Applying Intelligent Algorithms In Short-Term Electrical Load Forecasting

Trong Nghia Le
Ho Chi Minh City University of Technology and Education, Vietnam
trongnghia@hcmute.edu.vn (corresponding author)

Ngoc An Nguyen
Ho Chi Minh City University of Technology and Education, Vietnam
2230602@student.hcmute.edu.vn

Thi Ngoc Thuong Huynh
Ho Chi Minh City University of Technology and Education, Vietnam
thuonghtn@hcmute.edu.vn

Quang Trung Le
Lilama 2 International Technology College, Dong Nai, Vietnam
qtrungttc2@gmail.com

Thi Thu Hien Huynh
Ho Chi Minh City University of Technology and Education, Vietnam
hienhtthu@hcmute.edu.vn

Thi Thanh Hoang Le
Ho Chi Minh City University of Technology and Education, Vietnam
hoangltt@hcmute.edu.vn

Received: 5 July 2024 | Revised: 16 July 2024 and 19 July 2024 | Accepted: 21 July 2024

 DOI: https://doi.org/10.48084/etasr.8304

ABSTRACT

This study presents short-term electricity load forecasting for the New England area by processing initial data through correlation assessment and data clustering. This method is combined with artificial neural networks to improve accuracy and forecast performance. Data preprocessing focuses on two main issues: identifying correlations between variables to eliminate less relevant factors and retaining highly correlated variables to reduce noise, as well as reducing the data sample size before training the neural network. This evaluation aims to determine the factors that have a significant impact on electricity load. These factors can include previous load values, weather conditions, time, types of electricity usage, and others. This technique ensures that reducing the size in both dimensions of the large dataset does not result in the loss of critical information, maintaining the accuracy of computational programs and the performance of neural network training at high levels. The neural network is trained to classify and cluster data based on previously identified correlated characteristics. As a result, the forecasting model can make more accurate predictions about future electricity loads. Experimental results show that the proposed method achieved more than 97% accuracy, outperforming traditional methods in both speed and load forecasting accuracy. The new dataset had 63% fewer samples compared to the initial dataset.

Keywords-load forecasting; K-means algorithm; correlation assessment; neural networks; data processing

I. INTRODUCTION

Data management and processing in the energy sector require significant investment in technology, personnel, and processes. Organizations must adopt advanced tools and techniques to overcome these challenges, ensuring that data can be used effectively to support load forecasting and energy system management. However, accurate load prediction faces many difficulties due to rapid development, and traditional load

prediction tools have become somewhat outdated and inadequate, necessitating newer and more advanced powerful tools. Widely applied traditional methods, such as the regression method and various data transformation techniques, play a crucial role in forecasting electricity loads. However, regression must be applied with caution, especially when selecting and implementing data transformations, to ensure the effectiveness and accuracy of the forecast [1-2]. Furthermore, the multiple regression method is a powerful and efficient tool in electricity load forecasting. In [3-4], variant models with unique advantages were developed, including the introduction of a new regression method capable of estimating nonlinear loads. The exponential smoothing method, where the initial load is based on the previous model and is used to construct the future model, was proposed in [5]. In [6], an improved method was proposed that incorporated power spectrum analysis and autoregressive modeling. Furthermore, various other methods have been proposed, such as iterative reweighted least-squares [7], adaptive demand forecasting [8], stochastic time series [9], Autoregressive Moving-Average (ARMA) [10], Autoregressive Integrated Moving-Average (ARIMA) [11], fuzzy logic [12], and so on.

In general, each forecasting method has its own advantages. For example, regression is simple and quick to implement, fuzzy logic can handle ambiguous signals, and ARIMA is flexible and models data well, responding quickly to recent changes. However, each method also has its shortcomings, particularly when dealing with nonlinear input data. In this context, neural networks, a remarkable achievement of artificial intelligence, have emerged as a powerful and flexible tool for addressing complex problems. The ability of neural networks to learn from data, recognize patterns, and make predictions has been demonstrated in various fields. Specifically, in the field of load forecasting, neural networks offer significant advantages over traditional models. Regarding their self-learning abilities [13-14], neural networks can automatically learn and adjust from historical data without the need for manual intervention. This ability enables them to quickly adapt to new trends and handle non-linear relationships, which traditional methods such as linear regression, moving averages, and ARIMA models often struggle with. This ability allows neural networks to forecast loads with higher accuracy, especially in situations with significant fluctuations and heterogeneous data.

