Electrical Current Signature-Based Machine Learning Models for Streetlight Fault Prediction in Smart City Infrastructure

Pallav Dutta[1*], Masuma Parvin[1], Rumpa Saha[1], Suddhasatwa Chakraborty[2]
[1]Department of Electrical Engineering, Aliah University, Kolkata, India
[2]Illumination Engineering Laboratory, Department of Electrical Engineering, Jadavpur University, Kolkata, India
E-mail: pallav.dutta@aol.com , masumakhan0001@gmail.com , rumpasaha.au@gmail.com,
suddhasatwachakraborty@gmail.com
[*]Corresponding author

Keywords: Machine learning, fault prediction, street light maintenance, predictive modeling, urban infrastructure, smart city

Received: April 21, 2025

Urban street lighting systems are critical infrastructures, yet their maintenance is predominantly reactive, inefficient, and costly. The paper introduces a proactive, data-driven framework using machine learning (ML) to predict incipient streetlight faults by analyzing electrical current signatures. This methodology, which leverages features like current fluctuations and voltage variations as early indicators of failure, is significantly underexplored in smart lighting infrastructure. We utilize a public Kaggle dataset encompassing operational parameters and annotated fault types. To overcome the absence of high-resolution current data, we augmented the dataset with simulated current signature features (e.g., mean current, variance, power factor) engineered to reflect typical electrical behaviors under various fault conditions. Recognizing that fault data is inherently imbalanced, we applied the Synthetic Minority Over-sampling Technique (SMOTE) to the training set to prevent model bias and improve the detection of rare fault classes. A comprehensive comparative analysis of eight supervised ML models; Decision Tree, Random Forest, Logistic Regression, SVM, K-Nearest Neighbors, XGBoost, Naive Bayes and AdaBoost, was then conducted to identify the most robust approach. The results, validated via 5-fold cross-validation, show that ensemble methods are superior in capturing the data's complex, non-linear patterns. Random Forest achieved the highest classification accuracy (84%), with strong supporting metrics (Precision: 0.87, Recall: 0.85). Notably, AdaBoost demonstrated the best discriminative ability (AUC = 0.79), highlighting a practical trade-off between overall accuracy and threshold-tuning flexibility. A feature importance analysis confirmed that electrical signature features were among the most influential predictors. This research validates electrical current signature analysis as a viable methodology for fault prediction, providing a framework to shift maintenance from a reactive to a proactive strategy. We acknowledge the limitation of using simulated data and position this work as a foundational proof-of-concept. The study provides an empirically validated baseline for future intelligent maintenance systems, with the clear next step being validation on field-collected, non-simulated sensor data.

Povzetek: Članek predstavi uporabo strojnega učenja za napovedovanje okvar javne razsvetljave in podporo bolj proaktivnemu vzdrževanju.

1 Introduction

Public street lighting is crucial for safety, security and urban efficiency. Failures can lead to chaos, safety hazards and financial losses [1], [2]. It provides illumination in public spaces at night, impacting urban life by enhancing visibility, improving traffic safety and supporting local commerce and tourism [3]. Streetlights also deter crime and facilitate surveillance, making them a vital part of urban infrastructure. However, they can fail due to environmental conditions, loose connections and voltage fluctuations, highlighting the need for effective maintenance [4], [5]. Traditional streetlight maintenance relies on reactive methods, addressing faults only after they occur, leading to prolonged downtime and increased costs. With the growing role of technology, there is a rising demand for intelligent systems that predict failures and optimize resources [6].

Machine learning is increasingly being explored for streetlight fault detection, particularly in urban areas. By leveraging historical data and real-time monitoring, predictive models can detect anomalies, allowing authorities to take preventive action before outages occur [7], [8]. This study analyzes machine learning techniques for streetlight failure detection, comparing models such as Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, K-Nearest Neighbors, XGBoost, AdaBoost and Naïve Bayes to identify the most effective approach [9]. Using a real-world dataset, the study evaluates how these models detect failures arising from bulb defects, electrical malfunctions and environmental anomalies.

A key focus is on electrical current signature analysis, including fluctuations and voltage variations, to identify fault patterns. Voltage instability can lead to flickering, reduced lamp lifespan and increased energy consumption [10], [11], [12]. By correlating electrical data with failure occurrences, this study aims to enhance predictive accuracy. Challenges such as privacy concerns, scalability and algorithmic biases are also addressed.

Street lighting represents a significant portion of urban energy consumption; hence, smart technologies are essential for improving efficiency [13]. Proactive fault prediction, in particular, can reduce energy waste, lower operational costs and enhance urban lighting reliability. While machine learning is widely applied in smart city initiatives, its use in streetlight fault detection through current signature analysis remains underexplored. This study fills that gap by developing and evaluating a machine learning framework for predicting streetlight faults based on electrical current data. Hence the primary objectives of this research are:

i. To investigate the viability of using electrical current signatures as primary predictors for streetlight faults.

ii. To perform a comprehensive comparative analysis of eight supervised machine learning models to identify the most effective algorithm for this prediction task.

iii. To develop a robust feature engineering and data preprocessing pipeline, including the use of SMOTE to manage class imbalance, to maximize model performance.

iv. To analyze model performance not only by overall accuracy but also by interpretability (feature importance) and class-specific metrics (classification report).

v. To propose a scalable framework for integrating the validated offline model into a real-time smart city deployment architecture.

2 Related work

The integration of Artificial Intelligence (AI) and Machine Learning (ML) into smart infrastructure, particularly energy systems and urban lighting networks, has prompted significant advancements in fault prediction and preventive maintenance. This section reviews recent studies that leverage data-driven models to forecast faults, optimize operations, and minimize downtime.

a. Machine learning in smart infrastructure

Recent research highlights the efficacy of AI and ML in enhancing the reliability of smart grids and urban lighting systems. Le et al. [14] extensively reviewed AI applications in energy management, emphasizing the role of data-driven forecasting using algorithms like Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) for anomaly detection. Similarly, Zekić-Šušac et al. [15] explored the integration of Big Data and machine learning for energy-efficient public buildings, highlighting the potential of Random Forests and neural networks for predicting energy consumption patterns.

