

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET  
POPULAIRE**

**MINISTRE 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

- **Hidoussi Karim**
- **Bekri Mehdi Abdelouahab**

Intitulé

**Implémentation d'un système multi classificateur basé sur la méthode des  
valeurs moyennes**

*Soutenu le : .....*

*Devant le Jury composé de :*

<i>Nom &amp; Prénom</i>	<i>Grade</i>	<i>Qualité</i>	<i>Etablissement</i>
<i>M. S. Mezzache</i>	<i>MCB</i>	<i>Président</i>	<i>Univ-BBA</i>
<i>M. S. Azzoug</i>	<i>MCB</i>	<i>Examineur</i>	<i>Univ-BBA</i>
<i>Mme. A. Laouamri</i>	<i>MAA</i>	<i>Encadreur</i>	<i>Univ-BBA</i>

*Année Universitaire 2022/2023*

## Résumé

Au cours des dernières années, la demande croissante d'énergie électrique a poussé les fournisseurs d'électricité à trouver des solutions pratiques pour une gestion efficace de la consommation d'énergie. Dans ce contexte, nous avons proposé un système d'identification des appareils électriques qui fournit des informations sur les appareils. Ce système est conçu en utilisant une méthode de classification basée sur des classifieurs tels que KNN, ANN, GMM, ARD et ADI.

Notre projet est basé sur deux systèmes d'identification d'appareils électriques : le premier repose sur les paramètres statistiques du courant électrique et le deuxième utilise une combinaison de classifieurs. Nous avons évalué les performances du système en termes de taux de classification et de combinaison des signaux de courant électrique en utilisant la base de données PLAID.

## Abstract

In recent years, the growing demand for electrical energy has prompted electricity suppliers to find practical solutions for efficient management of energy consumption. In this context, we have proposed an electrical appliance identification system that provides information on appliances. This system is designed using a classification method based on classifiers such as KNN, ANN, GMM, ARD and ADI.

Our project is based on two electrical device identification systems: the first is based on the statistical parameters of electrical current, and the second uses a combination of classifiers. We evaluated the system's performance in terms of classification rate and combination of electric current signals, using the PLAID database.

## ملخص

في السنوات الأخيرة، دفع الطلب المتزايد على الطاقة الكهربائية مزودي الكهرباء إلى إيجاد حلول عملية للإدارة الفعالة لاستهلاك الطاقة. في هذا السياق، اقترحنا نظامًا لتعريف الأجهزة الكهربائية يوفر معلومات عن الجهاز. تم تصميم هذا النظام باستخدام طريقة تصنيف تعتمد على المصنفات مثل KNN و ANN و GMM و ARD و ADI.

يعتمد مشروعنا على نظامين لتعريف الأجهزة الكهربائية: الأول يعتمد على المعلومات الإحصائية للتيار الكهربائي، والثاني يستخدم مجموعة من المصنفات. قمنا بتقييم أداء النظام من حيث معدل التصنيف ومجموعة إشارات التيار الكهربائي باستخدام قاعدة بيانات PLAID.

# REMERCIEMENTS

*Tout d'abord, nous remercions Dieu « الله », notre créateur, le tout puissant, de nous avoir donné la force, la volonté et le courage pour avoir accomplir ce modeste travail.*

*Nous voulons exprimez à notre directrice de projet notre gratitude, **Mme A. LAOUAMRI** pour son soutien constant, sa patience et ses précieux conseils tout au long de ce projet, son expertise et sa expérience sont indispensables pour les hommes en même temps.*

*Nous remercions également nos enseignants de l'université, pour leur soutien et leurs encouragements. Leurs expertises ont été une source d'inspiration tout au long de nos études.*

*Enfin, nous tenons à remercier tous ceux qui ont participé à ce projet, que ce soit en fournissant des données ou en offrant leur aide d'une autre manière.*

*Leurs contribution est une contribution essentielle à ce projet.*

*Nous remercions toutes ces personnes pour leur soutien et leur contribution à ce travail. Merci beaucoup.*

# *DEDICACES*

*Je dédie ce travail à mes très chers parents qui ont été une source de soutien inébranlable tout au long de mon parcours universitaire. Leurs encouragements et leurs amours inconditionnels m'ont permis de réaliser ce projet.*

*A Mes Frères et mes Sœurs 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*

*Enfin, je dédie ce travail à tous mes professeurs et mentors qui ont su partager leur savoir et leur expérience avec moi. Vos enseignements m'ont inspiré et guidé vers ce moment où je peux enfin voir le fruit de mes efforts.*

***[KARIM]***

*Je dédie ce projet à mes parents, pour l'amour constant qu'ils m'ont toujours donné, leurs encouragements et toute l'aide précieuse qu'ils m'ont apportée durant mes études. Aucun mot, aucune dédicace ne suffirait à exprimer pleinement le respect, la considération et l'amour que j'ai pour eux, ainsi que pour les nombreux sacrifices consentis pour mon instruction et mon bien-être.*

*Je tiens également à remercier mon encadreur, M. Asma Laouamri, pour son soutien et son encadrement tout au long de ce projet.*

*Je souhaite exprimer ma gratitude à tous les membres de ma famille, qu'ils soient petits ou grands, pour leur soutien constant et leurs encouragements qui ont toujours été présents à mes côtés.*

*À mes sœurs bien-aimées, je dédie ce travail en reconnaissance de tous les sacrifices qu'elles n'ont cessé de faire pour moi tout au long de mes années d'études.*

*Enfin, je tiens à remercier chaleureusement tous mes chers amis et collègues pour leur soutien, leurs conseils et leur présence encourageante tout au long de ce projet.*

*Votre appui et votre présence ont été essentiels à mon parcours académique, et je suis profondément reconnaissant envers vous tous.*

***[MEHDI.ABDELOUAHAB]***

## ***Sommaire***

Introduction Générale .....	1
<b>Chapitre I : « Théorie de la classification et de combinaison des classifieurs » .....</b>	
I.1 Introduction .....	3
I.2 Classification .....	3
I.2.1 Classification supervisé .....	3
I.2.2 Classification non supervisé .....	4
I.3 Classificateur .....	4
I.3.1 Classification par Arbre de décision .....	4
A. Définition .....	4
A.1 Arbre .....	4
A.2 Arbre de décision .....	5
B. Principe de construction .....	5
C. Algorithmes de construction .....	6
I.3.2 Classification par les Modèles de Mélange Gaussiennes (GMM) .....	7
A. Définition .....	7
A.1 Modèle de mélange .....	7
A.2 Distribution Gaussienne .....	7
A.3 Modèle de mélange de gaussienne GMM .....	8
B. Approche GMM pour la classification .....	8
B.1 Vraisemblance des paramètres .....	8
B.2 Estimation .....	9
B.3 Classification .....	9
B.4 Maximisation .....	9
C. Avantages .....	9
I.3.3 Classification KNN .....	9
A. Description de l'algorithme KNN .....	9
B. Configuration de classification KNN .....	10
C. Distances utilisées .....	10
C.1 Distance Euclidienne .....	10
C.2 Distance de City block .....	11

C.3 Distance Cosinus .....	11
C.4 Distance de corrélation .....	11
I.3.4 Classification ANN .....	11
A. Définition .....	11
B. Algorithme de réseau de Neurones Artificiels .....	12
C. Propriétés .....	13
I.3.5 Classification discrimination .....	13
A. Définition .....	13
B. Analyse Discriminante Linéaire LDA .....	14
C. Fonctionnement de LDA .....	15
I.3.6 Taux de classification .....	15
I.4 Approches de combinaison des classifieurs .....	16
I.4.1 Contexte et motivations .....	16
I.4.2 Types de combinaison .....	17
A. Combinaison séquentielle .....	17
B. Combinaison parallèle .....	18
C. Combinaison hybride .....	18
I.4.3 Méthode de combinaison moyenne .....	19
I.5 Conclusion .....	19
<b>Chapitre II : «Architecture du système d'identification des appareils électriques à implémenter»</b> .....	
II.1 Introduction .....	21
II.2 Architecture du système de classification proposé .....	21
II.2.1 Base de données PLAID (plug load appliance identification dataset) .....	22
II.2.2 Phase d'apprentissage .....	23
II.2.3 Phase de test .....	24
II.2.4 Extraction des caractéristiques .....	25
II.2.5 Unité de classification .....	26
II.2.6 Performances .....	26
II.3 Structure du système multiclassifieurs proposé .....	27
II.3.1 Base d'entrée .....	28

II.3.2 Les classifieurs .....	28
II.3.3 Module de combinaison .....	29
II.3.4 Décision finale .....	29
II.4 Conclusion .....	29
<b>Chapitre III : Implémentation et résultats</b> .....	
III.1 Préliminaire .....	31
III.2 Implémentation du système de classification des appareils électriques .....	31
III.3 Résultats de classification et discussions .....	34
III.3.1 Classification par l'arbre de décision .....	34
III.3.2 Classification avec le classifieur KNN .....	35
III.3.3 Identification avec le classifieur ADI .....	36
III.3.4 Identification avec le classifieur ANN .....	38
III.3.5 Identification avec le classifieur GMM .....	39
III.4 Implémentation du système multi-classificateur (SMC) .....	40
III.5 Résultats de combinaison .....	43
III.6 Conclusion .....	46
Conclusion générale .....	47
Référence .....	48



## *Liste des figures*

<b>Figure I.1</b> : Structure d'arbre de décision .....	5
<b>Figure I.2</b> : Courbe de distribution Gaussienne .....	8
<b>Figure I.3</b> : Fonctionnement de la classification ANN .....	12
<b>Figure I.4</b> : Schéma de combinaison séquentielle de classifieurs .....	17
<b>Figure I.5</b> : Schéma de combinaison parallèle de classifieurs .....	18
<b>Figure I.6</b> : Schéma de combinaison hybride de classifieurs .....	19
<b>Figure II.1</b> : Structure du système de classification proposé .....	22
<b>Figure II.2</b> : Schéma synoptique de système de combinaison proposé .....	27
<b>Figure III.1</b> : Fenêtre des résultats de classification ARD avec un nombre de vecteurs = 150...	35
<b>Figure III.2</b> : Fenêtre des résultats de classification KNN avec un nombre de vecteurs = 15....	36
<b>Figure III.3</b> : Fenêtre des résultats de classification ADI avec un nombre de vecteurs = 10.....	38
<b>Figure III.4</b> : Fenêtre des résultats de classification ANN avec un nombre de vecteurs = 50.....	39
<b>Figure III.5</b> : Fenêtre des résultats de classification GMM avec un nombre de vecteurs = 40....	40
<b>Figure III.6</b> : L'architecture du système multi-classificateur.....	41
<b>Figure III.7</b> : fenêtre des résultats de combinaison avec un nombre de vecteurs = 50.....	44
<b>Figure III.8</b> : fenêtre des résultats de combinaison après élimination de ADI avec un nombre de vecteurs = 25 .....	45
<b>Figure III.9</b> : fenêtre des résultats de combinaison après élimination de ADI et GMM avec un nombre de vecteurs = 25.....	46

## ***Liste des tableaux***

<b>Tableau.II.1</b> : Récapitulative des appareils trouvés sur la base de données PLAID....	23
<b>Tableau III-1</b> : Variation de taux de classification avec le classifieur TREECLASS...	34
<b>Tableau III-2</b> : Sélection de la configuration optimale de classificateur KNN.....	35
<b>Tableau III-3</b> : Variation de taux de classification avec le classifieur KNN.....	36
<b>Tableau III-4</b> : Sélection de la configuration optimale de classificateur ADI.....	37
<b>Tableau III-5</b> : Variation de taux de classification avec le classifieur ADI.....	37
<b>Tableau III-6</b> : Variation de taux de classification avec le classifieur ANN.....	39
<b>Tableau III-7</b> : Variation de taux de classification avec le classifieur GMM.....	40
<b>Tableau III-8</b> : Variation de taux de combinaison en fonction de nombre de vecteurs.	43
<b>Tableau III-9</b> : Variation de taux de combinaison en fonction de nombre de vecteurs après élimination de ADI.....	44
<b>Tableau III-10</b> : Variation de taux de combinaison en fonction de nombre de vecteurs après élimination de ADI et GMM.....	45

## *Introduction générale*

L'énergie électrique est l'une des formes d'énergie importantes qui sont utilisées dans divers domaines et qui sont indispensables dans notre vie quotidienne dans les usages domestiques tels que l'éclairage, le chauffage, le fonctionnement des appareils électroménagers.

