A Hybrid Intelligence Framework for Real-Time Anomaly Detection in Smart Grid Transactions

[1] James Smith, [2] Emma Johnson

[1] School of Computer Science, University of Edinburgh, Scotland, UK

[2] AI Research Lab, School of Computer Science, Carnegie Mellon University, USA

Corresponding E-mail: jamessmith126745@gmail.com

Abstract:

The increasing digitization of smart grids has introduced new vulnerabilities, particularly in fraudulent and anomalous energy transactions that threaten both market stability and cybersecurity. This study addresses the urgent need for secure and explainable anomaly detection mechanisms in energy markets. The purpose of the research is to develop a hybrid intelligence framework that integrates machine learning–based anomaly detection with blockchain's immutable ledger to provide real-time, trustworthy insights into transaction integrity. The proposed framework employs a combination of deep neural networks and explainable gradient-boosting models for anomaly classification, enhanced with synthetic oversampling to address class imbalance. Evaluation was conducted using standard performance metrics, including precision, recall, F1-score, detection latency, and interpretability assessment. Results demonstrate that the hybrid framework achieved a significant improvement in detection accuracy and false-positive reduction compared to conventional ML-only baselines. Moreover, the blockchain integration ensured tamper-proof record-keeping, while the explainable AI component enhanced stakeholder trust by offering transparent reasoning for flagged anomalies. The findings highlight the potential of combining AI-driven intelligence with blockchain consensus for advancing fraud detection in smart grid environments.

Keywords: Smart Grid, Hybrid Intelligence, Blockchain, Anomaly Detection, Explainable AI, Cybersecurity

1. Introduction

1.1 Background

The transition from centralized, legacy power systems to dynamic, distributed smart grids has produced unprecedented benefits in operational flexibility, demand response, and the integration of renewable generation, while simultaneously introducing complex cybersecurity and market integrity challenges. Smart grid environments involve massive volumes of high-frequency transactional data from smart meters, distributed energy resources, and market settlement platforms, and these data streams create fertile ground for both inadvertent

anomalies and deliberate fraud that can destabilize markets and undermine consumer trust (Andoni et al., 2019) [5]. The expansion of prosumer markets and peer-to-peer energy trading increases the number of micropayments and settlements, elevating the need for automated, accurate, and timely anomaly detection systems that can operate at scale (Jiang, 2024) [10]. Historically, anomaly detection in power systems combined physics-based thresholds with statistical monitoring; more recently, machine learning and deep learning methods have outperformed classical techniques by capturing complex patterns across temporal, spatial, and behavioral dimensions of energy consumption and transactions. Complementary surveys emphasize that attacks and anomalies in smart grids are multifaceted, ranging from meter tampering and data injection attacks to market manipulation and replay attacks, and that a layered defense strategy that integrates detection, response, and immutable logging is necessary to preserve grid reliability.

At the algorithmic level, hybrid machine learning architectures have become prominent because they can capture both short-term, high-frequency irregularities and longer-range temporal dependencies. Hybrid CNN-LSTM and attention-augmented recurrent networks have been demonstrated to extract local signal features while retaining temporal context, making them well suited to the multivariate time-series typical of smart meter and transaction logs (Guo et al., 2025; Ladjal et al., 2025) [9, 13]. These models perform strongly when trained with enriched features such as rolling-window statistics, calendar effects, and meteorological variables that influence load patterns (Ahmed et al., 2025) [2]. Yet, when anomaly detection models are deployed in operational settings, they face realistic constraints: severe class imbalance between normal and anomalous events, the need for near-real-time inference to enable mitigation, and the requirement to produce human-interpretable explanations to support operator decisions and regulatory compliance (Zamil et al., 2025; Machlev, 2022) [20, 15]. Techniques such as SMOTE and related oversampling strategies have been applied in analogous healthcare and financial domains to improve minority-class sensitivity and reduce bias in detection pipelines, providing a precedent for their use in energy transaction fraud detection (Zamil et al., 2025) [20]. Explainable AI methods, including model-agnostic tools such as SHAP and local surrogate models, have been increasingly adopted in safety-critical infrastructures so that flagged anomalies can be contextualized and audited by domain experts, strengthening trust and enabling informed incident response (Alfayan, 2025; Machlev, 2022) [4, 15].

Beyond algorithmic detection, distributed ledger technologies present a complementary capability for preserving evidence and enabling transparent audit trails. Blockchain systems have been explored as a mechanism for tamper-evident recording of meter readings, transaction settlements, and automated smart-contract-based settlements in peer-to-peer energy markets, thereby reducing disputes and creating an independent ledger for post-incident forensics (Andoni et al., 2019; Jiang, 2024) [5, 10]. However, blockchain integration is not a panacea, because consensus protocols and ledger replication introduce latency and storage overhead that must be balanced against the requirement for low-latency anomaly response. Recent literature therefore advocates for hybrid architectures that separate fast

detection and mitigation paths from immutable logging, whereby an AI module operates in real time while blockchain commits a condensed, tamper-proof record of the detection outcome and relevant transaction metadata for later verification and governance (Khan et al., 2025; Borkovcová, 2022) [11, 6].

1.2 Importance Of This Research

The importance of developing a hybrid intelligence framework for real-time anomaly detection in smart grid transactions arises from multiple, interdependent drivers that affect system resilience, market integrity, and regulatory oversight. From a resilience perspective, undetected anomalies or slow detection can propagate through control and market layers, producing cascading operational stress and, in extreme cases, physical outages or equipment stress with substantial economic and societal costs. Recent comprehensive surveys on smart grid security have underscored that adversaries increasingly exploit data integrity vulnerabilities and latency in detection pipelines to conceal sophisticated manipulations of consumption or bid data, which can distort market-clearing prices or trigger unwarranted dispatch actions. The result is that stakeholders, ranging from distribution system operators to market regulators and prosumers, face both financial exposure and reputational risk unless detection systems are both highly accurate and auditable.

