Machine Learning-Based Next-Day National Electricity Demand Forecasting in Bangladesh: A Comparative Study of Random Forest, XGBoost, and LightGBM

1st Wakil Mahmud Rahat
Dept. of CSE
American International University-Bangladesh
Dhaka, Bangladesh
23-51954-2@student.aiub.edu

2nd Md. Thalha Jobayer
Dept. of CSE
American International University-Bangladesh
Dhaka, Bangladesh
22-49261-3@student.aiub.edu

3rd Nowshin Akter
Dept. of CSE
American International University-Bangladesh
Dhaka, Bangladesh
22-48259-2@student.aiub.edu

***Abstract*—Accurate short-term electricity demand forecasting is important for grid planning in Bangladesh, where national demand exhibits strong day-to-day variability. A national-level, machine-learning-only comparison of lightweight tree-based ensemble models for this task was not identified in our review of existing Bangladesh-focused forecasting literature, which is concentrated on deep learning approaches and sub-national data. This study addresses next-day national demand forecasting using a real-world Bangladesh Power Development Board-derived dataset (2019–2024), subjected to a documented, leakage-free, non-interpolative cleaning protocol including calendar-indexed lag construction and explicit exclusion of identified data-extraction errors. Random Forest, XGBoost, and LightGBM are compared against a naive persistence baseline using seven engineered features (three demand lags, a rolling mean, day-of-week, month, and a holiday indicator) under a chronological train/validation/test split, with hyperparameters selected on validation data only. On an untouched 2024 test set, XGBoost obtained the lowest mean absolute error (MAE) at 671.57 MW and a mean absolute percentage error (MAPE) of 5.65%, compared to the naive baseline's MAE of 749.24 MW and MAPE of 6.58%; however, Diebold-Mariano tests found no statistically significant difference between any pair of models, including each machine-learning model versus the baseline, at $\alpha = 0.05$. Across all three models, the most recent day's demand consistently dominated feature importance. This work contributes a reproducible data-cleaning protocol for this dataset alongside an empirical, significance-aware comparative evaluation of tree-based ensemble methods for Bangladesh electricity demand forecasting.**

***Index Terms*—electricity demand forecasting, Bangladesh, Random Forest, XGBoost, LightGBM, time series, data cleaning**

I. Introduction

A. Background and Motivation

Reliable short-term electricity demand forecasting supports grid operations, generation scheduling, and supply-demand balancing in any national power system. In Bangladesh, electricity demand forecasting has been addressed using both classical statistical models and machine learning approaches, applied at national, divisional, and city levels [1]–[5]. Prior work spans autoregressive statistical models with exogenous variables [1], comparisons of statistical and deep learning methods [2], feature-engineered hybrid pipelines [3], deep learning architectures applied to national and divisional data [4], and a deep learning hybrid architecture applied at city level [5]. This body of work indicates that demand forecasting remains an actively studied problem for the Bangladesh power system, motivating continued empirical evaluation of forecasting approaches at the national level.

B. Problem Statement

This study addresses **1-day-ahead forecasting of national electricity demand at daily frequency**, using a single aggregate national demand series rather than divisional or zonal data. Restricting the scope to national-level, single-horizon forecasting with a small set of lightweight, interpretable tree-based ensemble models – Random Forest, XGBoost, and LightGBM – yields a bounded, practical comparative study rather than an open-ended architecture search. This scope deliberately excludes deep learning, classical statistical models, and sub-national forecasting, which are addressed elsewhere in the literature discussed in Section II.

C. Research Questions and Objectives

This study is guided by the following research questions:

- **RQ1:** How accurately can lightweight tree-based ML models (Random Forest, XGBoost, LightGBM) forecast

next-day national electricity demand in Bangladesh, relative to a naive persistence baseline?
- **RQ2:** Among these three models, which achieves the lowest test-set error, and is this difference statistically distinguishable (via the Diebold-Mariano test) from the others or from the baseline?
- **RQ3:** Which engineered features (lag, rolling, calendar, holiday) contribute most to prediction accuracy in this setting?

