Future Technology
Open Access Journal
ISSN 2832-0379

August 2026| Volume 05 | Issue 03 | Pages 252-262
Journal homepage: https://fupubco.com/futech
https://doi.org/10.55670/fpll.futech.5.3.22

Future Publishing LLC

Article

A unified mathematical and computational framework for predictive analysis of complex dynamic systems

Pradeep Kumar H S[1]*, Harsha S[2]

[1]Department of Information Science and Engineering, The National Institute of Engineering, Mysuru, Affiliated to Visvesvaraya Technological University, Belagavi, India
[2]Department of Computer Science and Engineering (AI&ML), RNS Institute of Technology, Bengaluru, Affiliated to Visvesvaraya Technological University, Belagavi, India

ARTICLE INFO

Article history:
Received 08 January 2025
Received in revised form
03 May 2026
Accepted 12 June 2026

Keywords:
Power-system forecasting, Dynamic systems, Hybrid predictive modeling, Residual learning, Load shedding, Vector autoregression

*Corresponding author
Email address:
pradee.nie@gmail.com

DOI: 10.55670/fpll.futech.5.3.22

ABSTRACT

Interdependent operating states, rather than individual load curves, are increasingly required for short-term power system prediction. This study formulates and tests a common mathematical-computational approach for one-step-ahead prediction of demand, generation, load shedding, and imbalance derived from them within a dynamic grid environment. Operational data were archived as hourly data, prepared chronologically, transformed into a multivariate feature space, and split into training, validation, and test sets. The proposed framework is tested against naïve persistence, Ridge regression, Random Forest, XGBoost, and the VAR models separately and is based on a vector autoregressive mathematical core and an XGBoost residual-correction layer. In the results, the model's effectiveness is target-dependent. The hybrid framework consistently performed best for generation and was statistically similar to the advanced models for demand, load shedding, and derived imbalance. The mathematically derived variables, source-composition features, and short-term dynamic indicators were found to have different contribution values for each target in ablation and feature-importance analyses. Stressed load-shedding conditions were also found to have lower predictive accuracy in regime-specific testing. These results show that mathematically constrained residual learning is useful for coherent forecasting of continuous operating states, while sparse stress-related variables necessitate extensions to the model to learn them in an event-aware fashion. The study offers a replicable methodology for predictive analysis of a shorter time horizon in the context of grid operation.

1. Introduction

Electric power systems are becoming data intensive and stressed, coupled infrastructures whose dynamics of demand, availability of generation, dependence on imports, fuel composition, and operations are changing in time. From this broader perspective, short-term forecasting has emerged as a focal analytical process, as dispatch planning, reserve scheduling, congestion management, reliability evaluation and load shedding mitigation all rely on forecasting the system conditions prior to any propagation of imbalances through it. The requirement of forecasting has shifted more time from the aggregate load forecasting to include the system-level, distribution-level and substation-level applications [1]. This shift brings up a larger issue of predictive analysis for dynamically changing energy systems, in which the goal is not just to minimize numerical error but also to capture the relationships between different operating states, in which one state affects the other.

As the requirements to forecast become more complicated, hybrid forecasting approaches have been encouraged. Recent short-term load forecasting research often uses a hybrid of statistical decomposition, feature extraction and machine-learning methods to enhance robustness in nonlinear, time-varying and varying operating conditions [2]. These developments suggest that there is no universal modelling family that can represent all temporal relationships, particularly if there is a complex interplay of technical, behavioral and supply-side relationships that determine the operating states. Hybrid approaches thus indicate the necessity to mix complementary representations instead of by using one approach. An issue, however, is that as hybrid models become more prevalent there are also an increasing number of studies that use algorithms to improve the accuracy, but fewer of these provide an explicit linkage of the hybrid model as a coherent mathematical-computational model explicitly related to the systems behavior.

This worry is now further bolstered by the quick advancement of deep learning. Recent advances on short-term load forecasting have led to the development of recurrent, convolutional, attention-based and transformer-based approaches for better performance, especially in the case of large and high-frequency datasets [3]. However, such techniques could still be problematic due to their lack of interpretability, long tuning times, and unpredictable response to changes in structure. In power systems, however, merely the accuracy of predictions is not enough because the model's impact on the relationship between supply and demand, and stress indicators, is weak. The forecasting framework should, therefore, be able to capture the essential system-level consistency while learning the residual behavior of the system from the data in a nonlinear fashion.

The problem can be particularly manifested in renewable integrated and operationally heterogenous environments. Comparative studies of the performance of machine learning techniques for short-term net-load forecasting in renewable-integrated microgrids demonstrate that the data structure, resource composition, and the specific target predicted influence the performance of machine learning [4]. Such inconsistency suggests that there is a very important gap in research. While existing studies may focus on individual targets, particularly demand or net load, they do not consider the interdependence of states as simultaneously being looked at by the practical operation of the grid, such as demand, generation, load shedding, and imbalance. The models that work well under normal operating conditions might not be as effective during a stress event. Thus, the assessment of forecasting frameworks must not only consider the general level of accuracy, but the behaviour over a range of continuous and stress-related outputs which are targeted.

This need is further strengthened by new research on interpretability. The usefulness of a model is found to depend on the quality of predictions, as well as on how the model can be explained in situations of extreme or limited data, including for small-sample heatwaves [5]. Wider discussions of AI-powered energy analytics additionally reveal that load forecasting, anomaly detection, and demand response are evolving into a single task, instead of separate analytic issues [6]. The shifts indicate that the future studies on forecasting should shift from the accuracy comparison of a single target to a framework that can capture the interdependence of operations, temporal memory, the varying composition of supply, and stress-sensitive behavior. In multi-state power-system prediction, however, it has not yet been sufficiently developed to be mathematically structured and clear enough to be included in the combination of the two.

More recently it has been shown that meaningful temporal patterns can be extracted to boost the predictive capabilities while also providing better explanation of load behaviour through interpretable multi-scale temporal decomposition [7]. Transformer-based temporal fusion models for forecasting power-stations are also shown to be competitive and interpretable, as they are able to capture temporal salience and variable contribution [8]. Meanwhile, explainability in electric load forecasting has been reviewed, and it is still unclear what is meant by explainability, how it can be measured, and how it can be linked to model structure [9]. The fragmentation is a methodological matter as some advanced forecasting studies are model-driven (not framework-driven). The developers typically demonstrate the performance of a specific algorithm but less detail on the relationship between the mathematical structure, engineered features and residual correction in the whole forecasting process.

