Journal of Renewable Energy and Sustainable Development (RESD) Volume 12, Issue 2, December 2026, e-ISSN 2356-8569
http://dx.doi.org/10.21622/RESD.2026.12.2.2146

Received: 11 May 2026 Accepted: 02 September 2026 Published: 14 September 2026

# Robust Ensemble Learning for Smart Grid Stability Prediction Under Operating Variability

**P. Sirish Kumar [1*], K. Kavitha [2], Ch. Rajasekhar [3], M.D.R. Siva Santhosh [4], M.S.R. Naidu [5], and Sai Kiran Oruganti [6]**

[1*,4,5] Aditya Institute of Technology and Management, K.Kothuru, Tekkali, Srikakulam, Andhra Pradesh, India

[2] GMR Institute of Technology (GMRIT) - Deemed to be University, Rajam, Andhra Pradesh, India

[3] Aditya University, Surampalem, Andhra Pradesh, India

[1*,6] Lincoln University College, Petaling Jaya, Selangor Darul Ehsan, Malaysia

sirishdg@gmail.com, kavitha.akki1514@gmail.com, chukkarajasekhar@gmail.com, shivadasarad@gmail.com, naidu.msr@gmail.com, saisharma@lincoln.edu.my

## Abstract

*The rapid integration of renewable energy sources has made modern smart grids more dynamic and sensitive to fluctuations in generation and demand. Under such conditions, maintaining grid stability requires prediction models that remain reliable not only during normal operation but also under increased variability. This study presents a comparative perturbation-robustness assessment of four advanced ensemble learning models for smart grid stability prediction: Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), a Hybrid LightGBM–XGBoost model (Hybrid-LXGB), and an Out-of-Fold Stacking ensemble (OOF-Stacking). To represent perturbation-driven operating uncertainty, Gaussian perturbation-based variability was introduced at multiple severity levels. Model performance was evaluated using predictive accuracy, relative degradation, error amplification behavior, and sensitivity analysis. The results indicate that all models perform strongly under standard conditions; however, clear differences emerge as variability increases. XGBoost delivers marginally higher overall predictive performance, while LightGBM demonstrates greater resistance to performance degradation under more challenging scenarios. The Hybrid-LXGB and OOF-Stacking models also achieve competitive and stable results, confirming the effectiveness of ensemble learning approaches for smart grid stability prediction applications. Further analysis reveals that severe variability is the primary source of performance loss, leading to accuracy reductions exceeding 22% across the evaluated models. This observation highlights the importance of assessing prediction systems beyond clean benchmark settings. The findings suggest that model selection for smart grid stability prediction applications should consider perturbation robustness alongside conventional accuracy metrics. Overall, this study provides a variability-oriented evaluation framework for analyzing ensemble learning behavior in smart grid stability prediction applications.*

**Index-words:** *Ensemble Learning, LightGBM, Perturbation Robustness, Smart Grid Stability, Variability Analysis, XGBoost.*

* Corresponding author



Copyright © 2026, authors

http://dx.doi.org/10.21622/RESD.2026.12.2.2146

## Nomenclature

| Symbol / Abbreviation | Description |
|---|---|
| Acc | Classification accuracy |
| EAF | Error Amplification Factor |
| $E_{base}$ | Baseline prediction error |
| $E_{yar}$ | Prediction error under disturbed conditions |
| Hybrid-LXGB | Hybrid LightGBM–XGBoost Model |
| LightGBM | Light Gradient Boosting Machine |
| OOF | Out-of-Fold |
| OOF-Stacking | Out-of-Fold Stacking ensemble |
| $P_{Hybrid}$ | Hybrid prediction probability |
| $P_{LGBM}$ | Predicted probability obtained from LightGBM |
| $P_{LGBM}^{OFF}$ | Out-of-fold prediction probability generated by LightGBM |
| $P_{XGB}$ | Predicted probability obtained from XGBoost |
| $P_{XGB}^{OFF}$ | Out-of-fold prediction probability generated by XGBoost |
| $\alpha$ | Fusion weight used in the Hybrid-LXGB model |
| $\epsilon$ | Gaussian random noise |
| $\sigma$ | Variability intensity (standard deviation of Gaussian perturbation) |
| X | Original input feature vector |
| X′ | Perturbed input feature vector |
| XGBoost | Extreme Gradient Boosting |
| Z | Meta-feature matrix |

## I. Introduction

The global shift toward renewable energy has significantly changed the operating behavior of modern power systems. Large-scale integration of solar and wind resources introduces natural variability because these sources depend on weather conditions and intermittent generation [1], [2]. At the same time, rising electricity demand, expanding distributed generation, and bidirectional power flow have added further complexity to grid operation [3,4]. Together, these factors make it more difficult to maintain stable and secure system performance using conventional monitoring and control methods. As a result, accurate and timely grid stability prediction has become an essential requirement in modern smart grids [5].

In recent years, machine learning has become an effective tool for solving power system prediction problems because it can learn complex nonlinear relationships directly from operational data [6]. Methods such as Random Forest, Support Vector Machine, Artificial Neural Networks, and gradient boosting models have been widely used for load forecasting, fault diagnosis, energy management, and stability assessment [7,8]. Among these methods, Light Gradient Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost) have received strong attention due to their high accuracy, fast computation, and strong performance on structured tabular data [9-11].

Despite this progress, many existing studies evaluate models only under static or ideal operating conditions, which may not reflect real grid environments [12]. In practical renewable-powered systems, measurements can be affected by noise, communication delays, sensor disturbances, and sudden operating changes [13,14]. Under such conditions, models trained on clean datasets may lose prediction quality. This shows that robustness should be considered alongside accuracy, especially in safety-critical applications where unreliable predictions may lead to poor decisions [15]. In renewable-powered smart grids, these disturbances arise from several practical sources, including intermittent renewable generation, sensor measurement noise, communication uncertainties, and rapid load variations. Therefore, evaluating model performance under disturbed operating conditions is essential for reliable real-world deployment.

## II. Related work

To address these challenges, recent studies have explored hybrid and ensemble learning methods that combine multiple models to improve performance and reduce sensitivity to individual model weaknesses [16-20]. Stacking-based approaches, in particular, have shown promise in improving generalization through the integration of diverse learners [21,22]. However, limited attention has been given to how these advanced models behave under progressively increasing variability conditions [23]. It is still unclear whether models with higher baseline accuracy always remain robust, or whether simpler models may perform more consistently under disturbance.

Table 1 summarizes recent studies, highlighting their main methods, reported outcomes, and limitations in terms of robustness evaluation.

Table 1: Summary of recent works on grid stability prediction.

| Reference | Method / Model | Key Focus | Key Result | Limitation |
|---|---|---|---|---|
| [1], [2], [3] | Reviews | Renewable integration and AI in grids | Summarized recent advances and challenges in renewable-powered smart grids | No experimental validation |
| [6], [7] | ML Models | Stability prediction | Demonstrated the effectiveness of machine learning for grid stability assessment | Limited robustness analysis |
| [8], [13] | LightGBM / XGBoost | High prediction accuracy | Reported high predictive accuracy using gradient boosting models | No variability testing |
| [10], [11] | Data-driven ML | Stability under uncertainty | Improved prediction performance under uncertain operating conditions | Static evaluation settings |
| [14], [16] | Ensemble Learning | Improved prediction | Enhanced predictive performance through ensemble learning | No degradation analysis |
| [17], [18] | Stacking Models | Better generalization | Achieved improved generalization using stacking strategies | Limited sensitivity study |
| [20] | Hybrid Boosting | Performance improvement | Improved predictive performance using hybrid boosting models | No robustness comparison |

As shown in Table 1, most previous studies primarily focus on improving prediction accuracy under nominal operating conditions. Only a limited number of studies investigate how prediction performance changes when operating uncertainty increases. Furthermore, direct comparisons between standalone, hybrid, and stacking ensemble models under the same variability conditions remain limited. In particular, the relationship between prediction accuracy and robustness remains insufficiently explored.

Motivated by these limitations, this study presents a comparative perturbation-robustness evaluation of four ensemble learning models: LightGBM, XGBoost, a Hybrid-LXGB model, and an Out-of-Fold Stacking (OOF-Stacking) framework. Instead of proposing a new machine learning algorithm, this work introduces a structured robustness evaluation framework that compares different ensemble strategies under identical perturbation-driven operating conditions. Gaussian perturbation-based variability is progressively introduced into the test data to simulate increasing operating uncertainty, allowing both predictive accuracy and robustness degradation to be systematically analyzed. This provides additional insight into model behavior beyond conventional benchmark evaluations and helps identify models that achieve a better balance between predictive performance and operational robustness.

The main contributions of this work are summarized as follows:

- A comparative benchmark of four advanced ensemble learning models, namely LightGBM, XGBoost, Hybrid-LXGB, and OOF-Stacking, for renewable-powered smart grid stability prediction.
- A perturbation-based robustness evaluation framework that assesses model performance under progressively increasing variability rather than relying only on nominal operating conditions.
- A comprehensive analysis of predictive accuracy, relative degradation, error growth, and sensitivity to variability to understand the trade-off between prediction accuracy and operational robustness.
- Practical guidance for selecting ensemble learning models based on both predictive performance and robustness for renewable-integrated smart grid applications.

Overall, the findings support the development of more reliable artificial intelligence tools for sustainable and secure power system operation under uncertain operating environments.

The remainder of this paper is organized as follows. Section 2 presents the dataset, preprocessing steps, proposed methodology, and experimental framework. Section 3 discusses the experimental results and robustness analysis under different variability levels. Finally, Section 4 concludes the paper and outlines directions for future research.

