

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 :

Électronique Industrielle (MCIL)

Réalisé par :

Regaigui Abir

Bensbaa Nour El Houda

Intitulé

Identification des appareils électriques basée sur le classificateur KNN combiné avec la règle de vote

Encadreur: Mr Hacine Gharbi Abdenour

Co-encadreur: Mr Ghazali Fateh

Année Universitaire 2019/2020

Résumé

Ces dernières années, la demande croissante de l'énergie électrique exige aux distributeurs d'électricité de réfléchir sur des solutions pratiques pour la bonne gestion de la consommation énergétique. L'installation des compteurs électriques intelligents devient une nécessité pour bien maîtriser la consommation et d'économiser de l'énergie. Une fonction importante de ces compteurs consiste à fournir des informations en temps réel sur l'identité de l'appareil en état de fonctionnement ainsi que l'intervalle et la quantité de l'énergie consommée. Dans cette optique, nous avons proposé un système d'identification des appareils électriques domestiques basée sur l'analyse harmonique de Fourier et l'algorithme de classification KNN combiné avec la stratégie de règle de vote. Les performances du système sont évaluées en termes du taux de classification des signaux du courant électrique en utilisant la base de données COOLL.

Les résultats nous ont montré de bonnes performances avec un taux de classification de 100% obtenu en utilisant l'algorithme KNN configuré avec un nombre k de vecteurs voisins égal à 1 et le type de distance 'Cityblock', combiné avec la règle de vote. Plus particulièrement, les résultats nous ont montré que l'extraction des vecteurs de trois paramètres d'harmoniques STFS (Short Time Fourier Séries) est suffisante pour obtenir ce taux de classification.

Abstract

In recent years, the growing demand for electrical energy requires electricity distributors to think about practical solutions for the good management of energy consumption. The installation of smart meters is becoming a necessity in order to properly control consumption and save energy. An important function of these meters is to provide real-time information on the identity of the device in operation as well as the interval and amount of energy consumed. In this context, we have proposed an identification system for household electrical appliances based on harmonic analysis and KNN classification algorithm combined with voting rule strategy. System performance is evaluated in terms of the classification rate of electric current signals using the COOLL data.

The results showed us good performance with a classification rate of 100% obtained using the KNN algorithm configured with a number k of neighboring vectors equal to 1 and the distance type 'Cityblock', combined with the voting rule. More specifically, the results showed us that the extraction of vectors of three harmonic parameters STFS (Short Time Fourier Series) is sufficient to obtain this classification rate.

ملخص

في السنوات الأخيرة فرض الطلب المتزايد للطاقة على الموردين إيجاد حلول عملية من أجل التسيير الجيد للاستهلاك الطاقوي. أصبح تركيب العدادات الكهربائية الذكية حاجة ملحة من أجل التحكم في الاستهلاك واقتصاد الطاقة.

تتمثل إحدى الوظائف المهمة لهذه العدادات في توفير معلومات في الوقت الفعلي عن هوية الجهاز قيد التشغيل بالإضافة إلى المجال الزمني وكمية الطاقة المستهلكة. في هذا السياق اقترحنا نظام لتحديد الأجهزة الكهربائية المنزلية يقوم على التحليل الترددي لفورييه وخوارزمية التصنيف KNN مرفقة باستراتيجيه قاعدة التصويت. يتم تقييم أداء النظام من حيث معدل تصنيف إشارات التيار الكهربائي باستعمال قاعدة البيانات COOLL.

تظهر النتائج أداء ممتاز مع معدل تصنيف يساوي 100% متحصل عليه باستعمال الخوارزمية KNN معدلة بحيث يكون عدد الأسهم المجاورة K يساوي 1 ويكون نوع المسافة 'Cityblock', جنباً إلى جنب مع قاعدة التصويت. بصفة خاصة تظهر لنا النتائج ان استخراج اسهم مكونة من 3 عناصر توافقية STFS فقط كافية للحصول على معدل التصنيف هذا.



Remerciements

*Tout d'abord, Nous tenons à exprimer un profond remerciement à notre promoteur à Monsieur **Hacine Gharbi Abdenour** Pour l'aide inestimable, la disponibilité, la compréhension, la gentillesse, les conseils et les encouragements avec lesquelles il a bien voulu diriger ce travail, il était vraiment responsable compétant et patient.*

*Nous adressons tous nos remerciements les plus chaleureux à **Mr. Ghazali Fateh** pour son disponibilité par son assistance et pour son précieux suivi et l'intérêt.*

Nous nous acquittons, volontiers d'un devoir de gratitude et de remerciement à tous nos enseignants pour les enseignements qu'ils ont bien voulu nous donner durant notre cursus. Ils ont fait de leurs mieux afin de nous fournir un enseignement efficient et nous espérons que nos conduites.

Nos remerciements s'adressent aussi au président de jury et à ses membres qui ont accepté de juger notre travail.

Dédicaces

Je dédie Ce Modeste travail à

*Je remercie Allah le tout puissant de m'avoir donné le courage
jusqu'à l'achèvement de ce mémoire*

*A l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et
de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, que dieu te garde
dans son vaste paradis, à toi mon papa.*

*A la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma
vie et mon bonheur, tu a fait plus qu'une mère puisse faire pour que ses enfants
suivent le bon chemin dans leur vie et leurs études, aucune dédicace ne saurait
êtres assez éloquente pour exprimer ce que tu mérites pour tous les sacrifices que
tu n'as cessé de me donner depuis ma naissance, durant mon enfance et même à
l'âge adulte, a toi maman.*

*En témoignage de l'attachement, de l'amour et de l'affection que je porte pour
vous, à mon frères Abd Elhadi et mes sœurs Roukia, Thouria , Soundous*

A mes beaux frères Omar , AbdElhak

A Anges de l'âme Mosaab , Ghofrane

*A tous les membres de ma famille, petits et grands, qui m'ont toujours aidé et
encouragé, qui étaient toujours à mes cotés, et qui m'ont accompagnaient durant
mon chemin d'études supérieures, mes aimables amis, collègues d'études .*

A tous les familles : Bensbaa et Boussoualim

A ma binôme Abir

A tous ceux qui, par un mot, m'ont donné la force de continuer.



NOUREL HOUDA

Je dédie Ce Modeste travail à

Avant tout, je remercie ALLAH le tout puissant de m'avoir donné le courage et la patience pour réaliser ce travail malgré toutes les difficultés rencontrées.

Tous d'abord Je dédie ce modeste travail

À ma très cher père Aissa qui m'a encouragé par tous les moyens à avoir la confiance et l'espoir au cours

de la rédaction de ce travail. adorée et qui m'a aidé, grâce à leur prière et à leur bénédiction.

À mon cher mère

A la famille Regaigui et la famille Madhkeur

.Je le dédie aussi À ma très chers sœurs Soumaia ,Walida,ET Wiame.

A mes très cher frère Badre eldinne

À mes meilleurs amies Nour El Houda, Rouida, Lamia.

À tous les étudiants Mcil 5.



ABIR

Sommaire

Chapitre I : Etat de l'art sur les systèmes d'identification des appareils électriques domestiques

I.1 INTRODUCTION.....	3
I.2 METHODES D'IDENTIFICATION DES APPAREILS ELECTRIQUES	3
I.3 EVOLUTION DES APPROCHES NILM.....	4
I.4 PRINCIPE DE FONCTIONNEMENT DES SYSTEMES D'IDENTIFICATION DES APPAREILS ELECTRIQUES.....	6
I.4.1 Extraction des paramètres	7
I.4.2 Grandeurs caractéristiques	7
I.4.2.1 Domaine temporel	7
I.4.2.2 Domaine fréquentiel	8
I.4.2.3 Fréquence d'échantillonnage	8
I.5 ETAT DE L'ART SUR LES BASES DE DONNEES	8
I.6 CLASSIFICATION DES SIGNAUX	9
I.6.1 Classification supervisée	10
I.6.1.1 Etat de l'art sur les algorithmes de classification supervisé	11
I.7 CONCLUSION	11

Chapitre II : Système d'identification des appareils électriques proposé

II.1 INTRODUCTION	13
II.2 ARCHITECTURE DU SYSTEME PROPOSE	13
II.3 BASE DE DONNEES COOLL.....	14
II.3.1 Organisation de la base de données COOLL	14
II.3.2 Préparation de données.....	15
II.4 EXTRACTION DES PARAMETRES STFS.....	15

