A Machine Learning Approach for NILM based on Superimposed Current Profiles

A. M. Harsha S. Abeykoon
Department of Electrical Engineering
University of Moratuwa
Moratuwa, 10400, Sri Lanka
harsha@uom.lk

A. Paveen Senura Perera
Department of Electrical Engineering
University of Moratuwa
Moratuwa, 10400, Sri Lanka
paveenperera@gmail.com

R. K. Sanjeewika Prabodhanie
Department of Electrical Engineering
University of Moratuwa
Moratuwa, 10400, Sri Lanka
sanjirathnayake@gmail.com

M. D. N. V. Matharage
Department of Electrical Engineering
University of Moratuwa
Moratuwa, 10400, Sri Lanka
nisalvm@gmail.com

A. P. Abeysinghe
Department of Electrical Engineering
University of Moratuwa
Moratuwa, 10400, Sri Lanka
asitha169gmail.com

***Abstract*—This research focuses on identifying a new implementation of a machine learning approach for Non-intrusive load monitoring (NILM). We mathematically superimpose current profiles of individual appliances and compare against the actual combinational current profiles. This simple yet effective method is tested on combinations of 6 household devices in a typical low voltage residential installation and the high accuracy of correct identification confirms the proposed method is feasible. The proposed method eases the burden of the training phase which is considered as an inherent limitation of all supervised deep learning NILM models. We deploy the method on a Raspberry Pi 3 providing a solution to increase the scalability of NILM**

Keywords— Power signature analysis, Machine learning, NILM, Load identification

I. INTRODUCTION

The world is facing a challenging situation in meeting the energy demand due to the gradual depletion of fossil fuel reserves and increased threat on global warming caused by emissions. Research indicates that if the households can intelligently organize the usage of appliances in the house that would reduce the energy consumption by about 5% to 15% [1].

Secondly Demand Side Management (DSM) programs which apply more efficient Demand Response (DR) schemes to a power grid can play a significant role in bringing down the peak demand, thus reducing the capacity cost of the system. However, an average resident does not have sufficient expertise to keep detailed records of all of the electrical parameters.

Both these cases require the total energy consumption of a household to be disintegrated into device specific energy consumption allowing customers to make substantial savings in electricity consumption as well as allowing the Distribution System Operators (DSOs) to make decisions on activity basis (e.g. cleaning or cooking) and convert consumers to become more active elements of the Smart Grid [2].

Out of the load disaggregation schemes, the most feasible option appears to be Non-Intrusive Load Monitoring (NILM) where load measurements are performed by a single metering device installed at the main feeding panel [3] and not at the appliance or at each socket [4].

The topic of NILM has been widely discussed for more than three decades starting from G.W. Hart [3] considering only the active power. An appliance turning ON or OFF is detected when the difference in the power consumption exceeds a predefined threshold. The shortcoming is that the method can only be used to disaggregate ON/OFF state appliances and cannot distinguish appliances with similar active power. Later in research by combining reactive power and active power such devices were also identified. Further it failed to identify appliance in which the power rating is below 150W. Followed by this research NILM branched out primarily according to the frequency range of measurement a) Low frequency (1Hz or below) b) mid frequency (kHz range) c) high frequency (MHz range) [5].

A complete new high frequency feature is introduced by Patel et al with the examination of the noise of the voltage signal [6] [7].Studying these electromagnetic interferences (EMI) need high frequency signal acquisition and such measuring devices in a domestic setting are too costly and prone to error. Superimposing of EMI signals, effect of nearby houses and effect of radio signal interferences for the measurement have not yet been determined. In contrast, mid frequency provides the best compromise between the cost and the accuracy. In mid frequency NILM, event based and eventless methods are proposed [5]. However, these event based methods suffer from high complexity in event detection. Computational intensity in labeling the actual combination and further complex device clustering imposes a bottleneck in deploying these methods in an embedded platform in real time. In contrast, using machine learning methods to eventless NILM has been identified as an efficient way and the latest trend.