# III. Methodology

## A. Dataset description & preprocessing

This study uses the publicly available Smart Grid Stability dataset, which is widely adopted for evaluating machine learning models in power system applications [24,25]. The dataset captures different operating states of a renewable-integrated smart grid under varying conditions, making it suitable for stability prediction.

The dataset contains 60,000 samples with 12 numerical input features and one binary target label (stabf). The input variables are grouped into three categories: time-related parameters (tau1–tau4), power-related variables (p1–p4), and control gain parameters (g1–g4). These features represent key dynamic characteristics that influence grid stability, as shown in Figure 1. The target label indicates whether the operating condition is stable or unstable, resulting in a binary classification task. The dataset consists of 21,720 stable samples (36.2%) and 38,280 unstable samples (63.8%), indicating a moderately imbalanced class distribution. To reduce the influence of this imbalance during model evaluation, a stratified train–test split was adopted to preserve the original class proportions in both subsets. This ensures a fair and consistent comparison of all evaluated models. Since all variables are numerical, no categorical encoding was required. The dataset was selected because it combines realistic operating variability, balanced class structure, and public accessibility, making it a suitable benchmark for fair and reproducible evaluation of ensemble learning models despite its moderately imbalanced class distribution.

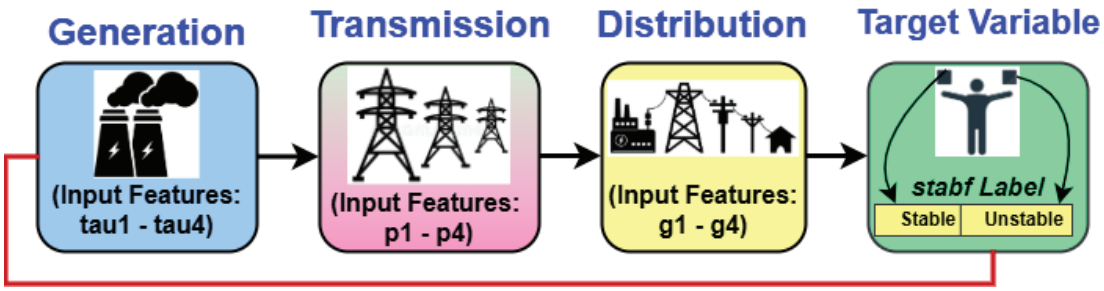



Figure 1: Structure of input feature categories and grid stability target variables.

The dataset was first examined for consistency and usability. All input features are numerical, so no categorical transformation was needed. The data is already well structured, which simplifies preprocessing. The target variable (stabf) was converted into binary numerical form for classification. The dataset was also checked for missing values and duplicate records. No major issues were found, so all samples were retained. This helps preserve the original data distribution.

Tree-based models such as LightGBM and XGBoost do not require feature scaling. Therefore, the original feature values were used without normalization. This also preserves their physical meaning [26,27]. Finally, the dataset was separated into input features and target labels to prepare it for training, testing, and robustness evaluation.

## B. Training and testing strategy

To ensure a fair evaluation, the dataset was divided into training and testing subsets using a stratified split. This preserves the proportion of stable and unstable samples in both sets. In this study, 80% of the data was used for training, while 20% was reserved for testing. The training set was used to train the models. The test set was used only for the final evaluation. A fixed random seed was applied during splitting to ensure reproducibility. For the stacking framework, an out-of-fold strategy was used within the training set. Predictions were generated using unseen folds, and these predictions were used to train the meta-model. This prevents information leakage and improves reliability [28,29]. The complete experimental framework adopted in this study is illustrated in Figure 2.

http://dx.doi.org/10.21622/RESD.2026.12.2.2146

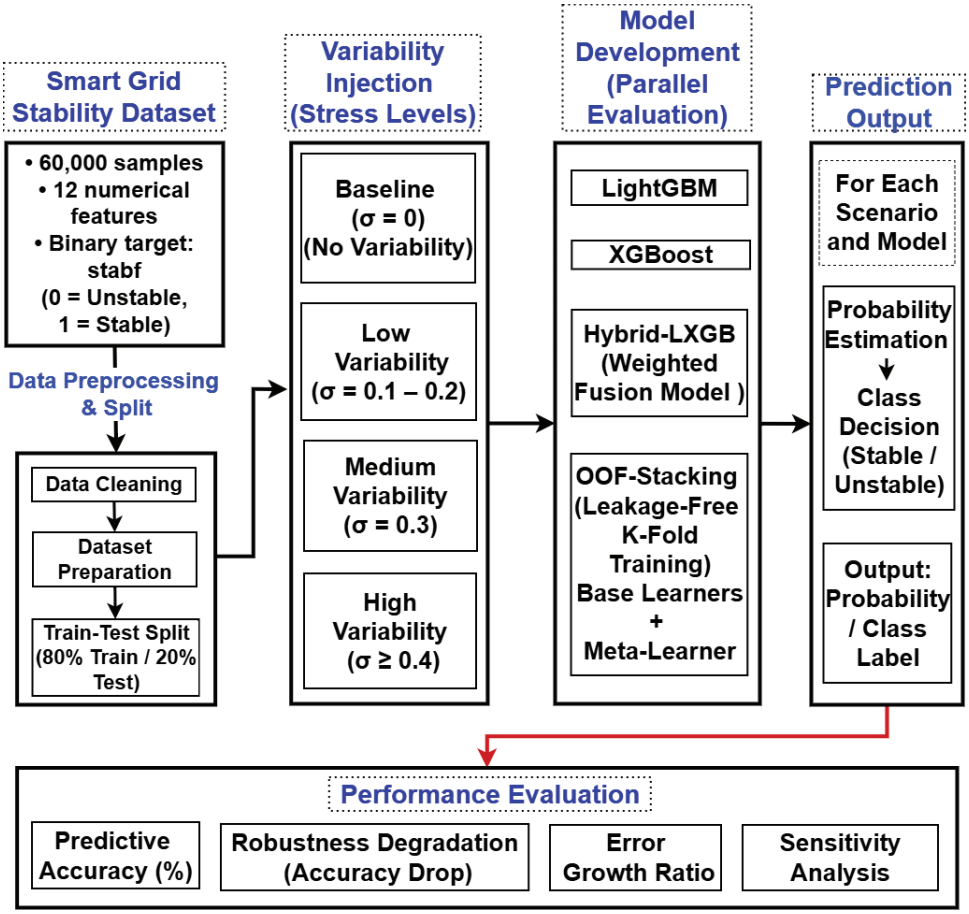



Figure 2: Overall comparative framework for robustness-oriented smart grid stability prediction.

Figure 2 presents the overall experimental framework adopted in this study. The process begins with dataset preprocessing and train–test splitting, followed by the generation of different variability levels to simulate uncertain operating conditions. The prepared data are then evaluated using four ensemble learning models, namely LightGBM, XGBoost, Hybrid-LXGB, and OOF-Stacking. Finally, the prediction results are assessed using predictive accuracy, robustness degradation, error growth ratio, and sensitivity analysis to compare the robustness of all models under increasing variability.

## C. *Baseline models*

Two widely used gradient boosting models were selected as baseline methods: LightGBM and XGBoost. Both are well-suited for structured tabular data and can capture nonlinear relationships effectively [26,27]. LightGBM is a tree-based boosting method that uses leaf-wise growth. This often leads to faster training and strong predictive performance. XGBoost is an optimized boosting algorithm that includes regularization and efficient tree construction. This helps reduce overfitting while maintaining high accuracy. Both models were trained using identical data splits and evaluation settings. Their results serve as reference points for comparison with the proposed ensemble methods.

### 1. Hybrid-LXGB model

The Hybrid-LXGB model combines LightGBM and XGBoost at the probability level. Although both are boosting models, they follow different learning strategies. As a result, they may capture different patterns in the data. Combining their outputs helps improve prediction stability [28]. Both models were first trained independently using the same training data. For each sample, probability estimates were obtained from both models. The hybrid probability is computed as given in Eq. (1).

$$P_{\text{Hybrid}} = \alpha P_{\text{LGBM}} + (1-\alpha)P_{\text{XGB}}, \alpha \in [0,1] \quad (1)$$

Where $P_{Hybrid}$ denotes the fused prediction probability, $P_{LGBM}$ and $P_{XGB}$ denote the predicted probabilities obtained from LightGBM and XGBoost, respectively, and $\alpha$ is the fusion weight ($0 \le \alpha \le 1$).

The internal structure of the proposed Hybrid-LXGB framework is shown in Figure 3, which visualizes the weighted fusion process.

In this study, the fusion weight was fixed at $\alpha = 0.80$ to assign slightly higher importance to the LightGBM probability while retaining the contribution of XGBoost. The same fusion weight was used for all experiments to ensure a consistent comparison across different variability levels.

The final class label is obtained using a threshold of 0.5 as shown in Eq. (2).

$$\hat{y} = \begin{cases} 1, & P_{\text{Hybrid}} \geq 0.5 \\ 0, & P_{\text{Hybrid}} < 0.5 \end{cases} \quad (2)$$

where $\hat{y}$ denotes the predicted class label, this approach helps reduce individual model errors and provides a balance between accuracy and robustness.

