Next Energy 11 (2026) 100556

Contents lists available at ScienceDirect

Next Energy

journal homepage: www.sciencedirect.com/journal/next-energy

Research article

Ensemble learning for event detection and disturbance classification in power quality data from solar energy systems

Nirmalkumar. J. Shiroya[a], Maitreyee Dey[a,*], Soumya P. Rana[b]

[a] Department of Computer Science and Applied Computing, London Metropolitan University, 166-220 Holloway Rd, London N7 8DB, UK
[b] Department of Computer Engineering, University of Greenwich, Medway Campus, Central Ave, Gillingham, Chatham ME4 4TB, UK

A R T I C L E I N F O

Keywords:
Solar energy sources
Big data analysis
Ensemble learning
Disturbance type event detection
Power quality data

A B S T R A C T

Integrating renewable energy sources into power grids introduces complex challenges, particularly in accurately detecting and classifying power quality (PQ) disturbances due to the variability and intermittency of renewable energy generation. This study proposes a classification framework that employs data balancing techniques and ensemble learning models to classify key PQ events, such as voltage sags, waveform distortions, and over-under frequency disturbances. A comprehensive dataset was collected from a solar farm in Norfolk, England, covering January to December 2023, to perform this analysis. By investigating this high-resolution, high-fidelity big PQ data, the research explored real disturbances and provided insights into the solar site's operational behavior, contributing to improved grid reliability. This work offers valuable insights into solar farm operations, helping utility-scale owners and operators implement more effective and cost-efficient condition monitoring strategies.

1. Introduction

With the shift toward renewable energy, monitoring solar grids has become crucial. Integrating more solar power introduces challenges in maintaining operational stability due to its variability, heavily influenced by environmental conditions [1]. This variability causes power quality (PQ) issues such as voltage dips, frequency imbalances, and waveform changes [2]. A solar farm's performance is closely linked to power grid stability. Operational issues like equipment failures or PQ disturbances can disrupt grid balance, causing energy demand shifts and uneven power frequencies that strain the grid [3]. This added stress increases maintenance costs and, in severe cases, may lead to power outages [4]. Reliable solar farm operation is essential for both site efficiency and overall grid stability. Effective monitoring helps operators anticipate disruptions, reduce downtime, and ensure a stable energy supply to the grid [5].

To ensure solar farms operate effectively and support the grid, operators rely on diverse data, continuously measuring voltage, frequency, current, power, and harmonics, along with solar panel performance metrics [6]. Advanced tools like PQ meters, phasor measurement units (PMUs), and Supervisory Control And Data Acquisition (SCADA) systems gather real-time data, giving operators valuable insights into site performance [7,8]. This data helps operators detect issues early and respond swiftly to prevent major problems. However, managing this data poses challenges: disturbances are rare compared to normal conditions, creating dataset imbalances that complicate event detection and classification. Additionally, labeling data requires expertise, often resulting in incomplete datasets [9].

2. Background study of ML-based PQ event detection

Maintaining PQ in grids with high solar integration requires advanced monitoring systems to detect disturbances in real-time. Studies have adapted various data acquisition tools, like power quality meter (PQM) and micro- phasor measurement units (μPMUs), which capture high-resolution data on voltage, frequency, current, etc., enabling the detection of transient events such as voltage sags and harmonic distortions [10]. μPMUs, in particular, have proven effective for tracking voltage and frequency fluctuations in distributed energy systems [8].

Traditional Machine Learning (ML) techniques have proven highly effective for detecting anomalous events in PQ monitoring. Models like Random Forest, Support Vector Machines (SVM), and k-Nearest Neighbors show success in distinguishing PQ disturbances like voltage sags, frequency deviations, and harmonic distortions [11,12]. De Yong et al. have shown that wavelet transforms combined with SVM can effectively detect voltage disturbances, such as sags and swell events, by extracting essential waveform features [13]. Axelberg et al. also demonstrated that SVM's robustness and simplicity enable accurate classification of voltage events, supporting real-time applications. However, as a binary classifier, SVM struggles with multi-class disturbances

* Corresponding author.
E-mail address: m.dey@londonmet.ac.uk (M. Dey).

https://doi.org/10.1016/j.nxener.2026.100556
Received 17 March 2025; Received in revised form 17 December 2025; Accepted 18 February 2026

and compound PQ issues in complex grids, underscoring the need for more advanced solutions in modern power systems [14]. Unsupervised methods, such as clustering and Isolation Forests, further support anomaly detection without the need for labeled data [15]. To deal with data and class imbalance, methods such as Synthetic Minority Oversampling Technique (SMOTE) and SMOTE combined with Edited Nearest Neighbors (SMOTE + ENN) have been explored to generate synthetic samples of rare events, enhancing ML model performance [16].

Traditional ML has some limitations; thus, researchers have increasingly turned to explore deep learning (DL) models, which offer the advantage of automatic feature extraction and improved classification accuracy. DL models have largely been used in the classification of PQ events. Chiam et al. [17] demonstrated the potential of applying Long Short-Term Memory (LSTM) networks to detect synthetic disturbance signals, showcasing these models' value in PQ device monitoring. Hybrid LSTM networks were used for analyzing time-series phasor data [18], making them suitable for identifying time-varying PQ disturbances [19]. Li et al. introduced a 1D-Convolutional Neural Network (CNN) model with multi-resolution analysis that demonstrated high noise resilience but relied on simulated data, raising questions about real-world applicability [20]. Expanding on the capabilities of DL, Vinayagam et al. proposed an ensemble learning model that combines multiple classifiers through voting and stacking techniques alongside DWT-based feature extraction [21]. This model effectively classifies PQ events in a wind-integrated microgrid. Despite its accuracy, this model's complexity and computational demands limit scalability. However, hybrid models, such as Cai et al.'s CNN-Gated Recurrent Unit (GRU) fusion model, effectively capture both spatial and temporal features, achieving high accuracy and strong noise immunity [22]. Similarly, research by Xiao et al. advanced multi-label PQ classification by combining CNN-GRU, ResNet-GRU, and Inception-GRU architectures, which effectively identify simultaneous disturbances, though they face similar scalability challenges [23]. Some studies further enhance the accuracy of PQ disturbance classification [12]. A Deep Auto-Encoder (DAE)-based method is proposed for PQ disturbance classification using Gabor features and a sparse DAE network with SoftMax, which achieves $> 97\%$ accuracy at 20 dB Signal-to-Noise Ratio (SNR), better performing than traditional SVM in both speed and complexity [24]. Along with ML models, several signal processing techniques are also used to address PQ issues, such as unbalanced and distorted grid voltages in 3-phase distribution systems [25]. In addition to a dynamic static compensator (DSTATCOM), a dual fundamental component extraction technique has been proposed to improve PQ in 3-phase systems under unbalanced and distorted grid conditions [26].

2.1. Problem statement

While existing research offers valuable insights, it often depends on simulated data that fails to reflect the complexity and variability of real-world solar farm operations. As a result, many models underperform when deployed in practical scenarios. Obtaining real, high-resolution data is inherently difficult due to the infrequency of actual disturbance events and the limited availability of expert-labeled datasets. The primary research gap, therefore, lies in the scarcity of real-world, high-resolution, accurately labeled, and balanced datasets—combined with the limited capability of conventional machine learning models to effectively handle such complexity in PQ disturbance detection and classification.

Combining this challenge, PQ disturbances in solar farms occur less frequently than in normal operations, creating highly imbalanced datasets. This combination of unrealistic training data and imbalanced real-world data presents significant difficulties for machine learning models, which often struggle to accurately detect and classify these rare events.

Also, traditional SCADA provides data every 15-minute interval, which makes it inefficient to capture some electrical events that may last less time. Consequently, critical disturbances may go undetected or trigger false alarms, weakening the effectiveness of monitoring and delaying responses to real issues. Without granular data along with precise identification and classification of disturbances, data reliability diminishes, limiting the efficacy of monitoring systems and increasing the risk of overlooked disturbances, elevated maintenance costs, and potential instability in the power grid.

2.2. Contribution

Analyzing PQ data is essential for solar farm operators, as it provides insight into the operational performance of the solar farm and enables the early detection of issues like voltage drops or frequency fluctuations. Prompt identification and resolution of these issues can prevent them from escalating, ensuring stable and efficient operations.