Keeping the first step towards implementing NILM in an embedded platform A.K Jain [8] has proposed a current peak based simplified method but it has not been successfully tested for real time load disintegration and establishing the ground truth for the supervised learning in a practical household has not been revealed. The closest inspiration to this paper is found in the work of E. P. Loukas et al who has successfully used a machine learning approach and proven that higher harmonic current features in the FFT domain make the device combinational classification accuracies higher [2]. Attempting to extend the concept to a separate new house this paper has reported a declining accuracy. Thus generalizing the concept for exponentially increasing number of combinations and multiple houses still remains arguable. Following that research we built the hypothesis that the harmonic features visible in FFT domain should be visible in time domain signals and linear combinations of time domain signals can artificially create a large enough dataset with all possible combinations to successfully train a supervised deep learning modal resulting a practically implementable solution for NILM. The originality of this research remains with mathematically superimposing the current profiles of individual devices to create the combinational waveforms and validating the similarity, randomizing the capture window of the steady state current profiles of individual devices to generate a larger training dataset, validating the deep learning model with experimental data and deploying the deep learning modal in an embedded system, allowing to identify device combinations real time, finally achieving benchmark accuracies in comparison to the results in [2] using the same deep learning library.

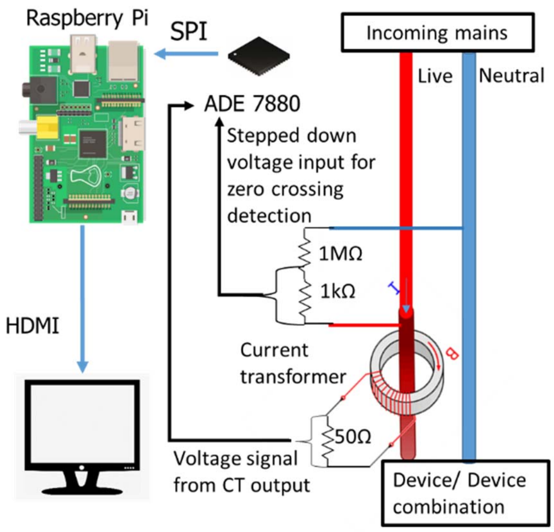


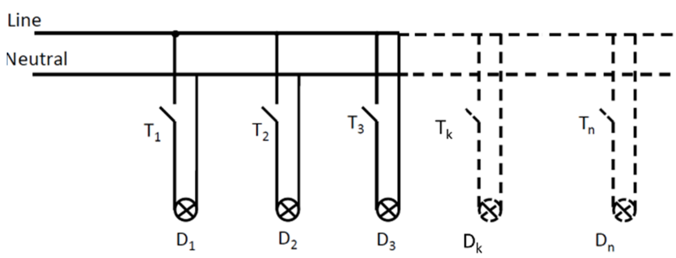
Fig. 1. Measurement harware setup

II. Measurement

All data in this work has been recorded by measurements via a customized measurement set-up as shown in Fig.1 in a LV residential installation of supply frequency 50Hz and 230V. This consists of a current transformer (measuring range ±10A with accuracy ±1%,). Digital conversion of the current transformer output is through, ADE7880 energy measurement chip from which the full range signal corresponding to 3.53A is divided into 23 bit allowing the sensitivity to be 50µA. This data is read through a Raspberry pi 3 at 3kHz allowing to capture up to 50th harmonics as for the Nyquist theorem. Four hundred data points are acquired in the steady state for duration of 133ms allowing the frequency domain resolution to be 7.5Hz. For the purpose of validating the method we used data from Pico scope model 3204A (range ±2V per division).

III. Modeling

For the purpose of illustrating the methodology in obtaining the superimposed waveforms, the normalized amplitude measurement in time domain is modeled by the notation D_k^q.



$$D_k^q(r) \rightarrow k \text{ - \{Device number(1-n)\}}$$
$$q \text{ - \{Capture window\}}$$
$$r \text{ - \{Data point number (1-400)\}}$$

Fig. 2. Naming convention for the device combinations

As shown in Fig. 2 let D denote a device and k is the device number.

$D_k^{\forall q} \rightarrow$

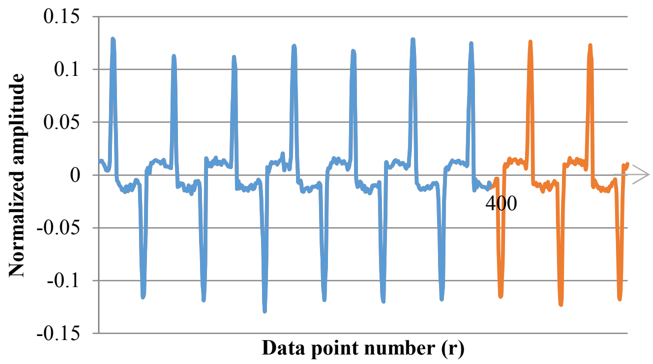


Fig. 3. Steady state current profile of a monitor.

