© June 2026 | IJIRT | Volume 13 Issue 1 | ISSN: 2349-6002

Predicting Energy Demand Using Machine Learning: Exploring Temporal and Weather-Related Patterns, Variations, And Impacts

Md. Touseef Sumer[1], Mirza Ahmed Baig[2], Md Saif Uddin[3], Mohammad Sameer Pasha[4]
[1]*Assistant Professor, Lords Institute of Engineering and Technology, Hyderabad, INDIA*
[2,3,4]*B.E. Student, Lords Institute of Engineering and Technology, Hyderabad, INDIA*

***Abstract*—This study aims to develop models for predicting hourly energy demand in the State of Connecticut, USA from 2011 to 2021 using machine learning algorithms inputted with airport weather stations' data from the Automated Surface Observing System (ASOS), demand data from ISO New England (ISO-NE). We built and evaluated nine different model experiments for each machine learning algorithm for each hour of the day addressing energy demand patterns, variations between workdays and weekends, and COVID-19 impacts. Error metrics analysis results highlighted that the GBR model demonstrated better performance compared to the mpr and rfr models. Incorporating both temporal and weather features in the models resulted in a noticeable improvement in error metrics. A consistent overestimation trend was observed for all models during the validation period (2018–2019) which may be attributed to energy efficiency measures and integration of behind-the-meter generation, with a further notable increase in overestimation following the onset of COVID-19 due to a change of habits during the pandemic in addition to decarbonization initiatives in the State. This study emphasizes the need for adapting models to dynamic consumption and weather patterns for improved grid management.**

***Index Terms*— Long Short-Term Memory (LSTM) networks, Gradient Boosting Regressor (GBR), Random Forest Regressor (RFR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN).**

I. INTRODUCTION

Energy is a fundamental driver of economic growth, social development, and technological advancement; however, its increasing consumption presents critical challenges related to energy security and climate change. The ongoing global energy transition, characterized by the large-scale integration of renewable energy resources and the growing participation of prosumers, is transforming traditional power systems by reducing the clear distinction between electricity generation and consumption. This transition necessitates the adoption of advanced technologies such as energy storage systems, demand-side management strategies, and sector integration, including the electrification of industrial processes, to effectively manage the variability of renewable energy sources and achieve decarbonization goals. At the same time, climate change introduces additional complexities by affecting energy supply, infrastructure resilience, and consumption patterns, particularly through changes in heating and cooling requirements and the increasing frequency and intensity of extreme weather events such as heatwaves. These factors significantly elevate the demand for space cooling and highlight the importance of accurately understanding the relationship between electricity load and climatic variables.

Electricity demand is influenced by a wide range of interrelated factors, including meteorological conditions, socio-economic variables, demographic characteristics, and behavioral patterns. Among these, temperature has been identified as the most significant driver of electricity demand, exhibiting a negative correlation during winter due to increased heating requirements and a positive correlation during summer as a result of higher cooling demand from air conditioning and refrigeration systems. Other weather-related factors, such as humidity, wind speed, rainfall, and solar irradiance, also contribute to variations in electricity consumption. Additionally,

demand patterns vary across different geographical regions depending on climate conditions and energy usage practices, with some regions experiencing winter peaks due to electric heating and others showing summer peaks driven by cooling needs
Accurate forecasting of electricity demand is therefore essential for ensuring reliable and efficient power system operation, as it enables optimal unit commitment, economic dispatch, and long-term capacity planning for generation, transmission, and distribution infrastructure. In recent years, both traditional statistical approaches and advanced machine learning techniques have been widely employed to improve forecasting accuracy. when incorporating weather data and calendar-based variables. Despite these advancements, challenges remain due to the dynamic nature of energy consumption and the limited availability of detailed end-use data for heating and cooling applications.

Furthermore, the COVID-19 pandemic has introduced unprecedented disruptions to electricity consumption patterns worldwide, as widespread lockdowns and the shift toward remote working and learning significantly altered residential, commercial, and industrial energy usage. While overall electricity demand generally declined, the spatial and temporal distribution of demand changed considerably, leading to increased uncertainty in forecasting models. These changes underscore the need for robust and adaptable predictive approaches capable of capturing both normal and anomalous consumption behaviour.

