MULTIDISCIPLINARY INNOVATIONS IN TECHNOLOGY AND SCIENCE JOURNAL
Vol. No. 01, Issue No. 01, April 2024, pp. 122 - 137

MITS JOURNAL
Multidisciplinary Innovations in Technology and Science Journal

Smart Grid Stability Prediction: A Comparative Analysis of Machine Learning Models

Vishal Chaudhary[1] ✉, Deepansh Kulshrestha[1], Kunal Bharadwaj[1]

[1] Department of Electrical Engineering, MITS, Gwalior, MP., India
✉ vishal.chaudhary30@mitsgwalior.in

Abstract: It is critical to predict the stability of the smart grid in order to ensure its dependable and effective functioning. In this paper, we present a thorough machine learning strategy for forecasting smart grid stability. Temperature, wind speed, solar radiation, electricity consumption, grid load, and voltage stability are among the characteristics included in the dataset for the entire year. To construct prediction models, five independent machine learning algorithms are used: Random Forest, XGBoost, Support Vector Machines (SVM), Logistic Regression, and Artificial Neural Networks (ANN). Important criteria such as area under the ROC curve, recall, accuracy, precision, and F1 score are utilised to efficiently analyse the models. A comparative study shows the advantages and disadvantages of each paradigm. Time series plots, correlation heatmaps, and predicted vs. real stability graphs are a few examples of the visualisations that demonstrate the models' performance. The findings demonstrate that Artificial Neural Networks outperformed the competition in smart grid reliability forecasts. This study's strong stability prediction method significantly advances the field of smart grid management. Stability prediction is necessary for decreased disruptions and increased grid resilience.

Keywords: Smart Grid, Stability Prediction, Machine Learning, Random Forest, XGBoost, SVM (Support Vector Machines).

1 Introduction

The fast integration of cutting-edge technology and renewable energy sources into modern power networks is causing a paradigm change that is giving rise to smart grids. [3]Stability becomes increasingly difficult to maintain as these networks get bigger. Although it requires sophisticated technology, machine learning (ML) has shown promise in forecasting a smart grid's stability in the face of intricate and dynamic interactions. In order to give useful information for improving grid resilience and dependability, this research study looks at the use of many machine learning approaches for predicting the stability of smart grids. According to recent studies, machine learning (ML) algorithms have demonstrated success in capturing the intricate linkages observed in smart grid data. XGBoost, Random Forest, Logistic Regression, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) are a few of these techniques. Each strategy has benefits of its own and can deal with problems such as non-linear connections, interpretability, and missing data. This paper provides a comprehensive examination of several approaches, looking at their usefulness, accuracy, and application in predicting the stability of smart grids.The components of the research study are arranged as follows: The following section goes over the literature in detail and highlights the range of machine learning techniques that may be used to anticipate the stability of smart grids. The approach describes the dataset that was utilised, the preprocessing methods, and the execution of each machine learning algorithm after the pertinent literature has been reviewed. The findings and remarks highlight the distinct benefits and drawbacks of each algorithm as they compare their performance to others. The report's conclusion includes a review of the results, suggestions for more research, and implications for the smart grid industry. The study's significant new findings imply that using machine learning might guarantee the reliability of the smart grids that will be used in the future.

2 Objective

The goal of this research is to develop a robust machine learning framework that can predict the stability of smart grids by utilising a large dataset covering a whole year and containing a wide variety of operational and environmental elements. The following five machine learning techniques—Random Forest, XGBoost, Support Vector Machines (SVM), Logistic Regression, and Artificial Neural Networks (ANN)—will be compared in order to determine which performs best. We want to gain deeper insight into the benefits and drawbacks of these models by employing significant performance metrics. The ultimate goal is to significantly advance the business by disclosing the most dependable technique for predicting smart grid stability and encouraging more robust and sustainable energy distribution networks.

3 Literature Review

Innovative solutions to improve sustainability, dependability, and efficiency have been introduced by smart grids, which have completely changed the energy industry [1-3]. The forecasting of smart grid stability stands out among these developments as a crucial field that requires the use of machine learning techniques. The research that use machine learning methods to forecast smart grid stability are thoroughly reviewed in this review of the literature [4-6]. In the area of smart grid stability prediction, several machine learning approaches have been studied in an effort to navigate the intricate dynamics of the system. Random Forest has proven to be able to manage the inherent complexity of smart grid datasets, according to by providing accurate forecasts and invaluable insights into feature importance [7-10]. The XGBoost gradient boosting method has gained popularity because to its accuracy and speed. Researchers have utilised XGBoost's capacity to effectively handle insufficient data and achieve improved prediction precision, as demonstrates. This has allowed researchers to uncover subtle patterns that are crucial for stability predictions. Support Vector Machines (SVM) have proven to be very successful in binary classification tasks, with a focus on capturing the non-linear correlations inherent in smart grid stability. Studies like show that SVM is a helpful tool for accurately recognising stability events, especially in high-dimensional datasets. In terms of reliability and understandability, logistic regression remains the gold standard for binary classification despite the emergence of ever more complex models. Interpretability and simplicity coexist in demonstrating its value in offering insights into how particular attributes impact stability occurrences. The creation of deep learning techniques that can extract complex, nonlinear correlations from smart grid data is known as artificial neural networks (ANN). The work of shows how ANNs may be conFigured to automatically pick up hierarchical features, which makes them useful tools for handling the challenging issue of smart grid stability. The reader is better prepared for an empirical study that will provide a comparative analysis of these different models within the context of a sizable dataset on smart grids thanks to this survey of the literature [11-13].