Depuis la fin du siècle dernier, la croissance économique est confrontée à des évolutions opposées qui sont d'une part la nécessité de développer encore plus d'activité pour assurer la richesse des populations et d'autre part la nécessité de réduire au maximum les émissions

De nombreux travaux de recherche scientifique abordent cette question en proposant des solutions de mise en œuvre de systèmes de rationalisation de la consommation énergétique globale. 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.

Nous avons proposé un système de classification et un système multiclassifieurs basée sur les paramètres statistiques et les algorithmes de classifications (KNN, ANN, ARD, GMM, ADI). Les performances du système sont évaluées en termes du taux de classification des signaux du courant électrique en utilisant la base de données PLAID.

Ce travail est divisé en trois chapitres, qui sont :

Le premier chapitre présente la théorie de la classification, plus particulièrement, ce chapitre présente des généralités sur le concept de la classification et les différents classificateurs tel que KNN, ANN, GMM, ARD, ADI, après ça on décrit la combinaison des classificateurs et les approches de combinaison.

Le deuxième chapitre est consacré à la description de notre système de classification et le système multiclassifieurs en présentant les différentes étapes de conception des systèmes proposés. Un système d'identification est principalement basé sur trois choses : une base de données, un algorithme d'extraction de paramètres (caractéristiques) et un classificateur. Nous avons choisi la base PLAID pour laquelle nous donnons une brève description.

Le troisième chapitre sera consacré à l'implémentation de notre système et discuter les résultats obtenus en réalisant différentes expériences.

# **CHAPITRE I :**

## **Théorie de la classification et de combinaison des classifieurs**

## **I.1 Introduction**

L'homme représente le plus parfait système de reconnaissance de forme et de classification des objets. Il peut facilement distinguer le bruit d'une voiture de celui du train, le visage d'une femme de celui d'un enfant etc. Aujourd'hui, les progrès scientifiques et techniques permettent d'essayer d'imiter certaines de ces reconnaissances à l'aide des machines.

Dans ce contexte, une théorie de la classification et des classificateurs sera présentée à travers ce chapitre. Des approches de combinaison des classificateurs seront également expliquées.

## **I.2 Classification**

La classification est une opération de structuration qui vise à organiser un ensemble d'observation en groupes homogènes et contrastés afin de faciliter l'analyse des informations et d'effectuer des prédictions [1]. La classification devra déterminer, parmi un ensemble fini de classes, à laquelle appartient un objet donné. Un classifieur doit être capable de modéliser au mieux les frontières qui séparent les classes les unes des autres [2].

La plupart des algorithmes de classification tentent de trouver un modèle (une fonction mathématique) qui explique le lien entre les données d'entrée et les classes de sortie. Un ensemble d'apprentissage (ensemble de classes) est donc utilisé par l'algorithme. Cette méthode de raisonnement est appelée inductive car on induit la connaissance (le modèle) à partir des données d'entrée (objets à classifier) et des sorties (leurs classes). Grâce à ce modèle, on peut alors prédire les classes de nouvelles données. Le modèle est bon s'il permet de bien prédire [3].

Il existe deux types de classification, la classification supervisée et la classification non supervisée. Ces deux approches se différencient par leurs méthodes et leurs buts.

### **I.2.1 Classification supervisée**

La classification supervisée est l'une des techniques les plus utilisées dans l'analyse des bases de données. Elle permet d'apprendre des modèles de décision qui permettent de prédire le comportement des exemples futures.

La classification supervisée consiste à inférer à partir d'un échantillon d'exemples classés une procédure de classification. Un système d'apprentissage effectue la recherche

d'une telle procédure selon un modèle. Les systèmes d'apprentissage peuvent être basés sur des hypothèses probabilistes (classifieur naïf de Bayes, méthodes paramétriques) ; sur des notions de proximité (plus proches voisins) ; sur des recherches dans des espaces d'hypothèses (arbres de décision, réseaux de neurones) [4].

### **I.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 [5].

## **I.3 Classificateur**

Un classificateur 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 classificateurs 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 classificateurs sont disponible pour l'expérimentation [5].

### **I.3.1 Classification par Arbre de décision**

#### **A. Définitions**

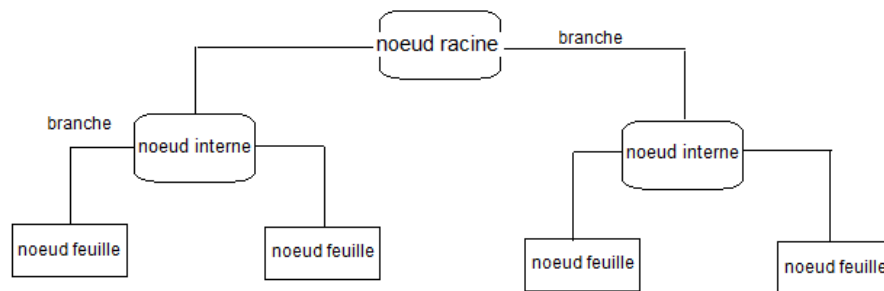
##### **A.1 Arbre**

En théorie des graphes, un **arbre** est un graphe non orienté, acyclique et connexe. L'ensemble des nœuds se divise en trois catégories :

- Nœud racine (l'accès à l'arbre se fait par ce nœud),
- Nœuds internes : les nœuds qui ont des descendants (ou enfants), qui sont à leur tour des nœuds.
- Nœuds terminaux (ou feuilles) : nœuds qui n'ont pas de descendant.

## A.2 Arbre de décision

Un arbre de décisions est un algorithme d'apprentissage supervisé non paramétrique, qui est utilisé pour les tâches de classification. Il y-a une structure hiérarchique et arborescente (**figure I.1**), qui se compose d'un nœud racine, de branches, de nœuds internes et de nœuds feuille, chaque nœud possède une condition qui amène à plusieurs réponses, ce qui dirige à un prochain nœud. Les tests s'effectuent dans les nœuds internes et les décisions sont prise dans les nœuds feuille.



**Figure I.1** : Structure d'arbre de décision.

## B. Principe de construction

- ✓ **Etape1** : définir la problématique en demandant quelle est la décision à prendre et en formulant la par une question. C'est le nœud racine (ou variable cible).
- ✓ **Etape2** : partitionner la variable cible (nœud racine) en des variables explicatives (nœuds internes) qui donnent plusieurs réponses à la question racine.
- ✓ **Etape3** : partitionner chaque nœud interne en des sous-nœuds qui expliquent en plus la variable cible et calculer la valeur d'un attribut qui est testé à chaque itération.
- ✓ **Etape4** : Le processus est répété plusieurs fois pour chaque sous-nœud jusqu'à ce que le processus de séparation s'arrête, si on obtient la meilleure valeur de l'attribut, c-à-d la décision est homogène à la variable cible. Ces derniers nœuds sont les nœuds feuilles.
- ✓ **Etape5** : La décision finale est sélectionnée selon le critère de meilleur attribut qui peut être calculé par les paramètres : **entropie** et **gain d'information**.
  - **Entropie** : L'entropie est un concept issu de la théorie d'informations, qui mesure l'écart de la distribution de la variable cible par rapport à la distribution uniforme. Pour un nœud  $S$  (interne ou feuille), on calcule l'entropie par rapport à la cible comme suit :

- Partitionner  $S$  sur les valeurs de la cible en  $m$  groupes :  $C_1, \dots, C_m$ .
- Calculer  $p_i$ ,  $i=1\dots m$ , la probabilité qu'un élément de  $S$  se retrouve dans  $C_i$ , tel que :  $p_i \approx \frac{|C_i|}{|S|}$  ou  $|C_i|$  est la taille du groupe  $C_i$ .

$$\text{entropie}(S) = - \sum_{i=1}^m p_i \log p_i \quad (\text{I.1})$$

Les valeurs d'entropie peuvent être comprises entre 0 et 1. Afin de sélectionner la meilleure fonction pour effectuer le fractionnement et trouver l'arbre de décisions optimal, il convient d'utiliser l'attribut avec le plus petit total d'entropie.

- **Le gain d'informations** : il représente la différence en entropie avant et après un fractionnement sur un attribut donné. L'attribut avec le gain d'informations le plus élevé produira le meilleur fractionnement car il fait mieux le travail de classement des données d'apprentissage selon sa classification cible. Le gain d'informations est donné par la formule suivante :

$$GI(S) = \text{entropie}(S) - \sum_{i=1}^m p_i \text{entropie}(S_i) \quad (\text{I.2})$$

### C. Algorithmes de construction

Plusieurs algorithmes automatiques sont utilisés pour construire les arbres de décision.

- **L'algorithme ID3** (Iterative Dichotomiser 3) publié en 1986, ne permet que de créer des arbres de classification mais ne construit pas seulement des arbres binaires (un nœud peut avoir plus de deux enfants).
- **L'algorithme C4.5** est basé sur ID3 mais ajoute quelques améliorations, notamment en se basant sur la mesure de l'entropie pour construire l'arbre de décision.
- **L'algorithme CART** dont l'acronyme signifie « Classification And Regression Trees » a pour rôle de construire des arbres binaires de classification et de régression. L'algorithme CART est très avantageux par rapport aux autres algorithmes, il est non seulement capable de gérer les arbres de régression mais aussi les arbres de classification en utilisant des métriques spécifiques.



### I.3.2 Classification par les Modèles de Mélange Gaussiennes (GMM)

#### A. Définitions

##### A.1 Modèle de mélange

Le modèle de mélange fini de lois de probabilité consiste à supposer que les données proviennent d'une source contenant plusieurs sous-populations homogènes appelées composants. La population totale est un mélange de ces sous-populations. [6]

Les modèles de mélange sont utilisés pour estimer les paramètres d'une distribution de variables aléatoires en les modélisant par une densité mélange. Usuellement, ils sont employés en classification et on considère que chaque composant du mélange caractérise une classe.

Soit  $X = (X_1, \dots, X_n)$  un échantillon de variables aléatoires indépendantes identiquement distribués de loi mélange fini à  $K$  composants, de densité  $f$  dont la forme générale est :

$$f(x) = \sum_{k=1}^K \Pi_k f_k(x) \quad (I.3)$$

Avec :  $\Pi_k$  les proportions respectives des sous populations telle que :  $0 < \Pi_k \leq 1$  et  $\sum_{k=1}^K \Pi_k = 1$

$f_k$  la densité du  $k^{\text{ième}}$  composant.

Le mélange (I.3) peut être vu comme la loi marginale de la variable  $X$  pour le couple  $(X, K)$ .

##### A.2 Distribution Gaussienne

En probabilité, on dit qu'une variable aléatoire réelle  $X$  suit une **loi normale** (ou **loi gaussienne**) d'espérance (ou moyenne)  $\mu$  et d'écart type  $\sigma$  strictement positif (de variance  $\sigma^2$ ), si cette variable aléatoire réelle  $X$  admet pour densité de probabilité la fonction  $p(x)$  définie, pour tout nombre réel  $x$ , par :

$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (I.4)$$

Une telle variable aléatoire est alors dite variable gaussienne. Sa densité de probabilité dessine une courbe dite **courbe en cloche** ou courbe de Gauss. La figure I.2 représente des distributions gaussiennes en fonction de  $\mu$  et  $\sigma^2$ .

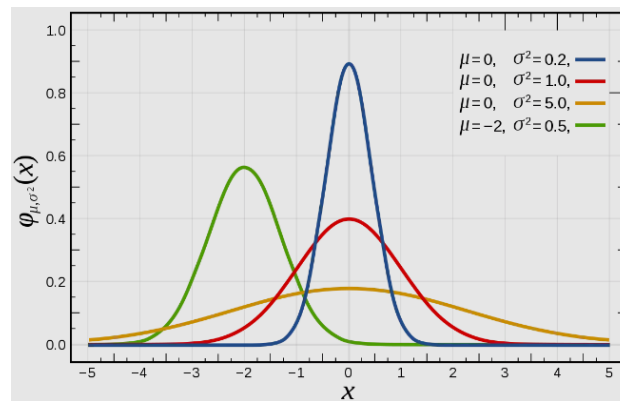


Figure I.2 : Courbe de distribution Gaussienne.