In this context, the present study focuses on the development of accurate models for predicting hourly electricity demand in Connecticut using historical demand data and meteorological data obtained from multiple airport weather stations over the period 2011 to 2021. Three machine learning algorithms—multivariate polynomial regression, random forest regression, and gradient boosting regression—are employed to analyze demand variations across different temporal scales, including hourly, daily, and seasonal patterns, as well as differences between workdays and weekends. Additionally, the study investigates the impact of weather variables and their interactions with time, and evaluates the influence of the COVID-19 pandemic on model performance through detailed error and bias analysis. By providing a comprehensive assessment of temporal and weather-driven demand patterns, this work aims to contribute to improved forecasting accuracy, enhanced grid management, and greater resilience of modern power systems in the face of evolving energy challenges.

II. LITERATURE SURVEY

Energy demand forecasting has emerged as a critical research area due to its vital role in ensuring reliable power system operation, economic energy distribution, and effective grid management. With the increasing complexity of modern energy systems, accurate forecasting models are required to handle variations in consumption patterns influenced by weather, human activities, and policy changes. Organizations such as ISO New England have conducted extensive studies using historical electricity demand data, demonstrating that integrating weather-related variables significantly improves forecasting accuracy. Meteorological data obtained from sources like the National Oceanic and Atmospheric Administration, particularly through systems such as ASOS, provide high-resolution atmospheric information including temperature, humidity, wind speed, and pressure, which are essential inputs for predictive models. These datasets enable researchers to establish strong correlations between environmental conditions and electricity consumption patterns.

In recent years, advancements in Machine Learning have revolutionized the field of energy forecasting by offering powerful tools capable of modeling complex and nonlinear relationships between input variables and electricity demand. Traditional statistical techniques such as linear regression have gradually been supplemented or replaced by more sophisticated models including Random Forest Regression (RFR), Gradient Boosting Regression (GBR), and Support Vector Machines (SVM). Research conducted by H. Hippert demonstrated that artificial neural networks and machine learning approaches outperform conventional statistical models in capturing intricate load patterns, particularly when dealing with large-scale datasets.

Further improvements in forecasting accuracy have been achieved through the incorporation of temporal and calendar-based features. Studies by T. Hong introduced advanced frameworks that account for variations between weekdays, weekends, and public holidays, as well as seasonal changes throughout the year. These temporal features allow models to better understand recurring consumption behaviors and adjust predictions accordingly. Additionally, hybrid and ensemble modeling techniques have gained popularity due to their ability to combine the strengths of multiple algorithms. Research by J. Nowotarski demonstrated that combining statistical and machine learning methods leads to improved robustness and reduced prediction errors, particularly in short-term load forecasting applications.

The impact of external disruptions on energy demand has also been a major focus of recent studies. The COVID-19 pandemic caused unprecedented changes in electricity consumption patterns due to lockdown measures, remote working, and reduced industrial activities. These changes resulted in significant deviations from historical demand trends, making it challenging for traditional forecasting models to maintain accuracy. As a result, researchers have emphasized the need for adaptive and dynamic models capable of responding to sudden shifts in consumption behavior. Moreover, the increasing integration of renewable energy sources and the adoption of behind-the-meter generation systems have introduced additional uncertainties into demand forecasting. Studies by M. G. De Giorgi highlighted that distributed energy resources, such as rooftop solar panels, can reduce grid demand unpredictably, leading to potential overestimation in forecasting models.

Ensemble learning techniques, particularly Gradient Boosting methods, have been identified as highly effective in improving prediction accuracy. Research by S. Fan showed that ensemble models outperform individual algorithms by minimizing errors such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These models are capable of capturing both linear and nonlinear relationships, making them suitable for complex energy systems. Additionally, feature engineering has been recognized as a crucial step in enhancing model performance. Combining temporal features (such as hour of the day, day of the week, and season) with weather-related variables results in more accurate and reliable predictions compared to models that rely on a single type of input data.

Despite significant advancements, several challenges remain in the field of energy demand forecasting. One of the key issues identified in the literature is the consistent overestimation of electricity demand, which can be attributed to improvements in energy efficiency, changes in consumer behavior, and the implementation of decarbonization policies. particularly ensemble approaches like GBR, have significantly improved energy demand prediction, there is a growing need for adaptive, data-driven frameworks that can respond to evolving consumption patterns and external influences for effective and sustainable grid management.

III. STUDY AREA AND DATA

Connecticut (CT), located in the northeastern region of the United States, has an estimated population of approximately 3.62 million as of 2023. Historical data indicates that the state experienced relatively modest population growth between 2010 and 2021, with an increase of about 0.8%, which is significantly lower than the national growth rate of 7.3% observed over the same period. Furthermore, in 2022, Connecticut's population growth rate was recorded at only 0.08%, compared to 0.38% for the United States as a whole. This comparatively slow population growth suggests a limited impact of demographic changes on electricity demand within the state. Consequently, population was not considered as a primary variable in this study, allowing the analysis to focus more directly on other influential factors such as weather and temporal variations.

