A Comparative Analysis of Enhanced Machine Learning Algorithms for Smart Grid Stability Prediction

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

23-10-2021 / 26-10-2021

CITATION

GHOSH, ANKIT; KOLE, ALOK (2021): A Comparative Analysis of Enhanced Machine Learning Algorithms for Smart Grid Stability Prediction. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.16863145.v1>

DOI

[10.36227/techrxiv.16863145.v1](https://doi.org/10.36227/techrxiv.16863145.v1)

A Comparative Analysis of Enhanced Machine Learning Algorithms for Smart Grid Stability Prediction

Ankit Ghosh, *Student, School of Nuclear Studies and Application, Jadavpur University, Kolkata, India*, Alok Koley, *Professor, Department of Electrical Engineering, RCC Institute of Information Technology, Kolkata, India*

Abstract—Smart grid is an essential concept in the transformation of the electricity sector into an intelligent digitalized energy network that can deliver optimal energy from the source to the consumers. Smart grids being self-sufficient systems are constructed through the integration of information, telecommunication, and advanced power technologies with the existing electricity systems. Artificial Intelligence (AI) is an important technology driver in smart grids. The application of AI techniques in smart grid is becoming more apparent because the traditional modelling optimization and control techniques have their own limitations. Machine Learning (ML) being a sub-set of AI enables intelligent decision-making and response to sudden changes in the customer energy demands, unexpected disruption of power supply, sudden variations in renewable energy output or any other catastrophic events in a smart grid. This paper presents the comparison among some of the state-of-the-art ML algorithms for predicting smart grid stability. The dataset that has been selected contains results from simulations of smart grid stability. Enhanced ML algorithms such as Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbour (KNN), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), Stochastic Gradient Descent (SGD) classifier, XGBoost and Gradient Boosting classifiers have been implemented to forecast smart grid stability. A comparative analysis among the different ML models has been performed based on the following evaluation metrics such as accuracy, precision, recall, F1-score, AUC-ROC, and AUC-PR curves. The test results that have been obtained have been quite promising with the XGBoost classifier outperforming all the other models with an accuracy of 97.5%, recall of 98.4%, precision of 97.6%, F1-score of 97.9%, AUC-ROC of 99.8% and AUC-PR of 99.9%.

Impact Statement — AI will play a key role in building a safer, more vibrant, transparent and empowered society. Green energy has the potential to change our world for the better. It is essential for the energy grids to go digital in order to initiate a green energy boom. Grid operations have become rather complex. It is necessary for today's utilities to respond to changing weather conditions and energy demands. An ever increasing number of distributed energy resources have led to network congestion, inconsistent energy delivery, device failures and blackouts. Managing all these complexities can be daunting and the providers are under a lot of pressure to deliver a consistent and profitable energy supply. That is where AI comes into the scenario. This paper presents a comparative analysis of enhanced ML algorithms for smart grid stability prediction. AI and ML techniques can revolutionise today's energy grids. The influx of AI and ML will make smart grids more powerful, reliable and economical.

Index Terms—artificial intelligence, comparative analysis, machine learning, smart grid, stability prediction

I. INTRODUCTION

THE global energy demand is rapidly increasing. Therefore, it is necessary to evolve and upgrade the energy systems in order to make them more efficient, flexible and sustainable. The integration of information and digital communication technologies with the conventional power grid systems is known as the smart grid. A smart grid can be defined as a network of self-sufficient systems that enable the integration of power generation sources (both conventional and non-conventional) to the electrical grid causing the reduction of workforce and offering safe, reliable, high quality and sustainable electricity to the consumers [1], [2]. Smart grids enable two-way flow of electricity and data with the aid of digital communication systems which include smart meters, smart appliances, renewable energy resources and energy efficient resources. The two way communication between the producers and the consumers of electricity and the bidirectional power flow enhances the security, reliability and the efficiency of power systems [3], [4]. The smart grids are interconnected and integrated with advanced metering infrastructure, control, and communication technologies. This

leads to the generation of an enormous amount of high-dimensional and multi-type data. Such massive amounts of data require collection, storage and management. The existing technologies have many limitations in processing the data. Therefore, it is necessary to implement Artificial Intelligence (AI) techniques in smart grid applications. AI techniques provide an efficient way to analyze data, learn from it and then make appropriate decisions to ensure that the performance of the grid is as intended. AI has the potential to revolutionize the energy industry. AI can act as the brain of the smart grid by continuously collecting and synthesizing overwhelming amounts of data from large number of smart sensors and by making timely decisions to enhance the stability and reliability of the smart grid. AI has an important role to play in making use of the large amount of potentially useful data and generating actionable insights from it. Conventional techniques involve huge and complex data, thereby increasing the computational time and in some cases reduction in accuracy as well, a challenge that can easily be addressed using AI and machine learning (ML) [5], [6]. AI and ML can help in an improved understanding of consumer behaviour allowing suppliers to gain access to their exact electricity requirements. This will enable generation of the right type of billing information as well. The integration of AI and ML with smart grids will help consumers gain access to their energy consumption and pricing data so that they can actively respond to the requests of cutting down energy consumption during the peak hours of energy demand. This in turn will result in an enhanced operational efficiency of smart grids. The incorporation of Information and Communication Technologies (ICTs) in smart grids allows both consumers and producers to become active agents in maintaining the intended functioning of the smart grid.