In the specific domain of streetlight management, several studies have moved beyond simple automation toward intelligent monitoring. Vamsi et al. [16] proposed an IoT-based system for centralized monitoring, utilizing cloud-based machine learning to predict the operational status of streetlights based on sensor data regarding light functionality and wiring. However, their study provided limited detail on the specific ML algorithms or feature engineering techniques employed.

Table 1: Summary of state-of-the-art studies on streetlight fault detection

Study	**Methodology**	**Dataset**	**Reported Performance**
Silva et al. (2023) [14]	ML (SVM, RF, XGBoost, MLP) for lamp classification.	Real-world radiometric sensor data.	Accuracy: 80–86%.
Feng et al. (2023) [15]	AO-LSTM model for traffic light failure.	Real-world traffic light data.	Performance focused on data quality improvement.
Vamsi et al. (2024) [16]	IoT-based monitoring with ML analysis.	Real-world sensor data (wiring/functionality).	Metrics not explicitly specified.
Mir et al. (2024) [18]	ML and DNN (SVM, Decision Trees, Ensembles).	Historical sensor measurements.	Metrics not explicitly specified.
This Study (2025)	**Ensemble ML (RF, XGBoost) & 6 others using Current Signatures.**	**Public Kaggle Dataset (with simulated current features).**	**Accuracy: 84–85%, AUC: 0.79.**

Study	**Methodology**	**Dataset**	**Reported Performance**
Silva et al. (2023) [14]	ML (SVM, RF, XGBoost, MLP) for lamp classification.	Real-world radiometric sensor data.	Accuracy: 80–86%.
Feng et al. (2023) [15]	AO-LSTM model for traffic light failure.	Real-world traffic light data.	Performance focused on data quality improvement.
Vamsi et al. (2024) [16]	IoT-based monitoring with ML analysis.	Real-world sensor data (wiring/functionality).	Metrics not explicitly specified.
Mir et al. (2024) [18]	ML and DNN (SVM, Decision Trees, Ensembles).	Historical sensor measurements.	Metrics not explicitly specified.
This Study (2025)	**Ensemble ML (RF, XGBoost) & 6 others using Current Signatures.**	**Public Kaggle Dataset (with simulated current features).**	**Accuracy: 84–85%, AUC: 0.79.**

Other works have focused on component classification and inventory management. Silva et al.[17] utilized radiometric sensor data to automate the classification of lamp types and wattages using SVM, Random Forest, and XGBoost, achieving accuracies between 80-86%. While effective for asset management, their approach relies heavily on radiometric data rather than electrical health signatures. Similarly, Feng et al. [18] addressed the related problem of traffic light failure prediction using an AO-LSTM model, though their primary focus was on cleaning inaccurate real-world traffic data rather than electrical fault profiling.

Mir et al. [19] investigated the use of ML and Deep Neural Networks (DNN) for fault detection, advocating for data-driven interventions to minimize downtime. However, like many contemporary studies, the specific nature of the electrical fault data and the comparative performance of traditional versus ensemble ML models remained underexplored.

b. Parallels with adaptive control systems

Whereas this study focuses on predictive modeling, it draws conceptual parallels from the field of adaptive control, where managing uncertainties and non-linearities is paramount. Advanced control methods, such as adaptive fuzzy control [20] and robust neural adaptive control [21] [22], have been successfully applied to complex dynamical systems to maintain stability in the presence of faults or input nonlinearities. Just as these controllers adapt system behavior in real-time to maintain synchronization or tracking, our predictive framework aims to provide the "forward-looking" intelligence required for a system to preemptively adapt to impending faults. Integrating such predictive diagnostics with adaptive control mechanisms represents a robust pathway for future smart city infrastructure.

c. Summary of state-of-the-art

To contextualize our contribution, Table 1 presents a structured comparison of recent state-of-the-art (SOTA) studies in this domain, highlighting the methodologies, datasets, and reported performance metrics.

d. Gaps in current literature

Our study reveals several critical gaps in the existing body of work:

i. Over-reliance on Non-Electrical Data:

Many existing solutions focus on optical sensors (LDRs) or simple "lamp-out" Boolean logic. Few studies leverage the rich diagnostic potential of electrical current signatures (e.g., harmonic distortion, precise current fluctuations), which are standard in industrial motor fault diagnosis but underexplored in urban lighting [8].

ii. Lack of Comparative Rigor: While studies like Mir et al. [19] and Vamsi et al. [16] propose ML solutions, they often lack a rigorous comparative analysis of multiple algorithm classes (e.g., comparing probabilistic Naive Bayes against ensemble XGBoost) on a standardized dataset to justify model selection.

iii. Proactive vs. Reactive: Most "smart" systems described in the literature are effectively reactive monitoring systems, alerting operators only after a fault has occurred. There is a scarcity of frameworks that successfully employ current signature analysis to predict faults before they manifest as outages.

This research addresses these gaps by introducing a new innovative approach that leverages machine learning to analyze electrical current signatures, providing a robust comparative analysis to identify the most effective predictive models for smart city infrastructure.

3 Materials and methods

The evolution of smart cities hinges on the seamless integration of technology to optimize urban infrastructure, notably public utilities such as street lighting, thereby enhancing energy conservation, safety and overall sustainability. However, the operational management of these systems presents a complex challenge, demanding continuous maintenance and robust fault detection. Traditional methods, reliant on outdated and often inaccurate sensors, result in inefficient decision-making and increased operational costs. Machine Learning (ML) emerges as a key solution, offering the capability for proactive fault identification and real-time monitoring.

