International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 6, December 2025, pp. 5067~5079
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i6.pp5067-5079

Impact of outlier detection techniques on time-series forecasting accuracy for multi-country energy demand prediction

Shreyas Karnick[1], Sanjay Lakshminarayanan[1], Madhu Palati[1], Prakash R[2]
[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, MS Institute of Technology and Management, Visvesvaraya Technological University, Belagavi, India
[2]Department of Electrical and Engineering, Acharya Institute of Technology, Bengaluru, India

Article Info

Article history:

Received Mar 13, 2025
Revised Jul 15, 2025
Accepted Sep 14, 2025

Keywords:

Energy demand prediction
Load forecasting
Outlier detection
Root mean squared error
Time-series forecasting

ABSTRACT

Accurate energy demand prediction is crucial for efficient grid management and resource optimization, particularly across multiple countries with varying consumption patterns. However, real-world energy demand data often contains outliers that can distort forecasting accuracy. This study evaluates the impact of five outlier detection techniques—Z-Score, density-based spatial clustering of applications with noise (DBSCAN), isolation forest (IF), local outlier factor (LOF), and one-class support vector machine (SVM)—on the performance of three time-series forecasting models: long short-term memory (LSTM) networks, convolutional neural network (CNN) Autoencoders, and LSTM with attention mechanisms. The models are tested using energy demand data from four European countries—Germany, France, Spain, and Italy—derived from real-time consumption records. A comparative analysis based on root mean squared error (RMSE) demonstrates that incorporating outlier detection significantly enhances model robustness, reducing forecasting errors caused by anomalous data. The findings emphasize the importance of selecting appropriate outlier detection strategies to improve the accuracy and reliability of energy demand forecasting. This research provides valuable insights into the trade-offs involved in outlier removal, with implications for policy and operational practices in energy management.

Corresponding Author:

Shreyas Karnick
Department of Electrical and Electronics Engineering, Faculty of Engineering, MS Institute of Technology and Management, Visvesvaraya Technological University
Belagavi 560018, Karnataka, India
Email: shreyaskarnick2022@gmail.com

1. INTRODUCTION

Accurate energy demand forecasting is a critical component of modern power system management. It facilitates optimal grid operation, ensures the reliability of power distribution, and supports the integration of renewable energy sources by predicting fluctuations in energy consumption. With the increasing complexity of global energy systems, forecasting models must account for the dynamic nature of energy demand across different regions, influenced by factors such as climate, socio-economic activities, and technological advancements. In particular, multi-country energy demand prediction presents a unique challenge, as each country exhibits its own consumption patterns and responds differently to various influencing factors. Effective forecasting models are therefore essential for achieving greater accuracy and efficiency in managing energy resources at the regional and international levels.

One significant challenge in time-series forecasting is the presence of outliers, which are abnormal data points that deviate significantly from expected patterns. Outliers in energy demand data may arise due to measurement errors, system failures, or rare events such as natural disasters or sudden changes in consumption patterns. These anomalous values, if left untreated, can severely impact the performance of forecasting models, leading to inaccurate predictions and inefficient decision-making. As such, addressing outliers is a vital preprocessing step for enhancing the accuracy and robustness of time-series forecasting models. Various outlier detection techniques, including statistical methods and machine learning approaches, have been developed to identify and mitigate the effects of outliers in time-series data.

In recent years, the application of machine learning models such as long short-term memory (LSTM) networks, convolutional neural networks (CNNs), and attention mechanisms has gained significant attention in energy demand forecasting. These models have demonstrated the ability to capture complex temporal dependencies and provide highly accurate predictions. However, the performance of these models is highly sensitive to the quality of the input data, particularly in the presence of outliers. Despite the advances in forecasting techniques, the impact of outlier detection methods on model performance, particularly for multi-country energy demand prediction, has not been extensively explored. In this context, the present study aims to investigate the role of different outlier detection techniques in improving the accuracy of energy demand forecasting for multiple countries.

This study focuses on evaluating five popular outlier detection techniques—Z-score, density-based spatial clustering of applications with noise (DBSCAN), isolation forest (IF), local outlier factor (LOF), and one-class support vector machine (SVM)—on their ability to improve the accuracy of energy demand predictions using three state-of-the-art time-series forecasting models: LSTM, CNN Autoencoders, and LSTM with attention. By analyzing real-time energy consumption data from four European countries—Germany, France, Spain, and Italy—the study provides insights into the effectiveness of these outlier detection methods in enhancing forecasting accuracy. The results of this study are expected to contribute valuable knowledge to the field of energy systems, offering a better understanding of the relationship between outlier detection and forecasting performance, with implications for both academic research and practical applications in the energy sector. The paper is organized as: section 2 reviews related work; section 3 describes the methodology; section 4 presents and discusses the results discussions; section 5 concludes the paper and suggests future directions.

2. LITERATURE REVIEW

Outlier detection in time series data is a critical component in enhancing the accuracy of forecasting models, particularly in applications like energy demand prediction. Many studies emphasize the impact of selecting effective outlier detection techniques on the performance of forecasting models. The study by Amalou *et al.* [1] presents the fast incremental support vector data description (FISVDD) algorithm for outlier detection, demonstrating its effectiveness in energy time series forecasting. The research highlights that choosing the appropriate kernel function for the FISVDD model significantly improves forecasting accuracy. This improvement is validated using the mean squared error (MSE) evaluation, which shows that FISVDD outperforms other outlier detection techniques. By selecting the right kernel function, the method effectively handles irregularities in energy consumption data, leading to superior results in multi-country energy demand forecasting.

Bandhan and Ganapati [2] have discussed outlier detection techniques. The study classifies outlier detection techniques into five major categories: statistical methods, distance-based approaches, density-based methods, clustering-based techniques, and ensemble methods. Each technique offers unique advantages for identifying anomalous data points. The adaptability of the methods for different domains of the study gains interest in further investigation. Richard [3] also explores the various techniques and their advantages and limitations. The authors in [4] have discussed an enhanced technique, called unsupervised outlier detection architecture with graph neural network (UOSC-GNN).

The authors of reference [5] have discussed anomaly detection techniques and compared the methods, IF, gaussian mixture model (GMM), and k-nearest neighbor (kNN) algorithms and concludes that IF outperforms both GMM and kNN in effectively isolating outliers from data. Outlier detection based on local density and natural neighbors have been discussed in [6], wherein a knowledge of knowing parameter K, for addressing challenges in existing methods related to parameter selection. In this work, manual parameter setting required for neighborhood parameter K is not required. Another work that integrates clustering and outlier scoring schemes, specifically using uncertainty soft clustering based on rough set theory is reported in [7]. The work discusses a Kernel Rough Clustering algorithm, demonstrating superior detection accuracy compared to five existing methods.

