-compacte–zoom

Figure I.4 – Deux exemples de signaux de la base PLAID

I.5.2 La base COOLL :

COOLL est un ensemble de données de mesures électriques au niveau de la prise. Il contient le courant et la tension de mise sous tension mesures de 42 appareils échantillonnés à $F_s = 100$ kHz (840 mesures de courant et 840 mesures de tension). Les appareils sont de 12 types différents avec un certain nombre d'exemples chacun (**tableau I.3**). [7]

N°	Type d'appareil	Type d'appareil	Nombre d'appareil par Type	Nombre de signaux (20 par appareil)
1	Perceuse	Drill	6	120
2	Ventilateur	Fan	2	40
3	Meuleuse	Grinder	2	40
4	Sèche-cheveux	Hair dryer	4	80
5	Taille-haie	Hedge trimmer	3	60
6	Lampe	Lamp	4	80
7	Décapeur thermique	Paint stripper	1	20
8	Robot électrique	Planner	1	20
9	Défonceuse	Router	1	20
10	Ponceuse	Sander	3	60
11	Scie	Saw	8	160
12	Aspirateur	Vaccum cleaner	7	140
			42	840

Tableau I.3: Résumé de la base de données COOLL

Les appareils du jeu de données COOLL sont mesurés individuellement, un à la fois. Le jeu de données ne contient pas de scénario où plusieurs appareils sont mesurés simultanément. En outre, les appareils sélectionnés sont choisis de manière à ce que le contrôle de l'instant de mise en marche soit possible, c.-à-d. en appuyant au préalable sur le bouton de mise en marche, l'appareil peut être commandé électroniquement (avec triacs contrôlés). Chaque mesure dure 6 secondes avec une durée de pré-déclenchement de 0,5 seconde (le pré déclenchement des premières mesures étant différent et égales à 1 seconde) et une durée post-

arrêt de 1 seconde. Ces durées correspondent respectivement à l'heure à laquelle l'appareil est éteint avant la mise sous tension et après la mise hors tension.

Pour chaque appareil, 20 mesures contrôlées sont effectuées. Chaque mesure correspond à un délai d'action spécifique allant de 0 à 19 ms avec un pas de 1 ms (on couvre ainsi l'ensemble durée du cycle de temps de la tension secteur de 50 Hz, soit 20 ms). Ce délai d'action correspond au temps avec lequel l'action de mise en marche est retardée par rapport au début d'un cycle de temps spécifique de la tension secteur 1.

I.6 Conclusion

Un aperçu sur les approches de surveillance de charge et d'identification des appareils électriques est exposé, dans ce chapitre. Un schéma synoptique de la conception de système d'identification est expliqué et deux méthodes d'extraction des caractéristiques sont détaillées.

On a terminé par un résumé des deux bases des signatures de courant électriques les plus utilisés pour l'évaluation et la validation des systèmes d'identification des appareils électriques.

Chapitre II : Méthodes de classification

et

Approches de Combinaison des classifieurs

II.1 Introduction

Dans ce chapitre, on présente une théorie sur les méthodes de classification dont on concentre sur les classifieurs utilisés dans notre implémentation, d'une part et on détaille les approches de combinaison des classifieurs, d'autre part. Le domaine de recherche sur les systèmes multi-classifieurs (MCS) (ou combinaison de classifieurs), s'attache à trouver comment plusieurs classifieurs peuvent être appliqués pour obtenir de **meilleurs systèmes de classification**. La méthode ou la stratégie de combinaison (appelée également fusion de décision) dépend fortement de la nature de ces classifieurs.

II.2 Classification

La **classification** est une étape importante dans l'**analyse des données** et consiste à regrouper ces données en classes homogènes. Il existe deux types : la classification **supervisée** et la classification **non supervisée**. Ces deux approches se différencient par leurs méthodes et leurs buts.

II.2.1 Classification supervisée :

Le processus de classification peut se découper en deux phases la première est une phase d'apprentissage, pendant laquelle l'algorithme cherche des règles de classification (au sens large). La seconde phase consiste à appliquer les règles de la classification découverte, à un ensemble d'objets afin d'identifier la classe d'appartenance de chaque objet.

La phase d'apprentissage d'un algorithme de classification supervisée est basée sur un ensemble d'objets (appelée ensemble d'apprentissage) dont la classe de chacun est connue.

Le but de la classification supervisée est de découvrir la structure des classes sur cet ensemble de données et de généraliser cette structure à un ensemble des données plus large.

II.2.2 Classification non supervisée :

La classification non supervisée consiste à diviser un ensemble de données en sous ensemble, appelés classes de sorte que les classes soient le plus homogènes possible suivant un critère définit. Les critères les plus utilisés sont : la distance entre les objets, la densité des classes ou des mesures probabilistes. Les objets sont regroupés selon le critère qu'utilise la méthode de classification employée [8].

II.3 Classifieurs

Un classifieur est tout algorithme qui trie les données dans des classes étiquetées, ou des catégories d'information. Un simple exemple pratique sont les filtres de spam qui

scannent les Emails reçus et les classifient soit en « spam » ou « non spam ». Ils ont un ensemble de règles dynamiques qui incluent les procédures d'interprétation qui s'occupe des valeurs vagues ou inconnues, la plupart des classifieurs utilisent les estimations de probabilité qui permettent aux utilisateurs de manipuler les données de classification avec des fonctions d'utilité.

Puisqu'il n'existe pas une seule forme de classification qui est appropriée pour toutes les bases de données, de vastes types de classifieurs sont disponible pour l'expérimentation.

II.4 Méthodes de classification

II.4.1 Classification KNN :

A. Définition : L'algorithme des k les plus proches voisins («k-nearest neighbors» ou kNN) est un algorithme de classification supervisée qui consiste à affecter une classe à un vecteur de paramètres de test en comparant ce dernier à un ensemble de vecteurs étiquetés, préalablement enregistré durant la phase d'apprentissage. Cette comparaison vise à sortir parmi cet ensemble les k vecteurs les plus proches voisins au vecteur considéré, en termes de **distances**. La classe affectée au vecteur de test est la classe la plus votée parmi les k classes obtenue dans l'étape de comparaison [3].

B. Principe de fonctionnement : Le principe de cet algorithme de classification est associé à une fonction de distance et à une fonction de choix de la classe majoritaire en fonction des classes des voisins les plus proches, qui constitue le modèle. La **figure II.1** montre ce principe qui a pour but de trouver la valeur de la classe où l'inconnu 'x' va être affecté [9].

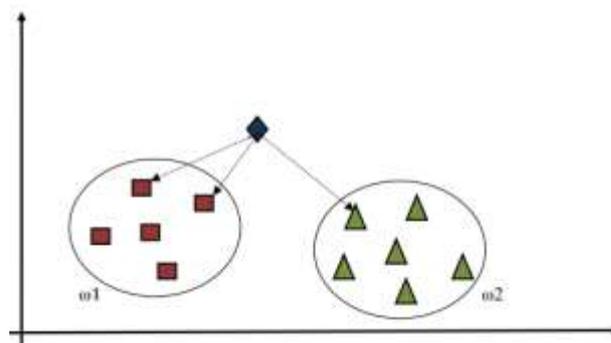


Figure II.1: Principe de fonctionnement de l'algorithme K-ppv.

Nous avons donc deux classes, et nous avons pris $k=3$. Parmi les 3 voisins, nous avons 2 qui appartiennent à ω_1 et 1 qui appartient à ω_2 donc 'x' sera affecté à ω_1 , la classe majoritaire.

Ce principe peut être formulé dans l'algorithme présenté par la **figure II.2**.

