

Article

Multi-Label Learning for Appliance Recognition in NILM Using Fryze-Current Decomposition and Convolutional Neural Network

Anthony Faustine ^{1,*}  and Lucas Pereira ² 

¹ Ireland's National Centre for Applied Data Analytics (CeADER) University College Dublin; Belfield Office Park, Unit 9, Clonskeagh, 4 Dublin, Ireland

² ITI, LARSyS, Técnico Lisboa; Av. Rovisco Pais, 1000 268 Lisboa, Portugal; lucas.pereira@tecnico.ulisboa.pt

* Correspondence: sambaiga@gmail.com or anthony.faustine@ucd.ie; Tel: +32493972982

Received: 18 May 2020; Accepted: 6 July 2020; Published: 11 August 2020



Abstract: The advance in energy-sensing and smart-meter technologies have motivated the use of a Non-Intrusive Load Monitoring (NILM), a data-driven technique that recognizes active end-use appliances by analyzing the data streams coming from these devices. NILM offers an electricity consumption pattern of individual loads at consumer premises, which is crucial in the design of energy efficiency and energy demand management strategies in buildings. Appliance classification, also known as load identification is an essential sub-task for identifying the type and status of an unknown load from appliance features extracted from the aggregate power signal. Most of the existing work for appliance recognition in NILM uses a single-label learning strategy which, assumes only one appliance is active at a time. This assumption ignores the fact that multiple devices can be active simultaneously and requires a perfect event detector to recognize the appliance. In this paper proposes the Convolutional Neural Network (CNN)-based multi-label learning approach, which links multiple loads to an observed aggregate current signal. Our approach applies the Fryze power theory to decompose the current features into active and non-active components and use the Euclidean distance similarity function to transform the decomposed current into an image-like representation which, is used as input to the CNN. Experimental results suggest that the proposed approach is sufficient for recognizing multiple appliances from aggregated measurements.

Keywords: multi-label learning; Non-intrusive Load Monitoring; appliance recognition; fryze power theory; V-I trajectory; Convolutional Neural Network; distance similarity matrix; activation current

1. Introduction

Recently, most of the world has witnessed a rapid increase in energy use in buildings (residential and commercial). Residential and commercial buildings consume approximately 60% of the world's electricity (The United Nation's Environment Programme's Sustainable Building and Climate Initiative (UNEP-SBCI)). Energy efficiency and conservation in buildings can be generally achieved through replacing devices with more efficient ones, improving the efficiency of the building (for example, using better insulation), or optimizing energy usage through behavior changes and application of cost-effective technologies [1]. Unlike other strategies for building energy saving, optimizing energy use through behavior changes is very fast and highly profitable. With the application of cost-effective technologies, this strategy can also provide end-use appliances consumption to households that give insight into what appliances are used, when they are used, how much power they consume, and why such consumption [2]. The end-use appliances consumption is also useful for estimating the amount of energy demand at consumer premises [3]. It further increases the awareness about the energy consumption behavior of consumers.

The recent advance in energy-sensing and smart-meter technologies has led to the rise of Non-Intrusive Load Monitoring (NILM) [4,5]. NILM is a computational technique that uses aggregate power data monitored from a single point source such as a smart meter or current or voltage sensor-plug to infer the end-appliances running in the building and estimate their respective power consumption [6]. It relies on signal processing and machine learning techniques that analyze appliance patterns from aggregate power measurements. NILM provides households with cost-effective monitoring of appliance-specific energy consumption, and it can be easily integrated into existing buildings without causing any inconvenience to inhabitants. Several machine learning techniques have been proposed to address the energy-disaggregation [7–12].

Recognizing appliances from aggregated power measurements is one of the vital sub-tasks of NILM [11,13]. It uses machine learning techniques to analyze the pattern of the electrical features vector extracted from aggregated measurements and classifies them into the respective appliance category. Feature vectors are obtained after the state-transitions of appliances have been detected. These features are extracted at different sampling rates (high-frequency or low-frequency) depending on the measurements and electrical characteristics needed by the NILM algorithm [14]. The high sampling frequency offers the possibility to consider fine-grained features such as voltage-current (V-I) trajectory, harmonics, wavelet coefficients from steady-state, and transient behavior. As a result, several techniques for appliance recognition applying high-frequency features such as V-I trajectory have been proposed [15,16].

It has been demonstrated that transforming the V-I trajectory into image representation and feeding it as the input to machine learning classifiers improves classification performance [11,14,17–21]. However, the presented works use single-label learning, thus assuming that only one appliance is active at a time. This strategy ignores the fact that multiple devices can be active simultaneously as well as dependencies between appliance usage. It further requires a perfect event detector to extract the appliance features just after an event has been detected, particularly in aggregated measurements [19]. In contrast to single-label learning, multi-label learning links multiple appliances to an observed aggregate power signal [22,23].

Several studies have demonstrated that multi-label learning represents a viable alternative to conventional NILM approaches [23–27]. For example, the work by [25], investigated the possibility of applying a temporal multi-label classification approach in non-event based NILM where a novel set of meta-features was proposed. In [26] an extensive survey for the multi-label classification and the multi-label meta-classification framework for low-sampling power measurements is presented. Recent studies have explored deep neural networks for multi-label appliance recognition in NILM [23,24], yet these approaches also rely on low-frequency data.

Instead, this paper presents a CNN-based multi-label appliance classification approach. The proposed method uses the current waveform generated from the aggregated measurements, taken in brief windows of time containing one or more than one event. The underlying assumption of this method is that the extracted aggregated current will be the summation of active appliances. Therefore, by training a classification model to learn the patterns of different combinations, it is possible to successfully identify such appliances when these appear in a future time window.

To improve the discriminating power of our method, we apply the *Fryze power theory*, which enables the decomposition of the current waveform into active and non-active components in time-domain [20,28]. Our research hypothesis is that the sum of the active and non-active components will exhibit unique and consistent characteristics based on the appliances that are running simultaneously, hence providing a distinctive feature for multi-label classification.

The decomposed current is then transformed into an image representation using the Euclidean-distance-similarity matrix [29] and fed into the CNN for multi-label classification. The proposed approach is evaluated against the PLAID dataset [30], which consists of aggregated voltage and current measurements at a 30 kHz sampling rate. The source code used in our experiments is publicly available on a GitHub repository (<https://github.com/sambaiga/MLCFCD>).

The main contribution of this paper is a multi-label learning strategy for appliance recognition in NILM. The proposed approach associates multiple appliances to an observed aggregated current signal. Overall, this contribution folds into four sub-contributions

1. We first demonstrate that for aggregated measurements, the use of activation current as an input feature offers improved performance compared to the regularly used V-I binary image feature.
2. Second, we apply the Fryze power theory and Euclidean distance matrix as pre-processing steps for the multi-label classifier. This pre-processing step improves the appliance feature's uniqueness and enhances the performance of the multi-label classifier.
3. Third, we propose a CNN multi-label classifier that uses *softmax* activation to capture the relations between multiple appliances implicitly.
4. Fourth, we conduct an experimental evaluation of the proposed approach on an aggregated public dataset and compare the general and per-appliance performances. We also provide an in-depth error analysis and identified three types of errors for multi-label appliance recognition in NILM. Finally, a complexity analysis of the proposed approach method is also presented.

The remainder of this paper is organized as follows: Section 2 summarizes related works while Section 3 introduces the methods utilized in this work. Section 4 describes the experimental design. Section 5 presents the results and discussion of the performed evaluations. Finally, Section 6 summarises the contributions of this paper and suggests future research direction.

2. Related Works

The concept of multi-label classification for NILM has gained momentum recently, as the systematic review finds in [26]. Besides an extensive survey of the topic, the authors present the multi-label meta-classification framework (RAkEL) and the bespoke multi-label classification algorithm (MLkNN), where both employ time-domain and wavelet-domain feature sets. Other approaches to multi-label NILM comprise restricted Boltzmann machines [31], and multi-target classification [32].

In [33], the authors present an algorithm that uses Sparse Representation based classification for multi-label NILM. Furthermore, the authors compare their algorithm to other cutting edge multi-label NILM approaches such as classification based on extreme learning machines (ELM) [34], graph-based semi-supervised learning [35], and an approach based on deep dictionary learning and deep transform learning [36]. Nalmpantis and Vrakas [37] present a multi-label NILM based on the Signal2Vec algorithm that maps any time series into a vector space. A deep neural network (DNN) based multi-label NILM applying active power features at low-sampling frequency is proposed in [23,24]. In [23], the authors propose an approach that builds on Temporal Convolutional Networks (TCNN). At the same time, Massidda et al. [24] applied Fully Convolutional Networks (FCNN) for multi-label-learning in NILM, adopting some methods used in semantic segmentation.