A single density network (SDN) and Z-score for outlier detection in analog tests is presented in [8], it introduces metrics like self-excluded fail rate (SE fail rate) and normalized area under curve (AUC) to quantify and visualize abnormality effectively. The techniques include IFs, which utilize binary decision trees to isolate anomalies, crucial for various applications [9], [10].

In [11], an electricity price forecasting of Danish electricity market, utilizing a GMM-lightweight gradient boosting machine hybrid detector and LSTNet-kernel density estimation (LSTNet-KDE) method, which enhances forecasting accuracy by effectively isolating and predicting outlier sequences is presented. On the other hand, RF algorithm for outlier detection is presented in [12]. Single-valued metric prediction is presented in [13], which enhances the accuracy of time-series forecasting [14], [15] in various applications, including energy demand prediction. Another algorithm, called fast incremental FISVDD is discussed in [16] for enhancing the forecasting accuracy.

Another significant contribution to this field is the hybrid model proposed by Songhua [17], which combines IF with outlier reconstruction (OR), CNN, and random forest (RF) for energy demand forecasting. This model, denoted as IF-OR-CNN-RF, demonstrates superior performance metrics, such as mean absolute error (MAE) and root mean squared error (RMSE), compared to other CNN-based models. The study underscores that integrating outlier detection methods with deep learning techniques enhances the robustness of forecasting models, particularly in the presence of outliers. This hybrid approach addresses challenges inherent in energy demand prediction by mitigating the influence of abnormal data points, leading to more reliable and accurate forecasts.

In a similar vein, Li *et al.* [18] proposes the CNN-gated recurrent unit (CNN-GRU) method, coupled with a random forest detection model optimized by grid search (CGA-RF), for anomaly detection in energy consumption data. Their study reveals significant improvements in performance metrics such as accuracy, precision, recall, and F1 score, compared to conventional methods. The use of a self-attentive mechanism in the CNN-GRU model helps in capturing dynamic changes in energy consumption, while the random forest model excels in detecting anomalies in residuals, ultimately boosting forecasting accuracy. The authors emphasize that handling anomalies effectively is crucial for enhancing energy management and operational efficiency in energy systems.

Fu *et al.* [19] contribute to this area by presenting a tree-based anomaly detection model, which was the winning solution in the large-scale energy anomaly detection (LEAD) competition. This method achieved a high ROC-AUC score of 0.9866, underscoring its efficacy in identifying outliers in energy time series. The study emphasizes the importance of feature engineering, particularly through value-changing features that capture variations in time series data. This research highlights the need for effective data preprocessing and anomaly detection to ensure the accuracy of energy consumption forecasting models.

Similarly, Gao *et al.* [20] explore outlier detection through correlation analysis based on graph neural networks (GNNs). Their proposed UOSC-GNN architecture improves anomaly detection by measuring the variance between expected and actual data states, showing improvements in accuracy and sensitivity. Although the study does not directly address energy forecasting, the techniques discussed are applicable in energy demand prediction, especially in identifying anomalous patterns that may influence forecasting models.

The integration of machine learning techniques for outlier detection in energy time series is further explored in the work of Ismaeel *et al.* [21], which investigates the scientific computing associates (SCA) statistical system for outlier detection in the context of water volume forecasting for the Dohuk Dam. While the primary focus is on water volume data, the principles of outlier correction in time series analysis can be applied to energy forecasting [22]. The paper demonstrates that outlier-adjusted forecasts perform better, enhancing the accuracy of time-series models by correcting for abnormal data points that would otherwise lead to forecasting errors.

Kyo [23] have presented a multi-objective optimization approach combining minimum index of symmetry and uniformity (ISU) and maximum likelihood autoregressive (AR) modeling for detecting outliers in nonstationary time series to decomposes trend and stationary components while balancing outlier detection and model selection. A deep learning framework using autoencoders and LSTM networks to detect anomalies in time series data is discussed in [24]. The hybrid model captures complex temporal patterns through reconstruction errors, enhancing reliability across applications. Kumar *et al.* [25] developed an ARIMA-DCGAN synergy that leverages ARIMA's linear modeling and DCGAN's nonlinear capabilities for outlier detection in time series. This approach outperforms existing methods, benefiting applications like fraud detection and predictive maintenance. Dani *et al.* [26] developed an ARIMA-DCGAN synergy that leverages ARIMA's linear modeling and DCGAN's nonlinear capabilities for outlier detection in time series. This approach outperforms existing methods, benefiting applications like fraud detection and predictive maintenance. Dani *et al.* [26] employs principal component analysis (PCA) for anomaly detection in time series by reducing dimensionality and highlighting deviations. This technique aids in timely risk mitigation and informed decision-making in organizational contexts.

Current literature typically focuses either on forecasting models (like LSTM or CNNs) or on anomaly detection methods in isolation. Our work bridges this gap by demonstrating how outlier detection directly improves model performance and quantifying these improvements across models and countries. It also reveals model-specific sensitivities — for example, LSTM-Attention models, while powerful, are more sensitive to outliers, a nuance not well documented before. Thus, this manuscript adds practical knowledge for both researchers and practitioners on how to select and combine anomaly detection and forecasting methods effectively, contributing to more resilient and accurate energy system operations.

3. METHODOLOGY

In this work, our methodological framework involves: i) real-world energy consumption datasets from four European countries, ii) systematic preprocessing (normalization, interpolation, outlier removal), iii) mathematical formulation and implementation of each forecasting and outlier detection technique and iv) evaluation metrics mean absolute percentage error (mean absolute percentage error (MAPE), RMSE, MAE) to quantify improvements in predictive accuracy.

The proposed methodology for energy demand forecasting, as illustrated in Figure 1, follows a structured approach to process multi-country energy demand datasets efficiently. The process begins with data preprocessing, where normalization and train-test splitting are performed to standardize the dataset for model training. Subsequently, outlier detection techniques such as Z-score, DBSCAN, IF, LOF, and SVM are applied to identify anomalous data points. A decision-making step determines whether outliers are present, influencing the choice of forecasting models. If outliers are detected, advanced models such as LSTM, CNN Autoencoder, and LSTM with Attention mechanism are utilized to enhance forecasting accuracy. These models leverage deep learning techniques to capture temporal dependencies and complex patterns within the dataset. Model performance is evaluated using RMSE, MAPE, and MSE to ensure robust predictive accuracy. The final step consolidates the forecasting results, providing insights into energy demand trends across multiple regions.