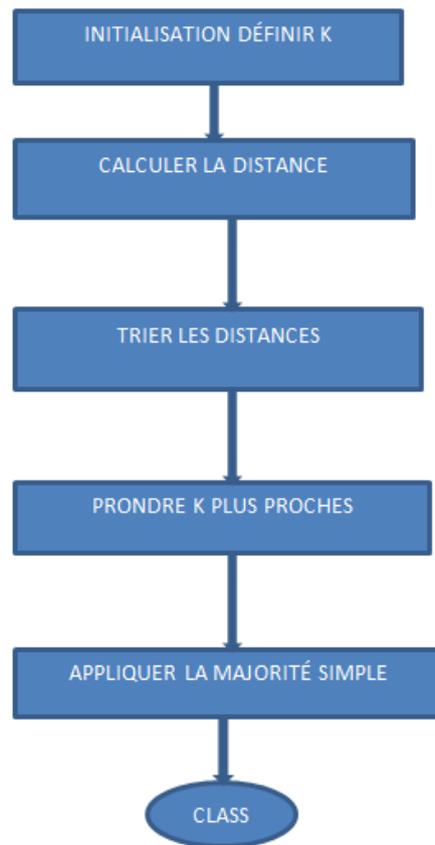


Figure II.2 : Fonctionnement de KNN [9]

C. Étapes de la classification KNN :

1. Phase de formation : un modèle est construit à partir des instances de formation.

- L'algorithme de classification trouve des relations entre les prédicteurs et les cibles.
- Les relations sont résumées dans un modèle.

2. Phase de test : tester le modèle sur un échantillon de test dont les étiquettes de classe sont connues mais non utilisées pour l'apprentissage du modèle. Les étiquettes sont connues mais n'ont pas été utilisées pour l'entraînement du modèle.

3. Phase d'utilisation : utiliser le modèle pour la classification sur de nouvelles données dont les étiquettes de classe sont inconnues.

D. Différentes distances utilisées :

Il existe différentes distances utilisées dans l'étape de comparaison de l'algorithme KNN telles que les distances : Euclidienne, Cityblock, Cosinus, etc

Soit X, Y deux instances, chacun se représente par un vecteur de p attributs, tel que :

$$X = \{ X_1, \dots, X_p \} \text{ et } Y = \{ Y_1, \dots, Y_p \}$$

- **Distances Euclidienne** : c'est La plus simple et la plus populaire des mesures de similarités entre des données multi variées, qui est un cas particulier de la famille des métriques.

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

- **Distance de City block** : Cette distance est définie comme suit [3]:

$$d_E(X, Y) = \sum_{l=1}^p |X_l - Y_l|$$

- **Distance Cosinus** : La distance cosinus, également appelée distance angulaire, est dérivée de la similitude cosinus qui mesure l'angle entre deux vecteurs. Elle est exprimée par [3]:

$$D_{cos} = 1 - \frac{\sum_{l=1}^p (X_l * Y_l)}{\sqrt{\sum_{l=1}^p X_l^2} * \sqrt{\sum_{l=1}^p Y_l^2}}$$

II.4.2 Classification par les Réseaux de neurones :

A. Définition : Les réseaux de neurones sont composés par des structures cellulaires artificielles, constituant une approche permettant de résoudre plusieurs problèmes concernant la perception, le mémoire, l'apprentissage et le raisonnement. Ils s'avèrent aussi des choix très prometteurs pour dépasser certaines limitations des ordinateurs classiques. Grâce à leur manipulation parallèle de l'information et à leurs mécanismes inspirés des cellules nerveuses (neurones), ils infèrent des propriétés émergentes permettant de résoudre des problèmes jadis qualifiés de complexes. Un neurone est une cellule cérébrale dont la fonction fondamentale consiste à collecter, traiter et transmettre des signaux électriques. Nous pensons que la capacité du cerveau à traiter les informations marque essentiellement la mise en réseaux de neurones. C'est pourquoi, quelques premières recherches en intelligence artificielle (IA) ont eu pour objectif d'élaborer des réseaux de neurones artificiels [10].

B. Les réseaux de neurones artificiels : Un neurone artificiel fait une sommation pondérée des potentiels d'actions qui lui parviennent (chacun de ces potentiels est une valeur numérique qui montre l'état du neurone qui l'a émis), puis s'active selon la valeur de cette addition (sommation) pondérée. Si cette addition dépasse un certain seuil, le neurone est activé et propage une réponse dont la valeur est celle de son stimulation (activation), sinon le neurone reste inactif et ne propage rien.

Chaque neurone formel (illustré par la **figureII.3**) accepte un nombre variable d'entrées. A chacune de ces entrées est lié un poids 'w' représentatif de la force de la connexion. Chaque neurone est accordé d'une sortie unique, qui permet d'alimenter un nombre variable de neurones avals. Le neurone traite la somme pondérée de ses entrées, puis il calcule sa sortie par une transformation non linéaire de cette addition. Les pondérations ou les poids représentent l'intensité synaptique de ce neurone [10]. Le fonctionnement d'un neurone formel est exprimé par les expressions suivantes:

$$E_i = \sum_{J=1}^n X_i W_{JI}$$

$$Y_i = F(E_i - \theta)$$

Tel que :

- X_i** : Signaux d'entrée du neurone artificiel i ;
- w_{ji}** : Poids (coefficients) des entrées ;
- Y_i** : La sortie du neurone artificiel i
- E_i** : Entrée globale ;
- θ** : Seuil d'activation du neurone ;

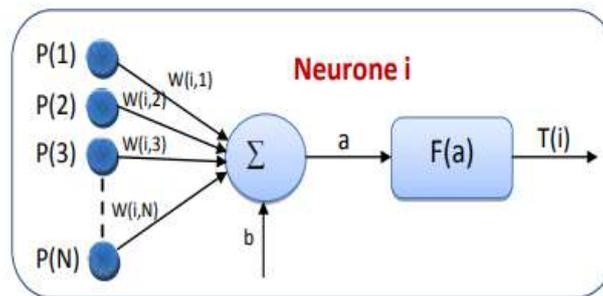


Figure II.3 : Structure d'un neurone formel.

C. Classification neuronale : Les réseaux de neurones artificiels ANNs (Artificial Neural Networks) sont apparus comme des outils intéressants dans divers domaines industriels et de recherche. Ils sont capables de résoudre les problèmes non linéaires et complexes comme la reconnaissance des formes, la classification ou l'optimisation. Un réseau neuronal artificiel est un modèle mathématique de calcul qui s'inspire de la structure des aspects fonctionnels des réseaux de neurones biologiques. L'apprentissage des ANNs est réalisé grâce à des algorithmes, ces derniers sont développés selon certaines règles d'apprentissages spéciaux.

D. Le perceptron multicouche: Le perceptron multicouche MLP (Multi Layer Perceptron) est un modèle non linéaire, le plus utilisé dans les réseaux de neurones. Il peut contenir deux ou plusieurs couches (**Figure II.4**). La couche d'entrée est la première couche dont le nombre de neurones égal au nombre de caractéristiques spécifiques sélectionnées. La couche de sortie est la dernière couche qui détermine la sortie, le nombre de neurones dans cette couche dépend du nombre de classes désirées. Les couches intermédiaires sont ajoutées afin de renforcer la capacité d'apprentissage du réseau [11].

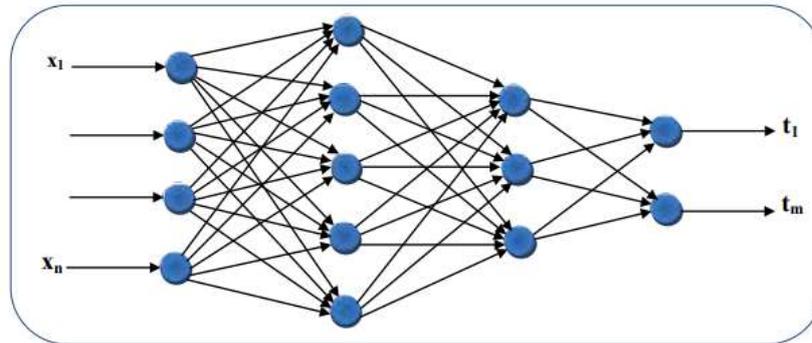


Figure II.4 : Un réseau perceptron multicouche.

II.4.3 Classification discriminante :

A. Définition : L'analyse discriminante est une technique de classification statistique bien connue. Elle utilise des données de formation pour estimer les paramètres des fonctions discriminantes des variables prédictes. Les fonctions discriminantes déterminent les frontières dans l'espace des prédictes entre différentes classes. Le classifieur qui en résulte discrimine entre les classes (les niveaux catégoriels de la réponse) en fonction des données prédictives. Il existe cinq différents types de classifieurs d'analyse discriminante à savoir : linéaire diagonale, linéaire, quadratique, quadratique diagonale et Mahalanobis.

B. Analyse Discriminante Linéaire (LDA) : C'est une approche probabiliste qui consiste à regrouper les données où une donnée sera considérée proche de l'autre si elle provient de la même densité de probabilité plutôt qu'en effectuant un simple calcul de distance déterministe. Chaque classe de données pourra donc être résumée par les paramètres de sa densité de probabilité. Chaque classe est supposée être Gaussienne.

