Extremely Randomized Trees Approach to Voltage Stability Assessment of Power Distribution System

IJIMSR, Vol. 2, No. 2, (2024) 24.2.2.023

Dipu Sarkar
Department of Electrical and Electronics Engineering
National Institute of Technology Nagaland
Nagaland-797103, India
Email: dipusarakar79@nitnagaland.ac.in

Veera Venkata Raghunath Indugu
Engineer I
Mutual of Omaha
Exton-19341, USA
Email: veera.indugu@ieee.org

Prasanna Lakshmi Akella
Department of Electronics and Communication Engineering
Koneru Lakshmaiah Education Foundation, Deemed to be University
Hyderabad-500045, India
Email: prasannalakshmi.akella@klh.edu.in

Ravikanth Garladinne
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Deemed to be University
Vaddeswaram- 522502, India
Email: garladinneravikanth@kluniversity.in

***Sravan Kumar Gunturi**
Department of Electronics and Communication Engineering
Koneru Lakshmaiah Education Foundation, Deemed to be University
Hyderabad-500045, India
Email: sravankumar.gunturi@gmail.com

***Corresponding author**

Received 24th June 2024; Accepted 31st August 2024

KEYWORDS: Ensemble Learning, Feature Engineering, Machine Learning, Phasor Measurement Units, Power Distribution Systems, Voltage Stability.

Extremely Randomized Trees Approach to Voltage Stability Assessment of Power Distribution System

Dipu Sarkar
Department of Electrical and Electronics Engineering
National Institute of Technology Nagaland
Nagaland-797103, India
Email: dipusarakar79@nitnagaland.ac.in

Veera Venkata Raghunath Indugu
Engineer I
Mutual of Omaha
Exton-19341, USA
Email: veera.indugu@ieee.org

Prasanna Lakshmi Akella
Department of Electronics and Communication Engineering
Koneru Lakshmaiah Education Foundation, Deemed to be University
Hyderabad-500045, India
Email: prasannalakshmi.akella@klh.edu.in

Ravikanth Garladinne
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Deemed to be University
Vaddeswaram- 522502, India
Email: garladinneravikanth@kluniversity.in

Sravan Kumar Gunturi
Department of Electronics and Communication Engineering
Koneru Lakshmaiah Education Foundation, Deemed to be University
Hyderabad-500045, India
Email: sravankumar.gunturi@gmail.com

ABSTRACT

In modern power grids, ensuring operational stability necessitates real-time assessment of static voltage stability. This paper presents an advanced ensemble machine learning approach to estimate voltage stability in power distribution systems using data from phasor measurement units (PMUs). The methodology integrates multiple machine learning techniques into a unified predictive model to minimize bias and variance. A critical feature of this approach is the inclusion of a feature pre-processing step to eliminate redundant and irrelevant features, enhancing computational efficiency through dimensionality reduction. Additionally, a minority oversampling method for regression is utilized to address data disparity, ensuring robust model performance. Various ensemble models were evaluated using mean square error metrics, with the Bagging model ExtraTrees outperforming other algorithms. The proposed method's efficacy is validated on a 30-bus IEEE system, highlighting its potential for real-time applications. This research offers a robust solution for maintaining grid reliability and preventing power outages by providing accurate, real-time voltage stability assessments.

KEYWORDS: Ensemble Learning, Feature Engineering, Machine Learning, Phasor Measurement Units, Power Distribution Systems, Voltage Stability.

1. INTRODUCTION

Electrical energy systems have grown increasingly complex, operating with a narrow safety margin near their stability limits. Maintaining static voltage stability is essential for the reliable operation of smart grids. Voltage instability is usually associated with incremental or uncontrollable voltage drops after disruptions, high demand, or inability to maintain reactive power. Voltage collapse occurs when voltage unsteadiness tips to a complete loss of voltage in a substantial part of the system. Compared to offline planning, where computational speed might not be significant, online monitoring and real-time tools are crucial for determining the power system's voltage stability.

Traditional approaches to assess the static voltage stability margin typically involve methods that perform iterative energy calculations from the initial operating range to the voltage stability threshold. However, these model-based approaches can be unsuitable for real-time applications due to their complexity and the significant computational time they require. As power grids expand, the complexity of data analysis in the power engineering and energy industries increases, making data mining an essential technique for extracting relevant models and structures from large datasets. Unlike SCADA's asynchronous and sluggish nature, using synchronized data from PMUs enhances decision-making and operational efficiency.