Even though multi-label learning was found to be competitive with state-of-the-arts NILM algorithms, none of the previous works have considered the V-I trajectory-based features for multi-label-classification. Existing NILM methods that use V-I based features for appliance classification uses single-label learning [11,14–21,38]. The use of V-I based features for appliance classification was first introduced in [15], where shape-based features extracted from V-I (e.g., number of self-interceptions) were used as input to a machine learning classifier. A review and performance evaluation of the seven load wave-shape is presented in [39]. The shape-based feature was found to have a direct correspondence to operating characteristics of appliances as contained in the current wave-shape. Several other features such as asymmetry, mean line and self-intersection assessment extracted from V-I waveforms were used to classify appliances in [16]. However, this approach compresses the information in the V-I-trajectory into a limited amount of features extracted solely based on deep engineering knowledge.

Against this background, other researchers demonstrated that transforming the V-I trajectory into a binary image representation improves classification performance [17,18] by leveraging on

state-of-the-art deep-learning algorithms for image recognition. For example, De Baets et al. [14,19] transforms the V-I trajectory into weighted pixelated V-I images and uses a CNN classifier. In another work, a hardware implementation of the appliance recognition system based on V-I curves and a CNN classifier is also proposed [21]. The work by [20,28] demonstrated that applying the Fryze power theory to decompose the current into active and non-active components could enhance the uniqueness of the V-I binary image and consequently improve classification performance. The work by Teshome et al. [28] applied the non-active current and non-active voltage ($V-I_f$) for appliance recognition. Liu et al. [20] further demonstrated that the visual representation of ($V-I_f$) is robust enough to be used in Transfer learning. Recently it has been shown that transforming the V-I into compressed distance similarity matrix consistently improves the appliance classification performance compared to the commonly used V-I image representation [11,13].

Motivated by these two works, we apply decomposed currents as input features for recognizing multiple running appliances. Still, unlike Liu et al. [20], and Teshome et al. [28], we transform the decomposed current into a 2D Euclidean-distance similarity matrix, which is later used as the input to the CNN model.

3. Proposed Methods

The goal of appliance recognition in NILM is to identify appliance states s_t^m from the aggregate measurements x_t composed of individual appliance measurements y_t^m such that $m = 1, 2, \dots, M$, where M indicates the number of appliances such that

$$x_t = \sum_{m=1}^M y_t^m \cdot s_t^m + \epsilon_t \quad (1)$$

where ϵ_t represents both any contribution from appliances not accounted for and measurement noise [40]. We refer to this as a multi-label NILM problem, where given observed aggregate measurements, the unobserved appliance states s_t of electrical appliances are estimated.

Specifically, the problem is formulated as follows: Let $\mathbf{X} \in \mathbb{R}^{T \times d} = \{x_1 \dots x_T\}$ denote a set of input features derived from the aggregate measurements of M appliances and $\mathbf{Y} \in \mathbb{R}^{T \times M}$ indicates the associated appliance measurements, where each appliance has k states denoted as $\mathbf{s}^m = \{s_1^m, \dots, s_k^m\}$ such that $s_k^m \in \{0, 1\}$. The matrix $\mathbf{S} \in \mathbb{R}^{T \times M}$ indicates the associated multi-label states. Thus, given $D = \{\mathbf{x}_t, \mathbf{s}_t | t = 1, \dots, T\}$ datasets, the goal is to learn a multi-label classifier that predicts the state vector $\mathbf{s}_t = \{s_t^1, \dots, s_t^M\}$ from the input aggregate power feature vector \mathbf{x}_t .

The proposed approach is summarized in Figure 1.

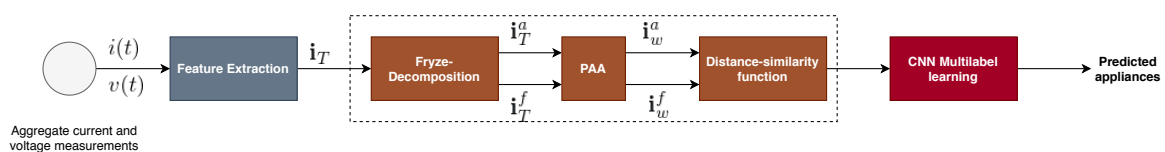


Figure 1. The block diagram of the proposed method. The dotted block is the pre-processing block where PAA stand for Piecewise Aggregate Approximation, a dimension reduction method for high-dimensional time series signal.

3.1. Feature Extraction from Aggregate Measurements

In this work, we consider the appliance feature extracted from high-frequency aggregate voltage and current measurements in brief windows of time. This feature contains one or more than one event and allows us to distinguish multiple appliances running simultaneously, as illustrated in Figure 2a. We define an activation current \mathbf{i} and voltage \mathbf{v} to be a one-cycle steady-state signal extracted from the aggregate current and voltage waveform.

To obtain the activation waveforms from aggregate measurements we measure $N_c = 20$ cycles of current and voltage $\{\mathbf{i}^{(a)}, \mathbf{v}^{(a)}\}$ after an event. As depicted in Figure 2b, the N_c circles correspond to steady-state behavior and is equivalent to $T_s \times N_s$ samples where $T_s = \frac{f_s}{f}$, f_s is sampling frequency and f is mains frequency. These cycles are aligned at the zero-crossing of the voltage and there-after one-cycle activation current \mathbf{i} and voltage \mathbf{v} with size T_s is extracted, as illustrated in Figure 2.

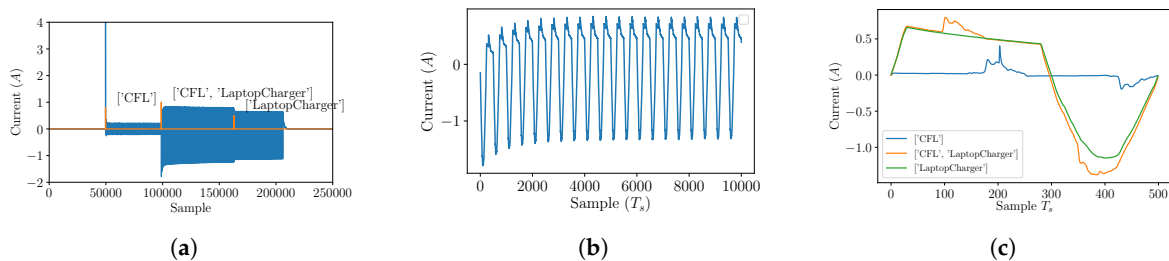


Figure 2. (a) Aggregated current signal for different events. (b) Twenty cycles of current after an event. (c) Extracted activation currents for different events. The activation current of the event at time \mathbf{i} is the summation of the activation current for the all running appliances.

3.2. Feature Pre-Processing

As discussed in the related work section, the V-I binary trajectory mapping has been one of the favored features for appliance classification in *single-label learning*. However, in this work, we consider the features derived from source current $i(t)$ in recognizing multiple running appliances from aggregate measurements. Through experimentation, it was found that the aggregated activation voltage $v(t)$ has an almost identical pattern for most of the events, as illustrated in Figure 3. This suggests that the activation current i reflects the electrical properties of an appliance.

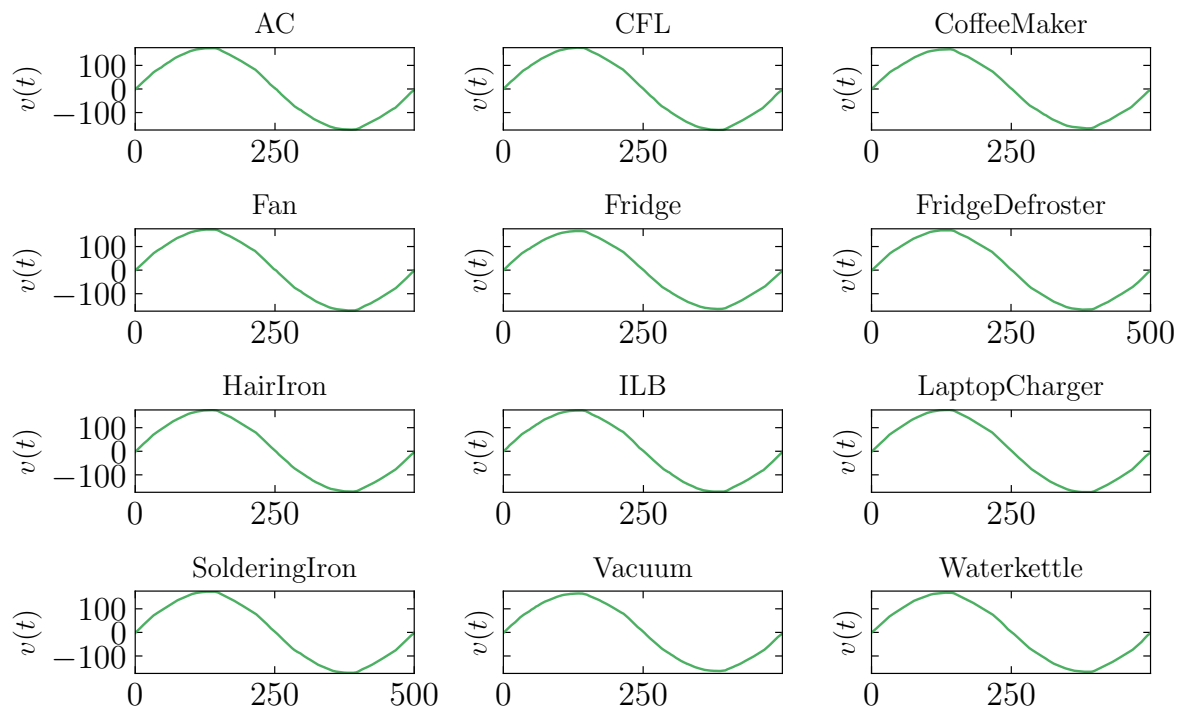


Figure 3. Activation voltage $v(t)$ for different appliances in the PLAID dataset. The voltage has an almost identical pattern for all the appliances.