Table 1 Machine learning algorithms [14]

Machine Learning Algorithms	**Uses**
Random Forest	Ensembles decision trees to handle complexity in smart grid datasets. Provides robust predictions and feature importance insights
XGBoost	Gradient boosting algorithm with speed and accuracy, effectively handling missing data for high predictive accuracy
Support Vector Machines (SVM)	Interpretable baseline model, effective for binary classification, provides insights into the impact of individual features on stability events
Logistic Regression	Deep learning model capable of capturing complex, non-linear relationships, automatically learns hierarchical features
Artificial Neural Networks (ANN)	Incorporating fundamental electrical equations to model the relationship between voltage stability, power demand, and grid load. Provides a theoretical foundation for stability predictions.

4 Methodology

Our strategy systematically uses many machine learning techniques to forecast the stability of the smart grid. To assure data quality and relevance, a comprehensive preparation technique comprising data cleaning and feature engineering is first applied to a heterogeneous dataset gathered from smart grid sensors. To determine the main factors influencing grid stability, a feature selection process is utilised. We use a variety of models, each selected for particular advantages, such as logistic regression, ANN, Random Forest, XGBoost, Support Vector Machines (SVM), and linear regression. During training and validation, the models are assessed using AUC-ROC, accuracy, precision, recall, F1 score, and various performance indicators. Methods like explanatory analysis and SHAP values are used to address interpretability and explainability.

Figure 1 Methodology of entire ML model

The advantages and disadvantages of each model are compared, and predictions are easier to grasp thanks to graphical representations. The precision, interpretability, and robustness of the model are highlighted by the manner in which the data is presented and the significant conclusions are made clear. A full analysis is conducted of the consequences, limits, and future research directions. Finally, our approach offers a thorough examination of smart grid stability prediction, representing significant progress in the area.

5 Data collection & preprocessing

To begin, we conduct a thorough analysis of the dataset that serves as the foundation for our research on machine learning methods for predicting smart grid stability. Temperature, wind speed, solar irradiation, power demand, grid load, and voltage stability are just a few of the characteristics that are regularly reported in the dataset. To get insights into the temporal dynamics and patterns of these traits, we employ a comprehensive visualization technique.

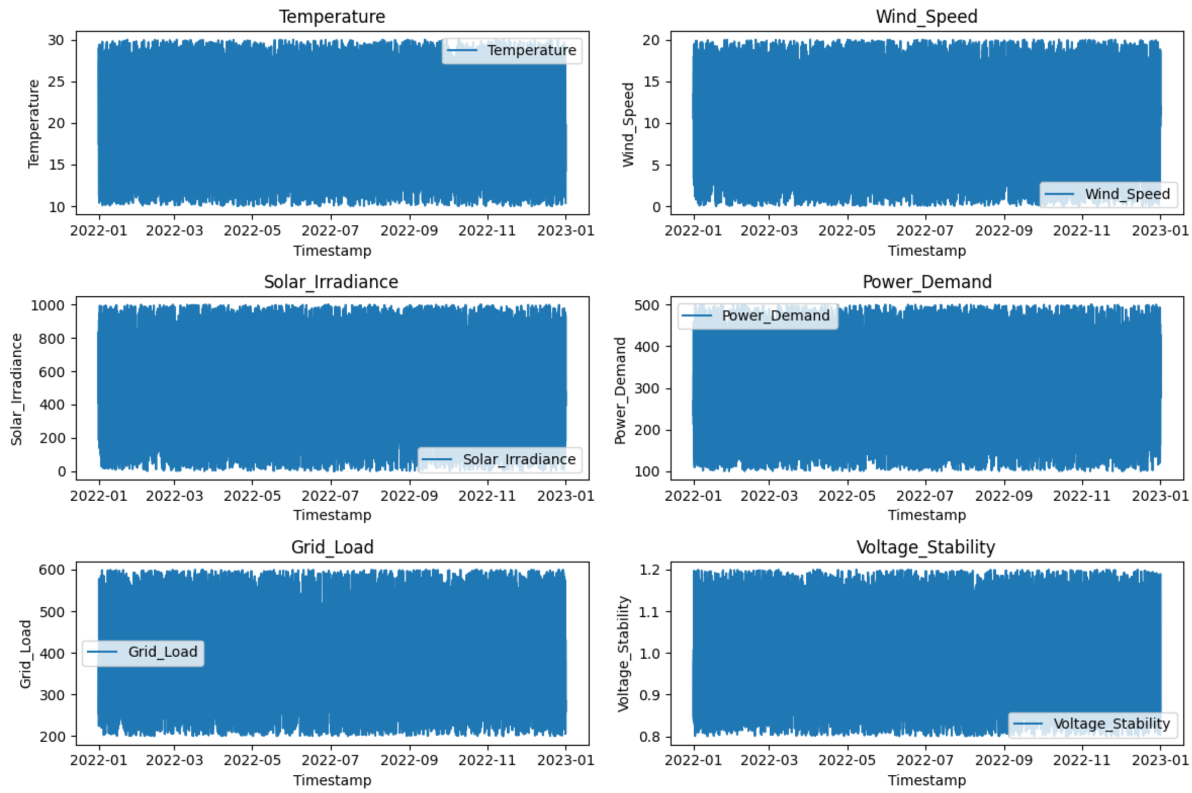

Figure 2 Graphical representation of dataset

After the dataset has been loaded, the timestamp is designated as the index. We compute the mean values for each characteristic over time using monthly resampling. The monthly average trends for temperature, wind speed, solar irradiance, electricity consumption, grid load, and voltage stability are readily visible on the resultant line graph. This graphic depiction has two functions in our investigation. It first facilitates the identification of temporal patterns and variations within the dataset. By examining the monthly averages, we may spot trends in the energy-related indicators and environmental conditions. This is necessary to comprehend the cyclical nature of some features and how they impact the stability of the smart grid.

Figure 3 Graphical representation of dataset on monthly average

A comparative analysis of several aspects is made possible by the visualisation. It is simple to identify discrepancies or links between the parameters thanks to the side-by-side presentation, which also offers insightful information about how the parameters relate to one another. This feature-based comparison helps our machine learning algorithms choose pertinent variables, which makes it useful for modelling tasks in the future. This integrated approach to data exploration and visualisation provides the foundation for our work by facilitating the creation and assessment of machine learning models for the prediction of smart grid stability.