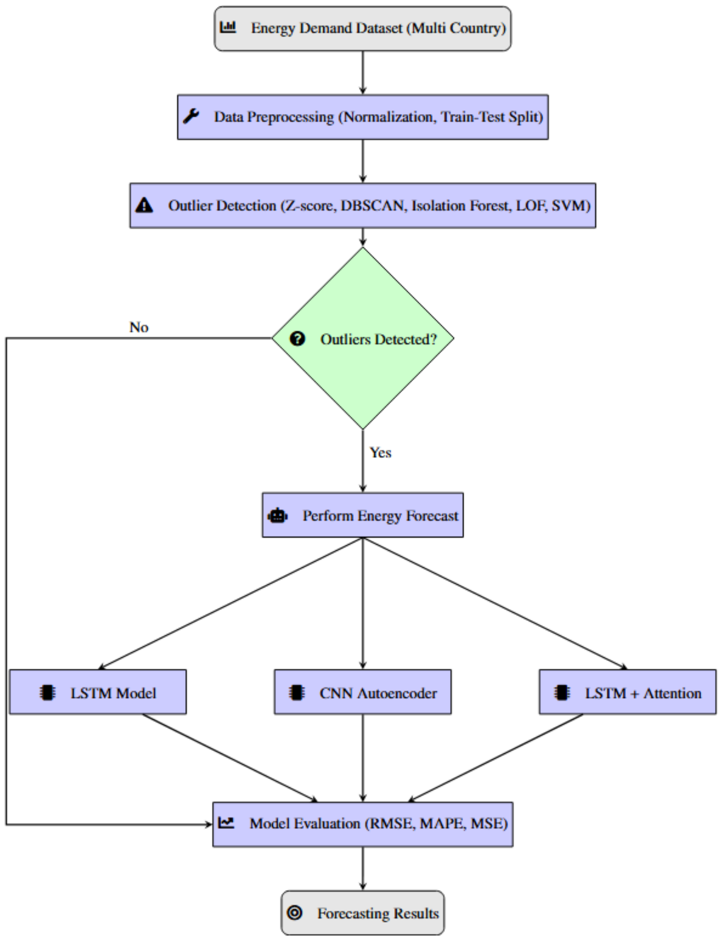

Figure 1. Flow chart of the proposed methodology for energy demand forecasting

3.1. Data collection and preprocessing
3.1.1. Dataset description
The dataset used in this study comprises hourly energy consumption data from multiple European nations, specifically Germany (DE), France (FR), Spain (ES), and Italy (IT). The data was acquired from the open power system data (OPSD) repository, a widely recognized source for energy-related time-series datasets. This dataset provides historical electricity demand values recorded at an hourly resolution, ensuring a granular understanding of energy consumption trends. Given the dynamic nature of energy demand, factors such as seasonal variations, economic activity, and climatic conditions influence consumption patterns. Thus, an in-depth exploration of these variations is necessary for enhancing forecasting accuracy and robustness.

3.1.2. Data preprocessing
Handling missing values: missing timestamps are filled using linear interpolation, while energy demand values are imputed using forward and backward filling. Data normalization: to ensure consistency across different countries and eliminate scale disparities, the raw time-series data undergoes normalization. The Min-Max normalization technique is applied, which scales values between 0 and 1, thereby facilitating stable convergence in deep learning models. The transformation of an energy demand value x_i is computed as (1).

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

In (1), x_{min} and x_{max} represent the minimum and maximum values within the dataset, respectively. This normalization mitigates the impact of large-scale discrepancies among different countries while preserving the relative magnitude of fluctuations, ensuring optimal model performance.

3.2. Outlier detection techniques
Outliers in time-series data arise due to various factors, including sensor malfunctions, erroneous recordings, grid failures, or unforeseen spikes in energy consumption. Failure to address these anomalies can lead to inaccurate predictions and model instability. This study examines five robust outlier detection methods Z-score, DBSCAN, IF, LOF, and one-class SVM. Each technique identifies anomalies based on distinct mathematical formulations and underlying principles.

3.2.1. Z-Score method
The Z-score method is a statistical technique that quantifies the deviation of each data point from the mean in terms of standard deviations. This approach assumes that energy consumption data follows a normal distribution, allowing the identification of extreme deviations. The Z-score for each value of x_i is computed as (2).

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (2)$$

In (2), μ is the mean and σ is the standard deviation. Data points with $|Z_i| > 3$ are classified as outliers.

3.2.2. DBSCAN
DBSCAN is a clustering-based anomaly detection technique that distinguishes normal and anomalous points based on data density. A point is considered an outlier if it does not belong to any high-density cluster. The algorithm relies on a neighborhood function (3).

$$(p) = \{q \in D \mid d(p, q) \leq \varepsilon\} \quad (3)$$

In (3), $d(p, q)$ denotes the distance between data points, and ε is a predefined threshold. Points with fewer than min samples neighbors are labeled as outliers. This method is particularly effective for detecting anomalies in datasets exhibiting nonlinear structures.

3.2.3. IF
IF is an ensemble learning technique that isolates anomalies by recursively partitioning data points. Unlike traditional methods that rely on distance metrics, IF constructs decision trees where anomalous points are identified through shorter path lengths. The anomaly score is given by (4).

$$s(x, n) = 2^{-\frac{E(h(x))}{(c(n))}} \quad (4)$$

In (4), $h(x)$ depends on x in the forest and $c(n)$ is the average path length for a dataset of size n. IF is computationally efficient and highly effective for high-dimensional data.

3.2.4. LOF

LOF assesses outliers by comparing the density of a point with its surrounding neighbors. A low-density point relative to its neighbors is flagged as an outlier. The LOF score is computed as (5).

$$LOK_k(p) = \frac{\sum_{o \in N_{k(p)}} \frac{LRD_k(o)}{LRD_k(p)}}{|N_{k(p)}|} \tag{5}$$

In (5), $LRD_k(p)$ represents the local reachability density, and $N_{k(p)}$ is the set of k-nearest neighbors. This approach is advantageous for detecting subtle anomalies in dynamic environments.

