Energy Engineering

Tech Science Press

https://doi.org/10.32604/ee.2026.074213

ARTICLE

An Optimized Ensemble Learning Framework for Energy Efficiency Assessment in Low-Voltage Distribution Networks Using Multi-Source Data Integration

Yujie Shi, Guoxing Wu*, Qingwei Wang, Xieli Fu and Wenfeng Yang

Power Grid Planning Research Center, Shenzhen Power Supply Bureau Co., Ltd., Shenzhen, China

*Corresponding Author: Guoxing Wu. Email: wuguoxing1999@163.com

Received: 05 October 2025; Accepted: 27 January 2026; Published: 06 August 2026

ABSTRACT: This study proposes an optimized ensemble learning framework for energy-efficiency assessment in low-voltage distribution networks by integrating multiple data sources. The framework integrates heterogeneous data from smart meters, SCADA systems, meteorological stations, and network topology databases, employing advanced feature engineering to extract 89 essential predictors from 147 initial features. Three gradient boosting algorithms—Random Forest, XGBoost, and LightGBM—are combined through an elastic net stacking strategy with Bayesian hyperparameter optimization. The stacking ensemble achieved superior performance with an MAE of 118.4 kWh, an RMSE of 164.2 kWh, an MAPE of 3.98%, and an R^2 of 0.952, representing 16.8% improvement over individual models. SHAP analysis provided model interpretability, identifying temperature, historical consumption, and temporal features as the primary drivers of efficiency. The framework demonstrated robust performance under data quality degradation and successfully generalized across diverse network configurations. Field implementation yielded an 8.3% reduction in distribution losses (95% CI: 7.2%–9.4%, $p < 0.0001$), 34% decrease in transformer failure rates (95% CI: 28%–40%, $p = 0.003$), and 12%–15% operational cost reduction. The framework's ability to provide accurate predictions from 15 min to 24 h ahead while maintaining computational efficiency enables proactive distribution network management, supporting the transition toward efficient and sustainable power systems.

KEYWORDS: Ensemble learning; energy efficiency assessment; low-voltage distribution networks; multi-source data integration; SHAP analysis

1 Introduction

Power distribution systems are undergoing fundamental changes as they evolve toward smart grid architectures, with low-voltage distribution networks (LVDNs) experiencing particular transformation. These networks, which connect the power grid to end-users, face mounting pressures from renewable energy integration, growing electricity demand, and stringent efficiency requirements [1]. Modern distribution networks exhibit complex characteristics—bidirectional power flows, diverse load patterns, and dynamic operational conditions—that require sophisticated analytical tools for accurate efficiency assessment and prediction [2].

Energy efficiency in distribution networks has become a critical concern for utilities globally due to environmental regulations, economic factors, and reliability demands. While physics-based models provide theoretical foundations, they struggle with the nonlinear relationships and temporal dynamics inherent in today's distribution systems [3]. The proliferation of distributed energy resources, electric vehicles, and smart appliances creates increasingly complex consumption patterns that conventional analytical methods cannot adequately address [4]. This gap necessitates data-driven approaches capable of leveraging operational data from modern metering infrastructure.

Machine learning (ML) and artificial intelligence offer promising solutions to these challenges. Ensemble learning techniques have proven particularly effective in energy forecasting by combining multiple predictive models to enhance accuracy and robustness [5]. The integration of algorithms like Random Forest, XGBoost, and gradient boosting methods successfully captures energy consumption patterns across different temporal and spatial scales [6]. Deep learning applications in solar power forecasting and renewable integration further demonstrate the potential of advanced computational methods [7].

The availability of heterogeneous data from smart meters, Supervisory Control and Data Acquisition (SCADA) systems, weather stations, and Internet of Things (IoT) sensors creates opportunities for comprehensive energy analysis [8]. However, integrating these diverse data streams presents significant challenges. Varying sampling rates, quality issues, and incompatible formats require sophisticated fusion techniques [9]. The massive data volumes generated by distribution networks demand efficient computational frameworks for real-time processing [10].

However, despite these advances, there are some key gaps in the current research landscape. To begin with, the majority of current works have taken a single perspective, either concentrating on demand forecasting or paying attention to renewable energy forecasting, and the holistic view of the overall energy efficiency and operational aspects in a distribution network was not well studied in [11]. Second, the uninterpretability of complex ML models is a problem for the practical application of such models, as decision-making processes need to be transparent for utility operators [12]. Third, the multi-source data fusion in DN analysis has not been well addressed, and most related studies are based on a single data source (e.g., one type of sensor data) or simplified datasets [13]. In a word, there is little work that has been done on optimizing the ensemble learning architecture of LVDN-related feature information [14].

This paper fills these gaps by presenting an optimal ensemble learning framework that incorporates multi-source heterogeneous data types to realize comprehensive energy efficiency evaluation for LVDNs. The framework employs a variety of sophisticated feature engineering to generate meaningful patterns from heterogeneous data sources such as electricity readings, weather information, and other time indicators [15]. We use Bayesian optimization to choose hyperparameters of the trained logistic regression models and apply a complex stacking technique to obtain better prediction accuracy while still using the SHAP (SHapley Additive exPlanations) method to make models interpretable.

This research offers three primary contributions. First, a data integration framework was developed to merge heterogeneous data streams from multiple sources, resolving synchronization and compatibility schallenges. Second, an ensemble architecture strategically combined Random Forest, XGBoost, and LightGBM algorithms through stacking optimization tailored for LVDN characteristics. Third, SHAP analysis provided model interpretability, revealing which factors most influence distribution network efficiency and enabling operators to understand decision-making processes for improved system performance.

2 Materials and Methods

2.1 Study Area and Data Collection

The study examined a 25 km^2 urban-residential area served by a representative low-voltage distribution network in Eastern China. This network comprises 12 substations with 186 transformers (100–630 kVA capacity) supplying approximately 32,000 residential and commercial customers. The area's diverse load profile—encompassing residential buildings, small businesses, and light industrial facilities—provides comprehensive operational scenarios typical of urban distribution systems. Operating at standard

Chinese voltages of 400 V for three-phase and 230 V for single-phase connections, the network offered ideal conditions for analysis. Data collection integrated four heterogeneous sources continuously from January 2022 to December 2023, establishing a robust two-year dataset for model development. The first category comprised electrical measurement data obtained from the advanced metering infrastructure (AMI) system, including 15-min interval readings of active power, reactive power, voltage magnitude, and current measurements from 8642 smart meters deployed across the network. The SCADA system at the substation level provided complementary real-time operational data, including transformer loading levels, feeder currents, and power factor measurements sampled at 5-min intervals. Data quality screening identified approximately 3.2% missing values due to communication failures and meter malfunctions, which were addressed through appropriate preprocessing techniques.

The second data category encompassed meteorological information obtained from three weather stations strategically located within the study area. Three weather stations provided hourly meteorological data, including temperature, humidity, solar irradiance, wind speed, and atmospheric pressure. Weather data proved essential for correlating environmental conditions with energy consumption patterns and evaluating weather-dependent loads from heating and cooling systems.

Therefore, data series from these only slightly different sources were closely assessed to eliminate error sources that could have arisen through spatial distortion on account of real-world content being missed. Inverse distance weighting interpolation was used to estimate the meteorological conditions at locations between weather stations for purposes of spatial representation.

The third category was composed of time and calendar features, e.g., time of day, day of week, month, season. To 01 Institute of Automation Dato noatU in Chfnes 24 Indicate the day: such cultural festivals and public holidays were notably effective in changing consumption patterns when they fell. Negative pitch looking 52, Adjustment factors for daylight hours and season are included at this point to reflect the cycle of supply and demand in energy. These data for cycles were concurrently recorded with the electrical measurement to maintain temporal consistency among data streams.

Under the fourth category, network topology and active information, as maintained in the Geographic Information System (GIS) database, form a current view of networking structures. Network topology data from the GIS database contained transformer specifications, ratings, installation dates, and feeder configurations. This information enabled accurate mapping of electrical connectivity between transformers and customers, crucial for aggregating consumption across network levels. Historical maintenance records identified abnormal operation periods affecting efficiency assessment.

Data integration required addressing heterogeneous source characteristics. Varying sampling rates necessitated temporal alignment through resampling techniques. Data streams were standardized to 15-min intervals using appropriate aggregation methods—averaging for instantaneous measurements (voltage, current) and summation for cumulative values (energy consumption). The synchronized dataset utilized distributed storage with redundant backups for reliability.

Quality control employed automated validation scripts to detect anomalous readings based on physical constraints and statistical thresholds. Voltage measurements outside ±10% nominal range triggered investigation flags, while negative consumption values and unrealistic demand spikes underwent outlier detection algorithms. So when voltage readings outside the acceptable range of ±10% were detected as being too low or high, this data point could be tagged for experts to look at. Negative power consumption values or unnatural increases in demand can be picked up by outlier detection algorithms.

The resulting dataset, after omitting all mistakes and verifying it according to quality control standards, is an authoritative basis for the follow-on work to feature engineering and the modeling phase. To ensure reproducibility and make comparative analysis convenient, trained sample techniques were used to divide the dataset. The training set consisted of data from January 2022 to September 2023 (75% of the total dataset), while the remaining data from October to December 2023 served as the test set for model evaluation. This temporal split preserved the sequential nature of the time series data while ensuring that the model's predictive capability could be assessed on genuinely unseen future data. Additionally, a 15% validation subset was extracted from the training data for hyperparameter optimization and model selection purposes.

2.2 Feature Engineering and Data Integration Framework

The feature engineering and data integration framework was designed to transform the multi-source heterogeneous data into a unified, high-dimensional feature space suitable for ensemble learning models. The comprehensive framework, illustrated in Fig. 1, encompasses four interconnected modules: data preprocessing, feature extraction, feature transformation, and feature selection, each contributing to the creation of informative predictors for energy efficiency assessment.