In addition, Recurrent Neural Network (RNN) [15], Convolutional Neural Network (CNN) [16], and Support Vector Machine (SVM) [17] models have also been applied in load forecasting. Although ARIMA models are powerful in handling time-series data, they face difficulties when dealing with nonlinear data and anomalous transformations. Deep learning techniques, such as RNN and CNN, have opened new opportunities to improve the accuracy and efficiency of load forecasting. RNNs, in particular, are effective in handling sequential data and can retain information from long sequences. CNNs excel in learning spatial features and have been successfully applied to various complex forecasting problems.

This study aims to present the advantages of applying neural networks in load forecasting and compare forecast results using improved data preprocessing with various network architectures. This study aims not only to provide a comprehensive view of the effectiveness of neural networks in load forecasting, but also to propose improved methods for optimizing training data in terms of sample size and variables. The effectiveness of the proposed method is demonstrated through the application of artificial intelligence and correlation algorithms to imitate intelligent or smart thoughts and actions of humans. All this process is performed automatically through computer programs and artificial intelligence software. Computer modeling and load forecasting were performed using data from the New England grid. After preprocessing steps to reduce variables and filter samples by clustering, simulations were carried out to test the effectiveness using a backpropagation neural network with 4 algorithms. The results were evaluated using the Mean Absolute Percentage Error (MAPE) to find the most suitable model.

II. METHODOLOGY

A. Data Preprocessing and Dimension Reduction

1) Introduction to the New England Power Grid Dataset

The data used for this study were sourced from the ISO New England power grid [18]. This data is openly accessible through the ISO New England website for quick access. The variables in the dataset include Hour, Temperature, Humidity, Electricity Consumption, Real-Time Price, and Time (considering weekdays/holidays). These variables form the basis of the neural network training dataset, initially comprising 5088 samples and 19 variables corresponding to the initial specific parameters. Figure 1 provides an overview of the data collection process.

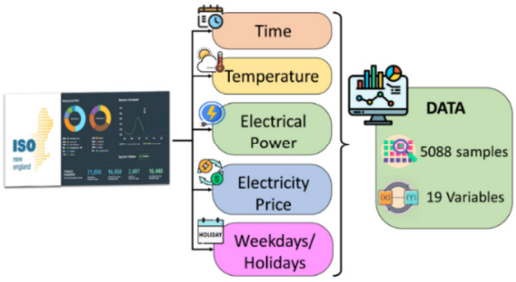

Fig. 1. The data collection process for training a neural network on the ISO New England power grid data.

2) Data Preprocessing

Data preprocessing is a critical step in the process of handling and analyzing data in the fields of data science and machine learning [19]. Its main task is to clean, normalize, and transform raw data to make them logical and suitable for analysis or model training. Data preprocessing helps improve the quality and performance of analyzing large datasets and training models. Data preprocessing can employ various methods, such as normalization, which involves transforming data into smaller value ranges. For instance, scaling data to range from -1.0 to 1.0 or from 0.0 to 1.0. Data correlation is an

important statistical measure to evaluate the relationship between two variables, indicating the extent and direction of the variability of one variable influencing another. In data normalization, there are three main methods: min-max scaling, z-score normalization, and decimal scaling.

3) Correlation Coefficient

The correlation coefficient is an important statistical measure to assess the relationship between two variables, indicating the degree and direction of change in one variable relative to another. Correlation is typically represented by a coefficient ranging from -1 to 1, where specific values carry distinct meanings. A positive value indicates that as one variable increases, the other variable also tends to increase. On the contrary, a negative value indicates that as one variable increases, the other variable tends to decrease. There are two approaches to correlation analysis: parametric and non-parametric. Parametric tests assume that the data follow standard distributions, such as normality, linearity, and homogeneity [20], whereas non-parametric tests do not require these assumptions [21].