In order to make the grid smart, a communication between the customers and the utility generating the electricity is needed to be established. The major objectives for the implementation of smart grids are

- improving demand side management
- increasing energy efficiency
- promoting a self-healing grid, thus improved reliability and resiliency

With the advancement of technology, the smart grids have been further modernized through the development of a concept known as demand response (DR). DR allows an increase in consumer control and cost savings. The consumers are provided an opportunity to reduce or shift their usage of electricity voluntarily during the peak hours with various forms of compensation like incentivizing with lower rates. The utilities and the operators are using DR programs to balance supply and demand of electricity in an evolving market. The main aims of DR analysis are

- minimization of the total electricity consumption
- reduction of the total required power generation
- promoting the idea of green and clean energy
- elimination of line overloading

DR is one of the most important concepts in smart grids.

DR allows end-use customers to play a significant role in the operation of the smart grid by changing their electricity consumption in response to changes in the price of electricity over a period of time [7], [8], [9]. If the consumption rate or tariff keeps on changing then the consumption pattern of the consumers is expected to change. Also, in DR incentives are paid to the consumers to induce the reduction of electricity use at times when the market prices are high or when the reliability of the smart grid is jeopardized. DR is performed at both the utility and the consumer level in order to achieve a low peak load curve and reduced electricity cost [10], [11], [12].

With the integration of newer advanced technologies such as AI and ML with the smart grids, it is feasible to predict the electricity demand of the consumers and automate DR [13], [14], [15]. As the smart grid involves vast data, the ML models can be improved over time and new predictive models can be deployed. So everything hinges on data and data is the key to unlock ML. Hence, AI and ML can be used to boost the performance of smart grids.

II. LITERATURE REVIEW

AI has been gaining a huge momentum and has been making a tremendous impact in the recent world. The transition of traditional electric grid system to the smart grid can be achieved through the integration of AI and ML techniques with the existing conventional methods. With the emergence of AI and ML the reliability and the resilience of smart grids can be improved. AI techniques can be applied to load forecasting, power grid stability assessment, faults detection, and security problems in the smart grid and power systems.

Azad et al. [16] have reviewed the application of different ML approaches aimed at enhancing the stability, reliability, security, efficiency and responsiveness of smart grid. They have also discussed some of the challenges in implementing ML solutions for smart grid.

Zheng et al. [17] have designed ML models based on historical data to investigate the influence of customer schedulable loads and the forecasted daily electricity price profile on aggregator profits. They have chosen K-Nearest Neighbor (KNN) and Gaussian process regression in their study because these algorithms have performed consistently and have produced accurate results in comparison to other ML classification and regression algorithms.

Bomfim [18] has investigated the evolution of the application of ML in the area of smart grids. He has presented an overview of research that has used ML in the area of smart grids, through the quantitative descriptive analysis of periodical articles and newspapers, registered in the IEEE Xplore Library database. He has conducted a survey that concludes that major topics related to smart grids on which research has been conducted are safety and reliability of the electrical network and energy management / forecasting.

You et al. [19] have presented how in comparison to conventional simulation-based approaches, AI can potentially save time on model development and numerical computation

of stability assessment in smart grids. They have also presented the development of a ML-based tool to assess power grid transient stability, frequency stability, and small signals stability. The test results from their proposed model have been quite encouraging.

Hossain et al. [6] have conducted a comprehensive study on the application of big data and machine learning in the electrical power grid introduced through the emergence of the next-generation power system-the smart grid. They have stated that the IoT-integrated Smart Grid system can provide efficient load forecasting and data acquisition technique along with cost-effectiveness. Smart grids generate a huge amount of data and such a massive data volume requires techniques far more superior to conventional methods for proper analysis and decision-making.

Günel and Ekti [20] have presented the advantages of the ML methods in extracting and analyzing features for designing modern industrial systems such as smart grids. They have discussed ML algorithms and given a brief study on the applications for smart grid.

Verma et al. [21] have presented how the scope of planning and operation of smart grids have broadened with the introduction of forecasting methods like artificial neural networks (ANNs), deep learning methods etc. Their paper has provided a survey of the works related to some of the smart grid components and have classified the works based on the computational intelligence tools used in solving the planning or operation problem.

