Revolutionizing Consumer Power Management: Unveiling Power Grid Feasibility Analysis Using Machine Learning

Minhazur Rahman[1], Sheik Erfan Ahmed Himu[2], Abdullah Al Shahid Chowdhury[3]
Arafat Ibne Ikram[4], Samiul Hoque Kashfi[5], Md Imtiaz Uddin [6], MD Faisal[7]
[134567]*Department of Electrical and Electronic Engineering, International Islamic University Chittagong. Chittagong, Bangladesh*
[2]*Department of Electrical and Mechanical Engineering, Nagoya Institute of Technology. Nagoya, Japan*
Email: [1]engr.minhazur@gmail.com, [2]erfanhimu@gmail.com, [3]shahidaac.eee@gmail.com, [4]arafatibne.ikram@gmail.com,
[5]samiulhoque81@gmail.com, [5]imtiazanam56@gmail.com, [7]faisalpes087@gmail.com

***Abstract*—As distributed energy sources become more prevalent, maintaining power grid stability is increasingly challenging. By integrating machine intelligence and communication technologies, traditional power networks could transform into smart grids. Through machine learning and artificial intelligence, these smart grids can adeptly respond to unexpected changes in consumer demand, power interruptions, surges in renewable energy production, and other crucial situations. Variations in power generation and loads, along with changes in the power system's structure, lead to varying shifts in the entire network's active power. Detecting these fluctuations through manual analysis is laborious, but machine learning can be highly impactful in this regard. This study aims to collect comprehensive data for analyzing grid stability, utilizing machine learning tools to thoroughly examine the data, and adopting a multimodal approach to compare model outcomes for an improved solution strategy. Our analysis reveals an accuracy exceeding 0.97, indicating strong potential for practical application and implications.**

***Index Terms*—Machine Learning, XGB, KNN, Decision Tree, SVM, GB, Grid, Power, Consumer, Dataset.**

I. Introduction

Maintaining power grid reliability involves ensuring both a consistent power supply and its quality [1]. Voltage, a key indicator of power quality, must be carefully controlled to avoid aging or damage to electrical equipment due to excessive or insufficient voltage levels. Reactive power is closely tied to voltage. Excessive reactive power can lead to a voltage rise, while insufficient reactive power can cause a voltage drop. To ensure safe and stable power grid operation, optimizing reactive power distribution is essential [2]. Current algorithms used in power systems struggle with the complex grid environment. To achieve real-time automatic optimization of reactive power and voltage, a more suitable algorithm is needed [3]. Reactive power optimization is complex due to nonlinear and hybrid characteristics, involving discrete (reactive power compensation devices, transformer taps) and continuous variables (transport voltage). Deep reinforcement learning, which combines perception and decision-making, is well-suited for handling intricate and layered issues [4]. Power system problems are often dynamic, multi-constrained, and nonlinear. Conventional measurement methods suffer from issues like poor convergence or low accuracy. Deep reinforcement learning offers a flexible and centralized approach to optimize reactive power, effectively controlling voltage balance [5]–[7]. Enhancements in distribution and energy systems are being accompanied by concurrent challenges. This necessitates a refined solution that is cost-effective, stable, dependable, efficient, and secure [8], [9].

These requirements culminate in the notion of a smart grid, which represents the fusion of ecological considerations and technological requisites [10]. Due to the expansion of smart grid infrastructure, effective monitoring, management, control, protection, and the equilibrium between demand and production under all conditions emerge as pivotal elements for the success and efficiency of forthcoming endeavors [11]. Furthermore, the manipulation and analysis of extensive data stemming from the intelligent infrastructure, characterized by the five V's of big data (Velocity, Variety, Veracity, Value, and Volume), necessitates careful consideration [12]. As a result, ensuring the stability of the smart grid remains intricate; however, emerging methodologies such as machine learning and artificial intelligence offer potent tools for stability management [13].