Corresponding objectives: (1) construct a cleaned, leakage-free, calendar-continuous national daily demand series from a real, imperfect BPDB-derived dataset, with all cleaning steps disclosed and non-interpolative; (2) implement and validation-tune Random Forest, XGBoost, and LightGBM for 1-day-ahead national demand forecasting against a naive persistence baseline; (3) evaluate all four methods once on an untouched 2024 test set using MAE, RMSE, MAPE, and Diebold-Mariano significance testing; (4) quantify each model's built-in feature importance to identify which predictors drive next-day demand.

D. Contributions

This paper makes the following contributions: (1) a documented, reproducible cleaning protocol for a real-world, imperfect national electricity dataset, including explicit handling of missing calendar dates and identified data-extraction errors; (2) an empirical, national-level comparison of three tree-based ensemble models against a naive baseline for Bangladesh electricity demand – based on our review of the literature discussed in Section II, a focused, national-level, machine-learning-only comparison of this specific model combination was not identified in our review; (3) a statistically grounded evaluation that distinguishes lowest numerical test error from statistically significant improvement, using the Diebold-Mariano test rather than ranking alone; (4) a feature-importance analysis identifying next-day demand's primary predictive driver, evaluated consistently across three independent model families; (5) transparent documentation and retention of real, extreme observations in the evaluation data, illustrating a disclosure-first approach to data integrity in applied forecasting work.

E. Paper Organization

The remainder of this paper is organized as follows. Section II reviews related work on Bangladesh electricity demand forecasting and positions this study relative to it. Section III describes the dataset and the data-cleaning protocol applied. Section IV defines the forecasting problem and modeling approach. Section V details the experimental setup. Section VI presents results, Section VII discusses their implications, Section VIII states limitations, and Section IX concludes the paper.

II. Related Work

A. General Short-Term Load Forecasting

Short-term electricity load forecasting has historically relied on classical statistical time-series models such as ARIMA and its extensions, which offer interpretability but assume linear demand dynamics. Machine learning approaches, including tree-based ensembles, have since been applied to capture nonlinear relationships between demand and calendar, weather, or historical features. More recently, deep learning architectures (e.g., LSTM, GRU) have been used to model sequential dependencies in demand data. This progression is reflected in the Bangladesh-specific literature reviewed below.

B. Bangladesh-Specific Studies

Rabbi et al. [1] applied an Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model to yearly national electricity demand in Bangladesh (1994–2018), using population and GDP per capita as exogenous variables. Piyal et al. [2] compared Long Short-Term Memory (LSTM), Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX), and Facebook Prophet for demand forecasting using Power Grid Company of Bangladesh (PGCB) data, reporting that LSTM outperformed the statistical baselines. A subsequent study [3] proposed a feature-engineered pipeline incorporating weather and economic indicators alongside FB-Prophet and LSTM for hourly PGCB demand data, achieving improved accuracy over earlier Bangladesh studies. A separate study [4] compared multiple deep learning architectures (LSTM, Gated Recurrent Unit (GRU), Bidirectional Long Short-Term Memory (BiLSTM), BiGRU) for long-term demand forecasting using web-scraped BPDB data at national and divisional daily granularity. At the city level, [5] proposed a hybrid CNN with stacked BiLSTM architecture for short-term demand forecasting in Dhaka, comparing it against LSTM, CNN-LSTM, and CNN-BiLSTM benchmarks using six years (2016–2021) of Dhaka-zone daily consumption data, and found the proposed CNN-stacked-BiLSTM model achieved the lowest error among these deep learning variants. Table I summarizes these studies.

