IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 April 2023, accepted 2 May 2023, date of publication 15 May 2023, date of current version 24 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3276475

RESEARCH ARTICLE

Deep Learning Based Non-Intrusive Load Monitoring for a Three-Phase System

B. GOWRIENANTHAN[1], N. KIRUTHIHAN[1], K. D. I. S. RATHNAYAKE[1], S. KIRUTHIKAN[1], V. LOGEESHAN[1], (Member, IEEE), S. KUMARAWADU[1], (Senior Member, IEEE), AND C. WANIGASEKARA[2], (Member, IEEE)

[1]Department of Electrical Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka
[2]Institute for the Protection of Maritime Infrastructures, German Aerospace Centre, 27572 Bremerhaven, Germany

Corresponding author: V. Logeeshan (logeeshanv@uom.lk)

ABSTRACT Non-Intrusive Load Monitoring (NILM) is a method to determine the power consumption of individual appliances from the overall power consumption measured by a single measurement device, which is usually the main meter. Increase in the adoption of smart meters has facilitated large scale implementation of NILM, which can provide information about individual loads to the utilities and consumers. This will lead to significant energy savings as well as better demand-side management. Researchers have proposed several methods and have successfully implemented NILM for residential sectors that have a single-phase supply. However, NILM has not been successfully implemented for industrial and commercial buildings that have a three-phase supply, due to several challenges. These buildings consume significant amount of power and implementing NILM to these buildings has the potential to yield substantial benefits. In this paper, we propose a novel deep learning-based approach to address some of the key challenges in implementing NILM for buildings that have a three-phase supply. Our approach introduces an ensemble learning technique that does not require training of multiple neural network models, which reduces the computational requirements and makes it economically feasible. The model was tested on a three-phase system that consists of both three- phase loads and single-phase loads. The results show significant improvement in load disaggregation compared to the existing methods and indicate its applicability.

INDEX TERMS NILM, neural networks, deep learning, ensemble learning, load disaggregation.

I. INTRODUCTION

The gradual depletion of fossil fuel reserves and the escalating threat of global warming caused by their emissions have driven the world towards energy conservation. Studies have indicated that optimal usage of appliances in the domestic sector can reduce the energy consumption by 20% to 15% [1]. However, consumers will not be able to optimize their usage pattern based on the monthly electricity bill that shows the aggregated power consumption.

Secondly, implementation of demand response schemes by the power utility can reduce the peak demand, which will flatten the demand curve and reduce the capacity cost of the power system [2]. Demand response schemes can also be used as a solution to mitigate the solar duck curve problem, which will enable more solar photovoltaic (PV) to be integrated into the electric grid [3]. However, demand response cannot be effectively implemented based on the aggregated power consumption data of every consumer.

In addition to this, changes in the power consumption pattern of a device due to improper maintenance, a possible fault or a mistake by the user cannot be identified from the overall consumption data.

These cases require the monitoring of appliance level power consumption. The most feasible method of monitoring is non-intrusive load monitoring (NILM), where power consumption is measured by a single metering device installed at the main feeding panel and is disintegrated into appliance level consumption [4]. Since only a single measurement device is utilized, it requires less hardware, which

The associate editor coordinating the review of this manuscript and approving it for publication was Akshay Kumar Saha.

results in lesser maintenance, and is less disruptive to the consumer.

The topic of NILM was introduced by George Hart in the 1980s, considering only the active power [4]. In his approach, switching on and off of an appliance is detected based on a predefined threshold difference in the power consumption. A major shortcoming in his approach is that it can only disaggregate ON/OFF state appliances and it cannot distinguish appliances that consume similar active power. Later on, by combining reactive power and active power, such devices were also identified and distinguished. However, it failed to identify loads that consumed less than 150W power.

Further research in NILM branched out based on the frequency range of measurement. They can be categorized as high frequency (MHz range), mid frequency (KHz range) and low frequency (less than 1Hz) [5]. Electrical devices are known to generate electrical noise at frequencies significantly higher than the fundamental frequency. This noise is produced briefly during device turn on and off cycles and in some cases, continuously, during the operation of the device due to the utilization of switched-mode power supplies that provide DC power to internal electronics. Some researchers have explored the utility of collecting these higher frequency signals(high and mid frequency signals) to implement NILM.

NILM based on high frequency was introduced by Patel et al. [6] based on the noise of the voltage signal. High frequency signal acquisition devices are expensive and prone to error due to radio signal interferences. Mid frequency based NILM provides good compromise between cost and accuracy [5]. Most of mid frequency based NILM are highly complex, and the computations required to label the load combination is a bottleneck to implement in an embedded platform

To mitigate these limitations and issues, machine learning based approaches were developed [18], [19]. J. Kelly and W. Knottbelt, developed a NILM model based on deep learning for the first time which showed significant improvements in load disaggregation compared to the existing machine learning based models [7]. Their model utilized sequence-to-sequence learning. The sequence-to-sequence model makes the predictions for all the time steps in the given input time series window. So, for a given range of time steps to predict, it has to iterate only once. This results in a neural network that is faster and consumes less resource but has higher error. In addition to this, it was developed on Vanilla Deep neural network architecture instead of convolutional neural network (CNN) or recurrent neural network (RNN), which loses the spatial data from the input data. This resulted in reduced performance of the output.

In 2017, C. Zhang proposed a CNN based sequence-to-point neural network model which performed far better than sequence to sequence models [8].The sequence-to-point model makes the prediction for one step at a time. This step will probably be located in the center time of the time series data window that was fed to the model. Therefore, for a given range of time steps to predict, the model has to iterate for each sample in the window for the prediction. This requires a higher number of calculations for the same amount of data input, since they need to predict multiple times (outputting a single point each time) for a window of data to get the predictions for the entire window. This process is time consuming and requires high computational power. In comparison, sequence-to-sequence models can be trained faster, and their prediction time is lower, compared to sequence to point models.

Regardless of the benefits and drawbacks, several sequence-to-sequence and sequence-to-point models were developed and implemented for the domestic sector, which has a single-phase supply. However, only a few models were developed for industrial and commercial buildings that have a three-phase supply. Most of the existing research on NILM for 3-phase systems focused only on the challenges related to implementation of NILM. These buildings consume significantly high power and implementing NILM can reap a lot of benefits. However, development of NILM models were limited due to various challenges and complexities [9].

The presence of multiple loads that are similar in nature poses a significant limitation to implement NILM in commercial and industrial buildings. In these scenarios, distinguishing between loads becomes challenging, and this can lead to errors in load identification. Additionally, the number and complexity of the loads are often much higher than those in residential sector, which makes it more challenging to develop and implement an accurate and robust NILM system.

Another obstacle in implementing NILM for a three-phase system is the identification of continuously operating loads. These types of loads do not have clear on/off cycles, which makes it extremely difficult to determine their energy consumption patterns. As a result, the identification of continuously operating loads requires sophisticated algorithms that can accurately differentiate them from other loads.