Unlike many studies that rely on simulated data, this work uses high-fidelity, high-resolution real-world big data collected from a 33 kV solar site using the patented grid data unit (GDU) developed by Neuville Grid Data [27]. Although this dataset provides detailed operational data, it lacks labeled fault information due to the limited availability of experienced operators. To address this, the study thoroughly analyzed the data patterns, consulting with power engineers when necessary, to identify and label 4 major grid disturbance events. This labeling enabled meaningful interpretation of the previously unstructured dataset. Given the infrequency and randomness of disturbances, which lead to an imbalance between disturbance and normal data, 3 data balancing techniques were evaluated to address this issue. After balancing, 4 ensemble learning models—Category Boosting (CatBoost), Adaptive Boosting (AdaBoost), Light Gradient Boosted Machine (LightGBM), and eXtreme Gradient Boosting (XGBoost)—were applied to classify the 4 types of disturbances. The models' performances were compared to identify the most effective approach for classifying these PQ disturbances in imbalanced, real-world data.

This approach not only enhances the detection and classification of PQ disturbances but also bridges a gap in existing research, which often overlooks real-world data complexities. By providing more accurate insights into solar farm operations, this research empowers operators to make data-driven decisions, improving the management of solar energy systems. Ultimately, these advancements support the integration of renewable energy into the power grid, contributing to a more sustainable energy future.

3. Proposed methodology

The proposed methodology aims to detect and classify PQ disturbances using real-world PQM data from a solar farm in England. The methodology encompasses essential stages to ensure that the data is accurately processed, labeled, and analyzed for robust disturbance detection. This approach consists of 4 main steps: (1) data collection and preprocessing, (2) data labeling, (3) data balancing, and (4) disturbance type event (DTE) classification and evaluation. This 4-step process flow is shown in Fig. 1, illustrating the sequential process from data collection to disturbance classification. By following this approach, the study achieves a structured, data-driven solution to PQ disturbance detection in solar farms, enhancing operational monitoring and contributing to improved solar farm reliability.

3.1. Data collection and pre-processing

The initial step begins with high-resolution PQ data accumulation and preprocessing to ensure it is clean and ready for analysis. This involves removing missing values, specific parameter selection, and standardizing measurements, creating a consistent dataset for further processing.

To gather this high-fidelity, high-precision data, engineers at Neuville Grid Data designed a specialized device, the GDU [27]. This

GDU was installed at 6 solar farm substations across England, connected to UK Power Networks. It integrates several essential components: a PMU, a PQM, a GPS antenna for time synchronization under 100 nanoseconds, solid-state memory for data buffering, and 3G/4G cellular telemetry for data transmission. Two main types of data were gathered: phasor measurements and PQ measurements [28]. This study examined PQ data collected by the PQM device from a 33 kV solar substation in Norfolk, England, accessed from a securely stored private cloud for analysis.

The data storage contains multiple files and metadata files daily. Among these, only the trend dataset—sampled at one-minute intervals—is selected for in-depth analysis. This dataset undergoes comprehensive preprocessing, including removing missing values and selecting relevant parameters. Key electrical parameters, such as Total Harmonic Distortion-voltage (THD-V), Upstream Voltage, Downstream Voltage, Frequency, and Total Demand Distortion-current (TDD-A), are analyzed as they capture both transient and steady-state PQ disturbances. Transient disturbances, which vary over time and are typically unbalanced, contrast with steady-state disturbances that remain balanced and stable, providing essential insights into PQ behavior.

After preprocessing, all daily data across the year 2023 were consolidated into a single file and used for machine learning training and testing purposes.

3.2. Data labeling

Since the data lacks label information on the differnt disturbance events, a critical step in this study is to label DTE from the data, yet the device itself could not automatically label the data. This labeling process was achieved through a combination of data-driven analysis from the meta files associated with the data containing information about daily electrical events, and consultations with power engineers. Through this process, 4 primary disturbance types were identified as these types of events are frequently occurring and potentially impactful to the distribution grid. The identified disturbance events include: (1) Type 1: Voltage sag, (2) Type 2: Under-frequency, (3) Type 3: Over-frequency, and (4) Type 4: Waveform/ Waveshape change.

Five key electrical parameters (upstream voltage, downstream voltage, THD-V, TDD-A, and frequency) were analyzed in this study. Their behavioral patterns under normal conditions and during various disturbance events are illustrated in Fig. 2. The line plots in this figure provide trend visualizations sampled at 1-minute intervals. Each subplot displays a 10-minute window, comprising 10 data points per parameter. These plots represent discrete time-series patterns associated with different disturbance types, offering a high-level overview of transient patterns rather than continuous waveform analysis.

To elaborate further, Fig. 2 (a)(i)-(v) illustrates the normal behavior patterns of all 5 electrical parameters. Specifically, Fig. 2 (a)(i) corresponds to THD-V, Fig. 2 (a)(ii) to upstream voltage, Fig. 2 (a)(iii) to downstream voltage, Fig. 2 (a)(iv) to frequency, and Fig. 2 (a)(v) to TDD-A. In contrast, Fig. 2 (b)-(e) depicts the characteristic variations of these parameters during the 4 identified disturbance events. Fig. 2 (b) shows the response during a voltage sag event, Fig. 2 (c) during an under-frequency event, Fig. 2 (d) during an over-frequency event, and Fig. 2 (e) during a waveshape change. Each subfigure presents the corresponding 10-minute time series behavior across the 5 parameters, highlighting the distinct patterns associated with each disturbance type.

This step was crucial for organizing the unlabeled data, establishing a strong foundation for effective model training, and enhancing the robustness of the models in detecting and classifying significant PQ events.

3.3. Data balancing

PQ disturbances in solar farms are infrequent compared to normal operating conditions, leading to an imbalanced dataset. To address this, data balancing techniques are applied to create a more representative dataset. This step is critical for improving the model's accuracy and reliability in detecting disturbances.

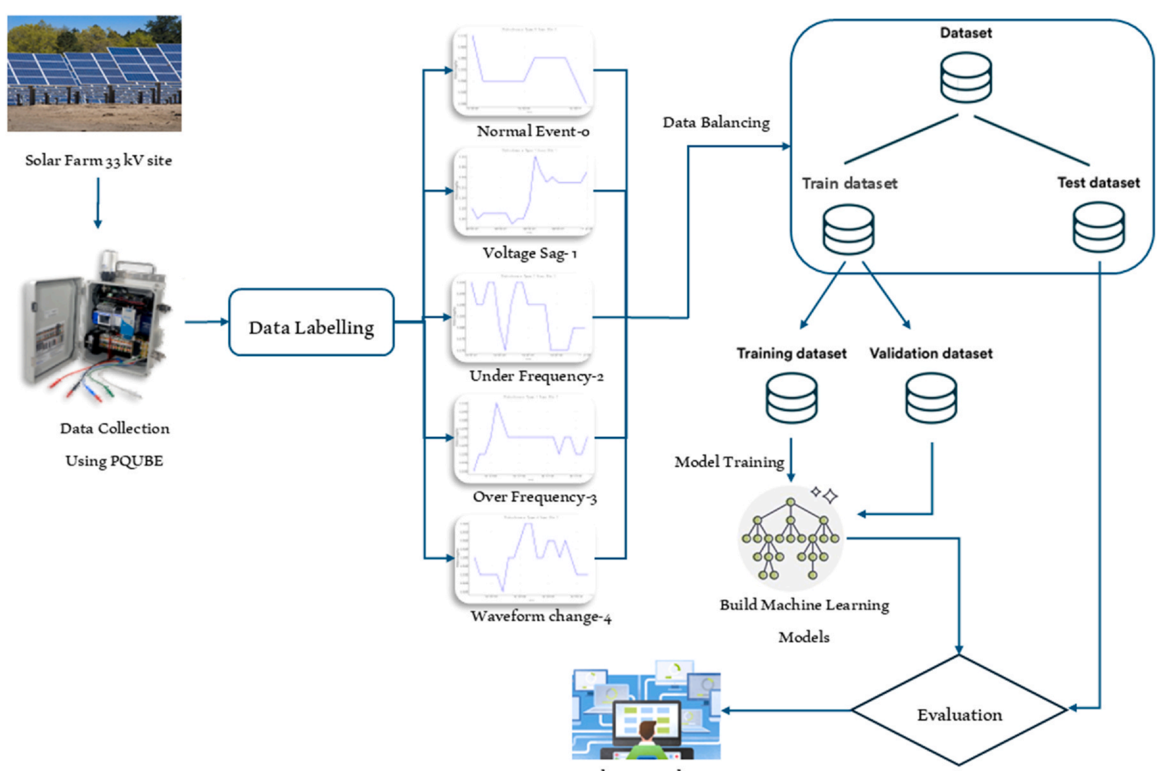

Fig. 1. Flowchart of the proposed methodology, outlining the sequential process from data collection to DTE classification. DTE = disturbance type event.