TABLE I
Summary of Reviewed Bangladesh Electricity Demand Forecasting Studies

Study	Year	Scope	Models	Granularity
[1]	2020	National	ARIMAX	Yearly
[2]	2023	Not reported	LSTM, SARIMAX, FB-Prophet	Not reported
[3]	2025	Not reported (PGCB)	FB-Prophet, LSTM	Hourly
[4]	2025	National + Divisional	LSTM, GRU, BiLSTM, BiGRU	Daily
[5]	2024	City (Dhaka)	CNN-Stacked-BiLSTM (proposed), LSTM, CNN-LSTM, CNN-BiLSTM	Daily

C. Positioning of This Study

Across the studies reviewed, Bangladesh electricity demand forecasting is addressed predominantly through deep learning architectures, frequently applied at sub-national (city or divisional) granularity, with classical statistical models serving

as the typical baseline comparator. None of the reviewed studies employ tree-based ensemble models (Random Forest, XGBoost, or LightGBM); the closest overlap in scope is a city-level deep learning hybrid architecture, as in [5], rather than a focused national-level tree-based ensemble comparison. A national-level, machine-learning-only comparison of Random Forest, XGBoost, and LightGBM at daily frequency, without a deep learning or classical statistical comparator, **was not identified in our review** of the literature summarized above.

III. Dataset and Data Preprocessing

A. Data Source

This study uses the "Structured Dataset of Daily Electricity Demand, Generation, Load Shedding, and Supply Constraints in Bangladesh (2019–2024)," published on Mendeley Data (DOI: 10.17632/x7r7wdb39k.1) [6]. The dataset is derived from daily reports published by the Bangladesh Power Development Board (BPDB), spanning November 21, 2019 to December 30, 2024, and is released in five progressively cleaned versions. This study uses Version 3, which adds documented holiday annotations to the underlying demand, generation, and supply-constraint fields without altering their values relative to Version 2. Only the national-level series is used; divisional/zonal fields present in the dataset are not used as predictors.

B. Target Variable

The forecasting target is "Max. Demand at eve. peak (Generation end)", reported in megawatts (MW). This field represents BPDB's own headline national evening-peak demand figure, measured at the point of generation. It is preferred over the dataset's parallel "Sub-station end" measurement, which additionally reflects transmission losses between generation and delivery, and over divisional demand fields, which represent sub-national aggregates rather than the national total.

C. Data Integrity

Five integrity issues were identified in the raw dataset and resolved prior to modeling.

Duplicate dates. No duplicate date entries were found in Version 3; this was verified programmatically as a hard-stop check prior to feature construction.

Missing calendar dates. Seventeen calendar dates, forming eleven contiguous gap events, are absent from the dataset. Of these, the six-day gap spanning July 18–23, 2024 began during a nationwide internet blackout in Bangladesh that started on July 18, 2024 amid the quota-reform protests [7]; the cited source documents the blackout's onset and continuation but does not establish its exact end date relative to this gap, and no claim is made here that the blackout explains every day within the gap. The causes of the remaining ten gap events are not documented and are not inferred here.

Zero-target extraction errors. Thirty dates in the raw target column contain a literal value of zero. Cross-field inspection shows that all other fields on these rows, including the closely related Sub-station-end demand figure, contain realistic, non-zero values, indicating a text-extraction failure specific to this one field rather than genuine zero demand. These 30 dates were excluded from the target series; they were not recovered, and the corresponding imputed replacement values present in dataset Versions 4 and 5 were not used at any stage of this study.

Retained observation (May 27, 2024). One date, May 27, 2024, records a target value of 4,554 MW, substantially below surrounding days. This value is supported by internal consistency: the sum of the dataset's nine divisional demand fields for this date equals the national Sub-station-end total exactly, and the ratio between Generation-end and Sub-station-end demand on this date falls within the normal range observed across the full dataset. This date also coincides with Cyclone Remal, which made landfall in Bangladesh on May 26, 2024 and caused widespread, multi-day power outages [8]. This observation is treated as a genuine, externally corroborated extreme event and is retained unmodified in the dataset; it is not classified as a data error.

