Master's Thesis in Artifical Intelligence

Radboud University

# Automatic Appliance Identification

*Investigation of Automatic Appliance Identification scenarios and methodologies*

*Author:*
Odysseas Krystalakos
s1016895
Radboud University

*External Supervisor:*
Ir A.C. van Rossum
Crownstone

*Internal Supervisor:*
Dr L. Ambrogioni
Radboud University

*Second Assessor:*
Dr U. Güçlü
Radboud University

June 2020

**Abstract**

Automatic Appliance Identification refers to the task of identifying household devices given measurements of its power consumption. Solving this problem is crucial for modern energy monitoring applications but, so far, it has been shown to be non-trivial. In addition, there seems to be confusion about the practical scenarios on which Appliance Identification can be deployed. In this research project we attempt to untangle the definition of Appliance Identification by proposing a distinction of three different scenarios. Among these, we describe the Appliance Load Identification scenario that, even though it had been implicitly mentioned in past works, it was never explicitly defined. With regards to experiments, we initially replicate results of noteable past works using open datasets. Next, we propose a novel set of techniques for Appliance Identification that use a mix of VI trajectory data, handpicked features and Multi-Modal Neural Networks. Finally, we propose three classifiers for the newly-defined Appliance Load Identification scenario. Through tests we find that most existing models are not robust to tests across datasets. We also find that combining VI trajectory representations with other features leads to increased performance. Last, we provide the results of our Appliance Load Identification models as baseline for future research.

# Preface/Acknowledgements

This research was part of the final Research Project in the Master's in Artificial Intelligence in Radboud University. The project was completed at Crownstone, a Rotterdam-based startup that greatly assisted during the entire research project, even during the times of COVID-19. The topic of Appliance Identification was introduced by Anne van Rossum and fit perfectly with my research interest in smart homes and environments. Anne also served as the external supervisor for this thesis and was extremely helpful in terms of domain knowledge, methodology and personal guidance. I would also like to thank the whole Crownstone team for hosting me and offering a creative environment for research. Finally I would like to thank Luca Ambrogioni, the internal supervisor, for the creative freedom during this research effort.

O. Krystalakos
June 2020

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Managing the electricity demands of a household has been a topic of great interest, over the last decades. Inspired by the high economic and ecological relevance of power consumption, individuals and organisations are increasingly eager to track and optimise their energy usage. As energy saving devices, home solar panel systems and green architecture are becoming more widespread, monitoring and understanding the consumption of a home becomes crucial. Such insight involves the total consumption, as well as fine-grained information about the contribution of each appliance.

In parallel with the timely concerns about sustainability, we also see a growing interest in smart houses and ambient intelligence. With cheaper hardware and portable computing devices, such as smartphones and wearables, the potential increases for infusing traditional houses with "smart" capabilities. This may include energy consumption tracking, remotely or automatically controlled houses and sophisticated security systems. Besides automation, all these upgrades on typical houses generate an abundance of data of high value. They can allow behavioural monitoring of the occupants, assist occupants in daily tasks and detect issues regarding electric devices, heating etc. As this notion of making our environment work for us becomes increasingly popular, energy monitoring can prove to be beneficial both in itself and as a general data acquisition technique.

## 1.1 Electric Load Monitoring

The potential of monitoring the electrical consumption has been recognised by the research community since the 1980's. One of the first proposed approaches for household-level monitoring is *Non-Intrusive Load Monitoring* (NILM)[1]. Under the NILM scenario, only one meter is used, which measures the total power consumption of the house. With this data, a suitable algorithm can infer the power consumption per-appliance using machine learning. The "Non-Intrusive" part comes from the notion that the meter does not interfere with the existing in-house circuit but is installed along the energy meter of the utility company.

An alternative scenario is the installation of several smart meters in-house that are interconnected through a network. This method is referred to as Intrusive Load Monitoring (ILM) and can be split in three sub-categories[2]. The first category, named ILM 1, involves splitting the house in regions and using one meter per region. This usually entails installing sensors at the circuit breaker panel. In ILM 2, more meters are deployed, one per outlet, thus receiving fine-grained data for each appliance. The final category, ILM 3, assumes that all devices come equipped with an embedded meter and a data feedback system. All of the three ILM solutions have their own use-cases and the choice usually depends on the trade-off between reliability and cost of installation. ILM 1 is considered the most cost effective, due to the low number of smart meters, but the process of separating the appliance consumption per-device is non-trivial. On the contrary, smart appliances are not always available or affordable, therefore rendering ILM 3 the most reliable but least cost-effective.

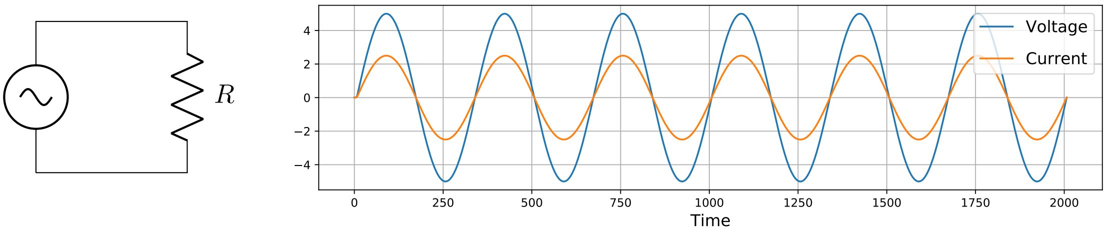
## 1.2 Automatic Appliance Identification

All of the aforementioned approaches are, in theory, capable of providing power consumption data, for each appliance separately. However, even though power data are available, there is no information about the type of the device. This is an important component for home automation and device-specific control. Manual labelling of appliances is not a viable solution for most Energy Load Monitoring scenarios. For example, for NILM, the user would have to match predicted loads to the devices manually, every time there is a new result. For ILM 1 and 2 it is required to pair the devices to specific meters. Such a setup is time consuming, not user-friendly and prevents reconfiguration of the house circuitry (eg. plugging out the kettle to in order to plug in the toaster). The ILM 3 approach is the only one that is not affected by this issue but assumes a common protocol of data collection for all devices, which is not realistic at the moment. For all these reasons, it is desirable to be able to infer a label for the appliance using the available consumption data. This task is referred to as Automatic Appliance Identification and is considered a crucial component of any load monitoring system.

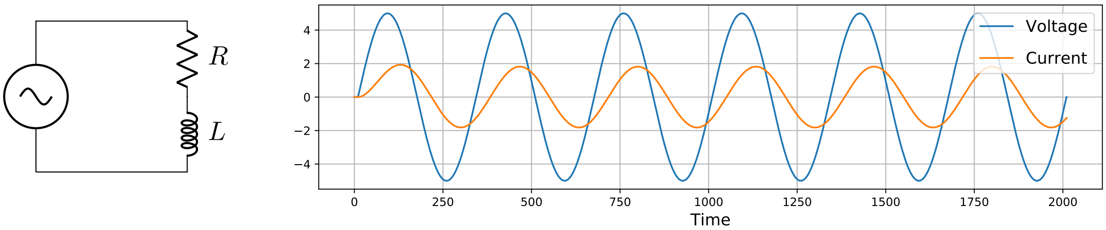
On a first glance, the feasibility of such a task may not be self-evident. The basis for the conception of Automatic Appliance Identification becomes clear when one examines the behaviour of basic AC electric circuits. The simplest of all can be considered a circuit comprised of a source and a resistor. In this case, the resistor component draws power that is converted into heat. To create more circuits, there is a set of components that can be placed alongside the resistor, the most common of which are capacitors, coils and diodes. By combining these components in various ways, we can get different circuits, each of which displays different behaviour.

In order to monitor these simple circuits with respect to their power consumption, we can measure the voltage that is supplied by the source and the current that is drawn by the components. It is expected that the design of the circuit, referring to its components and topology, will directly affect the measurements. An example of this can be seen in Figure 1.1 where we can examine the voltage and current waveforms for three circuits. It is visible, that there are differences in the measured current. For the R circuit, voltage and current have the shape of a sine wave. The same applies for the RL circuit but this time there is phase difference between the two. In the third example, the current has no negative values. In general, every component draws current in a specific way that depends on its natural properties. Therefore every circuit will have a characteristic current signal, according to its design.

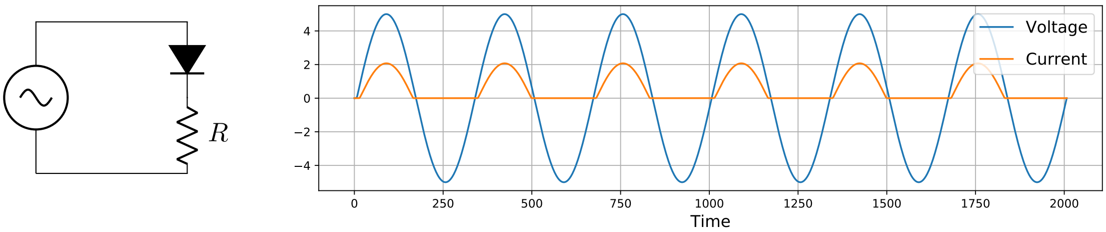
The main idea behind Automatic Appliance Identification is that the principle of identifying small circuits can be generalised to household electric appliances. Since each appliance has a unique circuit design, it will also have a distinctive consumption pattern. We can therefore use the characteristics of the voltage and current signals to infer the device label. Realistically however, similar devices will display similar power consumption patterns. As such, this task becomes a classification problem for which the input are the voltage and current measurements and the target classes are the device labels.




(a) R circuit




(b) RL Circuit




(b) Diode-Resistor Circuit

Figure 1.1: Voltage and Current measurements for three circuits with different components. While the Voltage waveform remains the same among all three, the Current shape changes depending on the components.

## 1.3 Scenarios for Appliance Identification

Similarly to energy monitoring, Appliance Identification is a quite broad term and can be distinguished into categories, according to the target classes. These categories differ mainly with respect to the use-cases to which they can be applied. In order to make the problem statement clearer and allow for a fair comparison between approaches we propose the following categorisation. To the best of author's knowledge, no such categorisation has been formulated before.

### 1.3.1 Appliance Type Identification

Appliance Type Identification refers to the task of inferring the *type* of the target appliance with regards to its functionality. Example target classes for this category would be "fridge", "kettle", "iron" etc. This means that, for example, fridges of different manufacturers should be recognised as the same type.

This approach towards appliance identification assumes low within-class variance and high between-class variance. In other words, it is expected that, in terms of energy consumption, different models of the same type would behave more similar to each other than to other appliance types. In theory, such a model, should be able to recognise any appliance available anywhere, even though it is trained on a small subset of them. As such, high generalisability is crucial.

### 1.3.2 Appliance Instance Identification

Appliance Instance Identification considers each device a separate entity. The target classes in this case are not appliance types but the specific model of a device, referred to as *instance*. This means that a particular appliance model should be considered distinct from another model of the same type. Consequently it allows for more detailed results, meaning that it can infer the appliance type and the model.

An Appliance Instance Identification system would require data for each specific instance. Ideally, the manufacturer can provide this type of information. If not, sampling can be done in-house during installation. Consequently, a model that classifies appliance instances can only work with the specific set of devices on which it was trained. As a result, it can be deployed only in homes with the same exact appliance set. Hence, generalisability is of low priority.