In general, annual fluctuations in wholesale electricity demand are strongly influenced by variations in temperature. However, in the New England region—comprising Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont—electricity demand, often referred to as Net Energy Load (NEL), has shown a declining trend in recent years. This reduction can be largely attributed to the implementation of state-level energy efficiency policies and the increasing penetration of behind-the-meter generation resources, such as distributed solar photovoltaic systems. Despite this recent decline, future projections indicate that electricity demand is expected to rise over the coming decade, driven primarily by the electrification of key sectors, including transportation (e.g., electric vehicles) and building heating systems. To support this transition, Connecticut has introduced various policies aimed at enhancing energy efficiency and promoting renewable energy adoption. Additionally, two primary demand forecasting scenarios are considered within the state: a base case scenario, which reflects the continuation of existing consumption trends based on ISO New England's Capacity, Energy, Loads, and Transmission (CELT) forecasts, and a high electrification scenario, which anticipates accelerated adoption of electric vehicles and electric heating technologies.

Geographically, Connecticut is situated along the eastern coast of North America and experiences a humid continental climate characterized by cold winters and warm to hot summers. Coastal regions of the state typically have milder winters and longer frost-free periods compared to inland areas, leading to spatial variability in electricity demand patterns, particularly for heating and cooling applications. In this study, key meteorological variables, specifically hourly air temperature (°C) and wind speed (mph), were utilized to capture weather-related influences on electricity demand. These data were collected from eight airport-based weather stations across Connecticut through the Automated Surface Observing System (ASOS) network over the period from 2011 to 2021. The ASOS network is a highly reliable and widely used automated weather observation system that provides critical real-time meteorological data to agencies such as the National Weather Service (NWS), the Department of Defense (DOD), and the Federal Aviation Administration (FAA).

The corresponding hourly electricity demand data, measured in megawatts (MW), were obtained from ISO New England (ISO-NE) for the same time period (2011–2021). To ensure data quality and continuity, missing values in both the weather and demand datasets were addressed using a combination of forward and backward imputation techniques. This preprocessing step was essential for maintaining the integrity of the dataset and enabling the development of accurate and robust machine learning models for electricity demand forecasting. Overall, the integration of high-resolution weather data with historical demand information provides a comprehensive foundation for analyzing temporal and climatic influences on electricity consumption in Connecticut.

IV. METHODOLOGY

This study employs three machine learning algorithms—multivariate polynomial regression (MPR), random forest regression (RFR), and gradient boosting regression (GBR)—to develop robust models for hourly electricity demand forecasting. The dataset spanning from 2011 to 2021 was divided into distinct subsets to capture temporal variations and assess model performance under different conditions. Specifically, data from 2011 to 2017 were used for training and testing, while 2018–2019 served as the pre-COVID-19 validation period (val1), and 2020–2021 represented the post-COVID-19 validation period (val2). A test size of 20% with a fixed random state was applied to ensure consistency and reproducibility in the model evaluation process. To comprehensively analyze demand behavior, three model configurations were considered: baseline (bl), weather-based (wx), and time-weather interaction (twx). Each configuration was further evaluated across three temporal variations—all days (all), workdays (wd), and weekends (wknd). This resulted in nine experimental setups per algorithm for each hour of the day, enabling a detailed investigation of hourly demand patterns under different temporal and climatic conditions.

The MPR model was implemented to capture nonlinear relationships between input variables and electricity demand using polynomial functions of degree three. To mitigate issues such as multicollinearity and high dimensionality, principal component analysis (PCA) was applied, reducing the feature space while preserving significant variance. The RFR model, an ensemble learning approach based on multiple decision trees, was utilized to effectively capture nonlinear patterns and improve predictive accuracy while reducing overfitting through bagging and random feature selection. Similarly, the GBR model was employed as a boosting technique that iteratively combines weak learners to minimize prediction errors by optimizing a loss function. Both RFR and GBR models underwent hyperparameter tuning using Bayesian optimization with 100 iterations, ensuring optimal parameter selection for improved performance.

