

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

- **BENKACI IMAD**
- **BELMOUMENE RABAH**

Intitulé

*Conception d'un système multi classifieurs en utilisant la méthode
multinomiale*

Soutenu le :

Devant le Jury composé de :

<i>Nom & Prénom</i>	<i>Grade</i>	<i>Qualité</i>	<i>Etablissement</i>
<i>Meguellati Sabrina</i>	<i>MAA</i>	<i>Président</i>	<i>Univ-BBA</i>
<i>Mme.Laouamri Asma</i>	<i>MAA</i>	<i>Encadreur</i>	<i>Univ-BBA</i>
<i>Yousfi Abderahime</i>	<i>MAB</i>	<i>Examineur</i>	<i>Univ-BBA</i>

Année Universitaire 2021/2022

Résumé

Notre projet porte sur l'étude des systèmes d'identification des appareils électriques, la conception de ces systèmes dépendant de la mise en œuvre de la méthode de classification. Les méthodes de classification peuvent être distinguées par des méthodes intrusives et non intrusives, nous avons mis en place un système pour identifier les appareils électriques et nous nous concentrons sur la recherche de la façon de combiner plusieurs classificateurs pour obtenir de meilleurs systèmes de classification. Le système dépend des paramètres statistiques du courant électrique. Un système multi-classificateur a été conçu qui combine la sortie des quatre classificateurs (KNN, ADI, ANN et arbre de décision) et le taux d'identification obtenu est satisfaisant.

Abstract

Our subject concerns the study of identification systems for electrical appliances, the design of these systems depends on the implementation of the method of classification. Classification methods can be distinguished by intrusive and non-intrusive methods, we have implemented a system to identify electrical appliances, and we focus on researching how to combine multiple classifiers to achieve better classification systems, the system depends on the statistical parameters of the electric current. A multi-classifier system was designed which combines the output of the four classifiers (KNN, ADI, ANN and decision tree) and the identification rate obtained is satisfactory.

تلخيص

تُعكف موضوعنا بدراسة أنظمة تحدد الأجهزة الكهربائية ، وعتمد نصمم هذه الأنظمة على تنفيذ طريقة التصنيف. نمكن نمز طرق التصنيف بالطرق الدخلة وغر الللنحامة ، نمذ طبننا نظاما لنعرف على الأجهزة الكهربائية ، وركز على البحث لنعكف الجمع بين المصنفات المعددة لتحكم أنظمة تصنيف أفضل ، وعتهد النظام على المعلومات الإحصائية لتأار الكهربائية . نم نصمم نظام من عدد المصنفات لجمع بين ناچ المصنفات الأربعة (KNN و ADI و ANN وشجرة

المرار) ومعدل التعرف الذي نم الحصول عليه مرض.

REMERCIEMENTS

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

Nous adressons nos vifs remerciements à notre encadreur **Dr. A.LAOUAMRI** pour ses conseils et ses dirigés du début jusqu'à la fin de ce projet de fin d'études.

Nous tenons également à remercier tous les responsables de département de d'Electronique de l'université de Mohamed El-Bachir El-Ibrahimi «Bordj Bou Arreridj», pour toute l'aide et l'accompagnement qu'ils nous ont fourni durant toute la période de notre cursus.

Finalement, on remercie également nos professeurs pour la qualité de enseignement qui 'ils nous ont prodigué au cours de ces années passées à l'université de Bordj Bou Arreridj. Enfin, on adresse nos sincères remerciements aux membres du jury. On les remercie pour leur patience ainsi que pour leur lecture attentive de ce travail.

Merci à vous tous !

DEDICACES

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

Mes amis, AREZKI, MOHAMMED, MOURAD, BADREDDINE, pour leur réconfort moral et leur disponibilité, à ceux qui étaient toujours présents pour moi, je les remercie pour leur motivation et surtout de ne jamais perdre l'espoir !

Merci beaucoup !

SOMMAIRE

Résumé.....	I
REMERCIEMENTS	II
Dédicaces	III
SOMMAIRE	IV
Liste des tableaux	VII
Liste des figures	VIII
Introduction Générale.....	1

Chapitre I : « Classifieurs et systèmes multi-classifieurs »

I.1	Introduction.....	3
I.2	Classification.....	3
	I.2.1 Classification supervisé.....	3
	I.2.2 Classification non supervisé.....	4
I.3	Classifieurs.....	4
I.4	Methode de Classification.....	4
	I.4.1 Classification par arbre de décision	4
	A. Definition.....	4
	B. Les algorithmes automatiques pour construire les arbres de décision.....	4
	I.4.2 Classification KNN.....	9
	A. Definition.....	9
	C. Etape de la classification KNN.....	9
	D. Différentes distances utilisés	10
	I.4.3 Classification ANN	11
	A. Definition.....	11
	B. Algorithme de réseau de neurones artificiels	12
	I.4.4 Classification Discriminante	13
	A. Definition.....	13
	B. Analyse Discriminante Linéaire(LDA)	13
	I.4.5 Taux de classification.....	14
I.5	Autre méthodes de classification.....	15

I.6 Méthodes de combinaison des classifieurs	15
I.6.1 Combinaison séquentielle	15
I.6.2 Combinaison parallèle.....	16
A. Approches de combinaison parallèle.....	16
B. Fonction de combinaison parallèle sans apprentissage si la sortie est de type « classe »	16
C. Fonction de combinaison parallèle avec apprentissage si la sortie est de type « classe »	17
I.7 Conclusion	18
Chapitre II : «Proposition des systèmes de classification et de combinaison des classifieurs pour l'identification des appareils électrique.»	
II.1 Introduction	20
II.2 Principe de fonctionnement des systèmes de classification des appareils électriques	20
II.2.1 Détection d'événements	21
II.2.2 Extraction des paramètres.....	21
II.2.3 Méthodes de classification des appareils électriques	21
II.2.4 Évaluation des performances	22
II.3 Structure de classification du système proposé	22
II.4 Organisation de la base de données PLAID	23
II.5 Extraction des paramètres statistiques.....	24
II.6 Architecture du système multi classifieurs proposé	25
II.7 Paramètres d'évaluation de combinaison.....	26
II.8 Conclusion	27
Chapitre III : «Résultats d'implémentation du système multi classifieurs proposé.»	
III.1 Introduction	29
III.2 Implémentation du système de classification.....	29
III.3 Résultats de classification	30
III.3.1 Classification par TREECLASS	30
III.3.2 Classification par le classifieur ANN.....	31
III.3.3 Classification par le classifieur ADI.....	32
III.3.4 Classification par le classifieur KNN	33
III.4 Implémentation du système multi-classifieur (SMC).....	34
III.4.1 Parties fonctionnelles du SMC	35
III.4.2 Fonction de combinaison ou méthode de MULTINOMILE.....	35
III.5 Résultats de combinaison	36

III.6 Conclusion.....	38
Conclusion générale	40
Références bibliographiques	41

Liste des Tableaux

Tableau II.1 : Résumé des appareils trouvés dans la base PLAID	24
Tableau III-1 : Variation de taux de classification avec le classifieur TREECLASS..	31
Tableau III-2 : Variation de taux de classification avec le classifieur ANN.	32
Tableau III-3 : Variation de taux de classification avec le classifieur ADI	33
Tableau III-4 : Variation de taux de classification avec le classifieur KNN	34
Tableau III.5 : Taux de classification des vecteurs après combinaison des classifieurs.	37
Tableau III.6 : Taux de classification des signaux après combinaison des classifieurs.	37
Tableau III.7 : Comparaison des performances de deux SMCs pour l'identification des appareils électriques.	38

Liste des Figures

Figure I.1: L'arbre de décision en utilisant l'algorithme ID3	5
Figure I.2: Principe de fonctionnement de l'algorithme K-ppv.	9
Figure I.3 : Fonctionnement de KNN [5]	10
Figure I.4 : Fonctionnement de la classification ANN	11
Figure I.5 : Schéma de combinaison séquentielle de classifieurs [9] [10] [11].	15
Figure I.6 : Combinaison parallèle de classifieurs [9] [10] [11] [12].	16
Figure I.7. Arbre binomial.....	18
Figure I.8. Arbre multinomial	18
Figure II.1 : Schéma général d'un système d'identification des appareils électriques.	21
Figure II.2 : Structure générale du système d'identification proposé.....	23
Figure II.3 : Deux exemples de signaux de la base PLAID.....	24
Figure II.4 : Architecture du système multi classifieurs proposé	26
Figure III.1 : Implémentation du système de classification.....	30
Figure III.2: Fenêtre des résultats de simulation de classification TREECLASS avec un nombre de vecteurs = 150.	31
Figure III.3: Fenêtre des résultats de simulation de classification ANN avec un nombre de vecteurs = 80.	32
Figure III.4: Fenêtre des résultats de simulation de classification ADI avec un nombre de vecteurs = 20.	33
Figure III.5: Fenêtre des résultats de simulation de classification KNN avec un nombre de vecteurs = 50	34
Figure III-6: Architecture du système multi-classificateur proposé	35

Introduction Générale

Récemment, de nombreuses recherches se sont développées sur la gestion intelligente de la consommation d'énergie des appareils électriques domestiques. Une partie de ces recherches a été menée sur la conception des systèmes de classification d'appareils électriques, dans le but de réduire la consommation d'énergie.

