SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I8P121

Volume 11 Issue 8, 236-247, August 2024

Original Article

Predicting Transient Stability of Power Systems Using Machine Learning: A Case Study on the IEEE New England 39-Bus Test System

Diaa Salman[1*], Yonis Khalif Elmi[2], Abdulaziz Ahmed Siyad[1], Abdirahman Abdullahi Ali[2]

[1]*Faculty of Engineering, Jamhuriya University of Science and Technology, Mogadishu, Somalia.*
[2]*Department of Electrical and Electronic Engineering, Cyprus International University, Nicosia, Northern Cyprus, Mersin.*

[*]*Corresponding Author: d.salman@just.edu.so*

Received: 10 June 2024 Revised: 12 July 2024 Accepted: 10 August 2024 Published: 31 August 2024

Abstract - *The need to evaluate the transient stability of power systems is inevitable and crucial in order to ensure that they will continue to operate efficiently after interruptions. In the present investigation, an attempt is made to use machine learning methods, particularly the XGBoost and the Random Forest models, with the objective of predicting the stability of the power systems after a fault has occurred. Thus, the dataset used by the models has several generator and bus parameters, as well as pre and post-fault conditions; the objective is to identify if the system stability is stable or unstable. In general, it is possible to conclude that the use of a hybrid model, combining the XGBoost and Random Forest techniques, outperforms each model separately. This is the case because it has the merits of both methods as it combines the two methods to identify the similarities of the two plans. In this case, the effectiveness of the proposed approach is assessed using evaluation parameters like accuracy, precision, recall rate, and F1-score. Furthermore, the study gives an understanding of the stability measures most affected by the characteristics and those that affect stability predictions. By applying more complex modes of predictive modeling, this work contributes to advancing the reliability and efficacy of power grid management.*

Keywords - *Transient stability, Machine learning, Power systems, XGBoost, Random Forest.*

1. Introduction

The transient stability of power systems is one of the most important aspects to ensure grid reliability. It becomes even more critical with the increased penetration of renewable energy sources. Traditional assessment of transient stability based on time-domain simulation and direct methods has been proven to suffer from shortcomings in both accuracy and computational effectiveness when power systems, equipped with distributed generations and uncertain renewable inputs, turn to be more complex and dynamic [1, 2]. This has raised interest in Machine Learning (ML) and Artificial Intelligence (AI) methods to improve transient stability prediction and management in modern power grids. The ensemble learning technique, particularly with Extreme Gradient Boosting (XGBoost) and Random Forest, presents efficacy in accurate system stability prediction under diverse operating conditions and fault scenarios utilizing large datasets and advanced feature selection techniques [3].

More recent advances in ML for the assessment of the stability of power systems have been on the integration of deep learning models like CNNs and RNNs, which capture complex temporal and spatial patterns within power system data [3, 4]. Moreover, these models are improved in adaptiveness and efficiency by utilizing transfer learning techniques and AutoML, reducing manual effort on feature engineering and parameter tuning within real-time stability assessments [5]. The additional support that such technologies will give in revolutionizing transient stability assessment is the deployment of real-time monitoring systems with emergency control using AI for timely and accurate information in the operation and prevention of blackouts [5].

Consistent with these advances, Transient Stability Assessment (TSA) in power systems using machine learning techniques has advanced dramatically. The higher penetration of renewable sources and the complexity of today's power grids demanded robust and accurate TSA. One of the exciting studies used machine learning techniques, in particular, Extreme Gradient Boosting and Random Forests, to develop dynamic stability analysis for power systems. The study illustrated flexibility in applying machine learning algorithms in the face of variations in parameters and structure for power systems by deriving significant improvements in the accuracy of results when taken concerning system topology [4]. Another study presented an improved form of TSA by combining feature selection using mRMR and ensemble learning implemented by WTA. The model demonstrated high accuracy in transient stability prediction through real-time data from comprehensive Area Monitoring Systems (WAMS), complemented with features based on electromagnetic power and voltage amplitude [5]. Another hybrid approach based on automated machine learning through auto-feature selection and Bayesian optimization has been proposed in a generalized form. The rationale behind this approach is to enhance the interpretation and deployment effectiveness of TSA models. Some other investigation reports have extended this by introducing CatBoost, a gradient boosting algorithm, combined with SHAP analysis for feature importance in testing the power systems [6].

Some studies have investigated deep learning models using stacked autoencoders combined with voting ensemble classifiers to address TSA challenges. These models have shown the capacity to handle big data by predicting reliable stability, which is an essential feature in the real-time operation of a power system. In addition to these, a hierarchical approach has also been developed to predict transient stability, which uses multiple support vector machines. This method achieves a compromise between accuracy in prediction and time of response, given misjudgments for unstable instances. This shortcoming is addressed by constructing different ensemble classifiers for each layer of the hierarchical scheme [7]. Furthermore, the work of a critical review on data-driven TSA approaches also presented the principles, prospects, and challenges in the field. The feature extraction, selection, model construction, and online learning toward effective TSA models significantly rest on this work [1].

There is another innovative way of using trajectory clusters to define the most relevant features towards TSA better. Applying Support Vector Machine (SVM) models enhances the robustness of the TSA system in various conditions, therefore improving the accuracy of the prediction [8]. On the other hand, other recent reviews concern the application of artificial intelligence techniques to machine, deep, and reinforcement learning for TSA. Some of these include the benefits and challenges of different AI approaches, emphasizing the need for advanced data generation, processing, and the deployment of models. Finally, techniques have been reviewed to improve the transient stability of renewable-rich power systems. These research works discuss the challenges that result from the low inertia features of renewables but proffer amelioration towards enhancing stability through advanced converter/inverter topologies, and their control means [9].

Consequently, transient stability is becoming a highly pertinent area of concern with the increasing complexity and penetration level of renewable energy sources. Traditional tools for transient stability studies include both time domain-based simulators and direct methods that have typically revealed limitations in accuracy, computational efficiency, and, more importantly, applicability for recent power grids with high variability and distributed generations. Previous studies have focused on applying many machine learning models to enhance the prediction accuracy of their TSA models and strive for efficiency. In this regard, Extreme Gradient Boosting (XGBoost) and Random Forest (RF) are ensemble learning methods that have shown great promise. Efforts must be made to investigate hybrid approaches that combine strength from a wider variety of machine learning models for better robustness and performance.

In this regard, the study aims to close this gap by developing and evaluating a hybrid XGBoost-RF model in predicting the transient stability of power systems; specifically, it will use the IEEE New England 39-Bus Test System. It includes the benefits of two models: XGBoost and Random Forest. The hybrid model ensures better predictive accuracy and reliability than the single model. The methodology involves data acquisition, pre-processing, feature selection, and model development, followed by training and validation using appropriate metrics. The results demonstrated that the hybrid XGBoost-RF model outperforms other individual models for improved accuracy, precision, recall, and F1 scores. The crucial information provided through this research will advance the assessment of the stability of a power grid, provide insights into the critical stability features, and enhance the reliability and management of the grid.