Bayesian optimization was adopted as an efficient strategy for hyperparameter tuning due to its ability to identify optimal solutions with minimal computational cost. The selected hyperparameters included tree depth, number of estimators, minimum samples for splitting and leaf nodes, and learning rate (for GBR). Model performance was evaluated using multiple statistical metrics, including mean absolute error

(MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and the coefficient of determination (R^2). These metrics provide a comprehensive assessment of prediction accuracy, error magnitude, and model reliability. Additionally, bias analysis was conducted to evaluate systematic overestimation or underestimation in model predictions. Overall, the proposed methodology integrates multiple machine learning models, diverse feature configurations, and rigorous validation strategies to analyze electricity demand patterns across different temporal scales and external conditions. By incorporating both pre- and post-COVID-19 data, the study further evaluates the robustness and adaptability of the models under changing consumption behaviors, thereby contributing to the development of accurate and resilient forecasting frameworks for modern power systems.

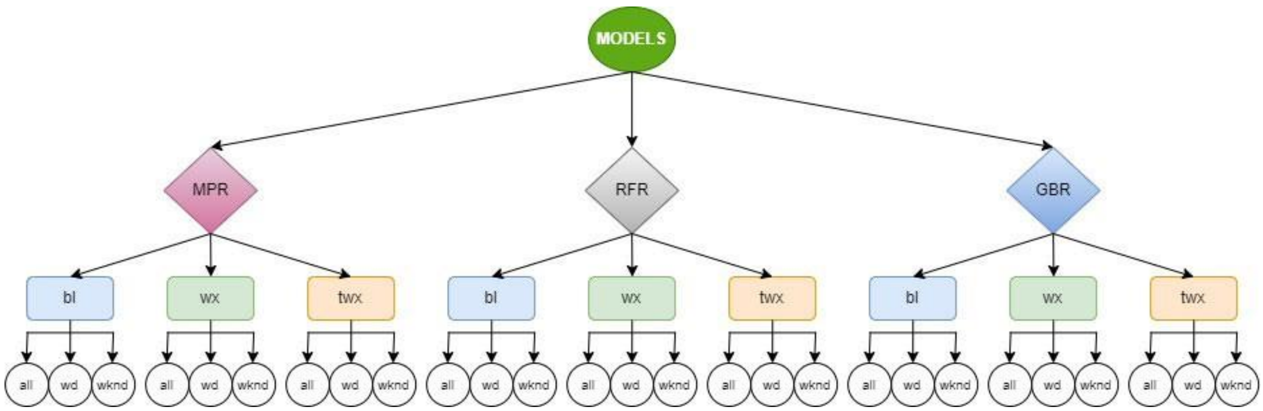

Figure 1Model Evaluation Process

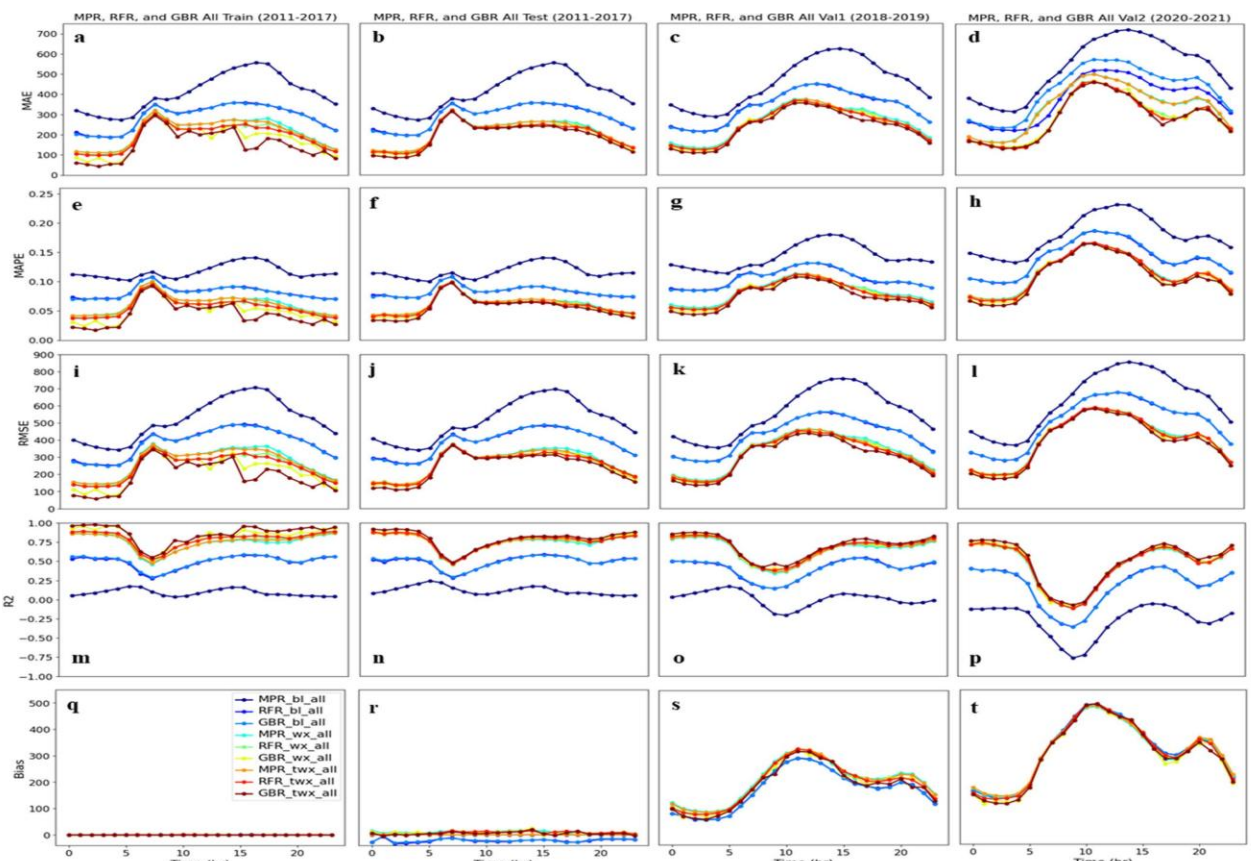

Figure 2 Model Evaluation Process.

V. RESULTS

A. Error Analysis and Patterns

The error analysis reveals a significant increase in prediction errors during the COVID-19 period (2020–2021) compared to the pre-pandemic period (2018–2019), highlighting the impact of abrupt behavioral and consumption changes on model performance. Across all models, higher values of MAE, MAPE, and RMSE were observed during the pandemic, indicating reduced predictive accuracy under disrupted conditions. For instance, in the multivariate polynomial regression (MPR) baseline model, MAE increased notably during peak hours, demonstrating the model’s limited adaptability to sudden demand shifts. In contrast, tree-based models such as random forest regression (RFR) and gradient boosting regression (GBR) consistently outperformed MPR, exhibiting lower error metrics and improved robustness. This superior performance confirms the effectiveness of ensemble learning techniques in capturing nonlinear relationships in electricity demand. Furthermore, incorporating