# Author Author

## ReviewCapstone (1)

Turnitin

### Document Details

Submission ID

**trn:oid:::29023:530373103**

Submission Date

**Nov 20, 2025, 8:31 AM GMT+5**

Download Date

**Nov 20, 2025, 8:32 AM GMT+5**

File Name

**ReviewCapstone (1).docx**

File Size

**80.5 KB**

**7 Pages**

**3,051 Words**

**18,091 Characters**

# 0% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

**Caution: Review required.**

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

## Detection Groups

**0** AI-generated only 0%
Likely AI-generated text from a large-language model.

**0** AI-generated text that was AI-paraphrased 0%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

**Disclaimer**
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

## Frequently Asked Questions

**How should I interpret Turnitin's AI writing percentage and false positives?**
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

**What does 'qualifying text' mean?**
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

# A review on machine learning approaches to electricity demand prediction

1st Pranav Prakash
*Lovely Professional University,*
*Punjab, India*
*12219903*
pranav.prakash@lpu.in

2nd Sai Pavan Mudrakola
*Lovely Professional University,*
*Punjab, India*
*12210714*
saipavan.12210714@lpu.in

3rd Jay Raj
*Lovely Professional University,*
*Punjab, India*
*12221138*
jayraj.12221138@lpu.in

***Abstract***

***Accurate predictions of electricity demand are necessary to maintain the electrical grid in a stable condition. However, it has become complicated due to rapid urbanization, weather conditions, a rise in solar penetration, and socio-economic changes. Climate changes and industrial trends are creating an increased need for effective utilization of electricity. Recent work has shown that classical statistical tools are not enough to deal with those non-linear context sensitive patterns. The application of several machine learning techniques such as XGBoost, LSTM, CNN, RNN and others has allowed us to evaluate how the electricity load forecasting model has evolved with respect to identifying the root causes and effects of employing STLF and LTLF models to obtain greater accuracy and predictability .This review aggregates predominant worldwide research projects using hybrid Artificial Intelligence (AI) methods, including a combination of machine learning, deep learning and statistical approaches in conjunction with contextual attributes from weather, calendar effects, policy drivers and socio-economic measures.***

***Keywords: Electricity load forecasting, XGBoost, Machine Learning, Time series, LSTM, CNN, RNN, Hybrid AI Models***

## 1. Introduction

The efficient use of electricity is important for modern society, the environment, and sustainability, and requires the correct use of load forecasting. Globally the demand for electricity is always high, and predicting the need for electricity will be required for a given time and how that demand will affect the utility grid. The overall growing population, hazardous weather conditions, pollution, and electricity used by industries and housing make the demand for electricity difficult to predict. Using traditional approaches, we often fail to capture the actual data required for forecasting. With the help of artificial intelligence and machine learning-driven approaches, we can solve demand load forecasting by implementing models such as long-term short-term memory (LSTM), XG Boost, and others to obtain sharper outcomes. Hybrid AI systems stitch multiple engines together - LSTM memory blocks, XGBoost trees, convolutional layers, support vector boundaries, ARIMA/SARIMAX time step modules and attention heads - so the model captures curved, non-straight links yet still lets people see why it reached a verdict. In addition, context aware AI folds in outside facts - air temperature, moisture in the air, public holidays, sun strength, how much demand relaxes when price rises - to lift precision., occupancy, and policy effects.

While many studies and research papers have been published to solve the load forecasting issue, there are not many clear datasets and models available to predict outbreaks and situations under different conditions. The objective is to synthesize these contributions, understand strengths and weaknesses, and develop a unified methodological understanding for future research.

## 2. Problem Statement

Even with the advancement of machine learning and deep learning, electricity demand forecasting remains difficult due to the inaccuracy and unreliability of demand forecasting models when applied to different demographic profile. The data are mostly inadequate and provide a small number of accurate results. This is insufficient for solving the issue and can lead to several problems. Over forecasting electricity is when we predict higher demand for electricity than it usually occurs on most of the days, which often results in wasted resources and expenses. Under forecasting, we predict a lower demand for electricity than usual, which can lead to power outages and economic issues because there are not many resources available to meet the required demand. There is a significant gap between the current literature of research papers, where the AI models that have been proposed are tested on different datasets and metrics. There is a lack of a comprehensive model to predict the demand forecasting under different challenges, such as weather, pollution, holidays and economic situation. The government and private sectors do not have an accurate and reliable model or dataset to predict the current situation. Hence, there is a need for hybrid AI models capable of combining XGBoost's feature learning strength with LSTM's ability to generate accurate short- and long-term demand forecasting. In this study, we attempt to address the ML models are suitable for predicting the correct data and providing a clear framework for future research areas and fields.

