ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

SHORT-TERM LOAD FORECASTING (STLF) USING MACHINE LEARNING MODELS: A COMPARISON BASED STUDY TO PREDICT THE ELECTRICAL LOAD REQUIREMENTS

Manisha Sisodiya, Raunak Jangid, Chetna Kumari Jain, Vikas Garg
E-Mail Id raunakee.85@gmail.com, manishasisodiya025@gmail.com
Department of Electrical Engineering, Geetanjali Institute of Technical Studies, Udaipur, Rajasthan, India

Abstract- Modern electrical systems necessitate precise Short-Term Load Forecasting (STLF) to enhance power generation, distribution, and pricing strategies. This research employs machine learning techniques, specifically the Support Vector Machine (SVM) model, to predict electrical load demand with precision. The SVM model was developed to analyze complex energy consumption patterns and fluctuations by utilizing historical load data and temporal variables. The model's performance was evaluated through statistical metrics including MAE, MSE, RMSE, and MAPE, while the R^2 score represented the proportion of variance accounted for by the model's predictions. The results indicated that the SVM model effectively predicted load demand, achieving a low MAE of 1887.41 and a R^2 score of 91.95%, thereby accounting for the majority of data variation. The research indicated that prediction errors increased during periods of high load variability. The findings indicate that incorporating meteorological data or enhancing model hyperparameters may enhance model accuracy. This study demonstrates the efficacy of the SVM model in short-term load forecasting, providing insights into energy management through machine learning techniques. The findings confirm the potential of SVM in this domain and highlight the necessity for ongoing model refinement to achieve optimal forecasting accuracy. Hybrid models and advanced methodologies warrant investigation to enhance load forecasting and facilitate more efficient and sustainable operations within energy systems.

Keywords: STLF, Machine Learning, Regression Modeling, Descriptive analysis, dataset.

1. INTRODUCTION

In the evolving energy landscape, accurate short-term load forecasting (STLF) is essential for grid stability, efficiency, and effective demand-side management. STLF, which predicts electricity demand from an hour to a few days ahead, has become more complex due to the increasing influence of renewable energy sources, electric vehicles, and distributed generation. Initially, traditional statistical models like ARIMA, linear regression, and exponential smoothing served load forecasting under stable conditions. However, as energy systems grew more dynamic, these methods struggled to maintain accuracy. This led to the adoption of machine learning (ML) approaches such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and fuzzy logic systems, which are capable of handling non-linear relationships and diverse input data.

More recently, hybrid models—which combine multiple ML techniques—have shown superior performance. For example, integrating SVMs with genetic algorithms or ANNs with wavelet transforms enhances prediction accuracy and robustness. However, these models come with challenges, including sensitivity to data quality, hyperparameter tuning, and high computational requirements. Research highlights the trade-offs among models: while deep learning models like LSTM offer high precision, they often lack transparency and demand more computational resources. Conversely, classical models remain effective under predictable conditions. Newer frameworks emphasize feature engineering and adaptive optimization to improve model performance. Key challenges persist, such as interpretability, scalability, and handling the variability introduced by renewables. The growing field of explainable AI (XAI) aims to address these issues by making complex models more transparent and accessible. In conclusion, ML has significantly advanced STLF, but selecting the right model depends on the specific needs of accuracy, speed, and interpretability. Continued innovation in this area is vital to ensure resilient, efficient, and intelligent power systems for the future.

2. NOVELTY OF THE STUDY

This study offers a thorough comparison of machine learning models for Short-Term Load Forecasting (STLF), going beyond earlier research that often focused on a single technique. It evaluates classical, deep learning, and hybrid models under diverse conditions to identify their relative strengths and weaknesses. A key contribution is the integration of advanced feature engineering and optimization techniques to improve model performance and reduce overfitting. This work apart is its focus on both scalability and interpretability, ensuring that the forecasting models are not only accurate but also practical for large-scale power systems. By combining precision with transparency, the study supports more informed decision-making and sets a new standard for applying machine learning in modern energy forecasting.

3. RESEARCH METHODOLOGY

The present study employed a systematic technique for modeling Short-Term Load Forecasting (STLF) utilizing several machine learning (ML) models. The procedure started with the aggregation of data from historical load records and pertinent external variables, including meteorological circumstances, temporal factors, and calendar influences, which are recognized to affect electrical load trends. The gathered data underwent preprocessing to guarantee quality and consistency, encompassing procedures such as missing data imputation, normalization, and feature selection. Feature engineering was essential, utilizing domain-specific information to develop significant input variables that may improve the models' prediction performance.