Therefore, we propose the decomposed current features obtained by applying the Fryze power theory [41].

The Fryze power theory decomposes activation current into orthogonal components related to electrical energy in time-domain [41]. According to this theory, it is possible to decompose the activation current \mathbf{i} into active $i(t)_a$ and non-active components $i(t)_f$, such that:

$$i(t) = i(t)_a + i(t)_f \quad (2)$$

The active current $i(t)_a$ is the current of the resistive load, having the same active power at the same activation voltage. In Fryze's theory, the active power is calculated as the average value of $i(t) \cdot v(t)$ over one fundamental cycle T_s defined as follows:

$$p_a = \frac{1}{T_s} \sum_{t=1}^{T_s} i(t)v(t) \quad (3)$$

The active current is therefore defined as

$$i_a(t) = \frac{p_a}{v_{rms}^2} v(t) \quad (4)$$

where v_{rms} is the rms voltage, expressed as follows:

$$v_{rms} = \sqrt{\frac{1}{T_s} \sum_{t=1}^{T_s} v(t)^2} \quad (5)$$

The current $i_a(t)$ represents the resistance information and is purely sinusoidal. The non-active component is then equal to

$$i(t)_f = i(t) - i(t)_a \quad (6)$$

Figure 4 presents the source currents and the corresponding active and non-active components for the twelve appliances in the PLAID dataset. It can be observed from Figure 4 that the active component approaches a pure sine wave even for non-periodic load currents like a Compact Fluorescent Lamp (CFL) and Laptop.

Once the activation-current has been decomposed, the Piece-wise Aggregate Approximation (PAA) is used to reduce the dimensional of the decomposed signal \mathbf{i}_a and \mathbf{i}_f from T_s to a predefined size w . PAA is a dimension reduction method for high-dimensional time series signal [42]. This is a crucial pre-processing step as it reduces the high-dimensionality of the extracted activation current feature with minimal information loss.

To further enhance the uniqueness of the decomposed-current feature, a Euclidean distance function $d_{u,v} = \|i(t)_u - i(t)_v\|_2$ that measures how similar or related two data points are is applied on the active and non-active current. The distance similarity function is widely used as a pre-processing step for many of the machine learning approaches such as K-means clustering and K-nearest neighbor algorithms [29,43]. The distance similarity matrix $D_{w,w}$ for points $i(t)_1, i(t)_2, \dots, i(t)_w$ is the a matrix of squared euclidean distances representing the spacing of a set of w points in euclidean space [29] such that

$$D_{w,w} = \begin{bmatrix} 0 & d_{1,2} & \cdots & d_{1,w} \\ d_{2,1} & 0 & \cdots & d_{2,w} \\ \vdots & \vdots & \ddots & \vdots \\ d_{w,1} & d_{w,2} & \cdots & 0 \end{bmatrix} \quad (7)$$

Figure 5 depicts the activation current, its components and their corresponding distance similarity matrix when a CFL and a laptop charger are active.

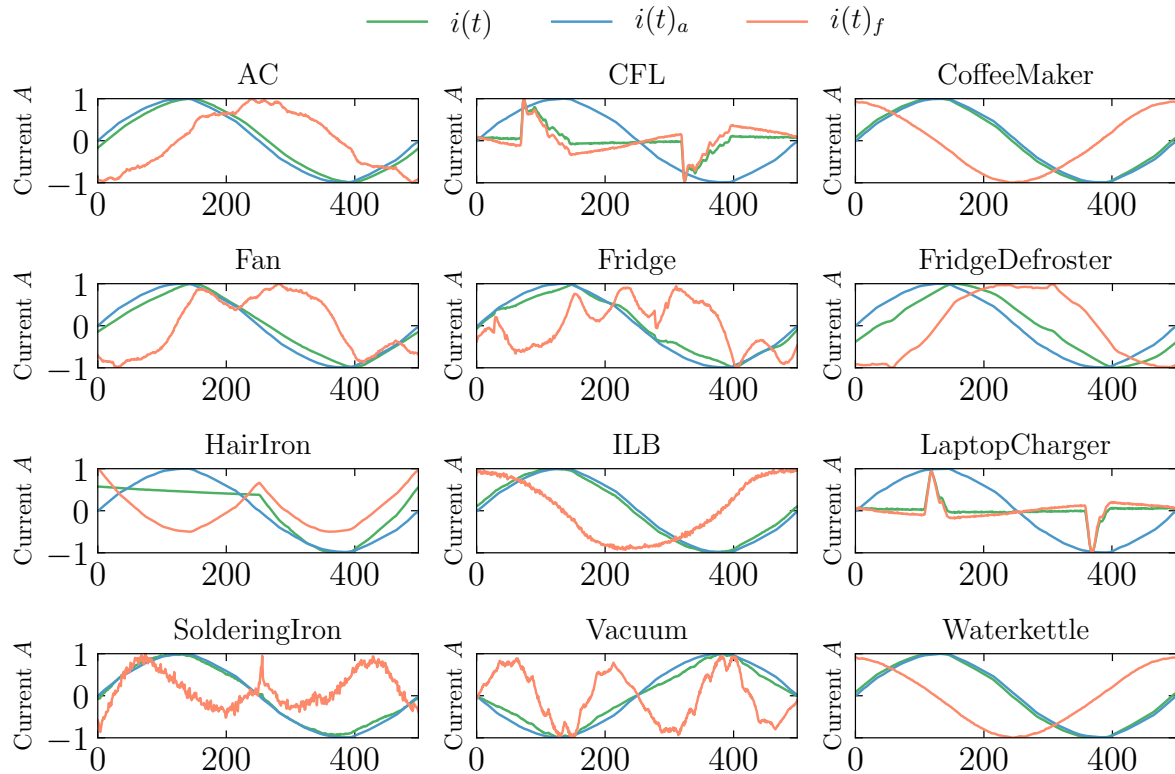


Figure 4. Normalized source current $i(t)$ and their respective active $i(t)_a$ and reactive components $i(t)_f$ after applying Fryze power theory. The current is normalized for visualization purposes.

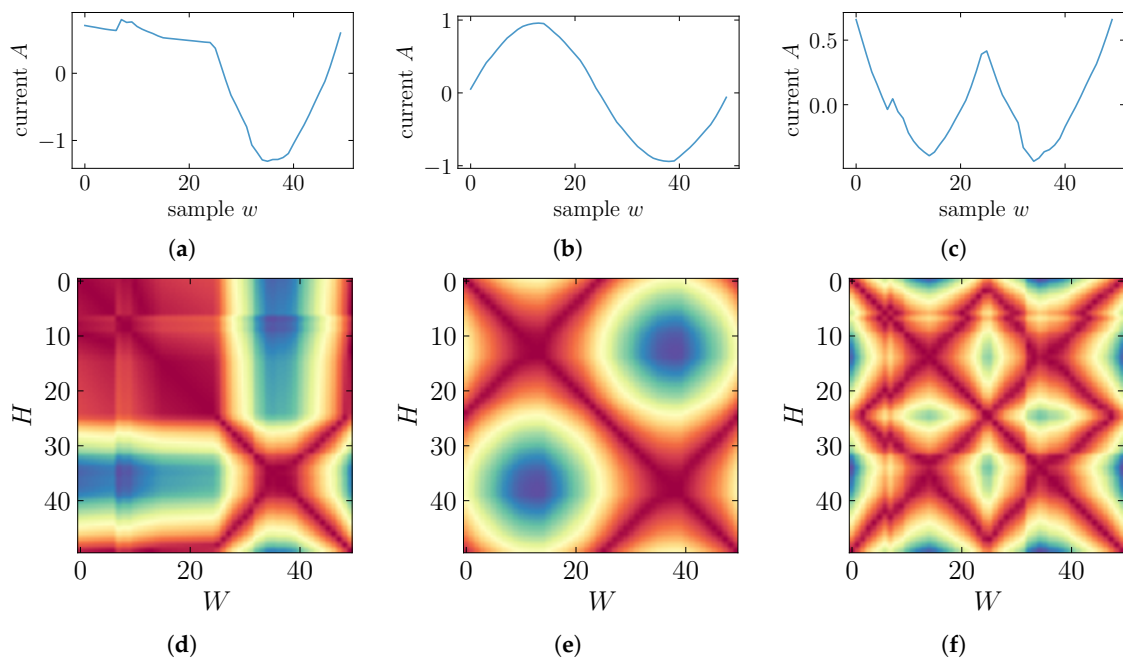


Figure 5. Currents and distance matrix when Compact Fluorescent Lamp (CFL) and laptop charger are active. (a) Source current $i(t)$. (b) Active current $i_a(t)$. (c) Non-active current $i_f(t)$. (d) Distance matrix for source current. (e) Distance matrix for active current (f) Distance matrix for non-active current.

