Short-term Power Load Forecasting in the Tennessee Valley Region Using XGBoost and Neural Networks

Abby Roberts, Stiven Lavrenov
Middle Tennessee State University
DATA 6990

***Abstract*—Accurate short-term electricity load forecasting is essential for efficient power grid operation, economic dispatching, and demand-side planning. While deep learning models dominate recent load forecasting literature, most empirical studies rely on private datasets in non-U.S. regions, limiting reproducibility. This research presents a comparative study of two forecasting approaches—XGBoost and a Multi-layer Perceptron Neural Network (MLP)—applied to hourly demand data from the Tennessee Valley Authority (TVA). Using two years of publicly sourced power and weather observations, we evaluate model performance using RMSE, MAPE, and R^2. Results demonstrate strong predictive capabilities for both approaches, with XGBoost achieving outstanding accuracy (MAPE = 0.0079, RMSE = 203.61 MW, R^2 = 0.9971), outperforming the MLP model (MAPE = 0.0222, RMSE = 541.85 MW, R^2 = 0.9788). This work provides a reproducible benchmark for U.S.-based short-term electricity forecasting using public data and an open codebase.**

***Index Terms*—Short-term load forecasting, XGBoost, neural network, multi-layer perceptron, TVA, power systems, time series, machine learning.**

I. INTRODUCTION

Short-term power load forecasting is a critical component of modern energy management. Utilities rely on hourly demand prediction to determine generation schedules, regulate power markets, and prevent grid instability. Even small forecasting errors may result in significant economic costs or reliability challenges. Increasing weather-driven variability and electrification of loads demand adaptable forecasting systems, motivating the integration of machine learning techniques.

Neural network architectures (e.g., LSTM, CNN-LSTM) have shown strong forecasting capabilities but often require extensive computing resources and careful tuning. Ensemble tree-based methods, such as XGBoost, frequently excel on structured tabular data and offer faster training and easier deployment. However, comparative empirical studies on publicly available U.S. grid data remain limited. This study compares XGBoost and a feed-forward multi-layer perceptron (MLP) neural network architecture for hourly TVA load forecasting using openly accessible power and weather data.

II. RELATED WORK

A growing body of literature examines machine learning for load forecasting. Jia et al. [3] and Jia et al. [4] demonstrate the value of deep learning and hybrid preprocessing for short-term load prediction but rely on private datasets. Wu et al. [8] shows that XGBoost often outperforms neural networks on tabular datasets, and Cai et al. [9] applies XGBoost to user-level load forecasting. Chen et al. [1] [2] discusses broader ML opportunities in power systems but lacks reproducible empirical results. The only example of power load forecasting done in the TVA region was completed in 1962, a date period that will have little to no significance on today's power demands and was done to perform a 15-year forecast rather than short-term load forecasting [11]. This work addresses the identified gap by benchmarking XGBoost and neural network performance using public TVA-region data.

III. DATA AND PREPROCESSING

A. Data Sources

We used two primary public data sources: (1) hourly TVA regional electricity demand obtained from the U.S. Energy Information Administration (EIA) Open Data API [12], and (2) hourly weather observations (temperature, relative humidity, precipitation, wind speed, cloud cover, etc.) from Open-Meteo [13] for representative TVA cities (Nashville, Memphis, Knoxville). The study covers 2023–2024, producing approximately 17,500 hourly observations.

B. Preprocessing and Feature Engineering

Raw JSON responses were parsed into Pandas DataFrames, merged on timestamp, and transformed. Key preprocessing steps included:

- Time decomposition: hour of day, day of week, month, season, and day/night indicator.
- Lag features: load(t-1) and load(t-24) are included to capture autocorrelation.
- Weather features are aggregated and selected from station-level data.
- Outlier and impossible-value checks (negative demand and humidity removal).
- Normalization and scaling where appropriate for MLP training.

No missing-value imputation was required for the final merged dataset.

IV. Models

A. XGBoost Regressor

The XGBoost model used the `XGBRegressor` module with GridSearchCV for hyperparameter tuning. Tuned parameters included `n_estimators`, `max_depth`, `learning_rate`, `subsample`, and `colsample_bytree`.