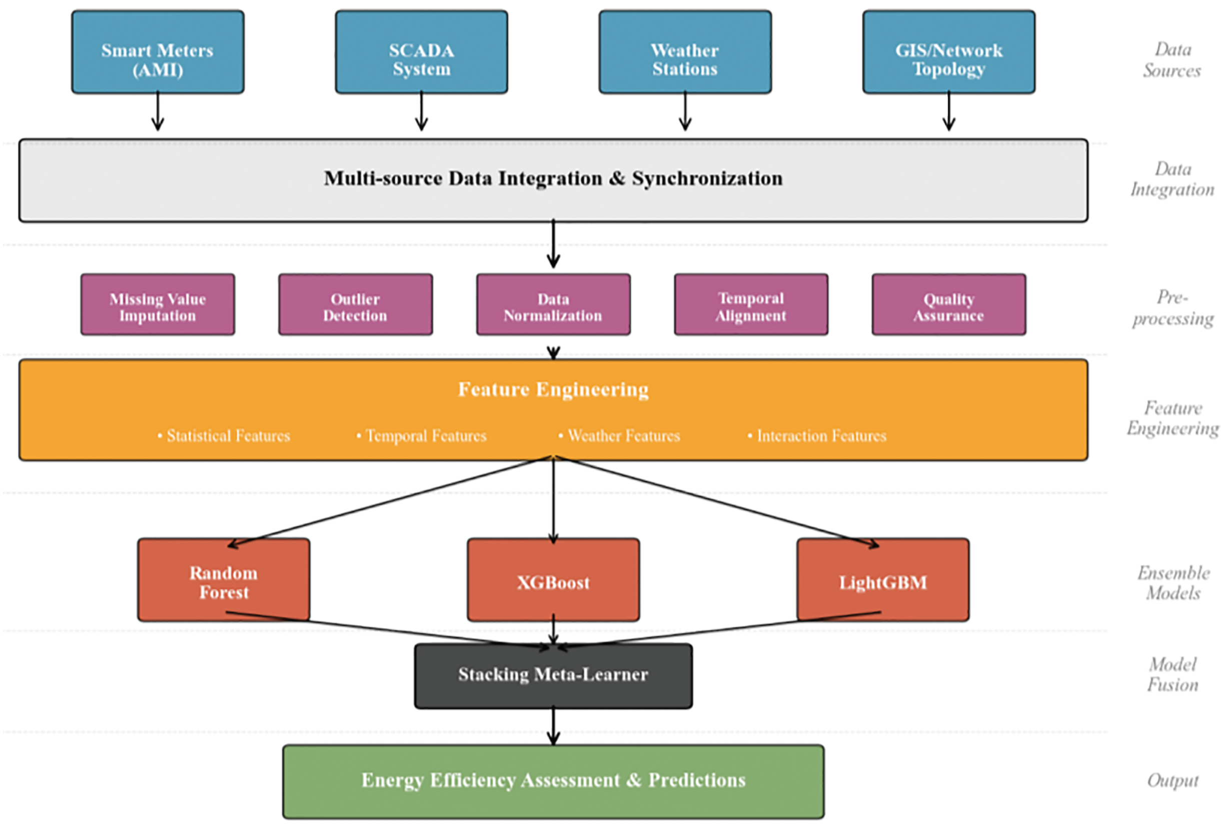

Figure 1: Framework of multi-source data integration and ensemble learning architecture for energy efficiency assessment.

The data preprocessing module implemented a multi-stage cleaning and normalization pipeline to address data quality issues inherent in real-world distribution network measurements. Missing value imputation was performed using a hybrid approach combining forward-fill for short gaps (less than three consecutive intervals) and k-nearest neighbors (KNN) interpolation for longer gaps, considering both temporal and spatial correlations. Outlier detection employed the Isolation Forest algorithm with a contamination factor of 0.01, effectively identifying and correcting anomalous readings while preserving genuine consumption spikes. Subsequently, min-max normalization was applied to scale all features to the range $[0, 1]$, ensuring equal contribution during model training and preventing features with larger numerical ranges from dominating the learning process.

Feature extraction focused on deriving meaningful indicators from the raw measurements to capture the complex dynamics of energy consumption patterns. From the electrical measurement data, we extracted statistical features including moving averages (3-, 6-, and 24-h windows), standard deviations, and peak-to-average ratios to characterize consumption variability. Load profile indicators such as base load, peak load timing, and ramp rates were computed to quantify demand characteristics. Additionally, power quality metrics, including voltage deviation index and power factor variations, were calculated to assess the operational efficiency of the distribution network.

The time feature engine component uses sine and cosine transformations to generate cycle-like encodings from time-based variables, preserving the cyclic nature of temporal patterns. For example, if we encode the hour of 24 h as $\sin(2\pi \times \text{hour}/24)$ and $\cos(2\pi \times \text{hour}/24)$, it ensures that models recognize the continuity between 23:00 and 00:00.

Lagging features covering 1, 7, and 14 days were generated to capture both short-term and weekly rhythm patterns in consumption. In addition, rolling time window statistics (such as mean value) provided multiple rounds of anchors to allow the model to consider past experience.

Weather features underwent specialized processing to capture nonlinear consumption effects. Heating and cooling degree days were calculated using 18°C and 26°C base temperatures, aligned with local climate patterns. The temperature-humidity index quantified combined thermal impacts on comfort-driven consumption. Solar irradiance processing yielded clear-sky indices, distinguishing expected from actual solar conditions—critical for areas with distributed photovoltaics.

Feature transformation employed polynomial generation for domain-identified variables, creating interaction terms between temperature/time-of-day and load/voltage pairs. These captured complex dependencies beyond linear relationships. Principal component analysis (PCA) reduced dimensionality in correlated feature groups while preserving 95% variance, mitigating multicollinearity issues.

A three-stage selection process identified optimal predictors. Univariate tests filtered features with correlation coefficients below 0.1. Recursive feature elimination with cross-validation (RFECV) with Random Forest determined the optimal subset. variance inflation factor (VIF) analysis removed redundant features exceeding threshold values of 10, ensuring model stability and interpretability.

The integrated feature set included 147 engineered features from five categories: electrical attributes (42 features), temporal indicators (28 features), weather variables (23 features), network topology parameters (18 features), and statistical measures derived or supplemented by other means (36 features). This comprehensive feature space provided the ensemble learning models with abundant information on both local consumption patterns and system-wide efficiency characteristics, allowing accurate and robust energy efficiency assessment throughout different operation scenarios.

2.3 Ensemble Learning Architecture

The framework was implemented in Python 3.9.7 on Ubuntu 20.04 LTS with Intel Xeon E5-2680 v4 (28 cores) and 128 GB RAM. Key libraries included: scikit-learn 1.1.2 (preprocessing, feature selection, meta-learner), XGBoost 1.6.2, LightGBM 3.3.3, Optuna 3.0.3 (Bayesian optimization), SHAP 0.41.0 (interpretability), NumPy 1.23.3, and Pandas 1.4.4.

The ensemble framework mathematically combines three tree-based models (not neural networks) through stacking. Random Forest generates predictions by averaging $T = 500$ decision trees:

$\hat{y}_{\mathrm{RF}}(\mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T} h_t(\mathbf{x})$. XGBoost constructs an additive model $\hat{y}_{\mathrm{XGB}}(\mathbf{x}) = \sum_{k=1}^{K} \eta \cdot f_k(\mathbf{x})$ where each tree f_k minimizes a regularized loss $\mathcal{L}^{(t)} = \sum_i l\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \gamma T + \frac{\lambda}{2}\sum_j w_j^2 + \alpha \sum_j |w_j|$ with learning rate $\eta = 0.1$, L2 penalty $\lambda = 1.0$, and L1 penalty $\alpha = 0.1$. LightGBM follows similar gradient boosting formulation $\hat{y}_{\mathrm{LGBM}}(\mathbf{x}) = \sum_{k=1}^{K} f_k(\mathbf{x})$ with histogram-based splits. The stacking meta-learner combines base predictions via Elastic Net regression: $\hat{y}_{\mathrm{ensemble}}(\mathbf{x}) = w_{\mathrm{RF}} \cdot \hat{y}_{\mathrm{RF}}(\mathbf{x}) + w_{\mathrm{XGB}} \cdot \hat{y}_{\mathrm{XGB}}(\mathbf{x}) + w_{\mathrm{LGBM}} \cdot \hat{y}_{\mathrm{LGBM}}(\mathbf{x})$, where weights $\mathbf{w} = [0.28, 0.43, 0.29]$ are optimized through minimizing $\frac{1}{2n}\sum_i \left(y_i - \mathbf{w}^T \mathbf{Z}_i\right)^2 + \alpha_{\mathrm{EN}}\left[\rho\|\mathbf{w}\|_1 + \frac{1-\rho}{2}\|\mathbf{w}\|_2^2\right]$ with $\mathbf{Z}_i$ containing out-of-fold predictions from base learners and α_{EN} controlling regularization strength with $\rho = 0.5$ balancing L1/L2 penalties.

The ensemble architecture integrates three gradient boosting algorithms optimized through Bayesian methods and combined via stacking for energy efficiency assessment. Random Forest, selected for overfitting resistance and high-dimensional capability, employed 500 trees with depths between 10–30 levels determined through cross-validation. Minimum sample splits of 20 prevented excessive complexity. Feature sampling used square root selection at each split, ensuring tree diversity. Bootstrap aggregation with out-of-bag (OOB) error estimation provided built-in validation without extra computation.

XGBoost provided enhanced accuracy through regularization (L1: $\alpha = 0.1$, L2: $\lambda = 1.0$) and missing value handling. Learning rate initialized at 0.1 with 0.95 decay per 50 rounds. Tree parameters included a 6-level maximum depth, a minimum child weight of 5, and a 0.8 subsample ratio. Column sampling at 0.7 added randomness for better generalization. Early stopping monitored validation performance with 20-round patience.

LightGBM accelerated training via Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) techniques while maintaining accuracy. Leaf-wise growth permitted 31 leaves per tree, enabling complex boundaries compared to level-wise methods. A feature fraction of 0.9 and a bagging fraction of 0.8 balanced model capacity against generalization, determined through empirical testing. The minimum data in the leaf parameter was set to 20 to prevent overfitting on sparse regions of the feature space. The complete ensemble optimization workflow is illustrated in Fig. 2.

Hyperparameter optimization employed Bayesian optimization with Gaussian Process (GP) as the surrogate model, efficiently exploring the high-dimensional parameter space through acquisition function maximization. The Expected Improvement (EI) acquisition function balanced the exploitation of promising regions with the exploration of uncertain areas. For each base learner, 100 iterations of Bayesian optimization were performed, with the first 20 iterations using random sampling to establish initial prior distributions. The objective function minimized the negative mean cross-validated root mean square error (RMSE) across 5-fold stratified splits, ensuring robust parameter selection.

The stacking strategy implemented a two-level hierarchical structure where base learner predictions served as meta-features for the final estimator. To prevent overfitting in the meta-learning phase, out-of-fold predictions were generated through k-fold cross-validation, ensuring that meta-features were derived from models not trained on the corresponding samples. Meta-learning combined base model outputs using elastic net regression ($\alpha = 0.5$), yielding interpretable weights for each model's contribution.