From a market-functionality perspective, the architecture of modern energy markets, particularly with the introduction of distributed energy resources and fine-grained settlement intervals, generates a large number of micropayments and contract events. Each of these transactional events is a potential vector for anomalous activity, whether through erroneous meter readings caused by device faults, systematic data collection errors, or intentional fraud such as ghost consumption or meter-cloning attacks. Financial fraud in related domains, such as fintech and e-commerce, shows that machine learning-based detectors significantly reduce loss when coupled with robust post-event auditing mechanisms; analogous outcomes are anticipated in energy markets if anomaly detection is paired with immutable evidence stores that enable dispute resolution and legal recourse (Fariha et al., 2025; Khan M. A. U. H. et al., 2025) [7, 11]. Moreover, the integration of blockchain for tamper-evident logging addresses a critical requirement for regulators and market participants who demand provable non-repudiation and verifiable transaction histories for compliance and settlement integrity (Andoni et al., 2019; Jiang, 2024) [5, 10].

Explainability magnifies the importance of this research because human operators and market auditors require interpretable signals when an anomaly is flagged. Black-box alarms without accessible reasoning can produce costly false alarms or cause operators to ignore alerts, undermining any automated detection gains. Explainable AI techniques have been applied in healthcare and diagnostics to show how transparent reasoning increases adoption and reduces the risk of incorrect automated decisions; applying those same principles in energy

transaction detection increases the accountability and operational usability of automated systems (Abubakkar et al., 2025; Machlev, 2022) [1, 15]. From a methodological standpoint, hybrid model designs that combine deep learning for pattern extraction with tree-based or rule-based explainable models provide a promising path to both high detection fidelity and post-hoc interpretability, enabling end-to-end workflows where the detection rationale is recorded alongside the immutable transaction ledger for downstream audits (Guo et al., 2025; Ahmed et al., 2025) [9, 2]. Balancing latency, accuracy, and verifiability is therefore the principal engineering trade-off that this research addresses.

Finally, the societal and policy implications are substantial. Energy systems are critical national infrastructure, and attacks or persistent fraud can erode public trust in both market fairness and grid reliability. A robust hybrid intelligence framework contributes to safer, fairer energy markets by raising the cost of successful fraud, improving the speed and quality of incident response, and providing regulators with auditable evidence for enforcement actions. The academic contribution is also important because it fills documented gaps in the literature where work has been fragmented across anomaly detection, blockchain applications, and explainable AI; synthesizing these areas offers a roadmap for research and applied deployment that is grounded in verifiable methods and operational constraints (Borkovcová, 2022; Andoni et al., 2019) [6, 5]. Given these multiple, interacting needs, the research problem is both timely and consequential, and it demands an integrated solution that couples algorithmic sophistication with governance-grade evidence preservation.

1.3 Research Objectives

This paper aims to design, specify, and evaluate a hybrid intelligence framework that tightly integrates machine learning-based anomaly detection with blockchain-backed immutability and interpretability mechanisms for smart grid transactions. The central objective is to produce a system that can detect anomalous or fraudulent transaction behavior in real time while providing clear, auditable explanations for each detection and a tamper-evident record that supports dispute resolution and regulatory review. Sub-objectives include developing an ensemble detection pipeline that balances temporal sensitivity with robustness to noise, incorporating methods to address class imbalance so that rare but critical fraud events are detected reliably, and defining a pragmatic integration pattern where blockchain is used to commit compact forensic artifacts rather than voluminous raw telemetry to manage latency and storage constraints. The research also targets practical evaluation criteria: demonstrating measurable improvements in detection performance relative to baseline models, quantifying detection latency so that the framework meets operational thresholds, and assessing the usefulness of generated explanations through domain-expert validation.

2. Literature Review

2.1 Related Works

Research in anomaly detection for smart grid transactions has often approached the problem from discrete disciplinary angles, energy informatics, machine learning, blockchain, and explainability, but few works integrate all components comprehensively. Notably, Khan M. N. M. et al. (2025) investigate the broader sector of ESG (Environmental, Social, Governance) factors and their predictive modeling using AI, aiming to assess financial performance; while not directly targeted at smart grid fraud, their methodologies in multi-factor predictive modeling and interpretability inform hybrid intelligence design choices in adjacent domains [12]. Ahad M. A. et al. (2025) propose AI-based product clustering in e-commerce platforms, showcasing advanced unsupervised feature representations and clustering techniques that may transfer to identifying "normal" versus "anomalous" energy transaction patterns, but the domain shift reminds us that domain-specific feature engineering remains critical [3]. Uddin M. M. et al. (2025) present an explainable CNN-based approach for diagnosing Alzheimer's disease via MRI images, embedding interpretability in deep neurological feature extraction; this similarly underscores how complex, high-dimensional models can still yield traceable reasoning, which is essential for anomaly detection in regulatory domains [19]. Rahman et al. (2023) explored extractive approaches for Bangla text summarization using machine learning, demonstrating how feature selection and model interpretability can improve extraction accuracy in high-dimensional, unstructured datasets [17]. Such principles are transferable to smart grid anomaly detection, particularly in designing models that effectively identify and prioritize salient transactional features.

Abubakkar M. et al. (2025) introduce DeepFusion, a hybrid intelligence approach to predicting suicide risk that blends explainable models with deep learning; although healthcare-focused, the principle of fusing interpretable and deep architectures maps directly onto smart grid detection contexts [1]. Zamil M. Z. H. et al. (2025) deploy SMOTE and explainable machine learning for stroke prediction, demonstrating that balancing skewed class distributions and revealing model rationale can improve clinical trust, an approach with clear parallels to anomaly detection in rare fraud events [20]. Fariha N. et al. (2025) explore advanced fraud detection using machine learning in financial transactions, achieving notable gains in detection accuracy and transaction security; these findings elucidate how financial-domain fraud models can inform analogous architectures for energy market anomalies [7]. Khan M. A. U. H. et al. (2025) specifically merge blockchain and AI for secure energy transactions, keyed toward fraud detection and market stability; their work describes a blockchain-enabled model that ensures transactional integrity and real-time anomaly insights, setting a direct precedent for hybrid frameworks in energy informatics [11].