$D_k^q \rightarrow$

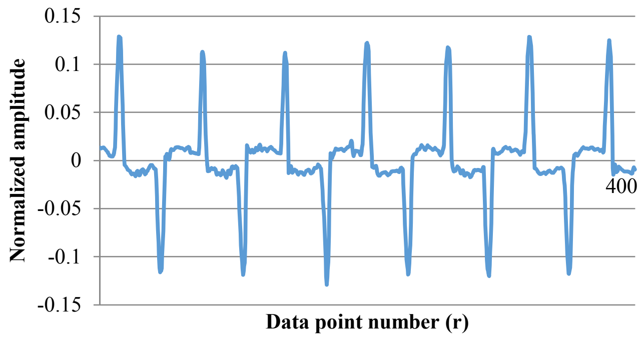


Fig. 4. Graphical representation of the random capture window of the steady state current profile.

D_k^q is a random capture window triggered at zero crossing (Fig. 3) from the steady state current profile $D_k^{\forall q}$ (Fig. 2) of the kth device. Thus r denotes a momentary, relative time instance in q. The captured window q contains 400 data points belonging to the k^{th} device that can be represented in a matrix as shown in (1)

$$D_k^q = [D_1^q(1) \;\; D_1^q(2) \;\; D_1^q(3) \ldots\ldots \; D_1^q(400)]_{1\times 400} \quad (1)$$

Assuming that the house has n number of devices, a collection of (1) with randomized q, is represented by D as shown in (2).

$$D = \left[D_1^{\{q\}} \;\; D_2^{\{q\}} \;\; D_3^{\{q\}} \;\ldots\ldots\ldots\ldots\ldots\ldots\; D_n^{\{q\}}\right]_{1\times n} \quad (2)$$

As shown in (3), a binary matrix which denotes the ON/OFF state of the device mix at a certain instance is represented by S. If the k^{th} appliance is ON S_k=1 otherwise S_k = 0.

$$S=[S_1\ \ S_2\ \ S_3\ \ \ldots\ldots S_k\ \ldots\ldots S_n]_{1\times n} \tag{3}$$

α is defined by converting the consecutive decimal numbers from 1 to 2^n-1 to binary and represented as shown in (4) where n is the number of devices.

$$\begin{array}{cc} & \begin{array}{cccc} 2^0 & 2^1 & \ldots & 2^n \end{array} \\ \alpha= & \begin{bmatrix} 1 & 0 & . & 0 \\ 0 & 1 & . & 0 \\ 1 & 1 & : & 0 \\ . & . & . & . \\ . & . & . & . \\ 1 & 1 & \cdot & 1 \end{bmatrix}_{n\times(2^n-1)} \end{array} \begin{array}{cc} \leftarrow & 1 \\ \leftarrow & 2 \\ \leftarrow & 3 \\ & . \\ & . \\ \leftarrow & (2^n-1) \end{array} \tag{4}$$

Let i & j denote rows and columns of a matrix respectively. Equation (5) is a representation of a row in α being equal to an instantaneous device combination S. Thus allowing α to contain all possible combinations of devices.

$$S=[\alpha_i^1\ \ \alpha_i^2\ \ .\ .\ \alpha_i^j\ \ .\ \ \alpha_i^{n-1}\ \ \alpha_i^n\ ]_{1\times n} \tag{5}$$

The combination waveform is obtained as shown in (6).

$$C=D[\alpha]^T$$

$$=\left[D_1^{\{q\}}\ D_2^{\{q\}}\ \ D_3^{\{q\}}\ldots\right]_{1\times n}\begin{bmatrix} \alpha_1^1 & . & \alpha_n^1 \\ \alpha_1^2 & . & \alpha_n^2 \\ . & . & . \\ \alpha_1^j . & \alpha_i^j & \\ . & . & . \\ \alpha_1^{2n-1} & . & \alpha_n^{2n-1} \end{bmatrix}_{(2^n-1)\times n} \tag{6}$$

$$=\left[\sum_{j=1}^{n} D_j^{\{q\}}.\alpha_1^j \qquad \sum_{j=1}^{n} D_j^{\{q\}}.\alpha_2^j \qquad \ldots\ldots \qquad \sum_{j=1}^{n} D_j^{\{q\}}.\alpha_n^j\right]$$

