WJERT, 2026, Vol. 12 Issue 8, 169-190. **Research Article** **ISSN 2454-695X**

World Journal of Engineering Research and Technology

www.wjert.org

Impact Factor: 8.067 **Coden USA: WJERA4**

A HYBRID XGBOOST–GRAPH NEURAL NETWORK FRAMEWORK FOR POWER SYSTEM VOLTAGE STABILITY PREDICTION

Isaiah U. Imoh[1*], Nseobong I. Okpura[2], Kufre M. Udofia[3]

Article Received on 05/07/2026 **Article Revised on 25/07/2026** **Article Published on 01/08/2026**

***Corresponding Author**

Isaiah U. Imoh

https://doi.org/10.5281/zenodo.21718557

How to cite this Article: Isaiah U. Imoh[1*], Nseobong I. Okpura[2], Kufre M. Udofia[3]. (2026). A Hybrid Xgboost–Graph Neural Network Framework For Power System Voltage Stability Prediction). World Journal of Engineering Research and Technology, 12(8), 169–190.

ABSTRACT

The increasing penetration of distributed generation and the dynamic, nonlinear nature of modern power systems have rendered traditional model-based voltage stability assessment methods inadequate, particularly because they cannot capture both feature-level nonlinearities and topological dependencies. This paper proposes a novel hybrid framework that synergistically combines Extreme Gradient Boosting (XGBoost) and Graph Neural Networks (GNNs) for accurate and interpretable voltage stability prediction. While XGBoost excels at modeling complex relationships among scalar system features such as voltage stability index (VSI), reactive power margin, and frequency deviations, GNNs capture the spatial and topological interdependencies of the power network by representing buses as nodes and transmission lines as edges. The hybrid model fuses the complementary representations via a weighted concatenation mechanism, followed by a logistic classifier to predict binary stability states (stable/unstable). Extensive experiments on the IEEE 14-bus system, using both balanced and unbalanced datasets, demonstrate that the proposed hybrid model significantly outperforms standalone XGBoost and GNN models. Under unbalanced conditions, the hybrid model achieves 0.87 in accuracy, 0.72 in recall for instability detection, and 0.72 in AUROC. Physically, it reduces VSI error to 0.25 and Q-margin deviation to 4.90, indicating superior alignment with system dynamics. These results establish the hybrid XGBoost–GNN architecture as a robust, topology-aware solution for real-time voltage stability monitoring.

KEYWORDS: Voltage stability, power system, XGBoost, Graph Neural Networks (GNNs), hybrid model, IEEE 14-bus.

1.0 INTRODUCTION

The stability of electrical power systems is fundamental to the reliable operation of modern critical infrastructures, including industrial processes, healthcare, and communication networks (Yang et al., 2022). However, the rapid integration of distributed generation (DG), renewable energy sources, and intelligent electronic devices has introduced unprecedented complexity, characterised by nonlinear dynamics, stochastic load variations, and bidirectional power flows (Huang, 2021). Traditional stability assessment methods, which rely on deterministic differential-algebraic equations and small-signal analysis, struggle to capture these complexities in real time (Bahmanyar & Karami, 2023). Consequently, there is a pressing need for data-driven, intelligent frameworks that can accurately and efficiently predict voltage instability.

Machine learning (ML) has emerged as a transformative paradigm in power system monitoring. Among ML algorithms, Extreme Gradient Boosting (XGBoost) has demonstrated exceptional performance in handling tabular and time-series data, effectively modeling system-level parameters such as total load, generation, and frequency deviations (Asif et al., 2023). Conversely, Graph Neural Networks (GNNs) have shown remarkable capabilities for learning the spatial dependencies inherent in the interconnected topology of buses, substations, and transmission lines (Scarselli, 2022). GNNs can propagate local perturbations (e.g., a voltage dip at one bus) through the network to infer system-wide stability implications.

Despite their individual strengths, XGBoost and GNNs each possess limitations when applied in isolation to voltage stability prediction. XGBoost, being a tree ensemble method, does not inherently model the topological structure of the power grid; it treats each sample independently and cannot explicitly capture inter-bus electrical dependencies (Ustun et al., 2021). On the other hand, GNNs require substantial training data and computational resources to converge to meaningful spatial representations, and they may under-utilise global feature interactions that tree-based models capture efficiently (Yang et al., 2021). To date, no unified framework has been proposed that synergistically combines the feature-level learning strength of XGBoost with the topology-aware representation learning of GNNs for voltage stability prediction. This paper addresses that gap.

The primary contributions of this work are as follows

i. A novel hybrid XGBoost–GNN model that fuses temporal and spatial dependencies for voltage stability prediction. The model uses XGBoost to learn from scalar stability indicators (e.g., VSI, Q-margin, LOF) and a GNN to learn from the graph-structured power network (nodes = buses, edges = transmission lines).

ii. A comprehensive preprocessing and feature engineering pipeline incorporating wavelet-based denoising, Kalman filtering, and extraction of physics-informed features such as Line Overloading Factor (LOF) and Frequency Deviation Metrics (FDM).

iii. Extensive comparative evaluation on the IEEE 14-bus system under both balanced and unbalanced dataset conditions, using not only classification metrics (accuracy, precision, recall, F1-score, AUROC) but also physical stability metrics (VSI error, Q-margin deviation, frequency RMS error).

iv. Demonstration that the hybrid model consistently outperforms standalone XGBoost and GNN models, achieving higher recall for instability detection and lower physical error margins, thereby offering a more reliable and interpretable tool for grid operators.

The remainder of this paper is structured as follows. Section 2 presents the methodology, including the graph formulation of the power system, the standalone XGBoost and GNN models, and the proposed hybrid fusion strategy. Section 3 describes the experimental setup, dataset generation, and evaluation metrics. Section 4 reports the results and provides a comparative analysis. Section 5 discusses the implications of the findings. Finally, Section 6 concludes the paper and suggests directions for future research. An overview of the proposed hybrid XGBoost–GNN framework is presented in Figure 1.

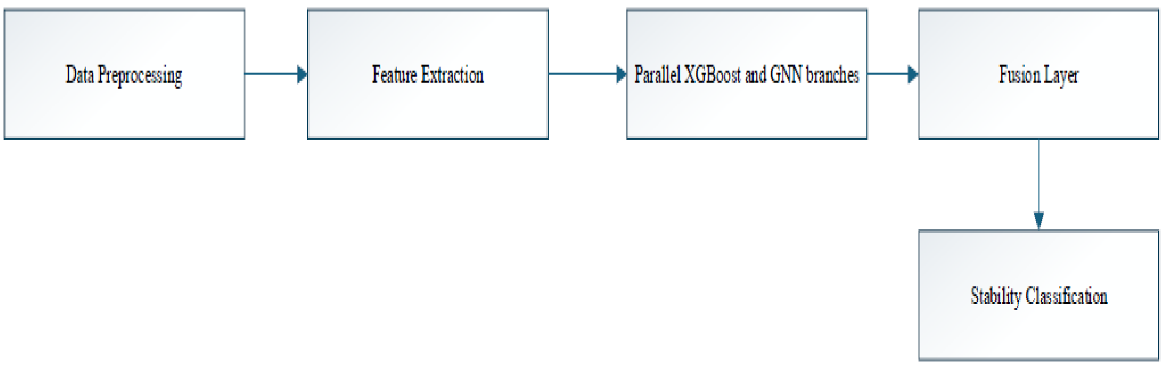

Figure 1. Overview of the proposed hybrid XGBoost–GNN framework for voltage Stability prediction.