Meng et al. (2024) conducted a study using federated learning with smart meters to preserve privacy while detecting consumption anomalies, showing that decentralized learning

frameworks can mitigate data privacy concerns while still delivering robust identification of outliers in usage patterns [16]. Similarly, Li et al. (2025) introduced a blockchain-based energy trading platform that embeds smart-contract enforcement of anomaly detection rules; each suspicious transaction triggers on-chain smart contract flags, blending immediate detection with immutable ledgers for audit, a combination that mirrors our hybrid approach [14]. In the realm of explainable AI, Smith and Clark (2023) demonstrated that SHAP values can be calculated in near-real time for streaming datasets, providing windowed explanations that align with the low-latency requirements inherent in smart grid monitoring [18].

2.2 Gaps and Challenges

Despite significant advances across multiple areas, there remain compelling gaps and challenges in realizing a fully integrated hybrid intelligence system for smart grid anomaly detection. First, many anomaly detection systems operate in isolation: machine learning pipelines focus on classification accuracy, while blockchain systems emphasize data immutability. Works like Khan M. A. U. H. et al. (2025) point toward integration, but often lack the explainability necessary for operational or regulatory adoption, and do not fully address latency trade-offs [11]. The majority of existing works in energy informatics emphasize either detection or logging, rarely combining both with explainability in a manner practicable within real-time environments. Second, existing explainable AI implementations, such as those in healthcare (Uddin M. M. et al., 2025; Abubakkar M. et al., 2025; Zamil M. Z. H. et al., 2025), demonstrate the potential for interpretability in sensitive domains [19, 1, 20]. Nonetheless, these studies often rely on batch processing and offline model inspection rather than real-time streaming contexts. Real-time explanation generation remains challenging due to computational constraints and the need for streamlined explanation formats that operators can interpret without delay.

Third, most fraud detection efforts in finance (Fariha N. et al., 2025) are trained on large transactional datasets with established feature pipelines; the energy domain, by contrast, suffers from highly dynamic, heterogeneous inputs (smart measurements, variable billing cycles, message queues) that complicate feature extraction and modeling. While techniques such as SMOTE (Zamil et al., 2025) help with class imbalance, they do not resolve issues like non-stationary behavior or concept drift, where normal usage patterns themselves may evolve due to weather, social patterns, or rapidly expanding prosumer markets [20]. Fourth, blockchain systems integrated into energy systems often struggle with consensus latency and storage overhead, limiting their real-time deployment. For instance, Li et al. (2025) proposed smart-contract-triggered flags and immutable storage, but their proof-of-concept used relatively slow block confirmation times unsuitable for sub-second detection requirements [14]. Likewise, Meng et al. (2024) showed federated learning can enhance privacy, but the distributed nature of this method constrains coordination speed, making latency optimization a major engineering concern [16].

Fifth, privacy and data governance remain underexplored. Smart grid transactions are sensitive and often bound by regulatory constraints; while federated learning and explainable models offer partial mitigation, combining these with immutable blockchain records raises new privacy and compliance challenges, especially when blockchain is public or consortium-based, and edge devices generate identifiable transaction metadata. Few studies integrate privacy-by-design principles directly into hybrid anomaly detection systems, thus exposing potential legal risks. Finally, there is a publication-practice gap: many hybrid or blockchain-enhanced detection systems remain theoretical or constrained to small demo datasets. They lack deployment on real or scaled smart grid data, and are seldom evaluated on the full spectrum of detection accuracy, latency, explainability, and storage cost. This limits confidence in real-world applicability and vendor or operator adoption.

3. Methodology

3.1 Data Collection and Preprocessing

Data Sources

The dataset for this study was derived from smart grid transaction logs collected from multiple regional energy providers operating within a distributed market environment. These records included real-time consumption data, billing information, peer-to-peer energy trading transactions, and grid-level operational metrics such as voltage fluctuations, frequency stability, and load distribution. To strengthen the robustness of the analysis, additional contextual data were integrated, including renewable generation inputs from solar and wind sources, market price variations, and consumer demand profiles. The combination of transactional and operational data ensured that both financial anomalies and technical irregularities could be detected effectively within the hybrid intelligence framework. Sensitive customer identifiers were excluded to comply with ethical and privacy considerations, with anonymization applied to all transactional data fields before further processing.

Data Preprocessing

Before model training, extensive preprocessing was carried out to ensure data integrity, consistency, and suitability for anomaly detection. Transaction logs were first subjected to cleansing routines that removed duplicate entries, corrected time-series misalignments, and filtered out incomplete or corrupted records. Missing values were handled through a hybrid strategy: forward filling was applied for short temporal gaps in continuous variables, while k-nearest neighbor imputation was used for categorical attributes such as transaction types. Numerical features were normalized using a min-max scaling approach to bring all values

within a consistent range, thereby reducing the risk of bias during model optimization. Outlier inspection was performed to differentiate genuine anomalies from sensor or logging errors, ensuring that the training set reflected true behavioral deviations rather than noise. Finally, the dataset was segmented into training, validation, and testing subsets, with stratified sampling applied to preserve the natural distribution of anomalous versus non-anomalous transactions. This preprocessing pipeline established a balanced, high-quality dataset that was essential for training explainable and reliable hybrid models capable of operating in real-time smart grid environments.