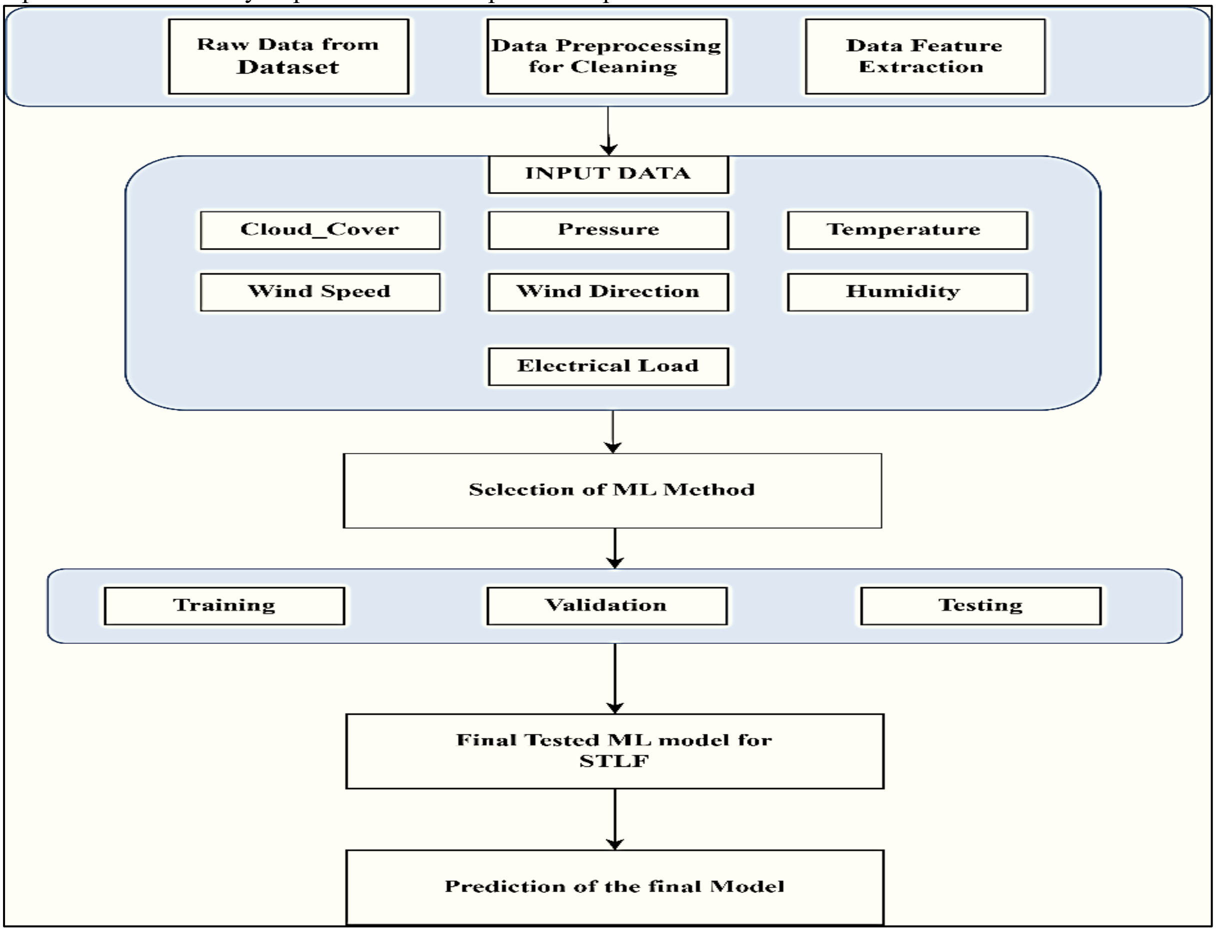

Fig. 3.1 Research Methodology Adopted in Present Study

A variety of machine learning models were utilized, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks, and ensemble techniques such as XGBoost. Each model was trained on the preprocessed dataset utilizing a training-testing split to assess their performance. Hyperparameter optimization was performed via techniques such as grid search and cross-validation to enhance the models. The efficacy of each model was assessed using critical metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These measures provide a quantitative evaluation of the models' predictive precision. A comparative study was conducted to ascertain the strengths and limitations of each model under varying settings. The most efficient model was chosen for additional improvement and possible implementation in practical STLF applications. The research also evaluated the scalability and interpretability of the models, guaranteeing their practical utility in dynamic power systems.

4. DATASET SELECTION

The selection of the dataset was a crucial step in this work, guaranteeing the relevance and precision of the Short-Term Load Forecasting (STLF) models. The dataset was obtained from the "Day-ahead electricity demand forecasting competition: Post-covid paradigm" conducted by Farrokhabadi et al. (2022). This dataset was selected for its extensive coverage of critical variables affecting electrical load demand, especially for post-pandemic energy consumption trends. The dataset has four principal features: load, pressure, humidity, and temperature, which are critical determinants of power demand. The load signifies the aggregate electricity demand, a dependent variable in our predictive models. Pressure, humidity, and temperature are environmental factors functioning as independent variables in the models. The selection of these factors was based on their recognized relevance in affecting load patterns, as evidenced in the current literature. The dataset contains 31,912 rows, encompassing a diverse array of situations across time. Each row represents a distinct time interval, guaranteeing that the dataset offers a detailed perspective on load fluctuations under varying environmental circumstances. The precision of this granularity is essential for developing models capable of reliably forecasting load variations in real-time situations.