If the same NILM model is deployed across multiple buildings, which consists of a larger number of more complex appliances, the performance of a NILM model decreases significantly. To mitigate this problem, custom NILM models must be developed for each client building, which can be computationally expensive and economically infeasible for three-phase systems.

Lastly, collecting data from commercial and industrial buildings without disrupting their operations is another significant challenge. These buildings are typically in use throughout the day, which makes it challenging to collect data without causing disruption. The process of data collection must be carefully planned and executed to minimize disruption. This research focuses on addressing some of these challenges and develop a NILM system that is accurate and robust.

Our work has contributed to Non-Intrusive Load Monitoring by addressing challenges related to disaggregating single phase and three-phase loads, improving accuracy and efficiency of demand-side managementÂ systems.

This paper is organized as follows. Section II describes in detail, the challenges that are addressed in this research. Section III describes the proposed approach. Section IV describes the preparation of data for training and evaluation. Section V illustrates the experimentation carried out and the results. Section VI concludes the outcome of this research.

II. CHALLENGES ADDRESSED

This research addresses several challenges related to disaggregation of both single-phase and three-phase loads in a three-phase system. This section is divided into two subsections. The first subsection discusses the challenges related to disaggregation of single-phase loads and the second subsection discusses the challenges related to disaggregation of three-phase loads.

A. CHALLENGES IN DISAGGREGATION OF SINGLE-PHASE LOADS

NILM for single-phase systems is a well-researched topic and several deep learning-based models with good accuracy already exist. However, the available single-phase NILM models take only one phase as input. This becomes a problem in three-phase systems, as the model must determine the phase to which the target appliance (appliance that is meant to be detected by NILM) is connected. One solution to overcome this problem is to request the consumer to manually determine the phase and give the aggregate power consumption of that phase as input to the single-phase NILM model. However, an average resident does not have sufficient expertise to determine the phase to which the appliance is connected. In addition, if the appliance is moved to a different location of the building, it might be connected to a different phase and it will not be detected.

To mitigate this problem, neural network for a target appliance can be implemented to all three phases separately. Hence, there will be three neural networks to detect and determine the power consumption of one target appliance. This is not computationally efficient since the target appliance will be detected by only one of the three neural networks and the other two neural networks operate without having the target appliance in their respective phases.

Another problem arises due to the presence of three-phase loads. Generally, single-phase appliances tend to have significantly lower power consumption in comparison to 3-phase appliances. Hence, their activations may get drowned-out amongst the larger 3-phase activations in the aggregate waveform. This makes it harder for the neural network to identify the presence of the single-phase activations in the crowded aggregate power consumption waveforms.

B. CHALLENGES IN DISAGGREGATION OF THREE-PHASE LOADS

Three key issues related to implementation of NILM for three-phase systems are addressed in this research. The first one is the presence of multiple similar loads. A three-phase system might consist of multiple appliances of the same type and even the same model as well. There are two different cases related to this challenge. The first case is when there are two appliances of the same type (such as two air conditioners) that have different power ratings or from different manufacturers. In this case, there will be unique patterns in the activations of these similar appliances that can be detected by the neural network. The second case is where there are multiple appliances of the same power rating and manufacturer. This case is more difficult to handle since the activation patterns of these appliances will be nearly the same.

The second challenge is to identify continuously operating loads. In NILM research related to single-phase systems, continuously operating loads such as routers and smoke detectors are usually ignored since it is difficult to detect whether they are ON or OFF and they don't consume significant power. It is difficult to detect their ON/OFF state because they are ON for most of the day and will have very random ON or OFF events. In 3-phase systems, there may be continuously operating loads such as exhaust fans that consume a significant amount of power. They cannot be ignored and the NILM model should be able to determine their power consumption as well.

The general idea in developing a deep learning based NILM model for a single phase system is to train a dedicated neural network for each target appliance, which can be deployed across multiple consumers. This paradigm is accepted under the assumption that the power consumption pattern and the power rating of a target appliance is similar, regardless of the manufacturer. However, this paradigm is not valid for a three-phase system. Appliances connected to a three phase supply in commercial and industrial buildings are much higher in number and are more complex in nature [9]. Therefore, the decrease in performance when the same NILM model is deployed across multiple buildings will be significant. To mitigate this problem, NILM models must be custom made for each client building, which incurs high computational cost. This makes NILM for a three phase system economically infeasible. These three issues are addressed in this research.

III. PROPOSED APPROACH

In our proposed method, a measurement device is attached to the three-phase supply which records the power consumption data of each phase separately and sends it to a server in the cloud.This raw aggregate power consumption data is preprocessed in this server and is disaggregated into appliance level power consumption by a deep learning model.

The deep learning model consists of dedicated neural networks that are trained for a target appliance. The pre-processed aggregate data is given as input to each neural network, which estimates the power consumption of the target appliance that it was trained for. The estimated power consumption of each appliance is sent to a web application where the consumer can check the power consumption of each of the target appliances in their building. The overview of this model is illustrated in Fig. 1.

