



Pázmány Péter Katolikus Egyetem
Információs Technológiai és Bionikai Kar

TDK Dolgozat

**Smart Plug with multi-load detection
capabilities**
**Több fogyasztó együttes felismerése okos
konnektorral**

Dániel István Németh
Computer Science Engineering MSc

2023

Advisor:
dr. Kálmán Tornai

Contents

1	Introduction	5
2	Literature review	6
2.1	Electrical Load Classification	6
2.1.1	NILM	6
2.1.2	Multi-label classification	6
2.1.3	ILM	8
2.2	Use cases	9
2.3	Summary	10
2.4	Previous work – Dimmer-based Smart Plug	10
3	Measured loads and dataset	15
3.1	Motivation	15
4	Process design	18
4.1	Data collection	18
4.2	Multi-class classification	19
4.3	Performance measures	20
5	Discussion of results	21
5.1	Multi-class classification	21
5.2	Evaluation of generalization capabilities	25
5.2.1	From single-load to multi-load	25
5.2.2	From multi-load to single-load	31
5.2.3	From multi-load to multi-load	34
6	Conclusions	35
6.1	Future work	36
	Acknowledgements	36

Kivonat

A megújuló energiaforrások elterjedésével egyre nehezebb feladat az elektromos hálózat terhelésének és termelésének kiegyensúlyozása. A folyamatosan változó terhelés mellett a termelés oldalán is megjelentek a nem vezérelhető nap- és szélenergiaforrások. Ezen megújuló energiaforrások kihasználásának maximalizálásához szükség van a hálózat fogyasztói összetételének ismeretére, valamint a fogyasztók vezérlésére, ütemezésére. A hálózat fogyasztóinak ismerete lehetőséget ad a hálózat terhelésének előrejelzésére. Bizonyos fogyasztók működésének ütemezésével a hálózat terhelése időzíthető, lehetővé téve a megújuló energiaforrások kihasználásának maximalizálását.

A kutatás korábbi lépéseiben megépítésre került egy okos konnektor prototípus, mely egy dimmer segítségével karaktersztikus adatokat gyűjt a csatlakoztatott fogyasztóról, majd ez alapján felismeri a fogyasztót. Az korábbi eredményekből TDK dolgozat és nemzetközi konferenciára cikk is készült. Ezen TDK dolgozatban az kerül megvizsgálásra, hogy hogyan lehetséges több, azonos okos konnektorra csatlakoztatott fogyasztót felismerni. Ez egy fontos probléma a valós környezetben történő alkalmazáshoz, hiszen nem várható el, hogy egy okos konnektorra mindig csak egyetlen eszköz legyen csatlakoztatva. Emellett olcsóbb megoldás, ha kevesebb okos konnektorral is megvalósítható a rendszer a fogyasztószintű vezérlési képesség csökkentése mellett.

A multi-class klasszifikációhoz először mérési adatok gyűjtésére volt szükség. Összesen 45 különböző fogyasztó-kombinációval készültek mérések. A klasszifikációhoz egy CNN alapú háló került alkalmazásra, ahol a kimeneten nem csak az egyes fogyasztókhoz voltak neuronok, hanem a bemeneti mérésben jelen lévő fogyasztók számára is adott becslést a háló. A két predikciót kombinálva jön létre a multi-class klasszifikációra képes megoldás mely 97% feletti átlagos pontosságot (accuracy) ért el. A bemutatott módszer korlátait és képességeit is elemeztem.

Abstract

With the spread of renewable, non-controllable energy resources, balancing the production and demand in the electrical grid becomes even more challenging. In addition to the constantly changing demand, non-controllable energy supplies like wind and solar power introduce changes and unknowns to the production-side as well. In order to best utilize these resources, grid load composition data is needed to forecast grid load. The ability to schedule individual loads is also required for Demand-Side Management. By scheduling certain loads, the grid load can be shifted to enable optimal utilization of renewable non-controllable resources.

In the previous steps of the research, we built a Smart Plug prototype. The plug uses a dimmer to extract characteristic information from the connected appliance by measuring its reaction to the dimming. This characteristic data can be used to detect the load that is connected to the plug. Previous results were presented both in conference papers and in TDK reports. The topic of this TDK report is detecting multiple loads connected to the same plug. In a real-world scenario, multiple loads may be plugged into the same Smart Plug (e.g. by using an extension cord). Multi-load detection also allows the reduction of installation costs by not requiring a Smart Plug for each load at the expense of individual load-level controllability.

The first step to achieve multi-load detection was to collect measurement data. 45 different load combinations were measured in total. A CNN-based Neural Network was used for classification, but the output was modified. In addition to the neurons corresponding to each load, the network also predicts the number of loads present. Multi-class classification is achieved by combining the two predictions. The method achieved over 97% average accuracy. In addition to these results, the generalization capabilities and limitations of the method were also examined.

1 Introduction

With the recent spread of renewable energy resources in the energy market, balancing the production and consumption-side is becoming more and more challenging. On the grid-level, energy providers need information about the load composition of the grid to enable load forecasting. Demand-Side Management is also needed to allow for better utilization of renewables.

On the household level, rooftop solar panel installations strain the grid more. Scheduling the loads on a household level, considering renewable energy production, requires the ability to control individual loads and identify these loads connected to the grid. Smart meters measure at a single point, and thus disaggregation is needed. Individual load control is also not possible this way. Our previous work proposed a Smart Plug capable of identifying the connected electrical load within seconds using a dimmer. The dimmer also enables the Smart Plug to turn the load off and on. In a real-world scenario, however, more than one load may be connected to the Smart Plug. This report examines what happens if more than one load is connected to the Smart Plug. We evaluate the possible methods to distinguish individual loads if multiple are connected to the same Smart Plug. The structure of a system in which our Smart Plug and load detection method can be utilized is shown in Figure 1.1.

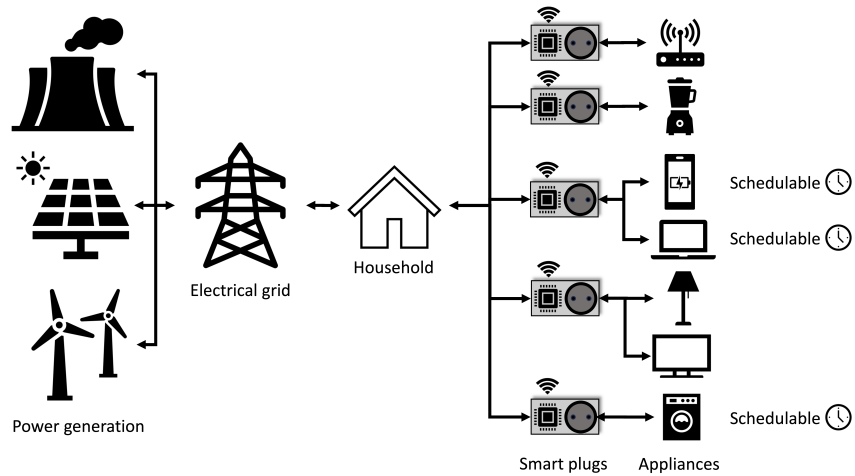


Figure 1.1: Structure of a Demand-Side Management system utilizing Smart Plugs

2 Literature review

This chapter contains a review of relevant publications in the literature. The first part focuses on electrical load classification methods, while the second part summarizes the application potentials of such systems.

2.1 Electrical Load Classification

In electrical load classification methods, two major categories are distinguished. Non-Intrusive Load Monitoring (NILM) typically utilizes smart meters, where the load is measured at a single point. Load disaggregation methods are needed to detect individual electrical loads from the aggregated load curve. In contrast, individual loads are measured with Intrusive Load Monitoring (ILM), typically using Smart Plugs.

2.1.1 NILM

Since with NILM, loads are measured at a single point or a group of loads is measured together, load control is impossible without extra hardware and detection capabilities. Therefore, although it is a cheap way to collect data about the load composition of the grid, it is not suitable if we aim to control individual loads.

2.1.2 Multi-label classification

Although NILM is unsuitable for our use case, the disaggregation methods used to detect different loads are relevant. Since the topic of this paper is detecting multiple loads connected to the same Smart Plug, multi-load detection methods utilized for NILM systems can offer insight into the possible methods that can be utilized in our use case.

In [1], a detailed review of different NILM methodologies and approaches in the literature is presented. The challenge of multi-label classification is also discussed. The simplest method is to create a label power set and train a single-class classifier with this power set as the label for the output. With this method, however, the number of classes to distinguish becomes very large, and there may be combinations of loads for

which no training data is available. Two other approaches are examined in [1]. Random k-labelsets (RAkEL) and ML-kNN. RAkEL [2] is an ensemble method where multiple single-class classifiers are trained on random subsets of the label power set. ML-kNN [3] uses kNN with an additional maximum a posteriori (MAP) step. This last step is used for providing the label set as the output.

Random Forests are used in [4] for achieving multi-label classification.

An LSTM autoencoder-based multi-label classification method is presented in [5]. The deepest layer of the autoencoder contains information about the loads that are present.

In [6], a Neural Network-based multi-label classification method, called BP-MLL, is introduced. Where a special error function is defined for backpropagation:

$$\left(\frac{1}{|Y||\bar{Y}|} \sum_{(k,l) \in Y \times \bar{Y}} \exp(-(c_k - c_l)) \right) \quad (2.1)$$

Where Y is the set of output labels belonging to a training sample, $\bar{Y}$ is the set containing all other labels, $|\cdot|$ measures the cardinality of the set, and c_j are the outputs of the Neural Network. The above error function compares the outputs of neurons representing output labels and non-output labels and aims to promote a high difference between these two sets of neuron outputs. During inference, a threshold was used to determine which j to consider (based on c_j) as belonging to the output label set. The threshold was learned using a threshold learning algorithm. In [6], several evaluation metrics are presented for multi-label classification. One such metric is the Hamming loss:

$$\text{loss}(h) = \frac{1}{p} \sum_{i=1}^p \frac{1}{Q} |\Delta(h(\mathbf{x}_i), Y_i)|, \quad (21)$$

Where Q is the cardinality of the set of all distinguished labels, Y_i is the true label set belonging to the i -th sample, Δ is the symmetric difference function, and $h(x_i)$ is the output of the classifier h . There are other loss metrics, such as One-error, Coverage, Ranking loss, and Average precision, which all focus on different aspects of the output of the classifier.

In [7], an autoencoder-like deep neural network structure is presented where the deepest layer contains high-level features while the entire net is trained with the labels' output. They utilized bidirectional association-based pooling layers.

Another method to achieve multi-label classification is to use a binary classifier for each target label [8].

Real-time NILM

An interesting new area of NILM research is the real-time detection of new loads in the load composition. In [9], [10], active power is measured with a relatively high frequency compared to other NILM methods. The sudden change in the active power can be used to detect if a new load is connected or turned on.

2.1.3 ILM

In [11], several challenges of ILM are pointed out, like the different types of loads that need to be distinguished, and the presence of these different categories can affect the accuracy rates of different ILM methods presented in the literature. Several use cases of ILM are also introduced, such as detecting human activity, monitoring appliance usage patterns, and detecting irregularities. ILM can also be used to gather data about the energy consumption of households. Smart Plugs with load detection capabilities are only present in the literature; there are no commercially available solutions [12].

In [13] an occupancy-driven load control scheme is presented. The system schedules and turns off loads based on room occupancy information. The plug load detection part of the system is presented in [14]. Only the power consumption data was used, one sample every minute, and the minimum time window examined was 5 minutes. The best accuracy rate was 93%. Although the load detection capabilities of the system would need improvement, the energy-saving potentials of occupancy detection are worth examining. Mainly since plug loads significantly contribute to the energy consumption of office spaces [13].

In addition to time-series data-based classification methods, there are faster ILM methods using high-frequency measurements while offering faster detection times. In [15] Voltage-Current trajectories are collected from instantaneous voltage and current readings, which were used to categorize the electrical load. The authors achieved high accuracy rates using Supervised Self-Organizing Map (SSOM)-based classification. However, the target classes consisted of typical load characteristics (like resistive and reactive loads) instead of distinguishing individual loads. In [16], Electric Load Signatures are recorded for electrical loads that are used to classify the load. Manipulating the electrical load's supply to extract characteristic information is another way to achieve fast ILM [17]–[19].

With ILM, the problem is twofold: In the literature, it is assumed that only a single load will be plugged into the smart outlet. This assumption rarely holds in real-world

scenarios. Similarly, unseen loads may be plugged into the outlet in a real-world scenario. This is also often overlooked in the literature.