3.2.5. One-class SVM

One-class SVM constructs a hyperplane that differentiates normal instances from outliers using kernel transformations. The objective function for anomaly detection is formulated as (6).

$$\min_{w,\xi,\rho} \frac{1}{2}\left||w|\right|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho \tag{6}$$

In (6), ν regulates the proportion of outliers. one-class SVM is particularly useful for datasets with complex distributions.

3.3. Forecasting models

3.3.1. LTSM

The LSTM networks are an advanced variant of recurrent neural networks (RNNs) specifically designed to address the vanishing gradient problem that hinders traditional RNNs in capturing long-term dependencies. LSTMs have gained significant traction in time-series forecasting, natural language processing, and sequential data modeling due to their ability to retain essential information over extended time intervals. Unlike conventional RNNs, LSTMs utilize memory cells and specialized gating mechanisms that selectively store or discard information, enabling more effective learning from long-range dependencies.

The architecture of an LSTM as shown in Figure 2 consists of memory cells regulated by three fundamental gates: the forget gate, which determines the retention of past information; the input gate, which controls the integration of new information; and the output gate, which dictates the transmission of relevant information to the next time step. These gates collectively manage the flow of information within the network, thereby mitigating issues associated with long-term dependencies. The mathematical formulation of these gates ensures that the model learns and adapts effectively to sequential patterns in the data, making LSTMs particularly suitable for applications involving temporal dependencies.

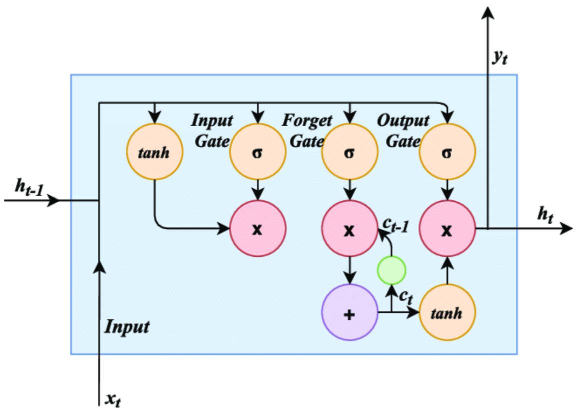

Figure 2. Architecture of LSTM layer

Forget gate: the forget gate regulates whether information from previous time steps should be retained or discarded based on the current input and the previous hidden state. Mathematically, it is defined as (7).

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad (7)$$

In (7), f_t represents the forget gate activation, and σ is the sigmoid activation function that constrains values between 0 and 1. A value close to 0 results in forgetting past information, while a value near 1 retains it. This mechanism ensures that irrelevant information does not accumulate in the memory cell, thereby improving model efficiency.

Input gate: The input gate governs the extent to which new information is incorporated into the cell state. It operates by generating a candidate memory value through a $tanh$ activation function and scaling it using a sigmoid gate to regulate its influence. The corresponding (8) and (9).

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (8)$$

$$\widetilde{C_t} = tanh(W_C.[h_{t-1}, x_t] + b_C) \quad (9)$$

In (8)–(9), i_t represents the input gate activation, and $\widetilde{C_t}$ denotes the candidate memory content. This controlled update mechanism ensures that only relevant new information is added to the memory, preventing unnecessary fluctuations in the learning process.

Cell state update*:* The cell state serves as the memory of an LSTM unit and is updated by combining retained past information with new inputs. The update equation is given by (10).

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \quad (10)$$

In (10), $\odot$ represents the element-wise product. The inclusion of the forget gate ensures that long-term dependencies are preserved while allowing new, relevant data to be incorporated effectively. This dynamic balance between memory retention and update is key to LSTM's superior performance in handling sequential data.

Output gate: the output gate determines how much of the updated cell state contributes to the hidden state and, consequently, the final output of the network. This process is governed by (11) and (12).

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (11)$$

$$h_t = o_t \odot tanh(C_t) \quad (12)$$

In (11) and (12), o_t are the output gate activation. The tanh activation ensures that the output values remain within a manageable range, thereby preventing extreme fluctuations. This selective information transfer enhances the model's ability to generate meaningful representations of sequential data.

3.3.2. CNN Autoencoder

CNN Autoencoder is a type of neural network that learns efficient representations of input sequences through an encoder-decoder structure. The encoder extracts important temporal features from the input time-series data and compresses them into a lower-dimensional latent space, while the decoder reconstructs the original input from this compressed representation. Given a time-series sequence $X = \{x_1, x_2, x_3, \ldots, x_T\}$, the encoder applies a series of convolutional operations to generate feature maps. The convolutional transformation for each filter h_k is given by (13).

$$h_k = \sigma(W_k * X + b_k) \quad (13)$$

In (13), W_k represents the filter weights, b_k is the bias term, $*$ denotes the convolution operation, and σ is a non-linear activation function (e.g., ReLU). The encoded feature representations are further passed through max-pooling layers to reduce dimensionality while preserving the most significant features. The latent representation, Z, is obtained as (14). In (14), max-pooling helps to retain dominant spatial-temporal features and reduces computational complexity.

$$Z = MaxPool(h_k) \quad (14)$$

The decoder reconstructs the input sequence from the latent representation by applying transposed convolution (deconvolution) layers, ensuring that the learned features accurately capture underlying time-dependent patterns. The reconstructed output sequence $\hat{X}$ is generated as (15).

$$\hat{X} = \sigma(W_d * Z + b_d) \quad (15)$$

In (15), W_d and b_d are the weights and bias of the decoder network, respectively. To fine-tune the model for time-series forecasting, the final layer is modified to predict the future time steps $\hat{Y}$ based on the learned latent features is given by (16).

$$\hat{Y} = Dense(Z) \tag{16}$$

In (16), the dense layer maps the compressed representation to the output space. By leveraging CNN-based feature extraction, the autoencoder improves forecasting accuracy by capturing intricate temporal dependencies while effectively handling noise and outliers in the dataset.