### 1.3.3 Appliance Load Identification

Appliance Load Identification focuses on the type of *load* that an appliance adds to the grid. In this context, with the term *load* we refer to the type of behavior a circuit displays, when regarded as a small circuit. This means that, in order to identify an appliance, we regard it as a much simpler circuit, based on its most prominent components. For example an iron and a toaster both fall into the class of 'resistive loads' as their main components are the resistors that generate heat. Similarly, an LCD TV and a laptop charger have electronic power supplies and can be grouped as such. With this categorisation, appliance types are abstracted into more general categories. In parallel, appliances such as air conditioners or washing machines have more than one modes of operation, and therefore, can be characterised by more than one loads. With a load identification system, we can get valuable insight on the way that the device operates and what are its most characteristic components.

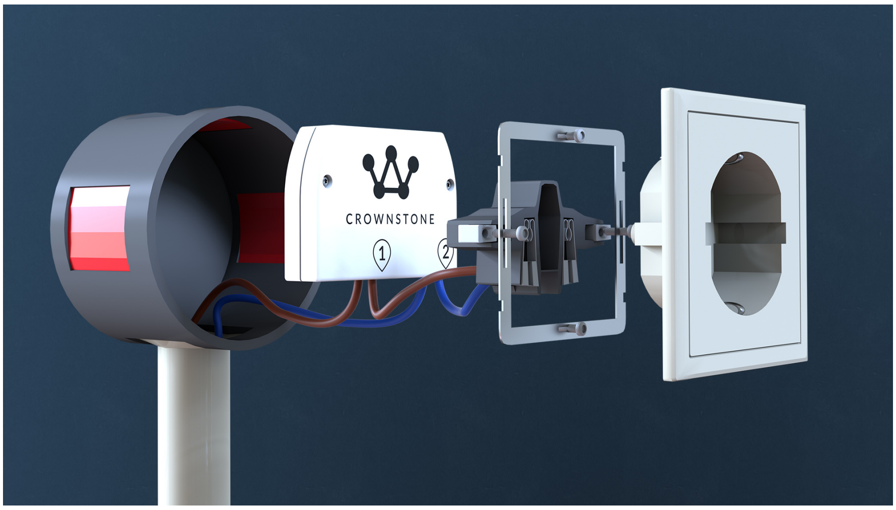



Figure 1.2: The Crownstone chip. It is installed behind the traditional power socket and controls the power flow.

## 1.4 Crownstone

Crownstone [1] is chip that can add smart capabilities to a traditional electric outlet. As seen in Figure 1.2, it can be installed behind the socket and is capable of measuring current and voltage from the devices that are plugged into it. Among other features, it can react to the position of occupants in the house, switch devices on/off and dim lights. It also offers communication via Bluetooth and is technically capable of high-frequency sampling of voltage and current data. To sum up, the use of Crownstones in a home, creates an ILM 2 scenario.

Crownstones can offer easy control of devices but they lack the ability to identify the connected appliance. The addition of this functionality would allow for intelligent control of devices depending on their type. Some example scenarios for this are:

1. Different types of lamps require different dimming methods. By differentiating LED from fluorescent lamps this process can be made automatic.

2. Plugging in a dangerous device (e.g. iron, power drill) when no adult is present near the outlet can be detected and prevented.

To achieve the above, it is necessary to identify appliances in real-time using high-frequency measurements of voltage and current.

## 1.5 Research Questions and Overview

Given the previous description of the Appliance Identification problem, this research aims to tackle a number of tasks. Initially, we investigate how existing appliance identification algorithms generalise in unseen datasets and we evaluate their performance. Next, we examine, to the best of our ability, if visual representations of the collected data can be used to improve performance, especially when combined with convolutional neural networks. Finally, we expand on

[1] https://crownstone.rocks

our definition of appliance load identification, and explore the feasibility of a load identification system.

With the above questions in mind we hypothesise the following:
$h_1$: Existing algorithms perform worse on unseen datasets compared to the dataset from which the training set was collected.
$h_2$: CNN architectures that use raw timeseries or visual representations of the data as input, outperform algorithms that use handpicked features.
$h_3$: A load identification system is feasible given labelled data.

In the following chapters we attempt to explore the topic and address the questions above. Initially, in Chapter 2, we explore the related literature on Appliance Identification and detect issues with the task itself and the existing methodology. In Chapter 3 we establish the methodology for experimentation, namely the datasets that are used, the necessary data preprocessing and finally the set of features deployed in this project. In Chapter 4 we apply this methodology on techniques retrieved from past literature, attempting to replicate their results and discover strengths and weaknesses. Then, we explain our proposed models and list their performance for Appliance Type Identification in Chapter 5 and Appliance Load Identification in Chapter 6. In Chapter 7 we answer the research questions based on the results and comment on interesting findings of the experimental process. Finally, in Chapter 8 we summarise the contribution of this research work and propose steps for the future, on the task of Appliance Identification.

This manuscript is accompanied by two Appendices that contain technical details about the experiments. Specifically, Appendix A contains circuit schematics for the artificial dataset introduced in Chapter 3 and Appendix B lists devices that are taken from an open dataset and are used in the experiments of Chapters 5 and 6.

# Chapter 2

# Related Work

## 2.1 Related Work

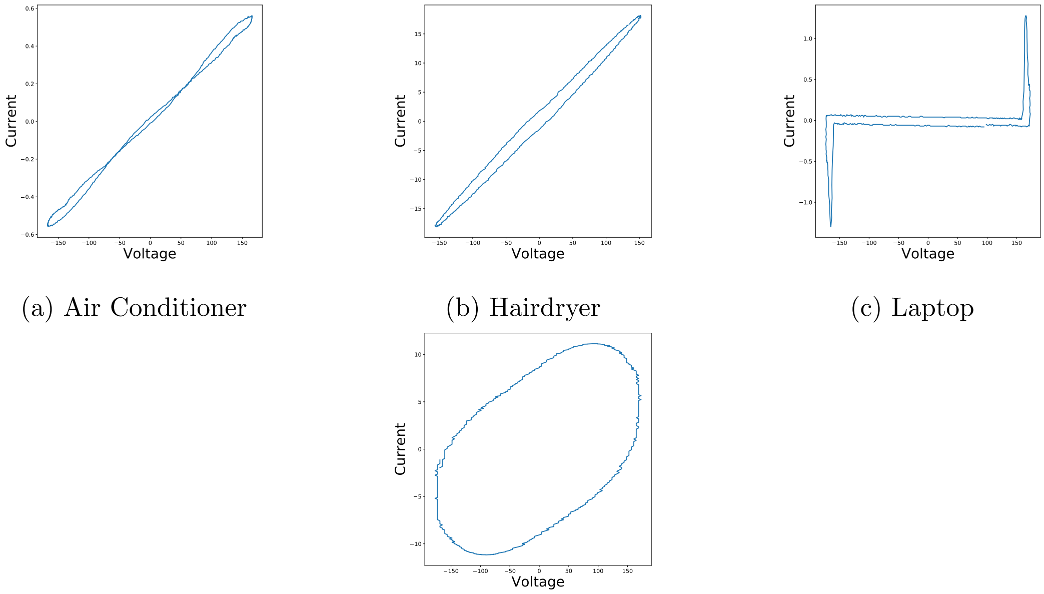
Several attempts have been published in the past on the problem of automatic appliance identification with varying degrees of success. Since it is closely coupled with the engineering task of developing a smart meter, most past literature considers the Appliance Identification system as a sub-component of a NILM or ILM solution[3–5]. The fundamental element of all approaches is that appliance identification is considered a classification problem that exploits voltage and current measurements.

## 2.2 High Frequency Appliance Identification

The main target of this research project is appliance identification in real time. This means that there is a need for obtaining descriptive data in a short amount of time. For this reason, we focus on literature that works with high-frequency measurements (>1kHz) and exclude bibliography that processes timeseries collected over hours of monitoring.

The common component of all Automatic Appliance Identification works is the fact that Machine Learning is deployed as a solution to the classification problem. Often, there is an extensive preprocessing phase to transform the raw signals into a descriptive feature vector. Kato *et al.* [3] opted for automatic feature extraction with PCA and classified appliances using an SVM. On the contrary, Reinhardt *et al.* [6] handpicked a set of features from the time and frequency domain. The performance of the these features was then examined through popular machine learning techniques such as Random Forest, Bagging, Bayesian Networks etc. Other pieces of literature tried to extract features such as Active and Reactive power [7] or used FFT [5]. Nevertheless, the main methodology remains the same and involves three step: data collection, feature extraction and classification.

In the more recent years, with the advancement of Deep Learning especially on the field of Computer Vision, researchers turned to Neural Networks for appliance identification systems. Barsim *et al.* [8] experimented with an ensemble of Neural Networks that classify appliances using a window of raw voltage and current signals. This way, the feature extraction part is omitted. Another methodology was proposed by De Baets *et al.* [9] that involves Convolutional Neural Networks on images of plots of the Voltage-Current Trajectories (VI trajectories). VI trajectory images have been established by several studies [10], [11] as a descriptive visualisation of raw voltage and current signals. In Figure 2.1 you can see example VI trajectories for four appliance




(a) Air Conditioner (b) Hairdryer (c) Laptop

(d) Washing Machine

Figure 2.1: VI trajectories of four appliances. VI trajectories vary in terms of shape, depending on the current waveform. Some differences that are clearly visible is that the Air Conditioner (a) shows self-crossings that are not present in the even shape of the Hairdryer (b). The Laptop (c) seems to form the shape of a rotated “Z” while the Washing Machine (d) encircles bigger area compared to the rest.

types. It is noticeable that the trajectories can vary greatly in terms of shape, depending on the current waveform. Apart from conveying a lot of information about the original signals, VI trajectories are also invariant to changes in the main voltage from region to region[12]. Therefore they are a robust representation, regardless of the outlet voltage.

Given the above information we see that existing work is split into two methodologies. One focuses on extracting descriptive features from the data that are used to create a “fingerprint” for the appliance. The alternative is to use Neural Networks that use either the raw signal or the VI representation as input. Both approaches seem to perform very well in their respective datasets. However, direct comparison is not possible due to different test sets across literature.

## 2.3 Relevant problems in Appliance Identification

Inspecting the literature pertaining to Appliance Identification, we notice a few issues. These mainly concern the definition of the problem itself as well as the data that are available for experimentation.

### 2.3.1 Insufficient datasets

Appliance identification has been consistently tackled using machine learning. This methodology requires high quality measurements for training and testing. These measurements are difficult to acquire, mainly because the setup of a data collection environment is a challenging and time consuming task. Firstly, it is necessary to acquire a high-frequency smart meters and a wide selection of devices to sample from. Secondly, collecting representative samples is not straightforward for many devices, as they may have parameters that alter their behaviour (e.g. washing machine cycles, hair dryer settings). As such, acquiring data to train and test models is non-trivial.

Fortunately, there are several open energy consumption datasets suitable for appliance identification. Unfortunately, each dataset comes with its own properties that make it quite difficult to merge them or cross-test. The first distinction between datasets is the sampling frequency used to collect energy data. Low frequency datasets usually sample values with a resolution of seconds and mainly focus on wattage. High frequency measurements are necessary to acquire voltage and current measurements, due to the periodic nature of the signals. Another point of discrepancy between datasets lies in the variances of mains electricity supplies. This refers to the voltage and frequency used by each country. For example, most European countries have a mains supply of 240V/50Hz while North America uses 120V/50Hz. This difference implies that the same device models will display slightly different behaviour when used in different regions. In addition, this variation affects the frequency response of the signal, therefore making models trained on one mains network, incompatible with the devices that use the other. Kholeif *et al.* [12] examined the most popular features for appliance identification and has found that several of them are independent of the supply voltage. Through these specific features, it is possible to combine data with different origins.