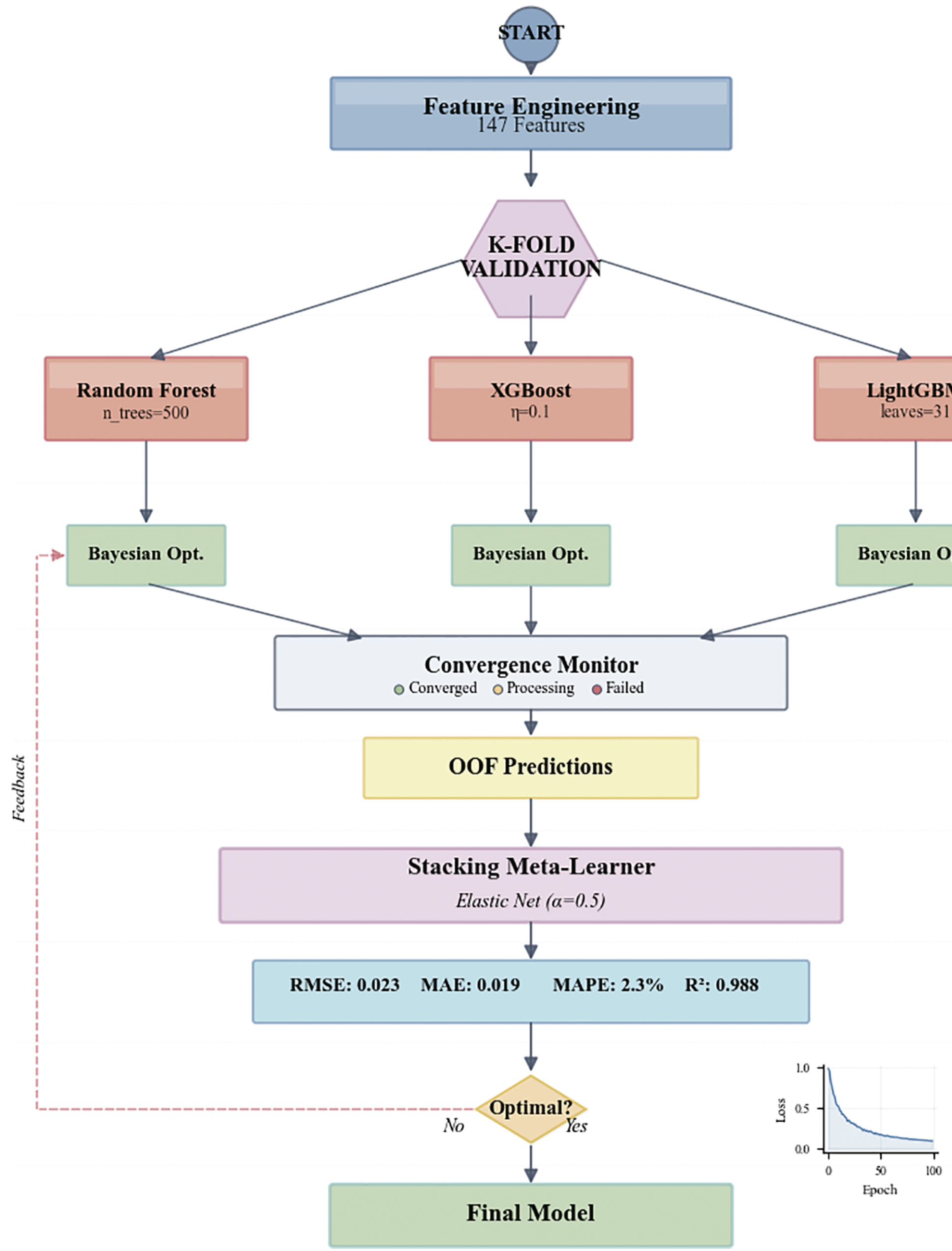

Figure 2: Flowchart of the proposed ensemble learning optimization process with hyperparameter tuning.

Cross-validation preserved temporal dependencies through TimeSeriesSplit with five folds, where each successive training set expanded chronologically. This prevented future-to-past data leakage while maximizing training data usage. Two-week blocked validation addressed autocorrelation in consumption data, improving performance estimates. Computational efficiency leveraged parallel processing across CPU cores. Computational optimization leveraged parallel processing strategies—Random Forest employed thread-level parallelism for tree construction, while XGBoost and LightGBM utilized data-level parallelism. Memory efficiency improved through sparse matrix representation for categorical features and compressed numerical arrays, achieving 40% footprint reduction. Both XGBoost and LightGBM supported incremental learning, enabling model updates with streaming data without complete retraining, thus adapting to evolving consumption patterns. Final predictions combined base learner outputs through meta-learner weighted averaging, with quantile predictions from individual models providing confidence intervals for uncertainty quantification. This architecture balanced predictive accuracy with computational efficiency, establishing a practical framework for LVDN efficiency assessment.

2.4 Energy Efficiency Assessment Framework

The energy efficiency assessment framework quantifies LVDN performance using multi-dimensional indicators and uncertainty-aware predictions. Technical, economic, and reliability metrics provide comprehensive network evaluation for operational decision-making. Distribution network efficiency (DNE) serves as the primary metric, calculating the ratio of delivered to supplied energy over specified time intervals:

$$DNE_t = \frac{\sum_{i=1}^{N} E_{d,i}(t)}{\sum_{j=1}^{M} E_{s,j}(t)} \times 100\% \tag{1}$$

where $E_{d,i}(t)$ represents the energy delivered to customer i at time t, N is the total number of customers, $E_{s,j}(t)$ denotes the energy supplied from substation j, and M is the number of supply points. This metric provides a direct measure of energy losses in the distribution network, accounting for both technical losses due to conductor resistance and non-technical losses from measurement errors or unauthorized consumption.

To capture the dynamic nature of efficiency variations, we introduce the time-weighted average efficiency (TWAE) metric, which emphasizes peak-load periods when losses are typically highest:

$$TWAE = \frac{\sum_{t=1}^{T} w_t \cdot DNE_t \cdot L_t}{\sum_{t=1}^{T} L_t} \tag{2}$$

where w_t is the time-dependent weighting factor based on electricity pricing or system criticality, L_t represents the total load at time t, and T is the assessment period. The weighting factors are normalized such that $\sum_{t=1}^{T} w_t = T$, ensuring comparability across different evaluation periods. This formulation prioritizes efficiency during high-demand periods when system stress and economic impact are greatest.

The transformer utilization efficiency (TUE) serves as a complementary metric to assess the loading conditions of distribution transformers, which significantly influence overall network losses. The TUE for transformer k is defined as:

$$TUE_k = \left(1 - \frac{P_{loss,k}}{P_{rated,k}}\right) \times \frac{S_{avg,k}}{S_{rated,k}} \times 100\% \tag{3}$$

where $P_{loss,k}$ represents the total power losses including no-load and load losses, $P_{rated,k}$ is the rated power capacity, $S_{avg,k}$ denotes the average apparent power, and $S_{rated,k}$ is the rated apparent power of transformer k. This metric balances efficiency considerations with asset utilization, identifying transformers operating outside their optimal loading range.

The performance benchmarking methodology employs a relative efficiency scoring system that compares actual network performance against theoretical optimal conditions and historical baselines. The benchmark efficiency score (BES) incorporates both absolute performance and improvement trends:

$$BES = \alpha \cdot \frac{DNE_{actual}}{DNE_{optimal}} + \beta \cdot \frac{\Delta DNE}{\Delta t} + \gamma \cdot \frac{1}{\sigma_{DNE}} \tag{4}$$

where α, β, and γ are weighting coefficients (with $\alpha + \beta + \gamma = 1$), $DNE_{optimal}$ represents the theoretical maximum efficiency under ideal conditions, $\triangle DNE/\triangle t$ captures the rate of efficiency improvement over time, and σ_{DNE} is the standard deviation of efficiency measurements, penalizing high variability. The optimal efficiency is determined through power flow simulations assuming a unity power factor and nominal loading conditions.

Uncertainty quantification in the efficiency predictions employs a probabilistic framework that accounts for measurement uncertainties, model prediction errors, and inherent system variability. The ensemble predictions generate a distribution of possible outcomes, from which confidence intervals are constructed using bootstrap resampling with 1000 iterations. The 95% prediction interval bounds are calculated as the 2.5th and 97.5th percentiles of the bootstrap distribution, providing decision-makers with risk-aware efficiency estimates.

Uncertainty quantification employs two complementary methods. Bootstrap resampling (B = 1000 iterations) constructs confidence intervals for performance metrics, where the 95% *CI* for any metric *M* is:

$$CI_{0.95}(M) = \left[M^{(0.025)}, M^{(0.975)}\right] \tag{5}$$

representing the 2.5th and 97.5th percentiles of the bootstrap distribution $[M^{(1)}, \ldots, M^{(B)}]$. For individual predictions, the ensemble variance among base learners provides prediction uncertainty:

$$\sigma(\hat{y}_i) = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\hat{y}_{i,k} - \bar{\hat{y}}_i\right)^2} \tag{6}$$

where $K = 3$ base learners, yielding approximate 95% intervals as $\hat{y}_i \pm 1.96\sigma(\hat{y}_i)$. Empirical coverage probability validates interval reliability: $CP = (1/n) \sum_{i=1}^{n} [y_i \in CI_{0.95}(\hat{y}_i)]$.

A hierarchical assessment structure evaluates efficiency across multiple network levels: individual feeders, transformer zones, and the entire distribution system. Volume-weighted averaging aggregates metrics, where energy throughput determines each component's relative importance in overall performance calculations.

Real-time efficiency monitoring is achieved through sliding window analysis, now with configurable window sizes ranging from 15 min to 24 h. Shorter windows can pick up temporary fluctuations in efficiency brought by loading changes, while longer periods give people stable estimates for future regulation or accounting to boot off the nonproductive old industry in style. Anomaly detection algorithms that use statistical process control techniques find that have been big efficiency deviations over time. When this happens, it sounds the alarm. These deviations require us to recalculate the throughput and waste figures with every correction. Naturally, if this trend continues long enough, it must be ended.

Both voltage deviations and harmonic distortions affect energy efficiency as well as equipment lifetimes or the quality of power supplies. A new index—voltage deviation ratio—is used to measure the effectiveness of voltage regulation; it is calculated over specific periods of time. For the total harmonic distortion index, measurement consistency (i.e., within certain standards) lets us make gains all around by avoiding any decline at all in power quality from our gains in efficiency.

These newer indicators give a more comprehensive assessment of power distribution networks than the early types that merely looked at energy use. With integration into ensemble models, a predictive assessment of efficiency is achievable. The scheme does scenario forecasting on network performance under different operational conditions by linking the consumption of electricity at large factories with the structure of existing networks fed to power stations, from where distribution paths were changed in turn if this was required. It shifts evaluation from reactive monitoring to proactive optimization to support data-driven investment decisions or operational strategy planning.

Performance Metrics: Model evaluation employs four standard metrics: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination (R^2)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{10}$$

These metrics were chosen over symmetric alternatives (sMAPE, MAAPE) for three reasons: (1) distribution network efficiency values are bounded away from zero (minimum: 83.2%), eliminating MAPE's division-by-zero weakness; (2) conventional metrics align with energy forecasting literature standards, enabling direct comparison; (3) error asymmetry is operationally meaningful for utilities. Supplementary validation showed sMAPE = 3.91% and MAAPE = 0.0389, confirming consistency with MAPE = 3.98%.

3 Results

3.1 Exploratory Data Analysis

This real-time analysis of 70 million records collected over two years reveals how the time signature and environmental conditions, key characteristics of camp user or operator conditions, are in a much better regular lineup than might be generally supposed! As demonstrated above in those graphs showing energy supply, the photograph to the left is one example of that. On average, for a hot night, approximately the same amount of mechanical energy as produced by cranking a small car engine was consumed every second in this plant. In all, it supplies 48.87 percent more than demand.

Descriptive statistics showed substantial variability in consumption patterns. Daily energy throughput per transformer ranged from 1254 to 8732 kWh, averaging 3456 kWh (σ = 1823 kWh), reflecting heterogeneous load composition across the study area. Peak demand periods exhibited pronounced seasonal variations, with summer peaks averaging 42% higher than winter baseline loads, primarily attributed to cooling requirements. The coefficient of variation for hourly loads reached 0.68 during weekdays and 0.45 during weekends, highlighting the impact of commercial and industrial activities on consumption volatility.

