Appliance Identification in NILM Applications by means of a Convolutional Auto-Encoder

Laura de Diego-Otón, Álvaro Hernández, Daniel Pizarro
Electronic Department
University of Alcalá)
Alcalá de Henares (Madrid), Spain
laura.diego@uah.es

Rubén Nieto
Electronics Technology Department
Rey Juan Carlos University
Móstoles (Madrid), Spain

***Abstract*—In energy efficiency applications, Non-Intrusive Load Monitoring techniques (NILM) are typically used to deduce which electrical loads are being used in a building at a given time. The identification of household appliances, in particular manually operated ones, is relevant information that can also be applied to infer the routines of tenants in Active and Assisted Living environments (AAL). These tools and applications are becoming increasingly interesting, especially in Western countries, where the ageing population is putting a strain on public social and health services. In this context, this work aims to classify the on/off events of the devices considered in the BLUED database. For this purpose, an architecture is presented, consisting of a Convolutional Auto-Encoder (CAE) followed by a classifier neural network. The CAE is used to implement a dimensionality reduction process after the encoder. Input data are formatted as images, created with extracted sections of the high-frequency electric current signal captured around the switching events. It is noteworthy that this dimensionality reduction also allows a decrease in the computational load of the classifier. Regarding the CAE functionality, the reconstruction error reaches a value of $1.579 \cdot 10^{-3}$, whereas in the validation stage a weighted average classification F1-score of 87 % is obtained for the whole architecture.**

***Index Terms*—Non-Intrusive Load Monitoring (NILM), Convolutional Auto-Encoder (CAE), Appliance Identification, Active and Assisted Living (AAL).**

I. Introduction

The concept of Active and Assisted Living (AAL) [1] is defined as the use of novel technologies to create a supportive environment, whose purpose is to improve the independent life of the elderly and people with mild cognitive impairments, allowing them to remain active in society for longer. Within these technologies, Remote Patient Monitoring (RPM) [2] is a key aspect. The functionality of these systems is to provide information on the health state of patients in their own homes or even in remote areas. Furthermore, they increase access to medical care whereas decreasing its cost and improving the detection of deterioration by making it faster.

This work was supported in part by the Spanish Ministry of Science, Innovation and Universities MCIN/AEI/10.13039/501100011033 (PoM project, PID2019-105470RA-C33, ALONE project, ref. TED2021-131773B-I00, and ATHENA project, ref. PID2020-115995RB-I00), by the Comunidad de Madrid (RACC project, ref. CM/JIN/2021-016) and by the University of Alcalá (FPU-UAH scholarship).

Within the methods to develop sustainable monitoring systems for social and medical care, there are direct and indirect methods. Direct monitoring is when the state of health is deduced directly from the biomechanical and/or physiological variables that are measured through the use of Body Sensor Networks (BSN) [3]. However, their intrusive nature makes them less attractive. On the other hand, indirect monitoring seeks to obtain information about the patient's health through other parameters that are not related to biomedical variables and, hence, are less or even non-intrusive. An illustrative example of these indirect methods is the Non-Intrusive Load Monitoring (NILM) [4] techniques, whose initial objective is to identify the electrical loads by analysing the aggregate consumption of a building or household. This information may be also used for RPM, although it is important to highlight that in these scenarios the number of factors involved is large and it will be necessary to implement other types of evaluations to know the real patient's health state.

To measure the degree of dependency or health status, the term Activities of Daily Living (ADL) [5] is widely used. It refers to the fundamental skills that a person must show to care for herself/himself independently. In this way, NILM techniques allow information to be obtained about the activities that involve the use of electrical appliances. More precisely, typical activities that might be monitored are cooking, housekeeping, laundry or home entertainment activities, among others. It should be mentioned that some of them are defined as Instrumental Activities of Daily Living (IADLs) [6].