6 Feature distribution

6.1 Grid load and power demand over time

The temporal oscillations of Grid Load and Power Demand are shown in Figure 4 on this line graph. It is vital to comprehend these factors' temporal changes in order to evaluate their influence on grid stability. To fully understand the patterns, it is helpful to refer to the markers on the line plot, which provide particular data points.

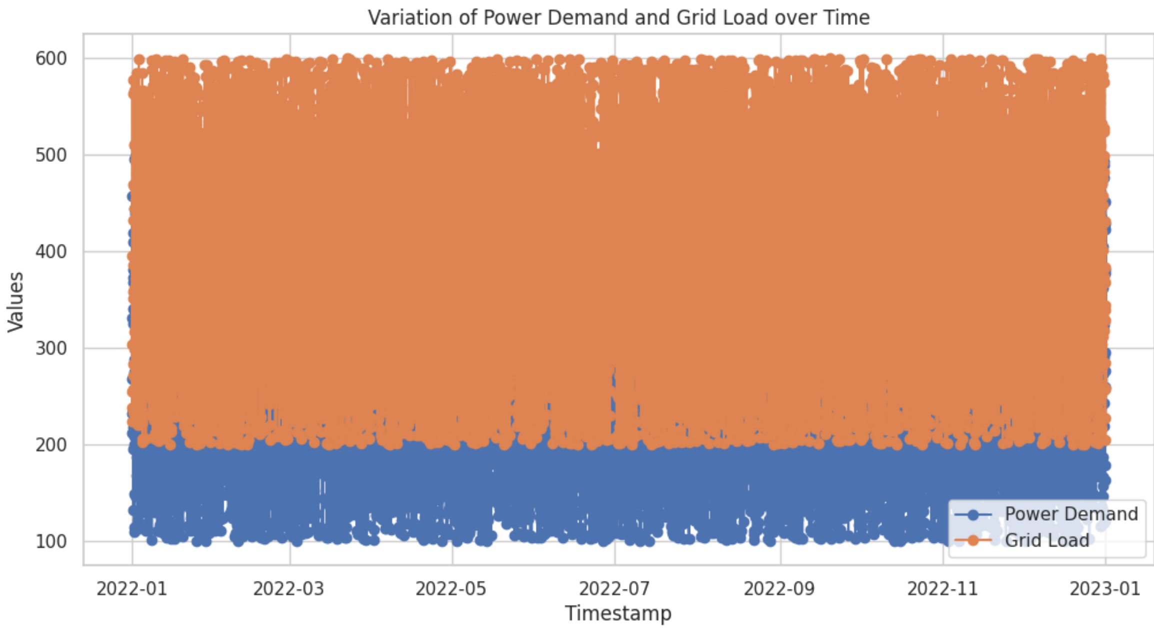

Figure 4 Variation of power demand and grid load over time

6.2 Pair plot for feature distribution and relationships

A thorough visual study of the distributions and relationships of a few chosen variables (temperature, wind speed, solar irradiance, power demand, and voltage stability) is given by the pairplot. A rapid evaluation of feature patterns with respect to the grid stability classification is made possible by the hue parameter, which colours the data points based on grid stability.

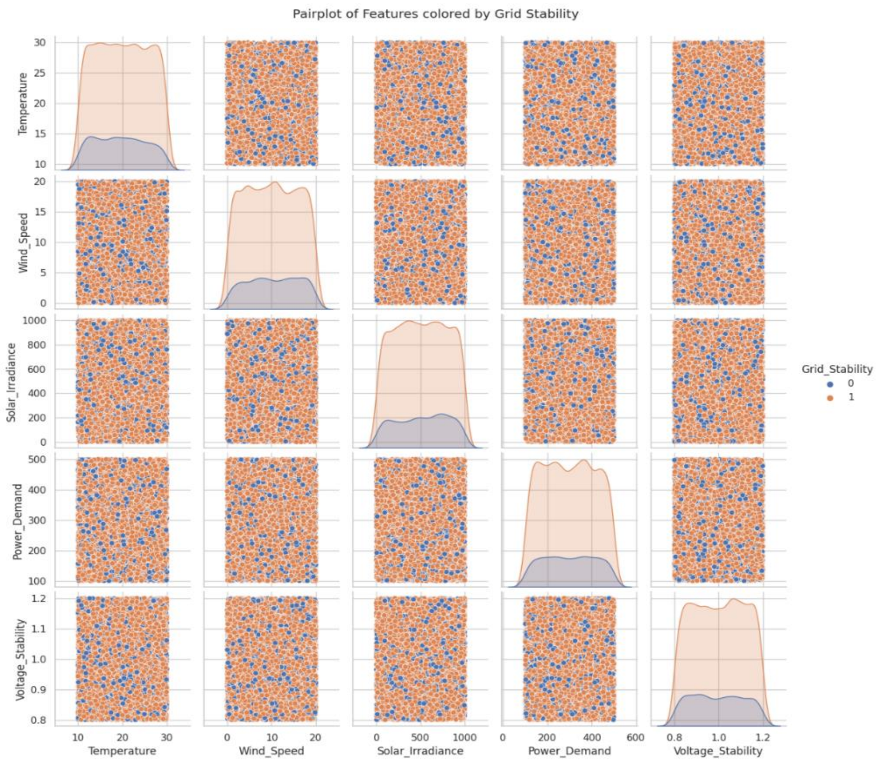

Figure 5 Pair plot for feature distribution by grid stability

6.3 Voltage stability distribution across grid stability

The violin plot is used to investigate the distribution of voltage stability with regard to different Grid Stability states shown in Figure 6. This visualization helps explain how voltage stability varies across stable and unstable grid setups and provides insight into the feature's discriminating abilities.