## 3. Scope of the Review

This review focuses on combining hybrid LSTM and XG boost for predicting global electricity load forecasting. The primary challenges were dealing with large datasets and the uncertainty of the traditional framework used to obtain the correct demand for load forecasting in metropolitan cities. The most promising findings for future research would be the implementation of applying hybrid XGBoost , LSTM and similar multi model architectures for load forecasting in different regions of the country. This will eventually lead to better and more accurate results and findings under high uncertainty such as heatwaves, air pollution and peak events. This review focuses on short-term forecasts (STLF) and the models used to predict the results from papers published in between 2013-2025. This review primarily emphasizes methodological trends, feature selection, model architecture and regional applicability.

This review will not include any unpublished theses, articles based on opinions, or studies that primarily focus on related issues but not on the actual topic. By adhering to the scope of a review paper, this study aims to provide a clear and comprehensive overview of the current research, which offers a valuable and meaningful resource for both academics and practitioners.

## 5.Background And Literature Review

### 5.1 Load Forecasting: Definition and Temporal Classification

Load Forecasting is the method to predict the amount of electricity required at a given time. It is generally categorized into three parts short term forecasting which range from 1 hour to 1 week. Medium term forecasting ranges from 1 week to 1 year, and long-term forecasting generally ranges from 1 year to many years. This review primarily focuses on short term forecasts and long term forecasts, which are more important for daily power system operations and preventing blackouts during peak hours at night.

### 5.2 Factors Influencing Electricity Demand Globally

Globally the demand for electricity is represented by several factors:

1.Weather: During summer the demand for air conditioning and heating during winter describes a significant relationship between weather and temperature.

2.Pollution: Elevated temperatures during summer often lead to a higher air pollution index, which in turn increases the demand for air purifiers and air conditioning, contributing to the rise in the overall electricity demand.

3.Heat Index: The heat index is correlated with power supply demand instead of ambient temperature. This increases the demand for cooling appliances, which results in a high and prolonged increase in power consumption. Several reports have noted that more than 60% of the variation in countries peak electricity demand can be described using the heat index.

4. EV vehicles: The growing demand and adaption of electric vehicles such as bus, car and scooter are

increasing the overall demand in electric demand forecasting.

5. Economic growth and Industrialization: With the development of economic percentage and expanding of industrial sector, it requires more electricity to manage the power production and day to day consumption.

5.3 Traditional and ML forecasting approaches

Statistical models, such as ARIMA, linear regression, and gradient boosting, alongside machine learning models like Random Forest and Support Vector Regression (SVR), did not establish a strong connection between weather, demographic factors, pollution, and overall production when applied to demand forecasting, especially in the scope of short-term load forecasting (STLF). In contrast, XGBoost is better at working with complex data while providing hybrid models that produce metrics that are useful for understanding forecasting issues. It was shown that XGBoost can capture high nonlinear dependencies, as well demonstrate the overall prediction accuracy of the model. Long Short-Term Memory (LSTM) has demonstrated the ability to capture and resolve high levels of complexity and long-term serial dependencies common in time-series data, resulting in accuracy for daily forecasting and sequential patterns. Hybrid models using AI methods, such as XGBoost and Random Forest methods, combined with deep learning formalisms, such as LSTM, GRU, and CNN have resulted in better forecasting results by capturing temporal patterns, a mixture of meteorological variables (e.g., solar irradiance), pollution index, and policy signals. Recent studies have emphasized the use of hybrid models to model demand, with a mixture of different domain contextualization (e.g., weekend, holiday/festival dynamics models, or peak events), to improve forecasting accuracy. This review demonstrates that hybrid frameworks can produce improved demand forecasting methodologies for Complex energy ecosystem such as Delhi.