If U is the number of random D matrices we make, let L (7) be a multi-label dataset consisting U x $(2^n\text{-}1)$ multi-label examples where X_i (8) considered as the features and Y_i (9) as labels.

$$L=\{(X_i,Y_i)\}_{i=1}^{U\ x\ (2^n-1)} \tag{7}$$

$$X_i=[C]^T{}_i \tag{8}$$

$$Y_i=\alpha_i \tag{9}$$

During the training phase we build function h:X→Y & we obtain the prediction Z_i:

$$Z_i=h(X_i) \tag{10}$$

During the validation we compare with the corresponding Y_i(9).

Extending the time domain analysis into the frequency domain the Fast Fourier transform is utilized to convert X_i into $f(X_i)$ and converted values are used in deep learning as features.

IV. Data Set

In building the dataset according to the proposed methodology, n=5 and n=6 cases taken into consideration out of the following devices shown in the Table I.

During the scope of the research, only one state out of multistate devices has been considered in defining the device mix. The selection of devices was based on covering all types of typical household devices, less than 150W power rating and for the purpose of comparing with results from [2].

TABLE I. Summary Of The Recorded Devices

Device	Power Rating	State	Type
CFL bulb	17.6W PF - 0.9159	ON/OFF state	Electronic ballasts
Fan	18.34W PF-0.5869	Multi state (SP1, SP2,Off)	Inductive load
Monitor	15.05 W PF - 0.9794	Multi state (Displaying, Stand by, OFF)	Switch mode power supply
Laptop	42.1 W PF - 0.9952	Multi state (charging /Idling/Off)	Switch mode power supply
Heater	34.13 W PF - 1	ON/OFF state	Pure Resistive
Phone	8.09W PF - 0.9929	ON/OFF state	Switch mode power supply
Hair Dryer	104.7W PF-1	Multi state (SP1/SP2/SP3 /OFF)	Pure Resistive

Prior to the training phase, data preprocessing was done by mathematically combining the waveforms of five devices. We repeat (2) for 1200 randomized D matrices. Thus U=1200 means that for any given combination, the dataset includes 1200 samples. This dataset is uploaded and available online [9]. Size is selected based on the best performance accuracies in a 70%-30% train-test split validation and also to maintain the computational time to the minimum. A multi-dimensional matrix of 1200x31 is being used as a feature matrix (8) of the deep learning model and corresponding ON/OFF state combinations used as labels (9). For validating purposes of the deep learning modal, we capture 120 samples each, of the actual combinations using the measurement setup and label them with their actual combination manually and create a validation dataset.

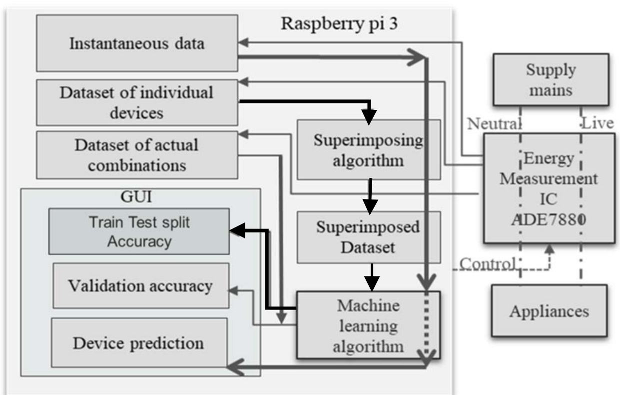


Fig. 5. Design Architecture

The entire functionality and the implementation of the proposed method for a practical household could be elaborated as shown in Fig. 5

As shown in Fig. 6 the measuring device generates device specific steady state current signatures and the dataset is captured with a trigger at zero crossing. The addition of the two waveforms in Fig. 6 is the resultant superimposed current waveform illustrated in Fig. 7. The actual combinational waveform is also plotted in the same axis which signifies the similarity of the two signals and thus provides a visual interpretation of the feasibility of the proposed method of superimposing. In the Fig. 8 the cross correlation between the actual waveform data and the mathematically combined waveform data of the monitor and heater is presented. The R^2 value of 0.93 confirms that the data is nearly equal.