As a result of the above issues with most datasets, literature regarding appliance identification mostly focuses on training and testing on one dataset using a splitting strategy. While this is a valid experimentally, we find that it is not sufficient in order to fully investigate the capacity of the models. Firstly, most datasets provide a very limited set of appliances instances or types. Secondly, sets collected with the same sensors may contain biases or characteristic noise. In conclusion, we find that experimenting with different data sources can be of great use when evaluating Appliance Identification techniques.

A major point that most datasets are lacking in is sufficient labelling. As mentioned in Section 1.3.1 each approach to the task requires specific labels. Some datasets provide device model names[13], which are useful for appliance instance identification, while others just mention the type[14]. The type of the appliance is often ambiguous (PC monitor vs TV, space heater vs air conditioner etc) which may also cause issues. Finally, according to the author's knowledge, there is no open dataset with labelled appliance loads.

### 2.3.2 Ambiguous testing scenarios

The testing scenario, as described in Section 1.3, is not always made clear in the literature. On one hand, Kato *et al.* and Reinhardt *et al.* [3, 6] regard appliance identification as identification of previously seen devices. On the other hand, Barsim *et al.* and De Baets *et al.* [8, 9] attempt to generalise to unseen devices by inferring classes such as "fridge", "hair dryer", "laptop" etc. Gisler *et al.* [15] attempted to separate these two tasks into two testomg protocols. However, Du *et al.* [16] and Lam *et al.* [17] reformulated the task once again and focused on groupin similar devices into general categories. Similarly, Iksan *et al.* [18] attempted to group devices semantically, essentially focusing on the load.

As we can see there is confusion with respect to the definition of Appliance Identification. Comparison between techniques of different scenarios is not only unhelpful but also misleading. For this reason it is crucial to clearly state the task at hand and the target classes.

### 2.3.3 Within and Between class variance

A major problem in appliance identification, especially when it comes to Appliance Type Identification, is the high variance in behaviour between devices of the same type. This means that, especially in more complex devices such as washing machines and air conditioners, the operation cycles may differ a lot. As such, it is a difficult task to generalise sufficiently when training on a small dataset. In addition to this, there are often major contradictions even between states of operation of the same device. For example a microwave has several settings, sometimes works only as a timer or may come equipped with an oven function. Similarly, a washing machine goes through the stages of spinning and pumping water for which it uses very different components. Overall, defining a coherent class is not always possible when working with electric appliances.

At the same time, there are often similarities between devices of different classes. For example a kettle and an iron are devices with just heating elements. Heating usually comes from a resistor which will display a very generic pattern of operation. This problem is widespread considering that most electric devices are comprised by a limited set of components: heating elements, motors, rectifiers, evaporators, electronic power supplies etc. All in all, from a physical perspective, the appliance identification task seems to suffer from high within-class variance and low between-class variance.

# Chapter 3

# Methods

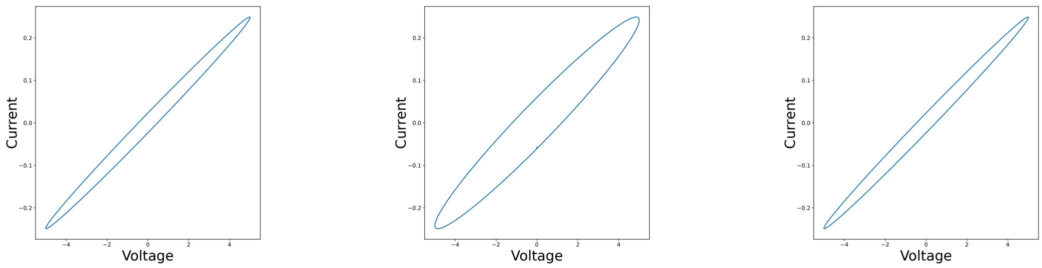
In order to tackle the questions posed in Section 1.5, it is necessary to conduct several experiments, using existing and novel methods. Since we investigate two different scenarios, namely Appliance Type Identification and Appliance Load Identification, it is necessary to formulate two different methodologies. In the next sections we explain the parts of the methodology that apply to both scenarios including the data and the handpickced features.

## 3.1 Open Energy Consumption Datasets

Due to the high research interest in smart houses and energy consumption, there are several open datasets on the topic. For the needs of this project, it is necessary that the measurements are on a per-device basis. It is also mandatory that samples are labelled in terms of appliance type. Since generalisation is crucial for Appliance Type Identification, the samples need to come from different instances of the same type, to ensure that the model captures the characteristics of the entire class. Given the above constraints, we focus on the datasets PLAID [14], the extension to PLAID (PLAID 2) [19] and WHITED [13]. The relevant properties of these datasets can be found in Table 3.1. PLAID offers high variation regarding the number of instances per type. The same holds for PLAID 2, that, besides adding more data, it caters for balancing the number of samples for each type, which was an issue in PLAID 1. Finally, WHITED has a very small set of devices per type which make it unsuitable for training but useful for cross-dataset tests.

| **Name** | **Sampling Frequency** | **# Houses** | **# Types** | **# Instances per Type** | **# Samples per Instance** |
|---|---|---|---|---|---|
| PLAID | 30kHz | 55 | 11 | 7-38 | 26-92 |
| PLAID 2 | 30kHz | 9 | 11 | 5-9 | 75-248 |
| WHITED | 44.1kHz | No house data | 46 | 1-5 | 1-20 |

Table 3.1: Properties of energy consumption datasets. All of them provide high-frequency measurements. The number of available appliance Types directly affects the difficulty of identification. The number of Instances per Type indicate the variety of devices for each type which is important for generalisation. A high number of samples provides a more complete picture of the device behaviour.




(a) No noise (b) Artificial phase shift (c) Additive sine noise

Figure 3.1: VI trajectory of RL circuit taken with different types of artificial noise. In the left, there is no added noise. In the centre, the VI trajectory encloses a greater area due to the phase shift in the current signal. In the right, the addition of high-frequency sine noise causes high-frequency fluctuations, creating an irregular line.

### 3.1.1 Artificial Dataset

With regards to Appliance Load Identification, to the author's knowledge, there are no available labelled data. For this reason, an artificial dataset is generated using a circuit simulation software. The design and simulation of the circuit of a realistic device is not an easy task. Since the data should be labelled with respect to the load and not the type of the appliance, there is no need to implement the actual circuit. Instead, it is sufficient to acquire data from equivalent, small circuits. With this information we can verify the classes are indeed separable and cohesive to allow for high accuracy classification.

For this purpose, use the NGSPICE [20] software that allows for the design, simulation and inspection of electrical circuits. We design 8 different circuits, consisting of Resistors, Capacitors, Coils and Diodes. Details regarding the designed circuits are listed in Appendix A. The circuits are simple enough to be designed manually and hand-labelled in terms of load. We randomise the values of the components, within a range, in order to get slightly different results. Next, realistic noise was added randomly. Specifically we introduce:

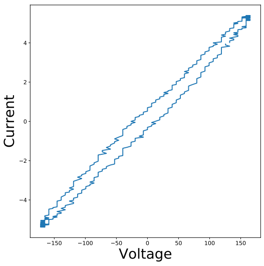
(a) a slight phase-shift representing capacitive coupling between wires

(b) a low-voltage, high-frequency sine wave representing fluctuations in the source or resistive components

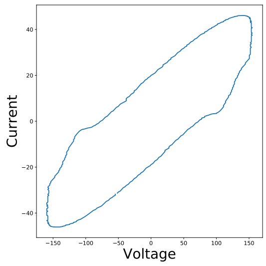
The impact of the noise techniques on the VI trajectories can be seen in Figure 3.1. Specifically, we see that the artificial phase shift results in a trajectory that encloses a greater area whereas the high frequency wave causes slight fluctuations in the shape. Using this procedure, we obtain 800 voltage-current samples labelled by load.

### 3.1.2 Labelling PLAID and WHITED for Load Identification

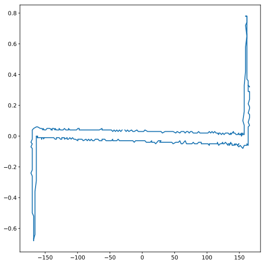
To test the models on realistic data, we also label PLAID and part of WHITED manually. This is a time consuming but nevertheless feasible task. The load of most samples can be identified manually based on (a) the physical properties of the appliance type and (b) the shape of the VI trajectory. Knowledge about the internal components and the general design of an appliance was very helpful to categorise them. For example, devices that mostly produce heat (light bulbs, toasters) typically contain a large resistor which is the most characteristic component of the



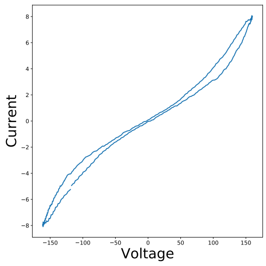

(a) Resistive Load



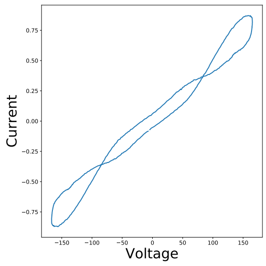

(b) Reactive Load



(c) Electronic Load




(d) Electronic Load with Power Factor Correction




(e) Complex Load

Figure 3.2: Examples of manually labelled VI trajectories. Each load displays distinct shapes. Resistive Loads (a) typically resemble an anti-diagonal line with a small enclosed area. Reactive loads (b) contain a greater area. Electronic Loads have a rotated "Z" shape with a middle segment that remains close to zero in the y-axis. Electronic Loads with Power Factor Correction also have the zero middle segment which is much smaller. Complex Loads display irregular shapes often containing self-crossings.

circuit. Similarly, electronic devices have comparable power supplies and can be easily grouped. For the rest, the VI trajectory is investigated. This involves looking at the shape and noticing the following criteria, inspired by Du *et al.* [16]:

(a) Trajectories that form clearly defined circles with big enclosed area fall into the reactive category.

(b) When trajectories are composed of a part where current remains zero, followed by a sudden increase of amperage, it is an indication of a power supply for a DC circuit.

(c) Self-crossings in the VI trajectory, or abrupt modulations, indicate complex loads.

Following this procedure we are able to label PLAID in terms of load. Example trajectories for each of the above criteria are displayed in Figure 3.2. In these trajectories we can see that resistive devices resemble an anti-diagonal line while reactive have a circular shape. As indicated by criterion (b), the electronic loads have a middle part that remains stable towards the centre of the shape. And finally the complex load of Figure 3.2e has two self-crossings.

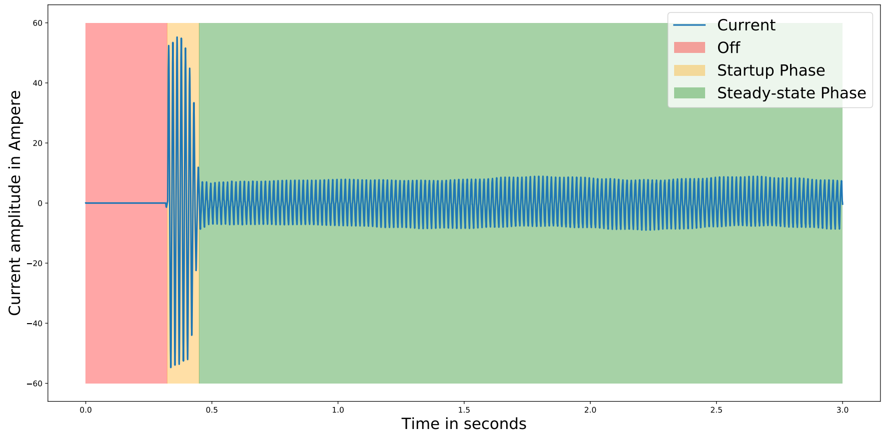



Figure 3.3: Phases of operation for an Air Conditioner. With the blue line we see the current measurement over time. In the red area on the left, the device is off and therefore the current amplitude is zero. When the device is powered-on there is an abrupt increase in the current amplitude, signifying the start-up phase in the orange segment. Finally, the current reaches the Steady-state Phase where device consumption remains relative stable, as seen in the green segment.