Shi et al. [22] have given a comprehensive and clear picture of the recent advances in the field of AI techniques for stability analysis and control in smart grids. They have presented a general overview of AI, including the definitions, history and state-of-the-art methodologies. And then, they have given a comprehensive review of AI applications to security assessment, stability assessment, fault diagnosis, and stability control in smart grids. The implementation of AI based techniques has generated impressive results. In addition to that they have also discussed the major challenges that emerge during the deployment of AI based solutions in terms of high data requirements, imbalanced learning, interpretability of AI, difficulties in transfer learning, the robustness of AI to communication quality, and the robustness against attack or adversarial examples. They have also provided potential solutions to overcome these obstacles in order to bridge the gap between research and practice, thereby encouraging AI applications for the stability control and analysis in smart grids.

In view of the complexities involved, and the uncertainty and the large volume of data associated with smart grids, AI techniques can be deployed for the future development and the success of smart grids. In their paper [23] Zhang et al. have summarized the potential of Deep learning (DL), reinforcement learning (RL) and their combination-deep reinforcement learning (DRL) for application in smart grids. They have also provided an overview of the recent advances in the research work on their applications in smart grid technology. AI 2.0 is expanding exponentially owing to the

decreasing cost of computing power, the profusion of data, and the availability of better algorithms in AI. Smart grids being the developmental trend of power systems, the integration of AI with the existing technologies is likely to enhance the performance of smart grids.

The emergence of smart grids promises a more effective and efficient generation, transmission and consumption of power in contrary to conventional electrical systems. DR, an important concept in smart grid encourages the participation of the customers in energy savings. In order to maintain the balance between the supply and electricity demand, the accurate forecasting of electricity usage is a necessity. ML based predictive models can be employed to achieve this forecasting. The smart meters incorporated with the smart grids send user statistics to the server. Using ML techniques, predictive models can be designed for processing the smart meter data. Ali and Choi [24] have provided a comprehensive review of the state-of-the-art AI techniques for supporting real-time and quick demand response. Owing to the rapid advancement of AI in the recent years, expert systems ANN and fuzzy logic have been utilized in the energy sector for energy management in residential areas, inside a smart home leveraging DR program, and overall demand-side management (DSM).

DR is a key aspect in smart grid for improving its reliability in a cost effective manner. Over the years, there has been increased interest of research on DR. DR involves high complexity, usage of large-scale data and real-time decision making. AI and ML play a vital role for enabling demand side response. In their research paper [25], Antonopoulos et al. have provided an overview of AI methods utilised for DR applications. Their work is based on a systematic review of over 160 papers, 40 companies and commercial initiatives, and 21 large-scale projects. They have illustrated the rising interest in AI based solutions in the DR sector. AI methods have provided tools for the prediction, real-time efficient control of distributed systems, and decision-making for the application in DR. Prediction and estimation of load being an integral part of power system operation; AI and ML techniques have been used in load forecasting in DR. AI methods have also been employed for the prediction of electricity prices. Hence, the applications of AI in DR allow the grid operators to maintain the balance of the grid's operations.

The installation of numerous smart meters generates huge amount of data on electricity consumption from the customer side for analysis. In order to construct energy-efficient and reliable smart grids it is essential to develop probabilistic load forecasting (PLF). Yang et al. [26] have employed a technique called the Bayesian deep learning to address this challenging problem. They have proposed the development of a novel multitask PLF framework based on Bayesian deep learning to quantify the shared uncertainties across distinct customer groups while accounting for their differences. They have designed a clustering-based pooling method to increase the diversity and the volume of data, thereby reducing overfitting and improving the performance of the model for

prediction. The test results that have been obtained demonstrated that their proposed model exhibited a better performance in comparison to the conventional methods.

Energy load forecasting is an important aspect in the development of future smart grids. The use of traditional statistical and ML approaches involves forecasting error and high degree of overfitting. In their paper [27] Mohammad and Kim have proposed an energy load forecasting (ELF) model based on deep neural network architectures for the management of energy consumption in smart grids. The applicability and the performances of deep feed-forward neural network (deep-FNN) and deep recurrent neural network (deep-RNN) have been investigated. Both the architectures have been simulated using multi size training set. The performances of the architectures using various activation functions and different combinations of hidden layer have also been tested and the simulation results have been compared in terms of mean absolute percentage error. From the experimental results they have been able to infer that their proposed model has outperformed the existing load forecasting models.

Kuo and Huang [28] have introduced an accurate deep neural network algorithm for short-term load forecasting (STLF). The performance of the proposed load forecasting model has been compared with five other commonly used AI algorithms namely, Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), Multi-layer Perceptron (MLP) and Long Short Term Memory network (LSTM). Their model has produced Mean Absolute Percentage Error (MAPE) and Cumulative Variation of Root Mean Square Error (CV-RMSE) of 9.77% and 11.66%, respectively, thus a very high forecasting accuracy.

Effective and accurate price forecasting are quite essential in smart grids in order to avoid the serious consequences of price dynamics. Atef and Eltawil [29] have proposed two intelligent techniques to tackle the Electricity Price Forecasting (EPF) problem using ML. At first, they have implemented a Support Vector Regression (SVR) model for predicting the hourly-price and then compared the results with a deep learning model. The deep learning approach has outperformed the SVR model with an average root mean square error value of 1.1165 and 0.416 respectively.