Pearson's correlation is an example of parametric correlation. Spearman correlation provides examples of non-parametric ordinal rank correlation without assuming any distribution of the data [22]. The benefits of using non-parametric methods include that they can operate on incomplete data (where only ordinal information is available), the Spearman rank correlation equals 1 when Y is a monotonic increasing function of X, and Spearman is robust against outliers compared to Pearson correlation [23].

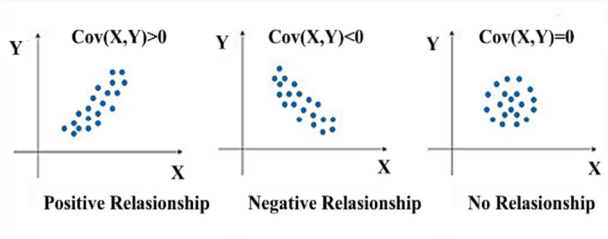

Fig. 2. Graphical visualization of correlation.

Pearson's correlation coefficient is given by

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_x \sigma_y} \quad (1)$$

where $cov(X,Y)$ is the covariance of X and Y, and σ_X and σ_Y is the standard deviation of X and Y, respectively. Figure 2 illustrates three scenarios based on the covariance:

- When $\rho(X,Y) > 0$, X and Y have a positive linear relationship. As X increases, Y tends to increase.
- When $\rho(X,Y) < 0$, X and Y have a negative linear relationship. As X increases, Y tends to decrease.
- When $\rho(X,Y) = 0$, X and Y have no relationship with each other.

The Spearman rank correlation coefficient plays a crucial role in assessing the degree of association between random variables. It is particularly useful in cases where data do not follow a normal distribution or when there are multiple outliers present. This method provides an effective non-parametric approach to measure nonlinear correlations.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (2)$$

where ρ is the Spearman correlation coefficient, Σd^2 denotes the sums of squares of rank differences, and n is the number of observations. The value of ρ lies within the range [-1, 1], indicating the degree of relationship between two variables [24]:

- $\rho > 0$ represents a positive monotonic correlation, where both variables increase or decrease together.
- $\rho = 0$ indicates no correlation between the variables.
- $\rho < 0$ represents a negative monotonic correlation, where one variable increases as the other decreases.

B. Application of K-means Algorithm in Load Forecasting

Data reduction is a process to reduce the size of the original dataset while retaining important and necessary information for analysis. The purpose of data reduction is to minimize the volume of data that needs to be processed, thereby saving computational resources and speeding up processing. K-means is a popular clustering algorithm that partitions data into K clusters based on similarity between data points. Each cluster is represented by its centroid, and the goal of K-means is to minimize the sum of squared distances between data points and their respective centroids.

Figure 3 depicts the data processing model of the proposed method using the K-means algorithm with 5088 initial samples after filtering significant variables using Spearman correlation. After clustering, only clusters that meet the requirements and pass secondary criteria for selecting suitable training data clusters are retained to train the neural network. In [25], issues related to determining the number of clusters in the K-means algorithm were discussed. Several indices are used to determine the number of clusters, such as the Duda and Hart index, the Calinski and Harabasz index, the Davis-Bouldin index, the Silhouette index, and the SD index.

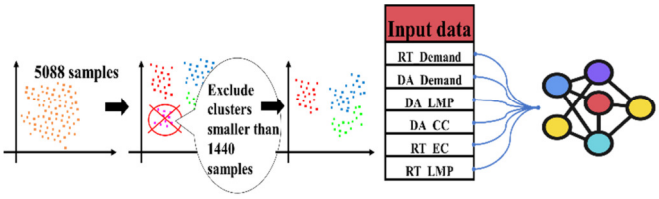

Fig. 3. Clustering and data reduction in the proposed method.