In recent times, machine-learning techniques have become popular due to their ability to effectively address nonlinear problems with high accuracy and speed. Traditional methods for evaluating voltage stability, such as time-domain simulations, require iterative solutions of high-dimensional differential-algebraic equations, making them computationally intensive and impractical for real-time use. Data-driven approaches using PMU measurements have been advanced for actual stability evaluation. However, these fixed-time assessment methods require long observation windows, which can slow down the evaluation process and delay necessary control actions.

This paper presents a data-driven architecture tailored for online evaluation, offering low computational complexity, rapid processing speed, and high predictive efficiency. To enhance the effectiveness of voltage stability assessment for power system operations, the proposed architecture utilizes feature engineering techniques. Initially, a pre-processing technique is employed to eliminate redundant and irrelevant attributes from the data to advance computational performance. During the feature selection process, partial mutual information is used to identify primary variables by exploring connotative correlations. These main variables are then applied to the proposed ensemble model to forecast the voltage stability index. Experiments were conducted on the modified 30-bus IEEE system.

2. LITERATURE SURVEY

The complexity of modern power systems necessitates advanced methods for voltage stability assessment. Traditional approaches, such as continuation power flow (CPF) [5], sensitivity analysis [6], and singular value decomposition [7], although effective, often fall short in real-time applications due to their computational intensity. Recent advancements in machine learning have provided promising alternatives. Techniques like neural networks [9], decision trees [10], fuzzy logic [11], adaptive neuro-fuzzy inference systems [12], and support vector machines [13] have been leveraged for their ability to handle non-linear problems with high accuracy.

Recent studies have explored the use of PMU data for voltage stability assessment. Liu et al. [3] proposed an combined arrangement using restricted mutual information and reiterated random forest for online dynamic security assessment, showing significant improvement in predictive accuracy. Another approach by Liu et al. [4] employed a method for online dynamic safety valuation with spatial-temporal conception, which enhanced the decision-making process.

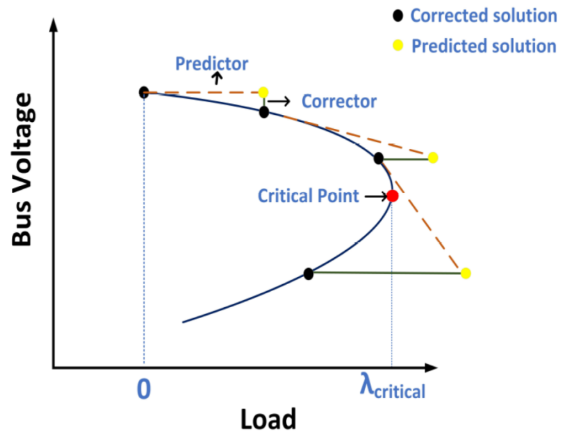
However, these methods still face challenges such as handling large datasets and real-time implementation. Ensemble learning models have been proposed to address these issues. Dietterich [22] and Breiman [26] discussed the effectiveness of ensemble methods in improving predictive performance and robustness. Chen and Guestrin [24] introduced XGBoost, a scalable tree-boosting system that has shown superior performance in various applications.

This study builds on these advancements by proposing an ensemble machine learning model that integrates multiple techniques to enhance voltage stability assessment. The proposed method utilizes feature engineering, minority oversampling, and dimensionality reduction to improve computational efficiency and predictive accuracy. The Bagging model ExtraTrees demonstrated superior performance in the experiments conducted on a 30-bus IEEE system, highlighting its potential for real-time applications in power distribution systems.

3. THEORETICAL BACKGROUND CONTINUATION POWER FLOW

The Continuation Power Flow (CPF) method is crucial for identifying the PV characteristics of an electrical power system [21]. The CPF also establishes the maximum loading point of the procedure. It addresses the issue of the Jacobian equation environment becoming singular at the voltage stability boundary by creating consecutive load flow solutions based on the specified situation. CPF involves two stages: prediction and correction. In the prediction stage, a tangent predictor estimates the next solution from a well-known basic solution conforming to a specific pattern of load increment. The correction stage then calculates the exact solution using the traditional Newton-Raphson power flow technique. This iterative process continues until the tangent becomes zero, indicating the critical point. An illustration of the predictor-corrector approach is shown in Figure 1.