Lu et al. [30] have proposed a reinforcement learning-based decision system for assisting the selection of electricity pricing plans, so that electricity payment and consumption dissatisfaction for individual smart grid end user can be minimized. They have modelled the decision problem as a transition probability-free Markov decision process (MDP) with improved state framework and have been solved using a Kernel approximator-integrated batch Q-learning algorithm, thereby improving the computational and prediction performance. Their proposed algorithm has the ability to extract the hidden features behind the time-varying pricing plans from a continuous high-dimensional state space. The test results have been quite promising. Thus, a precise predictive policy for individual user can be constructed reducing their cost and energy consumption dissatisfaction.

AI being the future vision of the electric power systems, computational intelligence tools can be effectively used in solving the planning or operational problems in smart grids.

III. PROPOSED METHODOLOGY

The proposed methodology has been described in this section.

A. Dataset used

The dataset that has been used corresponds to an augmented version of the “Electrical Grid Stability Simulated Dataset”, created by Vadim Arzamasov (Germany) and donated to the University of California (UCI) Machine Learning Repository.

The dataset contains the simulation results of grid stability for a reference 4-node star network as shown in Fig.1.

Fig. 1. Reference 4-node star network

The model has taken into consideration input features related to

- the total power balance (nominal power produced or consumed at each grid node)
- the response time of the participants to adjust consumption and/or production in response to the price changes, referred to as the reaction time
- energy price elasticity, which is a measure of the percentage change in the energy consumption relative to the percentage changes in the prices, all the other factors being equal

The dataset consists of 60,000 observations and 12 predictive features. The predictive features are

- Reaction time- Energy producer
- Reaction time- Consumer 1
- Reaction time- Consumer 2
- Reaction time- Consumer 3
- Power balance- Energy producer
- Power balance- Consumer 1
- Power balance- Consumer 2
- Power balance- Consumer 3
- Price elasticity co-efficient (gamma)- Energy producer
- Price elasticity co-efficient (gamma)- Consumer 1
- Price elasticity co-efficient (gamma)- Consumer 2

- Price elasticity co-efficient (gamma)- Consumer 3
- The prediction is a categorical (binary) label
- stable
 - unstable

B. Data pre-processing

The stable grid condition has been labeled as '1' while the unstable grid condition has been labeled as '0'.

The dataset contains features of various dimensions and scales altogether. This might adversely affect the model fitting and might lead to a biased outcome of the predictions in terms of misclassification errors and accuracy rates. Therefore, to eliminate this potential problem, standardization has been performed prior to the model fitting.

Standardization:

$$z = (x - \mu) / \sigma, \text{ mean } (\mu) = 0 \text{ and standard deviation } (\sigma) = 1.$$

C. ML models implemented

The following ML algorithms have been implemented for predicting the stability of the smart grid:

1. Support Vector Machine (SVM)
2. Logistic Regression
3. K-Nearest Neighbour (KNN)
4. Naïve Bayes (NB)
5. Decision Tree (DT)
6. Random Forest (RF)
7. Stochastic Gradient Descent (SGD) classifier
8. XGBoost
9. Gradient Boosting classifier

The proposed workflow has been illustrated in Fig.2.

Fig. 2. Proposed methodology

IV. RESULTS AND DISCUSSION

A performance comparative analysis of the different ML models has been performed based on the following evaluation metrics such as

1. Accuracy
2. Precision
3. Recall
4. F1-score
5. Area Under Curve- Receiver Operating

Characteristics (AUC-ROC) curve

6. Area Under Curve- Precision Recall (AUC-PR) curve

A. Metrics used for evaluating the performance of the different ML models

The performance metrics that have been used to evaluate the different ML models have been explained below.

1. Confusion matrix

True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)

2. Accuracy:

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN)$$

3. Precision:

$$\text{Precision} = TP / (TP + FP)$$

4. Recall:

$$\text{Recall} = TP / (TP + FN)$$

5. F1-score:

$$\text{F1-score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

6. AUC-ROC: AUC-ROC curve is an evaluation metric for the binary classification problems. The area under the curve (AUC) is a measure of the ability of the classifier to distinguish between the two classes. False Positive Rate is plotted along the x-axis and True Positive Rate is plotted along the y-axis of the AUC-ROC curve. The True Positive Rate (TPR), also known as the sensitivity helps in understanding what proportion of the positive class got correctly classified.

$$\text{TPR} = TP / (TP + FN)$$

The False Positive Rate (FPR) helps in understanding what proportion of the negative class got correctly classified by the classifier.

$$\text{FPR} = FP / (TN + FP)$$

$$\text{FPR} = 1 - \text{Specificity}, \text{Specificity} = TN / (TN + FP)$$

An excellent model should have an AUC near to 1 which is an indication that the model has a good measure of separability.