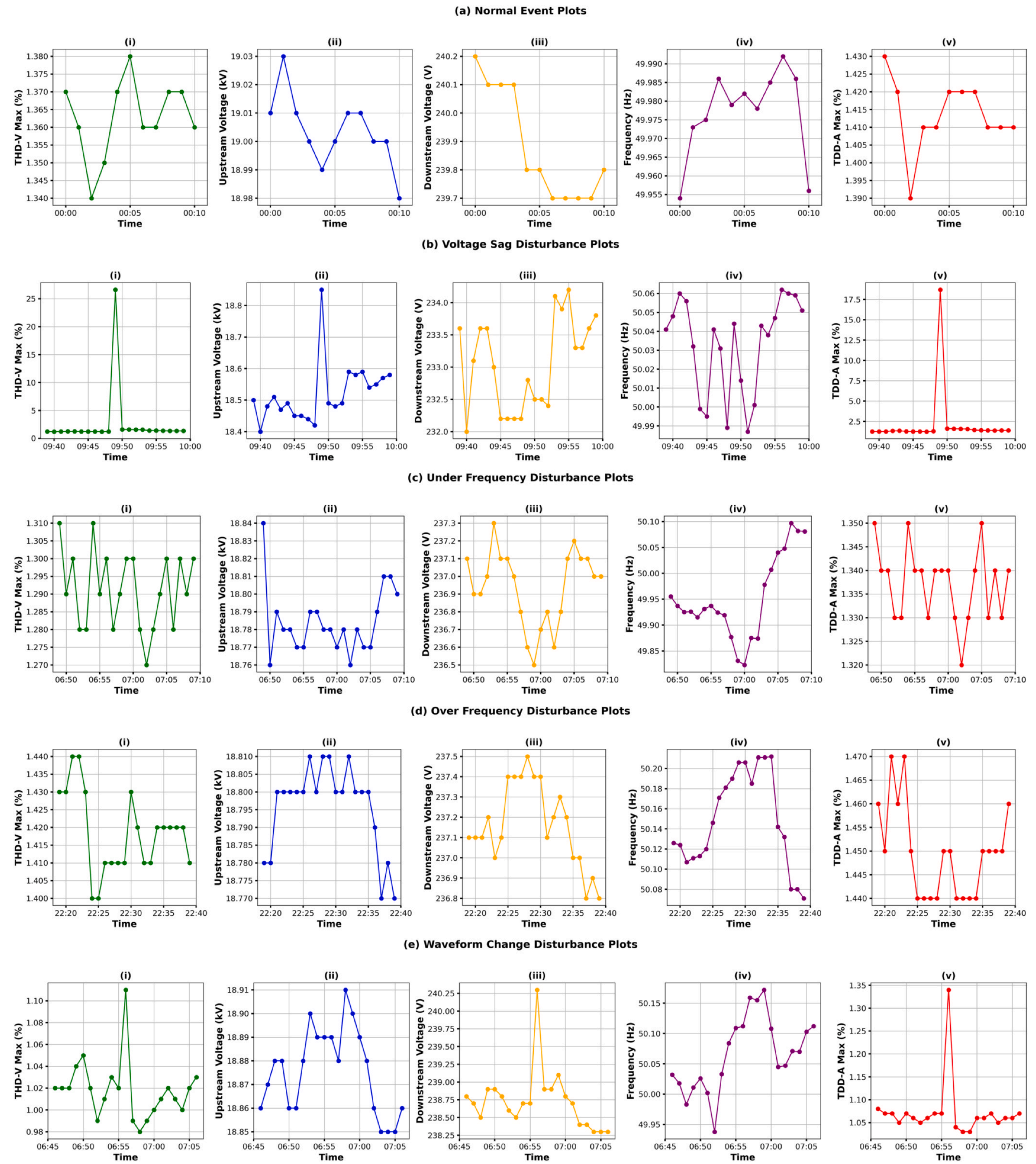

Fig. 2. Ten-min sample windows illustrating the behavior of 5 electrical parameters—THD-V, upstream voltage, downstream voltage, frequency, and TDD-A—under different event conditions: (a) normal operation, (b) voltage sag, (c) under-frequency, (d) over-frequency, and (e) waveshape change. TDD-A = total demand distortion-current; THD-V = total harmonic distortion-voltage.

To ensure rare events are not overshadowed by normal conditions in the classification process, this study evaluates 3 different data balancing methods: (1) SMOTE, (2) SMOTE + ENN, and (3) Adaptive Synthetic Sampling (ADASYN). These techniques help to balance the dataset by generating synthetic instances for minority class samples, ensuring the machine learning models are adequately trained to recognize rare PQ disturbances.

3.4. DTE classification and evaluation

In the final step, 4 ensemble machine learning models—XGBoost, AdaBoost, LightGBM, and CatBoost—are employed to classify the labeled DTE from the PQ data. Ensemble models have been considered here as they perform better for handling large, complex datasets than

Table 1
Performance comparison obtained from different ensemble models' outcomes without the data balancing techniques

Models	Precision	Recall	F1-Score
XGBoost	0.58	0.43	0.45
AdaBoost	0.65	0.40	0.45
LightGBM	0.25	0.22	0.23
CatBoost	0.65	0.41	0.46

Table 2
Average performance comparison of all experimented ensemble models with 3 different data balancing techniques

Model	Data balancing technique	Precision (macro avg)	Recall (macro avg)	F1-score (macro avg)
XGBoost	ADASYN	0.66	0.68	0.67
	SMOTE + ENN	0.60	0.74	0.65
	SMOTE	0.65	0.68	0.66
Adaboost	ADASYN	0.64	0.62	0.62
	SMOTE + ENN	0.62	0.70	0.65
	SMOTE	0.66	0.64	0.64
LightGBM	ADASYN	0.64	0.67	0.65
	SMOTE + ENN	0.60	0.73	0.66
	SMOTE	0.63	0.67	0.65
CatBoost	ADASYN	0.60	0.75	0.66
	SMOTE + ENN	0.58	0.78	0.66
	SMOTE	0.61	0.75	0.67

ADASYN = adaptive synthetic sampling; ENN = edited nearest neighbors; SMOTE = synthetic minority over-sampling technique.

the traditional ML model, reducing the risk of overfitting and underfitting, and helping to identify rare events more accurately.

The performance of each model is evaluated to determine the most effective approach for classifying PQ disturbances. This comprehensive evaluation provides insights into model performance, helping identify the optimal model for real-world applications.

3.5. Evaluation

Precision, recall, F1 score, and the Receiver Operating Characteristic (ROC) curve have been used as evaluation metrics for PQ disturbance event classification. They provide a comprehensive understanding of model performance, especially in imbalanced datasets, and improve operational decisions.

4. Result analysis and discussion

In this results analysis, the dataset size and data partitioning approach used for the study have been described. Following this, the outcomes from experiments using 3 data balancing techniques combined with 4 ensemble learning models are thoroughly analyzed across all 4 disturbance types to determine which model yields the best performance. Finally, the results are presented with figures and explained in detail.