La conception de tels systèmes est basée sur l'implémentation d'une ou plusieurs méthodes de classification, dont le rôle est de permettre d'affecter des identités (ou classes) aux mesures qui appartiennent à une base de données contenant les courants des appareils électriques.

Notre travail s'inscrit dans la conception de système multi-classificateur (MCS), qui concentre sur la façon de combiner plusieurs classificateurs pour obtenir un bon système de classification. Il existe deux manières principales de rechercher des systèmes MCS. Le premier est de savoir comment créer des combinaisons uniques, et le second vise à trouver différents types de combinaisons pour obtenir de meilleures performances.

Afin de mettre en œuvre une structure à ce système, d'identifier les appareils électriques et d'améliorer les performances de classification; Notre contribution portera d'abord sur l'étude, l'implémentation et la vérification de quatre types de classifieurs séparément. Ensuite, un système multi-classifieur sera conçu en sélectionnant une architecture et en implémentant une méthode de combinaison.

Ce manuscrit est divisé en trois chapitres :

Le premier chapitre présentera des généralités sur la classification et les classifieurs et donnera plusieurs structures de combinaison des classificateurs, d'une part, et plusieurs fonctions de combinaison, d'autre part.

Le deuxième chapitre couvrira les derniers développements dans le domaine d'identification des appareils électriques, avec exposition des systèmes de classification et de combinaison proposés.

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

Chapitre I :
Classifieurs et systèmes
Multi classifieurs

I.1 Introduction

La reconnaissance de formes est une branche du traitement d'image numérique et de l'intelligence artificielle. La recherche et les techniques dans cette science visent à trouver ou à développer des techniques de reconnaissance de motifs ou de structures spécifiques dans des images numériques.

La classification est effectuée à l'aide d'un modèle obtenu à l'aide d'une procédure d'apprentissage. Selon le type d'apprentissage utilisé, il existe deux catégories de classification, l'une utilisant l'apprentissage supervisé et l'autre utilisant l'apprentissage non supervisé. L'apprentissage supervisé utilise un ensemble d'exemples auxquels les étiquettes de classe sont déjà attribuées.

L'apprentissage non supervisé tente de trouver des structures inhérentes dans les données. L'apprentissage semi-supervisé utilise un petit nombre de données étiquetées et un grand nombre de données non étiquetées pour apprendre le classifieur.

I.2 Classification

La classification est une étape importante dans l'analyse des données. Elle consiste à regrouper ces données en classes homogènes. 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 : 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.

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 [1].

I.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.

I.4 Méthodes de classification

I.4.1 Classification par Arbre de décision :

A. Définition : Un arbre de décision est un modèle très simple. Etant donnée plusieurs caractéristiques, la décision se commence par un de ces caractéristiques; si ce n'ai pas suffisant, on utilise une autre, ainsi de suite. Il est largement connu et utilisé dans de nombreuses entreprises pour faciliter le processus de prise de décision et l'analyse des risques. Il a été largement utilisé dans les années 1960-1980 pour la construction de systèmes experts. Les règles sont introduites manuellement, pour cette raison ce modèle a perdu sa popularité après les années 80. L'apparition des méthodes mathématiques pour construire les arbres de décision fait revenir ce modèle à la bataille des algorithmes de l'apprentissage automatique [2].

B. Les algorithmes automatiques pour construire les arbres de décision:

B.1 Algorithme ID3 : L'algorithme ne fonctionne que sur des caractéristiques nominales. Donc, si on a des caractéristiques continues, il faut appliquer la discrétisation. Aussi, il est utilisé pour le classement seulement [3].

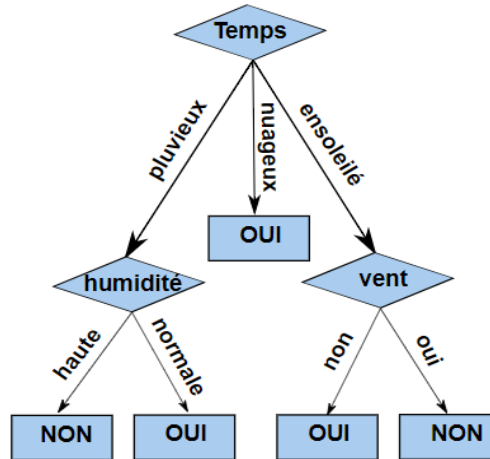


Figure I.1: L'arbre de décision en utilisant l'algorithme ID3.

Cet algorithme utilise la fonction entropie et le gain d'information pour décider quelle est la meilleure caractéristique. Étant donné un ensemble de classes C , l'entropie de l'ensemble de données S est exprimée par:

$$H(S) = \sum_{c_i \in C} -P(c_i) \log_2 P(c_i)$$

Où (en divisant le nombre des échantillons d'une certaine classe sur le nombre de tous les échantillons dans les données d'entraînement):

$$P(c_i) = \frac{|c_i|}{|S|}$$

Étant donné un vecteur de caractéristiques \vec{f} , en utilisant les valeurs d'une caractéristique f_j , on peut diviser l'ensemble de données S en plusieurs sous-ensembles groupés dans un ensemble S_j .

Le gain d'information est mesuré en se basant sur la différence entre l'entropie originale de S et celle après sa division en se basant sur une caractéristique f_j .

$$IG(S, f_j) = H(S) - \sum_{S_{jk} \in S_j} P(S_{jk})H(S_{jk})$$

$$\text{Où: } P(S_{jk}) = \frac{|S_{jk}|}{|S|}$$

La caractéristique ayant plus de gain d'information est celle sélectionnée comme meilleure. Aussi, la valeur avec entropie null est considérée comme feuille de l'arbre.

B.2 Algorithme C4.5 : Cet algorithme est une amélioration de l'algorithme **ID3**. Parmi les améliorations:

- Transformer les caractéristiques continues (numériques) en caractéristiques nominales dynamiques.
- Les caractéristiques sans valeurs sont ignorées lors du calcul de l'entropie et de gain d'information.
- Élagage des arbres après la création.

Pour décider quelle est la meilleure caractéristique afin de diviser l'ensemble de données, cet algorithme utilise une extension du gain d'information connue par rapport au gain. Lorsqu'on a une caractéristique ait un grand nombre de valeurs, elle sera favorisée par le gain d'information. Le rapport de gain fait face au problème de biais en normalisant le gain d'informations à l'aide de l'information de division.

Etant donnée:

- C : un ensemble de classes
- S : un ensemble de donnée d'entraînement
- \vec{f} : un vecteur de caractéristiques
- S_j : un hyper-ensemble contenant des ensembles avec les mêmes valeurs de la caractéristique f_j .
- $IG(S, f_j)$: le gain d'information en divisant l'ensemble de données S avec la caractéristique f_j .

L'information de division **SI** (Split Information) peut être calculée comme suit :

$$SI(S, f_j) = - \sum_{S_{jk} \in S_j} \frac{|S_{jk}|}{|S|} * \log_2 \left(\frac{|S_{jk}|}{|S|} \right)$$

Et le rapport de gain **GR** (Gain Ratio) est calculé comme suit :

$$GR(S, f_j) = \frac{IG(S, f_j)}{SI(S, f_j)}$$

Donc, la caractéristique avec le plus grand rapport de gain sera prise comme caractéristique de division.

L'algorithme ID3 ne supporte pas les caractéristiques avec des valeurs continues; comme l'âge, le prix, etc. Tandis que l'algorithme **C4.5** introduit le support de ce type de caractéristiques, en cherchant le meilleur seuil qui peut diviser l'ensemble des valeurs d'une caractéristique en deux.

Afin de sélectionner la bonne division, on suit l'algorithme suivant à chaque fois qu'on veut comparer une caractéristique avec d'autres:

- Pour chaque valeur **V_{jk}** d'une caractéristique **f_j** :
 - Diviser l'ensemble de données **S** en deux sous-ensembles, les données avec $f_j > V_{jk}$ et celles avec $f_j \leq V_{jk}$.
 - Calculer le rapport de gain **GR** de cet ensemble en le divisant avec un seuil **V_{jk}** sur la caractéristique **f_j**
- La valeur qui maximise le rapport de gain est prise comme seuil de division.

B.3. Algorithme CART : L'algorithme **CART** est similaire à celui de **C4.5** avec quelques différences, dont on cite :

- Il supporte la régression.
- Il utilise d'autres critères pour sélectionner la meilleure caractéristique. Il essaye de minimiser une fonction de coût.
- Il utilise le pré-élagage en utilisant un critère d'arrêt.

- Il crée des arbres binaires (je ne sais pas comment il le fait avec des caractéristiques nominales ayant plusieurs valeurs).