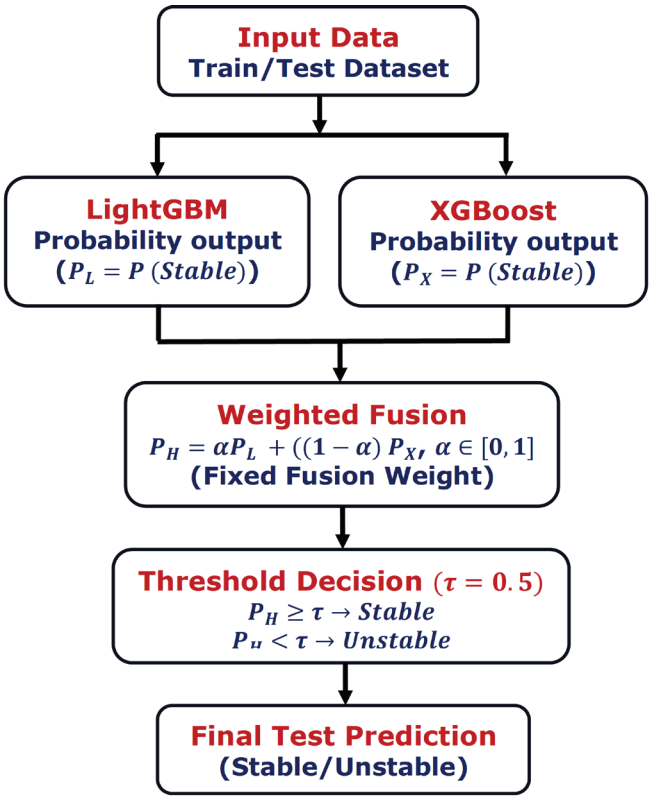



Figure 3: Hybrid-LXGB architecture using weighted fusion of LightGBM and XGBoost outputs.

## 2. OOF-Stacking model

The OOF-Stacking Model combines multiple learners using a two-stage framework. LightGBM and XGBoost are used as base models, while a meta-model produces the final prediction [29]. An Out-of-Fold (OOF) strategy was used during training. The training data was divided into $K$ folds [30]. In each iteration, $K$-1 folds were used for training, and the remaining fold was used for validation. This process continued until every fold had been used once for validation. Out-of-fold predictions were generated for all training samples using unseen data. These predictions were then combined to form the meta-feature matrix as given in Eq. (3).

$$Z = \left[P_{\text{LGBM}}^{\text{OOF}}, P_{\text{XGB}}^{\text{OOF}}\right] \quad (3)$$

Where $Z$ represents the meta-feature matrix, and $P_{LGBM}^{OFF}$ and $P_{XGB}^{OFF}$ denote the out-of-fold prediction probabilities generated by LightGBM and XGBoost, respectively. The final prediction is given in Eq. (4).

$$\hat{y} = f_{\text{meta}}(Z) \quad (4)$$

where $f_{meta}(\cdot)$ denotes the logistic regression meta-model, and $\hat{y}$ represents the final predicted class label.

During testing, the base models were retrained on the full training set. Their predictions were passed

http://dx.doi.org/10.21622/RESD.2026.12.2.2146

to the meta-model to obtain the final output. This framework improves generalization and reduces overfitting.

The two-level leakage-free stacking framework used in this work is illustrated in Figure 4, which summarizes the generation of OOF meta-features and final meta-learning.

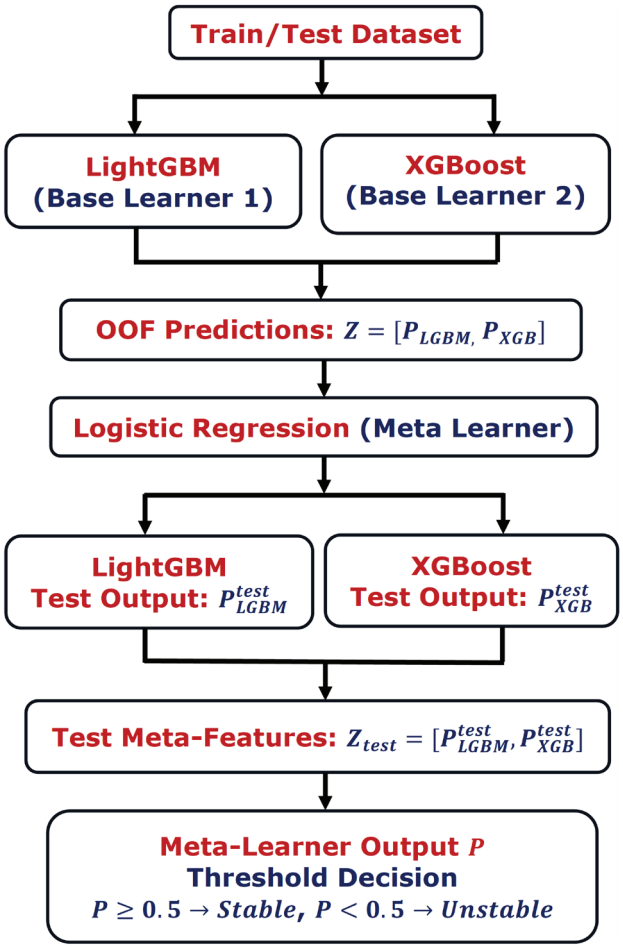



Figure 4: OOF-Stacking architecture using LightGBM and XGBoost base learners with logistic regression meta-learning.

## D. Hyperparameter settings

A manual validation-based tuning strategy was adopted, where multiple parameter combinations were experimentally evaluated on the validation set, and the configuration providing the best trade-off between accuracy and stability was selected [28]. For both LightGBM and XGBoost, the number of estimators was set to 500, the learning rate to 0.05, and the maximum depth to 6. Early stopping with a patience of 20 rounds was additionally applied during training to reduce overfitting and prevent unnecessary boosting iterations. Using the same settings ensures a fair comparison [26,27]. For the Hybrid-LXGB model, the fusion weight was fixed at $\alpha$ = 0.80. For the OOF-Stacking framework, 5-fold cross-validation was used to generate out-of-fold predictions. A logistic regression classifier was used as the meta-model to combine the out-of-fold predictions generated by the base learners. All hyperparameters were kept fixed across all variability scenarios. This ensures that performance changes are due to input disturbances rather than parameter tuning.

## E. Variability injection framework

To evaluate perturbation robustness under uncertain conditions, Gaussian perturbation-based variability was applied to the testing data after model training. This preserves a clean training process while allowing performance to be tested under disturbance [32]. For an input vector $X$, the disturbed sample $X'$ is defined as represented in Eq. (5).

$$X' = X \odot (1 + \epsilon), \epsilon \sim \mathcal{N}(0, \sigma^2) \quad (5)$$

http://dx.doi.org/10.21622/RESD.2026.12.2.2146

where $X$ represents the original input feature vector, $X'$ denotes the perturbed input vector, $\odot$ represents element-wise multiplication, $\epsilon$ is Gaussian random noise, and $\sigma$ controls the variability intensity.

Three levels of variability were considered:

- Very Low variability: $\sigma$ = 0.05
- Medium variability: $\sigma$ = 0.15
- High variability: $\sigma$ = 0.40

Continuous values of $\sigma$ were additionally evaluated to observe gradual performance changes and construct robustness curves. For continuous robustness analysis, the variability intensity $\sigma$ was gradually increased from 0 to 0.40 using uniform incremental steps, and model accuracy was recorded at each disturbance level. To ensure fair comparison, identical random seeds and variability configurations were consistently maintained across all model evaluations. This perturbation-based evaluation strategy is motivated by the increasing uncertainty and instability challenges observed in renewable-integrated smart grids [33-38].

The adopted Gaussian perturbation model provides a simplified representation of uncertainty and was selected to enable a consistent comparison of model robustness under controlled conditions. In practical smart grid environments, different feature groups, such as time-related parameters, power-related variables, and control gain parameters, may experience different types and levels of uncertainty. Similarly, communication delays, sensor faults, missing measurements, and non-Gaussian disturbances may affect system operation in different ways. Therefore, the present perturbation framework should be regarded as a general robustness evaluation approach rather than an exact representation of all real-world operating conditions. Future work will investigate feature-specific disturbance models and more realistic uncertainty sources using operational smart grid data.

### F. Performance assessment

Model performance was evaluated using classification accuracy under both baseline and disturbed conditions. To assess robustness, the Error Amplification Factor (EAF), relative accuracy drop, error growth ratio, and sensitivity to variability were also evaluated as defined in Eq. (6). The Error Amplification Factor (EAF) is defined as

$$\mathrm{EAF} = \frac{E_{var}}{E_{base}} \tag{6}$$

where $E_{var}$ represents the prediction error under disturbed conditions, and $E_{base}$ represents the baseline prediction error obtained under clean operating conditions.

The relative accuracy drop is computed as represented in Eq. (7).

$$\text{Accuracy Drop (\%)} = \frac{Acc_{base} - Acc_{var}}{Acc_{base}} \times 100 \tag{7}$$

where $Acc_{base}$ and $Acc_{var}$ denote the classification accuracies under baseline and disturbed conditions, respectively.

The error growth ratio is calculated as shown in Eq. (8).

$$\text{Error Growth Ratio} = \frac{1 - Acc_{var}}{1 - Acc_{base}} \tag{8}$$

Sensitivity to variability is estimated using the normalized performance degradation at each variability level as given in Eq. (9).

$$\text{Impact Weight} = \frac{\mathrm{Drop_i}}{\mathrm{Drop_{max}}} \tag{9}$$

where $\mathrm{Drop_i}$ is the relative accuracy drop at the $i^{th}$ variability level, and $\mathrm{Drop_{max}}$ is the maximum observed accuracy drop among all evaluated variability levels.

