Outliers Detection and Imputation in Wind Speed Data and Forecasting Using LSTM

Bolla Vamsikrishna
Electrical Engineering
National Institute of Technology
Andhra Pradesh
Tadepalligudem, India
bollavamsi04@gmail.com

K. R. N. Manikanta
Electrical Engineering
National Institute of Technology
Andhra Pradesh
Tadepalligudem, India
manikantakesana0@gmail.com

D V N K Sai Kiran
Electrical Engineering
National Institute of Technology
Andhra Pradesh
Tadepalligudem, India
vnksaikiran@gmail.com

Krishna Mohan Reddy Pothireddy
Electrical Engineering
National Institute of Technology
Andhra Pradesh
Tadepalligudem, India
krishnamohanreddyp@gmail.com

Sandeep Vuddanti
Electrical Engineering
National Institute of Technology
Andhra Pradesh
Tadepalligudem, India
sandeep@nitandhra.ac.in

2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET) | 979-8-3503-8399-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/SEFET61574.2024.10718190

***Abstract*—In response to the global shift from non-renewable to renewable energy sources (RES), the utilization of wind turbines has emerged as a prominent solution which offers substantial environmental benefits. This research focuses on enhancing the reliability and efficiency of wind turbine systems through the application of machine learning (ML). Specifically, the authors aim to detect outliers and perform imputations in the data generated by these systems. Outliers detection allows to identify irregularities or unexpected deviations in the data generated by these systems. By leveraging the power of ML, the power systems engineers can proactively identify issues such as equipment malfunctions, sensor inaccuracies, or external factors affecting performance. Moreover, imputation techniques are employed to handle missing or incomplete data, ensuring that the dataset remains comprehensive and reliable for analysis. This method ensures the sustained performance and durability of these RES, supporting the transition to a more sustainable and eco-friendly energy.**

Keywords— Outliers, Imputation, Renewable Energy, Outlier Removal, stability

I. Introduction

Due in large part to extensive electrification and electric motors, the electrical grid expansion of the 20th century was significantly dependent on non-renewable resources. Renewable energy sources (RES) like wind, solar, and hydro power are becoming more and more important due to factors including population increase, rising energy consumption, and depleting fossil fuel reserves [1]. Although the environment benefits from such RES, meeting real-time energy demand becomes more difficult due to their inherent volatility [2]. For grid stability and efficient resource use, accurate forecasting is especially important when it comes to wind power [3, 4]. However, outliers such as unforeseen meteorological phenomena, equipment failures, and missing data points make it difficult to produce accurate forecasts [5]. Thus, improving the accuracy and dependability of renewable energy forecasting models requires strong outlier detection and imputation strategies [6]. The authors of this research explore how deep learning is used to overcome these difficulties in wind energy forecasting [7]. Through the efficient resolution of these problems, deep learning aids in the advancement of more precise and dependable wind energy forecasting models [8]. Thus, better grid energy management is enabled, making it easier to incorporate RES smoothly and opening the door to a more robust and sustainable energy future.

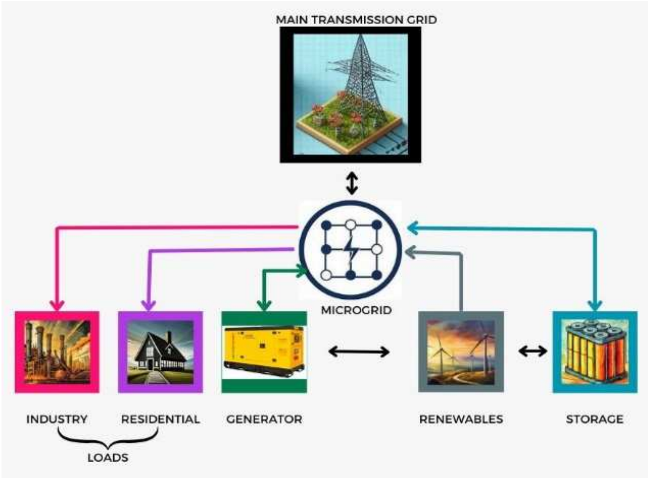

Fig. 1. Structure of a simple grid connected microgrid.