Dans le cas de l'algorithme **CART**, la meilleure caractéristique est celle qui minimise une fonction de coût $E(S)$. Chaque caractéristique f_j doit diviser un ensemble de données sur deux sous-ensembles: gauche (S_g) et droit (S_d). Donc, celle avec la plus petite valeur de l'Indice de diversité sera choisie.

$$ID(S, f_j) = \frac{|S_g|}{|S|} * E(S_g) + \frac{|S_d|}{|S|} * E(S_d)$$

Dans le cas de classement, **CART** utilise l'index de diversité Gini pour mesurer l'erreur de classification. Etant donnée un ensemble de classes C , la fonction Gini Index de l'ensemble de donnée S est exprimée par :

$$E(S) = \sum_{c_i \in C} P(c_i)(1 - P(c_i)) = 1 - \sum_{c_i \in C} P^2(c_i)$$

Où (en divisant le nombre des échantillons d'une certaine classe sur le nombre de tous les échantillons dans les données d'entraînement) :

$$P(c_i) = \frac{|c_i|}{|S|}$$

Dans le cas de régression, on utilise la somme des carrés résiduelle. Etant donné un ensemble de données S , la somme des carrés résiduelle est calculée par :

$$E(S) = \frac{1}{|S|} \sum_{i=1}^{i=|S|} (y_i - \bar{y})^2$$

Où y_i sont les valeurs attendues, et $\bar{y} = \frac{1}{|S|} \sum_{i=1}^{i=|S|} y_i$ est la valeur de sortie estimée.

I.4.2 Classification KNN :

A. Définition : L'algorithme des k les plus proches voisins (« k -nearest neighbors» ou k NN) 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 obtenues dans l'étape de comparaison [4].

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 I.2** montre ce principe qui a pour but de trouver la valeur de la classe où l'inconnu ' x ' va être affecté [5].

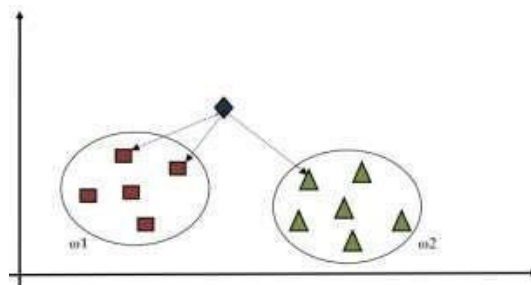


Figure I.2: Principe de fonctionnement de l'algorithme K-ppv.

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 I.3**.

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.

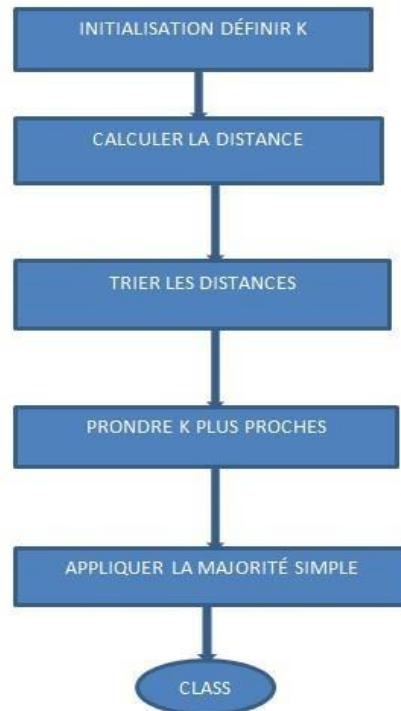


Figure I.3 : Fonctionnement de KNN [5].

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 [4]:

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

I.4.3 Classification ANN :

A .Définition : Les réseaux de neurones artificiels constituent des systèmes de traitement distribués, composés de unités de traitement interconnectées et adaptatives, caractérisées par une propension 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 [6].

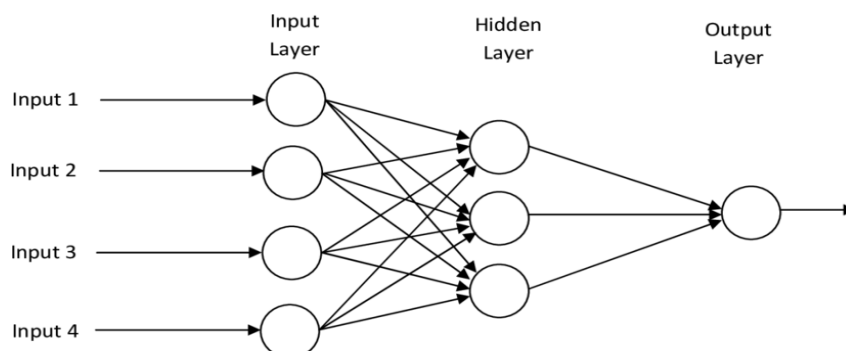


Figure I.4 : Fonctionnement de la classification ANN.

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. Les méthodes d'apprentissage sont données comme suit :

1. Apprentissage supervisé : Cette stratégie implique un formateur plus intelligent que le réseau.

2. Apprentissage non supervisé : cette stratégie est utilisée lorsqu'il n'y a pas de données d'exemple défini avec une réponse connue.

3. Apprentissage par renforcement : cette stratégie prend des décisions en fonction des commentaires de l'environnement.

Le réseau de neurones artificiels est un exemple d'apprentissage supervisé. Il a acquis les connaissances sous la forme d'unité de réseau connecté. C'est difficile pour l'homme d'extraire cette connaissance. Ce facteur a motivé l'extraction de la règle pour la classification dans l'exploration de données. 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. La formation l'échantillon est utilisée pour l'apprentissage du réseau tandis que l'échantillon de test est utilisé pour mesurer la précision du classificateur. La division de l'ensemble de données peut être effectuée par diverses méthodes telles que méthode d'exclusion, validation croisée, échantillonnage aléatoire. En général, les étapes d'apprentissage de réseau neuronal est le suivant :

- 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.

La capacité du réseau neuronal à ajuster la structure du réseau est sa capacité d'apprentissage en modifiant le poids et le rend utile dans le domaine de l'intelligence artificielle [7].

I.4.4 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 prédictes. Il existe cinq différents types de classificateurs 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 : [8]

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 :

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, c.-à-d. affecter chaque donnée de test \mathbf{x}_i à la classe \hat{z}_i ayant la plus grande probabilité a posteriori [8] :

$$r_{ik} = P(z_i = k | \mathbf{x}_i; \mu_k, \Sigma) = \frac{\pi_k f_{ik}(\mathbf{x}_i; \mu_k, \Sigma)}{\sum_{l=1}^K \pi_l f_{il}(\mathbf{x}_i; \mu_l, \Sigma)}$$

Et on a donc :

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

I.4.5 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.

I.5 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.

I.6 Méthodes de combinaison des classifieurs

I.6.1 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 I.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.

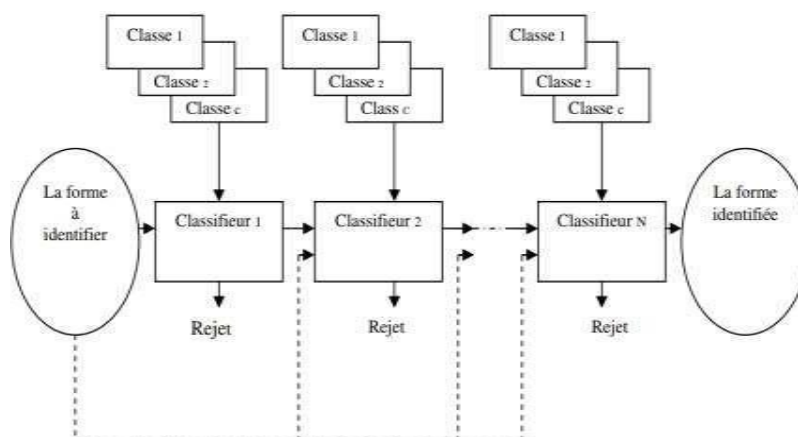


Figure I.5 : Schéma de combinaison séquentielle de classifieurs [9] [10] [11].

I.6.2 Combinaison parallèle: La figure I.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.

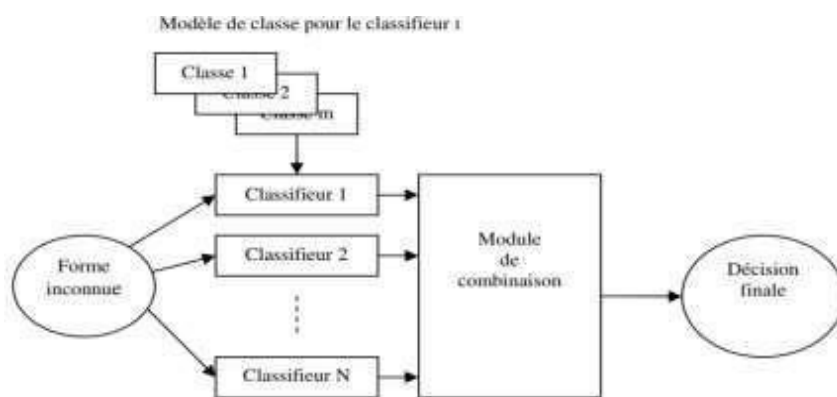


Figure I.6 : Combinaison parallèle de classifieurs [9] [10] [11] [12].

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 [12][13].
- **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 [12].

B. Fonctions de combinaison parallèle 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 [9] [12] [13]:

$$E(x) = \begin{cases} C_i \text{ si } \sum_i e(i) = \max_{c_i \in \{1..m\}} \sum_j e(j) \geq aK \\ \text{rejet sinon} \end{cases}$$

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 :

- **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.

C. Fonctions de combinaison parallèle avec apprentissage si la sortie est de type « classe » :

C.1 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 C_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.