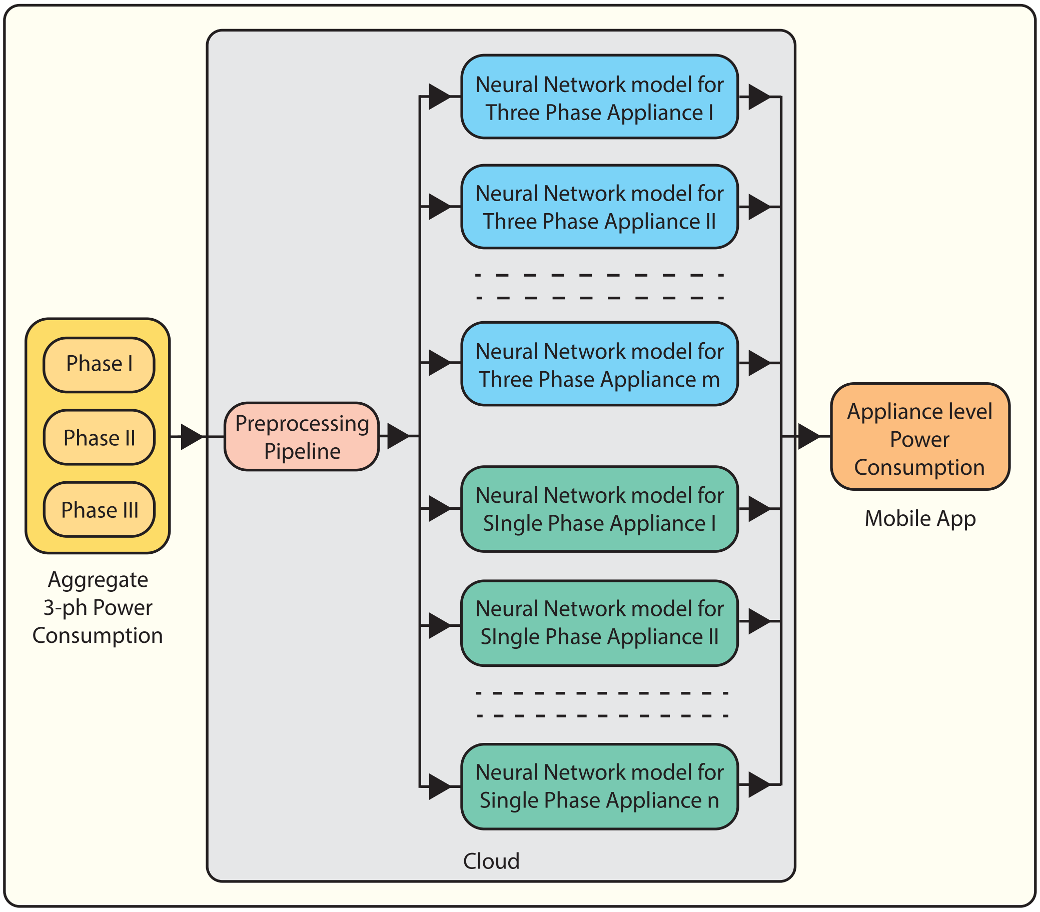

FIGURE 1. **Model overview.**

In order to solve the issues related to disaggregation of single-phase loads discussed in subsection A of section II, we feed the aggregate power consumption data of all the three phases as input to a single neural network. This neural network model learns to determine the power consumption of that target appliance no matter which phase it is connected to. By passing all three phases as input to the neural network, it will learn to use the differences between the aggregates of each phase to accurately predict the power consumption of the target single-phase appliances. This will improve the accuracy and is computationally efficient as well.

The second issue due to presence of three-phase loads is automatically solved by feeding all three phases as input to the neural network. The neural network learns to use the differences between the aggregates of each phase to separate the total power consumption of all the 3-phase loads from the aggregate waveform, leaving only the power consumption of the single-phase loads. This makes it much easier to detect the power consumption of the target single phase device.

Presence of multiple similar loads and continuously operating loads are two key challenges in disaggregation of three-phase loads, as discussed in subsection B of section II. This is mitigated by including continuously operating loads and loads that have very similar activation functions, in the training dataset. The performance of the neural network in these particular scenarios are evaluated and the network architecture is modified to improve the accuracy.

In order to reduce the computational cost and make NILM for three phase systems economically feasible while maintaining the desired level of accuracy, a novel ensemble learning method is introduced, together with a sequence-to-sequence model.

The proposed method is non-intrusive as it does not require any additional hardware to be installed. Instead, it makes use of easily accessible smart meter data. As no extra equipment or monitoring devices need to be built, this method is economical and practical. Without directly measuring each appliance's energy use, the model is able to break down each individual energy usage. This is accomplished by examining the unique patterns and signatures of each appliance's energy use as they are reflected in the data from smart meters. Thus, the suggested paradigm is a viable and non-intrusive approach to energy disaggregation, with the potential to be widely adopted byÂ theÂ industry.

This section is divided into three subsections. Subsection I describes preprocessing of the dataset. Subsection

II describes the neural network architecture. Subsection III describes the novel ensemble learning method proposed to improve the performance of disaggregation by the neural network.

A. DATA PREPROCESSING

Data preprocessing is carried out on the raw power measurement data, to make it suitable for neural networks. The preprocessing pipeline does the following process [7], [8], [10].

1) Extract the appliance and main power consumption data from the dataset
   In this step, we extract the relevant data from the raw power measurement data that we need for our analysis. We only keep the data related to the main power consumption and the appliance power consumption.
2) Take the records only when both appliance and main power sensor are active.
   Here, we filter out the data that is not relevant to us. We only keep the data where both the appliance and main power sensors are active. This ensures that we have accurate data to work with.
3) Break the time series data when there is a gap in sensor reading.
   When there is a gap in sensor readings, we need to break the time series data into separate chunks. This is because the neural network needs a continuous stream of data to work with. Breaking the time series data ensures that we have continuous data without any gaps.
4) If the time sample length is less than the desired length; pad the remaining time points with zero padding.
   Sometimes, the time series data that we have may not be long enough for our analysis. In such cases, we pad the remaining time points with zero padding. This simulates an all-power-off scenario and ensures that we have enough data to work with.
5) Resample the data to get a sample from aggregate power and appliance power on a constant sampling time interval.
   Resampling is the process of changing the rate at which data is sampled. In this step, we resample the data to get a sample from aggregate power and appliance power on a constant sampling time interval. This ensures that we have a consistent sample rate to work with.
6) Normalize the data.
   Normalization is the process of scaling the data to a common range. In this step, we normalize the data to ensure that it is within a common range. This ensures that the neural network can work with the data effectively.
7) Split the data for training and validation.
   In this step, we split the data into training and validation sets. This ensures that we have enough data to train the neural network and that we can validate its performance on data that it has not seen before.
8) Split the data into clusters of window size time-series data.
   In this step, we split the data into clusters of window size time-series data. This ensures that we have enough data to work with and that the neural network can analyze the data effectively.
9) Group the data into batches.
   In this final step, we group the data into batches. This ensures that we can feed the data into the neural network efficiently and that the neural network can analyze the data effectively.

B. NEURAL NETWORK ARCHITECTURE

Since the aggregate power is measured over time, it is a time series data. CNN can be used for time series data by treating the time dimension as the spatial dimension and applying 1D convolutions. The mathematical model for a CNN used for time series data is as follows.

Let X be an input time series with T time steps and D dimensions, represented as a tensor of size T × D. The CNN consists of L layers, where each layer l is defined by a set of parameters (weights and biases).

The output of the l^{th} layer, denoted as Z_l, is computed as shown in equation 1.

$$Z_l = f_l(W_l * A_{l-1} + b_l) \tag{1}$$

Here, A_{l-1} is the output of the previous layer ($A_{l-1} = Z_{l-1}$ for $l > 1$, and $A_0 = X$). W_l is a set of learnable convolutional filters of size $K_l \times D_{l-1} \times D_l$, where K_l is the kernel size, D_{l-1} is the number of dimensions in the input feature map, and D_l is the number of filters in the current layer. b_l is a set of learnable biases of size D_l. f_l is the activation function.

The output of the final layer L, denoted as Y, is computed as shown in equation 2.

$$Y = softmax(W_f * A_{L-1} + b_f) \tag{2}$$

Here, W_f is a set of learnable weights of size $D_{L-1} \times C$, where C is the number of classes. Softmax is the activation function that converts the output into a probability distribution over the classes. b_f is a set of learnable biases of size C.

Based on this mathematical model, we developed a 2-stage sequence-to-sequence CNN based on the Wavenet (developed by Google) [17]. This wavenet model has been utilized in both stages of our neural network model. In the first stage, the wavenet model is deployed as an ON-OFF classifier. The classifier predicts whether the target appliance is ON or OFF with a probability value. Such a stage exists in previous NILM research as well [11]. However, the power consumption predicted by the NILM model is multiplied by the output (0 or 1) of the ON-OFF classifier. Therefore, when the ON-OFF classifier wrongly predicts that the appliance is OFF, the output of the NILM model as a whole will be zero, even if the regression model correctly predicts the power consumption of the target appliance. Therefore, any errors in the ON-OFF classifier prediction will heavily affect the accuracy of the NILM model. In addition, it is assumed that when the appliance is in OFF state it won't consume any power, which may not be the case in some situations.