Artificial Intelligence (AI) denotes computer models designed to simulate human thought processes. It addresses challenges using intelligent methods that don't necessarily rely on human involvement. In everyday scenarios, numerous issues demand resolution, often necessitating human judgment to identify the optimal approaches, thereby ensuring successful outcomes for proposed solutions. Machine learning falls within the domain of Artificial Intelligence, enabling the acquisition of knowledge from data without explicit programming. This is achieved through algorithms designed to enhance the learning process.

979-8-3503-1873-9/23/$31.00

Many researchers have looked into power system stability. Here, we'll highlight some of the latest studies and discuss them. Mamoun Salazar et. al. [14] proposed a cunning MLSTM model that is familiar with predicting the sufficiency of the sagacious networks. The proposed model is researched the splendid grid dataset from UCI man-made intelligence Vault. The introduction of MLSTM is differentiated and regular ML models like LSTM, GRU, RNN. The overall assessment shows the prevalence of the proposed model with respect to precision, exactness, hardship, and ROC twist estimations. Vuk Malbasa and his co-author introduced a new method for using power system data to predict voltage stability, considering uncertainties in machine learning models and computational challenges [15]. Their experiments demonstrate a notable advantage in terms of shorter training and prediction times, along with a reduced number of required measurements to achieve accurate predictions. The paper suggests a deep learning model using a combination of Densely Connected Convolutional Networks and Residual Network structures to identify smart grid stability. The model's outcomes are contrasted with those of widely used classifier models from various studies, all employing the same dataset used in our analysis. In [16], the authors introduced a fresh model for assessing transient stability. They employed and further advanced machine learning techniques to identify early signs of potential blackouts. In [17], the authors investigated machine learning techniques to overcome the challenge of maintaining a simple representation. They found that system stability could be achieved even when some consumers adjusted their power consumption patterns.

In [18], the prediction of grid stability through diverse machine-learning approaches is reported. The authors employed a multilayer neural network along with Binary Particle Swarm and Binary Kangaroo Mob feature selection techniques. They conducted a comparison with results from Random Forest Classifier, Decision Tree classifier, Gradient Boosting Tree, and logistic regression methods.
Variations in generator and load outputs, along with changes in power system topology, result in diverse shifts in network active power [19]. Manually detecting these fluctuations is laborious. Machine Learning is crucial here, swiftly identifying changes through data analysis. This enhances power system management efficiency and accuracy, especially in the face of evolving challenges like renewable integration and dynamic demand [20].

This study focuses on three main objectives: first, the compilation of an extensive dataset for grid stability analysis with an adequate number of data nodes; second, the application of machine learning tools to conduct thorough and in-depth data analysis; and third, the adoption of a multimodal approach to compare outcomes from different models for an improved solution strategy. Supervised learning models are employed in conjunction with an evaluation process that considers data visualization and feedback. The analysis reveals accuracy percentages for various models: XGB - 97.86%, GB - 93.02%, SVM - 79.42%, Random Forest - 92.56%, KNN - 81.22%, LR - 81.43%, and Decision Tree - 80.37%. Of particular note, the XGB algorithm stands out for its exceptional adaptability in addressing regression, classification, ranking challenges, and user-defined objective functions.

II. Methodology

Fig. 1 depicts the main flowchart of the suggested approach. The process begins with library imports and comprises three key stages: (1) data visualization, (2) dataset classification, and (3) application of machine learning. These stages are elaborated further in the subsequent sections.