C.2. Méthode multinomiale : La méthode multinomiale est également l'une des méthodes en treillis pour prédire le mouvement des prix des actifs à partir d'aujourd'hui jusqu'à l'échéance. Pour mieux comprendre le modèle multinomial, voici ce qui suit des différences entre l'arbre binomial et l'arbre multinomial du prix des actifs S_0 , introduit dans [14].

Le treillis multinomial est la transformation du treillis binomial en supprimant certaines périodes intra journalières. Dans cette méthode, il faut une certaine probabilité de hausse et de

baisse du prix des actifs pour chaque nœud. La probabilité peut être calculée en utilisant l'équation ci-dessous :

$$P_{n,m} = \binom{n}{m} p^{n-m} (1-p)^m, m = 0, 1, 2, \dots, n$$

Avec : $P_{n,m}$ = probabilité pour chaque nœud, n = nombre de nœuds et m = situé du nœud.

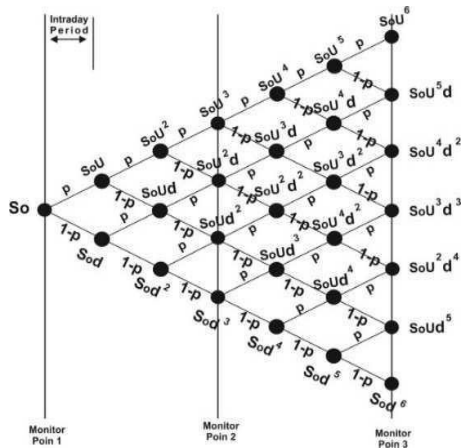


Figure I.7. Arbre binomial.

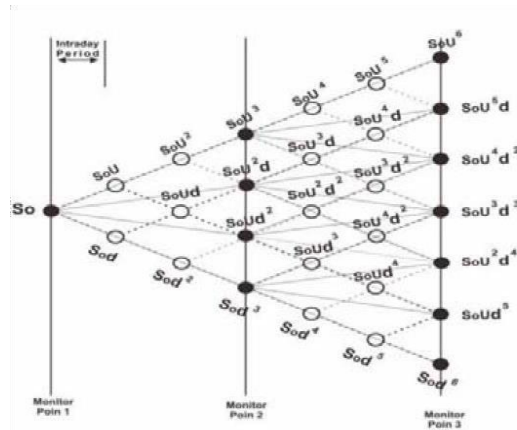


Figure I.8. Arbre multinomial.

Après avoir obtenu la probabilité pour chaque nœud, il est nécessaire de faire une modélisation du prix des actifs pour chaque nœud en utilisant un modèle multinomial, comme illustré à la **Figure I.8**.

Une définition de la méthode multinomiale destinée à la combinaison des classifieurs est donnée à la référence [15]. Cette définition est : dans cette méthode, nous estimons les probabilités a posteriori $P(m_k/s)$ pour tout $k = 1, \dots, c$ et toute combinaison de votes $S \in \Omega^L$. La probabilité a posteriori la plus élevée détermine l'étiquette de classe pour s . Alors, étant donné un $x \in R^n$, d'abord les labels s_1, \dots, s_L sont assignés par les classifieurs dans l'ensemble D , puis le label final est récupéré pour $s = [s_1, \dots, s_L]^T$

I.7. Conclusion

A travers ce chapitre, des méthodes supervisés de classification sont définies pour qu'on puisse les utilisées par la suite, pour la conception d'un système multi-classificateur, en se basant sur la méthode multinomiale pour la combinaison des classifieurs considérés.

Chapitre II :
*Proposition des systèmes de
classification et de
combinaison*

II.1 Introduction

Dans ce chapitre, on présente une proposition des systèmes de classification et de combinaison des classifieurs pour l'identification des appareils électrique en utilisant quelques méthodes de classification et une approche de combinaison des classifieurs.

II.2 Principe de fonctionnement des systèmes de classification des appareils électriques

Le fonctionnement d'un système d'identification des appareils électriques se déroule en plusieurs étapes. La première phase a pour objectif de modéliser les différentes classes d'appareils électriques en utilisant des références ou des modèles à entraîner par des algorithmes d'apprentissage standards (KNN, ANN, ADI, ...), appliqués sur des signaux d'entraînement extraits à partir d'une base de données d'apprentissage.

La deuxième phase consiste à évaluer les performances du système en utilisant des mesures de performances telles que le taux d'identification, la complexité du point de vue temps de calcul et d'espace mémoire. Cette phase exige une étape d'identification qui consiste à utiliser les références ou les modèles entraînés durant la phase d'apprentissage, pour classifier chaque signal de la base de test en utilisant des algorithmes de classification standards tels que **KNN, ADI, ANN**. [16][17]

Chaque étape nécessite une étape d'extraction de paramètres pour la transformation Chaque signal dans une série de vecteurs qui se compose chacun d'un ensemble de Paramètres extraits au cours d'une courte fenêtre d'analyse [18].

Plusieurs bases de données connues dans le domaine de la recherche scientifique, contenant des signaux électriques de courant et de tension, ont été élaborées et sont destinées à évaluer les performances du système d'identification des appareils électriques. Parmi ces bases de données, on peut citer la base de données **PLAID** [19].

Le fonctionnement du système d'identification des appareils électriques à partir du signal électrique d'entrée s'effectue en plusieurs étapes :

Détecter l'événement de démarrage de l'appareil à sélectionner.

➤ Extraire les paramètres (caractéristiques) en convertissant le signal d'entrée de Le système est dans une série de paramètres vectoriels.

➤ Classer les séquences vectorielles des paramètres extraits pour obtenir La catégorie d'appareils pour l'identification à l'aide de références ou de modèles formés pendant la phase d'apprentissage.

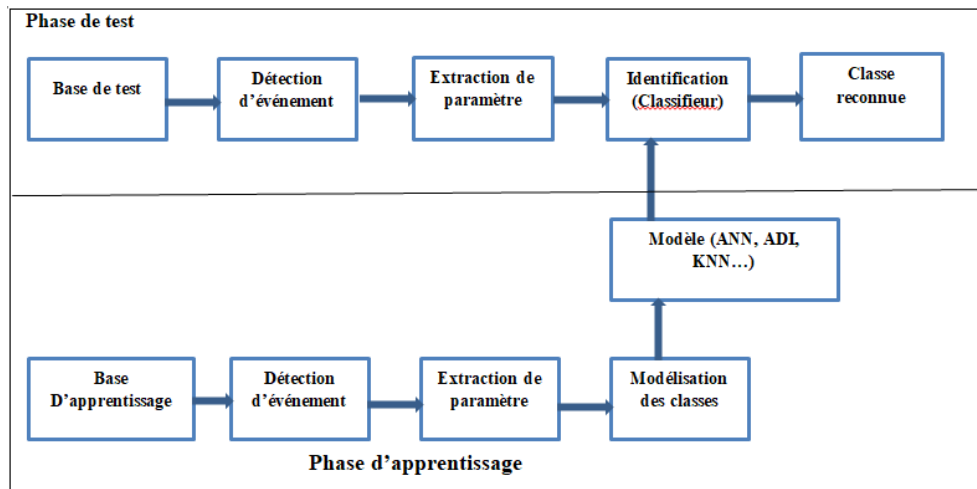


Figure II.1 : Schéma général d'un système d'identification des appareils électriques.

II.2.1) DETECTION D'ÉVÉNEMENTS

L'événement est défini par la transition qui se produit, sur le courant et/ou la tension, lorsque l'appareil est allumé (transition ON ou enclenchement) ou éteint (passage à OFF ou déclenchement). Cette étape consiste à détecter chaque démarrage ou arrêt d'un appareil qui se produit sur le réseau, c'est-à-dire à identifier les moments de prélèvement [20].

II.2.2) EXTRACTION DES PARAMÈTRES

La phase d'apprentissage et la phase de test nécessitent toutes deux une étape d'extraction de paramètres. En règle générale, cette étape divise chaque signal en une série de fenêtres d'analyse imbriquées, puis convertit chaque fenêtre d'analyse en un vecteur constitué d'un ensemble de paramètres en appliquant des techniques d'analyse telles que l'analyse de Fourier et l'analyse par ondelettes [21].

II.2.3) MÉTHODES DE CLASSIFICATION DES APPAREILS ÉLECTRIQUES

Le but des méthodes de classification est de déterminer les classes auxquelles appartiennent les objets en fonction de certains attributs descriptifs. Elle s'applique à un grand nombre d'activités humaines et est particulièrement adaptée au problème de la prise de décision automatisée. La plupart des algorithmes de classification essaient de trouver un modèle (fonction mathématique) qui explique la relation entre les données d'entrée et les classes de

sortie, et les méthodes de classification des modèles peuvent être divisées en deux types. Classification supervisée et classification non supervisée [22].

Généralement, la plupart des systèmes d'identification des appareils électriques sont basés sur des algorithmes de classification supervisés tels que KNN, ANN, ADI, etc.

II.2.4) EVALUATION DES PERFORMANCES

L'évaluation du système de classification repose généralement sur le découpage de la base de données globale en trois bases :

- Base d'apprentissage utilisée pour construire des modèles de classe.
- Base de données de validation utilisée pour définir et valider les paramètres des modèles de classe, obtenus lors de la phase d'apprentissage.
- Testez la base de données pour évaluer les performances du système de notation.