In our architecture, the ON-OFF classifier's output (which is a probability value) is fed as an additional input to the Regression stage (that is, the stage that predicts the power consumption of the target appliance), as shown in equation 3.

$$G(x) = R(x, C(x)) \tag{3}$$

Here x is the input 3-phase aggregate power consumption data, C(x) is the output of the ON-OFF classifier, R(x,C(x)) the output of the regression stage, and G(x) is the overall output of the NILM model.

The intuition behind this is that this additional ON-OFF input allows the regression stage to learn how to use the ON-OFF probability prediction as a guide, rather than blindly multiplying its output with this binary ON-OFF prediction. This way, even if the ON-OFF model gives an erroneous prediction the regression model can still output the correct value. This creates a stronger connection between the ON-OFF stage and the regression stage.

For this classifier, sigmoid activation function is used. The mathematical function of the sigmoid activation is shown in equation 4.

$$Sigmoid = \frac{1}{1+e^{-x}} \tag{4}$$

As this function would limit the output between 0 to 1 with an infinite order of continuity, this can be easily presented as a probability of the appliance being switched on.

In the second stage, the wavenet model is deployed as a regresser, which predicts the power consumption of individual appliances. The aggregate power consumption data and the output of the ON-OFF classifier are fed as input to this regression model. The regression model processes these two information and predicts the power consumption of individual devices.

C. ENSEMBLE LEARNING

Ensemble learning is utilized to improve the accuracy of disaggregation. Traditional ensemble learning methods require multiple neural networks to be trained. The predictions from the neural networks are combined using methods such as bagging, boosting, and stacking [8]. This requires high computational power and longer training time. For a deep learning based NILM, separate neural networks must be trained for each target appliance that needs to be monitored. Therefore, implementing ensemble learning for each target appliance is not feasible.

To mitigate this problem, we propose a novel ensemble learning method that does not require multiple neural networks to be trained. A single neural network is allowed to train even after the loss has plateaued. The loss will oscillate around the same value but the model instances at different points in training will have varying characteristics. A certain set of model instances is chosen out of those with losses around the plateaued value. Then, the weights of these model instances are averaged, and a new model is created with the same architecture but with the averaged weight values. Specifically, for each layer in the model architecture, the weights of each of the chosen model instances for that layer are averaged element wise. Then, they are assigned to the corresponding layer of the averaged model.

Depending on the specific characteristics of the individual model instances, different combinations of them result in averaged models that have different loss values. Through trial and error, the best combination of model instances from the set of model instances with similar losses is chosen.

The advantage of this technique is that it not only produces improved performance like in other ensembling methods but also requires very little additional time and computation since it requires only a single neural network to be trained for a target device.

IV. DATASET

The NILM model was trained and evaluated on four datasets. The datasets used are

- UK Domestic Appliance-Level Electricity (UK-DALE) dataset [13].
- Reference Energy Disaggregation Data Set (REDD) [14].
- Domestic electricity demand dataset of individual appliances in Germany (DEDDIAG) [15].
- Industrial Machines Dataset for Electrical Load Disaggregation (IMDELD) [16].

This dataset contains five 3-phase appliances (Washing machine, air conditioner, exhaust fan and two refrigerators) and three 1-phase appliances (Washing machine, dishwasher, and microwave oven). The exhaust fans in this dataset are perfect examples of large continuously operating appliances. Such continuously operating appliances are generally very difficult to detect since they do not have consistent repeating activation patterns, but rather have very random ON/OFF events. Training and testing the NILM model on such complex data will determine the effectiveness and practical viability of the model.

In addition to these appliances which are target appliances, "distractor loads" are added, which are smaller single-phase loads such as laptops, light bulbs, chargers, and routers. These distractor loads are obtained from real-world single-phase datasets.

V. EVALUATION METRICES

Mean absolute error (MAE), f1-score, and estimation accuracy (EA) are used to evaluate the performance of the NILM model. The mean absolute error is calculated according to equation (2).