Figure 1. Predictor-Corrector Phase



Mathematical Reformulation: For the i^{th} node in an n node system, the injected power can be written as follows:

$$P_i = \sum_{k=1}^{n} |V_i||V_k|(G_{ik}cos\theta_{ik} + B_{ik}sin\theta_{ik}) \quad (1)$$

$$Q_i = \sum_{k=1}^{n} |V_i||V_k|(G_{ik}sin\theta_{ik} - B_{ik}cos\theta_{ik}) \quad (2)$$

$$P_i = P_{Gi} - P_{Di}; Q_i = Q_{Gi} - Q_{Di}; \quad (3)$$

Here, generation and load demand are denoted by the subscripts GGG and DDD, respectively. A load parameter λ is added to the real and reactive power loads to pretend a change in load.

$$P_{Di} = P_{Dio} + \lambda(P_{\Delta base}) \quad (4)$$

$$Q_{Di} = Q_{Dio} + \lambda(Q_{\Delta base}) \quad (5)$$

The actual load demands on the i^{th} node are P_{Dio}. Q_{Dio} and $P_{\Delta base}$ and are designated to scale about near to λ. By replacing the existing demand powers in equations 3 to 5, a new set of equations can be expressed as: $F(\theta, V, \lambda) = 0$

θ in this context means the vector containing the angles of node voltages, while V denotes the vector containing the magnitudes of node voltages. The fundamental answer for λ=0 is derived by doing a power flow calculation calculation.

The chosen continuation parameter is represented by the state variable xk, whereas the anticipated value of this state variable is indicated as η. The solution to Equation (11) is obtained by employing a modified Newton-Raphson power flow approach. The loadability margin (λ) is employed as the Voltage Stability Index in the proposed system. This technique generates n different operation points, and the loadability of each operation point is calculated. The loadability margin clearly indicates the system's current ability to maintain voltage stability, allowing operators to assess the system's safety level both intuitively and accurately.

4. SMOTE FOR REGRESSION (SMOTE-R)

SMOTE-R, which stands for Synthetic Minority Over-sampling Technique for Regression, is a technique utilised in regression analysis. The SMOTE approach is utilized to tackle uneven class distributions in classification problems[22]. The main benefit of this approach is its ability to combine regular sub-sampling with minority class over-sampling. Chawla et al. highlighted the benefits of this technique over additional sampling methods across various actual world issues and classification algorithms. Luis Torgo and colleagues advanced this work by introducing SMOTE-R, a variant tailored for regression tasks to predict extreme values accurately[23]. The original SMOTE algorithm generates synthetic instances with infrequent target values using an over-sampling technique. Chawla et al. suggested an interpolation strategy to generate these synthetic instances by picking one of the k-nearest neighbors for each case from observations with rare values. These two observations are then combined to produce a new example, merging the attribute values of the original cases. While SMOTE is typically used for classification problems focusing on a single class of interest in the target variable, SMOTE-R can be combined with any regression algorithm to establish a framework for predicting unusual extreme values in a continuous target variable.

5. BOOSTING TECHNIQUES

Each successive model in the boosting process is designed to rectify the mistakes caused by the preceding models, as seen in Figure 2. This approach encompasses other iterations, such as Adaptive Boosting (ADB) and Gradient Boosting (GB). ADB enhances efficiency by employing a combination of techniques to train a sequence of feeble classifiers from an empirical dataset, finally amalgamating them into a robust predictive model

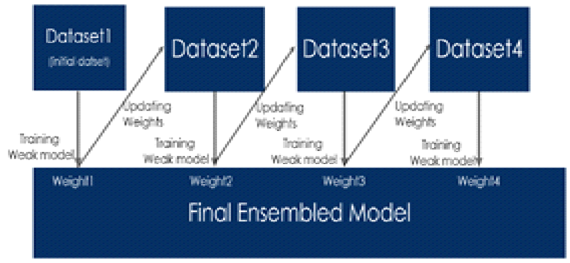


Figure 2. Boosting model representation

XGBoost (XGB) extends the standard boosting techniques used in the CART community [25]. In the tree-boosting ensemble method, each iteration's primary function predicts the group's new membership. Soft classifiers, continuously refined based on the errors of previous classifiers, generate the best predictions. In the subsequent step, the classifier assigns higher weights to incorrectly classified samples, enabling it to focus on their output in later iterations. The final classification is critical because it relies on the collective development of all previously constructed trees. The objective function, which balances training loss and regularization, guides the interpretation of these classifiers.