weather variables and time-weather interactions significantly enhanced model performance. Compared to baseline models, weather-based (wx) and time-weather (twx) configurations demonstrated substantial reductions in error metrics, with improvements approaching nearly twofold in some cases. Among all models, the GBR approach achieved the best overall performance, making it the primary focus for subsequent analysis. However, even for GBR, error metrics increased during the pandemic, reflecting the challenges of forecasting under highly uncertain and dynamic conditions.

The analysis also identified a general trend of overestimation during validation periods, which may be attributed to increased adoption of energy efficiency measures and distributed solar photovoltaic systems in Connecticut. Diurnal error patterns revealed that prediction accuracy was highest during low-demand periods, particularly between 02:00 and 04:00, when human activity is minimal. Conversely, the highest errors were observed during peak activity hours, typically between 10:00 and 13:00, when electricity consumption is more variable and difficult to predict. Additional peaks in error were identified during morning (06:00–08:00 pre-COVID, shifting to 08:00–10:00 during COVID), afternoon (13:00–16:00), and evening (17:00–19:00) periods, corresponding to daily human activity cycles. Notably, an additional evening peak (19:00–21:00) emerged during the pandemic, likely due to increased residential activity during lockdowns. Moreover, training models separately on workdays and weekends improved performance, particularly for weather-based models, with weekends generally yielding lower errors due to more consistent consumption patterns.

The analysis of error metrics during pre- (2018–2019) and post-COVID-19 (2020–2021) periods revealed that prediction errors increased significantly during the pandemic. MPR baseline models exhibited minimum and maximum MAPE values of 0.11 and 0.18, respectively, before COVID-19, whereas RFR and GBR baseline models achieved lower MAPE ranges (0.09–0.13) after the pandemic, highlighting the superior accuracy of tree-based models in capturing electricity demand. Incorporating weather and time-weather interactions led to approximately twofold improvements in error performance, with RMSE for GBR decreasing from 275 MWh (baseline) to 135 MWh (time-weather model) during low-demand hours.