II.4.1 Prétraitement	16
II.4.2 Fenêtrage	16
II.4.3 Analyse de Fourier à court terme	16
II.5 ALGORITHME DE CLASSIFICATION KNN	17
II.5.1 Description de l'algorithme KNN	17
II.5.2 Configuration de classification KNN	18
II.5.3 Distances utilisées.....	18
II.6 COMBINAISON DE LA REGLE DE VOTE AVEC L'ALGORITHME DE CLASSIFICATION KNN	19
II.6.1 Mesures de performances.....	20
II.7 CONCLUSION.....	20
 Chapitre III : Implémentation et résultats	
III.1 INTRODUCTION	23
III.2 IMPLEMENTATION DU SYSTEME D'IDENTIFICATION DES APPAREILS ELECTRIQUES	23
III.3 EXPERIENCES	25
III.3.1 Configuration optimale de classificateur KNN.....	25
III.3.2 Ordre optimal des harmoniques (pair ou impair).....	26
III.3.3 Classification KNN combinée avec la méthode de règle de vote	26
III.3.4 Effet de la taille de la base de données d'apprentissage.....	28
III.3.5 Performances du système d'identification en tenant compte la classification des signaux en 42 classes d'appareils	29
III.4 CONCLUSION	31
CONCLUSION GENERALE	32
REFERENCES	34

Liste des Figures

Figure I.1. Principe de mesures de l'approche non intrusive NILM.....	2
Figure I.2. Le schéma synoptique représentant le principe de fonctionnement d'un système d'identification.....	5
Figure II.1. Organigramme générale du système d'identification proposé	13

Liste des Tableaux

Tableau I.1. Résumé des méthodes NILM organisées par fréquence d'échantillonnage.....	4
Tableau I.2. Etat de l'art sur les bases de données.....	8
Tableau I.3. Etat de l'art sur les algorithmes de classification supervisée.....	10
Tableau II.1. Résumé de la base de données.....	14
Tableau III.1. Configuration de classificateur KNN.....	26
Tableau III.2. Influence de l'ordre des harmoniques sur le taux de classificateur.....	27
Tableau III.3. Taux de classification TCS et TCV en fonction du nombre de paramètres ...	28
Tableau III.4. Effet de la taille de la base d'apprentissage	29
Tableau III.5. Taux de classification TCV et TCS en fonction du nombre de paramètres STFS, en tenant compte les 42 classes d'appareils.....	29



Introduction générale



Ces années récentes, le concept de "smart grid" ou réseaux intelligents est arrivé dans le secteur de distribution d'électricité à travers la problématique de la demande croissante de l'énergie.

L'installation des compteurs communiquant intégrés à un système de communication en réseau est la pierre angulaire des réseaux intelligents. Les compteurs électriques traditionnels ne fournissent que la quantité d'énergie totale consommée, tandis que la fonction des compteurs intelligents est le relevé en temps réel et à distance de la puissance électrique consommée ou la communication des informations de consommation aux fournisseurs. De plus, il peut aussi identifier les appareils électriques présents sur le réseau résidentiel par l'utilisation des méthodes de surveillances des courbes de charges (courant et tension) des appareils domestiques. Ces méthodes se divisent en deux catégories, intrusive et non intrusive [3].

L'utilisation des compteurs électriques intelligents dans des installations domestiques permet de mieux gérer l'offre à la demande par les technologies de l'information et de la communication.

Dans notre étude, on s'intéresse aux méthodes non intrusives qui consistent à extraire des informations telles que : l'identité de l'appareil en état de fonctionnement, l'intervalle de consommation et la quantité de consommation, à partir d'un seul point de mesure du signal électrique au niveau du compteur.

Généralement, le fonctionnement d'un système d'identification des appareils électriques domestiques exige une étape d'extraction de paramètres caractérisant les différentes classes d'appareils ainsi qu'une étape de classification des signaux électriques. La conception d'un tel système se base sur une phase d'apprentissage et une phase de test. La phase d'apprentissage a pour but de modéliser les différentes classes d'appareils, alors que la phase de test basée sur l'étape de classification permet d'évaluer les performances du système d'identification. Ces deux phases exigent une étape d'extraction de paramètres qui consiste à convertir chaque signal électrique en une séquence de vecteurs de paramètres. La conception de ce système

Introduction générale

exige ainsi, une base de données à répartir en une base d'apprentissage et une base de test. Différentes bases de données sont utilisées pour l'évaluation des systèmes d'identification des appareils électriques, telles que les bases de données PLAID et COOLL, etc. Dans les travaux de [4][5], les auteurs ont proposé des systèmes d'identification basée respectivement sur les classificateurs HMM et KNN combinés avec la méthode d'extraction des paramètres harmoniques. Ces systèmes d'identification sont évalués sous la base de données PLAID. Cette base de données est constituée de 1074 signaux échantillonnés à une fréquence de 30 Khz. Dans [4], les auteurs ont proposé l'utilisation de la règle de vote associée à l'algorithme de classification KNN, appliquée sur des séquences des vecteurs de paramètres statistiques des coefficients STFS (Short Time Fourier Series).

Notre travail consiste à proposer un système d'identification des appareils électriques basée sur l'algorithme de classification KNN associé à la règle de vote et appliqué sur des séquences de vecteurs de paramètres STFS. Cette étude consiste également à évaluer ce système sous la base de données COOLL qui est constituée de 840 signaux échantillonnés à une fréquence de 100 Khz. Les performances sont évaluées en termes de taux de classification.

Notre manuscrit est organisé en trois chapitres :

- Le premier chapitre permet de donner un état de l'art sur les systèmes d'identifications des appareils électriques domestiques. Plus particulièrement, ce chapitre présente des généralités sur le concept d'identification des appareils électriques, et décrit les différentes méthodes d'extraction des paramètres (caractéristiques) à partir du signal du courant électrique ainsi que les méthodes de classification.
- Le deuxième chapitre est consacré à la description de notre système d'identification des appareils électriques en présentant les différentes étapes de conception du système proposé.
- Le troisième chapitre a pour but d'implémenter notre système et discuter les résultats obtenus en effectuant différentes expériences pour répondre aux questions posées sur la bonne configuration du système.
- Finalement, on termine notre manuscrit par une conclusion générale.



Chapitre I

*Etat de l'art sur les systèmes d'identification
des appareils électriques domestiques*



I.1 INTRODUCTION

Les compteurs électriques ne fournissent que la quantité d'énergie totale consommée. Cette unique donnée ne permet qu'une identification sommaire des appareils électriques. Cependant les récents développements en matière de compteurs électriques « intelligents » permettent de disposer en temps réel de la consommation électrique d'un foyer, d'un bâtiment ou d'une entreprise. Ces compteurs permettent donc de disposer du profil temporel de la consommation, qui constitue un outil puissant pour l'analyse de l'électricité consommée [8].

L'identification des appareils électriques domestiques a un intérêt croissant dans ces dernières années, elle permet d'identifier des appareils domestiques parcourus par un courant électrique dans le but de communiquer des informations de consommation d'énergie de chaque appareil au consommateur ainsi qu'au fournisseur, la gestion intelligente de la consommation électrique est ainsi une des préoccupations majeures non seulement pour les gestionnaires et les fournisseurs mais aussi pour les consommateurs d'énergie [8].

I.2 METHODES D'IDENTIFICATION DES APPAREILS ELECTRIQUES

Les méthodes d'identification des appareils électriques consistent à surveiller les courbes de charges afin de définir chaque consommation par usage à partir d'une consommation générale. Fait par l'extraction des signatures qui désigne chaque courbe de charge en fonctionnement de manière intrusive ou non intrusive.

Les méthodes de suivi intrusives ILM (Intrusive Load Monitoring), consiste à installer des semi compteurs attachées à chaque appareil, qui envoient l'information à une unité centrale de traitement pour l'inclure dans le rapport de consommation générale. Il existe de nombreux problèmes résultant de cette méthode du fait que chaque appareil surveillé nécessite un capteur individuel. Premièrement, avec l'augmentation du nombre d'appareils surveillés, le coût du système de mesure augmente considérablement. Deuxièmement, les données de mesure doivent être collectées dans une unité centrale, qui est un autre élément du système. Troisièmement, les données des compteurs doivent être envoyées à une unité centrale, de sorte que les compteurs d'électricité doivent être équipés d'une interface appropriée, par exemple une interface radio. De plus, le courant traversant les appareils surveillés doit également traverser le compteur, ce qui réduit la fiabilité de l'alimentation électrique. Enfin, les appareils de mesure consomment de l'énergie [1].

La deuxième méthode est le suivi non intrusif NILM (Non Intrusive Load Monitoring) qui représente une solution pour les approches ILM, tel que la décomposition de la consommation

générale à une consommation par usage se fait de manière aveugle et non intrusive et à partir d'un seul capteur met en amont de l'installation électrique.

En contrepartie il faut des algorithmes complexes pour le traitement et l'analyse des signaux. Les méthodes des NILM doivent être capables de détecter les changements d'état provoqués par une mise en marche ou en arrêt d'une charge, ainsi que de renseigner sur la nature et le nombre des charges en présence. Les grandeurs à mesurer sont les signaux correspondant à la tension d'alimentation totale $u(t)$ et le courant total de consommation $i(t)$ [1].

L'avantage principal lié à ce type d'approche réside dans sa simplicité d'installation et son faible coût.