### A.3 Modèle de mélange de gaussienne GMM

Le modèle de mélange gaussien est une combinaison linéaire de plusieurs composantes gaussiennes. Il est particulièrement utilisé si la structure de données est composée de plusieurs groupes, il faut les représenter par un modèle de mélange gaussien plutôt qu'une simple distribution gaussienne [6]. Il s'agit alors de déterminer la variance, la moyenne et l'amplitude de chaque gaussienne. Ces paramètres sont optimisés selon un critère de maximum de **Vraisemblance** pour approcher le plus possible la distribution recherchée. Cette procédure se fait le plus souvent itérativement via l'algorithme **EM** (Estimation-Maximisation). [7]

### B. Approche GMM pour la classification

On considère comme échantillon un tableau  $X = (X_1, \dots, X_n)$  de  $n$  variables aléatoires avec la densité  $\varphi(x; \theta)$  définie par :

$$\varphi(x_i; \theta) = \sum_{k=1}^s p_k \cdot \varphi(x_i; \alpha_k) \quad (I.5)$$

Tel que :

- $\varphi(x_i; \alpha_k)$ : Densité de probabilité de la classe  $k$ .
- $p_k$ : Probabilité qu'un élément de l'échantillon suive la loi  $\varphi_k$
- $\theta = (p_1 \dots p_s; \alpha_1 \dots \alpha_s)$ : Paramètre du modèle de mélange

Pour classifier le groupe de données  $X$ , le modèle GMM suit les approches suivantes :

#### B.1 Vraisemblance des paramètres

Estimation des proportions des composants (les probabilités  $p_k$  et les paramètres  $\alpha_k$ . Puis l'utilisation de log-vraisemblance classifiant suivante :

$$L_c(x_1, \dots, x_n; z, \theta) = \sum_{i=1}^n \sum_{k=1}^s z_{ik} \log(p_k \varphi(x_i; \alpha_k)) \quad (I.6)$$

### B.2 Estimation

Estimation des probabilités a posteriori :

$$t_{ik} = \frac{p_k \cdot \varphi_k(x_i; \alpha_k)}{\sum_{l=1}^s p_l \cdot \varphi_l(x_i; \alpha_l)} \quad (\text{I. 7})$$

### B.3 Classification

classification des individus avec la méthode du maximum a posteriori (MAP) :

$$z_k = \{i | t_{ik} = \max_{l=1, \dots, s} t_{il}\} \quad (\text{I. 8})$$

### B.4 Maximisation

calcul des paramètres du mélange :

$$p_k = \frac{\text{Card}(z_k)}{n} \quad (\text{I. 9})$$

Itérer les étapes **B.2** à **B.4**, jusqu'à la convergence (évolution très faible de  $L_c$ ).

## C. Avantages

Les modèles GMM présentent deux avantages principaux :

- Il s'agit d'une méthode probabiliste permettant d'obtenir **une classification** des observations. Une probabilité d'appartenance à chacune **des classes** est calculée et une classification est généralement obtenue en affectant chacune des observations à la classe la plus probable.
- Ils offrent une grande flexibilité de modélisation et permettent donc de modéliser un grand nombre de phénomènes.

### I.3.3 Classification KNN

#### A. Description de l'algorithme KNN

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.

Les différentes étapes de l'algorithme KNN sont données comme suit :

**Début Algorithme :**

Données (ou vecteurs) en entrée :

- Un ensemble de données (base de données).
- Une fonction pour la définition de la distance.
- Un nombre entier K.

Pour une nouvelle observation (vecteur) dont on veut prédire sa classe Faire :

1. Calculer toutes les distances de cette observation avec les autres observations du jeu de données (base de données d'apprentissage).
2. Retenir les observations du jeu de données les proches en utilisant la fonction de calcul de distance.
3. Prendre les valeurs des classes d'observations retenues.
4. Retourner la classe la plus dominante ou la plus votée.

**Fin Algorithme****B. Configuration de classification KNN**

La valeur de K et le choix de la distance sont les paramètres les plus importants de l'algorithme de classification KNN.

- Généralement, le choix de la valeur optimale de K consiste à chercher parmi un ensemble de valeurs K, celle qui permet d'obtenir un taux de classification maximal évalué sur une base de données de test.

- Le choix de la distance optimale consiste à suivre la même stratégie du choix de la valeur de K en appliquant différentes distances utilisées dans l'algorithme KNN.

**C. Distances utilisées**

Il existe différentes distances utilisées dans l'étape de comparaison de l'algorithme KNN.

Soient deux vecteurs X et Y de dimension d :  $X = (x_1, \dots, x_m)$  et  $Y = (y_1, \dots, y_m)$  .

**C.1 Distance Euclidienne**

Cette distance a été utilisée dans plusieurs systèmes d'identification basée sur l'algorithme KNN [papier Ghazali, all, 2019] [Nait Meziane, all, 2017]. La distance Euclidienne entre les deux vecteurs X et Y est donnée par :

$$d_E(X, Y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (\text{I. 10})$$

### C.2 Distance de City block

Cette distance est définie par la relation :

$$d_E(X, Y) = \sum_{i=1}^m (x_i - y_i) \quad (\text{I. 11})$$

### C.3 Distance Cosine

La distance cosinus, également appelée distance angulaire, est dérivée de la similitude cosinus qui mesure l'angle entre deux vecteurs [8]. Cette distance est définie comme suit :

$$d_{cos}(X, Y) = 1 - \frac{\sum_{i=1}^m X_i Y_i}{\sqrt{\sum_{i=1}^m X_i^2} \sqrt{\sum_{i=1}^m Y_i^2}} \quad (\text{I. 12})$$

### C.4 Distance corrélation

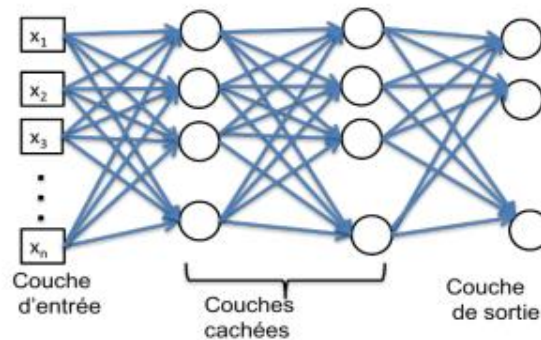
Cette distance est donnée par la formule suivante :

$$d_{cos}(X, Y) = 1 - \frac{\sum_{i=1}^m (X_i - \bar{Y}_i)}{\sqrt{\sum_{i=1}^m (X_i - \bar{Y}_i)^2}} \frac{(X_i - \bar{Y}_i)}{\sqrt{\sum_{i=1}^m (X_i - \bar{Y}_i)^2}} \quad (\text{I. 13}) \quad [3].$$

## I.3.4 Classification ANN (Artificial Neural Networks)

### A. Définition

Les réseaux de neurones artificiels constituent des systèmes de traitement distribués, composés d'unités de traitement interconnectées et adaptatives, caractérisées par une persistance inhérente à l'apprentissage de l'expérience et aussi découvrir de nouvelles connaissances. L'excellente capacité d'auto-apprentissage et l'auto-adaptation de ces systèmes d'apprentissage les a établis comme de puissants outils de reconnaissance de formes et en tant que composante essentielle de nombreux systèmes de classification. Ainsi, ils ont été largement étudiés et largement utilisés dans de nombreuses applications de l'intelligence artificielle. Dans la littérature, bien que de nombreux modèles différents aient été proposés.



**Figure I.3** : Fonctionnement de la classification ANN.

### B. Algorithme de réseau de Neurones Artificiels

L'algorithme ANN est un système adaptatif complexe qui peut modifier sa structure interne en fonction des informations y transitent. Il est obtenu en ajustant le poids de la connexion.

- Chaque connexion est associée à un poids.
- Un poids est un nombre qui contrôle le signal entre deux neurones.
- Les poids sont ajustés pour améliorer le résultat.

La procédure de classification commence par un jeu de données. L'ensemble de données est divisé en deux parties : échantillon d'apprentissage et échantillon de test.

Le premier échantillon est utilisé pour l'apprentissage du réseau tandis que l'échantillon de test est utilisé pour mesurer la précision du classificateur.

En général, les étapes d'apprentissage de réseau neuronal sont :

- La structure du réseau est définie avec un nombre fixe de nœuds en entrée, sortie et couche cachée.
- Un algorithme est utilisé pour le processus d'apprentissage par exemple : l'apprentissage par la règle de retro propagation qui se déroule selon les étapes suivantes [9]:

**Etape 1** : Initialiser les poids avec les valeurs au hasard.


**Etape 2** : Appliquer une donnée d'entraînement au réseau de neurones et calculer la sortie.

**Etape 3** : Calculer l'erreur de chacun des nœuds de sortie.

$$\text{Erreur au nœud } K \delta_K = (t_K - O_K)(1 - O_K)O_K \text{ sortie obtenue (I. 14)}$$

**Etape 4:** Calculer la correction des poids pour les nœuds de sorties

$$w_{jk} = w_{jk} + \eta \delta_k Z_j \text{ Valeur du nœud } j \text{ de la couche cachée (I. 15)}$$

  
 Facteur d'entraînement

**Etape 5 :** Propager l'erreur sur les poids des nœuds de la couche cachée.

L'erreur d'entrée du  $j^{\text{ième}}$  nœud de la couche cachée est  $\delta_{inj} = \sum_{K=1}^M \delta_K w_{jk}$  (I. 16)

L'erreur du  $j^{\text{ième}}$  nœud de la couche cachée est  $\delta_j = O_j(1 - O_j) \sum_{K=1}^M \delta_K w_{jk}$  (I. 17)

**Etape 6:** Calculer la correction des poids pour les nœuds de la couche cachée :

$$w_{ij} = w_{ij} + \eta \delta_j x_i \quad (\text{I. 18})$$

**Etape 7 :** Vérifier la convergence. Arrêter l'algorithme si une condition de convergence a été atteinte. Sinon, on reprend à l'étape 2

### C. Propriétés

cette méthode possède plusieurs propriétés fascinantes, on cite :

- **Le parallélisme :** Le parallélisme permet une rapidité de calcul supérieure.
- **La capacité d'adaptation :** C'est la capacité d'apprentissage qui permet au réseau de tenir compte de nouvelles contraintes ou de nouvelles données.
- **La capacité d'auto organisation :** cette propriété assure une stabilité autant que système dynamique.
- **Facilité de construction :** La simulation informatique d'un réseau de neurone pour une petite application est simple et ne nécessite qu'un temps de développement assez court.

### I.3.5 Classification discriminante

#### A. Définition

Une analyse discriminante est une technique statistique qui vise à décrire, expliquer et prédire l'appartenance à des groupes prédéfinis (classes) d'un ensemble d'observations (individus) à partir d'une série de variables prédictives (descripteurs).

La classification discriminante est la tâche qui consiste à discriminer des données, de façon supervisée, un ensemble d'objets, de telle manière que les objets d'un même groupe (appelé classes) sont plus proches (au sens d'un critère de (dis) similarité choisi) les uns à l'autres que celles des autres groupes [8].

**B. Analyse Discriminante Linéaire LDA**

L'analyse discriminante linéaire est un algorithme de référence en classification supervisée. Il peut être appréhendé de deux façons complémentaires :

- une approche **géométrique** qui revient à chercher des hyperplans qui séparent au mieux les groupes ;
- une approche **modèle** qui fait l'hypothèse que les lois des covariables sont des vecteurs gaussiens avec des valeurs de paramètres différentes pour chaque groupe.

On considère  $(x_1, y_1), \dots, (x_n, y_n)$  un échantillon où  $x_i$  est à valeurs dans  $\mathbb{R}^d$  et  $y_i$  dans  $\{0,1\}$ .

**L'approche géométrique** revient à chercher une droite de  $\mathbb{R}^d$  d'équation :

$$a_1x_1 + \dots + a_dx_d = 0 \text{ telle que : (I. 19)}$$

- les centres de gravité de chaque groupe projeté sur cette droite soit au mieux séparé  $\Rightarrow$  maximiser la distance inter-classe.
- les observations projetés soient proches de leur centre de gravité projeté  $\Rightarrow$  minimiser la distance intra-classe.