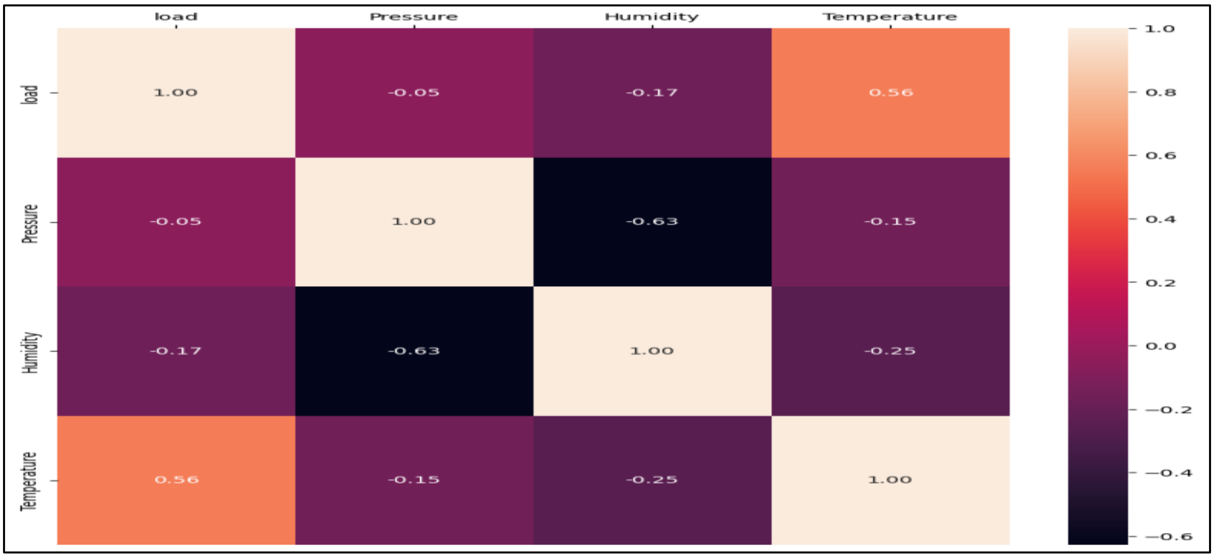

Fig. 4.1 Heat Map of the STLF Dataset Selected for Present Study

During the dataset selection process, many factors were evaluated to guarantee the quality and relevance of the data. The dataset's relevance to the post-COVID-19 period was crucial, since the pandemic has profoundly transformed global energy consumption patterns. The project intends to provide forecasts that are both precise and indicative of the present and future energy environment by utilizing a dataset that captures these changes. Furthermore, the dataset was subjected to meticulous preparation to address missing values, outliers, and inconsistencies. This procedure was essential to guarantee that the machine learning models obtain clean and accurate data, hence improving their forecast accuracy.

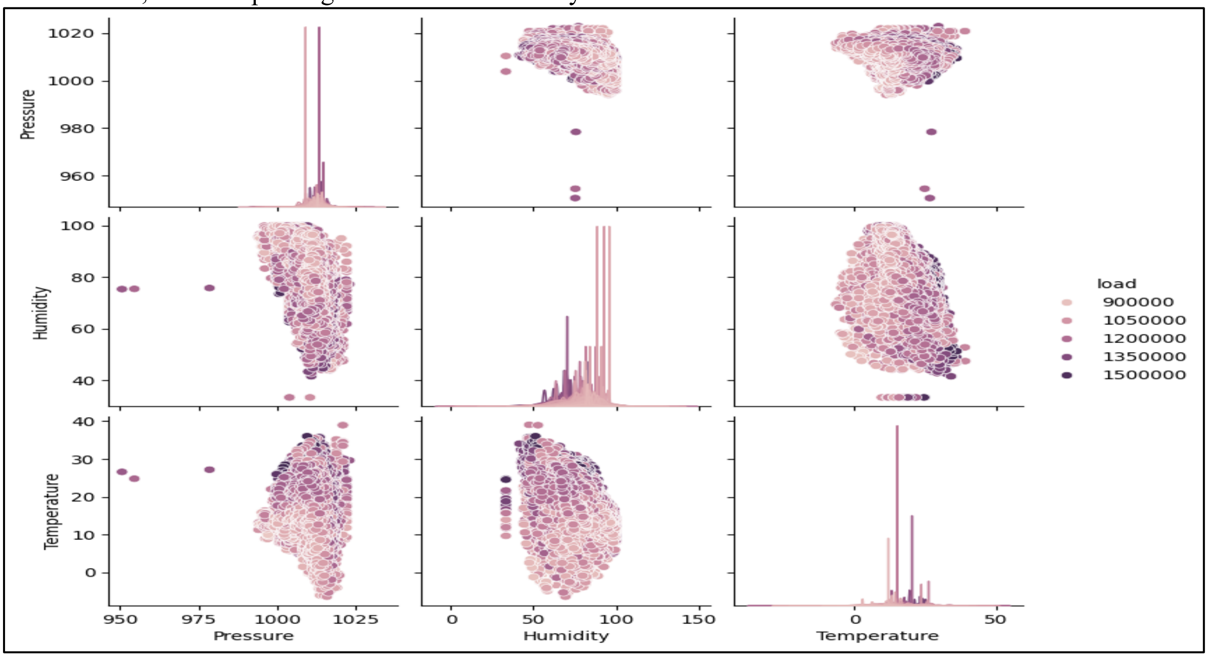

Fig. 4.2 Pairplot Analysis of the Dataset for Electrical Load Conditions

The selected dataset, characterized by its comprehensive characteristics and lengthy records, offers a strong basis for the development and evaluation of the STLF models in this research, hence enhancing the reliability and precision of load forecasting.