These metrics provide a comprehensive assessment of how model performance changes under increasing variability [30,31]. All models were evaluated using the same training–testing partitions, identical preprocessing steps, and consistent variability scenarios to ensure a fair comparison. Hyperparameter settings were fixed after tuning, and all reported results were obtained under the same experimental framework.

## IV. Results & Discussion

The proposed models were evaluated under both baseline and disturbed operating conditions to assess predictive performance and perturbation robustness. Four models were considered: LightGBM, XGBoost, Hybrid-LXGB, and OOF-Stacking. The analysis focuses on how predictive quality changes as uncertainty increases, since high baseline accuracy alone may not ensure reliable operation in renewable-powered smart grids.

### *A. Predictive performance & robustness*

Table 2 presents the classification accuracy under different variability levels. Under baseline conditions, all models achieved high accuracy (0.980–0.989), confirming the effectiveness of gradient boosting approaches for smart grid stability prediction. XGBoost recorded the marginally highest baseline accuracy, while Hybrid-LXGB and OOF-Stacking also produced competitive results. Although the baseline differences are relatively small, the relative ranking of the models changes slightly under stressed operating conditions. As disturbance intensity increases, classification accuracy declines for all methods. However, the rate of degradation differs across models. In contrast, LightGBM shows more stable behavior under uncertainty. Hybrid-LXGB remains consistently close to the leading models, indicating that probability-level fusion helps preserve competitive performance. OOF-Stacking also shows reliable behavior, particularly under moderate variability.

Although the numerical differences between models are relatively small under some operating conditions, the degradation trends provide useful descriptive information about their robustness as variability increases.

Table 2: Classification accuracy under different variability levels.

| Variability Level | LightGBM | XGBoost | Hybrid-LXGB | OOF-Stacking |
|---|---|---|---|---|
| Baseline | 0.9797 | 0.9808 | 0.9803 | 0.9798 |
| Very Low | 0.9644 | 0.9643 | 0.9644 | 0.9656 |
| Medium | 0.9200 | 0.9198 | 0.9202 | 0.9206 |
| High | 0.7588 | 0.7588 | 0.7592 | 0.7579 |

To assess whether the observed differences between models are statistically significant, pairwise McNemar tests were performed for all six model pairs under baseline, very low, medium, and high variability conditions. Holm correction was applied to account for multiple pairwise comparisons. None of the six pairwise comparisons was statistically significant at $p<0.05$ after correction under any of the evaluated conditions. The maximum observed accuracy difference was 0.108 percentage points at baseline, 0.125 percentage points under very low variability, 0.075 percentage points under medium variability, and 0.133 percentage points under high variability. Therefore, the small numerical differences between models should be interpreted as descriptive performance trends rather than statistically established predictive superiority.

Table 3: Summary of pairwise statistical significance tests.

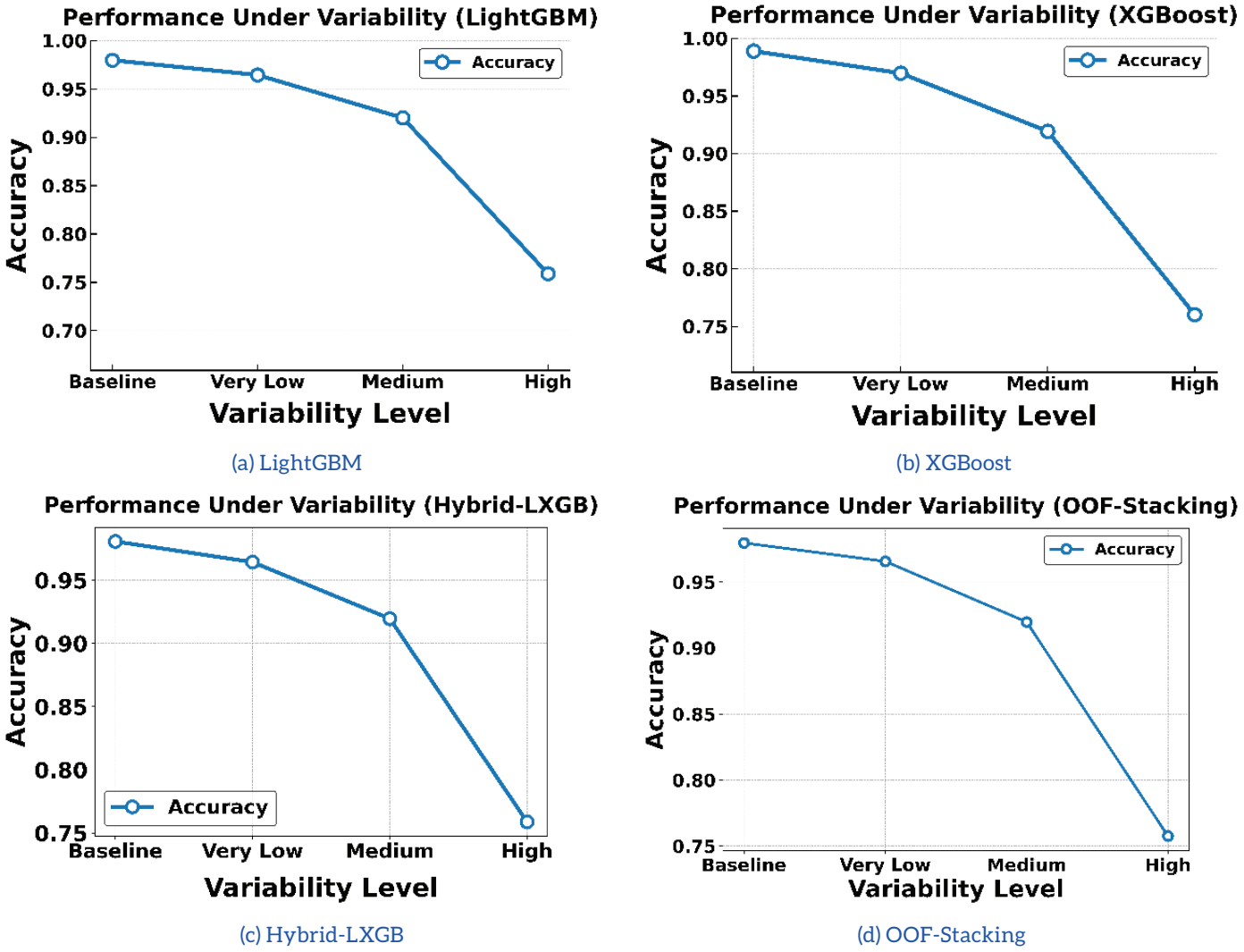
| Variability Level | Pairwise Comparisons | Maximum Accuracy Difference (%) | Significant Comparisons ($p<0.05$) |
|---|---|---|---|
| Baseline | 6 | 0.108 | 0/6 |
| Very Low | 6 | 0.125 | 0/6 |
| Medium | 6 | 0.075 | 0/6 |
| High | 6 | 0.133 | 0/6 |

**Note:** The maximum accuracy differences are calculated using the unrounded accuracy values.

The statistical analysis indicates that none of the evaluated models demonstrates a statistically significant accuracy advantage over the others under the tested variability conditions. Therefore, subsequent comparisons focus on descriptive robustness trends and degradation behavior rather than claims of statistically significant predictive superiority.

The comparative accuracy trends across all variability scenarios are illustrated in Figure 5.

Volume 12, Issue 2, December 2026, e-ISSN 2356-8569
http://dx.doi.org/10.21622/RESD.2026.12.2.2146




(a) LightGBM (b) XGBoost

(c) Hybrid-LXGB (d) OOF-Stacking

Figure 5: Accuracy comparison of all models under baseline and increasing variability conditions.

As observed in Figure 5, the performance gap between models becomes more visible at intermediate disturbance levels, while under severe variability, all models converge toward similar accuracy values.

This convergence occurs because severe perturbations affect the input features of all models to a similar extent, making class separation increasingly difficult. As a result, the advantage of individual learning strategies becomes less pronounced, leading to comparable prediction performance under high variability.

To complement the accuracy analysis, additional classification metrics, including precision, recall, and F1-score, were evaluated under baseline conditions, as presented in Table 4. All four models achieved consistently high values across these metrics, indicating reliable classification performance for both stable and unstable operating states. XGBoost records the highest numerical values across the evaluated metrics, while Hybrid-LXGB, LightGBM, and OOF-Stacking also demonstrate competitive results with only marginal differences. These findings confirm that the evaluated models maintain balanced classification capability under normal operating conditions.

Table 4: Baseline classification performance of the evaluated models.

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| LightGBM | 0.980 | 0.980 | 0.980 | 0.980 |
| XGBoost | 0.981 | 0.981 | 0.981 | 0.981 |
| Hybrid-LXGB | 0.980 | 0.980 | 0.980 | 0.980 |
| OOF-Stacking | 0.980 | 0.980 | 0.980 | 0.980 |

The high precision, recall, and F1-score values further confirm the strong predictive capability of all four models under baseline conditions. The close agreement among these metrics indicates that the models achieve balanced classification performance.

Volume 12, Issue 2, December 2026, e-ISSN 2356-8569
http://dx.doi.org/10.21622/RESD.2026.12.2.2146

The subsequent analysis, therefore, focuses on differences in robustness under increasing variability.