Le compromis entre ces deux distances s'obtient en maximisant le **coefficient de Rayleigh** qui est le quotient entre ces deux distances :

$$J(a) = \frac{B(a)}{W(a)} = \frac{a^t B a}{a^t W a} \quad (\text{I. 20})$$

où B et W sont les matrices inter et intra classes définies par :

$$B = \frac{1}{n} \sum_{k=1}^K n_k (g_k - g)(g_k - g)^t \text{ et } W = \frac{1}{n} \sum_{k=1}^K n_k V_k \text{ avec } V_k = \frac{1}{n_k} \sum_{i:Y_i=k} (X_i - g_k)(X_i - g_k)^t \quad (\text{I. 21})$$

Ici g désigne le centre de gravité du nuage  $x_i, i = 1, \dots, n$  et  $g_k, k = 0,1$  les centres de gravité des deux groupes. La solution est donnée par un vecteur propre associé à la plus grande valeur propre de  $W^{-1}B$ .

**L'approche modèle** fait l'hypothèse que les vecteurs  $X|Y = k, k = 0,1$  sont des vecteurs gaussiens d'espérance  $\mu_k \in \mathbb{R}^d$  et de matrice de variance covariance  $\Sigma$ . Ces paramètres sont estimés par maximum de vraisemblance et on déduit les probabilités a posteriori par la **formule de Bayes** :



$$P(Y = k|X = x) = \frac{\pi_k f_{X|Y=k}(x)}{f(x)} \quad (I.22)$$

### C. Fonctionnement de LDA

LDA se concentre principalement sur la projection des caractéristiques dans un espace de dimensions supérieures vers des dimensions inférieures. Vous pouvez y parvenir en trois étapes :

- Tout d'abord, vous devez calculer la séparabilité entre les classes qui est la distance entre la moyenne des différentes classes. C'est ce qu'on appelle la variance interclasses.

$$S_b = \sum_{i=1}^g N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (I.23)$$

- Deuxièmement, calculez la distance entre la moyenne et l'échantillon de chaque classe. On l'appelle aussi la variance intra-classe.

$$S_w = \sum_{i=1}^g (N_i - 1) S_i = \sum_{i=1}^g \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T \quad (I.24)$$

- Enfin, construisez l'espace de dimension inférieure qui maximise la variance interclasses et minimise la variance intra-classe. P est considéré comme la projection spatiale de dimension inférieure, également appelée critère de Fisher. [10]

$$S_w = \arg \max \frac{|P^T S_b P|}{|P^T S_w P|} \quad (I.25)$$

#### I.3.6. 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 :

$$TC = \frac{N_{SC}}{N_{ST}} * 100 \quad (I.26)$$

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 :

$$TCV = \frac{N_{VC}}{N_{VT}} * 100 \quad (I. 27)$$

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. [11]

## I.4 Approches de combinaison des classifieurs

### I.4.1 Contexte et motivations

Les chercheurs se rendent compte qu'il n'existe aucune méthode de classification pouvant manifester une supériorité sur les autres méthodes dans tous les problèmes et toutes les situations. Si aucune d'entre elles ne peut satisfaire entièrement les exigences d'une application envisagée, l'utilisation simultanée de plusieurs méthodes en même temps peut, éventuellement, permettre d'en cumuler les avantages sans en cumuler les inconvénients [1].

Plusieurs chercheurs ont fait les constats suivants :

- Il n'existe pas de « meilleur » classifieur capable de traiter (apprendre) n'importe quelle distribution des données d'apprentissage.
- Aucun classifieur ne peut discriminer suffisamment correctement un ensemble important de classes. [12]
- 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. Donc, deux raisons majeures sont derrière la combinaison [11]

**Précision** : une décision plus fiable peut être obtenue en combinant l'avis (les sorties) de plusieurs experts (classifieurs).

**Efficacité** : un problème complexe peut être décomposé en plusieurs sous-problèmes qui sont plus faciles à comprendre et à résoudre (diviser pour mieux régner...). [2]

Cette précision et cette efficacité proviennent essentiellement des faits suivants :

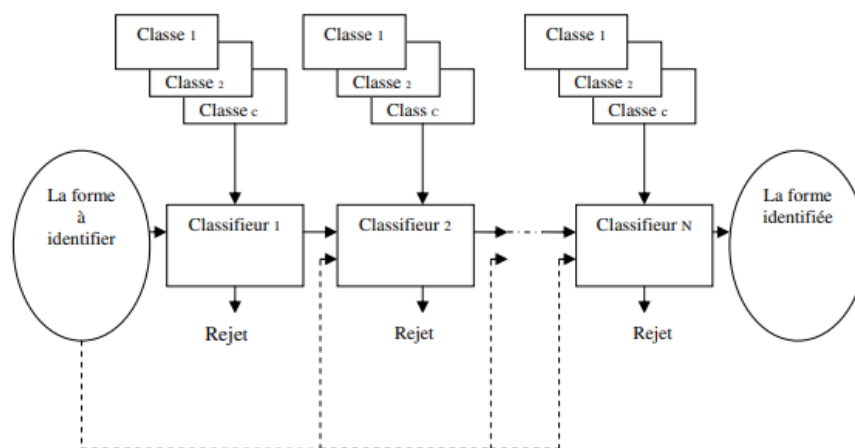
- Distribuer les caractéristiques sur des classifieurs adaptés
- Exploiter la complémentarité entre classifieurs
- Prendre en compte les performances de chacun des classifieurs
- Réduire l'importance des choix initiaux [12]

#### I.4.2 Types de combinaison

Différents schémas de combinaison de classifieurs ont été proposés dans la littérature selon l'ordre d'exécution des classifieurs où on observe trois types de combinaison : séquentielle, parallèle et série parallèle. [1]

##### A. Combinaison séquentielle

La combinaison séquentielle, appelée également combinaison série, consiste à positionner les classifieurs les uns à la suite des autres. Elle est donc organisée en niveaux successifs de décision qui permettent, soit de conforter la décision prise auparavant, soit de traiter les rejets ou les solutions jugées peu fiables obtenues aux niveaux précédents. Dans ce type de combinaison l'ordre d'exécution des classifieurs est primordial et influe sur le résultat final [11]. Cette méthode est surtout utilisée lorsqu'un classifieur fournit plusieurs solutions qu'on ne peut pas départager. Le classifieur intervenant en aval sert à lever les ambiguïtés. 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.



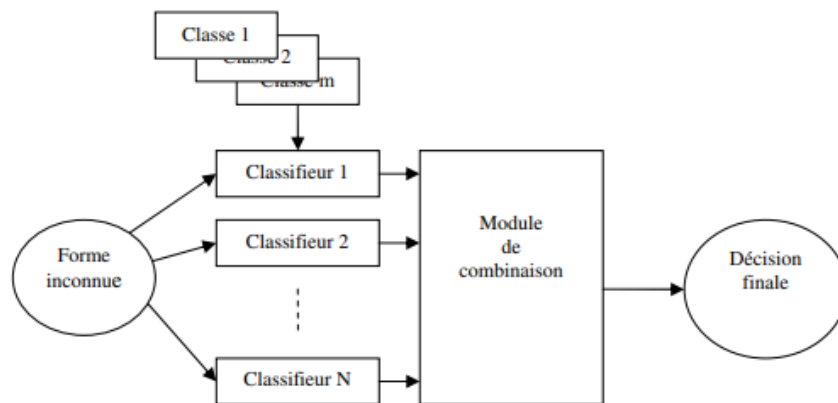
**Figure I.4** : Schéma de combinaison séquentielle de classifieurs.

### B. 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. Ce type de combinaison est très utilisé, parce que chaque classifieur qui satisfait le problème de la classification peut être utilisé normalement dans un système multiclassifieurs. [2]

Ce type de combinaison présente les avantages suivants :

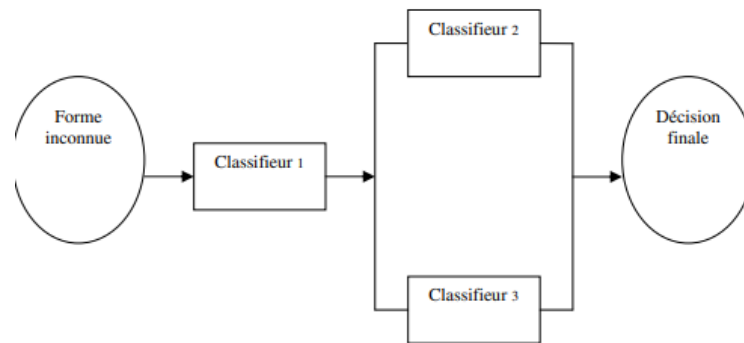
- Très facile à mettre en œuvre
- Ne nécessite pas une reparamétrisations des autres classifieurs en cas de modification de l'ensemble
- Activation de tous les classifieurs (coûteux en temps de calcul) [12]



**Figure I.5 :** Schéma de combinaison parallèle de classifieurs.

### C. 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. 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. [11]



**Figure I.6 :** Schéma de combinaison hybride de classifieurs.

### I.4.3 Méthode de combinaison moyenne

Le combineur moyen donne  $\hat{P}_1 = \frac{1}{L} \sum_{j=1}^L P_j$ . Si  $P_1, \dots, P_L$  sont normalement distribués et indépendants, alors  $\hat{P} \sim N(p, \frac{\sigma^2}{L})$ . La probabilité d'erreur pour ce cas est

$$P_e = P(\hat{P} < 0.5) = \Phi\left(\frac{\sqrt{L}(0.5 - p)}{\sigma}\right) \quad (I.28)$$

Le calcul de  $P_e$  pour le cas d'une distribution uniforme n'est pas si simple. Nous pouvons supposer que la somme de  $L$  variables indépendantes aboutira à une variable de distribution approximativement normale. Plus le  $L$  est élevé, plus l'approximation est précise. Sachant que la variance de la distribution uniforme pour  $P_j$  est  $\frac{b^2}{3}$ , on peut supposer que  $\hat{P} \sim N(p, \frac{b^2}{3L})$ . Alors [13]

$$P_e = P(\hat{P}_1 < 0.5) = \Phi\left(\frac{\sqrt{3L}(0.5 - p)}{b}\right) \quad (I.29)$$

## Conclusion

Dans ce chapitre, on a exposé le fonctionnement et les algorithmes de quelques classifieurs supervisés, qu'on utilisera par la suite, pour la conception d'un système d'identification des appareils électrique. On a également, expliqué l'utilité et l'objectif des approches de combinaison des classifieurs. La méthode de combinaison moyenne sera implémentée dans notre système multiclassifieurs.

# **CHAPITRE II :**

**Architecture du système d'identification  
des appareils électriques à implémenter**

## Chapitre II : Architecture du système d'identification des appareils Électriques à implémenter

### II.1 Introduction

Dans cette section, on expose une proposition pour la conception d'un système d'identification des appareils électrique avec cinq classificateurs combinés par l'approche des valeurs moyenne.

### II.2 Architecture du système de classification proposé

Un appareil électrique peut être reconnu en fonction de la caractéristique de ses courants électriques. Cependant, le système de classification de ces dispositifs doit inclure les éléments suivants :

- une phase d'apprentissage pour comprendre et simuler les différentes classes d'appareils.
- une phase de test pour catégoriser un groupe de signaux de test et évaluer l'efficacité du système d'identification.
- L'extraction des caractéristiques du signal que l'on pense noyées dans le signal électrique.
- L'étape suivante consiste à classifier l'ensemble des paramètres extraits pour obtenir l'identité ou la classe de l'appareil considéré en utilisant les classificateurs KNN, GMM, ANN, ADI, ARD.

Afin de pouvoir extraire les caractéristiques des signaux nous devons décomposer le signal du courant en utilisant les paramètres statistiques.