2. Models Description

The reliability of power systems is also very important, and they need to run continuously in order to maintain the stability and adequacy of today's power systems. As the power systems become more complicated and the load demands higher, transient stability, which refers to the system's ability to remain in phase with a sudden disturbance such as a fault, becomes important. Transient stability is an essential factor in power system security and its behavior during disturbances. It determines the system's capacity to remain stable and regain operating stability after a disturbance without a chain reaction of failures or blackouts.

The main research aim of this study is to build and evaluate the performance of XGBoost, RF, and the XGBoost-RF hybrid model for the assessment of transient stability in the power system. To evaluate the models' performance, accuracy, precision, recall, and F1-score indicators are applied. Furthermore, the study seeks to establish some of the main attributes that impact on stability predictions so as to help power system operators and engineers. A novel ensemble learning technique for TSA is put forth by Kunac et al. (2020) [10]. This has been presented as an ML approach to classification as a binary problem, and it is suggested that this approach be employed. They discuss the problems that are intrinsic to datasets that are imbalanced and derived from network measurements that have been obtained from PMUs installed in the distribution network during disturbances. It employs a wide assortment of base learners who are grouped in a voting committee to enhance robustness and accuracy. This is done in order to address the fact that there are issues of duplication and the problem of multicolinearity with many features. It is a method that integrates a pre-processing stage of feature selection that employs importance analysis based on various decision tree-based models. This further validates the efficiency of the ensemble learning model and its ability to work for the TSA. Their approach is theoretically sound, and when evaluated on the IEEE New England 39-bus test system, they are able to perform reasonably well, especially when dealing with imbalanced sample distributions and differential misclassification costs during testing.

Zhang et al. (2021) propose an active learning-based approach to power system TSA that deals with the issues of data acquisition for the offline training phase and the problem of frequent updates of the models due to changes in the grid [11]. Building on the highly widespread usage of PMUs, the method begins generating unlabeled samples through short-time simulations of various operational conditions and faults. Important TSA characteristics are chosen, and some of the samples are kept for long-term imitation to indicate their transience. These labeled samples are used to train a random forest model. It involves the high information entropy data from the remaining unlabeled samples are picked and labeled, used in the retraining of the model until the accuracy converges. When applied to a power system, the proposed method achieves drastic improvement in offline simulation time, model efficiency, and insensitivity to wide-area noise, proving its applicability for real-time TSA.

In the context of the growing demand for electricity, Wang et al. (2021) describe a method to predict transient stability based on the Long Short-Term Memory (LSTM) network, a type of recurrent neural network [12]. This method seeks to determine in advance the possibility of transient stability in order to maintain system stability. To compare the performance of the proposed LSTM-based approach, the work compares it with a multi-layer support vector machine on the IEEE 9-bus system and tests it on the New England 39-bus system. The training and testing data for the LSTM network is created through time-domain simulation with the aid of the power system analysis toolbox. Numerical analysis of simulation data suggests that the proposed LSTM-based method enhances the classification accuracy for stability prediction, thus validating its potential for enhancing the transient stability assessment of power systems.

2.1. Extreme Gradient Boosting (XGBoost)

One of the most famous efficient and scalable algorithms of gradient boosting machines is Extreme Gradient Boosting (XGBoost). This ensemble learning method proposes combining the predictive power of several weak learners into constructing a strong learner. XGBoost optimizes the objective function using gradient descent, and it also uses a regularization term that prevents model overfitting [13].

$$\mathcal{L}(\Theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{1}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{2}$$

$$g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i}, h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \tag{3}$$

$$w_j = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{4}$$

XGBoost has proved to be a very dominant and efficient algorithm in machine learning. Equation 1 defines the overall objective function for the case of XGBoost, where l represents the differentiable convex loss function measuring the difference between predictions $\hat{y}_i$ and its target y_i. Ω is the regularization term for the penalty associated with model complexity f_k In particular, the regularization term Ω has been defined with equation 2, where T is the number of leaves in the tree, λ is the regularization parameter, and w_j are the weights in the leaf, thus ensuring that it is productive and generalizes well for unknown data.

As Equation (3) shows, gradient g_i and Hessian h_i are correspondingly the first and second derivatives of the loss function, which are used to update model parameters and play a large part in model optimization. The leaf weight w_j is optimized using the formula provided in Equation (4), where I_j represents the set of data points assigned to leaf j, ensuring each leaf in the decision tree is weighted appropriately based on the contribution of the data points it contains. Figure 1 shows a schematic representation of the XGBoost model architecture, illustrating the process flow and key components involved in model building and training. Table 1 describes the structure and summary statistics of the dataset used in this study, which comprises 3120 instances with 350 features, where 80% of the instances are labelled as stable and 20% as unstable. The fault types are distributed as follows: single-phase faults constitute 70%, double-phase faults account for 20%, and three-phase faults make up 10%.

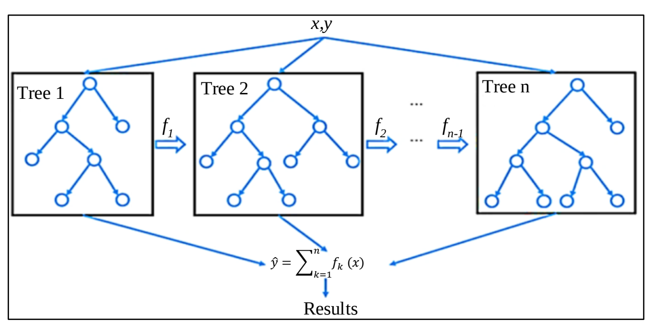

Fig. 1 Schematic representation of the XGBoost model architecture [14]

Table 1. Description of the dataset structure and summary statistics

Description	Value
Number of instances	3120
Number of features	350
Class distribution (stable)	80%
Class distribution (unstable)	20%
Fault Types Distribution	
Single-phase	70%
Double-phase	20%
Three-phase	10%

2.2. Random Forest

Random Forest (RF) is a collective learning technique that generates numerous decision trees during the training phase and produces the most frequent class occurrence for classification objectives or the average prediction for regression tasks [15]. Building each tree using a random subset of the training data and features, RF reduces overfitting and improves generalization. The prediction for an instance x using a Random Forest is given by:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} h_i(x) \quad (5)$$

Equation 5 illustrates how the prediction is made by averaging the predictions $h_i(x)$ of all N trees in the forest, thus providing a robust and reliable outcome. The trees are built using the following steps:

- Bootstrap Sampling: Randomly sample n instances from the training set with replacement.
- Feature Selection: At each node, randomly select m features from the total p features.
- Node Splitting: Split the node using the feature that provides the best split according to a certain criterion (e.g., Gini impurity, information gain).