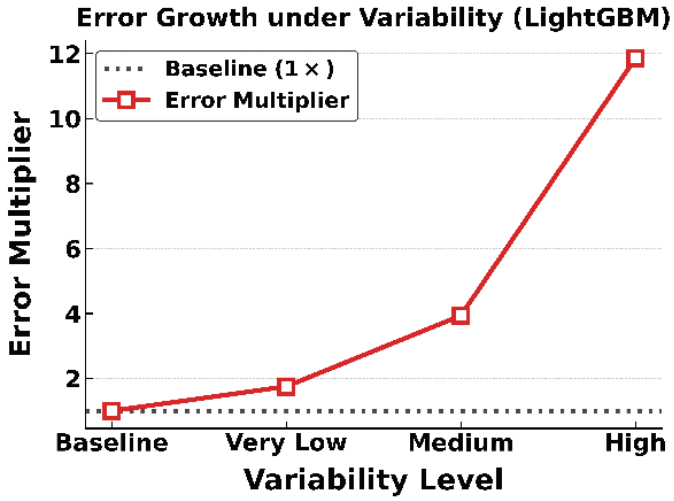
To better understand model stability, the relative accuracy drop is analyzed in Table 5. Under very low variability, all models show only minor degradation of approximately 1.5–1.7%. As disturbance increases, differences in the observed degradation trends become more apparent. At medium variability, the accuracy drop ranges from 6.047% to 6.211%, while high variability results in degradation of approximately 22.5–22.6% across all models. OOF-Stacking shows the lowest degradation at very low and medium variability, whereas LightGBM shows the lowest degradation at high variability. XGBoost exhibits the largest degradation at very low and medium variability. These differences are descriptive, as the statistical analysis shows no statistically significant pairwise differences after Holm correction.

Hybrid-LXGB maintains a competitive balance between predictive accuracy and robustness across the evaluated variability levels. Overall, the similar degradation ranges indicate that increasing variability has a much stronger effect on performance than the small differences between individual models.

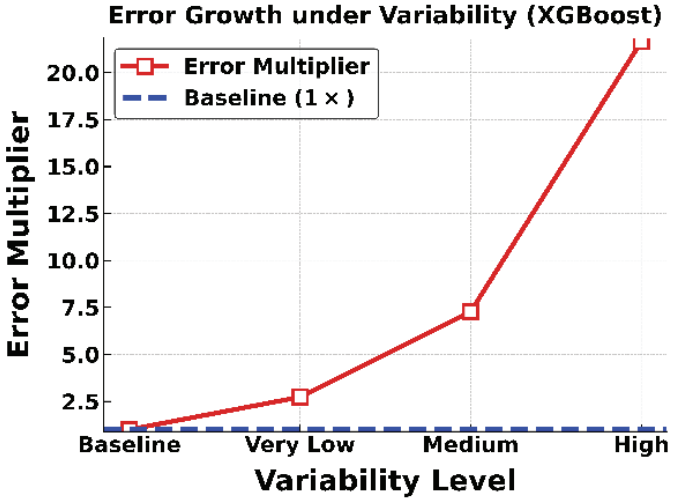
Table 5: Relative accuracy drop (%) under increasing variability levels.

| Variability Level | LightGBM | XGBoost | Hybrid-LXGB | OOF-Stacking |
|---|---|---|---|---|
| Very Low | 1.557 | 1.674 | 1.624 | 1.454 |
| Medium | 6.091 | 6.211 | 6.137 | 6.047 |
| High | 22.542 | 22.636 | 22.552 | 22.648 |

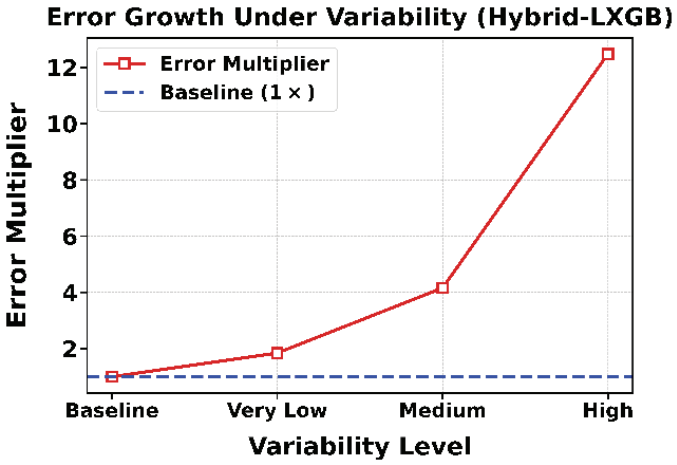
The degradation behavior of each Model under increasing disturbance is further illustrated in Figure 6.



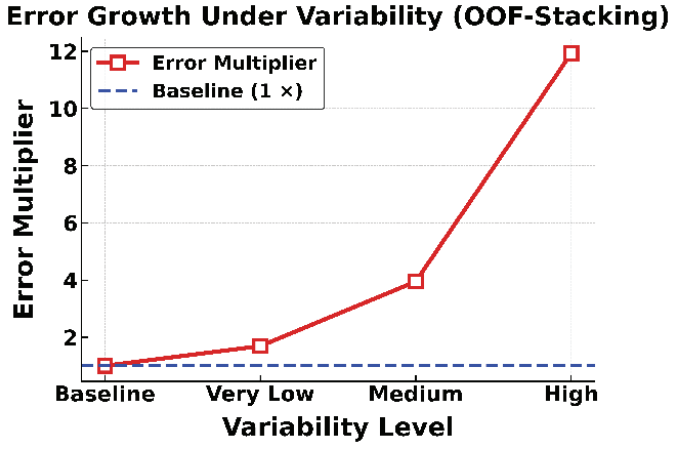

(a) LightGBM




(b) XGBoost




(c) Hybrid-LXGB




(d) OOF-Stacking

Figure 6: Comparison of accuracy degradation trends across models under varying disturbance levels.

Volume 12, Issue 2, December 2026, e-ISSN 2356-8569
http://dx.doi.org/10.21622/RESD.2026.12.2.2146

These results indicate that robustness differs across models, even when their baseline performance is similar.

### *B. Error growth and sensitivity analysis*

Table 6 and Table 7 summarize the error rate and error growth ratio across different variability levels. As expected, prediction errors increase for all models as the disturbance becomes stronger. The error growth ratio also increases substantially with variability, indicating progressively greater degradation relative to the clean baseline. Under high variability, the error growth ratio reaches approximately 12 times the baseline error across the evaluated models. XGBoost shows the highest error growth at very low, medium, and high variability, while LightGBM and OOF-Stacking exhibit comparatively lower error growth. Hybrid-LXGB remains between the standalone models in most conditions. These results indicate that severe variability has a much stronger effect on prediction error than the small differences between the models.

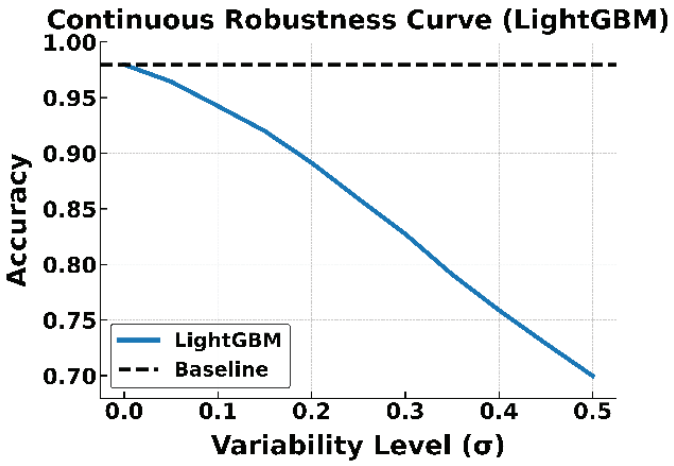
Table 6: Error rate under increasing variability levels.

| Variability Level | LightGBM | XGBoost | Hybrid-LXGB | OOF-Stacking |
|---|---|---|---|---|
| Baseline | 0.0203 | 0.0193 | 0.0197 | 0.0202 |
| Very Low | 0.0356 | 0.0357 | 0.0356 | 0.0344 |
| Medium | 0.0800 | 0.0802 | 0.0798 | 0.0794 |
| High | 0.2412 | 0.2413 | 0.2408 | 0.2421 |

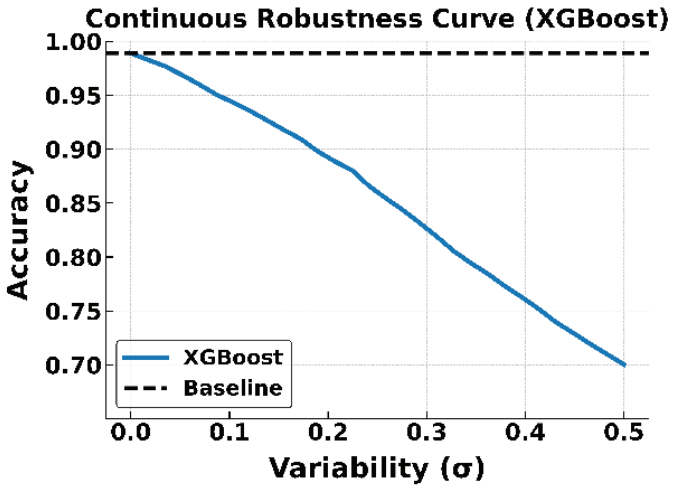
Table 7: Error growth ratio under increasing variability levels.

| Variability Level | LightGBM | XGBoost | Hybrid-LXGB | OOF-Stacking |
|---|---|---|---|---|
| Baseline | 1.000 | 1.000 | 1.000 | 1.000 |
| Very Low | 1.753 | 1.855 | 1.807 | 1.702 |
| Medium | 3.934 | 4.166 | 4.055 | 3.932 |
| High | 11.887 | 12.534 | 12.228 | 11.988 |

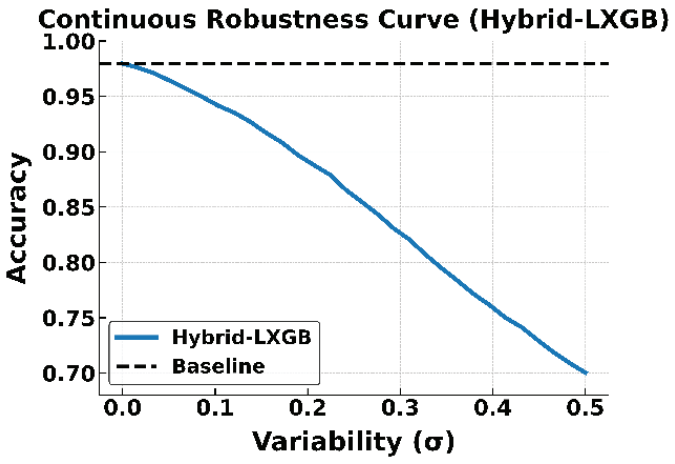
The escalation of prediction error with disturbance intensity is visualized in Figure 7.