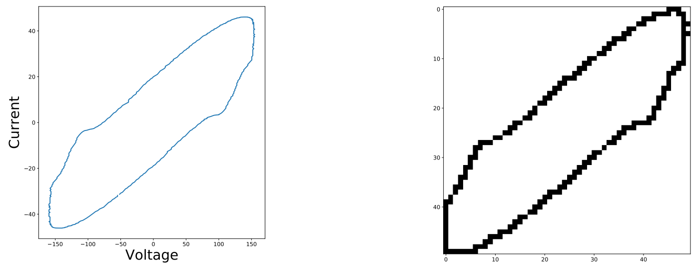
## 3.2 Data Preprocessing

All of the sources mentioned in Section 3.1 provide timeseries of voltage and current, longer than 5 seconds. The first part of the timeseries is usually a short period for which the device is off and therefore the measured current is 0. As soon as the device is switched on, there is an initial state of unexpected operation, referred to as the start-up phase[1]. During this period, the device is not functional but is preparing parts of the circuit (motor spinning up, capacitors charging etc). After some arbitrary time has passed, the circuit is expected to have reached the steady-state phase during which it operates normally. Looking at an example of an Air Conditioner in Figure 3.3 we can notice that the phases can be easily distinguished by the difference in current Amplitude.

Most previous attempts on appliance identification skip the start-up phase and focus on extracting features from the steady-state phase, which is considered to be the characteristic part of the device. Our investigation of the data seems to hint that there are fluctuations of the current even within the steady-state phase. This hints the appliance may go through different states during this operation. This is known to be true with several appliances that change behaviour as time passes. For example, an air conditioner cycles through the states of circulating cooling fluid and spinning the fans while a washing machine may be washing, spinning or pumping water. Therefore, there are cases for which we can distinguish several steady-states. This is especially useful, for Appliance Load Identification, as different states may represent different loads.

To detect transitions between states in the timeseries, we found points at which the consumption (Wattage) of the appliance changes abruptly, by an amount that exceeds a threshold. This threshold was handpicked for each appliance. A quick and clearly defined change in wattage indicates that different parts of the circuit are in-use and therefore the device has transitioned to another state. By detecting these state changes, we split the timeseries and thus get a more

[1]This phase is also mentioned as *inrush current* in the literature.




(a) Initial VI trajectory (b) Discretized VI trajectory

Figure 3.4: Mapping of a VI trajectory to a fixed a matrix of size 50x50. This process converts the continuous line into a rasterised representation which allows the image to be fed to a convolutional layer.

varied representation of the device behaviour.

## 3.3 Feature Extraction

Voltage and current signals are periodic with a frequency that depends on the mains supply of electricity (e.g. 50Hz for Europe, 60Hz for the USA). Since the signal remains relatively stable within a state, examining a small number of cycles should be enough to capture the characteristic performance of the device. For this reason, a significant part of existing literature [9, 12, 16, 21] uses a single AC cycle of data to extract features and representations. In this research project we combine the VI trajectory of a single cycle with metrics from the whole measured signal. These features are grouped in three categories: VI trajectory shape, time-domain and frequency-domain features.

VI trajectory shape features are features extracted not from the shape of the trajectory. The form of the trajectory depends on the waveform of current cycles in relation to mains voltage signal. It is a visual representation that indicates properties of the device operation. For usage with Convolutional Neural Networks, the image of the trajectory is discretized into a binary matrix of fixed dimension, as explained in [16] (see Figure 3.4). For machine learning models that do not typically handle images directly, feature extraction is necessary. In this case, a list of measures were calculated to capture the shape of the trajectory into an 1-dimensional vector.

- **Area** The area enclosed by the trajectory. It is proportional to the phase difference between voltage and current (proposed by [10]).

- **Self-Intersection** The number self-intersections that appear in the trajectory (proposed by [10]).

- **Asymmetry** A measure of detecting whether the positive and the negative part of the cycle have the same shape. This can be calculated by multiplying anti-diagonal cells (proposed by [16]).

- **Curvature of the mean line** A measure of harmonic distortion in the signal and can be calculated by measuring the vertical distance between the VI trajectory and the trajectory

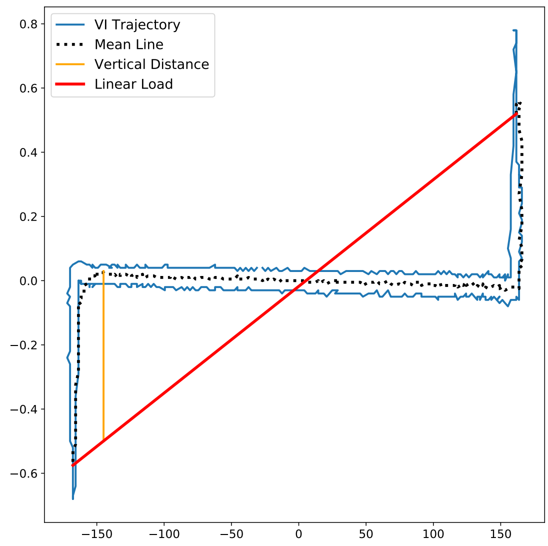



Figure 3.5: Graphical description of the curvature of the mean line for an electronic load. The mean line of the trajectory is depicted as the dotted black line. The curvature is measured as the maximum distance between a theoretical linear load (in red) and the mean line of the trajectory.

of a perfectly linear device. An example is shown in Figure 3.5 (proposed by [18]).

- **Slope of the middle segment** The middle segment of a VI trajectory is defined as the part where voltage is taking values in the range $(-\frac{1}{2}max(V), \frac{1}{2}max(V))$. The slope of the VI in this segment can distinguish electronic devices, that typically have a near-zero slope (proposed by [10]).

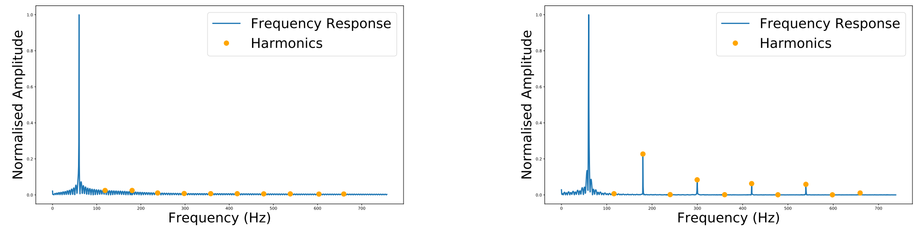
- **Peak of the middle segment** Similarly to the slope, this feature can differentiate reactive loads.

- **Width variance** The width of a segment is defined as the horizontal distance between the lines of the VI trajectory. The width can be measured in various points on the trajectory. High variance in these values is an indication of an "uneven", complex load.

The VI trajectory representation is not affected by the amplitude or the frequency of the mains supply [12]. By extension, the features extracted from its shape are very robust and easy to use across different datasets.

Time-domain features refer to properties of the signal that appear over multiple AC cycles, and are not conveyed in the VI trajectory representation.

- **Current Amplitude** This refers to the max current value that is measured within a cycle of operation. This is quite important as it can help distinguish appliances according to their consumption. One crucial point to notice here is that the max current is inversely proportional to the mains voltage, given that a device puts a steady load on the supply. As such, to make datasets from different regions compatible, we normalised the max current




(a) Light Bulb (b) Hairdryer

Figure 3.6: Frequency response of the current signal for two appliance types. On the left we see that the amplitude of all harmonics is very low relative to the fundamental frequency. This is a hint of a mainly resistive circuit. On the right we notice high amplitudes in every other harmonic, indicating the presence of a DC motor.

values with the mains voltage.

- **Phase shift** A measure that refers to the phase difference phase between voltage and current. The absolute value is conveyed in the area of the VI trajectory. However, using trajectories it is not possible to detect its sign. A positive phase shift (voltage lagging current) indicates an inductive circuit while the opposite (voltage leading current) a capacitive one. Therefore, information about the sign of the phase shift can aid in the differentiation between capacitive and inductive loads.

In the frequency domain, Kahl *et al.* [11] detected some interesting properties for some devices types. The intensity of specific harmonics imply presence of certain components. For example linear loads have very low harmonic content while motor-equipped appliances typically show high amplitudes for odd harmonics. An example of this phenomenon can be seen in Figure 3.6. Therefore, we found fitting to extract a small number of features using the Fast Fourier Transform.

- **Total Harmonic Distortion (THD)** A measure of harmonic distortion which is proportional to the amount of noise or high order components in the signal. With $f_0$ the amplitude of the fundamental frequency and $f_i$ the amplitude of the $i$-th harmonic, THD is defined as:
$$THD = \sqrt{\frac{\sum_{i=1}^{5} f_i^2}{f_0}} \tag{3.1}$$

- **Odd-Even Ratio (OER)** The ratio of odd to even harmonics can also aid the identification of a device as certain components introduce only odd or even harmonics. The OER is defined as:
$$OER = \frac{mean(f_1, f_3, f_5)}{mean(f_2, f_4, f_6)} \tag{3.2}$$

- **Spectral Flatness (SPF)**. Represents the distribution of energy in the frequency spectrum, with high values representing white noise and low values strong individual components. It is defined as:

$$SPF = \frac{\sqrt[N]{\prod_{i=1}^{5} f_i}}{\frac{1}{5}\sum_{i=1}^{f_i}} \tag{3.3}$$

The frequency response of a device is dependent of the frequency domain of the mains supply. In fact, the fundamental frequency of current wave is the frequency of the energy grid. Consequently, the features are incompatible between samples collected in different regions. To account for this, we acquire the fundamental frequency using the mains frequency and then calculate the harmonics in relation to that. With this technique, the aforementioned features are suitable for comparisons across data of different regions.

# Chapter 4

# Replication of existing studies

The first stage of this research project involves the replication of existing approaches to Appliance Identification. Besides validating the findings of other research works, this step is an opportunity to gain insight on the data and the algorithms. For this purpose, three pieces of past literature are implemented and tested using the PLAID dataset. The selection of studies encompasses techniques that use various methods for feature extraction, namely automatic feature extraction, handpicked features and VI trajectories. Moreover, they are all techniques that are considered central to the domain literature. In Table 4.1 there is some key information about each body of work. In the following sections, we outline the methodology for the replication and the results.

## 4.1 Replication of Reinhardt *et al.*

The purpose of this study is to formulate a system that is comprised of a smart meter that samples and preprocesses data, accompanied by a computational system that classifies the appliances. The application scenario of this attempt is not clear. The data used for the research are labelled according to the device type, therefore posing the problem as Appliance Type Identification. However, there is only one appliance instance per category which is present both in test and train sets. Hence, it is more akin to Appliance Instance Identification. However, in order to remain true to the original vision, we also regard this attempt as a Type Identification scenario.