Gini impurity for a node, used for selecting the best feature, is defined as:

$$G = \sum_{i=1}^{C} p_i(1 - p_i) \quad (6)$$

In Equation 6, C is the number of classes and p_i is the proportion of instances of class i in the node. This impurity measure helps identify the feature that best separates the classes, thereby enhancing the tree's decision-making process.

Random Forest operates by constructing multiple decision trees during training. The output is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Figure 2 shows a schematic representation of the Random Forest model architecture, highlighting the ensemble of decision trees and their contribution to the final prediction [16].

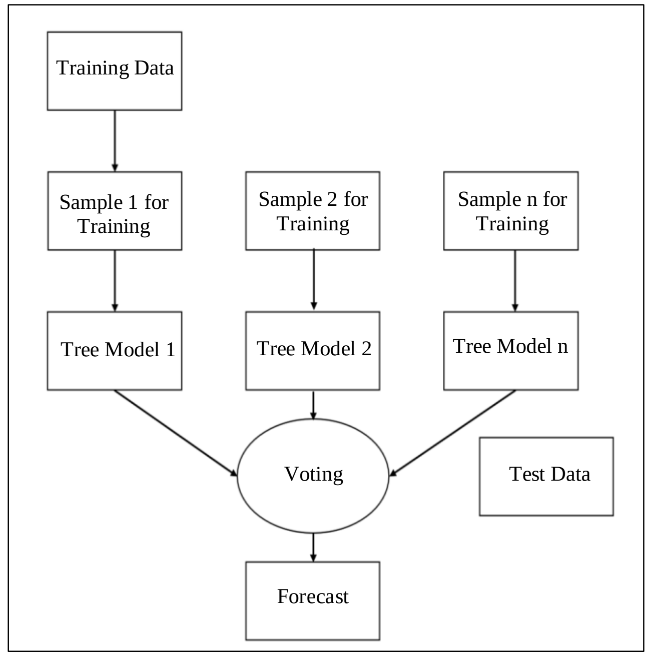

Fig. 2 Schematic representation of the Random Forest model

2.3. Hybrid XGBoost - RF Model

The hybrid XGBoost-RF model combines the strengths of both XGBoost and Random Forest to improve prediction accuracy and robustness [17]. This approach leverages the sequential boosting nature of XGBoost with the parallel ensemble strategy of RF. The hybrid model follows these steps:

- Feature Importance with RF: Train a Random Forest model to rank the features based on their importance, which helps in selecting the most relevant features for the subsequent XGBoost model.
- Feature Selection: Select the top-ranked features identified by the Random Forest model.
- Boosting with XGBoost: Use the selected features to train an XGBoost model, which refines the predictions by focusing on the misclassified instances.

The combined prediction y of the model is given by:

$$\hat{y} = \alpha\hat{y}_{\text{XGBoost}} + (1 - \alpha)\hat{y}_{RF} \quad (7)$$

According to Equation 7, $\hat{y}_{\text{XGBoost}}$ is the prediction made by the XGBoost model, while $\hat{y}_{\text{RF}}$ is the prediction resulting from the Random Forest model. For this case, having α enough as a weighting factor ensures sufficient generalized cross-validation to balance contributions by both models.

With the hybrid approach, the model will help in reducing variance because of RF's low value and bias due to XGBoost's low bias value; hence, high accuracy and reliability can be attained in the predicted system designed. Thus, a hybridizedXGBoost–RF model will incorporate the two major strengths of XGBoost and Random Forest. It is expected that such a hybrid model will leverage these strengths from both modelling approaches to strive ahead in improving overall prediction accuracy and robustness. Figure 3 Schematic representation of the hybrid model architecture for the integration of XGBoost-RF to enhance predictive performance.

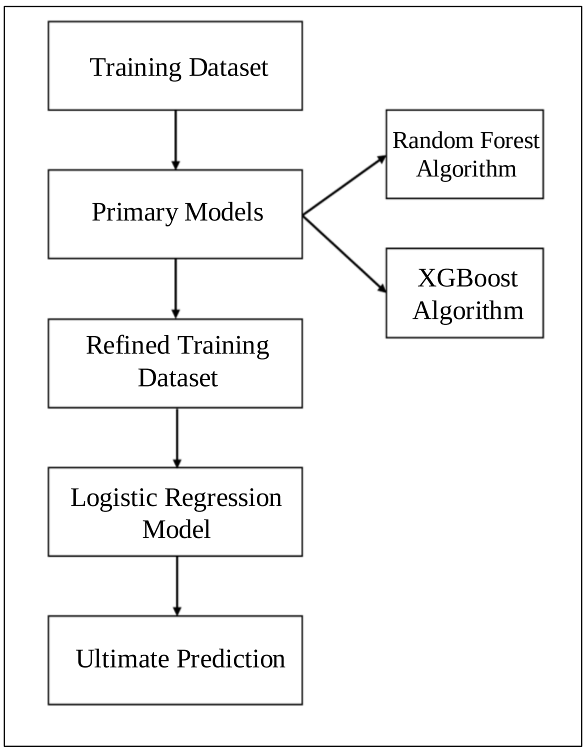

Fig. 3 Schematic architecture of the hybrid XGBoost-RF model

Table 2 compares the hyperparameters of an XGBoost model with those of a Random Forest model. What is shown in the table is different settings for each model to fine-tune its performance. For example, on the side of the XGBoost, it uses a learning rate of 0.1 with 100 boosting rounds, while on the side of Random Forest, there will be the implementation of 100 trees without maximum depth, allowing the trees to grow until all leaves are pure.

The two models have their different ways of feature selection and fighting overfitting. Among parameters for fine-tuning the boost process, there exists the learning_rate and colsample_bytree in the case of XGBoost, whereas Random Forest has n_estimators and max_features as critical parameters leading to the robust ensemble of decision trees in its form. All of this proves how each model improves prediction accuracy and model performance distinctly.

Table 2. Configurations for XGBoost and Random Forest models

Model	Parameter	Explanation	Best Value
XGBoost	Learning Rate	Shrinkage step to avoid overfitting	0.1
	Boosting Rounds	Number of iterations for boosting	100
	Tree Depth	Maximum depth a tree can reach	6
	Child Weight	Minimum instance weight sum required in a child	1
	Sample Ratio	Proportion of training data used in each boosting iteration	0.8
	Column Ratio	Proportion of columns used in tree construction	0.8
Random Forest	Tree Count	Total number of trees	100
	Max Tree Depth	Maximum allowable tree depth	None
	Split Requirement	Minimum samples needed to split a node	2
	Leaf Requirement	Minimum samples needed at a leaf node	1
	Feature Count	Number of features considered for the best split	The square root of the total features

2.4. Data Description

The dataset comprises 350 features derived from PMU-type signals combined with phasor measurements. These features were generated through 9360 systematic electro-mechanical transient simulations in MATLAB®/Simulink, based on the IEEE New England 39-bus power system test case network. This database was created to be open and accessible, facilitating experimentation with various machine learning techniques for Transient Stability Assessment (TSA) of electrical power systems [18].