4.1. Data description

Typically, data from traditional SCADA systems is collected at 15-min intervals. In contrast, the GDU provides high-resolution PQM data

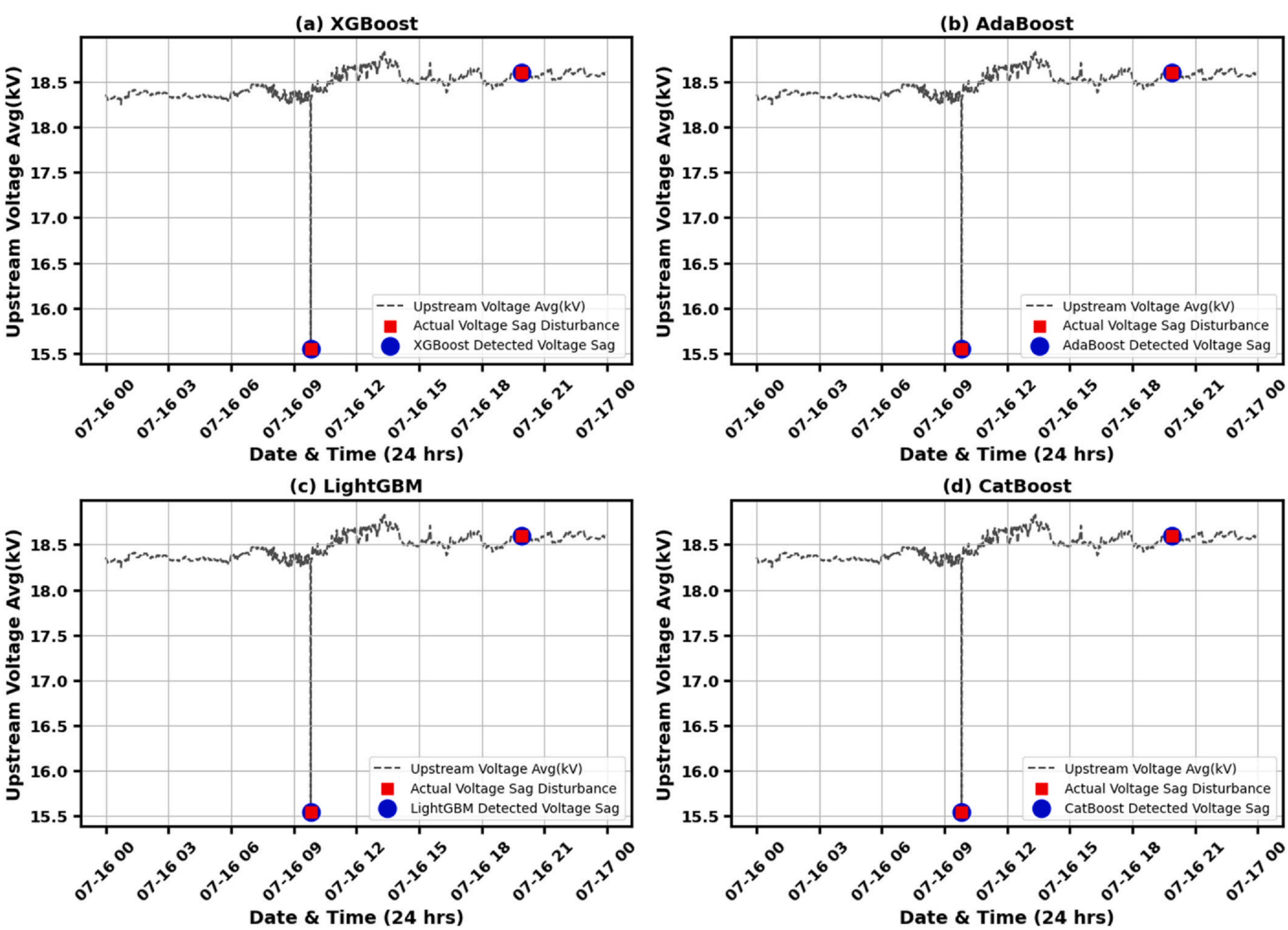

Fig. 3. Comparison of actual vs. predicted voltage sag events using the SMOTE + ENN data balancing technique across 4 ensemble models: (a) XGBoost, (b) AdaBoost, (c) LightGBM, and (d) CatBoost. ENN = edited nearest neighbors; SMOTE = synthetic minority over-sampling technique.

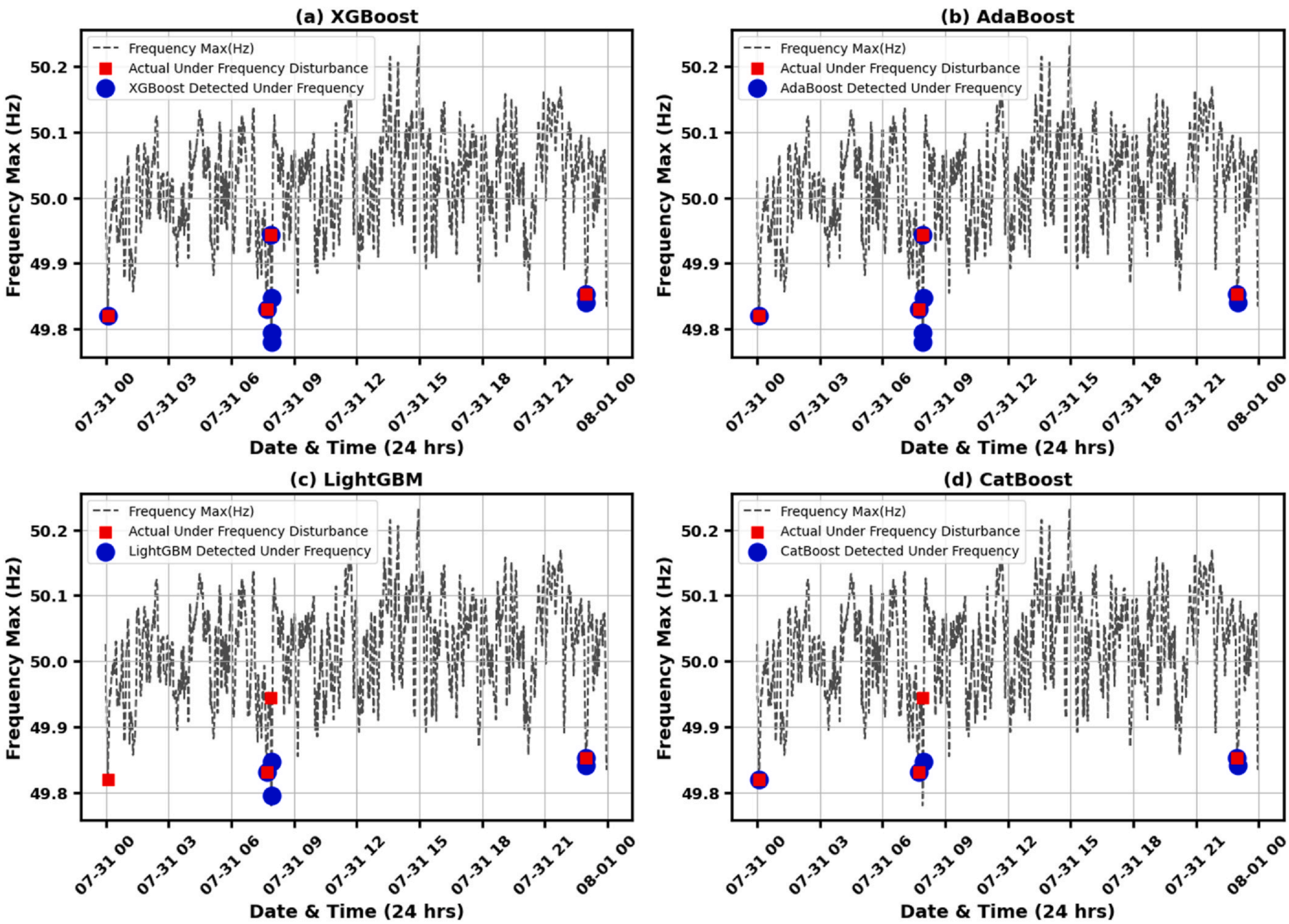

Fig. 4. Comparison of actual vs. predicted under-frequency events using the SMOTE + ENN data balancing technique across 4 ensemble models: (a) XGBoost, (b) AdaBoost, (c) LightGBM, and (d) CatBoost. ENN = edited nearest neighbors; SMOTE = synthetic minority over-sampling technique.

at one-minute intervals, allowing for the capture of more granular information about the substation's performance. This dataset depicts the characteristics of big data. In this study, data from a 33 kV substation connected with a utility-scale solar farm of 13-megawatt capacity was explored, utilizing a total of 1 year of historical data from 2023. From this year's worth of dataset, 4 PQ events were identified within the data and labeled accordingly, as described in Section 3.2.

A total of 525,600 data points per parameter were collected, and with 5 key electrical parameters (upstream voltage, downstream voltage, THD-V, TDD-A, and frequency), approximately 2.63 million data points were processed for training and testing purposes. Of this, roughly 1.3 million data points, covering 6 months, were allocated for training, while the remaining 6 months, containing around 1.3 million data points, were reserved for testing the machine learning models.