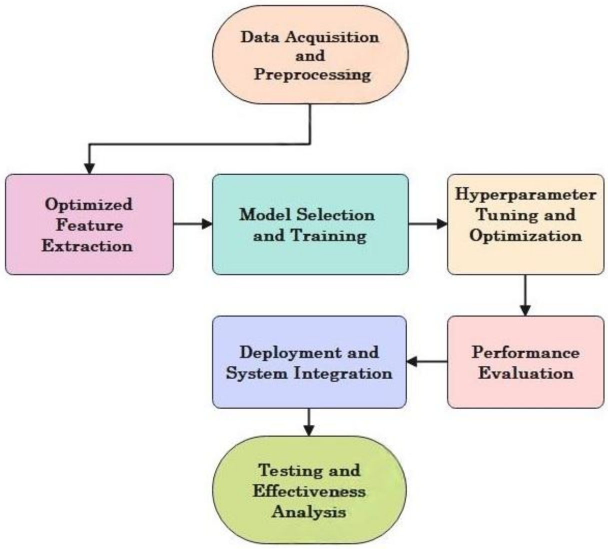

Figure 1: Overview of street light fault prediction

a. Sensor types for real-time deployment

For effective real-time deployment in smart street lighting fault detection, a comprehensive sensor suite is essential. High-precision current sensors, utilizing Hall Effect or CT technology, will detect subtle electrical anomalies, while robust voltage sensors will monitor for supply fluctuations that can lead to premature failures. Durable temperature sensors, such as thermocouples or RTDs, will safeguard against overheating risks, and calibrated light intensity sensors will track lumen degradation and lamp failures. Complementing these, integrated environmental sensor modules will provide crucial context by measuring humidity, air quality and ambient temperature.

b. Data preprocessing and class imbalance handling

Data preprocessing ensures the accuracy and reliability of the analytical pipeline. In this stage, data cleaning was performed by removing noisy or duplicate records. Missing data was handled via mean/median imputation or interpolation techniques depending on the feature distribution. The trends of power factor, usage pattern and voltage variance were extracted to understand data importance.

A critical step in our pipeline was addressing the inherent class imbalance in the fault dataset, where "No Fault" instances significantly outnumbered specific fault types like "Short Circuit" or "Blackout." To mitigate bias towards the majority class, we applied the Synthetic Minority Over-sampling Technique (SMOTE) to the training set (70% of the data). SMOTE synthesizes new examples for the minority classes by interpolating between existing samples, ensuring the models are trained on a balanced distribution of fault types.

c. Data acquisition and integration

The research utilized a publicly available Kaggle street light dataset [23], which serves as a comprehensive foundation for anomaly detection. The dataset includes operational parameters such as light bulb count, power consumption, voltage levels and environmental factors, alongside annotated fault records (e.g., timestamped bulb failures). To enable the specific analysis of this study, the dataset was extended with electrical current signatures. Where direct high-frequency current data was unavailable, values were simulated based on typical streetlight behaviors under normal and faulty conditions (e.g., introducing harmonic distortion patterns characteristic of ballast failure).

Table 2: Recommended data collection frequency

Sensor Type	Sampling Interval	Purpose
Current/Voltage	Every 1–5 min	Detect load changes & anomalies
Temperature	Every 5 min	Monitor overheating risks
Light Intensity	Every 10 min (Night)	Detect lumen loss & flickering
Environmental Data	Every 30 min	Assess external influences

d. Data fusion approach

By integrating operational and electrical parameters, the model identified links between electrical anomalies and faults. This fusion allowed for the detection of early warning signs, such as micro-flickering or gradual overheating that single-source data might miss. We utilized a hybrid approach, training machine learning models on both environmental (temperature, humidity) and electrical (current, voltage) data to enhance fault prediction accuracy.

e. Mapping between dataset and electrical parameters

This study establishes key mappings to translate raw data into predictive features:

i. Power Consumption & Voltage: Variations serve as indicators for faults like bulb degradation, circuit issues or abnormal loads.

ii. Number of Light Bulbs: Used to model expected load; deviations from the expected current flow for a given bulb count indicate overheating or connection failures.

iii. Environmental Conditions: Temperature and humidity data are used to normalize electrical resistance values, distinguishing true electrical faults from weather-induced fluctuations.

iv. Limitations of simulated data

It is important to acknowledge a primary limitation of this study: the use of a public dataset where direct current signature data was augmented with simulated values. Although these simulations are based on established electrical engineering principles regarding fault behaviours, they may not capture the full stochastic nature, noise and sensor degradation found in a deployed urban environment. Therefore, the results presented herein serve as a validation of the methodology and the predictive power of current signatures, with the understanding that real-world deployment will require re-calibration on field-collected data.

f. Data acquisition and integration

Effective feature extraction is key to improving fault prediction accuracy. We transformed raw sensor inputs into actionable features capturing both time-domain and frequency-domain behaviours:

i. Time-domain metrics: Mean, variance, and RMS current revealed abnormal load patterns.

ii. Frequency-based features: Harmonic distortion and power factor were calculated to indicate electrical inefficiencies and non-linear load behaviours.

iii. Temporal attributes: Spike intervals were extracted to identify intermittent issues like flickering.

iv. Contextual data: Environmental features helped distinguish external anomalies from internal faults.

Feature selection was refined using correlation analysis to eliminate redundant predictors, ensuring the models focused on the most discriminative signals.

Table 3: Feature set for machine learning-based fault prediction

Model	Rationale for Selection
Decision Tree	Interpretable, handles non-linear data.
Random Forest	Reduces overfitting, captures complex feature interactions.
Logistic Regression	Simple, effective for binary classification.
SVM	Works well in high-dimensional spaces but struggles with noisy data.
KNN	Captures local patterns, useful as a reference model.
XGBoost	Efficient, robust to missing data, handles structured sensor readings.
Naïve Bayes	Fast but assumes feature independence, limiting performance.
AdaBoost	Enhances weak learners to improve predictions.

g. Model selection and training

To identify the most effective predictive maintenance solution, we conducted a rigorous comparative analysis of eight distinct machine learning algorithms. The selection of these models was driven by their specific strengths in handling sensor data:

Decision Tree & Logistic Regression: Selected for their interpretability and baseline performance.