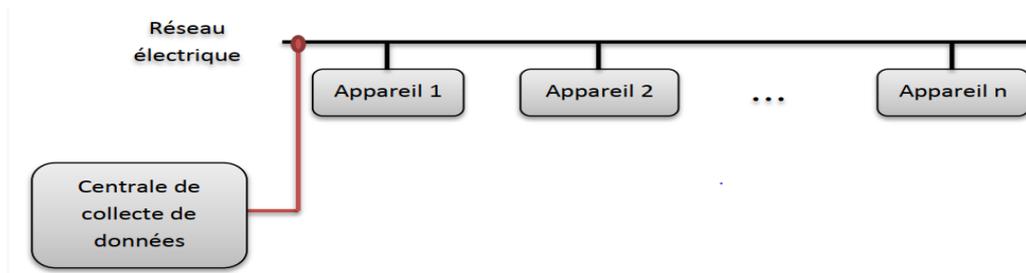


Figure I.1: principe de mesures de l'approche non intrusive NILM [1]

I.3 EVOLUTION DES APPROCHES NILM

La mise en œuvre la plus basique des NILM se base sur la variation de la puissance active. Cette approche a été étudiée par Hart [20], mais ne permet pas d'identifier certains appareils. Une première approche, développée par F. SULTANEM dans les années 90 [23, 22] basée sur la variation du courant efficace, des puissances actives et réactives. Elle est suivie par le développement d'une seconde approche par M. BONS utilisant les chaînes de Markov cachées [21] en 1994. Cette dernière permet de reconnaître la nature structurelle des appareils, ainsi que la combinaison des appareils existants dans la même installation. Par ailleurs T. ONADA [25] a utilisé les harmoniques du courant (jusqu'au 13ième) pour distinguer des charges similaires du point de vue des puissances active et réactive.

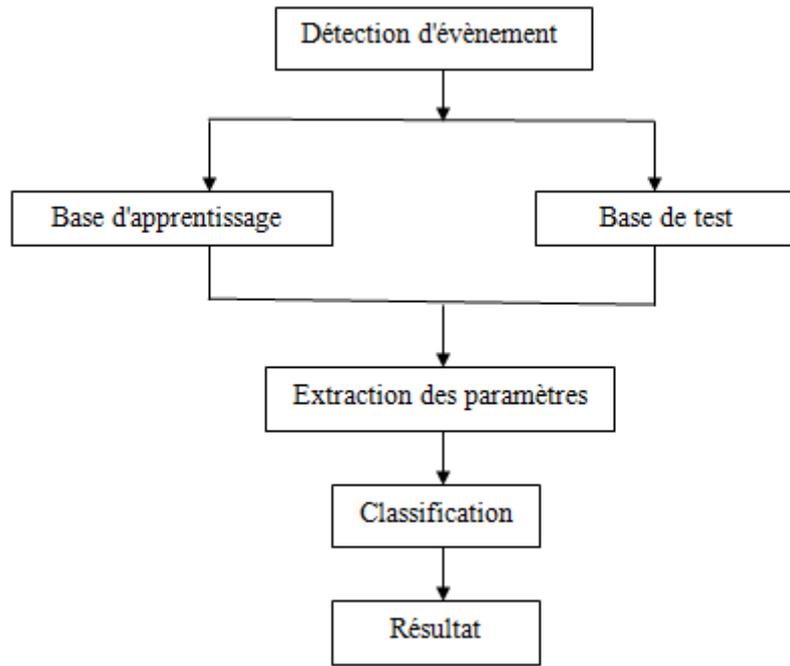
Cette technique s'est avérée performante dans les secteurs industriels. En outre, ces harmoniques renseignent sur le bon fonctionnement des appareils. Une autre méthode, développée par LEEB en 1993 [25, 26], permet d'identifier les charges à partir des formes des puissances active et réactive en régime transitoire [5]. Le tableau (I.1) nous présente un état de l'art sur la méthode NILM.

Table I-1 – Résumé des méthodes NILM organisées par fréquence d'échantillonnage [34]

Nom	Fréquence d'échantillonnage	Caractéristiques principales	Etudes
Très basse	Moins d'une fois par minute	Souvent plusieurs évènements par intervalle	(Powers et al., 1991) (Prudenzi, 2002) (Kolter et al., 2010) (Huang et al., 2013) (Curtis et al., 2014)
Moyenne	Plus rapide que 1 Hz mais plus faible que la fréquence fondamentale	Les formes des transitoires sont analysées	(Norford and Leeb, 1996) (Berges et al., 2009) (Goncalves et al., 2011) (Barsim et al., 2014) (Du et al., 2015)
Haute	Fréquence fondamentale à 2 kHz	Harmoniques d'ordre faible sont analysées	(Sultanem, 1991) (Roos et al., 1994) (Leeb et al., 1995) (Lee et al., 2005) (Srinivasan et al., 2006) (Akbar and Khan, 2007) (Wichakool et al., 2015) (Dong et al., 2013)
Très haute	Plus rapide que 2 kHz jusqu'à 40 kHz	Forme d'onde avec haut détail ou harmoniques hautes sont en général analysées	(Chan et al., 2000) (Laughman et al., 2003) (Lee et al., 2004) (Lam et al., 2007) (Shaw et al., 2008) (Suzuki et al., 2008) (Liang et al., 2010a,b) (Inagaki et al., 2011) (Wang et al., 2012) (Popescu et al., 2014)
Ultra haute	Plus rapide que 40 kHz	Bruit électrique et très hautes fréquences sont analysés	(Patel et al., 2007) (Gupta et al., 2010)

I.4 PRINCIPE DE FONCTIONNEMENT DES SYSTÈMES D'IDENTIFICATION DES APPAREILS ÉLECTRIQUES.

Le schéma synoptique de fonctionnement d'un système d'identification des appareils électriques est illustré sur la figure I.2.



FigureI.2: Le schéma synoptique qui représente le principe de fonctionnement d'un système d'identification

Le fonctionnement d'un système d'identification des appareils électriques se déroule en plusieurs étapes. Le système détecte premièrement l'événement de la mise en marche de l'appareil à identifier. Ensuite, une étape d'extraction de paramètres caractérisant les différentes classes ou signatures d'appareils, s'effectue à partir du signal du courant 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 un des classificateurs standards tel que KNN, GMM, etc. Cette classification tient en compte des modèles des différentes classes, entraînés préalablement dans une phase appelée phase d'apprentissage.

La conception d'un système d'identification des appareils électrique exige une phase d'apprentissage pour apprendre et modéliser les différentes classes d'appareils, ainsi qu'une phase de test pour classifier un ensemble de signaux de test et évaluer les performances du système d'identification. Ces deux phases exigent la préparation d'une base de données d'apprentissage et une base de données de test.

I.4.1 Extraction des paramètres

La phase d'apprentissage et la phase de test exigent chacune une étape d'extraction des paramètres. Généralement, cette étape découpe chaque signal en une séquence de fenêtres d'analyse chevauchées, ensuite elle convertit chaque fenêtre d'analyse en un ensemble de paramètres en appliquant des techniques d'analyse telles que l'analyse harmonique [Nait Meziane, 2016 et 2017, Hacine Gharbi 2020, Ghazali 2019] et l'analyse par ondelettes [Hacine Gharbi, 2018].

Les courbes de charges concerné de l'extraction des signatures sont les puissances actives et réactives, la tension ou le courant, parfois la puissance active de démarrage est la même pour deux appareils différents, en ce qui concerne la tension, elle ne représente pas une variation significative lors de changement d'état, alors elle ne donne pas une information utile pour l'identification de consommation par usage.

En plus, les caractéristiques du régime transitoire de courant électrique sont directement reliées à la tâche physique exécutée par la charge Pour ces raisons nous avons utilisé le courant électrique.

I.4.2 Grandeurs caractéristiques

Nous avons vu qu'il existe de nombreuses grandeurs, aussi bien dans le domaine temporel que fréquentiel, parmi elle, certaines vont nous servir à caractériser les évènements qui surviennent sur le réseau étudié.

Ces grandeurs seront stockées dans le vecteur d'informations, qui regroupe les grandeurs qui vont nous permettre de connaître les évènements qui ont été causés par une même charge ou type de charge. A chaque évènement est donc associé un vecteur d'informations [2].

I.4.2.1 Domaine temporel

Le vecteur d'information doit contenir l'information sur l'instant de transition, si les instants de transition sont estimés de façon peut préciser nous risquons de fournir des estimations sur la consommation qui ne correspond pas à ce qui est réellement produit sur le réseau.

Il est intéressant d'ajouter un paramètre qui indique le sens de la transition (ON ou OFF), pour savoir si l'amplitude estimée après la transition est plus faible ou plus élevée que ce qu'elle était avant la transition.

Les transitions courtes et longues sont définies par leur durée et leur forme, ces paramètres permet de connaître si l'évènement détecté a été provoqué par l'enclenchement d'un appareil résistif ou un moteur[2].

I.4.2.2 Domaine fréquentiel

La valeur de fréquence fondamentale n'est pas une grandeur caractéristique car elle ne dépend pas de l'appareil mais du fournisseur d'électricité et de perturbations qui surviennent sur le réseau. Par contre l'évolution de son amplitude complexe est une grandeur discriminante, elle permet de quantifier la transition[1].