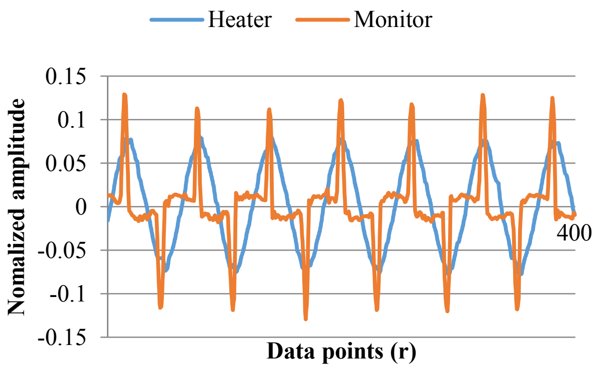


Fig. 6. Waveforms of devices obtained from the measurement setup

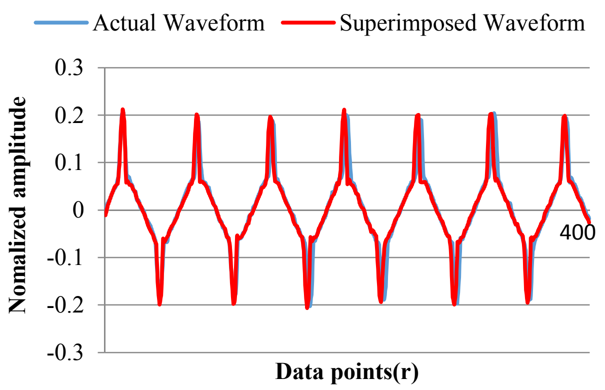


Fig. 7. Actual and superimposed waveform of Monitor and Heater combined

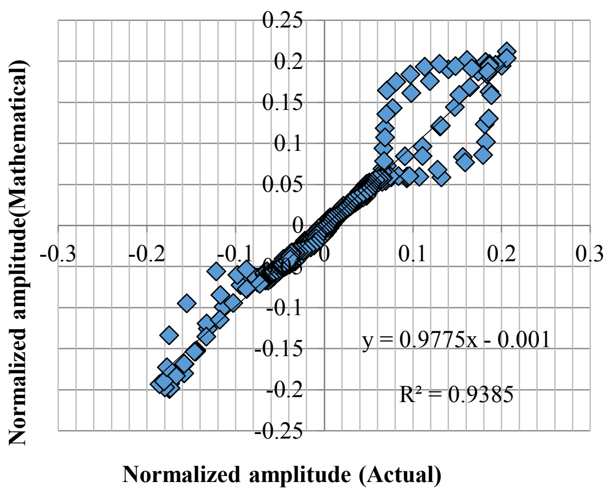


Fig. 8. Crosscorellation between actual and superimposed data

Further examining the Fig. 8 it is evident that the non-uniform sampling of the measuring device and the frequency fluctuations of the supply current imposes a shift in the time domain waveform thus acquiring the waveform for longer time duration and superimposing the waveform can lead into errors. This can be identified as the reason for the hypothetical base of converting the signal to the frequency domain expecting a higher accuracy.

V. Machine Learning

This research is based on a multi label classification. The dataset is used with the scikit-multilearn library [10] and using the standard evaluation matrices results were obtained. There are two main methods for tackling a multi-label classification which can be identified as problem transformation method and algorithm adaptation method. Both these methods are tested with the experimental data.

Problem transformation methods transform the multi-label problem into a set of binary classification problems, which can then be handled using single-class classifiers. This method resulted a higher computation time whereas algorithm adaptation methods adapt the algorithms to directly perform multi-label classification. In our current study we tested seven different standard binary classifiers namely Logistic Regression, Random Forest, Stochastic Gradient Descent, Naive Bayes, K-Nearest Neighbor, Support Vector Machines, Decision Trees.