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |\hat{y_t} - y_t| \tag{5}$$

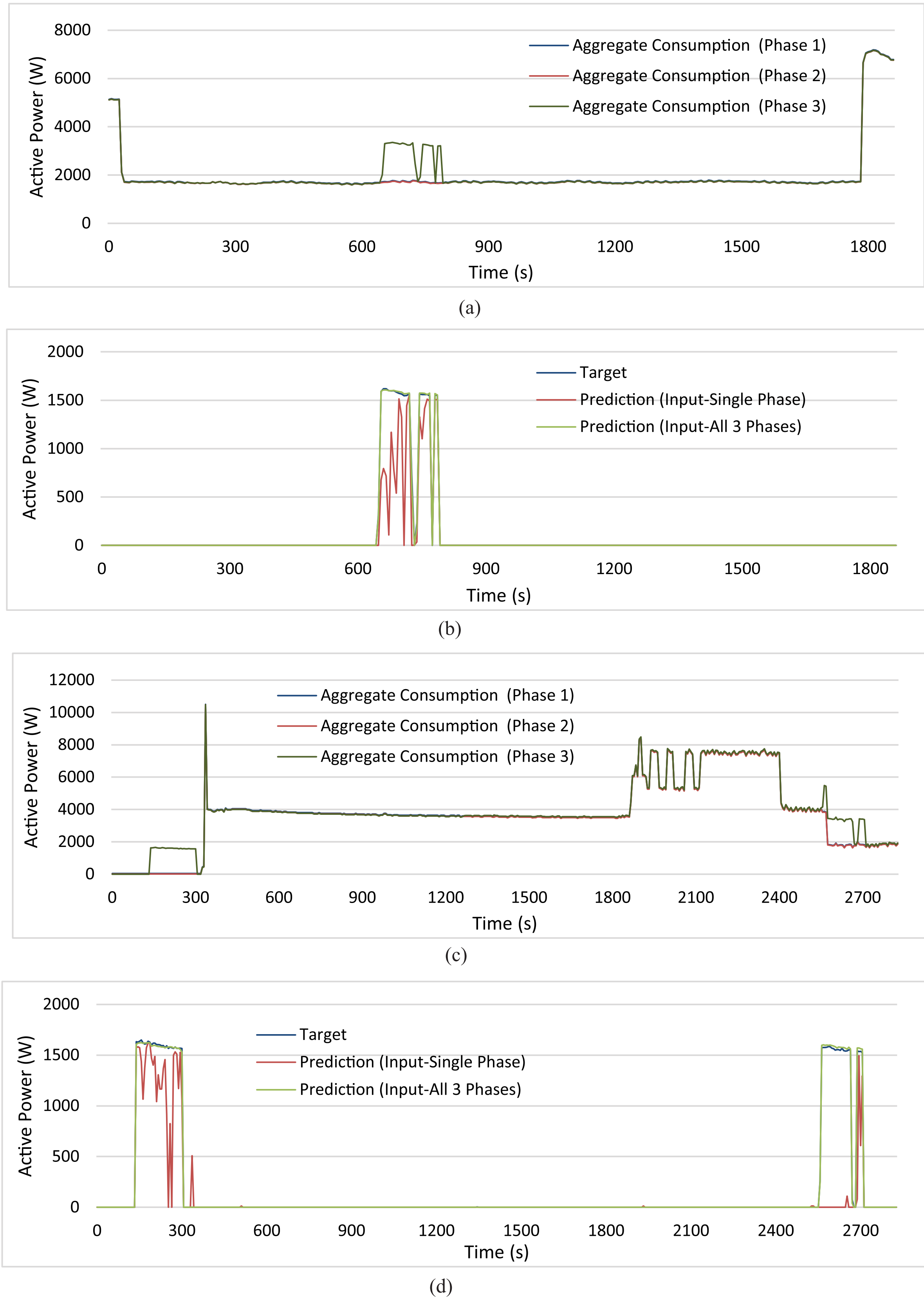

FIGURE 2. Performance evaluation of NILM model with single phase input and 3 phase input. (a), (c) - Aggregate power consumption of each phase. (b), (d) - Predicted and actual power consumption with washing machine as the target device.

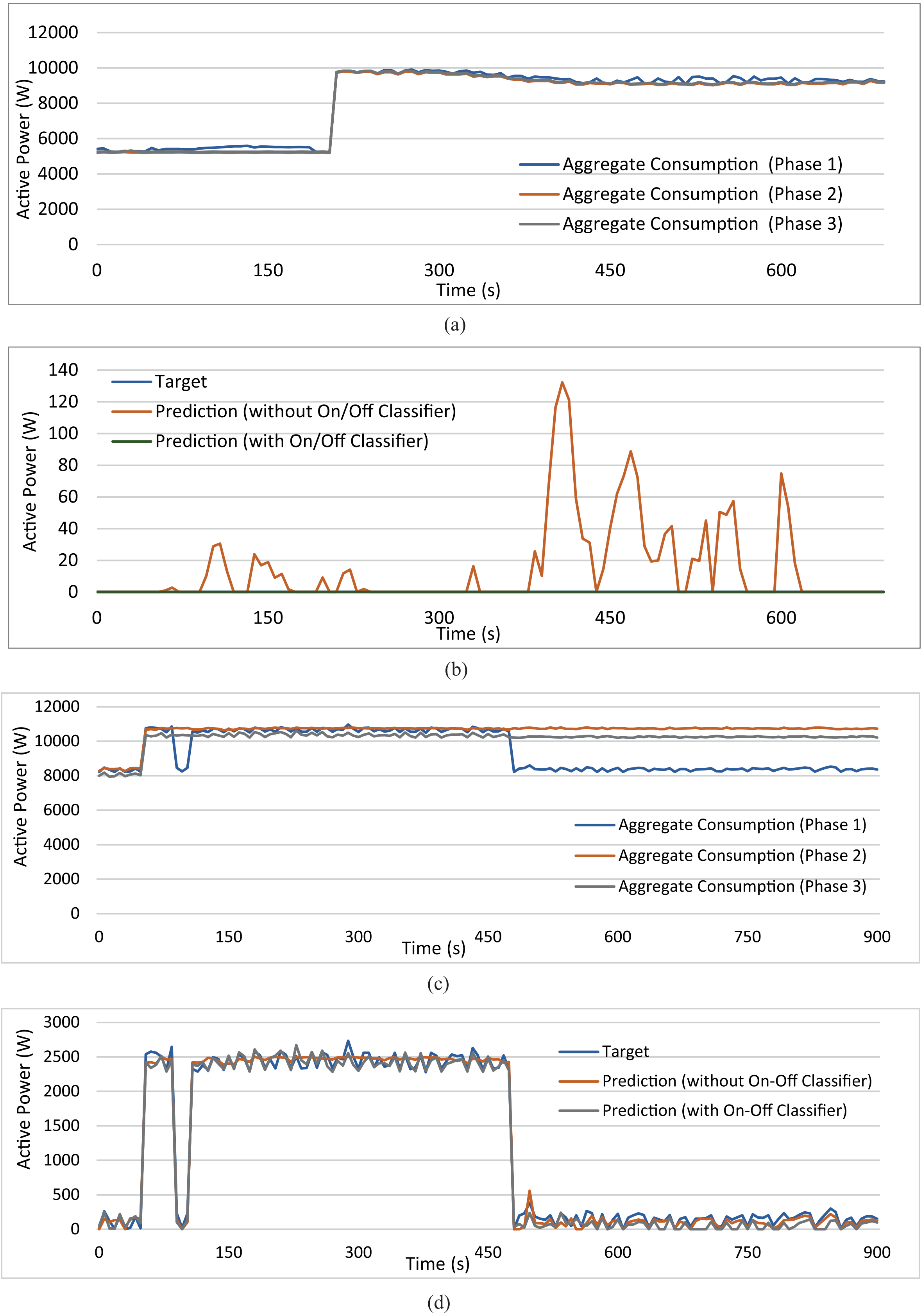

FIGURE 3. **Performance evaluation of NILM model with and without ON-OFF classifier. (a), (c) - Aggregate power consumption of each phase. (b), (d) - Predicted and actual power consumption with milling machine as the target device.**

where $\hat{y_t}$ is the prediction of the model at time t, and y_t is the actual power consumption of the appliance at time t. The f1-score is calculated according to equation (3)

$$f1 = \frac{2 * precision * recall}{precision + recall} \tag{6}$$

The precision and recall is calculated according to equation (4).

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN} \tag{7}$$

here TP, FP, and FN stand for true positives, false positives, and false negatives, respectively. When the prediction indicates that the appliance is ON when the appliance is actually ON, it is a true positive. When the prediction indicates that the appliance is ON when the appliance was actually OFF, it is a false positive. When the prediction indicates that the appliance was OFF when the appliance was actually ON, it is a false negative. We choose an appropriate threshold value, where if the predicted value is above the threshold, it is taken as an ON prediction, and OFF otherwise.

The F1 score combines precision and recall into a single score and is useful for evaluating classification models, particularly for imbalanced datasets. Precision is the fraction of true positive results among predicted positive results, while recall is the fraction of true positive results among actual positive results.

In the F1 score formula, precision and recall have equal importance, and F1 score can be used to determine overall model performance while precision and recall can be used to diagnose specific issues with model predictions. For example, a high precision and low recall may indicate that the model is too conservative and is missing too many positive examples, while a high recall and low precision may indicate that the model is too aggressive and is incorrectly classifying too many negative examples as positive.

The estimation accuracy is calculated using equation (5)

$$EA = 1 - \frac{\sum_{t=1}^{T} |\hat{y_t} - y_t|}{2\sum_{t=1}^{T} y_t} \tag{8}$$

where $\hat{y_t}$ is the prediction of the model at time t and y_t is the actual power consumption of the appliance at time t.