The dataset comprehensively covers different load and generation levels for the New England 39-bus benchmark power system. It includes all three primary types of short-circuit events at any line location: three-phase, two-phase, and single-phase faults. The network's power consumption was adjusted to 80%, 90%, 100%, 110%, and 120% of the fundamental system load levels. Short circuits were positioned either on the busbar or the transmission line (TL), occurring at 20%, 40%, 60%, and 80% of the line length. Features were effectively extracted through direct analysis of time-domain signals at the pickup time (pre-fault value) and trip time (post-fault value) of the corresponding distance protection relays.

The dataset is stochastic, comprising 3120 cases constructed from 9360 systematic simulations. It includes a statistical distribution of fault types: single-phase faults (70%), double-phase faults (20%), and three-phase faults (10%). Additionally, the dataset exhibits class imbalance, with less than twenty percent of instances belonging to the unstable class. Table 3 lists the feature names in the dataset.

Table 3. Dataset features description

No.	Feature	Description
1	WmGx	Rotor speed for each generator Gx, from G1 to G10
2	DThetaGx	Rotor angle deviation for each generator Gx, from G1 to G10
3	ThetaGx	Rotor mechanical angle for each generator Gx, from G1 to G10
4	VtGx	Stator voltage for each generator Gx, from G1 to G10
5	IdGx	Stator d-component current for each generator Gx, from G1 to G10
6	IqGx	Stator q-component current for each generator Gx, from G1 to G10
7	LAfvGx	Pre-fault power load angle for each generator Gx, from G1 to G10
8	LAlvGx	Post-fault power load angle for each generator Gx, from G1 to G10
9	PfvGx	Pre-fault value of the generator active power for each generator Gx, from G1 to G10
10	PlvGx	Post-fault value of the generator active power for each generator Gx, from G1 to G10
11	QfvGx	Pre-fault value of the generator reactive power for each generator Gx, from G1 to G10
12	QlvGx	Post-fault value of the generator reactive power for each generator Gx, from G1 to G10
13	VAfvBx	Pre-fault bus voltage magnitude in phase A for each bus Bx, from B1 to B39
14	VBfvBx	Pre-fault bus voltage magnitude in phase B for each bus Bx, from B1 to B39
15	VCfvBx	Pre-fault bus voltage magnitude in phase C for each bus Bx, from B1 to B39
16	VAlvBx	Post-fault bus voltage magnitude in phase A for each bus Bx, from B1 to B39
17	VBlvBx	Post-fault bus voltage magnitude in phase B for each bus Bx, from B1 to B39
18	VClvBx	Post-fault bus voltage magnitude in phase C for each bus Bx, from B1 to B39
19	Stability	Binary indicator (0/1) that determines if the power system was stable or unstable (0 - stable, 1

2.5. Working Structure

The analysis is initiated with data acquisition; the study ensures that it acquires massive data on features like rotor speed, rotor angle deviation, stator voltage, pre and post-fault power, and stability factors.

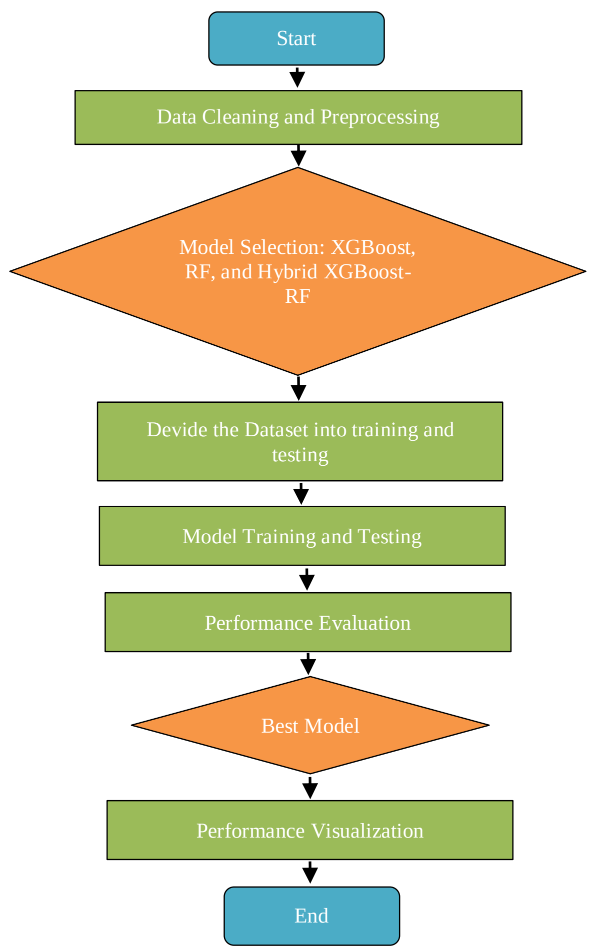

Fig. 4 Workflow of the research

After data collection, data pre-processing is done to handle any missing values and scale the data if needed. The data is then divided into train and test data sets for the purpose of model performance assessment. Feature selection is performed to select the most important features for classification and sort them according to their importance using an initial model. After that, the model-building process is done, where each of the XGBoost and Random Forest models is constructed individually and then integrated into a single, hybrid XGBoost-RF model. The models are trained with the training set. As for evaluating the models, there are several measures like accuracy, precision, recall, and F1-score. Further, the evaluation of the model is done by analyzing the ROC curve and the Precision-Recall curve. Last but not least, the variables that are most influential in supporting the model’s decision are extracted and displayed.

3. Results and Discussions

This paper compares three models, namely, XGBoost, RF, and a combined model of XGBoost-RF, for the purpose of assessing their proficiency in identifying transient stability in the IEEE New England 39-bus test system. The performance of these models is assessed using several metrics: Precision, recall, F1 score, and overall accuracy of the model. These metrics are vital in assessing the capability of the models to distinguish between stable and unstable states and, hence, accuracy and reliability in power system stability analysis. Table 4 illustrates the overall metrics results of the three models.

The model that was chosen for this task – XGBoost – performs quite well with an overall accuracy of 0. 89. For class 0 (stable), recall is 0. 95; thus, the stability of the prediction is highly likely to be true in 95 percent of the cases. The recall is 0. 89, which indicates the model correctly detects 89 percent of all actual stable instances. The F1 score, which is the average of the precision and recall rates and is a more accurate measure than the individual rates, is 0.89, which is almost equal to the Flesch-Kincaid Grade Level and Flesch Reading Ease. For class 1, which is unstable, the precision of the model is 0.86, and a recall of 0. 90, giving an F1 score of 0.88. It is important to note that the macro average of these metrics is equal to 0.89, which indicates reasonable pass rates for both classes. Another measure that is also equal to 0 is the weighted average of 0.89, which shows the model maintains a fairly steady performance over the set.