In this study, this deficiency is overcome by developing a single, mathematical, computational tool for predictive complex power-system state analysis. The study concentrates on short-term forecasting (one-step-ahead) and is based on operational data of the grid, which is a hybrid model consisting of a core VEC and a residual-correction layer using machine learning. Its selection of demand, generation and load shedding are chosen to be consumption requirement, supply response and stress intervention, respectively. Imbalance is not learned as an output but rather it is derived from predicted demand and generation, in order to create mathematical consistency. Other stability indicators, such as frequency, voltage, congestion, reserve violations or outage records are not modeled because they are not available in the dataset analyzed in this study.

The study has methodological and practical importance. Methodologically, it creates a hybrid forecasting method that is a structured approach instead of an ad hoc mix of methods. It is used practically to assess grid-relevant outputs not only in terms of demands, but also to analyse the behaviour of the models with respect to other continuous operational variables and with respect to other stress-related variables which are only sparsely represented. New ultra-short term forecasting studies have demonstrated that the temporal properties of the systems, as well as their recent motion, can have a significant impact on the predictability of the system [10]. This will help support the use of lagged variables, ramp rates, rolling statistics, and temporal indicators in the residual-correction layer. The study also considers the outputs of the scenarios as predictive sensitivity results from the learned model space, rather than causal grid simulations.

Accordingly, the current paper aims at achieving three goals: developing a unified mathematical-computational forecasting model that guarantees consistency between interconnected power-system states; comparison of forecasting capabilities of the proposed model with naïve, linear, nonlinear, and mathematical control models; evaluating model contribution, regime sensitivity and residual-learning behavior by ablation analysis, event-specific testing, and scenario-based predictive sensitivity. The study also introduces a repeatable framework to perform short-term predictive study of the dynamics of the power system operation without claiming a superiority of one model over all targets. This framing also helps ensure that the methodological claims are brought into line with the end product, the target-specific results, of the revised computational workflow, while not explicitly overstating novelty.

2. Literature review

In recent years, short-term load forecasting has been a field that has been subject of many researchers' studies, and a recent literature often highlights the need for hybridization as a remedy in nonlinear and nonstationary energy environments. Recent studies do not consider load behavior as a cue that can be mastered by just one family of algorithms; instead, a mixture of boosting, deep learning, feature extraction, decomposition, and optimization is used to enhance the predictive robustness. For instance, hybrid CatBoost and XGBoost are shown to achieve better short-term load forecasting accuracy if engineered features describe the non-linear nature of the load demand [11]. These improvements are often target-specific and rely on the input feature space design. Likewise, hybrid approaches using feature-extraction statistics and deep-learning pipelines reveal that preprocessing and representation design can play a significant role in forecasting [12]. Empirical evidence of the value of combining predictors has also been found in the optimized hybrid ensemble models used for very short-term load forecasting, which add complexity to the architecture but also diminish the level of interpretability [13].

The other significant line of work is decomposition-based forecasting. These methods assume that observed load series consist of several interacting components such as periodic, stochastic, and transient components. The idea of decomposed representation for improving predictive learning has been demonstrated with an aggregated hybrid modal decomposition of integrated energy systems in Ref [14]. But decomposition-based models are susceptible to the quality of segmentation, reconstruction, and mode-selection rules. Multi-energy forecasting systems based on hierarchical multi-task learning and spatiotemporal attention build on this idea by jointly forecasting the correlated outputs of the various energy products [15]. These methods are effective when there are strong relationships between different types of energy signals, but are not sufficient to completely address the problem of maintaining explicit system-state relationships between outputs of different operational meanings. The same, albeit in terms of signal regularization instead of mathematical consistency among interdependent states of the power system, can be said of EMD-BI-LSTM smart-grid energy management methods [16].

The other set of research is on optimization improved and information-aware deep learning. It is worth noting that by using particle swarm optimization and convolutional neural networks, the accuracy of short-term load forecasting will be enhanced in an automated search of the parameters [17]. But optimization benefits might lead to models that are more sensitive to the calibration data and less interpretable in terms of operations. Fisher-information based decomposition and selection strategies also demonstrate that the quality of forecasting using the information-guided representation can be enhanced [18]. However, these approaches tend to focus on prediction of the target signal, not on maintaining mathematical relationships between the demand, generation, stress and imbalance. Multivariate forecasting systems using feature selection and multi-objective intelligent optimization get closer to Integrated prediction as multiple variables and objectives are taken into account [19]. However, the focus remains primarily on forecasts at the global level, not on a common model that explicitly integrates a structured dynamic model with residual correction.

The work most related to the present study is found in the literature of the error-correction and staged hybrid forecasting. The error correction and hybrid ensemble learning models for industrial load forecasting show that a second-stage model can recover the information that was lost by a first prediction model, particularly if the demand behavior is more complex [20]. The use of multi-energy forecasting based on complementary ensemble empirical mode decomposition and composite reconstruction also demonstrates that layered prediction could enhance the forecasting accuracy of heterogeneous energy signals [21]. Multi-stage ensemble models that utilize decomposition and error factors as well as multi-objective optimization also substantiate the empirical worth of staged correction [22]. But these methods are mostly accuracy-oriented and/or architecture-intensive. They do not usually develop the hybrid model as a mathematically computable entity in which a structured multivariate dynamics core is explicitly made nonlinear by residual learning.

The literature reviewed indicates that in general, the hybrid short-term forecasting, decomposition-based learning, optimization-based modeling and error-correction methods have made significant advances. Meanwhile there are a number of weaknesses. First many studies are devoted primarily to load or net-load forecasting, whereas multiple interdependent states must be addressed for an operational power system analysis. Secondly, hybrid approaches sometimes achieve better accuracy but not always maintain mathematical consistency of related variables. Third, not all targets for stress are necessarily assessed as discrete targets, but as continuous operating variables. Fourth, there is an absence of discussion of statistical significance and risk for overfitting and redundancy in high-dimensional feature construction. The present study aims to fill these gaps by proposing and testing a combined VAR-XGBoost residual-correction approach for demand, generation, load shedding and derived imbalance forecasting.

3. Methodology

3.1 Research design

In this work, a quantitative predictive modeling design was applied to design and test a single mathematical–computational framework for one-step-ahead prediction of power-system operating states. It is a combination of a structured multivariate time-series model and a nonlinear residual correction. A temporal forecasting design was applied to ensure that the observations made at an earlier point in time forecasted the observations made at a later point in time, avoiding temporal leakage that can result from random shuffling of a series of observations. The proposed framework was compared against naïve persistence, Ridge regression, Random forest, XGBoost and a stand-alone vector autoregressive model. No human subject consent or ethical approval was necessary for the use of publicly available archival operational data used in the study.