On peut, à partir de la mesure des variations de l'admittance de l'installation, ou des variations de l'impédance de l'installation, identifier les charges. Cette méthode consiste à effectuer sur le courant électrique d'alimentation de mesures appropriées pour déterminer ces variations. La comparaison de ces variations avec une table de référence contenant les admittances et/ou les impédances des différents appareils disponibles sur le marché, permet de reconnaître les charges [5].

I.4.2.3 Fréquence d'échantillonnage

Les approches NILM classé en deux catégories en termes de fréquence:

- L'approches à basse fréquence d'échantillonnage (BFE) (1 Hz ou moins) est plus réponde à cause de mesure accessible et base de données facilement disponibles, utilise les puissances réelles et réactive comme entrée pour l'étage de détection.
- L'approche à haute fréquence d'échantillonnage (HFE) (quelques centaines de Hz au MHz, l'analyse à fréquence élevé et beaucoup plus complexe, c'est pourquoi il récent.

I.5 Etat de l'art sur les bases de données

Pour la conception du système d'identification on utilise une base de données de référence, Le Tableau (I.2) illustre un état de l'art sur les bases de données existantes.

Tableau I-2: Etat de l'art sur les bases de données à haute fréquence d'échantillonnage [3]

Références	Auteurs	La base de données	Fréquence d'échantillonnage
[25]	J. Zico Kolter Matthew J. Johnson	REDD	15 KHz
[26]	Hani jneid jeffry L. Anderson	BLUED	12KHz
[27]	Peng Gao Simon Mathew	PLAID	30KHz
[28]	Gulati	HFED	10KHz à 5MHz
[29]	Jack Kelly William Knottenbelt	UK-DALE	16 KHz
[30]	Andrew W.Robert S.Kahl	WHITED	44.1 KHz

I.6 CLASSIFICATION DES SIGNAUX

Les méthodes de classification ont pour but d'identifier les classes auxquelles appartiennent des objets à partir de certains traits ou paramètres descriptifs. Elles s'appliquent à un grand nombre d'activités humaines et conviennent en particulier au problème de la prise de décision automatisée. La procédure de classification sera extraite automatiquement à partir d'un ensemble d'exemples. Un exemple consiste en la description d'un cas avec la classification correspondante.

Un système d'apprentissage doit alors, à partir de cet ensemble d'exemples, extraire une procédure de classification, il s'agit en effet d'extraire une règle générale à partir des données observées. La procédure générée devra classer correctement les exemples de l'échantillon et avoir un bon pouvoir prédictif pour classer correctement de nouvelles descriptions[33].

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 [11].

I.6.1 Classification supervisée

L'apprentissage automatique est un domaine dont l'intérêt majeur est le développement des algorithmes permettant à une machine d'apprendre à partir d'un ensemble de données. La motivation originale de ce domaine était de mettre en œuvre des systèmes artificiels intelligents. Les algorithmes issus de ce domaine sont utilisés par plusieurs autres domaines, tels que la vision par ordinateur, la reconnaissance de forme, la recherche d'information, la bioinformatique, la fouille de données et beaucoup d'autres.

Il existe plusieurs types d'apprentissage automatique qui se distinguent essentiellement par leur objectif. Les deux types les plus connus sont : l'apprentissage supervisé et l'apprentissage non supervisé. Si l'on dispose d'un ensemble de points étiquetés, nous parlerons de classification supervisée. Dans le cas contraire, nous effectuons une classification non supervisée [10.13].

L'objectif, dans le cas d'apprentissage supervisé, est déterminé explicitement par la prédiction d'une cible. Dans ce cas, pour chaque exemple d'apprentissage x_i de Ω , une cible (étiquette) y_i est associée. La tâche d'un algorithme d'apprentissage est alors d'entraîner un modèle qui puisse prédire, pour une entrée x quelconque, la valeur de la cible y . La nature de l'ensemble de cibles (noté Y) dépendra du type de problèmes à résoudre.

Deux types de problèmes fréquents sont les problèmes de classification et de régression. Pour un problème de classification, Y correspond à un ensemble fini de classes auxquelles peuvent appartenir les différents x . Par contre pour celui de la régression, Y correspond à un ensemble de valeurs continues [33].

I.6.1.1 Etat de l'art sur les algorithmes de classification supervisé

Le tableau (I.3) illustre un état de l'art sur les algorithmes de classification supervisée.

Tableau I-3: Etat de l'art sur les algorithmes de classification supervisée [33]

Le classificateur	Etat de l'art
KNN	(Indyk et Motwani [1998])
Arbres de decision	CART(Breiman et al.[1984]), ID3 (Quinlan [1986]), C4.5(Quinlan [1993]) ...),
Séparateurs à vastes marges	V. V apnik (Vapnik [1998])
SVM	(Burges [1998])
Approche bayésien	(Jensen [2001], Neapolitan [2003])
Réseaux de Neurons	(Dietz <i>et al.</i> [1989], Broadbent et Lucas [1988])

I.7 CONCLUSION

Dans ce chapitre, nous avons présenté un état de l'art sur les systèmes d'identification des appareils électriques domestiques. Les différentes étapes de fonctionnement et les phases de conception d'un tel système d'identification sont décrites brièvement. De plus, des généralités sur l'apprentissage automatique et les méthodes de classification supervisée sont présentées. Egalement, nous avons présenté un état de l'art sur les bases de données utilisées dans l'évaluation des systèmes d'identification des appareils électriques.



Chapitre II
Systeme d'identification des appareils
électriques proposé



II.1 INTRODUCTION

Dans ce chapitre nous allons décrire notre système d'identification des appareils électriques en utilisant une méthode d'extraction de paramètres basée sur l'analyse harmonique, combinée avec l'algorithme de classification KNN associé à la règle de vote.

II.2 ARCHITECTURE DU SYSTEME PROPOSE

Le schéma synoptique de notre système d'identification des appareils électriques est illustré sur la figure (II.1). Le fonctionnement de notre système se déroule en une étape d'extraction de paramètres (caractéristiques) et une étape de classification. L'étape d'extraction consiste à convertir le signal du courant électrique en une séquence de vecteurs de paramètres, obtenue par l'application de l'analyse harmonique sur des fenêtres d'analyse chevauchées de 50%. Chaque fenêtre d'analyse est représentée ainsi par un vecteur de coefficients STFS (Short time Fourier Series). L'étape de classification consiste à appliquer premièrement l'algorithme KNN pour classer chaque vecteur en une classe de type d'appareil, ensuite appliquer la règle de vote sur la séquence des indices de classes obtenus pour décider sur la classe de signal considéré. La conception de notre système se base sur une phase d'apprentissage et une phase de test. La phase d'apprentissage a pour but de modéliser les différentes classes de types d'appareils électriques. Plus particulièrement, cette phase consiste à enregistrer pour chaque classe un ensemble de vecteurs avec leurs indices de classes. L'ensemble des vecteurs de paramètres STFS est enregistré dans une matrice appelée matrice de vecteurs d'apprentissage. Alors que les indices des classes des vecteurs sont enregistrés dans une matrice de type colonne appelée tableau d'indices d'apprentissage.

La phase de test consiste à évaluer les performances du système d'identification en terme de taux de classification des signaux. Ainsi la conception de notre système exige d'utiliser une base de données de référence et de la répartir en une base d'apprentissage et une base de test utilisées respectivement dans la phase d'apprentissage et la phase de test.

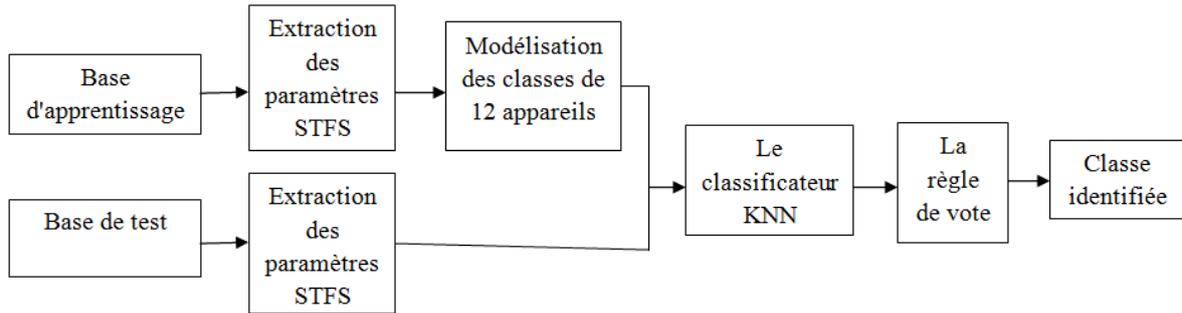


Figure II.1 : Structure générale du système d'identification proposé

II.3 BASE DE DONNEES COOLL

Dans notre travail, on s'intéresse à utiliser la base de données COOLL qui est constituée de signaux de courants et de tension, échantillonnés à une haute fréquence égale à 100Khz. Cette fréquence d'échantillonnage permet d'étudier l'analyse fréquentielle du signal jusqu'à une fréquence maximale égale à 50 Khz. Ainsi, le nombre maximal des paramètres d'harmoniques STFS est égal à 1000 paramètres en tenant compte la fréquence fondamentale de 50 hz et la bande fréquentielle de 50 Khz.