4.2. Ensemble learning results with and without data balancing

The advanced data balancing techniques—SMOTE, SMOTE + ENN, and ADASYN have been employed to address the class imbalance challenge inherent in PQ datasets, where normal operational states vastly outnumber critical disturbances. For evaluation, we focused on key performance metrics, including precision, recall, and F1-score.

However, to understand the effect of data balancing techniques on the PQUBE dataset, all models were also evaluated using the 4 ensemble methods without applying any data balancing. Table 1 shows the evaluation metrics for all 4 ensemble models prior to applying any data balancing techniques.

As shown below in Table 2, the F1-score, precision, and recall (macro average) metrics of the models with data balancing exhibit substantial improvement compared to those without, as presented in Table 1. This clearly demonstrates the positive impact of balancing the dataset on detecting minority class events, i.e., disturbances.

After evaluating each data balancing technique and ensemble model for each disturbance type, Table 2 summarizes the performance matrices of all models with different balancing techniques, which indicates that the recall values for all models are highest when using the SMOTE + ENN data balancing technique compared to the other techniques (ADASYN and SMOTE). Recall is a crucial metric that measures the model's ability to correctly identify all relevant instances of a disturbance, reflecting the effectiveness of detecting actual PQ events. In addition to achieving higher recall, the SMOTE + ENN technique also helps maintain the F1-scores of the models, which are generally slightly above those achieved with other balancing techniques across all models. This trend highlights the effectiveness of SMOTE + ENN in enhancing the detection capabilities for PQ disturbances, ensuring that fewer relevant events are missed while balancing precision and recall effectively. Overall, the combination of improved recall and maintained F1-scores with the SMOTE + ENN technique emphasizes its value in optimizing the performance of machine learning models for PQ monitoring.

The strong performance of SMOTE + ENN comes from its 2-step approach. First, SMOTE creates synthetic samples for the minority classes, increasing their representation in the data. However, SMOTE alone can sometimes create overlapping or confusing samples, making

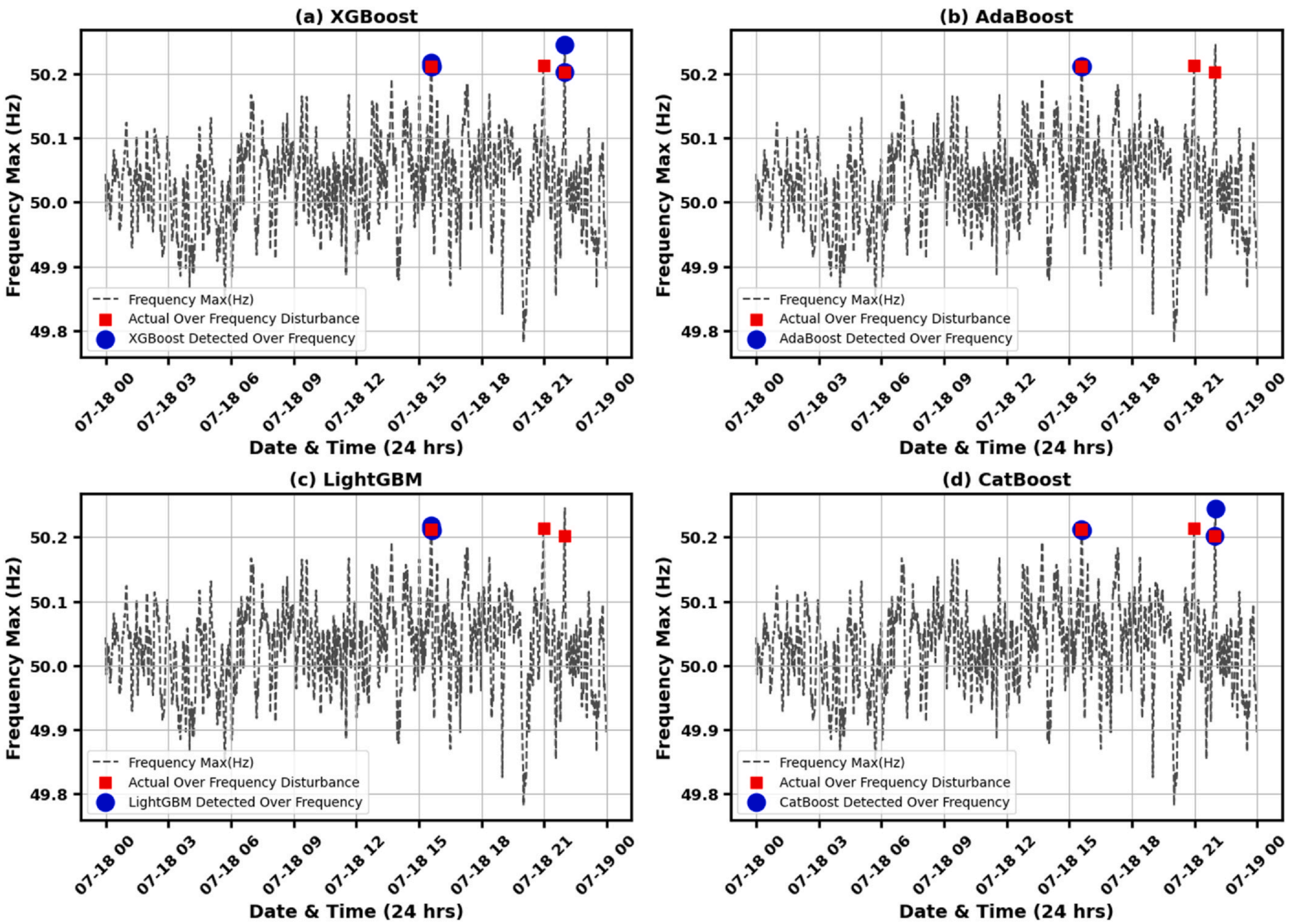

Fig. 5. Comparison of actual vs. predicted over-frequency events using the SMOTE + ENN data balancing technique across 4 ensemble models: (a) XGBoost, (b) AdaBoost, (c) LightGBM, and (d) CatBoost. ENN = edited nearest neighbors; SMOTE = synthetic minority over-sampling technique.

it harder for models to differentiate between different classes. This is where ENN helps by cleaning up the data and removing noisy or unnecessary samples from the majority class. This process is especially useful for detecting PQ disturbances, like voltage sags, which sometimes look similar to normal conditions. Using SMOTE + ENN, the models more effectively recognize the unique characteristics of each disturbance type, resulting in improved accuracy and consistency in predictions. Therefore, the DTE classification outcomes presented here are based solely on the SMOTE + ENN technique.

Fig. 3(a)-(d) illustrates the detection of voltage sag events using 4 ensemble machine learning models—XGBoost, AdaBoost, LightGBM, and CatBoost—applied with the SMOTE + ENN data balancing technique. The time series graphs of upstream voltage (kV) data for a specific date (July 16, 2023) highlight events for 24 h and are presented here solely for visualization purposes. Red squares mark the actual occurrence of voltage sag events, while blue circles indicate events detected by the models. As shown in Fig. 3, all the models accurately identified voltage sag events.

Fig. 4(a)-(d) illustrates the detection of under-frequency events using the same proposed models using the SMOTE + ENN method only. Each time series graph presents frequency data for 24 h on July 31, 2023, where events occurred within the specified range. Actual event occurrences are marked by red squares, while model-detected events are shown as blue circles. In Fig. 4, some models (XGBoost and AdaBost) effectively identify frequency disturbances, and some models (LightGBM and CatBost) incorrectly flag false events, complicating data interpretation. This variability in model performance highlights the need for careful evaluation and refinement to enhance accuracy in detecting frequency disturbances.

Fig. 5(a)-(d) shows the detection of over-frequency events using the same ensemble models, with frequency data displayed for a full day on July 18, 2023, when over-frequency events were recorded. Actual event occurrences are marked by red squares, while model-detected events are indicated by blue circles. In Fig. 5, it has been observed that all frequency disturbances are accurately detected; however, some false positives are also identified by the AdaBoost & LightGBM models, suggesting that while the models are generally effective in capturing over-frequency events, further refinement may be needed to reduce false event detections.

