A Hybrid Attention Model for Non-intrusive Load Monitoring based on Time Series Feature

Zhaorui Meng*, Xiaozhu Xie, Yanqi Xie, Jinhua Sun

Abstract—**Non-intrusive load monitoring (NILM) has become a widely used approach to monitor energy consumption by installing monitoring equipment at the power supply entrance. However, the accuracy of traditional deep neural network decomposition models falls short of meeting practical demand. To address this limitation, the present study introduces a novel hybrid neural network model, which integrates temporal feature extraction and an attention mechanism. The proposed model is designed to discern the salient attributes within power time series signals, thereby reducing the dimensionality of the resulting characteristic temporal signals via the application of Principal Component Analysis (PCA). Next, a Gated Recurrent Unit (GRU) neural network with an attention mechanism extracts the features of the generated information vector, and generates the load decomposition model after multiple iterations of learning. The experimental outcomes on the REDD public dataset substantiate the superiority of the proposed model over alternative deep learning techniques, including CNN, GRU, and GRU with attention. The proposed model demonstrates a significantly elevated degree of precision within the domain of load identification.**

Index Terms—Non-intrusive load monitoring; Attention mechanism; Feature extraction; Gated recurrent unit.

I. INTRODUCTION

The issue of global warming and environmental pollution has resulted in increasing attention towards scientific electricity usage and promoting efficient power consumption [1]. Non-intrusive load monitoring (NILM) represents an intelligent power utilization technology that aims to analyze the power data at the user's power supply and obtain load category and power utilization information in the user's area. The primary function of NILM is to decompose and classify loads, specifically to manage user's electricity consumption behavior, achieve refined energy management, improve energy utilization efficiency, and reduce energy consumption costs. These goals are of significant importance for reducing carbon emissions. As a result, NILM has emerged as an effective approach to promoting energy conservation and reducing environmental pollution [2-3].

In recent years, artificial intelligence and advanced measurement technology have developed rapidly. Scholars began to apply new technologies to the research of new electricity consumption [4], load forecasting [5] and load monitoring [6]. In 2015, Professor Kelly [7] introduced the application of deep learning algorithms to NILM for the first time. Compared with traditional algorithms, this method has higher accuracy and is more easily scalable. Currently, numerous scholars are applying neural network methods to NILM. For instance, Zhang et al. [8] proposed a sequence-to-point model that utilizes convolutional neural networks to train a load monitoring model. Experiments conducted on real-world household datasets demonstrated that the proposed neural network model outperforms other methods. Bonfigli et al. [9] introduced a NILM algorithm that uses the Denoising Autoencoder model to approach the NILM problem as noise reduction. Experiments conducted on three NILM datasets confirmed that the proposed model outperforms other comparison models. Shin et al. [10] have developed a deep network to improve the ability to analyze appliance on/off status. The proposed Subtask Gated Networks achieved superior performance compared to most of the benchmark cases.

Extensive research has been conducted to enhance the performance of NILM algorithms by analyzing features extracted from current and voltage signals. Wang et al. [11] successfully applied the Voltage-Current (V-I) trajectory in the NILM field, yielding positive outcomes. Jimenez et al. [12] investigated a combined approach employing the S-Transform and Support Vector Machines for load monitoring. Khalid et al. [13] developed a time-time-transform-based method to improve the accuracy of load identification. Moreover, various scholars have extensively studied the application of wavelet transform in NILM [14-16]. Collectively, experimental results from these studies have consistently demonstrated the effectiveness of hybrid models based on wavelet transform in NILM.

As an increasingly popular technique in various fields such as image recognition and speech recognition, the attention mechanism [17] has also been studied in the context of NILM. Wang et al. [18] developed a NILM model that integrates the sequence-to-sequence model with attention mechanism. Piccialli et al. [19] proposed a deep neural network that incorporates a tailored attention mechanism for NILM. Both studies show that the use of attention mechanism improves

Manuscript received March 20, 2023, revised February 1, 2024;

Zhaorui Meng is a senior engineer in School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian,361024, China (Corresponding author; Email: mengzhaorui@ xmut.edu.cn).

Xiaozhu Xie is a lecturer in School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian,361024, China (Email: xiaozhuxie@ xmut.edu.cn).

Yanqi Xie is a lecturer in School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian,361024, China (Email: yqxie@ xmut.edu.cn).