Dans certains cas, la taille de la base de données ne permet pas une participation en trois étapes, dans ce cas, seule une participation en deux parties est utilisée : la phase d'apprentissage et la phase de test [22].

II.3 Structure de classification du système proposé

Les appareils électriques peuvent être identifiés par la signature de leurs courants électriques. Cependant, le système d'identification de ces dispositifs illustré à la **Figure II.2** doit contenir :

- Une phase d'apprentissage pour apprendre et modéliser différentes classes d'appareils.
- Phase de test pour classer un ensemble de signaux de test et évaluer les performances du système d'identification. Ces deux phases nécessitent la préparation d'une base de données d'apprentissage et d'une base de tests.
- Une étape pour extraire les paramètres statistiques du signal, qui est la signature du signal électrique.
- Une étape d'extraction de la modélisation des différents types d'appareils électriques lors de la phase d'apprentissage.
- L'étape de classification est un acte ou une procédure cognitive qui permet d'attribuer une famille à laquelle il appartient (classe) l'appareil. Cette étape utilise l'un des classificateurs standards tels que : ADI, KNN, ANN, etc.

Afin de pouvoir extraire les caractéristiques des signaux, nous devons déconstruire le signal courant à l'aide de paramètres statistiques.

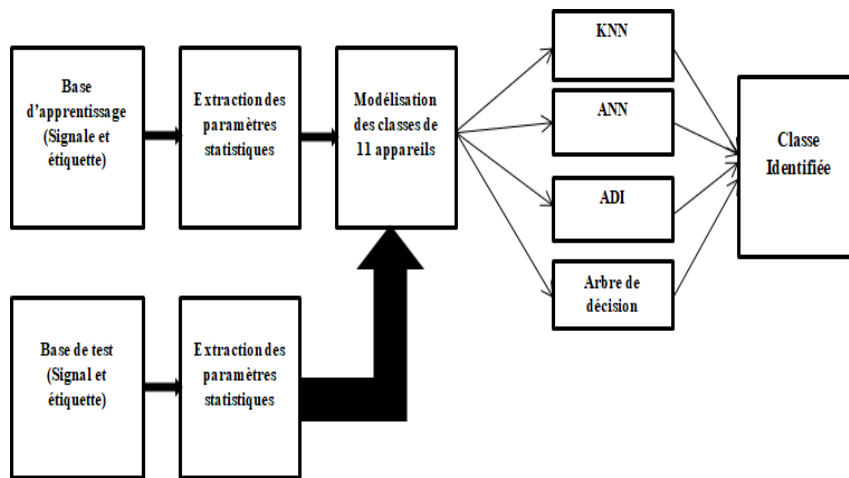


Figure II.2 : Structure générale du système d'identification proposé.

II.4 Organisation de la base de données PLAID

L'ensemble Appareil connecté Identification Data (PLAID) est une base de données publique de signatures électriques. Ces signatures sont des mesures de courants et de tensions électriques. Cette base de données contient 11 types d'appareils et chaque appareil a trois à six exemples (l'ensemble de données contient un total de 1074 signatures). Chaque signature de base de données dure quelques secondes.

Ces signaux sont échantillonnés à une fréquence de 30 kHz. L'objectif de PLAID est de fournir une bibliothèque publique de mesures d'appareils de haute précision pouvant être intégrées dans des algorithmes d'identification d'appareils Existants ou nouveaux [23].

Le tableau II.1 résume les dispositifs qui composent la base de données : les différents types d'appareils et le nombre d'exemples dans la base de données 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 II.1 : Résumé des appareils trouvés dans la base PLAID [24].

Deux exemples de signaux de PLAID sont donnés dans la figure II.3.

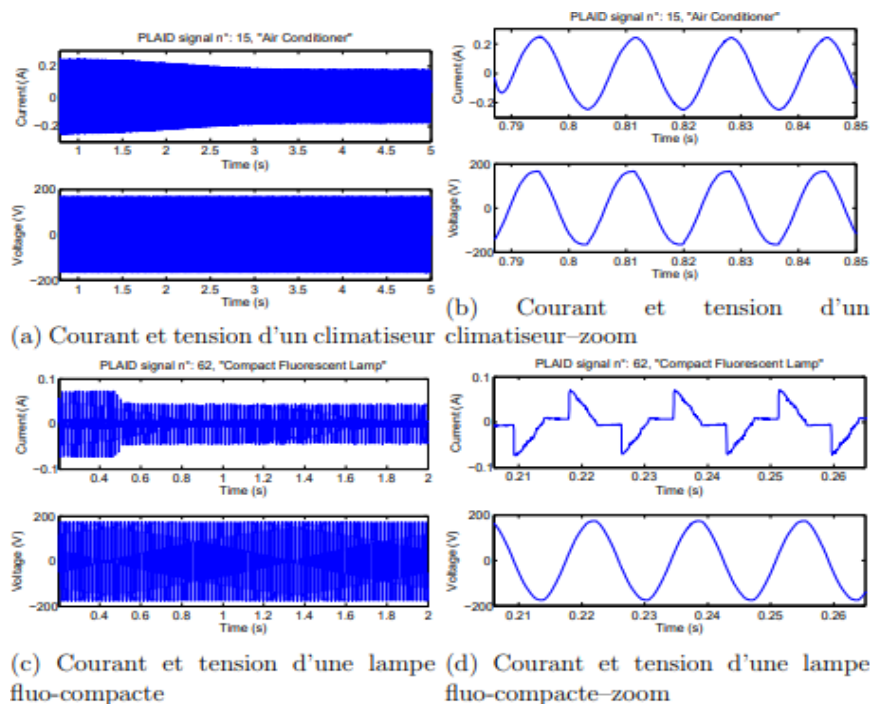


Figure II.3 : Deux exemples de signaux de la base PLAID.

II.5 Extraction des paramètres statistiques [16]

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

➤ La moyenne : $\mu = E[MH]$
est calculée par la formule : $\mu(p) = \frac{1}{R} \sum_{r=1}^R MH(p, R), p = 1, \dots, P$

➤ L'écart type : $\sigma = \sqrt{E[(MH - \mu)^2]}$
est calculée par la formule : $\sigma(p) = \sqrt{\frac{1}{R} \sum_{r=1}^R [(MH(p, R) - \mu(p))^2]}$

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

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

II.6 Architecture du système multi classifieurs proposé

L'objectif principal de nos travaux est de concevoir et mettre en œuvre un système d'identification des appareils électriques utilisant la base de données **PLAID** et les paramètres statistiques du signal.

L'idée principale est d'utiliser quatre méthodes de classification (**KNN**, **ANN**, **Decision Tree**, **ADI**)

La composition proposée (**Figure II.4**) comprend les sous-systèmes suivants :

- Règle d'entrée pour les modèles inconnus à sélectionner.
 - Classificateur KNN.
 - Classificateur ADI.
 - Classificateur ANN.
 - Classificateur d'arbre de décision.
 - L'unité composée qui donne la décision finale de sélection.
- **La base d'entrée:** On prend la règle des paramètres statistiques disposés dans les matrices d'apprentissage et de test.
- **classifieurs:** les classifieurs sont configurés en fonction de leurs paramètres optimaux décrits précédemment.
- **combinaison:** L'étape de combinaison consiste à prendre la décision finale en appliquant les fonctions du groupe aux résultats de classification fournis par chaque classifieurs.

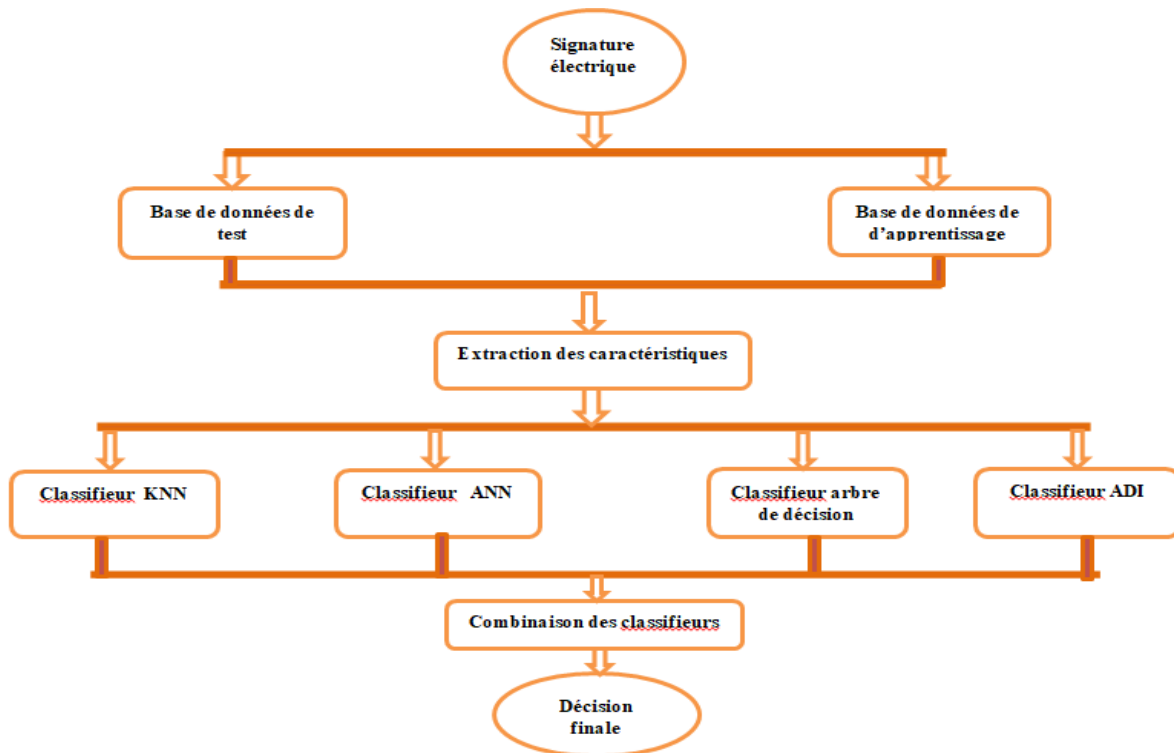


Figure II.4 : Architecture du système multi classifieurs proposé.