Retained observation (October 24, 2022). One additional date, October 24, 2022, records a target value of 5,652 MW, likewise substantially below surrounding days. As with the May 27, 2024 observation, this value is internally consistent: the sum of the nine divisional demand fields equals the national Sub-station-end total exactly, and the ratio between Generation-end and Sub-station-end demand falls within the normal range observed across the full dataset. Unlike the May 27, 2024 observation, no independently verified external event was identified for this specific date. This observation is retained without modification, and no cause is inferred.

D. Calendar-Indexed Feature Construction

Lag and rolling-window features require a continuous daily calendar to be well-defined; computing them from row position rather than calendar date silently produces incorrect values whenever rows are missing, as is the case here. To avoid this, the cleaned target series was reindexed onto a complete daily calendar spanning the dataset's full date range. This reindexing step expands the 1,850 originally recorded rows to **1,867 calendar-day positions**, by inserting the 17 missing calendar dates as explicit missing-value positions rather than adding new observations; no data is created at this step. Lag and rolling features were then computed on this 1,867-row calendar-indexed series, ensuring that each lag corresponds to an exact number of calendar days rather than an exact number of available rows. No missing value was interpolated, filled, or estimated at any stage. As a consequence, any candidate row whose target or required lag/rolling input fell on a missing or excluded date was removed: **203 rows were dropped** from the 1,867-row calendar-indexed series, yielding **1,664 usable rows** for modeling ($1{,}867 - 203 = 1{,}664$).

E. Holiday Feature

The holiday indicator feature was derived from the dataset's documented `Holiday_cat` field, with values 1 (single-day holiday) and 2 (multi-day holiday) mapped to a posi-

tive indicator. Per the dataset's own column documentation, `Holiday_cat = 3` denotes Ramadan as a distinct "special occasion" category rather than a holiday, and is therefore excluded from the indicator. Three rows (July 1–3, 2020) contain a documented inconsistency between `Holiday_cat` (multi-day holiday) and the free-text holiday name field ("No Holiday"); for these rows, `Holiday_cat` was treated as authoritative, and this inconsistency is disclosed here rather than silently resolved.

F. Final Dataset Summary

Table II summarizes dataset size at each cleaning stage described above.

TABLE II
DATASET SIZE BEFORE AND AFTER CLEANING

Stage	Rows
Original (Version 3)	1,850
Calendar-indexed (17 missing-date positions added)	1,867
Cleaned/usable (post feature construction)	1,664
Train (2019–2022)	1,113
Validation (2023)	297
Test (2024)	254

Fig. 1 shows the cleaned national demand series over the full study period, with the retained May 27, 2024 observation marked. The October 24, 2022 observation discussed above is not separately annotated in the figure.

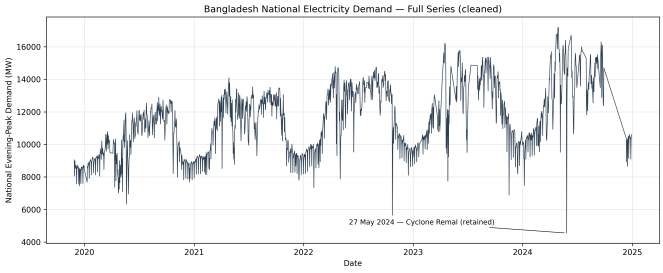


Fig. 1. Bangladesh national electricity demand, full cleaned series (2019–2024), with the retained May 27, 2024 observation annotated.

IV. METHODOLOGY

A. Forecasting Problem

This study addresses a one-step-ahead forecasting problem: given information available up to and including day t, predict national electricity demand on day $t+1$. Formally, let y_t denote observed demand on day t and X_t denote a feature vector constructed from information available by the end of day t or calendar facts about day $t+1$ known in advance. The task is to estimate $\hat{y}_{t+1} = f(X_t)$. Each row in the modeling dataset corresponds to a target day $t+1$; its associated features are constructed exclusively from day t and earlier, ensuring no information from the target day itself enters the feature vector.