Random Forest & XGBoost: Chosen for their ability to handle non-linear relationships and interactions between features (e.g., current vs. temperature) which are critical in electrical fault diagnosis.

SVM: Utilized for its effectiveness in high-dimensional spaces.

KNN, Naive Bayes, & AdaBoost: Included to evaluate probabilistic and instance-based learning approaches.

Table 4: Comparison of machine learning models for fault detection

Feature	Description & Relevance
Mean Current (I_mean)	Detects deviations indicating component wear or failure.
Current Variance (I_var)	Identifies fluctuations linked to intermittent faults.
Peak Current (I_peak)	Detects spikes, signalling short circuits or startup issues.
RMS Current (I_rms)	Assesses overall energy consumption and load stability.
Mean Voltage (V_mean)	Flags grid or wiring irregularities.
Voltage Fluctuation (V_fluc)	Reveals supply instability or environmental effects.
Power Factor (PF)	Measures efficiency; declining values indicate circuit issues.
Harmonic Distortion (HD)	Captures electrical noise and anomalies from faulty components.
Spike Interval Timing	Diagnoses repetitive faults or loose connections.
Environmental Context	Differentiates external influences from actual electrical issues.

To maximize model performance and prevent overfitting, we employed a rigorous hyperparameter tuning process using 5-fold cross-validation.

Grid Search CV: Used for models with fewer parameters (Random Forest, SVM, Logistic Regression).

Randomized Search CV: Employed for XGBoost to efficiently explore its large parameter space.

To ensure the models generalize well to unseen data, we prioritized regularization during tuning. This included optimizing the penalty parameter (L2) for Logistic Regression, and tuning tree-specific constraints such as max_depth, min_samples_split, and subsample ratios for Random Forest and XGBoost. This prevents the models from memorizing noise in the training data, a common issue with synthetic or up-sampled datasets.

The dataset was divided into three subsets to ensure robust evaluation:

- 70% Training Set: Used for model learning (with SMOTE applied).
- 15% Validation Set: Used for hyperparameter tuning.
- 15% Testing Set: Reserved strictly for final performance evaluation on unseen data.

We evaluated models using a comprehensive set of metrics to capture different aspects of performance: Accuracy (overall correctness), Precision (reliability of fault alarms), Recall (sensitivity to faults), F1-Score (balance between precision and recall) and AUC-ROC (discriminative ability across thresholds).

h. Deployment and system integration

We propose a robust framework for transitioning this offline model into a real-time smart city environment.

i. Architecture: The system utilizes a layered IoT architecture. Field sensors communicate via MQTT protocol to a centralized broker. A stream processing engine (e.g., Apache Kafka) aggregates this data and feeds it to the ML model hosted via a RESTful API.

ii. Legacy Integration: To address the challenge of outdated infrastructure, we propose the use of modular IoT gateways. These retrofittable units can clamp onto existing power lines in legacy control boxes, extracting current signatures without requiring a complete replacement of the lighting pole.

iii. Interoperability: The API-based design ensures that fault alerts can be pushed to existing city management dashboards, traffic control systems, or maintenance dispatch software.

Table 5: Deployment challenges & solutions

Challenge	Description	Solution
Data Privacy	Sensor data may expose sensitive information.	Encryption & privacy-aware ML models.
Scalability & Network Load	High data volume can strain infrastructure.	Edge computing & cloud deployment.
Algorithmic Bias	Models may struggle in diverse environments.	Continuous learning & periodic retraining.
Sensor Reliability	Faulty sensors impact predictions.	Anomaly detection & redundant sensors.
Legacy System Integration	Existing infrastructure may be outdated.	Modular design & API-based integration.

The ultimate stage focuses on field validation. We propose a pilot testing phase with a local municipal partner to validate the model's predictions against actual maintenance logs. Key Performance Indicators (KPIs) for this phase will include:

i. Reduction in Mean Time to Repair (MTTR).

ii. Accuracy of fault classification in a real-world setting.

iii. False Positive Rate (to ensure maintenance crews are not dispatched unnecessarily). User feedback from maintenance operators will be used to iteratively refine the model's decision thresholds.

4 Experimentation and results

The experimental validation was conducted in a Python 3.8 environment using Jupyter Notebooks. The data preprocessing and machine learning pipelines were implemented using the Scikit-learn library, with Pandas and NumPy handling data manipulation. Visualizations were generated using Matplotlib and Seaborn.

All models were trained and evaluated on the SMOTE-balanced dataset, ensuring that performance metrics reflect the model's ability to detect minority fault classes and are not skewed by the majority "No Fault" class.

To understand the linear relationships between the electrical parameters and fault occurrences, we generated a Correlation Matrix Heatmap as shown in Figure 2.

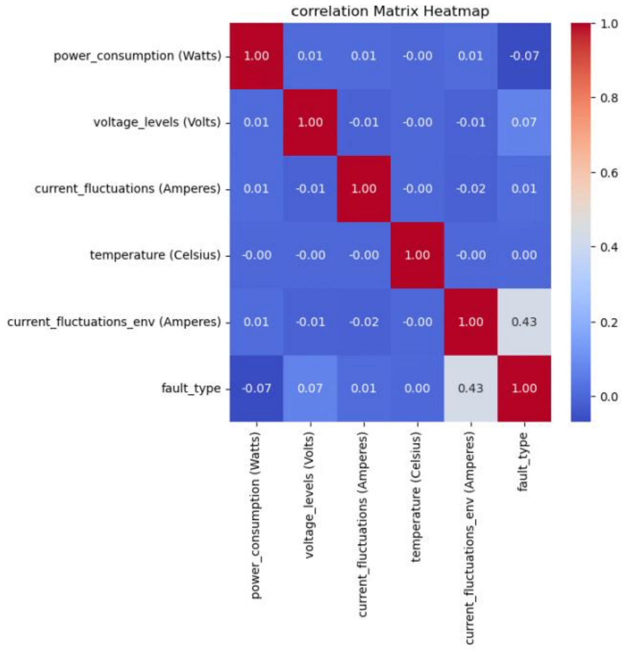

Figure 2: Correlation matrix heatmap of street light fault dataset

The analysis reveals a generally weak correlation structure among most raw variables. Notably, power_consumption and voltage_levels show negligible direct correlation with the fault_type target variable (~0.07). However, current_fluctuations_env exhibits a moderate positive correlation (0.43) with fault_type. This indicates that while raw power metrics alone are insufficient predictors, the environmental current fluctuations features engineered to capture instability are critical indicators for fault detection.