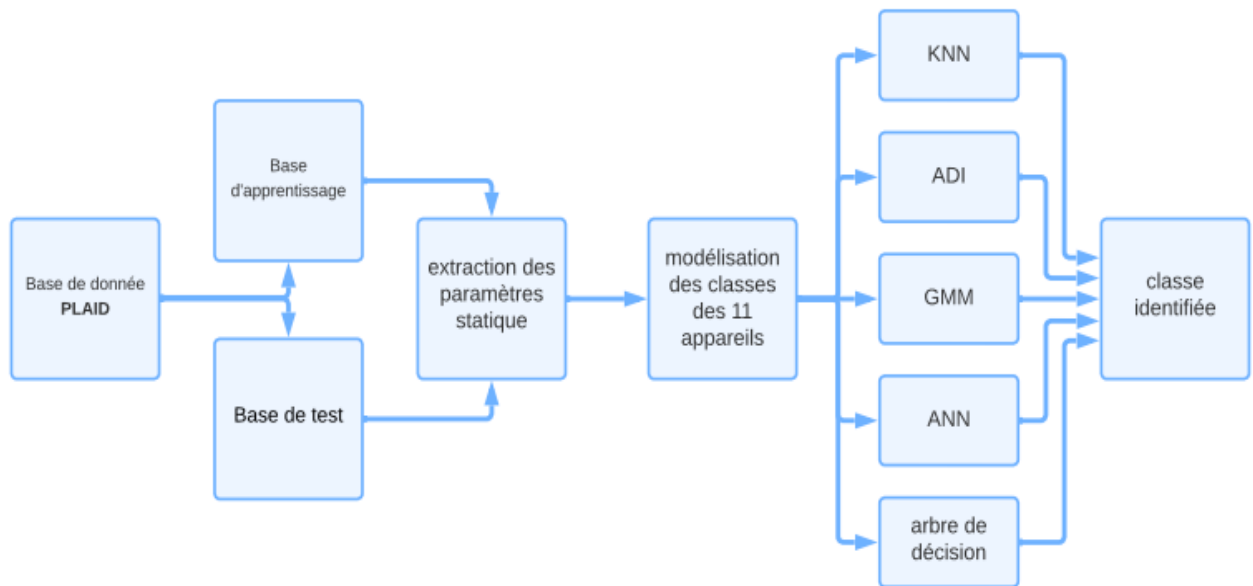
Le schéma synoptique de notre système de classification des appareils électriques est illustré sur **la figure II.1**. Ce système est conçu de :

- Une base de données d'entrée des courants électriques.
- Un bloc de division de la base de données en deux phases : apprentissage et test.
- Un bloc d'extraction des caractéristiques.
- Une unité de classification.
- Une évaluation des performances.

## Chapitre II : Architecture du système d'identification des appareils Électriques à implémenter

### II.2.1 Base de données PLAID (plug load appliance identification dataset)

Le jeu de données PLAID (Appliance Plug Load Identification) est un ensemble de données public qui contient des signatures électriques. Ces signatures ont été collectées pendant l'été 2013 à Pittsburgh, Pennsylvanie, aux États-Unis, à partir de 55 foyers. L'ensemble de données comprend 11 types d'appareils, avec trois à six instances pour chaque appareil, totalisant ainsi 1074 instances au total.



**Figure II.1:** Structure du système de classification proposé.

Chaque signature représente un signal de quelques secondes comprenant une phase transitoire d'activation (si disponible, car certains appareils, comme les réfrigérateurs qui fonctionnent en continu, peuvent ne pas avoir de phase transitoire) et une partie de la phase en régime permanent correspondant à la consommation stable. Les signaux sont échantillonnés à une fréquence de 30 kHz.

**Le Tableau 1** récapitule les types d'appareils présents dans le jeu de données, ainsi que le nombre d'instances pour chaque type. Pour plus de détails sur l'ensemble de données, vous pouvez vous référer à l'article de référence (Gao et al, 2014). [14]



## Chapitre II : Architecture du système d'identification des appareils Électriques à implémenter

Type d'appareil	Nombre d'occurrences
Climatiseur	66
Fluorescent compact Lampe	175
Ventilateur	115
Réfrigérateur	38
Sèche-cheveux	156
Chauffage	35
Lumière incandescente Ampoule	114
Ordinateur portable	172
Four micro-onde	139
Aspirateur	38
Machine à laver	26
<b>Total</b>	<b>1074</b>

**Tableau II.1** : Récapitulative des appareils trouvés sur la base de données PLAID.

### II.2.2 Phase d'apprentissage

Pour obtenir une identification précise des charges électriques, la phase d'apprentissage requiert souvent plusieurs séries de mesures. Cependant, ces mesures peuvent impliquer différents degrés d'intrusion. Elles nécessitent l'intervention d'un spécialiste qui se rend dans la résidence pour installer des dispositifs de mesure sur chaque appareil, afin de collecter les signatures électriques distinctes.

Parfois, pour améliorer la phase d'apprentissage, des tests supplémentaires sont réalisés en laboratoire. Ces tests peuvent comprendre des scénarios spécifiques ou des mesures supplémentaires sur les appareils dans des configurations contrôlées. Ces approches sont mises en place afin de garantir la qualité des données recueillies pendant la phase d'apprentissage.

Il est essentiel de noter que cette phase intrusive de mesure est effectuée en accord et avec la coopération des résidents, dans le but d'obtenir les informations nécessaires à une identification précise des charges électriques.

## Chapitre II : Architecture du système d'identification des appareils Électriques à implémenter

Après la détection des changements d'état, la deuxième étape de l'identification des charges électriques consiste à utiliser les signatures des appareils stockées dans **une base de données**. Ces signatures sont obtenues à l'aide de techniques d'apprentissage. Ainsi, les méthodes NIALM (Non-Intrusive Appliance Load Monitoring) peuvent être classées en deux catégories

**1. Les méthodes NIALM avec apprentissage automatisé** : Ces méthodes utilisent un processus d'apprentissage automatique pour identifier les charges électriques. Elles ne nécessitent pas d'intervention manuelle pour collecter les signatures des appareils.

**2. Les méthodes NIALM avec apprentissage manuel** : Ces méthodes nécessitent une phase d'apprentissage manuelle où les signatures des appareils sont collectées spécifiquement pour chaque installation. Ces méthodes manuelles ont tendance à être plus précises que les méthodes automatisées, car elles permettent de recueillir des signatures spécifiques des appareils présents dans l'installation.

Cependant, cette phase d'apprentissage manuel implique une certaine intrusion chez le client, ce qui peut être inconfortable pour lui. De plus, cela peut également ne pas être attrayant pour les distributeurs ou les fournisseurs d'énergie électrique. [15]

### II.2.3 Phase de test

La base de test est un ensemble d'exemples qui partage les mêmes caractéristiques que ceux de la base d'entraînement et qui sont mis de côté au début du processus d'entraînement pour être utilisés ultérieurement lors des tests. Elle est utilisée pour évaluer les performances et la précision du modèle prédictif développé.

La méthode de prédiction utilisée dépend du type d'information que l'on souhaite prédire, c'est-à-dire du type d'attribut de classe. Si l'attribut de classe est catégorique ou symbolique, ce qui signifie qu'il appartient à un ensemble fini de catégories distinctes, on parle de problème de classification. Dans ce cas, le modèle est entraîné à classer les exemples de la base de test dans différentes catégories prédéfinies.

En revanche, si l'attribut de classe est continu, c'est-à-dire qu'il est numérique et non catégorique, on parle alors d'un problème de régression. Dans ce cas, le modèle cherche à prédire une valeur numérique ou une tendance pour chaque exemple de la base de test.

La base de test est essentielle pour évaluer la capacité du modèle à généraliser et à faire des prédictions précises sur de nouvelles données qui n'ont pas été utilisées lors de l'entraînement. En utilisant cette base de test, on peut mesurer la performance du modèle, telle

## Chapitre II : Architecture du système d'identification des appareils Électriques à implémenter

que son taux de précision, son rappel, sa précision ou son erreur moyenne, selon le type de problème de prédiction. [7]

### II.2.4 Extraction des caractéristiques

Les caractéristiques extraites des signaux électriques sont utilisées pour identifier et différencier les appareils électriques en fonction de leurs schémas de consommation d'énergie. Elles doivent être soigneusement choisies pour capturer les informations essentielles permettant de distinguer les différents types d'appareils. L'objectif est de trouver **des caractéristiques pertinentes** qui peuvent expliquer de manière concise les variations de consommation électrique spécifiques à chaque classe d'appareils.

La sélection des caractéristiques est cruciale car elles doivent fournir une **représentation compacte** des données sans compromettre la qualité de l'identification. Cela signifie qu'elles doivent être aussi informatives et distinctives que possible tout en étant réduites en nombre. En d'autres termes, nous voulons identifier les caractéristiques les plus pertinentes tout en minimisant leur nombre.

L'utilisation de caractéristiques pertinentes et concises présente plusieurs avantages. Tout d'abord, cela facilite l'analyse des données en réduisant la complexité et la dimensionnalité des informations à traiter. De plus, cela permet de concevoir des systèmes d'identification plus efficaces, car une représentation compacte peut être plus rapidement comparée et utilisée pour classer les appareils électriques.

En résumé, les caractéristiques extraites des signaux électriques doivent être pertinentes, informatives et minimales pour pouvoir identifier les appareils électriques en fonction de leurs schémas de consommation d'énergie de manière efficace et précise. [16]

Dans notre cas, pour pouvoir extraire les caractéristiques des signaux nous devons décomposer le signal du courant en utilisant les paramètres statistiques.

#### • Les paramètres statistiques :

Les caractéristiques statistiques calculées sur les lignes de la matrice harmonique résultent en vecteurs statistiques de longueur égale à P, où P représente le nombre d'harmoniques. Ces vecteurs contiennent des mesures statistiques jusqu'au deuxième ordre, calculées sur les valeurs absolues de la matrice harmonique.

- La moyenne :  $\mu = E[M H]$  (II.1)

La moyenne est calculée par la formule :  $\mu(p) = \frac{1}{R} \sum_{r=1}^R M H(p, R)$ ,  $p = 1, \dots, P$ ; (II.2)

## Chapitre II : Architecture du système d'identification des appareils Électriques à implémenter

- L'écart type :  $\sigma = \sqrt{E[(M H - \mu)^2]}$  (II.3)

L'écart type est calculé par la formule :  $\sigma(p) = \sqrt{\frac{1}{R} \sum_{r=1}^R [(M H(p, R) - \mu(p))^2]}$  (II.4)

Enfin, chaque courant électrique d'appareil est représenté par un vecteur composé de deux valeurs statistiques temporelles caractérisant l'harmonique d'ordre p [17] :

$$V(p) = [\mu(p) \quad \sigma(p)] \quad (\text{II.5})$$

### II.2.5 Unité de classification

L'unité de classification fait référence à l'utilisation de plusieurs classifieurs dans le cadre d'un problème de classification. L'objectif est d'améliorer les performances de prédiction en combinant les résultats de plusieurs modèles plutôt qu'en utilisant un seul classifieur. Dans notre projet on utilise 5 classificateurs (KNN, ANN, GMM, ADI, ARD).

L'utilisation de plusieurs classifieurs dans une unité de classification présente plusieurs avantages. Elle permet de réduire le biais et la variabilité inhérents à un seul classifieur en combinant les forces de plusieurs modèles. Cela peut améliorer la robustesse et la précision des prédictions, en particulier dans les cas où les classifieurs individuels peuvent avoir des erreurs ou des lacunes spécifiques.

En outre, l'unité de classification peut faciliter l'apprentissage en fournissant une diversité dans les modèles utilisés. Les classifieurs individuels peuvent être formés avec différentes techniques, utiliser différents ensembles de données ou avoir différentes configurations, ce qui permet d'explorer davantage l'espace de recherche et d'obtenir une meilleure représentation des données.

En résumé, l'utilisation de plusieurs classifieurs dans une unité de classification permet de combiner les prédictions de différents modèles pour améliorer les performances de prédiction, réduire les erreurs spécifiques et faciliter l'apprentissage en explorant davantage l'espace de recherche.

### II.2.6 Performances

Le taux de classification des vecteurs est une mesure couramment utilisée pour évaluer les performances des classifieurs individuels. Il représente la précision avec laquelle un classifieur attribue la classe correcte à un échantillon. C'est une mesure simple et intuitive qui permet d'évaluer la capacité d'un classifieur à classifier correctement les données.