To identify the appliances, the switching events should be detected and classified. Currently, a large number of works in this field are focused on using Artificial Neural Networks (ANNs). As their name suggests, they are a structure that seeks to mimic how biological neurons connect to each other. These neurons are arranged in layers, where they work together such that each unit processes a specific input, using a particular function, and the extracted information is sent to the next neuron. According to different applications, it is possible to find diverse topologies in the definition of ANNs: Recurrent Neural Networks (RNNs) include feedback connections, which make them useful for processing sequential or time data series;

Convolutional Neural Networks (CNNs) use linear algebra principles to identify patterns within an image, such as matrix multiplication; and finally, Auto-Encoders (AEs) work in such a way that the decoder aims to reconstruct the input from a reduced dimensionality encoding created by the encoder. For example, in [7] it is described a 1-D convolutional combined with a long short-term memory RNN for a load disaggregation application, and in [8] authors proposed a structure based on the one-dimensional Convolutional Auto-Encoder.

About ANNs, there are two approaches depending on the learning process: supervised and unsupervised learning [9]. During the training process, the supervised algorithms have information about the correct result and make adjustments to their configuration based on the comparison of the obtained and expected results. However, the alternative topologies (unsupervised) learn on their own without information about the expected result. The baseline ANNs with unsupervised learning are the AEs. Considering the above, they are able to modify their structure by adjusting the parameters according to the reconstruction error.

Considering this context, this work presents a neural network architecture focused on identifying different appliances by using as input the switching events on the electrical current signal sampled at high frequency (in the range of kHz). Input data pre-processing includes the transformation of the temporal signal into images for the proposed convolutional layers. After that, the main structure comprises a first stage for the reduction of dimensionality and a final classification stage. The proposal has been verified thanks to the recorded high-frequency signals obtained from Building-Level fUlly labelled Electricity Disaggregation (BLUED) database [10] which achieved an experimental classification performance of 87 %. The main contribution of this work is the definition of an architecture for the identification of household loads in order to be used as a starting point for RPM. This architecture includes a CAE for the prospective creation of a system capable of solving the problem of classifying in an unsupervised method. In connection with this part of the model, a value of $1.579 \cdot 10^{-3}$ for the maximum reconstruction error in the CAE is obtained. The rest of the manuscript is structured as follows: Section II describes the proposed architecture; Section III presents its performance making use of the experimental data; and, finally, conclusions are discussed and possible future developments are commented in Section IV.

II. Proposed Architecture

The proposed architecture consists of a first pre-processing stage, where the input samples from the electric current are prepared, a second stage where the CAE plays a leading role, and a final stage for appliance identification, where the intermediate input samples are processed by a plain neural network classifier. The details for each stage are specified thereupon.

A. Pre-Processing Stage

Firstly, the sampling frequency of the aggregate electrical current signal is adjusted to the operation of a determined commercial Analog Front-End (AFE) [11]. In this specific experimental case, the frequency is fixed at 4 kHz. This allows that all the definitions and design decisions made hereinafter for the data coming from BLUED dataset might be easily extended to future experimental developments based on an AFE with the same sampling frequency. Additionally, temporal windows are captured around the switching events in order to reduce the amount of data to be processed. To avoid possible errors, the event detector stage has been omitted and the events are extracted directly from the database. The duration of these temporal windows is 1.024 seconds; bearing in mind that the frequency is 4 kHz, the total number of samples per window is 4096.

Since the AE topology is constituted by convolutional layers and they are usually used for feature extraction from images, the temporal windows are divided into sections and then organized as rows of a matrix to create the input images. It is worth mentioning that this procedure was previously described in [12]. In this work, three CNN architectures are proposed for load identification in the field of NILM techniques. The different versions aim to reduce the complexity of the initial architecture without affecting its classification capability. Finally, the second version has significantly fewer parameters with similar results to the other two proposed alternatives. Specifically, it achieves an F1 score of 92.85 %, around the values of previous works using similar classification techniques, with a number of trainable parameters of 523560. The BLUED database was also used to validate the proposal. The resulting images for training present a square size of 64×64. Furthermore, before inserting them into the ANN model, their values are normalized between 0 and 1 as the model's functionality saturates at those limits, thus providing vanishing gradient problems. Fig. 1 shows an example of the resulting images.