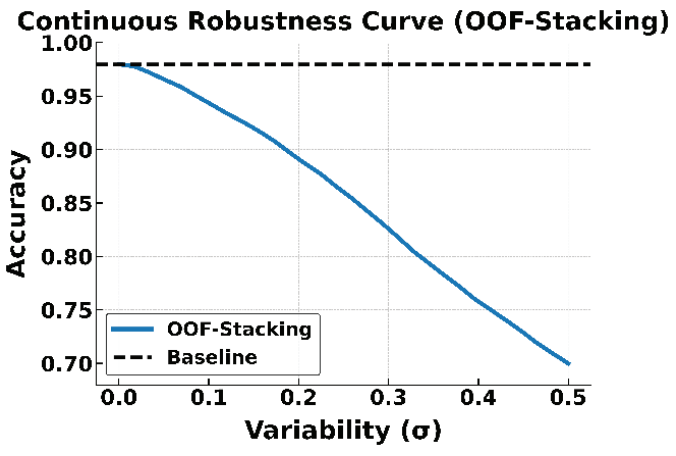

(a) LightGBM




(b) XGBoost




(c) Hybrid-LXGB




(d) OOF-Stacking

Figure 7: Error growth behavior of models showing error amplification with increasing variability.

Volume 12, Issue 2, December 2026, e-ISSN 2356-8569
http://dx.doi.org/10.21622/RESD.2026.12.2.2146

The trends in Figure 7 highlight that error growth is not linear. Instead, it accelerates as variability increases. This behavior confirms that models become more vulnerable under severe disturbance. These findings show that relying only on baseline accuracy can be misleading. A model with strong initial performance may degrade more rapidly under realistic operating uncertainty.

Tables 8 and 9 provide normalized and proportional interpretations of the observed degradation trends under increasing variability. High variability accounts for approximately 74–75% of the total degradation across all models, while medium variability contributes approximately 20%. Very low variability contributes less than 6%. This confirms that severe variability is the dominant contributor to overall performance degradation.

### *C. Continuous response and classification Behavior*

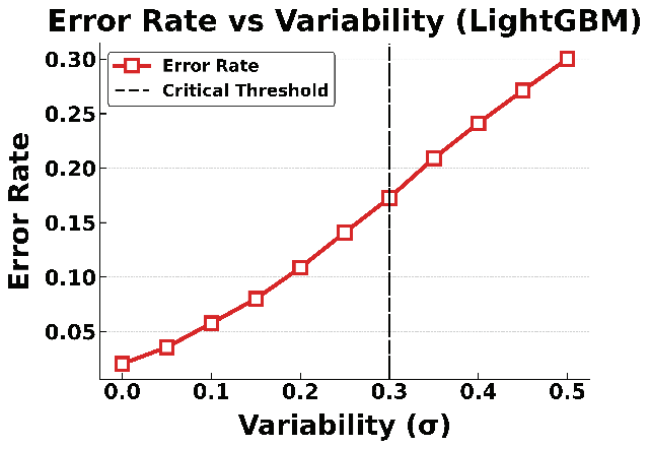
Figure 8 presents the continuous robustness curves, showing how classification accuracy changes as disturbance intensity gradually increases.

Table 8: Normalized impact weight under increasing variability levels.

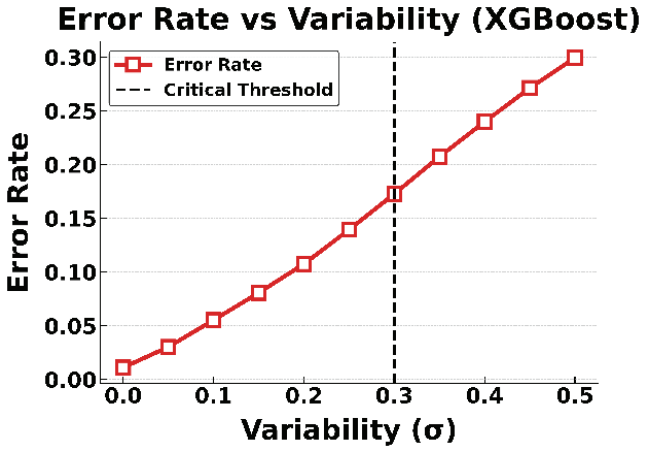
| Variability Level | LightGBM | XGBoost | Hybrid-LXGB | OOF-Stacking |
|---|---|---|---|---|
| Very Low | 0.069 | 0.074 | 0.072 | 0.064 |
| Medium | 0.270 | 0.274 | 0.272 | 0.267 |
| High | 1.000 | 1.000 | 1.000 | 1.000 |

Table 9: Contribution to total degradation (%) under increasing variability levels.

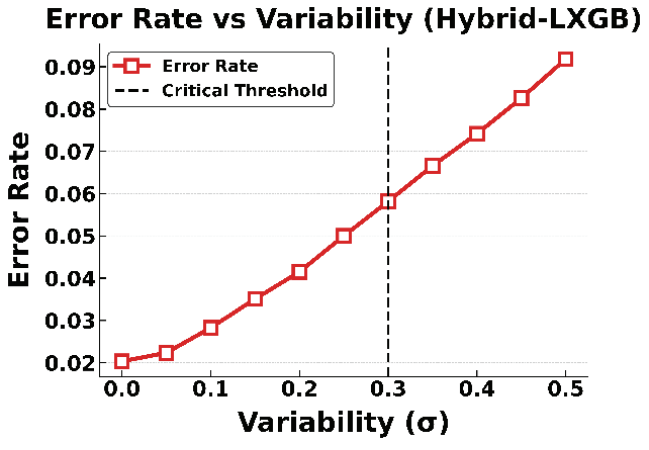
| Variability Level | LightGBM | XGBoost | Hybrid-LXGB | OOF-Stacking |
|---|---|---|---|---|
| Very Low | 5.157 | 5.485 | 5.357 | 4.823 |
| Medium | 20.176 | 20.350 | 20.245 | 20.057 |
| High | 74.667 | 74.165 | 74.397 | 75.120 |