2.2 Use cases

According to the U.S. Energy Information Administration’s annual report [20], by 2050, in the U.S., the electricity generation from solar and wind power is expected to grow significantly and, in total, to match or surpass electricity produced using natural gas. With the increase of non-controllable renewable resources’ presence in the composition of electricity production, it becomes increasingly difficult to match the production levels to instantaneous electricity demand. Demand-Side Management (DSM) aims to solve the problem of the difficult task of balancing the grid by controlling parts of the demand-side as well. Utilities can indirectly influence energy consumption by using different pricing schemes [21]. A key metric is the Peak-to-Average Ratio (PAR) which is the ratio of the peak electricity demand and the average. As the electric grid must handle the peaks, part of the production capacity is not utilized off-peak. This issue is further worsened by introducing non-controllable power sources like solar energy, which introduce more uncertainty to grid operations.

One application of Demand-Side Management is controlling Heating, Ventilation, and Air Conditioning (HVAC) equipment as a short interruption to their operation of them does not have an immediate effect [22]–[25]. Apart from HVAC, there are other loads whose operation can be interrupted or scheduled. The authors in [26] categorize household loads into schedulable (i.e., charging devices with built-in battery) and non-schedulable (i.e., Lighting, Television) categories. Several approaches are presented in the literature to best utilize the available solar energy or energy storage capacity [27]–[30]. In addition to reducing electricity-related costs ([29] – 29%, [28] – 19%), PAR can also be reduced significantly ([28], [31] – 46%, [27] – 8%).

Although there are plenty of papers related to DSM and the utilization of non-controllable renewables, key assumptions are being made which prevent their real-world applications. Most methods presented in the literature do not consider that electrical loads are moved around the household and that plug loads also significantly contribute to a household’s overall energy consumption.

2.3 Summary

Achieving Demand-Side Management requires information about the load composition of the electrical grid and capabilities to schedule different loads. Electrical load classification methods can provide the necessary data and control capabilities to enable the real-world application of the DSM methods presented in Section 2.2.

Research into load classification can be summarized in the following way:

- NILM

Individual load control is not possible

Disaggregation of load curves is required

Most research focuses on time-series data analysis

Some novel methods for Real-Time NILM

- ILM

Measures each individual load

The electrical load's power consumption characteristics are used typically

Individual load control can be achieved

2.4 Previous work – Dimmer-based Smart Plug

Previous work of this research is published in [18], [19], [32] as well as two previous TDK reports. Since this paper utilizes the same concept and measurement methodology, this section briefly introduces the measurement process and the Smart Plug prototype. A photo of the Smart Plug is shown in Figure 2.1. The Smart Plug utilizes a transformer and current transformer to measure instantaneous current ($U[k]$) and voltage ($I[k]$) values. An AC dimmer with a built-in power supply is used as the power source, and the dimmer is used to cut the supply of the load plugged into the Smart Plug. An RP2040 microcontroller controls the measurement process and records the load's response to the dimming. The ESP32 is responsible for WiFi communication and on-device CNN-based classification.

The measurement methodology is the following. Using the dimmer, the voltage supply of the connected load is cut with a dimming ratio of $\%_{min}$ for h number of AC cycles. The dimming control procedure can be seen in Figure 2.2. This is followed by d AC periods where the load receives uninterrupted power. Then the cutoff ratio is increased

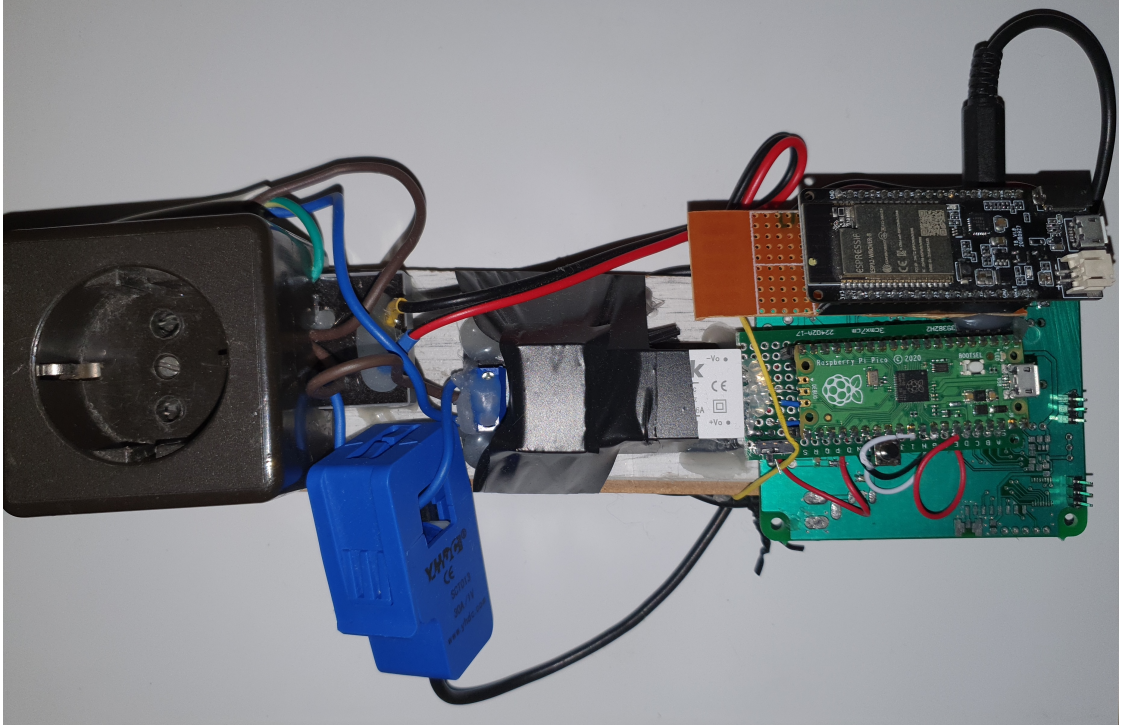


Figure 2.1: Smart Plug featuring the Raspberry Pi Pico and the ESP32.

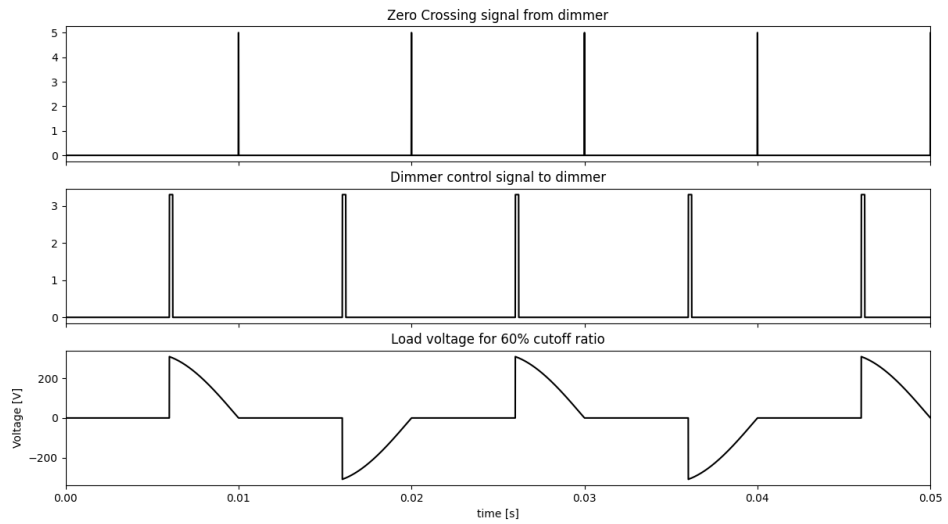


Figure 2.2: Signals between the dimmer and RP2040 and the resulting dimmed voltage curve when using a 60% cutoff ratio.

for another h number of AC cycles followed by another d number of non-dimmed cycles. This procedure is done with r number of different cutoff ratios, the last cutoff ratio being $\%_{max}$. The parameters of the measurement ($r, \%_{min}, \%_{max}, h, d$) define the measurement profile. In this report, measurement profiles will be displayed in the following format:

$$\{< r, \%_{min} - \%_{max} >, h, d\}$$

The cutoff ratios are evenly spaced between $\%_{min}$ and $\%_{max}$ in increasing order. During the measurement, for each AC period where dimming was used, three values were calculated from 360 multisampled (8-sample multisampling) instantaneous voltage and current measurements: Real Power, the RMS voltage, and the RMS current. These are calculated on the RP2040 using the following formulas:

$$P = \frac{1}{n} \cdot \sum_{k=1}^n U[k]I[k] \quad (2.2)$$

$$V_{RMS} = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^n U[k]^2} \quad (2.3)$$

$$I_{RMS} = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^n I[k]^2} \quad (2.4)$$

Additionally, Apparent Power can be calculated from V_{RMS} and I_{RMS} :

$$S = V_{RMS} \cdot I_{RMS} \quad (2.5)$$

The Real Power, RMS voltage, and current values are assembled into matrices of size $r \times h$. Figure 2.3 shows an example of the measurement process and construction of the Real Power measurement matrix.

Based on simulation results, it was determined that decreasing the measurement matrix sizes, and thus the measurement time and collected data, only resulted in slightly worse accuracy rates. This means that it is possible for computationally constrained environments, like NN inference on a microcontroller, to reduce the computational complexity while achieving high accuracy rates. A smaller measurement matrix also has the added benefit of faster measurement time. The number of AC cycles needed for a measurement with a measurement profile $\{< r, \%_{min} - \%_{max} >, h, d\}$ can be calculated the following way:

$$n_{cycles} = r \cdot h + (r - 1) \cdot d$$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	h=20	1	...	d=16
%min = 10%	11.73	10.95	11.16	12.06	10.72	11.99	12.52	12.12	11.63	11.23	11.37	11.92	11.57	12.45	12.02	11.7	11.64	11.74	11.98	12.73	IDLE	...	IDLE
15%	11.28	11.78	12.11	12.75	12.29	12.66	12.17	12.99	12.21	12.31	11.94	12.14	12.64	12.18	12.23	12.15	12.65	12.23	12.1	13.12	IDLE	...	IDLE
20%	12.42	16.38	15.64	16.64	16.28	16.76	16.14	15.79	17.57	15.36	15.38	15.87	16.68	15.49	14.46	15.16	16.64	15.54	15.85	15.52	IDLE	...	IDLE
25%	15.08	15.27	15.76	17.11	15.22	15.63	15.3	15.59	17.6	15	15.37	15.91	16.26	17.17	14.71	16.89	16.11	15.57	15.66	14.34	IDLE	...	IDLE
30%	15.14	15.65	16.57	16.22	15.84	15.47	15.59	15.07	16.71	14.71	16.66	15.28	16.22	15.96	14.77	15.82	16.24	17.07	18.04	14.21	IDLE	...	IDLE
35%	16.18	14.44	16.43	17.95	14.97	17.59	15.35	15.73	16.75	15.2	15.7	17.5	15.79	16.76	16.96	16.67	16.35	15.86	17.39	15.11	IDLE	...	IDLE
40%	16.02	16.46	15.75	17.68	16.89	16.73	16.64	17.21	16.68	16.22	16.34	16.73	16.37	18.25	16.18	16.92	16.88	17.37	16.97	15.94	IDLE	...	IDLE
45%	15.08	17.85	17.22	16.17	17.65	16.25	16.87	17.57	17.09	16.73	16.57	16.96	17.36	16.42	17.4	17.1	17.51	16.61	17.36	16.63	IDLE	...	IDLE
50%	15.31	17.47	17.78	16.57	17.29	17.59	17.25	17.79	16.63	17.49	17.28	17.72	16.67	17.06	18.07	17.1	17.53	17.2	16.22	16.78	IDLE	...	IDLE
55%	15.21	17.07	18.16	17.85	16.62	15.56	17.7	18.34	16.55	18	17.76	15.58	20.22	14.34	18.78	15.38	19.16	16.92	15.55	15.87	IDLE	...	IDLE
60%	16.89	16.76	17.87	16.42	16.36	16.73	17.37	18.48	16.51	17.2	17.17	16.53	18.93	15.6	17.42	16.84	17.29	18.68	16.27	16.63	IDLE	...	IDLE
65%	15.2	19.44	18.68	17.93	18.08	18.07	19.23	18.57	18.31	17.7	18.93	17.86	18.8	17.46	18.55	18.55	18.07	19.65	18.04	18.19	IDLE	...	IDLE
70%	11.27	15.39	15.53	15.03	15.49	16.09	15.74	15.84	15.58	16.09	15.63	16.05	16.06	15.83	16.28	14.81	15.93	15.78	15.17	15.95	IDLE	...	IDLE
%max = 75%	5.8	9.86	8.86	9.78	9.96	10.57	9.61	9.8	9.54	9.92	9.67	9.84	9.26	10.03	9.58	9.85	9.58	10.07	9.59	9.49	End of measurement		

Figure 2.3: Real Power measurement matrix of an LED Light bulb using the TEST_ORIG measurement profile. The power measurement values [W] are displayed for each dimmed AC period.