II.3.1 Organisation de la base de données COOLL

Dans la base de données COOLL, les signaux du courant et de la tension sont échantillonnés à une fréquence F_s égale à 100 kHz. Ces signaux sont mesurés sur 42 appareils électriques qui sont regroupés en 12 types. Pour chaque appareil, 20 mesures de signaux sont effectuées en contrôlant chacune par un délai d'action de mise en marche, fixé par rapport au passage par zéro de la tension du secteur, variant du 0 à 19 ms avec un pas de 1 ms. En total, la base de données COOLL contient 840 enregistrements de signaux du courant et 840 enregistrements de signaux de la tension [3]. Le tableau (II.1) présente un résumé sur les différents types d'appareils, le nombre d'appareils et le nombre de signaux par chaque type.

Chapitre II: Système d'identification des appareils électriques proposé

Tableau II-1: Résumé de la base de données COOLL [35].

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

II.3.2 Préparation de données

La base de données se divise en une base d'apprentissage et une base de test en prenant alternativement les signaux de chaque appareil. La base de données d'apprentissage est constituée des signaux du rang impair de chaque appareil, alors que la base de données de test est constituée des signaux du rang pair de chaque appareil. Ce processus de répartition permet d'obtenir 420 signaux pour chaque base de données (apprentissage et test).

II.4 EXTRACTION DES PARAMETRES STFS

L'étape d'extraction a pour objectif de convertir chaque signal d'entrée en une séquence de vecteurs constitués d'un ensemble de paramètres obtenus par l'analyse harmonique. Cette étape s'effectue en plusieurs étapes secondaires à savoir : le prétraitement, le fenêtrage, l'analyse de Fourier à court terme.

II.4.1 Prétraitement

Une étape de prétraitement de signal à pour objectif d'avoir un signal convenable pour la tâche de classification. Nous avons pris pour chaque signal l'intervalle de fonctionnement de l'appareil en extrayant le début et la fin du fonctionnement à partir des fichiers texte de la base de données COOLL.

II.4.2 Fenêtrage

Cette étape consiste à découper le signal en une séquence de fenêtre d'analyse de duré 20 ms (1/50Hz) avec un chevauchement de 50%, et à convertir chaque fenêtre d'analyse en un vecteur de N_f (nombre de coefficients) coefficients STFS. Ainsi, cette étape convertit chaque signal en une séquence de vecteurs de N_f paramètres STFS. Le range des coefficients varie de 0 à $N_f - 1$. Le nombre maximal des paramètres est donné par :

$$N_f = \frac{F_s}{2} * \frac{1}{F_0} = 1000 \quad (II.1)$$

F_s : Fréquence d'échantillonnage qui égal à 100 KHz.

F_0 : Fréquence de réseau qui égal à 50 Hz.

II.4.3 Analyse de Fourier à court terme

Le signal du courant est découpé en une séquence de fenêtres d'analyse de durée T égale à 20 ms, qui représente la période du signal de la tension du secteur.

Supposons que le signal du courant des appareils électriques est un signal discret périodique de période égale à N_A échantillons représentant la durée T. Ce nombre d'échantillons est donnée par :

$$N_A = T.F_s = 2020.10^{-3}.100.10^5 = 2000 \text{ échantillons} \quad (II.2)$$

La méthode la plus utilisée pour l'analyse des signaux périodiques est la méthode de Décomposition en Série de Fourier (DSF).

Soit un signal $x(n)$ discret périodique de période égale à N échantillons.

La DSF du signal $x(n)$ consiste à le décomposer en une série de fonctions linéairement orthogonales de formes exponentielles :

$$x(n) = \sum_{k=0}^{N-1} C_k \exp\left(\frac{j2\pi kn}{N}\right) \quad n = 0, \dots, N-1 \quad (\text{II.3})$$

Où les coefficients C_k sont les coefficients de la DSF calculés comme suit :

$$C_k = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \exp\left(-\frac{j2\pi kn}{N}\right), k = 0, \dots, N-1 \quad (\text{II.4})$$

Généralement le nombre de coefficients C_k correspondants aux fréquences normalisés $f_k = \frac{k}{N}$ est pris égal au nombre d'échantillons N de la fenêtre d'analyse.

L'ensemble des N coefficients C_k ($k = 0, 1, \dots, N-1$) représente le spectre fréquentiel discret de la fenêtre d'analyse considérée. Ce spectre a une étendue de 0 à F_s . La bande fréquentielle maximale du signal est $[0, \frac{F_s}{2}]$ et le nombre de points fréquentiels représentant cette bande fréquentielle est égal $\frac{N}{2}$. Ainsi, le spectre utile de la fenêtre d'analyse est constitué des coefficients C_k avec k varie de 0 à $\frac{N}{2} - 1$. Le coefficient C_1 correspond à la fréquence fondamentale, alors que les coefficients C_k ($k = 2, \dots, \frac{N}{2} - 1$) correspondent aux fréquences harmoniques. Le rang $k = 0$ correspond à la composante continue.

Dans notre travail, on a pris le module de chaque coefficient C_k ($k=0, \dots, \frac{N_A}{2} - 1$) pour extraire l'ensemble des paramètres représentant la fenêtre d'analyse. Ainsi, chaque fenêtre d'analyse est représentée par un ensemble de $\frac{N_A}{2}$ paramètres (1000 paramètres). Ces coefficients sont appelés également les coefficients de série de Fourier discrète à court terme (Short Time Fourier Series : **STFS**).

II.5 ALGORITHME DE CLASSIFICATION KNN

II.5.1 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 [9] [6] [8].

Chapitre II: Système d'identification des appareils électriques proposé

Dans notre travail, l'algorithme KNN est appliqué pour classifier chaque vecteur de paramètres, les différentes étapes de l'algorithme KNN sont données comme suit [8]:

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

II.5.2 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. En règle générale, le meilleur choix de k dépend des données. La valeur optimale de k peut être choisie par des techniques heuristiques [12]. 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. Les différentes distances seront décrites dans la section suivante.

II.5.3 Distances utilisées

Il existe différentes distances utilisées dans l'étape de comparaison de l'algorithme KNN telles que les distances : Euclidienne, Minkowski, Hamming, Cityblock, Cosinus, Corrélation, Gaussienne, etc. Les distances utilisées dans notre travail sont données comme suit: Euclidienne, Cityblock, Cosinus et Corrélation.

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

a). 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 $d_E(X, Y)$ entre les deux vecteurs X et Y est donnée par

$$d_E(X, Y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (II.5)$$

b). Distance de Cityblock

Cette distance est définie comme suit :

$$d_E(X, Y) = \sum_{i=1}^m |x_i - y_i| \quad (II.6)$$

c). Distance Cosinus

La distance cosinus, également appelée distance angulaire, est dérivée de la similitude cosinus qui mesure l'angle entre deux vecteurs [9]. Cette distance est définie comme suit :

$$d_{cos}(X, Y) = 1 - \frac{\sum_{i=1}^m X_i Y_i}{\sqrt{\sum_{i=1}^m X_i^2} \sqrt{\sum_{i=1}^m Y_i^2}} \quad (II.7)$$

d). Distance corrélation

Cette distance est donnée par la formule suivante :

$$d_{cor}(X, Y) = 1 - \frac{\sum_{i=1}^m (X_i - \bar{Y}_i)}{\sqrt{\sum_{i=1}^m (X_i - \bar{Y}_i)^2}} \frac{(X_i - \bar{Y}_i)}{\sqrt{\sum_{i=1}^m (X_i - \bar{Y}_i)^2}} \quad (II.8)$$

II.6 COMBINAISON DE LA RÈGLE DE VOTE AVEC L'ALGORITHME DE CLASSIFICATION KNN

La méthode de vote est l'une des techniques les plus efficaces de reconnaissance de formes, où la classification finale est la décision majoritaire prise entre un ensemble de plusieurs systèmes simultanés

Contrairement à l'application des règles de vote entre plusieurs classificateurs, nous proposons dans notre travail d'appliquer la règle de vote pour un seul classificateur k-NN, entre différents nombres de vecteurs correspondant au même signal. Plus particulièrement, la règle de vote est appliquée sur la séquence d'indices des classes obtenue durant l'étape de classification des vecteurs de paramètres STFS du même signal de test. L'application de la règle de vote a pour objectif de déterminer la classe du signal du test. La classe du signal correspond à la classe la plus votée parmi les classes de la séquence d'indices.

II.6.1 Mesures de performances

Les performances du système d'identification proposé sont évaluées en termes du taux de classification (ou d'identification) des signaux de la base de test. Le taux de classification TCS des signaux est défini par la relation suivante:

$$TCS = \frac{Ns_C}{Ns_T} \cdot 100 \quad (II.9)$$

Où Ns_T est le nombre Total des signaux de la base de test, Ns_C est le nombre de signaux correctement identifiés.