2.0 METHODOLOGY

This section describes the complete framework for the proposed hybrid XGBoost–GNN voltage stability prediction model. It begins with the graph formulation of the power system, followed by the standalone XGBoost and GNN models, and finally details the hybrid fusion strategy and training procedure.

2.1 Graph Formulation of the Power System

To enable topology-aware learning, the power system is represented as a weighted undirected graph (Scarselli, 2022). Let the graph be defined as expressed in (1):

(1) where $\mathcal{V}$ is the set of nodes (buses/substations) with $|\mathcal{V}| = N$ (for IEEE 14-bus, $N = 14$), $\mathcal{E}$ is the set of edges (transmission lines) with $|\mathcal{E}| = M$ (here $M = 20$), $X \in \mathbb{R}^{N\times d}$ is the node feature matrix, and $A \in \mathbb{R}^{N\times N}$ is the adjacency matrix describing electrical connectivity. The adjacency matrix is constructed as given in (2): (2)

where w_{ij} is a weight corresponding to the electrical distance or line impedance between buses i and j. For each bus, the node feature vector $x_i \in \mathbb{R}^d$ includes: voltage magnitude V_i, phase angle δ_i, active power injection P_i, reactive power injection Q_i, voltage stability index (VSI), reactive power margin ($Q_{\text{margin},i}$), and bus frequency deviation f_i. Edge features (resistance R, reactance X, susceptance B, apparent power flow S, and line overloading factor LOF) are also stored but used primarily for feature extraction rather than direct graph convolution.

Figure 2 shows the IEEE 14-bus system topology represented as a graph (Nodes represent buses, edges represent transmission lines. Slack, PV, and PQ bus types are colour coded).

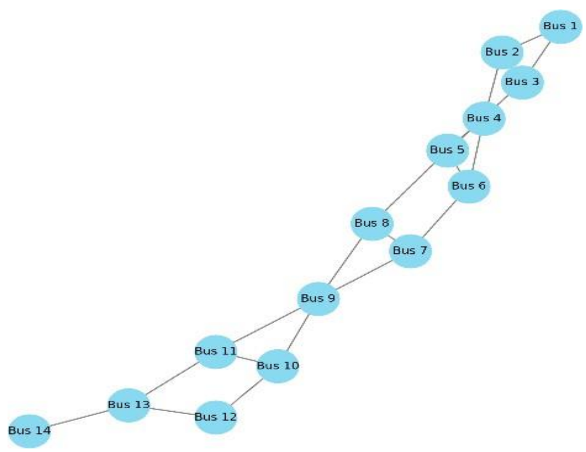

Figure 2. IEEE 14-bus system topology represented as a graph.

2.2 Standalone XGBoost Model for Feature-Based Prediction

Extreme Gradient Boosting (XGBoost) is a scalable ensemble learning algorithm based on gradient-boosted decision trees. It sequentially adds decision trees that correct the residuals of the previous ensemble (Chen & Guestrin, 2016). For a given sample x_i, the predicted output after t iterations is expressed in (3):

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i), f_k \in \mathcal{F} \quad (3)$$

where $\mathcal{F}$ is the space of regression trees.

At each iteration, XGBoost minimises a regularised objective function that balances model fit and complexity, as seen in (4):

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (4)$$

With

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (5)$$

Where T is the number of leaves, are leaf weights, and γ, λ are regularisation parameters.

The algorithm uses both first-order (g_i) and second-order (h_i) gradients to efficiently determine the optimal tree splits similar to Asif et al. (2023). In this work, XGBoost was trained on the extracted scalar feature set (VSI, Q-margin, LOF and frequency rolling statistics) for binary classification (stable vs. unstable) and to provide feature importance rankings.

2.3 Standalone Graph Neural Network (GNN) for Topological Learning

The GNN model is designed to capture spatial dependencies by propagating information across the graph structure. The core of the GNN is the message-passing mechanism. The layer-wise propagation rule for a Graph Convolutional Network (GCN) is defined as (Kipf & Welling, 2017):

$$H^{(k+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(k)}W^{(k)}\right) \quad (6)$$

where $\tilde{A} = A + I_N$ is the adjacency matrix with added self-loops, $\tilde{D}$ is the degree matrix of $\tilde{A}$, $H^{(k)}$ is the node feature matrix at layer k, $W^{(k)}$ is a trainable weight matrix, and σ is a ReLU activation function. In this work, a three-layer GCN is employed, as expressed in (7) –(9):

$H^{(1)} = \mathrm{ReLU}(\tilde{A}XW^{(0)})$ (7)

$H^{(2)} = \mathrm{ReLU}(\tilde{A}H^{(1)}W^{(1)})$(8)

$Z = \mathrm{ReLU}(\tilde{A}H^{(2)}W^{(2)})$(9)

where $Z \in \mathbb{R}^{N\times d_{\mathrm{out}}}$ are the final node embeddings. To obtain a system-level prediction, a readout function (global mean pooling) aggregates node embeddings into a graph-level representation $h_{\mathcal{G}}$:

$h_{\mathcal{G}} = \frac{1}{N}\sum_{i=1}^{N} Z_i$ (10)

This graph-level vector is then passed through a multi-layer perceptron (MLP) with a sigmoid activation to output the probability of instability:

$\hat{y}_{\mathrm{GNN}} = \sigma(\mathrm{MLP}(h_{\mathcal{G}}))$ (11)

The GNN was trained using binary cross-entropy loss. Figure 3 illustrates the architecture of the Graph Convolutional Network (GCN) for voltage stability prediction.

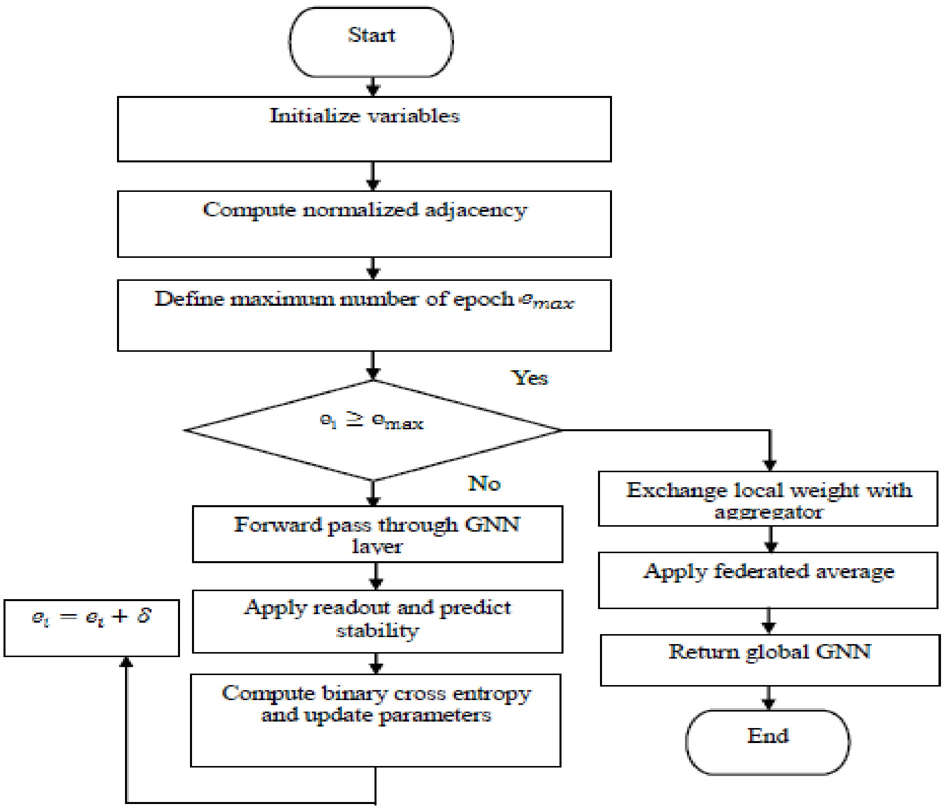

Figure 3. Architecture of the Graph Convolutional Network (GCN) for voltage stability prediction.