As mentioned in the work of Kolter and Johnson [14], estimation accuracy denotes how correctly the energy consumption of an appliance has been estimated, relative to the actual energy consumption of the appliance. The f1-score metric denotes the accuracy of the prediction of activeness of the appliance. The mean absolute error is an indication of the error of the prediction at each point in time, as mentioned in [6].

VI. EXPERIMENTATION AND RESULTS

Three novel techniques are introduced in this research. The first technique is to feed the power consumption data of all three phases as input to disaggregate single-phase loads. The second technique is to feed the ON-OFF classifier's output as an additional input to the Regression stage of the neural network. The third technique is the novel ensemble learning to obtain better predictive performance of the NILM model. The effectiveness of each these three techniques are discussed in this section.

This section is divided into three subsections. The first subsection discusses the evaluation of the first technique, which is to feed power consumption data of all three phases to disaggregate single-phase loads. The second subsection describes the effectiveness of the ON-OFF classifier in the regression stage of the neural network. The third subsection describes the performance of the novel ensemble learning.

A. DISAGGREGATION OF SINGLE PHASE LOADS

To evaluate the performance of disaggregation of single-phase loads when all three phases are fed as input, we conducted an experiment. In this experiment, two variations of a neural network NILM model were trained and evaluated: one that takes only one phase as input (the phase to which the target single-phase appliance is connected), and one that takes all three phases as input. Power measurements were taken with a sampling time of 6 seconds and the results are shown in Fig. 2.

Fig 2. (a) and (b) represents a scenario where the target device gets switched on for a brief period of time (about 200 seconds) while the power consumption of all the other loads remain nearly the same throughout the whole period. It can be seen in Fig. 2 (b) that during the period when the target device is switched on, the prediction when all three phases are input is highly accurate compared to when only one phase is input. Fig 2. (c) and (d) represents a scenario where a three-phase device switched on and the aggregate power consumption increased significantly. Even in that case, the model was able to predict correctly when all three phases are input while the prediction is erroneous when only single phase is given as input.

This demonstrates that when all three phases are fed, performance increases significantly. The reason for this is that the model was able to successfully learn to omit all the three phase loads by comparing the three-input phase power consumption values.

B. ON-OFF CLASSIFIER

In order to evaluate the performance of the ON-OFF classifier, we designed an experiment with two neural network architectures. The first neural network included the ON-OFF classifier whereas the second one did not include it. Other hyper parameters of the two neural networks were the same. The IMDELD dataset [14] was utilized for this experiment. Since it is a real-world dataset with aggregate data that is very chaotic at times, it can validate the performance of the ON-OFF classifier in such practical scenarios. The target appliance for this experiment is Milling Machine 1 (there are two milling machines in the IMDELD dataset). The results are shown in Fig. 3.

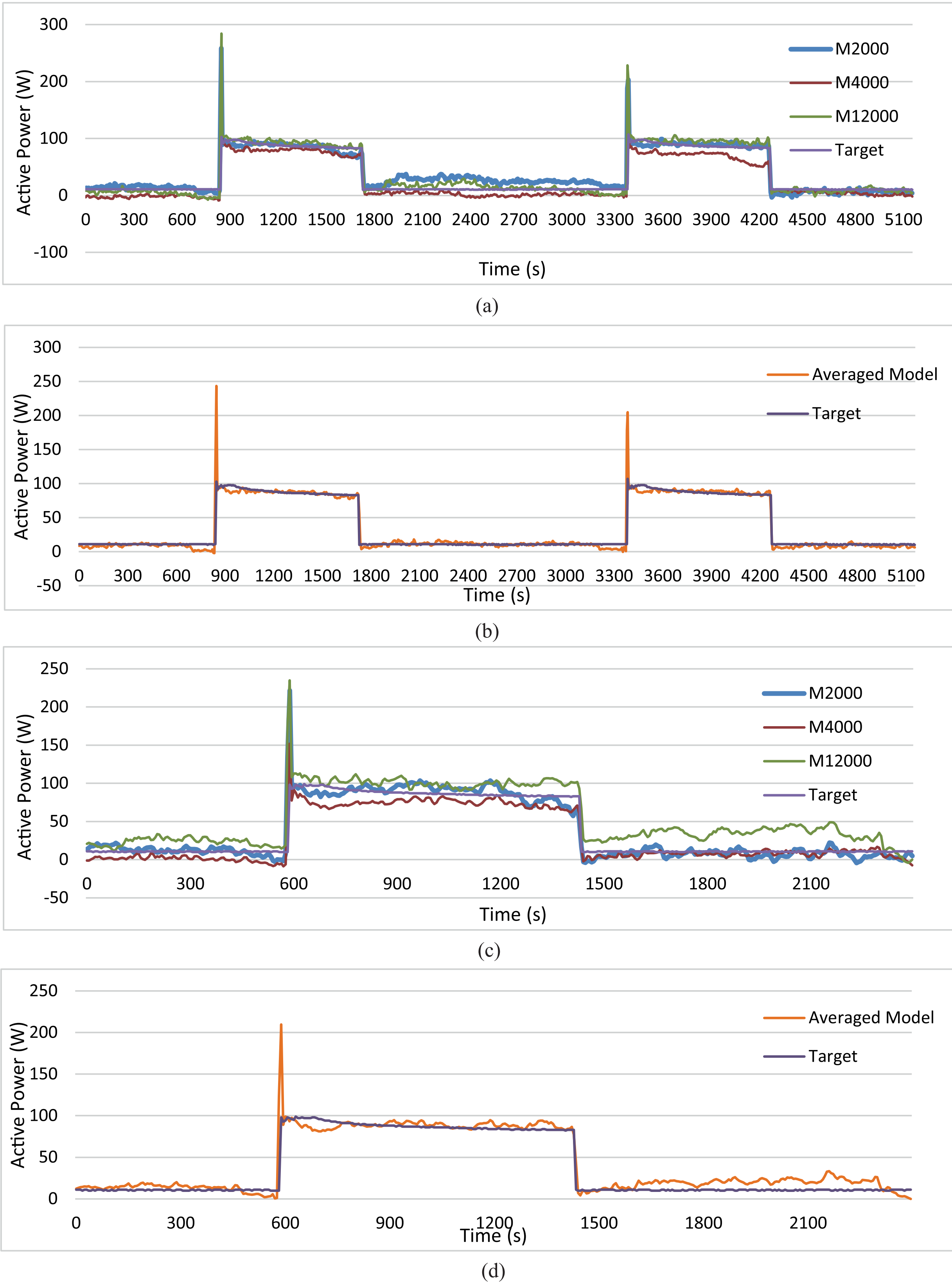

FIGURE 4. Performance evaluation of ensemble learning model with fridge as the target appliance. (a), (c) - Actual power consumption of target appliance and prediction by M2000, M4000, M12000 models. (b), (d) - Actual power consumption of target appliance and prediction by ensemble learning model.