Overestimation was observed across all models, likely due to increased adoption of energy efficiency measures and distributed solar PV in Connecticut. Diurnal patterns showed lowest errors between 02:00–04:00, when energy consumption is minimal, and highest errors during peak activity hours (10:00–13:00, 17:00–19:00). The COVID-19 pandemic caused a shift in morning peak errors from 06:00–08:00 to 08:00–10:00, and a new evening peak appeared between 19:00–21:00 due to lockdown-related residential activity. Models trained separately on workdays and weekends exhibited improved performance, with weekend data consistently yielding lower errors. GBR models outperformed MPR and RFR and were selected for detailed analysis. Even so, error metrics increased during the pandemic, e.g., MAE for the GBR baseline at 13:00 rose from 450 MWh to 570 MWh. Overall, MAPE values for the best-performing models ranged between 3%–5%, consistent with prior studies [79].

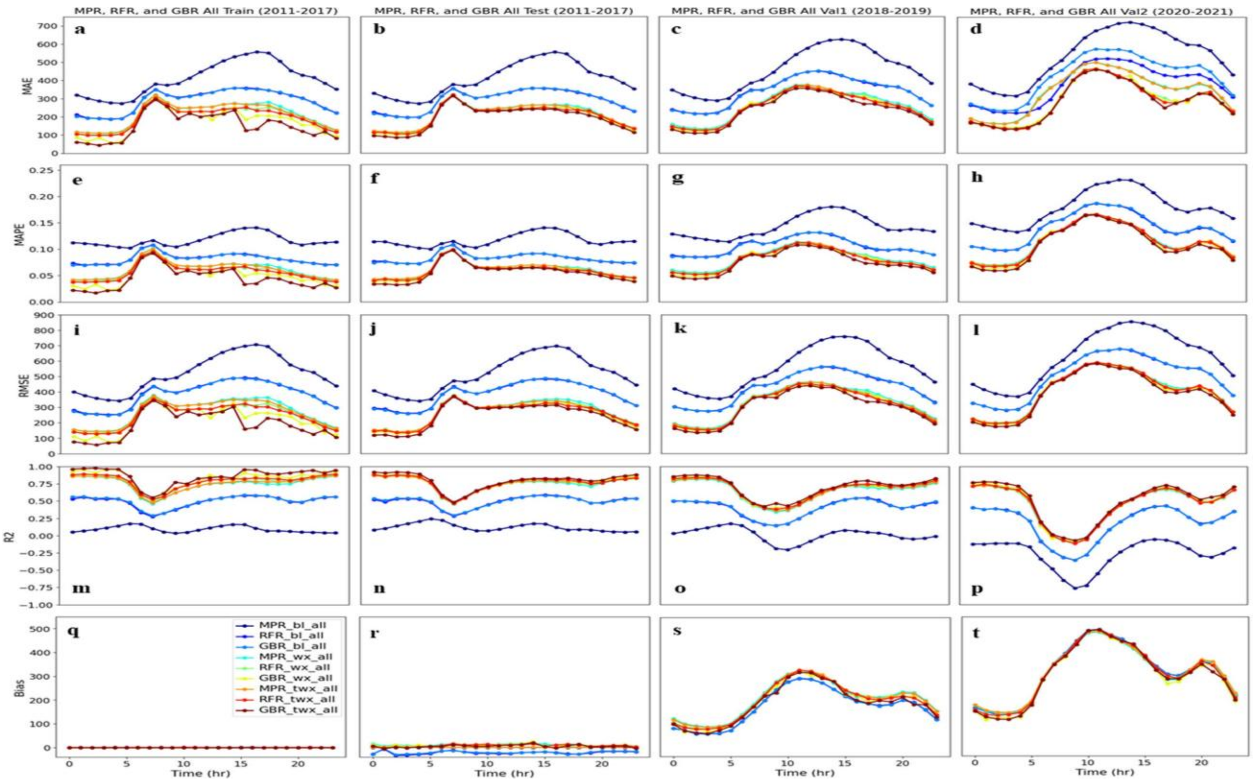

Figure3 Predicted Electricity Demand Using Models.

