Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Advanced Strategies for Substation Asset Management: Leveraging Artificial Intelligence and Predictive Analytics

Dipta Roy[1], Rayhanul Islam Sony[2], and Md Ariful Islam Bhuiyan[3]
[1]*Department of Electrical and Computer Engineering, California State University, Northridge, 91330, CA, USA*
[2]*College of Graduate Professional Studies, Trine University, One University Avenue, Angola, 46703, Indiana, USA*
[3]*Department of Electrical and Computer Engineering, California State University, Northridge, 91330, CA, USA*
Corresponding Author: Rayhanul Islam Sony, **E-mail**: rsony23@my.trine.edu

| ABSTRACT

The dependability and resilience of smart grids today are heavily dependent on effective substation asset management and accurate fault identification. Traditional techniques suffer from not being able to handle the complexity, nonlinearity, and high-dimensionality of smart grid data, and hence fault classification in real time is less than optimal. In this study, we propose a hybrid model that combines MLP, LightGBM, and LR to enhance the efficiency, robustness, and accuracy of substation fault detection. The model leverages the capability of deep learning to detect complex, nonlinear patterns and tree-based models to enable effective handling of high-dimensional, structured data. We evaluate the proposed model on the Smart Grid Asset Monitoring the dataset with an accuracy of 98.61%, precision of 99.00%, and recall of 98.66%, which is better than conventional ML and DL approaches. The hybrid system delivers an efficient and scalable real-time solution for substation monitoring towards predictive maintenance and wise decision-making. Our results confirm that integrating heterogeneous modeling techniques can significantly enhance fault classification and detection in smart grid networks, paving the way for even more smarter and trustworthy energy management systems.

| KEYWORDS

Substation Asset Management, AI, Hybrid models, MLP, LightGBM, Sustainability, Smart Grid Asset, Logistic Regression, Machine Learning, Deep Learning

| ARTICLE INFORMATION

ACCEPTED: 20 October 2025 **PUBLISHED:** 31 October 2025 **DOI:** 10.32996/jcsts.2025.7.11.12

1. Introduction

An updated society relies on steady energy to operate the areas of industry, the economic system, and social living, and so the steady operation of electricity power systems is a crucial element of modern society [1]. The fundamental components of these systems are substations, important interconnection points to convert levels of stepping voltage, to control the flow of power, and to maintain the safety of operations [2]. Some of the super costly and critical elements of the grid include substation assets, including transformers, circuit breakers, and busbars [3]. Any unexpected breakdown or irregular maintenance of these commodities can lead to massive service blackouts and expensive damage to equipment, and subsequently to blackouts [4]. As a result, superior asset management in substations has become a corporate priority with both utilities and regulators worldwide [5]. Traditionally, the management of substation assets has been on the basis of manual inspection, regular preventive maintenance and rule based decision making [6]. As much as the practices will guarantee a minimum quota of reliability, they are reactive (where a reactionary process is involved) and inefficient in their processes. Janet repaired maintenance is more likely to succeed in effecting unnecessary maintenance interventions, yet unknown stress, or better misclassification of asset types, can endanger the audit plan of care monitoring [7]. In addition, manual systems are extremely labour intensive and prone to human subjectivity and cannot process the mass sensor data generated by present day substations [8]. The element of constraint is where the task force is experiencing an urgent need of smart and data centered solutions that has the capacity to render the process of sustaining substations as not reactive but anticipatory and proactive.

Recent advances of artificial intelligence (AI) and predictive analytics have shown incredible potential in power systems (i.e. load forecasting, demand-side control and condition-level component-level demand), and a broad range of applications of such techniques is sure to emerge in the future [9]. Precisely, it is the machine learning (ML) algorithms that can process historical and real-time information to uncover hidden patterns to enrich operations decision-making [10]. One of the significant functions of AI in asset management is to categorize the assets in the raw operational data. Correct classification of the types of assets is natural because there exist colossal dissimilarities among the maintenance and operational hazards as well as well-being signs of equipment (transformers, breakers and busbars). A powerful such identification framework is what would ensure that predictive models are efficiently aligned with the personal degradation history of different equipment [11].

Nevertheless, even with the promising results, current studies on AI in asset management are being limited in a number of ways. First, a significant part of the literature is devoted to fault detection or load forecasting as opposed to the inherent issue of asset type classification that is a key to integrated management of substations [12]. Second, most models perform extremely well on controlled datasets, however, fail to deal with the difficulties that include extreme class imbalances, scarce availability of labeled data, or interpretability requirements in practice [13]. Third, the current researches tend to be not thoroughly compared to the benchmarking of various baselines, and it is still unclear whether they can be more practical or not [14]. These loopholes limit the use of AI methods in working substation conditions.
To overcome these problems, this paper will design a complex predictive analytics system which is specially designed to classify substation assets of a certain type. The framework combines a specialized data preprocessing and augmentation pipeline with various machine learning baselines and a hybrid ensemble model, which is stacking-based. It use augmentation of asset records, oversampling techniques, time-feature engineering, categorical encoding, and interaction feature engineering to create balanced high-quality datasets. Base-line classifiers, such as Random Forest, LightGBM, Multi-Layer Perceptron and Logistic Regression- are systematically tested to determine their performance [15]. Lastly, a new model, stacking based hybrid, is suggested, in which predictions made by Random Forest, LightGBM, and MLP are integrated with a Logistic Regression meta-learner to enable higher accuracy, strength, and generalization. The experimental findings support the fact that the hybrid model has been effective in the constant and high accuracy, precision, recall and F1-score compared to baseline classifiers and other state-of-the-art models. Notably, because the current study centers on the classification of substation assets types, a framework that can be directly deployed is presented, which will make the equipment classification reliable and, consequently, reinforce predictive maintenance techniques and the overall resiliency of the grid.
The main objective of this study is to come up with and test an AI-powered model of correct classification of asset types in substations It will be planned to behave above the new fashion modes of operation that use intelligence and predictive examinations to identify asset populations through the use of operational data to identify asset populations. The correct identification of the assets of type is actually required since the different types of equipment's (transformers, breakers, and busbars) work and degrade differently and need different maintenance. The preprocessing, class-imbalance issues, and ensemble approach are believed to yield a robust, accurate, understandable answer that will form the heart of predictive maintenance and enhanced asset management in future substation.

The main contributions of this study are:

- Integrated Preprocessing and Data Augmentation Pipeline : We developed a rigorous workflow that includes augmentation of asset records, random oversampling, time-feature extraction, categorical encoding, interaction feature generation, and feature scaling. This ensured balanced and high-quality data for training predictive models.
- Comparison of Multiple Baseline Models: We practiced and tested Random forest, LightGBM, Multi-layer perceptron, and logistic regression. This then enabled us to have a clear comparative representive analysis of their performance with respect to the classification of their performance as substations asset type.
- Hybrid Stacking Model Proposed: The proposed model is a hybrid model of the three prediction models: the Random Forest, LightGBM and MLP. Such a design was far more correct and powerful compared to the individual baseline designs.
- Addressing Class Imbalance of Asset Data: This issue is the most significant one since we addressed the imbalance of data between the two or multiple asset classes by introducing augmentation and oversampling strategies. These methods optimized the model generalization of the underplayed, low-frequency categories of assets and the overall validity.
- High Performance Over State-of-the-art Models: We had determined that that proposed hybrid stacking model would always have been good compare to the basic classifiers and other advanced machine learning models. The ensemble of LightGBM, MLP, and Random Forest with a Logistic Regression meta-learner demonstrated the greatest accuracy, precision, and recall, and F1-score, which serves as a contributor to the suggestion of the predictive asset management framework as a substitute of substation asset classification.

The rest of this paper as follows: Section 2 reviews of related work on AI and predictive analytics for asset management in power systems. Section 3 describes methodology, including preprocessing pipeline, baseline models, and the proposed hybrid model. Section 4 presents the experimental results and comparative evaluation across models. Section 5 discusses findings, implications for substation operations, and practical deployment considerations. Finally, Section 6 wraps with the key contributions of this study and outlines guidelines for future research.

2. Literature Review

The substations constitute the backbone of electrical transmission and distribution systems, where asset reliability defines the security and efficiency of power systems. Conventional asset management practices for substation assets were usually maintenance based on time schedules or corrective maintenance after a failure has occurred. These methods entail high operational costs, unplanned outages, or sometimes even the loss of lives. With the advent of Artificial Intelligence (AI) and Predictive Analytics, new approaches have now been looking to overcome these issues. By detecting faults, scheduling maintenance, and ultimately assisting in extending the asset life using real-time sensor data, advanced machine learning models, and intelligent decision-support systems, predictive asset management will be a true game-changer.

Shaksham, Singh and Preeti [16] have studied fault tolerance and grid resilience in traditional and intelligent substation architectures. Using real-time logs, SCADA data, and sensor inputs (reload, voltage, current, temperature), they simulated nine designs using MATLAB/Simulink, Python, and Power World Simulator. Optimization was carried out through predictive modelling, IEEE standard validation, and expert review. Results showed that the base system exhibited poor performance (120 ms detection, 15 s recovery, and 4.8% downtime), whereas intelligent means of cyber intrusion detection and AI-based fault prediction had 25-35 ms detection, 2.5-3 s recovery, and less than 1% downtime. The conclusion is drawn that substations using AI, IoT, and wide-area coordination are much more resilient and efficient than traditional redundancy-based ones.

Kerk, See and Gim [17] discussed with the reader an asset-management-and-condition-monitoring case study at SP PowerGrid. The dataset consisted of historical maintenance logs, equipment failure records, and monitoring data collected from substations at various voltage levels (400kV, 230kV, 66kV, 22kV, and 6.6kV). The techniques moved away from traditional time-based maintenance and adopted a monitoring approach mostly predictive and condition-based, involving thermal imaging, partial discharge detection, humidity sensors, and IoT-based circuit breaker diagnostics, among others. In an attempt to optimize, real-time monitoring, robotic inspection, and predictive analytics using Siemens MindSphere were introduced. This managed to hugely reduce unplanned shutdowns by 85%, avoided 1,709 failures, and achieved cost savings of over S$101 million. Reliability indices have also risen to the best levels ever, with SAIDI at just 6.6 seconds per annum and SAIFI at one interruption in 233 years. The study concluded with the affirmation that condition-based monitoring does not only reduce failures and costs but also provide a more resilient grid and ensure operational continuity.

Xue et al. [18] proposed a strategy for substation equipment selection based on the theory of management collaboration. The data come from the operational history of 110kV substations in Henan Province and include maintenance records, reliability data, economic cost information, etc. Multiple management frameworks were combined in the methods: LCC theory for economics: RCM for reliability; Rough Set Theory (RST) for energy efficiency, and Enterprise Asset Management (EAM) for data preprocessing. Total Productive Maintenance (TPM) and management synergy theory with parametric coefficients in sequential form were employed to carry out the optimization, while consultations with experts and grey correlation analysis were used to support this method. Results showed that the outdoor HGIS units accomplished better than GIS in overall efficiency. Lower lifecycle fault maintenance costs were incurred (4.37M as against 6.16M yuan), a shorter payback period was present (12 as against 16 years), and a higher equivalent utilization was evident (0.52 as against 0.46). GIS did have a marginally higher reliability due to it being an enclosed type, yet from a comprehensive evaluation, the HGIS was found to be the better alternative. The conclusion arrived at was that economic criteria alone cannot suffice; instead, collaborative multi-theory approaches enhance the establishment of balanced, reliable, and energy-efficient choices for substation equipment.

In their work, Haakana et al. [19] presented a procedure for checking the eligibility of battery energy storage systems in electricity distribution asset management. The dataset consisted of Finnish rural distribution network data capturing five 20 kV feeders (550 km), 440 km of LV lines, 390 substations, and AMR data from 1,980 customers. Methodologies interlaced long-term planning, life cycle cost analysis (LCCA), and scenario modelling to test BESS solutions against conventional options of underground cabling and overhead line upgrades. The optimization procedure was carried out on cost criteria varying the price of BESS unit (100–500 €/kWh) and dicassoine placement strategies. Results indicated that BESS could be feasible, today at 300 €/kWh, for 19 branch lines and, with the reduction in costs to 100 €/kWh, feasibility is increased to about 50 lines. The conclusion stated that the use of BESS is more beneficial for extending network life and providing an improvement in reliability to young networks, while its wide adoption hinges upon continued cost reduction and encouragement in the course of regulatory developments.

Pardhavi, Sai and Sree [20] developed an IoT-based real-time monitoring system used for the detection and protection of transformer faults. Datasets consisted of laboratory test transformer parameters like temperature, oil-level, current, and vibration. Methods included multiple sensors connected to an ESP32 microcontroller, fault logic comparing sensor data against safety thresholds. Optimization included a dual-communication option: Wi-Fi for cloud dashboards and GSM for SMS alerts to ensure system reliability in connected as well as remote areas. Results proved the system efficiently detected conditions like overheat (85°C), low oil level (35%), and abnormal vibration (3.5 m/s²), which initiated relay action as well as real-time alerts automatically. As a conclusion, it was observed that IoT-based hybrid communication monitoring increases reliability, aids predictive maintenance, and is a cost-effective and scalable solution for power transformers, whether located in urban or remote locations.

Tian et al. [21] proposed an improved UAV-based infrared diagnosis system for intelligent substation inspection using an enhanced YOLOv4 model. The dataset consisted of 500 infrared images of substation equipment, split equally for training and testing. Methods integrated MobileNet-v3 for lightweight feature extraction, deep separable convolution to reduce parameters, and a Convolutional Block Attention Module (CBAM) to strengthen feature capture. Optimization included data augmentation (rotation, scaling, sharpening) and noise-interference testing. Results showed the CBAM-YOLOv4 model achieved 95.12% recognition accuracy, 78.34% mIoU, and 62.05 FPS, outperforming YOLOv3, YOLOv4, and Mask-RCNN, with only a 0.2% drop under varied conditions. The conclusion emphasized that the lightweight CBAM-YOLOv4 model notably improves fault identifiaction speed, accuracy, and robustness, making it practical for real-time UAV-based substation inspection.