3.3. Multi-Label Modeling

A common approach that extends neural networks to multi-label classification is to use one neural network to learn the joint probability of multiple labels conditioned to the input features representation. The final predicted multi-label is obtained by applying a *sigmoid* activation function [23]. This process

requires an additional thresholding mechanism to transform the sigmoid probabilities to multi-label outputs. However, building such a threshold function is very challenging. Therefore a default threshold of 0.5 is often employed [44].

To address this challenge, we propose a CNN multi-label classifier that uses softmax to implicitly capture the relations between multiple labels. As shown in Figure 6, the proposed CNN multi-label classifier consists of a four-stage CNN layer each with 16, 32, 64, and 128 feature maps, 2×2 strides.

The first two CNN layers use a 5×5 filter size, while the last two layers use a 3×3 filter size. The four CNN layers are followed by a batch normalization layer and the ReLU activation function. The last CNN layer is followed by an adaptive average pooling layer with an output size of 1×1 . The CNN layer takes current-based features as inputs and produces a latent-feature vector \mathbf{z}_i .

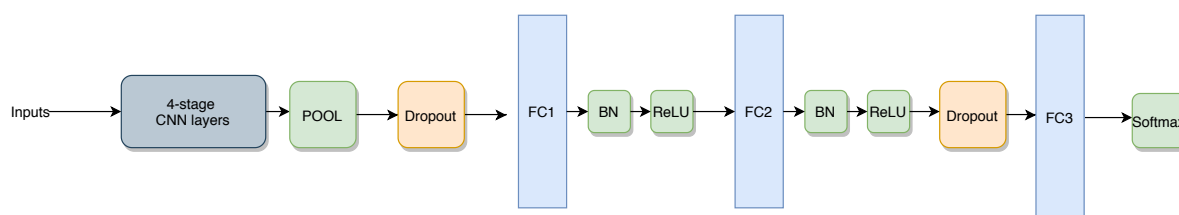


Figure 6. Block diagram of the Convolutional Neural Network (CNN) multi-label classifier. It consists of a CNN encoder to learn feature representation from the input feature, and the output layer to produce the predicted labels.

The output layer consists of three FC layers with a hidden size of 502, 1024, and $2M$, respectively. The last layer is followed by an adaptive average pooling layer and three linear layers with a hidden size of 5012, 1024, and M , respectively. M is the maximum number of appliances available. This layer receives the output of the CNN layer to produce an output \mathbf{O}_s of size $(2 \times M)$. The final predicted multi-label states, $\hat{\mathbf{s}}_t$, is obtained by applying the softmax activation function $\hat{\mathbf{s}}_t = \text{softmax}(\mathbf{O}_s)$. Thus, the proposed multi-label classifier learns the joint representation of multiple appliances states conditioned on activation-based input features.

To learn the model parameters, a standard back propagation is used to optimize the cross-entropy between the predicted softmax distribution and the multi-label target of each input feature.

$$\mathcal{L}(\hat{\mathbf{s}}, \mathbf{s}) = -\frac{1}{N} \sum_{t=1}^N \sum_{i=1}^M s_{ti} \cdot \log \frac{\exp(\hat{s}_{ti})}{\sum_j^2 \exp(\hat{s}_{tj})} \quad (8)$$

The joint cross-entropy loss implicitly captures the relations between labels.

The CNN multi-label classifier is trained for 500 iterations using the Adam optimizer with an initial learning rate of 0.001, betas of (0.9, 0.98), and a batch size of 16. A factor of 0.1 reduces the learning rate once the learning stagnates for 20 consecutive iterations. To avoid over-fitting, early stopping with patience is used where the training model terminates once the validation performance does not change after 50 iterations. The dropout is set to 0.25.

4. Evaluation Methodology

4.1. Dataset

The proposed method is evaluated on the PLAID dataset [30] that contains aggregate voltage and current measurements. The PLAID aggregated measurement data include measurements of more than one concurrently running appliances sampled at 30 kHz. It includes 1478 aggregated activations and deactivations for 12 different appliances measured at one location. Since we are interested in recognizing multiple active appliances, we only select activations and deactivations with at least one running appliance in the background resulting in 1154 samples. The distribution of the number of active appliances and appliances on the extracted 1154 activations is depicted in Figure 7.

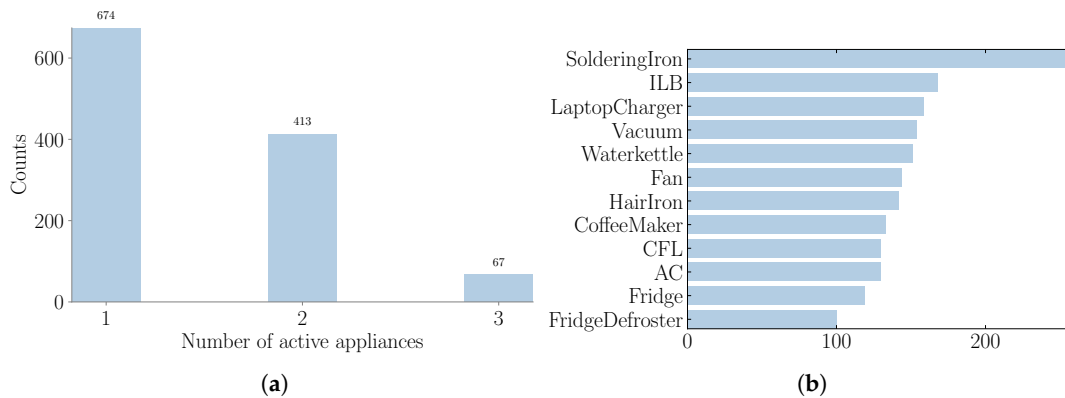


Figure 7. (a) Active appliances distributions. (b) Appliances distribution on the extracted 1154 activations. The soldering iron has large number of activations because it has two start-up events.

4.2. Performance Metrics

We quantitatively evaluate the classification performance with label-based and instance-based metrics. Label-based metrics work by evaluating each label separately and returning the average (micro or macro) value across all appliances. In contrast, the instance-based metrics evaluate bi-partition over all instances. To this end, two metrics, namely example-based F_1 (F_1 -eb) and macro-averaged F_1 (F_1 -macro) measures are used. Example-based F_1 (F_1 -eb) is an instance-based metric that measures the ratio of correctly predicted labels to the sum of the total true and predicted labels such that:

$$F_1\text{-eb} = \frac{\sum_{i=1}^M 2 \cdot t_p}{\sum_{i=1}^M y_i + \sum_{i=1}^M \hat{y}_i} \quad (9)$$

The F_1 -macro is derived from F_1 score and measures the label-based F_1 score averaged over all labels and is defined as:

$$F_1\text{-macro} = \frac{1}{M} \sum_{i=1}^M \frac{2 \cdot t_{pi}}{2 \cdot t_{pi} + f_{pi} + f_{pi}} \quad (10)$$

where t_p is true positive, f_p is false positive and f_n is false negative. High F_1 -ma usually indicate high performance on less frequent labels [45].

4.3. Experiment Description

To benchmark our approach, we adopt multi-label stratified 10-fold cross-validation with random shuffle [46]. This evaluation approach provides stratified randomized folds for multi-label while preserving the label's percentage in each fold. We compare the performance of the proposed CNN model against the commonly used multi-label k-nearest-neighbor (MLkNN) [47] and Binary relevance k-nearest-neighbor (BRkNN) [48] model.

To evaluate the proposed activation current feature, we first establish a baseline in which the V-I binary image is used as an appliance feature. The VI binary image of size $w \times w$, is obtained by meshing the $V - I$ trajectory and assigning a binary value that denotes whether the trajectory traverses it as described in [14]. This experiment setup helps us to answer an essential question on whether the proposed approach is sufficient for recognizing multiple appliances from aggregated measurements. We analyze this by altering the type input features and compare the obtained performance. To gain more insight into the proposed approach, we further examine the individual appliance performance and misclassification errors.

To analyze the computational complexity of the proposed approach, we also assess the training and inference times as a function of the number of data samples. This was achieved by training the MLkNN baseline and CNN-based multi-label classifier while varying the training and testing size.