Jinhua Sun is an associate professor in School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian,361024, China (Email: jhsun@ xmut.edu.cn).

the effectiveness of load monitoring compared to conventional neural networks.

Inspired by the aforementioned deep learning models and feature extraction methods, a hybrid model for NILM using time series feature extraction and attention mechanism is proposed in this study.

II. METHODOLODGY

2.1 Workflow of hybrid NILM model: TS-Attention-GRU

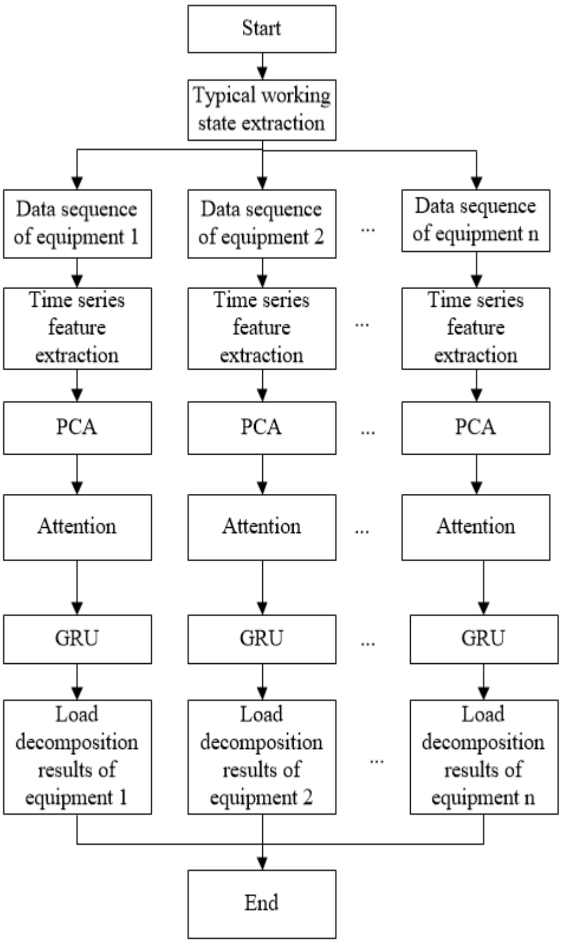


Fig. 1. Flowchart of TS-Attention-GRU NILM approach

The flow chart of the hybrid NILM model TS-Attention-GRU used in this paper is shown in Fig. 1. The load identification algorithm mainly includes the following steps:

1) Extraction of Typical Operational States: The process involves the collection of household electricity data, encompassing the active power consumption of the entire household as well as individual electrical appliances. By aggregating the active power and total power measurements of distinct electrical appliances, a comprehensive dataset is created for each specific appliance.

2) Time Series Feature Extraction: To enhance the effectiveness of load decomposition, the extraction of time series features is employed to extract the active power data of the load. In this study, the time series feature extraction method utilized the Python package "tsfresh," developed by Chris [20], as the preferred tool for this purpose.

3) Principal Components Analysis (PCA): The dimensionality of the feature vectors obtained from the time series analysis of electrical appliances is typically quite high, often reaching several hundreds. Training these feature vectors directly using neural networks would result in excessive training time. Hence, it becomes necessary to reduce the dimensionality of the previously extracted time series features through the application of PCA.

4) Neural Network Training: The load features acquired in the initial three steps are inputted into a Gated Recurrent Unit (GRU) neural network, incorporating an attention mechanism, for the purpose of training.

5) Load decomposition: The unidentified load is subjected to the same processing steps as outlined in steps 1 to 3. Subsequently, it is inputted into the trained neural network to undergo load decomposition.

. 2.2. Time series feature extract methods

Time series feature extraction holds significant importance within data science projects. Investigating and assessing the statistical characteristics of time series features play a crucial role in time series prediction. In this regard, Chris et al. have developed a Python package named "tsfresh" specifically designed for time series feature extraction. This software amalgamates 63 time series characterization methods to provide a total of 794 time series features. Tsfresh facilitates the computation of various time series features that encompass fundamental characteristics such as peak count, average value, maximum value, and time reversal symmetric statistics, among others. Furthermore, these features are subjected to hypothesis testing to identify the subset of features that best elucidate the underlying trend, a process referred to as decorrelation. The resulting feature sets can be effectively utilized in training machine learning models, regardless of whether the problem at hand involves time series regression or classification.