## Chapitre II : Architecture du système d'identification des appareils Électriques à implémenter

Le taux de classification est calculé en divisant le nombre d'échantillons correctement classés par le nombre total d'échantillons dans l'ensemble de test ou d'évaluation. Le résultat est généralement exprimé en pourcentage pour indiquer la précision du classifieur.

Un taux de classification élevé indique que le classifieur performe bien en prédisant correctement les étiquettes de classe des échantillons. Cependant, il est important de noter que la précision seule ne peut pas fournir une image complète des performances du classifieur, en particulier dans les cas de classes déséquilibrées ou lorsque les coûts des différentes erreurs varient.

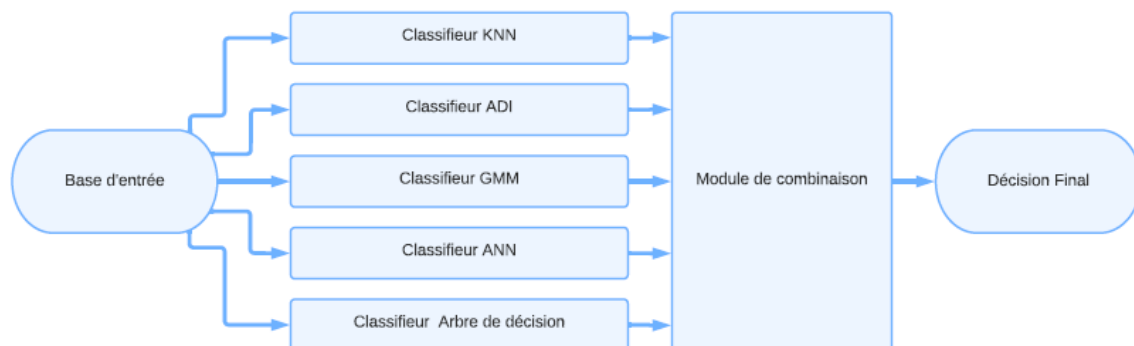
Dans les cas où les classes sont déséquilibrées (c'est-à-dire qu'il y a une différence significative dans le nombre d'échantillons entre les classes), le taux de classification peut être trompeur. Un classifieur peut obtenir un taux de classification élevé en prédisant simplement la classe majoritaire la plupart du temps, tout en se comportant mal sur la classe minoritaire.

### II.3 Structure du système multiclassifieurs proposé

Notre objectif est la conception et l'implémentation d'un système basé sur une combinaison de cinq classifieurs pour l'identification des appareils électriques avec un taux de combinaison acceptable.

Le système montré à la **figure II.2** est composé de :

- Une base d'entrée.
- Les classifieurs
- Module de combinaison.
- Décision finale.



**Figure II.2** : Schéma synoptique de système de combinaison proposé.

## Chapitre II : Architecture du système d'identification des appareils Électriques à implémenter

### II.3.1 Base d'entrée

Dans le contexte d'un système multiclassifieurs, une base d'entrée fait référence à l'ensemble de données utilisé pour entraîner et tester les classifieurs. Cette base d'entrée est généralement composée de deux parties : les exemples d'entraînement et les exemples de test.

**a) Exemples d'entraînement :** Il s'agit d'un sous-ensemble de la base d'entrée utilisé pour entraîner les classifieurs. Les exemples d'entraînement sont étiquetés avec les classes correspondantes, ce qui signifie que la classe de chaque exemple est connue. Les classifieurs apprennent à partir de ces exemples pour effectuer la classification.

**b) Exemples de test :** Il s'agit d'un autre sous-ensemble de la base d'entrée utilisé pour évaluer les performances des classifieurs. Les exemples de test sont également étiquetés avec les classes correspondantes. Après l'entraînement des classifieurs, ils sont utilisés pour prédire les classes des exemples de test et ces prédictions sont comparées aux étiquettes réelles pour évaluer la précision et les performances des classifieurs.

Il est important que la base d'entrée soit représentative du problème réel que l'on cherche à résoudre. Cela signifie qu'elle devrait inclure des exemples de toutes les classes que l'on souhaite classifier, et les exemples devraient être diversifiés et représentatifs de la variabilité du problème. Une base d'entrée de qualité est cruciale pour obtenir des classifieurs performants et fiables.

### II.3.2 Les classifieurs

Dans un système multiclassifieurs, les classifieurs travaillent en parallèle pour fournir des prédictions individuelles pour chaque échantillon, qui sont ensuite combinées pour obtenir une décision finale sur la classe d'appartenance. Chaque classifieur est responsable de la classification entre une paire spécifique de classes. Chaque classifieur est entraîné à distinguer entre ces deux classes uniquement. La sortie de chaque classifieur est un vecteur de classes, où chaque dimension représente la probabilité ou la confiance associée à une classe spécifique.

Il est important de noter que les classifieurs parallèles doivent être conçus de manière à ne pas interférer les uns avec les autres et à fonctionner de manière indépendante. Ils doivent être entraînés et utilisés en parallèle pour maximiser l'efficacité du système multiclassifieurs.

En fin de compte, la sortie de chaque classifieur est combinée pour obtenir une décision finale sur la classe d'appartenance de l'échantillon.

## **Chapitre II : Architecture du système d'identification des appareils Électriques à implémenter**

### **II.3.3 Module de combinaison**

Dans un système multiclassifieurs fonctionnant en parallèle, le module de combinaison est responsable de fusionner les sorties des classifieurs parallèles afin d'obtenir une prédiction finale pour chaque échantillon. Son objectif est de tirer parti des perspectives diverses fournies par les classifieurs individuels pour générer une prédiction plus précise et robuste. Le module de combinaison peut utiliser différentes techniques pour agréger les sorties, telles que le vote majoritaire, la somme pondérée, la méthode min-max etc...

Dans notre cas, on implémente la technique des valeurs moyenne qui combine les classes avec leurs probabilités des cinq classifieurs considérés.

### **II.3.4 Décision finale**

La décision finale dans un système multiclassifieurs est la classe ou l'étiquette attribuée à un échantillon donné après la combinaison des sorties des classifieurs parallèles. Cette étape est cruciale car elle détermine la prédiction finale du système pour chaque échantillon.

Le choix de la méthode de décision finale dépend des caractéristiques spécifiques du système multiclassifieurs, telles que le nombre de classifieurs, leur performance individuelle et leurs biais. Il est généralement recommandé de comparer différentes approches et d'évaluer leurs performances sur des ensembles de données de test pour choisir la méthode la plus appropriée pour un problème de classification donné.

## **II.4 Conclusion**

Les systèmes de classification et les systèmes de multiclassifieurs sont des approches clés en apprentissage automatique pour la catégorisation de données dans des classes prédéfinies. Ils jouent un rôle essentiel dans l'analyse des données et la catégorisation précise des données dans différentes classes.

# **CHAPITRE III :**

## **Implémentation et résultats**



### III.1 Préliminaire

Notre travail est divisé en deux parties. Dans la première partie, on implémente un système de classification des appareils électriques et le tester en utilisant à chaque fois un type de classifieur, à savoir : le **KNN**, **ADI**, **ANN**, **GMM** et le classifieur **d'arbre de décision**. La base d'entrée est conçue des paramètres statistiques (du courant électrique) extraites de la base de données **PLAID**. L'évaluation est réalisée en calculant le taux de classification des vecteurs.

Dans la deuxième partie de ce travail, on implémente un système multiclassifieurs qui combine les sorties des cinq classificateurs avec la fonction de décision **des valeurs moyenne**, dans le but d'implémenter une nouvelle structure de combinaison pour les systèmes d'identification des appareils électrique et d'améliorer les performances de classification.

### III.2 Implémentation du système de classification des appareils électriques

Cette implémentation est basée sur :

- **La base de données PLAID prétraitée** (enlèvement des durées de mise en marche et d'arrêt). Cette base contient 1074 courants électriques de 11 appareils différents enregistrés avec une fréquence d'échantillonnage de 30 KHz dans un intervalle de 5 secondes. Ensuite, on divise la base de données en :

→ une base d'apprentissage avec **537 signaux** qui représentent les **11 types** d'appareils électriques.

→ et une base de test qui contient les autres **537 signaux** de tous les types d'appareils figurant dans la base de données **PLAID** d'origine.

→ Chaque signal, des deux bases obtenues, est référencié par un **indice** et il est découpé en **250 vecteurs**.

- **L'étape d'extraction des paramètres** caractéristiques est réalisée selon la méthode suivante:

→ Chaque séquence de vecteurs (c.-à-d. chaque signal) est convertie en un seul vecteur qui représente les **paramètres statistiques** (moyenne et écart type) par l'utilisation des expressions données au deuxième chapitre.

→ On obtient, donc, pour chaque base de données (apprentissage et test) **537** valeurs pour la moyenne et **537** valeurs pour l'écart type.

→ A chaque vecteur de ces paramètres caractéristiques, on associe une 'classe'. Rappelant qu'on a **11** classes représentant les 11 types d'appareils électriques.

→ Enfin, l'ensemble des vecteurs obtenus des deux cas:  $V_i$  (**moyenne<sub>i</sub>**; **écart type<sub>i</sub>**; **classe<sub>i</sub>**) tel que  $i = 1$  à **537**, sont regroupés dans deux matrices : d'apprentissage et de test.

- **L'étape de classification** consiste à classer les vecteurs de paramètres caractéristiques de la base de test selon les modèles sauvegardés dans 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 :

### 1. L'algorithme KNN :

#### Début Algorithme

Données en entrée :

- Un ensemble de données D (base de donnée).
- Une fonction de définition de la distance Cityblock d.
- Un nombre entier K.

Pour une nouvelle observation X (vecteur) dont on veut prédire sa variable de sortie y (classe) Faire :

1. Calculer toutes les distances de cette observation X avec les autres observations du jeu de données D (base de données d'apprentissage).
2. Retenir les K plus proches observations du jeu de données D de X en utilisant la fonction de calcul de distance d.
3. Prendre les valeurs de y (classe) des K observations retenues.
4. Retourner la valeur y (classe) comme étant la valeur qui a été prédite par K-NN pour l'observation.

#### Fin Algorithme

### 2. L'algorithme arbre de décision :

**Étape 1** : Démarrez l'arbre avec le nœud racine, appelé S, qui contient l'ensemble de données complet.

**Étape 2** : Trouvez le meilleur attribut dans l'ensemble de données à l'aide de la mesure de sélection d'attribut (ASM).

**Étape 3** : Divisez le S en sous-ensembles contenant les valeurs possibles pour les meilleurs attributs.

**Étape 4** : Générez le nœud de l'arbre de décision, qui contient le meilleur attribut.

**Étape 5** : Créez de nouveaux arbres de décision de manière récursive à l'aide des sous-ensembles de l'ensemble de données créé à l'étape 3. Continuez ce processus jusqu'à

ce qu'un stade soit atteint où vous ne pouvez plus classer les nœuds et appelé le nœud final en tant que nœud feuille.

### 3. L'algorithme ANN :

**Entrée :** jeu de données D, base d'apprentissage, réseau.

**Sortie :** un réseau de neurones formé.

**Étape 1 :** recevez l'entrée.

**Étape 2 :** pesez l'entrée. Chaque entrée envoyée au réseau doit être pondérée, c'est-à-dire multipliée par une valeur aléatoire entre -1 et +1.

**Étape 3 :** additionnez toutes les entrées pondérées.

**Étape 4 :** générer une sortie : la sortie du réseau est produite en faisant passer cette somme par la fonction d'activation

### 4. L'algorithme GMM :

**Étape 1 :** Initialisez la moyenne  $\mu_k$ , la matrice de covariance  $\Sigma_k$  et les coefficients de mélange  $\pi_k$  par certaines valeurs aléatoires (ou autres valeurs).

**Étape 2 :** Calculez le  $C_k$ , valeurs pour tous les k.

**Étape 3 :** Estimez à nouveau tous les paramètres en utilisant les valeurs actuelles  $\mu_k$  et  $\Sigma_k$ .

**Étape 4 :** Calculez la fonction log-vraisemblance.

**Étape 5 :** Mettre un certain critère de convergence.

**Étape 6 :** Si la valeur log-vraisemblance se converge vers une certaine valeur (ou si tous les paramètres convergent vers certaines valeurs), arrêtez, revenez à l'étape 2.