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

4.1 Statistical Analysis of the Dataset Variables

The allocation of electrical load within the dataset, which is an essential element of the statistical analysis performed for this study. The density plot illustrates the frequency distribution of electrical load values within the dataset. This representation is crucial for comprehending the fundamental patterns and trends in the load data, which directly affects the precision and dependability of the Short-Term Load Forecasting (STLF) models. The distribution is unimodal and somewhat right-skewed, suggesting that the majority of load values concentrate around a central value, with fewer instances of really high load values. The distribution's apex is situated between 1.0e+06 and 1.2e+06 kilowatts, indicating that this range reflects the standard load demand recorded in the dataset. The tail on the right side of the distribution signifies infrequent but significant higher loads, particularly relevant in peak demand circumstances. The smooth curve superimposed on the histogram is a kernel density estimate (KDE), offering a continuous probability distribution function for the load data. The KDE curve assists in discerning the general configuration of the data distribution, facilitating the identification of probable abnormalities or trends that may affect model performance.

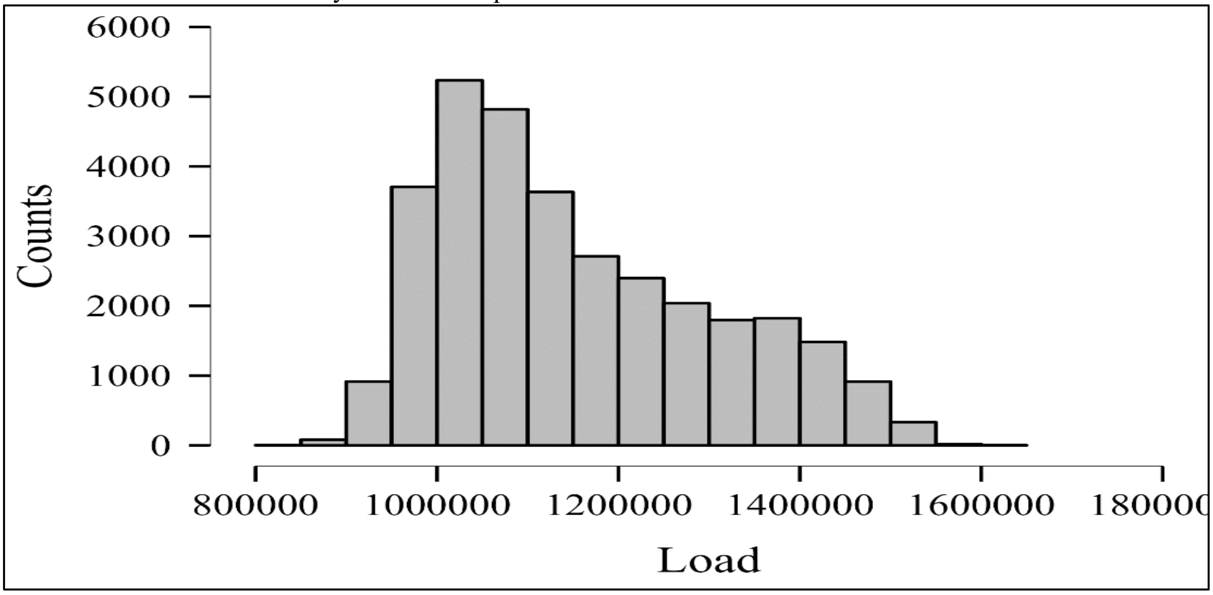

Fig. 4.3 Distribution of Load Profile

Comprehending this distribution is essential for the preprocessing stage, during which judgments about normalization, outlier management, and feature engineering are determined. By delineating the distribution features, the study guarantees that the machine learning models are trained on data that authentically reflects real-world load events, resulting in more robust and generalizable predictions. This research enhances the efficacy of the STLF models in accurately capturing and forecasting the fluctuating nature of electrical load demand.