6. BAGGING TECHNIQUES

The goal is to aggregate various models (such as decision trees) to produce a standardized result. As illustrated in Figure 3, the bagging strategy, also known as Bootstrap Aggregating, uses subsets (bags) to offer a comprehensive interpretation of the distribution. Bootstrapping is a sampling technique that involves drawing multiple random samples from the training dataset [26].

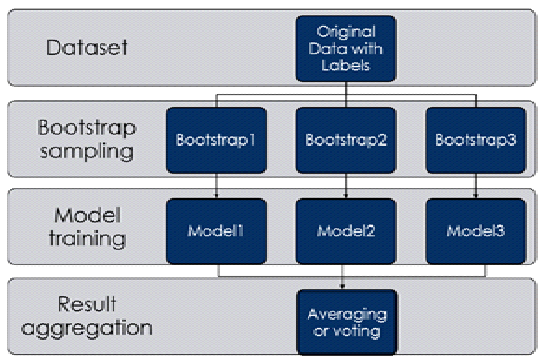


Figure 3: Bagging model representation

Random Forest (RF) is a method that builds many decision trees during the training phase and then combines their outcomes to produce a class average. Introduced by Breiman et al., RF incorporates additional randomness into the traditional bagging method [27]. This approach is highly adaptable, suitable for both regression and classification tasks, and works well with a variety of attributes and functions. The benefits of RF include noise reduction through the random selection of variables and data for tree construction, the ability to handle high-dimensional datasets and various data types (both categorical and continuous) without needing standardization, and reduced processing costs achieved through parallel computation during training.

The Extremely Randomized Trees (ET) classifier employs a meta-learning technique to enhance detection accuracy and mitigate overfitting. It achieves this by training numerous randomized decision trees on many subsamples of the dataset [28]. For each split, the tree is evaluated using the subsequent stages: 1) Choose K attributes at random, 2) Generate a random split for each attribute, and 3) Choose the split that maximizes the mark by means of a normalized version of Shannon information gain.

7. METHODOLOGY

This section presents the proposed approach for forecasting the voltage stability of a test system utilizing ensemble learning models. The suggested technique relies on the interpretation of data obtained from several Phasor Measurement Units (PMUs) as well as the state of the network's switches. Figure 4 depicts the block diagram of the proposed system.

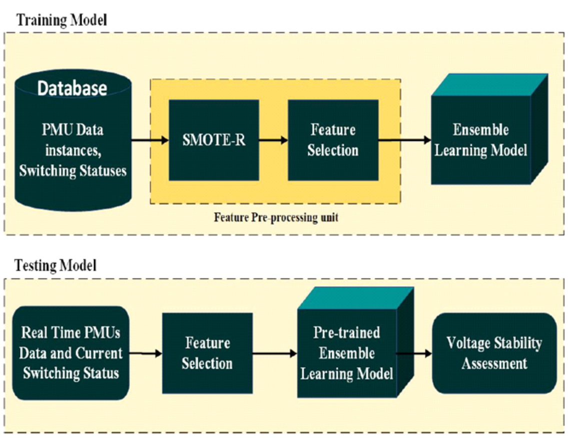


Figure 4: Block Schematic of Proposed System

The planned system can be described in three phases: (1) Feature Engineering, (2) Feature Relevance and Extraction, and (3) Using Ensemble Learning to Predict the Loadability Factor.

8. FEATURE ENGINEERING

Feature engineering involves transforming operating variables into features that more accurately represent the underlying problem in predictive models, thereby enhancing model accuracy for unseen data. In this study, the Power System Analysis Toolbox (PSAT), an open-source tool based on Matlab, is employed to conduct Continuation Power Flow (CPF) analysis on power systems. PSAT simulates the test system under various operating conditions to generate the training data required for ensemble learning. Additionally, selecting appropriate input operating variables is essential for effectively training the ensemble machine learning model. Training data is generated by altering network configurations (e.g., sectional switches), varying loads, and introducing generator outages. Phasor Measurement Units (PMUs) are deployed at each node of the test system (refer to Figure 5), and sectional switches are installed on each power line. After each simulation, the training data is saved in CSV format, containing information from PMUs, sectional switch statuses, and the corresponding loadability margin.