B. Dimensional Reduction Stage

The approach proposed for dimensional reduction is a CAE. This architecture has two main parts: an encoder which maps the image to a code thanks to convolutional layers, and a decoder which reconstructs the image from the code by convolutional transpose layers. The specific structure of the CAE, shown in Fig. 2, is configured as follows:

- **Encoder**:
  - **Convolutional 2D Layer**: The number of channels produced by the convolution operation is 16, the size of the convolution kernel is 3×3 and the value of the stride parameter for the cross-correlation is 2. Henceforward, the kernel and stride settings will stay the same, unless otherwise indicated.
  - **ReLU Activation Layer**.
  - **Convolutional 2D Layer**: The number of output channels is 32.

- **ReLU Activation Layer**.
- **Convolutional 2D Layer**: The number of output channels is 64.
- **ReLU Activation Layer**.
- **Convolutional 2D Layer**: The number of output channels is 128.
- **ReLU Activation Layer**.
- **Convolutional 2D Layer**: The number of output channels is 256. However, in this case, the size of the convolution kernel is 4×4 to obtain a code of one dimension and the stride parameter is 1.
- **ReLU Activation Layer**.

- **Decoder**:
  - **Convolutional Transpose 2D Layer**: The number of channels produced by the transposed convolution operation is 128. Moreover, to follow the same sequence as in the encoder, the size of the convolving kernel is 4×4 and the stride parameter is 1.
  - **ReLU Activation Layer**.
  - **Convolutional Transpose 2D Layer**: The number of output channels is 64. However, the size of the convolution kernel is again 3×3 and the value of stride for the cross-correlation is 2. Additionally, for this type of layer, an additional size is added to one side of each dimension in the output shape. Once again, this configuration will stay the same, unless otherwise indicated.
  - **ReLU Activation Layer**.
  - **Convolutional Transpose 2D Layer**: The number of output channels is 32.
  - **ReLU Activation Layer**.
  - **Convolutional Transpose 2D Layer**: The number of output channels is 16.
  - **ReLU Activation Layer**.
  - **Convolutional Transpose 2D Layer**: The number of output channels is 1.
  - **Sigmoid Activation Layer**.

Within the encoder, the images are processed by five *Convolutional 2D* layers, which extract the most significant features to finally obtain the specific code for each image; whereas in the decoder, the rest of the information in the images is derived from the code, by using five *Convolutional Transpose 2D* layers. All these layers are followed by a *ReLU* activation function, except for the last layer of the decoder that uses the *Sigmoid* function. Both functions are used to introduce non-linearity. However, the *ReLU* function is exclusive with values equal to or less than zero (making them equal to zero), which implies a loss of information for the reconstruction of the images at the end of the decoder.

Fig. 1. Examples of a 64×64 input image for the CAE, containing on-switching events of the a) refrigerator, b) printer, c) air compressor and d) garage door.

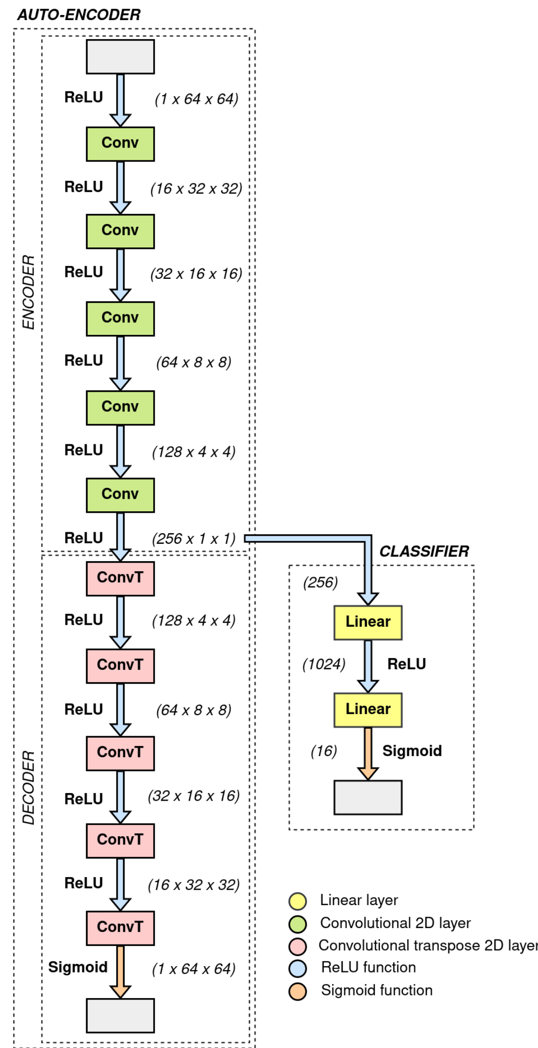

Fig. 2. Structure of the proposed CAE for the dimensional reduction stage.