3.3.3 LSTM with attention mechanism

The LSTM networks are widely used for time-series forecasting due to their ability to retain long-term dependencies while mitigating the vanishing gradient problem. However, traditional LSTMs treat all time steps with equal importance, which can lead to suboptimal performance in complex datasets where certain past time steps contribute more significantly to future predictions. The attention mechanism enhances LSTM by dynamically weighing the importance of past observations is presented in Figure 3, allowing the model to focus on the most relevant time steps. Given an input sequence $X = \{x_1, x_2, x_3, \ldots, x_T\}$, the LSTM processes the sequence iteratively using the (17).

$$\begin{aligned} f_t &= \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right)_t = tanh(W_C h_{t-1} + U_C x_t + b_C) C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t h_t \\ &= o_t \odot tanh(C_t) \end{aligned} \tag{17}$$

In (17), f_t , i_t and o_t denote the forget, input, and output gates, respectively, C_t is the cell state, h_t is the hidden state, and σ represents the sigmoid activation function. The attention mechanism is then applied to enhance the LSTM's ability to focus on critical time steps. The attention score α_{t} is computed using an alignment function that determines the relevance of each hidden state h_t with respect to the target output. The attention weights are computed as (18)–(20).

$$e_t = v^T \tanh(W_a h_t + b_a) \tag{18}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_t \exp(e_t)} \tag{19}$$

$$c_t = \sum_t \alpha_t h_t \tag{20}$$

In (18)–(20), e_t represents the attention score, W_a and v are learnable parameters, and c_t is the context vector obtained by taking the weighted sum of hidden states. The final output is then computed as (21).

$$y_t = softmax\left(W_y [c_t; h_t] + b_y\right) \tag{21}$$

By incorporating attention, the model selectively focuses on informative time steps, leading to improved forecasting accuracy. This approach is particularly beneficial for energy demand prediction, where external factors such as seasonal variations and peak demand periods exert varying levels of influence on future consumption. The attention-enhanced LSTM provides greater interpretability and adaptability, making it a robust choice for time-series forecasting tasks in energy management systems.

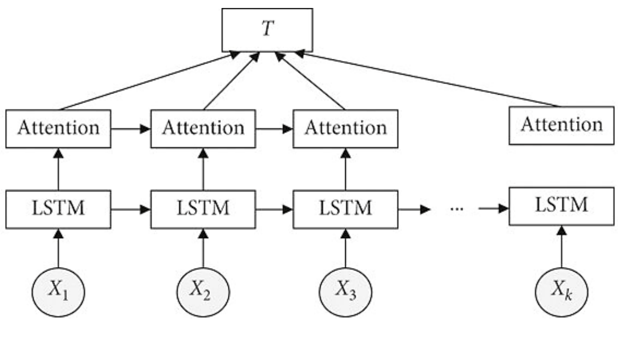

Figure 3. Architecture of LSTM with attention mechanism

4. RESULTS AND DISCUSSION

The impact of outlier detection techniques on time-series forecasting accuracy was analyzed using three deep learning models—LSTM, CNN-Autoencoder, and LSTM with attention—across multiple countries, including Germany (DE), France (FR), Spain (ES), and Italy (IT). The experimental results revealed that outlier removal significantly improved model performance by reducing error metrics such as RMSE and MAE. Among the five outlier detection techniques applied, IF and LOF demonstrated superior capability in detecting anomalous patterns, leading to the most noticeable improvements in forecasting accuracy. Specifically, in the case of Germany, the RMSE for LSTM without outlier removal was 17.34 MW, which reduced to 12.21 MW after applying IF. Similarly, Spain exhibited a substantial reduction in forecasting error, where the RMSE improved from 19.87 to 14.02 MW post outlier removal. These results indicate that eliminating outliers effectively mitigates noise, enabling the models to learn more representative energy consumption patterns.

The comparative analysis of various forecasting methods with and without outlier detection techniques demonstrates significant variations in predictive accuracy across different countries. The Table 1 presents MAPE values for Germany (DE), France (FR), Spain (ES), and Italy (IT), employing different forecasting approaches such as LSTM, CNN-Autoencoder, and LSTM with attention mechanism. When no outlier detection method is applied, the CNN-Autoencoder consistently exhibits the lowest MAPE values across all regions, indicating its robustness in handling raw data. Conversely, LSTM-Attention performs the worst among the three forecasting models, yielding the highest MAPE values, particularly in Italy (19.70%) and Spain (14.08%). This suggests that while attention mechanisms enhance LSTM models in certain scenarios, they may be more sensitive to anomalies present in the dataset.

The implementation of outlier detection techniques leads to a substantial improvement in forecasting accuracy, with one-class SVM emerging as the most effective method for noise reduction. Under this approach, CNN-Autoencoder attains the lowest MAPE values across all countries, particularly in France (2.12%) and Italy (3.17%), underscoring its efficiency in feature extraction and denoising capabilities. Similarly, the application of LOF and IF also improves prediction accuracy, albeit to a slightly lesser extent. Notably, LSTM's performance significantly benefits from these methods, with MAPE values dropping from 10.73% (without outlier detection) to as low as 3.18% (one-class SVM) in France. This highlights the importance of outlier handling in improving the predictive reliability of recurrent neural networks.

Among the outlier detection methods, DBSCAN and Z-Score filtering also exhibit promising results, though their effectiveness varies by forecasting model. DBSCAN, for instance, helps reduce MAPE values in LSTM models considerably, bringing them down to 5.81% (Germany) and 4.47% (France). Likewise, CNN-Autoencoder benefits from DBSCAN, attaining a MAPE of 2.88% in Spain, which is a considerable improvement from the baseline. However, the LSTM-Attention model, despite some improvements, continues to exhibit relatively higher error rates across most countries, suggesting that attention-based architectures might require more sophisticated anomaly handling techniques for optimal performance.

Overall, the findings place crucial role of outlier detection in enhancing forecasting accuracy, with one-class SVM and CNN-Autoencoder emerging as the most effective combination. While LSTM-based models benefit from anomaly filtering, the choice of forecasting model and outlier detection method must be tailored to the specific dataset and application context. Future research could explore hybrid approaches that integrate multiple anomaly detection strategies or leverage adaptive filtering mechanisms to further improve predictive performance in time series forecasting.

The effect of individual outlier detection methods varied across countries due to differences in data characteristics and energy consumption trends. Z-Score and one-class SVM, while effective in detecting extreme deviations, struggled with subtle anomalies present in non-Gaussian distributions. On the other hand, DBSCAN, which clusters data based on density, demonstrated mixed results shown in Figures 4, 5 and 6 performing well in structured datasets like France but underperforming in Italy due to irregular fluctuations in demand patterns. Overall, the study confirms that selecting an appropriate outlier detection technique is crucial for optimizing forecasting accuracy, and the best choice often depends on the underlying data distribution. The findings also highlight the necessity of adaptive anomaly detection strategies that can dynamically adjust to seasonal variations and long-term trends in energy consumption.