TABLE 1. **Performance evaluation of the model with ON-OFF classifier and without ON-OFF classifier.**

	Model with ON-OFF classifier	Model without ON-OFF classifier
Lowest validation loss (mean squared error)	0.0117	0.0133

TABLE 2. **Performance evaluation of Wavenet model and sequence-to-point model.**

	Wavenet model	Sequence-to-Point model
Number of trainable parameters	45,194	41,998,449
Time taken to train for 1 epoch	1 second	226 second
Lowest validation loss (mean squared error)	0.0117	0.0186

Fig. 3 (a) and (b) represents a scenario where there is a change in the aggregate power consumption but the target appliance remains switched off for the whole period of time. It can be observed in Fig. 3 (b) that the model without the ON-OFF classifier wrongly predicts a power consumption value, while the model with the ON-OFF classifier gives a highly accurate prediction of nearly zero power consumption for the whole period. Fig3. (c) and (d) represents a scenario where the target appliance is switched off but consumes a small amount of power. In this scenario, the model with the classifier predicts with a higher accuracy compared to the model without the classifier. The predicted values are much smoother as well, without any random spikes. This clearly shows that the NILM model architecture with the ON-OFF classifier fed as an input to the regression stage performs significantly better. The comparison of validation loss of the two model versions can be seen in table 1.

Since sequence-to-point models perform better than sequence-to-sequence models [9], we conducted another experiment to compare the performance of sequence-to-point models with our wavenet model (including the ON-OFF classifier), which is a sequence-to-sequence model. The results are shown in Table 2.

It can be observed from the validation losses, that the Wavenet model is more accurate than the sequence-to-point model. An additional advantage of this architecture is that it can be trained much faster than the three-channel sequence-to-point model. This is because the number of trainable parameters in the Wavenet model is much less compared to the sequence-to-point model and also, many of the repetitive calculations are avoided. It must be noted that even though the Wavenet model must be trained for hundreds of epochs while the sequence-to-point model can be trained within about 50 epochs, the overall time taken to train the Wavenet model remains much lower.

C. ENSEMBLE LEARNING

The performance of ensemble learning was evaluated using the fridge and kettle as the target appliance. Prediction by neural networks trained for different number of epochs and the prediction by ensemble learning model for the fridge is shown in Fig. 4.

Fig. 4 illustrates two different scenarios where the target device is switched on for a certain period of time. The models M2000 and M12000 has random spikes (Fig. 4 (a)). M4000 does not have any random spikes but the error is comparatively high when the device is switched on (Fig. 4 (a)). M12000 has higher error when the device is off (Fig. 4 (c)). However, the ensemble model has higher accuracy regardless of whether the target appliance is switched on or off. It has some random spikes but it is comparatively smaller compared to the spikes generated by models M2000 and M12000. (Fig. 4 (b), (Fig. 4 (d)).

The model was trained for 2 epochs. The model weights were saved for every 2000 batches, and the model instances at each of these checkpoints were evaluated on the test data to get the validation loss. Different combinations of the model instances were evaluated and it was observed that the combination of model instances at batch 2000, 4000, and 12000 of the second epoch resulted in the best performance. These model instances are termed as M2000, M4000, and M12000 respectively.

It can be observed that the prediction of the averaged model is a combination of the predictions of the individual models. Each of the individual model instances chosen has unique characteristics and perform better in certain sections of the dataset and perform relatively worse in other sections. The result of averaging the weights is that the erroneous predictions of each of the individual models are balanced out by the other models. For the sections where each of the individual models performs very well, the averaged model also seems to perform well, even though the effect of the other models balancing out the effect of each individual model applies to this case as well. We can also notice that the predictions of the averaged model are less erratic and smoother.

This same process was done for the kettle model as well. The comparison of the metrics for the model instance with the lowest loss versus the averaged model can be seen in Tables 3 and 4. For both target appliances, there is a noticeable improvement in every single metric after the averaging process is done. Especially for the models trained with kettle as the target appliance, we can see a large improvement in

TABLE 3. **Performance evaluation of ensemble learning model with fridge as target appliance.**

	Mean Absolute Error	Estimation Accuracy	F1 Score
Best model instance	13.655	0.848	0.926
Ensemble model	12.278	0.864	0.937

TABLE 4. **Performance evaluation of ensemble learning model with kettle as target appliance.**

	Mean Absolute Error	Estimation Accuracy	F1 Score
Best model instance	13.655	0.848	0.926
Ensemble model	12.278	0.864	0.937

both MAE and EA metrics after our ensembling technique is applied. These improvements prove the effectiveness of our proposed ensembling technique.

VII. CONCLUSION

In this research, we have developed a deep learning-based model to perform non-intrusive load monitoring (NILM) on industrial and commercial buildings that have a 3-phase supply. Through experimentation, and implementation of innovative ideas, we were able to mitigate the following challenges related to disaggregation of both single phase and three-phase loads.

- Difficulty in determining the phase to which a single-phase load is connected.
- Single phase loads being drowned out by high power three-phase loads.
- Difficulty in detecting multiple similar loads and continuously operating loads.
- High computational cost to train a custom NILM model for each client building.

We have proven that our model can identify many appliances in such buildings with high accuracy. The experiments conducted have given promising results and proves that it can be practically implemented. Through our efforts to reduce the size of the neural network without compromising on prediction accuracy, the NILM model we developed gives many practical advantages if it is to be deployed in large real-world buildings.

Our NILM system has the capability to enable significant energy savings as well as intelligent data-centric algorithms for demand side management for electricity. After extensive testing, we believe that the model we have developed can be deployed and make significant impacts to the world. In the future, we plan to enhance our NILM model by integrating non-linear single-phase and three-phase loads into our model.

REFERENCES

[1] E. Aydin, D. Brounen, and N. Kok, "Information provision and energy consumption: Evidence from a field experiment," *Energy Econ.*, vol. 71, pp. 403–410, Mar. 2018, doi: 10.1016/j.eneco.2018.03.008.

[2] M. H. Albadi and E. F. El-Saadany, "Demand response in electricity markets: An overview," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Jun. 2007, pp. 1–5, doi: 10.1109/PES.2007.385728.

[3] M. A. Rahman, I. Rahman, and N. Mohammad, "Implementing demand response in the residential energy system to solve duck curve problem," in *Proc. 2nd Int. Conf. Adv. Inf. Commun. Technol. (ICAICT)*, Nov. 2020, pp. 466–470, doi: 10.1109/ICAICT51780.2020.9333449.

[4] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992, doi: 10.1109/5.192069.

[5] T. Bernard and N. Fuhr, "Non-intrusive load monitoring (NILM) combining multiple distinct electrical features and unsupervised machine learning techniques," M.S. thesis, Univ. Duisburg-Essen, Duisburg, Germany, 2018.

[6] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd, "At the flick of a switch: Detecting and classifying unique electrical events on the residential power line," in *UbiComp 2007: Ubiquitous Computing*. 2007, doi: 10.1007/978-3-540-74853-3_16.