C. Appliance Identification Stage

A plain neural network classifier is proposed to classify the codes obtained at the middle point after the encoder in the CAE (between the encoder and the decoder). The configuration of the network, shown in Fig. 2, is described below:

- **Linear Layer**: The size of each output sample is four times the input size, which is 256.
- **ReLU Activation Layer**.
- **Linear Layer**: The final size of the output samples corresponds to the number of classes, which is 16 due to the number of appliances considered from the dataset.
- **Sigmoid Activation Layer**.

The values from the reduction space of the encoder are inserted into the classifier to be processed by two *Linear* layers: the first layer provides more output features than inputs since this potentially leads to better results (it has more values to learn patterns from); and, finally, the second layer computes the scores for each considered class or appliance, where the label with the highest score reveals the identified appliance.

III. Experimental Results

Since the electrical signals obtained from the BLUED database [10] are sampled at 12 kHz, the first step is down-sampling to 4 kHz. Besides that, for the extraction of the temporal windows, the location of the events is provided by the ground truth of the database.

It should also be noted that to make up for the class imbalance, the total set of input images includes samples artificially created by introducing noise. This disproportion is caused due to the fact that there may not be enough switching events for all the considered appliances. Likewise, the images have been divided into three independent subsets: the training set with the 50 % of the images, the testing set with half of the remaining samples and the validation one with the other half.

The validation of the model's performance has been done by the consumption signals extracted from the BLUED database. The involved appliances are also electrical devices which are not relevant for the estimation of routines, but it is important to identify them in order to be able to discard them in a later process where the routines are extracted. In view of the above, the total number of electronic devices that have been considered is sixteen, as can be observed in Table I.

Before detailing the metrics of the experimental results, it is worth including some information about the training settings. The learning algorithm for the optimization process is the Adam algorithm [14] with an initial learning rate of 10^{-4} for the case of the CAE, whereas for the classifier is 10^{-5}. This parameter is modified using a decay rate of 10^{-6} and 10^{-7}, respectively. Furthermore, some regulation techniques to avoid over-fitting are used. In addition to the data augmentation procedure mentioned previously, the early stopping technique is employed to stop the training process when validation loss does not improve. Another key factor that may influence the model performance is the batch size. For this case, the number of samples or images per batch is 32. Finally, every subset is shuffled before using it, so the order does not affect the effectiveness of the algorithm.

To evaluate the reconstruction efficiency, the Mean Squared Error (MSE) loss function is used [15]. Particularly, the testing MSE obtained for the CAE achieves a value of $1.579 \cdot 10^{-3}$. Furthermore, the similarity between the input images and the reconstructed ones is calculated by using the Deep Image Structure and Texture Similarity (DISTS) index [16]. In this case, the values obtained for the test set are between 0.170 and 0.230, which means that the obtained images are highly similar, since the lower this metric the better the reconstruction is.

For the classification, the F1-score is used to combine the precision and recall metrics: the precision evaluates the quality of the model for the classification task, whereas the recall reports on the quantity the model can identify. It is noteworthy that, for problems with unbalanced classes, instead of using accuracy, it is better to use precision, recall and F1-score. Table II presents the results for every class where a weighted average F1-score of 87 % is attained. The worst performance is obtained for class 13 (LCD Monitor 1).