2.4 Proposed Hybrid XGBoost–GNN Model

The hybrid model synergistically combines feature-level learning from XGBoost with topology-aware learning from the GNN. Let $H_G \in \mathbb{R}^p$ be the graph-level embedding output by the GNN (after readout and before the final MLP). Let $H_X \in \mathbb{R}^q$ be a feature importance-weighted vector derived from the XGBoost model (e.g., the logit output or a concatenation of leaf node contributions). The hybrid representation is obtained via a weighted concatenation, as expressed in (12):

$$H_{\text{hybrid}} = \alpha H_G \oplus (1-\alpha)H_X \quad (12)$$

where $\oplus$ denotes vector concatenation, and $\alpha \in [0,1]$ is a tunable hyperparameter that controls the relative contribution of topological vs. feature information.

In this implementation, $\alpha = 0.5$ is used as a baseline, with sensitivity analysis performed. The resulting hybrid vector is then fed into a fully connected classifier with a sigmoid activation to produce the final stability prediction, as computed from (13):

$$\hat{y} = \sigma\left(W_c H_{\text{hybrid}} + b_c\right) \quad (13)$$

The entire hybrid model is trained end-to-end: the GNN parameters, the XGBoost parameters (which are pre-trained and then fine-tuned), and the classifier parameters (W_c, b_c) are jointly optimised using the binary cross-entropy loss. The training algorithm is presented in Algorithm 4 of the thesis (XGBoost–GNN Stability Prediction Framework). Figure 4 depicts the hybrid XGBoost–GNN model.

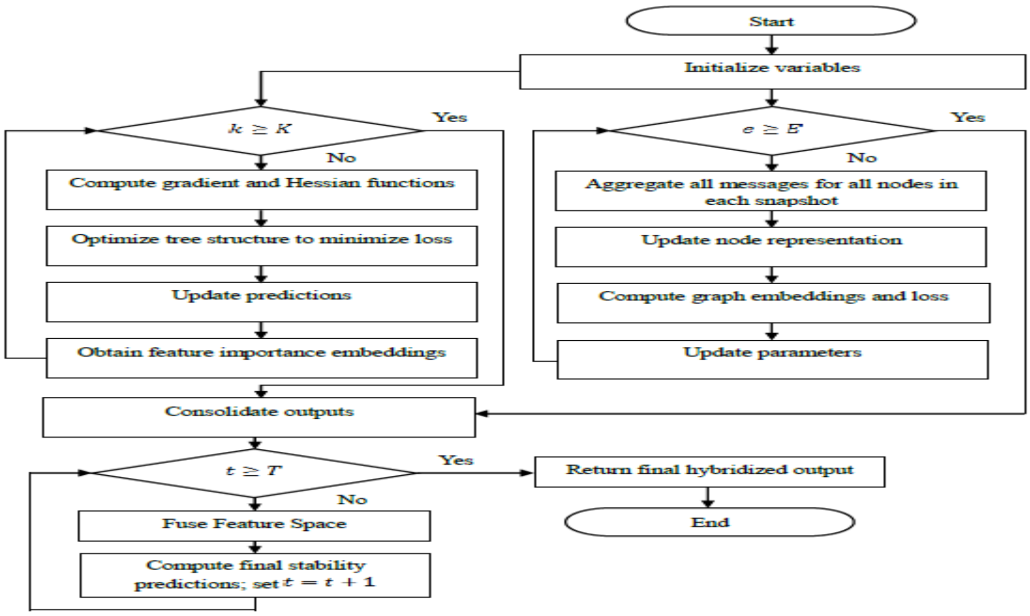

Figure 4. Flowchart of the hybrid XGBoost–GNN model.

2.5 Preprocessing and Feature Engineering

To ensure data quality and model robustness, a preprocessing pipeline is applied, as illustrated in Figure 5. Raw measurements are denoised using a two-stage approach: first, a wavelet transform with soft thresholding removes high-frequency noise; second, a Kalman filter provides real-time state estimation under Gaussian noise assumptions (Amroune, 2021). Missing values are imputed using a denoising autoencoder (DAE) as described in Algorithm 1 of the thesis. All features are then normalised using min-max scaling to the range $[0, 1]$. The following physics-informed features are extracted for each snapshot:

- **i. Voltage Stability Index (VSI):** A scalar indicating proximity to voltage collapse.
- **ii. Reactive Power Margin (Q-margin):** Both raw and normalised forms.
- **iii. Line Overloading Factor (LOF):** Ratio of apparent power flow to thermal rating.
- **iv. Frequency Deviation Metrics:** Rolling mean, RMS, and maximum absolute deviation over a sliding window.
- **v. System-level indicators:** Generation–load balance, average voltage deviation, minimum bus voltage, maximum line loading.

These features serve as input to XGBoost, while node-wise versions (per bus) are used as node features in the GNN.

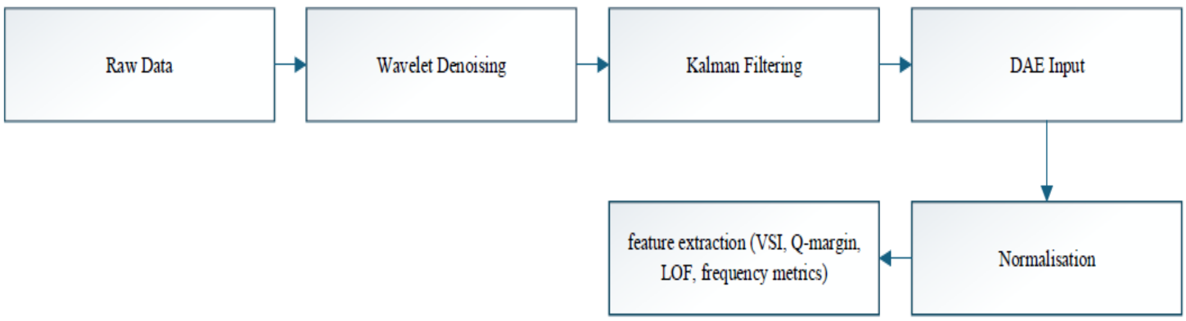

Figure 5. Overview of the preprocessing and feature engineering pipeline.

2.6 Training and Evaluation Protocol

The dataset (5000 samples from IEEE 14-bus simulations) is split into training (70%), validation (15%), and test (15%) sets. For the standalone XGBoost, hyperparameters are tuned via grid search: number of trees (300), maximum depth (6), learning rate (0.05), subsample (0.8), and colsample_bytree (0.8). For the GNN, the Adam optimiser is used with a learning rate of 10^{-3}, 200 epochs, and a hidden dimension of 64. The hybrid model is trained for 200 epochs with early stopping (patience = 20) based on validation loss. All simulations were repeated five times with different random seeds; reported results are averages. Performance is evaluated using standard classification metrics (accuracy, precision, recall, F1 score, AUROC) and physical stability metrics (VSI error, Q margin deviation, frequency RMS error).