B.1 Apprentissage du modèle LDA : [12]

L'apprentissage du modèle LDA consiste en l'estimation des paramètres μ_k et Σ à partir des données d'apprentissage. Ces estimations sont données par $\pi_k = \frac{n_k}{n}$ où n_k est le cardinal de la classe k , z_i est l'étiquette de classe de l'exemple x_i et $z_i = k$ désigne le fait que l'exemple x_i

appartient à la classe k . L'estimation des paramètres μ_k (les moyennes) et Σ (matrice de covariance) à partir des données d'apprentissage sont données par

$$\mu_k = \frac{1}{n_k} \sum_{i/z_i=k} x_i$$

$$\Sigma = \frac{1}{n-k} \sum_{k=1}^k \sum_{i/z_i=k} (x_i - \mu_k)(x_i - \mu_k)^T$$

B.2 Test du modèle : [12]

Dans L'analyse LDA, une fois les paramètres sont estimés, on peut prédire les classes des données de test par la règle du MAP (maximum a posteriori). Cette règle consiste à maximiser les probabilités a posteriori, ç-à-d affecter chaque donnée de test x_i à la classe z_i^* ayant la plus grande probabilité a posteriori :

$$\tau_{ik} = P(z_i = k | x_i; \mu_k, \Sigma) = \frac{\pi_k f_{ik}(x_i; \mu_k, \Sigma)}{\sum_{l=1}^k \pi_l f_{il}(x_i; \mu_l, \Sigma)}$$

et on a donc :

$$z_i^* = \arg \max_{k=1}^K \tau_{ik} \quad (i = 1, \dots, n)$$

II.4.4 Autres méthodes de classification :

Il existe de vastes autres types de classifieurs qui sont disponible pour l'expérimentation dont on cite : HMM, GMM, SVM, etc.

II.5 Taux de classification

Un taux de classification est utilisé pour évaluer les résultats de test de n'importe quel système de classification. Donc, il représente la mesure de performance et de validité des systèmes d'identification et/ou de classification.

Le taux de classification TC est défini, en général, par la relation suivante:

$$TC = \frac{N_{SC}}{N_{ST}} * 100$$

Où N_{ST} est le nombre Total des signaux de la base de test, N_{SC} est le nombre de signaux correctement identifiés.

Ainsi, une autre mesure de performances consiste à considérer le taux de classification des vecteurs de paramètres de tous les signaux de la base de données sans tenir en compte la classification des signaux. Ce taux est défini comme suit :

$$TCV = \frac{N_{VC}}{N_{VT}} * 100$$

Où N_{VT} est le nombre total des vecteurs de paramètres de tous les signaux de la base de test, N_{VC} est le nombre de vecteurs de paramètres correctement classifiés. Ce dernier taux de classification nous permet d'évaluer globalement les performances du système en tenant compte la classification d'un grand nombre de vecteurs.

II.6 Combinaison des classifieurs

II.6.1 Pourquoi combiner les classifieurs ? [13]

Plusieurs chercheurs ont fait les constats suivants :

- Il n'existe pas de «meilleur» classifieur capable de traiter n'importe quelle distribution des données d'apprentissage.
- Aucun classifieur ne peut discriminer suffisamment et correctement un ensemble important de classes.
- Lorsqu'on utilise plusieurs classifieurs séparément pour le même ensemble de formes, les ensembles des formes mal classées ne sont pas forcément les mêmes.

De ces constats et de bien d'autres aussi, a émergé l'idée de faire *coopérer* les classifieurs.

L'idée principale derrière la combinaison de classifieurs est **l'augmentation de performances**.

Cette augmentation de performance peut avoir plus de fiabilité dans les réponses, ou moins de rejet, ou bien à la fois. Donc, deux raisons majeures sont derrière la combinaison:

- **La Précision**: une décision plus fiable peut être obtenue en combinant l'avis (les sorties) de plusieurs classifieurs.
- **L'Efficacité**: un problème complexe peut être décomposé en plusieurs sous-problèmes qui sont plus faciles à comprendre et à résoudre.

II.6.2 Types de sortie du classifieur :

C'est le type de sortie proposé par les différents classifieurs qui va influencer en partie sur la manière de combiner les résultats des classifieurs. Un classifieur peut avoir l'une des sorties suivantes [14]:

- **Sortie de type classe** : C'est le type le plus général mais qui apporte le moins d'informations. Le classifieur ne donne que la proposition du type (la classe) de l'entrée à reconnaître sans aucune autre information $e_j(x) = C_i (i \in \{1..m\})$.

- **Sortie de type ensemble** : Le classifieur donne sa réponse sous forme d'un ensemble de classes candidates sans préciser ses préférences $e_j(x) = \{C_i / i \leq m\}$.
- **Sortie de type rang** : Ce type de sortie reflète l'ordre de préférence des propositions fournies par le classifieur, cela se traduit par l'attribution d'un rang pour chaque classe ; Plus cette dernière est probable moins le rang est élevé. La liste des propositions peut contenir toutes les classes possibles ou seulement les mieux classées. $e_j(x) = [r_{j1}, r_{j2}, \dots, r_{jm}]$ où r_{ji} est le rang attribué à la classe (i) par le classifieur (j).
- **Sortie de type mesure** : Cette sortie est la plus riche en informations puisque le classifieur dans ce cas associe à chaque classe une mesure de confiance qui peut être, par exemple, une probabilité. $e_j(x) = [M_{j1}, M_{j2}, \dots, M_{jm}]$ où M_{ji} est la mesure attribuée à la classe (i) par le classifieur (j).

II.6.3 Mesures de performances d'un classifieur [14]:

Pour une entrée donnée, un classifieur peut générer les réponses suivantes :

- **Un rejet** : pour indiquer que le classifieur n'a pas pu identifier cette entrée.
- **Une reconnaissance** : dans ce cas, il identifie bien l'entrée, et il lui attribue sa classe appropriée.
- **Une substitution** : le classifieur attribue une autre classe à l'entrée.

La performance d'un classifieur peut être mesurée en calculant les trois taux suivant :

$$\text{taux de rejet} = \frac{\text{Nombre de formes rejetées}}{\text{Nombre totale de formes}}$$

$$\text{taux de reconnaissance} = \frac{\text{Nombre de formes reconnus}}{\text{Nombre totale de formes}}$$

$$\text{taux de substitution} = \frac{\text{Nombre de formes males reconnus}}{\text{Nombre totale de formes}}$$

II.6.4 Stratégies de combinaison des classifieurs [15][16]

II.6.4.1 Combinaison séquentielle :

Elle permet, soit de conforter la décision prise auparavant, soit de traiter les rejets. Dans ce type de combinaison l'ordre d'exécution des classifieurs est primordial et influe sur le résultat final.

L'utilisation de la méthode de combinaison séquentielle ne fait intervenir les classifieurs avals que si les classifieurs utilisés en amont n'ont pas apporté une décision précise.

Dans le cas de la **Figure II.5**, si le classifieur 1 reçoit une liste de C classes possibles, il compare les résultats de ses extractions de caractéristiques aux C prototypes et génère une liste de candidats ordonnée de manière **décroissante** vis à vis **des probabilités**.

Il propose au classifieur suivant une liste $L1$ contenant les p premières propositions de la liste avec $p < C$. Le classifieur suivant est donc chargé de déterminer la classe de l'élément à identifier parmi les p classes qui lui sont proposées. Il générera à son tour une liste plus réduite ($L1 > L2 > L3 > \dots$). L'enchaînement des classifieurs continuera jusqu'à ce que le dernier classifieur prenne la **décision finale PN** en proposant le meilleur candidat de la liste qu'il aura lui-même établi.