The binary classifier which resulted the fastest computation of 15 seconds and best classification is identified as Binary Relevance K nearest neighbors' classifier and results are presented below. Following are brief explanation on the tested models with use of the Sci-kit multi learn library [10]

A. *Binary Relevance Schemes*

This can be identified as the simplest among all the schemes. As explained above, the original problem is decomposed into N no of sub-classifiers. These classifiers will individually produce 0 or 1 in output binary string for a data instance

B. *Classifier Chain Schemes*

Sometimes, class labels are interdependent. Just like Binary Relevance, the first N no of response variables for each class will be created. The output of one response variable for a sub-classifier will be used as an extra feature in the next sub-classifier.

The output from each classifier will be captured like the Binary Relevance Scheme and will determine class labels at the end.

C. *Label Power-set Schemes*

The fundamental concept of the Binary Relevance & Classifier chain is more or less the same. Label power-set works in a different way. It considers each combination of labels in the training dataset as a separate label. In general, a class-space of dimension N, there can be 2^N no of total possible label combinations. It does not decompose into any sub problems, but it directly predicts the combination of class labels as a whole.

VI. Evaluation Matrics

In traditional classification such as multi-class problems, accuracy is the most common evaluation criteria. Additionally, there exists a set of standard evaluation metrics that includes precision, recall, F-measure, and ROC area defined for single label multi-class classification problems [11]. However, in multi-label classification, predictions for an instance is a set of labels and, therefore, the prediction can be fully correct, partially correct (with different levels of correctness) or fully incorrect. None of these existing evaluation metrics capture such notion in their original form. This makes evaluation of a multi-label classifier more challenging than the evaluation of a single label classifier. Following matrices are defined to overcome the notion of partial correctness.

A. *Exact Match Ratio (MR)*

Evaluation of a multi-label classification algorithm is difficult mostly because multi-label prediction has an additional notion of being partially correct. One trivial way around would be just to ignore partially correct (consider them as incorrect) and extend the accuracy used in a single label case for multi-label prediction. This is called the Exact Match Ratio.

$$\text{Exact Match Ratio, MR}=\frac{1}{n}\sum_{i=1}^{n} I(Y_i=Z_i) \qquad (11)$$

Where, I is the indicator function. While Y_i represents the labels as described in (9) Z_i represents the predicted labels (11). A disadvantage of this measure is that it does not distinguish between completely incorrect and partially correct result which might be considered as harsh. In order to account for partial correctness, Godbole et in [12] proposed a new set of definitions for accuracy, precision, recall, and F1 measure. As in single label multi-class classification, the higher the value of accuracy, precision, recall and F1- score, the better the performance of the learning algorithm.

B. *Hamming Loss (HL)*

Hamming Loss reports how many times on average, the relevance of an example to a class label is incorrectly predicted. Therefore, hamming loss takes into account the prediction error (an incorrect label is predicted) and the missing error (a relevant label not predicted), normalized over total number of classes and total number of examples

Hamming Loss, HL

$$=\frac{1}{n(2^n-1)u}\sum_{i=1}^{(2^n-1)u}\sum_{n=1}^{n}[I(l\in Z_i\cap l\notin Y_i)+I(l\notin Z_i\cap l\in Y_i)] \qquad (12)$$

Where I is the indicator function. Ideally, we would expect hamming loss, HL = 0, which would imply no error; practically the smaller the value of hamming loss, the better the performance of the learning algorithm [13].

Label based measures evaluate each label separately and then averages over all labels. Therefore, any known measure, used for evaluation of a binary classifier (e.g. accuracy, precision, recall, F1, ROC etc.), can be used here. Any of these scores can be computed on individual class labels first and then averaged over all classes. In contrast, they can be computed globally over all instances and all class labels. The first one is called macro averaging and the second one is called micro averaging [14]. It is important to note here that by definition, macro averaged F1 would be more affected by the performance of the classes which has fewer examples and in contrast, micro averaged F1 would be more affected by the performance of the classes which has more examples [15]. But during this research for all combinations we use 1200 examples thus the number of examples are balanced for all classes making the accuracies better

VII. Results and Discussion

The overall accuracy and the hamming loss obtained for each device combination through the BrKNN model is as shown in Table II. The prediction time represents timing after the deep learning modal is trained completely, the time taken to capture an instantaneous waveform, pre-process it and predict the combination.