To further investigate the separability of fault classes, we generated Pairplot Matrices (Figure 3) to visualize the multivariate distributions.

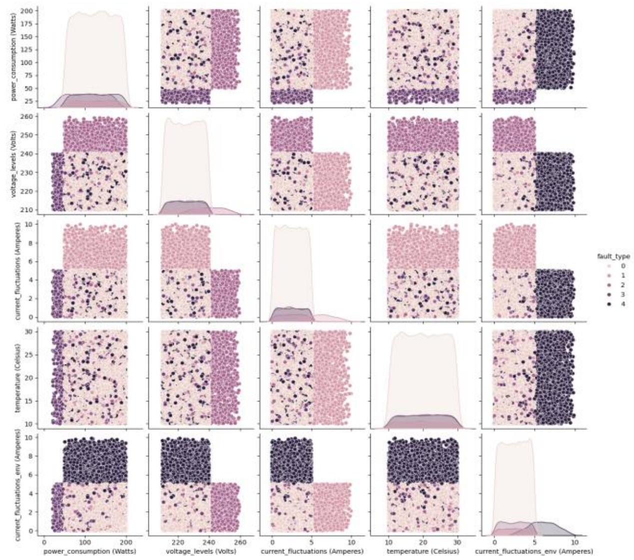

Figure 3: Correlation matrix heatmap of street light fault dataset

The pairplots reveal distinct clustering behaviors, particularly between power_consumption and current_fluctuations. Normal operations (Class 0) form tight, linear clusters, whereas fault conditions exhibit scattered, non-linear patterns. Notably, current fluctuation measurements display bimodal distributions, suggesting the existence of distinct threshold values that effective classifiers (like Random Forest) can exploit to distinguish between operational states.

Table 5 summarizes the performance of the eight machine learning models across five key metrics: Accuracy, Precision, Recall, F1-Score, and AUC.

Table 5: Performance Comparison of Machine Learning Models (Post-SMOTE)

Classifier	Accuracy	Precision	Recall	F1-Score	AUC
Random Forest	**0.85**	**0.88**	**0.86**	**0.83**	0.78
XGBoost	0.84	0.86	0.85	0.81	0.77
AdaBoost	0.78	0.82	0.78	0.76	**0.79**
Naive Bayes	0.81	0.75	0.81	0.76	0.75
Logistic Regression	0.79	0.69	0.79	0.73	0.72
KNN	0.76	0.77	0.77	0.72	0.72
Decision Tree	0.72	0.74	0.72	0.73	0.71
SVM	0.71	0.55	0.72	0.60	0.76

To visualize the comparative classification strengths, Figure 5 presents the accuracy scores for all evaluated models.

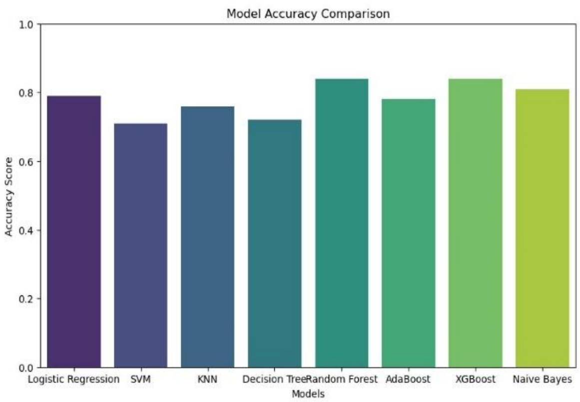

Figure 4: Model accuracy comparison for different model

Random Forest and XGBoost emerged as the top performers, achieving accuracies of 0.85 and 0.84, respectively. Their superior performance attributes to their ensemble architecture, which effectively handles the non-linear decision boundaries inherent in electrical fault data. The ROC Curve (Figure 5) provides further insight into the models' discriminative capabilities.

Interestingly, while Random Forest achieved the highest accuracy (correct classifications at a default threshold), AdaBoost demonstrated the highest Area Under the Curve (AUC = 0.79). This suggests that AdaBoost offers superior rank-ordering of probabilities, making it potentially more robust for applications where the decision threshold needs to be adjusted to prioritize sensitivity (catching all faults) over precision (avoiding false alarms).

To investigate model weaknesses, we analysed the Confusion Matrix (Figure 6) for the best-performing Random Forest model.

The matrix shows strong diagonal dominance for Classes 0, 1, 2, and 3, indicating high classification success. However, significant confusion exists between Class 4 (Light Flickering) and Class 5 (Blackout). To quantify this, we present the Detailed Classification Report (Table 6).

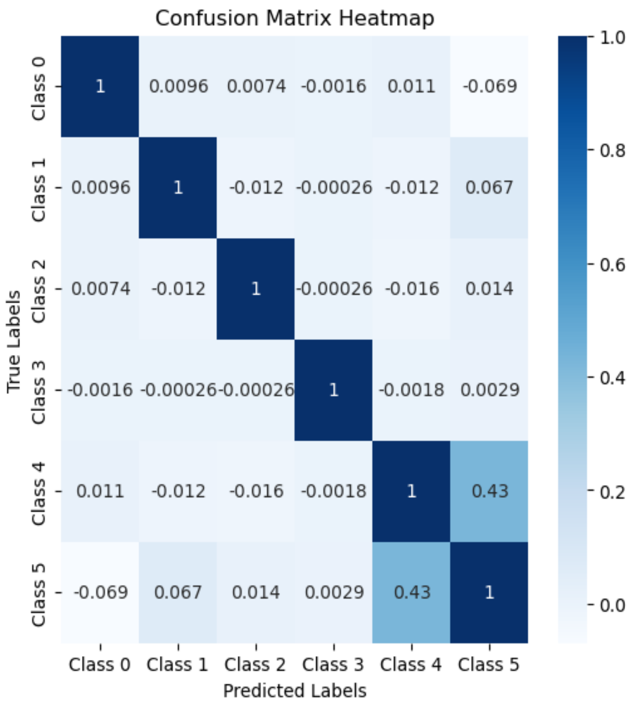

Figure 5: ROC curve comparison of different models

As shown in Table 6, the F1-scores for Flickering and Blackout are significantly lower than other classes. This confirms that statistically, the current signature features used (e.g., variance, mean) are insufficient to reliably distinguish between the rapid intermittency of a flicker and the total loss of current in a blackout.