The overall accuracy found in the RF model is higher compared to previous models with a value of 0. 99. Since class 0 has no true positive instances, precision is 0 for this class. 91, recall is 0. 90, and F1-score is 0. 90, which means that the model is able to select proper instances that will remain stable and give the correct prediction. In class 1, the precision is 0. 85, recall is 0. 90, and F1-score is 0.88, slightly outperforming XGBoost by increasing the capacity of detecting unstable instances. The average of macro precision, recall, and F1-score is 0. 90, 0. 89, and 0. 12 and 89, respectively, which shows more attendance balance between both classes. The weighted averages, which incorporate the distribution of instances, remain at 0.90, 0.89, and 0.89, which attest to the reliability and usefulness of the proposed model.

The combined XGBoost-RF model performs even better than the standalone XGBoost and RF models by attaining a high accuracy of 0. 99. It has a zero false positive rate for class 0, meaning that it has a perfect precision of 1. 00, showing that all of the stable instances that we predicted are correct and that recall is 0. 99 as well as an F1-score of 0.99. This shows how the model can accurately determine the stable state, which is an important feature of the model.

For class 1, precision is 0. 96, and recall is 1. For the actual unstable instances, it has an accuracy of 00, meaning that all the actual unstable instances are correctly captured. The F1-score is 0. 98. As for the macro average of the hybrid model, the precision is 0, the recall is 0, and the F1-score is 0. 98, 0. 99, and 0. 99 in the case of the model compared to 87 and 93, respectively, in case of the two different classes, which shows the model has better recognition capability in both the classes. All the weighted average values were equal to 0. 99%, thereby ascertaining the high reliability of the model in determining the stability of the power system.

The result of the comparison shows that both XGBoost and RF models offer good predictive capabilities; however, the hybrid XGBoost-RF model offers perfect predictive performance in terms of precision, recall, and the F1-score. It can be seen that the performance of the hybrid model is better in terms of accuracy, and the model is balanced for the stable as well as unstable classes, which makes it the most suitable model for the prediction of transient stability in the IEEE New England 39-bus test system. This shows that it can effectively be used in areas that need high accuracy and reliability in managing power system stability.

The most important features that affect the accuracy of the transient stability prediction model in the IEEE New England 39-bus test system are presented in the form of a bar chart in Figure 5. Feature importance in the context of machine learning refers to the significance of the features for building decision trees in the context of the ensemble model. The feature importance in the XGBoost-RF hybrid model is based on the XGBoost and Random Forest algorithms, and with the help of both algorithms, these important features are accurately identified and weighted.

The features captured in the chart are ordered based on the importance score, which estimates the impact of the feature on the prediction error. The features with the highest coefficients are more important for the model, as they allow the stability or instability of states to be determined. It is expected that the most relevant features are rotor speed

(WmGx), rotor angle deviation (DThetaGx), stator voltage (VtGx), as well as active and reactive power values before and after the fault (PfvGx, PlvGx, QfvGx, QlvGx). It also aids in knowing which factors of the power system contribute most to transient stability, thus enabling the monitoring of the system and the formulation of control measures. With regard to the features that are most salient, engineers and researchers can direct their attention to the most relevant aspects of the power system, which might result in increased stability and reliability.

Table 4. Performance metrics comparison of XGBoost, RF, and XGBoost-RF hybrid models

Model	Class	Precision	Recall	F1-Score
XGBoost	0	0.95	0.89	0.89
	1	0.86	0.90	0.88
	Accuracy	-	-	0.89
	Macro avg	0.88	0.89	0.89
	Weighted avg	0.89	0.89	0.89
RF	0	0.91	0.90	0.90
	1	0.85	0.9	0.88
	Accuracy	-	-	0.99
	Macro avg	0.90	0.89	0.89
	Weighted avg	0.90	0.89	0.89
XGBoost-RF	0	1.00	0.99	0.99
	1	0.96	1.00	0.98
	Accuracy	-	-	0.99
	Macro avg	0.98	0.99	0.99
	Weighted avg	0.99	0.99	0.99

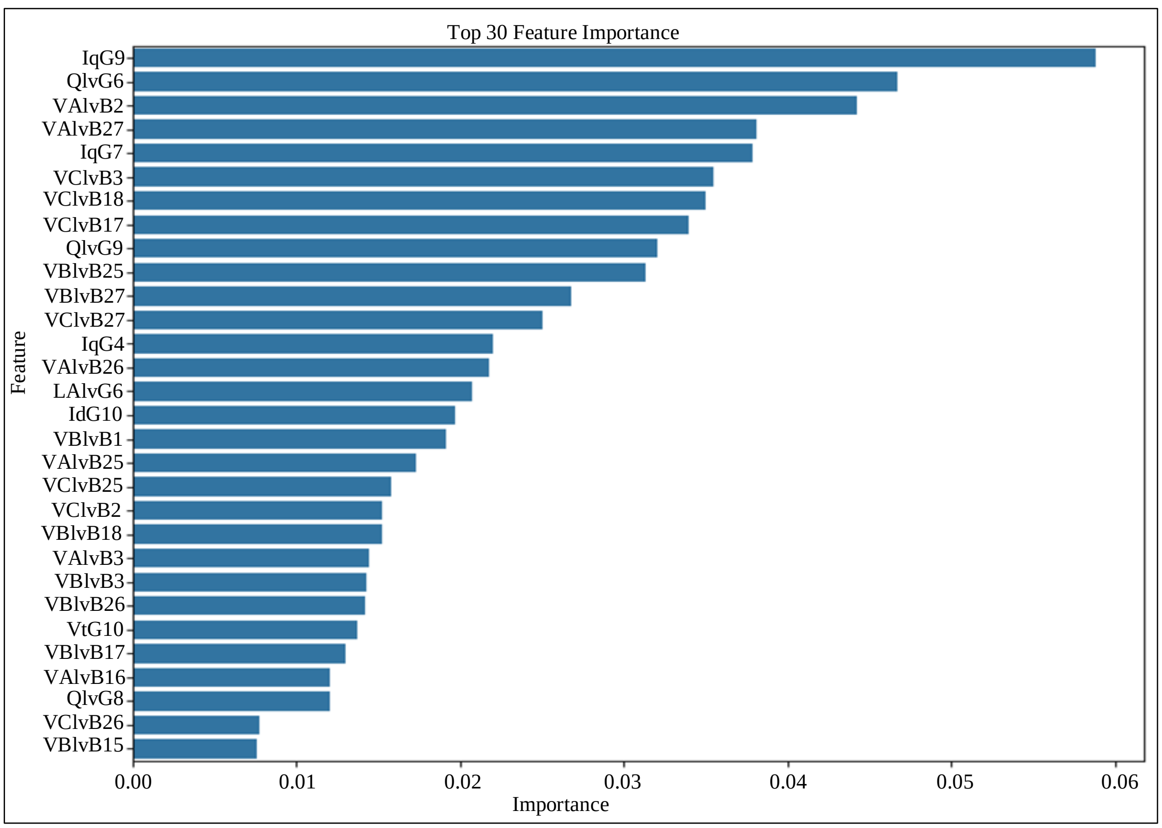

Fig. 5 The top features of importance in the prediction model

Figure 6 represents the ROC curve of the XGBoost-RF hybrid model for the prediction of the transient stability of the IEEE New England 39-bus test system, as shown in the figure below. The ROC curve is one of the most common and basic statistical measures used for the assessment of binary classification models. It maps TPR against FPR for different threshold levels, presenting a holistic picture of the model. In this figure, the hybrid model is proved to be close to the top left corner of the ROC curve plot, which means the model has almost perfect classification performance. The TPR (sensitivity or recall) indicates the share of actual positives that are classified correctly by the model. In contrast, the FPR (1-specificity) indicates the share of actual negatives that are classified as positive.