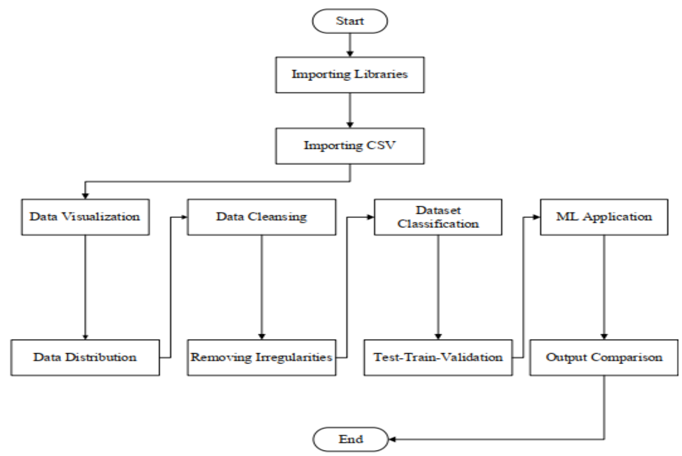

Fig. 1. Block Diagram of the System

A. Data Visualization

The section focuses on the sequential steps of data exploration in machine learning. Fig. 2 shows the flowchart of the data visualization process. It begins with Data Visualization, which aids in comprehending data distribution and variations. Following this, Scatter plots are employed, and subsequently, Pair plots are generated, enhanced with a stability-oriented hue. This series of processes culminate with a comprehensive understanding of the data's intricacies and marks the conclusion of the exploration phase.

B. Data Cleansing

The flowchart (Fig. 3) within the Data Cleansing section involves a series of essential steps. Initially, the process addresses data anomalies like spikes or outliers, ensuring their removal. Next, it focuses on rectifying missing values present in specific cells. Additionally, a thorough assessment of irregularities in the data's shape is conducted. Lastly, the dataset is transformed to its appropriate data types, optimizing it for subsequent analysis. These interconnected actions collectively contribute to the refinement and preparation of the data for further processing.

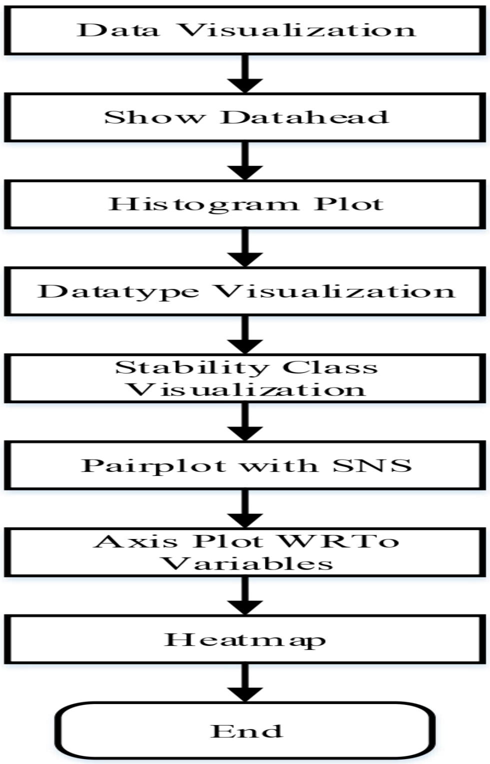

Fig. 2. Data Visualization

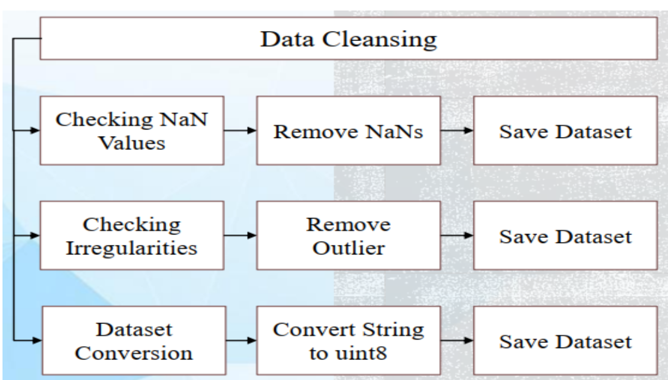

Fig. 3. Data Cleansing Process

C. Dataset Classification

The Dataset Classification section's flowchart involves a streamlined process which is shown in Fig. 4

Initially, the dataset is divided into training and testing subsets, serving their respective purposes. Further, individual data frames are established using a 70-30 split ratio, systematically designed to accommodate various algorithms. This methodical approach ensures the suitability of the dataset for diverse algorithmic applications, contributing to effective classification outcomes.