For further analysis, the confusion matrix is depicted in Fig. 3. The largest number of errors are introduced by class 10 (Lights). In particular, the number of samples from class 10 misclassified as class 13 is nearly half of the total number of samples from this last mentioned class. The reason for the misclassifications in this class 10 is because the switching

TABLE I
Selected Devices For Proposal Evaluation

Class	Device
00	Refrigerator
01	Hair Dryer
02	Iron
03	Television
04	Air Compressor
05	Kitchen Aid Chopper
06	Garage Door
07	Computer 1
08	Printer
09	Lamps (Living Room Desk)
10	Lights
11	Laptop 1
12	Basement Receiver, DVR, Blue-ray Player
13	LCD Monitor 1
14	Monitor 2
15	Living Room A/V System

TABLE II
Experimental Results For The Proposed Architecture

Class	Precision	Recall	F1-score
00	0.91	0.93	0.92
01	0.92	0.92	0.92
02	0.89	0.80	0.84
03	0.79	0.77	0.78
04	0.99	1.00	1.00
05	0.95	0.95	0.95
06	0.94	0.98	0.96
07	0.68	0.66	0.67
08	0.92	0.97	0.94
09	0.86	0.71	0.77
10	0.81	0.68	0.74
11	0.81	0.98	0.89
12	0.95	0.96	0.96
13	0.52	0.57	0.54
14	0.69	0.80	0.74
15	0.95	1.00	0.97
Weighted average	**0.87**	**0.87**	**0.87**

events of lights are not noticeable in comparison with other appliances, whether there is another appliance switched on in the image, as is shown in Fig 4. Conversely, class 4 scores are close to perfection even though this class has a large number of samples compared to other classes. This may be because the level of consumption and the type of electrical load (resistive, inductive, capacitive) of the air compressor differs to a larger extent from the rest of the loads. Furthermore, looking at the number of samples per class shown in Fig. 3, it can be observed that there exist an unbalanced number of events for the validation. In these situations, the minority class is more difficult to classify because there are few samples, as well as the abundance of the majority class can cause the model to ignore the minority class. However, as mentioned above, the worst results are obtained for class 10, which has a considerable volume of samples. Therefore, as this class is also misclassified with a large part of the other classes, the cause of these results may be that the model is introducing all mislabelled samples in this class.

Finally, a comparison with other previous works is considered here. In [17] it is proposed a CAE, which uses advanced training techniques such as batch normalization (BN) and hill climbing (HC), to disaggregate the appliances' power consumption. In this case, NILM is modelled as a regression problem and results show that the method performs the best in comparison to other schemes for the fridge and microwave cases, with a maximum MAE of 9.594 and a maximum SAE of 0.082 for the REDD database. On the other hand, in [18] it is proposed to model the dynamic behaviour of the NILM problem by a Long Short-Term Memory (LSTM) auto-encoder,

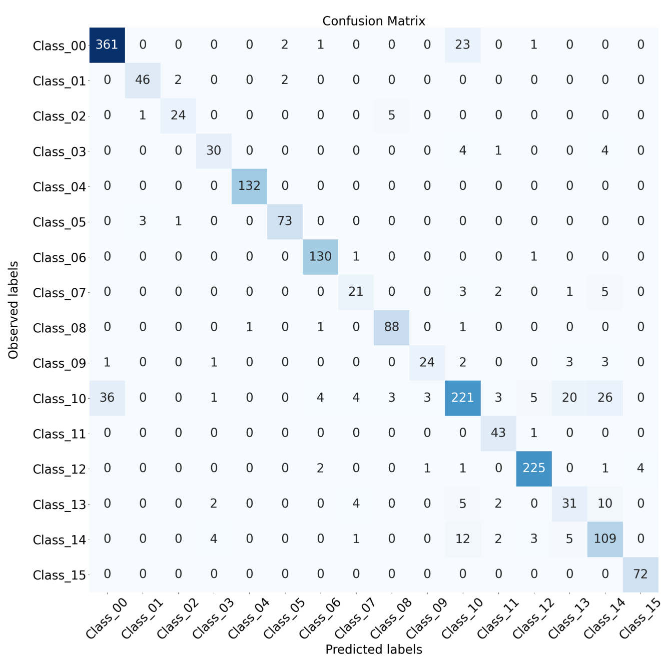

Fig. 3. Confusion matrix obtained in the test process.