The nature of the curve, especially at the top left corner, where it almost touches the axes, depicts an almost perfect classifier with a relatively high AUC. An AUC close to 1 indicates that the model is extremely effective in properly classifying the stable and unstable states in the power system with overall low false positive rates and high true positive rates. The ROC curve of nearly 90 degrees is enough to support the reliability and accuracy of the model in identifying the transient stability of the power systems and thus can be considered a good tool for power system stability analysis. The ROC curve showing high performance proves the efficiency of the proposed hybrid method, combining the best features of the XGBoost and Random Forest algorithms.

Figure 7 shows the Precision-Recall (PR) curve obtained for the XGBoost-RF hybrid model, which was applied in predicting the transient stability of the IEEE New England 39-bus test system. PR curve is considered to be more effective for analyzing the performance of binary classifiers, especially in the case of large amounts of unbalanced data. It graphs Precision, which is the proportion of the total number of positive predictions made by the model to the total number of actual positive cases, against Recall, which is the ratio of the total number of correct positive predictions to the total number of actual positive cases for different threshold levels.

In this figure, the PR curve for the hybrid model appears to have excellent performance, with an Average Precision (AP) of 1. This curve demonstrates that when Recall is 1, meaning all actual positives are included in the sample, Precision is still extremely high, near 0. 99. This means that for all the true positive instances, the model does not only identify them but also incurs very minimal false positive errors. The high precision at full recall indicates that the proposed model is very reliable for identifying the stable and unstable states that exist in the power system without having to sacrifice the accuracy of the model. A high AP score and the shape of the PR curve, which is in its favour, proved that the proposed hybrid model strikes the right balance between precision and recall and, therefore, can be recommended for use in predicting the transient stability of power systems. This performance metric again validates that the proposed XGBoost-RF hybrid model outperforms the other models and proves its ability to predict accurate and reliable results, which is more important for balancing the stability and reliability of the power systems.

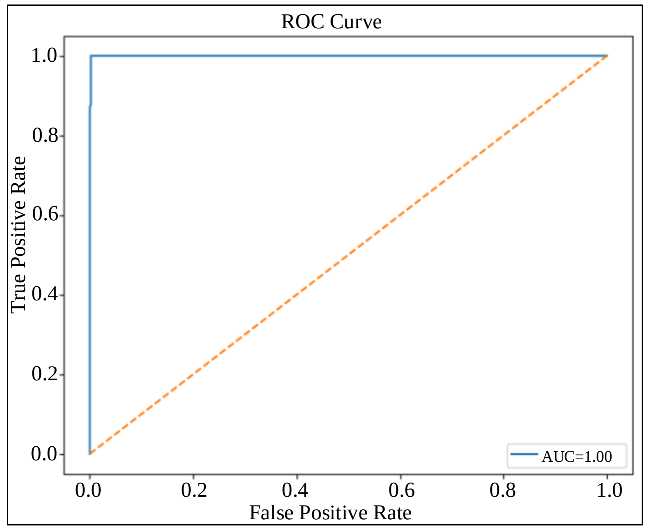

Fig. 6 Binary classifier performance evaluation of the model using receiver operating characteristic curve

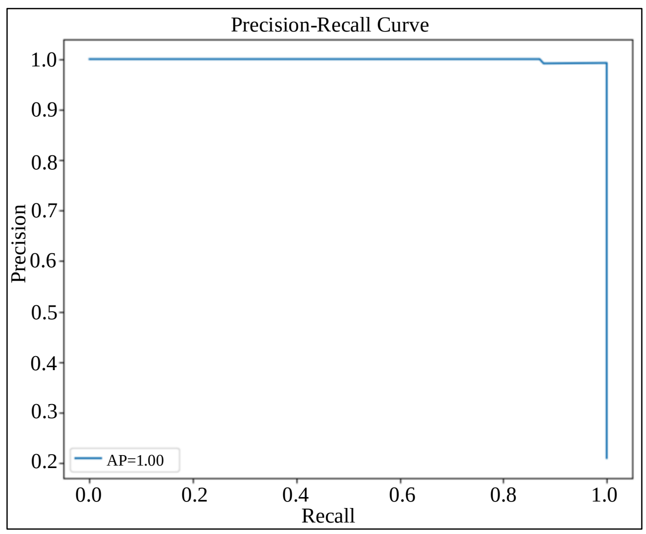

Fig. 7 Precision-recall curve of the model

Figure 8 describes the performance of the XGBoost-RF hybrid model in terms of the stability of the IEEE New England 39-bus test system in terms of the confusion matrix. The model achieved high accuracy by classifying 488 of 493 stable instances correctly (True Positives) and all the 131 unstable instances as False Positives (True Negatives). It committed only five classification mistakes of assigning stable instances as unstable (False Positives), while it did not misclassify any unstable instances as stable (False Negatives). This shows a high value of True Positives and True Negatives, proving the efficiency of the model and the low False Positives and False Negatives. The absence of false positives and no false negatives also indicates the effectiveness of the model and its importance in maintaining the stability of the power supply by detecting instabilities.

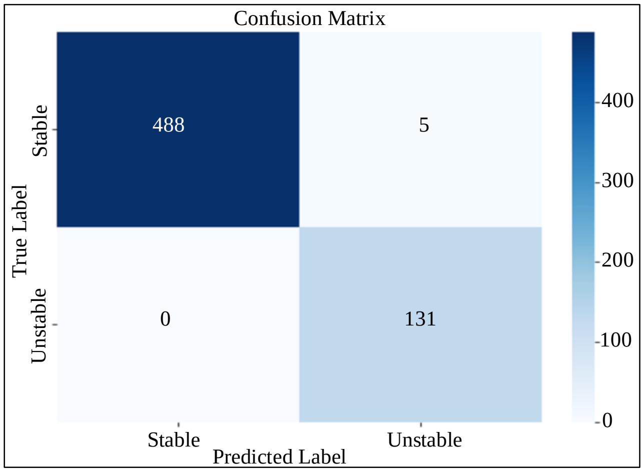

Fig. 8 Confusion matrix heatmap

Figure 9 captures the fluctuation in the training and validation of the XGBoost-RF model in relation to a specific hyperparameter. The graph shown below depicts how the model performance behaves when this parameter is varied, providing information about the right tuning to be done in order to maximize the generalization of the model to unseen data. The training score tends to be initialized high and may slightly drop as the parameter value increases; on the other hand, the validation score initially increases, reaches the maximum, and then may decrease, which implies overfitting in case of a high parameter value. Finding the highest value of the validation score assists in defining a proper parameter to reduce overfitting while increasing the model's prediction capability and robustness. This analysis is important in enhancing the model and making it perform better in the IEEE New England 39-bus test system data.