B. Features

Seven features were used, constructed on a calendar-indexed daily series (Section III-D) to guarantee correct temporal alignment: `demand_lag1`, `demand_lag2`, `demand_lag7` (demand on days t, $t-1$, and $t-6$ respectively, capturing short-term persistence and weekly seasonality); `rolling7_mean` (the mean of demand over the seven calendar days ending on day t); `day_of_week`, `month` (calendar attributes of the target day $t+1$, safely usable because calendar dates are known in advance); and `holiday_indicator` (a binary indicator of holiday status on the target day $t+1$, derived from the documented holiday category field). No feature derived from same-day (t+1) demand, generation, or supply-constraint data is used.

C. Baseline

A naive persistence baseline is defined as $\hat{y}_{t+1} = y_t$. This baseline requires no fitting. Persistence is a demanding baseline because daily electricity demand is strongly autocorrelated.

D. Models

Three tree-based ensemble models were evaluated: Random Forest, XGBoost, and LightGBM, selected for their established use in tabular regression, native feature importance support, and comparatively low computational cost relative to deep learning. No deep learning or classical statistical time-series models are included.

E. Metrics

Model performance is evaluated using Mean Absolute Error (MAE), root mean square error (RMSE), and Mean Absolute Percentage Error (MAPE):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

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

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3}$$

MAPE is well-defined throughout this study because the retained target series contains no zero or near-zero demand values.

F. Statistical Testing

The Diebold-Mariano test [9] is applied to pairs of forecast error series using squared-error loss. For two models with errors $e_{1,t}$ and $e_{2,t}$, the loss differential is $d_t = e_{1,t}^2 - e_{2,t}^2$; a positive mean loss differential indicates Model 1 has higher average loss than Model 2. The Harvey-Leybourne-Newbold small-sample correction [10] is applied, with significance assessed at $\alpha = 0.05$.

V. Experimental Setup

A. Chronological Split

The cleaned dataset was partitioned chronologically into training (2019–2022, $n = 1{,}113$), validation (2023, $n = 297$), and test (2024, $n = 254$) periods. No shuffling was applied. The test set remained untouched until the single final evaluation in Section VI.

B. Hyperparameter Tuning

Each candidate model was tuned via grid search, fit on training data only and evaluated on validation data only, using validation MAE as the sole selection criterion. The test set was not accessed during tuning. Table III lists the search space for each model.

TABLE III
Hyperparameter Search Space

Model	Parameter	Values
Random Forest	n_estimators	{100, 300, 500}
	max_depth	{None, 10, 20}
XGBoost	n_estimators	{100, 300}
	max_depth	{3, 6, 9}
	learning_rate	{0.05, 0.1}
LightGBM	n_estimators	{100, 300}
	num_leaves	{31, 63}
	learning_rate	{0.05, 0.1}

C. Final Training

For each model, the configuration achieving the lowest validation MAE was retrained on combined training and validation data (2019–2023) and evaluated exactly once on the untouched 2024 test set.

D. Reproducibility

A fixed random seed (42) was used for all stochastic components across all three models. The experiment was executed using the software environment recorded at the time of the run; exact library version numbers are not independently retained as a saved artifact and are therefore not reported here. No feature scaling or normalization was applied, as all three candidate models are tree-based ensembles invariant to monotonic feature transformations.

VI. Results

A. Selected Hyperparameters

Table IV reports the selected configuration and corresponding validation MAE for each model.

B. Test-Set Forecasting Performance

Table V reports test-set performance for all four methods, evaluated once on the untouched 2024 test set ($n = 254$).

XGBoost obtained the lowest numerical MAE, RMSE, and MAPE among the four evaluated methods. Whether these numerical differences reflect a statistically meaningful improvement is addressed next.