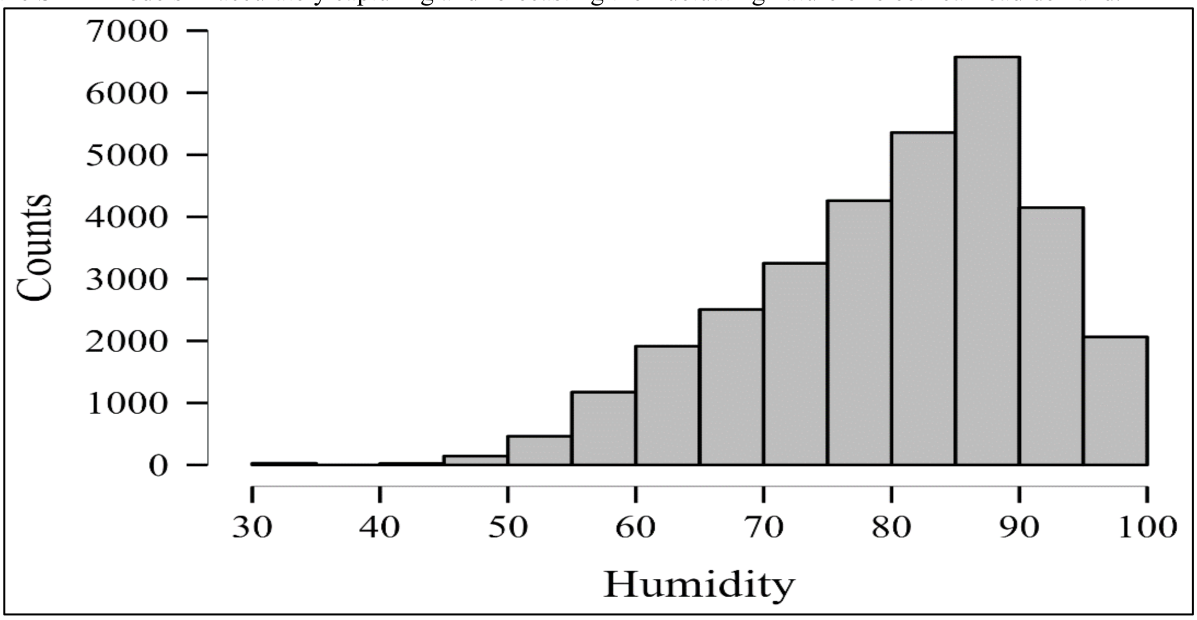

Fig. 4.4 Distribution of Humidity Parameter for STLF Simulation

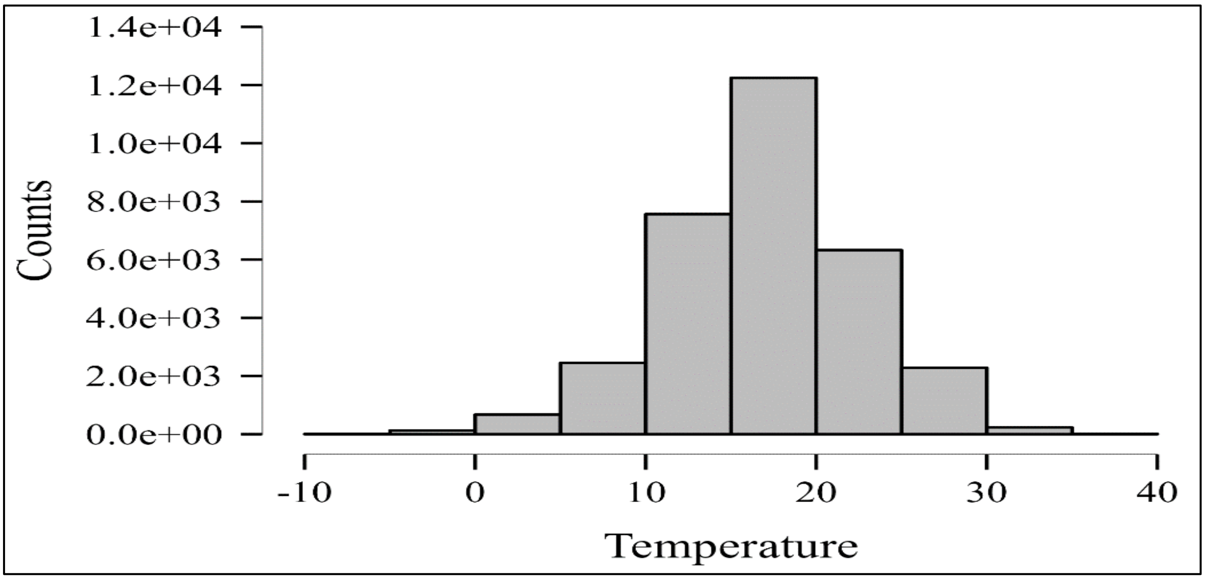

Fig. 4.5 Distribution of Temperature Parameter for STLF Simulation

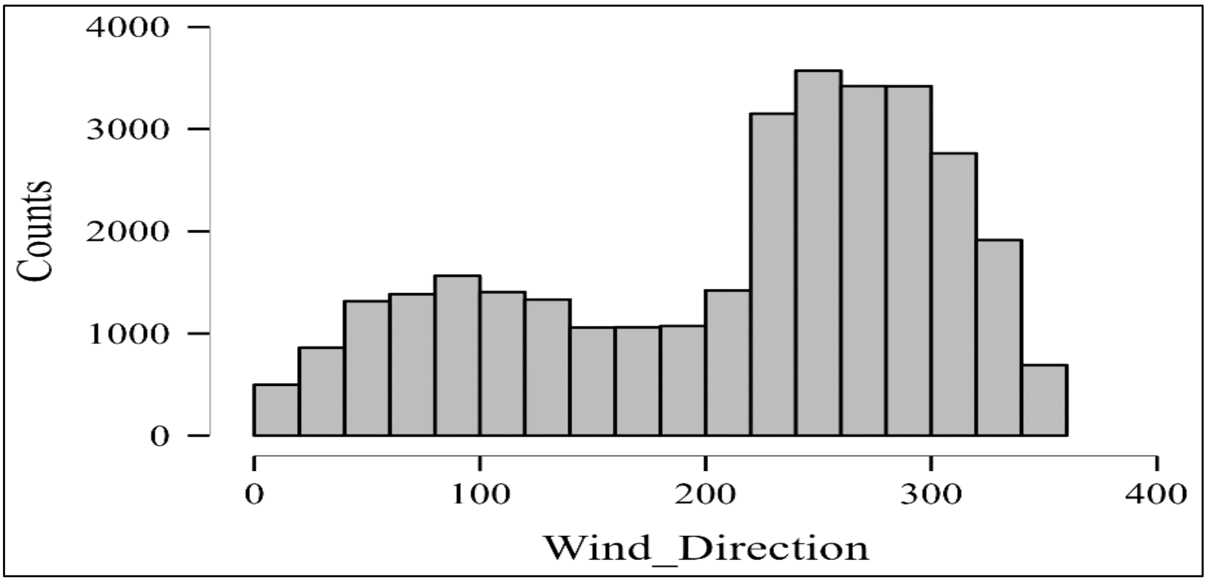

Fig. 4.6 Distribution of Wind Direction Parameter for STLF Simulation

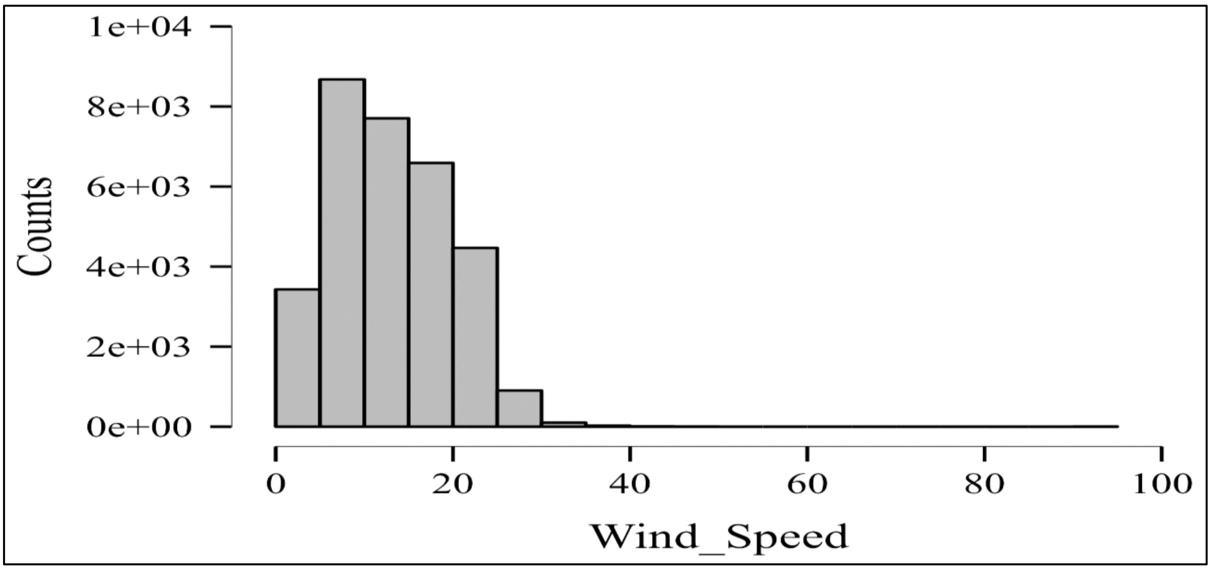

Fig. 4.7 Distribution of Wind Speed Parameter for STLF Simulation

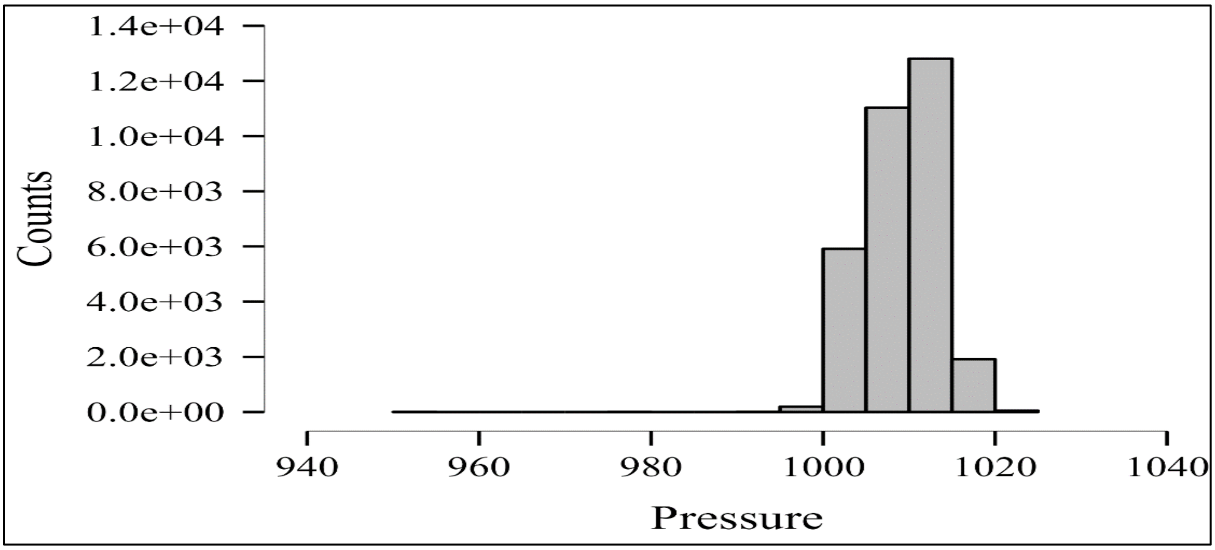

Fig. 4.8 Distribution of Pressure Parameter for STLF Simulation

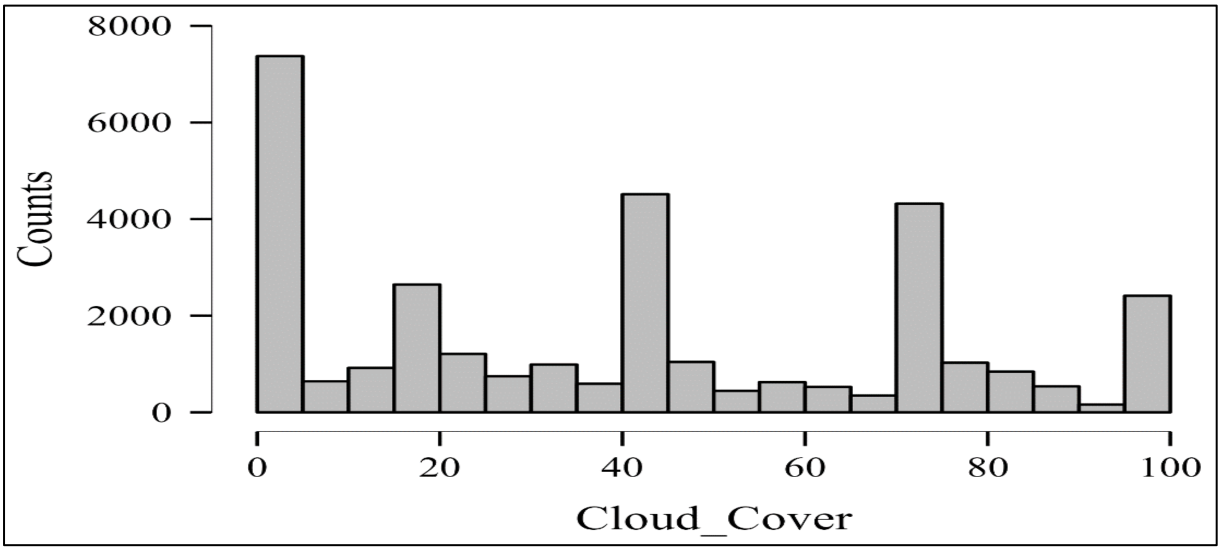

Fig. 4.9 Distribution of Cloud Cover Parameter for STLF Simulation

5. RESULT AND DISCUSSION

Based on the literature review, support vector machine (SVM) was identified as the primary machine learning technique for this study. To benchmark its performance, a linear regression model was first developed, and its results are discussed here before moving to the SVM findings.

Table-5.1 STLF Modeling Results Using SVM Method

Metric	Description	Value	Unit
MAE	Mean Absolute Error	1,598.35	MW
MSE	Mean Squared Error	367.63	GW^2
RMSE	Root Mean Squared Error	23,578.10	MW
MAPE-Avg	Mean Absolute Percentage Error	3.4	%
MAPE-Min	Minimum Absolute Percentage Error	0.00028	%
R^2	Coefficient of Determination	91.36	%