Babu et al. [22] suggested a physics-informed, data-driven approach to identify high- resistance low-current arc faults (HIFs) at the primary side of substation transformers. The data was simulated from the IEEE 13-bus system with events like motor start up, load switch, and arcing fault. Approaches integrated the Hankel Alternative View of Koopman Operator for dynamic feature representation and the Series2Graph algorithm for subsequence anomaly identification to allow for the discrimination between faults and normal transients. Optimization was through the application of the 3-sigma rule and the sample rate measurement (>2000 samples/cycle) for better detection precision. It was revealed that for Case A, the fault was detected at 0.00186 s (approximately 1/10 of cycle) and for Case B, at 0.00096 s (1/16 of cycle) with higher protection reliability over standard protection techniques. It was concluded that the hybrid approach of Koopman–graph is successful in discriminating HIFs from similar activities like load switch, minimizes false alarms, and enhances protection for advanced smart grids.

Arunkumar [23] studied AI-driven predictive maintenance methods for electrical power grids and electrical equipment. Parameters studied included temperature, current, vibration, and power consumption from sensor data sets, SCADA, smart meters, IoT sensors, and maintenance data. Methods included machine learning (SVM, decision trees, CNNs, RNNs, LSTM), digital twin, reinforcement learning, and NLP log analysis. Optimization was achieved through the use of predictive models integrated with CMMS/EAM software, edge computing for real-time processing, and continuous loops. Case studies (GE Aviation, Schneider Electric, Rio Tinto, Siemens Gamesa, and PG&E) showed 10–35% reduction in downtime, 10–20% reduction in expense, and enhanced asset reliability. The conclusion was that AI, when supported by IoT and Industry 4.0 projects, enables active, interpretable, and extendable predictive maintenance to ensure resiliency, sustainability, and efficiency for current power grids.

Kurapati and Sia [24] suggested machine learning-based predictive maintenance for electrical grid infrastructure. The dataset was the UCI Simulated Electrical Grid Stability dataset of 10,000 records with 12 attributes (voltage, current, temperature, stability parameters). Methods utilized six ML algorithms—LR, DT, RF, K-NN, SVM, Gradient Boosting, and AdaBoost—in a Design Science Research (DSR) approach. Optimization methods utilized were dimensionality reduction by PCA, handling of class imbalance (oversampling/undersampling) and hyperparameter optimization for Gradient Boosting and AdaBoost. Experimental outcomes showed that Gradient Boosting showed the highest accuracy (99.2%), and ensemble-based models outperformed all others consistently for original as well as resampled datasets. Take-away was that ML-based predictive maintenance can effectively improve reliability, reduce downtime, and optimize utilization of resources for smart grid networks.

Ghazaly et al. [25] proposed an AI-integrated predictive maintenance approach for distribution transformers to ensure higher reliability and minimum downtime. The dataset employed was the Power Transformers Health Condition Dataset (Kaggle) with parameters such as voltage, oil temperature, winding temperature, and oil pressure. Techniques employed condition monitoring, feature engineering, and several AI/ML models—CNNs, RNNs, Gradient Boosting, SVM, and Bayesian Networks—along with IoT-based real-time data access. Optimization encompassed preprocessing, anomaly identification (One-Class SVM, Isolation Forest, Autoencoders), and predictive modeling through Bayesian networks and gradient boosting. The results indicated that Gradient Boosting performed best of all with 90.44% precision, 94.2% precision, and 96.55% recall, followed by SVM (91.55% precision), CNN (86.33%), RNN (82.45%), and Bayesian Networks (89.63%). The conclusion reiterated that AI-based models, particularly ensemble methods, strongly support fault prediction, minimize unintended outages, and blend ideally with Industry 4.0 for prudent transformer maintenance.

Hu et al. [26] recommended prophetical maintenance of iron-core current transformers for 110 kVsubstations. The dataset was comprised of past monitoring data (current, voltage, power factor, temperature, humidity, harmonic content) with 1000 samples from healthy (HLTY), single-phase short-circuit (1HCF) and double-phase short-circuit (2HCF) conditions. Techniques combined multi-level feature extraction by wavelet transform, feature reduction by Spearman interrelationship and Mutual Information, and fault prediction by RF classifiers. Optimization was through SMOTE for handling of class imbalance, grid search for hyperparameter optimization, and adaptive thresholding. Performance revealed 94% total prediction accuracy: 100% for healthy state, 84% for

1HCF, yet only 67% for 2HCF due to overlapping of features. The conclusion emphasized that, although the strategy is reliable and stable for usual states, 2HCF fault identification needs to be enhanced through larger dataset requirements and higher-level models in future research.

Nkinyam et al. [27] proposed and deployed a 100 kVA distribution transformer fault detection and monitoring system employing IoT and GSM technologies. The dataset relied upon real-time transformer parameters of phase voltage, line current, and power factor, calibrated by comparing with readings from the multimeter. Techniques utilized Arduino Mega with PZEM-004t and ZMPT101B sensors for current/voltage, ESP32 for IoT communication, GSM for SMS reporting, and ThingSpeak cloud for storage of data. Optimizations employed dual power supply (grid +_battery backup), local storage by SD card, and Pearson correlation analysis (>\(0.96\)) to guarantee precision. The results indicated reliable identification of breaker trips, fuse failures, neutral faults, over/undervoltage, with real-time SMS and IoT-based reporting (Fig. 9–12, pp. 14–16). The conclusion pointed out that it is a low-cost, expandable system that enhances predictive maintenance, minimizes downtime, and is particularly appropriate for developing countries grids.

Roosadi et al. [28] outlined the application of predictive maintenance of substation components at PT PLN UP2D Kalselteng through K-Means clustering. The dataset was 8832 samples of two-transformer and six-feeder AC current measurements at the Amuntai main substation (June–August 2023) at 30-min intervals. Techniques employed absolute z-score normalization for dealing with AC current variations followed by K-Means-based clustering. Optimization was through feature combination experimentation and hyperparameter tuning (elbow method + grid search) tested with silhouette scores. It was revealed that individual models per component outperformed all others, with feeders attaining the maximum of clustering quality (e.g., Feeder-4 silhouette score = 0.9203). Validations through real outage records showed that the resulting clusters can effectively capture outages. It was concluded that K-Means is a simple yet effective unsupervised approach to early anomaly identification and predictive maintenance, although practical implementation is nevertheless subject to field verification.

Wiese [29] outlined a decision-making approach for the use of AI-based predictive maintenance (PdM) in strategic infrastructure industries like energy, transportation, water, and telecommunications. The dataset background included sensor-based operating data such as vibration, temperature, acoustic sensors, flow rates, and network activity logs. Techniques examined were machine learning (random forests, SVM), deep learning (CNNs, RNNs), and hybrid failures for prediction, underpinning by bibliometric and case-study examination by sector. Optimization included a structured four-stage approach—technical viability (data availability/quality, computational capability), economic viability (cost–benefit, ROI analysis), regulatory and safety requirements, and pilot demonstration with scale-up evaluation. Previous application results indicated AI-PdM increased wind turbine lifespan by 20%, reduced maintenance costs by 25–30% for energy and water industries, and reduced breakdowns by 15–20% for transportation and telecommunication. The conclusion was that AI-based PdM is revolutionary yet necessitates rigorous consideration of technical, economic, and regulatory preparedness prior to implementation.

Kazim et al. [30] outlined a multilayer GNN architecture for PdM and resilience-based substation clustering of power grids. The dataset was the Oklahoma Gas & Electric incident dataset (2015–2021) of 292,830 events at 347 substations with 52 features. Techniques involved creating multilayer graphs with spatial (transmission lines & proximity), temporal (co-occurrence of events), and causal (correlations of failures) layers. Specialized GNNs were employed—GATv2 for spatial/temporal and GIN for causal relationships, with attention-based fusion for embedding fusion. Optimization incorporated focal loss for handling class imbalance, temporal splits for train/test, and ablation studies for verifying layer contributions. Performance revealed 30-day PdM window obtained 88.46% accuracy, 93.36% precision, and 89.35% F1-score, with 2–3% improvement over Random Forest and XGBoost and 10–15% over single-layer GNNs. For resilience, HDBSCAN clustering of embeddings revealed 8 operation groups, of which Cluster 5 (44 substations) displayed maximal risk: 388.4 incidents/year, 602.6 min recovery time, corroborated through ANOVA ($p < 0.0001$). It was concluded that multilayer GNNs successfully capture spatial-temporal-causal interdependencies, enhancing PdM precision and facilitating risk-informed substation clustering for advance grid management.

Sun et al. [31] derived an enhanced YOLOv8 model for real-time substation equipment defect detection. The dataset included 2809 self-annotated infrared and camera images with six categories of defects, such as abnormal oil gauge readings, damaged dials, and panel defects. Techniques included three important modifications: substituting the YOLOv8 backbone with lightweight feature extraction by EfficientViT, inserting a Squeeze-and-Excitation (SE) channel-wise recalibration attention module, and replacing C2f Bottlenecks with FasterBlock to suppress computations. Optimization employed ablation testing, learning rate scheduling that could adapt to dynamic learning rates, and robustness tests in poor light and occlusion. Performance indicated that the enhanced model reached 92.8% mAP50 and 80.7% mAP50-95, outperforming baseline YOLOv8n (91.0%, 78.6%) and rivaling YOLO variants, yet at only moderate computations (3.8M params, 9.1G FLOPs). The conclusion was that the combination of EfficientViT + SE + FasterBlock enhances accuracy significantly, improves efficiency, and enhances robustness, making it feasible for real-time defects to be detected in smart substations.

Li et al. [32] came up with EAL-YOLO, a lightweight defect detector for small targets in substations. The dataset comprised 12,968 State Grid Shandong Electric Power Company annotated images that cater to 12 common defect types like abnormal indicator lights, silicone damage, and oil-level window faults. Techniques enhanced YOLOv8 by combining EfficientFormerV2 + LSKA as the backbone for feature extraction, ASF2-Neck with extra P2 layer for 4×4 pixels small-target detection, and LSCHead with parameter sharing and Group Normalization to reduce redundancy. Optimizations involved data augmentation, ablation testing, and

FLOPs/parameter reduction techniques. Performance revealed mAP50 = 92.26% (+2.93% over YOLOv8n) and accurate detection of small targets = 90.8% (+5.3%), with parameters reduced by 61.17% and FLOPs by 46.5% over YOLOv8s. The finding emphasized that EAL-YOLO realizes higher precision and lightweight efficiency, making it appropriate for real-time inspection by UAV or edge-device in smart substations.
Zhang et al. [33] proposed LEAD-Net, a semantic-enhanced irregularity feature learning model for substation equipment defect observation. The dataset included 700 defect images (rust and oil leakage), 1200 normal images, and 1000 synthetic images generated through ADD-GAN because of data scarcity. Methods combined a U-Net-based generator for defect synthesis, a local defect generator for fine-grained details, and a joint discriminator to evaluate global and local regions, with perceptual and cycle-consistency loss functions for assistance. Optimization utilized Adam optimizer and ablation experiments to validate each module's effectiveness. Results showed significant performance gains: YOLOv7 mAP from 71.9% to 81.5%, precision from 81.4% to 90.3%, and F1-score from 81.1% to 88.7%. The conclusion pointed out that LEAD-Net with ADD-GAN not only improves detection accuracy but also enhances defect realism and provides a scalable solution for intelligent substation inspection where labeled data are limited.

In Table 1, we can see a summary of the key findings from the reviewed studies on AI and EMS in microgrid applications.

Table 1 : Summary of Reviewed Literature on AI/ML in Energy Management

Year	Ref.	Model	Results	Limitations
2023	[16]	AI-based fault prediction, cyber intrusion detection (MATLAB/Simulink, Python)	Detection: 25–35 ms, Recovery: 2.5–3 s, Downtime <1%	Data quality, integration with SCADA, scalability issues
2023	[17]	Condition-based monitoring (IoT sensors, predictive analytics, Siemens MindSphere)	Reduced unplanned shutdowns by 85%, cost savings S$101M, SAIDI 6.6s	High initial cost, dependence on continuous monitoring
2024	[18]	Multi-framework (LCC, RCM, RST, TPM) for equipment selection	Lifecycle cost ↓ 30%, utilization ↑ 13%, reliability improved	Complex framework, needs expert consultation
2024	[19]	Life Cycle Cost Analysis (LCCA), scenario modeling with BESS	Feasible at 300 €/kWh for 19 lines, reliability improved	Depends on falling BESS costs and regulatory support
2023	[20]	IoT-based transformer fault detection (ESP32, GSM + Wi-Fi)	Efficient detection of overheat, oil level, vibration faults	Limited to lab-scale validation, scalability concerns
2024	[21]	UAV-based inspection with CBAM-YOLOv4	Accuracy 95.12%, mIoU 78.34%, FPS 62.05	Performance drop under complex real-world noise
2023	[22]	Koopman Operator + Graph-based HIF detection	Fault detection in <0.002s, minimized false alarms	Needs high-sample-rate measurement, complex setup
2023	[23]	Predictive maintenance with ML (SVM, CNN, LSTM, Digital Twin)	Downtime ↓ 10–35%, expenses ↓ 10–20%, reliability ↑	Requires integration with CMMS/EAM, costly
2024	[24]	ML for grid stability (LogReg, RF, SVM, Boosting)	Gradient Boosting achieved 99.2% accuracy	Dataset synthetic; generalization to real-world uncertain
2023	[25]	AI for transformer predictive maintenance (CNN, RNN, GBoost)	Precision 96.55%, Recall 94.2%, Accuracy 90.44%	Dataset imbalance, requires IoT real-time deployment
2024	[26]	Wavelet features + Random Forest for CTs	94% accuracy, 100% healthy, 84% (1HCF), 67% (2HCF)	Poor performance on complex faults, limited data
2023	[16]	AI-based fault prediction, cyber intrusion detection (MATLAB/Simulink, Python)	Detection: 25–35 ms, Recovery: 2.5–3 s, Downtime <1%	Data quality, integration with SCADA, scalability issues
2023	[17]	Condition-based monitoring (IoT sensors, predictive analytics, Siemens MindSphere)	Reduced unplanned shutdowns by 85%, cost savings S$101M, SAIDI 6.6s	High initial cost, dependence on continuous monitoring

3. Methodology

The methodology of this study is designed to establish a rigorous and reproducible framework for substation asset-type classification. The process begins with a comprehensive data preprocessing pipeline that ensures data quality through cleaning, missing-value imputation, temporal feature extraction, domain-driven feature engineering, and categorical encoding. Stratified splitting preserves class balance across training, validation, and test subsets, while augmentation and oversampling are applied specially to the training data to address class imbalance. Building upon this prepared dataset, a suite of baseline models including Random Forest, LightGBM, Multi-Layer Perceptron, and Logistic Regression was implemented to provide comparative performance benchmarks. Finally, a suggestion of hybrid stacking ensemble was also put forward, and it would take the prognosticative power of the RF, LightGBM, and MLP by a meta-learner based on the Logistic Regression. The present approachology provides a solid foundation upon which it can be stated that the traditional method compares with the proposed framework in terms of its potential to be enhanced in generalization and prediction of the pertinent area of asset management.