TABLE II. Tested Device Combination, Corresponding Accuracy Matrices and Response Times

Device combination	Exact match ratio	Hamming loss	Prediction Time
CFL bulb + Monitor + Laptop + Heater + Phone	84.82%	0.034	14.8(s)
CFL bulb + Monitor + Laptop + Heater + fan	96.61%	0.0087	15(s)
CFL bulb + hair dryer + Laptop + Heater + Phone	76.58%	0.046	15.4(s)
CFL bulb + Monitor + Laptop + Heater + Phone + fan	70.39%	0.083	15.8(s)
CFL bulb + Monitor + Laptop + Heater + Hair dryer + fan	80.3%	0.003	15.5(s)

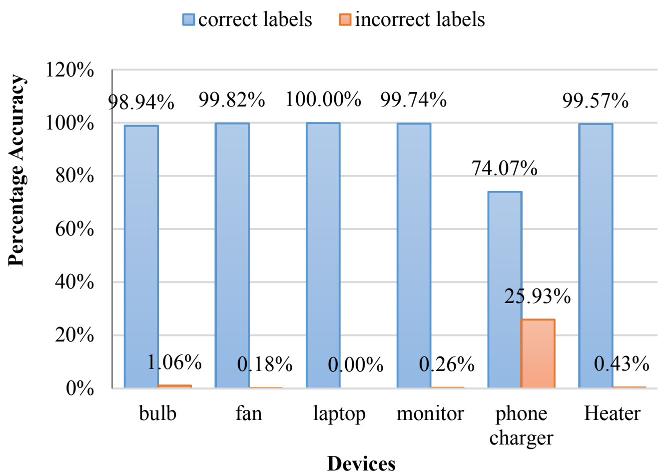


Fig. 9. Percentage identification accuracies of the individual devices

As illustrated from Fig. 9 the relatively high number of incorrect identifications of the phone denotes the range of measuring in the measuring device imposes a significant impact on the device disintegration. The results summary of [2] is included in Table III. In comparison with Table IV which represents the overall precession, recall and F1 scores of our research, showcases a significant improvement. Bearing in mind that the same deep learning library has been used and the number of devices is being increased up to six, using real time, 3kHz sampling data, as opposed to 50kHz non real time data and commercial software for analyzing waveforms it is safer to claim our approach has comfortably outnumbered the performance of [2].

To analyze the effect of extended measuring range and effect of high-power devices in the product mix the results in Table II for device combinations 2 and 3 could be compared. The phone results in a low accuracy whereas when replaced by a fan increases the accuracy. This increase in accuracy can be attributed to the existence of a significant feature in the measuring range due to the higher amplitude of the current consumption of the fan. By adding more devices, the time taken for training phase increases which is expected due to exponentially increasing number of combinations but the accuracies remain better as long as all the current signatures are noticeable in the measuring range. Further prediction times are not affected drastically and still remain in average range of 15 seconds even at a limited computation power of a Raspberry pi 3 and prediction time does not display any significant pattern relative to the device combination.

TABLE III. SUMMARY OF A MACHINE LEARNING APPROACH FOR NILM BASED ON ODD HARMONIC CURRENT VECTORS [2]

Device combination	Value		
	Precision	***Recall***	***F1-score***
TV with DVD + Laptop	0.667	0.999	0.800
TV with DVD + Laptop + Hairdryer	0.851	0.845	0.847

TABLE IV. RESULTS OBTAINED FROM THE PROPOSED METHOD

Device combination	Value		
	Precision	***Recall***	***F1-score***
CFL bulb + Monitor + Laptop + Heater + Phone (macro avaraged)	0.97	0.96	0.97
CFL bulb + Monitor + Laptop + Heater + Phone (micro avereged)	0.97	0.95	0.96

TABLE V. COMPARISON OF ACCURACY IN FREQUENCY AND TIME DOMAIN FOR COMBINATION 1 IN TABLE 1.

Type	Training combinations manually		Proposed method	
	Frequency domain	***Time Domain***	***Frequency domain***	***Time Domain***
Oscilloscope data	**98.38%**	97.58%	***91.53%***	*8*6.37%
Device data	**97.93%**	94.70%	***75.47%***	70.39%

For the purpose of evaluating the deep learning modal we trained for all combinations manually, the time domain data gave accuracy close to 95% for all combinations which again is a better performance than [2]. Those results are elaborated in the Table V. When the proposed method of superimposing waveforms also converted to FFT domain the accuracies increased slightly confirming that better results can be obtained by extending the proposed method into the frequency domain.