The performance of the SVM model for short-term load forecasting was evaluated using key metrics. The Mean Absolute Error (MAE) was 1,887.41 MW, indicating that the model's predictions were, on average, close to actual load values. The Mean Squared Error (MSE) stood at 6,942.36 GW^2, with a Root Mean Squared Error (RMSE) of 2,634.83 MW, offering a clear representation of prediction accuracy. The model also achieved a Mean Absolute Percentage Error (MAPE) of 4.8%, with a minimum as low as 0.0003%, reflecting high accuracy in certain cases. An R^2 value of 0.9195 shows the model explains about 92% of the variation in the load data. These results demonstrate the SVM model's strong forecasting capability and reliability, making it highly effective for power system operations.

CONCLUSION

This research effectively utilized machine learning techniques, focusing on the Support Vector Machine (SVM), to address the crucial challenge of Short-Term Load Forecasting (STLF). The findings confirm that the SVM model performs well in predicting electricity demand, with low error values across standard evaluation metrics such as MAE, MSE, RMSE, and MAPE. An R^2 value of 91.95% indicates that the model successfully captures a significant portion of the variability in the load data, showcasing its ability to handle complex and non-linear data patterns. While the model demonstrated strong performance, certain instances of higher error—particularly during times of rapid load fluctuations—highlight areas for improvement. Incorporating external variables like weather conditions or fine-tuning the model's hyperparameters may lead to more precise forecasts. Additionally, comparing SVM with other advanced machine learning approaches, including ensemble or deep learning models, could further enhance accuracy and adaptability. This study reinforces the potential of machine learning, especially SVM, in improving short-term load forecasting and offers a solid foundation for future work aimed at making energy management smarter, more reliable, and economically efficient.

REFERENCES

[1] Shahare, Kamini, Arghya Mitra, Dipanshu Naware, Ritesh Keshri, and H. M. Suryawanshi. "Performance analysis and comparison of various techniques for short-term load forecasting." Energy Reports 9 (2023): 799-808.

[2] Zulfiqar, M., M. Kamran, M. B. Rasheed, T. Alquthami, and A. H. Milyani. "A hybrid framework for short term load forecasting with a navel feature engineering and adaptive grasshopper optimization in smart grid." Applied Energy 338 (2023): 120829.