3.2 Data source and temporal preparation

The data for the empirical analysis was retrieved from the UCI Machine Learning Repository for hourly generation and demand in Bangladesh, respectively, for the Power sector. The empirical data comprised hourly generation and demand data for the Power sector of Bangladesh, obtained from the UCI Machine Learning Repository [23]. The data set contains operational data, such as demand, generation, load shedding, source-wise generation, cross-border imports and operational remarks per hour for the time frame 19 April 2015 to 17 June 2025. Temporal preparation encompassed datetime parsing, chronological sorting, removing duplicate timestamps, handling missing values, correcting invalid values, and spike capping. Extreme peaks in demand, generation and load shedding were capped with the use of the 0.001 and 0.999 empirical quantiles, respectively, and missing data in inactive or newly added source categories were considered to represent structural zeros. A summary of the main preprocessing audit is provided in Table 1. Table 1 presents the number of the hourly observations available after the temporal cleaning, feature building and target filtering in the final modeling dataset. It also demonstrates that the highest level of preprocessing was structural zero assignment for source variables with no activity or observations for a portion of the observation period.

Table 1. Preprocessing audit summary

Item	Value	Treatment
Raw records	92,650	Original dataset
Unique hourly records	92,218	After timestamp deduplication
Duplicate timestamps removed	432	First timestamp retained
Missing text-context entries	85,845	Filled as “normal”
Structural zero assignments	267,208 entries	In inactive or newly introduced source columns
Invalid negative numeric values	0	None detected
Extreme spike-capped entries	464	0.001/0.999 quantile limits
Final usable modeling observations	92,193	After feature and target filtering

In order to illustrate the influence of cleaning on the key operating variables, the before and after descriptive statistics were calculated for demand, generation and load shedding. These statistics are presented in Table 2. Table 2 shows that preprocessing mainly reduced physically implausible extreme values while preserving the central tendency of the main operational variables. The largest correction occurred in generation and load shedding, where extreme maximum values were capped using empirical quantile limits.

3.3 Target construction and system-state definition

The directly learned forecasting targets were next hour demand, generation and load shedding. The choice of these variables was made because they are related to the consumption requirement, supply response, and stress-state intervention. No other Grid stability indicators – such as voltage, frequency, Outages Registers, Congestion, Reserve-Violation Information – were modeled as these data was not available in the dataset. Imbalance was not learned independently, but rather derived from predicted demand and generation to ensure mathematical consistency:

$$\hat{I}_{t+1} = \hat{D}_{t+1} - \hat{G}_{t+1} \tag{1}$$

where $\hat{D}_{t+1}$ is predicted demand, $\hat{G}_{t+1}$ is predicted generation, and $\hat{I}_{t+1}$ is derived imbalance. This construction ensures that imbalance remains mathematically linked to demand and generation, although errors in both components can propagate into the derived imbalance estimate. A continuous stress indicator was also constructed by combining observed load shedding with positive imbalance.

3.4 Feature engineering

Temporal patterns, system memory, operating transitions and source composition were all represented by feature engineering. The variables were generated at 1, 2, 3, 6, 12 and 24 hours and the variables obtained were rolling 3, 6, 12 and 24 hour variables. The selections were justified by ACF/PACF diagnostics revealing a high level of short-term dependence and daily periodicity for demand, generation and load shedding. The final features groups are briefly summarized in Table 3. As seen in Table 3, a majority of the predictors were made up of rolling statistics and lags, indicating a keen temporal memory for power-system operation, at the hourly level. Three VAR forecast meta-features added to the feature matrix, raising it from 322 to 325.

Table 2. Summary statistics before and after preprocessing

Variable	Stage	Mean	Std	Min	Max
Demand	Before preprocessing	8,826.0009	2,772.0308	6.0000	156,050.0000
Demand	After preprocessing	8,819.5167	2,613.0540	3,283.6510	16,286.7450
Generation	Before preprocessing	9,438.8056	212,471.7646	73.0000	64,526,500.0000
Generation	After preprocessing	8,739.6570	2,486.3382	3,425.8680	15,666.0000
Load shedding	Before preprocessing	81.6464	443.5100	0.0000	65,359.0000
Load shedding	After preprocessing	78.7078	278.0580	0.0000	2,556.6980

Twenty-eight near-zero-variance predictors and 519 pairs of highly-correlated features (with absolute correlations above 0.90) were found, as well as 281 pairs with absolute correlations above 0.95 and 79 pairs with absolute correlations above 0.99. Although several engineered variables were mathematically related, their correlations were expected and generalization was addressed by chronological validation, regularization, tree-based baselines, and out-of-sample testing, with the removal of all the correlated variables not automatically performed.

Table 3. Feature-group summary

Feature group	Count	Examples
Base operational variables	3	generation, demand, load shedding
Mathematical dynamic-system features	8	imbalance, reserve margin, ramps
Temporal/calendar features	11	hour, month, weekend, peak hour
Source-composition features	18	thermal total, import share, source diversity
Lag features	90	demand_lag_1, gas_lag_24
Rolling statistical features	192	demand_roll_mean_24, generation_roll_std_12
Total baseline predictors	322	—
VAR forecast meta-features	3	VAR demand, VAR generation, VAR load shedding
Total hybrid predictors	325	—

3.5 Baseline model development

To get a wide range of comparison, five models were implemented. The naïve baseline was the current observed value used as the forecast for the next hour. Ridge was a linear baseline, Random Forest and XGBoost were nonlinear baselines, and VAR was the mathematical base without any baseline. The settings for the model are provided in Table 4.

Table 4. Model specification summary

Model	Main settings
Naïve persistence	Current observation used as next-hour forecast
Ridge regression	Ridge regression with $\alpha = 1.0$
Random Forest	200 trees; max_depth = 20; min_samples_split = 5; min_samples_leaf = 2
XGBoost baseline	300 trees; max_depth = 6; learning_rate = 0.05; subsample = 0.8; colsample_bytree = 0.8
VAR mathematical core	Candidate lags = 1, 2, 3, 6, 12; selected lag = 12
Hybrid residual XGBoost	400 trees; max_depth = 6; learning_rate = 0.05; subsample = 0.8; colsample_bytree = 0.8

The key settings of the main model for reproducibility are presented in Table 4. The parameters used in the hyperparameters were set and validated in the same way as they were previously tested in chronological order and according to the same test structure. Demand or generation data and load shedding data, gas, liquid fuel, coal and total imports were used to estimate the VAR model. The AIC, BIC, and HQIC were computed for each candidate lag and compared as shown in Table 5. This table indicates that the best AIC among all possible lag orders was obtained for lag 12 and thus it was chosen as the VAR mathematical core. Stationarity tests (ADF and KPSS) were also applied for the examination of stationarity, but several level variables exhibited high temporal dependence and nonstationarity, and hence VAR was treated as a short-horizon forecasting model core, not as a long-run structural model of the economy.