### 5. L'algorithme discrimination linéaire:

**Input :** Données de test;  $X^{\text{test}} = (X_1^{\text{test}}, \dots, X_n^{\text{test}})$ ; 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  $x_i$  en utilisant l'équation  $f_{ik}(x_i; \mu_k, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \Sigma^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (x_i - \mu_k)^T \Sigma^{-1} (x_i - \mu_k)\right)$

**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  $\tau_{ik}$  données par l'équation (5) :  $\hat{z}_i = \arg \max_{k=1}^K \tau_{ik}$

**End**

**Résultat :**  $\hat{z}_i$  ( $i = 1, \dots, n$ ) : classes de test estimés.

A la sortie de chaque classifieur, on obtient un **vecteur de classes** qui nous permet de calculer le **taux de classification**.

- L'étape d'évaluation : pour évaluer les méthodes de classification utilisées, un taux de classification des vecteurs est calculé pour chaque expérience d'identification selon l'expression:

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

### III.3 Résultats de classification et discussions

#### III.3.1 Classification par l'arbre de décision

Des simulations sont exécutées sous MATLAB. Avec la variation du nombre de vecteurs statiques, on obtient les taux de classifications donnés au **tableau III.1**. On remarque que, le meilleur taux prend la valeur **98.7671%**, il est obtenu pour un nombre de vecteurs égal à 150. La figure III.1 montre la fenêtre des résultats de cette simulation.

Nombre de vecteur	TC_ARD
10	98.4293
20	98.6489
50	98.2942
70	98.6658
100	98.5898
125	98.7164
150	98.7671
200	98.7249
250	98.7502

**Tableau III-1:** Variation de taux de classification avec le classifieur TREECLASS.

Name ▲	Value
nn	11x11042 double
ra	1x150 double
sa	150
t	1x1 classregtree
TABap	11911x150 double
TABte	11842x150 double
targets	11842x1 double
TC_ADI	70.0895
TC_ANN	78.1709
TC_ARD	98.7671
TC_GMM	36.8772
TC_knn	87.3839
tind	1x11842 double
VC_ADI	11842x1 double
VC_ANN	11842x1 double
VC_ARD	11842x1 double
VC_knn	11842x1 double
vect_par_mach	537x1 double
vpm	537x1 double

Figure III.1 : Fenêtre des résultats de classification ARD avec un nombre de vecteurs = 150.

### III.3.2 Classification avec le classifieur KNN

Le tableau III.2 illustre les résultats de simulation, 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 95.4315 % est obtenu avec la configuration (K = 3, distance = 'Cityblock').

K	Euclidean	Cityblock	Cosine	Correlation
1	95.3471	95.0600	83.5247	83.2461
2	95.3471	95.0600	83.5247	83.2461
3	95.3977	95.4315	84.0736	84.1327
4	95.3893	95.3977	84.3016	84.2341
5	95.2035	95.1697	84.2594	84.3185
6	95.1613	95.1106	84.2510	84.2510
7	94.6377	94.7053	84.0061	84.1159
8	94.6293	94.7222	84.1074	84.0990
9	94.3422	94.4604	84.0905	83.9976
10	94.2999	94.4942	84.0061	83.9723

Tableau III-2 : Sélection de la configuration optimale de classificateur KNN.

Pour simuler notre système d'identification, on utilise la configuration (K = 3, distance = 'Cityblock'). Les résultats de cette expérience sont représentés dans le tableau III.3. Ils représentent la variation de taux de classification en fonction de nombre de vecteurs statistiques. On peut remarquer que, la valeur la plus maximum obtenue est :

TC\_KNN = 95.4315 % avec un nombre de vecteur égale à 15 Cette valeur est soulignée dans la fenêtre de simulation indiquée à la figure III.2.

Nombre de vecteur	TC_KNN
5	94.8066
15	95.4315
20	94.2577
30	93.3879
50	92.6533
70	90.8546
100	89.2670
125	88.0594
150	87.3839
200	86.0328
250	84.9688

**Tableau III-3:** Variation de taux de classification avec le classifieur KNN.

Name ▲	Value
nn	1x11842 double
ra	1x15 double
sa	15
t	1x1 classregtree
TABap	11911x15 double
TABte	11842x15 double
targets	11842x1 double
TC_ADI	75.3589
TC_ANN	80.9576
TC_ARD	98.5982
TC_GMM	79.7163
TC_knn	95.4315
tind	1x11842 double
VC_ADI	11842x1 double
VC_ANN	11842x1 double
VC_ARD	11842x1 double
VC_knn	11842x1 double
vect_par_mach	537x1 double
vpm	537x1 double

**Figure III.2 :** Fenêtre des résultats de classification KNN avec un nombre de vecteurs = 15.

### III.3.3 Identification avec le classifieur ADI

Avec ce classifieur, en variant le type de la fonction discriminante de classifieur ADI. Vu que le meilleur taux de classification égal à 75.8993 %, la configuration 'type = quadratique' sera considérée dans les prochaines expériences. Le tableau III-4 illustre les résultats obtenus.

Type de fonction	Quadratic	Linear	Mahalanobis
Taux de classification	75.8993	64.5921	52.6347

**Tableau III-4 :** Sélection de la configuration optimale de classificateur ADI.

Pour simuler notre système d'identification, on utilise la configuration (type = 'Quadratic'). Les résultats de cette expérience sont représentés dans le tableau III.4. Ils représentent la variation de taux de classification en fonction de nombre de vecteurs statistiques. On peut remarquer que, La valeur la plus maximum obtenue est :

TC\_ADI= 95.4315 % avec un nombre de vecteur égale à 10 Cette valeur est soulignée dans la fenêtre de simulation indiquée à la figure III.5.

Mais, ce type de classifieur ne donne pas les bonnes valeurs de taux de classification par rapport à la classification KNN et TREECLASS.

Nombre de vecteur	TC_ADI
10	75.8993
20	71.9811
50	69.7771
70	69.8953
100	70.0980
125	70.2162
150	70.0895
200	71.0015
250	70.6637

**Tableau III-5 :** Variation de taux de classification avec le classifieur ADI.

Name ▲	Value
nin	11x11842 double
ra	[1,2,3,4,5,6,7,8,9,10]
sa	10
t	1x1 classregtree
TABap	11911x10 double
TABte	11842x10 double
targets	11842x1 double
TC_ADI	75.8993
TC_ANN	83.2207
TC_ARD	98.4293
TC_GMM	79.5220
TC_knn	94.8573
tind	1x11842 double
VC_ADI	11842x1 double
VC_ANN	11842x1 double
VC_ARD	11842x1 double
VC_knn	11842x1 double
vect_par_mach	537x1 double
vpm	537x1 double

**Figure III.3** : Fenêtre des résultats de classification ADI avec un nombre de vecteurs = 10.

### III.3.4 Identification avec le classifieur ANN

Le temps d'exécution avec ce type de classificateur est le plus long par rapport aux autres classificateurs utilisés, dans cette implémentation.

Compte tenu de la vitesse et du temps de calcul et après plusieurs simulations avec ce type de classifieur, la configuration optimale est la suivante

La division des données pour l'Apprentissage, la validation et les tests de ce classifieur est :

- Apprentissage 70%
- Validation 15%
- Test 15%

La structure finale 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-6, le meilleurs taux de classification obtenue est 85.0025 avec un nombre de vecteur égale à 50.



Nombre de vecteur	TC_ANN
15	80.9576
20	83.6092
25	79.1336
30	75.1731
50	85.0025
70	77.4362
100	80.1216
150	78.1709
200	74.9536
250	80.6029

**Tableau III-6 :** Variation de taux de classification avec le classifieur ANN.

Name	Value
nn	11x11042 double
ra	1x50 double
sa	50
t	1x1 classregtree
TABap	11911x50 double
TABte	11842x50 double
targets	11842x1 double
TC_ADI	69.7771
TC_ANN	85.0025
TC_ARD	98.2942
TC_GMM	79.4629
TC_knn	92.6533
tind	1x11842 double
VC_ADI	11842x1 double
VC_ANN	11842x1 double
VC_ARD	11842x1 double
VC_knn	11842x1 double
vect_par_mach	537x1 double
vpm	537x1 double

**Figure III.4 :** Fenêtre des résultats de classification ANN avec un nombre de vecteurs = 50.

### III.3.5 Identification avec le classifieur GMM

Afin d'évaluer l'effet du nombre de vecteurs statiques sur les taux de classification. Les résultats, présentés dans le tableau III-7, montrent que le taux de classification varie avec le nombre de vecteurs. Il est à noter que le meilleur taux de classification, atteignant 79,6487 %, et atteint lorsque le nombre de vecteurs est égal à 40. La figure III.5 donne une représentation visuelle des résultats obtenus à partir de cette simulation.

Avec 150 vecteurs, le taux de classification chute à 36.8772%, suggérant une diminution substantielle lorsque le nombre de vecteurs devient élevé.

Nombre de vecteur	TC_GMM
10	79.5220
20	79.1589
40	79.6487
50	79.4629
70	78.1963
100	77.8838
125	78.1287
150	36.8772

Tableau III-7 : Variation de taux de classification avec le classifieur GMM.

Name	Value
nin	11x11042 double
ra	1x40 double
sa	40
t	1x1 classregtree
TABap	11911x40 double
TABte	11842x40 double
targets	11842x1 double
TC_ADI	70.0811
TC_ANN	77.4616
TC_ARD	98.6067
TC_GMM	79.6487
TC_knn	93.0079
tind	1x11842 double
VC_ADI	11842x1 double
VC_ANN	11842x1 double
VC_ARD	11842x1 double
VC_knn	11842x1 double
vect_par_mach	537x1 double
vpm	537x1 double

Figure III.5 : Fenêtre des résultats de classification GMM avec un nombre de vecteurs = 40.

### III.4 Implémentation du système multi-classificateur (SMC)

L'architecture du système multi-classificateur représentée à la figure III.6 est une architecture de combinaison parallèle largement utilisée dans la littérature. Elle présente les caractéristiques suivantes :

- Les classifieurs utilisent les mêmes données d'apprentissage et de test en entrée.

- Les classifieurs fonctionnent simultanément et de manière indépendante les uns des autres.

- Il n'y a pas d'interaction entre les classifieurs.

- Les classifieurs sont fixes et ne sont pas modifiés.

Dans notre implémentation, nous avons les composants fonctionnels suivants :

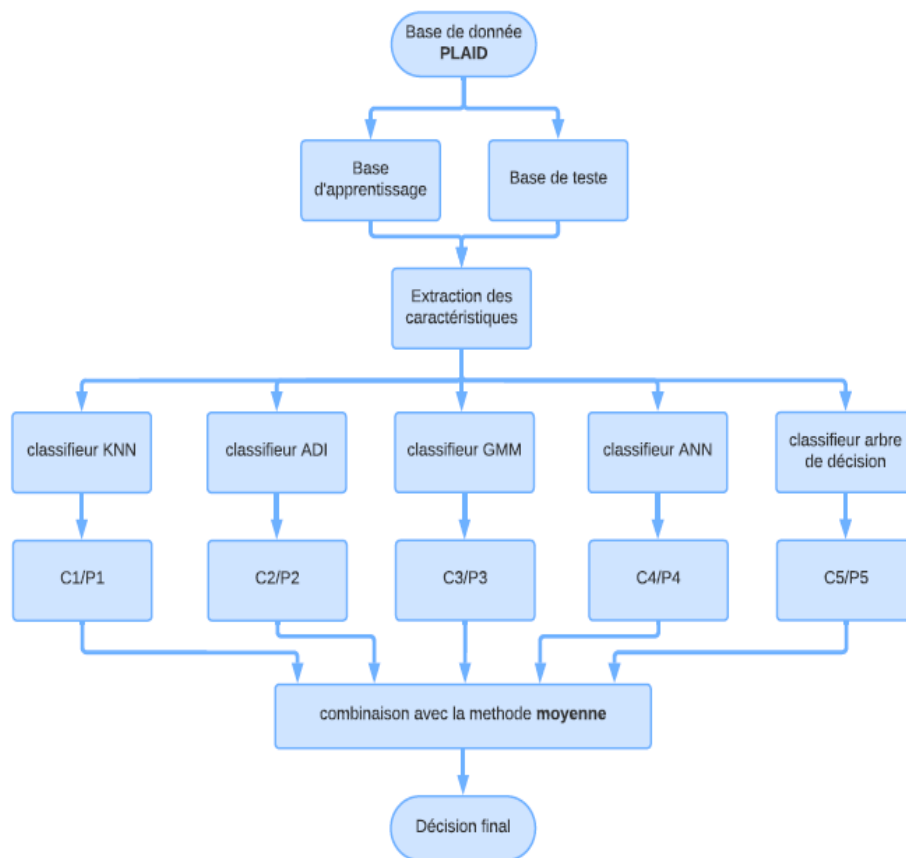
- Une base de données de formes à identifier en entrée du système.