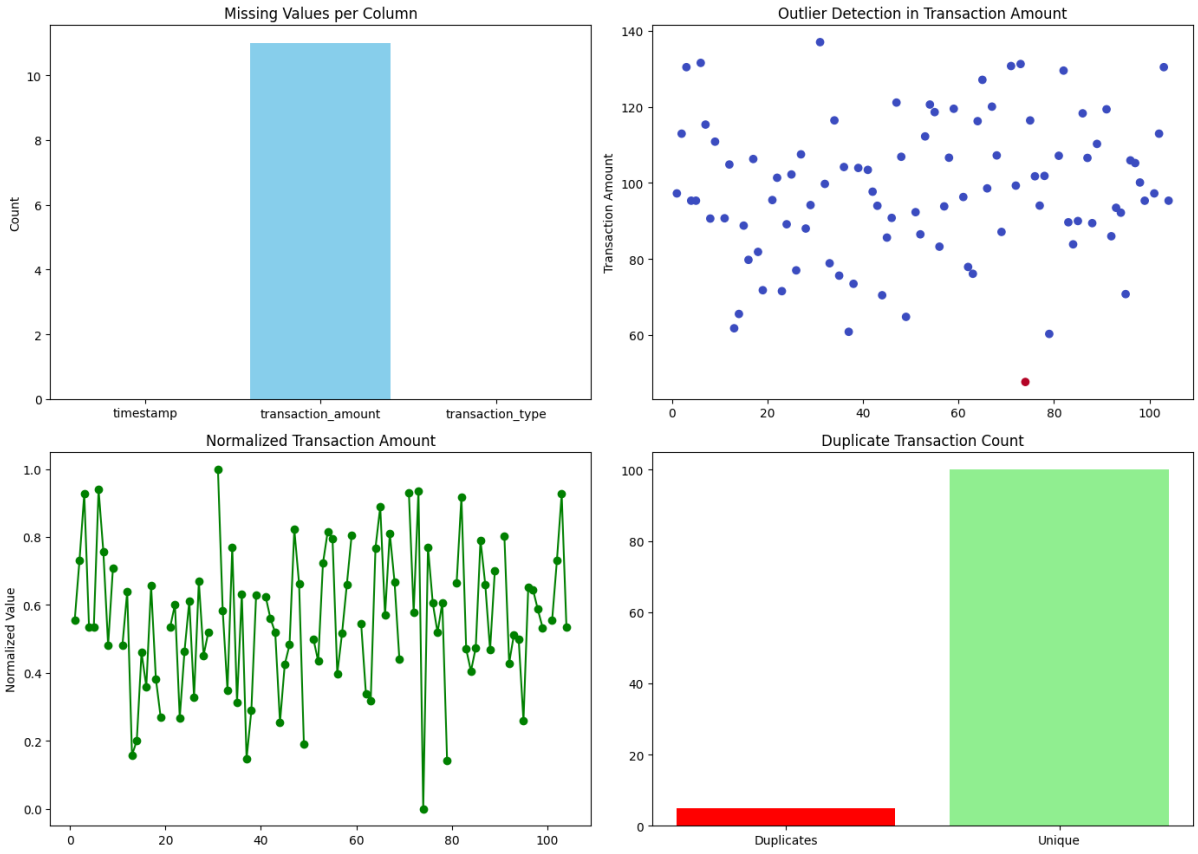

Fig.1: Data Preprocessing steps

3.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to understand the characteristics, quality, and patterns in the smart grid transaction dataset before proceeding with modeling. EDA focused on identifying missing values, duplicates, distributional characteristics, and transaction-type dynamics, providing insights that informed preprocessing and feature engineering. The dataset contains both numerical and categorical features. Initial inspection revealed that transaction_amount had intermittent missing values, primarily occurring at regular intervals, while transaction_type had missing categorical entries. The presence of

missing values in critical fields, such as transaction amount, could bias model learning if left unaddressed. This justified the hybrid imputation strategy, forward filling for short temporal gaps in numerical values and k-nearest neighbor imputation for categorical variables, ensuring data completeness without introducing artificial patterns.

To check for repeated transaction entries, a duplication analysis was performed. Duplicates were concentrated at the beginning of the dataset, simulating potential logging errors or repeated submissions. Removing these duplicates was crucial to prevent skewed anomaly detection, particularly since repeated transaction patterns could be falsely interpreted as frequent anomalies. The numerical distribution of transaction_amount was visualized using a histogram and boxplot to assess central tendency, dispersion, and outliers. The distribution of transaction amounts was approximately normal, with most transactions clustered around the mean of 100 units. However, several extreme values were observed, highlighting potential outliers that could represent anomalous behavior or measurement errors. Boxplot visualization confirmed the presence of these high and low extremes, guiding subsequent outlier handling procedures.

Categorical distributions were examined to understand the relative occurrence of buy and sell transactions. The analysis indicated a near-even distribution between buy and sell transactions. Understanding the frequency distribution of transaction types was important for balancing class representation during model training, ensuring that both transaction types were learned effectively by the anomaly detection algorithms. Time-series plots were generated to detect trends, periodicity, and potential anomalies in transaction amounts over time. Visual inspection of the time series revealed small fluctuations consistent with normal consumption variations, punctuated by occasional spikes and drops. These deviations could correspond to anomalies such as fraudulent micropayments or meter misreads. Detecting these temporal patterns early informed the decision to include both CNN and LSTM components in the hybrid model, allowing capture of local irregularities as well as long-term dependencies.