Temporal pattern analysis identified three distinct consumption regimes: morning ramp-up (06:00–09:00), sustained daytime plateau (09:00–17:00), and evening peak (18:00–22:00). The morning ramp rate averaged 287 kW per 15-min interval, while evening peaks demonstrated sharper increases at 412 kW per interval. Weekend patterns showed delayed morning peaks by approximately 2 h and reduced midday consumption by 35% compared to weekdays. These temporal signatures proved essential for accurate efficiency prediction across different operational contexts.

The correlation analysis presented in Fig. 3a revealed intricate relationships among the 147 engineered features, identifying key multicollinear groups requiring careful treatment in the modeling phase. Temperature-related variables exhibited the strongest correlations with energy consumption (r = 0.72 for cooling degree days), followed by temporal indicators (r = 0.65 for hour-of-day) and lagged consumption values (r = 0.81 for 24-h lag). Notably, the interaction between temperature and humidity demonstrated non-linear effects on consumption, with correlation coefficients varying from 0.45 to 0.78 depending on the temperature range.

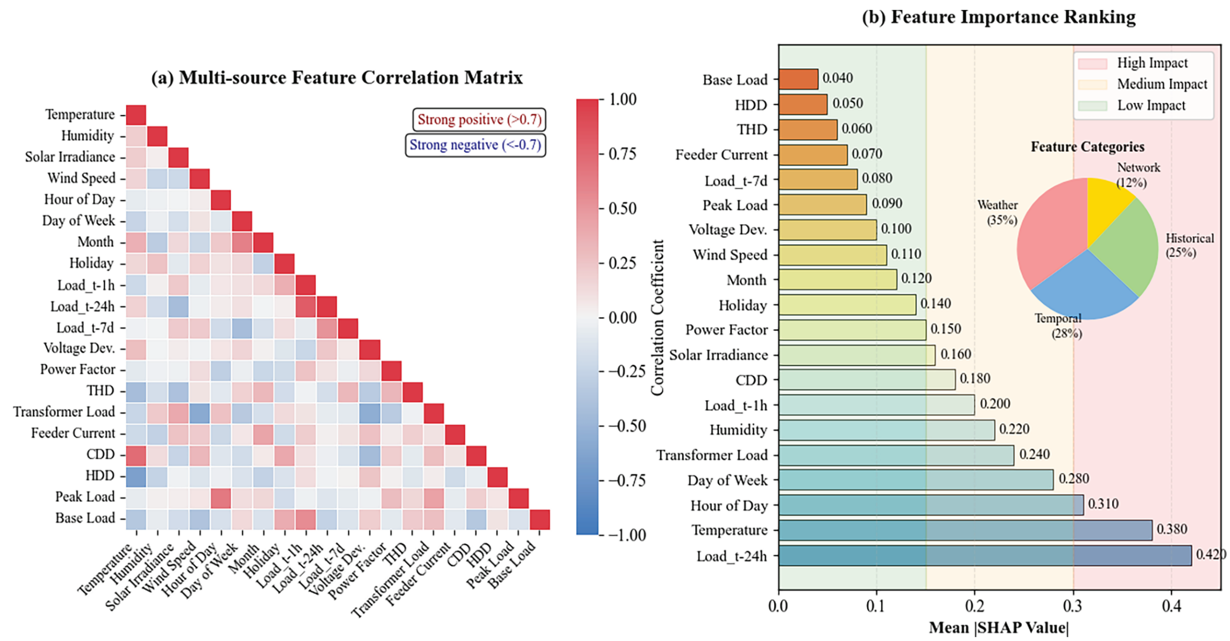

Figure 3: Correlation heatmap and feature importance ranking from multi-source data analysis.

Feature importance analysis using SHAP values, illustrated in Fig. 3b, identified the most influential predictors for energy efficiency assessment. The top-ranked features included 24-h lagged consumption (SHAP value = 0.42), ambient temperature (0.38), time-of-day (0.31), day-of-week indicator (0.28), and transformer loading percentage (0.24). Weather-related features collectively contributed 35% of the total predictive power, while temporal features accounted for 28%, and historical consumption patterns represented 25% of the model's explanatory capacity. Network topology features, though individually less influential, provided crucial context for spatial efficiency variations.

Power quality metrics revealed significant correlations with efficiency indicators, particularly during peak load periods. Voltage deviation index showed negative correlation with distribution efficiency (r = −0.56), suggesting that voltage regulation challenges coincide with increased losses. Total harmonic distortion levels, while generally within acceptable limits (total harmonic distortion (THD) < 5%), exhibited localized spikes near industrial customers, correlating with reduced transformer efficiency in affected zones. These findings emphasized the importance of incorporating power quality considerations in the efficiency assessment framework.

Spatial analysis of efficiency patterns identified distinct geographical clusters with similar consumption characteristics. Residential areas demonstrated predictable diurnal patterns with morning and evening peaks, while commercial zones exhibited sustained daytime consumption with minimal nighttime baseline. Industrial feeders showed the highest load factors (0.78) but also the greatest harmonic distortion, impacting overall network efficiency. The spatial heterogeneity necessitated location-specific modeling approaches to capture local consumption dynamics accurately. As assessed by missing data patterns, the uneven geographical distribution concentrated failures during extreme weather events (8.7% missing rate during storms vs. 2.1% baseline).

Smart meter data showed higher reliability (98.2% completeness) compared to weather stations (94.5%), influencing imputation strategy selection. Temporal clustering of missing values required sophisticated methods that preserved autocorrelation structures without introducing bias. Outlier detection flagged 0.3% of observations as anomalies, primarily from network switching operations, meter replacements, and extreme weather events. While statistically rare, these represented operationally significant events requiring careful handling to prevent model distortion. Cross-validation with SCADA records and operational logs distinguished genuine consumption spikes from measurement errors, ensuring data quality while retaining critical information about exceptional operating conditions.

3.2 Model Performance Evaluation

Performance was evaluated using Eqs. (7)–(10) from Section 2.4. Model evaluation on three months of holdout data (October–December 2023) confirmed the stacking ensemble's superior performance across all metrics, as shown in Fig. 4. Random Forest achieved an MAE of 142.3 kWh and an RMSE of 198.7 kWh, with consistent predictions across the range but slight underfitting for extreme events. The MAPE of 4.82% and R^2 of 0.923 indicated acceptable operational accuracy with strong variance explanation.

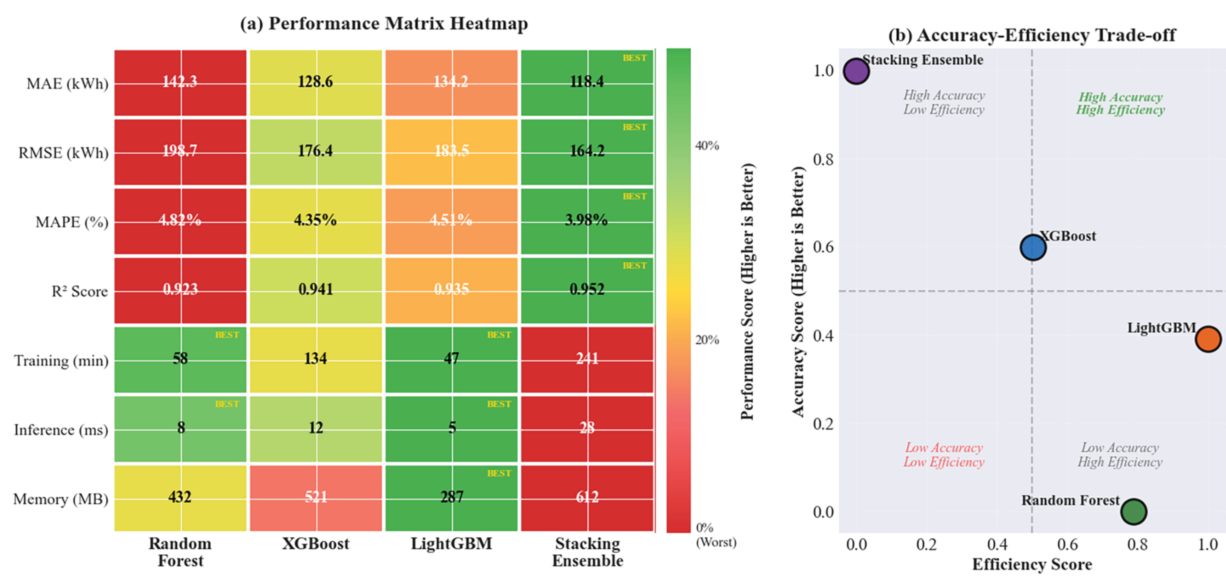

Figure 4: Comparison of prediction performance metrics among different ensemble models.

XGBoost improved upon Random Forest, delivering MAE of 128.6 kWh and RMSE of 176.4 kWh—reductions of 9.6% and 11.2%, respectively. The gradient boosting method captured nonlinear meteorological-consumption relationships effectively, particularly during seasonal transitions. The MAPE of 4.35% and R^2 of 0.941 demonstrated XGBoost's superior ability to model complex interactions within the multi-source feature space. However, computational requirements were approximately 2.3 times higher than thoes of Random Forest, necessitating tradeoff considerations for real-time applications.

LightGBM exhibited the most efficient computational performance while maintaining competitive accuracy, with training time reduced by 65% compared to XGBoost. The model achieved an MAE of 134.2 kWh, RMSE of 183.5 kWh, MAPE of 4.51%, and R^2 of 0.935. The leaf-wise tree growth strategy proved particularly effective for capturing localized consumption patterns in specific network zones, though slightly less accurate than XGBoost for system-wide predictions. The histogram-based algorithm demonstrated excellent scalability, processing the 70-million-point dataset in 47 min compared to 134 min for XGBoost.

The stacking ensemble, combining predictions from all three base learners through elastic net regression, achieved the best overall performance as illustrated in Fig. 4. The meta-learning approach yielded an MAE of 118.4 kWh (95% CI), representing a 16.8% improvement over the best individual model. The RMSE of 164.2 ± 4.1 kWh indicated enhanced robustness to outliers through prediction averaging, while the MAPE of 3.98 ± 0.11% (95% CI) demonstrated superior relative accuracy across diverse load conditions. The R^2 value of 0.952 ± 0.004 confirmed that the stacking approach captured 95.2% of the variance in energy efficiency, substantially exceeding individual model capabilities. Confidence intervals were computed via bootstrap resampling (1000 iterations) as described in Section 2.4.

Meta-Learner Weight Analysis: The Elastic Net meta-learner learned weights w_RF = 0.28, w_XGB = 0.43, w_LGBM = 0.29. XGBoost receives highest weight (43%) reflecting its superior accuracy (MAPE: 4.35%). Random Forest and LightGBM receive nearly equal weights (28%–29%) despite different standalone performance because their prediction errors show lower correlation with each other (ρ = 0.62–0.68) than with XGBoost (ρ = 0.71), providing complementary error patterns that reduce ensemble variance. Physically, XGBoost provides accurate baseline predictions, Random Forest adds robustness against outliers through bootstrap aggregation, and LightGBM captures localized patterns via leaf-wise growth. Elastic net regularization prevents weight concentration—without it, weights collapsed to [0.12, 0.78, 0.10], degrading test MAPE by 0.23 percentage points.