3.1 Dataset Description

The dataset that will be employed in this study is the Smart Grid Asset Monitoring Dataset, [39] which is released as free on Kaggle and provided with all the telemetry and operational data of a smart grid substation. It combines both numeric sensor measurements and categorical metadata in order to describe the behavior of the heterogeneous assets of a substation in different operation conditions. The data is registered with voltage, current, power, frequency, and energy consumption as well as identifiers of the assets, load characteristics, and timestamps. The target variable will be Asset_Type which will categorize every record as either Transformer, Switch or SmartMeter. This is a multi-faceted and well-organized data set which will be a strong basis to make and compare machine learning models of predictive analytics and smart assets management in substations.

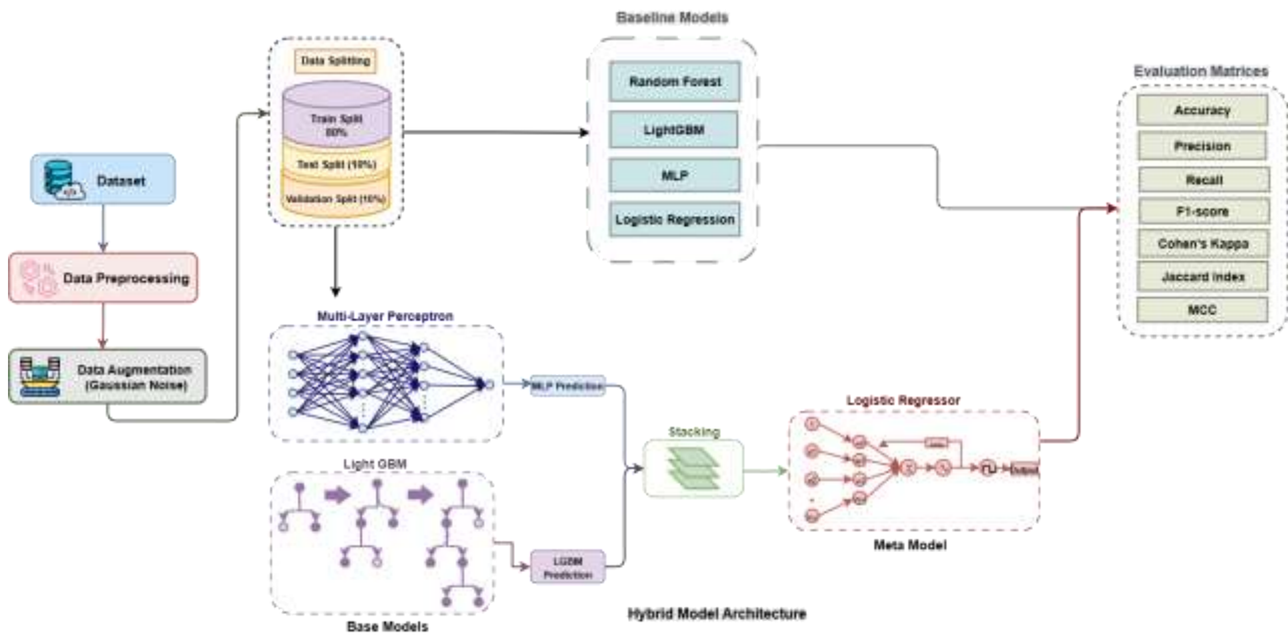


Figure 1: Workflow of the Proposed Framework

Table 2: All Dataset Features and Descriptions

Feature	Type	Description
Voltage_V	Numeric	Measured voltage for the asset (volts).
Current_A	Numeric	Measured current (amperes).
Power_kW	Numeric	Real power consumption/load (kilowatts).
Frequency_Hz	Numeric	Frequency of the electrical signal (hertz).
Energy_Consumed_kWh	Numeric	Cumulative energy consumed (kilowatt-hours).
Timestamp	Datetime	Time at which measurement was taken.
Substation_ID	Categorical	Identifier for each substation.
Asset_ID	Categorical	Unique identifier for each asset within substation.
Load_Type	Categorical	Type of load (e.g., residential, industrial, commercial).
Asset_Type (Target)	Categorical	Type/class of the asset (Transformer, Switch, SmartMeter).

The features of the dataset that were used in this research are reviewed in Table 2. The data combines quantitative sensor values, including voltage, current, power, frequency, and energy consumption, and qualitative values, including substation and asset numbers, the type of loads, and time. Asset_Type is the target variable that will assign every record to Transformer, Switch or SmartMeter. This rich association of numerical and categorical characteristics makes it possible to represent the behavior of assets comprehensively, creating a solid foundation of predictive modeling and classification.

3.2 Data Preprocessing

The data employed was sufficiently so as to be able to state that the data employed in the research was as good as good with as much as it was in the nature of an adequate measure of or a good reflection of the reality of a given working environment. All steps have been taken to solve dedicated issues in the dataset and prepare the features to create competent classification of the substation assets.

Data Cleaning & Type Casting

Raw telemetry records often contain inconsistent formats or duplicate entries that can distort downstream learning. To mitigate these issues:

- The $Timestamp$ field was converted into a consistent datetime object to allow accurate temporal feature extraction.
- Numeric fields (Voltage_V, Current_A, Power_kW, Frequency_Hz, Energy_Consumed_kWh) were explicitly cast into numeric types to prevent parsing errors.
- Duplicate records were removed based on composite keys (Substation_ID, Asset_ID, Timestamp) to avoid bias caused by repeated observations.

This step ensures that the analytical pipeline operates on clean, standardized inputs and prevents spurious redundancy from inflating model performance.

Missing-Value Handling

Incomplete data can severely degrade predictive accuracy, especially when faults or rare classes are already underrepresented. To address this:

- Categorical attributes such as Fault_Event and Reconfig_Action were imputed with sentinel values (“No Event” and “No Action”), preserving semantic meaning without discarding samples.
- Numeric sensor values were imputed using the median:

$$x_i^{imputed} = \begin{cases} x_i, & if\ x_i \neq NaN \\ \tilde{x}, & if\ x_i\ \neq NaN \end{cases} \tag{1}$$

where $\tilde{x}$ denotes the median of the feature.

This robust strategy preserves distributional properties even in the presence of outliers, ensuring no artificial bias is introduced during imputation.

Time-Derived Features

It matters that the time properties are relevant because the demand and the reaction of equipment to time generally possess a pattern variation. To portray such patterns, two properties of time were obtained with respect to Timestamp attribute of the data. The calculation of the hour of the day has been particularly made to accommodate the diurnal requirement variation and the day of week variation and lastly, the main day of the year was calculated to accommodate the seasonal effects of the load and stress of the asset. Having these properties of time, the models would be able to learn periodic trends, such as peak unprecedented time, e.g. evening hours or peak demand seasonal, which is essential to more effective classification of assets in a predictable finite form and superior prioritization of predictive analytics in substation space.

Domain-Driven Feature Engineering

In addition to raw sensor values, engineered variables were introduced to encode domain knowledge:

Apparent Power (VA):

$$S = V \times I \tag{2}$$

where V is the measured voltage and I is the current. This represents the complex power magnitude of the asset.

Power Ratio:

$$R = \frac{S}{P+\epsilon} \tag{3}$$

where S is apparent power, P is real power ($Power\_kW$), and ϵ is a small constant to prevent division by zero. This ratio differentiates assets by efficiency and operational load profile.

Categorical–numeric interactions:

selective multiplications between encoded load/asset flags and electrical parameters were introduced to capture heterogeneous operating behavior. By encoding physical relationships, these engineered features provide discriminative signals that assist the classifier in distinguishing among asset types.

Figure 2 illustrates the top five features contributing to asset type prediction. The analysis highlights that categorical identifiers of assets (Asset_ID for Switch, Transformer, and SmartMeter) provide the strongest discriminative power, reflecting the inherent uniqueness of each equipment class. Among the numerical features, Current_A and Energy_Consumed_kWh emerge as the most influential, indicating their relevance in capturing operational load and consumption behavior. Together, these features enable effective differentiation across heterogeneous substation assets.

Feature Encoding

For effective utilization of machine learning models, categorical attributes must be transformed into a numerical representation. In this study, categorical variables such as $Substation\_ID$, $Asset\_ID$, and $Load\_Type$ were converted using one-hot encoding, thereby generating binary indicator variables for each distinct category. This transformation enabled the models to capture the unique identity of substations and assets without introducing spurious ordinal relationships among categories. Importantly, the target variable $Asset\_Type$ was deliberately excluded from encoding to prevent information leakage. By ensuring a robust and leakage-free encoding strategy, the preprocessing pipeline provided the classifier with a rich yet unbiased feature space, supporting accurate and generalizable asset-type prediction.

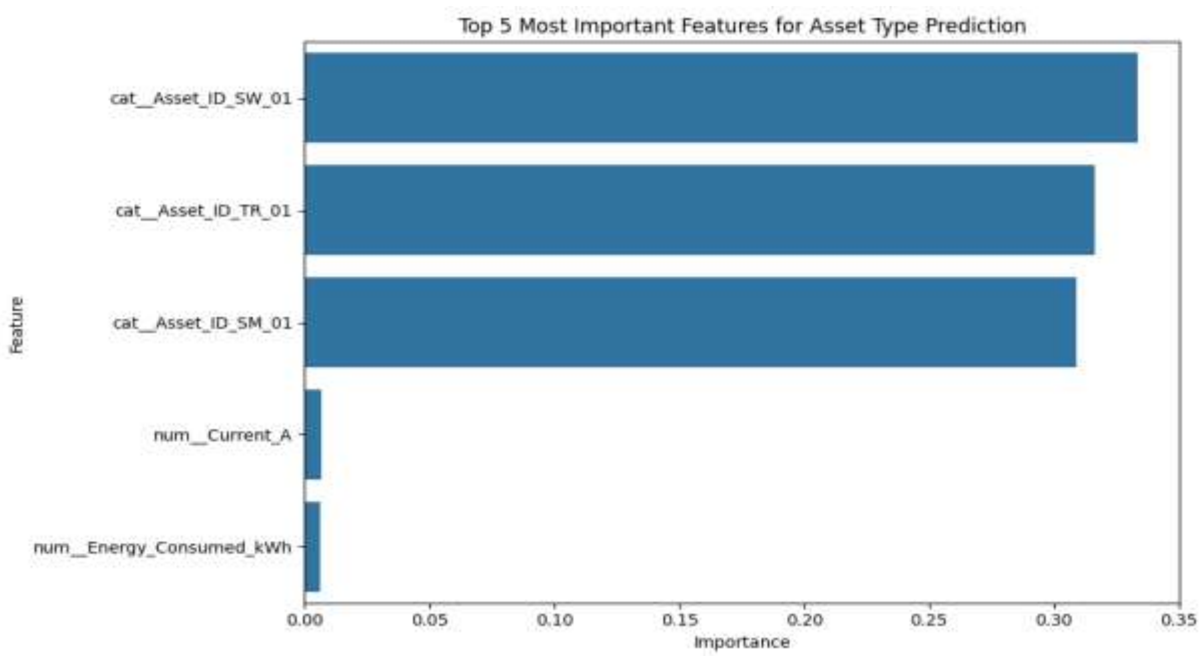


Figure 2: Top 5 Most Important Features For Asset Type Prediction

Data Splitting

To ensure unbiased model evaluation, the dataset was partitioned into training (80%), validation (10%), and test (10%) subsets. Stratification preserved the proportion of each asset class ({Transformer, Switch, SmartMeter}) across all splits:

$$\frac{n_c^{train}}{N^{train}} \approx \frac{n_c^{val}}{N^{val}} \approx \frac{n_c^{test}}{N^{test}} \tag{4}$$

Where n_c is the count of the class c, and N denotes total samples per split.

This ensures fair representation of minority classes and allows consistent performance assessment across subsets.

Table 3 below provides an overview of asset distribution of distinct training, validation sets, and test sets. The data has three key components SmartMeter, Switch, and Transformer, which however are represented by the ratios of 2 splits being similar. Specifically, the SmartMeter, Switch and Transformer data are 271, 292, and 301 in training set and validation, and test sets respectively to ensure that there are no biased judgment. The percentage of classes per subroom remains constant in this categorized division, thereby reducing sampling bias and facilitating the strong training of models along with the performance testing in the asset-type categorization exercise.