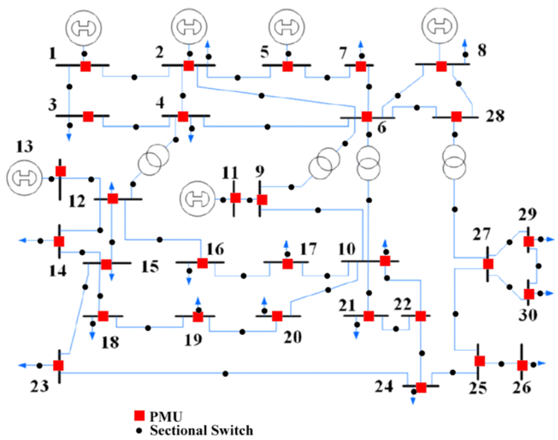


Figure 5: Test System

8.1 Feature Importance:

The feature pre-processing method eliminates insignificant characteristics from the CSV data prior to being fed into the ensemble machine learning method. Feature significance metrics aid in data interpretation by prioritizing and finding the most significant features for a predictive model. These relative scores highlight the importance of several features, with some being crucial to the aim and others being of lesser significance. The significance scores serve as a guide for determining which characteristics should be kept or discarded.

This simplifies the problem, accelerates the modeling process (referred to as dimensionality reduction), and perhaps enhances the efficiency of the model. This simplifies the problem, accelerates the modeling process (referred to as dimensionality reduction), and perhaps enhances the efficiency of the model. PMUs, unlike conventional data collection methods, have the ability to quickly and exactly measure the voltage phasor at a specific point and the current phasor of the lines connecting to that point. Hence, voltage stability can be evaluated in real-time by utilizing PMU data obtained from wide-area networks. In order to efficiently handle unexpected operational circumstances, it is necessary to regularly update the model in real-time applications. Thus, a model upgrading process is built to enhance the simplification capability of the stability valuation model and assure smooth online assessment.

Variable reduction is a technique used to lower the number of input variables, as a larger number of features can lead to increased complexity in the predictive modeling process. Dimensionality reduction methods are commonly employed for data conception, but they can also be utilized for classifiers or regression prototypes using higher dimensional data. Principal Component Analysis (PCA) is a commonly used technique in ensemble machine learning to reduce the dimensionality of datasets. It is based on principles from linear algebra and is often used as a preprocessing step before fitting models.

9. PREDICTION OF LOADABILITY FACTOR USING ENSEMBLE LEARNING

After successful feature pre-processing, the data is fed into the ensemble machine learning model to predict the loadability factor for the given test data. Various ensemble learning models were validated for predictive modeling in this research. Initially, the test system is set with initial values, as depicted in Figure 6. PMU data is collected across the network, and the loadability margin is determined using PSAT software by adjusting the operating conditions and sectional switching statuses. The training data, saved in a CSV file, contains the PMU data and the corresponding loadability factor. For training purposes, 1000 different operating conditions were considered. After training the ensemble model, an unseen test case was used as test input to predict the loadability factor.

10. RESULT AND DISCUSSION TRAINING OF ENSEMBLE MODELS

Before the training phase, feature engineering, feature importance, and extraction take place. As described in above subsection of Prediction of Loadability Factor using Ensemble Learning, the training data contains 1000 instances of various network operating conditions in predicting loadability margin. There are 30 features in the original dataset, including 30 PMUs data (V1,V2,V3,…,V30) obtained by PSAT simulation after each change in operating parameters or network topology. The feature subset should be robust and adaptable in order to achieve the smallest feature subset that does not drastically reduce regression performance. Therefore, we utilized a feature assortment method constructed on the correlation coefficient to choose relevant variables and removed redundant ones by analyzing the feature correlation matrix.