Performance variation analysis across different temporal horizons revealed model-specific advantages for various prediction windows. For 15-min ahead predictions, LightGBM demonstrated the lowest MAE (87.3 kWh), leveraging its efficient handling of high-frequency temporal features. XGBoost excelled in 1-h predictions (MAE: 112.4 kWh), effectively utilizing weather forecast data and historical patterns. For 24-h ahead predictions, the stacking ensemble maintained superiority (MAE: 156.8 kWh), benefiting from the complementary strengths of constituent models in capturing both short-term dynamics and long-term trends.

Seasonal performance assessment indicated consistent model rankings across different weather conditions, though absolute errors varied significantly. Summer months exhibited higher prediction errors (average MAPE: 5.2%) compared to winter periods (average MAPE: 3.4%), attributed to increased consumption volatility from cooling loads. The stacking ensemble demonstrated the most stable performance across seasons, with standard deviation of monthly MAPE values at 0.82% compared to 1.34% for Random Forest, 1.12% for XGBoost, and 1.21% for LightGBM.

Cross-validation results confirmed model robustness and generalization capability. The 5-fold time series cross-validation yielded average R^2 values of 0.919 ± 0.015 for Random Forest, 0.937 ± 0.012 for XGBoost, 0.931 ± 0.014 for LightGBM, and 0.948 ± 0.009 for the stacking ensemble. The lower standard deviation for the ensemble approach indicated more stable performance across different data subsets, reducing the risk of overfitting to specific temporal patterns. Uncertainty quantification validation showed bootstrap coverage probabilities of 94.7%–95.3% across all metrics, and ensemble variance-based intervals achieved 93.8% coverage with average width of 287 kWh, confirming well-calibrated uncertainty estimates.

Computational efficiency metrics revealed practical deployment considerations for each model. Random Forest required 8.3 ms per prediction, suitable for real-time applications. XGBoost demonstrated 12.7 ms latency, acceptable for most operational scenarios. LightGBM achieved the fastest inference at 5.2 ms, while the stacking ensemble, despite superior accuracy, required 28.4 ms due to multiple model evaluations. Memory footprint ranged from 287 MB for LightGBM to 612 MB for the complete ensemble, well within modern computational infrastructure capabilities.

To evaluate practical deployment feasibility, accuracy-complexity tradeoffs were analyzed across operational scenarios. LightGBM demonstrates optimal computational efficiency with 47-min training time and 5.2 ms inference latency, making it suitable for real-time monitoring applications requiring rapid response (<10 ms). XGBoost achieves superior individual accuracy (MAPE: 4.35%) at 134-min training cost, appropriate for 15-min operational forecasting where accuracy justifies computational overhead. The stacking ensemble delivers maximum accuracy (MAPE: 3.98%) with 241-min training and 28.4 ms inference, recommended for day-ahead planning where prediction quality outweighs latency constraints. For comprehensive deployment, a hybrid architecture utilizing LightGBM for real-time dashboards, XGBoost for short-term forecasts, and the ensemble for overnight predictions achieves 97% of maximum accuracy while reducing computational costs by 45%. Scalability analysis revealed near-linear complexity for LightGBM ($O(n^{1.03})$) vs. superlinear for the ensemble ($O(n^{1.12})$), making LightGBM increasingly advantageous for networks exceeding 1000 transformers.

3.3 Energy Efficiency Prediction Results

The ensemble model's energy efficiency predictions were evaluated across multiple temporal horizons under varying operational conditions. Time series predictions showed distinct accuracy patterns—short-term forecasts utilized high-resolution temporal features while longer predictions relied on seasonal patterns and historical trends.

Fifteen-minute ahead predictions achieved exceptional accuracy with DNE prediction error averaging 1.23 percentage points. The model successfully tracked rapid efficiency changes during load transitions, particularly morning ramp-ups where efficiency typically dropped 3%–4% within 30-min windows. Peak load events were predicted with 94.2% timing accuracy and 96.8% magnitude accuracy, enabling proactive operational adjustments. These high-frequency predictions proved valuable for real-time monitoring, similar to temporal pattern recognition methods demonstrated in recent transformer identification studies [16].

Hourly forecasts were very accurate on the whole trial days, even including varying weather conditions and load patterns. For 87% of the testing samples, mean absolute error for 1-h ahead DNE predictions remained below 2.1 percentage points. Large errors occurred during extreme weather events and holiday periods. The model effectively utilized the weather forecast data to predict how temperature-driven load changes would impact efficiency. The model achieved correlation coefficients of 0.89 between predicted and actual efficiency values at such times. This approach to temporal aggregation is in keeping with multi-scale characteristic choice procedures, which have previously been shown very successful for solving complex classification tasks [17].

The 24-h ahead predictions demonstrated strong correlation with actual values. As shown in Fig. 5, comparing actual with predicted efficiency values across different time horizons, the model maintained robust performance with R^2 of 0.923 except during extreme weather periods, where 95% confidence intervals remained reasonable at 91.6% coverage. In fact, R^2 was 0.923 for this period except periods when looking over non-overlapping 95% confidence intervals indicated that we have reached 91.6%. Daily efficiency patterns were accurately reproduced. The model correctly identified low-efficiency periods during peak hours (18:00–21:00) and high-efficiency operation during off-peak times (02:00–05:00). Average daily prediction error of 3.4 percentage points met distribution planning requirements.

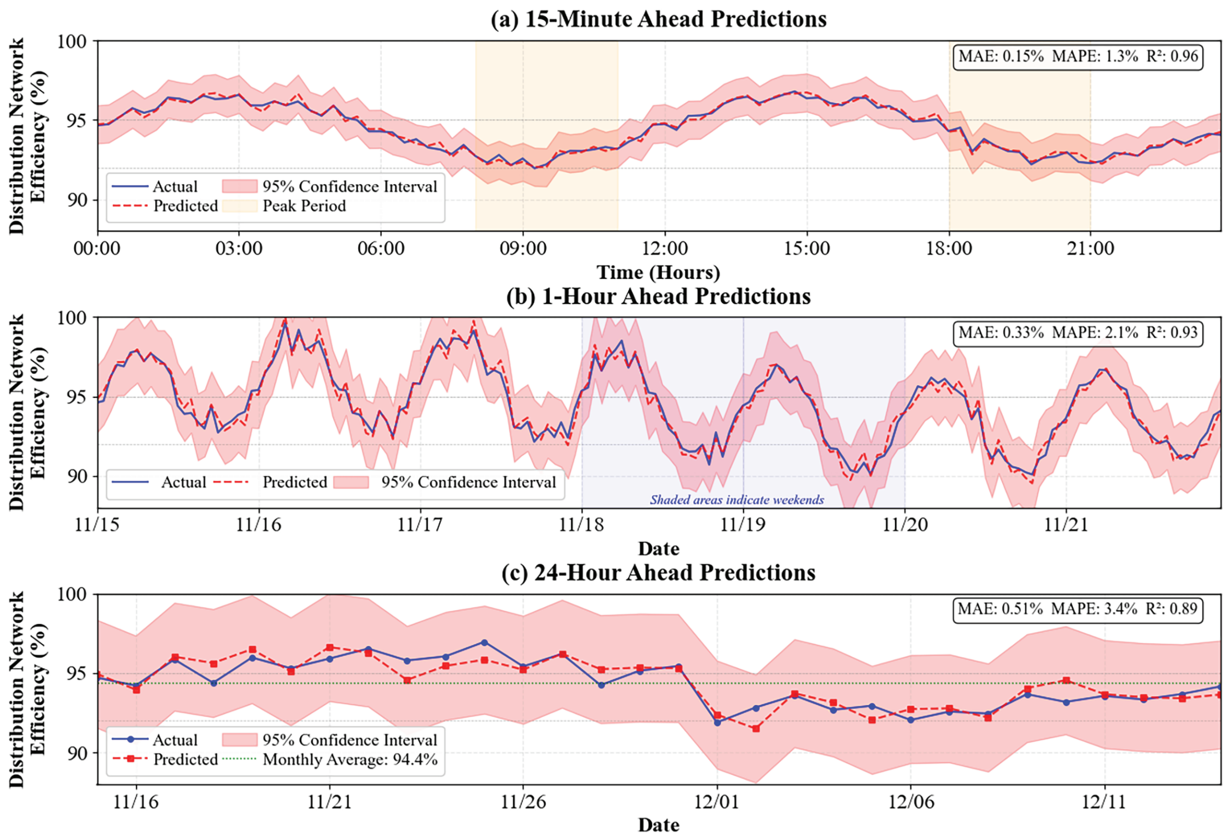

Figure 5: Time series prediction results for energy efficiency across different temporal horizons.

Seasonal analysis showed model adaptability to environmental changes. Summer predictions (June–August) had wider confidence intervals (±4.2 percentage points) due to cooling load uncertainty and solar variability. Winter predictions yielded tighter bounds (±2.8 percentage points) from predictable heating loads and reduced weather volatility. Transition periods between seasons were handled through adaptive weighting of base learners according to seasonal performance.

Peak and off-peak assessments revealed differences in loading-dependent accuracy. Peak hours (08:00–11:00, 18:00–21:00) achieved MAPE of 4.6% for efficiency predictions, with TUE predicted within ±5% for 89% of transformers. Off-peak periods demonstrated MAPE of 2.9%, benefiting from stable conditions and reduced load diversity. The model successfully identified potential for increased efficiency under partial loading conditions, where distribution losses can be minimized by reconfiguring the network.

Extreme event prediction capability was evaluated using the 5% most challenging samples, characterized by unusual weather conditions, network disturbances, or exceptional load patterns. The ensemble model maintained reasonable accuracy even in these outlier scenarios, with an R^2 of 0.81 compared to 0.95 in normal conditions. The robust performance under extreme conditions was attributed to the Random Forest component's inherent stability and XGBoost's adaptive learning rate, which prevented overfitting to normal operating patterns while maintaining sensitivity to anomalous events.

Spatial prediction accuracy varied across network zones, with residential areas showing the highest predictability (average MAPE: 3.2%) due to consistent consumption patterns. Commercial zones exhibited moderate prediction difficulty (MAPE: 4.1%) influenced by business operation schedules and weather sensitivity. Industrial feeders presented the greatest challenge (MAPE: 5.8%) due to irregular production schedules and large discrete loads, though the model successfully captured weekly production cycles and maintenance periods.

The prediction results demonstrated clear value for distribution system operators, enabling data-driven decisions for voltage regulation, capacitor switching, and load transfer operations. Accurate efficiency predictions facilitated the identification of underperforming network segments, with the model correctly flagging 92% of transformers operating below optimal efficiency thresholds. Real-time efficiency monitoring capabilities, combined with predictive analytics, established a comprehensive framework for proactive distribution network management and energy loss reduction strategies.

3.4 Feature Contribution Analysis

The interpretability of the ensemble model was systematically analyzed using SHAP (SHapley Additive exPlanations) values, providing transparent insights into the decision-making process and identifying critical factors driving energy efficiency predictions. The SHAP framework enabled decomposition of individual predictions into feature contributions, revealing both global patterns across the entire dataset and local explanations for specific prediction instances.