Classification method	SVM	FC	CNN	CNN
Number of samples (train-test)	30 - 220	100 - 150	150 - 100	50 - 200
TEST_ORIG	99.56%	99.51%	99.92%	99.84%
TEST_HALVED	98.74%	98.52%	99.90%	99.86%
TEST_TINY	97.40%	97.88%	-	-
TEST_FOUR	99.35%	98.50%	99.35%	98.81%

Figure 2.4: Summary of different classification results. (Average accuracy rates, based on 100 runs.) [18]

The following measurement profiles were selected to be used for measurements:

- TEST_ORIG : $\{< 14, 10\% - 75\% >, h = 20, d = 16\}$ Measurement time: 488 AC cycles (9.76s)
- TEST_HALVED : $\{< 7, 15\% - 75\% >, h = 10, d = 8\}$ Measurement time: 118 AC cycles (2.36s)
- TEST_TINY : $\{< 2, 10\% - 75\% >, h = 12, d = 4\}$ Measurement time: 28 AC cycles (0.56s)
- TEST_FOUR : $\{< 4, 15\% - 75\% >, h = 8, d = 4\}$ Measurement time: 44 AC cycles (0.88s)

Using these four different measurement profiles and different classification methods, the average accuracy rates achieved can be seen in Figure 2.4. We can see that CNN offers the best accuracy rates, although it required more training samples to achieve it.

The performance of on-device classification was validated using the TEST_HALVED measurement profile, and it was shown that an accuracy rate of 99.44% can be achieved.

3 Measured loads and dataset

The different electrical loads measured previously and used in this report are shown in Table 3.1. In total, 14 loads were measured and grouped into 11 categories. Figure 3.1 shows real power measurement matrices of some previously measured loads. We can see that in some cases, like with the 4A 'smart' lead acid battery charger, there is a periodicity that is independent of the measurement process. It is also apparent that the three USB chargers have very similar measurement matrices. These matrices can also change in time depending on the usage of the load (e.g., laptop) or the charge level (e.g., battery chargers, USB chargers).

3.1 Motivation

As discussed in Chapter 2, ILM and NILM both have advantages and disadvantages and are suited for different purposes. Combining these methods could be beneficial and could solve some of the challenges of using Smart Plugs in a real-world scenario. In [33], for example, a hybrid ILM-NILM approach is presented where the Smart Plugs help improve the accuracy of the load disaggregation. We can also approach this problem from the other side: **Enable Smart Plugs to detect multiple loads simultaneously.**

To our best knowledge, this hybrid approach of integrating NILM methods into a Smart Plug has not been examined in the literature. All Smart Plug prototypes capable of load class-level distinction that we have seen in the literature expect a single load to be connected to the Plug. This means a Smart Plug is needed for every load we aim to control and/or monitor. The use case for our original Smart Plug prototype was to measure and control loads that are moved around within the household since those are the scenarios where load detection is required. In a real-world environment, an extension cord may be plugged into a smart plug, allowing multiple loads to be connected to the same Smart Plug. In those scenarios, our original method may not work correctly since it can only detect a single load. This is problematic both for load forecasting and load control. If only the dominant load is detected, the Plug may incorrectly assume the

Load description	Load name	Load category
10W Apple USB adapter charging an iPad	ipad10W	USB
5W Apple USB adapter	usbapple5V1A	
5W generic USB adapter	usb5V1A	
4A "smart" lead-acid battery charger	batterycharger4A	batterycharger4A
800mA traditional lead-acid battery charger	batterycharger800mA	batterycharger800mA
A fan	fan	fan
A hairdryer	hairdryer	hairdryer
LED Light Bulb	ledbulb	ledbulb
LED-based Spotlight	ledspotlight	ledspotlight
Incandescent Light Bulb	incandescentbulb	INCANDESCENTS
Infrared Heat Lamp	irlamp	
Laptop charger charging the laptop	laptop	laptop
LCD screen	monitor	monitor
Soldering iron	solderingiron	solderingiron

Table 3.1: Electrical loads measured and their categorization

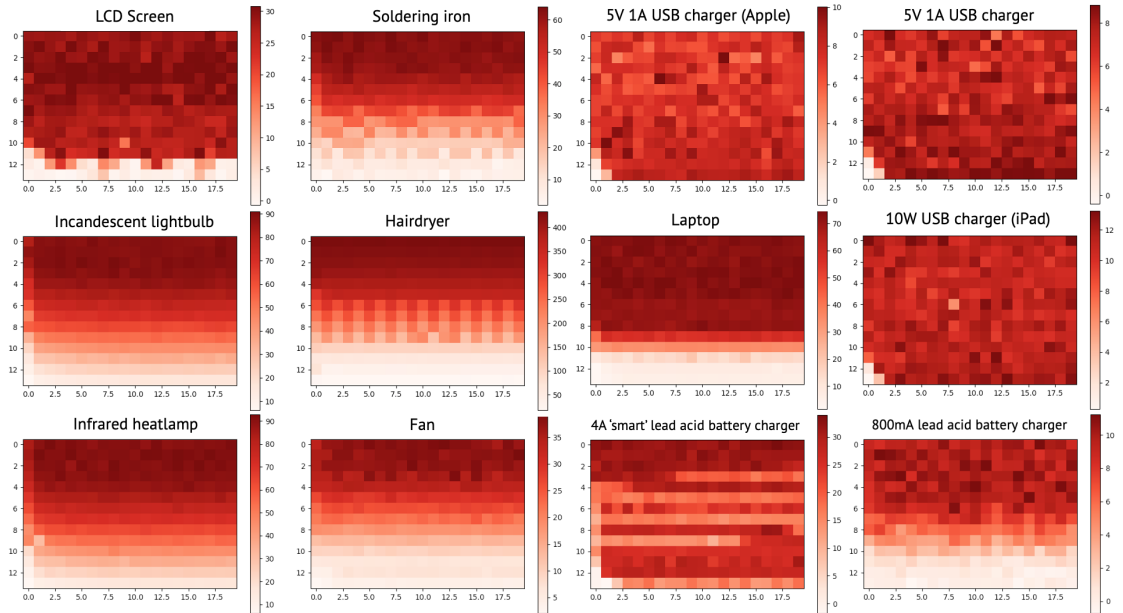


Figure 3.1: Measurement matrices of some of the previously measured loads. (Measurement unit: Watts)

schedulability based on only that load. In a worse case, the combination of the loads may cause the Plug to detect it as something completely different from any of the loads plugged into the Smart Plug.

Creating a Smart Plug capable of detecting multiple loads simultaneously solves the misclassification issues arising when multiple loads are present. Still, it also enables the adjustment of the granularity of the control capabilities while reducing the installation cost of a Smart Plug system. It becomes possible to deploy fewer Smart Plugs, which enables cost reduction at the expense of individual load controllability (e.g., if a non-schedulable and a schedulable load are connected to the same Smart Plug, the Plug cannot turn off the loads temporarily).

4 Process design

There are established ways for NILM methodology and classification accuracy evaluation, as shown in Chapter 2, but our use case is different. We expect that there will only be, at most, a few devices connected to the same Smart Plug. Our approach is first to determine if the network can recognize two loads and then check if more is possible. Since, in the past, CNN-based classification proved to be effective at both generalization and accurate load detection; we aim to use it for multi-load classification. There are, however, two challenges that arise with creating such a multi-class solution:

1. For even the 11 different load categories we examined, there are $\binom{11}{2} = 55$ different combinations of two loads and $\binom{11}{3} = 165$ combinations for three loads
2. The interaction between the different loads connected to the same Smart Plug may result in a measurement matrix similar to a different load's matrix

The goal is to create a method that can generalize as well as possible so that the size of the training dataset can be minimized.

4.1 Data collection

In order to have sufficient data to evaluate any method, we needed to measure different combinations of loads across all load categories. We have measured 43 of the 55 possible two-load category combinations. With each combination shown in Figure 4.1, at least 100 measurements were recorded. In addition to the two-load combinations, we have collected another 100 measurements for a few three-load category combinations:

- fan + ledbulb + INCANDESCENTS
- fan + ledspotlight + solderingiron

With this data and the original 250 measurements for each of the 14 loads, we had both single and multi-load measurements for all 11 load categories. The 100 multi-load measurements were recorded for the combinations for all four measurement profiles.

	solderingiron	monitor	laptop	INCANDESCENTS	ledspotlight	ledbulb	hairdryer	fan	batterycharger800mA	batterycharger4A
USB	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
batterycharger4A	✗	✗	✓	✓	✓	✓	✗	✓	✗	
batterycharger800mA	✓	✓	✓	✓	✓	✓	✗	✓		
fan	✓	✓	✓	✓	✓	✓	✓			
hairdryer	✓	✗	✗	✓	✓	✓				
ledbulb	✓	✗	✓	✓	✓					
ledspotlight	✓	✓	✓	✓						
INCANDESCENTS	✓	✗	✓							
laptop	✗	✗								
monitor	✓									

Figure 4.1: All possible two-load category combinations. ✓ marks the combinations for which we have recorded at least 100 measurements.

4.2 Multi-class classification

The easiest method to achieve multi-class classification would be to add new neurons to the final output layer of the NN, each corresponding to a combination of loads. However, this would result in a lot of output neurons and lacking generalization capabilities between the same load present in multiple combinations. Although this method performed well on the dataset, we do not consider it a true multi-class classification. Instead, a different method was chosen.

Traditionally, using the Softmax activation function in the last layer, only one output is selected. The target output looks something like this: $[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0]$ which results in the smallest error if the 5th neuron's output (before the Softmax activation) is significantly larger than any other neuron's in the last layer. But we can extend it and use a target output like this: $[.5, 0, 0, 0, 0, .5, 0, 0, 0, 0, 0]$ in which case the error is the smallest when the 1st and 6th neuron's output is equally high (before the Softmax activation) compared to any other neuron's output in the final layer. This way, both neurons learn to fire when the input is a multi-load measurement. This ensures that if the network learned the data correctly, then for an N-load multi-load input matrix, the top N neurons with the most significant output should be the N loads present in the multi-load input. However, this leaves an important question: how can we determine N?

As seen in Chapter 2, threshold learning algorithms were used in the literature for this problem. We have chosen a similar but slightly different approach: We wanted the same NN to learn to predict the number of loads present in the measurement by adding three additional neurons to the final layer. The first of these neurons should fire if only one load is present, the second if there are two loads, and the third if there are three loads. This way, the network predicts N (the number of loads), and we can select the top N neurons from the first part of the output layer. So, for example, the target output for some training samples may look like:

- One-load measurement: $[0, 0, 0, 0, \mathbf{1}, 0, 0, 0, 0, 0, 0 \mid \mathbf{1}, 0, 0]$
- Two-load measurement: $[\mathbf{.5}, 0, 0, 0, 0, \mathbf{.5}, 0, 0, 0, 0, 0 \mid 0, \mathbf{1}, 0]$
- Three-load measurement: $[\mathbf{.33}, 0, 0, \mathbf{.33}, 0, \mathbf{.33}, 0, 0, 0, 0, 0 \mid 0, 0, \mathbf{1}]$

A custom Softmax function was also needed to ensure the training worked correctly: The Softmax activation was calculated separately for the output neurons' first part (load categories) and the second part (number of loads).

4.3 Performance measures

As we have seen in Chapter 2, there are several different error measures for multi-class classification methods. Our network predicts two separate variables: 1) the probability distribution of electrical loads; 2) the top N loads to consider in the distribution. So it makes sense to evaluate the performance of these two methods separately as well. For this, we have selected the accuracy measure. An output is considered correct for the distribution if the top N loads are the N loads present in the training sample as outputs. For this part, we assumed to know the true N. For predicting the number of loads, we simply compared if the correct neuron corresponding to the true N had the most significant output value.

To measure the performance of the two parts of the prediction combined, we introduce the **strict accuracy** measure: a sample is considered correctly classified if both the distribution and prediction of the number of loads are correct, i.e., if the output of the multi-class classification solution matches precisely the expected output. If either the top N loads in the distribution or predicting the number of loads don't match with the true values, the sample is considered incorrectly classified. We call this strict accuracy since it does not reward scenarios when part of the prediction is correct.

5 Discussion of results

In this chapter, we present the results of applying the multi-class classification method introduced in Section 4.2. to the dataset presented in Section 4.1. In addition to showing and explaining the results, the generalization capabilities of our method are also evaluated. The results of the tests presented here are crucial for understanding what the model learned about the loads and their combinations and what to expect when it encounters an unseen combination of (previously seen) loads.

5.1 Multi-class classification

The network used for all classification tasks was a Convolutional Neural Network consisting of two convolutional and two hidden fully connected layers. Softmax-provided normalization was used between the convolutional and fully connected layers, while all other activation functions were leaky ReLU. The input of the CNN was the Real Power and Apparent Power measurement matrix for a measurement.

Unless stated otherwise, the measurements used for the classification were recorded using the TEST_ORIG measurement profile (14×20 measurement matrix size).

The first results are shown in Figures 5.1, 5.2 and 5.3. Since multi-class classification is a complex task, these figures need some explanation for correct interpretation. Figure 5.1 contains the raw results. In the top left corner, we can see that for the first part of the classification (correct distribution/class detection), the average accuracy rate (of 100 runs) is 98.17%, while the worst was 79.23%. Similarly, for the second part (predicting the number of loads), the average was 99.13%, while the worst was 92.62%. Finally, the strict accuracy rate is displayed: the average is 97.39% while the worst is 73.39%. In the first column, we see the true loads present in the sample. For training the model, 70 samples were used from each multi-class combination and 70 from the single-class measurements. For testing, 30 measurements were used for each single-load and multi-load combination. For every single load or multi-load combination in the first column, the corresponding row shows the average of the CNN's output over the 100 runs. The

Class detection accuracy: AVG: 98.17% Worst: 79.23% Load number prediction accuracy: AVG: 99.13% Worst: 92.62% Combined strict accuracy: AVG: 97.39% Worst: 73.39%														
1 load	USB	30	0	0	0	0	0	0	0	0	0	30	0	0
	batterycharger4A	0.19	29.74	0	0	0	0.05	0	0	0	0.02	29.05	0.95	0
	batterycharger800mA	0	0	30	0	0	0	0	0	0	0	30	0	0
	fan	0	0	0	30	0	0	0	0	0	0	29.67	0.33	0
	hairdryer	0	0	0	0.04	29.79	0.17	0	0	0	0	29.09	0.91	0
	ledbulb	0	0.02	0	0	0	29.98	0	0	0	0	30	0	0
	ledspotlight	0	0	0	0	0	0	30	0	0	0	30	0	0
	INCANDESCENTS	0	0	0.02	0.03	0	0	0	29.95	0	0	25.42	4.58	0
	laptop	0.12	0	0	0	0	0	0	29.65	0.23	0	29.9	0.1	0
	monitor	0.29	0.15	0.01	0	0.01	0	0	0.01	29.53	0	29.61	0.39	0
	solderingiron	0	0	0	0	0	0	0.19	0	0	29.81	29.56	0.44	0
	batterycharger4A+fan	0.02	29.85	0.01	29.73	0	0.14	0.12	0.01	0.1	0.02	30	0	0
	batterycharger4A+INCANDESCENTS	0.02	29.6	0	0.55	0.01	0.28	0	29.4	0.01	0.03	0.02	29.98	0
	batterycharger4A+laptop	0.01	29.74	0.12	0.07	0.01	0.25	0.1	29.58	0.12	0	30	0	0
2 loads	batterycharger4A+ledbulb	0	30	0	0	0.01	29.96	0.03	0	0	0	0.06	29.94	0
	batterycharger4A+ledspotlight	0.19	29.8	0.06	0.04	0	0.14	29.7	0	0.01	0.06	0.03	29.97	0
	USB+batterycharger4A	30	30	0	0	0	0	0	0	0	0	0.2	29.8	0
	batterycharger800mA+fan	0	0	29.99	30	0	0	0.01	0	0	0	0.03	29.97	0
	batterycharger800mA+INCANDESCENTS	0	0	27.88	1.51	0	0.59	29.92	0	0	0.1	0.64	29.36	0
	batterycharger800mA+laptop	0.01	30	0	0	0	0.01	0	29.96	0.02	0	0	30	0
	batterycharger800mA+ledbulb	0	0	30	0.01	0	29.99	0	0	0	0	0	30	0
	batterycharger800mA+ledspotlight	0	0	29.98	0.02	0	0	30	0	0	0	0.02	29.98	0
	batterycharger800mA+monitor	0.33	0.04	29.3	0.04	0	0.03	0.36	0	0.42	29.48	0.04	29.96	0
	batterycharger800mA+solderingiron	0	0	29.37	0.3	0	0	0.33	0	0	30	0.19	29.81	0
	USB+batterycharger800mA	30	30	0	0	0	0	0	0	0	0	0	30	0
	fan+hairdryer	0	0	0	25.71	29.74	0	0	0.38	0	4.17	0	30	0
	fan+INCANDESCENTS	0	0	0.59	28.89	0	0	0.2	29.85	0	0.47	0	29.95	0.05
	fan+laptop	0	0	0	29.63	0	0	0.13	0.02	29.85	0.34	0.03	30	0
	fan+ledspotlight	0	0	0.64	29.98	0	0	29.37	0	0	0.01	0	30	0
	fan+monitor	0.09	0.02	0.01	29.5	0.01	0.01	0.02	0.32	0.36	29.49	0.17	0.01	29.99
	fan+solderingiron	0	0	0.09	29.01	0	0	1.14	0	0	29.76	0	28.94	1.06
	USB+fan	30	30	0	30	0	0	0	0	0	0	0	30	0
	hairdryer+INCANDESCENTS	0	0	0	0.03	29.98	0	0	29.58	0	0.41	0	30	0
	hairdryer+ledbulb	0.01	0	0	0.01	29.97	29.98	0.02	0.01	0	0	0.34	29.64	0.02
	hairdryer+ledspotlight	0	0	0	0	30	0	0	0	0	0	0	30	0
	hairdryer+solderingiron	0	0	0	3.07	29.69	0	0	0.69	0	26.55	0	30	0
	INCANDESCENTS+laptop	0.01	0.01	0	0.02	0	0.01	0	29.93	29.91	0.08	0.03	30	0
	ledbulb+INCANDESCENTS	0	0	0	0	30	0	0	30	0	0	0	29.74	0.26
	ledspotlight+INCANDESCENTS	0	0	0.42	0.59	0	0	28.38	29.98	0	0.63	0.4	29.6	0
	INCANDESCENTS+solderingiron	0	0	1.15	0	0	0	1.82	29.43	0	27.6	0	29.67	0.33
	USB+INCANDESCENTS	29.7	0	0	0	0	0	0.3	30	0	0	0.08	29.92	0
	ledbulb+laptop	0.06	0.23	0	0.04	0	29.81	0.01	0.01	29.84	0	0	30	0
	ledspotlight+laptop	0	0	0	0.02	0	0	29.98	0	30	0	0	30	0
	USB+laptop	29.96	0.02	0.02	0	0	0.02	0	0	29.98	0	0	30	0
	fan+ledbulb	0	0	0	29.71	0	0	30	0	0	0.29	0	30	0
	ledbulb+ledspotlight	0	0	0	0	30	30	0	0	0	0	0	30	0
	ledbulb+solderingiron	0	0.07	0	0.07	0	29.94	0	0.02	0	29.9	0	30	0
	USB+ledbulb	29.92	0.04	0	0	0	29.97	0.07	0	0	0	0	30	0
	ledspotlight+monitor	0.41	0.03	0.07	0.16	0	0.16	29.65	0.01	0.17	29.34	0.03	29.97	0
	ledspotlight+solderingiron	0	0	0	0.01	0	0	29.99	0	0	30	0.3	29.7	0
	USB+ledspotlight	29.88	0	0	0	0	0	29.99	0	0	0.13	0	30	0
	monitor+solderingiron	0.15	0.02	0.08	0.34	0.02	0	0.02	0.38	0.13	29.48	29.38	0.08	29.92
	USB+monitor	29.58	0.36	0.11	0.03	0	0.03	0.04	0	0.35	29.49	0.01	0.89	29.11
	USB+solderingiron	29.47	0	0	0.3	0	0	0	0.3	0	29.93	0	30	0
3 loads	fan+ledbulb+INCANDESCENTS	0.15	0	0	30	0	29.85	30	0	0	0	0.01	29.99	0
	fan+ledspotlight+solderingiron	0	0	0.01	30	0	0	29.99	0.02	0	29.98	1.85	28.15	0

Figure 5.1: Multi-class classification results using 70 training samples from each load combination category and 30 test samples. The extended confusion matrix shows the averages of 100 runs.

Class detection accuracy: AVG: 98.17% Worst: 79.23%	solderingiron	monitor	laptop	INCANDESCENTS	ledspotlight	ledbulb	hairdryer	fan	batterycharger800mA	batterycharger4A	USB
One load	99.37%	98.43%	98.83%	99.83%	100.00%	99.93%	99.30%	100.00%	100.00%	99.13%	100.00%
Two loads											
USB	98.00%	97.07%	99.80%	99.00%	99.57%	99.70%		100.00%	100.00%	100.00%	
batterycharger4A			97.87%	97.07%	98.40%	99.87%		98.67%			
batterycharger800mA	97.90%	96.50%	99.87%	92.67%	99.93%	99.97%		99.97%			
fan	95.90%	97.33%	98.63%	95.80%	97.83%	99.03%	84.83%				
hairdryer	87.47%			98.53%	100.00%	99.83%					
ledbulb	99.53%		98.97%	100.00%	100.00%						
ledspotlight	99.97%	96.67%	99.93%	94.53%							
INCANDESCENTS	90.10%		99.60%								
laptop											
monitor	97.10%										
Three loads	fan+ledbulb+INCANDESCENTS			fan+ledspotlight+solderingiron							
	99.50%			99.90%							

Figure 5.2: Summary of class detection accuracy results using 70 training samples from each load combination category and 30 test samples. The accuracy rates are averages of 100 runs. The empty white cells show the combinations for which no measurements were recorded.

Combined strict accuracy: AVG: 97.39% Worst: 73.39%	solderingiron	monitor	laptop	INCANDESCENTS	ledspotlight	ledbulb	hairdryer	fan	batterycharger800mA	batterycharger4A	USB
One load	97.90%	97.30%	98.50%	84.60%	100.00%	99.93%	96.83%	98.90%	100.00%	96.13%	100.00%
Two loads											
USB	98.00%	94.67%	99.80%	98.73%	99.57%	99.70%		100.00%	100.00%	99.33%	
batterycharger4A			97.87%	97.03%	98.40%	99.73%		98.67%			
batterycharger800mA	97.27%	96.37%	99.87%	90.80%	99.87%	99.97%		99.87%			
fan	93.47%	97.33%	98.63%	95.73%	97.83%	99.03%	84.83%				
hairdryer	87.47%			98.53%	100.00%	98.67%					
ledbulb	99.53%		98.97%	99.13%	100.00%						
ledspotlight	98.97%	96.57%	99.93%	93.20%							
INCANDESCENTS	90.10%		99.60%								
laptop											
monitor	97.10%										
Three loads	fan+ledbulb+INCANDESCENTS			fan+ledspotlight+solderingiron							
	99.50%			93.83%							

Figure 5.3: Summary of strict accuracy results using 70 training samples from each load combination category and 30 test samples. The accuracy rates are averages of 100 runs. The empty white cells show the combinations for which no measurements were recorded.

	Class. Accuracy		Num. Pred. Accuracy		Strict Accuracy	
	AVG	Worst	AVG	Worst	AVG	Worst
TEST_ORIG	98.17%	79.23%	99.13%	92.62%	97.39%	73.39%
TEST_HALVED	96.17%	82.68%	97.84%	84.94%	94.18%	73.69%
TEST_FOUR	88.07%	59.11%	94.39%	88.45%	83.20%	55.77%

Table 5.1: Multi-class classification accuracy rates for different measurement profiles (Train samples: 70, test samples: 30)

green part shows the distribution of the top N predicted load categories. The places with box borders show the true classes. We can see that for the two-load measurements, there are two boxes in each row. The blue part shows the results for predicting the number of loads present in a measurement. The places corresponding to the correct output have bold values. We can see that for some of the single load measurements in the INCANDESCENTS category, the second part of the network incorrectly predicted an additional load present.

Figure 5.2 shows the class distribution accuracy rates for each single load case, two-load combination, and three-load case. It essentially compresses the class detection matrix part in Figure 5.1 to only the values in boxes. Although Figure 5.2 shows less information, it is less cluttered and easier to understand as the accuracy rates for the two-load combinations can be displayed in an upper left triangular matrix form.

Similarly to Figure 5.2, Figure 5.3 shows accuracy rates for each single load case, two-load combination, and three-load case. Still, it displays strict accuracy rates considering the load count prediction of CNN. We can see that each accuracy rate in Figure 5.3 is less than or equal to its corresponding value in Figure 5.2. It is important to note that the values shown in Figure 5.3 cannot be calculated from the matrices in Figure 5.1, so these were separately calculated during the testing phase when the data for Figure 5.1 was collected.

Since measurement data was recorded with the TEST_ORIG, the TEST_HALVED, the TEST_FOUR, and the TEST_TINY profiles, we can also evaluate the same method as described above for the measurements recorded with the TEST_HALVED and the TEST_FOUR profiles. (The 3×3 convolution does not apply to the TEST_TINY profile’s 2×12 matrices.) The results are shown in Table 5.1. We can see that the accuracy rates are noticeably worse with smaller measurement matrices and, thus, less data.

5.2 Evaluation of generalization capabilities

Examining the generalization capabilities of our method is crucial for determining the feasibility of using this method for a larger list of loads. Knowing the generalization capabilities and limitations also helps the multi-load measurement planning (e.g., all or just some combinations need to be measured) to be more efficient. It also allows us to further improve the method by highlighting its weaknesses. We want to explore and answer multiple generalization-related questions:

- whether we can use single-load measurements to classify multi-load measurements
- whether it is possible to reduce the number of multi-load measurements needed
- whether by learning multi-load measurements, the network learns something about single-load measurements
- whether multi-load measurements can help when encountering other multi-load combinations during inference

5.2.1 From single-load to multi-load

For single-load to multi-load generalization, the goal is to evaluate if the network can infer something about the multi-load samples if the training set only consisted of single-load samples. Our choice of the measure was the percentage of (N-load) multi-load samples for which there is at least one correct load category in the top N prediction. For this case, predicting the number of loads was not considered. The results are shown in Figure 5.4. As can be seen, on average, 78.66% of predictions contained at least one correct load category. Interestingly, the fan, ledspotlight, INCANDESCENTS and solderingiron load categories were frequent false positives.

The results are bad for considering the classification accuracy as the measure (i.e., when the top N categories match exactly the N true load categories). The only multi-load combinations reaching a rate of 30% are the following:

- USB + ledbulb 72.85%
- INCANDESCENTS + solderingiron 56.85%
- batterycharger4A + ledbulb 49.76%
- ledbulb + ledspotlight 30.95%

Prediction (output) - TOP N														
Average ratio of samples having at least one correct load category in the TOP N: 78.66%														At least one correct load in top N:
		USB	batterycharger4A	batterycharger800mA	fan	hairdryer	ledbulb	ledspotlight	INCANDESCENTS	laptop	monitor	solderingiron		
1 load	USB	30												
	batterycharger4A		30											
	batterycharger800mA			30										
	fan				30									
	hairdryer					30								
	ledbulb						30							
	ledspotlight							30						
	INCANDESCENTS								30					
	laptop									30				
	monitor										30			
	solderingiron											29.96	30	
N = 2 loads	batterycharger4A+fan	1.27	2.96	1.02	63.9	4.39	13.66	66.45	1.97	20.63	22.61	1.14	66.37%	
	batterycharger4A+INCANDESCENTS	0.01	1.07	1.37	93.01	13.35	4.9	20.47	6.71	12.9	7.2	39.01	7.74%	
	batterycharger4A+laptop	0.06	38.84	1.33	21.64	4.03	12.73	41.06	4.54	50.14	24.04	1.59	74.50%	
	batterycharger4A+ledbulb	0.15	95.31	1.6	10.9	2.8	53.26	16.67	3.54	11.64	3.29	0.84	98.81%	
	batterycharger4A+ledspotlight	19.65	12.89	3.9	5.96	2.94	83.72	32.7	1.95	16.93	19.07	0.29	44.35%	
	USB+batterycharger4A	3.82	100	2.35	7.46	2.26	1.61	3.4	3.18	24.72	47.98	3.22	100.00%	
	batterycharger800mA+fan	0.03	0	0	99.97	0	0	0	0.55	0	0	99.48	99.97%	
	batterycharger800mA+INCANDESCENTS	0.03	0	0.03	0	3.63	0	0	100	0	0.52	95.79	100.00%	
	batterycharger800mA+laptop	7.17	0.11	0.2	37.13	18.37	10.76	45.53	7.41	44.13	26.6	2.59	44.33%	
	batterycharger800mA+ledbulb	0.06	3.23	10.98	38.73	3.64	31.09	78.27	0	9.92	23.82	0.26	42.07%	
	batterycharger800mA+ledspotlight	0	0	0	98.99	0.27	1	96.88	0	2.86	0	0	96.88%	
	batterycharger800mA+monitor	7.8	0	1.82	33.94	6.04	13.52	54.35	2.25	43.77	34.77	1.74	36.18%	
	batterycharger800mA+solderingiron	0	0	0	9.19	0.73	0.02	0	90.06	0	0	100	100.00%	
	USB+batterycharger800mA	2	0	41.4	56.53	1.01	0.15	89.09	0	3.93	5.89	0	43.40%	
	fan+hairdryer	0	0	2.98	91.03	7.93	2.97	4.06	9.46	4.02	0.05	77.5	93.04%	
	fan+INCANDESCENTS	0.87	1.06	0.55	0	6.83	0	0	100	0.23	3.33	87.13	100.00%	
	fan+laptop	2.1	0.01	1.3	53.28	48.99	5.29	22.58	14.88	21.73	9.84	20	70.72%	
	fan+ledspotlight	0	0	0	100	0	0	0	2.98	0	0	97.02	100.00%	
	fan+monitor	1.01	0.01	0.7	83.43	19.08	5.78	30.71	6.71	17.15	10.17	25.25	87.42%	
	fan+solderingiron	0	0	0	0	0	0	0	100	0	0	100	100.00%	
	USB+fan	0	0	2	94.92	6.32	1.02	2.19	6.14	1.3	1.01	85.1	94.92%	
	hairdryer+INCANDESCENTS	0.09	1.05	0	67.5	3.86	3.75	0.41	32.6	1.1	1	88.64	35.03%	
	hairdryer+ledbulb	3.59	0.35	1.62	25.65	100	4.57	2.33	38.35	10.17	0.02	13.35	100.00%	
	hairdryer+ledspotlight	0	0	3.12	90.8	9.95	1	10.96	8.32	4.99	1	69.86	20.91%	
	hairdryer+solderingiron	0	0.12	2	87.27	5.52	2.77	3.27	12.73	3.88	0.43	82.01	86.53%	
	INCANDESCENTS+laptop	0.81	0	0	49.64	43.36	4.4	0.52	32.6	5.99	7.23	55.45	38.48%	
	ledbulb+INCANDESCENTS	0	0	0.01	44.79	6.46	1	0	59.54	0.01	2	86.19	59.54%	
	ledspotlight+INCANDESCENTS	1.83	2	1	0	22.5	0	0	100	0	3.06	69.61	100.00%	
	INCANDESCENTS+solderingiron	2	3.01	2	0	28.31	1.81	0	100	1.99	4.03	56.85	100.00%	
	USB+INCANDESCENTS	0	0.32	1	5	21.03	1	0	100	0.96	5.52	65.17	100.00%	
	ledbulb+laptop	0.01	14.27	0.06	34.02	15.02	11.53	27.04	9.78	67.5	13.44	7.33	71.31%	
	ledspotlight+laptop	34.46	0	1.73	32.93	34.36	5.33	33.11	9.51	33.35	10.91	4.31	58.90%	
	USB+laptop	2.53	2.91	0.05	22.24	12.57	5.81	37.57	4.22	95.12	12.06	4.92	95.12%	
	fan+ledbulb	0.01	1.01	2.9	93.26	11.78	7.92	49.52	3.91	8.78	8.53	12.38	94.22%	
	ledbulb+ledspotlight	0.99	3.3	0.7	28.95	3.27	48.3	75.24	2.14	9.54	26.31	1.26	92.59%	
	ledbulb+solderingiron	0	0.97	1.98	95.97	10.55	4	11.01	4.74	3.67	3.44	63.67	67.67%	
	USB+ledbulb	72.86	0.02	0.02	1.36	4.43	99.99	10.91	3.93	4.38	1.61	0.49	100.00%	
	ledspotlight+monitor	14.61	0.11	0.71	38.84	15.85	20	54.13	3.96	25.62	23.99	2.18	71.26%	
	ledspotlight+solderingiron	0	0	0	0	0	0	0	100	0	0	100	100.00%	
	USB+ledspotlight	4.26	0	0	69.9	6.29	2.05	94.97	1	13.38	7.11	1.04	94.98%	
	monitor+solderingiron	0.25	0.05	0.04	82.07	30.67	2.69	4.53	14.27	5.89	4.07	55.47	59.53%	
	USB+monitor	13.47	10.17	0.67	4.64	0.38	13.3	14.34	3.03	41.65	96.78	1.57	97.24%	
	USB+solderingiron	0	0.03	0	62.98	0.99	0.01	0	40.23	0	0	95.76	95.76%	
N = 3 loads	fan+ledbulb+INCANDESCENTS	3.11	7.97	2.07	15.16	55.32	10.39	1	98.64	7.23	17.61	81.5	100.00%	
	fan+ledspotlight+solderingiron	8.38	12.22	3.04	5.27	42.13	9	2.01	100	2.02	15.93	100	100.00%	
	Sum of misclassifications into category:	110.32	64.3	43.82	1258	509.95	230.05	637.28	615.74	315.26	336.49	1203.2		
Sum of classifications into category:		209.26	315.37	98.25	1958.3	637.21	502.1	1037.3	1345.8	633.22	506.27	1957		

Figure 5.4: Multi-load classification results (average of 100 runs) containing the top N (N=1, 2 or 3 depending on the number of true target load categories) predictions of the CNN. **The training data did not contain any multi-load samples.** The training data consisted of 220 single-load measurements from each load category.

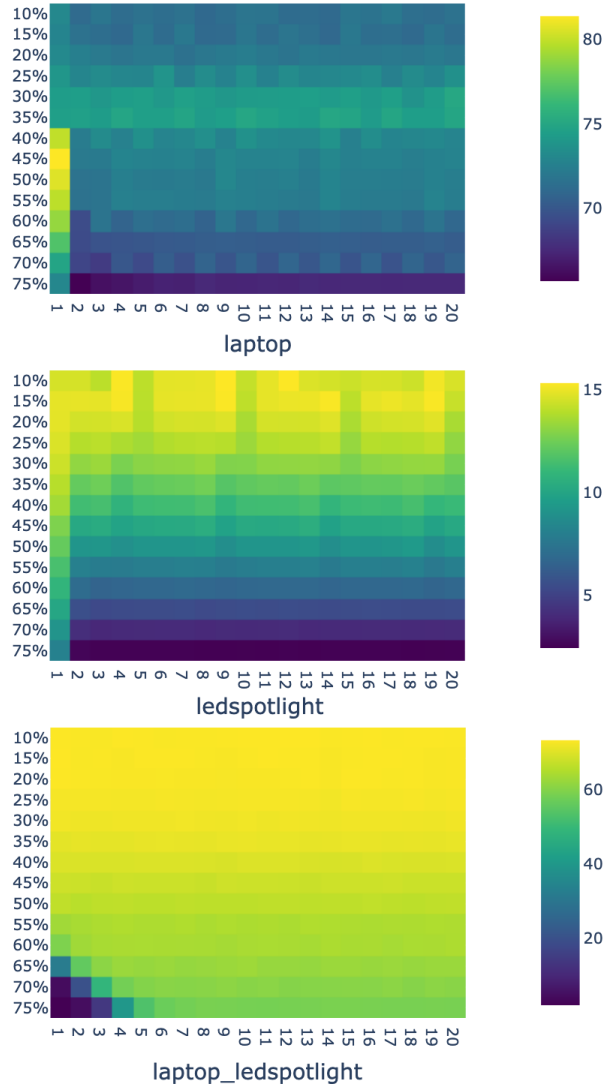


Figure 5.5: The mean of all measured Real Power measurement matrices for the laptop and ledspotlight loads (top 2 figures) and the mean Real Power measurement matrix of the combined laptop + ledspotlight measurements (bottom figure).

Although these results allow us to conclude that at least some generalization is happening, relying on only single-load measurements for multi-load detection would result in poor performance. The foremost reason for the poor performance is that when measuring multiple loads simultaneously using the method introduced in Section 2.4, the loads are connected (even during dimming) and act as one circuit. As such, the measurement is not simply a sum of two separate measurements but a measurement of this new circuit formed by connecting the two loads in parallel. Both loads can have capacitive and/or inductive characteristics. It can be seen in Figure 5.5 where the means of all Real Power measurement matrices are shown for the single-load measurements for the laptop and ledspotlight loads and the multi-load measurements where the two loads are measured simultaneously. It is worth noting that the laptop’s charge cycle and usage may have resulted in different mean power consumption values, but this does not significantly affect the relative structure of these mean matrices.

Multi-load classification with small training set

We have seen that only relying on single-load measurements would result in inaccurate classification. Another practical issue with many load combinations is the measurement time. Taking 70 measurements with all 45 combinations introduced in Section 4.1 would take around 9 hours using the TEST_ORIG measurement profiles, and it would increase rapidly if more loads are added in the future.

If reasonable accuracy rates can be achieved with less than 70 measurements per combination, then it would allow the method to scale better when new loads are added. To test if using less than 70 measurements would be feasible, we used the same method introduced in Section 5.1, with the only difference being the training and test sample sizes. From each single-load category, 160 were randomly selected for the training set, while the other 90 were allocated to the test set. For the multi-load combinations, for each combination, 10 samples were randomly chosen to be in the training set, while the remaining 90 were assigned to the test set. Similarly to the previous method, this process, along with the model training and test set evaluation, was run 100 times. The large confusion matrix is shown in Figure 5.6 while the classification part’s accuracy rates and the strict accuracy rates are shown in Figures 5.7 and 5.8. We can see cases where strict accuracy rates are low, but the overall average is still high, 88.44%. The results of this method and the previous one introduced in Section 5.1 are compared in the table below:

Class detection accuracy: AVG: 91.59% Worst: 69.27% Load count prediction accuracy: AVG: 95.61% Worst: 80.69% Combined strict accuracy: AVG: 88.44% Worst: 60.95%												Num. pred. loads: 1			Num. pred. loads: 2			Num. pred. loads: 3		
1 load	USB	90											90							
	batterycharger4A	0.06	89.87										89.47		0.53					
	batterycharger800mA			90									90							
	fan				90								89.96		0.04					
	hairdryer				0.04	89.62							89.15		0.85					
	ledbulb						90						89.98		0.02					
	ledspotlight							90					90							
	INCANDESCENTS								90				88.72		1.28					
	laptop	0.03								89.93		0.04	89.94		0.06					
	monitor	0.07	0.22		0.01		0.02			0.01	89.67		89.94		0.06					
	solderingiron							0.01			89.99		89.7		0.3					
2 loads	batterycharger4A+fan	0.34	88.25	0.17	85.68	0.03	1.93	2.17	0.2	1.12	0.1	0.01		90						
	batterycharger4A+INCANDESCENTS	2.65	73.42	0.07	18.57	0.3	11.14	0.31	65.67	2.43	1.4	4.04	0.35	89.65						
	batterycharger4A+laptop	0.65	87.73	1.7	0.85	0.01	3.43	0.9	0.02	83.86	0.84	0.01	0.09	89.91						
	batterycharger4A+ledbulb	0.19	89.94	0.05		0.08	88.9	0.64	0.03	0.13	0.02	0.02	1.47	88.53						
	batterycharger4A+ledspotlight	4.18	83.42	0.66	0.08	0.05	7.92	82.66	0.13	0.4	0.49	0.01	1.16	88.84						
	USB+batterycharger4A	89.79	89.98							0.22	0.01		6.72	83.28						
	batterycharger800mA+fan			87.64	89.81			2.04				0.51	3.2	86.8						
	batterycharger800mA+INCANDESCENTS	0.01		48.47	14.45			22.45	89.9			4.72	45.83	44.17						
	batterycharger800mA+laptop	0.02	0.05	89.91		0.01	0.02	0.08		89.83	0.06	0.02	0.02	89.98						
	batterycharger800mA+ledbulb	0.01		89.96			89.92	0.11						90						
	batterycharger800mA+ledspotlight			89.71	0.48			89.81					0.69	89.31						
	batterycharger800mA+monitor	4.91	0.27	82.57	0.25	0.03	0.12	2.25	0.05	4.52	85.01	0.02	0.29	89.71						
	batterycharger800mA+solderingiron			80.62	0.32			9.26				89.8	10.67	79.33						
	USB+batterycharger800mA	90		90									0.1	89.9						
	fan+hairdryer		0.74		65.63	88.9	0.01	0.01	2.15			22.56		90						
	fan+INCANDESCENTS	0.04		6.31	70.39	0.16		7.61	88.64			6.85	7.47	82.45	0.08					
	fan+laptop	0.22	0.01		88.48	0.04	0.04	0.14	0.09	87.33	3.5	0.15	0.01	89.99						
	fan+ledspotlight			6.65	87.47			82.69				3.19	4.31	85.69						
	fan+monitor	0.75	0.36		85.94	0.09	0.21	0.14	2.14	4.69	83.24	2.44	0.1	89.9						
	fan+solderingiron			3.64	77.15	0.01		12.24	1.17			85.79	0.04	81.35	8.61					
	USB+fan	89.96		0.03	88.98				0.91			0.12	0.04	89.96						
	hairdryer+INCANDESCENTS			0.01	2.78		87.38	0.79		83.74		5.3		90						
	hairdryer+ledbulb	0.22	0.02	0.06	0.34		88.77	87.21	0.34	2.21	0.15	0.02	0.66	17.65	72.34	0.01				
	hairdryer+ledspotlight						89.99		89.93					90						
	hairdryer+solderingiron				25.39		89.47			4.22		60.92		90						
	INCANDESCENTS+laptop	0.33	0.14		0.74	0.02	0.32		88.96	87.82	0.97	0.7		90						
	ledbulb+INCANDESCENTS	0.04			0.02		89.82			89.87	0.11	0.14		89.08	0.92					
	ledspotlight+INCANDESCENTS	0.03		3.24	3.81				66.83	89.96		16.13	36.02	53.98						
	INCANDESCENTS+solderingiron			1.78	15.01			28.07	83.55			51.59	6.8	83.1	0.1					
	USB+INCANDESCENTS	86.18			1.35		0.86	0.05	90			1.56	2	87.98	0.02					
	ledbulb+laptop	0.19	1.4	0.04	0.11	0.03	88.85		0.1	89.28				89.99	0.01					
	ledspotlight+laptop	0.28		0.02	0.47			89.44		89.67	0.11			90						
	USB+laptop	89.3	0.16	0.03			0.52		0.01	89.96	0.02		0.02	89.97	0.01					
	fan+ledbulb		0.03		89.57	0.04	89.11		0.25		0.84	0.16	0.03	89.97						
	ledbulb+ledspotlight						90	90						90						
	ledbulb+solderingiron	0.01	0.06		0.99	0.08	89.89		1.58		0.03	87.36		90						
	USB+ledbulb	88.82	0.2	0.02			89.99	0.89			0.08		0.5	89.5						
	ledspotlight+monitor	9.52	0.4	1.19	0.97	0.05	3.95	87.78	0.08	4.32	71.74		0.28	89.72						
	ledspotlight+solderingiron			4.7	6.51			78.14	0.66			89.99	6.28	83.72						
	USB+ledspotlight	90						90						90						
	monitor+solderingiron	0.49	0.17	0.06	10.37	0.04	0.5	0.16	2.86	3.78	84.61	76.96	0.4	89.59	0.01					
	USB+monitor	82.75	2.65	1.36	0.09	0.04	0.18	0.72	0.05	3.69	88.46	0.01	15.56	74.44						
	USB+solderingiron	89.91			0.03		0.05	0.02	1.68			88.31		90						
3 loads	fan+ledbulb+INCANDESCENTS	5.54			83.97		88.32	0.9	89.07	0.9		1.3		4.97	85.03					
	fan+ledspotlight+solderingiron	0.01		9.78	88.13	0.17	0.04	77.99	4.22			89.66		35.46	54.54					

Figure 5.6: Multi-class classification results using 10 training samples from each load combination category and 160 from the single-load categories. The extended confusion matrix shows the averages of 100 runs.

Class detection accuracy: AVG: 91.59% Worst: 69.27%	solderingiron	monitor	laptop	INCANDESCENTS	ledspotlight	ledbulb	hairdryer	fan	batterycharger800m	batterycharger4A	USB
One load	99.99%	99.63%	99.92%	100.00%	100.00%	100.00%	99.58%	100.00%	100.00%	99.86%	100.00%
Two loads											
USB	98.08%	90.68%	99.18%	95.76%	100.00%	98.68%		98.82%	100.00%	99.74%	
batterycharger4A			91.20%	62.34%	85.37%	98.71%		93.60%			
batterycharger800mA	89.36%	88.14%	99.71%	53.84%	99.47%	99.87%		97.17%			
fan	81.06%	88.28%	95.46%	76.70%	89.31%	98.53%	72.52%				
hairdryer	67.10%			90.13%	99.91%	95.58%					
ledbulb	96.96%		98.04%	99.66%	100.00%						
ledspotlight	86.81%	77.49%	99.03%	74.21%							
INCANDESCENTS	50.62%		96.66%								
laptop											
monitor	81.61%										
Three loads											
	fan+ledbulb+INCANDESCENTS			fan+ledspotlight+solderingiron							
	90.40%			84.24%							

Figure 5.7: Summary of class detection accuracy results using 10 training samples from each load combination category and 160 from the single-load categories. The accuracy rates are averages of 100 runs. The empty white cells show the combinations for which no measurements were recorded.

Combined strict accuracy: AVG: 88.44% Worst: 60.95%	solderingiron	monitor	laptop	INCANDESCENTS	ledspotlight	ledbulb	hairdryer	fan	batterycharger800m	batterycharger4A	USB
One load	99.67%	99.57%	99.87%	98.58%	100.00%	99.98%	99.00%	99.96%	100.00%	99.31%	100.00%
Two loads											
USB	98.08%	79.02%	99.16%	93.68%	100.00%	98.36%		98.79%	99.89%	92.37%	
batterycharger4A			91.14%	62.23%	84.78%	97.22%		93.60%			
batterycharger800mA	78.92%	87.97%	99.69%	20.80%	98.71%	99.87%		93.77%			
fan	72.73%	88.23%	95.46%	72.84%	86.08%	98.50%	72.52%				
hairdryer	67.10%			90.13%	99.91%	78.22%					
ledbulb	96.96%		98.03%	98.63%	100.00%						
ledspotlight	80.78%	77.31%	99.03%	44.09%							
INCANDESCENTS	49.18%		96.66%								
laptop											
monitor	81.57%										
Three loads											
	fan+ledbulb+INCANDESCENTS			fan+ledspotlight+solderingiron							
	87.43%			57.07%							

Figure 5.8: Summary of strict accuracy results using 10 training samples from each load combination category and 160 from the single-load categories. The accuracy rates are averages of 100 runs. The empty white cells show the combinations for which no measurements were recorded.

		Predicted class										
True class	AVG Accuracy: 55.30%	USB	batterycharger4A	batterycharger800mA	fan	hairdryer	ledbulb	ledspotlight	INCANDESCENTS	laptop	monitor	solderingiron
	USB	224.12	2.73	2.42	0	0	0	1.67	0	0	19.06	0
	batterycharger4A	159.44	79.4	2.06	0	0	6.19	1.69	0	0	1.21	0.01
	batterycharger800mA	54.88	0	185.4	0	0	2.09	7.39	0	0	0.24	0
	fan	0	0	7.56	208.75	0	0	33.69	0	0	0	0
	hairdryer	0.01	0	0	0.33	120.86	125.35	0.3	2.66	0.04	0.22	0.23
	ledbulb	160.59	2.37	2.57	0	0	84.3	0.17	0	0	0	0
	ledspotlight	63.66	0	35.43	0	0	0	150.91	0	0	0	0
	INCANDESCENTS	0	0	30.22	0.03	0	0	19.52	198.23	0	0	2
	laptop	205.06	5.09	0.82	0	0	0.14	0.7	0	29.39	8.8	0
	monitor	164.41	1.56	6.72	0.01	0	0.16	5.18	0	0.03	71.93	0
	solderingiron	0	0	37.6	1.22	0	0	43.7	0	0	0	167.48
Sum of classifications into category:		1032.2	91.15	310.8	210.34	120.86	218.23	264.92	200.89	29.46	101.46	169.72

Figure 5.9: Single-load part of the confusion matrix (average of 100 runs) where the training data did not contain any single-load samples

	Class. Accuracy		Num. Pred. Accuracy		Strict Accuracy	
multi-load train samples	AVG	Worst	AVG	Worst	AVG	Worst
70	98.17%	79.23%	99.13%	92.62%	97.39%	73.39%
10	91.59%	69.27%	95.61%	80.69%	88.44%	60.95%

We can see that although there was a 7-fold reduction in the number of multi-load training samples used, the average accuracy rates are still high. Additionally, by only requiring 10 measurements for each multi-load combination, **the total measurement time is about 90 minutes instead of 9 hours.**

5.2.2 From multi-load to single-load

For this scenario, we aim to examine the CNN's output for the single-load measurements when it was trained on only the multi-load data. For obvious reasons, we are only interested in the first part of the prediction (the load categories) since the model would predict N to be either two or three as it has not encountered any one-load case during training. The training set consisted of 90 measurements from each multi-load category. The test set contained 10 measurements from the multi-load categories and 250 measurements

		Predicted class										
True class	Correct load in TOP 2 for 89.31% of samples	USB	batterycharger4A	batterycharger800mA	fan	hairdryer	ledbulb	ledspotlight	INCANDESCENTS	laptop	monitor	solderingiron
	USB	247.7	44.25	23.2	0	0.01	16.55	44.29	0	0	124	0
	batterycharger4A	234.95	222.24	6.59	0.01	0	23.4	2.64	0	0.14	10.01	0.02
	batterycharger800mA	177.67	0	239.53	0	2.5	2.48	73.25	0	0	4.57	0
	fan	0	0	48.17	246.16	0	0	205.67	0	0	0	0
	hairdryer	0.56	0.12	0.11	1.41	247.45	248.11	1.09	0.57	0.33	0.17	0.08
	ledbulb	244.58	10.71	0.8	0	0	236.47	5.5	0	0	1.94	0
	ledspotlight	163.05	0	127.74	0	0	0	208.8	0	0	0.41	0
	INCANDESCENTS	1.66	0	150.77	7.88	0	0	91.33	247.03	0	0	1.33
	laptop	236.65	40.92	16.74	0.43	0.01	0.93	5.82	0.91	123.79	73.8	0
	monitor	241.71	19.13	26.93	0.08	0	5.2	7.81	0	0.62	198.5	0.02
	solderingiron	0	0	145.39	4.05	0	0	112.16	0	0	0	238.4
Sum of classifications into category:		1548.5	337.37	785.97	260.02	249.97	533.14	758.36	248.51	124.88	413.4	239.85

Figure 5.10: Single-load part of the confusion matrix (average of 100 runs) containing the top 2 predictions of the CNN. The training data did not contain any single-load samples.

from each single-load category. The results (average of 100 runs) are shown in Figure 5.9. We can see an interesting pattern. Although for some loads like USB, fan, and INCANDESCENTS most of the samples were classified correctly, for others, a significant portion of the samples were classified into other categories, most notably to USB. We can observe that almost all misclassifications happen to loads that consume less power than the true load category. This is most visible with the USB category consisting of USB chargers that consume 5-12W. A reason for this strange phenomenon may be due to the training of the CNN. The network was trained to provide two (or three) highest outputs, and there was no significant penalty if there was a slight difference between these two (or three) highest firing levels. It may be the case that the correct output is present in the top 2 output classes. To verify this hypothesis, we can rerun the training, but this time record the top 2 output classes for each single-load test sample. The results are shown in Figure 5.10. As we can see, the true class is present in the top 2 predictions of the model for 89.31% of the test samples. We can see that there are still a lot of samples that are incorrectly classified to USB from the top 2 outputs, but for all of these single-class load categories, the true category is also present in the top 2 predictions. With these results, we have shown that from only multi-load samples, the network learned to a large extent, how to classify single-load measurements.

	battery charger 4A battery charger 800mA fan hair dryer led bulb led spotlight INCANDESCENTS laptop monitor soldering iron										Num. pred. loads: 1	Num. pred. loads: 2	Num. pred. loads: 3	At least one correct in top N	Class. Accuracy	Strict Accuracy
USB	battery charger 4A	battery charger 800mA	fan	hair dryer	led bulb	led spotlight	INCANDESCENTS	laptop	monitor	soldering iron						
battery charger 4A+fan	9.7	65.5	0.2	0.1		24.8	35.2	2.3	55.3	6.9		100		65.6%	0.0%	0.0%
battery charger 4A+INCANDESCENTS	13.6	1.5		61.1	1.7	56.6	0.4	0.6	7	10.4	47.1	0.5	99.5	0	2.1%	0.0%
battery charger 4A+laptop	12.5	56.2	10.4	11.9		57.7	4.1		39.8	7.4		0.1	99.9	0	86.1%	9.9%
battery charger 4A+led bulb	18.1	92	1.6	0.1		28	30.9		29.3			1.7	98.3	0	100.0%	20.0%
battery charger 4A+led spotlight	40.4	46.1	14.7	2.7		59.3	17.1	0.1	10.2	9.4		4.5	95.5	0	59.0%	4.2%
USB+battery charger 4A		100				11.3		4.7	70.4	13.6		100	0	100.0%	0.0%	0.0%
battery charger 800mA+fan			1.1	99.7			92.4				6.8	63.6	36.4	0	99.7%	1.1%
battery charger 800mA+INCANDESCENTS				54.6			37.6	100			7.8	71.8	28.2	0	100.0%	0.0%
battery charger 800mA+laptop	14	4.7	46.9	16.7	0.1	0.2	31.6	0.4	29.6	55.7	0.1	5.2	94.8	0	71.4%	5.1%
battery charger 800mA+led bulb	5	10	5	0.1		99.5	70			10.4		10	90	0	99.5%	5.0%
battery charger 800mA+led spotlight	0.4		55	74.6			70					100	0	0	100.0%	25.0%
battery charger 800mA+monitor	77.4	2.2	16.4				19.7		72	12.3		0.2	99.8	0	28.7%	0.0%
battery charger 800mA+soldering iron				42.3	10		47.7				100	39.8	60.2	0	100.0%	0.0%
USB+battery charger 800mA	75.5		100	0.4		4.1				20		13	87	0	100.0%	75.5%
fan+hair dryer					100			9.9		90.1			100	0	100.0%	0.0%
fan+INCANDESCENTS			52				26.7	100			21.3	2.9	96.9	0.2	100.0%	0.0%
fan+laptop				45.3		0.4	6.8	19.5	37.5	72.1	18.4		100	0	77.6%	5.2%
fan+led spotlight			75.3	99.8							24.9	71.8	28.2	0	99.8%	0.0%
fan+monitor	10.4	2.7		56.8	0.1	0.1	0.8	15.2	62.4	36.6	14.9	0.1	99.9	0	83.1%	10.3%
fan+soldering iron			1.2	37.3			70.9				90.6	9.2	17.3	73.5	99.9%	28.0%
USB+fan	58			20.1		3.6	17.9			21.2	79.2	10.9	89.1	0	64.3%	13.8%
hair dryer+INCANDESCENTS				19.6	83.1		0.1				97.2		93.2	6.8	83.1%	0.0%
hair dryer+led bulb	17.3		1.9	19.1	92.9	26.7	17.6	0.9	0.5	0.3	22.8	71.6	27.9	0.5	98.8%	20.8%
hair dryer+led spotlight				50.6	55		30.2				64.2		78.7	21.3	80.2%	5.0%
hair dryer+soldering iron				97.6	100		0.1	2.3					100	0	100.0%	0.0%
INCANDESCENTS+laptop	0.3	9.9		16.1		4.7		8.2	1	75.6	84.2		100	0	9.2%	0.0%
led bulb+INCANDESCENTS				26.6	5.5	91.5		67.2		9.2			41.9	58.1	100.0%	58.7%
led spotlight+INCANDESCENTS			3.7	3.7				100			92.6	49.2	50.8	0	100.0%	0.0%
INCANDESCENTS+soldering iron			3.1	30.4			91.5	75					89.1	10.9	75.0%	0.0%
USB+INCANDESCENTS		10	8	16.8	10	20.9	10.9	100			23.4	52.2	38.1	9.7	100.0%	0.0%
led bulb+laptop	3.6	86.3	0.2	30.8	0.1			11.3	67.7				100	0	67.7%	0.0%
led spotlight+laptop	9.2		6.6	32	2.7		17.1		70.9	61.5			100	0	80.5%	7.5%
USB+laptop	11.3	26.6	0.6		0.6	35.2	9.5	0.5	100	15.7		11.9	88.1	0	100.0%	11.3%
fan+led bulb		14.1		40	0.1	49.9	9.2	35.5	0.4	11.4	39.4		100	0	86.8%	3.1%
led bulb+led spotlight		50.1	1.8	3.7	5.1	84.8	39.5			15			100	0	96.0%	28.3%
led bulb+soldering iron		2.1		63.2	0.5	88.9		44.3		0.6	0.4		100	0	89.0%	0.3%
USB+led bulb	0.4	29.1	1.6		8.8	69.7	81.9			8.5		48.5	51.5	0	69.7%	0.4%
led spotlight+monitor	36.2	2.9	12.1	6.3	0.9	20	88.5		21.8	11.2	0.1	0.1	99.9	0	94.6%	5.1%
led spotlight+soldering iron			21.6	0.6			78				99.8	50.4	40.2	9.4	100.0%	77.8%
USB+led spotlight	35		15.9	1.8		4	90.2			53.1			100	0	97.8%	27.4%
monitor+soldering iron	5	0.4	0.3	89.5		11.1	0.7	16.1	35.2	40.9	0.8	0.1	99.9	0	41.7%	0.0%
USB+monitor	31.7	7.1	33.8	13.5		0.1	10.3		10.8	84.9	7.8	78.9	21.1	0	87.6%	29.0%
USB+soldering iron	89.8	0.7		35.4		10.3	4.7	46.5			12.6	8.9	91.1	0	90.5%	11.9%
fan+led bulb+INCANDESCENTS	47.5				29.2	78.4	21.2	100			23.7	0.1	99.9	0	100.0%	0.0%
fan+led spotlight+soldering iron			11.3	100			51	37.7			100		100	0	100.0%	51.0%
Averages:														84.11%	12.02%	6.41%

Figure 5.11: Multi-load classification results. Each row represents the sample predictions belonging to the omitted multi-load combination. The training set for each row consisted of 70 samples per single and multi-load combinations (except the omitted one). The top N predictions (averages of 10 runs) are displayed in the left green part, the class number predictions in the blue part, and the average accuracy rates for the samples belonging to the omitted class are displayed in the right green part.

5.2.3 From multi-load to multi-load

The goal of examining multi-load to multi-load generalization is to see how the model interprets a multi-load input that was not part of the training data. Still, the loads present in the input have been in the training data both individually and in combination with other loads (e.g., we have the USB+fan combination that we omit from the training data; both the USB and the fan is present in the training data in several other combinations and as single-load measurements as well). By examining how the model classifies an unseen but familiar (i.e., has seen components of it) combination, we understand the model's limitations and capabilities. Figure 5.11 displays the average results of this method run ten times for each combination. We can see that in 88.11% of the cases, at least one load was correct in the top N predictions of the model. In 12.02% of the cases, all N category predictions of the model were correct. In 6.41% of the cases, in addition to the N category predictions being accurate, the load count (number of loads) prediction was also correct. These last two percentages are low numbers, which warrant an explanation. Some of the error certainly comes from the network's inability to correctly model the interaction of two loads and their resulting measurement matrix. Another significant portion of the error may come from cases when the model assumes that only one load is present. If we look at the rows in Figure 5.11 where the first blue column's (*Num. pred. loads: 1*) values are high, most of the combinations consist of two loads that have significantly different power consumption levels (e.g., batterycharger800mA+fan, batterycharger800mA+INCANDESCENTS, USB+INCANDESCENTS, USB+monitor, USB+batterycharger4A, hairdryer+ledbulb). In these cases, it is understandable that the model predicts only one load and picks the one with the higher power consumption. It is an important observation to consider when collecting multi-load measurement data.

6 Conclusions

In this paper we have presented a Smart Plug capable of detecting multiple loads simultaneously. A large dataset was collected to enable the evaluation of our novel method. We have shown that the Neural Network we designed is capable of accurately detecting at least up to three different loads connected to the same Smart Plug simultaneously. With our strict accuracy measure, 97.39% average accuracy rate was achieved.

We have also evaluated the generalization capabilities of our solution by examining how the model responds to unseen data related to loads that it has encountered as part of the training data as either part of a different combination or as an individual load. By training the model on only single-load measurements and seeing how the model classifies multi-load measurements, we could see that in 78.66% of the cases, at least one load (of the N loads) was correctly identified in the top N prediction of the sample. We have also shown that only 10 samples per multi-load category are enough to reach an average strict accuracy rate of 88.44%. Classifying single-load measurement samples using a model trained on only multi-load samples, an accuracy rate of 55.3% was achieved while the correct category was in the top 2 prediction of the model’s output for 89.31% of samples. To see how the model would classify a new combination of previously seen loads (both single-load and multi-load measurements in the training set, excluding the selected specific combination), we trained several models where we excluded one specific multi-load combination from the training data. We could then evaluate how the model would classify that unseen combination. Although the average strict accuracy rate was low (6.41%), in 84.11% of the cases, at least one correct load was present in the top n prediction of the model.

To summarize, we have shown that multi-load detection and, thus, a Smart Plug with hybrid ILM-NILM detection capabilities is possible. With this, we are one step closer to enabling our Smart Plug to function in a real-world environment. Real-world applicability is rarely considered in the literature, and to the best of our knowledge, we are the first to successfully apply hybrid ILM-NILM detection methods to load detection using a Smart Plug.

6.1 Future work

In the future, we plan to deploy the model on the ESP32 microcontroller of the Smart Plug, allowing on-device multi-load detection capabilities. We also aim to explore disaggregation methods to determine the power usage of the individual loads from the aggregated measurements, as the current system only detects the loads present but does not estimate the individual power consumption of the loads.

Acknowledgements

Supported by the ÚNKP-22-2-II-PPKE-8 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.



KULTURÁLIS ÉS INNOVÁCIÓS
MINISZTERIUM

Bibliography

- [1] S. M. Tabatabaei, S. Dick, and W. Xu, “Toward non-intrusive load monitoring via multi-label classification”, *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 26–40, 2017. DOI: [10.1109/TSG.2016.2584581](https://doi.org/10.1109/TSG.2016.2584581).
- [2] G. Tsoumakas and I. Vlahavas, “Random k-labelsets: An ensemble method for multilabel classification”, in *Machine Learning: ECML 2007*, J. N. Kok, J. Koronacki, R. L. d. Mantaras, S. Matwin, D. Mladenič, and A. Skowron, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 406–417, ISBN: 978-3-540-74958-5.
- [3] M.-L. Zhang and Z.-H. Zhou, “Ml-knn: A lazy learning approach to multi-label learning”, *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007, ISSN: 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2006.12.019>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320307000027>.
- [4] X. Wu, Y. Gao, and D. Jiao, “Multi-label classification based on random forest algorithm for non-intrusive load monitoring system”, *Processes*, vol. 7, no. 6, 2019, ISSN: 2227-9717. DOI: [10.3390/pr7060337](https://doi.org/10.3390/pr7060337). [Online]. Available: <https://www.mdpi.com/2227-9717/7/6/337>.
- [5] S. Verma, S. Singh, and A. Majumdar, “Multi-label lstm autoencoder for non-intrusive appliance load monitoring”, *Electric Power Systems Research*, vol. 199, p. 107414, 2021, ISSN: 0378-7796. DOI: <https://doi.org/10.1016/j.epsr.2021.107414>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378779621003953>.
- [6] M.-L. Zhang and Z.-H. Zhou, “Multilabel neural networks with applications to functional genomics and text categorization”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 10, pp. 1338–1351, 2006. DOI: [10.1109/TKDE.2006.162](https://doi.org/10.1109/TKDE.2006.162).

- [7] M. Bello, G. Nápoles, R. Sánchez, R. Bello, and K. Vanhoof, “Deep neural network to extract high-level features and labels in multi-label classification problems”, *Neurocomputing*, vol. 413, pp. 259–270, 2020, ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2020.06.117>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231220311115>.
- [8] L. Zhang, Y. Jin, X. Yang, *et al.*, “Convolutional neural network-based multi-label classification of pcb defects”, *The Journal of Engineering*, vol. 2018, no. 16, pp. 1612–1616, 2018. DOI: <https://doi.org/10.1049/joe.2018.8279>. eprint: <https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/joe.2018.8279>. [Online]. Available: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/joe.2018.8279>.
- [9] C. L. Athanasiadis, T. A. Papadopoulos, and D. I. Doukas, “Real-time non-intrusive load monitoring: A light-weight and scalable approach”, *Energy and Buildings*, vol. 253, p. 111 523, 2021, ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2021.111523>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378778821008070>.
- [10] C. Athanasiadis, D. Doukas, T. Papadopoulos, and A. Chrysopoulos, “A scalable real-time non-intrusive load monitoring system for the estimation of household appliance power consumption”, *Energies*, vol. 14, no. 3, 2021, ISSN: 1996-1073. DOI: [10.3390/en14030767](https://doi.org/10.3390/en14030767). [Online]. Available: <https://www.mdpi.com/1996-1073/14/3/767>.
- [11] A. Ridi, C. Gisler, and J. Hennebert, “A survey on intrusive load monitoring for appliance recognition”, in *2014 22nd International Conference on Pattern Recognition*, 2014, pp. 3702–3707.
- [12] L. Gomes, F. Sousa, and Z. Vale, “An intelligent smart plug with shared knowledge capabilities”, *Sensors*, vol. 18, no. 11, p. 3961, Nov. 2018. DOI: [10.3390/s18113961](https://doi.org/10.3390/s18113961).
- [13] Z. D. Tekler, R. Low, C. Yuen, and L. Blessing, “Plug-mate: An iot-based occupancy-driven plug load management system in smart buildings”, *Building and Environment*, vol. 223, p. 109 472, 2022, ISSN: 0360-1323. DOI: <https://doi.org/10.1016/j.buildenv.2022.109472>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S036013232200703X>.

- [14] Z. D. Tekler, R. Low, Y. Zhou, C. Yuen, L. Blessing, and C. Spanos, “Near-real-time plug load identification using low-frequency power data in office spaces: Experiments and applications”, *Applied Energy*, vol. 275, p. 115 391, 2020, ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2020.115391>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S030626192030903X>.
- [15] L. Du, D. He, R. G. Harley, and T. G. Habetler, “Electric load classification by binary voltage–current trajectory mapping”, *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 358–365, 2016.
- [16] A. F. da S. Veloso, R. G. de Oliveira, A. A. Rodrigues, R. A. L. Rabelo, and J. J. P. C. Rodrigues, “Cognitive smart plugs for signature identification of residential home appliance load using machine learning: From theory to practice”, in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–6.
- [17] T. Petrović and H. Morikawa, “Active sensing approach to electrical load classification by smart plug”, in *2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2017, pp. 1–5.
- [18] D. Németh. and K. Tornai., “Sp4lc: A method for recognizing power consumers in a smart plug”, in *Proceedings of the 11th International Conference on Smart Cities and Green ICT Systems - SMARTGREENS*, INSTICC, SciTePress, 2022, pp. 69–77, ISBN: 978-989-758-572-2. DOI: [10.5220/0010982800003203](https://doi.org/10.5220/0010982800003203).
- [19] D. I. Németh and K. Tornai, “Electrical load classification with open-set recognition”, *Energies*, vol. 16, no. 2, 2023, ISSN: 1996-1073. DOI: [10.3390/en16020800](https://doi.org/10.3390/en16020800). [Online]. Available: <https://www.mdpi.com/1996-1073/16/2/800>.
- [20] S. Nalley and A. LaRose, “Annual energy outlook 2022 (aeo2022)”, *Energy Information Agency*, p. 23, 2022.
- [21] S. Iqbal, M. Sarfraz, M. Ayyub, *et al.*, “A comprehensive review on residential demand side management strategies in smart grid environment”, *Sustainability*, vol. 13, no. 13, p. 7170, Jun. 2021, ISSN: 2071-1050. DOI: [10.3390/su13137170](https://doi.org/10.3390/su13137170). [Online]. Available: <http://dx.doi.org/10.3390/su13137170>.
- [22] J. A. Rama Curiel and J. Thakur, “A novel approach for direct load control of residential air conditioners for demand side management in developing regions”, *Energy*, vol. 258, p. 124 763, 2022, ISSN: 0360-5442. DOI: <https://doi.org/>

- 10.1016/j.energy.2022.124763. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360544222016668>.
- [23] Y. Zhang and N. Lu, "Demand-side management of air conditioning cooling loads for intra-hour load balancing", in *2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, 2013, pp. 1–6. DOI: 10.1109/ISGT.2013.6497905.
- [24] P. S. N. R. Patne, and A. M. Jadhav, "A smart demand side management mechanism for domestic energy consumers with major hvac load", in *2016 International Conference on Electrical Power and Energy Systems (ICEPES)*, 2016, pp. 504–511. DOI: 10.1109/ICEPES.2016.7915982.
- [25] J. Ma, X. Ma, and S. Ilic, "Hvac-based cooperative algorithms for demand side management in a microgrid", *Energies*, vol. 12, no. 22, 2019, ISSN: 1996-1073. DOI: 10.3390/en12224276. [Online]. Available: <https://www.mdpi.com/1996-1073/12/22/4276>.
- [26] B. Mahapatra and A. Nayyar, "Home energy management system (hems): Concept, architecture, infrastructure, challenges and energy management schemes", *Energy Systems*, vol. 13, no. 3, pp. 643–669, Aug. 2022, ISSN: 1868-3975. DOI: 10.1007/s12667-019-00364-w. [Online]. Available: <https://doi.org/10.1007/s12667-019-00364-w>.
- [27] N. Javaid, G. Hafeez, S. Iqbal, N. Alrajeh, M. S. Alabed, and M. Guizani, "Energy efficient integration of renewable energy sources in the smart grid for demand side management", *IEEE Access*, vol. 6, pp. 77 077–77 096, 2018. DOI: 10.1109/ACCESS.2018.2866461.
- [28] S. Noor, W. Yang, M. Guo, K. H. van Dam, and X. Wang, "Energy demand side management within micro-grid networks enhanced by blockchain", *Applied Energy*, vol. 228, pp. 1385–1398, 2018, ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2018.07.012>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261918310390>.
- [29] A. Bouakkaz, A. J. G. Mena, S. Haddad, and M. L. Ferrari, "Efficient energy scheduling considering cost reduction and energy saving in hybrid energy system with energy storage", *Journal of Energy Storage*, vol. 33, p. 101 887, 2021, ISSN: 2352-152X. DOI: <https://doi.org/10.1016/j.est.2020.101887>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352152X20317242>.

- [30] S. M. Hosseini, R. Carli, and M. Dotoli, “Robust optimal energy management of a residential microgrid under uncertainties on demand and renewable power generation”, *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 618–637, 2021. DOI: 10.1109/TASE.2020.2986269.
- [31] D. Caprino, M. L. Della Vedova, and T. Facchinetti, “Peak shaving through real-time scheduling of household appliances”, *Energy and Buildings*, vol. 75, pp. 133–148, 2014, ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2014.02.013>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378778814001248>.
- [32] D. I. Németh and K. Tornai, “Detecting unknown electrical loads using open set recognition”, in *2022 IEEE 10th International Conference on Smart Energy Grid Engineering (SEGE)*, 2022, pp. 7–11. DOI: 10.1109/SEGE55279.2022.9889770.
- [33] B. Völker, P. M. Scholls, T. Schubert, and B. Becker, “Towards the fusion of intrusive and non-intrusive load monitoring: A hybrid approach”, in *Proceedings of the Ninth International Conference on Future Energy Systems*, ser. e-Energy '18, Karlsruhe, Germany: Association for Computing Machinery, 2018, pp. 436–438, ISBN: 9781450357678. DOI: 10.1145/3208903.3212052. [Online]. Available: <https://doi.org/10.1145/3208903.3212052>.