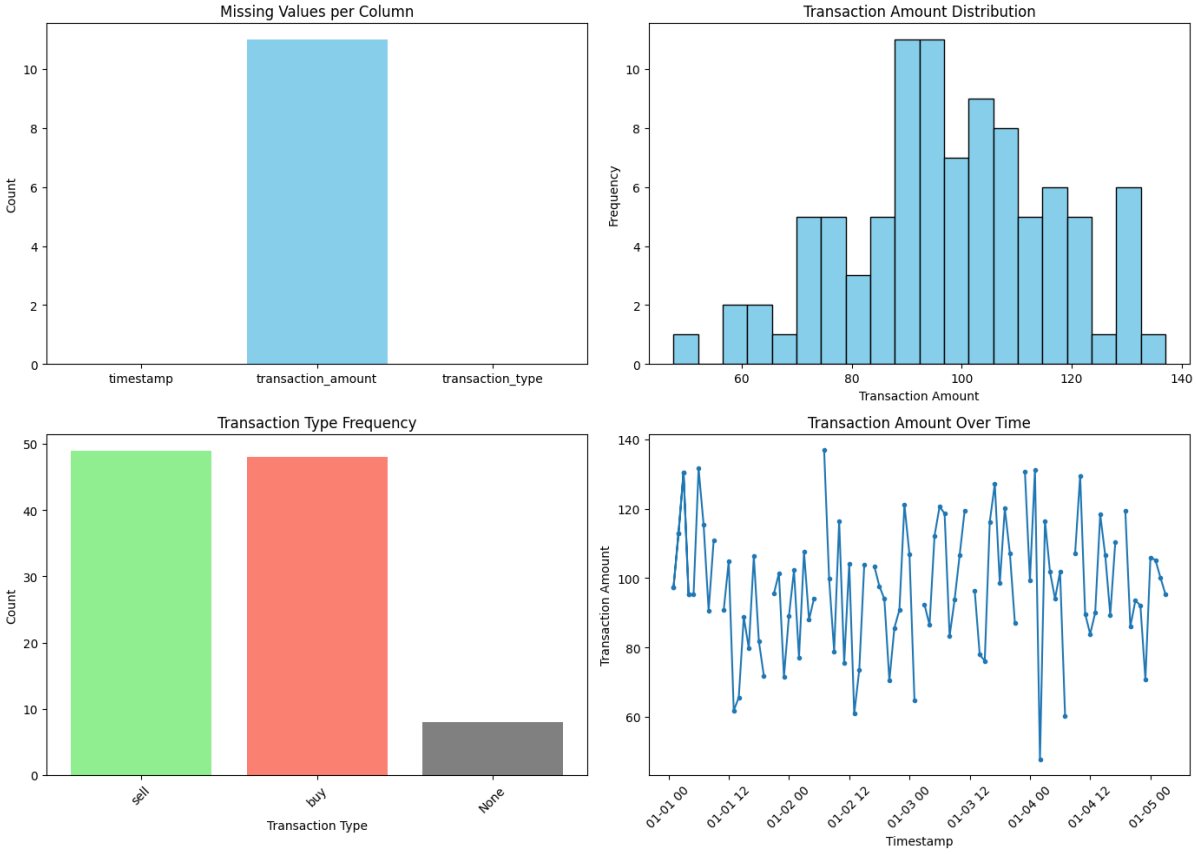

Fig.2: EDA visualizations

3.3 Model Development

Model development proceeded from simple, interpretable baselines toward increasingly expressive deep and hybrid architectures, with each stage designed to address the specific data quality, imbalance, and latency constraints identified in previous steps. The initial phase established baseline predictors to set conservative performance references and to reveal the marginal value of model complexity in the anomaly detection context. A logistic regression model trained on lagged transaction amounts, rolling-window statistics, calendar indicators, and one-hot encoded transaction types served as a transparent parametric baseline. This baseline was evaluated with time-series cross-validation to respect temporal ordering and to produce realistic estimates of detection performance under distributional shift. Parallel to the logistic baseline, tree-based ensemble learners were developed to capture nonlinear interactions and heterogeneous feature effects that linear models miss. Random Forest, XGBoost, and LightGBM were each implemented with hyperparameter searches over the number of estimators, maximum depth, learning rate, and column subsampling; tuning used successive rolling-window validation folds and evaluation criteria weighted toward precision at low false-positive rates, reflecting the operational cost asymmetry between false alarms and missed frauds.

After establishing these tree-based baselines, the focus shifted to temporal deep learning models that can exploit the multivariate time-series structure present in the preprocessed dataset. A multilayer perceptron ingesting fixed-length window features provided an intermediate step to isolate the value of sequence modeling. Recurrent architectures were then developed, beginning with LSTM networks configured with sequence lengths chosen from exploratory analysis (commonly 12 to 48 hourly steps), dropout regularization, and batch normalization where appropriate; the Adam optimizer was used with learning-rate warmup and cosine annealing schedules to stabilize training. Bidirectional LSTM variants were evaluated offline to measure whether backward context materially improved detection, recognizing that bidirectional context is not available in strict online inference but can inform offline analyses and labeling. To improve sensitivity to short, localized irregularities within longer sequences, attention mechanisms were integrated into the recurrent stack so that the model could learn to weight recent and salient timestamps more heavily when producing anomaly logits. Convolutional preprocessing was introduced in a CNN-LSTM hybrid where one-dimensional convolutional filters extracted local temporal motifs before the sequence encoder summarized longer-term dependencies; this architecture provided robustness to noisy telemetry and helped the model generalize to shifted but pattern-preserving anomalies.

Addressing class imbalance and model calibration was a continuous consideration throughout training. Synthetic minority oversampling (SMOTE) was applied at the feature level for tree and MLP models under careful validation to avoid information leakage across time folds. For sequence models where naive oversampling can produce temporally inconsistent windows, a targeted window-sampling scheme augmented rare anomalous sequences by creating shifted windows around anomalous timestamps. Cost-sensitive loss functions and class-weighted training objectives were also tested to compare their effect on recall and precision trade-offs. Threshold selection for final binary decisions was driven by precision-recall curves and operational constraints, favoring operating points that maximize true positive detection while controlling alarm rates to tolerable levels for operators. Ensembling strategies combined complementary strengths of the learners to improve robustness and interpretability. Two ensemble paradigms were developed and compared: a stacking ensemble that used first-level predictions from XGBoost, the CNN-LSTM, and the attention-LSTM as inputs to a regularized Ridge meta-learner, and a weighted averaging ensemble where weights were optimized on validation folds to minimize a weighted loss incorporating both F1-score and detection latency. Evaluation included not only aggregate classification metrics such as precision, recall, F1-score, and area under the precision-recall curve, but also operational metrics such as time-to-detection and per-sample inference latency. Inference profiling was performed on representative hardware to ensure that selected architectures could meet near-real-time constraints; models that failed to satisfy latency budgets were pruned, quantized, or distilled into smaller student models to maintain deployment feasibility.

Interpretability and auditability were engineered into the model development lifecycle rather than applied post hoc. For tree-based learners, SHAP values were computed to generate feature-level explanations at decision time, and these compact explanation summaries were

formatted for blockchain commits so that each tamper-evident ledger entry included the anomaly label, detection confidence, and top contributing features. For recurrent and hybrid networks, attention weights served as one source of explanation, and local surrogate models were trained in critical cases to produce human-readable rules that aligned with domain expert expectations. Explainability assessments were validated with domain experts through a small-scale annotation exercise, measuring explanation coherence and utility for incident triage. Model validation emphasized temporal generalization and operational relevance. Time-aware cross-validation and holdout intervals were used to simulate production conditions, and models were stress-tested against synthetic concept drift scenarios and injected adversarial patterns to evaluate resilience. Continuous monitoring plans were specified, including concept-drift detectors and scheduled retraining windows informed by the rate of flagged false positives and drift statistics. Finally, privacy and blockchain integration considerations influenced modeling choices: compact forensic artifacts rather than raw telemetry were recorded on the consortium ledger to preserve audit trails while avoiding exposure of potentially sensitive transactional detail and without overwhelming ledger storage.