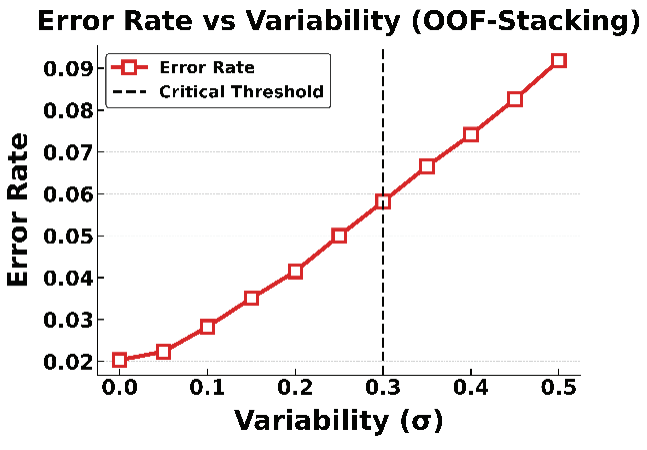

(a) LightGBM




(b) XGBoost




(c) Hybrid-LXGB




(d) OOF-Stacking

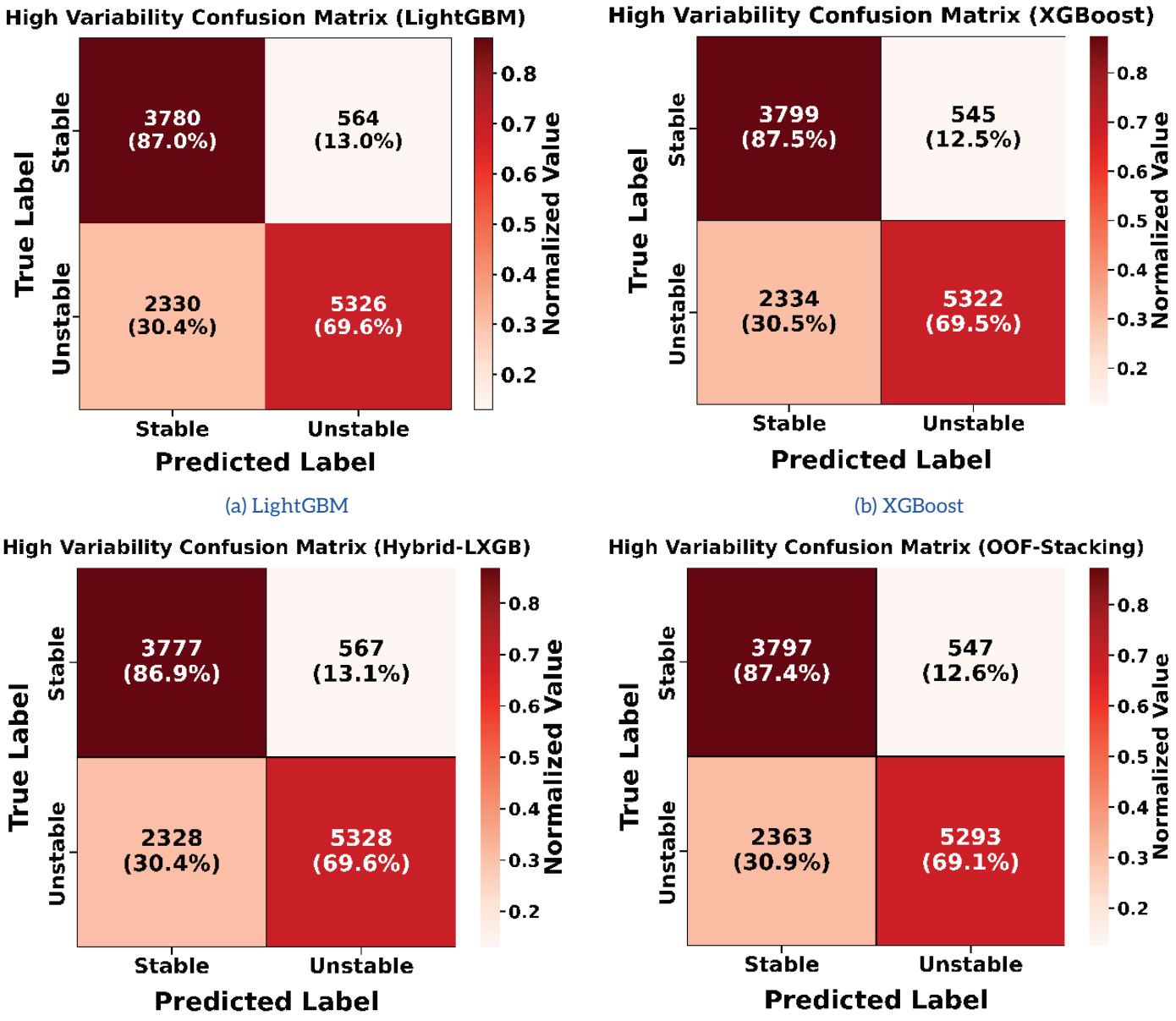
Figure 8: Continuous robustness curves showing accuracy variation with variability intensity (σ).

Volume 12, Issue 2, December 2026, e-ISSN 2356-8569
http://dx.doi.org/10.21622/RESD.2026.12.2.2146

At low disturbance levels, all models perform similarly and maintain high accuracy. As variability increases, accuracy declines in a nonlinear manner, with the reduction becoming more pronounced at higher disturbance levels. LightGBM shows a relatively smooth degradation trend, while XGBoost exhibits a slightly larger reduction in accuracy across the higher variability range. These observations describe the degradation trends rather than statistically significant differences between the models.

Hybrid-LXGB follows an intermediate trend and remains close to the other models across the evaluated range. OOF-Stacking also maintains consistent performance, particularly in the moderate variability region. At severe disturbance levels, all models converge toward similar accuracy. This suggests that extreme variability limits the performance differences among the evaluated methods.

Figure 9 compares the confusion matrices of all models under high variability conditions.




(a) LightGBM

(b) XGBoost

(c) Hybrid-LXGB

(d) OOF-Stacking

Figure 9: Confusion matrix comparison of all models under high variability conditions.

All models retain strong recognition of stable states, with correct classification rates ranging from 86.9% to 87.5%. However, the identification of unstable states is more challenging. The correct classification rates range from 69.1% to 69.6%, indicating reduced performance for these cases. The class-wise prediction differences under severe disturbance are visually summarized in Figure 9.

This result shows that severe disturbance mainly affects the detection of unstable operating conditions. In practical smart grid operation, misclassifying an unstable state as stable may delay corrective control actions, increasing the risk of voltage instability, frequency deviations, or cascading failures. In contrast, predicting a stable state as unstable may result in unnecessary preventive actions, which may generally be less critical than failing to detect an unstable condition. Therefore, improving the reliable detection of unstable operating conditions remains an important objective for real-world smart grid monitoring systems. Among the evaluated models, XGBoost achieves the highest stable-state recognition (87.5%). Hybrid-LXGB and LightGBM show slightly improved recognition of unstable states based on confusion matrix observations (69.6%). OOF-Stacking remains competitive but shows slightly higher confusion for unstable samples.

Overall, the differences between models are relatively small under high variability. This indicates that all models preserve a reasonable level of predictive capability even under severe uncertainty.

These findings reinforce an important conclusion. Evaluating models only under clean benchmark conditions is not sufficient for smart grid applications. Robustness to variability must be treated as a key requirement.

Among the tested methods, LightGBM and Hybrid-LXGB show favorable accuracy–robustness trends, while XGBoost records the highest baseline accuracy. However, the pairwise statistical analysis shows that the observed accuracy differences are not statistically significant after Holm correction. Therefore, these model differences should be interpreted as descriptive trends rather than statistically established superiority. From a practical perspective, XGBoost may be considered when nominal accuracy is prioritized, while LightGBM and Hybrid-LXGB provide competitive alternatives when robustness under increasing variability is also considered.

Computational efficiency was also considered to assess the model-development cost. Table 10 compares the approximate training time of the evaluated models under the same experimental environment. LightGBM required the shortest training time, followed by Hybrid-LXGB and XGBoost. OOF-Stacking required the highest training time because multiple base models were trained during cross-validation before fitting the meta-model. Although OOF-Stacking introduces a higher training cost, this overhead is incurred during model development. These results provide a comparative view of training efficiency and can support model selection when computational cost is also considered alongside predictive performance and robustness.

Table 10: Comparison of computational efficiency.

| Model | Training Time (s) |
|---|---|
| LightGBM | 3.49 |
| XGBoost | 8.46 |
| Hybrid-LXGB | 6.63 |
| OOF-Stacking | 38.40 |

## V. Conclusion

This work presented a perturbation-robustness-focused evaluation of four ensemble learning models for renewable-powered smart grid stability prediction under increasing operating variability. Instead of relying only on nominal accuracy, the study examines how predictive performance changes as uncertainty increases. All models achieve strong baseline performance, while their degradation trends differ under increasing variability. XGBoost records the highest baseline accuracy, while LightGBM shows a relatively lower degradation trend under increasing variability. Hybrid-LXGB provides a competitive balance between accuracy and robustness, while OOF-Stacking remains a stable alternative. However, pairwise McNemar tests with Holm correction show that none of the observed accuracy differences are statistically significant. Therefore, the model differences should be interpreted as descriptive robustness trends rather than statistically established superiority. Severe variability remains the dominant source of performance degradation across all models, highlighting the need to consider robustness alongside prediction accuracy. The study is limited by the use of a single benchmark dataset,

simulated Gaussian perturbations, and offline evaluation. Future work will consider diverse real-world datasets, feature-specific disturbances, sensor and communication uncertainties, and uncertainty-aware prediction methods.

## References

[1] M. B. Prasad, P. Ganesh, K. V. Kumar, P. A. Mohanarao, A. Swathi, and V. Manoj, "Renewable Energy Integration in Modern Power Systems: Challenges and Opportunities," in *International Conference on Renewable Energy Resources and Applications (ICRERA-2024)*, E3S Web of Conferences, Jan. 2024, pp. 1–10. Accessed: Sep. 08, 2026. [Online]. Available: https://typeset.io/papers/renewable-energy-integration-in-modern-power-systems-9byobzb5myno

[2] S. Singh and S. Singh, "Advancements and Challenges in Integrating Renewable Energy Sources Into Distribution Grid Systems: A Comprehensive Review," *Journal of energy resources technology*, vol. 146, no. 9, pp. 1–37, May 2024, doi: 10.1115/1.4065503.

[3] M. Cavus, "Advancing Power Systems With Renewable Energy and Intelligent Technologies: A Comprehensive Review on Grid Transformation and Integration," *Electronics*, vol. 14, no. 6, p. 1159, 2025, doi: 10.3390/electronics14061159.

[4] P. Das *et al.*, "Data-Driven Demand Response Optimization With Hybrid Forecasting and AC Power Flow Validation in Renewable-Integrated Distribution Systems," *Renewable Energy and Sustainable Development*, vol. 12, no. 1, p. 216, Jun. 2026, doi: 10.21622/resd.2026.12.1.1913.

[5] N. A. Sharma, "Current Trends and Future Directions in Renewable Energy Systems," *International Journal for Research Publication and Seminars*, vol. 15, no. 2, pp. 186–198, Jun. 2024, doi: 10.36676/jrps.v15.i2.1408.

[6] M. Khalil and S. A. Sheikh, "Advancing Green Energy Integration in Power Systems for Enhanced Sustainability: A Review," *IEEE Access*, vol. 12, pp. 151669–151692, Jan. 2024, doi: 10.1109/access.2024.3472843.