II.7 Paramètres d'évaluation de combinaison

a: Précision : C'est le rapport du nombre total de classifications correctes au nombre total de segment de données appliqué. La relation de ce paramètre est donnée par :

$$\text{Précision (\%)} = \frac{TP+TN}{TP+TN+FP+FN} \times 100$$

b: Sélectivité : C'est le faux rejet des segments critiques, il est calculé par l'expression suivante :

$$\text{Sélectivité (\%)} = \frac{TP}{TP+FP} \times 100$$

c: Sensibilité : La sensibilité est définie par la relation :

$$\text{Sensibilité (in \%)} = \frac{TP}{TP+FN} \times 100\mu$$

d: Spécificité : La spécificité est la détection du segment inter-critique, sa relation est :

$$\text{Sélectivité (in \%)} = \frac{TN}{TN+FP} \times 100$$

e: Taux de détection moyen : il représente la moyenne entre la sensibilité et la sélectivité :

$$\text{L'ADR} = \frac{\text{Sensibilité} + \text{Sélectivité}}{2}$$

Tel que :

(TP) : le nombre d'événements positifs correctement détectés.

(TN) : le nombre d'événements négatifs correctement détectés.

(FP) : le nombre d'événements positifs mal identifiés.

(FN) : le nombre d'événements négatifs mal identifiés.

II.8 Conclusion

Dans ce chapitre, nous avons décrit les différentes étapes de fonctionnement et les différentes phases de conception de notre système proposé. L'étape de préparation de données est décrite en introduisant la base de données PLAID et sa répartition en une base d'apprentissage et une base de test. Ensuite, l'étape d'extraction des paramètres est présentée en introduisant les paramètres statistiques. Nous avons terminé, ce chapitre, par la description de paramètres d'évaluation de combinaison.

Chapitre III: Résultats d'implémentation du système multi classifieurs proposé

III.1 Introduction

Notre travail consiste, premièrement, à implémenter 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** et le classifieur **d'arbre de décision**. Comme entrée à notre système, on introduit les paramètres statistiques (du courant électrique) extraites de la base de données **PLAID**. L'évaluation est réalisée en calculant le taux de classification.

En second lieu, on implémente un système multi-classificateur qui combine la sortie des classificateurs (KNN, ADI, ANN et arbre de décision) avec la fonction de décision **multinomiale**, dans le but d'améliorer les performances de classification.

III.2 Implémentation du système de classification

Un schéma synoptique de notre implémentation est donné à la **figure III.1**. Il se compose des étapes expliquées ci-dessous.

- Cette implémentation est basée, premièrement, 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 [16].
- 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) [16] 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.

Chapitre III Résultats d'implémentation du système multi classifieurs proposé.

→ Enfin, l'ensemble des vecteurs obtenus des deux cas: V_i ($moyenne_i$ $écart\ type_i$ $classe_i$) tel que $i = 1$ à 537, sont regroupés dans deux matrices : d'apprentissage et de test.

- L'étape de classification consiste à classifier 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, un des classifieurs suivants : KNN, ADI, ANN et Arbre de décision (TREECLASS), en respectant leurs algorithmes expliqués dans le premier chapitre.

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

- Pour évaluer les méthodes de classification utilisées, un taux de classification des vecteurs est calculé pour chaque expérience d'identification (ou de classification), selon l'expression:

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

En parallèle, un taux de classification des signaux peut être estimé selon l'équation :

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

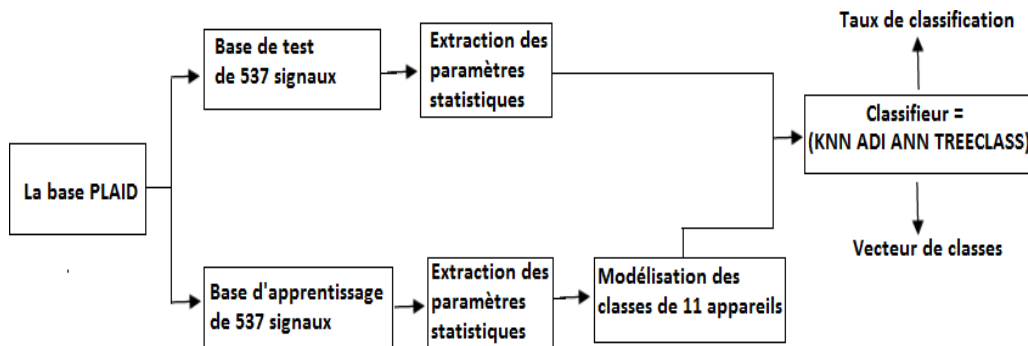


Figure III.1 : Implémentation du système de classification.

III.3 Résultats de classification

Au cours de ces expériences, quatre types de classificateurs sont utilisés pour identifier les appareils électriques par les paramètres statistiques, dans le but d'obtenir et de comparer les taux de classification. Les résultats de différentes simulations sont représentés dans les tableaux ci-dessous. Les paramètres pertinents, dans chaque tableau sont marqués.

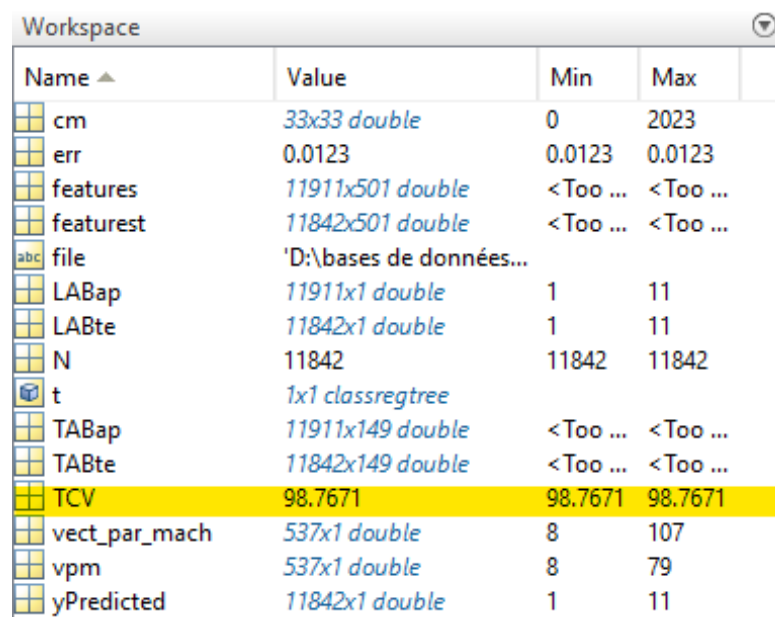
III.3.1 Classification par TREECLASS:

Chapitre III Résultats d'implémentation du système multi classifieurs proposé.

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 **98.76%**, il est obtenu pour un nombre de vecteurs égal à 150. La figure III.2 montre la fenêtre des résultats de cette dernière simulation.

Nombre de vecteurs statistiques	T _{vc} (%)
5	97,15
20	98,64
50	98,31
80	98,70
150	98,76
200	98,72

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



Name	Value	Min	Max
cm	33x33 double	0	2023
err	0.0123	0.0123	0.0123
features	11911x501 double	<Too ...	<Too ...
featurest	11842x501 double	<Too ...	<Too ...
file	'D:\bases de données...		
LABap	11911x1 double	1	11
LABte	11842x1 double	1	11
N	11842	11842	11842
t	1x1 classregtree		
TABap	11911x149 double	<Too ...	<Too ...
TABte	11842x149 double	<Too ...	<Too ...
TCV	98.7671	98.7671	98.7671
vect_par_mach	537x1 double	8	107
vpm	537x1 double	8	79
yPredicted	11842x1 double	1	11

Figure III.2: Fenêtre des résultats de simulation de classification **TREECLASS** avec un nombre de vecteurs = 150.

III.3.2 Classification par le classifieur ANN: Le temps d'exécution avec ce type de classifieur est le plus long par rapport aux autres classifieurs utilisés, dans cette implémentation. La configuration la plus optimale est la suivante :

→ La division des données introduite au programme de ce classifieur est :

- 70% pour l'apprentissage.
- 15% pour la validation.
- 15% pour le test.

→ 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.2**, on peut remarquer que les meilleurs taux de classification de vecteur et des signaux sont obtenus lorsque le nombre de paramètre égal à 80.