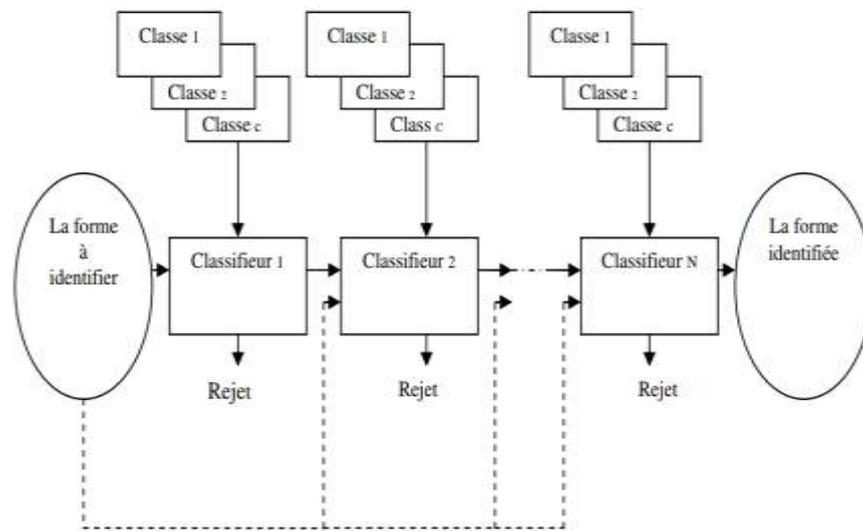


Figure II.5 : Schéma de combinaison séquentielle de classifieurs [13] [14] [15]

II.6.4.2 Combinaison parallèle :

La Figure II.6 illustre la combinaison parallèle. Les classifieurs opèrent indépendamment les uns des autres en une première phase, puis on fusionne leurs réponses respectives lors d'une dernière phase qui nécessite la définition d'une fonction de combinaison pour donner une décision finale. La **fonction de combinaison** de classifieur est liée d'une part à l'approche de combinaison utilisée et d'autre part au type de sorties générées par les classifieurs. La taxonomie des méthodes de combinaison est schématisée dans la Figure II.7.

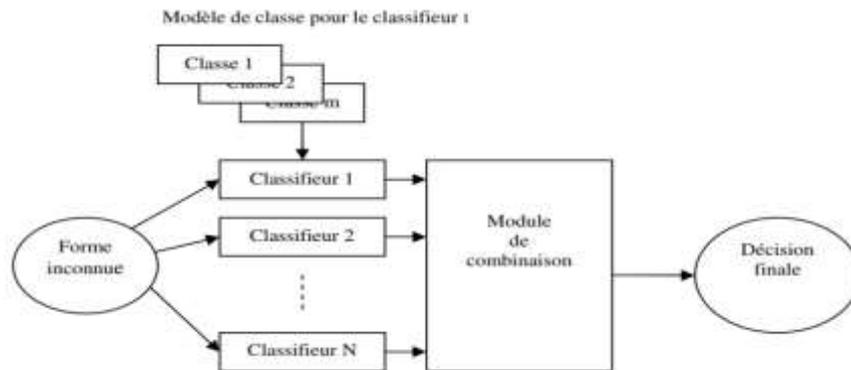


Figure II.6 : Combinaison parallèle de classifieurs [13] [14] [15] [16]

A. Approches de combinaison parallèle :

- ❖ **Sans apprentissage** : ces méthodes se contentent d'utiliser seulement et simplement les sorties des classifieurs sans intégrer d'autres informations a priori sur les performances de chacun des classifieurs [16][17].
- ❖ **Avec apprentissage** : ces méthodes de combinaison déterminent via une base d'apprentissage supplémentaire les paramètres de la combinaison. L'apprentissage s'effectue à partir d'une base de validation permettant notamment de prendre en compte l'efficacité de chacun des classifieurs de l'ensemble [16].

A.1 Fonctions de combinaison sans apprentissage si la sortie est de type « classe »

- **Méthodes de vote** : Toutes les méthodes de votes peuvent être dérivées de la règle de majorité avec seuil exprimée par [13] [16] [17]:

$$E(x) = \begin{cases} C_i & \text{si } \sum_i e(i) = \max_{C_i \in \{1, \dots, M\}} \sum_j e(j) \geq \alpha K \\ \text{rejet sinon} & \end{cases}$$

Où K est le nombre de classifieurs à combiner et α est le seuil qui représente la proportion du nombre de classifieurs devant répondre la même classe pour que cette classe soit la classe finale sinon rejet. On peut distinguer différents type de vote :

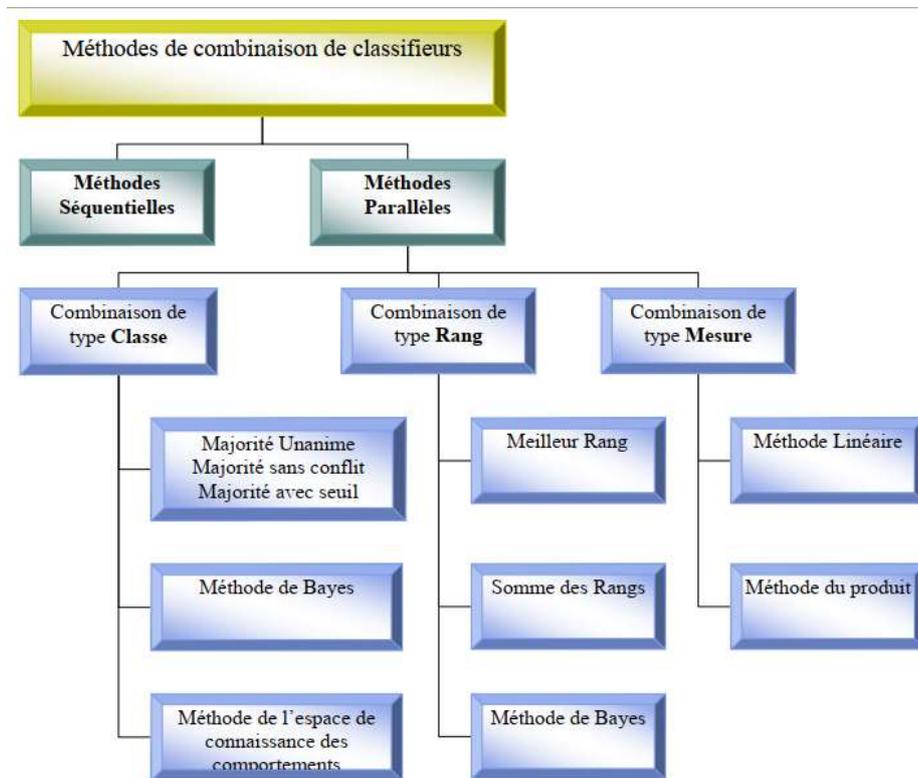


Figure II.7 : Taxonomie des méthodes de Combinaison [13]

- **La majorité unanime** : si $\alpha=1$, donc la classe finale est choisie si tous les classifieurs proposent cette réponse sinon la réponse finale est un rejet.
- **La majorité absolue** : la classe finale est décidée si plus de la moitié des classifieurs l'ont proposé, ici $\alpha=0,5$.
- **La majorité simple** : c'est la fonction de fusion la plus simple à implémenter. La sortie de chaque classifieur est considérée comme étant un vote pour une classe. Le nombre de votes pour chacune des classes est compté et l'ensemble choisit la classe en ayant remporté le plus ($\alpha=0$). Néanmoins, il y a un rejet si toutes les classes ont le même nombre de votes, dans ce cas, les risques de conflits sont particulièrement importants.

A.2 Fonctions de combinaison avec apprentissage si la sortie est de type « classe » :

- **Méthode de Bayes** : Elle introduit la matrice de confusion, cette dernière décrit toutes les erreurs que le classifieur est susceptible de faire en fonction des entrées qui lui sont

proposées, autrement dit, cette méthode utilise plus d'informations sur les classifieurs pour combiner leurs résultats. Le classifieur estime sa croyance avec une incertitude sur chacune des N classes par un nombre réel appelé croyance (ou belief). La méthode de combinaison connaît bien le comportement de chacun des classifieurs, c'est-à-dire qu'elle connaît la probabilité pour qu'une classe soit la vraie classe de x sachant que Cl_k a donné la réponse w_{jk} . La méthode Bayésienne repose sur la recherche de la classe w_i telle que la probabilité que x appartienne réellement à w_i connaissant la sortie de chaque classifieur soit la plus forte.