7. Similar to the ROC curve, the PR curve is used for evaluating the performance of the binary classification algorithms. The AUC-PR is constructed by plotting the precision against the recall for a single classifier at various threshold values. The nearer the AUC-PR score is to 1, the better the performance of the classifier.

B. Performance evaluation of the implemented algorithms

75% of the dataset has been used for training the models and the remaining 25% has been used for testing.

The attributes for determining the evaluation metrics for each of the nine ML models have been demonstrated in Table

I.

TABLE I
ATTRIBUTES FOR CALCULATING THE PERFORMANCE METRICS OF THE
DIFFERENT MODELS

Model	TP	FP	FN	TN
SVM	8407	1633	1218	3742
Logistic Regression	8408	1611	1217	3764
KNN	9094	1102	531	4273
NB	8855	1743	770	3632
DT	8840	818	785	4557
RF	9312	508	313	4867
SGD	8209	1555	1416	3820
XGBoost	9471	225	154	5150
Gradient Boosting	9290	742	335	4633

As it has been shown in Table I, the XGBoost classifier has detected 9471 data as true positives and 5150 true negatives accurately. It has misclassified only (225+154=) 379 data.

$$\text{Accuracy} = (9471 + 5150) / (9471 + 225 + 154 + 5150) = 0.975$$

$$\text{Misclassification} = (225 + 154) / (9471 + 225 + 154 + 5150) = 0.025.$$

Hence, it can be inferred that XGBoost has exhibited an accuracy of 97.5% and misclassification of 2.5%. Its performance has been better in comparison to the other classifiers.

To get a clear picture, the evaluation metrics of all the models have been tabulated.

The performance comparative analysis of the different ML algorithms based on accuracy, precision, recall and F1-score have been depicted in Table II.

TABLE II
PERFORMANCE COMPARISON BASED ON ACCURACY, RECALL, PRECISION
AND F1-SCORE

Model	Accuracy	Recall	Precision	F1-score
SVM	0.809	0.873	0.837	0.855
Logistic Regression	0.811	0.874	0.839	0.856
KNN	0.891	0.945	0.892	0.918
NB	0.832	0.920	0.836	0.876
DT	0.893	0.918	0.915	0.916
RF	0.945	0.967	0.948	0.957
SGD	0.802	0.852	0.841	0.846
XGBoost	0.975	0.984	0.976	0.979
Gradient Boosting	0.928	0.965	0.926	0.945

From the test results that have been shown in Table II, it can be concluded that XGBoost classifier has outperformed all the other models in terms of accuracy, recall, precision and F1-score. XGBoost has exhibited an accuracy, recall, precision and F1-score of 97.5%, 98.4%, 97.6% and 97.9% respectively.

After XGBoost, the performance of RF classifier has been noteworthy with an accuracy of 94.5%, recall of 96.7%, and precision of 94.8%, F1-score of 95.7%.

Gradient Boosting classifier follows XGBoost and RF with an accuracy, recall, precision and F1-score of 92.8%, 96.5%, 92.6% and 94.5% respectively.

After XGBoost, RF and Gradient Boosting the test results obtained with DT classification model have been quite promising. DT classifier has an accuracy of 89.3%, recall,

precision and F1-score of 91.8%, 91.5% and 91.6% respectively.

KNN follows DT with an accuracy, recall, precision and F1-score of 89.1%, 94.5%, 89.2% and 91.8% respectively.

Among the remaining classifiers, the performance of NB has been better than SVM, Logistic Regression and SGD classifiers.

NB classifier has achieved an accuracy of 83.2%, whereas SVM, Logistic Regression and SGD have accuracies of 80.9%, 81.1% and 80.2% respectively.

Therefore, XGBoost has been the most accurate classification algorithm while SGD classifier has been the least.

To enable a clear interpretation of the binary classification models, the ROC and the PR curves have been plotted. This has been described in the following sections.

C. ROC curves of the different models for performance comparison

The ROC curves corresponding to SVM, Logistic Regression, KNN, NB, DT, RF, SGD, XGBoost and Gradient Boosting classifiers have been shown in Fig. 3, 4, 5, 6, 7, 8, 9, 10 and 11 respectively.

Fig. 3. ROC curve for SVM

Fig. 4. ROC curve for Logistic Regression

Fig. 5. ROC curve for KNN

Fig. 8. ROC curve for RF

Fig. 6. ROC curve for NB

Fig. 9. ROC curve for SGD

Fig. 7. ROC curve for DT

Fig. 10. ROC curve for XGBoost

Fig. 11. ROC curve for Gradient Boosting classifier

The area under the ROC curves of all the ML algorithms have been tabulated in Table III.