Fig. 6(a)-(d) presents the detection outcomes of waveshape change events using the same 4 ensemble models with the SMOTE + ENN technique. The THD-V data, shown for a full day on September 18, 2023, highlights waveshape change disturbances over 24 h, as THD-V serves as a key parameter for identifying abrupt waveform changes. In Fig. 6, CatBoost demonstrates strong performance, accurately detecting most waveshape disturbances. However, certain models (XGBoost, AdaBoost, and LightGBM) exhibit some false detections, indicating that while the models generally perform well, additional optimization may be beneficial to improve accuracy in waveshape disturbance detection across all models in PQ monitoring.

The overall accuracy and error rates of the models remain similar when comparing results with and without the application of the SMOTE + ENN data balancing technique for detecting both normal events and disturbances, as shown in Table 3. However, the use of high-resolution data

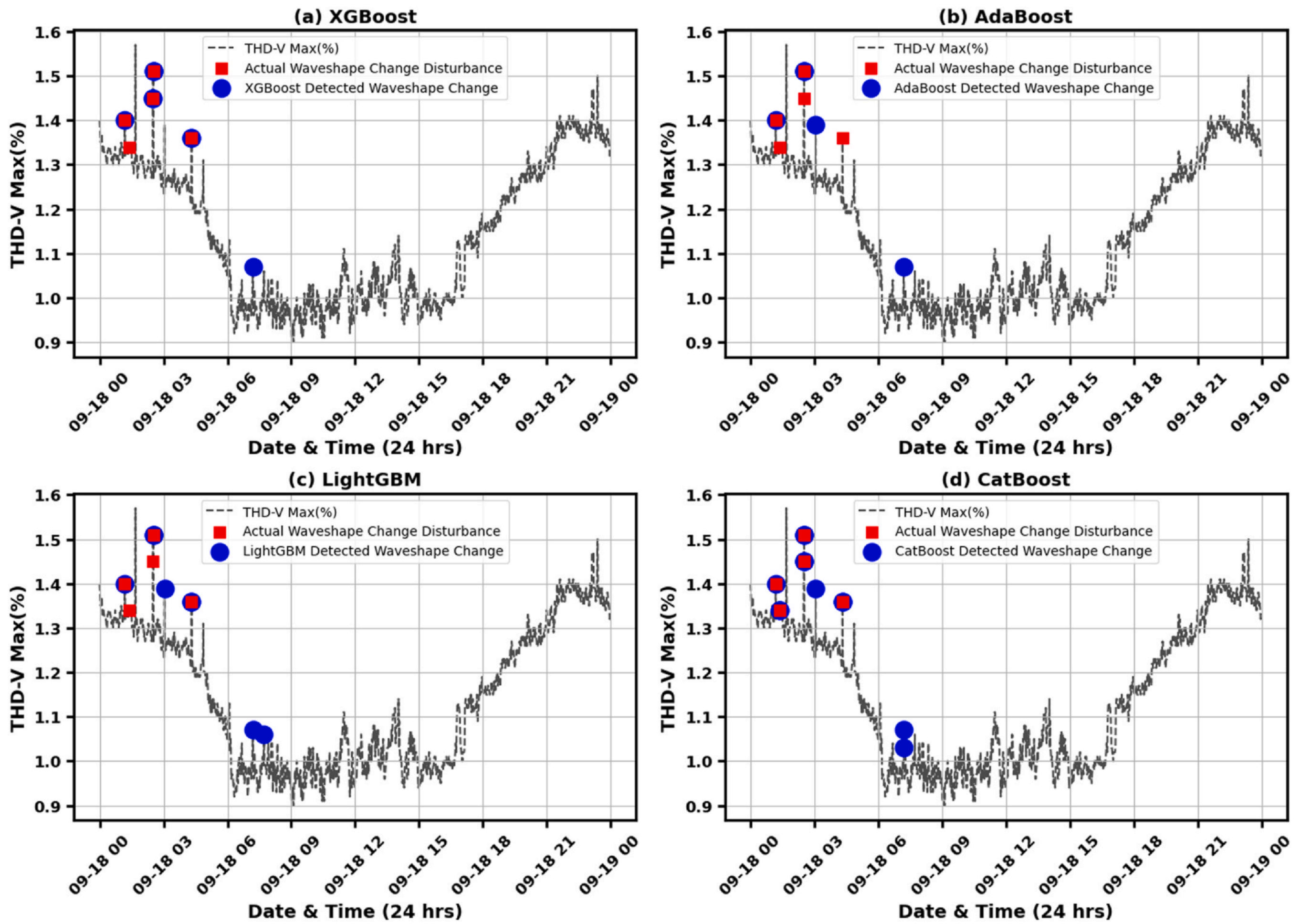

Fig. 6. Comparison of actual vs. predicted waveshape change events using the SMOTE + ENN data balancing technique across 4 ensemble models: (a) XGBoost, (b) AdaBoost, (c) LightGBM, and (d) CatBoost. ENN = edited nearest neighbors; SMOTE = synthetic minority over-sampling technique.

Table 3
The accuracy and error calculations comparison without any data balancing techiniques and using SMOTE + ENN data balancing for all the compared models

Models	Without data balancing		With data balancing (SMOTE + ENN)	
	Accuracy (%)	Error (%)	Accuracy (%)	Error (%)
XGBoost	99.76	0.24	99.76	0.24
AdaBoost	99.78	0.22	99.78	0.22
LightGBM	99.62	0.38	99.77	0.23
CatBoost	99.78	0.22	99.75	0.25

ENN = edited nearest neighbors; SMOTE = synthetic minority over-sampling technique.

enables more precise event labeling and classification. This is evident from the significant improvements observed in precision, recall, and F1-score for disturbance detection across all models when data balancing is applied. These results demonstrate that the combination of high-resolution data and SMOTE + ENN enhances the models' ability to correctly identify true disturbance events, effectively reducing both false negatives and false positives.

4.3. ROC analysis

The use of high-resolution data significantly improves the Area Under the ROC Curve (AUC) across all disturbance types. The Receiver Operating Characteristic (ROC) curve shown in Fig. 7 without data balancing and Fig. 8 with the SMOTE + ENN data balancing technique for all the experimented ensemble machine learning models, such as XGBoost, AdaBoost, LightGBM, and CatBoost, to detect the 4 various PQ disturbances.

Furthermore, in Fig. 7, it can be observed that LightGBM performs poorly under this condition, while the other boosting models appear to show near-perfect detection for both normal and disturbance events. However, this should not be misinterpreted as perfect detection of disturbances. While the ROC curves for both balanced and unbalanced datasets appear similar for boosting models, the true benefit of the data balancing technique is better captured through precision, recall, and F1-score, especially for the minority class. As a result, recall for rare classes is low, even though the overall accuracy and ROC-AUC appear high due to the overwhelming number of correctly classified normal events.

Fig. 8 (a), for Voltage Sag detection, all models achieve an AUC of 1.00, indicating perfect performance in distinguishing these events from normal conditions. This trend continues with under-frequency disturbances, where XGBoost, AdaBoost, and CatBoost also reach an AUC of 1.00, while LightGBM performs nearly as well with an AUC of 0.99. Similarly, all models achieve an AUC of 1.00 for Over Frequency detection, showcasing their strong ability to accurately identify true PQ disturbances. The detection of Waveshape Change, however, presents a greater challenge. In this case, LightGBM achieves the highest AUC of 0.99, followed by XGBoost with an AUC of 0.97, and both AdaBoost and CatBoost score slightly lower at 0.96. This suggests that while the models are highly capable overall, Waveshape Change events are more complex and harder to classify with complete accuracy compared to