3.6 Unified mathematical–computational framework

The proposed framework integrates a VAR mathematical core with an XGBoost residual-correction layer. Let y_t denote the multivariate operating-state vector at time t, and let X_t denote the engineered feature vector. The VAR component generates the structured one-step forecast:

$$\hat{y}_{t+1}^{VAR} = c + \sum_{i=1}^{p} A_i\, y_{t+1-i} \quad (2)$$

where p is the selected VAR lag order, c is the intercept vector, and A_i is the coefficient matrix for lag i. The residual between the observed and VAR-predicted state is:

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1}^{VAR} \quad (3)$$

The residual-correction layer estimates nonlinear residual structure using XGBoost:

$$\hat{e}_{t+1} = f_{XGB}(X_t, \hat{y}_{t+1}^{VAR}) \quad (4)$$

The final hybrid prediction is:

$$\hat{y}_{t+1} = \hat{y}_{t+1}^{VAR} + \hat{e}_{t+1} \quad (5)$$

The complete computational sequence was: prepare chronological data, engineer features, fit VAR, generate one-step VAR forecasts, compute residuals, train XGBoost residual correction, add predicted residuals to VAR forecasts, derive imbalance from corrected demand and generation, and evaluate on the chronological test set.

Table 5. VAR lag-selection diagnostics

Lag	AIC	BIC	HQIC
1	85.8398	85.8477	85.8422
2	85.3781	85.3929	85.3827
3	85.2494	85.2710	85.2561
6	85.1099	85.1522	85.1230
12	85.0241	85.1078	85.0500

3.7 Evaluation strategy

The data were split chronologically into training, validation and testing subsets of 70%, 15% and 15% respectively. The evaluation of the performance of the models was done by applying the following metrics: mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R-squared), symmetrical mean absolute percentage error (SMAPE) and selectively mean absolute percentage error (MAPE). The report of MAPE was suppressed if there was a high percentage of near-zero target observations and was reported only if there were not too many near-zero observations in the target data. The statistical robustness was evaluated by considering the RMSE confidence intervals and Diebold–Mariano tests.

3.8 Robustness, interpretability, and scenario analysis

Three types of robustness were assessed: regime-specific evaluation, ablation analysis, and scenario-based predictive sensitivity. Peak hour periods and load shedding events were the subjects of regime specific testing. Mathematically induced variables and feature related to the source were eliminated in ablation experiments. The XGBoost residual-layer feature importance method was used to evaluate interpretability. Selected demand and supply-composition inputs were perturbed and recomputed dependent variables prior to making predictions using scenario analysis techniques. The perturbations were viewed as bounded predictive sensitivity tests in the learned model space, rather than simulating physical changes to the sensor grid. The perturbations were seen as predictive sensitivity tests in the learned model space, rather than simulating physical changes to the sensor grid. The computer timing was recorded to measure the extra overhead added by VAR forecasting and residual correction.

4. Results

4.1 Data preparation and modeling readiness

The timing after the parsing of datetime, chronological indexing, removing duplicate timestamps, preprocessing, constructing features, and filtering the target, there were 92,193 valid hourly observations in the final modeling dataset. There were 92,650 records in the raw data, after deduping records with the same timestamp there were 92,218 unique hourly records. The hybrid feature matrix had 325 predictors with three VAR forecast feature meta-features added to the baseline feature matrix, which had 322 predictors. Three of the directly learned targets were the next hour demand, generation, and load shedding. Derived imbalance was measured downstream as predicted demand minus predicted generation. The final modeling structure and chronological split is presented in Table 6. This table shows that all models were evaluated on the same fixed chronological training, validation, and test partitions. This ensured that model comparison was based on identical temporal data structure and avoided random-split leakage.

Table 6. Analytical dataset structure and chronological partitioning

Component	Value
Raw observations	92,650
Unique hourly observations	92,218
Final modeling observations	92,193
Baseline predictor count	322
Hybrid predictor count	325
Learned targets	3
Derived evaluation target	1
Training observations	64,535
Validation observations	13,829
Test observations	13,829

4.2 Baseline performance across learned and derived targets

There was variability in baseline performance across targets. Random Forest performed the best with the lowest demand RMSE, and the hybrid framework performed best with the lowest generation RMSE for the directly learned continuous targets. Ridge regression did not work well, as a linear model was not enough in the high-dimensional nonlinear feature space. The test sets RMSE values for the demand and generation are shown in Table 7.

The Random Forest model yielded the minimum value of RMSE for the demand while the hybrid model yielded the minimum value of RMSE for the generation as presented in Table 7. For continuous operational variables, both nonlinear ensemble learning and residual-corrected mathematical forecasting are competitive, as the difference between the two forecasts is slight for demand. The ranking was significantly different for load shedding and derived imbalance due to their sparse, stress-related, and strong persistence-driven nature. The test-set RMSE values are given in Table 8. For both load shedding and derived imbalance, the lowest RMSE is obtained for naïve persistence as shown in Table 8. For these targets the percentage-error metrics were not as stable since many observations were either equal to zero or close to it, and therefore RMSE and MAE were more relevant for ranking than SMAPE, which was only interpreted with caution.

Table 7. Test-set RMSE comparison for learned targets

Model	Demand RMSE	Generation RMSE
Naïve	601.95	470.08
Ridge	1842.63	2437.46
Random Forest	588.95	435.08
XGBoost	804.76	482.51
VAR	803.01	842.14
Hybrid	599.59	427.18

Table 8. Test-set RMSE comparison for load shedding and derived imbalance

Model	Load Shedding RMSE	Derived Imbalance RMSE
Naïve	93.50	404.82
Ridge	491.50	732.67
Random Forest	363.10	483.94
XGBoost	416.07	563.17
VAR	352.28	473.41
Hybrid	383.77	646.56

4.3 Mathematical core performance

Demand, generation, load shedding, gas, liquid fuel, coal and total fuel imports were used to estimate the VAR mathematical core. The performance of candidate lags of 1, 2, 3, 6 and 12 were compared and lag 12 was chosen as it gave the lowest AIC. Table 9 indicates standalone test-set performance of VAR Core. Table 9 indicates that the temporal structure in demand and generation was captured – evidenced by positive R^2 values – by VAR. This did not, however, produce the lowest RMSE for any target, indicating that a nonlinear correction or some other method comparing the baseline is needed.