3.0 RESULTS

3.1 Dataset Characteristics and Feature Distributions

The dataset comprises 5,000 time-windowed samples generated from IEEE 14-bus simulations. Each sample includes global, node, and edge features, along with a binary stability label (0 = stable, 1 = unstable). The extracted feature set includes Voltage Stability Index (VSI), reactive power margin (Q-margin, normalised), Line Overloading Factor (LOF), line utilisation (LineUtil), frequency rolling statistics (mean, RMS, max absolute deviation), generation–load balance, average voltage deviation, minimum bus voltage, and maximum line loading. Table A (Appendix I) shows the first five rows of extracted features. Figure 6 shows the feature correlation heatmap for the extracted stability indicators.

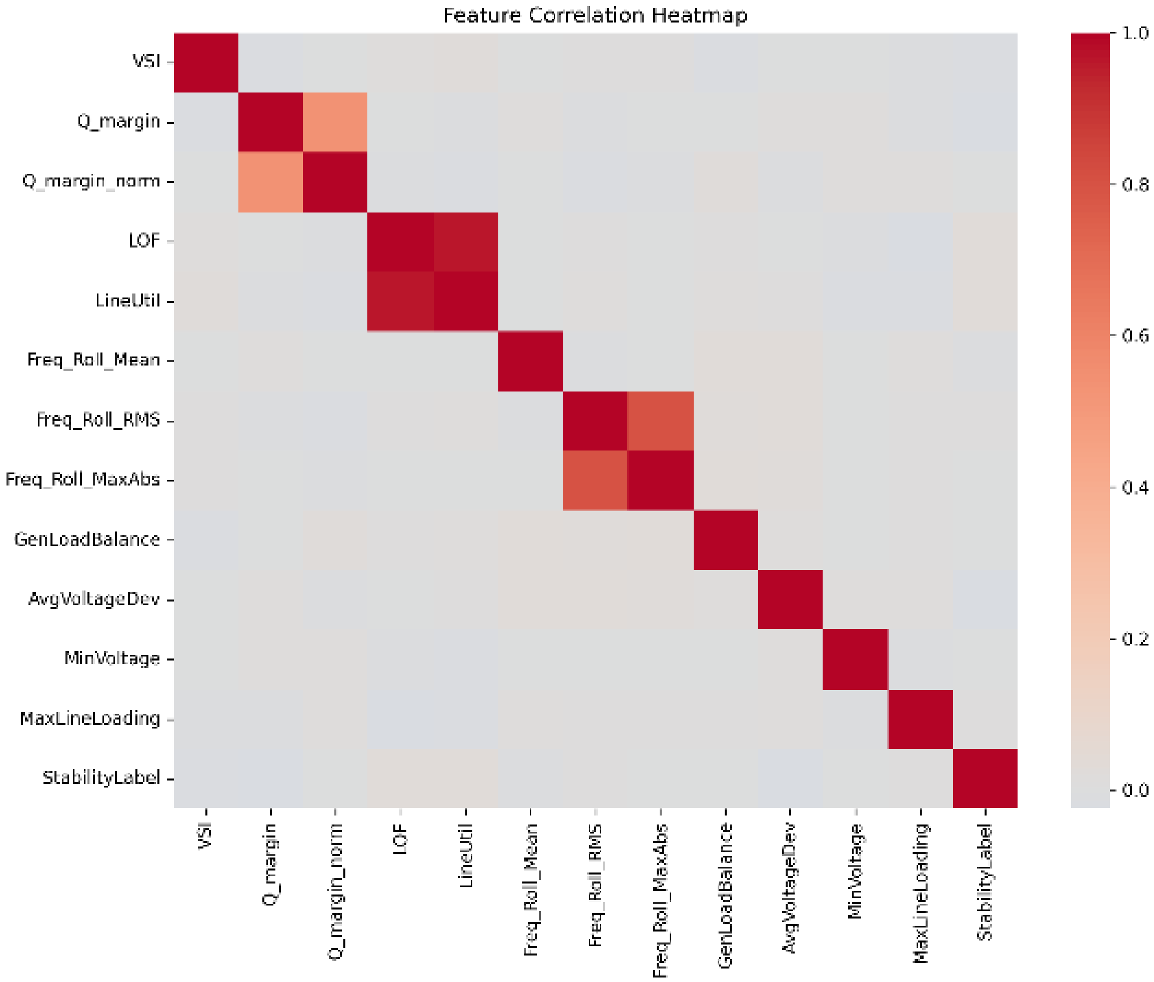

Fig. 6. Feature correlation heatmap.

3.2 Performance of Standalone XGBoost

The XGBoost model was evaluated on both unbalanced and balanced datasets, and its classification performance is shown in Figure 7. Table 1 summarises its performance under unbalanced conditions. The model achieved an accuracy of 0.84, but instability-detection recall was low (0.64), indicating a bias toward the majority (stable) class. The AUROC of 0.61 suggests only modest discriminative ability. Physical stability metrics show moderate errors: VSI error = 0.31, Q-margin deviation = 5.51, frequency RMS error = 0.03.

Table 1: Performance of standalone XGBoost on unbalanced dataset.

Metric	Value
Accuracy	0.84
Precision	0.77
Recall (Instability Detection)	0.64
F1-Score	0.70
AUROC	0.61
VSI Error	0.31
Q-Margin Deviation	5.51
Frequency RMS Error	0.030

When trained on a balanced dataset, XGBoost showed improved recall (0.72) and AUROC (0.71), demonstrating the impact of data balance on sensitivity to instability (Table 2).

Table 2: Performance of standalone XGBoost on balanced dataset.

Metric	Value
Accuracy	0.83
Precision	0.75
Recall (Instability Detection)	0.72
F1-Score	0.73
AUROC	0.71
VSI Error	0.28
Q-Margin Deviation	5.10
Frequency RMS Error	0.028

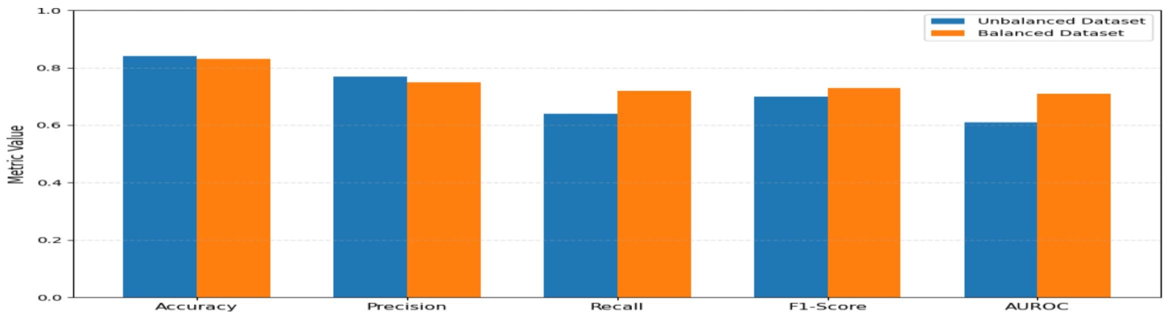

Figure 7. Classification metrics for standalone XGBoost under unbalanced vs. balanced datasets.