The overall improvements in model performance post outlier removal reinforce the importance of data preprocessing in time-series forecasting tasks. While deep learning architecture can capture complex temporal dependencies, their effectiveness is significantly influenced by data quality. This study demonstrates that integrating robust outlier detection mechanisms can substantially enhance forecasting reliability, making energy demand prediction models more applicable for real-world energy management and grid optimization. Future research should explore the combination of multiple anomaly detection methods using ensemble techniques and investigate the impact of incorporating external factors such as weather conditions and economic indicators to further refine predictions.

Table 1. Comparative results for various forecasting methods with and without outlier methods

Outlier method	Forecasting method	Germany (DE) MAPE (%)	France (FR) MAPE (%)	Spain (ES) MAPE (%)	Italy (IT) MAPE (%)
None	LSTM	10.73	6.96	10.00	15.49
	CNN-Autoencoder	5.03	4.69	4.13	7.23
	LSTM-attention	14.01	7.63	14.08	19.70
Z-Score	LSTM	9.01	5.36	7.16	12.65
	CNN-Autoencoder	3.70	3.40	3.18	4.48
	LSTM-attention	7.99	6.38	9.45	11.19
DBSCAN	LSTM	5.81	4.47	5.76	8.63
	CNN-Autoencoder	3.60	2.91	2.88	3.74
	LSTM-attention	7.88	5.38	8.59	10.63
IF	LSTM	5.79	3.85	5.49	9.17
	CNN-Autoencoder	3.48	2.78	3.44	5.49
	LSTM-attention	9.35	5.00	8.51	13.41
LOF	LSTM	4.82	3.41	4.66	7.64
	CNN-Autoencoder	3.12	2.70	2.70	3.76
	LSTM-attention	6.48	4.21	6.39	9.03
One Class SVM	LSTM	5.76	3.18	4.71	8.50
	CNN-Autoencoder	2.80	2.12	2.51	3.17
	LSTM-attention	5.27	3.26	4.35	8.11