Microgrids (MG) are used for many purposes in energy distribution and management [9]. Future communities may rely on MG during the event of main grid failure. MG have localized and self-contained power systems, mainly useful in remote areas where grid access is limited [10]. It supports in integrating RES such as solar, wind, and hydropower which is an environment-friendly energy mix. MG's able to meet the supply and demand of electricity more efficiently through advanced energy management which helps in cost saving and better energy utilization. They also help the grid by regulating frequency and voltage control indeed in grid stability. MGs are prepared for emergencies, in case of disasters, hospitals, and any emergency service centres with MG operate independently providing necessary services in crisis. MG adjust the usage of electricity based on pricing, grid conditions, or other factors.

Outliers are the points that depart from the rest of the data points in the dataset [11]. These are caused due to many reasons such as sensor malfunctions, data entry issues, human errors, etc. In some cases, outliers occur naturally due to the inherent variability in the data [12]. system malfunctions in industries and glitches in technological systems can also lead to outliers. Identifying the root cause of outliers is very important to understand their nature and decide whether to exclude them for analysis or investigate the data further.

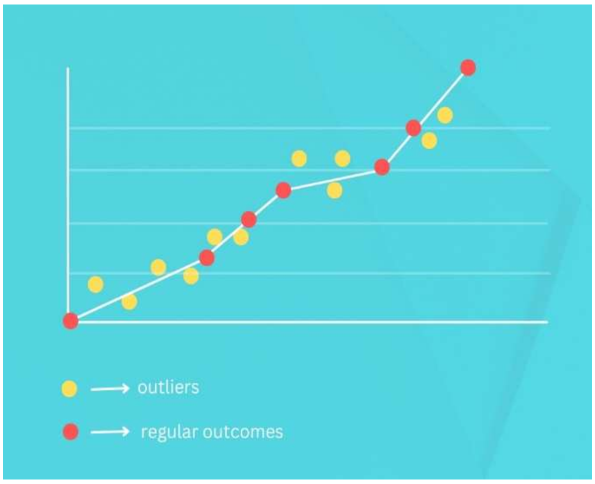

Fig. 2. Illustration of outlier and regular outcome.

Fig. 2 indicates the deviation of data points from other data points which is termed as outliers. In research, missing data is a common adversary that can undermine the most accurately compiled datasets. Ignoring these gaps taints analyses, creates bias, and imperils results. In these kinds of situations, imputation, the ability of filling in these missing values aids in correct forecasting. There are many different approaches, each with advantages and disadvantages, ranging from straightforward mean/median substitutes to complex machine learning operations. The sort of missingness, research issue, and type of data influences which approach is the best. Aforesaid that imputation [13], is about making refined decisions, to guarantee reliable results, it is necessary to perform the sensitivity analyses, and consider multiple imputations.

II. Impact of Outliers on the Forecasting Techniques

Outliers impact forecasting techniques, like deviating the accuracy and reliability of prediction. Outliers in the data can distort the overall pattern. It also increases the error metrics such as root mean square error (RMSE), mean square error (MSE) and mean absolute error (MAE) [14]. Outliers can cause artificial trends in the data which creates a disproportionate impact on some forecasting methods that are sensitive to extreme values. Outliers can introduce additional variability in data which leads to less stable and more uncertain predictions [15]. In an attempt to accommodate the outliers forecasting models may become overly complex and overfit the training data.

Statistical methods of forecasting can be seen as predicting future trends and outcomes using historical data [16]. This type of method mainly depends on mathematical models and statistical techniques to analyse the relation with historical data. Time series analysis such as time series decomposition and moving averages, regression analysis, ARIMA, Bayesian method, and Monte Carlo simulation [17]. The preference for forecasting methods depends on the nature of the data, the underlying patterns, forecasting horizon. sometimes combination of methods or model selection based on cross-validation is used to improve the accuracy of prediction [18].

III. Modelling of Microgrid

A. Dispatchable energy sources

Dispatchable energy sources are the on-demand powerhouses that can quickly modify their production to meet the constantly fluctuating demands for electricity. Consider them as adaptable members of the grid team. The power output of dispatchable sources, like fossil fuel, nuclear power, and hydro power-based generation, can be adjusted, in contrast to solar and wind power, which are dependent on sunshine and winds. While some, like gas and coal, have negative environmental effects, greener alternatives, including geothermal and biomass, have potential.

B. Non-dispatchable energy sources

The output power from non-dispatchable energy sources is uncontrollable. Further, these sources are sporadic, volatile and highly intermittent. Although they are eco-friendly, their unpredictability presents difficulties. Batteries store the excess energy from solar and wind energy and discharge it when needed. Smart grids adapt based on their predictions of weather patterns. By switching to off-peak hours for their electricity use, consumers even contribute to the solution.