TABLE III
PERFORMANCE COMPARISON OF THE MODELS BASED ON AUC-ROC

Model	AUC-ROC
SVM	0.887
Logistic Regression	0.887
KNN	0.953
NB	0.915
DT	0.882
RF	0.990
SGD	0.879
XGBoost	0.998
Gradient Boosting	0.981

From the AUC-ROC scores depicted in Table III, it can be observed that XGBoost classifier has the highest AUC-ROC score among all the other classifiers. XGBoost has produced an AUC-ROC score of 0.998 which is almost equal to 1. Therefore, it can be concluded that XGBoost best distinguishes between the two classes namely, stable (1) and unstable (0). The AUC-ROC score for RF is 0.990 which is less than 1 by only 0.010. The ability of RF classifier to distinguish between the two classes is also noteworthy. Gradient Boosting classifier does not fall behind with an AUC-ROC score of 0.981 which is only 0.019 short of 1. KNN and NB have AUC-ROC scores above 0.900. KNN and NB have AUC-ROC scores of 0.953 and 0.915 respectively. SVM and Logistic Regression algorithms have also shown promise with AUC-ROC scores of 0.887 respectively. DT classifier follows SVM and Logistic Regression with an AUC-ROC score of 0.882. SGD has the least AUC-ROC score of 0.879 which is less than 1 by 0.121. It can be inferred that the performance of all the classifiers have been quite encouraging.

In order to have a better understanding of the performances of each of the classification algorithms, the AUC-PR scores of all the models have also been evaluated. A comparison of the performances of each of the models based on AUC-PR has also been done. This has been described in the following section.

D. Performance evaluation of each classifier based on PR curves

The AUC-PR score of each algorithm has been demonstrated in Table IV.

TABLE IV
PERFORMANCE COMPARISON OF THE CLASSIFIERS BASED ON AUC-PR

Model	AUC-PR
SVM	0.932
Logistic Regression	0.932
KNN	0.974
NB	0.951
DT	0.943
RF	0.995
SGD	0.927
XGBoost	0.999
Gradient Boosting	0.990

From the test scores shown in Table IV, it can be seen that the AUC-PR score for XGBoost is 0.999 which is clearly exemplary. The AUC-PR for XGBoost is the highest among all the other models. RF follows with an AUC-PR of 0.995 which is less than that of XGBoost by 0.004. The AUC-PR score for Gradient Boosting classifier is 0.990 which is less than that of RF by a mere 0.005.

Fig. 12, 13 and 14 show the AUC-PR curves of XGBoost, RF and Gradient Boosting classifiers respectively.

Fig. 12. PR curve for XGBoost

Fig. 13. PR curve for RF

Fig. 14. PR curve for Gradient Boosting classifier

The AUC-PR score for KNN is 0.974 and that for NB is 0.951. DT has an AUC-PR score of 0.943. SVM and Logistic Regression both have AUC-PR scores of 0.932 respectively. SGD classifier has the least score of 0.927.

The PR curves for KNN, NB, DT, SVM, Logistic Regression and SGD have been displayed in Fig. 15, 16, 17, 18, 19 and 20 respectively.

Fig. 15. PR curve for KNN

Fig. 16. PR curve for NB classifier

Fig. 17. PR curve for DT classifier

Fig. 18. PR curve for SVM classifier

Fig. 19. PR curve for Logistic Regression

Fig. 20. PR curve for SGD classifier

The higher the AUC-PR, the better the classifier performs. From the AUC-PR plots it can be inferred that the performances of all the classification algorithms have been quite promising.

From the test results, it can be concluded that XGBoost is undoubtedly the best classification algorithm for predicting the stability of the smart grid from the dataset that has been considered. XGBoost has evidently surpassed all the other algorithms in terms of accuracy, recall, precision, F1-score, AUC-ROC and AUC-PR.

The metric scores of XGBoost classifier have been demonstrated in Table V.

TABLE V
METRIC SCORES EXHIBITED BY XGBOOST

Metric	Score
Accuracy	0.975
Recall	0.984
Precision	0.976
F1-score	0.979
AUC-ROC	0.998
AUC-PR	0.999

In addition to the metrics that have already been discussed, the training time and prediction time of each algorithm have also been analyzed. This has been discussed in the following section.

E. Performance comparison based on training time and prediction time taken by each model

The training time is a representation of the length of the time period taken by a classifier from the beginning of the model training to the moment it is ready to perform the task of classification.

The prediction time is the duration taken by a classifier to predict the outcome.

The training time and the prediction time for each classification algorithm have been extracted. Table VI depicts the performance comparison of the models based on training and prediction times.