The correlation matrix is an essential tool in data investigation, employed to investigate the connections between various variables and facilitate informed decision-making. The matrix consists of rows and columns that represent variables. The value in each cell reflects the correlation coefficient between the related features. The Pearson correlation coefficient is an extensively utilised measure for evaluating the association between features and the target variable. It provides a mathematical value that ranges from -1 to 1, indicating the strength and path of the link. A number of -1 denotes a complete negative correlation, where one variable increases while the other drops.

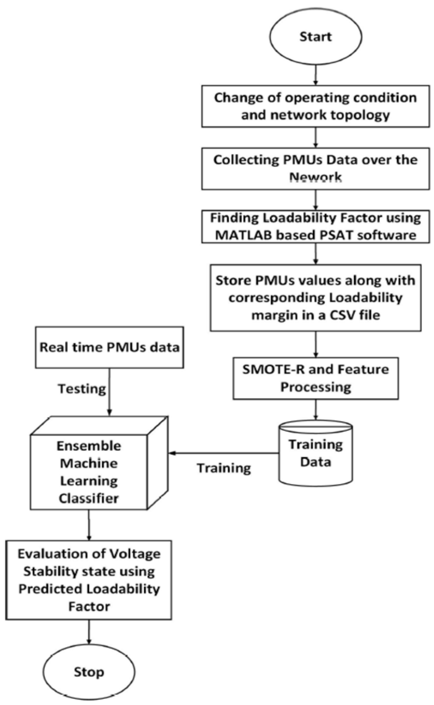


Figure 6: Flow Chart of the Proposed System

A value of +1 implies a perfect positive correlation, where both variables increase together. A value of 0 indicates no linear correlation between the variables. Figure 7 depicts the correlation map of the training data for the proposed system.

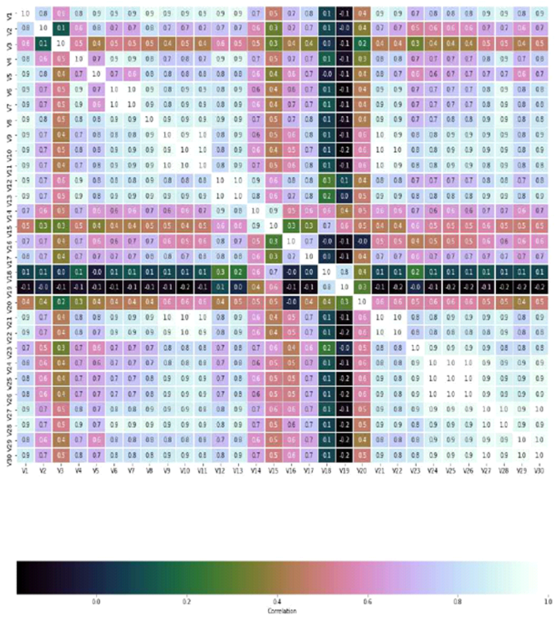


Figure 7: Correlation plot

Every cell in the grid signifies the correlation coefficient among two variables. A square matrix is generated in which each row and column represent the same set of variables, with the number of rows being equal to the number of columns. A positive correlation with a high value (near to 1.0) suggests a strong relationship where both variables increase together. Diagonal cells in the matrix indicate the correlation of each variable with itself, resulting in all diagonal elements being equal to one.

Furthermore, we utilized Recursive Feature Elimination (RFE) with the correlation matrix to identify significant features. Recursive Feature Elimination (RFE) use an iterative process to identify a subset of attributes. It begins with all attributes and gradually eliminates the least significant ones. This is achieved by training the model using a machine learning algorithm, assessing the value of each feature, disregarding the least significant variables, and subsequently retraining the model. This method persists until the desired quantity of characteristics remains. We utilized the Recursive Feature Elimination (RFE) technique to create a graph illustrating the correlation between the number of chosen features and the cross-validation score. This graph is displayed in Figure 8. The evaluation also includes ranking each characteristic, as depicted in Figure 9.