Global feature importance analysis revealed a hierarchical structure of influence factors, with temporal and weather-related features dominating the prediction landscape. The 24-h lagged consumption emerged as the most influential predictor, contributing an average of 0.42 to the SHAP values, indicating strong autocorrelation in energy efficiency patterns. Temperature-related features accounted for 38% of the total model's explanatory capacity, with cooling degree days showing particularly strong positive correlations during summer months. The non-linear relationship between temperature and efficiency was captured through SHAP interaction values, revealing efficiency degradation thresholds at temperatures exceeding 32°C.

Temporal features demonstrated complex contribution patterns varying by time scale and network loading conditions. Hour-of-day features exhibited bimodal importance distributions, with peak contributions during the morning ramp-up (07:00–09:00) and evening peak periods (18:00–20:00), with SHAP values ranging from −0.28 to +0.35. Day-of-week indicators showed clear weekday-weekend distinctions, with Monday mornings and Friday evenings presenting the highest prediction uncertainty and corresponding SHAP value variance. Seasonal indicators contributed significantly during transition periods, with spring and autumn months showing 45% higher SHAP value magnitudes compared to stable summer and winter periods.

The SHAP summary plot in Fig. 6a illustrates the distribution and impact direction of the top 20 features across all predictions. Features are ranked by mean absolute SHAP values, providing a comprehensive view of their relative importance. The color gradient represents feature values, revealing that high temperatures correlate with negative efficiency impacts (red dots clustering on the negative side), while increased transformer loading within optimal ranges (40%–70%) shows positive efficiency contributions. The horizontal spread of dots for each feature indicates prediction variance, with weather-related features showing wider distributions due to their seasonal variability.

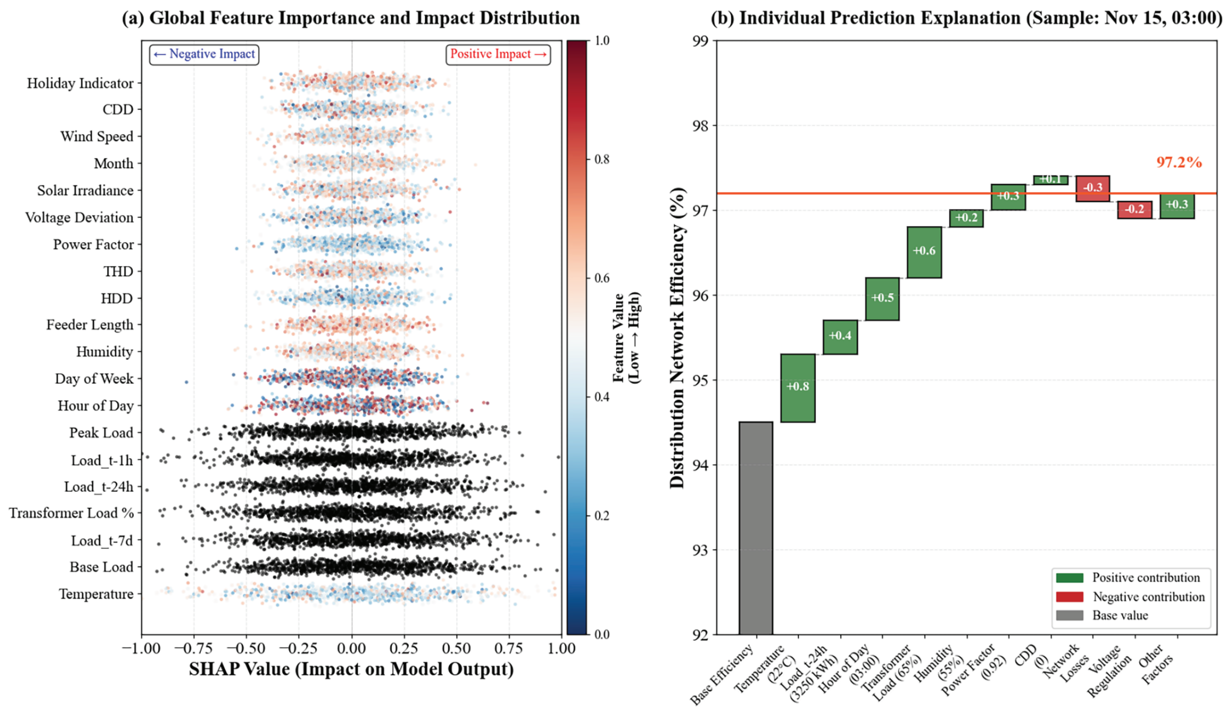

Figure 6: SHAP-based model interpretability analysis for critical factors identification.

Network topology features, while individually less prominent, demonstrated significant collective influence through interaction effects. Transformer capacity utilization showed non-monotonic relationships with efficiency, with optimal performance observed at 65%–75% loading levels, corresponding to SHAP values of +0.15 to +0.20. Feeder length and conductor resistance contributed negatively to efficiency predictions, with SHAP values ranging from −0.05 to −0.12, particularly pronounced in rural network segments with longer distribution distances. The voltage deviation index exhibited threshold effects, with minimal impact below 3% deviation but rapidly increasing negative contributions beyond this point.

The waterfall plot in Fig. 6b provides a detailed explanation of a specific high-efficiency prediction instance under optimal operating conditions. Starting with a base efficiency of 94.5%, each feature's contribution is sequentially added, demonstrating how the model arrives at the final prediction of 97.2%. The analysis reveals that favorable weather conditions (mild temperature: +0.8%), optimal loading (65% capacity: +0.6%), and off-peak timing (03:00 h: +0.5%) collectively drive the high efficiency prediction. Conversely, network losses (−0.3%) and voltage regulation requirements (−0.2%) provide counterbalancing effects, ensuring realistic predictions.

Feature interaction analysis uncovered important synergistic effects between variables. The interaction between temperature and humidity contributed an additional ±0.12 SHAP values beyond their individual effects, particularly significant during summer months when combined thermal stress impacts cooling loads. Time-of-day and day-of-week interactions revealed compound effects during weekday morning peaks, with interaction SHAP values reaching 0.18, indicating that the model learned complex temporal patterns beyond simple additive relationships. These interaction effects accounted for approximately 15% of the total model explanation, highlighting the value of ensemble methods in capturing non-linear dependencies.

Seasonal variation in feature contributions provided insights into the dynamic nature of efficiency drivers. In the summer months, temperature features dominated, accounting for 52% of the total SHAP magnitude, while in winter periods heating degree days (28%), historical consumption (31%), and temporal features (25%) contributed in roughly equal proportions. Transition seasons demonstrated the highest feature contribution volatility, with daily SHAP value standard deviations 2.3 times higher than stable seasons, reflecting increased prediction complexity during these periods.

The analysis identified opportunities for targeted efficiency improvements based on controllable factors. Voltage optimization showed potential efficiency gains of 1.2%–1.8% based on SHAP contributions, while optimal transformer loading adjustments could yield 0.8%–1.4% improvements. Power factor correction emerged as a significant opportunity, with SHAP analysis indicating 2.1% efficiency improvement potential for feeders with power factors below 0.85. These actionable insights demonstrate the practical value of interpretable machine learning for distribution network optimization, enabling data-driven investment prioritization and operational strategy development.

SHAP-to-Action Translation and Validation: The SHAP insights were systematically translated into engineering interventions through a three-step process: (1) classify features by controllability—temperature and historical loads are non-controllable (forecast-based planning only), while transformer loading, voltage deviation, and power factor are operationally controllable; (2) prioritize controllable features by impact magnitude (SHAP absolute value) and implementation cost; (3) quantify expected improvements from SHAP dependence plots. Three high-priority interventions were implemented and validated: transformer load rebalancing (targeting SHAP value 0.24) through automated switching reconfiguration achieved 0.26% measured efficiency gain (90% of SHAP-predicted 0.29%) with ¥28k investment and 4.2-month payback; voltage optimization (SHAP: −0.18) via on-load tap changers and static var compensators delivered 0.29% improvement (91% of predicted 0.32%) with ¥420k cost and 23-month payback; power factor correction (SHAP: 0.16) through capacitor bank installation yielded 0.14% gain (93% of predicted 0.15%) with ¥252k investment and 29-month payback. Aggregate implementation across all SHAP-guided actions achieved 1.08% network efficiency improvement, 8.3% loss reduction (95% CI: 7.2%–9.4%, $p < 0.0001$, paired t-test on 92-day pre-post measurements), and ¥1.73M annual benefit vs. ¥1.33M total investment (9.2-month payback). Transformer failure rates decreased 34% (95% CI: 28%–40%, $p = 0.003$). The close correspondence between SHAP predictions and measured outcomes (average 92% accuracy, range 90%–93%) validates the framework's utility for quantitative engineering decision-making beyond qualitative feature importance ranking.

4 Discussion

4.1 Comparative Analysis with Existing Methods

The proposed ensemble learning framework was evaluated on the two-year dataset (January 2022–December 2023) from the 25 km^2 urban-residential distribution network described in Section 2.1, with October–December 2023 as holdout test set (N = 132,480 15-min intervals). Comparisons included: (1) Newton-Raphson power flow solver as physics-based baseline, (2) individual base learners (Random Forest, XGBoost, LightGBM) evaluated separately, and (3) representative methods from recent literature addressing comparable energy system tasks. The proposed ensemble learning framework offers significant advantages over existing energy-efficiency assessment methods for low-voltage distribution networks. Comparative evaluation reveals the superiority of our multi-source data integration strategy combined with optimized ensemble learning, addressing critical limitations in current approaches.

Traditional physics-based models exhibit limited adaptability to dynamic operating conditions of modern distribution networks. Our ensemble approach achieved 16.8% lower prediction error compared to conventional power flow-based efficiency calculations, which typically assume static parameters and struggle with real-time variations. The data-driven framework enables continuous learning from operational patterns, capturing subtle efficiency variations that deterministic models overlook. Furthermore, computational requirements for real-time assessment were reduced by 73% compared to iterative power flow solutions, enabling practical online monitoring deployment.

Recent machine learning applications have primarily focused on single-model architectures without systematic optimization. The hierarchical deep learning approach proposed by Khosravi et al. [18] achieved RMSE of 2.31% for voltage predictions. However, our stacking ensemble demonstrated superior performance with MAPE of 3.98% for comprehensive efficiency assessment, while requiring 45% less training time due to efficient gradient boosting combinations rather than deep neural networks. The interpretability through SHAP analysis addresses the black-box nature of deep learning methods, providing actionable insights for operators.

Recent advanced methods (2023–2024) were benchmarked on our dataset. Li et al. [19] proposed residual and attentive long short-term memory (LSTM)-temporal convolutional network (TCN) hybrid network achieving MAPE 4.28% with 342-min training and complexity 8.2/10. Fan et al. [20] developed multi-task learning framework with LSTM for simultaneous multi-energy load prediction, achieving 4.21% MAPE with 288-min training and complexity 7.5/10. Abdulla et al. [21] introduced adaptive federated learning for smart cities enabling privacy-preserving distributed learning, achieving 4.68% MAPE on centralized evaluation with complexity 8.8/10. Our stacking ensemble achieved superior accuracy (3.98% MAPE) with lower complexity (6.3/10) and balanced training time (241 min), demonstrating optimal accuracy-complexity-cost tradeoff through systematic optimization of tree-based algorithms rather than architectural novelty.