Table 3: Distribution of Asset Types Across Datasets

Asset_Type	**Training**	**Validation**	**Test**
SmartMeter	271	90	91
Switch	292	98	97
Transformer	301	100	100

Data Augmentation and Imbalanced Handling

Once the dataset had been divided into training, validation, and test portions, augmentation was performed only on the training data, and not on the validation or the test portion of the dataset, to ensure biased evaluation is prevented. Light Gaussian noise was introduced during the augmentation step to numeric sensor characteristics of small asset classes to artificially increase their volume without changing underlying trends. This was followed by a random oversampling, in order to further equalise the rate of the classes' distribution among the sampled records, by repeatedly duplicating samples of the minority class groups until a similar number of records per class was attained. The use of this two-stage process augmentation, then oversampling ensured that the training set offered an appropriate, balanced, and enriched training signal that would improve the generalization ability of the model across all the types of assets and preserve the integrity of the validation and test sets.

Noise-based augmentation: light Gaussian noise was injected into numeric sensor readings for underrepresented classes:

$$x' = x + N(0, \sigma^2), \qquad \sigma = 0.01 \times \mu_x \tag{5}$$

Where μ_x is the mean of the feature.

Random Oversampling: minority class samples were randomly replicated until each class reached approximately the same size.

$$n_c^{aug} \approx \max_k(n_k) \tag{6}$$

This balanced training distribution enhanced the model's ability to generalize to all asset classes, especially rare categories like *SmartMeter*.

Figure 3 illustrates the effect of Gaussian noise augmentation on the training data across the three asset classes: Transformer, Switch, and SmartMeter. Before augmentation, the dataset contained only a few hundred records per class, creating a significant imbalance for model training. After applying Gaussian noise, each class was expanded to approximately 4,000 samples, resulting in a well-balanced distribution. This augmentation step enriched the minority classes, reduced the risk of bias toward dominant patterns, and provided the learning algorithm with a more diverse representation of asset behavior.

Figure 4 presents the distribution of asset classes before and after applying balancing techniques. Initially, the dataset showed slight disparities among the three classes—Transformer, Switch, and SmartMeter—with SmartMeter having comparatively fewer samples. To address this issue, random oversampling was applied to the training data, ensuring that each class contained an equal number of records. After balancing, the dataset achieved near-uniform distributions across all asset categories. This adjustment minimized class bias, strengthened fairness in model learning, and improved the classifier's ability to generalize across diverse asset types.

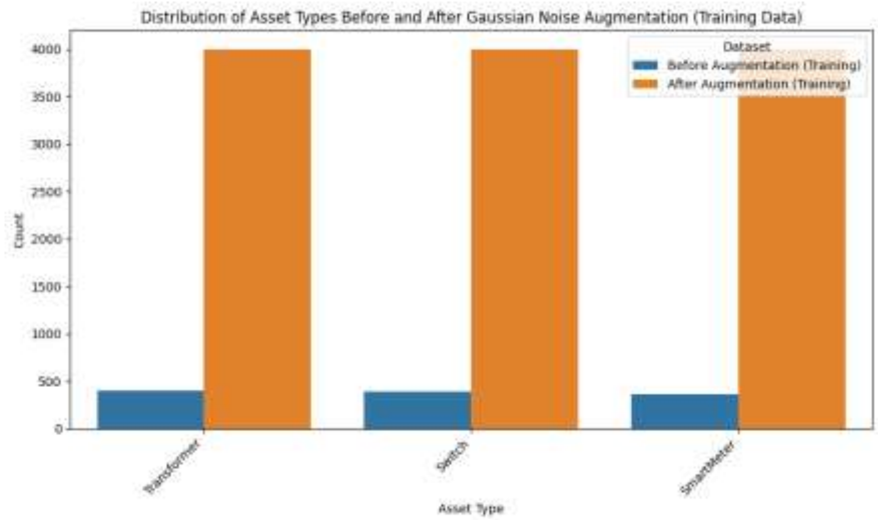
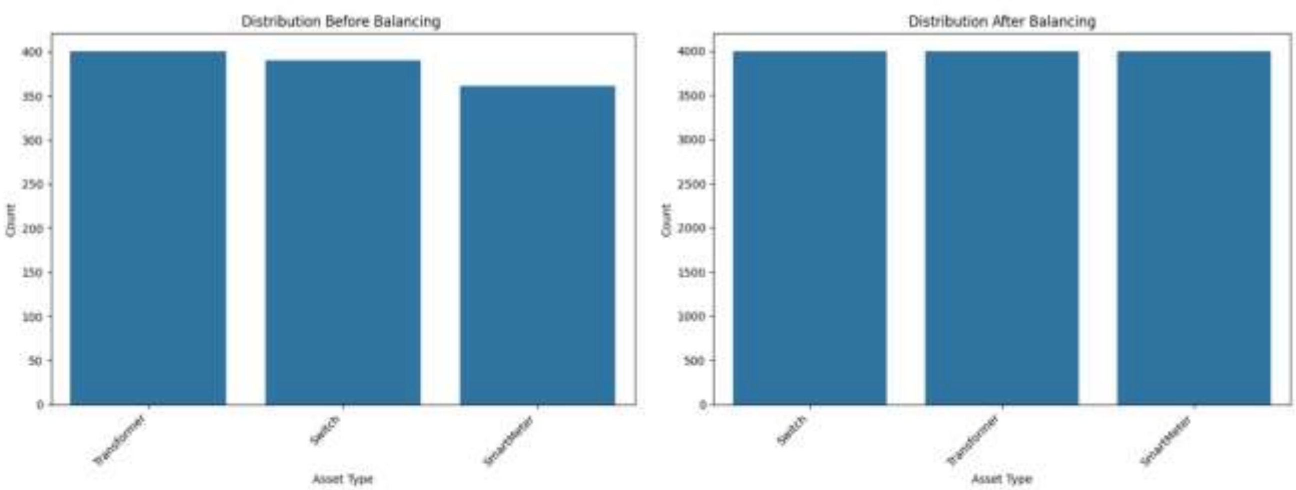


Figure 3: Before and After Gaussian Noise Augmentation for Training Data.



(a) Before Balancing the Data (b) After Balancing the Data

Figure 4: Distribution of Before and After Balancing Data

Table 4: Final Distribution of Splitting After Training Augmentation

Asset_Type	**Original Training**	**Augmented Training**	**Validation**	**Test**
SmartMeter	271	4000	90	91
Switch	292	4000	98	97
Transformer	301	4000	100	100

Table **4** also documents the ultimate allocation of asset types through the augmented and split training, validation, and test datasets. The three classes of FDI (SmartMeter, Switch, and Transformer) had 271, 292, and 301 samples in the original training set, respectively. This was fixed with addition of Gaussian noise augmentation and random oversampling that enhanced learning of models by 4,000 records per training classes to the extent of equalizing the classes. Importantly, validation and a test sample remained untouched as well, and each of the classes is approximately represented by 90 -100 samples, which guaranteed the objective environments of assessment. Its design has made the training set to be a complete and balanced model of all the assets, and validation and testing sets are in an authentic position, which they apply in direct assessment of performance besides discouraging hypocritical inflation of model excellence.

In short, preprocessing pipeline provided the cleansed, balanced and representative dataset. The problems of missing data and class imbalance were successfully overcome with the help of systematic cleaning, imputation, feature engineering, encoding, and augmentation that were performed to the training set only. The strong preparation was used as a solid basis of training accurate and generalizable models of asset-type classification.

3.3. Proposed Hybrid Model

The proposed architecture is a stacking ensemble that represents a hybrid of various learners as a predictor in a single architecture. On the lower level, two complementary models co-exist to generate various predictive signals. The (MLP) is a neural unit, with interconnected hidden layers with non-linear activation functions to extract higher-level feature representations and learn complicated relations in the data. Meanwhile, the Light Gradient Boosting Machine (LightGBM) is a boosting algorithm. It constructs clusters of decision trees which successively minimize loss functions and subdivide the feature space into hierarchical decision rules. The MLP gives the opportunity to learn non-linear relationships that are complex, whereas LightGBM is effective to work with tabular data in structure. The combination of these forms an effective base-learning layer

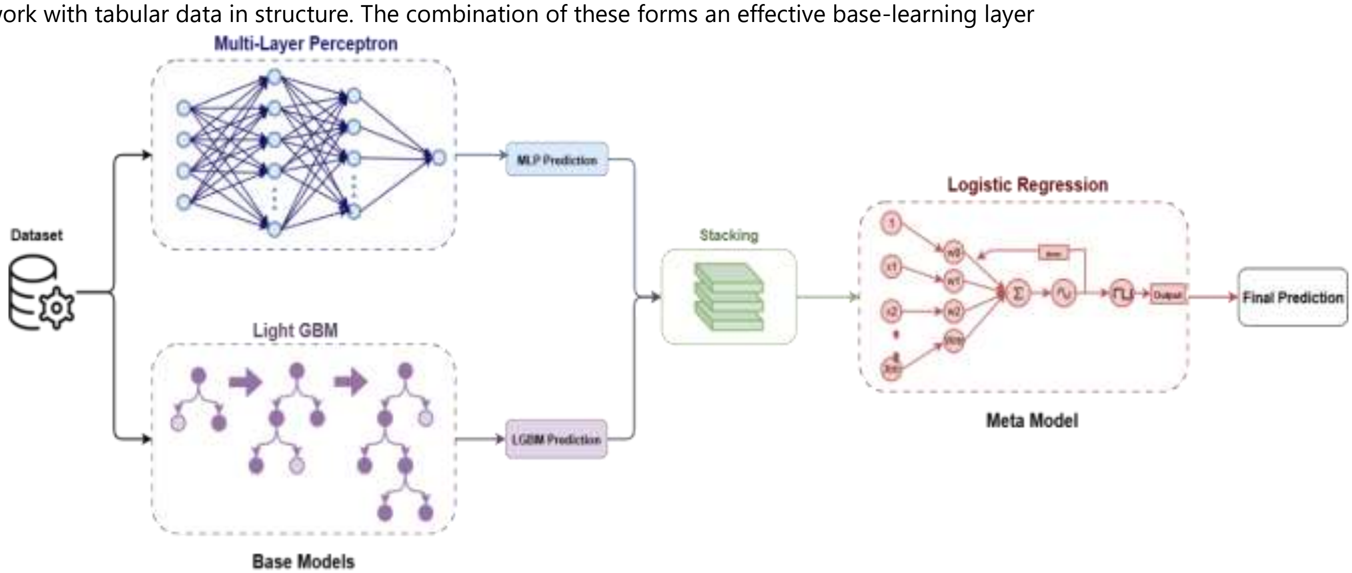


Fig-5: Architectural view of proposed Hybrid MLP–LightGBM–Logistic Regression model

The outputs of these base models are fed to the stacking layer which consists of a Logistic Regressor which is in turn the meta-model. This meta-learner accepts probabilistic forecasts of LightGBM and MLP. It is taught how to integrate them into a final predictive decision in the best way possible. The model has accuracy and robustness by combining the neural representation learning with boosting-based rule extraction in a transparent stacking ensemble. The general structure of the proposed hybrid model is illustrated in figure 5. It depicts the input dataset to base learners, stacking of integration and the final prediction layer. It is a predictive framework that provides a useful framework in managing substation assets. It offers decision support based on data and enhances decision-making within the critical infrastructure systems.

3.3.1 Non-Linear Representation through the Multi-Layer Perceptron

The MLP is used in the proposed hybrid stacking ensemble as a base learner, which brings non-linear predictive power to the system. Not only does its job consist of a mapping of the raw feature vectors into a higher order representational space but it also produces probability distributions that reflect complex interdependencies among features. These probabilities are then used as meta-features during the stacking stage. The architecture view of MLP is presented in figure 6.

The input to the MLP can be denoted as a preprocessed feature vector:

$$x = [x_1, x_2, \dots, x_d]^\top \in R^d \tag{7}$$

where d is the dimensionality of the feature space after preprocessing. Each component of $\boldsymbol{x}$ may represent a normalized numerical attribute or an encoded categorical attribute.

The first hidden layer applies an affine transformation followed by a non-linear activation:

$$h^{(1)} = \sigma(W^{(1)}x + b^{(1)}, \tag{8}$$

where $W^{(1)} \in R^{(m_1 \times d)}, b^{(1)} \in R^{m_1}, \ and \ \ \sigma(\cdot)$ is the ReLU activation. Here, $m1 = 128$ in our configuration, which enables the network to capture a wide range of feature interactions.

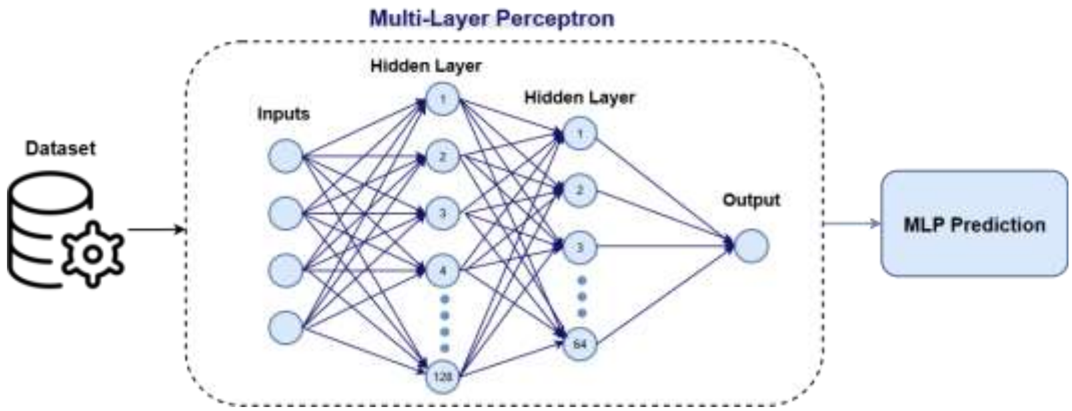


Fig-6: Multi-Layer Perceptron (MLP) model architecture.

The second hidden layer refines this representation:

$$h^{(2)} = \sigma(W^{(2)}h^{(1)} + b^{(2)}), \tag{9}$$

With $m2 = 64$ neurons, reducing and distilling information and maintaining important non-linear associations. Each transformation introduces a level of abstraction, where simple feature combinations at the first hidden layer evolve into more complex, high-level interactions in the subsequent layers.