Les valeurs les plus pertinents des deux taux d'évaluation sont marquées en jaune à la fenêtre de **la figure III.3**.

Nombre de vecteurs statistiques	T _{vc} %	T _{sc} %
5	78,42	78,39
20	81,91	82,68
50	84,04	82,86
80	90,42	90,68
150	82,56	82,49
200	77,69	76,53

Tableau III-2: Variation de taux de classification avec le classifieur ANN.

Figure III.3: Fenêtre des résultats de simulation de classification ANN avec un nombre de vecteurs = 80.

Name	Value	Min	Max
k	11	11	11
ki	13x1 double	11830	11842
kk	537	537	537
LABap	11911x1 double	1	11
LABte	11842x1 double	1	11
LABtei	13x1 double	11	11
Lc	11x1 double	1	11
Ls	537x1 double	1	537
M	7	7	7
net	1x1 network		
r	13x1 double	1	1
t	11x11911 double	0	1
TABap	11911x79 double	<Too ...	<Too ...
TABte	11842x79 double	<Too ...	<Too ...
targets	11911x1 double	1	11
Tc	11	11	11
TCS	90.6890	90.6890	90.6890
TCV	90.4206	90.4206	90.4206

III.3.3 Classification par le classifieur ADI: Les expériences sont exécutées avec la fonction discriminante de type „quadratique“. Les résultats obtenus dans le **tableau III.3** montre que le

meilleur taux de classification des signaux prend la valeur de 73.74 %. Mais, ce type de classifieur ne donne pas les bonnes valeurs de taux de classification par rapport aux autres classifieurs.

Nombre de vecteurs statistiques	Tvc %	Tsc %
5	67.27	69.08
20	72.53	73.74
50	69,92	72.62
80	70.06	71.81
150	70.06	70.39
200	70.97	71.69

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

Name	Value
D	12x1 double
features	11911x501 double
featurest	11842x501 double
file	'D:\bases de données...
i	537
ik	12x1 double
Ite	11842x1 double
j	537
k	11
ki	13x1 double
kk	537
LABap	11911x1 double
LABte	11842x1 double
LABtei	13x1 double
Lc	11x1 double
Ls	537x1 double
M	12
r	13x1 double
TABap	11911x19 double
TABte	11842x19 double
Tc	11
tk	12
Tk	11x1 double
Ts	537
TVC	72.5384
TVS	73.7430
vect_par_mach	537x1 double
vpm	537x1 double

Figure III.4: Fenêtre des résultats de simulation de classification ADI avec un nombre de vecteurs = 20.

III.3.4 Classification par le classifieur KNN: Pour simuler notre système d'identification, on utilise la configuration par défaut du classifieur KNN et qui est ($K = 1$, distance = "Euclidean"). Les résultats de cette expérience sont représentés dans le **tableau III.4**.

Ils représentent la variation des deux taux de classification en fonction de nombre de vecteurs statistiques. On peut remarquer que, le taux de classification des vecteurs est supérieur au taux de classification des signaux, **dans ce cas. Les valeurs les plus maximum obtenues sont :**

T_{vc} = 94.78 % et T_{sc} = 84.35%. Ces valeurs sont soulignées dans la fenêtre de simulation indiquée à la **figure III.5**.

Nombre de vecteurs statistiques	T _{vc} %	T _{sc} %
5	93,21	83.42
20	94,73	83,79
50	94,78	84.35
80	94,63	83,79
150	93,97	82.86
200	93,38	82,86

Tableau III-4: Variation de taux de classification avec le classifieur KNN.

Name	Value	Min	Max
features	11911x501 double	<Too ...	<Too ...
featurest	11842x501 double	<Too ...	<Too ...
file	'D:\bases de données...		
i	537	537	537
ik	13x1 double	1	13
lte	11842x1 double	1	537
j	537	537	537
k	11	11	11
ki	13x1 double	11830	11842
kk	537	537	537
LABap	11911x1 double	1	11
LABte	11911x1 double	1	11
LABtei	13x1 double	11	11
Lc	11x1 double	1	11
Ls	537x1 double	1	537
M	13	13	13
r	13x1 double	1	1
TABap	11911x49 double	<Too ...	<Too ...
TABte	11842x49 double	<Too ...	<Too ...
Tc	11	11	11
TCS	84.3575	84.3575	84.3575
TCV	94.7813	94.7813	94.7813

Figure III.5: Fenêtre des résultats de simulation de classification KNN avec un nombre de vecteurs = 50

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

L'architecture du système de combinaison ou le système multi-classificateur est donnée à la **figure III.6**. C'est une architecture de combinaison parallèle et qui représente l'approche la plus utilisée dans la littérature. 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.

Notre implémentation contient les parties fonctionnelles suivantes :

- Une base des formes à identifiées à l'entrée du système.
- Les classifieurs parallèle : arbre décision, ANN, ADI et KNN.
- Unité de combinaison qui donne la classification finale.

III.4.1 Parties fonctionnelles du SMC

- **Base des formes à identifiées** : cette base se compose des deux matrices d'apprentissage et de test utilisées pour valider notre système de classification.
- **Les classifieurs** : les quatre classifieurs sont configurés par leurs paramètres optimaux utilisés précédemment.
- **Unité de combinaison**: Cette unité représente le cœur de conception de notre système (SMC). Elle nécessite la programmation de l'algorithme de fonctionnement de la méthode **MULTINOMIALE**. Cette méthode permet de combiner les sorties des classifieurs de type „Classe“ selon l'approche parallèle présentée au premier chapitre.

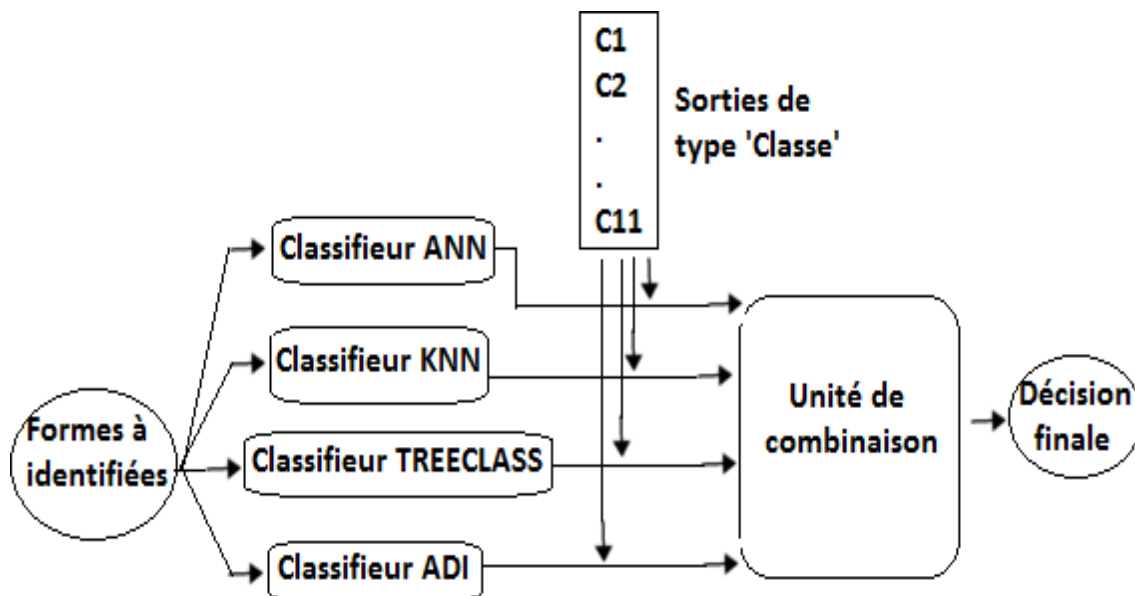


Figure III-6: Architecture du système multi-classificateur proposé.

III.4.2 Fonction de combinaison ou 'méthode de MULTINOMIAL' :

La combinaison multinomiale ou le combinateur "Behavior Knowledge Space" (BKS) est le combinateur optimal pour toutes les dépendances entre les sorties du classificateur [15].

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

Début d'algorithme

/*Résultats de la classification parallèle */

1. Obtenir un tableau de référence $E_{(N \times 4)}$ avec des sorties individuelles de 4 classifieurs pour N entrées de Test. L'entité $e(i, j)$ est la classe attribuée par le classifieur D_j à l'entrée de test i .
2. Un tableau $T_{(N \times 1)}$ de correctes classes (la référence) est également fourni.
3. Trouvez la classe prédominante dans le tableau T , appelé m_p .

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

1. Trouvez les classes attribuées à cette entrée de test, en exécutant les 4 classifieurs considérés et placez-les dans une vectrice ligne appelé „ r ”.
2. Comparez „ r ” avec chaque ligne du tableau de référence. Rangez dans un ensemble S les classes des entrées de test, dont les lignes correspondent à „ r ”.
3. S'il n'y a pas de correspondance ($S = \emptyset$), attribuez la classe m_p à la nouvelle entrée de test. Sinon, attribuez la classe prédominante dans S . S'il y a une égalité de classe, choisissez au hasard parmi les classes égales.
4. Maintenir la classe de nouvelle entrée.

Fin.