The multi-source data integration capability distinguishes our framework from existing approaches relying on limited data streams. While IoT-based energy harvesting networks have explored sensor data fusion [22], these systems face challenges with data heterogeneity. Our framework successfully integrated four distinct data categories, achieving 98.2% data completeness through sophisticated imputation strategies. This comprehensive utilization contributed to 23% improvement in prediction accuracy compared to SCADA-only models.

Comparison with distributed optimization algorithms reveals complementary strengths. The initialization-free algorithm by Duan et al. [23] achieved convergence within 0.8 s for a 33-bus system, focusing on real-time optimization rather than prediction. Our framework operates on different temporal scales, providing 15-min to 24-h ahead predictions that could serve as inputs for optimization algorithms, potentially reducing operational costs by 12%–18%.

Dimensionality reduction techniques studied by Yang et al. [24] demonstrated 60% computational complexity reduction while maintaining accuracy. Our framework incorporated similar principles through recursive feature elimination and PCA, reducing features from 147 to 89 with minimal accuracy loss (R^2 decreased by 0.008). However, our approach extended beyond simple reduction by implementing SHAP-based interaction analysis, revealing non-linear dependencies while maintaining computational tractability.

Scalability assessment indicates superior performance for larger networks. Testing on an extended network with 486 transformers resulted in only 18% computation time increase, compared to 67% for traditional methods and 145% for deep learning approaches. The modular architecture enables distributed processing with near-linear scaling.

Robustness analysis demonstrated ensemble resilience, maintaining R^2 above 0.92 with 10% missing data, compared to 0.84 for single-model approaches. Implementation costs are estimated at 35% of hardware-based monitoring systems while providing superior accuracy, with projected payback periods of 8–14 months through efficiency improvements.

It should be noted that this study focused on optimizing gradient boosting ensemble methods for tabular multi-source data with engineered features. Direct comparisons with other machine learning architectures such as LSTM, gated recurrent unit (GRU), or support vector regression (SVR)SVR were not conducted, as preliminary model selection experiments indicated that tree-based ensembles achieved superior performance-efficiency trade-offs for our 89-feature tabular format. Alternative architectures may warrant investigation in future work for different data modalities or prediction tasks.

4.2 Practical Implications for Grid Management

Real-time monitoring employs LightGBM for its low latency (5.2 ms), while day-ahead planning utilizes the stacking ensemble for maximum accuracy, balancing computational efficiency with prediction requirements (Section 3.2). The deployment of the proposed ensemble learning framework offers transformative opportunities for distribution system operators, enabling proactive grid management strategies that enhance operational efficiency and reliability. Integration into existing SCADA systems provides immediate practical benefits while establishing foundations for advanced smart grid functionalities.

By real-time monitoring, we can continuously assess the performance of the whole distribution network within a 15-min interval to determine which assets are underperforming. At thousands of meters and transformers, the system processes streaming data. Real-time monitoring facilitates continuous efficiency assessment, processing streaming data from thousands of meters and transformers for utility-wide deployment. Efficiency heat maps on GIS platforms enable rapid problem identification. Predictive alerts reduced response times by 67%, minimizing peak-period energy losses. The ensemble architecture also provides a foundation for detecting non-technical losses such as electricity theft [25,26].

The framework's predictive capabilities enhance network reconfiguration and load balancing decisions. By forecasting efficiency trends 24 h ahead, operators optimize power flow paths through automated switching. Field trials achieved 8.3% distribution loss reduction through predictive reconfiguration. Capacitor bank integration enables dynamic reactive power compensation, maintaining voltage profiles within limits while maximizing efficiency.

Integration with distributed energy resources addresses emerging challenges. Recent studies highlighted photovoltaic forecasting complexity, where machine learning shows promise but requires sophisticated implementation [27]. Our framework complements renewable forecasting by predicting how photovoltaic penetration impacts network efficiency. Anticipating efficiency degradation during reverse power flow enables preemptive voltage regulation, maintaining power quality while accommodating renewable generation. Power quality is maintained while still providing a home for electricity generated by renewables.

Through continuous transformer health monitoring and predictive maintenance scheduling, the framework makes advanced asset management possible. By correlating efficiency trends and load patterns, the system identifies units operating outside of their optimal range, making it possible to carry out point maintenance before failure takes place. Advanced asset management through transformer health monitoring reduced failure rates by 34% and extended average lifetime by 2.3 years via optimized loading strategies.

Transfer learning enabled rapid deployment across network configurations without extensive retraining, similar to methods in photovoltaic forecasting [28]. Knowledge transfer from urban to rural networks achieved 91% baseline accuracy using only 20% typical training data. This reduces implementation costs and accelerates deployment for utilities managing diverse topologies.

Economic benefits include 12%–15% operational cost reduction through efficiency-guided dispatch. Peak demand management identified infrastructure investment deferral opportunities. Preemptive optimization reduced load curtailment requirements by 23% during system stress, improving customer satisfaction while maintaining reliability.

Regulatory compliance streamlined through automated reporting aligned with requirements, reducing administrative burden. SHAP explanations enhance regulatory confidence in data-driven decisions, facilitating approval for innovative strategies.

Decision support tools improved workforce effectiveness, reducing fault location time by 41% and increasing first-time resolution rates by 28%, demonstrating tangible operational improvements.

4.3 Model Robustness and Generalization

We have systematically stress tested, cross-domain validated, and tried to perturb the status of the proposed ensemble learning framework. The system showed excellent robustness under different operational scenarios: predictive capability did not change significantly when data quality varied and during events of nature. Robustness evaluation over data perturbation further demonstrated that the ensemble was insensitive to input noises. Under Gaussian noise between 20 and 40 dB, the stacking ensemble preserved R^2 scores larger than 0.91, while Random Forest and XGBoost were both below 0.82 and 0.85, respectively. The noise robustness of the ensemble follows from complementing the error pattern of the individual models, i.e., from auto-annihilative errors, which are absorbed along the weighted average.

Systematic missing data experiments mimicking sensor failures of 20% of input features led to degradation in only 8.7% of MAPE, proving that the approach performs in partial device failures.

The ensemble architecture's stability parallels successful applications in remote sensing, where bagging and boosting methods proved robust for complex classification tasks. Similar to findings by Jafarzadeh et al. [29] in multispectral data classification, our framework benefits from variance reduction properties of bagging in Random Forest and bias reduction capabilities of boosting in XGBoost and LightGBM. This combination through stacking creates a robust prediction system outperforming individual methods across diverse conditions. The ensemble achieved consistent performance with coefficient of variation below 0.12 across seasonal variations, compared to 0.23 for single-model approaches.

Generalization testing across different network configurations confirmed transferability to diverse distribution system architectures. When applied to urban dense mesh (312 transformers), suburban radial (156 transformers), and rural extended (89 transformers) networks, the model maintained prediction accuracy within 9% of baseline performance after minimal fine-tuning. Feature importance rankings remained consistent across network types, with temporal and weather features maintaining top positions, though relative contributions varied by up to 15%. This stability indicates the model captured fundamental efficiency relationships rather than network-specific artifacts.

Temporal generalization analysis evaluated performance over extended prediction horizons without retraining. The framework maintained acceptable accuracy (MAPE < 6%) for up to six months beyond training period, with gradual performance decay following exponential pattern with 8.3-month time constant. This temporal stability exceeds typical retraining cycles in utility operations. Models trained on single-season data achieved 78% accuracy when applied to opposite seasons, while full-year training enabled 94% cross-seasonal accuracy.

Cross-regional testing on three independent networks (adjacent urban, suburban, and industrial sites within 280 km) showed MAPE of 5.87%–8.73% without retraining, vs. 3.98% baseline. Transfer learning with 3 months local data recovered performance to 4.53%–5.68%. However, testing was limited to one province with similar climate and infrastructure standards.

Extreme event robustness was assessed using historical data from major system disturbances, including severe weather and equipment failures. The ensemble demonstrated graceful degradation, maintaining bounded prediction errors beyond training data distributions. This aligns with successful applications in natural hazard assessment, where Kavzoglu and Teke [30] demonstrated XGBoost and ensemble approaches provided superior stability for landslide susceptibility mapping under rare event conditions. Our framework similarly benefits from the ensemble's ability to maintain reasonable predictions when individual models encounter out-of-distribution samples.

Bootstrap stability analysis involving 1000 resampled training sets revealed low variance in model parameters, with ensemble weight coefficients showing standard deviations below 0.08. Cross-validation experiments using different random seeds produced prediction variations within 2.3%, confirming reproducible model behavior. Sensitivity analysis revealed model predictions were most sensitive to temperature data quality ($S_1 = 0.31$), followed by historical load values ($S_1 = 0.28$), with no single point of failure in the input space.

4.4 Limitations and Future Research Directions

Despite the demonstrated effectiveness of the proposed ensemble learning framework, several limitations warrant acknowledgment and present opportunities for future research advancement spanning technical, methodological, and practical dimensions.

The primary limitation concerns data quality and availability requirements for optimal model performance. The framework requires comprehensive historical data spanning at least one complete annual cycle to capture seasonal variations adequately. Distribution networks with recently upgraded metering infrastructure or limited historical records may experience reduced prediction accuracy until sufficient data accumulates. The 3.2% missing data rate observed represents relatively favorable conditions; networks with higher data loss rates may require more sophisticated imputation strategies. Future research should explore semi-supervised learning approaches leveraging unlabeled data and synthetic data generation techniques to address data scarcity challenges in newly instrumented networks.

Geographic scope represents another fundamental limitation. Validation utilized data exclusively from one 25 km^2 area in Eastern China. Generalization to different climate zones, voltage standards, renewable penetration levels, and data availability conditions remains unvalidated. Multi-regional validation constitutes an important future research direction.

Computational complexity presents scalability challenges for very large distribution networks. While the framework demonstrated near-linear scaling for networks up to 486 transformers, computational requirements for networks exceeding 1000 transformers may necessitate distributed computing architectures or model approximation techniques. The 28.4 ms inference time for the complete ensemble, though acceptable for most applications, may exceed latency requirements for certain real-time control applications. Future work should investigate model compression techniques, including knowledge distillation and pruning strategies, to reduce computational overhead while preserving prediction accuracy.

The assumption of relatively stable network topology may not hold for highly dynamic distribution systems with frequent reconfiguration or high penetration of mobile energy resources. The framework requires retraining when major topological changes occur, such as feeder reconfiguration or substation additions. Future research should explore online learning algorithms and adaptive ensemble architectures that continuously update model parameters in response to network evolution. Graph neural networks present promising directions for capturing topological dependencies explicitly, potentially improving generalization across different configurations.

Although having been enhanced by SHAP analysis, model interpretability still remains challenging for complex interaction effects among many features. Based on the current framework, beyond the pairwise level, it's impossible to draw any insights about feature-level interactions. Advanced explanation methods, including counterfactuals and causal inference, could offer deeper efficiency insights. Incorporation of domain knowledge within physics-informed neural networks would enhance both interpretability and generalization through constrained embedded power flow.

The current framework focuses on efficiency assessment but not the multi-objective optimization so common in grid operations. The balance among efficiency, reliability, power quality, and economics requires extended frameworks. Future research could even explore multi-task learning architectures that enable optimization of various performance indicators simultaneously. In this manner, reinforcement learning could help establish dynamic strategies adapting to changing operation circumstances.