Table 9. Test-set performance of the VAR mathematical core

Target	RMSE	R^2
Demand	803.01	0.8889
Generation	842.14	0.8559
Load shedding	352.28	-0.0353
Derived imbalance	473.41	-0.0966

4.4 Unified hybrid framework performance

The hybrid architecture consisted of using VAR forecasts and adding an XGBoost residual-correction layer. It obtained the following RMSEs: 599.59 for the demand, 427.18 for the generation, 383.77 for the load shedding, and 646.56 for the derived imbalance. This hybrid model showed the best performance for generation and was competitive for demand, though did not outperform naïve persistence for stress-related targets. Demand and generation are shown in Figure 1 so that the behavior of the actual and predicted data can be seen graphically for the operational variables of the system as they are continuous. Figure 1 shows that the hybrid-predicted demand and generation series generally followed the observed temporal patterns over the representative test segment. The alignment of major fluctuations indicates that the hybrid framework captured dominant short-term operating structure, although local deviations remained during sharper transitions.

4.5 Comparative model ranking

The final ranking showed that no single model dominated all targets. Random Forest performed best for demand, the hybrid framework performed best for generation, and naïve persistence performed best for load shedding and derived imbalance. This target-specific ranking is summarized in Figure 2.

Figure 2 shows that continuous operational variables favored Random Forest or the hybrid framework, whereas sparse stress-related targets were more accurately captured by naïve persistence. Diebold–Mariano tests were conducted to assess whether selected differences were statistically meaningful, as shown in Table 10. The hybrid framework was statistically similar to the Random Forest method for demand and generation and significantly better than VAR for these continuous targets as shown in Table 10. In contrast, naïve persistence was significantly better than the hybrid model for load shedding and derived imbalance. This was further confirmed by the RMSE confidence intervals which were calculated for each model: for the demand model, the interval for the Random Forest was (555.59,625.07), and for the Hybrid model, it was (570.08, 632.38); for the generation model, the interval for the Random Forest was (424.61,447.46) and the interval for the Hybrid model was (414.99,440.41).

Table 10. Diebold–Mariano statistical comparison summary

Target	Key comparison	p-value	Interpretation
Demand	Hybrid vs. Random Forest	0.3682	Not statistically significant
Demand	Hybrid vs. VAR	<0.001	Hybrid significantly better
Generation	Hybrid vs. Random Forest	0.3660	Not statistically significant
Generation	Hybrid vs. VAR	<0.001	Hybrid significantly better
Load shedding	Naïve vs. Hybrid	<0.001	Naïve significantly better
Derived imbalance	Naïve vs. Hybrid	<0.001	Naïve significantly better

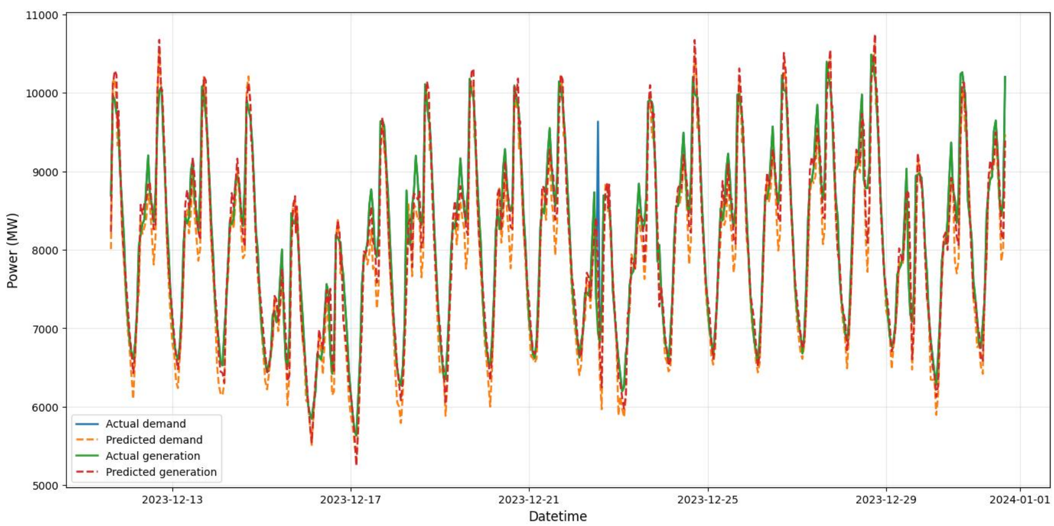

Figure 1. Actual and predicted demand and generation on the test set

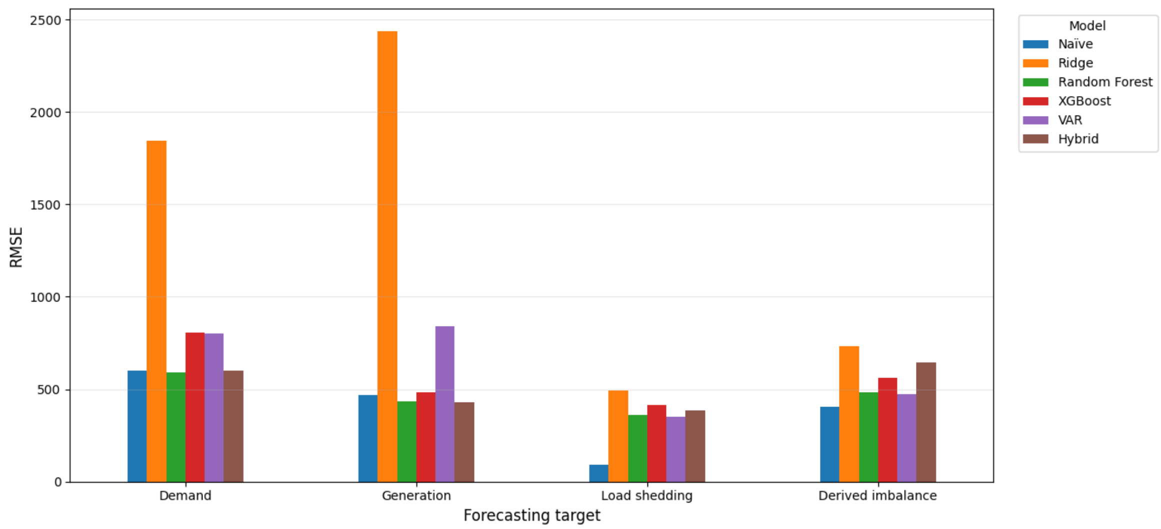

Figure 2. Test-set RMSE comparison across models and targets

4.6 Operational-regime evaluation

Operational-regime tests were used to assess the performance of hybrids during peak hour and load shedding event times. A test subset of 3,918 peak-hour observations and 7,092 load-shedding-event observations were collected. The RMSE values were 528.82, 396.92, 369.41 and 586.88 for demand, generation, load shedding and imbalance, respectively, during the peak hours. During load-shedding events, errors increased to 684.75, 484.47, 483.77, and 840.75, respectively. The results shown in Figure 3 are the regime-specific results. Figure 3 shows that errors were higher during load-shedding-event periods than during peak-hour periods for all four targets. These results are interpreted as diagnostic robustness evidence, indicating that stressed operating regimes were more difficult to forecast than recurring peak-hour conditions.