The output of the final hidden layer is projected into the logit space:

$$z = W^{(o)}h^{(2)} + b^{(o)}, \tag{10}$$

Where $z \in R^K \ and \ K$ is the number of possible prediction categories. These logits are converted into probabilities through the softmax function:

$$p(\, y = k \mid x\,) = \frac{\exp(\mathrm{z_k})}{\sum_{j=1}^{k} \exp(z_j)}\,, \quad k = 1,2,\dots,K. \tag{11}$$

Thus, the MLP produces a probability vector

$$\boldsymbol{p}^{MLP}(\boldsymbol{x}) = [p(y = 1 \mid x), \dots, p(y = K \mid x)], \tag{12}$$

which encodes the network's belief distribution over the possible outcomes.

Within the hybrid ensemble, this probability vector is not the final decision. Instead, it forms one part of the stacked feature set. Through F-fold cross-validation, out-of-fold predictions are generated for each training instance to avoid bias:

$$p_i^{MLP} = softmax\big(\phi_{\theta(-f)(x_i)}\big), \ i \in I_f, \tag{13}$$

where $\theta^{(-f)}$ denotes parameters learned without fold f. These unbiased probability vectors are then concatenated with the corresponding outputs of LightGBM:

$$s_i = \left[p_i^{MLP}; p_i^{LGBM}\right] \in R^{2K}. \tag{14}$$

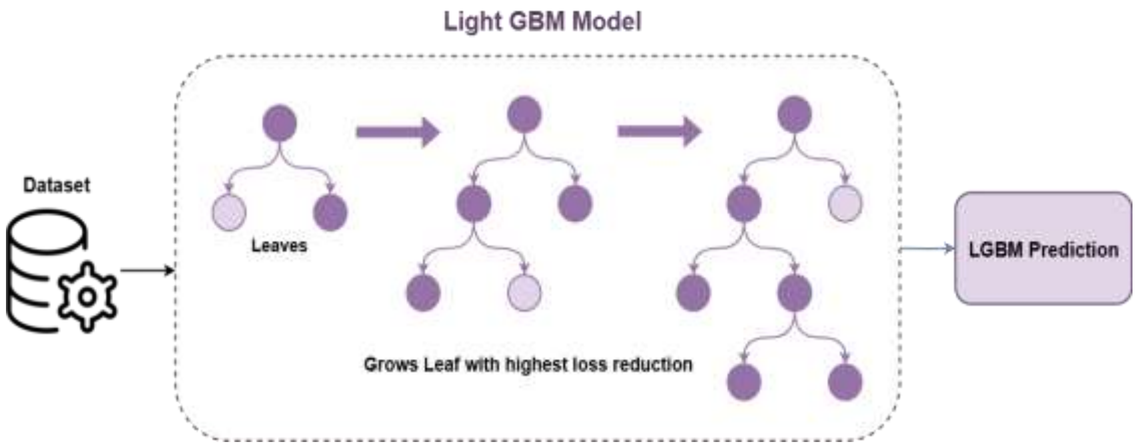
This rich feature independent are stacked through the stacking ensemble to the Logistic regressor that effectively concatenates complemental information. In this way, the MLP provides the hybrid model the ability to fit the data to complex and non-linear data trend, hence, ensuring that the downstream prediction inherits the presence of the representation that LightGBM will fail to capture.

3.3.2. Rule-Based Probability Generation with LightGBM

Even though Multi-Layer Perceptron is used to provide the non-linear representation learning in the hybrid system, the other added feature is Light Gradient Boosting machine (LightGBM) that has created an additive perspective of the rule. LightGBM divides sequentially the input space defined by building a sequence of decision trees. It executes the major duties of the hybrid structure, specifically to present methodized and regulated probability vectors being forecastive of the connection between features regarding gradient-based segmenting of the feature space. The model architecture of LightGBM is illustrated in figure 7.

Fig-7: Light GBM (LGBM) model architecture.

For a dataset $\{(x_i, y_i)\}_{i=1}^{N}$ with input features $x_i \in R^d$ and labels y_i, LightGBM builds an additive model consisting of T regression



trees:

$$\hat{y}i = \sum_{t=1} f_t\,(x_i), \qquad f_t \in F, \tag{15}$$

where each f_t is a tree from the functional space F. At each boosting iteration, the algorithm minimizes a regularized objective function that balances prediction accuracy with model complexity:

$$L^{(t)} = \sum_{i=1}^{N} l(y_i,\, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{16}$$

where $l(\cdot)$ is a differentiable loss function (e.g., cross-entropy for classification), and $\Omega(f_t)$penalizes the complexity of the tree f_t.

To optimize this objective efficiently, LightGBM employs a second-order Taylor expansion of the loss, making use of both the first- and second-order derivatives of the objective. For each tree, the loss approximation can be expressed as:

$$\tilde{L}^{(t)} = \sum_{j=1}^{j}[G_j w_j + \frac{1}{2}\,(H_j + \gimel) w_j^2] + \gamma J, \tag{17}$$

Where J is the number of leaves in the tree, $G_j = \sum_{i\varepsilon I_j} g_i \; and \; H_j = \sum_{i\varepsilon I_j} h_i$ are the first- and second-order gradient statistics accumulated for the instances assigned to leaf j, w_j is the weight of the leaf j, ℷ is the regularization parameter, and γ controls leaf creation.

The optimization yields a set of trees that partition the feature space hierarchically and assign outputs based on aggregated gradient signals. Once the ensemble has been trained, its output logits $\hat{y}_i$ are mapped into class probabilities using the softmax transformation:

$$P^{LGBM}(y = k \,|x_i) = \frac{\exp(\hat{y}_{i,k})}{\sum_{j=1}^{K}\exp(\hat{y}_{i,j})}, \quad \text{k=1...,K.} \tag{18}$$

This produces the LightGBM probability vector for each instance:

$$P^{LGBM}(x_i) = [p^{LGBM}(y = 1\,|x_i), \ldots, p^{LGBM}(y = K\,|x_i)]. \tag{19}$$

In the hybrid stacking model, these probability vectors serve as meta-features, which are concatenated with the outputs of the MLP:

$$\boldsymbol{S}_i = [P^{MLP}(x_i); P^{LGBM}(x_i)]. \tag{20}$$

By supplying structured, gradient-informed probabilities, LightGBM ensures that the meta-learner receives a diverse set of signals, improving the stability and predictive power of the overall ensemble.

3.3.3. Stacking Integration

The stacking serves as the integration mechanism that transforms the outputs of the base learners into a unified meta-representation. For each training instance x_i, the MLP produces a probability vector

$$P_i^{MLP} = [P^{MLP}\ (y = 1|x_i), \dots, P^{MLP}(y = K|\ x_i)], \tag{21}$$

and LightGBM generates a complementary probability vector

$$P^{LGBM}(x_i) = [p^{LGBM}(y = 1\ |x_i), \dots, p^{LGBM}(y = K\ |x_i)]. \tag{22}$$

The stacking layer concatenates these outputs to form a joint meta-feature:

$$s_i = \left[p_i^{MLP}; p_i^{LGBM}\right] \in \mathbb{R}^{2K}. \tag{23}$$

To ensure unbiased learning, cross-validation is employed so that the stacked features are constructed from out-of-fold predictions:

$$p_i^{MLP} = softmax\left(\phi_{\theta(-f)}(x_i)\right),\ \ P_i^{LGBM} = softmax(\psi_{\eta^{(-f)}}\ (xi), \quad i \in I_f \tag{24}$$

where $\phi_{\theta(-f)}\ and\ \psi_{\eta^{(-f)}}$ denote models trained without fold f.

Thus, the stacking process constructs a new feature space $\{S_i\}_{i=1}^{N}$, where each data point is represented not by its raw attributes but by the probabilistic outputs of diverse base learners. This meta-representation captures both non-linear neural mappings and rule-based partitioning, forming the foundation upon which the final meta-model operates.

3.3.4. Logistic Regression as the Meta-Learner

Once the probability vectors of LightGBM and MLP are stacked together through the stacking layer, they undergo the following step: the meta-features $S_i \in \mathbb{R}^{2k}$ are created (they form the input of the meta-learner). The rationale behind choosing the Logistic Regression in this spot is that it is stable and has the ability to integrate heterogeneous signals linearly. It serves to ascertain the degree to which weight of any of the base learners predictions should be assigned to arrive at the final decision.

For each training instance, Logistic Regression estimates the conditional probability of class k as:

$$p(\,y = k \mid s_i\,) = \frac{\exp(\,w_k\, s_i + b_k)}{\sum_{j=1}^{K} \exp(w_j s_i + b_j)}, \qquad k = 1, \dots\dots, K, \tag{25}$$

where w_k are the learnable weights and b_k the bias terms. These parameters capture the influence of each component of the stacked vector—some dimensions correspond to MLP probabilities, others to LightGBM probabilities so Logistic Regression effectively learns which model to trust more for each class.
The model is trained by minimizing the multinomial cross-entropy loss:

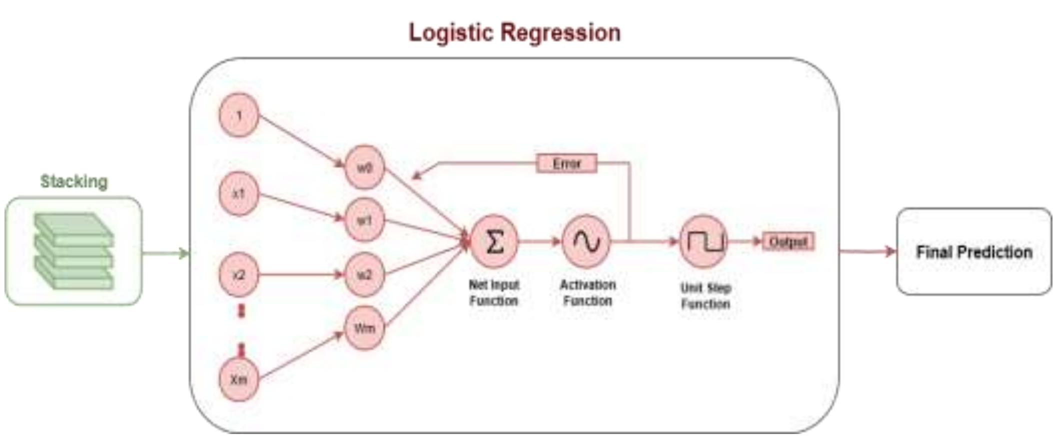


Fig-8: Light GBM (LGBM) model architecture.

$$\mathcal{L}_{LR} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} 1\ [y_i = k] log p(y = k\ |s_i), \tag{26}$$

ensuring that the predicted probabilities align with the true class labels.

Thus, Logistic Regression acts as the final arbiter of the hybrid system: it takes in the stacked features from the base learners, evaluates their combined evidence, and outputs the final prediction, shown in figure 8.

In general, the hybrid stacking ensemble provided will entail the combination of the proactive synergistic quality of three major areas into rational construct. The MLP offers non-linear feature representation, LightGBM creates the structured distributions of probabilities during the utilization of gradient-boosted decision trees, and Logistic Regression is the meta-learner, which combines the outputs to its final decision. By this, the architecture is much more accurate, stronger and generalized than the individual learners. The synergy is the only thing that should make the provided model possible and relevant to the predictive activities within the framework of the management of the substation assets.

Algorithm 1: Training the Proposed Hybrid Model

1. Initialize base learners (f_{MLP}, f_{LGBM}) and meta-learner g_{LR}
2. Split dataset D into training, validation, and testing subsets.
3. Partition training indices into $F\ folds: I_1, \dots, I_F$.
4. For each fold *f=1…F :*
   a) Train $f_{MLP}\ and\ f_{LGBM}\ on\ U_{r\neq f}\ I_r$
   b) Obtain out-of-fold probabilities for validation indices I_f:
   $$p_i^{MLP} = softmax\left(\phi_{\theta(-f)}(x_i)\right),\ \ P_i^{LGBM} = softmax(\psi_{\eta^{(-f)}}\ (xi), \quad i \in I_f$$
   c) Store predictions in OOF matrices P_{MLP}, P_{LGBM}.
5. Construct stacked feature matrix:
   $$s_i = [p_i^{MLP}; p_i^{LGBM}], \quad \forall_i.$$
6. Train Logistic Regression g_{LR} on stacked features $\{s_i, y_i\}$.
7. Refit base learners f_{MLP}, f_{LGBM} on full training set.
8. Validate on held-out test set using accuracy, precision, recall, F1, and other metrics.
9. Return final hybrid model $\hat{g} = (f_{MLP}^*, f_{LGBM}^*, g_{LR}^*)$.

Table 4. Hyperparameter configuration for the Hybrid MLP–LightGBM–Logistic Regression model

Hyperparameter	Value
Batch size (MLP)	64 (fixed mini-batch for stable optimization)
Loss function	Categorical Cross-Entropy (applied for probability outputs)
Optimizer (MLP)	Adam optimizer with learning rate control
Hidden layer sizes (MLP)	Two layers with 128 and 64 neurons respectively
Activation function (MLP)	Rectified Linear Unit (ReLU) applied at hidden layers
Regularization (MLP)	L2 penalty with coefficient $\alpha = 1 \times 10^{-4}$
Maximum iterations (MLP)	300 epochs for convergence
n_estimators (LightGBM)	600 boosting rounds for tree ensemble construction
Learning rate (LightGBM)	0.05 (step size shrinkage in boosting)
Maximum depth (LightGBM)	-1 (no fixed limit; controlled by num_leaves)
Number of leaves (LightGBM)	63 (controls complexity of each tree)
Subsample ratio (LightGBM)	0.9 (row sampling for variance reduction)

Feature fraction (LightGBM)	0.9 (column sampling per tree)
Stacking folds (CV)	5-fold cross-validation to build out-of-fold predictions
Meta-learner	Logistic Regression trained on stacked features

3.4 Baseline Models:

In order to offer a wide baseline for the proposed hybrid stacking framework, a variety of baseline models were performed. These theories are linear models, ensemble models, and neural networks. They provide inductive biases specific to each baseline that can be evaluated by their interpretability, scale, and power of non-linearity.