IV. Solution Methodologies

A. Convolutional neural networks (CNNs)

Traditionally CNNs have been used for image recognition. However, they can also be used to detect patterns within time series data. CNNs treat time series as a one-dimensional sequence, and they can extract local patterns as well as relationships between data points. This allows them to detect outliers, classify them, and even forecast them. Pre-processing steps such as normalization and preparation make the data ready for analysis. Dilated convolutions solve the problem of capturing long-distance dependencies. While CNNs are limited in interpretability and require careful data preparation, they offer several advantages for time series data. These include effective local pattern extraction, higher dimensional data handling, parallelization capabilities, and more. Fig. 2 shows the CNN based outlier detection.

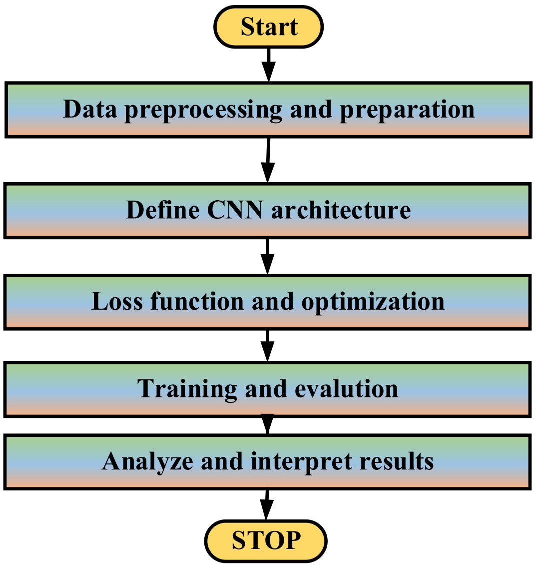

Fig. 3. Convolutional neural network based forecasting.

B. Recurrent neural network

Unlike their traditional counterparts, RNNs possess a unique ability to remember, allowing them to process information in the context of what came before. This memory" empowers them to grasp the nuances of human language, making them adept at understanding complex narratives, generating creative text formats, and even translating languages. Their ability to learn and adapt based on past experiences mirrors the human mind in its capacity for learning and growth. Biases inherent in training data could be amplified, leading to skewed or inaccurate outputs. Moreover, the very technology that grants RNNs their power could also pose a threat, potentially leading to the creation of harmful or manipulative content. Therefore, it is crucial to approach the integration of RNNs into my writing with a responsible and ethical mindset

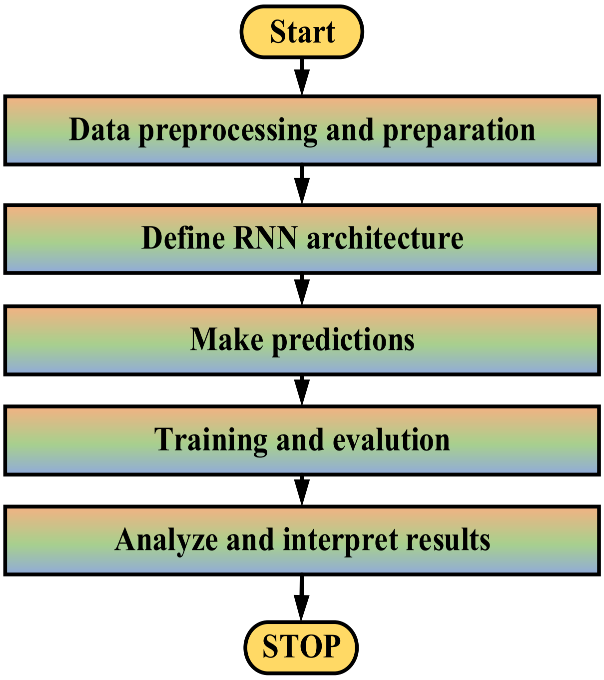

Fig. 4. Recurrent neural network-based forecasting.

C. Long-short term memory

LSTMs, or long transient memory organizations, address a jump forward in man-made consciousness. Dissimilar to customary brain organizations, they have the striking skill to acquire and adjust in light of past data, making them experts at handling successive information. This power converts into uncommon execution in assignments like language interpretation, discourse acknowledgment, and time series gauging. LSTMs succeed at grasping the specific circumstance and stream of data, meshing together apparently different pieces into a rational entirety. Envision a discussion where each sentence expands upon the last, prompting a more profound and more extravagant comprehension. LSTMs accomplish this by examining groupings of data, not as detached elements, but rather as interconnected strings woven into the texture of a bigger story. Their true capacity stretches out a long way past simple specialized accomplishments.