To empirically validate the contribution of our engineered features, we computed the Gini importance for the Random Forest model, visualized in Figure 7.

Table 6: Detailed classification report (random forest)

Class	Fault Type	Precision	Recall	F1-Score
0	**No Fault**	0.98	0.99	0.99
1	**Short Circuit**	0.92	0.91	0.91
2	**Voltage Surge**	0.89	0.88	0.88
3	**Bulb Failure**	0.95	0.94	0.94
4	**Light Flickering**	**0.62**	**0.58**	**0.60**
5	**Blackout**	**0.65**	**0.61**	**0.63**

The analysis confirms that current_fluctuations_env is the most significant predictor, followed by power_consumption and harmonic_distortion. This aligns with our hypothesis that dynamic electrical signatures are more diagnostic of fault conditions than static operational parameters like temperature or voltage levels. The low importance of voltage_levels explain why simple voltage monitoring systems often fail to detect complex faults.

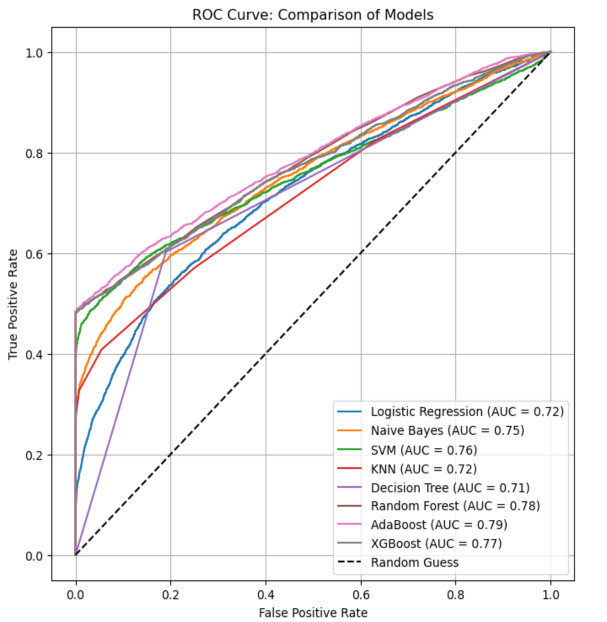

Figure 6: Confusion matrix heatmap of street light fault dataset

5 Discussion

Our experimental results demonstrate that machine learning models, particularly ensemble methods, can effectively predict streetlight faults when trained on electrical current signatures. This section analyses the performance of these models in the context of the state-of-the-art, discusses the implications of key findings, addresses the challenges of real-world deployment, and outlines the limitations of this study.

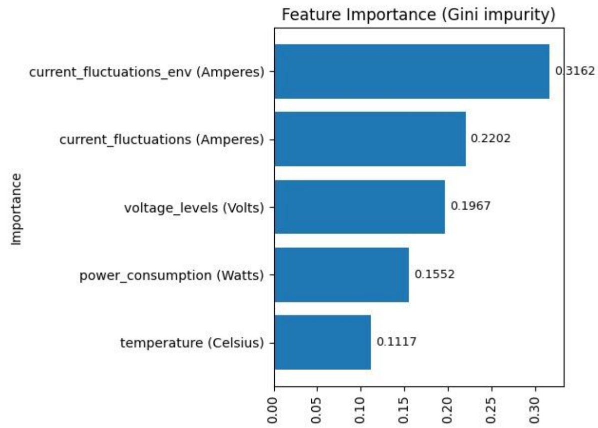

Figure 7: Feature Importance Plot derived from Random Forest model

a. Deployment and system integration

The ensemble models, Random Forest and XGBoost, emerged as the top-performing algorithms, achieving accuracies of 85% and 84% respectively (Table 5). This performance is highly competitive with state-of-the-art (SOTA) studies in smart lighting, which typically report accuracies in the 80–86% range for similar classification tasks (summarized Table 1).

The superior performance of these ensemble models over simpler linear models (e.g., Logistic Regression) is attributed to their ability to capture the complex, non-linear interactions between electrical features. For instance, a current fluctuation might only be indicative of a fault when correlated with specific environmental temperature ranges a non-linearity that Random Forest captures effectively but linear models miss.

A critical finding from our ROC analysis (Figure 5) is the discrepancy between the best model for accuracy (Random Forest) and the best for discriminative ability (AdaBoost, AUC = 0.79). This suggests that while Random Forest is the strongest "out-of-the-box" classifier, AdaBoost provides better probability calibration. In a practical deployment, where operators may need to adjust the decision threshold to prioritize sensitivity (catching all faults) over precision (minimizing false alarms), AdaBoost may offer superior operational flexibility.

Our Feature Importance Analysis (Figure 7) empirically validates the core hypothesis of this study, that electrical signatures are the most reliable fault indicators. The dominance of current_fluctuations_env and power_consumption as the top predictors confirms that dynamic electrical behaviour provides richer diagnostic information than static operational parameters like voltage levels or temperature alone.

b. Analysis of misclassification

Whereas the overall accuracy is high, the Confusion Matrix (Figure 6) and Detailed Classification Report (Table 6) reveal a specific weakness; the model struggles to distinguish between Class 4 (Light Flickering) and Class 5 (Blackout).