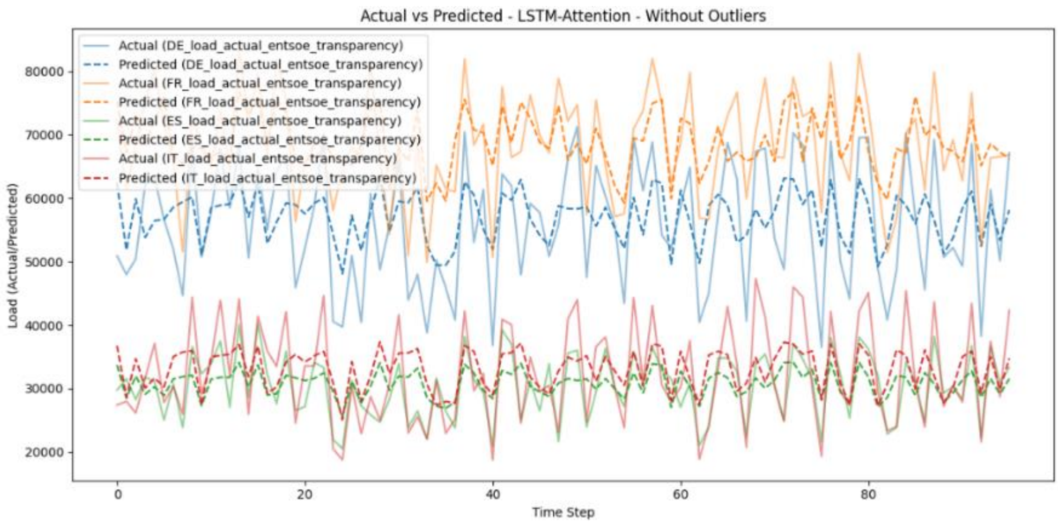

Figure 4. Load forecasting using LSTM without outliers removed

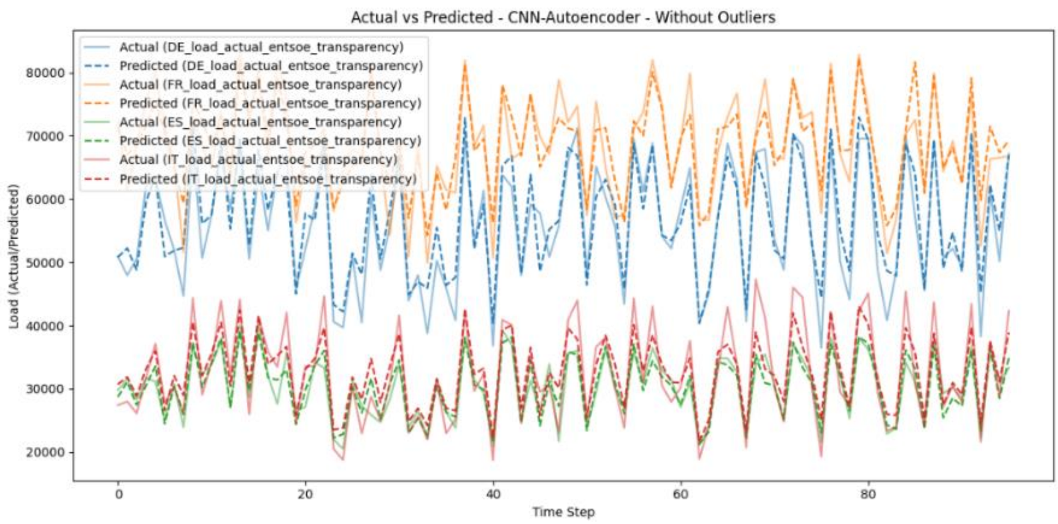

Figure 5. Load forecasting using CNN autoencoder without outliers removed

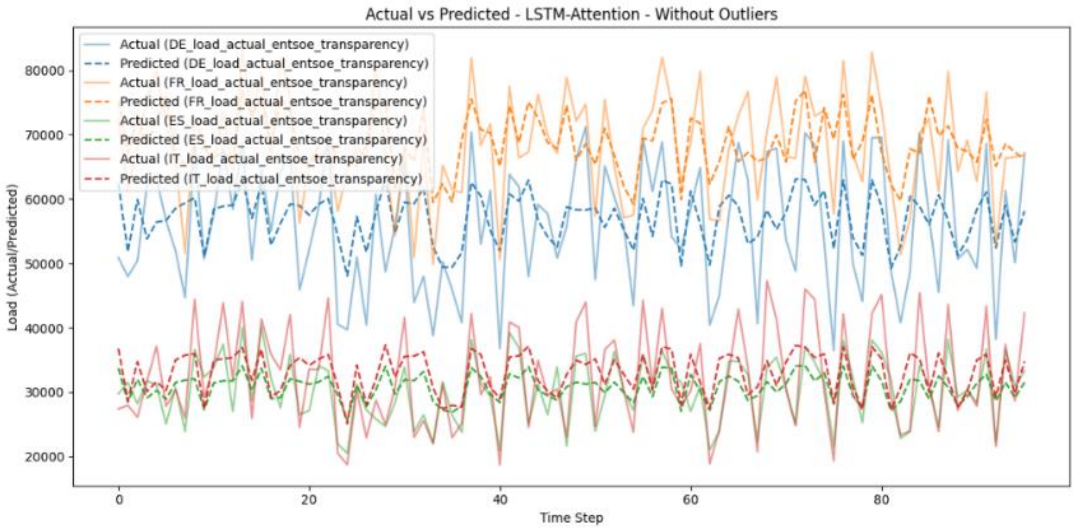

Figure 6. Load forecasting using LSTM-attention without outliers removed

The results indicate that outlier detection significantly enhances forecasting accuracy across all models and countries. For example, the CNN-Autoencoder paired with one-class SVM achieved the lowest MAPE of 2.12% in France, demonstrating the synergy between robust feature extraction and anomaly detection. These findings highlight that ignoring outlier handling can lead to suboptimal model performance, underscoring the necessity of robust preprocessing in energy forecasting pipelines.

5. CONCLUSION

This paper investigates the role of outlier detection techniques in improving the accuracy and robustness of deep learning-based energy demand forecasting for multiple countries. By integrating five prominent outlier detection algorithms with three state-of-the-art forecasting models, the study reveals several key findings.

Outlier removal significantly reduces forecasting errors, enhancing model reliability, especially in real-world, noisy datasets. Among the evaluated techniques, one-class SVM and LOF proved particularly effective at identifying anomalous data and improving model performance. The combination of CNN Autoencoder and one-class SVM achieved the best predictive accuracy, demonstrating the value of pairing strong feature extractors with robust anomaly detectors.

The LSTM-Attention model, while designed for capturing complex temporal dependencies, exhibited higher sensitivity to outliers, underscoring the need for careful preprocessing when deploying attention-based architectures. These findings advance present knowledge by demonstrating the tangible benefits of integrating anomaly detection into forecast pipelines, an aspect often overlooked in previous studies. The work also offers practical insights into selecting model-method combinations for energy demand forecasting, which are directly applicable in operational settings.

This research opens several promising directions: developing adaptive or hybrid outlier detection methods that respond to dynamic seasonal and regional variations; incorporating exogenous factors such as weather or economic indicators for improved forecasting; and extending the framework to other domains like various data analysis water resource management or renewable generation forecasting.

FUNDING INFORMATION

Authors state there is no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Shreyas Karnick	✓	✓	✓			✓	✓	✓	✓	✓	✓		✓	
Sanjay Lakshminarayanan		✓		✓		✓		✓		✓		✓		
Madhu Palati	✓			✓						✓	✓	✓	✓	
Prakash R					✓					✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, SK, upon reasonable request.

REFERENCES

[1] I. Amalou, N. Mouhni, and A. Abdali, “Kernel function selection for FISVDD applied to outliers detection application for energy time series datasets,” *Electric Power Components and Systems*, pp. 1–16, May 2024, doi: 10.1080/15325008.2024.2350058.

[2] A. Badhan and A. Ganpati, “Overview of outlier detection methods and evaluation metrics: A review,” in *Challenges in Information, Communication and Computing Technology*, London: CRC Press, 2024, pp. 736–741, doi: 10.1201/9781003559092-127.

[3] R. M. Montgomery, “Techniques for outlier detection: a comprehensive view,” *Preprint*, Oct. 2024, doi: 10.20944/preprints202410.1603.v1.

[4] K. Xu, Y. Li, Y. Li, L. Xu, R. Li, and Z. Dong, “Masked graph neural networks for unsupervised anomaly detection in multivariate time series,” *Sensors*, vol. 23, no. 17, p. 7552, Aug. 2023, doi: 10.3390/s23177552.

[5] K. T. Vasudev, M. M. Manohara Pai, and R. M. Pai, “Comparative analysis of generic outlier detection techniques,” in *International Conference on Data Analytics & Learning*, 2022, pp. 117–126.

[6] J. Yang, L. Yang, W. Wang, and R. Pu, “An outlier detection algorithm based on local density and natural neighbors,” in *2023 2nd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE)*, Nov. 2023, pp. 51–56, doi: 10.1109/CBASE60015.2023.10439072.

[7] W. Meng, C. Wenhang, and D. Hongyan, “Investigating outlier detection techniques based on kernel rough clustering,” *Recent Advances in Computer Science and Communications*, vol. 17, no. 1, pp. 61–71, Jan. 2024, doi: 10.2174/2666255816666230912153541.

[8] C.-K. Lin, C.-C. Lu, S.-W. Chang, Y.-H. Chu, K.-C. Wu, and M. C.-T. Chao, “Outlier detection for analog tests using deep learning techniques,” in *2023 IEEE 41st VLSI Test Symposium (VTS)*, Apr. 2023, pp. 1–7, doi: 10.1109/VTS56346.2023.10139998.

[9] G. Kalnoor, K. S. Dasari, S. S, N. Waddenkery, and B. Pragathi, “Enhanced brain tumor detection from MRI scans using frequency domain features and hybrid machine learning models,” *Journal of Modern Technology*, pp. 141–149, Jan. 2025, doi: 10.71426/jmt.v1.i2.pp141-149.

[10] A. V. Solanke, S. K. Verma, S. Kumar, B. Oyinna, and K. E. Okedu, “MPPT for hybrid energy system using machine learning techniques,” *Journal of Modern Technology*, vol. 1, no. 1, pp. 19–37, Sep. 2024, doi: 10.71426/jmt.v1.i1.pp19-37.

[11] X. Shen, H. Liu, G. Qiu, Y. Liu, J. Liu, and S. Fan, “Interpretable interval prediction-based outlier-adaptive day-ahead electricity price forecasting involving cross-market features,” *IEEE Transactions on Industrial Informatics*, vol. 20, no. 5, pp. 7124–7137, May 2024, doi: 10.1109/TII.2024.3355105.