II.6.4.3 Combinaison hybride :

L'approche hybride consiste à combiner à la fois des architectures séquentielles et parallèles afin de tirer pleinement avantage de chacun des classifieurs utilisés. La Figure II.8 présente un exemple de combinaison hybride dans laquelle on combine un classifieur en série avec deux classifieurs en parallèle. Ce type d'approche permet de générer de nombreux schémas de coopération qui peuvent rapidement devenir complexes à optimiser. Il illustre les deux aspects de la combinaison qui sont d'une part la réduction de l'ensemble des classes possibles et d'autre part la recherche d'un consensus entre les classifieurs afin d'aboutir à une décision unique. [13]

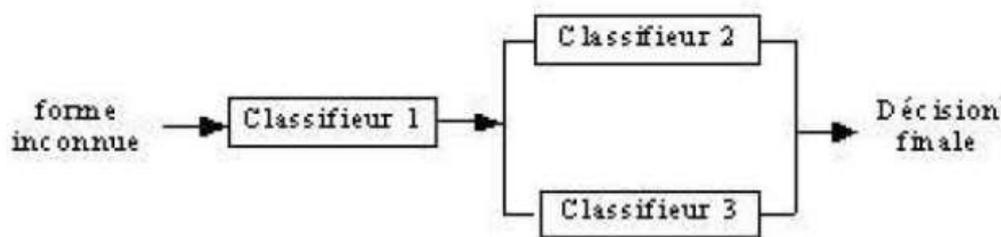


Figure II.8: Combinaison hybride de classifieurs [13] [17]

II.7 Conclusion

A travers ce chapitre, nous avons défini le fonctionnement de quelques classifieurs supervisés, qu'on utilisera par la suite, pour tester notre système d'identification des appareils électrique. La combinaison de classifieurs est un outil efficace pour avoir une grande performance sans augmenter la complexité des techniques de classification existantes. Elle est convenable pour les applications nécessitant une grande exactitude de classification. Parmi les approches de combinaison illustrées, dans ce chapitre, on considèrera la plus convenable à l'organisation de notre base vis-à-vis les entrées et les sorties du système.

Chapitre III: Implémentation et Résultats de Classification et de combinaison des classifieurs

III.1 Préliminaire

Notre travail est divisé en deux parties. Dans la première partie, on implémente deux systèmes d'identification des appareils électriques. Le premier système est basé sur l'analyse harmonique et le classificateur **KNN** puis le classificateur **ADI** (analyse discriminante) puis le classificateur **ANN** (Artificial Neural Network), avec l'utilisation d'une meilleure configuration – obtenue par expérience - pour chaque classificateur. Le deuxième système est basé sur les paramètres statistiques du courant électrique et les mêmes types de classificateurs. Ces systèmes sont testés en appliquant la base de données PLAID (décrite dans le premier chapitre) et ils sont évalués en calculant le taux de classification.

Dans la deuxième partie de ce travail, on conçu un système multi-classifieurs qui combine les sorties des trois classificateurs (KNN, ADI et ANN) avec une fonction de décision, dans l'objectif d'implémenter une structure de combinaison pour les systèmes d'identification des appareils électrique et d'améliorer les performances de classification.

III.2 Implémentation des systèmes d'identification

La conception des deux systèmes suit les mêmes étapes, mais elle se distingue dans l'extraction des paramètres caractéristiques : utilisation des coefficients de la série de Fourier pour le premier système et l'utilisation des paramètres statistiques pour le deuxième système.

La première étape de cette conception est la préparation de la base de données des courants électriques. On a utilisé ici la base **PLAID** prétraitée dans les travaux de recherche de [5] [18]. Ce prétraitement a permis d'obtenir que les signaux utiles de fonctionnement des appareils électriques en enlevant les durées de mise en marche et d'arrêt. Le nombre de signaux obtenu est de **1074** enregistrés dans un intervalle de **5 secondes** avec une fréquence d'échantillonnage de **30 KHz**.

La deuxième étape consiste à diviser la base de données en une base d'apprentissage et une base de test qui devront contenir tous les types d'appareils figurant dans la base de données PLAID d'origine. Donc, chaque base se compose de **537 signaux** qui représentent les **11 types** d'appareils électriques. A Chaque signal, on affecte un **indice** et on le découpe avec une durée de 20ms en **250 vecteurs**.

Dans **la troisième étape**, on extraire les paramètres caractéristiques comme suit :

- Pour le premier système, chaque vecteur (des deux bases : apprentissage et test) est converti en un ensemble de coefficients de Fourier par l'application de la formule d'analyse harmonique mentionnée dans le premier chapitre. Le nombre de coefficients est limité à **10** puisque, il est mentionné dans la référence [3] que ce nombre est suffisant pour

avoir de bonnes performances. Par conséquent, on obtient $10 * 250$ paramètres par un signal et $2500 * 537$ paramètres par base d'apprentissage ou de test.

- Pour le deuxième système, chaque signal (ou chaque séquence de vecteurs) est converti en un seul vecteur qui représente les paramètres statistiques (**moyenne et écart type**) (la meilleure combinaison mentionnée et prouvée dans la référence [5]) par l'utilisation des expressions données au premier chapitre. Ces paramètres statistiques sont calculés à partir des coefficients de l'analyse harmonique. Par conséquent, on obtient pour chaque base de données (apprentissage et test) **537 valeurs** pour la moyenne (μ) et **537 valeurs** pour l'écart type (σ). Dans ce cas, la taille des deux bases : apprentissage et test, après cette extraction, est réduite considérablement. Cela permettra de limiter le temps d'identification.

A chaque vecteur de ces paramètres caractéristiques, on associe une '**classe**'. L'ensemble des classes représente les 11 classes ou types d'appareils électriques. Aussi, l'ensemble des vecteurs de paramètres obtenus, dans les deux cas, sont regroupés dans deux matrices d'apprentissage et de test.

L'étape de classification consiste à classer les vecteurs de paramètres de la base de test selon les modèles de la base d'apprentissage pour **affecter** à chaque vecteur de test une '**classe**'. En appliquant, à chaque fois, l'un des algorithmes de classification suivants :

Algorithme KNN :

Arguments :

X : L'instance non classée

S : L'ensemble des instances (ensemble de test) qu'on devra classés

D : L'ensemble des instances (ensemble d'entraînement) déjà classés

Y_i : Les classes des instances de D

d : La distance euclidienne entre deux instances

Entrées : D, S, d, K

Sorties : Taux de classification

Début d'algorithme :

Pour chaque instance X dans S

Calculer $d(X, D)$ // calculer la similarité entre X et instances de l'ensemble D

Classer les distances par ordre croissant

Compter le nombre d'occurrences de chaque classe Y_i parmi les K plus proches selon l'ordre, assigner à la classe la plus fréquente.

Calculer le taux de classification

FIN retourner le taux de classification

Algorithme ADI :

Entrée: Données de test ; $\mathbf{X}^{\text{test}} = (\mathbf{x}^{\text{test } 1}, \dots, \mathbf{x}^{\text{test } n})$;

Paramètres d'apprentissage ($\mu_1, \dots, \mu_K, \Sigma$)

```

/* Calcul des densités pour les différentes classes */
for i ← 1 to n do
  for k ← 1 to K do
    Calculer la densité  $f_{ik}$  avec  $\mu_k$  et  $\Sigma$  pour chaque exemple de test  $\mathbf{x}_i$ 

$$f_{ik}(\mathbf{x}_i; \mu_k, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \Sigma^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\mathbf{x}_i - \mu_k)^T \Sigma^{-1} (\mathbf{x}_i - \mu_k)\right)$$

    End
  End
/* Classification */
for i ← 1 to n do
  Calculer la classe de chaque donnée de test  $x_i$  en maximisant les probabilités a
  posteriori  $\tau_{ik}$ 

$$\hat{z}_i = \arg \max_{k=1}^K \tau_{ik}$$

End
Sortie :  $\hat{z}_i (i = 1, \dots, n)$  : classes de test estimées.

```

Algorithme ANN :

1. Appliquez L'entrée.
2. Calculez les valeurs d'entrée 'de réseau de neurone' dans les unités de couche cachée.
3. Calculez-les sorties de la couche cachée.
4. Calculez les valeurs 'de réseau de neurone' d'entrée dans les unités de couche de sortie.
5. Calculez les sorties des unités de sortie.
6. Calculez le terme d'erreur pour les unités de sortie,
7. Calculez le terme d'erreur pour les unités cachées,
8. Mettez à jour les pondérations sur la couche de sortie.
9. Mettez à jour les poids sur la couche cachée.
10. Répétez les étapes de l'étape 1 à l'étape 9 jusqu'à ce que l'erreur soit suffisamment faible pour chaque paire de vecteurs d'apprentissage.