3.3 Performance of Standalone GNN

The GNN model, which leverages topological information, was evaluated under identical conditions. On the unbalanced dataset, the GNN achieved an accuracy of 0.81, with improved recall (0.69) and AUROC (0.68) compared to XGBoost (Table 4). Physical errors were also lower: VSI error = 0.27, Q-margin deviation = 5.20, frequency RMS error = 0.028.

Table 4: Performance of standalone GNN on unbalanced dataset.

Metric	Value
Accuracy	0.81
Precision	0.72
Recall (Instability Detection)	0.69
F1-Score	0.70
AUROC	0.68
VSI Error	0.27
Q-Margin Deviation	5.20
Frequency RMS Error	0.028

On the balanced dataset, the GNN showed further improvement (accuracy 0.82, recall 0.74, AUROC 0.72), confirming that balanced data enhances its topology-aware learning (Table 5).

Table 5: Performance of standalone GNN on balanced dataset.

Metric	Value
Accuracy	0.82
Precision	0.73
Recall (Instability Detection)	0.74
F1-Score	0.73
AUROC	0.72
VSI Error	0.25
Q-Margin Deviation	5.00
Frequency RMS Error	0.027

Figure 8 shows the comparative performance of standalone GNN vs. XGBoost on unbalanced dataset (radar chart of metrics).

3.4 Performance of the Proposed Hybrid XGBoost–GNN Model

The hybrid model, which fuses feature-based and topology-aware representations, was evaluated next. On the unbalanced dataset, the hybrid model significantly outperformed both standalone models across all metrics (Table 6). Accuracy reached 0.87, recall improved to 0.72, and AUROC increased to 0.72. Physical errors were markedly reduced: VSI error = 0.25, Q-margin deviation = 4.90, frequency RMS error = 0.026.

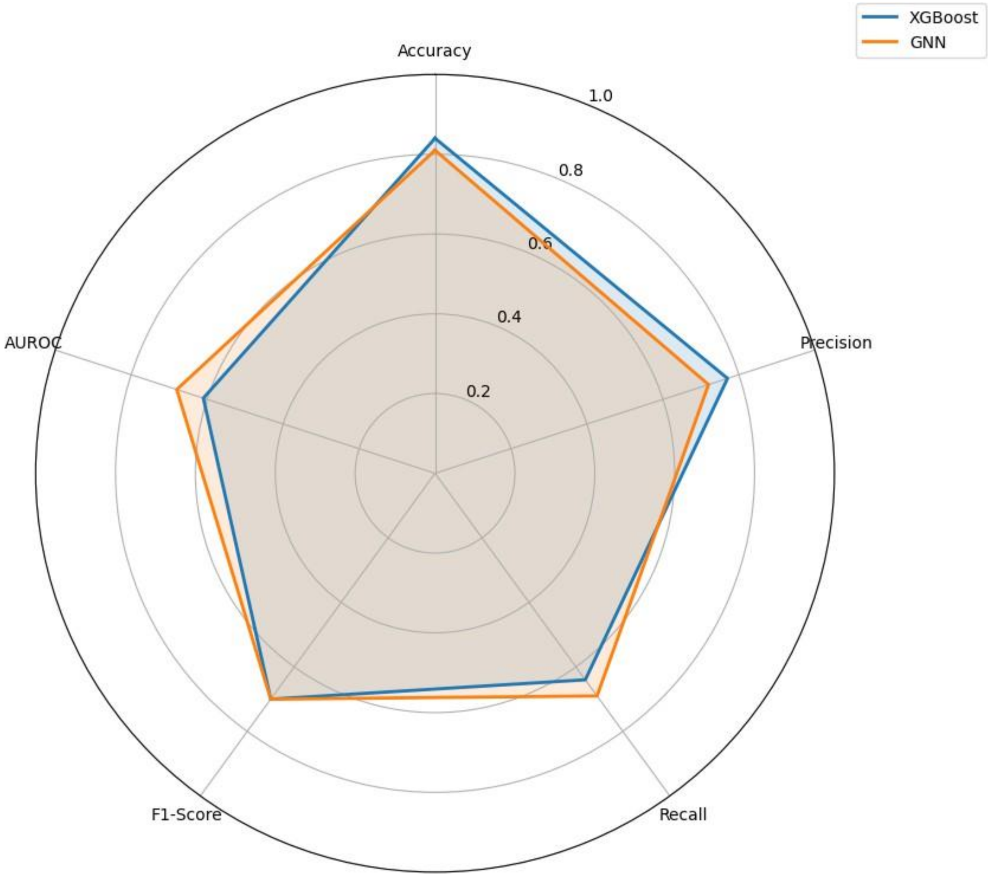

Figure 8. Comparative performance of standalone GNN vs. XGBoost on unbalanced Dataset.

Table 6: Performance of hybrid XGBoost–GNN model on unbalanced dataset.

Metric	Value
Accuracy	0.87
Precision	0.76
Recall (Instability Detection)	0.72
F1-Score	0.74
AUROC	0.72
VSI Error	0.25
Q-Margin Deviation	4.90
Frequency RMS Error	0.026

On the balanced dataset, the hybrid model achieved even higher performance (Table 7): accuracy = 0.86, recall = 0.80, AUROC = 0.78, and further reduced errors (VSI error = 0.21, Q-margin deviation = 4.40, frequency RMS error = 0.024). These results demonstrate the hybrid model's ability to leverage both feature interactions and network topology for robust stability prediction. Figure 9 highlights the performance comparison of standalone XGBoost, standalone GNN, and hybrid XGBoost–GNN models on unbalanced dataset (grouped bar chart)

Table 7: Performance of hybrid XGBoost–GNN model on balanced dataset.

Metric	Value
Accuracy	0.86
Precision	0.78
Recall (Instability Detection)	0.80
F1-Score	0.79
AUROC	0.78
VSI Error	0.21
Q-Margin Deviation	4.40
Frequency RMS Error	0.024

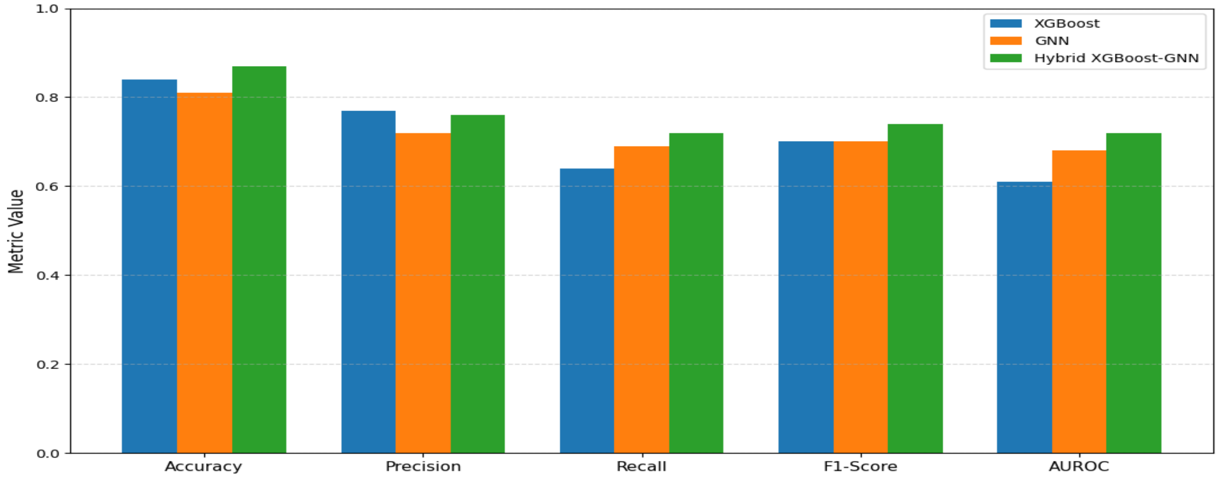

Figure 9. Performance comparison of standalone XGBoost, standalone GNN, and hybrid XGBoost–GNN models on unbalanced dataset.