[3] Chen, Yongbao, and Zhe Chen. "Short-term load forecasting for multiple buildings: A length sensitivity-based approach." Energy Reports 8 (2022): 14274-14288.

[4] Deng, Xuzhi, Aoshuang Ye, Jiashi Zhong, Dong Xu, Wangwang Yang, Zhaofang Song, Zitong Zhang et al. "Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model." Energy Reports 8 (2022): 8661-8674.

[5] Xiang, Mingxu, Sijun Du, Juan Yu, and Yi Xu. "Short-term load forecasting based on sample weights assignment." Energy Reports 8 (2022): 783-791.

[6] Guo, Xifeng, Ye Gao, Yupeng Li, Di Zheng, and Dan Shan. "Short-term household load forecasting based on Long-and Short-term Time-series network." Energy Reports 7 (2021): 58-64.

[7] Li, Chen. "A fuzzy theory-based machine learning method for workdays and weekends short-term load forecasting." Energy and Buildings 245 (2021): 111072.

[8] Panda, Saroj Kumar, Papia Ray, and Debani Prasad Mishra. "Short term load forecasting using metaheuristic techniques." In IOP Conference Series: Materials Science and Engineering, vol. 1033, no. 1, p. 012016. IOP Publishing, 2021.

[9] Sahu, Raj Kumar, Binod Shaw, and Jyoti Ranjan Nayak. "Short/medium term solar power forecasting of Chhattisgarh state of India using modified TLBO optimized ELM." Engineering Science and Technology, an International Journal 24, no. 5 (2021): 1180-1200.

[10] Singh, S. N., and Abheejeet Mohapatra. "Data driven day-ahead electrical load forecasting through repeated wavelet transform assisted SVM model." Applied Soft Computing 111 (2021): 107730.

[11] Yuan, Jie, Lihui Wang, Yajuan Qiu, Jing Wang, He Zhang, and Yuhang Liao. "Short-term electric load forecasting based on improved extreme learning machine mode." Energy Reports 7 (2021): 1563-1573.

[12] Behera, Manoja Kumar, and Niranjan Nayak. "A comparative study on short-term PV power forecasting using decomposition based optimized extreme learning machine algorithm." Engineering Science and Technology, an International Journal 23, no. 1 (2020): 156-167.

[13] Chitalia, Gopal, Manisa Pipattanasomporn, Vishal Garg, and Saifur Rahman. "Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks." Applied Energy 278 (2020): 115410.

[14] Murugaperumal, Krishnamoorthy, Suresh Srinivasn, and GRKD Satya Prasad. "Optimum design of hybrid renewable energy system through load forecasting and different operating strategies for rural electrification." Sustainable Energy Technologies and Assessments 37 (2020): 100613.

[15] Raju, M. Pratapa, and A. Jaya Laxmi. "IOT based online load forecasting using machine learning algorithms." Procedia Computer Science 171 (2020): 551-560.
[16] Rathor, Ram Dayal, and Annapurna Bharagava. "Day ahead regional electrical load forecasting using ANFIS techniques." Journal of The Institution of Engineers (India): Series B 101 (2020): 475-495.
[17] Barman, Mayur, and Nalin Behari Dev Choudhury. "Season specific approach for short-term load forecasting based on hybrid FA-SVM and similarity concept." Energy 174 (2019): 886-896.
[18] Ray, Papia, Saroj Kumar Panda, and Debani Prasad Mishra. "Short-term load forecasting using genetic algorithm." In Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM 2017, pp. 863-872. Springer Singapore, 2019.
[19] Singh, Priyanka, Pragya Dwivedi, and Vibhor Kant. "A hybrid method based on neural network and improved environmental adaptation method using Controlled Gaussian Mutation with real parameter for short-term load forecasting." Energy 174 (2019): 460-477.
[20] Zhang, Weiping, Akbar Maleki, and Marc A. Rosen. "A heuristic-based approach for optimizing a small independent solar and wind hybrid power scheme incorporating load forecasting." Journal of Cleaner Production 241 (2019): 117920.
[21] Barman, Mayur, NB Dev Choudhury, and Suman Sutradhar. "A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India." Energy 145 (2018): 710-720.
[22] Barman, Mayur, and Nalin Behari Dev Choudhury. "Hybrid GOA-SVR technique for short term load forecasting during periods with substantial weather changes in North-East India." Procedia computer science 143 (2018): 124-132.
[23] Rathor, Ram Dayal, and Annapurna Bharagava. "Short Term Load Forecasting of a Region of India using Generalized Regression Neural Network." Global Journals of Research in Engineering 17, no. F7 (2017): 15-22.
[24] Patel, Harsh, Mahesh Pandya, and Mohan Aware. "Short term load forecasting of Indian system using linear regression and artificial neural network." In 2015 5th Nirma University International Conference on Engineering (NUiCONE), pp. 1-5. IEEE, 2015.
[25] Farrokhabadi, M., Browell, J., Wang, Y., Makonin, S., Su, W., & Zareipour, H. (2022). Day-ahead electricity demand forecasting competition: Post-covid paradigm. IEEE Open Access Journal of Power and Energy, 9, 185-191.