L'algorithme proposé est arrêté lorsque les carrés des différences entre les valeurs réelles et les valeurs désirées sont suffisamment petits.

Dans **l'étape d'évaluation**, des mesures de performances de nos expériences d'identification sont calculées, à savoir :

Le taux de classification des vecteurs :

$$T_{VC} = \frac{\text{nombre de vecteurs correctement classifiés}}{\text{nombre de vecteur total}} \times 100$$

Le taux de classification des signaux, en appliquant la règle de vote:

$$T_{SC} = \frac{\text{nombre de signaux correctement classifiés}}{\text{nombre de signaux total}} \times 100$$

III.3 Configuration matérielle et logicielle

Dans nos expériences, on a utilisé les outils suivants :

- Un PC Intel(R) core (TM) **I5_5300U@2,30GHZ**
- Système d'exploitation 64 bits.
- 8Go de RAM.
- Microsoft Word Office 2010.
- MATLAB R2016a.

III.4 Résultats d'implémentation et discussions

III.4.1 Expériences sur le système d'identification (SI) en utilisant l'analyse harmonique

III.4.1.1 Identification avec le classifieur KNN

Le tableau III.1 illustre les résultats (*taux de classification des vecteurs en %*) de simulation de **SI** avec un nombre de paramètres harmonique égal à 5, en variant les variables de classifieur KNN : le nombre des vecteurs les plus proches voisins **K** ainsi que le type de **la distance**. On peut observer que le taux de classification maximale de **93.95 %** est obtenu avec la configuration (K = 5, distance = 'Cityblock'). Pendant l'exécution de ces simulations, on a constaté que l'expérience avec la distance 'Cosinus' ou 'Correlation' prend un temps très long pour donner les résultats.

Distance K	Euclidienne	Cosinus	Cityblock
1	93.024	82.05	93.66
3	93.70	84.01	93.85
5	93.79	84.77	93.95
7	93.88	85.05	93.95
10	93.88	85.25	93.94

Tableau III-1: Sélection de la configuration optimal de classificateur KNN

Considérant la configuration (K = 5, distance = 'Cityblock'), notre système est simulé, dans ce cas, en variant le nombre de paramètres harmonique. Les résultats de cette expérience sont représentés dans le tableau III.2. Ils représentent la variation des deux taux de classification en fonction de nombre d'harmonique. On peut remarquer que, le taux de classification des signaux est supérieur au taux de classification des vecteurs et cela dû à l'application de la règle de vote pour classifier les signaux de la base de test. Le taux de classification s'améliore

avec l'augmentation de paramètres harmoniques dans les vecteurs de test pour arriver aux valeurs $T_{cv} = 97.77\%$ et $T_{cs} = 98.51\%$.

Nombre d'harmoniques	Tcv %	Tcs %	Temps de simulation (s)
5	93.95	94.97	5.97
10	95.75	95.90	7.5
20	95.39	95.90	531.47
50	97.59	98.32	1282.47
100	97.68	98.32	4203.810930
250	97.77	98.51	5219.024180

Tableau III-2 : Taux de classification de KNN en fonction de nombre d'harmoniques

III.4.1.2 Identification avec le classifieur ADI

Afin de déterminer la bonne configuration de ce classifieur, une simulation de SI en variant le type de la fonction discriminante, a donné les résultats présentés dans le tableau III.3. Vu que le meilleur taux de classification égal à 75.71 %, la configuration 'type = quadratique' sera considérée dans les prochaines expériences.

Les résultats de tableau III.4 montre que le meilleur taux de classification des signaux prend la valeur de 76.16 %. Mais, ce type de classifieur ne donne pas les bonnes valeurs de taux de classification par rapport à la classification KNN.

Type de fonction	Linéaire	quadratique	mahalonobis
Taux de classification Tvc	66.71	75.71	53.73

Tableau III-3: Sélection de la configuration optimal de classificateur ADI, Avec un nombre d'harmonique = 25

Nombre d'harmoniques	Tcv %	Tcs %	Temps de simulation (s)
5	59.98	60.14	5.24 s
10	69.73	70.20	5.27 s
20	73.51	74.30	5.74 s
50	74.69	76.16	7.28 s
100	73.42	74.86	11.24 s
250	71.49	72.62	23.15

Tableau III-4 : Taux de classification de l'ADI en fonction de nombre d'harmoniques

III.4.1.3 Identification avec le classifieur ANN

En tenant compte de taux de classification et de temps de calcul et après plusieurs simulations avec ce type de classifieur, on a constaté que la configuration suivante est la plus optimale :

- La division des données pour l'Apprentissage, la validation et les tests :
 - Apprentissage 70%
 - Validation 15%
 - Test 15%
- La structure finale présentée dans la **figure III.1**, est composée de trois couches :
 - Une couche d'entrée : 250 neurones (les paramètres d'entrée).
 - Une couche cachée (hidden layer) : 80 neurones.
 - Une couche de sortie : 11 neurones (les classes).

A partir de tableau III.5, on peut conclure que le meilleurs taux de classification est obtenu lorsque le nombre de paramètre égal à 10.

Nombre d'harmoniques	Tcv %	Tcs %	Temps de simulation (s)
5	74.0197	75.9777	1648.548478
10	83.3686	85.8473	1598.759509
20	76.9812	78.9572	1694.508216
50	79.5814	82.4953	3276.653851
100	77.8157	81.7505	2181.950994
250	69.2432	72.2533	11501.059097

Tableau III-5 : Taux de classification de l'ANN en fonction de nombre d'harmoniques

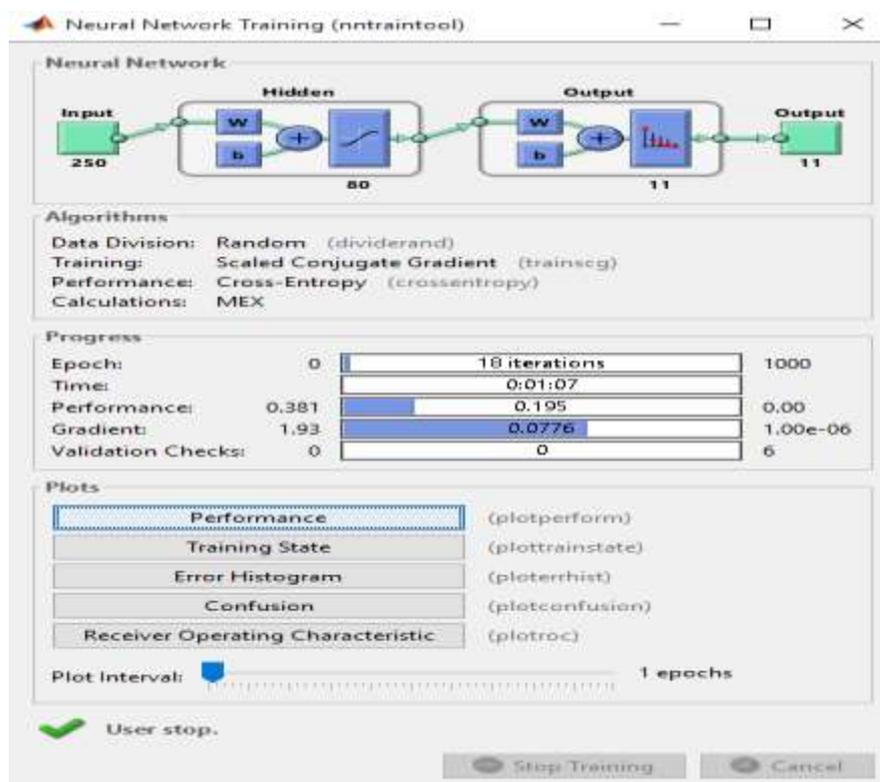


Figure III-1 : Visualisation du réseau réalisé sur la boîte à outils

III.4.2 Expérience sur le système d'identification en utilisant les paramètres statistiques

Durant ces expériences, on utilise les mêmes types de classifieurs, pour identifier les appareils électriques à travers les paramètres statistiques, dans l'objectif d'obtenir des meilleurs taux de classification. Les taux de classification, dans ce cas, sont évalués en fonction de nombre de vecteurs statistiques. Les valeurs maximales sont marquées dans les tableaux ci-dessous qui représentent les résultats de différentes simulations. Ce soulignement indique les paramètres pertinents pour les bonnes classifications.