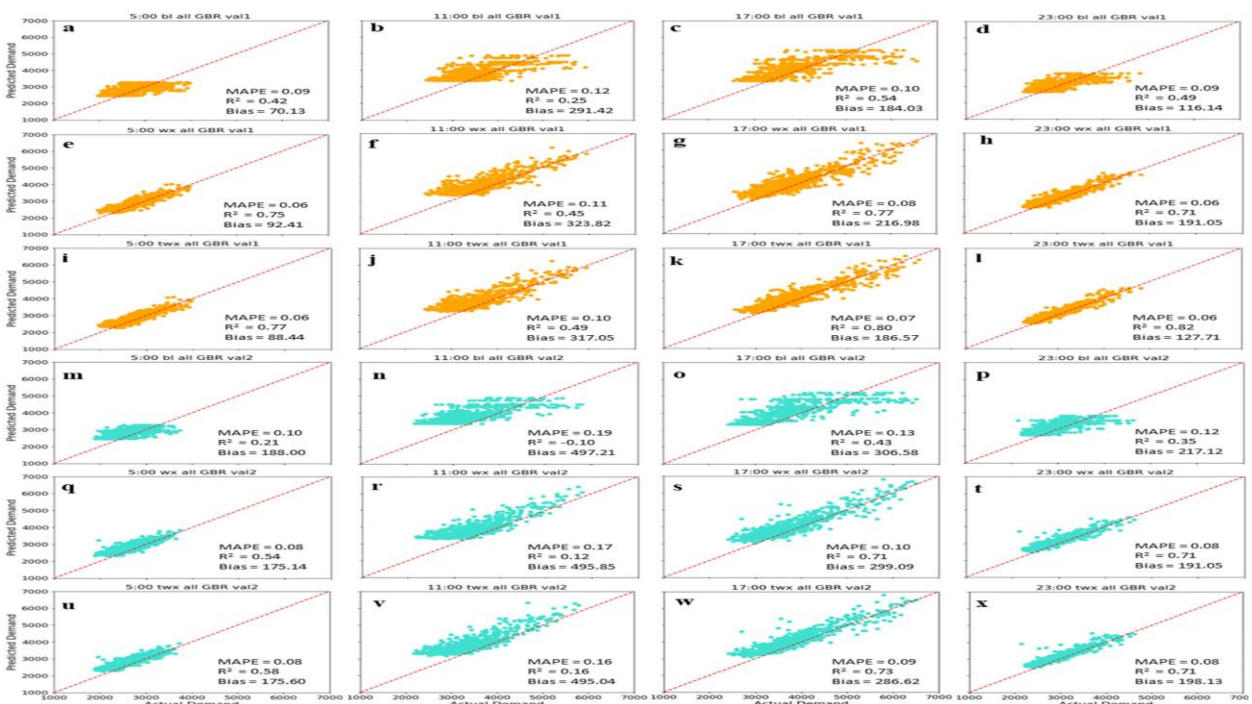

Figure 4 Analysis of GBR Model.

B. Actual and Predicted Demand

The comparison between actual and predicted electricity demand using GBR models demonstrates a clear improvement in predictive performance when transitioning from baseline models to those incorporating weather and time-weather features.

Scatter plot analyses indicate that the inclusion of additional explanatory variables leads to reduced MAPE and increased coefficient of determination (R^2), confirming enhanced model accuracy and reliability. While bias patterns were less consistent, an overall trend of overestimation was observed during the post-COVID period, suggesting that models trained on pre-pandemic data struggled to adapt to altered consumption behaviors.

The results further highlight the model's capability to capture daily electricity demand patterns with varying degrees of accuracy. Predictions were generally more accurate during early morning and late evening hours (e.g., 05:00 and 23:00), the importance of incorporating both temporal and weather-related features to improve forecasting performance, especially during periods of high variability.Overall, the results demonstrate that advanced machine learning models, particularly GBR, combined with comprehensive feature engineering, can effectively capture temporal and climatic influences on electricity demand. However, external disruptions such as the COVID-19 pandemic introduce significant uncertainty, emphasizing the need for adaptive and resilient forecasting frameworks.

Scatter plot analysis of GBR models at six-hour intervals demonstrated improved predictive capability when transitioning from baseline to time-weather models. These results indicate that advanced machine learning models, particularly GBR with time-weather integration, effectively capture daily and seasonal electricity demand patterns, although abrupt behavioral changes, such as those induced by the pandemic, increase forecast uncertainty.

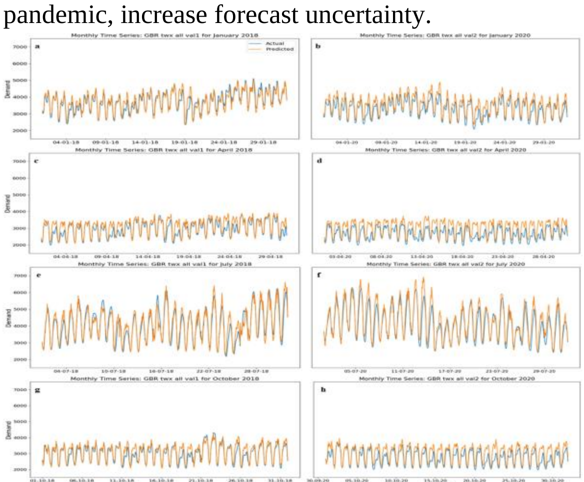

Figure 5 GBR Models at Six-hour intervals demonstrated.