VIII. CONCLUSIONS

This paper concludes that superimposed device signatures can provide accurate predictions for the combined device signatures and thus be a successful NILM method and it can be successfully deployed in an embed platform. Proving the significance of this method when the data of 5 devices given into the scikit multi-learn machine learning library, based on a Multi label classification BrKNN it has recorded an overall highest accuracy of 96% . However, it is evident the range of the measuring device and the device combination in consideration as well as the frequency of the measurement imposes a significant impact in device identification. Comparing with the results of [2] proposed method is a significant improvement. Measurements performed on a LV installation and an overall classification accuracy over 70% has been obtained for 6 devices thus it confirms that this is extendable to a higher number of devices. The frequency domain analysis improves the accuracies slightly due to the fact that it reduces the impact of sampling errors and zero crossing detection errors. In conclusion it is safe to admit that superimposed current signatures lead to a very successful load disintegration scheme and could be deployed on an embedded platform which will predict the results real time under 15 seconds.

REFERENCES

[1] S. Darby,"The effectiveness of feedback on energy consuption:A Review for DEFRA of the Literature on Metering, Billing and Direct Displays," *Environmental Change Institute, University of Oxford, Oxford 2006.*

[2] E. P. Loukas et al., "A Machine Learning Approach for NILM based on Odd Harmonic Current Vectors," *2019 8th International Conference on Modern Power Systems (MPS)*, Cluj Napoca, Romania, 2019, pp. 1-6, doi: 10.1109/MPS.2019.8759666.

[3] G. W. Hart, "Nonintrusive appliance load monitoring," in *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870-1891, Dec. 1992, doi: 10.1109/5.192069.

[4] A. Ridi, C. Gisler and J. Hennebert, "A Survey on Intrusive Load Monitoring for Appliance Recognition," *2014 22nd International Conference on Pattern Recognition*, Stockholm, 2014, pp. 3702-3707, doi: 10.1109/ICPR.2014.636.

[5] T. Bernard and N. Fuhr, "Non-Intrusive Load Monitoring (NILM) combining multiple distinct electrical features and unsupervised machine learning techniques," *thesis, University of Duisburg-Essen,* Duisburg, 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 (Nominated for the Best Paper Award)," *UbiComp 2007: Ubiquitous Computing Lecture Notes in Computer Scence,* pp. 271–288.

[7] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, and S. Patel, "Disaggregated End-Use Energy Sensing for the Smart Grid," *IEEE Pervasive Computing,* vol. 10, no. 1, pp. 28–39, 2011.

[8] A. K. Jain, S. S. Ahmed, P. Sundaramoorthy, R. Thiruvengadam and V. Vijayaraghavan, "Current peak based device classification in NILM on a low-cost embedded platform using extra-trees," *2017 IEEE MIT Undergraduate Research Technology Conference (URTC)*, Cambridge, MA, 2017, pp. 1-4, doi: 10.1109/URTC.2017.8284200.

[9] Paveen, "Superimposed current profiles," Kaggle, 18-Jun-2020. [Online].Available:https://www.kaggle.com/paveenperera/superimposed-current-profiles. [Accessed: 18-Jun-2020].

[10] P. Szymański and T. Kajdanowicz, "A scikit-based Python environment for performing multi-label classification," *Journal of Machine Learning Research 1(2016)*, vol. 20, Feb. 2017.

[11] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. Available: 10.1016/j.patrec.2005.10.010.

[12] S. Godbole and S. Sarawagi, "Discriminative Methods for Multi-labeled Classification," In *Proceedings of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2004)*, pp. 22–30, 2004.

[13] M. S. Sorower, "A literature survey on algorithms for multi-label learning," *Oregon State University, Corvallis*, vol. 18, pp. 1–25, Dec. 2010.

[14] Y. Yang, "An Evaluation of Statistical Approaches to Text Categorization," *Information Retrieval 1*, no. 1999, pp. 69–90, Apr. 1999.

[15] L. Tang, S. Rajan, and V. K. Narayanan, "Large scale multi-label classification via metalabeler," *Proceedings of the 18th international conference on World wide web*, pp. 211–220, Jan. 2009.