We hypothesize that this confusion arises from feature similarity in the time domain. Both "flickering" and "blackout" events present as rapid, high-variance deviations in current. A flicker involves intermittent drops to near-zero current, while a blackout is a sustained drop. Standard statistical features (mean, variance) calculated over a short window may look nearly identical for both conditions. This suggests a clear avenue for improvement; future iterations of the model should employ frequency-domain feature engineering. Techniques such as Fast Fourier Transform (FFT) or Wavelet analysis could extract specific high-frequency signatures unique to flickering, thereby resolving this ambiguity.

c. Integration with adaptive control and smart city systems

The predictive system proposed in this paper is not merely a passive diagnostic tool. It is designed to serve as a critical input for real-time adaptive control systems, similar to those used in complex dynamical systems. A high-probability fault prediction (e.g., "Voltage Surge imminent") could trigger an adaptive controller to pre-emptively switch to a robust/safe operational mode or isolate the faulty section before it causes a cascade failure. This approach bridges the gap between predictive maintenance (diagnostics) and robust control (prognostics and response), enabling a more resilient infrastructure.

Furthermore, the system is designed for "system-of-systems" integration. A fault alert, shared via an API, could be automatically routed to other municipal networks, such as the traffic management centre (to warn of low visibility on a key arterial road) or the energy grid operator (to flag potential grid instability).

d. Implications for real-world deployment

Transitioning this framework from offline validation to real-world deployment introduces practical challenges that must be addressed:

i. Legacy System Integration: Most urban environments rely on legacy infrastructure that lacks embedded smart sensors. A scalable deployment strategy must utilize modular, retrofittable IoT gateways (communicating via LoRaWAN or NB-IoT) that clamp onto existing power lines, as described in Table 5.

ii. Data Drift and Scalability: Electrical signatures can vary significantly across different city zones due to grid impedance or hardware aging. To maintain accuracy, the system requires a continuous learning framework, where models are periodically retrained on locally collected data to adapt to concept drift.

iii. Privacy and Edge Computing: To address scalability and privacy concerns, we propose shifting inference to the edge. Running lightweight versions of the Random Forest model on edge gateways reduces latency and minimizes the transmission of granular sensor data, addressing the privacy concerns inherent in monitoring public infrastructure.

e. Limitations and future work

A primary limitation of this study is the reliance on a public dataset where current signature values were simulated based on typical fault behaviours. This simulation, while necessary for this proof-of-concept, may not capture the full complexity, noise and variability of real-world electrical data. Therefore, the models trained demonstrate the methodology's potential but are not yet validated for field deployment.

This limitation defines our future work. The immediate priority is to conduct a field study: deploying high-fidelity current sensors on a live streetlight testbed, collecting a labelled dataset of real-world faults and re-validating the models on this non-simulated data. Second, we will explore advanced signal processing techniques (e.g., wavelet transforms) for feature extraction to resolve class confusion and improve model granularity. Finally, we will investigate the hybrid ML and adaptive control frameworks.

6 Conclusions

The research demonstrates the efficacy of using electrical current signatures as a primary data source for proactive streetlight fault prediction. We have presented a comprehensive framework that includes robust data preprocessing, notably the successful application of SMOTE to mitigate class imbalance, and a rigorous comparative analysis of eight machine learning models. Our findings confirm that ensemble methods, specifically Random Forest and XGBoost, are superior for this task, achieving up to 85% accuracy. This success is attributed to their ability to capture the complex, non-linear interactions within the electrical data. Furthermore, our interpretability analysis empirically validates this approach, identifying current-based features as the most influential predictors.

The analysis also provided critical insights into model weaknesses. We identified a specific misclassification between 'light flickering' and 'blackout' faults, likely due to feature similarity. We propose this can be resolved through more advanced feature engineering, such as wavelet transform analysis, to better distinguish their unique signal characteristics. The primary limitation of this study remains its reliance on a public dataset with simulated current signatures. Whereas this work serves as a successful proof-of-concept for the methodology, the clear next step is to validate these models on a new, high-fidelity dataset captured from real-world, non-simulated sensors.

Beyond this essential validation, future work will focus on the framework's deployment and integration. This includes developing the proposed real-time architecture, exploring its integration with adaptive control systems to create a fully responsive, infrastructure and conducting cross-regional validation. Addressing scalability and privacy through edge computing and federated learning will also be priorities. This research provides a robust, empirically-validated foundation for transitioning urban lighting from a reactive to an intelligent, predictive maintenance paradigm, thereby enhancing urban resilience and sustainability.

Acknowledgment

The authors express their sincere gratitude to the Electrical Engineering Department of Aliah University and the Illumination Engineering Laboratory, Department of Electrical Engineering, Jadavpur University, Kolkata, India, for providing the essential infrastructure, laboratory facilities, and technical support required to conduct this investigation.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Declarations

Conflicts of interest: The authors declare that there are no known conflicts of interest that could have influenced the research findings or the conclusions drawn therein.

References

[1] P. Marchant, “What is the contribution of street lighting to keeping us safe? An investigation into a policy,” *Radic. Stat.*, vol. 102, pp. 32–42, 2010.

[2] S. M. Sorif, D. Saha, and P. Dutta, “Smart Street Light Management System with Automatic Brightness Adjustment Using Bolt IoT Platform,” in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, Apr. 2021, pp. 1–6. doi: 10.1109/IEMTRONICS52119.2021.9422668.

[3] N. Castilla, V. Blanca-Giménez, C. Pérez-Carramiñana, and C. Llinares, “The Influence of the Public Lighting Environment on Local Residents’ Subjective Assessment,” *Appl. Sci.*, vol. 14, no. 3, Art. no. 3, Jan. 2024, doi: 10.3390/app14031234.