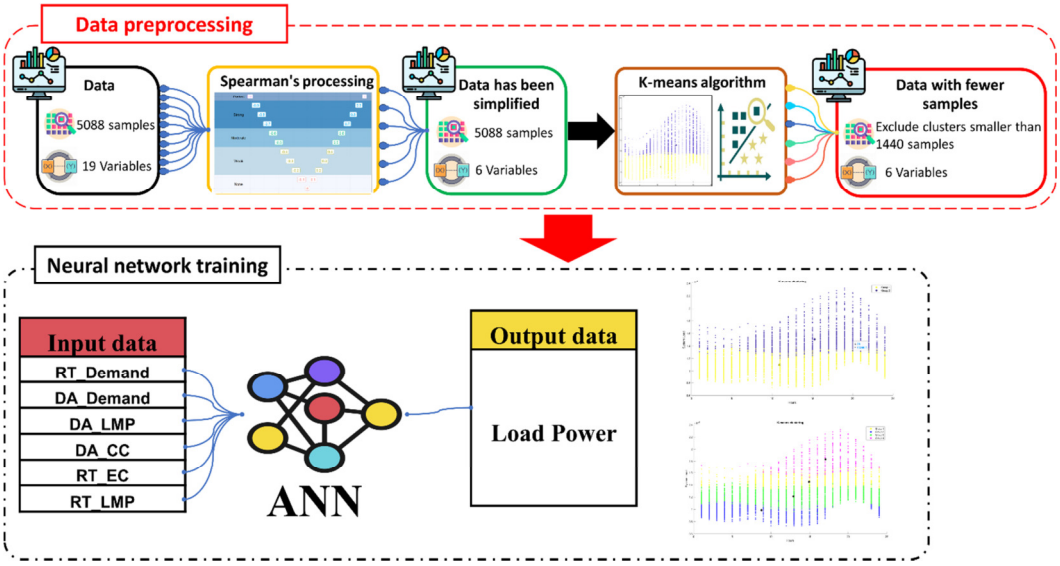

Fig. 4. Proposed forecasting model combining Spearman correlation and K-means.

After selecting important features, this study uses K-means for data clustering. The number of clusters is chosen to suit the complexity of the data and the computational capabilities of the model. Each cluster is represented by its centroid, which helps reduce the number of data samples while retaining important information. This process significantly reduces the size of the input data, improving computational efficiency for Artificial Neural Networks (ANNs).

$$Distance\ (a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \quad (3)$$

where $Distance(a,b)$ is the distance from data point a to cluster centroid b, a_i is the value of the i^{th} attribute in data point a, and b_i is the value of the i^{th} attribute in cluster centroid b.

III. PROPOSED LOAD FORECASTING MODEL

This study focuses on two main data processing issues: finding correlations among parameters to eliminate less relevant factors and retaining highly correlated factors to reduce noise, while also reducing the data sample size. Since daily electricity load data can repeat on certain days within a month or year, using only a portion of the data can achieve high forecasting performance. Additionally, this study uses ANNs, which outperform traditional methods in terms of speed and accuracy in load forecasting. Currently, research on load forecasting achieves very high accuracy, with errors below 3%. Figure 4 outlines the operational process of the proposed forecasting model based on these objectives. The data processing stage consists of four steps.

- Step 1: Data collection involves collecting 5068 samples and 19 variables.
- Step 2: Using the Spearman correlation to identify significant variables in the data. After this step, the dataset consists of 5068 samples and 8 variables. The dataset with the representative variables is used as input and output for the ANN. The network structure includes six input signals, which are the six representative electrical parameters of the grid (RT_Demand: Real-time Demand, DA_Demand: Day-Ahead Cleared Demand, DA_LMP: Day-Ahead Location Marginal Price, DA_CC: Congestion Component Day-Ahead, RT_EC: Energy Component of Real-time LMP, and RT_LMP: Real-time Location Marginal Price), and one output signal, which is the load power parameter of the grid.
- Step 3: The proposed method reduces the data by applying the K-means algorithm. The K-means algorithm divides the data set into smaller groups, ensuring that each group has at least one day of data (1440 samples) and each hour in the group has at least 60 samples.
- Step 4: The training dataset is fed into the ANN. The ANN's effectiveness is evaluated using benchmark results.