In each run, the model is trained on p samples of data for 100 iterations and tested on $(1 - p)$ samples data where $p \in [0.1, 0.9]$.

Finally, we compare the appliance classification results with related state-of-art methods. However, we should emphasize that due to the difficulty in producing fair comparisons as a result of different experimental settings (e.g., sampling frequency, measurements, learning strategy, dataset, and the metrics) these comparisons are merely illustrative of the potential of the proposed method.

5. Results and Discussion

5.1. Comparison with Baseline

The results of the comparisons between the baselines and the proposed CNN multi-label learning for the V-I binary image and activation current feature are depicted in Figure 8a.

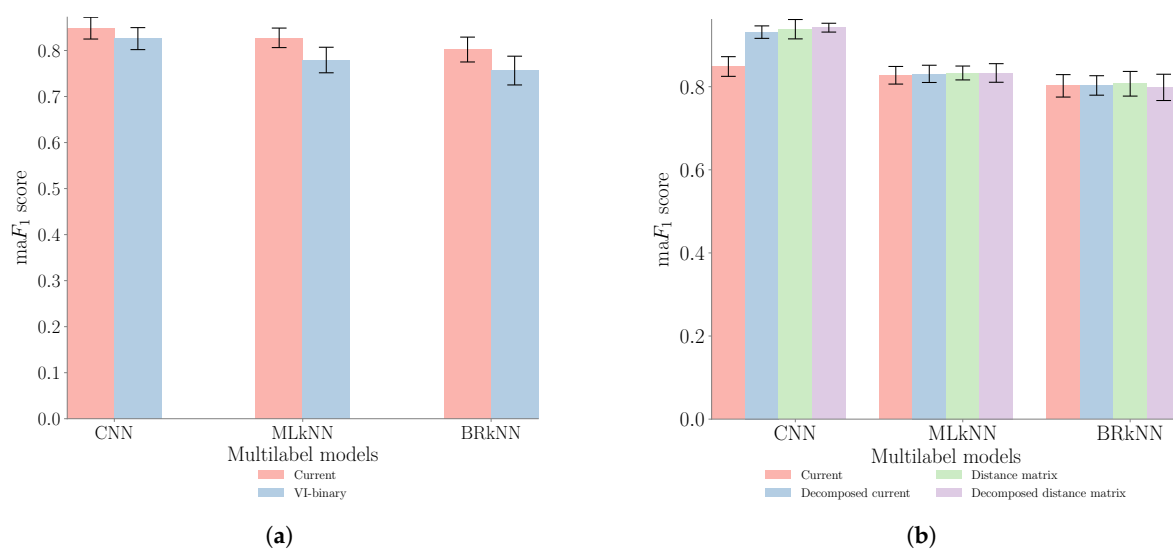


Figure 8. maF_1 score performance comparison between the proposed CNN model and the two baselines for different inputs features: (a) Comparison between voltage-current (V-I) binary image and current activation features; (b) Comparison between the activation current based features.

From Figure 8a, we see that the proposed CNN multi-label learning performs better than the baselines in both feature types. We also observe that compared to the current activation feature, the V-I binary feature representation yields low maF_1 scores in both the proposed CNN model and the two baseline algorithms. We see a slight increase in maF_1 score (from 0.826 ± 0.024 to 0.849 ± 0.024 for the CNN model and from 0.779 ± 0.028 to 0.827 ± 0.021 for the MLkNN model) when activation current is used as input features. This result suggests that features derived from activation current could be useful in recognizing appliances from total measurements.

We, therefore, analyzed three additional features derived from the activation current, namely decomposed current, current distance similarity matrix, and decomposed distance similarities. The results are presented in Figure 8b.

As it can be observed, the three current-based features significantly improve the classification performance in the CNN model, while achieving nearly the same performance on the two baselines. For the CNN model, the decomposed current feature attains an average 9.4% maF_1 score (from 0.849 ± 0.024 to 0.931 ± 0.015) increase over the activation current feature. This result is in line with the one obtained in [20], which suggested that decomposing the activation current into its active components enhances the uniqueness of the V-I trajectory. We also see about 10 percentage points increase in maF_1 (from 0.849 ± 0.024 to 0.94 ± 0.015) for the decomposed distance similarities. The decomposed current and current distance matrix features achieve comparable performance.

This result indicates that the decomposed current features could help increase the performance of appliance recognition in NILM.

Figure 9 presents the predicted multiple appliances from the CNN based classifier with different feature representation. We see that compared to the activation current in Figure 9a and the V-I image Figure 9c, the proposed Fryze's current-decomposition in Figure 9b,d is capable of detecting all multiple running appliances. This shows that the Fryze current decomposition-based feature alone is sufficient for the identification of multiple running appliances.

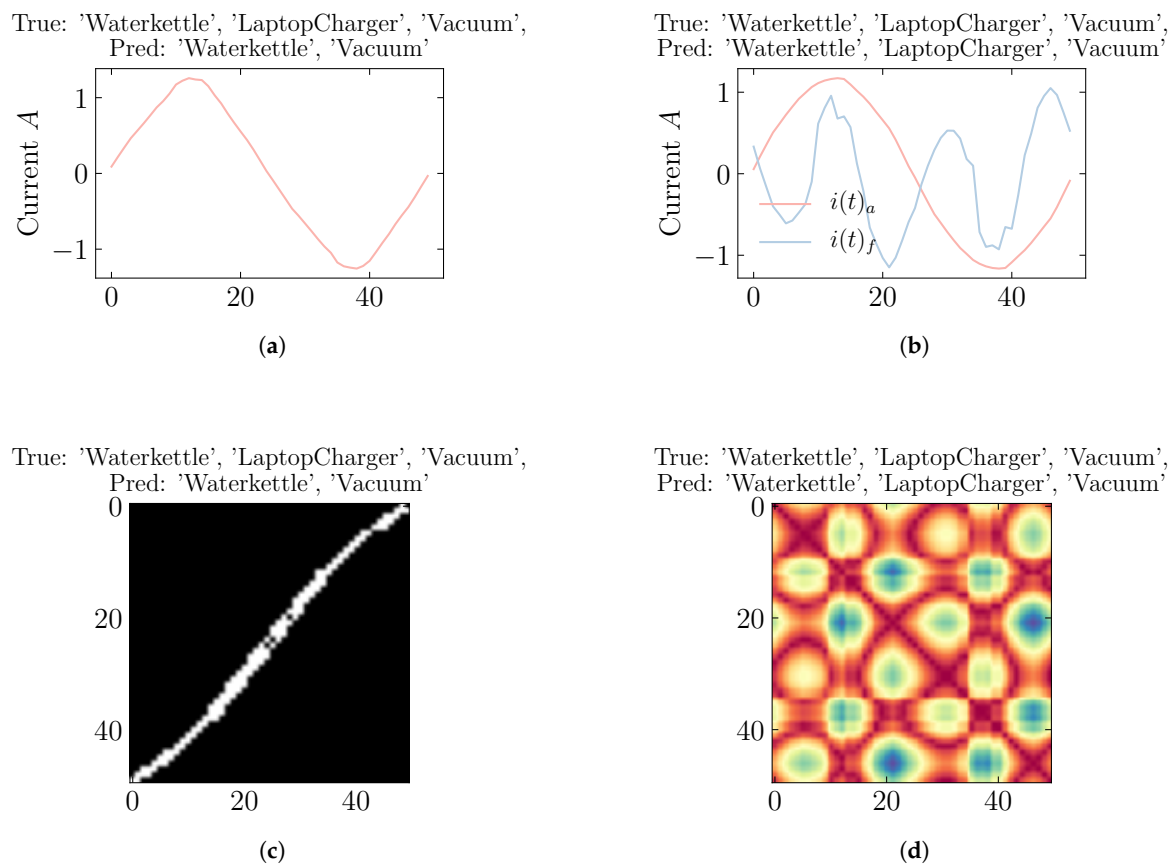


Figure 9. Prediction comparison for different feature representations with the proposed CNN multilabel classifier. (a) Action current. (b) Decomposed current. (c) V-I image. (d) Distance matrix.

To gain insights on the performance of individual appliances, we further analyze the per-appliance ebF_1 score for the MLkNN and the proposed CNN multi-label classifier, as depicted in Figure 10. The CNN model with decomposed current distance matrix feature obtains over 90% ebF_1 score for each appliance except for AC, ILB, and LaptopCharger. We also see that the MLkNN baseline with the same decomposed current distance matrix feature obtains over 90% ebF_1 score for only four appliances, namely FridgeDefroster, CoffeeMaker, Vacuum, and CFL. In both cases, we observe low scores for V-I binary features except for FridgeDefroster, CoffeeMaker, and Vacuum, which score above 90% ebF_1 .