TABLE VI
COMPARISON OF THE TRAINING TIME AND PREDICTION TIME OF EACH MODEL

Model	Training Time (s)	Prediction Time (s)
SVM	102.531	4.931
Logistic Regression	0.043	0.001
KNN	0.395	16.097
NB	0.078	0.031
DT	1.117	0.016
RF	19.022	0.447
SGD	0.771	0.005
XGBoost	11.135	0.047
Gradient Boosting	31.067	0.071

The training time and the prediction time for Logistic Regression is the minimum at 0.043s and 0.001s respectively. The training time for XGBoost classifier is 11.135s and its prediction time is 0.047s. SVM takes the maximum time for training. The training time is 102.53s which is over a minute. The prediction time for SVM is 4.931s. The prediction time for KNN is 16.097s which means KNN takes the most time for predicting the outcome among the other algorithms. Its training time is however 0.395s. NB has a training time and prediction time of 0.078s and 0.031s respectively. DT takes 1.117s for training and 0.016s for predicting. The training time and prediction time for RF classifier are 19.022s and 0.447s respectively. SGD classifier is quite fast in predicting the outcomes. Its prediction time is 0.005s. The training time for SGD is 0.771s. Gradient Boosting classifier has a training and prediction time of 31.067s and 0.071s respectively.

V. CONCLUSION

The application of AI and ML techniques in smart grids provides powerful technical support to the digital power systems. In this paper a comparative analysis of nine ML classification algorithms namely, SVM, Logistic Regression, KNN, NB, DT, RF, SGD, XGBoost and Gradient Boosting has been performed based on six evaluation metrics namely, accuracy, recall, precision, F1-score, AUC-ROC and AUC-PR for predicting smart grid stability from a dataset that has been obtained from the UCI machine learning repository. XGBoost classifier has attained an accuracy of 97.5%, recall of 98.4%, precision of 97.6%, F1-score of 97.9%, AUC-ROC of 99.8% and AUC-PR of 99.9% outperforming all the other classification algorithms that have been implemented. The training and prediction times of each ML model have also been analyzed. Logistic Regression has demonstrated the least training and prediction times among the other models. It takes 0.043s for training and 0.001s for predicting the outcome. The training time and the prediction time for XGBoost classifier are 11.135s and 0.047s respectively. In general the experimental results that have been obtained for each of the nine ML algorithms have been quite encouraging. The stability of smart grid is an essentiality for enabling efficient power distribution. ML plays a vital role in predicting the stability of a smart grid. As part of the future work, other ML models can be deployed for predicting the stability of the smart grids and enhancing their reliability. Also, hardware implementation of the ML models can be achieved using the Field Programmable Gate Array (FPGA) for the real time analysis.