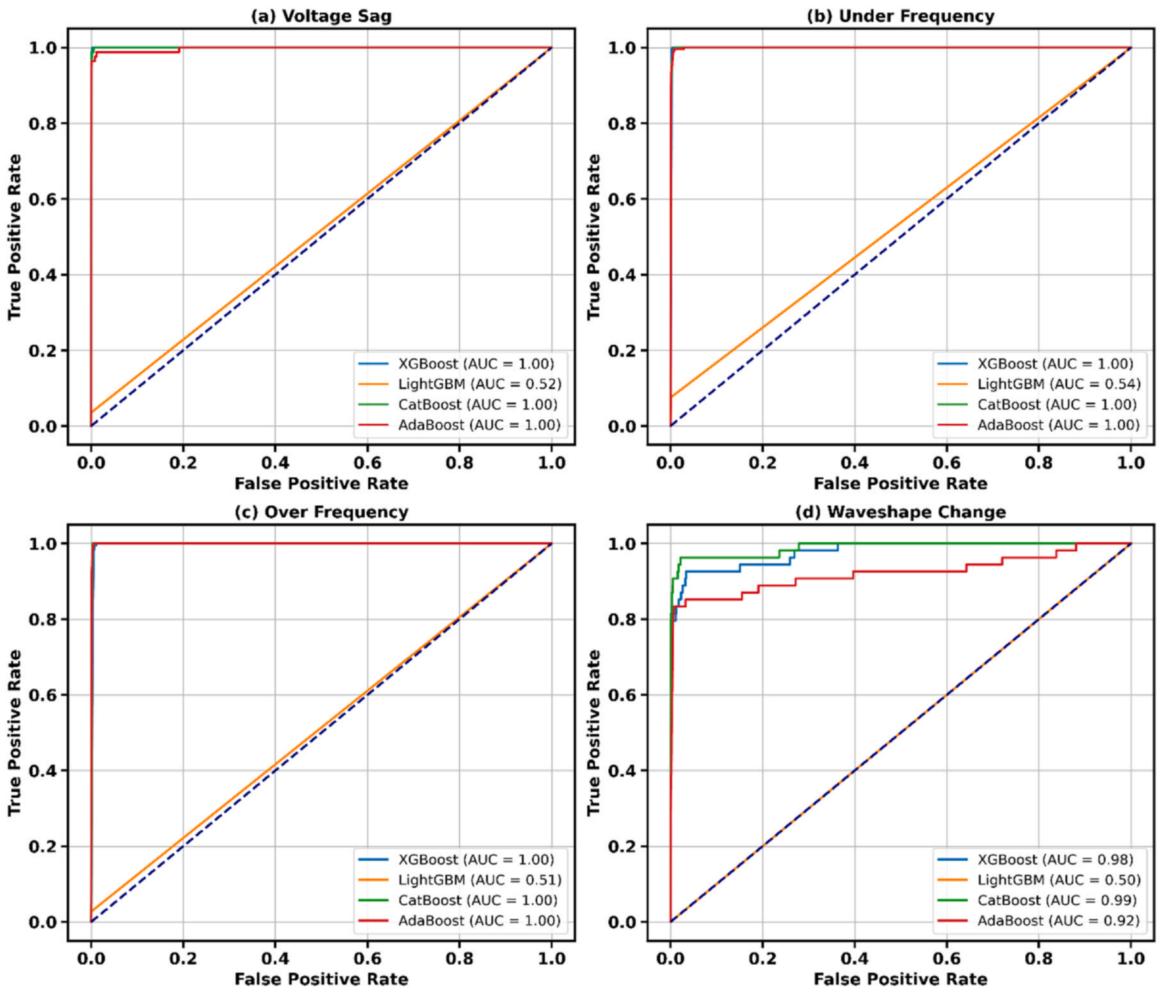

Fig. 7. ROC curves for detecting all disturbance types using various ensemble learning models without any balancing technique.

other disturbances. Additionally, the time complexity is also considered, and found that XGBoost performed faster model training among others.

Here, the use of high-resolution data significantly improves the AUC across all disturbance types. For instance, as shown in Fig. 8 of our paper, models trained on this 1-min resolution data achieve AUC scores close to or equal to 1.00 for Voltage Sag, Over- and Under-Frequency events—indicating near-perfect discrimination capability. In contrast, with coarser resolution, these brief events might go undetected, lowering the true positive rate and thus reducing AUC.

While ROC-AUC values were consistently high before and after data balancing, this metric alone does not reflect class-specific sensitivity. Precision, recall, and F1-score analyses revealed significant performance improvements in rare event detection post-balancing—underscoring the importance of using complementary evaluation frameworks when dealing with imbalanced PQ datasets.

4.4. Discussion

The dataset used in this study was obtained from a real operational solar substation, which captures high-resolution measurements under real grid conditions, including inherent system noise. This contrasts with many previous works that rely on simulated or noise-free data, and our results therefore reflect true noise conditions.

In this research, the ensemble learning models—XGBoost, AdaBoost, LightGBM, and CatBoost—consistently outperformed, offering a superior balance between classification accuracy and computational efficiency. The ensemble learning models inherently handle noisy data better than many traditional ML models due to their tree-based, nonlinear nature. This robustness is reflected in our high ROC-AUC scores (close to or equal to 1.0 for several disturbance types), even when real-world noise is present. Parameters like THD-V and TDD-A inherently contain signal noise and were used for detecting waveform/ waveshape disturbances. Our classification of Type 4: Waveshape Change events, which are more sensitive to noise, showed relatively lower AUC scores (0.96–0.99 compared to 1.0 for other types), indicating the impact of noise on performance. These results demonstrate that while the models are generally robust, noise still slightly degrades performance for certain disturbance types.

Among these, XGBoost combined with the SMOTE + ENN technique yielded the best overall results, achieving high recall and precision for rare disturbance events without incurring excessive training time. This makes ensemble models particularly suitable for real-world PQ monitoring systems, where rapid detection and low-latency inference are critical. These findings highlight the practical advantage of using ensemble methods over more computationally intensive DL approaches in the context of high-resolution, imbalanced PQ datasets.

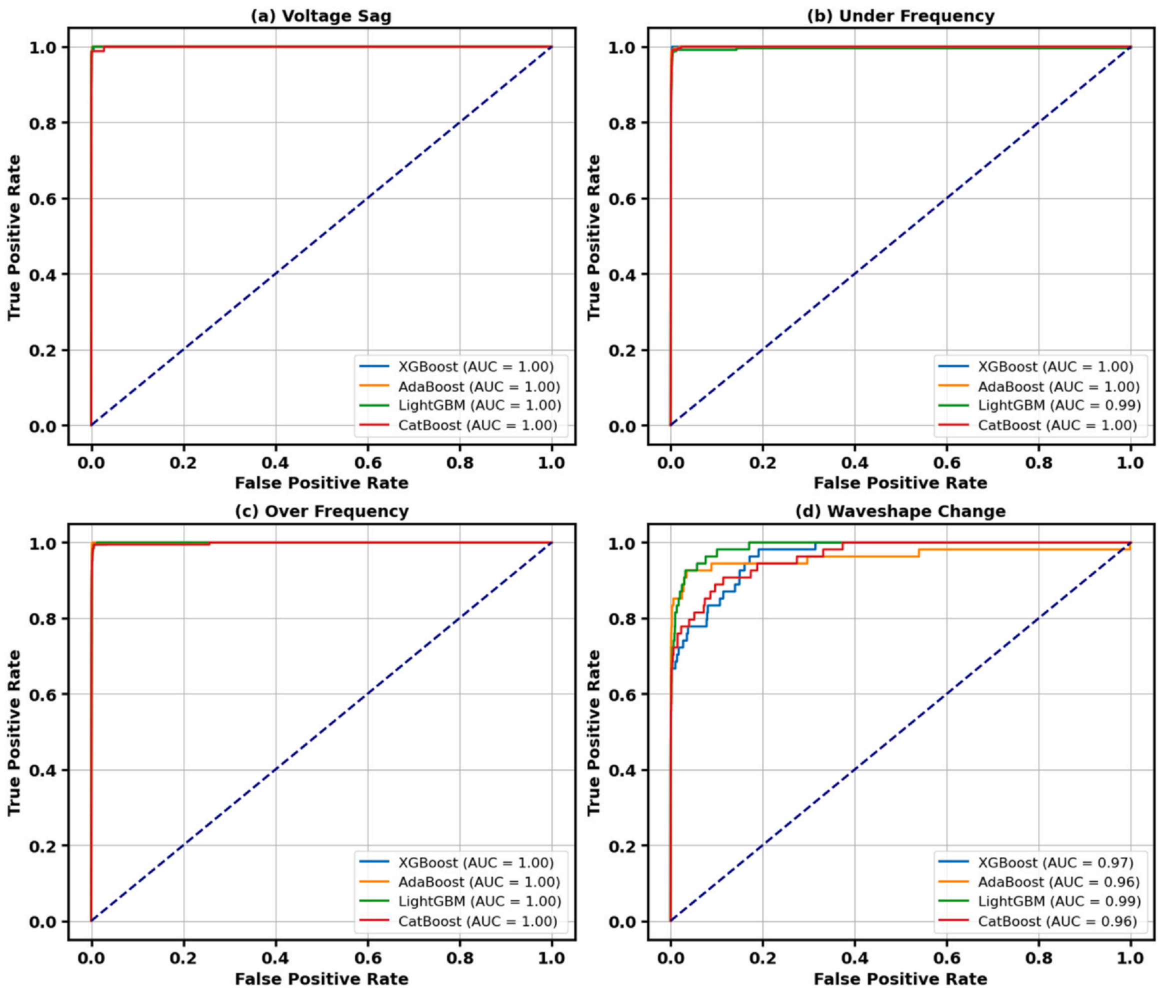

Fig. 8. ROC curves for detecting all disturbance types using various ensemble learning models with the SMOTE + ENN balancing technique. ENN = edited nearest neighbors; SMOTE = synthetic minority over-sampling technique.