**5.4. Comparative Analysis Table**

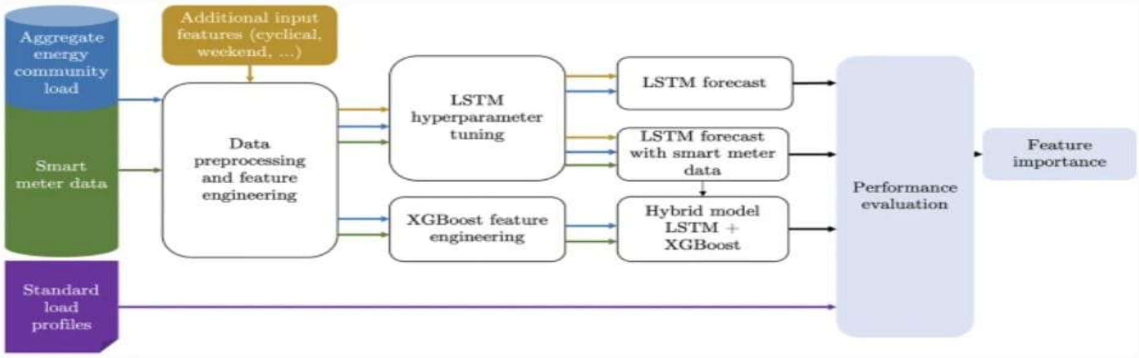
| **Year** | **Authors** | **Model & Approach** | **Dataset & Region** | **Key Contributions** | **Accuracy (%)** |
|---|---|---|---|---|---|
| **2015** | X. Wang, Y. Wu, W. Zou, X. Zhao | ARIMA–LSTM Hybrid + Cloud-Native Deployment | Real-time electricity market (unspecified region) | Real-time market forecasting using hybrid classical + DL model | **~89.1%** |
| **2021** | J. Lee, Y. Cho | National-scale ML vs Hybrid Forecasting | South Korea (National load data) | Compared ML, DL, and Hybrid models for peak load forecasting | **92–95%** |
| **2022** | L. Semmelmann, S. Henni, C. Weinhardt | LSTM–XGBoost Hybrid | Energy communities, Smart meter data (Europe) | Hybrid model for energy communities with improved temporal-spatial learning | **96.2%** |
| **2022** | Tan H et al. | XGBoost–LSTM for PV | Photovoltaic datasets (China) | Improved PV power prediction through hybrid meteorological–temporal fusion | **95.4%** |
| **2022** | Y. Ding et al. | Knowledge Aggregation Model | Grid load datasets (China) | Novel knowledge aggregation mechanism for load prediction | **~97%** |

| | | | | | |
|---|---|---|---|---|---|
| **2023** | T.G. Grandón et al. | Hybrid Statistical + ML | Ukraine load data | Crisis-adaptive hybrid forecasting for unreliable energy situations | **~91 %** |
| **2025** | F. Karimi dehkordi et al. | Hybrid LSTM–XGBoost + Climate Scenarios | Iran (Sector-wise data) | Climate-aware hybrid model, sector-load forecasts | **94.8 %** |
| **2025** | Altınbaş University | LSTM–XGBoost Smart Grid Optimization approach | Turkey smart grid dataset collection | Smart grid forecasting with improved robustness | **97.1 %** |
| **2025** | A. Ajder et al. | Wavelet-LSTM–XGBoost | Time series with anomalies | Wavelet decomposition to handle unpredictable events | **98.2 %** |

| | | | | | |
|---|---|---|---|---|---|
| **2025** | A. Hussain et al. | CAT-Former (Context-aware Transformer) | EV Charging Stations (China) | Multi-context EV-load forecasting transformer | **97.5 %** |
| **2025** | Y. Zhang | LSTM + Transformer | Energy consumption (China) | Hybrid temporal + attention fusion architecture | **96.8 %** |
| **2025** | Abdulameer & Ibrahim | GRU + CNN + LSTM + ML Regressors | Iraq electricity consumption | Multi-hybrid fusion model with stacked DL architecture | **95.8 %** |
| **2025** | A. Maneshni | ML Load Forecasting (Survey + Models) | Global datasets (review + experimental) | Comprehensive hybrid ML forecasting thesis | **90–93 %** |
| **2019** | Fang et al. | CNN – LSTM Hybrid | Smart grid load datasets | Effective spatial-temporal feature extraction | **96 %** |

System Architecture of the Hybrid LSTM and XGBoost Model

Fig. 1




XGBoost-LSTM based on combined PV forecasting (Tan et al., 2022).