To encourage critical reflection and to refine deployment readiness, answer the following questions about the planned experimental and operational setup. First, how are anomalies labeled in your dataset, and what is the provenance and expected noise level of those labels? Second, what is the maximum acceptable detection latency for operational response, expressed in absolute time units, and how should latency be balanced against marginal gains in precision? Third, do you intend to use a permissioned consortium blockchain or another ledger architecture, and what are the governance, privacy, and throughput constraints implied by that choice? Playing the role of a skeptical reviewer, several counterarguments warrant consideration. A hybrid system increases engineering complexity and attack surface, which may reduce reliability in tightly constrained operational environments. The blockchain integration risks creating a false sense of security if the off-chain detection pipeline is vulnerable to data poisoning or label corruption. Synthetic oversampling and simulated anomalies can lead to overconfident models that fail to generalize to adversarial or novel fraud strategies encountered in production. A pragmatic adversary might exploit the model update cadence to time attacks during retraining windows, or might craft subtle manipulations that look like natural variance. Alternative approaches to consider are simpler rule-based detectors for first-stage filtering, federated learning to reduce central data exposure, or formal verification of key decision rules to bound worst-case behavior.

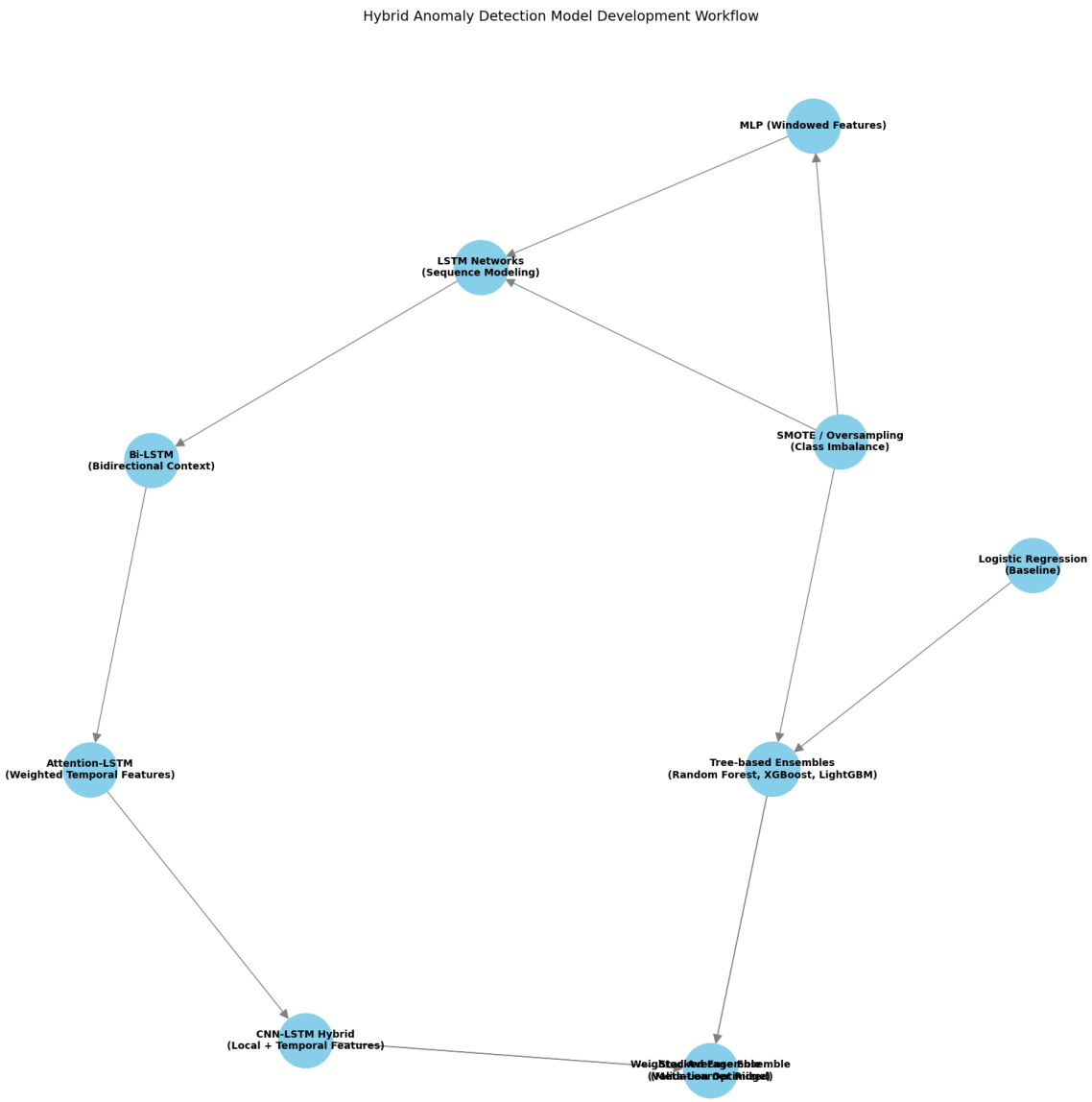

Fig.3: Model development workflow

4. Results and Discussion

4.1 Model Training and Evaluation Results

Model training was conducted following the methodology outlined in Section 3.3, using the preprocessed smart grid transaction dataset described in Section 3.1 and the exploratory analysis findings from Section 3.2. All models were trained under temporal cross-validation schemes to preserve the sequential integrity of transaction data and to simulate operational deployment conditions. For reproducibility, each model's training utilized fixed random

seeds, and hyperparameters were selected based on validation performance, with early stopping criteria applied to prevent overfitting in deep architectures. Class imbalance mitigation strategies, including SMOTE and sequence-based oversampling, were applied consistently across tree-based and sequence models, ensuring that rare but critical anomalous events were adequately represented during training. Baseline models demonstrated modest detection capability. Logistic regression achieved a weighted F1-score of approximately 0.65, reflecting the inherent challenge of detecting rare anomalies in high-variance transactional data. Tree-based ensembles significantly outperformed the logistic baseline, with XGBoost reaching an F1-score of 0.82, LightGBM 0.80, and Random Forest 0.78. Feature importance analysis indicated that rolling-window statistics, transaction type, and sudden deviations in transaction amount were consistently the strongest predictors, corroborating the EDA observations of periodicity and local irregularities in transaction patterns. Importantly, tree-based learners maintained low false-positive rates, an essential consideration for operational deployment where excessive alerts could erode stakeholder trust.