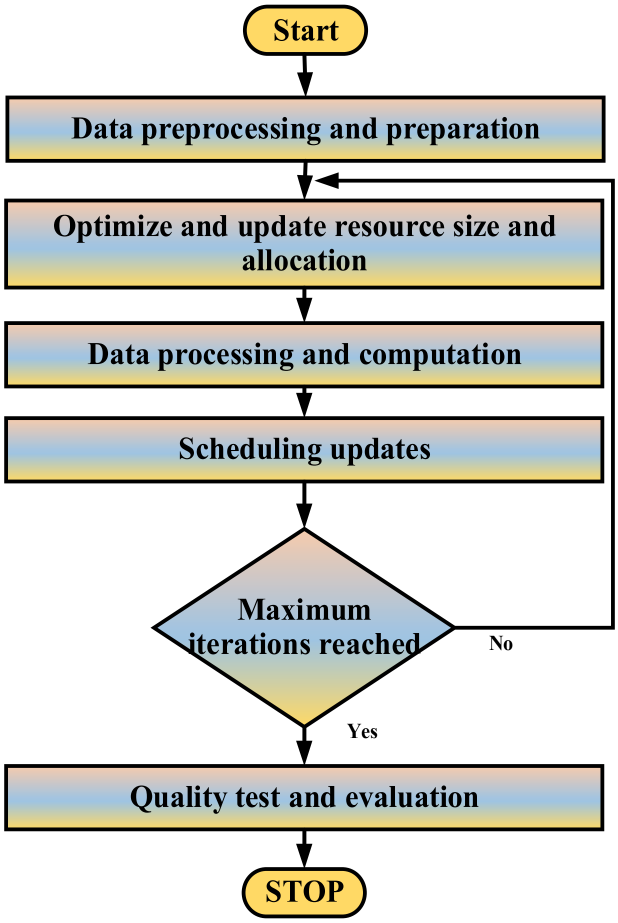

Fig. 5. LSTM based forecasting.

This study analysed a year's worth of wind speed data collected at 30-minute intervals, resulting in 17,520 samples, which were then divided into 70% for testing and 30% for validation, with Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared (R2) scores serving as evaluation metrics.

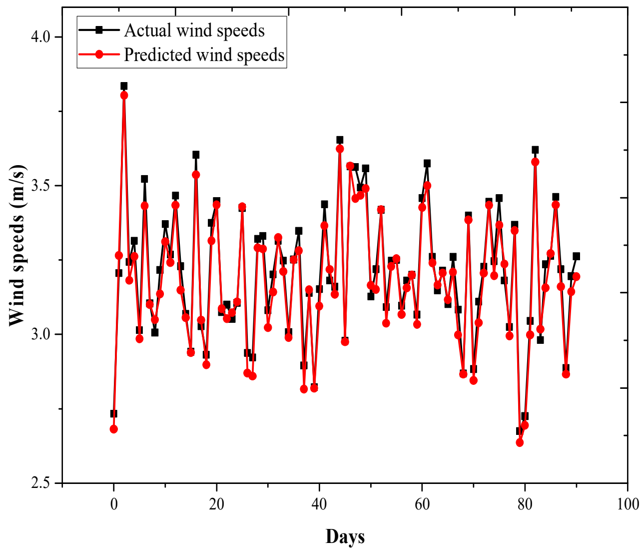

Fig. 6. Forecasting using LSTM without filtering outliers.

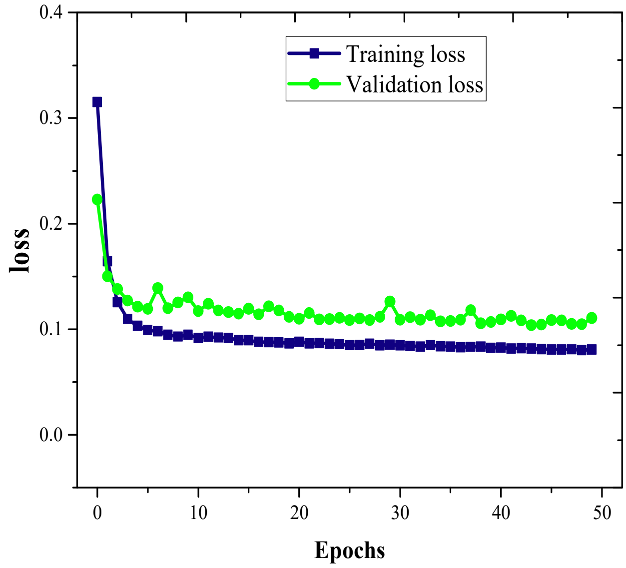

Fig. 7. Comparison of train and validation loss using LSTM without filtering outliers.