TABLE IV
Selected Hyperparameters and Validation Performance

Model	Selected Configuration	Val. MAE (MW)
Random Forest	n_estimators=100, max_depth=None	662.99
XGBoost	n_estimators=300, max_depth=3, lr=0.05	599.52
LightGBM	n_estimators=300, num_leaves=31, lr=0.05	636.48

TABLE V
Test-Set Forecasting Performance

Model	MAE (MW)	RMSE (MW)	MAPE (%)
XGBoost	671.57	1060.44	5.647
LightGBM	684.91	1066.95	5.775
Random Forest	695.21	1070.61	5.847
Naive Persistence	749.24	1230.25	6.580

C. Statistical Comparison

Table VI reports Diebold-Mariano test results ($\alpha = 0.05$).

No comparison reached statistical significance at $\alpha = 0.05$. This does not establish that the models perform identically, nor does it rule out a real difference that this sample size was insufficient to detect.

D. Feature Importance

Table VII reports each model's built-in feature importance (%).

`demand_lag1` was the highest-ranked feature in all three models. Across all three, `holiday_indicator` consistently received the lowest or near-lowest importance ranking in this dataset.

E. Summary of Figures

Fig. 2 presents actual versus predicted next-day national demand for all four methods over the 2024 test period. Figs. 3–

TABLE VI
Diebold-Mariano Test Results

Comparison	DM Statistic	p-value
Random Forest vs. Naive	-1.408	0.1604
XGBoost vs. Naive	-1.520	0.1298
LightGBM vs. Naive	-1.462	0.1448
XGBoost vs. LightGBM	-0.228	0.8197

TABLE VII
Feature Importance by Model (%)

Feature	RF	XGBoost	LightGBM
demand_lag1	82.03	71.75	21.58
rolling7_mean	6.50	9.82	18.21
demand_lag7	5.04	5.14	20.72
demand_lag2	3.36	4.57	20.04
day_of_week	1.82	4.22	9.00
month	0.95	1.91	9.08
holiday_indicator	0.31	2.59	1.37

5 present the per-model feature importance distributions summarized in Table VII.

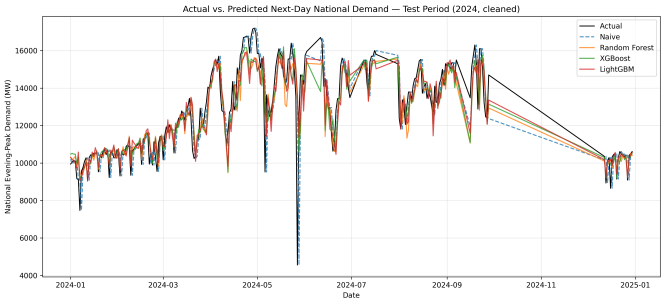


Fig. 2. Actual vs. predicted next-day national demand, 2024 test period.

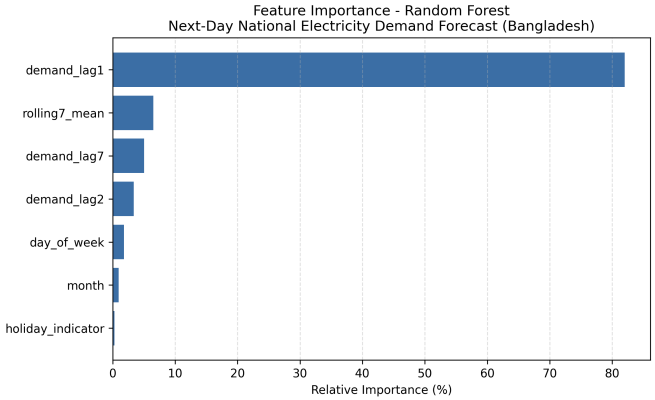


Fig. 3. Feature importance – Random Forest.