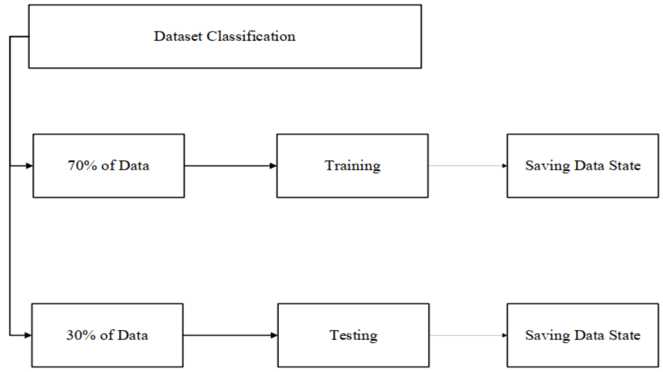

Fig. 4. Data Classification Process

D. Machine Learning Application

To apply a machine learning model, first prepare the model by selecting an algorithm and tuning its hyperparameters. Then, input the dataset for model training, split it into training and test sets, and calculate all performance parameters. Finally, it generates accuracy matrices from the test data to visualize the performance of the model. The whole process shown in Fig. 5

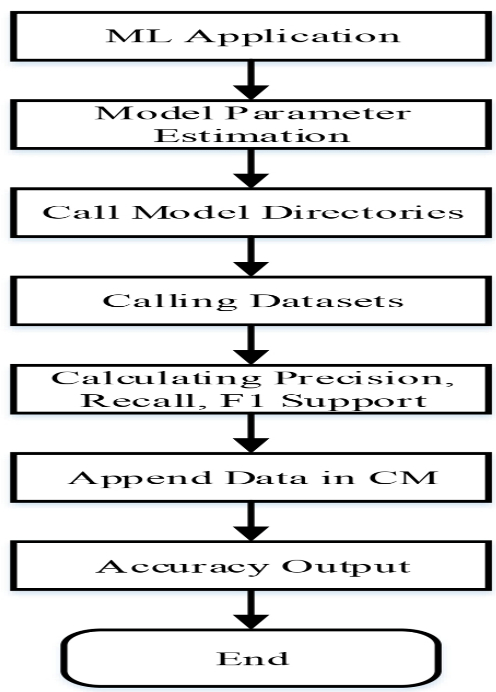

Fig. 5. ML Application

III. Dataset Description

The dataset shown in Table I, employed in the experiment is sourced from the UCI machine learning repository, comprising

TABLE I. Dataset Overview

Sl.	tau1	tau2	tau3	tau4	p1	p2	p3	p4	g1	g2	g3	g4	stab	stabf
0	2.959060	3.079885	8.381025	9.780754	3.763085	-0.782604	-1.257395	-1.723086	0.650456	0.859578	0.887445	0.958034	0.055347	unstable
1	9.304097	4.902524	3.047541	1.369357	5.067812	-1.940058	-1.872742	-1.255012	0.413441	0.862414	0.562139	0.781760	-0.005957	stable
2	8.971707	8.848428	3.046479	1.214518	3.405158	-1.207456	-1.277210	-0.920492	0.163041	0.766689	0.839444	0.109853	0.003471	unstable
3	0.716415	7.66966	4.4486641	2.340563	3.963791	-1.027473	-1.938944	-0.997374	0.446209	0.976744	0.929381	0.362718	0.028870	unstable
4	3.134112	7.608772	4.9443759	9.857573	9.525811	-1.125531	-1.845975	-0.554305	0.787110	0.45450	0.656947	0.820923	0.049860	unstable

60,000 samples and 14 features. It pertains to outcomes from grid stability simulations for four-node star references, with features encompassing factors like producer and consumer reaction times, nominal power generation and consumption, as well as consumer and producer gamma coefficients. Notably, the dataset includes variables such as tau1, tau2, tau3, and tau4 denoting reaction times for the electricity producer and three consumers, p1, p2, p3, and p4 representing nominal power, and g1, g2, g3, and g4 as gamma coefficients for both producers and consumers. The dataset also incorporates two dependent variables and a categorical "stabf" label indicating stability ("stable") or instability ("unstable"). The "stab" variable represents the maximal real part of the equation root, with positive and negative values correlating with system instability and stability, respectively. Consequently, the "stab" variable is not utilized as either input or output.