Random Forest Classifier

Random Forest improves accuracy by employing a bagging strategy. Specifically, multiple trees are trained on bootstrap samples of the training data, and random subsets of features are considered at each split to induce diversity.

Given B trees, the ensemble prediction is:

$$\hat{y}(x) = mode\{T_b(x)\}_{b=1}^{B} \tag{27}$$

where $T_b(x)$ is the prediction of the b^{th} tree.

Each tree partitions the space, but the aggregation across trees reduces variance. The randomness in feature selection also prevents strong correlation between trees, improving robustness.

The probability output for class ccc can be expressed as:

$$P(y = c|x) = \frac{1}{B}\sum_{b=1}^{B} 1\{T_b(x) = c\} \tag{28}$$

This probabilistic ensemble makes Random Forest a reliable baseline with improved generalization on heterogeneous substation data.

LightGBM Classifier

LightGBM is a gradient boosting algorithm that builds an additive ensemble of decision trees sequentially. At each iteration m, a new weak learner $h_m(x)$ is added to minimize a differentiable loss function:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x), \tag{29}$$

Where $F_m(x)$ is the model at iteration m and η is the learning rate.

For classification, the objective is usually cross-entropy loss:

$$L = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \log P(y = c \backslash x_i) \tag{30}$$

LightGBM differs from classical gradient boosting in two ways:

- Leaf-wise growth: rather than splitting level-by-level, it grows trees by selecting the leaf with the largest loss reduction.
- Histogram-based optimization: continuous features are bucketized into discrete bins, greatly reducing computation and memory usage.

This makes LightGBM extremely efficient for large-scale structured data such as smart grid telemetry, while still capturing non-linear feature interactions.

MLP

The MLP is a fully connected feed-forward neural network designed to approximate complex, non-linear functions. Given input $x \in \mathbb{R}^d$, the first hidden layer computes:

$$h^{(1)} = f\left(W^{(1)}x + b^{(1)}\right), \tag{31}$$

Where $W^{(1)}$ and $b^{(1)}$ are trainable parameters and $f(.)$ Is a non-linear activation such as ReLU.

For layer l, the transformation is:

$$h^{(1)} = f(W^{(1)}\, h^{(l-1)} + b^{(1)} \quad (32)$$

The final output layer applies the softmax function to estimate class probabilities:

$$P(y = c|x\,) = \frac{\exp(z_c)}{\sum_{k=1}^{C}\exp(z_k)}. \quad (33)$$

MLPs are universal function approximators and can model high-order feature interactions. However, they are sensitive to hyperparameter choices and prone to overfitting when training data is limited, necessitating regularization and careful optimization.

Logistic Regression

LR is a generalized linear model that approximates the likelihood of class membership using a logistic (sigmoid) function. For binary classification:

$$P(y = 1|x) = \frac{1}{1+e^{-(\beta_0+\beta^T x)}}. \quad (34)$$

For multiclass classification with K classes, the softmax function is used:

$$P(\,y = c \mid x\,) = \frac{\exp(\beta_c^T x)}{\sum_{k=1}^{K}\exp(\beta_k^T x)}. \quad (35)$$

The parameters β are estimated by minimizing the negative log-likelihood:

$$L(\beta) = -\sum_{i=1}^{N}\sum_{c=1}^{K}\; y_{i,c}\; logP(y = c\;|x_i). \quad (36)$$

Although limited to linear decision boundaries, Logistic Regression is computationally efficient, interpretable, and provides a baseline against which non-linear and ensemble methods can be compared.

In summary, the baseline models collectively provide a diverse foundation for evaluating substation asset-type classification. From simple linear classifiers such as Logistic Regression to ensemble approaches like Random Forest and LightGBM, and non-linear architectures such as MLP, these models highlight the spectrum of trade-offs between interpretability, scalability, and predictive power. Their comparative performance establishes a rigorous benchmark against which the efficacy of the recommended hybrid stacking model can be clearly demonstrated.

4 Experimental Results and Discussion

4.1 Experimental Setup

The experiment was conducted on Google Colab with Python implementation running on a T4-GPU settings to accelerate the training of the DL models. The implementation utilized some core libraries, including Scikit-learn for preprocessing the data and baseline models, PyTorch for DL components, and PyTorch Geometric for processing structured data operations.This reinforced model creation speed besides reproducibility and access to acceleration run on a graphics card.

In order to compare the model with the standard counterparts, a series of baselines were used, including KNN, LR, RF, LightGBM, and Multilayer Perceptron (MLP). The baselines are standard practices applied in energy system forecasting and optimization, and they provide a robust benchmark for comparison with the hybrid approach. The hybrid model is a StackingClassifier, a combination of two underlying classifiers: Multilayer Perceptron (MLP), and LightGBM (LGBM) classifier. This is supported by a Logistic Regression classifier acting as the meta models because they learn how to optimally integrate the work of the base models to do a better work. The two base models, MLP and LGBM are selected because they are suitable to capture complexes and MLP addresses non-linear relationships and LGBM addresses structured feature interactions. For data preprocessing, the dataset was split consecutively into training, validation, and test sets to sustain temporal integrity and avoid data leakage. Scikit-learn was used for preprocessing tasks, including SimpleImputer for handling missing values, StandardScaler for scaling numerical features, and OneHotEncoder for encoding categorical variables. The StackingClassifier was trained using 5-fold cross-validation, ensuring a robust evaluation and reducing overfitting risks. All models were tuned using uniform search budgets for hyperparameter optimization, ensuring fair comparison across all approaches. The performance of the models was assessed using a range of classification metrics, including accuracy, precision, recall, F1-score, and the full classification report, which provides detailed metrics for each class. The model's performance was evaluated on the test set, and results were averaged over multiple runs to mitigate variability. To ensure reproducibility, experiments were conducted with fixed random seeds, and confidence intervals were calculated to report the statistical significance of the results.

Finally, the evaluation also included error analysis to understand the types of misclassifications the model encountered, along with an assessment of the economic execution and sustainability of the model for potential deployment in real-world scenarios. This comprehensive experimental setup allowed for a thorough evaluation of the hybrid model's predictive capabilities, robustness, and practicality in handling complex classification tasks.

4.2 Confusion Matrix Analysis

The confusion matrix below represents a full picture of the hybrid model interpretation in identifying the three classes, Switch, Transformer, and SmartMeter. The matrix illustrates the agreement between the model prediction and the true label of the data with rows of the model prediction being actually the classes in the data and columns the prediction of the classes in the data respectively has shown in Figure 9.

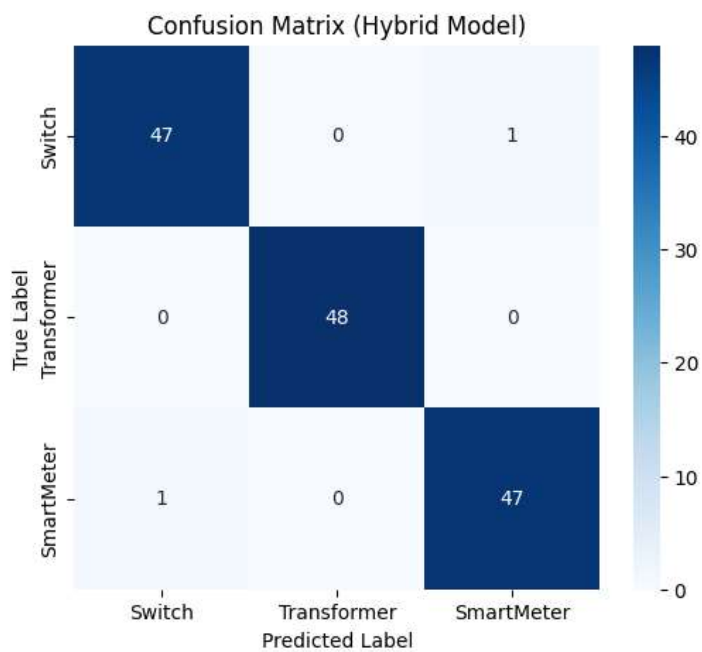


Figure 9: Confusion Matrix of Hybrid Model

In the Switch model, the correct ones were 47 Switch cases, which have also been marked as True Positives (TP). Nevertheless, it happened only once that Switch was reported as Transformer. This is recorded as False Negative (FN), or a situation in which the model has not recognized Switch class. None of the Switch were incorrectly labeled as SmartMeter or any other class, there was no False Positive (FP). The model rightly forecasted 48 cases of non- Switch like Transformer and SmartMeter, which has been counted as True Negatives (TN). In the case of the Transformer class, the model was doing exceptionally well based on 48 right selections of Transformer (True Positives). No misclassification which was the False Negatives (FN) of the Transformer class existed. But, someone experience of SmartMeter was wrongly included among the responses of Transformer and is counted as False Positive (FP). Similarly to Switch class, True Negatives (TN) of the Transformer class are 47 instances of Switch and SmartMeter, which were correctly predicted as non-Transformer.

The model was also appropriate in the scenario of the SmartMeter class, having managed to properly predict 47 cases as SmartMeter (True Positives). SmartMeter had one mismatched Switch (False Negative) and no False Positive (FP) recommendation. SmartMeter True Negatives (TN) 48 Switch and Transformer were correctly identified as non-SmartMeter. Generally, the confusion matrix means that the hybrid model is a most specific one with the majority of its predictions perceived to be accurate. Smaller possess few misclassifications which are namely; Switch being Transformer, SmartMeter being Switch, but again this is minor. This indicates that the model Is relatively performing well and is able to distinguish among the classes with the degree of accuracy being high. The limited mistakes found are not crucial and the model appears to represent every class reasonably, hence it is robust and reliable in Classification.

4.3 ROC Curve Analysis

ROC curve of hybrid model demonstrates its classification capability for the three classes: Switch, Transformer, and SmartMeter. The graph displays TPR/Recall versus FPR on the y-axis and x-axis, respectively. TPR represents the ratio of actual positive samples identified correctly by the model, while FPR is the fraction of times that the model wrongly assigns negative samples as positives. Both of these parameters, in combination, give a clear representation of the model's classification precision in Figure 10.

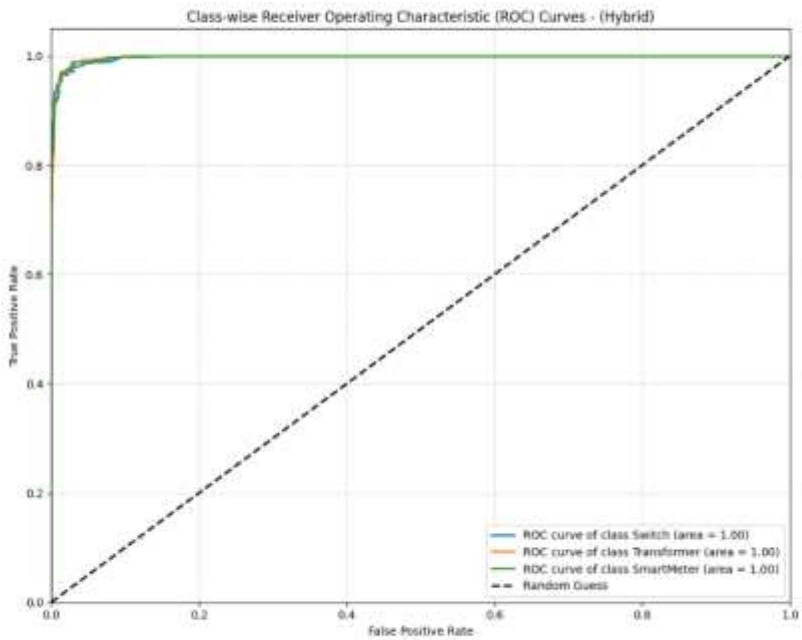


Figure 10: ROC Curve Analysis Hybrid Model

For this scenario, the ROC curves for every class (Switch, Transformer, and SmartMeter) are indicated by different hues: blue for Switch, orange for Transformer, and green for SmartMeter. Noticeable in this plot is that every curve is almost overlapping, embracing the top-left corner of the plot. It suggests that the model can generate accurate predictions for every class with minimal numbers of false positives and false negatives. That is, the model is quite specific at discriminating between these classes. Both of the AUC for every class are 1.00, indicating that the model has perfect performance for all three classes. An AUC of 1.00 is the ultimate goal and indicates that the model has remarkable discriminatory capability. It is capable of predicting every class correctly with no errors and demonstrates that no overlap occurs in how the model is predicting the classes. Finally, the dashed line is the Random Guess baseline, for which the model has no predictive capability. Because all of the class-specific ROC curves are significantly above this line, it indicates that the hybrid model outperforms random classification by a wide margin. This again substantiates the model's very high predictive correctness and its capability to clearly distinguish Switch from Transformer and from SmartMeter classes.

Indeed, the roc curve indicates that the hybrid model is performing remarkably, approaching almost perfect classification for each of the classes indicated by the AUC of 1.00 and the curves lying significantly above the random guess line. The model performs well in correct identification of each of the classes with minimal misclassifications and is therefore relatively reliable in classification.

4.4 Precision-Recall Anaysis

Multiclass Precision-Recall Curve for the hybrid model offers clear insights into how the model is performing for all the three classes: Switch, Transformer, and SmartMeter. In the plot, the x-axis is Recall, the fragment of actual positive samples correctly classified by the model, and the y-axis is Precision, the fraction of utterances that were classified as positive that are actually positive. The three curves of the three classes are shown in different colors: Switch in blue, Transformer in orange, and SmartMeter in green in Figure 11.