In Figure 10, the bar chart represents the identification of the 30 most critical features employed by the XGBoost-RF mixed model for the transient stability assessment of the IEEE New England 39-bus test system. For each feature, the probability density function is illustrated both for class 0, which corresponds to the stable state, and class 1, which corresponds to the unstable state, so that a comparison can be made of how these features differ between the two classes. By analyzing these distributions, it is possible to understand what differentiates the model and which features it uses to make the predictions. This visualization is useful in understanding the aspects that have high separability between the classes and, hence, are valuable in the classification task. It also provides an understanding of the underlying patterns of power system stability.

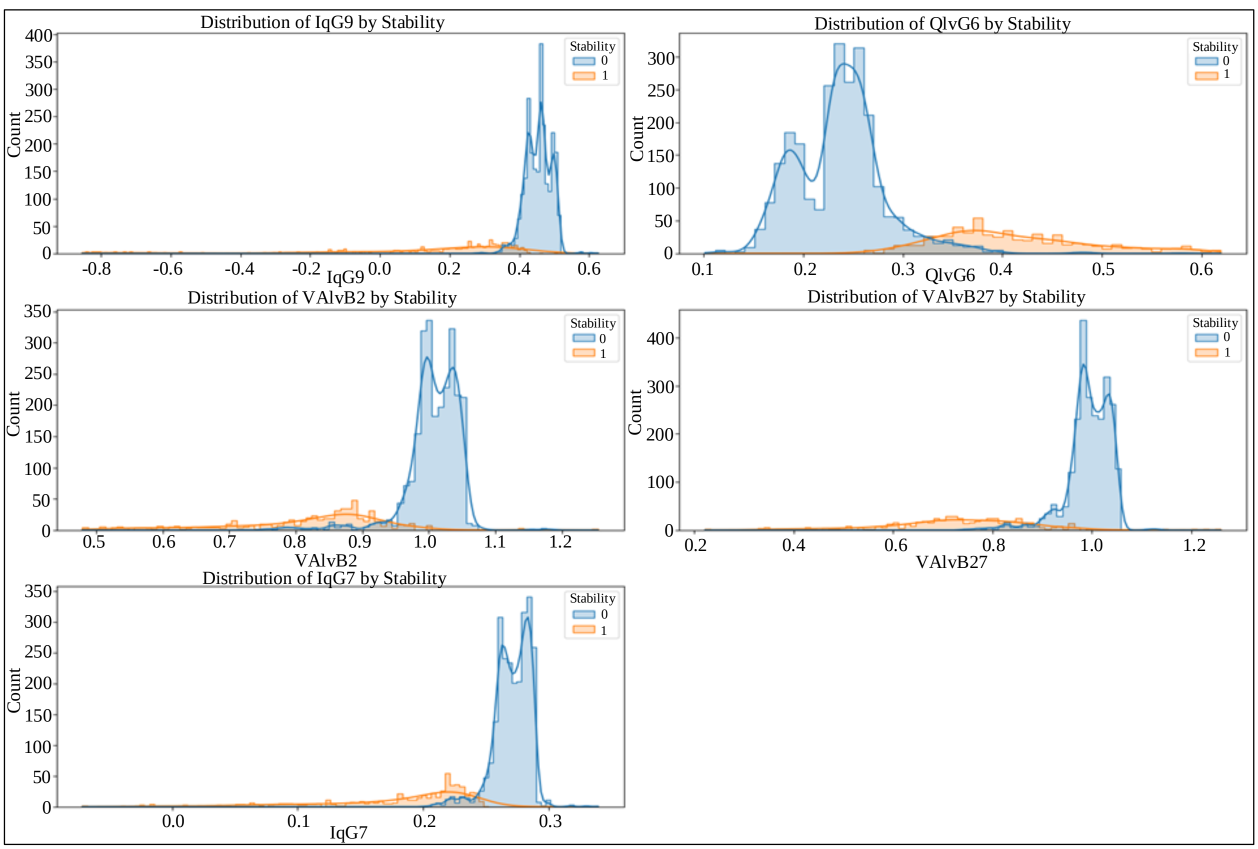

Fig. 9 Training and validation scores change with a specific parameter

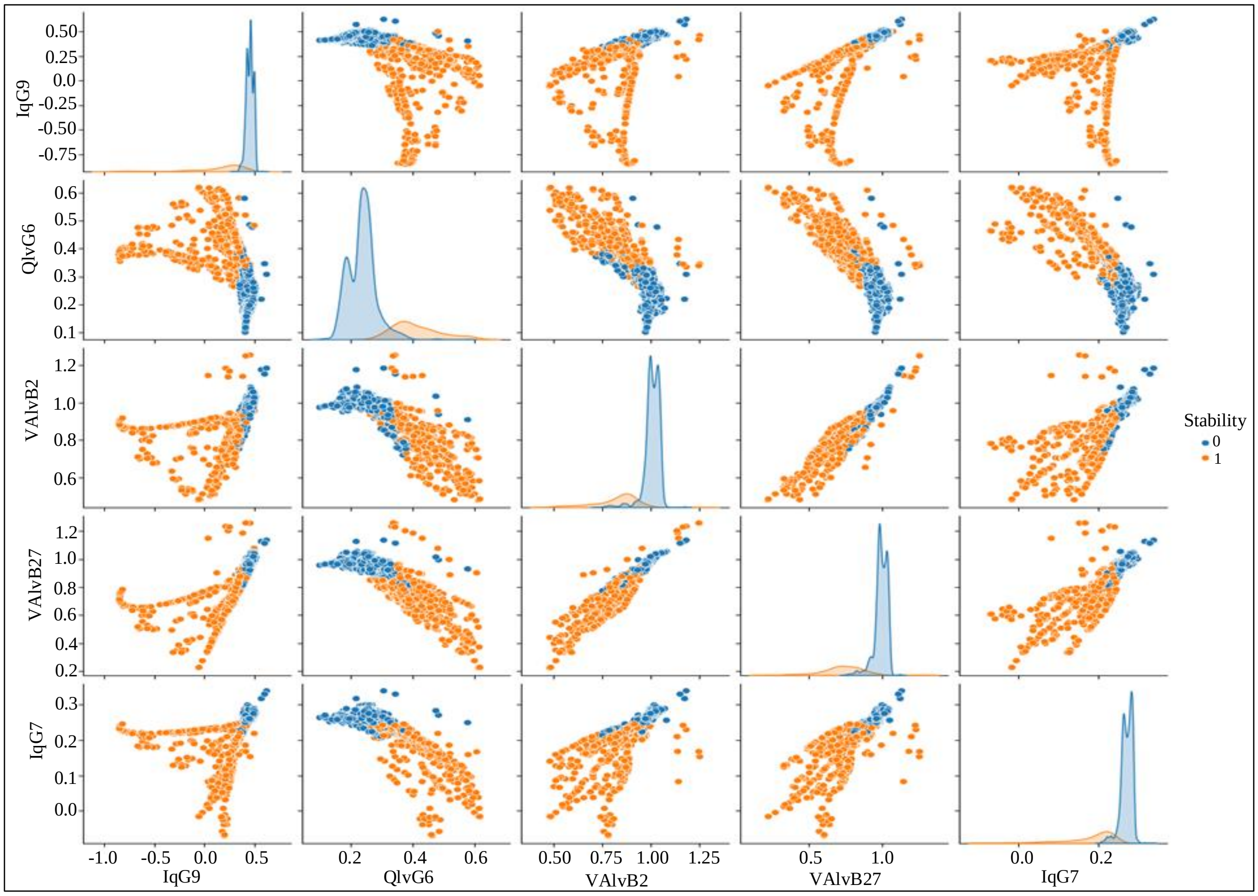

Fig. 10 The distribution of the most important features for the two classes

4. Conclusion

The current work shows how machine learning models, such as XGBoost, RF, and a combined XGBoost with RF model, can be used to forecast transient stability in the IEEE New England 39-bus test system. The chosen algorithms' ability to classify stable and unstable states was measured using such parameters as precision, recall, F1-score, and accuracy. The overall accuracy of the model was 89%, proving that the XGBoost model is quite stable and has a good predictive analysis.

However, while the RF model enhanced a bit in identifying unstable cases, the accuracy rate was higher at 99%. However, the proposed combined XGBoost-RF model performed even better than the individual models of XGBoost and RF with an accuracy level of 99%. This hybrid model outperformed the individual models, based on precision, recall, and F1-scores, which show the stable and unstable classes' performance. The study shows that the XGBoost-RF model is the most appropriate for transient stability prediction, hence its recommendation for power systems. The paper also showed that features affecting the model accuracy of the system are the rotor speed, rotor angle deviation, stator voltage, as well and active and reactive power values before and after the fault. Thus, applying modern approaches to machine learning, this research supports the continuous development of power grid management reliability and efficiency. The findings from this research could be useful in the enhancement of stability assessment techniques that assist in the stability and secure operation of contemporary power systems in view of growing demands and risks. Possible directions for future work include the use of real-time data feeds and more sophisticated deep learning methodologies to enhance the accuracy and timeliness of the stability assessment, which would be helpful in developing more complex and adaptive stability assessment frameworks for smart grid systems in the ever-changing environment.

Funding Statement

This study was performed with the support of Jamhuriya University of Science and Technology, Mogadishu, Somalia.

Acknowledgments

The authors would like to extend their gratitude to the engineering faculty of Jamhuriya University of Science and Technology for their invaluable assistance and support throughout this research. Special thanks are due to the Department of Electrical Engineering for their guidance and contributions. The authors also appreciate the cooperation and encouragement from their colleagues and peers.

References