The original data used for training and testing are unavailable. For this reason, PLAID is deployed as a substitute which means that there is a difference in the set of target classes. However, the features are not specific to the appliance types that Reinhardt *et al.* used, and

| **Reference** | **Identification Scenario** | **Preprocessing Methods** | **Algorithms** |
|---|---|---|---|
| Reinhardt *et al.* [6] | Type Identification | Handpicked Features | WEKA Algorithms |
| Kato *et al.* [3] | Instance Identification | PCA | SVM Classifier |
| De Baets *et al.* [19] | Type Identification | VI trajectory Images | Convolutional Neural Network |

Table 4.1: Key components of replicated studies. The studies are picked to reflect three different preprocessing and feature extraction pipelines. Each study also uses a different identification algorithm.

| Algorithm | Accuracy |
|---|---|
| Bagging | 0.86 |
| Bayesian Network | 0.78 |
| J48 | 0.86 |
| Jrip | 0.82 |
| LogitBoost | 0.82 |
| Naive Bayes | 0.64 |
| Random Committee | 0.93 |
| Random Forest | 0.93 |
| Random Tree | 0.84 |

Table 4.2: Results of replication of Reinhardt *et al.* [6] on PLAID.

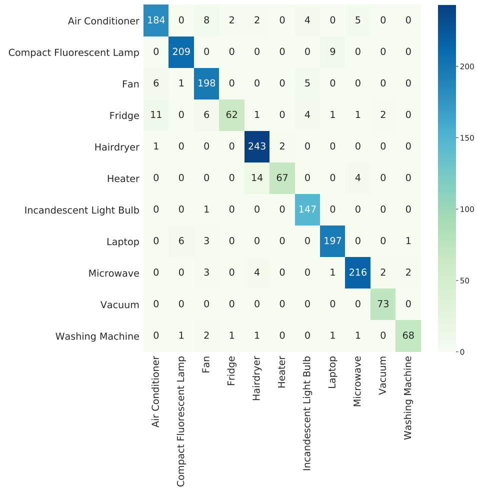



Figure 4.1: Confusion matrix for Random Forest experiment for Reinhardt *et al.* Notice that misclassification happens in patterns. This is visible in the 14 Heater examples that are identified as Hairdryer.

therefore this change should not be an significant. Next, the feature extraction process, involves extracting 10 handpicked features such as phase shift, the root mean square of the current and several features from the frequency domain. The models are created using machine learning algorithms, provided by the `WEKA`[22] software. We test this approach with 25-cross validation. Results are listed in Table 4.2 and the confusion matrix for the best attempt can be seen in 4.1.

The replication results show that there is a noticeable drop in accuracy. There are multiple reasons for this. Firstly, since we are using PLAID there are several instances per appliance type. This makes it necessary for the algorithms to learn the characteristics that apply to the whole class and not a specific device. Secondly, due to the nature of PLAID, there is a class imbalance that may have an impact on the prediction accuracy. However, from this study we deduct that the 10 handpicked features can sufficiently capture the components that differentiate appliance types. It should also be noted that, due to the train-test split of the data, samples from the same instances will be present in both sets. Therefore, we expect that the score of this experiment is higher than in the case of a test set with unseen instances.

## 4.2 Replication of Kato *et al.*

In this study, Kato *et al.* attempted to tackle the Appliance Instance Identification scenario. The novel part of this research is the automatic feature extraction method, that uses PCA on the raw current signal. The hypothesis is that PCA is able to extract the most important components from a single cycle of current data. The component vector that is extracted is then used to train an SVM classifier to differentiate between the target appliances.

The data used in the original experiments are not available and for this reason we use WHITED. WHITED offers a wide range of different appliance instances, hence it is suitable for Appliance Instance Identification. Since the specific number of PCA components is not men-

| Number of PCA Components | Accuracy |
|---|---|
| 20 | 0.85 |
| 40 | 0.87 |
| 60 | 0.88 |

Table 4.3: Results of replication of Kato *et al.* [3] on WHITED. Higher number of PCA components causes an increase in accuracy.

tioned, we execute the experiment with a range of values. All results can be found in Table 4.3. The PCA and SVM classifier were taken from the `sklearn` Python library [23].

Compared to the 99% accuracy reported in the original paper, in our experiments we find a maximum accuracy of 88%. This indicates a considerable drop of performance. Unfortunately, due to the different datasets, it is not easy to hypothesise on the cause of this. However, one point that we notice during the replication of this study is that there are often hidden biases in datasets. One critical bias that we found is that the samples of specific appliances were phase-locked. This means that all of their traces were collected with a specific phase shift. This was captured in the feature vector and resulted in very high classification accuracy. By randomising the phase, this bias is eliminated.

## 4.3 Replication of De Baets *et al.*

This approach proposed by De Baers *et al.* [9] takes advantage of the recent advances in deep learning, utilising a Covolutional Neural Network(CNN) for appliance identification. In this methodology, VI trajectory images of dimension 50x50 are created. These are then passed to a CNN of specified architecture in order to classify appliances into types. We test the network with the PLAID dataset with the “leave-one-house-out” strategy. This means that the classifier is trained on appliance types sampled from $N$ houses and tested on appliances from a different house. This is the same strategy deployed by the original research.

Averaging the results from all test houses produces an F1-score of the results in Figures 4.2 and 4.3. Our implementation scores an F1-score of 0.74 which is slightly lower than the 0.77 reported in the original paper. However the difference is not significant and may be due to the stochasticity in the training of the neural network.

While the study has been successfully replicated, we notice some warning signs in the implementation and training. The first one is that the network learns the train-set perfectly in just two epochs which is a sign of possible over-fitting. This is reinforced by the very big number of trainable parameters ($\sim$ 10 million) compared to the number of training examples ($\sim$ 20000). Therefore, we believe that this study can benefit from tests on other datasets or with more conservative architectures.

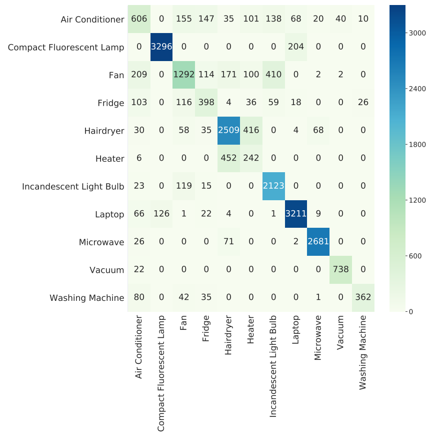



Figure 4.2: Confusion matrix for the replication experiment of De Baets *et al.*

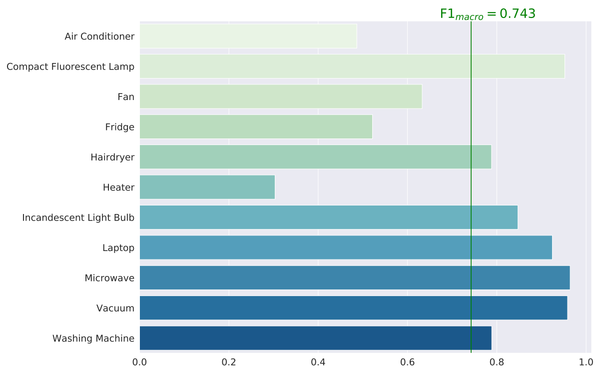



Figure 4.3: F1-score per class for the replication experiment of De Baets *et al.*

# Chapter 5

# Appliance Type Experiments

When it comes to Appliance Type Identification, this project expands on the previous literature by De Baets *et al.* [9] and Barsim *et al.* [8]. The focus in placed on high-frequency techniques that use visual representations of the source signal. For this purpose, VI trajectories and Neural Networks have been shown to perform well and are further investigated in these experiments. Working in the same manner as De Baets *et al.* in Chapter 4, we aim to improve accuracy and evaluate and ability to generalise to unseen datasets.

## 5.1 Models

Our attempts in this task stem from the architecture used in Section 4.3 and aim to remedy the weaknesses of the original. In this section we describe two models, a Multi-Modal Convolutional Neural Network and an Ensemble of Neural Networks.

### 5.1.1 Multi-Modal CNN

The first model that we propose is a Multi-Modal neural network. This architecture comes with three main enhancements. First, we fuse the VI trajectory with extra features. Second, we aim to tackle overfitting. Last, we take measures to account for class imbalances between datasets.

The first improvement, lies in the data representation itself. To overcome the limitations of the VI trajectory, we introduce the features *Maximum Wattage*, *Phase Shift*, *THD*, *OER* and *SPF*. To combine the numerical features with the 2D-trajectory matrix we propose a Multi-Modal Convolutional Network with two inputs. The trajectory matrix is fed into a stack of Convolutional and Max Pooling layers. The numerical features are fed to a fully connected layer, which is then concatenated with the output of the Convolutional layers. In this manner, the two input branches are merged into one and produce a latent space representation. This technique is widely used in sensor fusion, to combine data collected with from different sensors [24].

An important point that is mentioned in Section 4.3, is that the network seems to overfit to the training set, due to its disproportionately large number of parameters. In many cases, decreasing the layer sizes of a large network can act as a regularisation technique [25] as the network is forced to extract generic features from the data. Using this rationale, we reduce the sizes of dense layers and lower the number of filters of the convolutional layers. In addition, $l2$ regularisation is applied on Convolutional and Dense layers.

A small but substantial point in the design of the model, is noticing the class imbalance in the training set. While this has been pointed out before in the literature [19], no counter measures

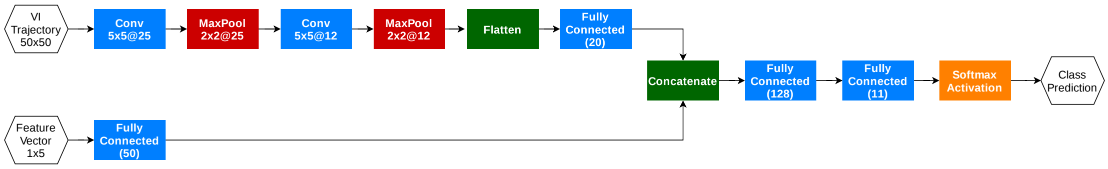



Figure 5.1: Architecture of proposed Multi-Modal Convolutional Neural Network. The network has two input layers that accept different data representations. The two branches are concatenated deeper in the network to produce a single prediction.

have been applied. For this reason, we introduce class weights depending on the number of samples per class. The weights are computed using

$$w_i = \frac{N_i}{\sum_{j=1}^{L} N_j} \tag{5.1}$$

where $w_i$ is the weight for class $i$, $N_i$ the number of examples per class and $L$ the total number of classes.

The resulting architecture can be seen in Figure 5.1. The network is trained using Stochastic Gradient Descent with a learning rate of 0.1 and momentum of 0.8.

### 5.1.2 Ensemble of Multi-Modal CNNs

Examining the Appliance Type Identification problem from a practical point of view, hints that the definition of the target classes is problematic. When working with a dataset there is a defined set of appliance types. However, in a realistic setting, every household has its own distinct set of appliances which may be larger or smaller than the one used during training. Having a model that can infer more appliance types than those that are present may be unnecessarily complex and cause false positives. At the same time, if there is a device in the test set that is unknown to the model, there is no way to reject it. Hence, it is convenient to have a flexible set of appliance types and, as such a variable number of outputs.

In order to get a model that can be adjusted to the set of appliance types, we assembled several binary classifiers. Each of the binary classifiers is trained to recognise only one type of appliance. In this way, it is possible to remove redundant classifiers or add more. This method of creating separate binary models is referred to as *one-vs-all* classification and it is commonly used in multi-class tasks [26]. The architecture of each binary classifier is the same as the one specified in Figure 5.1. The final class prediction is done by selecting the most confident positive prediction of all classifiers.