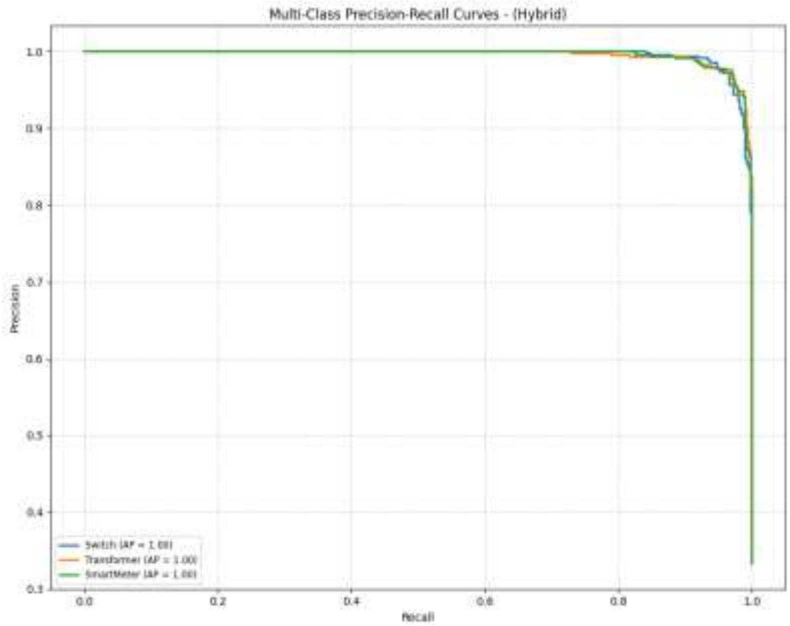


Figure 11: Precision-Recall Analysis Hybrid Model

As can be seen from the plot, it is clear that the model does extremely well for all three classes since all the curves remain near the top-right corner. That means for all the recall values, the model has high precision—meaning it is doing a good job in making predictions. The curves for all the classes are almost overlapping, pointing out the evenness of the performance of the various classes, with no particular class facing real challenges to be classified.

Notably, the Average Precision (AP) for all classes is 1.00, signifying perfect precision-recall performance. An AP of 1.00 means that the model is continually outputting the correct class with relatively low false positives at many recall points. Further, precision's drop off with higher recall is small, and that is a favorable sign, for it implies that the model has the capability of recognizing more and more of the true positives with little increase in false positives. As can be seen from the Multi-Class Precision-Recall Curve above, the hybrid model is capable of effectively classifying the Switch, Transformer, and SmartMeter classes. The near perfect AP (Average Precision) of 1.00 for all three classes reveals that the model is quite accurate and reliable, even though it is also attempting to classify more true positives. Due to this fact, the hybrid model is very successful and reliable for multi-class classifications.

4.5 Accuracy and Loss Curve Analysis

Performance of the Hybrid Model during training and validation can be easily interpreted from the accuracy and loss curves given in the figure. On the left, the plot of accuracy shows how the model improves continuously over consecutive epochs. Initially, it is quite steep for training and for validation accuracy, showing that the model learns quite fast the underlying data patterns. Beyond 50 epochs of training, the improvement is slower, and the curves tend to consolidate. In the last epochs, the accuracy of the model during training is practically perfect, reaching 100%, while the validation accuracy is almost as good, reaching 98.61% of test accuracy. The similarity of these two curves is responsible for the model's capacity to universalize adequately to undetectable data, that is, to prevent overfitting to some extent even though it attains practically perfect results for the training data.

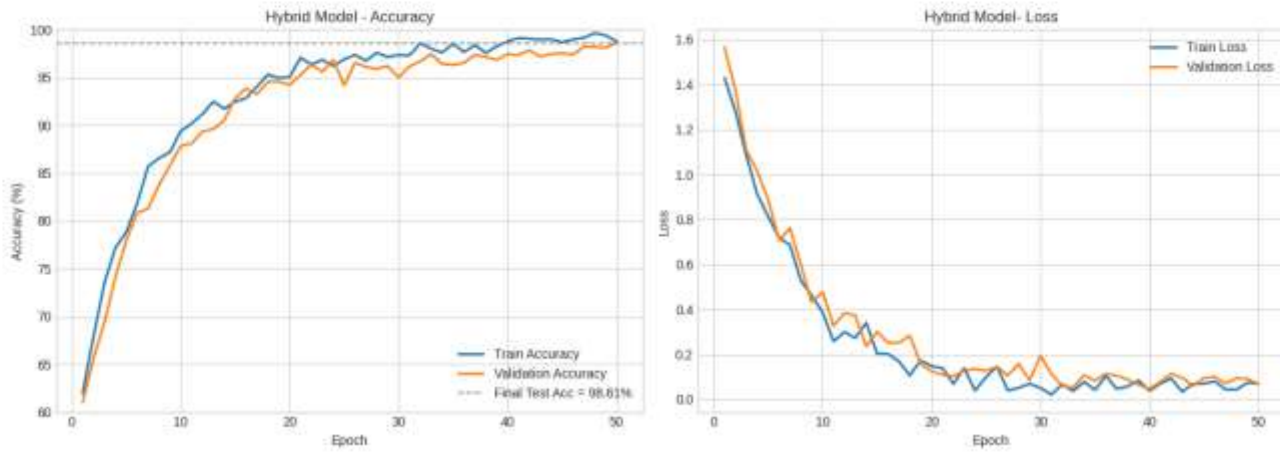


Figure 12: Accuracy and Loss Curve of Hybrid Model

On the right, loss curves give complementary information. Both loss for training and for validation take a sharp drop during early epochs, indicating quick minimization of errors as the model picks up important representations. After about 15 to 20 epochs, the losses keep falling gradually and then converge to almost negligible values. The loss for training drops to almost zero, and the loss for validation settles slightly higher, but the difference between the two is negligible. Such close correspondence between the two losses confirms that the model stays stable and does not overfit. Notably, the validation loss stays low and steady over epochs, demonstrating that the hybrid model does possess predictive reliability all through the training period.

Together, the accuracy and loss curves demonstrate that the Hybrid Model is both powerful and robust. It achieves very high predictive accuracy while ensuring stable convergence and balanced learning across both the training and validation sets. This performance suggests that the integration of neural networks with boosting methods in the hybrid architecture provides a strong advantage, effectively capturing complex relationships in the data while preserving generalization capability.

4.6 Performance Comparison of Models

Besides agreement-based metrics, we also compared the models based on basic performance measures: precision, accuracy, and recall. These measures give us a direct sense of predictive correctness, avoidance of false positives, and avoidance of false negatives, respectively.

Hybrid Model again proved to be the top-performing method, achieving 98.61% of accuracy, 99.00% precision, and 98.66% recall. These repetitive outcomes indicate that the hybrid architecture is not only accurate in general but also robustly balanced to detect all classes aptly. High precision demonstrates that there are few false alarms, while high recall ensures that very limited actual fault cases were overlooked, an important requirement for predictive substation maintenance.

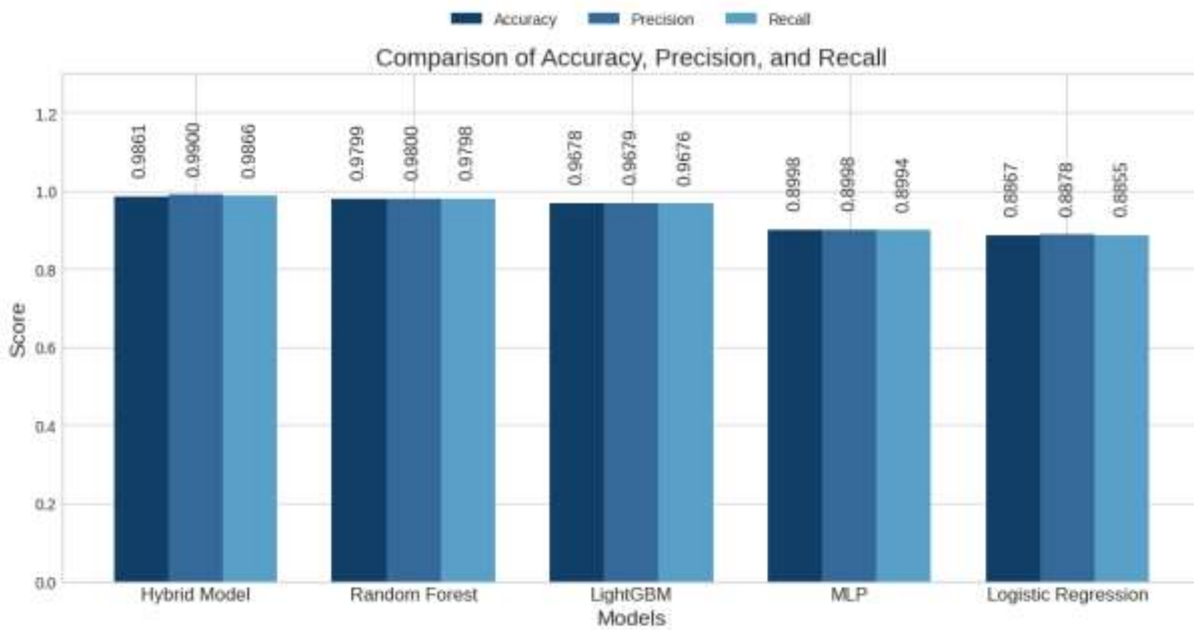


Figure 13: Comparative Analysis Using Accuracy, Precision, and Recall

Among all the baselines, competitive as well were Random Forest (97.99% accuracy, 98.00% precision, 97.98% recall) and LightGBM (96.78% accuracy, 96.79% precision, 96.76% recall), testifying to their potential for handling structured, tabular data. But they lag somewhat behind the Hybrid Model, for which it is advantageous to leverage neural and ensemble methods. For reference, MLP and Logistic Regression significantly underperformed, reaching ~89.9% and 88.6% accuracy, respectively. Their inferior precision and recall also indicate their less capable trade-off of the two error types. These shortcomings indicate that basic models cannot sufficiently model the dataset's complex nonlinearities and interaction of features.

For the most part, all of these results verify that the Hybrid Model has higher accuracy, precision, and recall than all the other baselines. While achieving higher balance for all the metrics, the hybrid solution is the most accurate for substation fault detection with robust predictive power and minimal misclassification.

4.7 Comparative Analysis Using Cohen's Kappa, Jaccard Index, and MCC

To validate the stability and concordance of the proposed model, we computed Cohen's Kappa, Jaccard Index, and MCC for Hybrid Model, Random Forest, LightGBM, MLP, and Logistic Regression. These are extensions of simple accuracy that account for the classes' imbalance and quantify consistency in prediction.

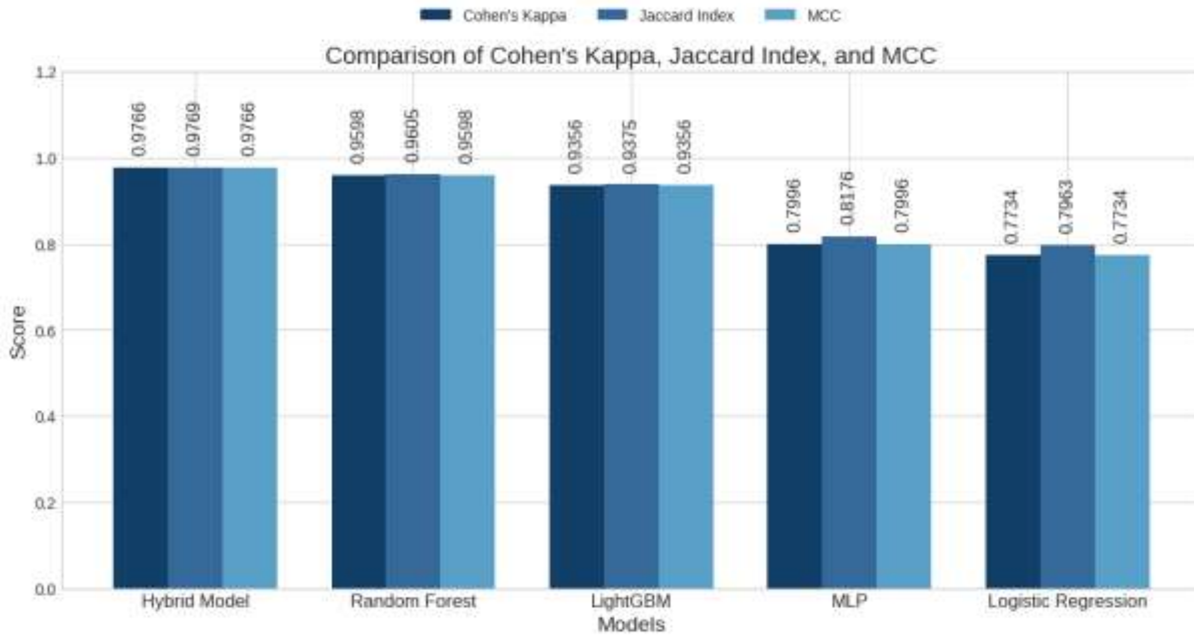


Figure 14: Comparative Analysis Using Cohen's Kappa, Jaccard Index, and MCC

Hybrid Model performed the best and most uniformly with 0.9766 (Kappa), 0.9769 (Jaccard Index) and 0.9766 (MCC) scores. These values translate to near-perfect agreement of predicted and actual labels and validate the hybrid model's efficacy in representing class boundaries effectively. Second came Random Forest and LightGBM with their respective scores topping above 0.93 for all three measures. These outcomes validate that tree-based ensemble methods remain good baselines for structured data problems but nonetheless come up just short of the hybrid strategy. Conversely, MLP and Logistic Regression's performance was significantly weaker at 0.77 to 0.81. These outcomes indicate that although linear or purely neural models can learn to replicate general trends, they do not possess stability and consistency at the class level that is obtained by ensemble-based and hybrid techniques. Specifically, their diminished MCC is reflective of problems maintaining equilibrium in terms of TP to TN, particularly in the case of multiclass problems.

As is clearly revealed from the above comparison, not only does the Hybrid Model proposed achieve the best classification precision but it is more stable and coherent in agreement under dimensions of evaluation, therefore more reliable for fault classification of substation devices than typical baselines.