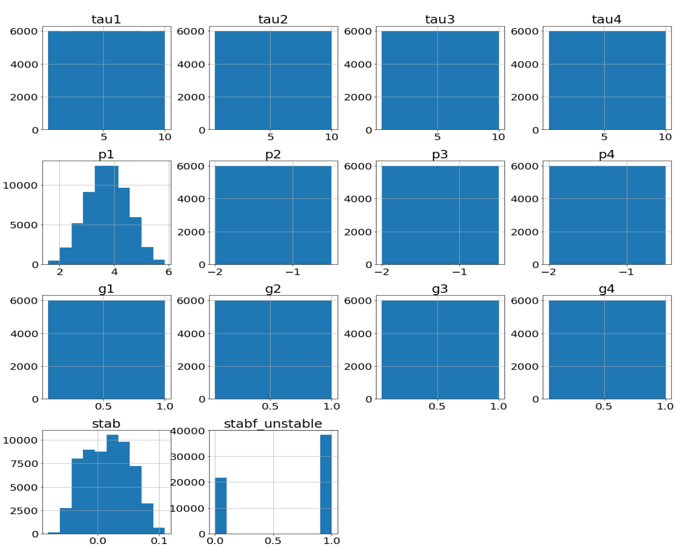

Fig. 6. Data Histogram

A histogram depicting a summary of the data is being generated. In this visualization (Fig. 6), the data is presented as it exists in the CSV file, allowing for the identification of any underlying patterns. The histogram encompasses all properties of the dataset, including the range of data points. The x-axis corresponds to the values of the data, while the y-axis represents the frequency or count of each data point.

Subsequently, a tally of data types is conducted, along with a count of features within each class. In terms of data type, there are 13 occurrences of the Float type and 1 occurrence of the Int type within the CSV. The total label count reveals 38,280 instances for the "Unstable" label and 21,720 instances for the "Stable" label. Following this, a Seaborn (SNS) pairplot is generated, utilizing the feature as a hue parameter. The