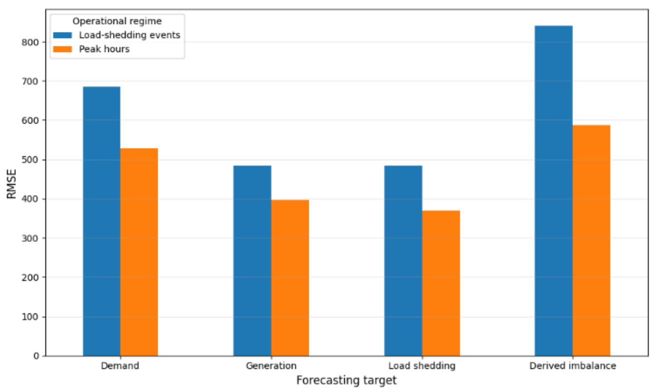

Figure 3. Hybrid model performance under peak-hour and load-shedding-event conditions

4.7 Ablation and component contribution analysis

Ablation analysis showed that feature contributions were target-specific. Removing mathematically derived variables produced RMSE values of 534.15 for demand, 536.65 for generation, 369.35 for load shedding, and 540.85 for imbalance.

Removing source-related variables produced RMSE values of 472.29, 513.03, 340.40, and 476.77, respectively. The full and ablated configurations are compared in Figure 4.

Figure 4 shows that removing feature groups did not affect all targets uniformly. Source-related variables were especially important for generation, while the contribution of mathematically derived and source-related features varied across the remaining outputs. These results are interpreted as component-sensitivity evidence rather than causal proof of feature-group importance.

4.8 Residual-model interpretation, scenario response, and computational cost

Feature-importance analysis showed that the residual-correction layer relied mainly on recent dynamic changes. The most important predictors included generation percentage change, generation ramp, demand ramp, solar rolling minimum, and peak-hour status. Figure 5 presents the ranked feature-importance profile. The residual-correction model was mainly explained by the recent short-term system dynamics and some temporal/sourcerelated variables, as shown in Figure 5. The analysis also implied that XGBoost provided a sort of "local nonlinear correction" mechanism, which is not too dissimilar from structured VAR forecasts.

The hybrid framework was found to add to the processing via VAR rolling forecasts and residual correction, as computational timing demonstrated that this was the case. The time it takes for VAR rolling forecast generation is 4.1192 s, while the time for hybrid residual-correction training is 22.6471 s, which is significantly shorter than the time required for baseline training using XGBoost (19.0064 s) or baseline training using Random Forest (869.1565 s). So, the hybrid model had added pipeline complexity, but its computational requirements were still manageable in the offline experiment setting. The "scenario analysis" was read as "predictive sensitivity" and not as "causal simulation". The greatest increase in predicted load shedding and derived imbalance occurred when the perturbation of gas reduction was applied, suggesting that the learned forecasting structure was more sensitive to the perturbation of gas reduction.

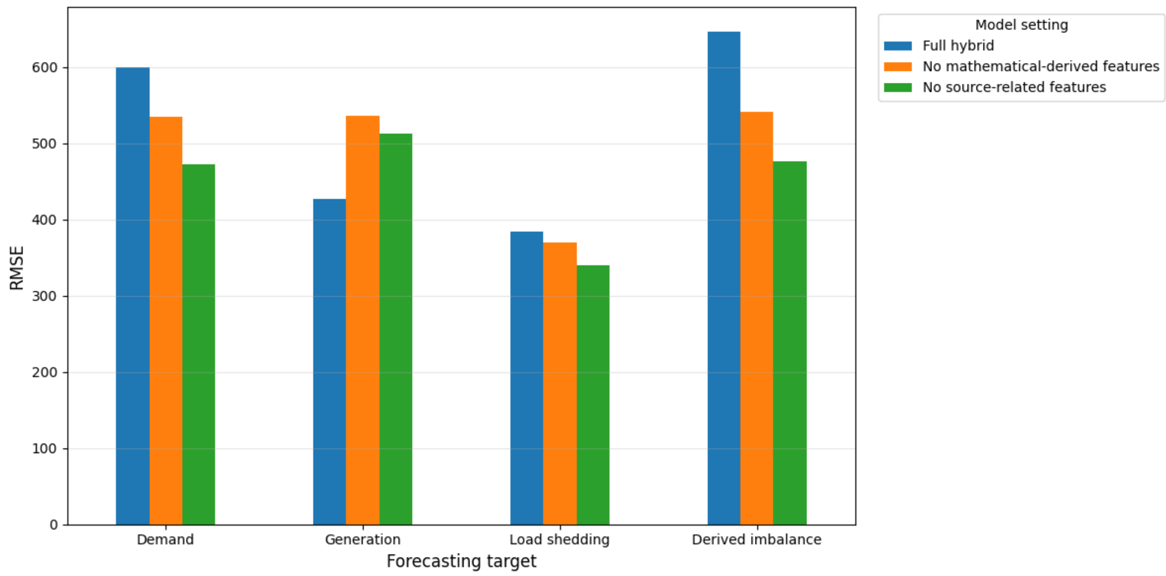

Figure 4. Ablation study showing the effect of removing major framework components

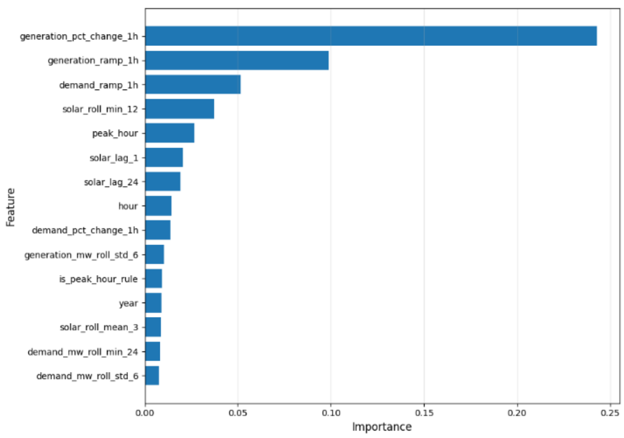

Figure 5. Top feature importances in the XGBoost residual-correction model

In summary, the results suggest that the hybrid framework can be beneficial for continuous operational forecasting, particularly generation; however, additional operational inputs and extensions of the modeling approach for events are required for sparse stress targets.