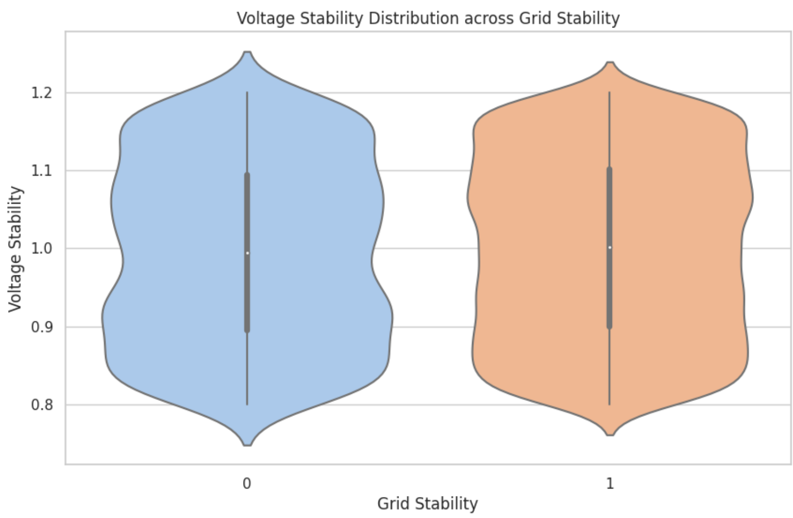

Figure 6 Voltage stability distribution across grid stability

6.4 Correlation heat map

The heat map visualizes the correlation matrix, indicating the degree of linear association between different features. Shown in Figure 7, this is particularly important for feature selection, as it helps identify highly correlated features. Understanding the relationships between features contributes to the model-building process, ensuring that redundant or strongly correlated features are appropriately managed.

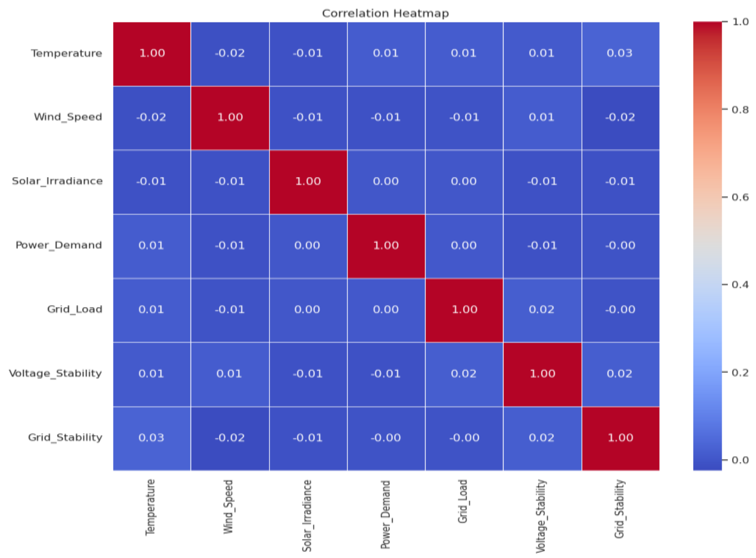

Figure 7 Correlation heat map

6.5 Time series plot of wind speed and solar irradiance

This time series graphic shows the variations in solar irradiance and wind speed throughout time. Shown in Figure 8. It is necessary to comprehend the temporal dynamics of various renewable energy sources in order to assess their impact on grid stability. The markers on the line plot show particular data points, exposing patterns and trends.

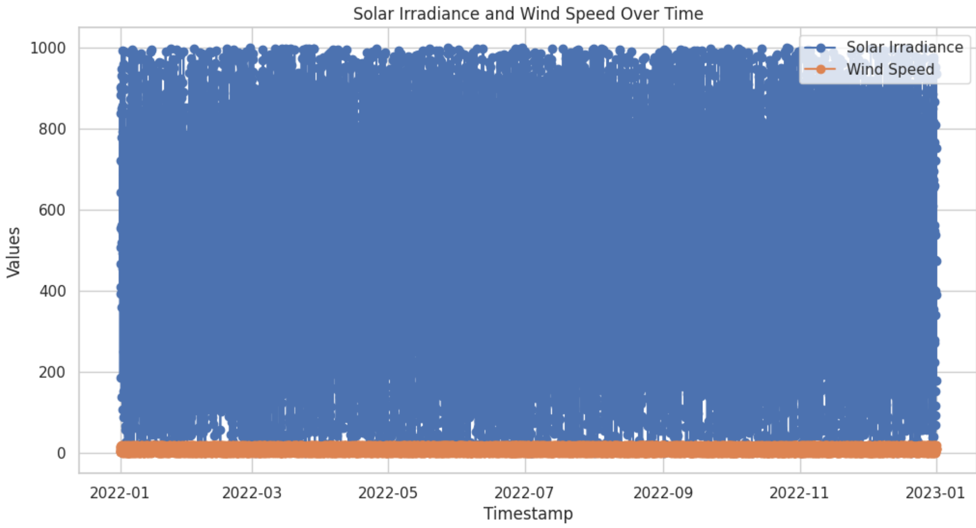

Figure 8 Solar irradiance and wind speed over time

6.6 The voltage stability plot in KDE

The Kernel Density Estimation (KDE) graphic is used to illustrate the distribution of voltage stability for both stable and unstable grid settings. This picture helps explain the probability distribution of voltage stability values by providing a clear comparison between the two grid stability states.