## 5.2 Experiment setup

To evaluate the performance of the two proposed models, we use the PLAID and WHITED datasets. We consider as target classes the 11 appliance types available in PLAID. We train and test using the leave-one-house-out method, as described by De Baets *et al.* This means that, we can pick one house for testing and then train on the rest of the data. There are 64 houses in datasets PLAID and PLAID 2 and therefore, 64 experiments are to be run. However, due to time constraints we test only on the 9 houses of PLAID 2. WHITED is also used to test

| Model | PLAID 2 | WHITED |
|---|---|---|
| Baseline by De Baets *et al.* | 0.57 | 0.32 |
| Multi-Modal CNN | 0.62 | 0.48 |
| Ensemble of Multi-Modal CNN | 0.66 | 0.47 |

Table 5.1: F1-score for Appliance Type Identification experiments. All models are trained on PLAID and tested on PLAID 2 and WHITED. The scores can be compared to evaluate the performance.

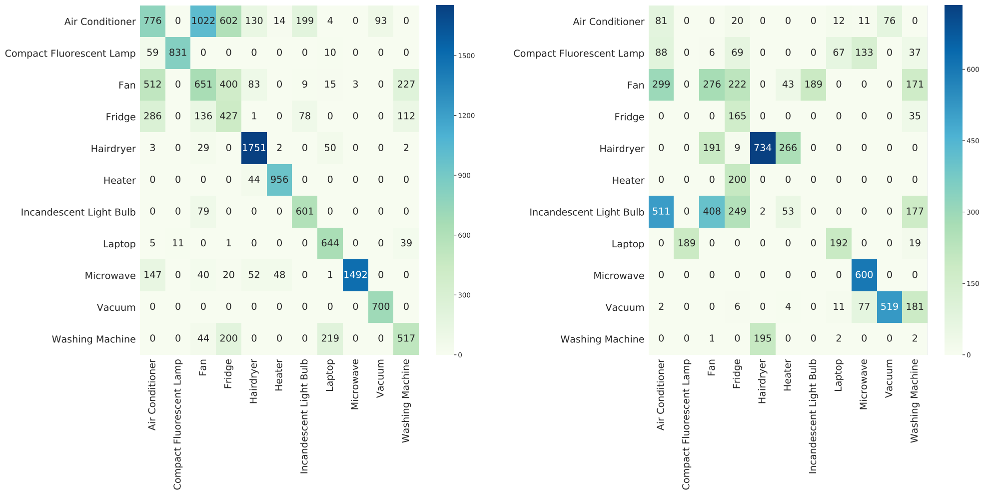
for generalizability to other datasets. The data in WHITED are not split in houses and so, we picked some of them to create an artificial house. The list of appliances in the artificial house can be found in Appendix B. The performance of all networks is evaluated using F1-score in order to take into account the imbalances of the test set.

## 5.3 Results

Executing the above experiments yields the scores that are listed in Table 5.1. In addition to the F1-score, we also include the confusion matrices that are displayed in Figures 5.2-5.4. These can help detect patters in the classification errors and allow for observations regarding the weak points of each model.

Inspecting the results we notice that there is a clear performance boost in the architectures that we proposed on both datasets, compared to the baseline. We manage to reach the highest score of 0.66 for PLAID 2 and 0.48 for WHITED. This is significantly higher compared to the scores of the baseline. Between the two classifiers that we propose, there is no clear winner as the Multi-Modal CNN performs better on WHITED while the Ensemble scores higher on PLAID 2. Overall, the WHITED set remains the hardest one to classify and, clearly, there is a large gap in the performances for each dataset.

Looking at the confusion matrices for WHITED we can notice interesting patterns. First, there are certain devices that are never recognised correctly over all experiments. Namely, the class “Compact Fluorescent Lamp” is misclassified in a distinct way for each of the classifiers. Similarly, the class “Heater” is universally regarded as “Hairdryer”. Second, the devices “Incandescent Light Bulb” and “Washing Machine”, are recognised much better with the proposed Multi-Modal Network (Figure 5.3b) than with the baseline (Figure 5.3b).




(a) PLAID (b) WHITED

Figure 5.2: Confusion matrices for baseline.


(a) PLAID (b) WHITED

Figure 5.3: Confusion matrices for Multi-Modal CNN.


(a) PLAID (b) WHITED

Figure 5.4: Confusion matrices for Ensemble of Multi-Modal CNN.

# Chapter 6

# Appliance Load Experiments

Until this point, Appliance Load Identification was investigated through clustering techniques [4, 17]. These methodologies do not explicitly identify loads. Instead they attempt define load categories as observed through clusters of collected data. This is to be expected as there are no available open datasets that provide labels. Since we have generated artificial data and hand-labelled an existing real-world dataset, it is now possible to deploy classification techniques on the task of Appliance Load Identification.

## 6.1 Models

For this problem, we construct three models using K-Nearest-Neighbors, Random Forest and Convolutional Neural Networks. The first step is to establish a baseline result, using a 10-Nearest Neighbours classifier. The input to this classifier is the VI shape, time-domain and frequency features, as described in Section 3.3. The second model is a Random Forest, as an attempt to achieve higher accuracy. This way we can evaluate the performance of a more sophisticated classifier that uses only the most descriptive features. Finally, a simple Convolutional Neural Network is implemented in order to see how a classifier would perform on Appliance Load Identification without explicit feature extraction. To experiment on this, we feed the network with VI trajectory images. The architecture in detail is outlined in Figure 6.1. It is trained using Stochastic Gradient Descent with a learning rate of 0.1 and momentum of 0.8.

## 6.2 Experiment setup

For Appliance Load Identification experiments we use the artificial dataset and the hand-labelled samples obtained as described in Section 3.1. The first step is to measure the performance on the artificial dataset which is split using 5-fold cross validation. For the real-world data, we follow

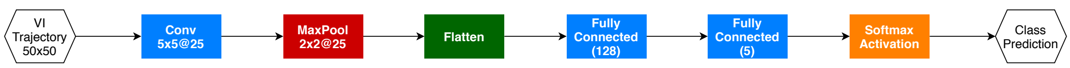
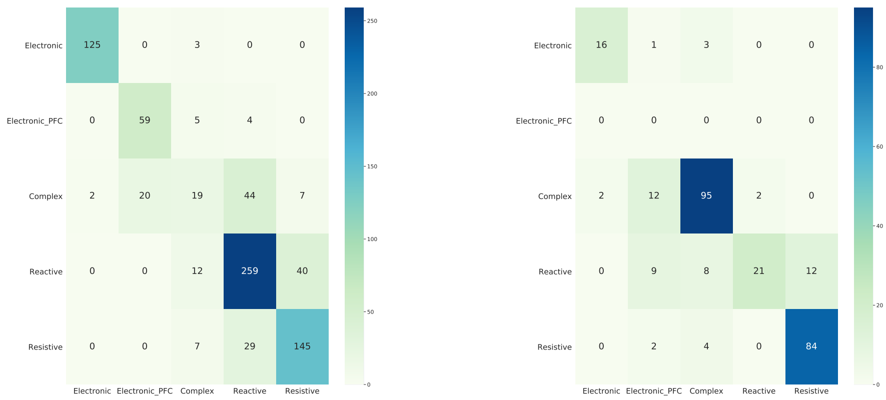



Figure 6.1: Architecture of proposed Convolutional Neural Network for Load Identification.

| **Model** | **Artificial Dataset** | **PLAID 2** | **WHITED** |
|---|---|---|---|
| 10-Nearest Neighbours | 0.94 | 0.73 | 0.63 |
| Random Forest | 0.98 | 0.84 | 0.77 |
| CNN | 0.94 | 0.87 | 0.73 |

Table 6.1: F1-score for Appliance Load Identification experiments.




(a) PLAID (b) WHITED

Figure 6.2: Confusion matrices for 10-Nearest Neighbours.

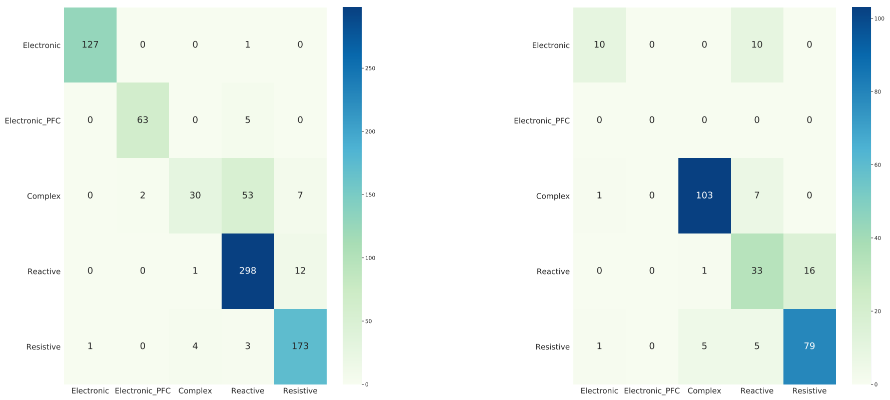
the same leave-one-house-out strategy that is mentioned in Chapter 5. Again, F1-score is used as the evaluation metric.

## 6.3 Results

Running the above experiments produces the results that are listed in Table 6.1. Confusion matrices are displayed in Figures 6.2-6.4.

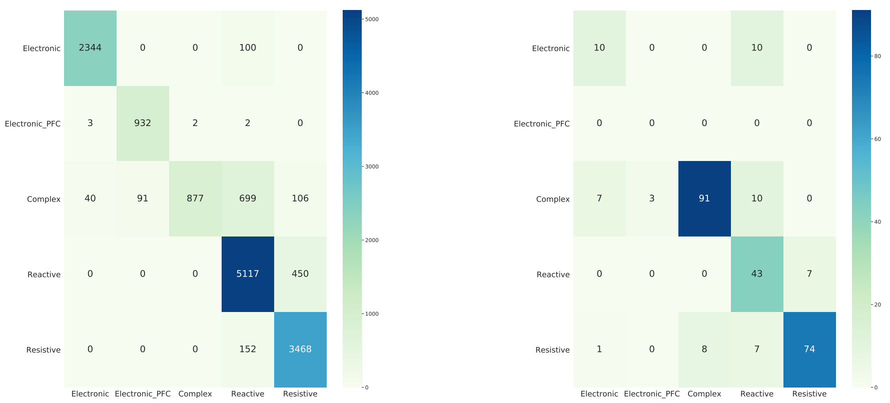
Studying the results dataset-wise, we initially see that all models achieve a near-perfect score on the artificial dataset. This is to be expected as it consists of simple circuits that should be easily recognisable. For PLAID 2 the best result is achieved by the Convolutional Neural Network with an F1-score of 0.87. Compared to the result of 10-Nearest-Neighbours, we see a considerable performance boost. For WHITED, the top score is 0.77 with the Random Forest Classifier which is again considerably higher that the baseline. Overall, we see that all performances in WHITED suffer, compared to PLAID 2, much like in Chapter 5. But even in this case, the Random Forest and the Neural Network perform similarly on each dataset.

Looking at the confusion matrices for PLAID 2, we see that the Random Forest and Neural Network models manage to tackle the misclassification for the “Resistive” class. Nevertheless, we see that all three models suffer with the “Complex” class. In the case of WHITED, we notice that all classifiers struggle with the “Electronic” and “Reactive” classes. This is probably an indication that the devices in the test set are significantly different than the devices of the same class in the train set.




(a) PLAID (b) WHITED

Figure 6.3: Confusion matrices for Random Forest.




(a) PLAID (b) WHITED

Figure 6.4: Confusion matrices for CNN.