III.4.2.1 Identification avec le classifieur KNN

Nombre de vecteurs statistiques	Tcv %	Tcs %
5	<u>95.07</u>	83.79
10	94.84	83.61
20	94.71	83.79
50	94.78	<u>84.35</u>
100	94.42	83.42
200	93.37	82.86

Tableau III-6 : Taux de classification de KNN en fonction de nombre de vecteurs statistiques

III.4.2.2 Identification avec le classifieur ADI

Nombre de vecteurs statistiques	Tcv %	Tcs %
5	73.84	69.64
10	<u>75.89</u>	<u>70.76</u>
20	71.98	67.03
50	69.77	67.03
100	70.09	65.54
200	71.00	64.80

Tableau III-7 : Taux de classification d'ADI en fonction de nombre de vecteurs statistiques

III.4.2.3 Identification avec le classifieur ANN

Nombre de vecteurs statistiques	Tcv %	Tcs %
5	82.33	74.48
10	85.66	78.39
20	77.74	75.04
50	<u>87.93</u>	<u>82.12</u>
100	85.95	80.81
200	78.69	74.48

Tableau III-8 : Taux de classification d'ANN en fonction de nombre de vecteurs statistiques

III.4.2.4 Comparaison entre les résultats des deux systèmes d'identification :

A travers les expériences sur les deux systèmes d'identification, on peut conclure que la taille de la base d'entrée est réduite et le temps de simulation est très rapide dans le deuxième système. Une comparaison de meilleur taux de classification des vecteurs (en %) obtenu dans les deux systèmes est donnée dans le tableau III.9. Les taux maximaux sont marqués en gris.

Type classifieur Base d'entrée	KNN	ADI	ANN
Paramètres de la série de fourrier	97.59	74.69	83.36
Paramètres statistiques	95.07	75.89	87.93

Tableau III.9 : Comparaison des résultats de classification des deux systèmes

III.4.3 Conclusion :

A travers les simulations précédentes, on a déterminé les configurations optimales et on a fixé le nombre de paramètres à utiliser pour obtenir la meilleure identification, pour chaque type de classifieurs. Les résultats des expériences sur les deux systèmes d'identification montrent que les paramètres statistiques sont plus pertinents par rapport aux séquences de vecteurs de paramètres de la série de fourrier, si on combine les trois facteurs de performances des simulations : taux de classification, espace mémoire et temps d'exécution.

III.5 Implémentation d'un système multi-classifieur (SMC)

Rappelons que notre objectif est la conception et l'implémentation d'un système basé sur une combinaison de trois classifieurs pour l'identification des appareils électriques avec un taux de classification appréciable.

III.5.1 Architecture du système proposé

Vu que, ce travail présente notre première contribution dans le domaine de combinaison, on propose d'implémenter l'approche la plus simple et la plus utilisée dans la littérature et qui est la combinaison parallèle. Elle possède les particularités suivantes :

- Les classifieurs utilisent les mêmes données d'apprentissage et de test en entrée.
- Les classifieurs opèrent en même temps et indépendamment les uns des autres.
- Aucune interaction n'existe entre classifieurs.
- Les classifieurs sont fixes et ne changent pas.

La combinaison proposée comprend (**figure III.2**) les sous-systèmes suivants :

- La base d'entrée des formes inconnues à identifiées.
- Le classifieur KNN.

- Le classifieur ADI.
- Le classifieur ANN.
- Module de combinaison qui donne la décision d'identification finale.

La base d'entrée : on prend la base des paramètres statistiques arrangés dans les matrices d'apprentissage et de test, déduite à partir de la première partie de ce chapitre.

Les classifieurs : les trois classifieurs sont configurés par leurs paramètres optimaux démontrés précédemment.

Module de combinaison : ce module nécessite la définition de la fonction de combinaison qui joue un rôle très important dans la conception d'un SMC. Selon la théorie mentionnée dans le deuxième chapitre, puisque l'approche considérée est une approche parallèle avec apprentissage et les sorties des classifieurs sont de type 'classe', on peut utiliser la méthode de Bayes pour combiner les sorties des classifieurs et donner une décision finale.

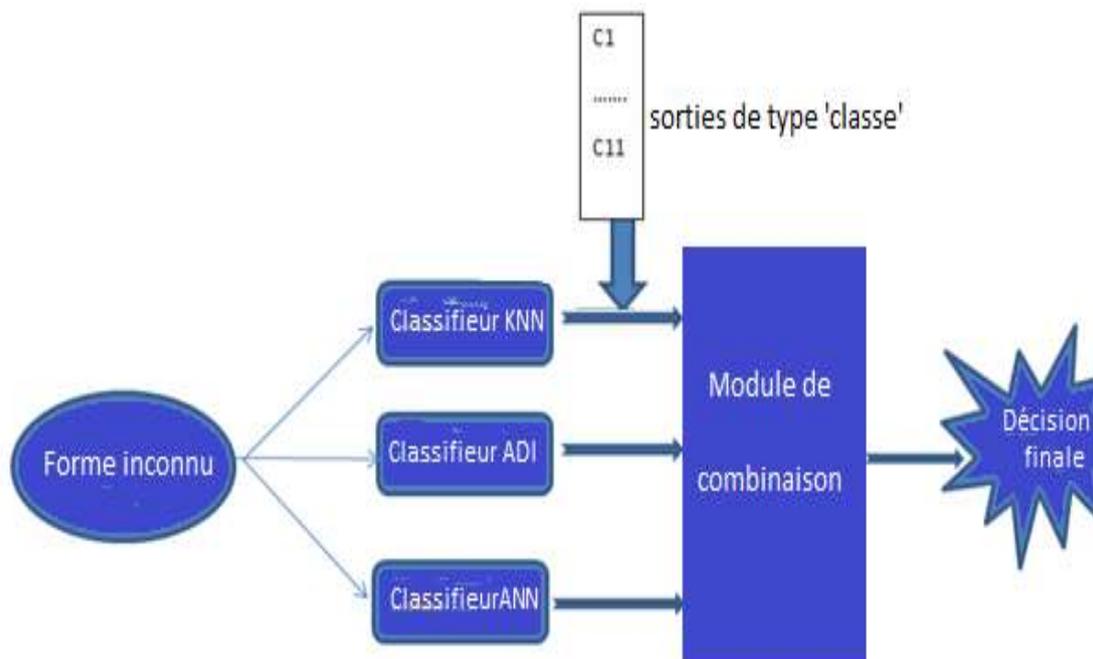


Figure III-2 : Architecture de la combinaison proposée

III.5.2 Fonction de combinaison ou 'méthode de Bayes'

La méthode Bayésienne repose sur la recherche de la classe w_i telle que la probabilité que l'entrée 'x' appartienne réellement à w_i connaissant la sortie de chaque classifieur, soit la plus forte. Par conséquent, on doit associer aux vecteurs de classes de sortie de chaque classifieur les vecteurs des probabilités d'appartenances de ces classes.

Pour chaque classifieur 'J', la probabilité $P(w_i/x)$ est calculée pour chaque identité (ou classe) w_i qui apparaît dans la liste générée par le classifieur [13] :

$$P\left(W_i/x\right) = \frac{1/d_i}{\sum \frac{1}{d_i}}; i = 1 \dots 11$$

Où d_i est la distance associée à la i ème identité dans la liste fournie par le classifieur.

Cette méthode de combinaison est programmée selon l'algorithme d'apprentissage et de fonctionnement suivant :

Début d'algorithme

/*Résultats de la classification parallèle */

1. Former un tableau $E_{(N \times L)}$ avec des sorties individuelles des 3 classifieurs. N est le Nombre des entrées de test. L'entité $e(i,j)$ est la classe attribuée par le classificateur D_j à l'entrée de test i . Un tableau $T(N \times 1)$ des correctes classes (ou les classes références) est également fournie.
2. Trouvez le nombre des entrées de test dans chaque classe dans T . Désignez ces nombres par N_1, N_2, \dots, N_c . ($c = 11$ classes possibles).
3. Pour chaque classificateur $D_i, i = 1, \dots, 3$, calculez une matrice de confusion C_i de dimension de 11×11 . La (j_1, j_2) ème composante est : $C_i(j_1, j_2) = \frac{K(j_1, j_2) + \frac{1}{c}}{N_{j_1} + 1}$
où $K(j_1, j_2)$ est le nombre des entrées de test avec la vrai classe j_1 , classé par le classificateur D_i en classe j_2 .