[12] Y. Zheng, “Dataset outlier detection method based on random forest algorithm,” in *2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, Oct. 2022, pp. 552–556, doi: 10.1109/AIAM57466.2022.00111.

[13] J. Li, L. Zheng, Y. Zhu, and J. He, “Outlier impact characterization for time series data,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 13, pp. 11595–11603, May 2021, doi: 10.1609/aaai.v35i13.17379.

[14] S. Salman Qasim and S. M. Nsaif, “Advancements in time series-based detection systems for distributed denial-of-service (DDoS) attacks: a comprehensive review,” *Babylonian Journal of Networking*, vol. 2024, pp. 9–17, Jan. 2024, doi: 10.58496/BJN/2024/002.

[15] A. K. Bhardwaj, P. K. Dutta, and P. Chintale, “AI-powered anomaly detection for kubernetes security: a systematic approach to identifying threats,” *Babylonian Journal of Machine Learning*, vol. 2024, pp. 142–148, Aug. 2024, doi: 10.58496/BJML/2024/014.

[16] H. Jiang, H. Wang, W. Hu, D. Kakde, and A. Chaudhuri, “Fast incremental SVDD learning algorithm with the Gaussian kernel,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 3991–3998, Jul. 2019, doi: 10.1609/aaai.v33i01.33013991.

[17] H. Songhua, “A novel outlier calendrical heterogeneity reconstruction deep learning model for electricity demand forecasting,” *Journal of Intelligent & Fuzzy Systems*, vol. 46, no. 2, pp. 3363–3394, Feb. 2024, doi: 10.3233/JIFS-235218.

[18] C. Li, D. Liu, M. Wang, H. Wang, and S. Xu, “Detection of outliers in time series power data based on prediction errors,” *Energies*, vol. 16, no. 2, p. 582, Jan. 2023, doi: 10.3390/en16020582.

[19] C. Fu, P. Arjunan, and C. Miller, “Trimming outliers using trees,” in *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, Nov. 2022, pp. 456–461, doi: 10.1145/3563357.3566147.

[20] Y. Gao *et al.*, "Outlier detection in temporal and spatial sequences via correlation analysis based on graph neural networks," *Displays*, vol. 84, p. 102775, Sep. 2024, doi: 10.1016/j.displa.2024.102775.

[21] S. S. Ismaeel, K. M. T. Omar, and S. A. Othman, "Detection of outlier in time series with application to dohuk dam using the SCA statistical system," *General Letters in Mathematics*, vol. 13, no. 2, pp. 36–48, 2023, doi: 10.31559/glm2023.13.2.2.

[22] A. Zabihi, M. Parhamfar, and M. Khodadadi, "Strengthening resilience: a brief review of cybersecurity challenges in IoT-driven smart grids," *Journal of Modern Technology*, pp. 106–120, Nov. 2024, doi: 10.71426/jmt.v1.i2.pp106-120.

[23] K. Kyo, "Identifying and estimating outliers in time series with nonstationary mean through multi-objective optimization method," *Big Data, Data Mining and Data Science: Algorithms, Infrastructures, Management and Security*, pp. 1–20, 2025.

[24] R. Sathe and S. Shinde, "A deep learning framework for effective anomaly detection in time series data," in *2024 4th Asian Conference on Innovation in Technology (ASIANCON)*, Aug. 2024, pp. 1–7, doi: 10.1109/ASIANCON62057.2024.10837697.

[25] M. K. PK, M. R. Gurram, A. A. Hossain, and F. Amsaad, "ARIMA-DCGAN synergy: a novel adversarial approach to outlier detection in time series data," in *NAECON 2024 - IEEE National Aerospace and Electronics Conference*, Jul. 2024, pp. 423–427, doi: 10.1109/NAECON61878.2024.10670660.

[26] S. K. Dani, C. Thakur, N. Nagvanshi, and G. Singh, "Anomaly detection using PCA in time series data," in *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Mar. 2024, pp. 1–6, doi: 10.1109/IATMSI60426.2024.10502929.

BIOGRAPHIES OF AUTHORS

Shreyas Karnick has a post graduation in power systems and pursing Ph.D. in electrical and electronics engineering. He has published 3 research papers in international journals, 2 research papers in international conferences, and 3 research papers in national conferences. He is currently working for Accenture Solutions Pvt. Ltd as service delivery operations specialist and works in the delivery of solutions and services for power system utilities. His areas of interest include power system analysis, electrical load forecasting, deep learning and python programming. He can be contacted at: shreyaskarnick2022@gmail.com.

Sanjay Lakshminarayanan studied at the Kendriya Vidyalayas, IISc and Hebbal Branches. He received the B.Tech. degree from the Indian Institute of Technology (IIT), Kharagpur, India, in 1990 and the M.Sc. (Engg.) degree in electrical engineering from the Indian Institute of Science (IISc) Bangalore, India, in 1995. He completed his Ph. D. in 2007 from the Indian Institute of Science (IISc), Bangalore, India. He has been in industry for about seven years. He was with Grentel technologies, Cochin, Hical magnetics Pvt. Ltd, Bangalore, and GE medical systems, Bangalore and has a teaching experience of about seventeen years. Currently he is a professor, in the Department of Electrical and Electronics Engineering, at BMS Institute of Technology, Bangalore. His research interests are in power converters, PWM strategies, and motor drives. He can be contacted at email: sanjay.eee@bmsit.in.

Madhu Palati received the B. Tech degree in electrical and electronics engineering from Sri Venkateshwara University, Tirupati, in 2003, M.E. from M.S. University, Baroda, in 2005 and Ph.D. from Jain University, Bangalore in June 2016. He has worked as a software engineer in Keane India Ltd, Gurgaon for a period of one and half years and in IBM Private Limited, Bangalore for a period of three years. He has worked as Assistant Professor in the Department of Electrical and Electronics Engineering, School of Engineering and Technology, Jain University, Bangalore for a period of seven and half years. He is currently working as assistant professor at BMS Institute of Technology and Management, Bangalore. He can be contacted at email: madhupalati@bmsit.in.

Prakash R received B.E in electrical and electronics engineering from Basaveshwara College of Engineering, Bagalkot, M.E in power system engineering from Walchand College of Engineering, Sangli and Ph. D. in power system engineering from VTU, Belgaum. He is currently working as a professor and head dept of EEE, Acharya Institute of Technology, Bengaluru, Karnataka. His research interests include power system, renewable energy, and micro-grids. He can be contacted at email: prakashrp1960@gmail.com.