Fig. 2 illustrates the three sequential stages of the tsfresh algorithm. Initially, the algorithm employs a thorough and precise feature mapping technique to represent the time series data, while simultaneously taking into account supplementary features that describe the meta information. Subsequently, each feature vector's significance in relation to the predicted target is independently evaluated. This evaluation yields a p-value vector, which quantifies the relative importance of each feature with respect to the predicted target. Finally, the p-value vector undergoes assessment using the Benjamini-Yekutieli (BY) multiple testing procedure [21] to determine the selection of features to be retained for further analysis.

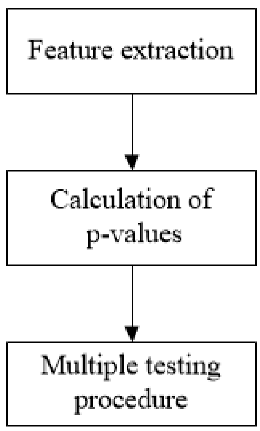


Fig. 2. Feature extraction workflow of tsfresh

2.3 GRU

Long Short-Term Memory (LSTM) is often used in time series prediction. Nevertheless, because of its complicated internal structure, it usually takes more time to train the LSTM model.

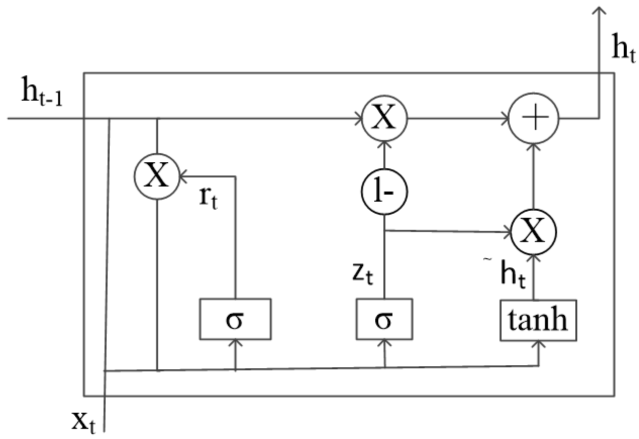


Fig. 3. Unit structure of GRU

Unlike LSTM, GRU [22] only has update door and reset door. The internal structural of GRU is shown in Fig. 3. Its internal relationship is:

$$z_t = \sigma(W_z h_{t-1} + U_z x_t + b_z) \tag{1}$$
$$r_t = \sigma(W_r h_{t-1} + U_r x_t + b_r) \tag{2}$$
$$\widetilde{h_t} = tanh(W_h(r_t h_{t-1}) + U_h x_t + b_h) \tag{3}$$
$$h_t = (1 - z_t)h_{t-1} + z_t \widetilde{h_t}) \tag{4}$$

where z_t is update gate, r_t is reset gate, $\widetilde{h_t}$ is the summary of the input x_t and output of the previous hidden layer h_{t-1}. σ is the sigmoid function. W_z, W_r, W_h, U_z, U_r, U_h, b_z, b_r, b_h are the weight matrix of the cell.

2.4 Attention mechanism

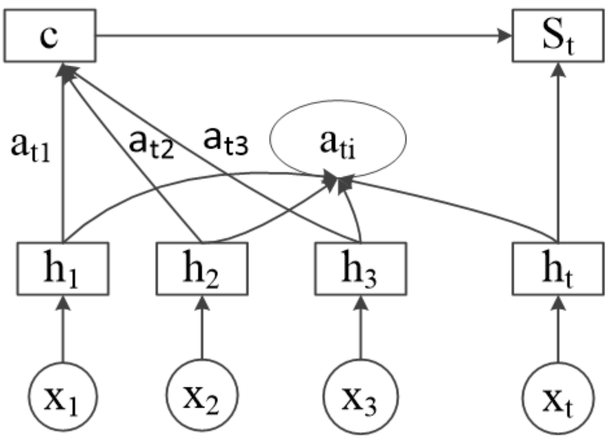


Fig. 4. Attention unit structure

The attention mechanism efficiently directs limited computational resources towards crucial information within the power sequence, thereby optimizing computing efficiency and expediently extracting the most effective information. Through weighted summation, the attention mechanism determines the attention weight assigned to each element within the power sequence, signifying the relative importance of the input information. The adaptive allocation of these weights enhances both the training efficiency and convergence speed of the model. Attention unit structure is shown in Fig. 4, where x_1, x_2,…, x_t are the power sequence data. h_1, h_2,…, h_t are the state value of the GRU hidden layer output corresponding to the input. $a_{t,i}$ are the current attention weight value corresponding to the GRU hidden layer output value. S_t are the final output hidden layer status value. The calculation steps of attention mechanism are as follows:

$$e_{t,i} = V^T \tanh(Wh_t + Uh_i), i = 1,2,\dots,t-1 \tag{5}$$
$$a_{t,i} = \frac{e_{t,i}}{\sum_{k=i}^{N_t} a_{k,i}}, i = 1,2,\dots,t-1 \tag{6}$$
$$C = \sum_{i=1}^{N_i} a_{t,i} h_i, i = 1,2,\dots,t-1 \tag{7}$$
$$S_t = \mathrm{f}(C, h_i) \tag{8}$$

where V, W and U are training parameters.

III. EXPERIMENTS

To verify the validity of the algorithm proposed, this paper selected the public dataset REDD [23] for experiment. The neural network was built in PyTorch 1.11 environment. The hardware platform is Intel Xeon platinum 8124 with 128G memory and RTX3080 graphics card.

3.1. Datasets

. The REDD dataset encompasses long-term power consumption data from six households, comprising primarily high-voltage current data (16.5 kHz) and low-power (1 Hz) sampling data. Specifically, the high-frequency sampling data is acquired from the current and voltage acquisition device connected to the household's power supply port, whereas the low-frequency power data is obtained from power acquisition devices installed on each load branch and two buses. For the experimental purposes of this paper, the low-frequency power data of household 1 within the REDD dataset is utilized.

3.2. Metrics

NILM has many evaluation metrics. In this paper, we choose metrics for status monitoring: Accuracy, and F_1 score. Meanwhile, we choose Mean Absolute Error (MAE) as the metric for power decomposition.

1) Status monitoring metrics:

$$PR = \frac{TP}{TP+FP} \tag{9}$$
$$RE = \frac{TP}{TP+FN} \tag{10}$$
$$F_1 = \frac{2\times PR\times RE}{PR+RE} \tag{11}$$
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

where true positive (TP) indicates the number of status monitoring are consistent with the actual status; false positive (FP) indicates the number of positive but actually negative status monitoring; false negative (FN) indicates the number of negative but actually positive status monitoring; true negative (TN) is the count of correctly captured negative status. The closer Accuracy and F_1 are to 1, the higher the accuracy of the model.

2) Metric for power decomposition:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y(t) - \hat{y}(t)| \tag{13}$$

where $y\,(t)$ is the power decomposition value at time t; $\hat{y}(t)$ is the actual power value at time t; T is the duration. The smaller the MAE, the higher the accuracy of the model.

3.3. Experimental results

In order to assess the algorithm's performance, the proposed model is compared with convolutional neural network (CNN), GRU, and GRU enhanced with attention mechanism. The experimental results are presented in Tables I-III. The tables reveal that the algorithm proposed in this paper achieves the highest values across all three performance indicators: accuracy, F_1 score, and MAE, with

Table I. Comparison of prediction accuracy of each model.

Accuracy.	Washing Machine	Dish Washer	Fridge	Micro wave	Sockets
CNN	0.9428	0.6599	0.8232	0.9572	0.9111
GRU	0.9379	0.6887	0.8377	0.9598	0.8722
Attention with GRU	0.9542	0.8299	0.8477	0.9570	0.9116
TS-Attention-GUR	0.9859	0.9557	0.8655	0.9914	0.9038

Table II. Comparison of prediction F_1 of each model.

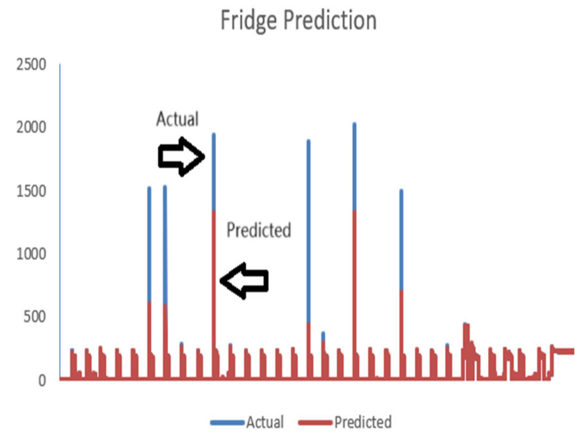
Accuracy.	Washing Machine	Dish Washer	Fridge	Micro wave	Sockets
CNN	0.4575	0.1860	0.7487	0.2353	0.9533
GRU	0.4399	0.2014	0.7597	0.2577	0.9314
Attention with GRU	0.5105	0.2354	0.7476	0.2498	0.9536
TS-Attention-GUR	0.7719	0.3518	0.7652	0.3019	0.9525

Table III. Comparison of prediction MAE of each model.