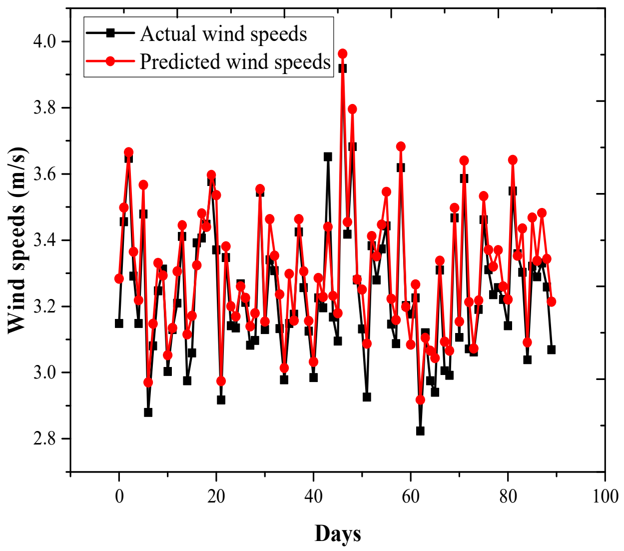

Fig. 8. Forecasting using LSTM filtering outliers.

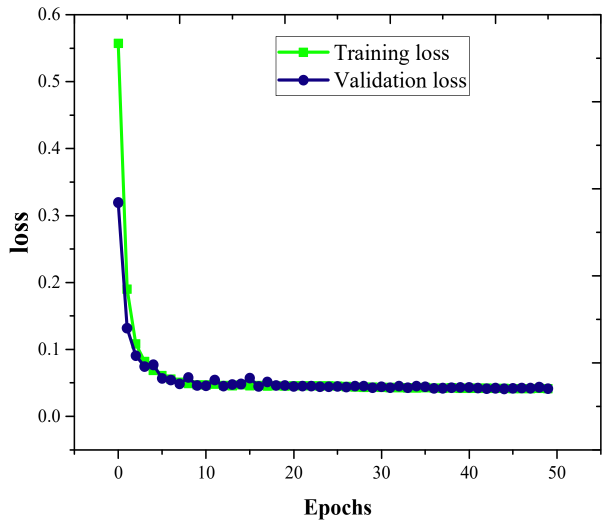

Fig. 9. Comparison of train and validation loss using LSTM filtering outliers.

A histogram was plotted using the raw dataset to visualize error distribution, identifying samples with errors exceeding 0.5 as outliers. These outliers were subsequently replaced with the mean value of the dataset for outlier treatment. After treating outliers in the raw dataset, the model's performance was reassessed, repeating the process of MAE and R2 score calculation, revealing improved accuracy with decreased MAE and MSE, and an increased R2 score, indicating the positive impact of outlier treatment on forecasting performance.

TABLE I. COMPARISON METRICS WITH AND WITHOUT OUTLIER DETECTION AND IMPUTATION

Metrics	**LSTM**	
	Synthetic dataset	***Dataset after outliers filtering***
MSE	0.04085	0.00126
MAE	0.13256	0.01861
R-squared	0.95264	0.98095

CONCLUSION

The results underscore the importance of outlier treatment in enhancing the accuracy of wind speed forecasting models, as outliers can distort underlying data patterns, leading to inaccurate predictions. By identifying and treating outliers, the model's ability to capture true trends in the data is enhanced, resulting in more reliable forecasts. Additionally, the methodology used for outlier treatment is simple and easy to implement, involving the replacement of outliers with the mean value of the dataset, a straightforward technique applicable to various datasets, without requiring complex algorithms or extensive computational resources, thereby making it accessible to researchers and practitioners alike, ultimately improving the reliability of wind speed forecasting models for applications such as renewable energy generation and weather forecasting.