- Des classifieurs parallèles tels que l'arbre de décision ARD, l'ANN (réseau de neurones artificiels), l'ADI (analyse discriminante linéaire), le KNN (k plus proches voisins) et le GMM.

- Une unité de combinaison qui fournit la classification finale.

Cette architecture parallèle permet de tirer parti de plusieurs classifieurs pour obtenir des décisions de classification plus robustes.

**La base d'entrée :** on utilise la base des paramètres statistiques arrangés dans les matrices d'apprentissage et de tests utilisés pour valider notre système de classification.



**Figure III.6 :** L'architecture du système multi-classificateur

**Les classifieurs** : les cinq 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. L'approche considérée est une approche parallèle avec apprentissage et les sorties des classifieurs sont de type 'classe', notre implémentation basé sur la méthode des valeurs Moyenne pour combiner les sorties des classifieurs et donner une décision finale.

Cette méthode de combinaison est programmée selon l'algorithme suivant :

#### **Début d'algorithme**

1- Initialisation des variables : Les variables « L, c, S, P, » sont définies et initialisées.

2- Calcul des compteurs de classes : À l'intérieur des boucles, les compteurs de classes « cM, cN1, cN2, cN3, cN4 et cN5 » sont mis à jour en comptant les occurrences des différentes classes dans les données.

3- Construction de la matrice « DP » : Les compteurs de classes « cN1, cN2, cN3, cN4, cN5 » sont utilisés pour construire la matrice « DP », qui est une combinaison de ces compteurs.

4- Calcul de « P » : La somme des valeurs dans la matrice « DP » est divisée par « L » pour obtenir « P », qui représente les proportions de chaque classe dans les données.

5- Calculer La différence entre les compteurs de classes « cM » et « P » est calculée pour obtenir « Def » qui présenter l'erreur dans chaque classe. Ensuite, la somme des valeurs absolues de Def est calculée et stockée dans la variable « som ».

6- Calcul du taux d'erreur : La variable « EROR » est mise à jour en divisant som par « S » (le nombre de lignes dans ree), ce qui donne le taux d'erreur.

7- Calcul du taux de combinaison : Le taux de combinaison « TC\_CMB » est calculé en soustrayant « EROR » de 1 et en multipliant le résultat par 100.

**Fin**

### **III.5 Résultats de combinaison**

Le tableau III.8 fourni semble représenter une comparaison entre le nombre de vecteurs et le taux de combinaison correspondant.

Les données montrent que lorsque le nombre de vecteurs augmente de 10 à 50, le taux de combinaison augmente également, passant de 90,7110 à 92,5942. Cela indique qu'il existe une corrélation positive entre le nombre de vecteurs et le taux de combinaison dans cette plage.

Cependant, de manière intéressante, lorsque le nombre de vecteurs augmente davantage de 50 à 100, le taux de combinaison diminue légèrement pour atteindre 91,8679. Cela suggère qu'il pourrait y avoir d'autres facteurs qui influencent le taux de combinaison au-delà d'un certain seuil.

Les données pour 150 vecteurs et 200 vecteurs montrent une chute significative du taux de combinaison à 75,2964 et 62,8846 respectivement. Cela pourrait indiquer que lorsque le nombre de vecteurs devient trop élevé, le taux de combinaison peut diminuer de manière substantielle.

<b>Nombre de vecteurs</b>	<b>Taux de combinaison</b>
10	90.7110
25	91.1704
50	92.5942
100	91.8679
150	75.2964
200	62.8846

**Tableau III-8 :** Variation de taux de combinaison en fonction de nombre de vecteurs.

Name ▲	Value
TABap	11911x50 double
TABte	11842x50 double
targets	11842x1 double
TARGETS	11911x1 double
TC_ADI	69.7771
TC_ADI_1	80.3795
TC_ANN	85.0025
TC_ANN_1	78.3897
TC_ARD	98.2942
TC_ARD_1	98.6903
TC_CMB	92.5942
TC_GMM	79.4629
TC_GMM_1	82.8646
TC_KNN	92.6533
TC_KNN_1	99.4543
tind	1x11842 double
TIND	1x11911 double
VC_ADI	11842x1 double
VC_ADI_1	11911x1 double

Figure III.7 : fenêtre des résultats de combinaison avec un nombre de vecteurs = 50.

Dans le **tableau III-9**, nous avons supprimé le classifieur ADI par ce que ce classifieur avait un taux de classification faible, on remarque que le taux de combinaison est amélioré.

Nombre de vecteurs	Taux de combinaison
10	92.3577
25	<u>93.0122</u>
50	92.7652
100	92.9340
150	71.7721
200	56.7049

Tableau III-9 : Variation de taux de combinaison en fonction de nombre de vecteurs après élimination d’ADI.

Name ▲	Value
TABap	11911x25 double
TABte	11842x25 double
targets	11842x1 double
TARGETS	11911x1 double
TC_ADI	71.4406
TC_ADI_1	79.5231
TC_ANN	79.1336
TC_ANN_1	74.3430
TC_ARD	98.6742
TC_ARD_1	98.7323
TC_CMB	93.0122
TC_GMM	79.2518
TC_GMM_1	82.6631
TC_KNN	93.4808
TC_KNN_1	99.4543
tind	1x11842 double
TIND	1x11911 double
VC_ADI	11842x1 double
VC_ADI_1	11911x1 double

**Figure III.8** : fenêtre des résultats de combinaison après élimination de ADI avec un nombre de vecteurs = 25.

Après l'élimination de classifieur ADI, on a éliminé aussi le classifieur GMM à cause de son taux de classification de moins de 80%. On remarque que le taux de combinaison augmente et il égale à 96.0254%.

Nombre de vecteurs	Taux de combinaison
10	96.0254
25	92.8390
50	94.5871
100	93.7651
150	93.7708
200	92.3774

**Tableau III-10** : Variation de taux de combinaison en fonction de nombre de vecteurs après élimination d'ADI et GMM.

Name ▲	Value
TABap	11911x10 double
TABte	11842x10 double
targets	11842x1 double
TARGETS	11911x1 double
TC_ADI	75.8993
TC_ADI_1	80.5306
TC_ANN	83.2207
TC_ANN_1	85.9374
TC_ARD	98.4293
TC_ARD_1	98.5979
TC_CMB	96.0254
TC_GMM	79.5220
TC_GMM_1	82.4364
TC_KNN	94.8573
TC_KNN_1	99.4543
tind	1x11842 double
TIND	1x11911 double
VC_ADI	11842x1 double
VC_ADI_1	11911x1 double

**Figure III.9** : fenêtre des résultats de combinaison après élimination de ADI et GMM avec un nombre de vecteurs = 25.

### III.6 Conclusion

A travers ce travail, on a réussi à mettre en place un système de classification et un système multiclassifieurs. Plusieurs expériences sont exécutées et les résultats obtenus ont été satisfaisants, en termes de taux de classification. Dans le système multiclassifieurs, il est à noter que le meilleur taux de combinaison atteint 92.5942%, lorsque le nombre de vecteurs égal à 50. Les performances s'améliorent lorsque nous supprimons les classifieurs ADI et GMM, et le taux de combinaison augmente à 96.0254%.

Nous considérons que notre objectif a été atteint puisque un système multi classifieurs est implémenté et expérimenté.



## ***Conclusion générale***

Dans ce travail, nous avons développé un système de classification pour les appareils électriques. Notre approche repose sur l'utilisation de cinq types de classificateurs : KNN, ANN, ADI, GMM et un Arbre de décision. Pour évaluer et tester notre système, nous avons créé deux ensembles de données : un ensemble d'apprentissage et un ensemble de test. Ces ensembles sont composés de paramètres statistiques extraits de la base de données des courants électriques PLAID. Nous avons obtenu des taux de classification acceptables.

Le taux de classification maximal a été atteint avec l'utilisation de l'arbre de décision, avec une précision de 98,7671%. Les quatre autres classificateurs étaient nécessaires pour concevoir un système de classification multi-classificateur. Nous avons intégré ces classificateurs en adoptant une approche parallèle, où chaque classificateur fonctionne indépendamment. Pour combiner les sorties des classificateurs, nous avons utilisé la méthode des valeurs Moyenne comme fonction de liaison. En termes de taux de classification des vecteurs, nous avons obtenu une valeur maximale de 92,5942% et 96,0254% après avoir éliminé les classificateurs ADI et GMM.

Ce système pourrait être utile dans diverses applications, telles que la reconnaissance d'images et le traitement du langage naturel, où une classification précise et efficace est cruciale.

Cette contribution peut être complétée par l'utilisation:

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

## Références

[1] : Guermoudi Mohammed el Amine, Fekih Mohammed el Amine, 'Fusion des classifieurs supervisés: Application sur la classification pixellaire des images microscopiques' Mémoire Master-Université Abou Bakr Belkaid– Tlemcen 2013.

[2] : Amirouche Radia, 'Une combinaison de classifieurs pour la reconnaissance des visages humains', mémoire de magister, Université de Annaba, 2006

[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] : [REF1\_classification combinaison\_conceptsdebase]

[5] : BENKACI IMAD, BELMOUMENE RABAH, 'Conception d'un système multi classifieurs en utilisant la méthode multinomiale' Mémoire de Master, Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj 2022.

[6] : Danho Djrobie, 'modèle de mélange et classification', thèse de doctorat, Dauphine université de paris, 2016.

[7] : Ghalem Dalila, Zegadi Yousra, ' Identification des appareils électriques basée sur les modèles GMM' Mémoire de Master, Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj 2021.

[8] : Faïcel Chamroukhi, 'Classification supervisée : Analyse discriminante' Université de Caen Lab of Mathematics Nicolas Oresme UMRCNRS6139.

[9] : Houacine Noura, 'Classification des textures par les réseaux de neurones convolutifs', Mémoire de Master, DEPARTEMENT D'AUTOMATIQUE, UNIVERSITE MOULOUD MAMMARI DE TIZI-OUZOU, 2018

[10]: Cite web: '<https://www.knowledgehut.com/blog/data-science/linear-discriminant-analysis-for-machine-learning>'.

[11] : Bensefia Adem, 'Identification des appareils électriques par combinaison des classifieurs', Mémoire de Master, Université de Bordj Bou Arreridj, 2021.

[12] : Laurent HEUTTE, 'Combinaison de Classifieurs Pourquoi et comment les combiner?' Université de Rouen.

[13]: LUDMILA I. KUNCHEVA, 'combining pattern classifiers: Methods and Algorithms' Published by John Wiley & Sons, Inc., Hoboken, New Jersey 2014.

[14]: Nait-Meziane, M., Hacine-Gharbi, A., Ravier, P., Lamarque, G., Le Bunetel, J-C. and Raingeaud, Y. « HMM-based transient and steady-state current signals modeling for electrical appliances identification », Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM), pp.670–677. (2016)

[15]: 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 ». Thèse, Spécialité : Electromagnétisme. Université Blaise Pascal - Clermont-Ferrand II, 9/12/2009. France.

[7]: GHALEM Dalila, ZEGADI Yousra [Identification des appareils électriques basée sur les modèles GMM] Mémoire de Master, Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj **2021**

[16]: Abdenour Hacine Gharbi and Philippe Ravier [Wavelet Cepstral Coefficients using for Electrical Appliances Identification Hidden Markov Models] LMSE laboratory, University of Bordj Bou Arréridj, Elanasser, 34030 Bordj Bou Arréridj, Algeria PRISME laboratory, University of Orleans, 12 rue de Blois, 45067 Orleans, France

[17]: 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).