IV. SIMULATION RESULTS AND EVALUATION

The initial dataset obtained from ISO New England [18] yielded a MAPE with values from 5088 samples and 19 feature values. Table I presents the results of variable reduction using Spearman correlation.

TABLE I. INFORMATION ON TYPICAL VARIABLES BASED ON SPEARMAN CORRELATION

Spearman value				Feature value	Variable component name
1 - 0.6				6	RT_Demand
Perfect	-1		1		
Strong	-0.9		0.9		DA_Demand
	-0 .8		0.8		
	-0.7		0.7		DA_LMP
Moderate	-0.6		0.6		
	-0.5		0.5		DA_CC
Weak	-0.4		0.4		
	-0.3		0.3		RT_EC
	-0.2		0.2		
None	-0.1		0.1		RT_LMP

For every six variables, the sample size should be at least ten times the number of variables. Several studies have addressed this issue. In [26], it was suggested that the sample size should be at least 10-20 times the number of feature values, especially in logistic regression models. In [27], the importance of having enough samples was emphasized to ensure the stability and accuracy of the model, recommending at least 10-20 samples per feature. The value of K when clustering using K-means is determined based on the principle of six variables, each variable corresponding to 10 samples. For a 24-hour dataset, the minimum dataset size is 6×10×24=1440 samples. Therefore, the clustering simulation is performed from $K = 2$ to $K = 7$ because when dividing into other K values, no cluster meets the required number of samples.

According to these studies, the minimum requirement is 1440 samples for the training dataset. After clustering, the resulting sample sets that meet this requirement are printed in bold. Subsequently, each sample within the K clusters is selected hourly to ensure a minimum of 60 training samples per hour, as shown in Table III. Following this selection process, Group 1 contains 3242 samples with $K = 2$, and Group 2 contains 1893 samples with $K = 4$. Data groups with hourly sample values that do not meet the minimum of 60 samples are excluded. Subsequently, the ANN is trained. The ANN was simulated using MATLAB version 2023a, on a PC with a 12th Gen Intel(R) Core(TM) i5-12450H at 2.00 GHz processor and 16 GB RAM.

TABLE II. INFORMATION ON TYPICAL VARIABLES BASED ON SPEARMAN CORRELATION

Number of Clusters (*K*)	***K* = 2**	***K* = 3**	***K* = 4**	***K* = 5**	***K* = 6**	***K* = 7**
Number of samples in each group of *K* clustering value	**1846**	572	340	702	1268	586
	3242	**2419**	**1505**	1357	1030	833
		2097	1350	**1687**	1153	1138
			1893	207	554	152
				1135	888	930
					195	281
						1168

The clustered data were filtered to create sample sets to train the ANN. The number of neurons in the hidden layer was 10. Various algorithms were used for training, including Levenberg-Marquardt, Bayesian, Scaled Conjugate Gradient, and Resilient Backpropagation. The results of the networks, corresponding to the clustering cases, show that for Group 1 with 3242 samples, the lowest MAPE was 0.63%, and for Group 2 with 1893 samples, it was 0.31%. These MAPE values are within the range allowed for accurate forecasting.

TABLE III. HOURLY SAMPLE INFORMATION FOR CLUSTERS IN THE CASE OF K=2 AND K=4 CLUSTERING

Sample group/ Algorithm	**Group 1 = 3242**			**Group 2 = 1893**		
	Min	**Max**	**MAPE**	**Min**	**Max**	**MAPE**
Bayesian	0%	9.10%	0.43%	0%	8.56%	0.72%
Lenvenberg-Marquardt	0.01%	7.84%	0.36%	0%	8.07%	0.31%
Scaled Conjugate Gradient	0.00%	6.32%	0.64%	0%	8.09%	0.73%
Resilient Backpropagation	0.00%	8.11%	1.44%	0%	10.82%	1.13%