III.5 Résultats de combinaison

Les **tableaux III.5** et **III.6** illustrent les changements des deux taux de classification selon le nombre de vecteurs statistiques et ceci en appliquant notre fonction de combinaison. On remarque sur le **tableau III.6** que le taux d'identification des signaux égal à 100% en utilisant le classifieur TREECLASS, quel que soit le nombre de vecteurs appliqué à l'entrée. D'après ces résultats, on voit bien que les meilleurs taux sont enregistrés pour un nombre de 100 vecteurs statistiques.

On peut observer que quelle que soit la taille de signaux d'entrée considérée, la combinaison permet l'obtention de performances de reconnaissance acceptable et qui varie entre 90.48% et 93.37% pour la reconnaissance des vecteurs et entre 91.80 et 95.15% pour la reconnaissance des signaux.

Il est intéressant de constater que la combinaison des quatre classifieurs, dont deux donne un taux de classification $< 88\%$, permet une amélioration de ce taux d'identification.

Afin de situer les résultats obtenus, une comparaison avec le système multi-classificateur implémenté dans la référence [24] peut être effectuée. Les deux systèmes exploitent la même base de forme à identifier. La comparaison des deux résultats est donnée

Chapitre III Résultats d'implémentation du système multi classifieurs proposé.

au tableau III.7. On remarque qu'il existe une différence de 1.52 % entre les deux systèmes dans le taux de classification de vecteurs, et on peut dire que la combinaison bayésienne a enregistré les meilleures performances.

Enfin, on peut conclure que les règles de combinaison utilisées par les deux systèmes, sont valables puisque elles influent considérablement sur les résultats finaux obtenus.

Classifieur Nombre de vecteurs de test	TREECLASS	ANN	ADI	KNN	Taux TVC de COMBINAISON
5	97,15	73,32	67,27	93,21	90,48
25	98,62	86,19	71,65	94,37	92,23
50	98,31	80,77	69,92	94,78	91,52
100	98,58	75,19	70,18	94,44	93,37
150	98,76	79,37	70,06	93,97	92,79

Tableau III.5 : Taux de classification des vecteurs après combinaison des classifieurs.

Classifieur Nombre de vecteurs de test	TREECLASS	ANN	ADI	KNN	Taux TSC de COMBINAISON
5	100	75.41	69.08	94.59	91,80
25	100	87.70	72.06	94.97	93,85
50	100	82.86	72.62	96.08	93,29
100	100	77.28	72.06	95.34	95,15
150	100	81.00	70.39	94.78	94,59

Tableau III.6 : Taux de classification des signaux après combinaison des classifieurs.

Système	Base de données	Caractéristique	Classifieurs utilisés	Fonction de combinaison	TVC%	TSC %
Système 1, [2021]	PLAID	Paramètres statistiques	KNN, ADI et ANN	Bayésienne	94.89	96.08
Système 2, [2022]	PLAID	Paramètres statistiques	KNN, ADI, ANN et TREECLASS	Multinomiale	93.37	95.15

Tableau III.7 : Comparaison des performances de deux SMCs pour l'identification des appareils électriques.

III.6 Conclusion

Dans ce chapitre, on a implémenté un système d'identification des appareils électriques basé sur la combinaison des classifieurs. Le système opère en trois grandes phases : prétraitement de la base d'entrée, classification et combinaison.

On a introduit deux matrices d'apprentissage et de test qui contiennent des vecteurs de paramètres statistiques, dans l'objectif de tester la fonctionnalité de notre système d'identification avec un type de classifieur. Plusieurs expériences sont exécutées et les résultats obtenus ont été satisfaisants, en termes de taux de classification.

Ensuite, des simulations avec notre système multi-classificateur ont été menées, elles ont montré l'utilité de la combinaison des classifieurs ; Vu qu'elles sont parvenues à enregistrer des taux de classification variant entre **91.80 et 95.15%** malgré qu'on a deux classifieurs dans notre architecture qui donnent des taux entre les **70 %** et les **80%**.

Conclusion générale & Références bibliographique

Conclusion générale

A travers ce travail, on a implémenté un système de classification pour les appareils électriques qui se base à chaque fois sur un type de classificateur : KNN, ANN, ADI puis Arbre de décision. Afin de valider et tester notre système, on a introduit deux matrices d'apprentissage et de test qui se composent des paramètres statistiques extraites de la base de données des courants électriques PLAID. Des taux de classification acceptables sont enregistrés.

Les taux maximaux sont obtenus à l'aide du classificateur **d'arbre de décision** avec des valeurs égales à **98,76%**. La prise en compte des quatre autres classificateurs était nécessaire pour concevoir un système multi classificateur. Nous avons pu intégrer les classificateurs en considérant l'approche parallèle comme une structure pour réaliser le système de combinaison et en utilisant la méthode **MULTINOMIAL** comme la fonction de liaison entre les sorties des classificateurs. Les résultats obtenus en termes de taux de classification des vecteurs et des signaux prennent respectivement des valeurs maximales de **93.37% et 95.15%**.

Ce travail nous a permis également d'acquérir des informations et des compétences et de découvrir un des domaines de recherche récent.

Cette contribution peut être poursuivi par :

- L'implémentation d'autres structures et d'autres méthodes de combinaison.
- L'utilisation de ces systèmes de classification et de combinaison pour d'autres applications dans le domaine d'informatique, d'électronique ou biomédicale.

Références bibliographique

- [1] 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.
- [2] <https://becominghuman.ai/decision-trees-in-machine-learning-f362b296594a>
- [3] <https://towardsdatascience.com/a-guide-to-decision-trees-for-machine-learning-and-data-science-fe2607241956>
- [4] 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.
- [5] Izabela Moise, Evangelos Pournaras, Dirk Helbing ; K-Nearest Neighbour Classifier ;DGESS.
- [6] Hamada, M.; Hassan, M. Artificial neural networks and particle swarm optimization algorithms for preference prediction in multi-criteria recommender systems. Inform. Multidiscip. Dig. Publ. Inst. 2018,5, 25.
- [7] K. Jain, J. Mao, and K. M. Mohiuddin, “Artificial neural networks: A tutorial,” Computer, vol. 29, no. 3, pp. 31–44, 1996.
- [8] Faïcel Chamroukhi Analyse discriminante par Classification supervisée ;Université de Caen Statistics & Data Science Lab of Mathematics Nicolas Oresme.
- [9] Amirouche Radia, „Une combinaison de classifieurs pour la reconnaissance des visages humains“, mémoire de magister, Université de Annaba, 2006.
- [10] 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.
- [11] YANNICK OUFELLA, “Evolution du concept de front roc et combinaison de classifieur“, mémoire de MASTER GÉNIE INFORMATIQUE, Université de Rouen, France, 2008.
- [12] Leila CHERGUI, „Combinaison de classifieurs pour la reconnaissance de mots arabes manuscrits“, Thèse de doctorat en science, Université de constantine, 2013

-
- [13] H. Zouari et al., „Un panorama des méthodes de combinaison de classifieurs en reconnaissance de formes“, in Proc ; RFIA“2002, Angers, France, vol. 2, pp 499-508, 2002.
- [14] Astridnindya I P and Lesmono J D 2011 *Proc. Sem. Nas. Matematika* vol 6 (Bandung: Universitas Katolik Parahyangan) p 456.
- [15] Ludmila i. Kuncheva, „COMBINING PATTERN CLASSIFIERS“, 2014, Book Published by John Wiley & Sons, Inc.
- [16] Fateh Ghazali, Abdenour Hacine-Gharbi, Philippe Ravier, Tayeb Mohamadi, Papier « 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) 27: 2980 – 2997.
- [17] De Paiva Penha D, Castro ARG. «Convolutional neural network applied to the identification of residential equipment in nonintrusive load monitoring systems ». In: 3rd International Conference on Artificial Intelligence and Applications; Chennai, India; 2017. pp. 11–21.
- [18] 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)
- [19] Thomas. Picon, Mohamed Nait-Meziane, Philippe Ravier, Guy Lamarque, Clarisse Novello, Jean-Charles Le Bunetel and Yves Raingeau, «Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification », arXiv:1611.05803v1 [cs. OH] 17 Nov 2016.
- [20] Nabil Amirach « Détection d’évènements simples à partir de mesure sur courant alternatif » thèse de doctorat, Spécialité : Traitement du Signal, Université de Toulon, France.10 Juin 2015.
- [21] Fateh Ghazali, Abdenour Hacine-Gharbi, Philippe Ravier « Statistical features extraction based on the discrete wavelet transform for electrical appliances identification » international conference on intelligent system and pattern recognition conference ISPR '20, October 16–18, 2020, Virtual Event, Tunisia.

[22] Application des algorithmes d'apprentissage automatique pour la détection de défauts de roulements sur les machines tournantes dans le cadre de l'Industrie 4.0, WendBenedo Arnaud Bienvenue ZOUNGRANA, 2020. Québec, Canada

[23] Gao J, Giri S, Kara EC, Berges M. « Plaid: A public dataset of high-resolution electrical appliance measurements for load identification research: demo abstract ». In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings; New York, NY, USA; 2014. pp. 198–199.

[24] : Bensefia Adem, „Identification des appareils électriques par combinaison des classifieurs“, Mémoire de Master, *Université de Bordj Bou Arreridj*, 2021.