# Chapter 7

# Discussion

In this section we discuss the results of the previous chapters and the main findings that can be derived from them. Subsequently, using these findings we give answers to the research questions. In addition, we reflect on other interesting points encountered during the research process that may be relevant for the problem of appliance identification.

## 7.1 Generalisability of Type and Load Identification

In Section 1.5 we set to find if existing appliance identification methods perform comparably in datasets other than the ones used in the original research. The purpose of this question is to see whether the reported results actually reflect the capabilities of the proposed methods on the task or specifically on the testing set. As explained in the Chapter 1, the effect of generalisation is important for both Appliance Identification scenarios but especially for Appliance Type Identification.

In Chapter 5 we evaluate the model by De Baets *et al.* on two additional datasets: PLAID 2 and WHITED. The network is trained in the same way as in the original literature and therefore it should appear to have comparable performance. Instead we see a significant drop in F1-score when comparing with results of Section 4.3. This means that, deploying the proposed model in real-world data would probably produce unreliable predictions. The techniques that we propose for Appliance Type Identification manage to reduce that gap of accuracy but there is still a noticeable drop from PLAID 2 to WHITED. We hypothesise that this may happen due to biases introduced by the sampling equipment or the potentially low within-class variety of devices. Another possible explanation is that the classifier fits to the distribution of appliance types of the train set. As such, testing with a different class imbalance is producing inaccurate prediction. That being said, both of these hypothesis need to be tested in order to find the cause of this weakness.

In regards to Appliance Load Identification we see comparable results in real-world datasets despite the simple feature extraction and algorithms. There is good reason to believe that this happens due to the clearly defined classes that reflect the physical properties of the device. The characteristics of each class are universal to all devices that belong to it. Hence we expect it to be is easy for the classifier to capture the properties of the whole set based on a small sample.

## 7.2 Visual representations for Appliance Identification

The second research question aims to examine the visual representations of consumption data and their capacity for Appliance Identification. In this research project we focus on the VI trajectory representation that is both intuitive for humans and conveys information that past literature considers crucial for the task. In the Type Identification scenario we see that most existing literature opt for this representation for use with Convolutional Network. We see, though, that the addition of certain features contributes to increased performance, as seen with the Multi-Modal CNN and its ensemble counterpart. This hints that enhancing the VI trajectory with extra information can improve performance.

For the Appliance Load Identification experiments in Chapter 6, we notice that there is negligible difference in F1-score between the CNN model and the techniques that use handpicked features. From this we can deduct that the visual representation sufficiently captures all important components in order to differentiate load classes. To conclude, we believe that for the Load Identification scenario, visual representations appear to perform on par with handpicked features.

## 7.3 Feasibility of Load Identification

The third and final research question posed in Section 1.5 pertains the feasibility of an Appliance Load Identification system. This is examined by first differentiating between small-circuit loads in the artificial dataset and then working with realistic appliance consumption data. We see that all classifiers manage to achieve F1-score higher than 0.9 which hints that the classes are sufficiently separable and cohesive. This is a first indication that separating loads of small circuits is task that can be easily tackled.

Experimenting with real-world data yields consistent results around 0.8 for PLAID 2 and 0.7. This means that Load Identification on realistic data is not perfect but possible. Inspecting the confusion matrices in Figures 6.3 and 6.3 we notice that there is a consistent error for both classifiers in the Reactive and Resistive classes. This is an indication that either the classes are ill-defined or that some of the defining characteristics of the class are not present in the training set. Nevertheless, the task seems feasible using simple techniques, either with handpicked features or Convolutional Networks.

## 7.4 Other findings

A finding that was neither expected nor a part of the research goals is the strong dependency of the results on the training and testing data. This is visible in the replication studies in particular, as the results are very different from the ones reported in the literature. In addition, we can also find evidence of this in the big disparity between F1-scores of similar datasets such as PLAID and PLAID 2. This effect, combined with the unclear appliance identification scenarios, makes it very hard to understand the findings and deploy appliance identification methods. Especially in the Type Identification scenario, all of the results seem to be highly dependent on the data distribution, the target classes and the variety of devices.

The second interesting finding regarding Appliance Type Identification is that there are hints that the "type" classes might be ill-defined. Grouping appliances in terms of their functionality might seem reasonable from the perspective of the user. However there is no guarantee that internal structure of appliances of the same type will be the similar. The classes might end-up being overly broad (e.g. space heaters might contain fans, thermostats etc) or extremely specific (e.g. LED TVs and PC Monitors are essentially the same device). This loose definition of "types" makes the task even harder, because different datasets can offer slightly different descriptions. However it is problem that does not seem to apply to Appliance Load Identification, in which, classes directly reflect the physical properties of the device.

## 7.5 Ethical Concerns on Appliance Identification

When one contemplates the type of data that are collected and generated through Energy Monitoring techniques, it is easy to notice a potential privacy risk. Energy consumption data can be used to infer behavioural patterns of the occupants by monitoring their interactions with their appliances and their location in their own home. This ever-present collection of measurements through sensors and smart devices results in an abundance of sensitive information. Often the user is unaware of this process. Even worse, the produced data are not typically under the user's ownership. Such technology deployed in large scale, either by energy companies or smart tech, can pose a major security threat and ethical issues on a personal and societal level.

More often than not, the burden of considering the ethical implications of technology falls onto philosophers, legislators and the companies. While the aforementioned parties have a important role to play, we find it problematic that researchers often fail to take a stance regarding their findings. In our case, Automatic Appliance Identification is of high interest as it aims to convert raw, unlabelled power consumption data into a detailed report of what devices are being used at any moment. More importantly, this is a highly applied research field and in close relationship with the commercial technology. For this reason, we believe privacy should be an essential part of the discussion when working with Appliance Identification.

Privacy consciousness needs to be reflected in the application scenarios as well as in research methodology. In this particular task, proposing computationally efficient models, that can be deployed locally, is crucial to enable privacy by design. This, combined with hardware such as Crownstones that respect the ownership and confidentiality of the data, can be of great benefit to the end user. On the contrary, Appliance Identification techniques that require huge computational resources are more likely to be a tool for the commercial institutions that can afford the cloud infrastructure. In this case, the data are more likely to end up outside the user's reach. Under the same premise, we believe that the use of open datasets will mitigate the incentive to massively collect data. Finally, publishing research findings with transparency is major component to democratise the knowledge. By following these simple steps, we believe that it is possible to research sensitive areas such as Appliance Identification while avoiding ethical complications.

# Chapter 8

# Conclusion

The goal of this thesis was to investigate the state of Appliance Identification, the limits of existing approaches and the potential for improvement. We have seen that while there is a wide range of proposed methods that tackle the task, there is also confusion on how these methods can be applied practically. We have proposed a distinction of three Appliance Identification scenarios that can serve as a useful tool for solidifying the experiment setup and allowing for fair comparisons.

In addition to the theoretical study and the replication experiments, we have proposed a set of novel models that improve upon the existing literature and achieve performance. More specifically, the Multi-Modal CNN managed to perform significantly higher than baseline. The Ensemble of Multi-Modal CNN not only had performance equivalent to its counterpart but also allowed a flexible classification scheme. Finally, the Appliance Load Identification act as noteworthy baseline models for the task. We believe that this contribution will be of great use to future research on the newly-introduced scenario.

To sum, the key findings of this research work are:

- Experiments on Appliance Type Identification are very dependent on the train/test data.
- Combining VI trajectories with features that are not implied by the trajectory can increase the performance.
- Appliance Load Identification is feasible, given a set of labelled data to train on.
- Appliance Load Identification techniques can be formulated both with handpicked feature extraction and VI trajectory images.

## 8.1 Future Work

Given the findings of this research project we find that there is great room for improvement. Our first proposal is to explicitly define the identification scenario. When the scenario is clearly mentioned it is easier for the research community to understand the proposed methods as well as the results. Moreover, it allows for a fair comparison since there are no incompatibilities between the target classes. For this reason we strongly propose that future works explore the boundaries of the three scenarios and attempt to establish a standard testing methodology for each one.

Our second suggestion is about the Appliance Load Identification scenario and how it can be a great addition to the Appliance Identification task. While we have showed that identifying appliances with these five specific classes is possible, we are interested in seeing what other loads can be found in household appliances. A more specific set of load classes might help understand better the internal elements of the identified appliance. This way, the end user can receive more fine-grained feedback than a generic class label.

Finally, we see that our proposed Load Identification methods aim to categorise the behaviour of the circuit without explicitly inferring its internal structure. However, we believe that there is potential to understand the device circuit using the data as input. This problem is very akin to the task of Blind System Identification that falls under the topic of dynamical systems. We know that circuits can be modelled by equations using the Laplace Transform. By defining this a set of equations, one per load class, we believe that it would be possible to identify appliances with blind system identification methodology and tools.

# Bibliography


[1] G. W. Hart. “Nonintrusive appliance load monitoring”. In: *Proceedings of the IEEE* 80.12 (Dec. 1992), pp. 1870–1891.

[2] A. Ridi, C. Gisler, and J. Hennebert. “A Survey on Intrusive Load Monitoring for Appliance Recognition”. In: *2014 22nd International Conference on Pattern Recognition.* Aug. 2014, pp. 3702–3707.

[3] Takekazu Kato et al. “Appliance Recognition from Electric Current Signals for Information-Energy Integrated Network in Home Environments”. en. In: *Ambient Assistive Health and Wellness Management in the Heart of the City.* Ed. by Mounir Mokhtari et al. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009, pp. 150–157.

[4] Liang Du et al. “Nonintrusive, Self-Organizing, and Probabilistic Classification and Identification of Plugged-In Electric Loads”. In: *IEEE Transactions on Smart Grid* 4.3 (Sept. 2013), pp. 1371–1380.

[5] Shyh-Jier Huang et al. “Classification of home appliance electricity consumption using power signature and harmonic features”. In: *2011 IEEE Ninth International Conference on Power Electronics and Drive Systems.* Dec. 2011, pp. 596–599.

[6] A. Reinhardt et al. “Electric appliance classification based on distributed high resolution current sensing”. In: *37th Annual IEEE Conference on Local Computer Networks - Workshops.* Oct. 2012, pp. 999–1005.

[7] S. K. K. Ng, Jian Liang, and J. W. M. Cheng. “Automatic appliance load signature identification by statistical clustering”. en. In: (Jan. 2009), pp. 38–38. (Visited on 09/18/2019).

[8] Karim Said Barsim, Lukas Mauch, and Bin Yang. “Neural Network Ensembles to Real-time Identification of Plug-level Appliance Measurements”. In: *arXiv:1802.06963 [cs, eess]* (Feb. 2018). (Visited on 09/24/2019).

[9] Leen De Baets et al. “Appliance classification using VI trajectories and convolutional neural networks”. In: *Energy and Buildings* 158 (Jan. 2018), pp. 32–36. (Visited on 09/23/2019).

[10] T. Hassan, F. Javed, and N. Arshad. “An Empirical Investigation of V-I Trajectory Based Load Signatures for Non-Intrusive Load Monitoring”. In: *IEEE Transactions on Smart Grid* 5.2 (Mar. 2014), pp. 870–878.

[11] Matthias Kahl et al. “A Comprehensive Feature Study for Appliance Recognition on High Frequency Energy Data”. In: *Proceedings of the Eighth International Conference on Future Energy Systems.* e-Energy ’17. New York, NY, USA: ACM, 2017, pp. 121–131. (Visited on 10/22/2019).

[12] Abeer A. Kholeif, Hossam A. Abd El-Ghany, and Ahmed M. Azmy. “Impact of supply voltage variation on V-I trajectory identification method”. In: *2017 Nineteenth International Middle East Power Systems Conference (MEPCON).* Dec. 2017, pp. 839–844.