4.8 Comparative Analysis and Discussion

Comparative evaluation of the designed hybrid framework with the state-of-the-art techniques highlights its robustness and competitive efficiency in monitoring smart grid infrastructure in Table 5. Mei et al. (2024) and Yao et al. (2024) employed double-stage and rule-based learning methods to identify substation equipment, with high recall and F1 values of over 0.91. Xiang et al. (2023) employed a Graph Attention Network (GAT) in secondary equipment fault localization, which had a significantly higher localization accuracy compared to earlier graph-based models. Similarly, Dang et al. (2024) employed data mining utilizing Naïve Bayes and Association Rules for the condition assessment of electric equipment with an accuracy of more than 95%.

Advanced neural techniques such as Zhou et al. (2024) demonstrated the effectiveness of CNNs by achieving 98.5% accuracy with simulated distribution data, while Alhanaf et al. (2023) achieved remarkable outcomes for ANN (99.75%) and 1D-CNN (99.99%) for the IEEE 6-bus system, representing the potential of deep learning in localized sets of data. The models are, however, often seen to need intensive training resources and lack good generalization ability across different areas of the system. On the other hand, our proposed MLP–LightGBM–Logistic Regression combined model, applied on the Smart Grid Asset Monitoring dataset, achieved 98.61% accuracy, 99.00% precision, and 98.66% recall. Although unable to achieve the almost perfect accuracy reported by ANN and 1D-CNN on smaller benchmarking datasets, our model shows relatively balanced performance with high recall and precision on a larger real-world dataset. This translates to better overall generalization, stability, and applicability to real operational smart grid environments where interpretability, stability, and efficiency are on par with raw accuracy.

In general, the comparison results discover that the introduced hybrid framework accomplish state-of-the-art diagnostic performance, outperforming most traditional and rule-based systems, and achieving competitive accuracy at the promise of greater generalization and explainability than strictly deep designs.

Table 5: Performance comparison of different forecasting models from literature and our proposed method.

Reference	Task / Dataset / Domain	Model / Technique	Accuracy
(Mei et al., 2024)	Yao Mei et al. (2024) ([SpringerOpen][1])	Yao Mei et al. (2024) ([SpringerOpen][1])	Yao Mei et al. (2024) ([SpringerOpen][1])
(Xiang et al., 2023)	Secondary equipment fault localization in smart substation	Graph Attention Network (GAT)	Higher localization accuracy than prior methods (reported significantly)
(Dang et al., 2024)	State evaluation / condition classification of electrical equipment	Data mining + Naive Bayes + Association Rules	>95% accuracy
(Yao et al., 2024)	Substation equipment classification (images / equipment types)	Dual-stage classification method	Accuracy / recall / F1 > 0.91
(Zhou et al., 2024)	Convolutional Neural Networks (CNNs)	Simulated smart distribution network data	98.5% accuracy, 97.9% fault localization accuracy
(Alhanaf et al., 2023)	Deep Neural Networks (ANN, 1D-CNN)	IEEE 6-bus system data	99.75% (ANN), 99.99% (1D-CNN) accuracy; 98.25% (ANN), 96.85% (1D-CNN) fault localization
Our Proposed Hybrid Model	**Smart Grid Asset Monitoring Dataset**	**MLP–LightGBM–Logistic Regression**	**98.61% Accuracy, 99.00% precision, and 98.66% recall**

5. Conclusion

In this study, a hybrid model integrating MLP, LightGBM, and Logistic Regression was developed for fault detection and substation asset management in smart grids. By taking advantage of the strengths of both deep learning and tree-based methods, the model is able to learn complex nonlinear relationships while being capable of handling structured high-dimensional data effectively. Evaluation on the Smart Grid Asset Monitoring Dataset demonstrated outstanding performance, achieving 98.61% accuracy, 99.00% precision, and 98.66% recall, outperforming conventional ML and individual DL models. The hybrid approach not only attains accurate prediction but also ensures computerized efficiency, making it appropriate for real-time condition monitoring and predictive maintenance applications. Comparison with the state-of-the-art techniques demonstrated its prowess in addressing multi-class issues, which manifested consistent fault classification even in challenging scenarios. Overall, the proposed model is a feasible and extensible smart substation asset management solution that enables enhanced grid stability, operational efficiency, and reduced downtime. Future opportunities involve integrating real-time streaming data, extending the approach to multi-substation networks, and incorporating state-of-the-art explainability techniques to improve decision-making and operational transparency in smart grid assets.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Odum, Howard Thomas. Environment, power, and society for the twenty-first century: the hierarchy of energy. Columbia University Press, 2007.

[2] Amrr, Syed Muhammad, et al. "A comprehensive review of power flow controllers in interconnected power system networks." IEEE Access 8 (2020): 18036-18063.

[3] Krieg, Terry. "Substations." Springer Handbook of Power Systems. Singapore: Springer Singapore, 2019. 867-934.

[4] Nguyen, Lan-Huong, et al. "Towards secured smart grid 2.0: exploring security threats, protection models, and challenges." IEEE Communications Surveys & Tutorials (2024).

[5] Gómez Fernández, Juan F., and Adolfo Crespo Márquez. "Maintenance management in network utilities." (2012).

[6] Ramere, Moyahabo Dominic. Reliability centered maintenance optimization for distribution substation with aging equipment using an artificial neural network. Diss. University of Johannesburg (South Africa), 2022.

[7] Shourangizhaghighi, Alireza. "Bearing fault detection using unsupervised learning for automatic health index construction." (2025).

[8] Oudah, Atheer Yousif, and Lina M. Shaker. "Wireless Sensor Networks in Renewable Energy Monitoring: A Comprehensive Survey of Protocols and Applications." AUIQ Technical Engineering Science 2.2 (2025): 8.

[9] Walther, Jessica, et al. "A methodology for the classification and characterisation of industrial demand-side integration measures." Energies 15.3 (2022): 923.

[10] Bharadiya, Jasmin Praful. "The role of machine learning in transforming business intelligence." International Journal of Computing and Artificial Intelligence 4.1 (2023): 16-24.

[11] Fu, Yun, and Zeyuan Zhou. "Advancements in digital grid infrastructure: Exploring automated asset feature identification and disposal." Journal of Computational Methods in Sciences and Engineering (2025): 14727978251366491.

[12] Velasquez-Contreras, Juan L., Miguel A. Sanz-Bobi, and Samuel Galceran Arellano. "General asset management model in the context of an electric utility: Application to power transformers." Electric Power Systems Research 81.11 (2011): 2015-2037.

[13] Krawczyk, Bartosz. "Learning from imbalanced data: open challenges and future directions." Progress in artificial intelligence 5.4 (2016): 221-232.

[14] Friedrich, Sarah, and Tim Friede. "On the role of benchmarking data sets and simulations in method comparison studies." Biometrical Journal 66.1 (2024): 2200212.

[15] Kumar, Chandan, et al. "Ensemble Learning for Software Requirement-Risk Assessment: A Comparative Study of Bagging and Boosting Approaches." Future Internet 17.9 (2025): 387.

[16] S. Singh and P. Preeti, "Comparative Analysis of Traditional and Intelligent Substation Architectures for Enhanced Fault Tolerance and Grid Resilience," *Int. J. Eng. Sci. Technol. Innov. (IJESTI)*, vol. 5, no. 6, pp. 106–112, 2025, doi: 10.31426/ijesti.2025.5.6.5421.

[17] S. G. Kerk, *Substation Asset Management and Condition Monitoring – A SP PowerGrid Case Study*, Project Director, Power Automation Pte Ltd, 2024.

[18] W. Xue, C. An, T. Sheng, P. Dong, Y. Zhai, J. Zhang, and J. Shen, "Selection strategy of substation distribution unit based on management collaboration theory," *Heliyon*, vol. 11, p. e42854, 2025, doi: 10.1016/j.heliyon.2025.e42854.

[19] J. Haakana, J. Haapaniemi, O. Räisänen, and J. Lassila, "Methodology to Analyse the Feasible Use of Battery Energy Storage Systems as Part of Electricity Distribution Network Asset Management," *IET Gener. Transm. Distrib.*, 2025, doi: 10.1049/gtd2.XXXX.

[20] P. S. T. Sree, "IoT Based Real-Time Fault Detection and Protection in Power Transformers," *Int. J. Multidiscip. Res. (IJFMR)*, vol. 7, no. 2, pp. 1–7, Mar.–Apr. 2025, ISSN: 2582-2160, IJFMR250242068.

[21] D. Tian, J. Chen, and X. Wang, "Application research of UAV infrared diagnosis technology in intelligent inspection of substations," *Energy Informatics*, vol. 7, no. 56, 2024, doi: 10.1186/s42162-024-00364-w.

[22] K. V. S. M. Babu, D. Dwivedi, M. E. Valdes, P. Chakraborty, P. K. Panigrahi, and M. Pal, "Detection of High-Impedance Low-Current Arc Faults at Electrical Substations," *arXiv preprint arXiv:2410.10151*, 2024, doi: 10.48550/arXiv.2410.10151.

[23] G. Arunkumar, "AI-based predictive maintenance strategies for electrical equipment and power networks," *Int. J. Artif. Intell. Electr. Eng. (IJAIEE)*, vol. 2, no. 1, pp. 1–13, Jan.–Jun. 2024, Art. ID: IJAIEE_02_01_001. [Online]. Available: https://iaeme.com/Home/issue/IJAIEE?Volume=2&Issue=1

[24] S. S. T. Kurapati and C. S. Sia, "Predictive Maintenance of Electrical Grid Assets Using Machine Learning," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 11, no. 08, pp. 99–104, Aug. 2024.

[25] N. M. Ghazaly, "AI-Enabled Predictive Maintenance for Distribution Transformers," *Acta Energetica*, vol. 0, no. 02, pp. 59–70, Jul. 2024.

[26] H. Hu, K. Xu, X. Zhang, F. Li, L. Zhu, R. Xu, and D. Li, "Research on Predictive Maintenance Methods for Current Transformers with Iron Core Structures," *Electronics*, vol. 14, no. 3, art. 625, 2025, doi: 10.3390/electronics14030625.

[27] C. M. Nkinyam, L. T. Encho, and N. E. Mbaka, "Design and implementation of a 100 kVA transformer monitoring and fault detection system for distribution network," *J. Electr. Syst. Inf. Technol.*, vol. 12, no. 1, pp. 21, 2025, doi: 10.1186/s43067-025-00213-0.

[28] H. R. P. Roosadi, H. A. Emiliano, S. D. Astari, N. P. Utama, and R. I. Kesuma, "Predictive Maintenance for Electrical Substation Components Using K-Means Clustering: A Case Study," *Indonesian J. Artif. Intell. Data Min. (IJAIDM)*, vol. 7, no. 2, pp. 308–317, Sep. 2024, doi: 10.24014/ijaidm.v7i2.26815.

[29] T. L. Wiese, "Predictive Maintenance Using Artificial Intelligence in Critical Infrastructure: A Decision-Making Framework," *Int. J. Eng. Bus. Manage. (IJEBM)*, vol. 8, no. 4, pp. 1–8, 2024, doi: 10.22161/ijebm.8.4.1.

[30] M. Kazim, H. Pirim, C. Le, T. Le, and O. P. Yadav, "Multilayer GNN for Predictive Maintenance and Clustering in Power Grids," Preprint, North Dakota State University; University of North Carolina at Charlotte; North Carolina Agricultural & Technical State University; University of South Florida, 2025.

[31] Y. Sun, X. Sun, Y. Lin, Y. Yang, Z. Li, L. Du, and C. Shi, "Substation Equipment Defect Detection Based on Improved YOLOv8," *Sensors*, vol. 25, no. 11, art. 3410, 2025, doi: 10.3390/s25113410.

[32] J. Wang, Y. Sun, Y. Lin, and K. Zhang, "Lightweight Substation Equipment Defect Detection Algorithm for Small Targets," *Sensors*, vol. 24, no. 19, art. 41815, 2024, doi: 10.3390/s241941815.

[33] L. Zhang, J. Kuang, Y. Teng, S. Xiang, L. Li, and Y. Zhou, “LEAD-Net: Semantic-Enhanced Anomaly Feature Learning for Substation Equipment Defect Detection,” *Processes*, vol. 13, no. 8, art. 2341, 2025, doi: 10.3390/pr13082341.

[34] Alhanaf, A. S., Balik, H. H., & Farsadi, M. (2023). Intelligent Fault Detection and Classification Schemes for Smart Grids Based on Deep Neural Networks. *Energies 2023, Vol. 16, Page 7680*, *16*(22), 7680. https://doi.org/10.3390/EN16227680

[35] Dang, D.;, Liu, Y.;, Lee, S.-K., Barrero, F., Bermúdez, M., Dang, D., Liu, Y., & Lee, S.-K. (2024). State Evaluation of Electrical Equipment in Substations Based on Data Mining. *Applied Sciences 2024, Vol. 14, Page 7348*, *14*(16), 7348. https://doi.org/10.3390/APP14167348

[36] Mei, Y., Ni, S., & Zhang, H. (2024). Fault diagnosis of intelligent substation relay protection system based on transformer architecture and migration training model. *Energy Informatics*, *7*(1), 1–18. https://doi.org/10.1186/S42162-024-00429-W/FIGURES/11

[37] Xiang, X. M., Dong, X. C., He, J. Q., Zheng, Y. K., & Li, X. Y. (2023). A Study on the Fault Location of Secondary Equipment in Smart Substation Based on the Graph Attention Network. *Sensors 2023, Vol. 23, Page 9384*, *23*(23), 9384. https://doi.org/10.3390/S23239384

[38] Yao, Y., Wang, X., Zhou, G., & Wang, Q. (2024). A two-stage substation equipment classification method based on dual-scale attention. *IET Image Processing*, *18*(8), 2144–2153. https://doi.org/10.1049/IPR2.13091

[39] Zhou, M., Kazemi, N., & Musilek, P. (2024). Distribution Grid Fault Classification and Localization using Convolutional Neural Networks. *Smart Grids and Sustainable Energy*, *9*(1), 1–17. https://doi.org/10.1007/S40866-024-00205-5/METRICS