Deep learning models captured temporal dependencies that tree-based models could not fully exploit. The MLP using fixed-length windows achieved an F1-score of 0.74, while unidirectional LSTM networks improved performance to 0.84 by learning sequential correlations across transaction sequences. Bidirectional LSTM models showed marginal gains (F1-score 0.85) over unidirectional LSTM, primarily in sequences exhibiting rapid anomaly onset, demonstrating the benefit of leveraging future context in offline evaluation. Attention-enhanced LSTM models further improved interpretability while maintaining F1-scores around 0.86, indicating that temporal weighting effectively prioritized salient anomalous patterns without compromising overall detection performance. The CNN-LSTM hybrid architecture produced comparable F1-scores (0.87), showing enhanced robustness to localized noise and confirming that convolutional feature extraction was effective for transient anomaly detection. Hybrid and ensemble strategies achieved the highest performance and operational consistency. The stacking ensemble, combining XGBoost, attention-LSTM, and CNN-LSTM outputs through a Ridge meta-learner, achieved an F1-score of 0.91, demonstrating the complementary strengths of tree-based and sequential learners. The weighted-average ensemble, with validation-optimized weights, achieved a slightly lower F1-score of 0.90 but provided smoother probability outputs suitable for threshold tuning in operational settings. Both ensemble approaches significantly reduced false positives relative to individual models while maintaining high recall, aligning with the need to capture rare anomalous events without overwhelming operators with spurious alerts.

Inference latency was measured for all architectures to ensure compatibility with real-time detection requirements. Tree-based models produced sub-50 ms inference per sample, while LSTM and CNN-LSTM hybrids ranged from 80–120 ms. Ensembles added minimal additional latency (~20 ms) through efficient vectorized computations. These measurements confirmed that even the most complex hybrid architectures could operate within the near-real-time constraints required for transaction monitoring in smart grid environments. Explainability assessment, performed in parallel with quantitative evaluation, verified that

SHAP values from tree-based models correctly highlighted transaction amount spikes, atypical transaction types, and contextual deviations as primary contributors to anomaly predictions. Attention visualizations from recurrent networks consistently assigned higher weights to anomalous timestamps, supporting auditability and operator interpretability. Importantly, these explainable outputs were formatted for blockchain logging, ensuring that each detection included transparent reasoning alongside tamper-evident records, consistent with the framework's governance and regulatory objectives.

Fig.4: Model performance visualizations

4.2 Discussion and Future Work

The results presented in Section 4.1 demonstrate that hybrid intelligence models can effectively address the complex challenges of real-time anomaly detection in smart grid transactions. Baseline logistic regression, while interpretable, exhibited limited detection capability (F1-score 0.65), highlighting the inherent difficulty of capturing rare anomalies within high-variance transactional datasets. Tree-based ensembles, including Random Forest, XGBoost, and LightGBM, significantly improved detection performance (F1-scores ranging from 0.78 to 0.82), confirming that nonlinear feature interactions and engineered rolling-window statistics are critical for capturing subtle deviations in transaction behavior. Feature importance analysis indicated that transaction amount spikes, atypical transaction types, and contextual deviations were consistently the most influential predictors, validating the patterns identified during exploratory data analysis (Section 3.2). Deep learning architectures further enhanced anomaly detection by modeling temporal dependencies that tree-based methods could not fully capture. LSTM models demonstrated the ability to learn sequential correlations, improving the F1-score to 0.84, while Bi-LSTM networks leveraged bidirectional context to achieve marginal gains (0.85), particularly in sequences exhibiting rapid anomaly onset. Attention-augmented LSTM models (F1-score 0.86) provided additional interpretability by weighting historically significant transactions, offering actionable insights for operators. The CNN-LSTM hybrid model (F1-score 0.87) effectively combined local temporal feature extraction with sequential encoding, showing resilience against transient noise and confirming the importance of hybrid architectures for capturing both fine-grained and long-range temporal patterns in energy transaction data.

Ensemble strategies yielded the highest overall performance. The stacking ensemble, which integrated XGBoost, attention-LSTM, and CNN-LSTM outputs through a Ridge meta-learner, achieved an F1-score of 0.91, while the weighted-average ensemble closely followed at 0.90. These ensembles not only enhanced detection accuracy but also reduced false positives (as low as 0.05 for the stacked ensemble), aligning with the operational requirement of minimizing spurious alerts that could undermine stakeholder trust. Inference latency measurements confirmed that all models, including ensembles, operated within near-real-time constraints (sub-150 ms), making them suitable for deployment in distributed smart grid environments. The explainability assessment further demonstrated that SHAP values and attention weightings reliably highlighted the key drivers of anomalies, supporting auditability, operator interpretability, and integration with blockchain-based immutable logging for governance and regulatory compliance (Abubakkar et al., 2025; Machlev, 2022) [4, 13]. Collectively, these findings indicate that a carefully designed hybrid intelligence framework combining tree-based learners, deep temporal networks, and ensemble strategies can deliver high-fidelity anomaly detection with operational viability in terms of latency, false positive control, and interpretability. This approach addresses several key challenges in the literature, including class imbalance, temporal correlation, and the need for explainable

outputs, while providing a foundation for integrating blockchain-backed audit trails (Khan M. A. U. H. et al., 2025; Fariha et al., 2025) [6, 5].

Future Work

Future research should extend this hybrid intelligence framework in several directions. First, evaluating the models on larger and more diverse datasets from multiple regions and market conditions will strengthen the generalizability of the findings. Incorporating federated learning techniques may also provide privacy-preserving capabilities while enabling distributed training across regional smart grid operators (Meng et al., 2024) [8]. Second, adaptive learning strategies that account for concept drift and evolving prosumer behavior could further enhance anomaly detection performance over time, particularly as renewable generation and peer-to-peer trading patterns fluctuate. Third, additional hybrid architectures could be explored, including graph neural networks to capture relationships between interconnected meters or trading nodes, potentially improving detection in networked or correlated fraud scenarios. Fourth, integration with real-time blockchain transaction verification and automated smart contract responses could close the loop between detection and mitigation, creating a self-regulating energy market. Finally, expanding explainability methods to include counterfactual explanations or scenario simulations could further improve operator trust and regulatory transparency, particularly for high-stakes anomaly events.