TABLE IV. HOURLY SAMPLE FOR REPRESENTATIVE CLUSTERS IN THE CASE OF *K* = 2 AND *K* = 4 CLUSTERING

	3242 (Receive)			**1893 (Receive)**	
***K* = 2**	1	193	***K* = 4**	1	82
	2	194		2	76
	3	200		3	66
	4	205		4	68
	5	204		5	83
	6	195		6	93
	7	151		7	98
	8	120		8	93
	9	120		9	84
	10	128		10	82
	11	124		11	76
	12	129		12	76
	13	125		13	67
	14	121		14	68
	15	117		15	62
	16	112		16	62
	17	91		17	65
	18	76		18	73
	19	66		19	68
	20	64		20	65
	21	67		21	73
	22	100		22	105
	23	160		23	123
	24	180		24	88

V. CONCLUSION

The experimental results indicate that with the initial ISO NewEngland dataset consisting of 5088 samples and 19 feature values, the proposed method achieved the best optimization with a MAPE of 0.31% using the Levenberg-Marquardt training model. The proposed method effectively reduced both the number of variables and the sample size, achieving a new dataset with a size of 1893 samples, reducing the total number of samples by 63% and the number of variables to 6. These results demonstrate that the proposed approach to data preprocessing using the Spearman correlation coefficient combined with K-means clustering is feasible and highly applicable for building datasets for machine learning models. This significantly improves the efficiency of load forecasting models.

ACKNOWLEDGMENT

This work belongs to project grant no. T2024-115, funded by Ho Chi Minh City University of Technology and Education, Vietnam.

REFERENCES

[1] N. T. Dung and N. T. Phuong, "Short-Term Electric Load Forecasting Using Standardized Load Profile (SLP) And Support Vector Regression (SVR)," *Engineering, Technology & Applied Science Research*, vol. 9, no. 4, pp. 4548–4553, Aug. 2019, https://doi.org/10.48084/etasr.2929.

[2] J. Chakravorty, S. Shah, and H. N. Nagraja, "ANN and ANFIS for Short Term Load Forecasting," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2818–2820, Apr. 2018, https://doi.org/10.48084/etasr.1968.

[3] N. H. Duong, M. T. Nguyen, T. H. Nguyen, and T. P. Tran, "Application of Seasonal Trend Decomposition using Loess and Long Short-Term Memory in Peak Load Forecasting Model in Tien Giang," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11628–11634, Oct. 2023, https://doi.org/10.48084/etasr.6181.

[4] M. Madhukumar, A. Sebastian, X. Liang, M. Jamil, and M. N. S. K. Shabbir, "Regression Model-Based Short-Term Load Forecasting for University Campus Load," *IEEE Access*, vol. 10, pp. 8891–8905, 2022, https://doi.org/10.1109/ACCESS.2022.3144206.

[5] A. A. El-Keib, X. Ma, and H. Ma, "Advancement of statistical based modeling techniques for short-term load forecasting," *Electric Power Systems Research*, vol. 35, no. 1, pp. 51–58, Oct. 1995, https://doi.org/10.1016/0378-7796(95)00987-6.

[6] D. G. Infield and D. C. Hill, "Optimal smoothing for trend removal in short term electricity demand forecasting," *IEEE Transactions on Power Systems*, vol. 13, no. 3, pp. 1115–1120, Dec. 1998, https://doi.org/10.1109/59.709108.

[7] J. Luo, T. Hong, Z. Gao, and S. C. Fang, "A robust support vector regression model for electric load forecasting," *International Journal of Forecasting*, vol. 39, no. 2, pp. 1005–1020, Apr. 2023, https://doi.org/10.1016/j.ijforecast.2022.04.001.

[8] Q. C. Lu, W. M. Grady, M. M. Crawford, and G. M. Anderson, "An adaptive nonlinear predictor with orthogonal escalator structure for short-term load forecasting," *IEEE Transactions on Power Systems*, vol. 4, no. 1, pp. 158–164, Oct. 1989, https://doi.org/10.1109/59.32473.

[9] S. Sharma, A. Majumdar, V. Elvira, and É. Chouzenoux, "Blind Kalman Filtering for Short-Term Load Forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4916–4919, Aug. 2020, https://doi.org/10.1109/TPWRS.2020.3018623.