[7] M. S. Eltohamy *et al.*, "Enhancing Electric Grid Flexibility for the Integration of Variable Renewable Energy: Challenges, Innovations, and Future Directions," *IEEE Access*, vol. 14, pp. 23080–23109, 2026, doi: 10.1109/access.2026.3660837.

[8] A. Joshan, "Emerging Trends and Advanced Techniques in Power System Optimization for Future Smart Grids POWER," *CONTROL AND DATA PROCESSING SYSTEMS*, vol. 2, no. 2, pp. 26–38, 2025, doi: 10.30511/pcdp.2025.2060096.1022.

[9] J. J. Moreno Escobar, O. Morales Matamoros, R. Tejeida Padilla, I. Lina Reyes, and H. Quintana Espinosa, "A Comprehensive Review on Smart Grids: Challenges and Opportunities," *Sensors*, vol. 21, no. 21, p. 6978, Oct. 2021, doi: 10.3390/s21216978.

[10] S. K. Fasogbon *et al.*, "Optimization of Energy Grid Efficiency With Machine Learning: A Comprehensive Review of Challenges and Opportunities," *Renewable and Sustainable Energy Reviews*, vol. 223, p. 115980, Jul. 2025, doi: 10.1016/j.rser.2025.115980.

[11] A. Sundaramurthy, K. Ramasamy, D. Velusamy, and C. Vaithiyalingam, "Enhancing Smart Grid Stability: Data-Driven Predictive Modeling in Distribution Systems," *International Journal of Electrical and Electronics Research*, vol. 12, no. 2, pp. 623–631, Jun. 2024, doi: 10.37391/ijeer.120239.

[12] M. Esmaeili Shayan, M. Petrollese, S. H. Rouhani, S. Mobayen, A. Zhilenkov, and C. L. Su, "An Innovative Two-Stage Machine Learning-Based Adaptive Robust Unit Commitment Strategy for Addressing Uncertainty in Renewable Energy Systems," *International Journal of Electrical Power & Energy Systems*, vol. 160, p. 110087, Sep. 2024, doi: 10.1016/j.ijepes.2024.110087.

[13] P. N. Papadopoulos, S. Chatzivasileiadis, and A. Marot, "Can Machine Learning Help Keep the System Secure?: Power Systems and Change Addressing the Increasing Complexity and Uncertainty During the Energy Transition," *IEEE Power and Energy Magazine*, vol. 22, no. 6, pp. 100–111, Nov. 2024, doi: 10.1109/mpe.2024.3421388.

[14] A. M. Ali, O. K. Dawoud, O. A. Ghoneim, and M. Abdel-Basset, "Machine Learning Models for Smart Grid Stability Prediction: A Comparative Analysis," *Scientific Reports*, vol. 16, no. 1, p. 12694, Apr. 2026, doi: 10.1038/s41598-026-47385-x.

[15] G. Porawagamage, K. Dharmapala, J. S. Chaves, D. Villegas, and A. Rajapakse, "A Review of Machine Learning Applications in Power System Protection and Emergency Control: Opportunities, Challenges, and Future Directions," *Frontiers in smart grids*, vol. 3, p. 1371153, Apr. 2024, doi: 10.3389/frsgr.2024.1371153.

[16] A. Al-Ataby and H. Attia, "Deep Learning-Based Stability Prediction for Electrical Grids: A Data-Driven Approach," in *2024 17th International Conference on Development in eSystem Engineering (DeSE), Khorfakkan*, United Arab Emirates, Nov. 2024, pp. 574–578. Accessed: Sep. 08, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10912018/

[17] J. Li, S. Zhou, Y. Xu, M. Zhu, and L. Ye, "A Multi-Band Uncertainty Set Robust Method for Unit Commitment With Wind Power Generation," *International Journal of Electrical Power & Energy Systems*, vol. 131, p. 107125, Oct. 2021, doi: 10.1016/j.ijepes.2021.107125.

[18] S. Chatzivasileiadis, A. Venzke, J. Stiasny, and G. Misyris, "Machine Learning in Power Systems: Is It Time to Trust It?," *IEEE Power and Energy Magazine*, vol. 20, no. 3, pp. 32–41, May 2022, doi: 10.1109/mpe.2022.3150810.

[19] M. Sasmal, P. Das, B. Pal, S. Ghosh, and S. Dutta, "Quantum-Based Hybrid Framework for Adaptive Congestion Management in Smart Grids With Stochastic Renewable Penetration," *Renewable Energy and Sustainable Development*, vol. 12, no. 1, p. 173, May 2026, doi: 10.21622/resd.2026.12.1.1910.

[20] R. R. Elbanna, M. H. ElMessmary, H. Diab, and M. Abdelsalam, "A Smart Hybrid Optimization Model for DSSE in Renewable Energy-Powered Distribution Networks," *Renewable Energy and Sustainable Development*, vol. 11, no. 2, p. 314, Sep. 2025, doi: 10.21622/resd.2025.11.2.1271.

[21] J. Oelhaf *et al.*, "A Scoping Review of Machine Learning Applications in Power System Protection and Disturbance Management," *International Journal of Electrical Power & Energy Systems*, vol. 172, p. 111257, Oct. 2025, doi: 10.1016/j.ijepes.2025.111257.

[22] M. K. Boutahir, A. Hessane, Y. Farhaoui, and M. Azrour, "An Effective Ensemble Learning Model to Predict Smart Grid Stability Using Genetic Algorithms," in *Advanced Technology for Smart Environment and Energy*, J. Mabrouki, A. Mourade, A. Irshad, and S. Chaudhry, Eds., Cham, Switzerland: Springer, 2023, pp. 129–137.

[23] E. Marevac, E. Kadusic, C. Ruland, and N. Zivic, "Application of Deep Learning and Neural Networks for Smart Grid Stability Predictions," in *2024 IEEE 12th International Conference on Smart Energy Grid Engineering (SEGE)*, Oshawa, ON, Canada, Aug. 2024, pp. 84–89.

[24] P. S. Kumar, C. Chandrika, P. K. Rao, P. K. Rao, and S. K. Oruganti, "Interpretable Hybrid Machine Learning Models for Renewable-Powered Smart Grid Stability Prediction," *Renewable Energy and Sustainable Development*, vol. 11, no. 2, p. 397, Oct. 2025, doi: 10.21622/resd.2025.11.2.1509.

[25] V. A. G. Raju, M. Mishra, J. G. Singh, J. Nayak, and P. B. Dash, "Smart Grid Stability Prediction Based on Bayesian-Optimised LGBM for Smarter Energy Management," *International Journal of Advanced Mechatronic Systems*, vol. 11, no. 4, pp. 226–241, 2024, doi: 10.1504/ijamechs.2024.143382.

[26] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 3146–3154.

[27] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, vol. 1, no. 1, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785.

[28] L. Rokach, "Ensemble-Based Classifiers," *Artificial Intelligence Review*, vol. 33, no. 1, pp. 1–39, Nov. 2009, doi: 10.1007/s10462-009-9124-7.

[29] V. Coscrato, M. H. de A. Inácio, and R. Izbicki, "The NN-Stacking: Feature Weighted Linear Stacking Through Neural Networks," *Neurocomputing*, vol. 399, pp. 141–152, Jul. 2020, doi: 10.1016/j.neucom.2020.02.073.

[30] C. Yarramsetty, T. Moger, and D. Jena, "A Hybrid Model of Convolutional Neural Network and an Extreme Gradient Boosting for Reliability Evaluation in Composite Power Systems Integrated With Renewable Energy Resources," *Electrical Engineering*, vol. 107, no. 3, pp. 3495–3510, Sep. 2024, doi: 10.1007/s00202-024-02683-3.

[31] J. Bergstra and Yoshua Bengio, "Random Search for Hyper-Parameter Optimization," *Journal of machine learning research*, vol. 13, no. 10, pp. 281–305, 2017, doi: 10.5555/2503308.2188395.

[32] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," *arXiv*, Dec. 2014, doi: 10.48550/arxiv.1412.6572.

[33] T. A. Rajaperumal and C. C. Columbus, "Enhanced Wind Power Forecasting Using Machine Learning, Deep Learning Models and Ensemble Integration.," *PubMed*, vol. 15, no. 1, pp. 20572–20572, Jul. 2025, doi: 10.1038/s41598-025-05250-3.

[34] P. Breviglieri, T. Erdem, and S. Eken, "Predicting Smart Grid Stability With Optimized Deep Models," *SN Computer Science*, vol. 2, no. 2, Feb. 2021, doi: 10.1007/s42979-021-00463-5.

[35] V. Arzamasov, K. Bohm, and P. Jochem, "Towards Concise Models of Grid Stability," in *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, Aalborg, Denmark, Oct. 2018, pp. 1–6.

[36] F. Meißner, T. Sous, C. Ziesemann, P. Wirtz, F. Klein-Helmkamp, and A. Moser, "Developing a Benchmark Grid Model for Future Congestion Management and Stability Analysis," in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Dubrovnik, Croatia, Oct. 2024, pp. 1–5.

[37] C. U. Osuagwu, "Advanced Grid Management Strategies for Mitigating Stability Challenges in High Renewable Energy Penetration Scenarios," *CogNexus*, vol. 1, no. 02, pp. 119–139, May 2025, doi: 10.63084/cognexus.v1i02.93.

[38] B. Schäfer, D. Witthaut, M. Timme, and V. Latora, "Dynamically Induced Cascading Failures in Power Grids," *Nature Communications*, vol. 9, no. 1, p. 1975, May 2018, doi: 10.1038/s41467-018-04287-5.