5. Discussion

The results show that the mathematical-computational framework proposed is the most effective when the target refers to a continuous operational state. The lowest test RMSE for generation was obtained by the hybrid framework while the lowest test RMSE for demand was obtained by the RF model, the two models were found to be statistically equivalent. This is a typical pattern, where the VAR core and XGBoost residual-correction layer are capable of fitting common multivariate structures and short-term deviations that are nonlinear. The framework, however, failed to outperform naïve persistence for load shedding and derived imbalance. The outcome is target dependent such that added complexity does not necessarily enhance every power-system outcome.

The comparative ranking gives a realistic interpretation of forecasting behaviour. In these continuous states, there was a strong short-term auto-correlation, periodicity within the day and smooth transitions, so nonlinear ensemble learning and residual correction proved helpful. This result is consistent with the insights gained from hybrid forecasting studies that complementary model combinations can deliver better forecasts when a signal has regular temporal structure and nonlinear residual variation [11]. Load shedding did not behave in that way. This target was deemed to be sparse, locally persistent, and affected by operational decisions that are not captured in the dataset, which is why naïve persistence is superior. Advanced models were then therefore, only marginally superior to the current-value benchmark. Therefore, complexity needs to be assessed on the dynamics it offers in relation to the target, not taken for granted [13].

Imbalances in the derived figures should be treated with caution. The imbalance was not learned separately, but was derived from the predicted demand and generation, thus maintaining mathematical consistency of the two main operating variables. However, this design also permitted errors in the forecast of both components to be transmitted to the estimated value. If the demand forecast and generation forecast were both reasonably accurate, but both had small errors, the errors could add together to give a large imbalance error. The lower imbalance performance is thus not necessarily interpreted as a failure of the framework to represent system structure. Instead, it demonstrates that compound prediction uncertainty is reflected in the derived stress indicators. This also accounts for the lack of consistency of percentage-based metrics for load shedding and imbalance and the caution shown in emphasizing absolute-error measures and SMAPE.

The results of the ablation and feature-importance also support the proposed framework. The mathematically derived variables and source-related variables did not all impact all targets equally, revealing that there was a target-specific contribution from features. In the source-related variables, generation was of particular interest, while generation percentage change, generation ramp, demand ramp, solar rolling minimum and peak-hour status were important in the residual-correction layer. This lends weight to the argument that XGBoost was a local nonlinear adjustment mechanism and not an alternative to the VAR core.

The result is consistent with the decomposition-oriented and feature-sensitive forecasting research in which gains in forecasting are often achieved when models are fed with variables that capture meaningful transitions in the systems being forecasted, not just raw historical information [14]. It is also in accord with error-correction research that supports the presence of residual structure that occurs in second-stage learners that does not occur in the first-stage learners who make the initial predictions [20].

The errors were found to be higher during the load-shedding-event times than the peak-hour times in the regime-specific evaluation. This means that operating regimes with high stress levels are harder to predict than repetitive high load operating. This is in line with research on multi-stage hybrid forecasting which suggests that while there may be a relationship between average forecasting accuracy and robustness in unusual or stressed situations, there is no guarantee of equality [22]. On the practical side, the framework seems to be most useful for forecasting in a continuous state, particularly generation forecasting, whereas the forecasting a stress state needs information on events, such as outage records, reserve status, dispatch instructions, voltage, frequency, or congestion indicators. Therefore, the outputs of the scenarios should be read as a prediction of sensitivity within the space of learned models and not as a causal simulation of physical interventions in the grid.

5.1 Limitations and future work

There are some restrictions to be taken into account. First of all, the analysis relies on a single national power-system dataset of Bangladesh; transferability to other power-system structures, market designs and levels of renewable penetration is therefore dependent on external validation. Secondly, the study was only concerned with one-step-ahead forecasting, but operational planning could involve multi-step and probabilistic horizons. Third, the data set was lacking of some variables, such as voltage, frequency, outage, congestion, reserve-violation, and dispatch-control, which may be necessary for a sparse set of stress targets. Fourth, while the hybrid framework was found to be impractical for real-time deployment in the case of EMS/SCADA, additional testing of latency, data availability, model updating, and uncertainty quantification would be needed for such a deployment. Multi-horizon forecasting, regime-adaptive learning, event-aware stress prediction, external validation and more interpretable tools like SHAP or partial-dependence analysis should be explored in future work.

6. Conclusion

This work was the development and testing of a single mathematical-computational approach for one-step-ahead predictive analysis of dynamic power-system operating states. The framework was a hybrid of a vector autoregressive mathematical kernel and an XGBoost residual-correction layer to predict demand, generation and load shedding; imbalance was then calculated from the predicted demand and generation to maintain mathematical balance. The results indicate that the framework was successful for continuous operational variables. Its generation performance was the best and it was statistically competitive for demand, though Random Forest performed best in terms of raw demand RMSE. Naïve persistence, on the other hand, was superior in the load shedding and derived imbalance tasks, suggesting that different modelling approaches must be used for different types of stress-related variables, as compared to continuous system states. The study also makes methodological contribution in conceptualizing hybrid forecasting as a well-defined mathematical-computational process, and not a mere "ad hoc" mix between algorithms. It also provides real-life practice through the assessment of several outputs, relevant to the grid, in a common time-staggered testing design. The results demonstrate that the predictive value is not only a function of the complexity of the model but also a function of each of the statistical and operational nature of the target variables. Hence, hybrid models need to be interpreted in a target-specific manner, particularly when stress events are few and far between, or when they are affected by unobserved operational decisions. Expansion to multi-step forecasting, probabilistic uncertainty estimation, regime-adaptive learning and general validation performed for various power systems should be extended in the future. Future enhancements could include event-aware features, physics-guided constraints, SCADA/EMS variables and more robust interpretability tools for more reliable forecasting in varying grid conditions.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, additional data will be provided by the corresponding author upon reasonable request.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] M. G. Pinheiro, S. C. Madeira, and A. P. Francisco, "Short-term electricity load forecasting—A systematic approach from system level to secondary substations," Applied Energy, vol. 332, Art. no. 120493, 2023. https://doi.org/10.1016/j.apenergy.2022.120493