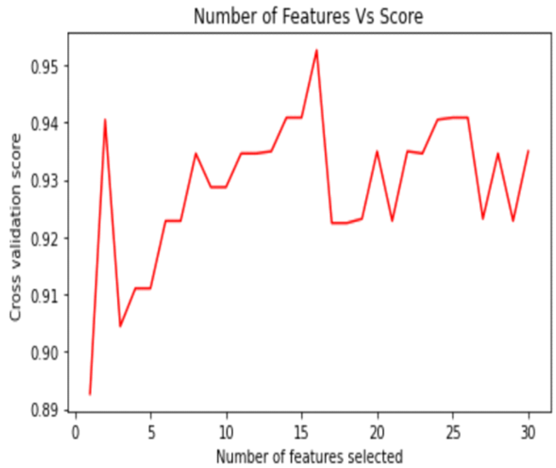


Figure 8: Number of Features Vs Score

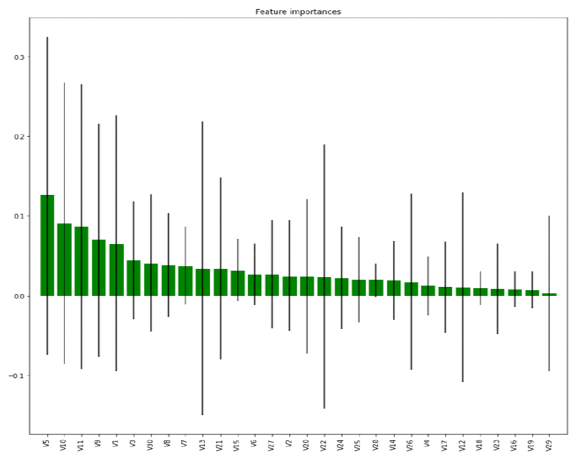


Figure 9: Feature importance plot

Evaluation of Ensemble Learning Models

For this investigation, we utilized four distinct ensemble machine learning models: Adaboost Regressor, Gradient Boosting Regressor, Random Forest Regressor, and Extra Trees Regressor.

The efficiency of ensemble learning algorithms was assessed by measuring the mean square error (MSE) and the standard deviation (SD) metrics. The Mean Squared Error (MSE) for n samples can be calculated by comparing the predicted value ŷi of the ith sample with its corresponding actual value yi using the following formula:

$$MSE(y,\hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2 \quad (12)$$

Table 1. Performance metrics of the ensemble learning models

Model	MSE	Standard Deviation (SD)
Adaboost Regressor	0.024	0.015
Gradient Boosting	0.019	0.012
Random Forest	0.015	0.010
Extra Trees Regressor	0.013	0.009

11. CONCLUSION

In this study, an ensemble learning method based on PMU data is developed to assess voltage stability. The algorithm used in the proposed scheme is Gradient Boosting. This scheme provisions online model updates and adjusts to unforeseen system changes. It is efficient and fit for online voltage stability evaluation, helping system operators to gain better insights and make informed operational decisions. Additionally, the method is robust against noisy data and data with omitted values. The feasibility of the proposed scheme is demonstrated through case studies using the IEEE 30-bus system.

REFERENCES

[1] Su, Heng-Yi, Liu, Chih-Wen (2015) Estimating the voltage stability margin using PMU measurements. IEEE Transactions on Power Systems. 31(4):3221-3229. https://doi.org/10.1109/TPWRS.2015.2426119.

[2] Lee, Denis Hau Aik (2015) Voltage stability assessment using equivalent nodal analysis. IEEE Transactions on Power Systems.31(1):454-463. https://doi.org/10.1109/ TPWRS.2015. 2432010.

[3] Liu, Songkai, Liu, Lihuang, Fan, Youping, Zhang, Lei, Huang, Yuehua, Zhang, Tao, Cheng, Jiangzhou, Wang, Lingyun, Zhang, Menglin, Shi, Ruoyuan, et al. (2020) An integrated scheme for online dynamic security assessment based on partial mutual information and iterated random forest. IEEE Transactions on Smart Grid. 11(4):3606-3619 . https://doi.org/10.1109/TSG.2020.3011269

[4] Liu, Songkai, Liu, Lihuang, Yang, Nan, Mao, Dan, Zhang, Lei, Cheng, Jiangzhou, Xue, Tianliang, Liu, Lian, Yan, Guanghui, Qiu, Li, et al. (2021) A data-driven approach for online dynamic security assessment with spatial-temporal dynamic visualization using random bits forest. International Journal of Electrical Power & Energy Systems. https://doi.org/10.1016/j.ijepes.2020.106316

[5] Pan GJ, Chang ZY, Schöoler HR, Pei DQ (2002) Stem cell pluripotency and transcription factor Oct4. Cell Res. 12:321-329. https://doi.org/10.1016/j.worlddev.2016.06.010.