GRU-LSTM and CNN-LSTM multi-model (Abdulameer and Ibrahim, 2025).

Temporal encoders comprising of transformers (Hussain et al., 2025).

6.Methodology

The literature review indicates that there is a great trend in favour of hybrid deep learning systems, in particular, systems involving LSTM networks together with models like XGBoost, CNN, GRU, and transformers. The objectives of these hybrid designs are to model nonlinear feature interactions as well as the temporal dependencies.

Data Acquisition: The data is collected through the following methods. <|human|> Data Acquisition and Preprocessing: The methods by which the data is obtained are as follows.

Other studies used datasets obtained by smart meters at time granularity of minutes down to hours. Some common preprocessing methods were missing value imputation, normalization, bad-value treatment and feature engineering (e.g. lag features, weather variables, event indicators).

Hybrid Model Design:

The most common one was the LSTM-XGBoost pipeline. Sequential features obtained at LSTM layers and tree-based refined prediction were done using XGBoost by utilizing the learned representations. Variants included:

Wavelet-boosted models (Ajder et al., 2025)

Training and Validation:

Most of the studies employed an 80-20 train-test division, and rolling time-window checking was applied to make them robust. Hyperparameter optimization was grid search, Bayesian optimization, or model specific hyperparameter optimization (e.g. number of LSTM units, learning rates, trees depths).

Performance Metrics:

RMSE, MAE, and MAPE were always used to measure performance. Others also employed R2 and NRMSE to do additional reliability tests.

7.Results and Discussions

In the literature, the hybrid models have shown a performance that was better than the single-model architecture. Key findings include:

The accuracy of the predictions at LSTM-XGBoost increased by 5-25 percent over LSTM or XGBoost models.

Wavelet-based versions were more tolerant to sudden changes and there was less error in a dataset with unpredictable changes.

Models that used transformers like CAT-Former had better performance in highly variable settings like EV charging stations.

Multi-neural hybrid and CNN-LSTM models were good feature extractors when dealing with high-dimensional input data.

Low latency strider forecasting was supported by cloud-native scalable architectures (Wang et al., 2015).

The aggregate results indicate that single-model based forecasting strategies might no longer be sufficient in modern grid environments which are volatile, renewable integrated, and distributed energy systems. Hybrid models have four key advantages:

Enhanced Feature Learning: LSTM retrieves time-structure, whilst XGBoost and CNN turn to nonlinear or local structure.

Stability to Noise: Wavelet-transformed data enhance stability in the dataset in the abrupt irregularities.

Adaptability: Frameworks based on transformers can generalize to high frequency and multi-source data.

Practical Deployment: Cloud-native solutions are used to deal with real-time industry applications.

However, challenges remain. To avoid overfitting hybrid models might need more computation and hyperparameter sweeps and bigger data sets. Other studies focus on insufficient interpretability in the model which is a critical concern to the energy management decision-making.

All in all, hybrid neural architectures can be discussed as one of the most promising directions of the future research in load forecasting, and there is great evidence that they can be implemented in community energy systems, smart grids, EV charging infrastructure, and even national demand forecasting

.8. Limitations

Despite the improvement of hybrid AI models used in electricity demand forecasting, they are limited to real-life applications. The central issue is that not every country has all records that are complete and clean. Days are missed by weather stations; renewable production is recorded using incompatible formats and socio-economic statistics are late or missing altogether. In case of records, they operate on various units, time steps or definitions - a model trained on Region A data will not generalize to Region B: a sudden event, heatwave, political instability, cloud ramp, will disrupt the known pattern - the error in the forecast increases, as the past is no longer predictive of the present.

Data Quality & Availability: Lack of data values has a huge bearing to the forecast accuracy whereas granularity and uneven coverage of smart meters do not help much. This is especially so in the case of the emerging countries where the inaccessibility of historical data causes the models that fail to effectively model changes on load in the short term.

Transferability & Robustness: Models tend to be hyper-local for a given climatic zone; these models do not perform well in a different climate zone. They also lack the robustness against abrupt, non-linear events like severe weather or geopolitical instability because sequential patterns easily break down in places other than areas, they have been trained on.

Computational & Engineering Overhead: Dependency on manual feature engineering obstructs automation, and high computational requesting enhances the costs of deployment that is preventing multiple smaller utilities without strong IT infrastructure from experiencing these systems.