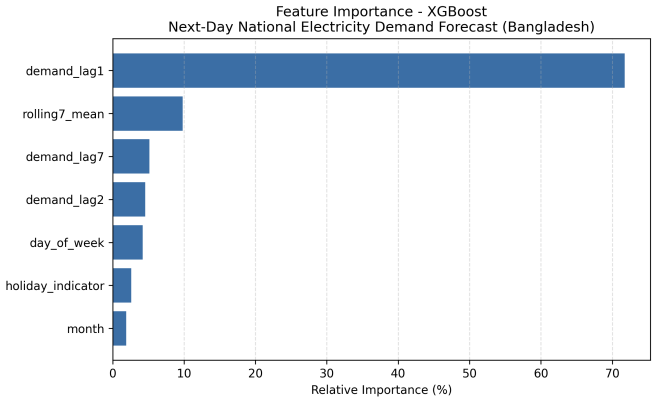


Fig. 4. Feature importance – XGBoost.

VII. Discussion

The results in Section VI indicate an observed ranking among the four evaluated methods, with XGBoost achieving the lowest MAE, RMSE, and MAPE, followed by LightGBM, Random Forest, and naive persistence. This ordering reflects the point estimates obtained on this particular test period rather than evidence that XGBoost is inherently superior for this task.

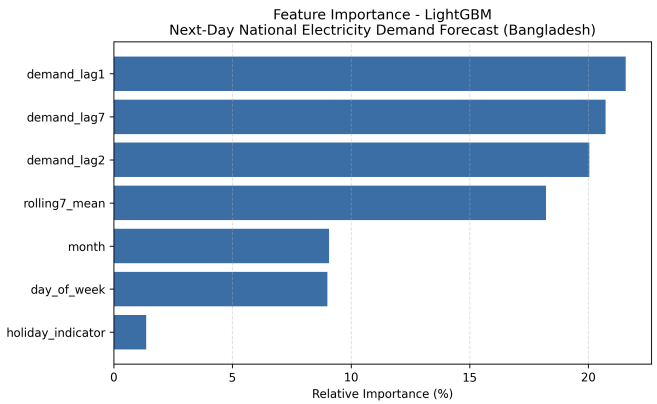


Fig. 5. Feature importance – LightGBM.

This distinction matters because the Diebold-Mariano tests found no statistically significant difference at $\alpha = 0.05$ for any evaluated pair. A numerically lower error and a statistically distinguishable improvement are separate claims: the former was observed, the latter was not established.

The consistent dominance of `demand_lag1` across all three models suggests that, in this national daily forecasting setting, the most recent day's demand carries the majority of predictive signal available to these models, consistent with the strength of the naive persistence baseline. The low ranking of `holiday_indicator` across all three models is an observation specific to this dataset, feature set, and time period.

These findings relate to the literature in Section II. Prior Bangladesh-focused studies [1]–[5] predominantly rely on deep learning architectures or classical statistical baselines, frequently at sub-national granularity, with none of the reviewed studies employing tree-based ensemble models. The present study's national-level, machine-learning-only comparison was not identified in our review of this literature.

With respect to the research questions: RQ1 is addressed by Table V, showing all three ML models achieving lower point-estimate error than naive persistence. RQ2 is addressed by the combination of Tables V and VI: XGBoost achieved the lowest numerical error, but this and all other pairwise differences were not statistically significant at $\alpha = 0.05$. RQ3 is addressed by Table VII: `demand_lag1` was the dominant feature across all three models.

VIII. Limitations

This study has several limitations. The dataset is derived from BPDB source records covering 2019–2024, which contained documented gaps and extraction errors, handled through explicit exclusion rather than interpolation, reducing usable observations. Final evaluation was conducted on a single test period (2024 only). The forecasting task was restricted to one-day-ahead national daily demand; intraday, multi-day-ahead, regional, and city-level forecasting were not evaluated. The model comparison was limited to naive persistence, Random Forest, XGBoost, and LightGBM; deep learning and classical

statistical models were outside scope. The feature set was restricted to seven lag, rolling, and calendar-based predictors; external variables such as weather or economic indicators were not included. The Diebold-Mariano tests found no statistically significant difference among models at $\alpha = 0.05$, which should not be interpreted as evidence of equivalence.