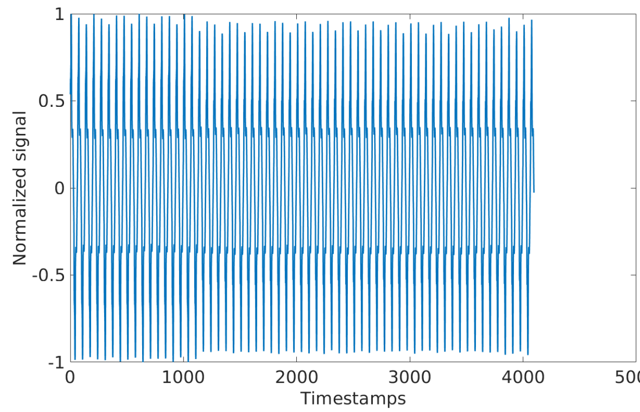

Fig. 4. Example of a temporal window, where an event from the lights is almost imperceptible.

which outperforms several state-of-the-art techniques with a micro F1-score of 0.74 for the REDD and 0.81 for the Pecan Street databases. Furthermore, in [19] a simple plug-and-play type smart measuring device is presented, which integrates a web application and a mobile application to visualize the estimated equipment-wise energy consumption by a denoising auto-encoder. The proposal has demonstrated its effectiveness for the refrigerator in house 1 of the REDD database and has shown promising results for the unseen data of the refrigerator in house 2. The MAE and RMSE values calculated with the proposed model are 16.66 and 47.47 for the seen data and 43.05 and 80.11 for the unseen data, respectively. Finally, in [20] a generative algorithm is proposed that integrates a variational autoencoder to detect unknown appliances and identify the known ones. Experimental F1-score on PLAID and WHITED databases for the known appliances are 0.9530 and 0.9831, respectively.

IV. Conclusions

In this work, a CAE-based classification architecture for the identification of electrical loads is proposed as a starting point to create an RPM system that may be able to provide information related to the health status of patients based on electricity consumption. In order to test the effectiveness of the proposal, the BLUED database has been used. A reconstruction error of $1.579 \cdot 10^{-3}$ and a weighted average classification F1-score of 87 % have been obtained. These preliminary results are encouraging, as a performance in the same range as previous works is achieved.

Future work will extend this study by applying unsupervised machine learning to avoid the need of human intervention for unknown electrical loads as is proposed in [20]. This feature will allow the proposal to be as close as possible to real-life situations, where household appliances are often replaced because their useful life is limited. Furthermore, the future purpose is to infer certain routines and daily activities through the information obtained from the appliances, in the context of developing supporting tools and RPM systems.

REFERENCES

[1] G. Cicirelli, R. Marani, A. Petitti, A. Milella & T. D'Orazio, "Ambient Assisted Living: A Review of Technologies, Methodologies and Future Perspectives for Healthy Aging of Population," in *Sensors*, vol. 21, pp. 3549, 2021, doi: 10.3390/s21103549

[2] L. P. Malasinghe, N. Ramzan & K. Dahal, "Remote patient monitoring: a comprehensive study," in *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 57-76, 2019.

[3] R. Gravina & G. Fortino, "Wearable Body Sensor Networks: State-of-the-Art and Research Directions," in *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12511-12522, 2021, doi: 10.1109/JSEN.2020.3044447.

[4] J. Revuelta Herrero, Á. Lozano Murciego, A. López Barriuso, D.Hernández de la Iglesia, G. Villarrubia González, J. M. Corbacho Rodríguez, & R. Carreira, "Non Intrusive Load Monitoring (NILM): A State of the Art," in *Proc. 15th International Conference on Practical Applications of Agents and Multi-Agent Systems*, pp. 125- 138, 2018, doi: 10.1007/978-3-319-61578-3_12.

[5] C. M. Giebel, C. Sutcliffe & D.Challis, "Activity of daily Living and Quality of Life across different Stages of Dementia: A UK study," in *Aging & Mental Health*, vol. 19, no. 1, pp. 63-71, 2015, doi: 10.1080/13607863.2014.915920.