[1] Shitu Zhang, Zhixun Zhu, and Yang Li, "A Critical Review of Data-Driven Transient Stability Assessment of Power Systems: Principles, Prospects and Challenges," *Energies*, vol. 14, no. 21, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Petar Sarajcev et al., "Artificial Intelligence Techniques for Power System Transient Stability Assessment," *Energies*, vol. 15, no. 2, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Minghua Chen et al., "XGBoost-Based Algorithm Interpretation and Application on Post-Fault Transient Stability Status Prediction of Power System," *IEEE Access*, vol. 7, pp. 13149-13158, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] Mihail Senyuk et al., "Power System Transient Stability Assessment Based on Machine Learning Algorithms and Grid Topology," *Mathematics*, vol. 11, no. 3, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Jun Liu et al., "An Improved Power System Transient Stability Prediction Model Based on mRMR Feature Selection and WTA Ensemble Learning," *Applied Sciences*, vol. 10, no. 7, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Paul Andrew Vlahutin, "*Power System Transient Stability Assessment by Pattern Recognition*," Dissertations & Theses, Western Reserve University, 1971. [Google Scholar]

[7] Yanzhen Zhou et al., "A Hierarchical Method for Transient Stability Prediction of Power Systems Using the Confidence of a SVM-Based Ensemble Classifier," *Energies*, vol. 9, no. 10, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[8] Luyu Ji et al., "Using Trajectory Clusters to Define the Most Relevant Features for Transient Stability Prediction Based on Machine Learning Method," *Energies*, vol. 9, no. 11, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[9] Albert Poulose, and Soobae Kim, "Transient Stability Analysis and Enhancement Techniques of Renewable-Rich Power Grids," *Energies*, vol. 16, no. 5, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] A. Kunac, and P. Sarajcev, "Ensemble Learning Approach to Power System Transient Stability Assessment," *2020 5th International Conference on Smart and Sustainable Technologies (SpliTech)*, Split, Croatia, pp. 1-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Yuqiong Zhang et al., "A Power System Transient Stability Assessment Method Based on Active Learning," *The Journal of Engineering*, vol. 2021, no. 11, pp. 715-723, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Qilin Wang, Chengzong Pang, and Hashim Alnami, "Transient Stability Prediction Based on Long Short-term Memory Network," *2021 North American Power Symposium (NAPS)*, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ramraj Santhanam et al., "Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets," *International Journal of Control Theory and Applications*, vol. 9, no. 40, pp. 651-662, 2016. [Google Scholar] [Publisher Link]

[14] Y. Wang et al., "A Hybrid Ensemble Method for Pulsar Candidate Classification," *Astrophysics and Space Science*, vol. 364, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[15] Emil D. Attanasi, and Timothy C. Coburn, "Random Forest," *Encyclopedia of Mathematical Geosciences*, pp. 1182-1185, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Harmandeep Kaur, and Munish Kumar, "On the Recognition of Offline Handwritten Word Using Holistic Approach and AdaBoost Methodology," *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 11155-11175, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Jabed Al Faysal et al., "XGB-RF: A Hybrid Machine Learning Approach for IoT Intrusion Detection," *Telecom*, vol. 3, no. 1, pp. 52-69, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Petar Sarajcev et al., IEEE New England 39-Bus Test Case: Dataset for the Transient Stability Assessment (2.0), Dataset, Zenodo, 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7350829