[13] Matthias Kahl et al. "WHITED-A Worldwide Household and Industry Transient Energy Data Set". In: 2016.

[14] Jingkun Gao et al. "PLAID: A Public Dataset of High-resoultion Electrical Appliance Measurements for Load Identification Research: Demo Abstract". In: *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. BuildSys '14. New York, NY, USA: ACM, 2014, pp. 198–199. (Visited on 09/19/2019).

[15] C. Gisler et al. "Appliance consumption signature database and recognition test protocols". In: *2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*. May 2013, pp. 336–341.

[16] Liang Du et al. "Electric Load Classification by Binary Voltage–Current Trajectory Mapping". In: *IEEE Transactions on Smart Grid* 7.1 (Jan. 2016), pp. 358–365.

[17] H. Y. Lam, G. S. K. Fung, and W. K. Lee. "A Novel Method to Construct Taxonomy Electrical Appliances Based on Load Signaturesof". In: *IEEE Transactions on Consumer Electronics* 53.2 (May 2007), pp. 653–660.

[18] Nur Iksan et al. "Appliances identification method of non-intrusive load monitoring based on load signature of V-I trajectory". In: *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*. Nov. 2015, pp. 1–6.

[19] Leen De Baets et al. "Handling imbalance in an extended PLAID". eng. In: *2017 Fift IFIP Conference On Sustainable Internet And ICT for Sustainability (SUSTAINIT 2017)*. 2017, pp. 32–36. (Visited on 10/21/2019).

[20] *Ngspice, the open source Spice circuit simulator - Intro.* URL: http://ngspice.sourceforge.net/ (visited on 03/26/2020).

[21] Darío Baptista et al. "Implementation Strategy of Convolution Neural Networks on Field Programmable Gate Arrays for Appliance Classification Using the Voltage and Current (V-I) Trajectory". en. In: *Energies* 11.9 (Sept. 2018), p. 2460. (Visited on 12/31/2019).

[22] Mark Hall et al. "The WEKA data mining software: an update". In: *SIGKDD Explorations* 11.1 (2009), pp. 10–18.

[23] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[24] Sebastian Münzner et al. "CNN-based sensor fusion techniques for multimodal human activity recognition". In: *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ISWC '17. Maui, Hawaii: Association for Computing Machinery, Sept. 2017, pp. 158–165. (Visited on 05/13/2020).

[25] Chiyuan Zhang et al. "Understanding deep learning requires rethinking generalization". In: *arXiv:1611.03530 [cs]* (Feb. 2017). (Visited on 10/25/2019).

[26] Mikel Galar et al. "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes". en. In: *Pattern Recognition* 44.8 (Aug. 2011), pp. 1761–1776. (Visited on 05/14/2020).

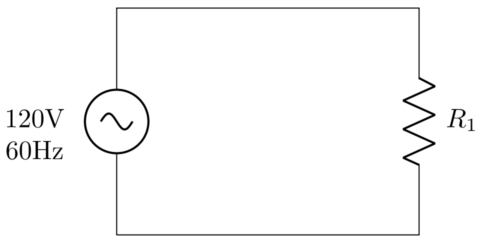
# Appendices

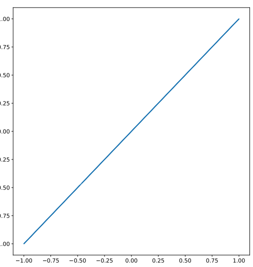
# Appendix A

# Circuits used in Artificial Dataset

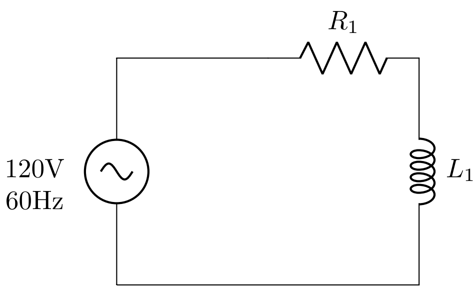
In this chapter you can find the schematics and characteristic VI trajectories for the 8 circuits, used in the artificial load dataset as described in Section 3.1.1.

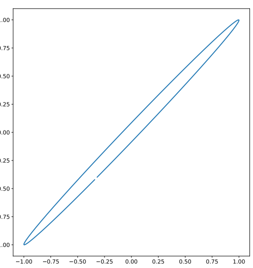
## Resistive Circuit - Resistive Load



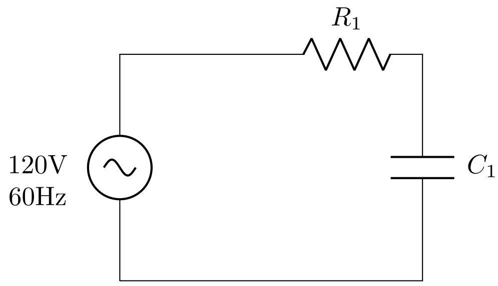


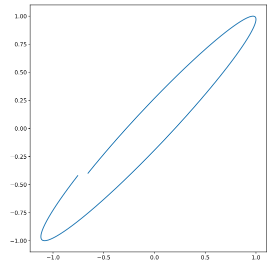
## RL Circuit - Reactive Load



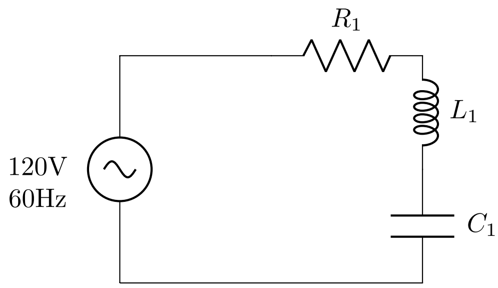


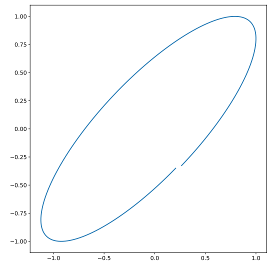
## RC Circuit - Reactive Load



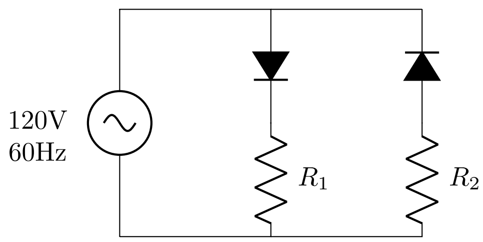



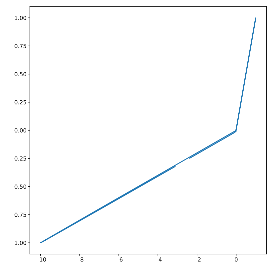
## RLC Circuit - Reactive Load



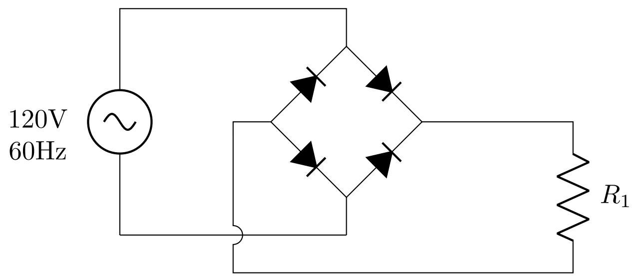



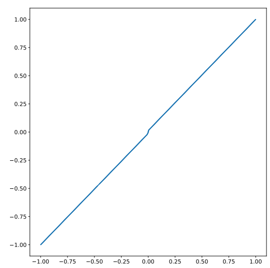
## Asymmetrical Resistive Circuit - Resistive Load



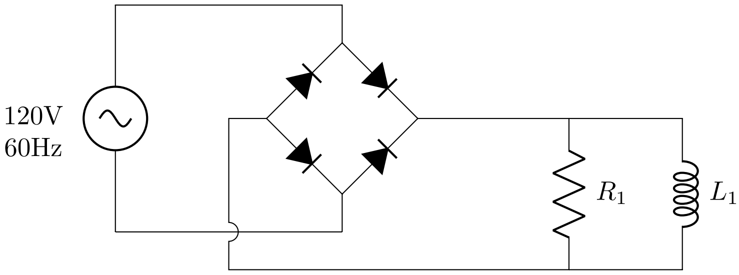



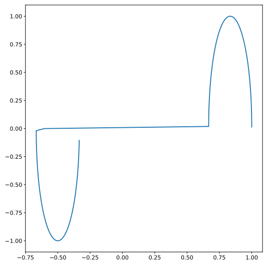
## Resistive Circuit with a rectifier - Electronic Load



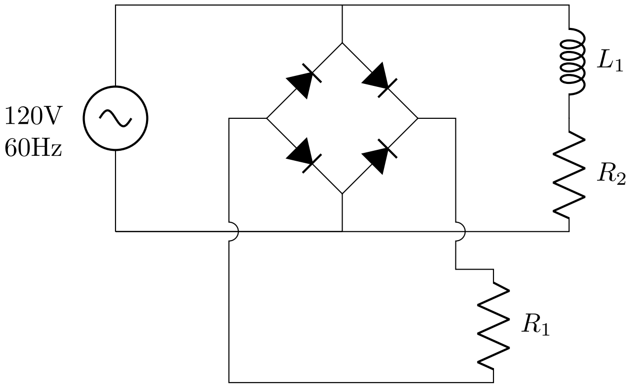



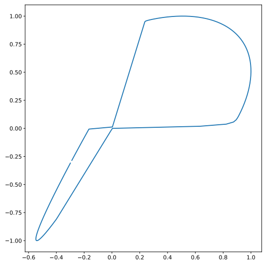
## RL with a rectifier - Electronic Load



## Multi-Draw Circuit - Complex Load



# Appendix B

# Artificial house in WHITED

Devices in the WHITED artificial house for Appliance Type Experiments.

| **Appliance Type** | **Device Name** | **Region** |
|---|---|---|
| Air Conditioner | Electrolux | r5 |
| Compact Fluorescent Lamp | IKEA15W | r1 |
| Compact Fluorescent Lamp | PhilipsGenie8W | r6 |
| Fan | Cyclone3000 | r4 |
| Fan | HoneywellCL25AE | r5 |
| Fan | Salco-STT23-1 | r2 |
| Fan | ChingHai35W | r6 |
| Fan | Krisbow50W | r5 |
| Fan | VOV-50W | r1 |
| Fridge | Danby | r8 |
| Hairdryer | Tedi | r3 |
| Hairdryer | RemingtonD5000 | r3 |
| Hairdryer | BaBylissPro | r6 |
| Hairdryer | RemingtonD3090 | r1 |
| Hairdryer | BraunSatinHair | r3 |
| Hairdryer | PhilipsSalonDryTravel | r5 |
| Heater | Heller | r1 |
| Incandescent Light Bulb | Osram-25W | r1 |
| Incandescent Light Bulb | PhilipsClassicTone40W | r7 |
| Incandescent Light Bulb | Tungsten-40W | r3 |
| Incandescent Light Bulb | Vintage-40W | r1 |
| Incandescent Light Bulb | Halogen-30W | r3 |
| Incandescent Light Bulb | Osram-100W | r1 |
| Laptop | Schenker-W503 | r1 |
| Laptop | Lenovo-IdeapadS11 | r8 |
| Microwave | Sharp | r3 |
| Microwave | Privileg8020 | r3 |
| Microwave | Whirlpool | r8 |
| Vacuum | Nilfisk | r1 |
| Vacuum | Maximus | r5 |
| Vacuum | Vento | r1 |
| Vacuum | Siemens2500W | r3 |
| Washing Machine | Privileg | r2 |