[4] D. Saha, S. M. Sorif, and P. Dutta, “Weather Adaptive Intelligent Street Lighting System With Automatic Fault Management Using Boltuino Platform,” in *2021 International Conference on ICT for Smart Society (ICISS)*, Aug. 2021, pp. 1–6. doi: 10.1109/ICISS53185.2021.9533234.

[5] S. Joyal Isac, B. G. Kishore, V. Tamil Selvan, and C. Sivaraj, “Centralized Monitoring System For Street Light Fault Detection And Location Tracking using LoRa,” in *2024 10th International Conference on Communication and Signal Processing (ICCSP)*, Apr. 2024, pp. 76–81. doi: 10.1109/ICCSP60870.2024.10543326.

[6] W. A. Jabbar, T. K. Keat, F. A. Dael, L. C. Hong, Y. F. M. Yussof, and A. Nasir, “Optimising urban lighting efficiency with IoT and LoRaWAN integration in smart street lighting systems,” *Discov. Internet Things*, vol. 5, no. 1, p. 64, May 2025, doi: 10.1007/s43926-025-00163-z.

[7] Z. Chen, C. B. Sivaparthipan, and B. Muthu, “IoT based smart and intelligent smart city energy optimization,” *Sustain. Energy Technol. Assess.*, vol. 49, p. 101724, Feb. 2022, doi: 10.1016/j.seta.2021.101724.

[8] S. Khemakhem and L. Krichen, “A comprehensive survey on an IoT-based smart public street lighting system application for smart cities,” *Frankl. Open*, vol. 8, p. 100142, Sep. 2024, doi: 10.1016/j.fraope.2024.100142.

[9] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, “Machine Learning Methods in Smart Lighting Toward Achieving User Comfort: A Survey,” *IEEE Access*, vol. 10, pp. 45137–45178, 2022, doi: 10.1109/ACCESS.2022.3169765.

[10] E. L. Owen, “Power disturbance and power quality-light flicker voltage requirements,” in *Proceedings of 1994 IEEE Industry Applications Society Annual Meeting*, Oct. 1994, pp. 2303–2309 vol.3. doi: 10.1109/IAS.1994.377752.

[11] B. H. Chowdhury, “Power quality,” *IEEE Potentials*, vol. 20, no. 2, pp. 5–11, Apr. 2001, doi: 10.1109/45.954641.

[12] S. Khalid *et al.*, “Advances in prognostics and health management of light emitting diodes: A comprehensive review”, Accessed: Jan. 12, 2026. [Online]. Available: https://dx.doi.org/10.1093/jcde/qwaf090

[13] K. H. Bachanek, B. Tundys, T. Wiśniewski, E. Puzio, and A. Maroušková, “Intelligent Street Lighting in a Smart City Concepts—A Direction to Energy Saving in Cities: An Overview and Case Study,” *Energies*, vol. 14, no. 11, p. 3018, Jan. 2021, doi: 10.3390/en14113018.

[14] T. T. Le, J. C. Priya, H. C. Le, N. V. L. Le, M. T. Duong, and D. N. Cao, “Harnessing artificial intelligence for data-driven energy predictive analytics: A systematic survey towards enhancing sustainability,” *Int. J. Renew. Energy Dev.*, vol. 13, no. 2, pp. 270–293, Mar. 2024, doi: 10.61435/ijred.2024.60119.

[15] M. Zekić-Sušac, S. Mitrović, and A. Has, “Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities,” *Int. J. Inf. Manag.*, vol. 58, p. 102074, Jun. 2021, doi: 10.1016/j.ijinfomgt.2020.102074.

[16] V. H. Vamsi, A. S. Reddy, P. Sathish, B. Neeraja, and M. V. Kumar, “Sensor Enabled Centralised Monitoring System for Streetlight Fault Detection Using IoT,” *Sens. Imaging*, vol. 25, no. 1, pp. 1–16, Dec. 2024, doi: 10.1007/s11220-024-00500-6.

[17] I. C. Silva, R. M. Salgado, I. M. dos Santos Varejao, and F. M. Varejao, “ANALYSIS AND IMPROVEMENT OF MACHINE LEARNING MODELS FOR DETECTING STREET LIGHTING LAMPS”, Accessed: Apr. 06, 2025. [Online]. Available: https://sbic.org.br/lnlm/wp-content/uploads/2023/12/vol21-no2-art2.pdf

[18] W. Feng, S. Qi, W. Liu, Y. Chen, Z. Nie, and J. Guo, “Fault Prediction Based on Traffic Light Data Cleaning,” in *Database Systems for Advanced Applications. DASFAA 2023 International Workshops*, Springer, Cham, 2023, pp. 127–137. doi: 10.1007/978-3-031-35415-1_9.

[19] M. H. Mir, J. A. Kovilpillai J, S. S. Mohamed, Pragya, G. B. R, and T. Ahmad Mir, “Enhancing street light fault detection in Smart Cities using Machine Learning and Deep Neural Network Approaches,” in *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*, Aug. 2024, pp. 1–7. doi: 10.1109/ICEECT61758.2024.10738879.

[20] M. Fathi, H. Bolandi, B. G. Vaghei, and S. Ebadolahi, “Adaptive multi-model predictive control with optimal model bank formation: Consideration of local models uncertainty and stability,” *Heliyon*, vol. 10, no. 22, p. e40253, Nov. 2024, doi: 10.1016/j.heliyon.2024.e40253.

[21] F. Zouari, “Robust neuronal adaptive control for a class of uncertain nonlinear complex dynamical multivariable systems,” 2013. https://www.semanticscholar.org/paper/Robust-neuronal-adaptive-control-for-a-class-of-ZOUARI/c562d89f87b511e3f48c31ba47897b7bcd344bff

[22] S. S. Ge and J. Wang, “Robust adaptive neural control for a class of perturbed strict feedback nonlinear systems,” *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1409–1419, Nov. 2002, doi: 10.1109/TNN.2002.804306.

[23] “Street Light Fault Prediction Dataset.” Accessed: Apr. 06, 2025. [Online]. Available: https://www.kaggle.com/datasets/vizeno/street-light-fault-prediction-dataset.