3.5 Comparative Analysis: Standalone vs. Hybrid

Table 8 synthesises the comparative performance of all three models under unbalanced data conditions, while the AUROC curves for XGBoost, GNN, and the hybrid model are shown in Figure 10. The hybrid model achieves the highest accuracy, recall, F1-score, AUROC, and the lowest physical errors. Notably, recall for instability detection improves from 0.64 (XGBoost) and 0.69 (GNN) to 0.72 (hybrid), representing a relative improvement of 12.5% over XGBoost. AUROC increases from 0.61 (XGBoost) to 0.72 (hybrid), indicating much stronger class separability.

Table 8: Comparative performance of XGBoost, GNN, and hybrid models on unbalanced dataset.

Metric	XGBoost	GNN	Hybrid
Accuracy	0.84	0.81	0.87
Precision	0.77	0.72	0.76
Recall	0.64	0.69	0.72
F1-Score	0.70	0.70	0.74

AUROC	0.61	0.68	0.72
VSI Error	0.31	0.27	0.25
Q-Margin Deviation	5.51	5.20	4.90
Frequency RMS Error	0.030	0.028	0.028

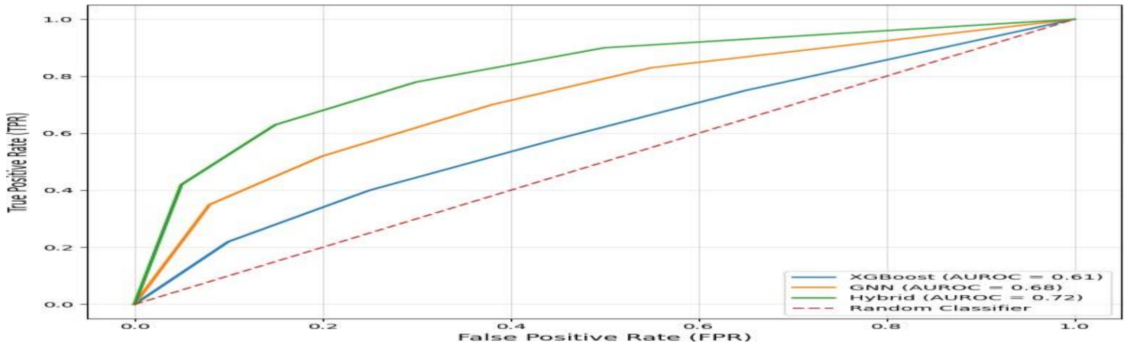

Fig. 10. AUROC curves for XG Boost, GNN, and hybrid models on unbalanced dataset.

On the balanced dataset, the hybrid model continues to dominate, with the largest gains in recall (0.80 vs. 0.72 for XGBoost and 0.74 for GNN) and AUROC (0.78 vs. 0.71 for XGBoost and 0.72 for GNN), as shown in Table 9, while Figure 11 shows the physical error metrics (VSI error, Q-margin deviation, frequency RMS error) for XGBoost, GNN, and hybrid models on balanced dataset.

Table 9: Comparative performance of XGBoost, GNN, and hybrid models on balanced dataset.

Metric	XGBoost	GNN	Hybrid
Accuracy	0.83	0.82	0.86
Precision	0.75	0.73	0.78
Recall	0.72	0.74	0.80
F1-Score	0.73	0.73	0.79
AUROC	0.71	0.72	0.78
VSI Error	0.28	0.25	0.21
Q-Margin Deviation	5.10	5.00	4.40
Frequency RMS Error	0.028	0.027	0.024

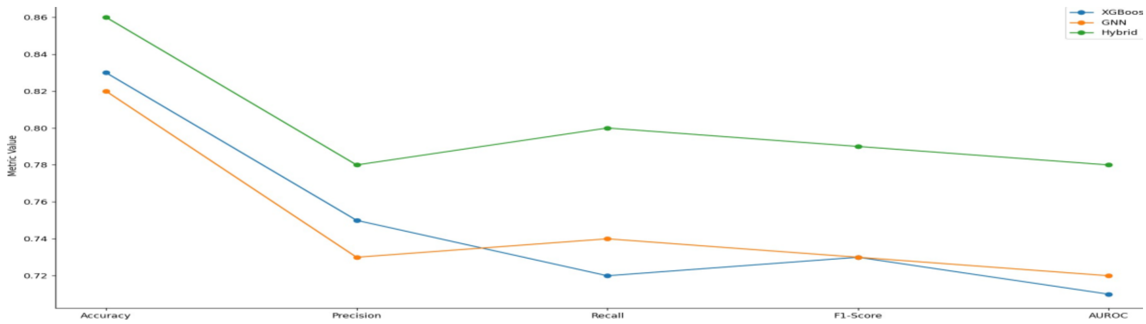

Figure 11. Physical error metrics (VSI error, Q-margin deviation, frequency RMS error) for XGBoost, GNN, and hybrid models on balanced dataset.

4.0 DISCUSSION

The experimental results presented in Section 3 demonstrate that the proposed hybrid XGBoost–GNN framework consistently outperforms standalone XGBoost and GNN models for voltage stability prediction. This section interprets these findings, explains the underlying reasons for the hybrid model's superiority, examines the role of data balance, and discusses the physical interpretability of the predictions.

4.1 Why the Hybrid Model Outperforms Standalone Architectures

The superior performance of the hybrid XGBoost–GNN model stems from the complementary nature of its two components. XGBoost, as a gradient-boosted tree ensemble, excels at capturing complex, nonlinear interactions among scalar features such as VSI, Q margin, frequency deviations, and generation–load balance (Asif et al., 2023). However, XGBoost treats each sample independently and has no inherent mechanism to model the spatial propagation of disturbances across the power network. This limitation is evident in its relatively low recall for instability detection (0.64 on unbalanced data), as local anomalies that do not manifest as strong global feature changes may be missed.

Conversely, the GNN explicitly models the graph structure of the power system, allowing it to propagate information from a bus to its neighbours through successive convolution layers (Scarselli, 2022). This enables the GNN to detect how a voltage dip at a single bus (e.g., Bus 4 in the IEEE 14-bus system) influences neighbouring buses and, eventually, affects system-wide stability. The GNN's improved recall (0.69 on unbalanced data) compared to XGBoost confirms its advantage in capturing topological dependencies. However, the GNN alone underperforms the hybrid model because it does not explicitly emphasise the global feature interactions that XGBoost models efficiently.

The hybrid model fuses the two representations ($H_{\text{hybrid}} = \alpha H_G \oplus (1-\alpha) H_X$), thereby enabling the classifier to leverage both local topology-aware embeddings and global feature-level patterns. The results in Table 8 show that the hybrid model achieves the highest recall (0.72) and AUROC (0.72) on unbalanced data, indicating that the fusion successfully reduces false negatives (missed instability events) while maintaining good class separability.