Dans notre travail, le nombre des signaux de test Ns_T est égal à 420 signaux et le nombre de classes considérées correspond aux 12 types d'appareils.

La classification d'un signal de test consiste à appliquer la règle de vote sur la séquence des indices de classes des vecteurs classifiés par l'algorithme KNN. La classe du signal de test correspond à la classe la plus votée parmi les classes des vecteurs de la séquence. 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 :

$$TCV = \frac{Nv_C}{Nv_T} \cdot 100 \quad (II.10)$$

Où Nv_T est le nombre total des vecteurs de paramètres de tous les signaux de la base de test, Nv_C est le nombre de vecteurs de paramètres correctement classifiés.

Ce dernier taux de classification nous permet d'évaluer globalement les performances du système en tenant compte la classification d'un grand nombre de vecteurs.

II.7 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 décrit en introduisant la base de données COOLL et sa répartition en une base d'apprentissage et une base de test. Ensuite, l'étape d'extraction des paramètres est présentée en introduisant l'analyse harmonique des signaux basée sur la décomposition en série de Fourier. Ensuite, l'étape de classification des vecteurs par l'algorithme KNN est décrite en présentant cet algorithme et ses paramètres de configuration. Une description abrégée sur la stratégie de la règle vote et sa combinaison avec l'algorithme KNN, est donnée pour bien comprendre le processus de classification des signaux. Enfin, nous avons terminé ce chapitre

Chapitre II: Système d'identification des appareils électriques proposé

par la description de l'étape de mesure de performances du système proposé en termes du taux de classification des signaux.



Chapitre III
Implémentation et résultats



III.1 INTRODUCTION

L'identification des appareils électriques domestiques est utilisée pour la conception des compteurs intelligents dans le but d'économiser l'énergie.

Dans notre travail, on s'intéresse à implémenter un système d'Identification des Appareils Electrique (IAE) basée sur l'analyse harmonique et le classificateur KNN combiné avec la règle de vote Les performances du système sont évaluées sous la base de données COOLL en termes de taux de classification.

Plusieurs questions peuvent se poser sur la conception d'un tel système d'identification des appareils électriques.

On peut citer les questions suivantes :

1. Comment implémenter un système d'identification ?
2. Quelle est la meilleure configuration de classificateur KNN ?
3. Quel est l'ordre optimal des coefficients STFS (pair / impair) ?
4. Est-ce que l'utilisation de la règle de vote contribue à l'amélioration des performances du système d'identification ?
5. Quel est l'effet de la taille de la base de données d'apprentissage sur le taux de d'identification ?

Plusieurs expériences peuvent être effectuées pour répondre à ces questions.

III.2 IMPLÉMENTATION DU SYSTÈME D'IDENTIFICATION DES APPAREILS ÉLECTRIQUES

Le processus d'identification des appareils électrique commence par une étape de prétraitement de signal du courant électrique, qui consiste à détecter l'intervalle temporel de fonctionnement de l'appareil. Cet intervalle peut être obtenu en cherchant le début et la fin du fonctionnement de l'appareil à partir des méthodes de détection appliquées sur le signal du courant électrique. Dans notre travail, le début et la fin du fonctionnement dans chaque enregistrement du courant électrique, sont fournis à partir des données de la base COOLL sous formes de fichiers de type texte. Ainsi, on n'utilise que la partie utile de chaque signal représentant l'état de fonctionnement de l'appareil.

La deuxième étape consiste à découper le signal obtenu en une séquence de fenêtres de durée égal à 20 ms (période : 1/50) avec un chevauchement de 50%. Chaque fenêtre d'analyse est convertie en un ensemble de paramètres STFS en appliquant l'analyse de Fourier. Ainsi,

Chapitre III: Implémentation et résultats

cette étape a pour objectif de transformer chaque signal en une séquence de vecteurs de paramètres STFS.

Le nombre maximal de paramètres STFS de chaque vecteur est égale à 1000. Dans la référence [3] les auteurs ont montré qu'un nombre de 10 paramètres est suffisant pour avoir de bonnes performances du système d'identification des appareils électriques évaluées sous la base de données PLAID. Ce choix a été appliqué sur les paramètres statistiques des coefficients STFS [4]. Ainsi, dans notre travail, on limite le nombre de paramètres STFS à 10. Ce choix permet de réduire l'espace mémoire ainsi que le temps de calcul.

La conception de notre système se base sur une phase d'apprentissage et une phase de test qui exige chacune une étape d'extraction des paramètres STFS. Ainsi, cette étape est appliquée sur les 420 signaux de la base d'apprentissage et également sur les 420 signaux de la base de test, en transformant chaque signal en une séquence de vecteurs de 10 paramètres STFS. L'ensemble des vecteurs de paramètres obtenus sont regroupés dans une matrice de vecteurs d'apprentissage. De plus, on associe à chaque vecteur un indice de classe. L'ensemble de ces indices est enregistré dans un vecteur de type colonne (tableau).

Dans la phase de test, chaque séquence de vecteurs de paramètres STFS est classifiée en appliquant l'algorithme de classification KNN sur chaque vecteur. Ensuite, on utilise la règle de vote sur la séquence des indices de classes des vecteurs pour obtenir la décision finale sur la classe du signal de test correspondant à cette séquence. L'ensemble des indices de classes représente les 12 classes ou types d'appareils électriques.

Plusieurs critères de mesures de performances sont considérés dans ce travail. Le premier critère 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:

$$TCV = \frac{Nv_C}{Nv_T} \cdot 100 \quad (III.1)$$

Où Nv_T est le nombre global des vecteurs de paramètres de tous les signaux de la base de test, Nv_C est le nombre de vecteurs de paramètres correctement classifiés. Ce taux de classification permet d'évaluer globalement les performances du système en tenant compte tous les vecteurs de la base de données de test.

Le deuxième critère consiste à mesurer le taux de classification ou d'identification des signaux en appliquant la règle de vote :

$$TCS = \frac{Ns_C}{Ns_T} \cdot 100 \quad (III.2)$$

Chapitre III: Implémentation et résultats

Où N_{S_T} est le nombre global des signaux de la base de test, N_{S_C} est le nombre de signaux correctement classifiés ou identifiés.

Le taux de classification TCV est utilisé dans le but d'évaluer plus précisément les performances du système d'identification en tenant compte un grand nombre de vecteurs de test (162832). Réellement, le taux de classification adoptée est le TCS qui permet d'identifier le type d'appareils à partir de la classification de chaque signal de test et non par chaque vecteur.

III.3 EXPERIENCES

III.3.1 Configuration optimale de classificateur KNN

La configuration du classificateur consiste généralement à fixer le nombre des vecteurs les plus proches voisins K ainsi que le type de la distance (voir tableau III-1). Ainsi, cette expérience a pour objectif de déterminer la bonne configuration du classificateur KNN en cherchant le K optimal parmi les 10 premières valeurs (K=1 à 10) ainsi que la meilleure distance en termes de taux d'identification. Les performances du système sont évaluées en tenant compte le taux de classification TCV des vecteurs donnés par la relation (III.1.).

Tableau III-1: Configuration de classificateur KNN

K \ Distance	Distance			
	Euclidean	Correlation	Cosine	Cityblock
1	97.93	94.11	95.64	98.16
2	97.93	94.11	95.64	98.16
3	97.68	94.16	95.58	98.02
4	97.72	94.28	95.70	98.04
5	97.48	93.91	95.32	97.84
6	97.51	93.98	95.41	97.86
7	97.32	93.69	95.12	97.73
8	97.33	93.69	95.20	97.75
9	97.12	93.39	94.91	97.59
10	97.13	93.40	94.95	97.60

Chapitre III: Implémentation et résultats

Les résultats obtenus à partir du tableau III-1, montrent que le taux d'identification maximal est obtenu en choisissant la configuration (distance = 'Cityblock', k=1) avec un taux TCV égal à 98.16%. Dans les expériences suivantes, on choisira cette configuration.

III.3.2 Ordre optimal des harmoniques (pair ou impair)

L'objectif de cette expérience est d'étudier l'influence de l'ordre des harmoniques (impair et pair) sur les performances du système d'identification. Le tableau III.2 nous montre le taux de classification TCV du système pour les deux types d'ordre (pair et impair). Pour chaque type, chaque vecteur est constitué de 5 paramètres STFS. Les performances sont évaluées de la même manière que celles de l'expérience 1.

Tableau III-2 : Influence de l'ordre des harmoniques sur le taux de classification TCV

L'ordre des harmoniques	Nombre d'harmoniques	Taux de classification TCV %
Pair	5	93.16
Impair	5	98.55
Pair et impair	10	98.16

Les résultats obtenus, dans le tableau III-2, montre clairement la pertinence des paramètres STFS d'ordre impair avec un taux de classification de 98.55%. Les paramètres STFS d'ordre pair donnent les mauvaises performances avec un taux d'identification égal à 93.16%. Ces résultats confirment la pertinence des paramètres STFS du rang impair discutée dans (Mohamed Nait-Meziane, All, 2016).