REFERENCES

[1] P. K. Mohan Reddy and M. Prakash, "Optimal Dispatch of Energy Resources in an Isolated Micro-Grid with Battery Energy Storage System," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 730-735, doi: 10.1109/ICICCS48265.2020.9121176.

[2] Pothireddy, K.M.R.; Vuddanti, S.; Salkuti, S.R. Impact of Demand Response on Optimal Sizing of Distributed Generation and Customer Tariff. *Energies* **2022**, *15*, 190. https://doi.org/10.3390/en15010190.

[3] K. M. R. Pothireddy and S. Vuddanti, "Optimal re-distribution of incentives to the demand responsive loads during overlap," 2023 9th International Conference on Electrical Energy Systems (ICEES), Chennai, India, 2023, pp. 240-243, doi: 10.1109/ICEES57979.2023.10110068.

[4] K. M. R. Pothireddy, A. R. Battula and S. Vuddanti, "Demand response: What is the optimal curtailment for economical benefit?," 2023 9th International Conference on Electrical Energy Systems (ICEES), Chennai, India, 2023, pp. 235-239, doi: 10.1109/ICEES57979.2023.10110210.

[5] S. R. Bull, "Renewable energy today and tomorrow," in Proceedings of the IEEE, vol. 89, no. 8, pp. 1216-1226, Aug. 2001, doi: 10.1109/5.940290.

[6] Wang, D.; Cui, X.; Niu, D. Wind Power Forecasting Based on LSTM Improved by EMD-PCA-RF. Sustainability 2022, 14, 7307. https://doi.org/10.3390/su14127307.

[7] Liu, Y.; Guan, L.; Hou, C.; Han, H.; Liu, Z.; Sun, Y.; Zheng, M. Wind Power Short-Term Prediction Based on LSTM and Discrete Wavelet Transform. Appl. Sci. 2019, 9, 1108. https://doi.org/10.3390/app9061108

[8] H. Wilms, M. Cupelli, A. Monti and T. Gross, "Exploiting Spatio-Temporal Dependencies for RNN-based Wind Power Forecasts," 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), Bangkok, Thailand, 2019, pp. 921-926, doi: 10.1109/GTDAsia.2019.8715887.

[9] K. Moharm, M. Eltahan and E. Elsaadany, "Wind Speed Forecast using LSTM and Bi-LSTM Algorithms over Gabal El-Zayt Wind Farm," 2020 International Conference on Smart Grids and Energy Systems (SGES), Perth, Australia, 2020, pp. 922-927, doi: 10.1109/SGES51519.2020.00169.

[10] V. S R and P. R, "Week Ahead Wind Power Forecast Using A Time-Series Forecasting Model," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 633-638, doi: 10.1109/ICACCS57279.2023.10112685.

[11] Y. Cao and L. Gui, "Multi-Step wind power forecasting model Using LSTM networks, Similar Time Series and LightGBM," 2018 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 2018, pp. 192-197, doi: 10.1109/ICSAI.2018.8599498.

[12] J. W.G and J. J.E, "Performance Analysis on Wind Power Forecasting in Machine Learning and Deep Learning," 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 2023, pp. 592-596, doi: 10.1109/ICCPCT58313.2023.10245893.

[13] K. M. San, J. G. Singh and K. Prakash N., "Wind Speed Forecasting using Hybrid Model of CNN and LSTM with Wavelets," 2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT), Bhubaneswar, India, 2023, pp. 297-301, doi: 10.1109/APSIT58554.2023.10201713.

[14] I. Koprinska, D. Wu and Z. Wang, "Convolutional Neural Networks for Energy Time Series Forecasting," 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1-8, doi: 10.1109/IJCNN.2018.8489399.

[15] Lakshminarayan, K., Harp, S.A. & Samad, T. Imputation of Missing Data in Industrial Databases. Applied Intelligence 11, 259–275 (1999). https://doi.org/10.1023/A:1008334909089

[16] Hodge, V., Austin, J. A Survey of Outlier Detection Methodologies. Artificial Intelligence Review 22, 85–126 (2004). https://doi.org/10.1023/B:AIRE.0000045502.10941.a9

[17] H. Wang, M. J. Bah and M. Hammad, "Progress in Outlier Detection Techniques: A Survey," in IEEE Access, vol. 7, pp. 107964-108000, 2019, doi: 10.1109/ACCESS.2019.2932769