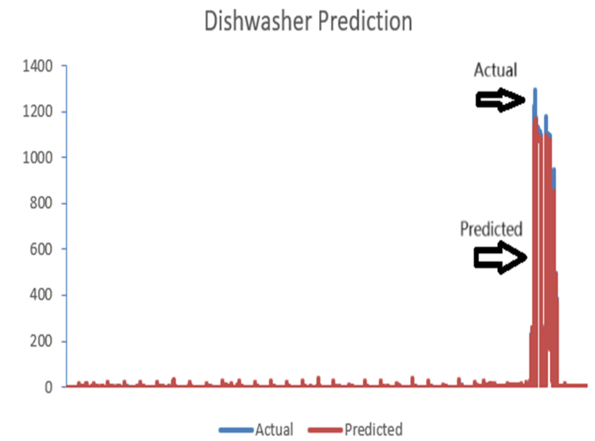
Accuracy.	Washing Machine	Dish Washer	Fridge	Micro wave	Sockets
CNN	38.14	79.15	41.85	52.66	1.52
GRU	36.97	77.46	43.09	44.442	1.6109
Attention with GRU	33.92	63.2	42.58	45.72	1.58
TS-Attention-GUR	31.47	29.36	38.07	40.35	1.5923

the exception of sockets. Overall, the proposed algorithm exhibits improvements ranging from 5% to 30% in comparison to other algorithms. However, it is worth noting that the algorithm proposed in this paper does not exhibit significant advantages for electrical equipment that is infrequently used, such as sockets.

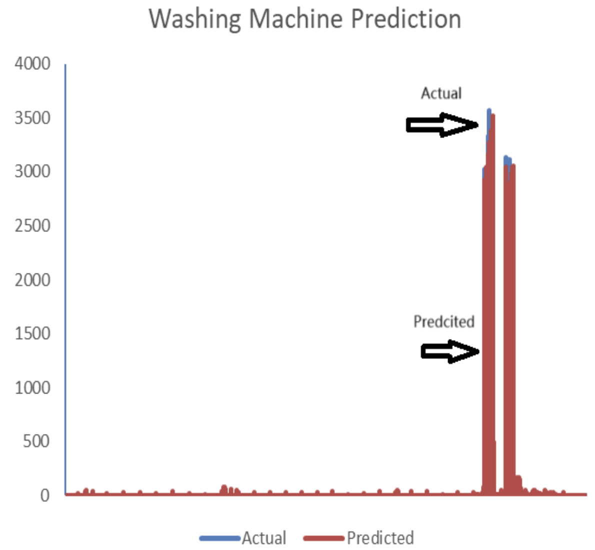
Fig.5 presents the outcomes of load power decomposition achieved through the algorithm proposed in this study. The figure showcases the accurate identification of the initiation and termination phases of electrical equipment, alongside a notable effectiveness in power decomposition. Consequently, the load decomposition model proposed in this paper successfully fulfills the requirements of the NILM task.



(a) Fridge



(b) Dishwasher



(c) Washing machine.

Fig. 5. Load power decomposition results

IV. Conclusion

This study presents a non-intrusive load monitoring (NILM) model integrating time series feature extraction and an attention mechanism. The methodology entails decomposing the time series features of large-scale load data, then inputting the PCA-reduced data into a gated recurrent unit (GRU) neural network with an incorporated attention mechanism. This approach facilitates non-intrusive load monitoring and power disaggregation. The experimental results indicate that the proposed method displays higher predictive accuracy relative to convolutional neural networks (CNNs), GRUs, and GRUs with an attention mechanism when applied to frequently used electrical devices. In contrast, for infrequently utilized electrical devices such as sockets, the predictive accuracy of the proposed method is analogous to that of the other benchmark models.