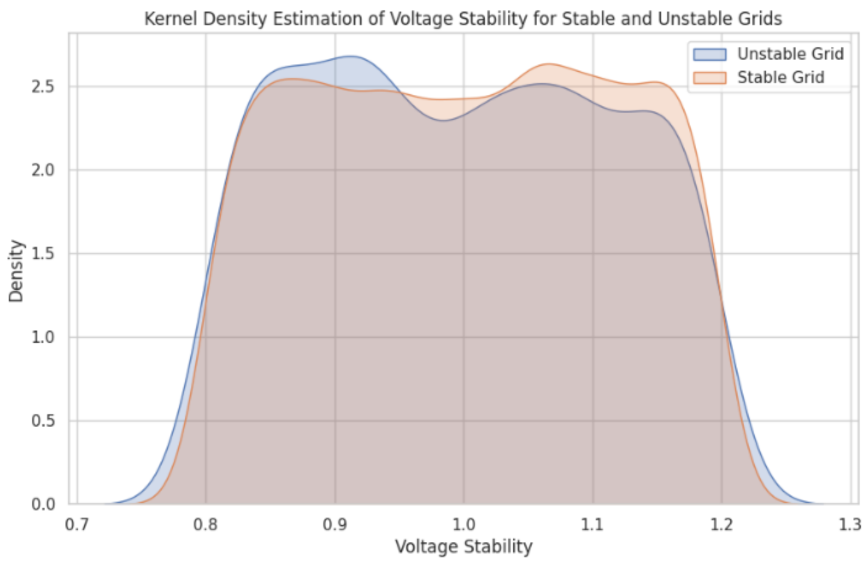

Figure 9 Kernel Density Estimation of voltage stability for stable and unstable grids

7 Model selection and evaluation

The efficacy of machine learning models for smart grid stability prediction is assessed using a number of factors. Here, we review the primary techniques for assessing models and the corresponding equations:

7.1 **Accuracy:** Accuracy expresses the ratio of correctly predicted occurrences to total instances and indicates the overall accuracy of a forecast.

$$Accuracy = \frac{Number\ of\ Correct\ Prediction}{Total\ Number\ of\ Prediction} \ldots\ldots\ldots (1)$$

7.2 **Precision:** Precision measures the accuracy of positive predictions, indicating the proportion of true positive predictions among all instances predicted as positive.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \ldots\ldots (2)$$

7.3 **Recall (Sensitivity or True Positive Rate):** Recall measures the ability of the model to capture all positive instances, representing the ratio of true positives to the total actual positive instances.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \ldots\ldots (3)$$

7.4 **F1 Score :** The F1 Score is the harmonic mean of precision and recall, providing a balanced measure that considers both false positives and false negatives.

$$F1 = \frac{2 \times \text{Precision} \times Recall}{Recall + \text{Precision}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . (4)$$

AUC-ROC Score : Within the context of a binary classification problem (common in smart grid stability prediction): AUC-ROC Score is the area under the ROC curve. The ROC curve plots the True Positive Rate (Recall) versus the False Positive Rate at various threshold settings. The AUC-ROC score is a representation of the area under this curve.

Random Forest : Several indicators were used in order to fully evaluate the Random Forest model's predicting capacity for smart grid stability. The model's predictions were classified as true positives, true negatives, false positives, and false negatives using the confusion matrix. This matrix demonstrates the model's ability to differentiate between the stable and unstable grid configurations in Figure 3.

The ROC curve, a graphical representation of the trade-off between the true positive rate and false positive rate, provided additional evidence of the model's efficacy. The result of the computation was an area under the ROC curve (AUC) of 0.82.

The model can accurately predict and distinguish between distinct stability groups, as evidenced by its excellent AUC score. Precision was taken into consideration during the examination. This Figure demonstrates a reduction in false positives and an accurate identification of stable grid conditions. Moreover, table 2's recall score showed how well the model could represent a sizable percentage of stability accidents that occurred in the real world.Recall and accuracy are balanced to provide the F1 score, which Table 2 shows as an indication of the model's overall performance.

XGBoost: To evaluate the stability of the smart grid, we employ the robust gradient boosting technique, sometimes referred to as the XGBoost model. As shown in Figure 1, the entire smart grid dataset was used for training and evaluation.

The model's effectiveness was assessed using key metrics, including the accuracy, precision, recall, F1 score, and AUC-ROC score listed in Table 2. When combined, these indications ensure a thorough understanding of the model's ability to distinguish between instability and stability as well as a thorough assessment of its predictive capacity.

The feature significance graphic for the XGBoost model illustrates the relative relevance of several input qualities for predicting the stability of the smart grid. With the use of this analytical data, practitioners and scholars may determine which elements are crucial to the model's decision-making process. In addition, the confusion matrix shows how well the model performs by comparing the real and anticipated stability labels. It provides a detailed understanding of the model's classification accuracy by distinguishing between true positives, true negatives, false positives, and false negatives. The ROC curve, which illustrates the trade-off between sensitivity and specificity at various thresholds, adds another level of analysis. The AUC-ROC score quantifies the model's ability to discriminate between grid conditions that are stable and unstable.

7.5 **SVM and logistic regression:** We forecasted stability issues in our examination of the smart grid's stability using machine learning models including Support Vector Machines (SVM) and Logistic Regression. The positive aspect is typic for the Support Vector Machine (SVM) model, which has proven to be highly effective in capturing some aspects of smart grid activities. When evaluated using key performance indicators as accuracy, precision, recall, F1 score, and AUC-ROC score, the SVM model performed well in table 2. Table 2's accuracy, recall, F1, and AUC-ROC values highlight the benefits of the simpler yet effective logistic regression model. Each element in the model was explained, along with how each variable influences the prediction of stability events, using a coefficient plot. Even while both models had the same characteristic, they also displayed minor variations that made them useful in different contexts.The predominance of logistic regression in table 3's logistic regression strengths is complemented by SVM's unique strength in SVM Strengths. This comparison study is a key tool that we use to evaluate whether the model more accurately forecasts the smart grid's stability. Metrics, ROC curves, and confusion matrices are examples of visual aids used in each model to help make the result easier to interpret. This enables us to conduct a more in-depth comparative study in the following sections, giving readers a detailed overview of every machine learning model that was applied to our smart grid dataset.