III.3.3 Classification KNN combinée avec la méthode de règle de vote

Dans les expériences 1 et 2, le taux de classification TCV se calcule à partir de l'ensemble des vecteurs classifiés de tous les signaux en utilisant l'algorithme KNN. Dans cette expérience, on applique la classification KNN combinée avec la stratégie de règle de vote dans le but de prendre une décision sur la classification de chaque signal. Le taux de classification des signaux se calcule à partir de la relation (III.2). Le tableau III-3 nous présente le taux d'identification TCS obtenu par l'application de la règle de vote sur les vecteurs classifiés de chaque signal par l'algorithme KNN. Le taux est donné en fonction du nombre de paramètres STFS en variant ce nombre de 1 à 10.

Chapitre III: Implémentation et résultats

Les classes sont les 12 types d'appareils. De plus, ce tableau illustre le taux de classifications TCV dans le but de détailler également les performances du système d'identification.

Tableau III-3: Taux de classification TCS et TCV en fonction du nombre de paramètres. Le TCS est obtenu par l'application de la règle de vote sur les vecteurs classifiés de chaque signal par l'algorithme KNN.

Nombre de Paramètres STFS	TCV	TCS
1	63.35	90.47
2	76.77	98.57
3	93.53	100
4	95.73	100
5	97.31	100
6	97.63	100
7	97.95	100
8	97.97	100
9	98.12	100
10	98.16	100

Plusieurs points peuvent être remarqués à partir de ce tableau :

- La règle de vote améliore le taux de classification.
- L'utilisation des 3 premiers paramètres STFS est suffisante pour obtenir de bonnes performances avec un taux de classification TCS égal à 100%.
- La prise des 5 premiers paramètres STFS est suffisante pour obtenir de bonnes performances en termes de taux de classification des vecteurs avec une valeur TCV égale 97.31%.

Ce résultat montre l'efficacité de la stratégie de combinaison de l'algorithme de classification KNN avec la règle de vote. Ce résultat confirme également les résultats de l'application de la règle de vote dans la référence [Ghazali et All] mais sans tenir compte le type de paramètres statistiques.

La prise des 3 premiers paramètres STFS est suffisante pour obtenir de bonnes performances du système d'identification de types d'appareils électriques.

Chapitre III: Implémentation et résultats

Ainsi, les résultats obtenus dans cette expérience nous permettent de répondre sur la question posée sur la contribution de la règle de vote à l'amélioration des performances du système d'identification des appareils électriques en tenant compte les 12 classes des types d'appareils.

III.3.4 Effet de la taille de la base de données d'apprentissage

Cette expérience a pour but d'étudier l'effet de la taille de la base d'apprentissage sur les performances du système d'identification. Cette expérience consiste à calculer le taux TCS pour différentes tailles de la base d'apprentissage. Ces tailles sont prises inférieures à 50% dans le but de voir la robustesse du système d'identification à la réduction de la taille de la base d'apprentissage sans changement de la taille de la base de test. Ainsi, la taille de la base de test égale à 50% (la même base utilisée dans les expériences précédentes). Les différentes tailles de la base d'apprentissage prises dans cette expérience sont 50%, 40%, 30%, 20% et 10%. Le tableau III.4 nous montre les taux de classification TCV et TCS en fonction de la taille de la base d'apprentissage.

Tableau III-4: Effet de la taille de la base d'apprentissage

La taille de la base d'apprentissage	Le nombre de signaux	Taux TCV %	Taux TCS %
50%	420	98.16	100
40%	336	97.19	99.52
30%	252	95.81	99.28
20%	168	92.88	98.33
10%	84	84.75	93.33

Les résultats du tableau (III.4) nous permettent de constater les remarques suivantes :

- La réduction de la taille de la base d'apprentissage diminue les performances du système d'identification en termes des taux de classification TCV et TCS.
- Une grande baisse des performances du système est apparue à partir des tailles inférieures à 20% en tenant compte le taux de classification TCV.

Chapitre III: Implémentation et résultats

- La taille de 10% provoque une diminution considérable des performances du système en termes du taux de classification TCS.

III.3.5 Performances du système d'identification en tenant compte la classification des signaux en 42 classes d'appareils

La base de données COOLL contient des enregistrements de 12 types d'appareils tels que : lampe, machine à laver, etc. Chaque type regroupe plusieurs appareils, chacun se caractérise par sa puissance nominale, sa marque, etc. Le nombre d'appareils de cette base est égal à 42 [34].

Toutes les expériences précédentes considèrent le taux d'identification des types d'appareils. Ainsi, 12 classes sont considérées dans ces expériences.

Dans cette expérience, on vise à identifier les 42 appareils à partir des signaux du courant électrique. La conception d'un système d'identification des 42 appareils se base sur les mêmes étapes du système précédent, sauf que, les indices des classes de types d'appareils doivent être changés par les indices des classes d'appareils. Les taux de classification TCV et TCS sont calculés en considérant 42 classes d'appareils. Le tableau (III.5) présente les taux de classification TCV et TCS en fonction du nombre de paramètres STFS qui varie de 1 à 10. Le système d'identification se base sur l'algorithme de classification KNN combiné avec la règle de vote. La configuration de l'algorithme KNN se base sur le choix de K égal à 1 et la distance Cityblock.

Cette expérience permet d'évaluer les performances du système d'identification de 42 appareils au lieu de 12 types d'appareils. Cette étude a pour objectif d'examiner la capacité du système à bien discriminer ce grand nombre d'appareils.

Chapitre III: Implémentation et résultats

Tableau III.5: Taux de classification TCV et TCS en fonction du nombre de paramètres STFS, en tenant compte les 42 classes d'appareils.

Nombre de Paramètres	TCV	TCS
1	52.78	86.90
2	68.24	96.90
3	91.39	99.76
4	93.79	100
5	95.89	100
6	96.27	100
7	96.68	100
8	96.72	100
9	96.98	100
10	97.06	100

A partir du tableau III.5, on peut citer les points suivants :

- La règle de vote corrige énormément les performances du système en considérant les 2 premiers paramètres STFS.
- La prise d'un nombre de paramètres supérieur ou égal à 4 est suffisante pour obtenir un taux d'identification TCS égal à 100%. Ce résultat montre l'efficacité d'utilisation de la règle de vote.
- En termes de taux de classification TCV, les performances du système sont acceptables à partir du 5^{ième} paramètre.

On peut constater que notre système d'identification des 42 classes d'appareils a pu classer correctement tous les signaux de la base de test en appliquant l'algorithme KNN combiné avec la règle de vote sur des vecteurs composés seulement de 4 paramètres STFS.

Chapitre III: Implémentation et résultats

En comparant les résultats des tableaux (III.3) et (III.5), on peut donner les remarques suivantes :

- Dans la classification de 12 classes, Le taux TCS prend sa valeur max de 100% à partir le troisième paramètre, tandis que dans la classification de 42 classes on trouve que le taux TCS prend sa valeur max de 100% à partir le quatrième paramètre STFS.
- Les valeurs de taux TCV en fonction de nombres de paramètres pour la classification des 12 classes sont supérieures à celle obtenus dans la classification de 42 classes

La classification de 12 classes plus performante que la classification de 42 classes en terme de taux de classification. Mais en contre partie la classification de 42 classes plus précise que la classification de 12 classes et a la capacité de discriminer un plus grand nombre de classes.

III.4 CONCLUSION

Dans ce chapitre nous avons décrit notre système d'identification des appareils électriques, de plus nous répondons aux questions posées dans l'introduction par des expériences et nous obtenons les points suivants:

- La bonne configuration de l'algorithme de classification est "type de distance= cityblock avec k=1"
- L'ordre optimal des harmoniques STFS en terme de taux de classification TCV sont les harmoniques de rang impairs
- La règle de vote améliore le taux de classification et que l'utilisation des 3 premiers paramètres STFS est suffisante pour obtenir de bonnes performances avec un taux de classification TCS égal à 100%.
- La réduction de la taille de la base d'apprentissage diminue les performances du système d'identification en termes des taux de classification TCV et TCS.
- système d'identification des 42 classes d'appareils a pu classier correctement tous les signaux de la base de test en appliquant l'algorithme KNN combiné avec la règle de vote sur des vecteurs composés seulement de 4 paramètres STFS.

Conclusion générale

L'utilisation des compteurs intelligents dans les installations électriques domestiques est devenue une préoccupation majeure des distributeurs de l'énergie électrique. Ces compteurs intègrent généralement une fonction importante qui consiste à fournir des informations en temps réel sur l'identité de l'appareil et l'intervalle de son fonctionnement ainsi que l'énergie consommée. Cette tâche peut être effectuée en concevant des systèmes d'identification des appareils électriques à partir des signaux électriques. Les méthodes d'identification des appareils électriques peuvent être regroupées en deux catégories intrusives (ILM) et non intrusives (NILM).