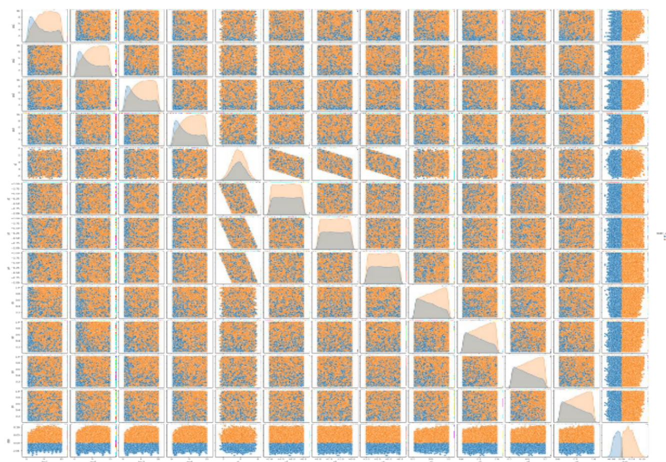

Fig. 7. Relation in Between Features and Label

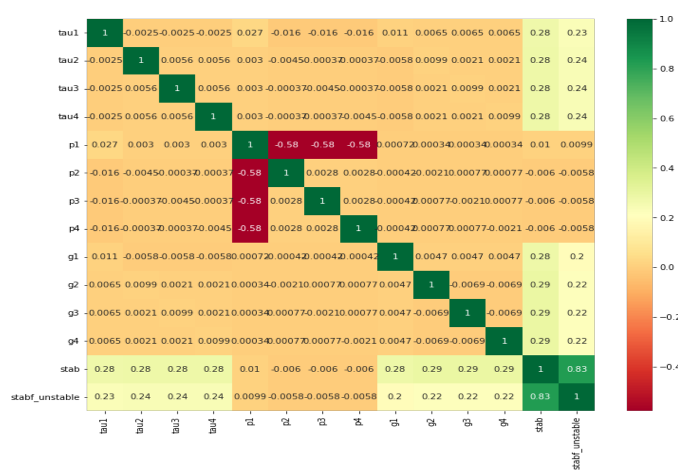

Fig. 8. Relation in Between Features and Label

correlation overview among features and the dependent variable is presented in Fig. 7 using the Python Seaborn package. The heatmap in Fig. 8 highlights a predominantly positive correlation among most variables, except for p3 and p4, which exhibit a distinct pattern.

IV. Result and Analysis

In this section, we present the outcomes of the applied model and draw comparisons. To assess the effectiveness of the machine learning algorithm, we employ various metrics. Given the distinct distribution of the dependent variable and the binary classification nature of the problem [21], we adopt the confusion matrix. This matrix elucidates the interplay between the labels [22]. A visual representation of the confusion matrix is provided in Table II.

TABLE II. Binary Classification Using Confusion Matrix

	True Actual Class		
Predicted Class	Classes	Positive	Negative
	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Furthermore, our evaluation metrics encompass accuracy, precision, recall, F1 score, and geometric mean, each of which is meticulously defined below.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

Table III presents the results of accuracy, precision, recall and F1-score derived from summing up the confusion matrix across all models. Among these metrics, the XGB model attains the most favorable outcomes.

Fig 9 Shows the output result of our experiment. To get the best results from the analysis, we applied several model from the Scikit learn toolbox. The results are varied based on the type of the algorithm. The XGB algorithm produces the best possible outcome. It is a highly flexible and versatile tool that can work through most regression, classification, and ranking problems as well as user-built objective functions. As an open-source software, it is easily accessible and it may be used through different platforms and interfaces. The Accuracy matrics comparison is also shown in Fig. 10