[10] G. N. Shilpa and G. S. Sheshadri, "Electrical Load Forecasting Using Time Series Analysis," in *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, Vijiyapur, India, Jul. 2020, pp. 1–6, https://doi.org/10.1109/B-HTC50970.2020.9297986.

[11] C. Tarmanini, N. Sarma, C. Gezegin, and O. Ozgonenel, "Short term load forecasting based on ARIMA and ANN approaches," *Energy Reports*, vol. 9, pp. 550–557, May 2023, https://doi.org/10.1016/j.egyr.2023.01.060.

[12] S. M. M. Alam and Md. H. Ali, "A New Fuzzy Logic Based Method For Residential Loads Forecasting," in *2020 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, Chicago, IL, USA, Oct. 2020, pp. 1–5, https://doi.org/10.1109/TD39804.2020.9299999.

[13] C. Ieracitano, N. Mammone, A. Hussain, and F. C. Morabito, "A Convolutional Neural Network based self-learning approach for classifying neurodegenerative states from EEG signals in dementia," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, Jul. 2020, pp. 1–8, https://doi.org/10.1109/IJCNN48605.2020.9207167.

[14] M. L. Abdulrahman *et al.*, "A Review on Deep Learning with Focus on Deep Recurrent Neural Network for Electricity Forecasting in Residential Building," *Procedia Computer Science*, vol. 193, pp. 141–154, Jan. 2021, https://doi.org/10.1016/j.procs.2021.10.014.

[15] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306, https://doi.org/10.1016/j.physd.2019.132306.

[16] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proceedings of the 24th International Conference on Artificial Intelligence*, Buenos Aires, Argentina, Apr. 2015, pp. 3995–4001.

[17] R. A. Khan, C. L. Dewangan, S. C. Srivastava, and S. Chakrabarti, "Short Term Load Forecasting using SVM Models," in *2018 IEEE 8th Power India International Conference (PIICON)*, Kurukshetra, India, Sep. 2018, pp. 1–5, https://doi.org/10.1109/POWERI.2018.8704366.

[18] "Reliable electricity, competitive prices, clean-energy transition - ISO New England." https://www.iso-ne.com/.

[19] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. Cham, Switzerland: Springer International Publishing, 2015.

[20] K. F. Weaver, V. C. Morales, S. L. Dunn, K. Godde, and P. F. Weaver, *An Introduction to Statistical Analysis in Research: With Applications in the Biological and Life Sciences*. John Wiley & Sons, 2017.

[21] M. Moran, "Selecting Between Parametric and Non-Parametric Analyses," *Statistics Solutions*, Dec. 19, 2016. https://www.statisticssolutions.com/selecting-between-parametric-and-non-parametric-analyses/.

[22] P. Sedgwick, "Pearson's correlation coefficient," *BMJ*, vol. 345, Jul. 2012, Art. no. e4483, https://doi.org/10.1136/bmj.e4483.

[23] Y. Kim, T. H. Kim, and T. Ergün, "The instability of the Pearson correlation coefficient in the presence of coincidental outliers," *Finance Research Letters*, vol. 13, pp. 243–257, May 2015, https://doi.org/10.1016/j.frl.2014.12.005.

[24] H. Akoglu, "User's guide to correlation coefficients," *Turkish Journal of Emergency Medicine*, vol. 18, no. 3, pp. 91–93, Sep. 2018, https://doi.org/10.1016/j.tjem.2018.08.001.

[25] A. M. Mehar, K. Matawie, and A. Maeder, "Determining an optimal value of K in K-means clustering," in *2013 IEEE International Conference on Bioinformatics and Biomedicine*, Shanghai, China, Sep. 2013, pp. 51–55, https://doi.org/10.1109/BIBM.2013.6732734.

[26] J. S. Long and J. Freese, Regression Models for Categorical Dependent Variables Using Stata, Second Edition. Stata Press, 2006.

[27] F. E. Harrell, Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer International Publishing, 2015.