[7] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ.*, Nov. 2015, pp. 55–64, doi: 10.1145/2821650.2821672.

[8] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," in *Proc. 32nd AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 1–8, doi: 10.1609/aaai.v32i1.11873.

[9] A. Meier and D. Cautley, "Practical limits to the use of non-intrusive load monitoring in commercial buildings," *Energy Buildings*, vol. 251, Nov. 2021, Art. no. 111308, doi: 10.1016/j.enbuild.2021.111308.

[10] C. Shin, S. Joo, J. Yim, H. Lee, T. Moon, and W. Rhee, "Subtask gated networks for non-intrusive load monitoring," in *Proc. 33rd AAAI Conf. Artif. Intell. (AAAI), 31st Innov. Appl. Artif. Intell. Conf. (IAAI) 9th AAAI Symp. Educ. Adv. Artif. Intell. (EAAI)*, 2019, pp. 1150–1157, doi: 10.1609/aaai.v33i01.33011150.

[11] J. Jiang, Q. Kong, M. D. Plumbley, N. Gilbert, M. Hoogendoorn, and D. M. Roijers, "Deep learning-based energy disaggregation and on/off detection of household appliances," *ACM Trans. Knowl. Discovery Data*, vol. 15, no. 3, pp. 1–21, Jun. 2021, doi: 10.1145/3441300.

[12] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," 2021, *arXiv:2104.02395*.

[13] J. Kelly and W. Knottenbelt, "The U.K.-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five U.K. homes," *Sci. Data*, vol. 2, no. 1, pp. 1–14, Mar. 2015, doi: 10.1038/sdata.2015.7.

[14] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. SustKDD Workshop*, no. 1, 2011, pp. 1–6.

[15] M. Wenninger, A. Maier, and J. Schmidt, "DEDDIAG, a domestic electricity demand dataset of individual appliances in Germany," *Sci. Data*, vol. 8, no. 1, p. 176, Jul. 2021, doi: 10.1038/s41597-021-00963-2.

[16] P. B. D. M. Martins, V. B. Nascimento, A. R. D. Freitas, P. B. E. Silva, and R. G. D. Pinto, "Industrial machines dataset for electrical load disaggregation," IEEE Dataport, Tech. Rep., Dec. 2018, doi: 10.21227/cg5v-dk02.

[17] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: A generative model for raw audio," 2016, *arXiv:1609.03499*.

[18] W. A. T. Wickramarachchi, P. H. Panawenna, J. Majuran, V. Logeeshan, and S. Kumarawadu, "Non-intrusive load monitoring for high power consuming appliances using neural networks," in *Proc. 3rd Int. Conf. Electr., Control Instrum. Eng. (ICECIE)*, Nov. 2021, pp. 1–5, doi: 10.1109/ICECIE52348.2021.9664681.

[19] B. Gowrienanthan, N. Kiruthihan, K. D. I. S. Rathnayake, S. Kumarawadu, and V. Logeeshan, "Low-cost ensembling for deep neural network based non-intrusive load monitoring," in *Proc. IEEE World AI IoT Congr. (AIIoT)*, Jun. 2022, pp. 551–556, doi: 10.1109/AIIoT54504.2022.9817165.

B. GOWRIENANTHAN received the B.Sc. degree (Hons.) in electrical engineering from the Department of Electrical Engineering, University of Moratuwa, Sri Lanka, in 2022. He is currently pursuing the master's degree in information technology with The University of Queensland, Brisbane, Australia. He was one of the authors of the paper titled "Low-Cost Ensembling for Deep Neural Network based Non-Intrusive Load Monitoring" (IEEE, 2022), which received the Best Paper Award in the category of Neural Networks from the 2022 IEEE World AI IoT Congress. His research interests include deep learning and the IoT technologies.

N. KIRUTHIHAN received the dual B.Sc. degree in electrical engineering and the B.I.T. degree from the University of Moratuwa, Sri Lanka, in 2023. From 2021 to 2022, he conducted research on 3-phase non-intrusive load monitoring utilizing deep learning technique and resulted to publish in IEEE. He is currently an Associate Software Engineer with Calcey Technologies, Colombo, Sri Lanka, one of a leading software development company. His research interests include artificial intelligence and deep learning. He was a recipient of the Best Paper Award from the IEEE AI IOT World Congress in Seattle, USA, in 2022, and the Best Project Award from the Department of Electrical Engineering, University of Moratuwa.

K. D. I. S. RATHNAYAKE is currently pursuing the B.Sc. degree in electrical engineering with the Department of Electrical Engineering, University of Moratuwa, Sri Lanka, and a minor in entrepreneurship with the Faculty of Business, University of Moratuwa.

She has contributed to the conference paper titled "Low-cost ensembling for deep neural network based on non-intrusive load monitoring." Her research interest includes improving the efficiency of high-power usage via a non-intrusive load monitoring technique. She was a recipient on the Best Paper Award in the category of neural networks from the IEEE AIIOT 2022.

S. KIRUTHIKAN received the B.Sc. degree in electrical engineering from the Department of Electrical Engineering, University of Moratuwa, Sri Lanka, in 2021, where he is currently pursuing the master's degree.

He is the first author of the paper titled "Pose estimation of a robot arm from a single camera." His research interests include deep learning, the IoT, robotics, and control.

V. LOGEESHAN (Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from the University of Peradeniya, Sri Lanka, in 2014, and the Ph.D. degree from North Dakota State University, USA, in 2019. He is currently a Senior Lecturer with the Department of Electrical Engineering, University of Moratuwa, Sri Lanka. He has published several research papers in leading international conferences and journals. His research interests include power systems, AI and the IoT applications, instrumentation, smart sensor development, and lab-on-a-chip development. He is a member of the ACM. He has received awards for his research contributions.

S. KUMARAWADU (Senior Member, IEEE) received the B.Sc. degree in engineering from the University of Moratuwa, Sri Lanka, in 1996, and the Ph.D. degree from Saga National University, in 2003. He is currently a Professor with the Department of Electrical Engineering, University of Moratuwa. His research interests include intelligent autonomous systems, system modeling and control, building automation and control systems, power systems stability and control, renewable energy systems, and artificial intelligence applications. He has published extensively in his field and received several awards for his contributions to research and teaching.

C. WANIGASEKARA (Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from the University of Peradeniya, Sri Lanka, in 2013, and the M.E. degree in electrical and electronic engineering and the Ph.D. degree in control engineering from The University of Auckland, New Zealand, in 2016 and 2020, respectively.

From 2020 to 2022, he was a Postdoctoral Researcher with the Centre for Industrial Mathematics, University of Bremen, Germany. He is currently a Control/Systems Engineer with the Institute for the Protection of Maritime Infrastructures, German Aerospace Centre, Bremerhaven, Germany. His current research interests include nonlinear system identification and control, machine learning, and networked control systems. He was a recipient of the Fowlds Memorial Prize for the most distinguished student in M.E. studies, in 2016.