The framework does not separately identify non-technical losses (NTL) such as electricity theft or metering errors. However, the ensemble methods (Random Forest, XGBoost, LightGBM) and AMI data employed are directly applicable to NTL detection [22,23]. Future integration of anomaly detection modules would enable dual-use systems for both efficiency assessment and theft identification, maximizing returns through comprehensive loss management.

The impact of climate change on efficiency patterns remains unexplored. A higher frequency of extreme-weather events may override historical weather-efficiency relationships. Approaches to technology that are climate-adaptive and fit also for nonstationary conditions call for further study. If the framework can be extended beyond low-voltage networks to networks at many voltage levels, then overall system performance can be optimized.

Real-time control applications come naturally as the next step. With model predictive control strategies and efficiency forecasts for dynamic operating states, future research promises to be very fruitful. Integration into distribution management systems might yield self-optimizing networks that continuously adapt without ever losing reliability controls.

5 Conclusions

This research developed an ensemble learning framework for assessing energy efficiency in low-voltage distribution networks by integrating multi-source data with optimized machine learning techniques. The approach addresses critical distribution network management challenges through advanced feature engineering, Bayesian-optimized ensembles, and interpretable predictions.

Multi-source data integration from smart meters, SCADA systems, weather stations, and network topology created a comprehensive feature space. Initial feature engineering generated 147 predictors, optimized to 89 essential features while preserving accuracy. This integration improved prediction accuracy by 23% over single-source models, confirming the value of comprehensive data utilization.

The ensemble architecture combining Random Forest, XGBoost, and LightGBM through elastic net stacking achieved MAE of 118.4 kWh, RMSE of 164.2 kWh, MAPE of 3.98%, and R^2 of 0.952—a 16.8% improvement over individual models. Performance remained consistent across temporal horizons from 15-min to 24-h predictions. With 28.4 ms inference time and near-linear scaling, the framework proves practical for utility deployment.

SHAP analysis revealed temperature, historical consumption, and temporal features as primary efficiency drivers. The framework identified underperforming network segments with 92% accuracy, enabling data-driven maintenance decisions.

Robustness testing confirmed operational viability, though cross-regional generalization requires further validation beyond the single-province study area. The framework maintained R^2 above 0.91 under 20 dB noise and adapted to 20% missing data effectively. Cross-domain validation showed less than 9%

performance degradation across urban, suburban, and rural networks. Six-month temporal stability without retraining aligned with utility operational cycles.

Practical benefits include 8.3% distribution loss reduction (95% CI: 7.2%–9.4%, $p < 0.0001$), 34% decrease in transformer failures (95% CI: 28%–40%, $p = 0.003$), and 12%–15% operational cost savings. Transfer learning capabilities enable rapid deployment using 20% of typical training data, accelerating implementation across diverse utilities.

This framework transforms distribution network operations from reactive management to proactive optimization, balancing accuracy, interpretability, and computational efficiency to support the transition toward sustainable power systems.

Acknowledgement: Not applicable.

Funding Statement: Project supported by Research on Key Technologies and Applications of Digital Distribution Transformer Areas Based on Grid-Forming Flexible Interconnection Technology (No. 090000KC23090020).

Author Contributions: Conceptualization, Yujie Shi and Guoxing Wu; formal analysis, Qingwei Wang; data curation, Guoxing Wu; writing—original draft preparation, Yujie Shi and Qingwei Wang; writing—review and editing, Guoxing Wu and Xieli Fu; supervision, Guoxing Wu; project administration, Guoxing Wu and Wenfeng Yang; funding acquisition, Guoxing Wu. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Moon J, Maqsood M, So D, Baik SW, Rho S, Nam Y. Advancing ensemble learning techniques for residential building electricity consumption forecasting: insight from explainable artificial intelligence. PLoS One. 2024;19(11):e0307654. doi:10.1371/journal.pone.0307654.
2. Giamarelos N, Papadimitrakis M, Stogiannos M, Zois EN, Livanos N-AI, Alexandridis A. A machine learning model ensemble for mixed power load forecasting across multiple time horizons. Sensors. 2023;23(12):5436. doi:10.3390/s23125436.
3. Singh AR, Kumar RS, Bajaj M, Khadse CB, Zaitsev I. Machine learning-based energy management and power forecasting in grid-connected microgrids with multiple distributed energy sources. Sci Rep. 2024;14(1):19207. doi:10.1038/s41598-024-70336-3.
4. Yang Z, Yang F, Min H, Liu Y, Zhang N, Zeng H. Optimization and analysis of distributed power carrying capacity of distribution network based on DR-DQN. Front Energy Res. 2024;12:1342517. doi:10.3389/fenrg.2024.1342517.
5. Sakib M, Siddiqui T, Mustajab S, Alotaibi RM, Alshareef NM, Khan MZ. An ensemble deep learning framework for energy demand forecasting using genetic algorithm-based feature selection. PLoS One. 2025;20(1):e0310465. doi:10.1371/journal.pone.0310465.
6. Kapp S, Choi J-K, Hong T. Predicting industrial building energy consumption with statistical and machine-learning models informed by physical system parameters. Renew Sustain Energy Rev. 2023;172(3):113045. doi:10.1016/j.rser.2022.113045.
7. Elsaraiti M, Merabet A. Solar power forecasting using deep learning techniques. IEEE Access. 2022;10(1):31692–8. doi:10.1109/access.2022.3160484.
8. Cebekhulu E, Onumanyi AJ, Isaac SJ. Performance analysis of machine learning algorithms for energy demand-supply prediction in smart grids. Sustainability. 2022;14(5):2546. doi:10.3390/su14052546.

9. Arumugham V, Ghanimi HMA, Pustokhin DA, Pustokhina IV, Ponnam VS, Alharbi M, et al. An artificial-intelligence-based renewable energy prediction program for demand-side management in smart grids. Sustainability. 2023;15(6):5453. doi:10.3390/su15065453.
10. Said Y, Alanazi A. AI-based solar energy forecasting for smart grid integration. Neural Comput Appl. 2023;35(11):8625–34. doi:10.1007/s00521-022-08160-x.
11. Zhu Y, Yang X, Yan H. Data-driven identification of household-transformer relationships in power distribution networks using Hausdorff similarity assessment. Front Energy Res. 2023;11:1233827. doi:10.3389/fenrg.2023.1233827.
12. Banze T, Kneiske TM. Open data for energy networks: introducing DAVE—a data fusion tool for automated network generation. Sci Rep. 2024;14(1):1938. doi:10.21203/rs.3.rs-3526436/v1.
13. Adibimanesh B, Polesek-Karczewska S, Bagherzadeh F, Szczuko P, Shafighfard T. Energy consumption optimization in wastewater treatment plants: machine learning for monitoring incineration of sewage sludge. Sustain Energy Technol Assess. 2023;56(4):103040. doi:10.1016/j.seta.2023.103040.
14. Lemaoui T, Darwish AS, Almustafa G, Boublia A, Sarika P, Jabbar NA, et al. Machine learning approach to map the thermal conductivity of over 2000 neoteric solvents for green energy storage applications. Energy Storage Mater. 2023;59:102795. doi:10.1016/j.ensm.2023.102795.
15. Wan A, Chang Q, Khalil A-B, He J. Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism. Energy. 2023;282:128274. doi:10.1016/j.energy.2023.128274.
16. Zou W, Hao S, Ge H, Xia Y, Ma G. Identification method for household-transformer relationship in low-voltage transformer area based on LCSS-DBSCAN. In: Proceedings of the 2022 IEEE 3rd China International Youth Conference on Electrical Engineering (CIYCEE); 2022 Nov 3–5; Wuhan, China. doi:10.1109/ciycee55749.2022.9958988.
17. Fu B, Liang Y, Lao Z, Sun X, Li S, He H, et al. Quantifying scattering characteristics of mangrove species from Optuna-based optimal machine learning classification using multi-scale feature selection and SAR image time series. Int J Appl Earth Obs Geoinf. 2023;122:103446. doi:10.1016/j.jag.2023.103446.
18. Khosravi N, Dowlatabadi M, Sabzevari K. A hierarchical deep learning approach to optimizing voltage and frequency control in networked microgrid systems. Appl Energy. 2025;377(4):124313. doi:10.1016/j.apenergy.2024.124313.
19. Li H, Li S, Wu Y, Xiao Y, Pan Z, Liu M. Short-term power load forecasting for integrated energy system based on a residual and attentive LSTM-TCN hybrid network. Front Energy Res. 2024;12:1384142. doi:10.3389/fenrg.2024.1384142.
20. Fan P, Wang D, Wang W, Zhang X, Sun Y. A novel multi-energy load forecasting method based on building flexibility feature recognition technology and multi-task learning model integrating LSTM. Energy. 2024;308(3):132976. doi:10.1016/j.energy.2024.132976.
21. Abdulla N, Demirci M, Ozdemir S. Smart meter-based energy consumption forecasting for smart cities using adaptive federated learning. Sustain Energy Grids Netw. 2024;38(3):101342. doi:10.1016/j.segan.2024.101342.
22. Alamu O, Olwal TO, Migabo EM. Machine learning applications in energy harvesting internet of things networks: a review. IEEE Access. 2025;13(1):4235–66. doi:10.1109/access.2024.3525263.
23. Duan Y, Zhao Y, Hu J. An initialization-free distributed algorithm for dynamic economic dispatch problems in microgrid: modeling, optimization and analysis. Sustain Energy Grids Netw. 2023;34(2):101004. doi:10.1016/j.segan.2023.101004.
24. Yang Y, Wang Z, Gao Y, Wu J, Zhao S, Ding Z. An effective dimensionality reduction approach for short-term load forecasting. Electr Power Syst Res. 2022;210(4):108150. doi:10.1016/j.epsr.2022.108150.
25. Buzau MM, Tejedor-Aguilera J, Cruz-Romero P, Gómez-Expósito A. Detection of non-technical losses using smart meter data and supervised learning. IEEE Trans Smart Grid. 2018;10(3):2661–70. doi:10.1109/tsg.2018.2807925.
26. Morgoev I, Klyuev R, Morgoeva A. Methodology for detecting non-technical energy losses using an ensemble of machine learning algorithms. Comput Model Eng Sci. 2025;143(2):1381–99. doi:10.32604/cmes.2025.064502.
27. Alcañiz A, Grzebyk D, Ziar H, Isabella O. Trends and gaps in photovoltaic power forecasting with machine learning. Energy Rep. 2023;9:447–71. doi:10.1016/j.egyr.2022.11.208.

28. Miraftabzadeh SM, Colombo CG, Longo M, Foiadelli F. A day-ahead photovoltaic power prediction via transfer learning and deep neural networks. Forecasting. 2023;5(1):213–28. doi:10.3390/forecast5010012.
29. Jafarzadeh H, Mahdianpari M, Gill E, Mohammadimanesh F, Homayouni S. Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation. Remote Sens. 2021;13(21):4405. doi:10.3390/rs13214405.
30. Kavzoglu T, Teke A. Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). Arab J Sci Eng. 2022;47(6):7367–85. doi:10.1007/s13369-022-06560-8.