B. Multi-layer Perceptron

The MLP used a feed-forward architecture with multiple dense layers and dropout regularization. The network was trained with the Adam optimizer minimizing MSE loss. The model was tuned using the Optuna, a hyperparameter optimization framework, by tuning the learning rate, batch size, dropout rate, and number of layers. All categorical features were encoded appropriately, and continuous inputs were scaled using standardization.

V. Evaluation Metrics

We adopt common regression metrics used in the literature:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2}$$

We also report the coefficient of determination R^2 for the model's explanatory power.

VI. Results

Table II presents test-set performance for both models.

TABLE I: Test performance comparison (lower is better for RMSE/MAPE).

Model	RMSE (MW)	MAPE	R^2
XGBoost	203.61	0.00788	0.99710
MLP	541.85	0.02221	0.97880

XGBoost produced significantly lower RMSE and MAPE compared to the MLP. XGBoost's MAPE of 0.788% indicates sub-1% average relative error on hourly demand predictions. The MLP, while still high-performing with R^2 near 0.98, had a roughly three-fold higher RMSE and an approximately three times larger MAPE.

A. Feature Importance

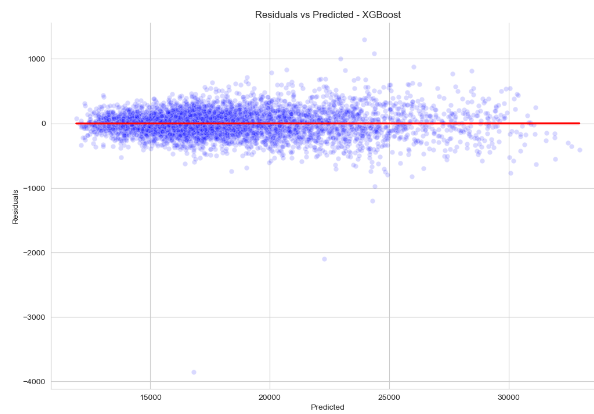
XGBoost's feature importance (gain-based) identifies lagged load features (t-1, t-24), hour-of-day indicators, and temperature as the most influential predictors. A table placeholder is included in Table II.

B. Residuals vs Predicted

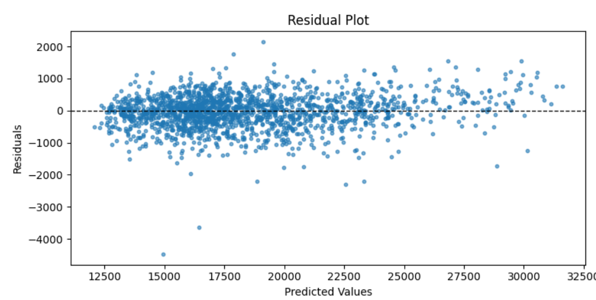
Figure 1 provides an example of a predicted vs. residual comparison for both models. Visual inspection shows that XGBoost more closely tracks rapid intra-day changes.

TABLE II: XGBoost Feature Importance

Features	Values
power_1hr	0.878887
nashville_temperature_2m	0.016130
hour	0.014821
nashville_is_day	0.008410
season_Summer	0.008037
knoxville_is_day	0.007625
memphis_is_day	0.007318
knoxville_apparent_temperature	0.006232
nashville_apparent_temperature	0.006134
power_1day	0.006015



(a) XGBoost



(b) MLP

Fig. 1: Sample predicted vs. residual load.

VII. Discussion

The results suggest that for TVA hourly load forecasting using engineered time-weather tabular data, XGBoost is an efficient and highly accurate method. Its strong performance likely stems from the ability of gradient-boosted trees to capture nonlinear interactions between engineered features without heavy data scaling or sequence modeling.

The MLP, while capable, required more extensive tuning and still produced larger forecast errors. This outcome is consistent with prior observations that deep networks do not always outperform tree ensembles on structured tabular tasks [8].

Limitations include: (1) single-region study (TVA) — results may not directly generalize to regions with different load patterns; (2) the MLP architecture tested was feed-forward (not sequence-specialized like LSTM/Transformer), leaving room to evaluate advanced temporal architectures; (3) additional exogenous variables (e.g., socioeconomic indicators, distributed generation) may improve performance.