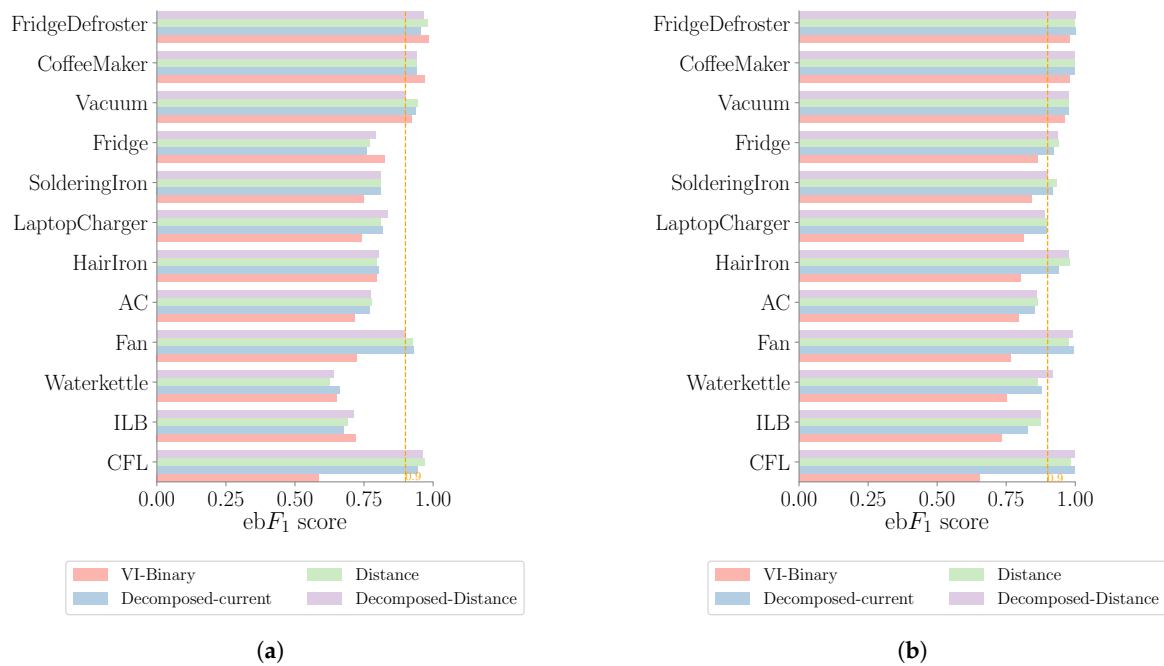


Figure 10. Per-appliance ebF_1 score on PLAID dataset. AC = air conditioning, CFL = compact fluorescent lamp, ILB = incandescent light bulb. (a) Multi-label k-nearest-neighbor (MLkNN) (b) CNN.

5.2. Error Analysis

We also analyze the miss-classification errors for the proposed CNN model. To this end, we identified three types of errors, namely *zero-error*, *one-to-one*, and *many-to-many* errors.

The *zero-type* mistake happens when a model predicts no appliance is running while there is at least one active appliance. It can be observed from Figure 11 that the number of zero-type mistakes is very low for the three feature types with decomposed-current distance making no such type of error.

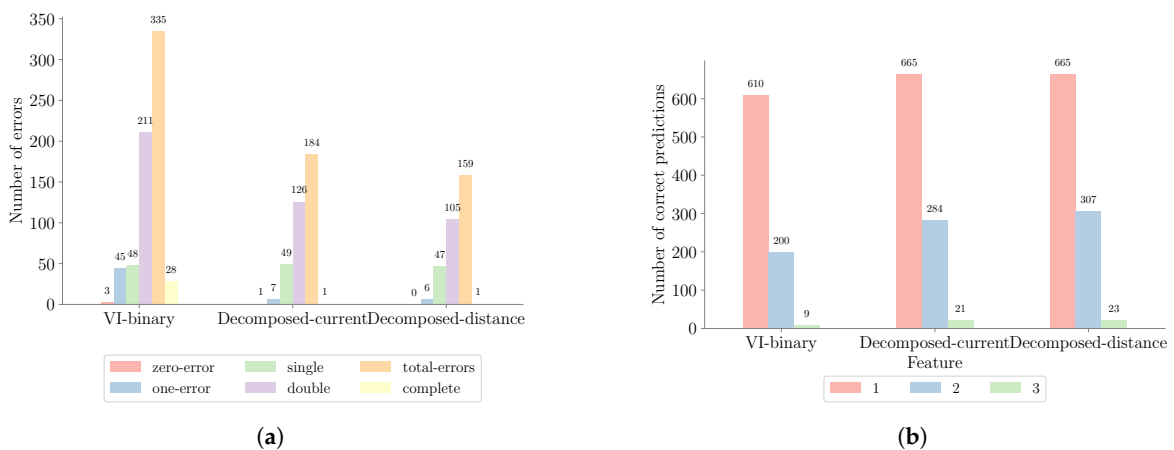


Figure 11. (a) Distributions of type errors the model makes. (b) Number of correct predictions for single, double and triple activations.

On the other hand, the *one-to-one* is the type of error that the model makes when there is only one active appliance running. We see from Figure 11 that the V-I binary image makes 45 one-type errors while the current based features reduce this to seven for the decomposed current, and six for the decomposed current distance feature. The low error rate when one appliance is running can be attributed to the high number of single activations, over 50%, as presented in Figure 7a. It further shows

the effectiveness of the proposed CNN multi-label learning in recognizing individually operating appliances, with over 98% accuracy, as shown in Figure 11b.

The *many-to-many errors* are confusions that a model makes when several appliances are active. Since the PLAID dataset used in our experiment consists of up to three simultaneous active appliances, we further categorized *many-to-many errors* into *single*, *double*, or *complete-error*. A *single error* occurs when a model confuses only one appliance when two or three appliances are active, whereas in *double fault*, the model confuses two appliances when three appliances are active. The *complete-error* is the case when the model produces incorrect predictions for all the active appliances. It can be inferred from Figure 11 that the proposed CNN multi-label model makes a higher number of *double errors* for the three input feature types used. This is likely to be caused by the fewer numbers of samples with more than two appliances running simultaneously at about 5.8%, as depicted in Figure 7a.

5.3. Complexity Analysis

The results for the complexity analysis between the baseline and the proposed CNN multi-label learning are presented in Figure 12. As expected, since the proposed method is an eager learner (i.e., a model is created in the training phase), it takes significantly longer to train than the MLkNN baseline (Figure 12a). In contrast, the proposed method has a much shorter inference time since the model was already created in the training phase. Furthermore, from Figure 12b it can be observed that the proposed method achieves better performance even with less training data, which is positive if one considers that labeled data is scarce and often hard to acquire.

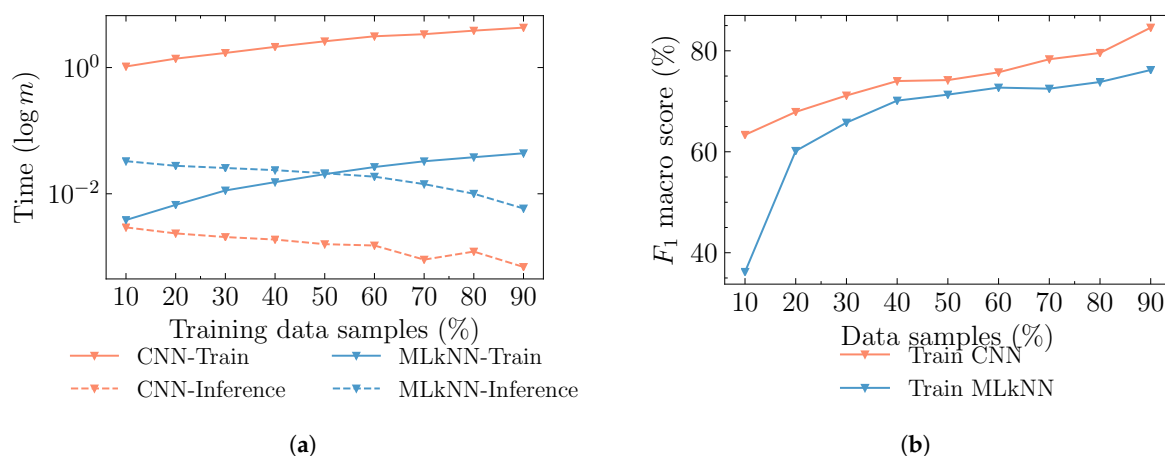


Figure 12. (a) Distributions of type errors the model makes. (b) Number of correct predictions for single, double and triple activations.

5.4. Comparison with State-of-the-Art Methods

Table 1 provides an overview of the results obtained in other related works. As it can be observed, there are many differences that make a fair and objective comparison impossible to achieve. For instance, while our approach uses current waveforms extracted from high-frequency power measurements, the results presented in [26] were obtained on low-frequency data, and on a different dataset. Yet, they also used the MLkNN multi-label classifier, achieving considerably lower results. Moreover, our results cannot be directly compared with the ones presented in [49], as these were obtained from a private dataset, besides the very different experimental settings including a different performance metric. In [23,37], the F_1 macro score for TCNN and FCNN DNN based multi-label classifiers are given; however, they use UK-DALE dataset making the comparison irrelevant.