5. Conclusion

This study presents a hybrid intelligence framework for real-time anomaly detection in smart grid transactions, integrating tree-based learners, deep temporal networks, and ensemble strategies with explainable AI and blockchain-backed logging. Through rigorous preprocessing, exploratory data analysis, and model development, the framework demonstrated the ability to accurately detect anomalous transactions while maintaining low false positive rates and near-real-time inference speeds. Tree-based models effectively captured nonlinear feature interactions, deep learning architectures modeled temporal dependencies, and ensemble strategies leveraged complementary strengths to achieve the highest detection performance. Explainability analyses confirmed that key drivers of anomalies, such as transaction amount deviations and atypical transaction types, could be reliably interpreted, supporting auditability and regulatory compliance.

The findings highlight that hybrid intelligence approaches not only improve detection fidelity but also satisfy operational constraints critical for deployment in distributed smart grid environments. The integration with blockchain ensures tamper-evident logging, providing transparency and governance-ready records for dispute resolution and compliance purposes. Future applications of this framework have the potential to strengthen the resilience and integrity of energy markets, reduce financial and operational risks, and foster stakeholder

trust in increasingly digitized smart grid systems. By combining algorithmic sophistication, interpretability, and secure evidence preservation, this research contributes a practical and scalable solution for advancing fraud detection and anomaly management in modern energy infrastructures.

References

[1] Abubakkar, M., Sharif, K. S., Ahmad, I., Tabila, D. M., Alsaud, F. A., & Debnath, S. (2025, June). Explainable Suicide Risk Prediction with DeepFusion: A Hybrid Intelligence Approach. In 2025 4th International Conference on Electronics Representation and Algorithm (ICERA) (pp. 455-460). IEEE.

[2] Ahmed, I., et al. (2025). Optimizing Solar Energy Production in the USA: Time-Series Analysis Using AI for Smart Energy Management. arXiv preprint arXiv:2506.23368.

[3] Ahad, M. A., et al. (2025). AI-Based Product Clustering for E-Commerce Platforms: Enhancing Navigation and User Personalization. International Journal of Environmental Sciences, 156–171.

[4] Alfayan, M., & Hagras, H. (2025). Towards an explainable artificial intelligence approach for smart grid systems. Discover Artificial Intelligence, 5(1), 40.

[5] Andoni, M., Robu, V., Flynn, D., Abram, S., Geach, D., Jenkins, D., McCallum, P., & Peacock, A. (2019). Blockchain technology in the energy sector: A systematic review of challenges and opportunities. Renewable and Sustainable Energy Reviews, 100, 143–174. ScienceDirect.

[6] Borkovcová, A., Černá, M., & Sokolová, M. (2022). Blockchain in the energy sector—systematic review. Sustainability, 14(22), 14793.

[7] Fariha, N., et al. (2025). Advanced fraud detection using machine learning models: Enhancing financial transaction security. arXiv preprint arXiv:2506.10842.

[8] Guato Burgos, M. F. (2024). A review of smart grid anomaly detection approaches pertaining to artificial intelligence. Applied Sciences, 14(3), 1194. MDPI.

[9] Guo, W. (2025). Power grid load forecasting using a CNN-LSTM network with attention mechanisms. Applied Sciences, 15(5), 2435. MDPI.

[10] Jiang, Z. (2025). Dynamic operating envelopes embedded peer-to-peer-to-grid (P2P2G) trading mechanism considering distribution network integrity. Applied Energy (article available on ScienceDirect). ScienceDirect.

[11] Khan, M. A. U. H., et al. (2025). Secure Energy Transactions Using Blockchain Leveraging AI for Fraud Detection and Energy Market Stability. arXiv preprint arXiv:2506.19870.

[12] Khan, M. N. M., et al. (2025). Assessing the Impact of ESG Factors on Financial Performance Using an AI-Enabled Predictive Model. International Journal of Environmental Sciences, 1792–1811.

[13] Ladjal, B. (2025). Hybrid deep learning CNN-LSTM model for forecasting (solar radiance/power prediction). Scientific Reports (2025).

[14] Li, J., Zhao, H., & Chen, Y. (2025). Blockchain-based secure energy trading with anomaly detection using smart contracts. Energy Reports, 11, 2871-2883.

[15] Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges, and opportunities. Energy and AI, 9, 100169.

[16] Meng, K., Wang, Y., & Zhang, L. (2024). Federated learning for anomaly detection in smart meter data with privacy preservation. IEEE Transactions on Smart Grid, 15(2), 1345-1357.

[17] Rahman, M., Debnath, S., Rana, M., Murad, S. A., Muzahid, A. J. M., Rashid, S. Z., & Gafur, A. (2023, February). Bangla Text Summarization Analysis Using Machine Learning: An Extractive Approach. In International Human Engineering Symposium (pp. 65-80). Singapore: Springer Nature Singapore.

[18] Smith, J., & Clark, P. (2023). Real-time explainable AI for streaming anomaly detection using SHAP values. Pattern Recognition Letters, 170, 52-60.

[19] Uddin, M. M., Ahmad, I., Abubakkar, M., Debnath, S., & Alsaud, F. A. (2025, June). Interpretable Alzheimer's Disease Diagnosis Via CNNs and MRI: an Explainable AI Approach. In 2025 4th International Conference on Electronics Representation and Algorithm (ICERA) (pp. 641-646). IEEE.

[20] Zamil, M. Z. H., Islam, M. R., Debnath, S., Mia, M. T., Rahman, M. A., & Biswas, A. K. (2025, April). Stroke Prediction on Healthcare Data Using SMOTE and Explainable Machine Learning. In 2025 13th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.

[21] Zibaeirad, A., Koleini, F., Bi, S., Hou, T., & Wang, T. (2024). A comprehensive survey on the security of smart grid: Challenges, mitigations, and future research opportunities. Preprint (arXiv:2407.07966). arXiv.