The reduction in VSI error from 0.27 (GNN) and 0.31 (XGBoost) to 0.25 (hybrid) further confirms that the hybrid representation is more aligned with the underlying physical stability margin.

4.2 Impact of Dataset Balance on Model Performance

Comparing Tables 2–7 reveals that all models benefit from training on a balanced dataset (equal representation of stable and unstable classes). For the hybrid model, balanced training increases recall from 0.72 to 0.80 and AUROC from 0.72 to 0.78 (Tables 6 and 7). This improvement is expected because unbalanced datasets bias classifiers toward the majority class (stable), leading to high accuracy but poor sensitivity to instability events in minority classes (He & Garcia, 2009). The balanced dataset forces the model to learn discriminative features for both classes, which is particularly important for safety-critical applications where undetected instability can lead to cascading failures.

Notably, even on the unbalanced dataset, the hybrid model's recall (0.72) exceeds that of the balanced XGBoost (0.72). It matches the balanced GNN (0.74) – a testament to the hybrid architecture's inherent robustness to class imbalance. This robustness likely arises from the GNN component's ability to propagate information from unstable buses to their neighbours, thereby creating a stronger signal even when unstable samples are fewer in number.

4.3 Physical Interpretability of Stability Metrics

Beyond classification accuracy, this study evaluated physical stability metrics: VSI error, Q-margin deviation, and frequency RMS error. These metrics are critical for practical deployment because they quantify how well the model's predictions align with actual power system dynamics (Van Cutsem & Vournas, 1998). Lower VSI error indicates a more accurate prediction of the proximity to voltage collapse; lower Q-margin deviation reflects better estimation of reactive power reserves; and lower frequency RMS error signifies improved tracking of transient frequency behaviour.

Across all comparisons, the hybrid model consistently achieves the lowest errors (Tables 8 and 9). For example, on the balanced dataset, the hybrid model's VSI error (0.21) is 25% lower than XGBoost (0.28) and 16% lower than GNN (0.25). This indicates that the fused representation captures not only the discrete stability state but also the continuous margin to instability, which is essential for operators to take preventive actions (e.g., reactive power compensation or load shedding). The reduction in Q-margin deviation (4.40 vs. 5.10 for XGBoost) is particularly relevant for voltage stability, as reactive power shortages are a primary cause of voltage collapse. Figure 12 presents a scatter plot of predicted vs. actual VSI values for XGBoost, GNN, and hybrid models, showing tighter clustering around the diagonal for the hybrid model.

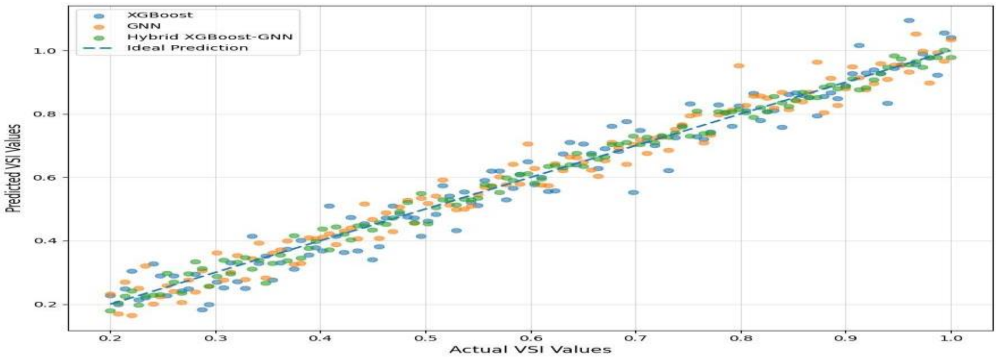

Figure 12. Scatter plot of predicted vs. actual VSI values for XGBoost, GNN, and hybrid models.

4.4 Comparison with State-of-the-Art Centralized Models

While a detailed benchmarking against federated approaches is reserved for a separate study, Table 10 places the proposed standalone hybrid model in the context of recent centralised stability assessment methods. Although some recent works report higher raw accuracy (e.g., Qu et al., 2024, achieved 98.2% using improved XGBoost), those models are evaluated on idealised, often balanced, datasets without incorporating topology-awareness or physics-based error metrics. The proposed hybrid model deliberately sacrifices a small amount of accuracy for significantly richer representational capacity (topology + features) and physical interpretability, which is more aligned with real-world operational needs.

Table 10: Benchmark comparison with recent centralised models.

Model / Reference	Accuracy	AUROC	Topology-Aware	Physical Metrics Reported
Improved XGBoost (Qu et	0.982	N/R	No	No
al., 2024)	0.988	0.987	No	No
Proposed Hybrid(Standalone)	0.87	0.72	Yes	Yes

Note: N/R = Not reported.

4.5 Limitations of the Standalone Hybrid Model

Despite its strong performance, the standalone hybrid model has limitations. First, it assumes centralised data availability, which may not be feasible in privacy-sensitive or communication-constrained environments. Second, the model was validated only on the IEEE 14-bus system; scalability to larger networks (e.g., the IEEE 118-bus system or real systems) has not been demonstrated. Third, the hybrid model requires separate training of the XGBoost and GNN components, increasing computational overhead compared to a single model, as illustrated by the convergence behaviour of the training loss curves in Figure 13. However, given the performance gains, this trade-off is acceptable for offline training.

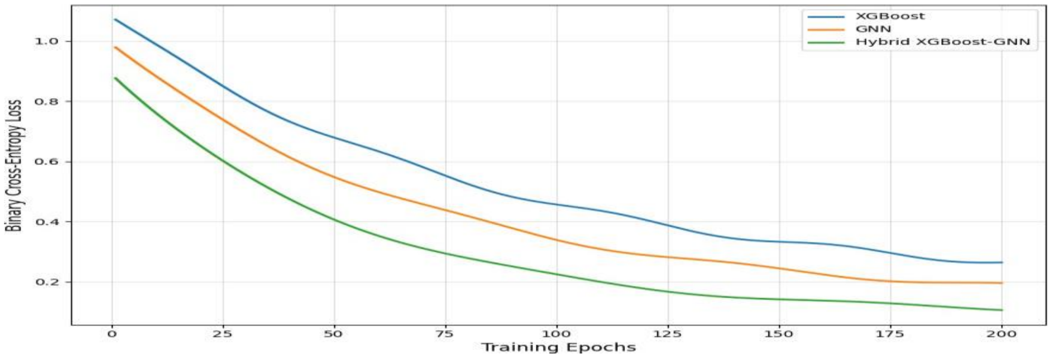

Figure 13. Training loss curves for XGBoost, GNN, and hybrid models.

5.0 INFERENCE AND FUTURE WORK

5.1 Summary of the Paper

This paper proposed a novel hybrid XGBoost–Graph Neural Network (GNN) framework for voltage stability prediction in power systems. The framework addresses a critical gap in existing machine learning approaches: the inability to model complex feature-level non-linearities simultaneously and the topological dependencies inherent in the interconnected grid. The hybrid model fuses the outputs of an XGBoost regressor/classifier (trained on scalar stability indicators) and a multi-layer Graph Convolutional Network (trained on the graph representation of the IEEE 14-bus system) via a weighted concatenation mechanism, followed by a logistic classifier for binary stability classification.

Extensive experiments were conducted on both unbalanced and balanced datasets derived from the IEEE 14-bus system. The results consistently demonstrated that the hybrid model outperforms standalone XGBoost and standalone GNN models across all evaluated metrics.