An almost direct comparison is only possible between our method and the results from [19] who have used the same dataset and performance metric. Still, it should be stressed that the performance

evaluation method was different since their work targets single-label classification. Yet, the results obtained with our approach are superior by six percentage points.

In short, for a fair comparison, we would have to re-implement all these approaches, which unfortunately is not always possible. Nevertheless, to make this task easier for other authors, we open-sourced the code necessary to replicate our experiments.

Table 1. Results comparison.

Approach	Learning Strategy	Model	Dataset	Sampling Frequency	Results (Metric)
De Baets et al. [19]	single	CNN	PLAID [30]	High	88.0% (F_1 macro)
Faustine et al. [13]	single	CNN	PLAID [30]	High	97.77% (F_1 macro)
Tabatabaei et al. [26]	multi	MLkNN	REDD-House1 [50]	Low	61.90% (F_1 macro)
Lai et al. [49]	multi	SVM/GMM	Private	-	90.72% (Accuracy)
Yang et al. [23]	multi	FCNN	UK-DALE-house 1 [51]	Low	93.8% (F_1 score)
Nalmpantis and Vrakas [37]	multi	TCNN	UK-DALE-house 1 [51]	Low	92.5% (F_1 score)
Proposed approach	multi	CNN	PLAID [30]	High	94.0% (F_1 score)

6. Conclusions and Future Work Directions

In this work, we have approached appliance recognition in NILM as a multi-label learning problem which links multiple appliances to an observed aggregate current signal. We first show that features derived from activation current alone could be useful in recognizing devices from total measurements. We later apply Fryze’s power theory, which decomposes the current waveform into active and non-active components. The decomposed current signal was then transformed into an image-like representation using the Euclidean-distance-similarity function and fed into the CNN multi-label classifier. Experimental evaluation on the PLAID aggregated dataset shows that the proposed approach is very successful at recognizing multiple appliances from aggregated measurements with an overall 0.94 F-score.

We further show the effectiveness of the proposed CNN multi-label learning in recognizing a single running appliance with over 98% accuracy. We will investigate the use of Fryze’s current decomposition and distance similarity matrix for single-label appliance recognition in future iterations of this work. Finally, we presented a detailed error analysis and identified three types of errors: *zero-error*, *one-to-one*, and *many-to-many* errors.

At this point, we acknowledge that the performance of the proposed approach is not yet satisfactory in detecting triple running appliances. A possible explanation for this issue is the small number of training samples with more than two running appliances. In the future, we would like to test our approaches against datasets with more training data. However, this may imply the development of such a dataset since the currently available ones are still scarce concerning high-frequency measurements [52,53].

Finally, it should be mentioned that the proposed method assumes that the appliance state transition (power event) is known in advance. However, in practice, this information has to be provided by an event detection algorithm (e.g., [54–56]). Therefore, future work should investigate how to integrate the proposed approach in the event-based NILM pipeline. Specifically, we plan to explore the use of the proposed Fryze current decomposition for event detection in multi-label appliance recognition.

Author Contributions: Conceptualization, A.F.; data curation, A.F.; formal analysis, A.F. and L.P.; methodology, A.F. and L.P.; resources, A.F.; software, A.F.; supervision, L.P.; validation, L.P.; writing—original draft, A.F.; writing—review and editing, A.F. and L.P. All authors have read and agreed to the published version of the manuscript.

Funding: Lucas Pereira has received funding from the Portuguese Foundation for Science and Technology (FCT) under grants CEECIND/01179/2017 and UIDB/50009/2020.

Acknowledgments: The authors thank Christoph Klemenjak and Shridhar Kulkarni for providing insightful comments and advice towards the completion of this work.

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

References

1. Monacchi, A.; Versolatto, F.; Herold, M.; Egarter, D.; Tonello, A.M.; Elmenreich, W. An Open Solution to Provide Personalized Feedback for Building Energy Management. *CoRR* **2015**, *abs/1505.0*, 1–28. [[CrossRef](#)]
2. Batra, N.; Singh, A.; Whitehouse, K. If You Measure It, Can You Improve It? Exploring The Value of Energy Disaggregation. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments—BuildSys '15, Seoul, Korea, 4–5 November 2015; pp. 191–200. [[CrossRef](#)]
3. Froehlich, J.; Larson, E.; Gupta, S.; Cohn, G.; Reynolds, M.; Patel, S. Disaggregated end-use energy sensing for the smart grid. *IEEE Pervasive Comput.* **2011**, *10*, 28–39. [[CrossRef](#)]
4. Reyes Lua, A. Location-aware Energy Disaggregation in Smart Homes. Master's Thesis, Delft University of Technology, Delft, Netherlands, 2015.
5. Klemenjak, C.; Jost, S.; Elmenreich, W. Yomopie: A user-oriented energy monitor to enhance energy efficiency in households. In Proceedings of the 2018 IEEE Conference on Technologies for Sustainability (SusTech), Long Beach, CA, USA, 11–13 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–7.
6. Hart, G. Nonintrusive appliance load monitoring. *Proc. IEEE* **1992**, *80*, 1870–1891. [[CrossRef](#)]
7. Zeifman, M.; Roth, K. Nonintrusive appliance load monitoring: Review and outlook. *IEEE Trans. Consum. Electron.* **2011**, *57*, 76–84. [[CrossRef](#)]
8. Zoha, A.; Gluhak, A.; Imran, M.A.; Rajasegarar, S. Non-intrusive Load Monitoring approaches for disaggregated energy sensing: A survey. *Sensors* **2012**, *12*, 16838–16866. [[CrossRef](#)]
9. Klemenjak, C.; Elmenreich, W. On the applicability of correlation filters for appliance detection in smart meter readings. In Proceedings of the 2017 IEEE International Conference on Smart Grid Communications (SmartGridComm), Dresden, Germany, 23–27 October 2017; IEEE: Piscataway, NJ, USA 2017; pp. 171–176.
10. Nalmpantis, C.; Vrakas, D. Machine learning approaches for non-intrusive load monitoring: From qualitative to quantitative comparison. *Artif. Intell. Rev.* **2018**, *52*, 217–243. [[CrossRef](#)]
11. Faustine, A.; Pereira, L. Improved Appliance Classification in Non-Intrusive Load Monitoring Using Weighted Recurrence Graph and Convolutional Neural Networks 2019. *Energies* **2020**, *13*, 3374. [[CrossRef](#)]
12. Gomes, E.; Pereira, L. PB-NILM: Pinball Guided Deep Non-Intrusive Load Monitoring. *IEEE Access* **2020**, *8*, 48386–48398. [[CrossRef](#)]
13. Faustine, A.; Pereira, L.; Klemenjak, C. Adaptive Weighted Recurrence Graphs for Appliance Recognition in Non-Intrusive Load Monitoring. *IEEE Trans. Smart Grid* **2020**, *1*. [[CrossRef](#)]
14. De Baets, L.; Ruysinck, J.; Develder, C.; Dhaene, T.; Deschrijver, D. Appliance classification using VI trajectories and convolutional neural networks. *Energy Build.* **2018**, *158*, 32–36. [[CrossRef](#)]
15. Lam, H.Y.; Fung, G.S.K.; Lee, W.K. A Novel Method to Construct Taxonomy Electrical Appliances Based on Load Signatures of. *IEEE Trans. Consum. Electron.* **2007**, *53*, 653–660. [[CrossRef](#)]
16. Wang, A.L.; Chen, B.X.; Wang, C.G.; Hua, D. Non-intrusive load monitoring algorithm based on features of V–I trajectory. *Electr. Power Syst. Res.* **2018**, *157*, 134–144. [[CrossRef](#)]
17. Du, L.; He, D.; Harley, R.G.; Habetler, T.G. Electric Load Classification by Binary Voltage–Current Trajectory Mapping. *IEEE Trans. Smart Grid* **2016**, *7*, 358–365. [[CrossRef](#)]
18. Gao, J.; Kara, E.C.; Giri, S.; Bergés, M. A feasibility study of automated plug-load identification from high-frequency measurements. In Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 14–16 December 2015; pp. 220–224. [[CrossRef](#)]
19. De Baets, L.; Dhaene, T.; Deschrijver, D.; Develder, C.; Berges, M. VI-Based Appliance Classification Using Aggregated Power Consumption Data. In Proceedings of the 2018 IEEE International Conference on Smart Computing (SMARTCOMP), Sicily, Italy, 18–20 June 2018; pp. 179–186. [[CrossRef](#)]
20. Liu, Y.; Wang, X.; You, W. Non-Intrusive Load Monitoring by Voltage–Current Trajectory Enabled Transfer Learning. *IEEE Trans. Smart Grid* **2019**, *10*, 5609–5619. [[CrossRef](#)]