/*Combinaison : Pour chaque nouvelle entrée de test*/

1. Trouvez les classes s_1, \dots, s_3 attribuées à cette entrée par les 3 classifieurs de base.
2. Pour chaque classe $\omega_k, k = 1, \dots, 11$
 - (a) Calculer la probabilité : $P(k) = N_k/N$.
 - (b) Pour $i = 1 \dots 3$, calculez $P(k) = P(k) \times C_i(k, s_i)$.
3. Attribuez la classe k^* à l'entrée de test, où : $k^* = \arg \max_{k=1}^{11} P(k)$
4. Maintenir la classe de nouvelle entrée.

Fin.

III.5.3 Résultats de combinaison

Les tableaux III.10 et III.11 suivants décrits les résultats de la combinaison par la théorie bayésienne. Les taux de classification (des vecteurs et des signaux) de chaque classifieur sont mentionnés puis le taux résultant de la combinaison est donné.

Classifieur Nombre de Vecteurs de test	KNN	ADI	ANN	Taux TVC de COMBINAISON
5	95.07	73.84	82.51	94.89
10	94.84	75.89	81.05	94.45
20	94.71	71.98	85.83	94.68
50	94.78	69.77	80.90	94.58
100	94.42	70.09	75.92	94.45

Tableau III.10 : Taux de classification des vecteurs après combinaison des classifieurs en fonction des paramètres statistiques

Classifieur Nombre de Vecteurs de test	KNN	ADI	ANN	Taux TSC de COMBINAISON
5	94.5996	55.8659	75.7914	94.4134
10	95.5307	62.1974	77.0950	95.1583
20	95.7169	64.8045	74.8603	94.7858
50	96.0894	66.8529	86.9646	96.0894
100	95.3445	70.0186	90.5028	95.1583

Tableau III.11 : Taux de classification des signaux après combinaison des classifieurs en fonction des paramètres statistiques

Pour chaque vecteur de test, on peut remarquer que la valeur de taux de la combinaison est acceptable qui varie entre 94.45% et 94.89%. Aussi, cette valeur est supérieure à la moyenne des taux de l'ensemble des classifieurs.

Sur la totalité des vecteurs de test appliqués, nous avons constaté que : les résultats obtenus pour l'identification ont été satisfaisants en terme de taux de classification.

III.5.4 Conclusion

Nous considérons que notre objectif a été atteint puisque un système multi classifieurs est implémenté et expérimenté, ainsi, une remarquable amélioration de taux d'identification est obtenue.

Conclusion générale et Perspectives

Dans notre projet, on a conçu des systèmes d'identification des appareils électriques basée sur le classificateur KNN, puis le classificateur ADI, puis le classificateur ANN. Des séquences de vecteurs des coefficients de la série de fourrier puis des séquences de vecteurs de paramètres statistiques du courant électriques sont appliquées aux entrées de nos systèmes. Les performances du système sont évaluées en termes du taux de classification, sous la base de données PLAID.

A travers les différentes expériences, on a précisé les configurations optimales des trois classifieurs et on a spécifié les paramètres les plus pertinents à utiliser pour obtenir des taux de classification satisfaisant avec un temps d'exécution acceptable et une taille de mémoire limitée. Les taux maximum sont obtenus avec le classificateur KNN avec des valeurs égal à 97.59% et 95.07%. La considération des deux autres classificateurs été nécessaire pour la conception d'un système multi classifieurs.

On a arrivé à combiner les classifieurs en considérant l'approche parallèle comme une structure de réalisation du système multi classifieurs et en utilisant la méthode Bayésienne comme une fonction de liaison entre les sorties des classificateurs. Les résultats obtenus en termes des taux de classification des vecteurs et des signaux prennent des valeurs maximum de 94.89% et 95.15% respectivement. Nous considérons que notre objectif a été atteint avec un taux d'identification acceptable.

Cette contribution peut être complétée par l'utilisation:

- D'une méthode de sélection des classifieurs les plus valable pour afin d'aboutir à des performances optimales.
- D'autres architectures de combinaison.
- Des autres fonctions de combinaison.

BIBLIOGRAPHE

- [1] : Hala Najmeddine, Méthode d'identification et de classification de la consommation d'énergie par usages en vue de l'intégration dans un compteur d'énergie électrique Electromagnétisme'. Université Blaise Pascal - Clermont-Ferrand II, 2009. Français. ffNNT : **2009**, CLF21993ff. fftel-00724570f
- [2] : Mohamed NAIT MEZIANE, 'Identification d'appareils électriques par analyse des courants de mise en marche', THÈSE de Doctorat de l'université d'Orléans ,**2016**.
- [3] : Regaigui Abir Bensbaa Nour El Houda; 'Identification des appareils électriques basée sur le classificateur KNN combiné avec la règle de vote' ; Mémoire de Master, Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj **2019**
- [4] : Mohamed Nait-Meziane , Abdenour Hacine-Gharbi , Philippe Ravier , Guy Lamarque , Jean-Charles Le Bunetel and Yves Raingeaud , 'HMM-based Transient and Steady-state Current Signals Modeling for Electrical Appliances Identification', **2016** - International Conference on Pattern Recognition Applications and Methods.
- [5] : Fateh GHAZALI , Abdenour HACINE-GHARBI , Philippe RAVIER; Tayeb MOHAMADI, 'Extraction and selection of statistical harmonics features for electrical appliances identification using k-NN classifier combined with voting rules method', Turkish Journal of Electrical Engineering & Computer Sciences (2019).
- [6]: MEDJDOUB Housseem eddine, GAAD AlaEddine; 'L'analyse harmonique du courant électrique pour l'identification des appareils domestiques par l'approche SVM'; Mémoire de Master, Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj, **2019**.
- [7]: Thomas Picon, Mohamed Nait Meziane, Philippe Ravier, Guy Lamarque, Clarisse Novello, Jean-Charles Le Bunetel, Yves Raingeaud ; 'COOLL: Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification'.
- [8] HALICHE AMEL ; classification supervisée à base de KNN avec pondération d'attributs par l'algorithme génétique ; mémoire magister université de sciences et de technologie houari Boumediene ALGER **2015**.
- [9] Izabela Moise, Evangelos Pournaras, Dirk Helbing ; K-Nearest Neighbour Classifier ;DGESS

- [10] Mme BABA HAMED Amel née BENSEMAIN ; Vers Un Modèle De Classification Neuronale Des Données Médicales A Base De La Technologie FPGA ; Thèse de DOCTORAT ; Université Aboubakr Belkaïd – Tlemcen –**2017**
- [11] Guer mou di Mohammed el Amine. • Fekih Mohammed el Amine ; Fusion des classifieurs supervisés: Application sur la classification pixellaire des images microscopiques ; Université Abou Bakr Belkaid– Tlemcen **2013**
- [12] Faïcel Chamroukhi Analyse discriminante par Classification supervisée ; Université de Caen Statistics & Data Science Lab of Mathematics Nicolas Oresme
- [13] : Amirouche Radia, ‘Une combinaison de classifieurs pour la reconnaissance des visages humains’, mémoire de magister, Université de Annaba, **2006**.
- [14] : Gasmi.I, Merouani.H et Souici-Meslati.L, ‘Combinaison de classifieurs’, 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, March 27-31, **2005** – TUNISIA.
- [15] : YANNICK OUFELLA, ‘Evolution du concept de front roc et combinaison de classifieur’, mémoire de MASTER GÉNIE INFORMATIQUE, Université de Rouen, France, **2008**.
- [16] : Leila CHERGUI, ‘Combinaison de classifieurs pour la reconnaissance de mots arabes manuscrits’, Thèse de doctorat en science, Université de constantine, **2013**
- [17] : H. Zouari et all, ‘Un panorama des méthodes de combinaison de classifieurs en reconnaissance de formes’, in Proc ; RFIA’2002, Angers, France, vol. 2, pp 499-508, **2002**.
- [18]: Abdenour Hacine-Gharbi, Philippe Ravier and Mohamed Nait-Meziane, ‘Relevant harmonics selection based on mutual information for electrical appliances identification’, Int. J. Computer Applications in Technology, Vol. 62, No. 2, **2020**.