Dans notre étude, nous avons conçu un système d'identification des appareils électriques domestiques basé sur le classificateur KNN combiné avec la stratégie de règle de vote appliquée sur des séquences de vecteurs de paramètres STFS. Les performances du système sont évaluées en termes du taux de classification, sous la base de données COOLL.

Les résultats obtenus montrent clairement l'efficacité de notre système d'identification. Pour répondre aux différentes questions posées sur la bonne configuration du système, nous avons effectué plusieurs expériences à savoir : la configuration optimale de classificateur KNN, l'ordre optimal des coefficients STFS (pair ou impair) ainsi que l'effet d'application de la stratégie de règle de vote sur les performances du système d'identification, puis nous avons étudié l'effet de la taille de base de données d'apprentissage sur les performances du système et finalement nous avons évalué les performances du système d'identification en tenant compte la classification des signaux en 42 classes d'appareils.

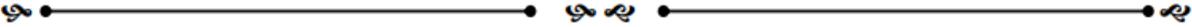
La configuration optimale du classificateur KNN est obtenue en choisissant la distance 'Cityblock' avec $k=1$. Cette configuration donne un taux de classification de vecteurs TCV égal à 98.16%. L'ordre impair des harmoniques donne également le meilleur taux d'identification TCV égal à 98.55%. Les résultats nous ont montré que l'application de la règle de vote sur les séquences de vecteurs constitués de plus de trois paramètres permet d'obtenir un taux de classification des signaux TCS égal à 100%. L'application de cette

Conclusion générale

configuration du système sur l'identification de 42 appareils de la base COLL en utilisant seulement 4 paramètres a donné également un taux TCS égal à 100%.

Pour des travaux futurs, nous pouvons donner les perspectives suivantes :

- L'utilisation d'autres méthodes d'extraction de paramètres telles que celles basées sur l'analyse par ondelettes.
- L'application de la règle de vote avec d'autres classificateurs tels que : ADI, GMM, etc.



Références



- [1] Nabil Amirach. « Détection d'évènements simples à partir de mesure sur courant alternatif » thèse de doctorat, Spécialité : Traitement du Signal ,10 Juin 2015, Université de Toulon, France.
- [2] Mathieu Sanquer, « Détection et caractérisation des signaux transitoires (application à la surveillance de courbes de charges » thèse, Spécialité : Sciences de l'ingénieur physics /Electronique, 15/03/2013, université de Grenoble, France.
- [3] Mohamed Nait Meziane."Identification d'appareils électriques par analyse des courants de mise en marche". Thèse, Spécialité: Traitement du signal et de l'image. Université d'Orléans, 9/12/2016. France
- [4] 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 .
- [5] 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.
- [6] Mohamed Nait Meziane, Philippe Ravier, Guy Lamarque, Karim Abed-Meraim, Jean-Charles Le Bunetel, et al. "Modeling and estimation of transient current signals". 2015 23rd European Signal Processing Conference (EUSIPCO), Aug 2015, Nice, France. pp.1960-1964.
- [7] Cite web : <https://www.xlstat.com/fr/solutions/fonctionnalites/k-nearest-neighbors-knn>.
- [8] Cite web : <https://mrmint.fr/> « Mr.Mint Machine Learning made easy »2018.
- [9] V. B. Surya Prasatha,b,c,d,*, Haneen Arafat Abu Alfeilate, Ahmad B. A. Hassanate, Omar Lasassmehe, Ahmad S. Tarawnehf, Mahmoud Bashir Alhasanatg,h, Hamzeh S. Eyal Salmane « Effects of Distance Measure Choice on KNN Classifier

Références

- Performance » Big Data Volume 00, Number 00, 2019 , Mary Ann Liebert, Inc. DOI: 10.1089/big.2018.017529 Sep 2019.
- [10] Faïcel Chamroukhi, « Classification supervisée : Les K-plus proches voisins » Spécialité : Statistics & Data Science University, Mathematics Nicolas Oresme laboratory.
 - [11] Chikouche Soumia, Mémoire Master « Système de détection d'intrusion basé sur la classification comportementale des processus » spécialité : l'informatique. Université de M'sila 28/06/2012.
 - [12] Barigou Fatiha, thèse « Contribution à la catégorisation de textes et à l'extraction d'information» thèse de doctorat, Spécialité : informatique, Université d'Oran. 2012/2013.
 - [13] Tariq Abdessettar, « Développement d'un système de classification automatique de l'activité physique » Mémoire Master, Spécialité : technologie de l'information, université de Boulevard du Maréchal, France Février 2018.
 - [14] Laurent Candillier , « Contextualisation, visualisation et évaluation en apprentissage non supervisé » thèse de doctorat, Spécialité: Informatique, université Charles de Gaulle, France, 3 septembre 2006.
 - [15] cite web : <https://mrmint.fr/introduction-k-nearest-neighbors> 2018.
 - [16] Christophe Chesneau, « Éléments de classification » Mémoire Master Université de Caen Normandie, 22 janvier 2016.
 - [17] Jianxin Wu, « Distance metrics and data transformations » National Key Lab for Novel Software Technology Nanjing University China 11 February 2020 .
 - [18] Tanmay Basu¹, C. A. Murthy¹, and Himadri Chakrabarty « A Tweak on k-nn Decision rule » Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata-700108, India 2 Department of Computer Science, Surendranath College, University of Calcutta, Kolkata - 700009, India.
 - [19] Dourdour Khedidja, Mouss Hayet « Combinaison de classifieurs KPPV pour la classification des chiffres imprimé Dourdour » Département de génie industriel, Laboratoire d'automatique et productique Université de Batna, Algérie.
 - [20] G. W. Hart- Minimum information estimation of structure- Ph.D. Dissertation, Massachusetts Institute of Technologie, Dept. of Electrical Engineering and Computer
 - [21] M. BONS, " Modèles à sources markovienne cachée appliqués à l'analyse non-intrusive des principaux usages domestiques de l'électricité", thèse de doctorat, Université de Rennes I, France, 1996

Références

- [22] F. SULTANEM. "Using appliance signatures for monitoring residential loads at meter panel level- IEEE Transactions on Power Delivery, Vol. 6, pp.1380-1385, October 1991.
- [23] F. SULTANEM " Procédé et appareil d'analyse de signaux de courant et de tension en vue du repérage de charges à usage domestique- Brevet Français 2645968, 19 October 1990.
- [24] T. ONADA, Y. NAKANO, K. YOSHIMOTO, brevet "System and Method for estimating power consumption of electric apparatus, and abnormality alarm system utilizing the same" US patent 6,816,078 B2, 9 November 2004.
- [25] Leeb S. A conjoint pattern recognition approach to nonintrusive load monitoring. PhD, Massachusetts Institute of Technology, Cambridge, MA, USA, 1993
- [26] S. B. Leeb and J. L. Kirtley, Jr., "A multiscale transient event detector for nonintrusive load monitoring," in Industrial Electronics, Control, and Instrumentation, 1993. Proceedings of the IECON '93., International Conference on, 1993, pp. 354-359 vol.1.
- [27] Kolter, J.Z.; Johnson, M.J. REDD : A Public Data Set for Energy Disaggregation Research. In Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability, San Diego, CA, USA, August 2011; pp. 1–6.
- [28] Kyle S. Dawson, David J. Schlegel, Christopher P. Ahn, Scott F. Anderson, Éric Aubourg, Stephen Bailey et al. "the baryon ascillation spectroscopic survey of SDSS" The Astronomical Journal, published in January 2013. doi 10.1088/0004-6256/145/1/10.
- [29] Mathew, S., Yella, A., Gao, P. et al. Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers. Nature Chem 6, 242–247 (2014). <https://doi.org/10.1038/nchem.1861>
- [30] Gulati, Ranjay & Gargiulo, Martin. (1999). Where Do Inter organizational Networks Come From? 1. American journal of sociology. 104. 1398–1438-1398–1438.
- [31] Kelly, Jack and Knottenbelt, William Neural NILM: Deep Neural Networks Applied to Energy Disaggregation 2015. Association for Computing Machinery, New York, NY, USA
- [32] Micheal E. Williams, Fangxin Hong, Randy D Gascoyne, Lynne I Wagner, John C Krauss, Rituximab Extended Schedule or Retreatment Trial (RESORT) for Low

Références

Tumor Burden non-Follicular Indolent B-cell Non-Hodgkin Lymphomas: Eastern Cooperative Oncology Group Protocol E4402, 2016

- [33] HASSAN CHOUAIB « Sélection de caractéristiques méthodes et applications » thèse de doctorat spécialité : informatique, université paris Descartes , 08/07/2011.
- [34] Clark, M. S. Improving the feasibility of energy disaggregation in very high-and low-rate sampling scenarios. Master's thesis, British University ,Columbia.(2015)
- [35] Thomas Picon, Mohamed Nait Meziane, Philippe Ravier, Guy Lamarque, Clarisse Novello,all. COOLL: Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification,