Article

Power Profile and Thresholding Assisted Multi-Label NILM Classification

Muhammad Asif Ali Rehmani ^{1,2} , Saad Aslam ^{2,*} , Shafiqur Rahman Tito ² , Snjezana Soltic ²,
Pieter Nieuwoudt ² , Neel Pandey ² and Mollah Daud Ahmed ³

¹ Department of Mechanical and Electrical Engineering, SF&AT, Massey University, Auckland 0632, New Zealand; m.a.a.rehmani@massey.ac.nz

² School of Professional Engineering, Manukau Institute of Technology, Auckland 2104, New Zealand; shafiqur.tito@manukau.ac.nz (S.R.T.); snjezana.soltic@manukau.ac.nz (S.S.); pieter.nieuwoudt@manukau.ac.nz (P.N.); neel.pandey@manukau.ac.nz (N.P.)

³ Research Office, Manukau Institute of Technology, Auckland 2104, New Zealand; daud.ahmed@manukau.ac.nz

* Correspondence: saad.aslam@manukau.ac.nz

Abstract: Next-generation power systems aim at optimizing the energy consumption of household appliances by utilising computationally intelligent techniques, referred to as load monitoring. Non-intrusive load monitoring (NILM) is considered to be one of the most cost-effective methods for load classification. The objective is to segregate the energy consumption of individual appliances from their aggregated energy consumption. The extracted energy consumption of individual devices can then be used to achieve demand-side management and energy saving through optimal load management strategies. Machine learning (ML) has been popularly used to solve many complex problems including NILM. With the availability of the energy consumption datasets, various ML algorithms have been effectively trained and tested. However, most of the current methodologies for NILM employ neural networks only for a limited operational output level of appliances and their combinations (i.e., only for a small number of classes). On the contrary, this work depicts a more practical scenario where over a hundred different combinations were considered and labelled for the training and testing of various machine learning algorithms. Moreover, two novel concepts—i.e., thresholding/occurrence per million (OPM) along with power windowing—were utilised, which significantly improved the performance of the trained algorithms. All the trained algorithms were thoroughly evaluated using various performance parameters. The results shown demonstrate the effectiveness of thresholding and OPM concepts in classifying concurrently operating appliances using ML.

Keywords: non-intrusive load monitoring (NILM); machine learning; smart building energy management systems (SBEM); multiclassification; computational complexity; energy efficiency; demand side management

Citation: Rehmani, M.A.A.; Aslam, S.; Tito, S.R.; Soltic, S.; Nieuwoudt, P.; Pandey, N.; Ahmed, M.D. Power Profile and Thresholding Assisted Multi-Label NILM Classification. *Energies* **2021**, *14*, 7609. <https://doi.org/10.3390/en14227609>

Academic Editors: Justyna Żywiołek and Muhammad Asghar Khan

Received: 20 October 2021

Accepted: 12 November 2021

Published: 14 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There has been a greater focus on utilising renewable energy resources since the Kyoto Protocol in order to reduce the effects of greenhouse gasses, global warming and reduce our carbon footprint. The integration of unpredictable renewable energy into the power grid acts as the driving force towards the evolution of the existing grid system to smart grid—characterised by bi-directional power flow, control and two-way communication. This provides an opportunity to achieve enhanced energy efficiency through user participation. To implement this, next-generation power systems intend to exploit artificial intelligence and machine learning to design sustainable energy systems. These systems are likely to work with smart grids to optimise energy consumption [1]. The concept of smart building energy management systems (SBEMs) has become popular and consistent with

the smart grid concept [2]. The purpose of SBEM is to optimise the energy consumption of a building. Home appliances can be controlled efficiently by monitoring the cost of a consumer's energy usage and yielding a better return on investment for the utility provider. To realise this, we need to obtain the energy consumption of individual appliances. It is understood that some modern appliances can communicate with smart meters and the per appliance energy consumption can be obtained automatically; however, these appliances are expensive. Additionally, existing appliances that do have this advantage require alternate methods/strategies to classify the energy consumption of individual appliances. There are two types of strategies: intrusive and non-intrusive. Intrusive load monitoring (ILM) requires additional hardware and/or complex network to be installed to measure the energy consumption of individual devices, which introduces complexity with an additional system cost that makes it infeasible in many circumstances. NILM, on the other hand, extracts appliance level information from aggregated energy consumption information measured at a single point. Therefore, NILM presents an effective solution that relies on computationally intelligent techniques such as machine learning [3]. This paper focuses on the second technique, commonly known as non-intrusive load monitoring (NILM). Typical time-synchronised power profiles of a few appliances with respect to main meters are highlighted in Figure 1. Randomly operating concurrent appliances generate multiple levels of power output. The operating ON and OFF state of appliances are labelled as sequences of 1 s and 0 s, respectively. The length of the code is determined by the number of appliances of the household.

To perform the segregation of the power consumption of appliances, one of the fundamental tasks is to obtain and understand the features of the appliances. These features can represent various types of appliance data such as on/off trends, voltage and current, power consumption (real, reactive and apparent) and its temporal variations. There are various types of appliances, each with its own power profile. It is easy to segregate two devices with very different power profiles; however, if several appliances have approximately similar power profiles, then segregation becomes a difficult task. Moreover, devices that consume very low power are also a problem for classification since such low-level power can often be regarded as noise. One of the ways to tackle this is to observe the temporal variations of the power profiles for long periods of time [2].

Most of the recent works that consider NILM techniques focus on utilising complex learning algorithms such as Deep Neural Networks (DNN) and Convolutional Neural Networks (CNN) [1–3]. Though these algorithms are effective, they require high computational resources and a considerable time for training. Since the research is moving towards the green machine learning paradigm, we require learning algorithms that are simpler and can be obtained from relatively smaller datasets as compared to the requirements of DNN and CNN. In this paper, we obtain satisfactory results from less complex algorithms. Moreover, several machine learning models are trained for a comprehensive comparative analysis. To tackle the classification problems discussed above and reduce the number of datasets required to train the model, we introduce two novel concepts—i.e., thresholding (TH) and occurrence per million (OPM)—which are detailed in Section 3 of this work. These concepts have two-fold advantages: (i) simplicity of training and (ii) effectively addressing the non-uniform distribution of the practical datasets.

Figure 1. Power profile of appliances in REDD House 1.

The proposed model is robust against a large variety of appliances as opposed to classifying only a few devices as normally showcased in the research. The ML models considered in this study utilise the concepts of TH/OPMs and power windowing for training and are thoroughly tested and verified on a publicly available dataset known as the Reference Energy Disaggregation Dataset (REDD) [4] (<http://redd.csail.mit.edu/> last accessed: 13 November 2021). The trained models are comprehensively evaluated using various and well-known performance metrics. The main contributions of this work are enumerated as follows:

- Without using a large number of datasets and conventional DNN and CNN approaches, we have trained, tested and validated several classifiers based on a real-world dataset (REDD) that can achieve accuracies up to 98%.
- As opposed to recent research works that consider only a few appliances and their combinations, we have considered over a hundred combinations of various appliances to train the ML classifiers. This reflects a more practical scenario.
- We introduce the concepts of TH and OPM to tackle the non-uniformity of the dataset. Various randomly selected values of TH and OPMs are used to demonstrate the effect of their usage on various performance parameters. The results show a significant performance improvement with the utilisation of these concepts.
- A comprehensive comparative study of various machine learning classifiers is presented.

2. Literature Review

There has been an increasing interest in devising energy-efficient techniques for load monitoring. NILM has been described as one of the techniques that can be leveraged to provide energy-efficient solutions [5]. The energy segregation is obtained via the signatures of each appliance, which include but are not limited to power consumption, on/off status and temporal variations. The goal of NILM is to study and understand these signatures. This represents a significant challenge since different manufacturers produce different signatures for the purpose of extracting energy. Moreover, a single appliance can have different operating modes in which the power consumption can be very different. Moreover, since the power consumption can vary, this is often confused with noise [6]. Various studies have been presented in the literature to circumvent these challenges and apply the NILM techniques effectively. Hidden Markov Models (HM) and Factorial Markov Models (FHM) [7–10] are also widely utilised [2,11]. However, one of the problems with these techniques is the requirement of pre-existing knowledge about the number of appliances, which is assumed to be fixed. This may be impractical for many scenarios.

More recently, researchers have produced well-performing algorithms based on deep learning algorithms (DLA). DLAs and the recurrent neural network have been applied for appliance disaggregation [12]. Energy segregation techniques based on CNN have been proposed as well [13,14]. Another research work based on long short-term memory (LSTM) and the recurrent neural network (RNN) has been proposed in [2], which improves the accuracy of segregation.

A few research works have also investigated the combination of DNNs for segregating appliances. A similar approach has been presented in [15], that combines CNNs with autoencoders and improves the performance as compared to using conventional CNN. Research work presented in [16] also utilises CNN for NILM but with a pre-processed input referred to as a differential input. The differential power utilisation of devices is obtained after pre-processing the data, which become the input to CNN. This improves the classification performance.

Another interesting study has been presented in [17], which distributes the learning network into two parts: the main network and sub-network. The main network uses regression and the sub-network performs the classification of the appliances.

Most of the research works discussed above and others such as those presented in [17–37] typically rely on large datasets to train the ML algorithms. Moreover, most of these solutions predominantly rely on DNN and CNN techniques. Therefore, the computational complexity is large and such solutions can become difficult to implement practically. It is also evident from the literature review that most of the research works consider only a few appliances and their combination to segregate the aggregated power consumption. However, this does not represent a practical scenario. Therefore, more appliances and their combinations need to be considered as they exist in a real-world scenario. It should be noted that many of these works focus on classifying a particular appliance working at various power levels, thus terming this as multi-classification. However, the work proposed

in this study takes a more practical approach, where all home appliances are working simultaneously and, irrespective of the power levels, these appliances are classified.

It should be noted here that most of the research works apply either data augmentation or data pre-processing techniques to clean and uniformly distribute the data. These techniques consume more computational resources and complicate the learning process. To cater for this, we introduce two concepts; TH and OPM. These concepts eliminate the requirement of the data augmentation/complex pre-processing techniques and also require less computational resource for a similar data size for learning.

3. Methodology

3.1. Training Data Set

The reference energy disaggregation dataset commonly known as REDD is a widely accepted and utilised dataset for NILM techniques [1]. Most of the articles discussed in this work consider the REDD dataset as well. Since it was one of the first openly accessible datasets, it has matured and presents a wider range of information. Information on appliances (labels, power usage etc) for several houses are available. The dataset provides appliance-level power consumptions along with aggregated power consumption. This dataset is purposely built for the development and evaluation of energy segregation techniques. It is effective because it provides insights into both instantaneous and temporal variations in loads and their power profiles. This dataset is exploited for the training and testing of the algorithms utilised in this paper.

3.2. Multiclassification Problem

We approach the per appliance power segregation as a multiclassification problem. Each appliance and combinations of several appliances have their labels and ML classifiers train on this information to provide classification labels. This work makes use of aggregated power consumptions (having the on/off state of different appliances). As opposed to binary classification, this technique requires a sample to be mapped to multiple outputs. In particular, this work applies a variety of ML algorithms and performs the multiclassification to a tremendous degree of accuracy.

3.3. Simulation Setup

We have used the REDD dataset of House 1 and time-synchronised the aggregate power of the house with the respective appliances. Once the dataset was time-synchronised, the labels were generated based on their on/off state. We marked the label as “0” if it was off or “1” otherwise. The catenated appliance states were then generated as a class of uniquely operating appliances in the dataset. Our experimental setup consisted of six core Intel® i7-8750H mobile processors with 32 GB RAM running on the Microsoft® Windows 10® operating system. All the processing was conducted in the Python® language (version 3.8.8) while using sklearn and tensorflow modules for machine learning. The house 1 REDD dataset key statistic attributes are listed in Table 1.

Table 1. REDD House 1 appliance parameters.

S.No.	Appliance	Ch. No.	Maximum Power	Minimum Power	Mean Power	Std. Dev.	On States	Off States
1	Oven	3	1725	3	241	577	10,496	396,252
2	Oven	4	2565	3	372	798	8461	398,287
3	Refrigerator	5	2359	1	56	89	406,712	36
4	Dishwasher	6	1422	1	274	422	37,033	369,715
5	Kitchen Outlet	7	59	7	21	2	406,748	0
6	Kitchen Outlet	8	1550	3	28	17	406,748	0
7	Lighting	9	363	1	46	46	382,361	24,387
8	Washer Dryer	10	4475	1	88	182	18,162	388,586
9	Microwave	11	2906	1	22	164	406,748	0
10	Bathroom	12	1686	1	7	100	382,665	24,083
11	Electric Heater	13	192	1	7	4	7394	399,354
12	Stove	14	36	1	5	3	9279	397,469
13	Kitchen Outlet	15	1118	1	6	69	406,483	265
14	Kitchen Outlet	16	1585	1	76	324	10,883	395,865
15	Lighting	17	112	1	66	11	111,425	295,323
16	Lighting	18	90	1	13	23	405,221	1527
17	Washer Dryer	19	18	3	5	5	6	406,742
18	Washer Dryer	20	3223	3	2562	598	5058	401,690

Table 2. Attributes of classification algorithms.

Parameters	CART	KNN	KNN (City Block)	LDA	NB	ET	RF
Criterion	gini	N/A	N/A	N/A	N/A	gini	gini
Splitter	best	N/A	N/A	N/A	N/A	random	N/A
Minimum sample split	2	N/A	N/A	N/A	N/A	2	2
Minimum sample leaf	1	N/A	N/A	N/A	N/A	1	1
Neighbours	N/A	5	10	N/A	N/A	N/A	N/A
Weight	N/A	uniform	distance	N/A	N/A	N/A	N/A
Leaf size	N/A	30	30	N/A	N/A	N/A	N/A
Distance metric	N/A	Minkowski	City Block	N/A	N/A	N/A	N/A
Distance function	N/A	Euclidean distance	Manhattan distance	N/A	N/A	N/A	N/A
Solver	N/A	N/A	N/A	SVD	N/A	N/A	N/A
Shrinkage	N/A	N/A	N/A	None	N/A	N/A	N/A
Priors	N/A	N/A	N/A	None	N/A	N/A	N/A
Tolerance	N/A	N/A	N/A	1.00×10^{-4}	N/A	N/A	N/A
Smoothing Variance	N/A	N/A	N/A	N/A	1.00×10^{-9}	N/A	N/A
Estimators	N/A	N/A	N/A	N/A	N/A	N/A	100

3.4. Dataset Optimisation

In order to efficiently perform machine learning classification based on multiclassification, the dataset needs to be tuned well for high prediction accuracies. In order to do so, it is important to understand the operation of appliances through their power profiles and gathered data. To implement better prediction accuracies through machine learning algorithms, we have introduced a hybrid approach to label the data.

The first approach defines the operating power window for the appliances. For instance, we have selected the power operating window of a microwave appliance from 20 watts to 1650 watts in the REDD House 1 dataset. The lower limit indicates that the microwave was either on standby mode or when the door of the microwave was open, it activated the light inside the microwave, generating a small power spike in the time domain. However, once the heating cycle started, the power consumption of the microwave

increased with respect to the settings selected by the user and was well above the power consumption of standby mode. The power window was selected through the survey of power profiles of various appliances reported in the literature [38–43] and by analysing the time domain appliance usage pattern from the House 1 REDD dataset.

Multiclassification labels were generated for the microwave at a particular time utilising the power profile operating window. The same strategy was followed for all the other appliances in the REDD House 1 dataset. Table 3 provides the power window bins for House 1 in the REDD database where the channel number indicates the measured power at a particular channel of the power meter. Power profile values that do not fall in the range of the power window are considered as not operating or turned off at that particular instance. This strategy has removed power spikes or outliers in the dataset for a particular appliance in the time domain.

Table 3. Power window of appliances in House 1 in the REDD database.

Appliances	Channel No.	Power Window (Watts)	
		Lower Bound	Upper Bound
Oven	3	1500	1800
Oven	4	1500	2600
Refrigerator	5	175	500
Dishwasher	6	30	1200
Kitchen outlets	7	10	41
Kitchen outlets	8	10	150
Lighting	9	20	400
Washer dryer	10	250	700
Microwave	11	20	1650
Bathroom GFI	12	1500	1700
Electric heater	13	1	21
Stove	14	1000	1500
Kitchen	15	500	1200
Kitchen	16	1200	1700
Lighting	17	20	115
Lighting	18	5	100
Washer dryer	19	1	25
Washer dryer	20	1000	3500

The second approach we have introduced is to identify the unique combination of appliances that seldom operate in the dataset. The operations of appliances is random as it is based on the resident's time spent during a particular day, weather conditions and variations in appliance usage due to the occupant's personal choice or due to some random event. The overall power consumed by a particular resident is the sum of power consumed by all the appliances. We noticed that there are certain unique combinations of devices that seldom operate simultaneously. Such unique combinations that seldom operate make the classification of appliances based on aggregate power difficult due to a lack of training data. In order to correctly classify the appliance based on aggregate power and to increase the overall classification accuracy, it is paramount to identify seldom occurring concurrent unique combinations of appliances and remove them from the dataset. To implement this, we have used a threshold or occurrence per million (OPM) frequency of such unique combinations. The threshold value from 5 to 50 with a step of 5 is chosen to remove unique combinations of seldom operating appliances that fall below the set threshold value or frequency. For instance, the House 1 dataset for REDD has 406,748 instances of operations, which are synced in the time domain with the main meter. Selecting a threshold value of 5 equates to the unique instances of the events being at least 12 per million for this particular dataset. Any unique combination of appliances that falls below the threshold of 12 per million, or in the case of the House 1 REDD dataset below 5 instances, was dropped from the dataset. This procedure resulted in the feature reduction of unique labels, which

makes the classification more effective at the cost of removing a particular set of events that is less probable to happen in future.

The above strategy resulted in the generation of 22 different datasets for House 1 in the REDD dataset. These datasets include 10 datasets for different OPMs that do not include any power window, and additional 10 datasets for which the power window has been applied along with the different OPMs and finally, two datasets in which one dataset contained only power window multiclassification label and another dataset in which neither the power window nor OPMs were applied to generate the classification labels.

4. Results

The data preprocessing mainly comprised of thresholding and windowing concepts, which have already been described in the methodology section. All the segments of the data obtained from House 1 were extracted from the aggregate power reading. These segments of aggregated values were used for training and testing purposes. The concept of window size was introduced to depict a practical operational window of every appliance since each appliance has a specific power profile. We removed all the data sequences where the power profiles readings were too large or small for the given window, practically representing noise. Moreover, the concept of thresholding was used as well to cater for the non-uniformity of data (i.e., imbalance data). Therefore, different values of thresholding were used to evaluate the performance of the ML algorithms.

4.1. Performance Metrics

Data distribution plays a significant role in the training process, specifically for a multiclassification problem [44–46]. The accuracy of the trained algorithms, in one way or the other, also depends on the data distribution. In general, we prefer machine learning algorithm to have higher weighted precision and recall scores for evaluating performance. However, there is a tradeoff between weighted precision and the recall score performance metric. Often, tuning a machine learning algorithm for high-weighted precision results in lower-weighted recall scores, and vice versa. The better approach is to consider the weighted $F1$ -score, which is a function of precision and recall. The $F1$ -score combines the precision and recall by calculating the harmonic mean of precision and recall. However, for class imbalanced data, even the weighted $F1$ -score of the machine learning algorithm can be misleading in most cases as it assigns weights to the classes based on their sample size in the dataset to calculate the weighted precision, recall and the resulting $F1$ -score, thus favouring the majority class. The weighted $F1$ -score can be mathematically represented by Equation (1):

$$F1_{Class1} * W_1 + F1_{Class2} * W_2 + \dots + F1_{ClassN} * W_N \quad (1)$$

4.2. Micro and Macro Averaging

In order to improve the prediction of class imbalance data for multi-label/multiclassification, it is better to investigate the micro and macro performance metrics in the Python sklearn library for machine learning. Micro performance metrics in sklearn compute the precision, recall and resulting $F1$ -score by considering the total true positives, false negatives and false positives without considering the proportion of predictions for each label in the dataset denoted by Equation (2).

$$F1_{Class1+Class2+\dots+ClassN} \quad (2)$$

On the other hand, macro performance metric in sklearn computes the precision, recall and resulting $F1$ -score for each label and returns the average of each performance metric by incorporating the proportion of each label in the dataset while evaluating the predicted value against the actual state in the dataset denoted by Equation (3):

$$F1_{Class1} + F1_{Class2} + \dots + F1_{ClassN} \quad (3)$$

In a nutshell, the macro-averaging performance metric computes a given metric (i.e., accuracy, precision, recall and *F1*-score) independent of the sample size of each class and then takes the average, while micro-averaging aggregates the metric of all classes and computes the average.

A simple weighted accuracy does not cater for the frequency of a particular class and therefore does not represent the true performance of the algorithm. Let us consider an example where 10 data samples are considered and a total of 3 classes exist in the data. It is assumed that 1 class appears 6 times in the data, whereas the frequencies of the other 2 are 3 and 1, respectively. If a machine learning algorithm is trained on such data, the algorithm will already be biased, with 1 class appearing 60% of the time.

Since the macro performance metric does not include the label proportions of the dataset to compute the performance metric, it is a better performance evaluation metric for class-imbalanced multi-label/multiclassification datasets.

4.3. Performance Evaluation

4.3.1. Accuracy

A comprehensive performance evaluation has been performed for the trained ML algorithms. The performance parameters such as accuracy, precision, *F1*-score and processing times have been considered. It should be noted that, since this work introduces the concept of windowing and thresholding, the impact of these concepts on various parameters has been detailed as well. Moreover, as discussed in the preceding section, due to the presence of class imbalances (different numbers of input instances for each class), considering simple weighted accuracy does not depict the correct evaluation of the algorithms. In this regard, weighted accuracies and macro accuracies for various machine learning algorithms have been computed for the class-imbalanced REDD House 1 dataset, as shown in Figures 2 and 3. It can be seen that there is a huge difference between the two accuracies which represents the stated facts. While the weighted accuracy is more than 90% for some algorithms. It drops down to only 37% when macro accuracy is considered. Ideally, these two values should be as close to each other as possible, but this is not the case as the REDD House 1 dataset is a class-imbalanced dataset. To improve the macro accuracy, we demonstrate the impact of utilising the concepts of OPMs/thresholding and power windowing on the trained algorithms by reevaluating the macro accuracy on the REDD House 1 dataset. The tabulated results in Table 4 present the quantitative advantages of employing OPMs/thresholding and windowing on the evaluated metrics of the various machine learning algorithms.

Table 4. Performance of the trained machine learning algorithms.

S.No.	Algorithm	Power Window	OPMs	Macro Precision	Macro Recall	Macro <i>F1</i> Score	Training Samples	Testing Samples
1	CART	No	123	0.66	0.67	0.66	323,560	80,891
		Yes	123	0.96	0.96	0.96	324,114	81,029
2	ET	No	123	0.69	0.68	0.68	323,560	80,891
		Yes	123	0.98	0.97	0.97	324,114	81,029
3	KNN	No	123	0.73	0.70	0.71	323,560	80,891
		Yes	123	0.99	0.97	0.98	324,114	81,029
4	KNN (CB)	No	123	0.73	0.71	0.71	323,560	80,891
		Yes	123	0.99	0.98	0.98	324,114	81,029
5	LDA	No	123	0.27	0.31	0.25	323,560	80,891
		Yes	123	0.42	0.44	0.40	324,114	81,029
6	NB	No	123	0.34	0.42	0.33	323,560	80,891
		Yes	123	0.52	0.57	0.51	324,114	81,029
7	RF	No	123	0.70	0.69	0.69	323,560	80,891
		Yes	123	0.97	0.97	0.97	324,114	81,029

Figure 2. Weighted accuracy of the trained algorithms.

Figure 3. Macro accuracy of the trained algorithms.

It should be noted that OPMs/thresholding and power windowing greatly improved the results of all the machine learning algorithms, irrespective of their working principles. Figure 4 provides the variation of macro precision with respect to OPMs/thresholding and the power windowing of considered algorithms.

Figure 4. Impact of the power profile and thresholding on the macro accuracy of the trained algorithms.

It can be observed from Table 4 that the KNN-City Block algorithm achieves the best accuracy as compared to the rest of the algorithms. However, the performance of a few of the other algorithms, such as KNN, RF and ET, remains very close to the best-performing algorithm. This result demonstrates that, even without deep learning techniques, significant performance can be achieved.

Another important aspect of this result is the impact of OPM, windowing and thresholding on the accuracy of the trained algorithms. It can be clearly seen that with the increase in thresholding, accuracy curves with OPMs and windowing significantly increase for all the implemented algorithms. The three curves from bottom to top represent the accuracy without OPM and windowing concepts (we call it baseline accuracy), accuracy with OPMs but without windowing and lastly accuracy with both OPM and windowing concepts, respectively. The percentage increase in the accuracy of all the implemented algorithms as a result of applying OPMs with windowing is shown in the Table 5. Similar trends are observed for precision and F1 score.

Table 5. Percentage increase in accuracy.

Algorithm	With Respect to Base Line			With Respect to OPM No Windowing		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
CART	145%	132%	146%	46%	42%	45%
ET	122%	113%	127%	42%	43%	43%
KNN	101%	96%	102%	36%	38%	38%
KNN (City Block)	108%	102%	109%	35%	38%	38%
LDA	169%	140%	154%	56%	43%	56%
NB	141%	145%	155%	53%	36%	53%
RF	105%	102%	108%	39%	41%	41%

4.3.2. Per Appliance Performance Evaluation

To evaluate the performance of the trained algorithm on individual appliances, the accuracy, precision and F1-score of five different appliances are detailed in Table 6. It can be seen that the trained algorithms can classify individual appliances with high confidence.

Table 6. Per appliance performance metrics.

	Algorithms	Accuracy	Precision	F1-Score
Bathroom GFI	LDA	0.9588	0.8299	0.957
	KNN	0.9864	0.9616	0.9861
	KNN City Block	0.9866	0.9627	0.9862
	CART DT	0.9786	0.9022	0.9787
	NB	0.9508	0.7843	0.9492
	RF	0.9859	0.9587	0.9856
	ET	0.9849	0.9487	0.9845
Refrigerator	LDA	0.7497	0.6277	0.6709
	KNN	0.9996	0.9992	0.9996
	KNN City Block	0.9996	0.9992	0.9996
	CART DT	0.999	0.9986	0.999
	NB	0.7363	0.563	0.6653
	RF	0.9994	0.999	0.9994
	ET	0.9994	0.9991	0.9994
Lighting	LDA	0.7378	0.68	0.6484
	KNN	0.9999	0.9999	0.9999
	KNN City Block	0.9999	0.9999	0.9999
	CART DT	0.9994	0.9992	0.9994
	NB	0.7412	0.6684	0.6723
	RF	0.9996	0.9996	0.9996
	ET	0.9997	0.9996	0.9997
Oven	LDA	0.9882	0.6208	0.9917
	KNN	0.9999	0.9967	0.9999
	KNN City Block	1	0.9967	1
	CART DT	0.9999	0.9919	0.9999
	NB	0.9824	0.5881	0.9885
	RF	1	0.9967	1
	ET	0.9999	0.9951	0.9999
Kitchen Outlets	LDA	0.9866	0.6148	0.9907
	KNN	0.9999	0.9911	0.9999
	KNN City Block	0.9999	0.9911	0.9999
	CART DT	0.9997	0.9823	0.9997
	NB	0.9695	0.5293	0.9808
	RF	0.9998	0.9853	0.9998
	ET	0.9998	0.9911	0.9998

4.3.3. Processing Time

Another parameter of interest is processing time. One of the important aspects of this study is achieving the desired accuracies without implementing computationally complex algorithms such as DNN. Moreover, as a result of applying OPM and windowing, the processing time decreases. Figure 5 highlights the variations in processing time for the four best algorithms. It can be seen from the results that the processing times for CART, LDA, RF and ET have tremendously decreased with the increase in thresholding. The percentage decreases in processing times from the baseline are 50%, 47%, 61% and 67% respectively. For both versions of KNN, the processing times do not significantly vary but improve substantially with respect to the baseline. The percentage decreases in processing times of all the algorithms are detailed in Table 7.

Table 7. Percentage decrease in processing time.

Algorithm	With Respect to Base Line	With Respect to OPM no Windowing
CART	50%	25%
ET	67%	48%
KNN	−7%	0%
KNN (City Block)	−9%	−1%
LDA	41%	−2%
NB	56%	−1%
RF	61%	33%

4.3.4. Class Variations

The samples obtained from the REDD dataset contain a huge number of appliance combinations. However, we believe that appliances can be classified even with a smaller number of combinations. More appliance combinations makes the learning process more complex. Therefore, another advantage of using OPM and windowing concepts is a decrease in the number of appliance combinations. The decrease in unique combinations of the classes reduces the complexity of machine learning and leads to an improvement of the respective accuracy and processing time, as shown in previous sets of results. Figure 6 highlights the impact of OPM and power windowing on the number of classes required to train the ML algorithms.

It can be seen that, for most of the classifiers, the number of classes required to segregate the appliances decreases drastically as we increase the thresholding. It is interesting to observe that this trend is valid for both curves representing OPM with and without windowing. The percentage decrease in the number of classes for all the algorithms is presented in Table 8.

Table 8. Percentage decrease in classes.

Algorithm	Percentage Decrease	
	With Respect to Base Line	With Respect to OPM No Windowing
CART		
ET		
KNN		
KNN (City Block)	80%	3%
LDA		
NB		
RF

Figure 5. Processing time variation of the trained algorithms.

Figure 6. Unique class variation with respect to OPMs and power window.

4.3.5. Training Samples

Another important aspect of utilising OPM and windowing is that, without an appreciable decrease in the data size/number of training samples, an improvement in performance metrics can be realised. Figure 7 illustrates the variation of the number of samples required to train the ML algorithm as a result of applying OPMs/thresholding and power windowing data optimisation techniques. It should be noted that at the expense of less than a 1% decrease in data size, there is an overall improvement in accuracy and processing time of more than 100% and 40%, respectively, for the best-performing machine learning algorithms.

Figure 7. Training sample variations for all algorithms.

4.3.6. Performance Comparison

The performance of the proposed methodology (the best-performing algorithm) is compared with the recent NILM approaches. Different parameters such as accuracy, precision and *F1*-score have been compared. It can be observed from Table 9 that the proposed method performs better compared to approaches in other relevant research works. This implies that, applying the concepts of OPM and TH, the NILM system can accurately classify different appliances. It should be noted that not all research works report every parameter considered; therefore, only the reported parameters have been presented in the table.

Table 9. Performance comparison.

Research Works	Accuracy	Precision	<i>F1</i> -Score
[47]	-	0.91	0.93
[48]	0.95	-	0.68
[49]	-	-	0.81
[50]	-	-	0.90
[51]	0.98	-	0.98
[52]	-	-	0.78
[53]	0.92	-	-
[54]	0.95	-	0.89
[3]	0.58	-	-
[55]	0.70	-	-
[56]	0.89	-	-
[27]	-	-	0.91
[24]	0.95	0.95	0.95
This work	0.99	0.98	0.98

5. Conclusions

In this paper, we investigated the problem of the classification of multiple appliances operating at a particular time through their aggregate power. For this purpose, various ML algorithms were employed. This research demonstrated that, without applying computationally complex algorithms such as DNN, CNN or RNN, we can still achieve an accuracy of close to 99% while utilising only a moderate number of samples. Moreover, in order to handle imbalanced data while improving metrics such as accuracy, processing time and the number of appliance combinations, the concepts of OPM and windowing were utilised. The results demonstrate that, with the application of OPM and windowing, the accuracy increased by more than 100% for all the algorithms and the processing time decreased by 40% for LDA, NB, RF and ET, respectively. This is significant not only for NILM approaches but also for the development of the futuristic approach of green ML, where a low-energy footprint of computational resources is expected in the prediction of load demand and appliance usage patterns.

In the future, we intend to consider the temporal variations of power profiles and train the ML algorithms accordingly. Moreover, we also aim to demonstrate a comparative study of segregating the appliances of different houses representing a different sample/data distribution.

Author Contributions: Conceptualisation, methodology, M.A.A.R., S.A., S.R.T., S.S. and P.N.; investigation, visualisation, writing—original draft preparation, M.A.A.R., S.A. and S.R.T.; writing—review and editing, M.A.A.R., S.A., S.R.T., S.S., P.N., N.P. and M.D.A.; resources, supervision, project administration, funding acquisition, S.R.T., N.P. and M.D.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by MIT Strategic Research Fund 2021 grant number SRF-2021-RP0034.

Data Availability Statement: Data used for processing of machine learning algorithms can be accessed from an open source REDD house repository through this link: <http://redd.csail.mit.edu/>.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CART DT	Classification and Regression Trees—Decision Tree
CNN	Convolutional Neural Network
DLA	Deep Learning Algorithms
DNN	Deep Neural Network
ET	Extra Trees (also referred to as Extremely Randomised Trees)
FHM	Factorial Markov Models
HM	Hidden Markov Models
KNN	k-Nearest Neighbors
KNN-CB	k-Nearest Neighbors—City Block
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
ML	Machine Learning
NB	Naïve Bayes
OPM	Occurrence per million
REDD	Reference Energy Disaggregation Dataset
RF	Random Forest
RNN	Recurrent Neural Network
SBEM	Smart Building Energy Management System
TH	Thresholding

References

1. Khan, M.M.R.; Siddique, M.A.B.; Sakib, S. Non-intrusive electrical appliances monitoring and classification using K-nearest neighbors. In Proceedings of the 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), Dhaka, Bangladesh, 23–24 December 2019; pp. 1–5.
2. Kim, J.; Le, T.T.H.; Kim, H. Nonintrusive load monitoring based on advanced deep learning and novel signature. *Comput. Intell. Neurosci.* **2017**, *2017*, 1–21. [CrossRef] [PubMed]
3. Tabatabaei, S.M.; Dick, S.; Xu, W. Toward non-intrusive load monitoring via multi-label classification. *IEEE Trans. Smart Grid* **2016**, *8*, 26–40. [CrossRef]
4. Kolter, J.Z.; Johnson, M.J. REDD: A public data set for energy disaggregation research. In Proceedings of the Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA, USA, 21 August 2011; Volume 25, pp. 59–62.
5. Hart, G.W. Nonintrusive appliance load monitoring. *Proc. IEEE* **1992**, *80*, 1870–1891. [CrossRef]
6. Shin, C.; Rho, S.; Lee, H.; Rhee, W. Data requirements for applying machine learning to energy disaggregation. *Energies* **2019**, *12*, 1696. [CrossRef]
7. Kolter, J.Z.; Jaakkola, T. Approximate inference in additive factorial hmms with application to energy disaggregation. In Proceedings of the Artificial Intelligence and Statistics, PMLR, La Palma, Spain, 21–23 April 2012; pp. 1472–1482. Available online: <http://proceedings.mlr.press/v22/zico12.html> (accessed on 13 November 2012).
8. Kim, H.; Marwah, M.; Arlitt, M.; Lyon, G.; Han, J. Unsupervised disaggregation of low frequency power measurements. In Proceedings of the 2011 SIAM International Conference on data Mining SIAM, Mesa, AZ, USA, 28–30 April 2011; pp. 747–758.
9. Shaloudegi, K.; György, A.; Szepesvári, C.; Xu, W. SDP relaxation with randomized rounding for energy disaggregation. *arXiv* **2016**, arXiv:1610.09491.
10. Zhong, M.; Goddard, N.; Sutton, C. Signal aggregate constraints in additive factorial HMMs, with application to energy disaggregation. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 3590–3598.
11. Jiang, J.; Kong, Q.; Plumbley, M.D.; Gilbert, N.; Hoogendoorn, M.; Roijers, D.M. Deep Learning-Based Energy Disaggregation and On/Off Detection of Household Appliances. *Acm Trans. Knowl. Discov. Data TKDD* **2021**, *15*, 1–21. [CrossRef]
12. Kelly, J.; Knottenbelt, W. Neural nilm: Deep neural networks applied to energy disaggregation. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, Seoul, Korea, 4–5 November 2015; pp. 55–64.
13. Zhang, C.; Zhong, M.; Wang, Z.; Goddard, N.; Sutton, C. Sequence-to-point learning with neural networks for non-intrusive load monitoring. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
14. Chen, K.; Wang, Q.; He, Z.; Chen, K.; Hu, J.; He, J. Convolutional sequence to sequence non-intrusive load monitoring. *J. Eng.* **2018**, *2018*, 1860–1864. [CrossRef]
15. Sirojan, T.; Phung, B.T.; Ambikairajah, E. Deep neural network based energy disaggregation. In Proceedings of the 2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 12–15 August 2018; pp. 73–77.
16. Zhang, Y.; Yang, G.; Ma, S. Non-intrusive load monitoring based on convolutional neural network with differential input. *Procedia CIRP* **2019**, *83*, 670–674. [CrossRef]
17. Shin, C.; Joo, S.; Yim, J.; Lee, H.; Moon, T.; Rhee, W. Subtask gated networks for non-intrusive load monitoring. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 1150–1157.
18. Rafiq, H.; Zhang, H.; Li, H.; Ochani, M.K. Regularized LSTM based deep learning model: First step towards real-time non-intrusive load monitoring. In Proceedings of the 2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 12–15 August 2018; pp. 234–239.
19. Rafiq, H.; Shi, X.; Zhang, H.; Li, H.; Ochani, M.K.; Shah, A.A. Generalizability Improvement of Deep Learning-Based Non-Intrusive Load Monitoring System Using Data Augmentation. *IEEE Trans. Smart Grid* **2021**, *12*, 3265–3277. [CrossRef]
20. Linh, N.V.; Arboleya, P. Deep learning application to non-intrusive load monitoring. In Proceedings of the 2019 IEEE Milan PowerTech, Milan, Italy, 23–27 June 2019; pp. 1–5.
21. Herrero, J.R.; Murciego, A.L.; Barriuso, A.L.; de La Iglesia, D.H.; González, G.V.; Rodríguez, J.M.C.; Carreira, R. Non intrusive load monitoring (nilm): A state of the art. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*; Springer: Berlin, Germany, 2017; pp. 125–138.
22. Nalmpantis, C.; Vrakas, D. Machine learning approaches for non-intrusive load monitoring: From qualitative to quantitative comparison. *Artif. Intell. Rev.* **2019**, *52*, 217–243. [CrossRef]
23. Devlin, M.A.; Hayes, B.P. Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data. *IEEE Trans. Consum. Electron.* **2019**, *65*, 339–348. [CrossRef]
24. Athanasiadis, C.; Doukas, D.; Papadopoulos, T.; Chrysopoulos, A. A Scalable Real-Time Non-Intrusive Load Monitoring System for the Estimation of Household Appliance Power Consumption. *Energies* **2021**, *14*, 767. [CrossRef]
25. Gopinath, R.; Kumar, M.; Joshua, C.P.C.; Srinivas, K. Energy management using non-intrusive load monitoring techniques-State-of-the-art and future research directions. *Sustain. Cities Soc.* **2020**, 102411. [CrossRef]
26. Yang, D.; Gao, X.; Kong, L.; Pang, Y.; Zhou, B. An event-driven convolutional neural architecture for non-intrusive load monitoring of residential appliance. *IEEE Trans. Consum. Electron.* **2020**, *66*, 173–182. [CrossRef]
27. Song, J.; Wang, H.; Du, M.; Peng, L.; Zhang, S.; Xu, G. Non-Intrusive Load Identification Method Based on Improved Long Short Term Memory Network. *Energies* **2021**, *14*, 684. [CrossRef]

28. Rafiq, H.; Shi, X.; Zhang, H.; Li, H.; Ochani, M.K. A deep recurrent neural network for non-intrusive load monitoring based on multi-feature input space and post-processing. *Energies* **2020**, *13*, 2195. [\[CrossRef\]](#)
29. Çavdar, İ.H.; Faryad, V. New design of a supervised energy disaggregation model based on the deep neural network for a smart grid. *Energies* **2019**, *12*, 1217. [\[CrossRef\]](#)
30. D’Incecco, M.; Squartini, S.; Zhong, M. Transfer learning for non-intrusive load monitoring. *IEEE Trans. Smart Grid* **2019**, *11*, 1419–1429. [\[CrossRef\]](#)
31. Piccialli, V.; Sudoso, A.M. Improving non-intrusive load disaggregation through an attention-based deep neural network. *Energies* **2021**, *14*, 847. [\[CrossRef\]](#)
32. Völker, B.; Pfeifer, M.; Scholl, P.M.; Becker, B. A Framework to Generate and Label Datasets for Non-Intrusive Load Monitoring. *Energies* **2021**, *14*, 75. [\[CrossRef\]](#)
33. Kong, W.; Dong, Z.Y.; Wang, B.; Zhao, J.; Huang, J. A practical solution for non-intrusive type II load monitoring based on deep learning and post-processing. *IEEE Trans. Smart Grid* **2019**, *11*, 148–160. [\[CrossRef\]](#)
34. Faustine, A.; Pereira, L. Improved appliance classification in non-intrusive load monitoring using weighted recurrence graph and convolutional neural networks. *Energies* **2020**, *13*, 3374. [\[CrossRef\]](#)
35. Faustine, A.; Pereira, L. Multi-Label Learning for Appliance Recognition in NILM Using Fryze-Current Decomposition and Convolutional Neural Network. *Energies* **2020**, *13*, 4154. [\[CrossRef\]](#)
36. Lazzaretti, A.E.; Renaux, D.P.B.; Lima, C.R.E.; Mulinari, B.M.; Ancelmo, H.C.; Oroski, E.; Pöttker, F.; Linhares, R.R.; Nolasco, L.D.S.; Lima, L.T.; et al. A Multi-Agent NILM Architecture for Event Detection and Load Classification. *Energies* **2020**, *13*, 4396. [\[CrossRef\]](#)
37. Klemenjak, C.; Kovatsch, C.; Herold, M.; Elmenreich, W. A synthetic energy dataset for non-intrusive load monitoring in households. *Sci. Data* **2020**, *7*, 1–17. [\[CrossRef\]](#)
38. Tsai, C.H.; Bai, Y.W.; Lin, M.B.; Jhang, R.J.R.; Chung, C.Y. Reduce the standby power consumption of a microwave oven. *IEEE Trans. Consum. Electron.* **2013**, *59*, 54–61. [\[CrossRef\]](#)
39. Raj, P.A.D.V.; Sudhakaran, M.; Raj, P.P.D.A. Estimation of standby power consumption for typical appliances. *J. Eng. Sci. Technol. Rev.* **2009**, *2*, 71–75. [\[CrossRef\]](#)
40. Issi, F.; Kaplan, O. The determination of load profiles and power consumptions of home appliances. *Energies* **2018**, *11*, 607. [\[CrossRef\]](#)
41. Pipattanasomporn, M.; Kuzlu, M.; Rahman, S.; Teklu, Y. Load profiles of selected major household appliances and their demand response opportunities. *IEEE Trans. Smart Grid* **2013**, *5*, 742–750. [\[CrossRef\]](#)
42. Fung, A.S.; Aulenback, A.; Ferguson, A.; Ugursal, V.I. Standby power requirements of household appliances in Canada. *Energy Build.* **2003**, *35*, 217–228. [\[CrossRef\]](#)
43. Lee, S.; Ryu, G.; Chon, Y.; Ha, R.; Cha, H. Automatic standby power management using usage profiling and prediction. *IEEE Trans. Hum.-Mach. Syst.* **2013**, *43*, 535–546. [\[CrossRef\]](#)
44. Nguyen, G.H.; Bouzerdoun, A.; Phung, S.L. Learning pattern classification tasks with imbalanced data sets. In *Pattern Recognition*; Intechopen: London, UK, 2009; pp. 193–208.
45. Thabtah, F.; Hammoud, S.; Kamalov, F.; Gonsalves, A. Data imbalance in classification: Experimental evaluation. *Inf. Sci.* **2020**, *513*, 429–441. [\[CrossRef\]](#)
46. Farrand, T.; Mireshghallah, F.; Singh, S.; Trask, A. Neither private nor fair: Impact of data imbalance on utility and fairness in differential privacy. In Proceedings of the 2020 Workshop on Privacy-Preserving Machine Learning in Practice, Virtual Event, USA, 9 November 2020; pp. 15–19.
47. Mauch, L.; Yang, B. A new approach for supervised power disaggregation by using a deep recurrent LSTM network. In Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 14–16 December 2015; pp. 63–67.
48. Singh, S.; Majumdar, A. Non-intrusive load monitoring via multi-label sparse representation-based classification. *IEEE Trans. Smart Grid* **2019**, *11*, 1799–1801. [\[CrossRef\]](#)
49. Dash, S.; Sodhi, R.; Sodhi, B. An appliance load disaggregation scheme using automatic state detection enabled enhanced integer programming. *IEEE Trans. Ind. Inform.* **2020**, *17*, 1176–1185. [\[CrossRef\]](#)
50. Azizi, E.; Beheshti, M.T.; Bolouki, S. Event Matching Classification Method for Non-Intrusive Load Monitoring. *Sustainability* **2021**, *13*, 693. [\[CrossRef\]](#)
51. Houidi, S.; Fourer, D.; Auger, F. On the use of concentrated time–frequency representations as input to a deep convolutional neural network: Application to non intrusive load monitoring. *Entropy* **2020**, *22*, 911. [\[CrossRef\]](#) [\[PubMed\]](#)
52. De Baets, L.; Ruyssinck, J.; Develder, C.; Dhaene, T.; Deschrijver, D. Appliance classification using VI trajectories and convolutional neural networks. *Energy Build.* **2018**, *158*, 32–36. [\[CrossRef\]](#)
53. Mukaroh, A.; Le, T.T.H.; Kim, H. Background Load Denoising across Complex Load Based on Generative Adversarial Network to Enhance Load Identification. *Sensors* **2020**, *20*, 5674. [\[CrossRef\]](#)
54. Zhou, X.; Li, S.; Liu, C.; Zhu, H.; Dong, N.; Xiao, T. Non-Intrusive Load Monitoring Using a CNN-LSTM-RF Model Considering Label Correlation and Class-Imbalance. *IEEE Access* **2021**, *9*, 84306–84315. [\[CrossRef\]](#)
55. Bonfigli, R.; Principi, E.; Fagiani, M.; Severini, M.; Squartini, S.; Piazza, F. Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models. *Appl. Energy* **2017**, *208*, 1590–1607. [\[CrossRef\]](#)
56. Welikala, S.; Dinesh, C.; Ekanayake, M.P.B.; Godaliyadda, R.I.; Ekanayake, J. Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting. *IEEE Trans. Smart Grid* **2017**, *10*, 448–461. [\[CrossRef\]](#)