21. Baptista, D.; Mostafa, S.; Pereira, L.; Sousa, L.; Morgado, D.F. Implementation Strategy of Convolution Neural Networks on Field Programmable Gate Arrays for Appliance Classification Using the Voltage and Current (V-I) Trajectory. *Energies* **2018**, *11*, 2460. [[CrossRef](#)]
22. Yeh, C.K.; Wu, W.C.; Ko, W.J.; Wang, Y.C.F. Learning Deep Latent Spaces for Multi-Label Classification. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), San Francisco, CA, USA, 4–9 February 2017; pp. 2838–2844.
23. Yang, Y.; Zhong, J.; Li, W.; Gulliver, T.A.; Li, S. Semi-Supervised Multi-Label Deep Learning based Non-intrusive Load Monitoring in Smart Grids. *IEEE Trans. Ind. Inform.* **2019**, *10*, 1. [[CrossRef](#)]
24. Massidda, L.; Marrocu, M.; Manca, S. Non-Intrusive Load Disaggregation by Convolutional Neural Network and Multilabel Classification. *Appl. Sci.* **2020**, *10*, 1454. [[CrossRef](#)]
25. Basu, K.; Debusschere, V.; Bacha, S.; Maulik, U.; Bondyopadhyay, S. Nonintrusive Load Monitoring: A Temporal Multilabel Classification Approach. *IEEE Trans. Ind. Inform.* **2015**, *11*, 262–270. [[CrossRef](#)]
26. 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](#)]
27. Buddhahai, B.; Wongseeree, W.; Rakkwamsuk, P. A non-intrusive load monitoring system using multi-label classification approach. *Sustain. Cities Soc.* **2018**, *39*, 621–630. [[CrossRef](#)]
28. Teshome, D.F.; Huang, T.D.; Lian, K. Distinctive Load Feature Extraction Based on Fryze’s Time-Domain Power Theory. *IEEE Power Energy Technol. Syst. J.* **2016**, *3*, 60–70. [[CrossRef](#)]
29. Dokmanic, I.; Parhizkar, R.; Ranieri, J.; Vetterli, M. Euclidean Distance Matrices: Essential theory, algorithms, and applications. *IEEE Signal Process. Mag.* **2015**, *32*, 12–30. [[CrossRef](#)]
30. Medico, R.; De Baets, L.; Gao, J.; Giri, S.; Kara, E.; Dhaene, T.; Develder, C.; Bergés, M.; Deschrijver, D. A voltage and current measurement dataset for plug load appliance identification in households. *Sci. Data* **2020**, *7*, 49. [[CrossRef](#)] [[PubMed](#)]
31. Verma, S.; Singh, S.; Majumdar, A. Multi Label Restricted Boltzmann Machine for Non-intrusive Load Monitoring. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 8345–8349.
32. Buddhahai, B.; Wongseeree, W.; Rakkwamsuk, P. An Energy Prediction Approach for a Nonintrusive Load Monitoring in Home Appliances. *IEEE Trans. Consumer Electron.* **2019**, *66*, 96–105. [[CrossRef](#)]
33. Singh, S.; Majumdar, A. Non-intrusive Load Monitoring via Multi-label Sparse Representation based Classification. *IEEE Trans. Smart Grid* **2019**, *11*, 1799–1801. [[CrossRef](#)]
34. Kongsorot, Y.; Horata, P. Multi-label classification with extreme learning machine. In Proceedings of the 2014 6th International Conference on Knowledge and Smart Technology (KST), Chonburi, Thailand, 30–31 January 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 81–86.
35. Li, D.; Dick, S. Residential household non-intrusive load monitoring via graph-based multi-label semi-supervised learning. *IEEE Trans. Smart Grid* **2018**, *10*, 4615–4627. [[CrossRef](#)]
36. Singhal, V.; Maggu, J.; Majumdar, A. Simultaneous detection of multiple appliances from smart-meter measurements via multi-label consistent deep dictionary learning and deep transform learning. *IEEE Trans. Smart Grid* **2018**, *10*, 2269–2987. [[CrossRef](#)]
37. Nalmpantis, C.; Vrakas, D. On time series representations for multi-label NILM. *Neural Comput. Appl.* **2020**. [[CrossRef](#)]
38. Li, L.; Zhao, Y.; Jiang, D.; Zhang, Y.; Wang, F.; Gonzalez, I.; Valentin, E.; Sahli, H. Hybrid Deep Neural Network–Hidden Markov Model (DNN-HMM) Based Speech Emotion Recognition. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2–5 September 2013; pp. 312–317. [[CrossRef](#)]
39. Hassan, T.; Javed, F.; Arshad, N. An Empirical Investigation of V-I Trajectory Based Load Signatures for Non-Intrusive Load Monitoring. *IEEE Trans. Smart Grid* **2014**, *5*, 870–878. [[CrossRef](#)]
40. Klemenjak, C.; Makonin, S.; Elmenreich, W. Towards Comparability in Non-Intrusive Load Monitoring: On Data and Performance Evaluation. In Proceedings of the 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), The Hague, The Netherlands, 25–28 October 2020; IEEE: Piscataway, NJ, USA, 2020.

41. Staudt, V. Fryze-Buchholz-Depenbrock: A time-domain power theory. In Proceedings of the 2008 International School on Nonsinusoidal Currents and Compensation, Lagow, Poland, 10–13 June 2008. [[CrossRef](#)]
42. Keogh, E.J.; Pazzani, M.J. Scaling Up Dynamic Time Warping for Datamining Applications. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2000; KDD '00, pp. 285–289. [[CrossRef](#)]
43. Ontañón, S. An overview of distance and similarity functions for structured data. *Artif. Intell. Rev.* **2020**. [[CrossRef](#)]
44. Mahajan, D.; Girshick, R.; Ramanathan, V.; Paluri, M.; Li, Y.; Bharambe, A.; Maaten, L.v.d. Exploring the Limits of Weakly Supervised Pretraining. In *Computer Vision—ECCV 2018*; Springer: Berlin/Heidelberg, Germany, 2018. [[CrossRef](#)]
45. Lanchantin, J.; Sekhon, A.; Qi, Y. Neural Message Passing for Multi-Label Classification. In *Machine Learning and Knowledge Discovery in Databases*; Brefeld, U., Fromont, E., Hotho, A., Knobbe, A., Maathuis, M., Robardet, C., Eds.; ECML PKDD 2019, Lecture Notes in Computer Science; Springer: Cham, Switzerland; Volume 11907. [[CrossRef](#)]
46. Sechidis, K.; Tsoumakas, G.; Vlahavas, I. On the Stratification of Multi-label Data. In *Machine Learning and Knowledge Discovery in Databases*; Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 145–158.
47. Zhang, M.L.; Zhou, Z.H. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognit.* **2007**, *40*, 2038–2048. [[CrossRef](#)]
48. Spyromitros, E.; Tsoumakas, G.; Vlahavas, I. An Empirical Study of Lazy Multilabel Classification Algorithms. In *Artificial Intelligence: Theories, Models and Applications*; Springer: Berlin/Heidelberg, Germany; pp. 401–406. [[CrossRef](#)]
49. Lai, Y.X.; Lai, C.F.; Huang, Y.M.; Chao, H.C. Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home. *Inf. Sci.* **2013**, *230*, 39–55. [[CrossRef](#)]
50. Kolter, J.Z.; Johnson, M.J. REDD : A Public Data Set for Energy Disaggregation Research. In Proceedings of the 1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD'11), San Diego, CA, USA, 21 August 2011; pp. 1–6.
51. Kelly, J.; Knottenbelt, W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Sci. Data* **2015**, *2*, 150007. [[CrossRef](#)]
52. Pereira, L.; Nunes, N. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1265. [[CrossRef](#)]
53. Klemenjak, C.; Reinhardt, A.; Pereira, L.; Makonin, S.; Bergés, M.; Elmenreich, W. Electricity Consumption Data Sets: Pitfalls and Opportunities. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, BuildSys '19, New York, NY, USA, 13–14 November 2019; ACM: New York, NY, USA, 2019. [[CrossRef](#)]
54. Pereira, L. Developing and evaluating a probabilistic event detector for non-intrusive load monitoring. In *2017 Sustainable Internet and ICT for Sustainability (SustainIT)*; IEEE: Funchal, Portugal, 2017; pp. 1–10. [[CrossRef](#)]
55. De Baets, L.; Ruysinck, J.; Develder, C.; Dhaene, T.; Deschrijver, D. On the Bayesian optimization and robustness of event detection methods in NILM. *Energy Build.* **2017**, *145*, 57–66. [[CrossRef](#)]
56. Houidi, S.; Auger, F.; Sethom, H.B.A.; Fourer, D.; Miègeville, L. Multivariate event detection methods for non-intrusive load monitoring in smart homes and residential buildings. *Energy Build.* **2020**, *208*, 109624. [[CrossRef](#)]