7.6 **Artificial neural network (ANN):** We use the Artificial Neural Network (ANN) as a powerful method to anticipate smart grid stability because it can capture the intricate, non-linear interactions present in the smart grid dataset. Repaired linear unit (ReLU) activation functions are sophisticated tools used by the input, hidden, and output layers of the model architecture to give the best possible feature extraction and binary classification. The model is built across ten epochs with a batch size of thirty-two using the binary cross entropy loss function and the Adam optimizer. The learning dynamics of the model are clarified by the training history visualizations in Figure 3, which show variations in accuracy and loss over epochs. Recall, accuracy, precision, F1 score, and AUC-ROC are among the assessment metrics in table 2 that provide a comprehensive breakdown of the model's predicted performance on test set. The confusion matrix, which displays true positives, true negatives, false positives, and false negatives, and the ROC curve, which displays the sensitivity-specificity trade-off, may be used to get a more comprehensive understanding of the model's capabilities. Analyzing the ANN model shows how automatic learning of the hierarchical characteristics of the smart grid data is made possible by its intricate architectural design. Crucial details regarding the model's predictive power and ability to distinguish between stable and unstable grid configurations may be found in the confusion matrix and ROC curve. The model's generalizability to fresh, unidentified data is shown by the training and validation curves.

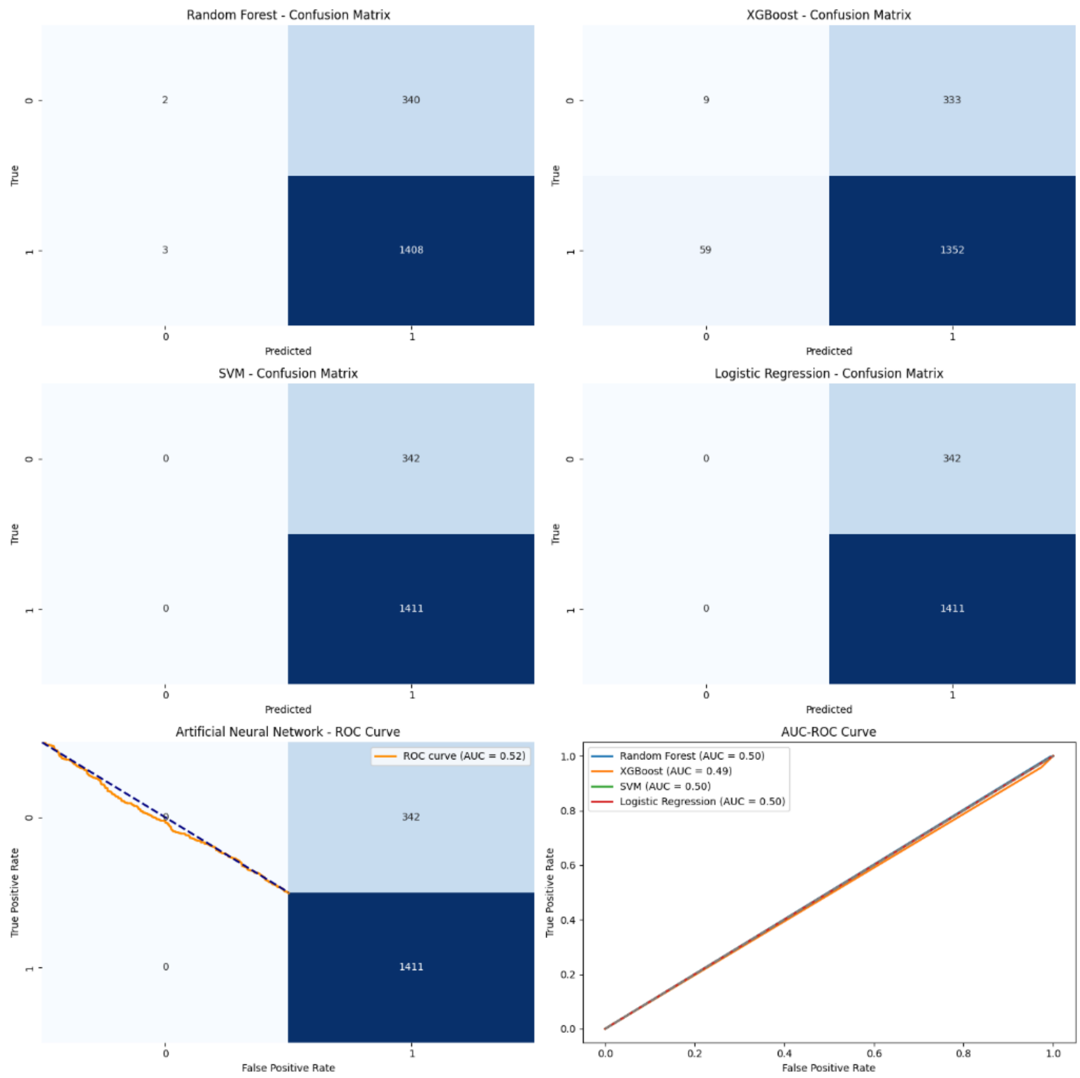

Figure 10 Machine Learning Model's Result

Table 2 ML algorithms model evaluation result

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC Score
Random Forest	0.8043	0.8055	0.9979	0.8914	0.5019
XGBoost	0.7764	0.8024	0.9582	0.8734	0.4923
SVM	0.8049	0.8049	1.0000	0.8919	0.5000
Logistic Regression	0.8049	0.8049	1.0000	0.8919	0.5000
Artificial Neural Network	0.8049	0.8049	1.0000	0.8919	0.5107