REFERENCES

- [1] C. Wietfeld, A. A. Cardenas, H. Chen, P. Popovski and V. W. S. Wong, "Smart Grids", *IEEE Wireless Communications*, vol. 24, no. 2, pp. 8-9, April 2017.
- [2] Luis M. Camarinha-Matos, "Collaborative smart grids – A survey on trends", *Renewable and Sustainable Energy Reviews*, vol.65, pp.283-294, Nov. 2016.
- [3] R. Morello, C. De Capua, G. Fulco and S. C. Mukhopadhyay, "A Smart Power Meter to Monitor Energy Flow in Smart Grids: The Role of Advanced Sensing and IoT in the Electric Grid of the Future", *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7828-7837, Dec. 2017.
- [4] T. Samad and A. M. Annaswamy, "Controls for Smart Grids: Architectures and Applications", *Proceedings of the IEEE*, vol. 105, no. 11, pp. 2244-2261, Nov. 2017.
- [5] B. K. Bose, "Artificial Intelligence Techniques in Smart Grid and Renewable Energy Systems—Some Example Applications", *Proceedings of the IEEE*, vol. 105, no. 11, pp. 2262-2273, Nov. 2017.
- [6] E. Hossain, I. Khan, F. Un-Noor, S. S. Sikander and M. S. H. Sunny, "Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review", *IEEE Access*, vol. 7, pp. 13960-13988, Jan. 2019.
- [7] S. Gupta, R. Kambli, S. Wagh and F. Kazi, "Support-Vector-Machine-Based Proactive Cascade Prediction in Smart Grid Using Probabilistic Framework", *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2478-2486, April 2015.
- [8] S. Ghosh, X. A. Sun and X. Zhang, "Consumer profiling for demand response programs in smart grids", in *Proc. IEEE PES Innovative Smart Grid Technologies*, Tianjin, China, 2012, pp. 1-6.
- [9] A. Almalaq and G. Edwards, "A Review of Deep Learning Methods Applied on Load Forecasting", in *Proc. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Cancun, Mexico, 2017, pp. 511-516.
- [10] J. S. Vardakas, N. Zorba and C. V. Verikoukis, "A Survey on Demand Response Programs in Smart Grids: Pricing Methods and Optimization Algorithms", *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 152-178, July 2014.
- [11] S. Bahrami, M. H. Amini, M. Shafie-khah and J. P. S. Catalão, "A Decentralized Electricity Market Scheme Enabling Demand Response Deployment", *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4218-4227, July 2018.
- [12] S. C. Chan, K. M. Tsui, H. C. Wu, Y. Hou, Y. Wu and F. F. Wu, "Load/Price Forecasting and Managing Demand Response for Smart Grids: Methodologies and Challenges", *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 68-85, Sept. 2012.
- [13] Fahad Javed, Naveed Arshad, Fredrik Wallin, Iana Vassileva, Erik Dahlquist, "Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting", *Applied Energy*, vol.96, pp.150-160, Aug. 2012.
- [14] Renzhi Lu and Seung Ho Hong, "Incentive-based demand response for smart grid with reinforcement learning and deep neural network", *Applied Energy*, vol.236, pp.937-949, Feb. 2019.
- [15] Muhammad Qamar Razaa and Abbas Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings", *Renewable and Sustainable Energy Reviews*, vol.50, pp. 1352-1372, Oct. 2015.
- [16] S. Azad, F. Sabrina and S. Wasimi, "Transformation of Smart Grid using Machine Learning," in *Proc. 2019 29th Australasian Universities Power Engineering Conference (AUPEC)*, Nadi, Fiji, 2019, pp. 1-6.
- [17] Y. Zheng, S. Suryanarayanan, A. A. Maciejewski, H. J. Siegel, T. M. Hansen and B. Celik, "An Application of Machine Learning for a Smart Grid Resource Allocation Problem," in *Proc. 2019 IEEE Milan PowerTech*, Milan, Italy, 2019, pp. 1-6.
- [18] T. S. Bomfim, "Evolution of Machine Learning in Smart Grids," in *Proc. 2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE)*, Oshawa, ON, Canada, 2020, pp. 82-87.
- [19] Shutang You, Yinfeng Zhao, Mirka Mandich, Yi Cui, Hongyu Li, Huangqing Xiao, Summer Fabus, Yu Su, Yilu Liu, Haoyu Yuan, Huaiguang Jiang, Jin Tan and Yingchen Zhang, "A Review on Artificial Intelligence for Grid Stability Assessment," in *Proc. 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, Tempe, AZ, USA, 2020, pp. 1-6.
- [20] K. Günel and A. R. Ekti, "Exploiting Machine Learning Applications for Smart Grids" in *Proc. 2019 16th International Multi-Conference on Systems, Signals & Devices (SSD)*, Istanbul, Turkey, 2019, pp. 679-685.
- [21] P. Verma, K. Sanyal, D. Srinivasan, K. S. Swarup and R. Mehta, "Computational Intelligence Techniques in Smart grid planning and operation: A Survey", in *Proc. 2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, Singapore, 2018, pp. 891-896.
- [22] Zhongtuo Shi, Wei Yao, Zhouping Li, Lingkan Zeng, Yifan Zhao, Runfeng Zhang, Yong Tang and Jinyu Wen, "Artificial intelligence techniques for stability analysis and control in smart grids: Methodologies, applications, challenges and future directions", *Applied Energy*, vol.278, pp.115733-115744, Nov. 2020.
- [23] Dongxia Zhang, Xiaoqing Han and Chunyu Deng, "Review on the research and practice of deep learning and reinforcement learning in smart grids", *CSEE Journal of Power and Energy Systems*, vol.4, issue.3, pp. 362 – 370, Sept. 2018.
- [24] Syed Saqib Ali and Bong Jun Choi, "State-of-the-Art Artificial Intelligence Techniques for Distributed Smart Grids: A Review", *Electronics*, vol.9, pp.1030-1055, June 2020.
- [25] Ioannis Antonopoulos, Valentin Robu, Benoit Couraud, Desen Kirli, Sonam Norbu, Aristides Kiprakis, David Flynn, Sergio Elizondo-Gonzalez and Steve Wattam, "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review", *Renewable and Sustainable Energy Reviews*, vol.130, pp. 109899-109934, Sept. 2020.
- [26] Yandong Yang, Wei Li, T. Aaron Gulliver and Shufang Li, "Bayesian Deep Learning-Based Probabilistic Load Forecasting in Smart Grids", *IEEE Transactions on Industrial Informatics*, vol.16, issue.7, pp. 4703 – 4713, July 2020.
- [27] Faisal Mohammad and Young-Chon Kim, "Energy load forecasting model based on deep neural networks for smart grids", *International Journal of System Assurance Engineering and Management*, vol.11, pp. 824–834, Sept. 2019.
- [28] Ping-Huan Kuo and Chiou-Jye Huang, "A High Precision Artificial Neural Networks Model for Short-Term Energy Load Forecasting", *Energies*, vol.11, pp.213-226, Jan. 2018.
- [29] S. Atef and A. B. Eltawil, "A Comparative Study Using Deep Learning and Support Vector Regression for Electricity Price Forecasting in Smart Grids," in *Proc. 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA)*, Tokyo, Japan, 2019, pp. 603-607.
- [30] T. Lu, X. Chen, M. B. McElroy, C. P. Nielsen, Q. Wu and Q. Ai, "A Reinforcement Learning-Based Decision System for Electricity Pricing Plan Selection by Smart Grid End Users," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2176-2187, May 2021.