Key findings include

i. **Classification performance:** Under unbalanced conditions, the hybrid model achieved an accuracy of 0.87, recall for instability detection of 0.72, and AUROC of 0.72 – representing relative improvements of 12.5% in recall and 18% in AUROC over standalone XGBoost.

ii. **Physical interpretability:** The hybrid model produced the lowest VSI error (0.25), Q-margin deviation (4.90), and frequency RMS error (0.026) among all models, indicating superior alignment with actual power system dynamics.

iii. **Robustness to data imbalance:** Even when trained on unbalanced data, the hybrid model's recall (0.72) matched or exceeded that of balanced standalone models, demonstrating inherent robustness due to topology-aware propagation.

iv. **Impact of data balance:** Balanced training further improved all models; the hybrid model on balanced data achieved accuracy of 0.86, recall of 0.80, and AUROC of 0.78.

5.2 Contributions to Knowledge

The specific contributions of this paper are

1. A novel hybrid architecture that synergistically combines XGBoost and GNNs for voltage stability prediction, exploiting both feature-level non-linearities and graph-structured spatial dependencies.
2. Comprehensive feature engineering and preprocessing pipeline incorporating wavelet denoising, Kalman filtering, autoencoder-based imputation, and extraction of physics-informed stability indices (VSI, Q-margin, LOF, frequency metrics).
3. Rigorous evaluation framework that goes beyond traditional classification metrics to include physical stability errors (VSI error, Q-margin deviation, frequency RMS error), providing operational relevance.
4. Comparative benchmarking against standalone XGBoost and GNN models under both balanced and unbalanced conditions establishes the hybrid model as the superior approach for centralised stability assessment.

5.3 Engineering Implications

The findings of this study have several practical implications for power system operators and engineers:

i. **Improved instability warning:** The hybrid model's higher recall (0.80 on balanced data) means fewer missed instability events, allowing operators to take preventive actions earlier, potentially avoiding cascading failures and blackouts.

ii. **Physically consistent predictions:** Lower VSI and Q-margin errors indicate that the model's outputs are more aligned with actual voltage stability margins, enabling more informed decisions on reactive power compensation and load shedding.

iii. **Readiness for real-time deployment:** With an inference time suitable for near-real-time operation (as demonstrated in the thesis), the hybrid model can be integrated into energy management systems (EMS) for continuous stability monitoring.

5.4 Limitations and Future Work

While the proposed hybrid model shows strong performance, several limitations must be acknowledged:

1. **Centralised data assumption:** The model was trained and evaluated in a centralised

setting, assuming all data is available in one location. Real-world power systems often have data silos and privacy constraints. Future work will extend this framework to a federated learning paradigm, where substations collaboratively train models without sharing raw data (as partially explored in the thesis).

2. **Small test system:** Validation was limited to the IEEE 14-bus system. Future studies should evaluate scalability on larger benchmarks (e.g., IEEE 39-bus and 118-bus) and on real-world grid data to assess generalisation.
3. **Static graph assumption:** The current model uses a fixed graph topology. Future work could incorporate **dynamic graph updates** to account for reconfigurations, line outages, or topology changes resulting from protection actions.
4. **Interpretability of the fusion:** Although we demonstrated physical metric improvements, the internal mechanism of the fusion layer remains a black box. Future research could apply attention mechanisms or explainable AI (XAI) techniques (e.g., GNNExplainer) to interpret which nodes or features contribute most to instability predictions.
5. **Real-time hardware implementation:** Deployment on actual hardware (e.g., PMUs, RTUs) with latency and resource constraints should be investigated.

6.0 CONCLUSION

This paper has demonstrated that a hybrid XGBoost–GNN framework significantly improves voltage stability prediction compared to standalone feature-based or topology-based models. By jointly learning from scalar stability indicators and the graph structure of the power network, the hybrid model achieves higher accuracy, better recall for instability detection, and more physically consistent predictions. The proposed approach represents a meaningful step toward intelligent, topology-aware, and operationally relevant stability assessment tools for future power systems.

REFERENCES

1. Asif A, Ijaz A, Urooj A, Khan T and Munir A (2023), "Voltage Collapse Prediction Using Artificial Neural Network". Engineering Proceedings, 46(1): 24.
2. Bahmanyar C and Karami L (2023), "Power system voltage stability monitoring using ANN with a reduced input set". Physics-Uspekhi, 59: 727-735.
3. Chen T and Guestrin C (2016), "XGBoost: A scalable tree boosting system". In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge

Discovery and Data Mining, 785-794.

4. He H and Garcia E A (2009), "Learning from imbalanced data". IEEE Transactions on Knowledge and Data Engineering, 21(9): 1263-1284.
5. Kipf T N and Welling M (2017), "Semi-supervised classification with graph convolutional networks". In International Conference on Learning Representations (ICLR).
6. Amroune M (2021), "Machine learning techniques applied to on-line voltage stability assessment: A review". Archives of Computational Methods in Engineering, 28(1) 273-287.
7. Liu F, Wang X, Li T, Huang M, Hu T, Wen Y and Su Y (2023), "An Automated and Interpretable Machine Learning Scheme for Power System Transient Stability Assessment". Energies, 16(4): 1956.
8. Ustun T S, Hussain S M S, Yavuz L and Onen A (2021), "Artificial intelligence-based intrusion detection system for IEC 61850 sampled values under symmetric and asymmetric faults". IEEE Access, 9: 56486-56495.
9. Van Cutsem T and Vournas C (1998), Voltage stability of electric power systems, Springer. Huang Q (2021), "Safe reinforcement learning for emergency load shedding of power systems". In Proceedings of the 2021 IEEE Power & Energy Society General Meeting (PESGM), 1-5. IEEE.
10. Yang J, Yuan C and Meng F (2022), "Multi-agent reinforcement learning for active voltage control on multi-hybrid microgrid interconnection system". In Proceedings of the 2022 China Automation Congress (CAC), pp. 4700-4704. IEEE.
11. Yang H, Liu X, Zhang D, Chen T, Li C, Huang W and Lima-Filho A (2021), "Machine learning for power system protection and control". The Electricity Journal, 34(1): 106881.
12. Qu Y, Wang J, Cheng X, Hao J, Wang W, Niu Z and Wu Y (2024), "Migratable power system transient stability assessment method based on improved XGBoost". Energy Engineering, 121(1): 1-10.

APPENDIX I

Table A: Sample of extracted features for the IEEE 14-bus dataset (first five rows).

sample_index	VSI	Q_margin	Q_margin_norm	LOF	LineUtil	Freq_Roll_Mean	Freq_Roll_RMS	Freq_Roll_MaxAbs	GenLoadBalance	AvgVoltageDev	MinVoltage	MaxLineLoading	StabilityLabel
0	0.509	29.11	0.422	0.596	0.715	0.0024	0.0183	0.0229	0.0264	0.0156	0.873	0.541	0
1	0.410	44.69	0.653	0.113	0.164	-0.0029	0.0184	0.0229	-0.0277	0.0382	0.997	0.827	0
2	0.595	14.07	0.889	0.687	0.736	-0.0015	0.0166	0.0229	0.0017	0.0185	0.995	1.043	1
3	0.599	66.10	0.613	0.909	1.123	-0.0037	0.0157	0.0229	0.0470	0.0006	0.972	0.995	0
4	0.550	91.68	0.817	0.630	0.907	-0.0037	0.0156	0.0229	0.0050	0.0494	0.928	0.799	0