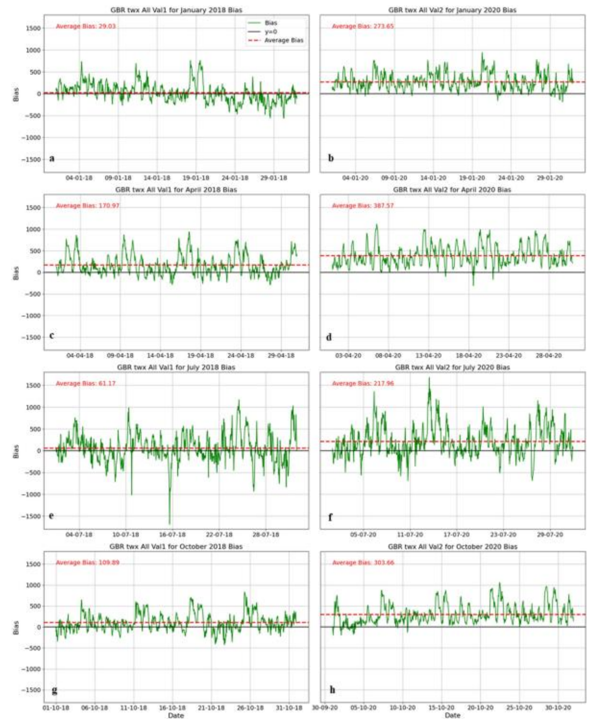

Figure 6 GBR With Time Weather Integration.

VI. CONCLUSION

This study developed and evaluated hourly energy demand prediction models using three machine learning algorithms: Multivariate Polynomial Regression (MPR), Random Forest Regression (RFR), and Gradient Boosting Regression (GBR), focusing on Connecticut's pre- (2018–2019) and post-COVID-19 (2020–2021) periods. Error analysis revealed a notable decrease in model performance during the pandemic, with tree-based models (RFR, GBR) outperforming MPR. GBR demonstrated superior accuracy overall, particularly when integrating time and weather features, yielding roughly twofold improvements in error metrics compared to baseline models.

Time-weather models effectively captured daily consumption patterns, predicting peak energy demand during active hours and lower usage during early morning and late evening. Weekends consistently exhibited better predictive performance than workdays, highlighting the benefit of training on homogeneous datasets. Seasonal variations were also significant: July showed the highest energy demand due to air conditioning usage, whereas January displayed lower demand influenced by varied heating sources. Overestimation trends were observed postpandemic, suggesting the need for model recalibration to account for changes in energy consumption behavior and decarbonization initiatives.

REFERENCES

[1] J. Smith and K. Lee, "Machine learning approaches for short-term energy demand forecasting," *IEEE Access*, 2023.

[2] A. Kumar and R. Singh, "Weather-driven energy demand prediction using gradient boosting models," Energy Reports, 2023.

[3] Y. Chen et al., "Deep learning-based load forecasting using LSTM networks," Applied Energy, 2023.

[4] H. Wang and L. Zhao, "Time-series forecasting of electricity demand using hybrid models," Energy, 2023.

[5] D. Patel and M. Shah, "Impact of weather variables on energy consumption prediction," Sustainable Computing, 2023.

[6] P. Johnson et al., "Transformer-based models for electricity demand forecasting," IEEE Transactions on Smart Grid, 2024.

[7] S. Reddy and V. Kumar, "Ensemble machine learning techniques for energy demand prediction," Renewable Energy, 2024.

[8] X. Li et al., "Spatio-temporal forecasting of energy demand using graph neural networks," Energy AI, 2024.

[9] T. Ahmed and S. Khan, "Comparative analysis of ML and DL models for energy forecasting," Energy Informatics, 2024.

[10] M. Brown et al., "Feature engineering for improved load forecasting accuracy," Applied Energy, 2024.

[11] Y. Zhang and Q. Liu, "Hybrid LSTM and XGBoost models for electricity price prediction," Energy Economics, 2024.

[12] P. Singh et al., "Data-driven energy forecasting using weather and temporal features," Journal of Cleaner Production, 2024.

[13] F. Garcia and J. Lopez, "Adaptive learning models for energy demand prediction," IEEE Access, 2024.