[6] Hirsimaki P, Arstuka AU, Trump BF, Marzella L. Autophagocytosis. In: Trump BF, Arstuka AU, eds.(1983) Pathobiology of cell membranes. New York: Plenum Press. 201-236.

[7] Kryger M, Roth T, Dement W, eds. (1994) The Principles and Practice of Sleep Medicine. 2nd Edition. Philadelphia, Saunders.

[8] Jelinic P, Stehle JC, Shaw P (2006) The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. PLoS Biol. 4: e355. DOI: 10.1371/ journal.pbio. 0040355.

[9] Orengo CA, Bray JE, Hubbard T, LoConte L, Sillitoe I (1999) Analysis and assessment of ab initio three-dimensional prediction, secondary structure, and contacts prediction. Proteins, 43(Suppl 3):149-170.

[10] Kharitonov SA, Barnes PJ. Clinical aspects of exhaled nitric oxide. Eur Respir J, in press.

[11] Zvaifler NJ, Burger JA, Marinova-Mutafchieva L, Taylor P, Maini RN (1999) Mesenchymal cells, stromal derived factor-1 and rheumatoid arthritis

[12] Jones X (1996) Zeolites and synthetic mechanisms. In Proceedings of the First National Conference on Porous Sieves: 27-30 June 1996; Baltimore. Edited by Smith Y. Stoneham: Butterworth-Heinemann.16-27.

[13] Ponder B, Johnston S, Chodosh L (Eds) (1998) Innovative oncology. In Breast Cancer Res. 10:1-72.

[14] Smith Y (Ed) (1996) Proceedings of the First

National Conference on Porous Sieves: 27-30 June 1996; Baltimore. Stoneham: Butterworth-Heinemann.

[15] Hunninghake GW, Gadek JE (1994) The alveolar macrophage. In Cultured Human Cells and Tissues. Edited by Harris TJR. New York: Academic Press.54-56. [Stoner G (Series Editor): Methods and Perspectives in Cell Biology, vol 1.]

[16] Kohavi R (1995) Wrappers for performance enhancement and oblivious decision graphs. PhD thesis. Stanford University, Computer Science Department.

[17] The Mouse Tumor Biology Database [http://tumor.informatics.jax.org/mtbwi/index.do].

[18] Junhua, W., Liu, C., Ma, T. (2018) Novel scheme for fault-induced delayed voltage recovery assessment. IEEE Transactions on Power Systems. 33(6):6542-6552. https://doi.org/10.1109/TPWRS.2018.2827485.

[19] Shah, S. M., Karimi, H. R., Sheikhi, A. (2017) Real-time stability assessment of microgrid using adaptive neuro-fuzzy inference system. IEEE Transactions on Smart Grid. 8(3):1347-1355. https://doi.org/10.1109/TSG.2016.2598959

[20] Ajjarapu, V., Christy, C. (1992) Continuation power flow: A tool for steady state voltage stability analysis. IEEE Transactions on Power Systems. 7(1):416-423. https://doi.org/10.1109/59.141737

[21] Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2002) Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 16:321-357. https://doi.org/10.1613/jair.953

[22] Torgo, L., Ribeiro, R. P., Pfahringer, B., Branco, P. (2013) Smote for regression. Progress in Artificial Intelligence. 2(1):21-30. https://doi.org/10.1007/s13748-013-0040-0

[23] Dietterich, T. G. (2000) Ensemble methods in machine learning. Multiple Classifier Systems. 1857:1-15. https://doi.org/10.1007/3-540-45014-9_1

[24] Chen, T., Guestrin, C. (2016) Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 785-794. https://doi.org/10.1145/2939672.2939785

[25] Breiman, L. (1996) Bagging predictors. Machine Learning. 24(2):123-140. https://doi.org/10.1007/BF00058655

[26] Breiman, L. (2001) Random forests. Machine Learning. 45(1):5-32. https://doi.org/10.1023/A:1010933404324

[27] Geurts, P., Ernst, D., Wehenkel, L. (2006) Extremely randomized trees. Machine Learning. 63(1):3-42. https://doi.org/10.1007/s10994-006-6226-1

[28] Phadke, A. G., Thorp, J. S. (2008) Synchronized phasor measurements and their applications. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-76537-2