In the future, it is anticipated that a broader range of load characteristic libraries encompassing diverse types and models will be established with the aim of enhancing the effectiveness of NILM algorithms.

REFERENCES

[1] Rajendiran, G., et al. "Energy management using non-intrusive load monitoring techniques - State-of-the-art and future research directions," Sustainable Cities and Society, vol. 62, no.1, 2020.

[2] P. A. Schirmer and I. Mporas, "Non-Intrusive Load Monitoring: A Review," IEEE Transactions on Smart Grid, vol. 14, no. 1, pp. 769-784, 2023.

[3] Lucas, P. and N. Nuno. "Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools-A review," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 8, no. 6,2018.

[4] Miaomiao Wang, Qingwen Luo, Lulu Kuang, and Xiaoxi Zhu, "Optimized Rolling Grey Model for Electricity Consumption and Power Generation Prediction of China," IAENG International Journal of Applied Mathematics, vol. 49, no.4, pp577-587, 2019.

[5] Zhaorui Meng, Yanqi Xie, and Jinhua Sun, "A Hybrid Model for Short-term Load Forecast Using Empirical Mode Decomposition and Transfer Learning," IAENG International Journal of Computer Science, vol. 48, no.1, pp134-141, 2021.

[6] Chen, Kunjin et al. "Scale- and Context-Aware Convolutional Non-Intrusive Load Monitoring," IEEE Transactions on Power Systems, vol. 35, no. 6, pp. 2362-2373, 2020.

[7] Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," in Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, pp. 55–64, 2015.

[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 Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[9] R. Bonfigli, A. Felicetti, E. Principi, M. Fagiani, S. Squartini, and F. Piazza, "Denoising autoencoders for non-intrusive load monitoring: improvements and comparative evaluation," Energy and Buildings, vol. 158, pp. 1461–1474, 2018.

[10] C. Shin, S. Joo, J. Yim, H. Lee, T. Moon, and W. Rhee, "Subtask gated networks for non-intrusive load monitoring," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp.1150–1157,2019.

[11] Wang, A. L, et al. "Non-intrusive load monitoring algorithm based on features of V–I trajectory," Electric Power Systems Research, vol. 157, pp.134-144,2018.

[12] Jimenez, Y, et al. "Feature extraction for nonintrusive load monitoring based on S-Transform," Power Systems Conference IEEE, pp.1-5,2014.

[13] Khalid, K., et al. "Nonintrusive load identification using extreme learning machine and TT-transform." International Conference on Advances in Electrical IEEE, 2017.

[14] Duarte, C, et al. "Non-intrusive load monitoring based on switching voltage transients and wavelet transforms." Future of Instrumentation International Workshop (FIIW), 2012 IEEE, 2012.

[15] Gillis, J. M, and W. G. Morsi. "Non-Intrusive Load Monitoring Using Semi-Supervised Machine Learning and Wavelet Design," IEEE Transactions on Smart Grid, vol. 8, no. 6, pp. 2648-2655, 2016.

[16] Gillis, et al. "Nonintrusive Load Monitoring Using Wavelet Design and Machine Learning," IEEE Transactions on Smart Grid, vol. 7, no. 1, pp. 320-328, 2016.

[17] Bahdanau, D., K. Cho, and Y. Bengio. "Neural Machine Translation by Jointly Learning to Align and Translate," Computer Science, 2014.

[18] K. Wang, H. Zhong, N. Yu, and Q. Xia, "Nonintrusive load monitoring based on sequence-to-sequence model with attention mechanism," Zhongguo Dianji Gongcheng Xuebao/Proceedings of the Chinese Society of Electrical Engineering, vol. 39, no. 1, pp. 75–83, 2019.

[19] Piccialli, V., and A. M. Sudoso. "Improving Non-Intrusive Load Disaggregation through an Attention-Based Deep Neural Network," Energies, vol. 14, no. 4, 2019.

[20] Christ, M, et al. "Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)," Neurocomputing, vol. 307, pp. 72-77, 2018

[21] Benjamini, Y., and D. Yekutieli. "The control of the false discovery rate in multiple testing under dependency," Annals of Statistics, vol. 29, no.4, 2001.

[22] Cho, Kyunghyun et al. "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," Computer Science ,2014.

[23] Kolter, J. Z, and M. J. Johnson. "REDD: A Public Data Set for Energy Disaggregation Research," Artificial Intelligence, vol. 25, 2011.