[6] M. Lee, R. K. Mishra, A. Momin, N. El-Refaei, A. B. Bagheri, M. K. York, M. E. Kunik, M. Derhammer, B. Fatehi, J. Lim, R. Cole, G. Barchard, A. Vaziri & B. Najafi, "Smart-Home Concept for Remote Monitoring of Instrumental Activities of Daily Living (IADL) in Older Adults with Cognitive Impairment: A Proof of Concept and Feasibility Study," in *Sensors*, vol. 22, pp. 6745, 2022, doi: 10.3390/s22186745

[7] Y. T. Quek, W. L. Woo & T. Logenthiran, "Load Disaggregation Using One-Directional Convolutional Stacked Long Short-Term Memory Recurrent Neural Network," in *IEEE Systems Journal*, vol. 14, no. 1, pp. 1395-1404, 2020, doi: 10.1109/JSYST.2019.2919668.

[8] H. Chen, YH. Wang, & CH. Fan, "A convolutional autoencoder-based approach with batch normalization for energy disaggregation," in *J Supercomput*, vol. 77, pp. 2961–2978, 2021, doi: 10.1007/s11227-020-03375-y.

[9] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, & A. J. Aljaaf, "A systematic review on supervised and unsupervised machine learning algorithms for data science," in *Supervised and unsupervised learning for data science*, Springer, pp. 3-21, 2020, 10.1007/978-3-030-22475-2_1.

[10] K. Anderson, A. F. Ocneanu, D. Benítez, D. Carlson, A. Rowe & M. Bergés, "Blued: a fully labelled public dataset for event-based nonintrusive load monitoring research," in *Proc. 2nd KDD Workshop on Data Mining Applications in Sustainability*, pp. 1-5, 2012.

[11] D. Ying & D. A. Hall, "Current Sensing Front-Ends: A Review and Design Guidance," in *IEEE Sensors Journal*, vol. 21, no. 20, pp. 22329-22346, 2021, doi: 10.1109/JSEN.2021.3094830.

[12] L. de Diego-Otón, Á. Hernández, R. Nieto & C. Pérez-Rubio "Evaluación de una Arquitectura CNN para la Identificación de Cargas en NILM," in *28th Seminario Anual de Automática, Electrónica Industrial e Instrumentación*, pp. 116-121, 2021.

[13] L. de Diego-Otón, Á. Hernández, R. Nieto & M. C. Pérez-Rubio, "Comparison of Neural Networks for High-Sampling Rate NILM Scenario," in *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1-6, 2022, doi: 10.1109/MeMeA54994.2022.9856406. 2018. Available online: ArXiv abs/1803.08375.

[14] D. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proc. 3th International Conference on Learning Representations*, pp. 1-15, 2015.

[15] N. Khare, P.S. Thakur, P. Khanna & A. Ojha, "Analysis of Loss Functions for Image Reconstruction Using Convolutional Autoencoder," in *Computer Vision and Image Processing: 6th International Conference*, vol 1568, pp. 338-349, 2022, doi: 10.1007/978-3-031-11349-9_30.

[16] K. Ding, K. Ma, S. Wang & E.P.Simoncelli, "Image quality assessment: Unifying structure and texture similarity," in *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 5, pp. 2567-2581, 2020.

[17] H. Chen, YH. Wang, YH. & CH. Fan, "A convolutional autoencoder-based approach with batch normalization for energy disaggregation," in *The Journal of Supercomputing*, vol. 77, pp. 2961–2978, 2021. doi: 10.1007/s11227-020-03375-y.

[18] S. Verma, S. Singh & A. Majumdar, "Multi-label LSTM autoencoder for non-intrusive appliance load monitoring," in *Electric Power Systems Research*, vol. 199, p. 107414, issn. 0378-7796, 2021, doi: 10.1016/j.epsr.2021.107414.

[19] W.A.H.M. Sewwandi, C.N.D. Anthony, D.M.A.M. Disanayaka, M.A.K.S. Boralessa & K.T.M.U. Hemapala, "Non-Intrusive Load Monitoring Using Denoising Autoencoder Neural Networks," in *IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 408-413, 2022, doi: 10.1109/R10-HTC54060.2022.9929710.

[20] Y. Han, K. Li, C. Wang, F. Si & Q. Zhao, "Unknown Appliances Detection for Non-Intrusive Load Monitoring Based on Conditional Generative Adversarial Networks," in *IEEE Transactions on Smart Grid*, pp. 1-1, 2023, doi: 10.1109/TSG.2023.3261271.