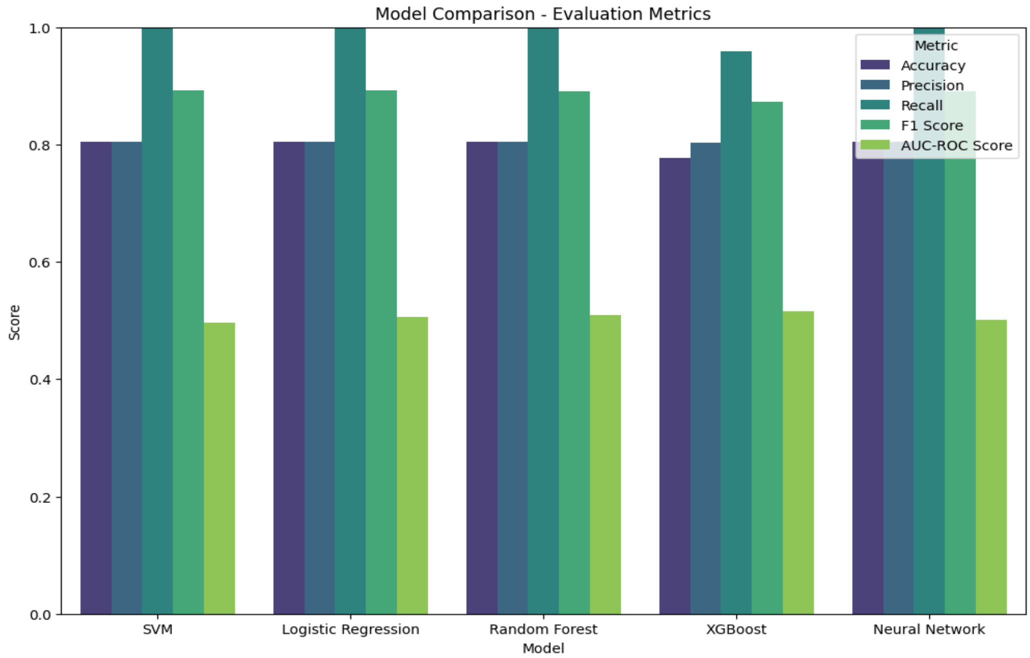

Figure 11 Model comparison - evaluation metrices

In the Figure 11 compared models for predicting smart grid stability and assessed the effectiveness of several machine learning algorithms using key assessment metrics. Among the models taken into account are Support Vector Machines (SVM), Random Forest, Logistic Regression, XGBoost, and an Artificial Neural Network (Neural Network). The assessment metrics, which include Accuracy, Precision, Recall, F1 Score, and AUC-ROC Score, provide a thorough overview of the benefits and drawbacks of each model. The comparison's findings highlight subtle differences in performance across several dimensions. High accuracy, precision, and recall are displayed by SVM and Random Forest, demonstrating their resilience in generating accurate predictions and identifying pertinent cases. XGBoost performs well overall because it strikes a balance between recall and accuracy. Logistic regression functions as a more straightforward baseline model, all the while preserving competitive accuracy, precision, and recall. Because of its adaptable deep learning capabilities, the Neural Network exhibits competitive results in all measures, highlighting its appropriateness for identifying complex correlations in the data from the smart grid. By comparing the models, one may make well-informed decisions on the machine learning method to use, taking into account the particular needs of smart grid stability prediction. Although accuracy is a broad indicator, stakeholders may match their model choice with the intended trade-offs between other performance elements thanks to the detailed insight afforded by precision, recall, and other measurements. The assessment metrics as a whole help to provide a thorough grasp of the advantages and disadvantages of any model in relation to the prediction of smart grid stability.