VIII. Contributions and Future Work

We provide a reproducible benchmark and an open codebase: https://github.com/abbyawesome/power_load_forecast_data6990. Future directions:

- Evaluate sequence models (LSTM, Transformer) and hybrid architectures.
- Expand geographic coverage and test cross-region generalization.
- Deploy model in near real-time to assess operational viability.

IX. Conclusion

This study compared XGBoost and a deep neural network for short-term hourly load forecasting in the TVA region using open data. XGBoost achieved superior performance (RMSE=203.61 MW, MAPE=0.00788, R^2=0.9971), making it a practical and accurate option for operational forecasting. The study contributes a transparent, reproducible reference for U.S.-based load forecasting using publicly available data.

X. Team Contribution Statement

This research project was completed by two individuals, Abby Roberts and Stiven Lavrenov, with an even distribution of work. Abby Roberts was responsible for creating the problem statement, finding relevant literature, acquiring power load data, developing the XGBoost model, and preparing the presentation for this project. Stiven Lavrenov was responsible for conducting the literature review, acquiring the weather data, developing the MLP neural network, and writing this paper.

References

[1] H. Chen, X. Feng, Q. Chen and R. Han, "The Application of Machine Learning in Power Grid," 2024 Boao New Power System International Forum - Power System and New Energy Technology Innovation Forum (NPSIF), pp. 568-576, 2024.

[2] X. Chen, X. Wang, A. Colacelli, M. Lee and L. Xie, "Electricity Demand and Grid Impacts of AI Data," 2025. [Online]. Available: https://arxiv.org/pdf/2509.07218.

[3] L. Jia, G. Li, Z. Zhang, Y. Wang, Y. Sun and S. Li, "Deep learning-based short-term load forecasting for power grids," 2024 4th International Conference on Energy, Power and Electrical Engineering (EPEE), pp. 188-191, 2024.

[4] W. Jia, J. Chen, S. Luo, Y. Huang and B. Zhang, "Research on Short-Term Power Load Forecasting Method Based on PCA-VMD-LSTM-MTL," 2023 IEEE/IAS Industrial and Commercial Power System Asia (ICPS Asia), pp. 440-444, 2023.

[5] R. Li, J. H. Li and H. M. Li, "The short-term electric load forecasting grid model based on MDRBR algorithm," 2006 IEEE Power Engineering Society General Meeting, p. 6, 2006.

[6] J. Lu, J. Qian, Q. Zhang, S. Liu, F. Xie and H. Xu, "Best practices in China Southern Power Grid competition of AI short-term load forecasting," 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), pp. 1-5, 2020.

[7] C. Xu, Y. Zhang, X. Yang, X. Liu, C. Wang and S. Chen, "Grid Power Load Forecasting Mechanism Based on Neural Network," 2023 IEEE 5th Eurasia Conference on IOT, Communication and Engineering (ECICE), pp. 548-551, 2023.

[8] J. Wu, Y. Li and Y. Ma, "Comparison of XGBoost and the Neural Network model on the class-balanced datasets," 2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC), pp. 457-461, 2021.

[9] J. Cai, H. Cai, Y. Cai, L. Wu and Y. Shen, "Short-term Forecasting of User Power Load in China Based on XGBoost," 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), pp. 1-5, 2020.

[10] L. Pan, P. Li, Y. Fu, J. Li, J. Lv and J. Yang, "An Ultra-short-term Power Load Forecasting Method Based on XGBoost and Fusion Feature Optimization," 2024 6th International Conference on Energy Systems and Electrical Power (ICESEP), pp. 537-540, 2024.

[11] W. R. New, "Load Forecasting on the TVA System Par I-Substation Loads," in Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems, vol. 81, no. 3, pp. 101-105, April 1962, doi: 10.1109/AIEEPAS.1962.4501275. keywords: Load forecasting;Weather forecasting;Substations;Marketing and sales;Economic forecasting;Power generation economics;Cultural differences;Business;Power system planning;Power system economics

[12] https://www.eia.gov/opendata/

[13] https://open-meteo.com/en/docs

A. Acknowledgement of AI Use

I acknowledge the use of ChatGPT to improve the organization and tone of this report. The use of AI was solely used to improve the verbiage and areas where this report's language could be strengthened.