[2] G. F. Fan, Y. R. Liu, H. Z. Wei, M. Yu, and Y. H. Li, "The new hybrid approaches to forecasting short-term electricity load," Electric Power Systems Research, vol. 213, Art. no. 108759, 2022. https://doi.org/10.1016/j.epsr.2022.108759

[3] Y. Eren and İ. Küçükdemiral, "A comprehensive review on deep learning approaches for short-term

load forecasting," Renewable and Sustainable Energy Reviews, vol. 189, Art. no. 114031, 2024. https://doi.org/10.1016/j.rser.2023.114031

[4] G. Tziolis, J. Lopez-Lorente, M. I. Baka, A. Koumis, A. Livera, S. Theocharides, et al., "Direct short-term net load forecasting in renewable integrated microgrids using machine learning: A comparative assessment," Sustainable Energy, Grids and Networks, vol. 37, Art. no. 101256, 2024. https://doi.org/10.1016/j.segan.2023.101256

[5] Y. Xiao, X. Hu, Y. Lin, Y. Lu, R. Jing, and Y. Zhao, "Interpretable short-term electricity load forecasting considering small sample heatwaves," Applied Energy, vol. 398, Art. no. 126417, 2025. https://doi.org/10.1016/j.apenergy.2025.126417

[6] X. Wang, H. Wang, B. Bhandari, and L. Cheng, "AI-empowered methods for smart energy consumption: A review of load forecasting, anomaly detection and demand response," International Journal of Precision Engineering and Manufacturing-Green Technology, vol. 11, no. 3, pp. 963–993, 2024. https://doi.org/10.1007/s40684-023-00537-0

[7] Y. Jiang, Y. Li, and Y. Chen, "Interpretable short-term load forecasting via multi-scale temporal decomposition," Electric Power Systems Research, vol. 235, Art. no. 110781, 2024. https://doi.org/10.1016/j.epsr.2024.110781

[8] A. B. Ferreira, J. B. Leite, and D. H. Salvadeo, "Power substation load forecasting using interpretable transformer-based temporal fusion neural networks," Electric Power Systems Research, vol. 238, Art. no. 111169, 2025. https://doi.org/10.1016/j.epsr.2024.111169

[9] L. Baur, K. Ditschuneit, M. Schambach, C. Kaymakci, T. Wollmann, and A. Sauer, "Explainability and interpretability in electric load forecasting using machine learning techniques–a review," Energy and AI, vol. 16, Art. no. 100358, 2024. https://doi.org/10.1016/j.egyai.2024.100358

[10] C. Wang, H. Zhao, Y. Liu, and G. Fan, "Minute-level ultra-short-term power load forecasting based on time series data features," Applied Energy, vol. 372, Art. no. 123801, 2024. https://doi.org/10.1016/j.apenergy.2024.123801

[11] L. Zhang and D. Jánošík, "Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches," Expert Systems with Applications, vol. 241, Art. no. 122686, 2024. https://doi.org/10.1016/j.eswa.2023.122686

[12] G. F. Fan, Y. Y. Han, J. W. Li, L. L. Peng, Y. H. Yeh, and W. C. Hong, "A hybrid model for deep learning short-term power load forecasting based on feature extraction statistics techniques," Expert Systems with Applications, vol. 238, Art. no. 122012, 2024. https://doi.org/10.1016/j.eswa.2023.122012

[13] M. Y. Junior, R. Z. Freire, L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. dos Santos Coelho, "Optimized hybrid ensemble learning approaches applied to very short-term load forecasting," International Journal of Electrical Power & Energy Systems, vol. 155, Art. no. 109579, 2024. https://doi.org/10.1016/j.ijepes.2023.109579

[14] H. Chen, H. Huang, Y. Zheng, and B. Yang, "A load forecasting approach for integrated energy systems based on aggregation hybrid modal decomposition and combined model," Applied Energy, vol. 375, Art. no. 124166, 2024. https://doi.org/10.1016/j.apenergy.2024.124166

[15] C. Song, H. Yang, J. Cai, P. Yang, H. Bao, K. Xu, and X. B. Meng, "Multi-energy load forecasting via hierarchical multi-task learning and spatiotemporal attention," Applied Energy, vol. 373, Art. no. 123788, 2024. https://doi.org/10.1016/j.apenergy.2024.123788

[16] N. Mounir, H. Ouadi, and I. Jrhilifa, "Short-term electric load forecasting using an EMD-BI-LSTM approach for smart grid energy management system," Energy and Buildings, vol. 288, Art. no. 113022, 2023. https://doi.org/10.1016/j.enbuild.2023.113022

[17] Y. Y. Hong and Y. H. Chan, "Short-term electric load forecasting using particle swarm optimization-based convolutional neural network," Engineering Applications of Artificial Intelligence, vol. 126, Art. no. 106773, 2023. https://doi.org/10.1016/j.engappai.2023.106773

[18] W. Xiao, L. Mo, Z. Xu, C. Liu, and Y. Zhang, "A hybrid electric load forecasting model based on decomposition considering fisher information," Applied Energy, vol. 364, Art. no. 123149, 2024. https://doi.org/10.1016/j.apenergy.2024.123149

[19] Q. Xing, X. Huang, J. Wang, and S. Wang, "A novel multivariate combined power load forecasting system based on feature selection and multi-objective intelligent optimization," Expert Systems with Applications, vol. 244, Art. no. 122970, 2024. https://doi.org/10.1016/j.eswa.2023.122970

[20] C. Fan, S. Nie, L. Xiao, L. Yi, and G. Li, "Short-term industrial load forecasting based on error correction and hybrid ensemble learning," Energy and Buildings, vol. 313, Art. no. 114261, 2024. https://doi.org/10.1016/j.enbuild.2024.114261

[21] K. Li, P. Duan, X. Cao, Y. Cheng, B. Zhao, Q. Xue, and M. Feng, "A multi-energy load forecasting method based on complementary ensemble empirical model decomposition and composite evaluation factor reconstruction," Applied Energy, vol. 365, Art. no. 123283, 2024. https://doi.org/10.1016/j.apenergy.2024.123283

[22] C. Fan, S. Nie, L. Xiao, L. Yi, Y. Wu, and G. Li, "A multi-stage ensemble model for power load forecasting based on decomposition, error factors, and multi-objective optimization algorithm," International Journal of Electrical Power & Energy Systems, vol. 155, Art. no. 109620, 2024. https://doi.org/10.1016/j.ijepes.2023.109620

[23] "PGCB Hourly Generation Dataset (Bangladesh)," UCI Machine Learning Repository, 2025. [Online]. Available: https://doi.org/10.24432/C59P6V.