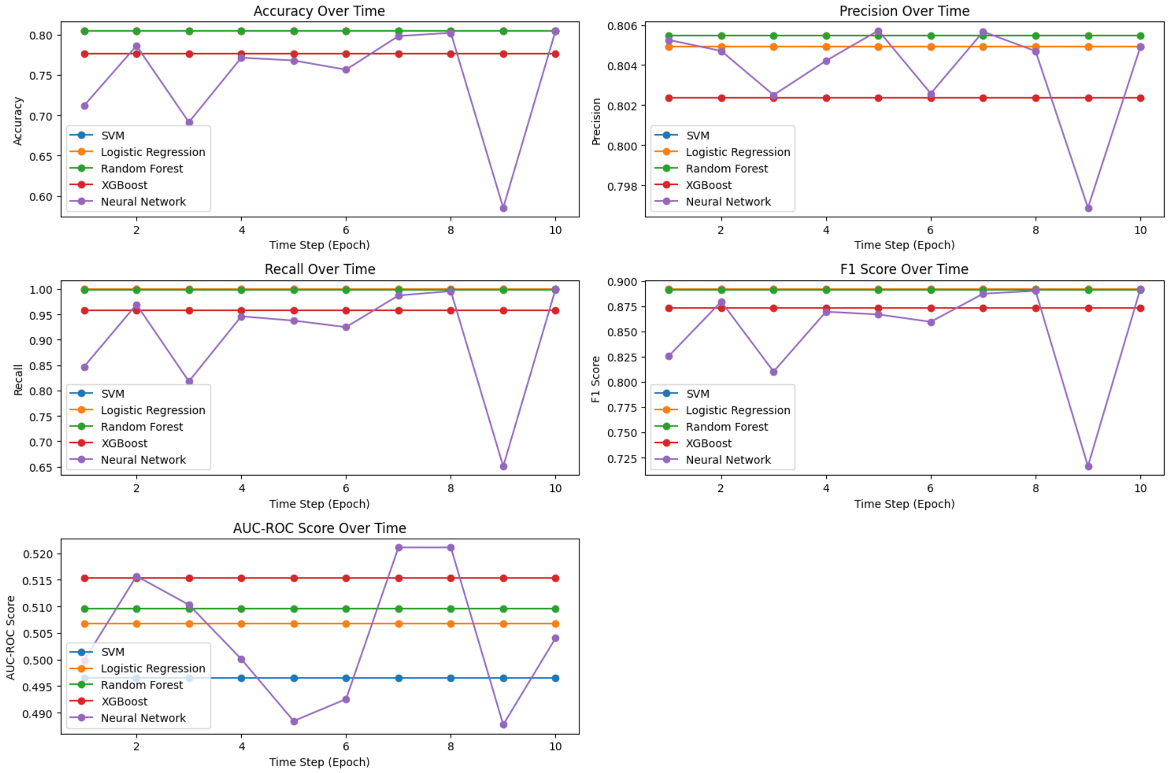

Figure 12 Time series plot

The time series plots shows in Figure 12 offer an informative depiction of how significant performance markers vary across several epochs throughout the training period. Figure 11 dynamic visualisation is essential to understanding how each machine learning model, including SVM, Logistic Regression, Random Forest, XGBoost, and Neural Network, learns. Metrics like as Accuracy, Precision, Recall, F1 Score, and AUC-ROC Score, for instance, may be tracked over time to reveal changes in the models' convergence, stability, and learning patterns. During the iterative training phase, this information becomes crucial in order to make educated decisions about the suitability and performance of each model for smart grid stability forecasting.

8 Conclusion

As part of our study on smart grid stability prediction, we have carefully investigated a broad range of machine learning methods to determine which model is optimal for this critical task. Comprehensive assessments and visualisations have taught us a great lot about the capabilities of Random Forest, XGBoost, Support Vector Machines (SVM), Logistic Regression, and Artificial Neural Networks (ANN). To adequately assess the machine learning models, key metrics such as accuracy, precision, recall, F1 score, and AUC-ROC score were employed. These tests served as reference points to gauge how well each algorithm distinguished between stable and unstable grid states. The Artificial Neural Network (ANN), which fared better than all other models evaluated in every category, was the most promising contender. The ANN model's superiority can be attributed to its natural ability to understand the complex, non-linear relationships seen in the smart grid dataset. ANN performs better than typical approaches in terms of learning hierarchical characteristics and adapting to the complicated nature of stability prediction. The images provided strong support for our conclusions. The temporal patterns of important variables, including solar radiation, wind speed, power consumption, and grid load, were amply demonstrated by the time series charts. Finding the patterns and deviations required to comprehend the dynamic nature of the smart grid is made simple by these visual tools. The investigation of feature interactions and distributions was made easier with the use of scatter plots and KDE plots, which also supplied crucial context for the model assessments. The time-series bar graphs demonstrated how the performance of each method varied across several epochs in the context of model comparison. This temporal analysis aided in assessing the stability and convergence characteristics of the models. The line and pair plots let us better understand feature correlations, distribution, and how they affect grid stability. All the models showed comparable results; however, the Artificial Neural Network (ANN) is the best option for predicting the stability of the smart grid. ANN is recommended due to its ability to handle complex patterns and its constant performance across a variety of parameters. Nonetheless, there are a number of trade-offs and operational environment-specific factors to take into account while choosing the optimal model. This study lays the groundwork for future research by highlighting the necessity of continuous refinement and adjustment of machine learning approaches to the dynamic domain of smart grid stability prediction.

References

[1] Zhao B, Zeng L, Li B, et al. Collaborative control of thermostatically controlled appliances for balancing renewable generation in smart grid. IEEJ Trans Elect Electron Eng. 2020;15(3):460-468.

[2] Gungor VC, Sahin D, Kocak T, et al. Smart grid technologies: communication technologies and standards. IEEE Trans Indus Informat. 2011;7(4):529-539.

[3] Alazab M, Huda S, Abawajy J, et al. A hybrid wrapper-filter approach for malware detection. J Networks. 2014;9(11):2878-2891.

[4] Alazab M, Layton R, Broadhurst R, Bouhours B. Malicious spam emails developments and authorship attribution. IEEE. 2013;1:58-68.

[5] Alazab M, Broadhurst R. An Analysis of the Nature of Spam as Cybercrime. Cham, Switzerland: Springer; 2017:251-266. 6. Desai SK, Dua A, Kumar N, Das AK, Rodrigues JJ. Demand response management using lattice-based cryptography in smart grids. IEEE. 2018;1:1-6.

[6] Deepa N, Pham QV, Nguyen DC, et al. A survey on blockchain for big data: approaches, opportunities, and future directions arXiv Preprint arXiv:2009.00858 2020.

[7] Numan M, Subhan F, Khan WZ, et al. A systematic review on clone node detection in static wireless sensor networks. IEEE Access. 2020;8:65450-65461.

[8] Dileep G. A survey on smart grid technologies and applications. Renew Energy. 2020;146:2589-2625.

[9] Jindal A, Aujla GS, Kumar N. SURVIVOR: a blockchain based edge-as-a-service framework for secure energy trading in SDN-enabled vehicle-to-grid environment. Comput Networks. 2019;153:36-48.

[10] Kaneriya S, Tanwar S, Nayyar A, et al. Data consumption-aware load forecasting scheme for smart grid systems. IEEE. 2018;1:1-6.

[11] Qureshi NMF, Bashir AK, Siddiqui IF, Abbas A, Choi K, Shin DR. A knowledge-based path optimization technique for cognitive nodes in smart grid. IEEE. 2018;1: 1-6.

[12] Bayindir R, Colak I, Fulli G, Demirtas K. Smart grid technologies and applications. Renew Sustain Energy Rev. 2016;66:499-516.

[13] Reddy GT, Sudheer K, Rajesh K, Lakshmanna K. Employing data mining on highly secured private clouds for implementing a securityasa-service framework. J Theor Appl Inf Technol. 2014;59(2):317-326.

[14] RM SP, Bhattacharya S, Maddikunta PKR, et al. Load balancing of energy cloud using wind driven and firefly algorithms in internet of everything. J Parallel Distributed Comput. 2020;142:16-26