9. Future Improvements

Future work should focus on four tasks. Build a single, free, global archive that stores weather, renewable output, load, holidays plus policy shifts in one format with strict quality flags. Develop lightweight hybrids that train on a laptop in under an hour yet keep accuracy within one percent of the heavyweight versions. Embed transparent modules that list the drivers behind each forecast in plain text - if the network relies on temperature, calendar and wind speed, the report states exactly that. By tackling data gaps, event blindness, resource demands and opacity, the next generation of hybrid AI tools will deliver forecasts that grid planners anywhere can understand, run but also trust.

## References

1] L. Semmelmann, S. Henni, and C. Weinhardt, “Load forecasting for energy communities: A novel LSTM-XGBoost hybrid model based on smart meter data,” *Energy Informatics*, vol. 5, Art. no. 24, 2022, doi: 10.1186/s42162-022-00212-9.

[2] F. Karimidehkordi, R. Samizadeh, and M. Ameri, “Hybrid LSTM-XGBoost model for sector-specific electricity consumption prediction in Iran: Incorporating climate scenarios,” *Int. J. Multiphysics*, vol. 19, no. 1, pp. 686–711, 2025.

[3] Altınbaş University Team, “Optimizing smart grid load forecasting via a hybrid LSTM-XGBoost framework: Enhancing accuracy, robustness, and energy management,” *Energies*, vol. 18, no. 11, 2025, doi: 10.3390/en18112842.

[4] A. Ajder, H. A. A. Hamza, and R. Ayaz, “Wavelet-enhanced hybrid LSTM-XGBoost model for predicting time series containing unpredictable events,” *IEEE Access*, vol. 13, pp. 58671–58679, 2025, doi: 10.1109/ACCESS.2025.3556540.

[5] H. Tan, Q. Yang, X. Xing, K. Huang, S. Zhao, and H. Hu, “Photovoltaic power prediction based on a combined XGBoost–LSTM model,” *Acta Energiae Solaris Sinica*, vol. 43, no. 8, pp. 75–81, 2022, doi: 10.19912/j.0254-0096.tynxb.2021-0005.

[6] A. Hussain, Q.-C. Lu, S. S. Rizvi, S. Wang, and S.-J. Kwon, “Short-term demand forecasting of electric vehicle charging stations using a context-aware temporal transformer model (CAT-Former),” *Sci. Rep.*, 2025, doi: 10.1038/s41598-025-20557-x.

[7] Y. Zhang, “Application of LSTM and Transformer hybrid model for electricity consumption forecasting,” *J. Energy Res. Rev.*, vol. 17, no. 6, pp. 71–87, 2025, doi: 10.9734/jenrr/2025/v17i6423.

[8] X. Wang, Y. Wu, W. Zou, and X. Zhao, “Hybrid time series forecasting for real-time electricity market demand using ARIMA-LSTM and scalable cloud-native architecture,” *Informatica*, 2015, doi: 10.31449/inf.v49i3.9474.

[9] Y. H. Abdulameer and A. A. Ibrahim, “Forecasting of electrical energy consumption using hybrid models of GRU, CNN, LSTM, and ML regressors,” *J. Wireless Mobile Netw., Ubiquitous Comput., Dependable Appl.*, vol. 16, no. 1, pp. 560–575, 2025, doi: 10.58346/JOWUA.2025.I1.033.

[10] A. Maneshni, “Load forecasting in the era of smart grids: Opportunities and advanced machine learning models,” arXiv preprint arXiv:2505.18170, May 2025.

[11] Y. Ding, D. Wu, Y. He, X. Luo, and S. Deng, “Highly-accurate electricity load estimation via knowledge aggregation,” arXiv preprint arXiv:2212.13913, Dec. 2022.

[12] T. G. Grandón, J. Schwenzer, T. Steens, and J. Breuing, “Electricity demand forecasting with hybrid statistical and machine learning algorithms: Case study of Ukraine,” arXiv preprint arXiv:2304.05174, Apr. 2023.

[13] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, “A hybrid neural network model for power demand forecasting,” *Energies*, vol. 12, no. 5, p. 931, 2019, doi: 10.3390/en12050931.

[14] J. Lee and Y. Cho, “National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model?” arXiv preprint arXiv:2107.06174, July 2021.