5. Conclusions and future work

In conclusion, this study presents a comprehensive approach for enhancing PQ disturbance detection employing ensemble machine learning models, integrated with data balancing techniques. Using this high-resolution PQM data collected over a year from a 33 kV substation, this research effectively classifies disturbances, including Voltage Sag, Under Frequency, Over Frequency, and Waveshape Change. Our findings highlight that the combination of SMOTE + ENN with the XGBoost model consistently outperforms other techniques, achieving superior accuracy in handling class imbalance and capturing the distinct characteristics of PQ disturbances. This approach significantly improves PQ monitoring practices, ensuring better and more proactive management of the grid. Timely corrective actions can reduce downtime, maintain system stability, and safeguard the operation of sensitive equipment, ultimately improving efficiency, reducing maintenance costs, and enhancing grid reliability.

Future work could further refine this framework by exploring additional balancing techniques, such as under-sampling or hybrid models, to optimize model performance. Additionally, the application of hybrid learning architectures tailored to time-series data could improve the accuracy and responsiveness of the model, better handling dynamic PQ disturbances.

CRediT authorship contribution statement

Nirmalkumar J Shiroya: Data curation; software; investigation; methodology; formal analysis; visualisation; writing - original draft. **Maitreyee Dey:** Conceptualisation, Resources, Investigation, Software, Methodology, Data curation, Validation, Visualisation, Writing and reviewing. **Soumya Prakash Rana:** Visualisation, Software, Investigation, Validation, Formal analysis, Reviewing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT to improve the English language style only. All the raw sentences and the work are originally written by the authors. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgments

The authors would like to thank Neuville Grid Data for their provision of the power quality data, which supported the successful completion of this research project.

References

[1] K. Nwaigwe, P. Mutabilwa, E. Dintwa, An overview of solar power (pv systems) integration into electricity grids, Mater. Sci. Energy Technol. 2 (3) (2019) 629–633.

[2] M. Dey, S.P. Rana, J. Wylie, C.V. Simmons, S. Dudley, Detecting power grid frequency events from μ pmu voltage phasor data using machine learning. In: Proceedings of the 11th IET International Conference on Renewable Power Generation - Meeting net zero carbon (RPG 2022), 2022, pp. 125-129, vol. 2022.

[3] C. Nirosha, P.S.K. Patra, Power quality issues of wind and solar energy systems integrated into the grid, Adv. Sci. Lett. 26 (5) (2020) 514–523.

[4] M.H.J. Bollen, F. Hassan, Power Quality Disturbances, 2011. pp. 223-298.

[5] A. Alshahrani, S. Omer, Y. Su, E. Mohamed, S. Alotaibi, The technical challenges facing the integration of small-scale and large-scale pv systems into the grid: a critical review, Electronics 8 (12) (2019) 1443.

[6] S. Ansari, A. Ayob, M.S.H. Lipu, M.H.M. Saad, A. Hussain, A review of monitoring technologies for solar pv systems using data processing modules and transmission protocols: progress, challenges and prospects, Sustainability 13 (15) (2021) 8120.

[7] K. Sayed, H.A. Gabbar, Scada and smart energy grid control automation, Smart Energy Grid Engineering, Elsevier, 2017, pp. 481–514.

[8] M. Dey, S.P. Rana, C.V. Simmons, S. Dudley, Solar farm voltage anomaly detection using high-resolution pmu data-driven unsupervised machine learning, Appl. Energy 303 (2021) 117656.

[9] R.A. de Oliveira, M.H. Bollen, Deep learning for power quality, Electr. Power Syst. Res. 214 (2023) 108887.

[10] S. Parameswaran, M. Dey, P. Patel, Micro-pmu data-driven anomalous voltage event detection for the power distribution system. In: Proceedings of the 12th International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA-2024), Springer, 2024.

[11] O.P. Mahela, A.G. Shaik, N. Gupta, A critical review of detection and classification of power quality events, Energy 65 (1) (2014) 495–505.

[12] I. Topaloglu, Deep learning-based approach for power quality disturbances classification in power transmission systems, J. Electr. Eng. Technol. 18 (1) (2023) 77–88.

[13] D.D. Yong, S. Bhowmik, F. Magnago, An effective power quality classifier using wavelet transform and support vector machines, Expert Syst. Appl. 42 (15) (2015) 6075–6081.

[14] P.G.V. Axelberg, I.Y.-H. Gu, M.H.J. Bollen, Support vector machine for classification of voltage disturbances, IEEE Trans. Power Deliv. 22 (3) (2007) 1297–1303.

[15] U. Singh, M. Dey, P. Patel, Enabling grid stability: harnessing μ pmu data for data-driven analysis of grid frequency events. In: Proceedings of the 12th International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA-2024), Springer, 2024.

[16] S. Chatterjee, Y.-C. Byun, Highly imbalanced fault classification of wind turbines using data resampling and hybrid ensemble method approach, Eng. Appl. Artif. Intell. 126 (2023) 107104.

[17] D.H. Chiam, K.H. Lim, K.H. Law, Lstm power quality disturbance classification with wavelets and attention mechanism, Electr. Eng. 105 (2023) 259–266.

[18] M. Dey, D. Wickramarachchi, S.P. Rana, C.V. Simmons, S. Dudley, Power grid frequency forecasting from μ pmu data using hybrid vector-output lstm network. In: Proceedings of the 2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE). IEEE, 2023. pp. 1-5.

[19] G.S. Chawda, A.G. Shaik, M. Shaik, S. Padmanaban, J.B. Holm-Nielsen, O.P. Mahela, P. Kaliannan, Comprehensive review on detection and classification of power quality disturbances in utility grid with renewable energy penetration, IEEE Access 8 (2020) 146807–146830.

[20] D. Li, I.A. Channa, X. Chen, L. Dou, S. Khokhar, N.A. Azar, A new deep learning method for classification of power quality disturbances using dwt-mra in utility smart grid, Comput. Electr. Eng. 117 (2024) 109290.

[21] A. Vinayagam, V. Veerasamy, M. Tariq, A. Aziz, Heterogeneous learning method of ensemble classifiers for identification and classification of power quality events and fault transients in wind power integrated microgrid, Sustain. Energy Grids Netw. 31 (2022) 100752.

[22] J. Cai, K. Zhang, H. Jiang, Power quality disturbance classification based on parallel fusion of cnn and gru, Energies 16 (10) (2023) 4029.

[23] X. Xiao, K. Li, Multi-label classification for power quality disturbances by integrated deep learning, IEEE Access 9 (2021) 2150–2160.

[24] P. Khetarpal, N. Nagpal, M.S. Al-Numay, P. Siano, Y. Arya, N. Kassarwani, Power quality disturbances detection and classification based on deep convolution auto-encoder networks, IEEE Access 11 (2023) 46026–46038.

[25] S. Prajapati, R. Garg, P. Mahajan, Honey badger algorithm based pi controller for dc link voltage control of solar photovoltaic system connected to grid for enhanced power quality, Electr. Power Compon. Syst. (2024) 1–20.

[26] A. Ram, P.R. Sharma, R.K. Ahuja, Y. Arya, A dual fundamental current extraction controller for dstatcom in the distribution system for power quality improvement under non-ideal source voltage condition, Electr. Power Compon. Syst. (2023) 1–24.

[27] C.V. Simmons, Methods and apparatus for the sensing, collecting, transmission, storage, and dissemination of high-resolution power grid electrical measurement data, uS Patent App. 17/057,611, 2021.

[28] D. Maitreyee, S.P. Rana, High-resolution electrical measurement data processing, uS Patent App. 18/248,519, 2024.