IX. Conclusion

This study evaluated next-day national electricity demand forecasting in Bangladesh using Random Forest, XGBoost, and LightGBM, compared against a naive persistence baseline. On the untouched 2024 test set, XGBoost achieved the lowest numerical MAE, RMSE, and MAPE, and all three machine learning models produced lower point-estimate error than naive persistence; however, Diebold-Mariano tests found no statistically significant difference at $\alpha = 0.05$ for any evaluated pair. Across all three models, `demand_lag1` was consistently the dominant predictive feature. This work contributes a documented, leakage-free, non-interpolative data-cleaning protocol together with a significance-aware empirical comparison of lightweight tree-based ensemble models for national daily forecasting – an evaluation setting not identified in our review of the existing Bangladesh-focused literature. These results reflect a single dataset, evaluation period, and experimental configuration as described in this paper, and do not establish general model superiority or deployment readiness.

References

[1] F. Rabbi, S. U. Tareq, Md. M. Islam, M. A. Chowdhury, and M. A. Kashem, "A Multivariate Time Series Approach for Forecasting of Electricity Demand in Bangladesh Using ARIMAX Model," *2020 2nd Int. Conf. on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, Bangladesh, 2020, pp. 418–422, doi: 10.1109/STI50764.2020.9350326.

[2] A. P. Piyal, S. Ahmed, K. F. Rahman, and A. S. M. Mohsin, "Energy Demand Forecasting Using Machine Learning Perspective Bangladesh," *2023 IEEE IAS Global Conf. on Renewable Energy and Hydrogen Technologies (GlobConHT)*, Male, Maldives, 2023, pp. 1–5, doi: 10.1109/GLOBCONHT56829.2023.10087679.

[3] M. T. Islam, S. A. Turja, and A. Habib, "Enhanced power demand forecasting for Bangladesh: using feature engineering associated with environmental and economic impact," *Journal of Data, Information and Management*, vol. 7, no. 1, pp. 1–19, Feb. 2025, doi: 10.1007/s42488-025-00140-9.

[4] S. Yeaser, T. Nafs, and Md. I. Hossain, "Comparative Analysis of Deep Learning Models for Long-Term Electricity Demand Forecasting in Bangladesh Using Web-Scraped Data," *2025 17th Int. Conf. on Electronics, Computers and Artificial Intelligence (ECAI)*, Târgovişte, Romania, 2025, pp. 1–7, doi: 10.1109/ECAI65401.2025.11095479.

[5] K. F. Bin Akhter, S. Mobasshira, S. N. Haque, M. A. K. Hesham, and T. Ahmed, "Short-Term Electricity Demand Forecasting of Dhaka City Using CNN with Stacked BiLSTM," arXiv:2406.06651, Jun. 2024.

[6] Md. I. Hossain, T. Nafs, and S. Yeaser, "Structured Dataset of Daily Electricity Demand, Generation, Load Shedding, and Supply Constraints in Bangladesh (2019–2024)," Mendeley Data, 2025, doi: 10.17632/x7r7wdb39k.1.

[7] "2024 Bangladesh internet blackouts," *Wikipedia*. [Online]. Available: https://en.wikipedia.org/wiki/2024_Bangladesh_internet_blackouts. Accessed: Sep. 10, 2026.

[8] Inter-Cluster Coordination Group (ICCG), Inter Sector Coordination Group (ISCG), and Bashan Char Rohingya Response, "Bangladesh: Cyclone Remal 2024 – Situation Report No. 02 (As of 27 May 2024)," *ReliefWeb*, May 27, 2024. [Online]. Available: https://reliefweb.int/report/bangladesh/bangladesh-cyclone-remal-2024-situation-report-no-02-27-may-2024. Accessed: Sep. 10, 2026.

[9] F. X. Diebold and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, Jul. 1995.

[10] D. Harvey, S. Leybourne, and P. Newbold, "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, vol. 13, no. 2, pp. 281–291, Jun. 1997, doi: 10.1016/S0169-2070(96)00719-4.