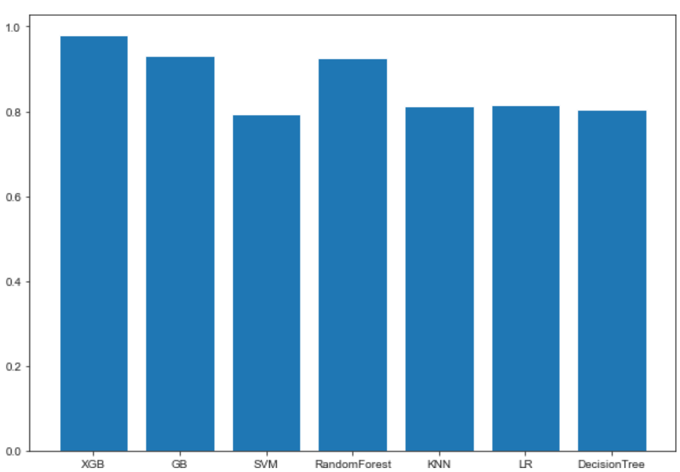

Fig. 9. Output Comparison

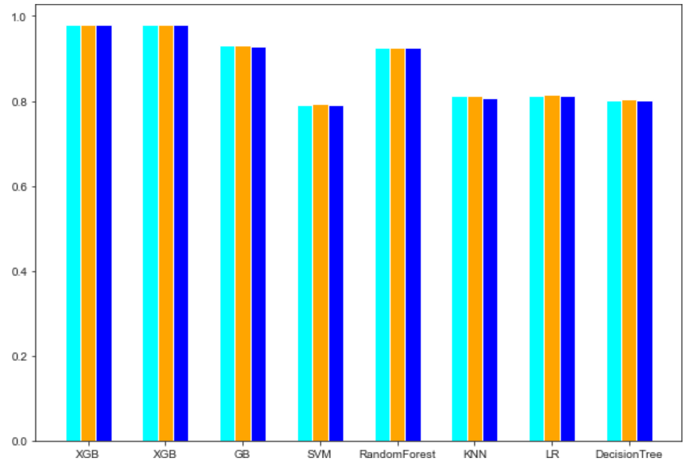

Fig. 10. Accuracy Metric Comparison

V. CONCLUSION

This research extensively explored the application of machine learning models on a simulated grid stability dataset. The study involved a meticulous comparison of output accuracy across various models, all evaluated using the same dataset. The models under scrutiny included XGB, GB, SVM, Random Forest, KNN, LR, and Decision Tree. The findings incontrovertibly showcased the superior performance of the XGB model, attaining an impressive accuracy score of 97.86% when contrasted with alternative models on the original dataset. This outcome underscores the feasibility and potential implications of practical utilization. It is worth noting that, akin to many other machine learning approaches, our study relied on commonly accessible datasets due to inherent data acquisition constraints. The dataset's specifically tailored distribution, while beneficial for experimental purposes, may introduce variations in output when extrapolated to real-time systems. Moving forward, the acquired insights hold promise for real-world implementation within active grid environments, fostering practical experimentation. Furthermore, an exciting avenue for future exploration involves the integration of more comprehensive machine learning algorithms, thereby extending the horizons of this research endeavor.

REFERENCES

[1] M. M. Gulzar, M. Iqbal, S. Shahzad, H. A. Muqeet, M. Shahzad, and M. M. Hussain, "Load frequency control (lfc) strategies in renewable energy-based hybrid power systems: A review," *Energies*, vol. 15, no. 10, p. 3488, 2022.

[2] A. Eid, S. Kamel, and L. Abualigah, "Marine predators algorithm for optimal allocation of active and reactive power resources in distribution networks," *Neural Computing and Applications*, vol. 33, no. 21, pp. 14327–14355, 2021.

[3] L. Yavuz, A. Önen, S. Muyeen, and I. Kamwa, "Transformation of microgrid to virtual power plant–a comprehensive review," *IET generation, transmission & distribution*, vol. 13, no. 11, pp. 1994–2005, 2019.

TABLE III. Confusion Matrix Result

No	Name	TP	FN	FP	TN	Accuracy	Precision		Recall		F1-Score	
							0	1	0	1	0	1
1	XGB	11326	165	220	6289	97.86%	0.98	0.97	0.99	0.97	0.98	0.97
2	GB	11134	357	900	5609	93.02%	0.93	0.94	0.97	0.86	0.95	0.90
3	SVM	10168	1323	2381	4128	79.42%	0.81	0.76	0.88	0.63	0.85	0.69
4	Random Forest	11112	379	960	5549	92.56%	0.92	0.94	0.97	0.85	0.94	0.89
5	KNN	10459	1032	2348	4161	81.22%	0.82	0.80	0.91	0.64	0.86	0.71
6	LR	10242	1249	2093	4416	81.43%	0.83	0.78	0.89	0.68	0.86	0.73
7	Decision Tree	10012	1479	2054	4455	80.37%	0.83	0.75	0.87	0.68	0.85	0.72

[4] A. Likmeta, A. M. Metelli, A. Tirinzoni, R. Giol, M. Restelli, and D. Romano, "Combining reinforcement learning with rule-based controllers for transparent and general decision-making in autonomous driving," *Robotics and Autonomous Systems*, vol. 131, p. 103568, 2020.

[5] R. El Helou, D. Kalathil, and L. Xie, "Fully decentralized reinforcement learning-based control of photovoltaics in distribution grids for joint provision of real and reactive power," *IEEE Open Access Journal of Power and Energy*, vol. 8, pp. 175–185, 2021.

[6] M. Choudhury, S. E. Ahmed Himu, M. U. Khan, M. Z. Hasan, M. S. Alam, and T. Soga, "Analysis of charge transport resistance of zno-based dsscs because of the effect of different compression temperatures," *AIP Advances*, vol. 13, no. 9, 2023.

[7] S. E. A. Himu, S. Sultana, M. S. H. Chowdhury, A. I. Ikram, H. R. Saium, and M. M. Hossain, "Modification of dynamic logic circuit design technique for minimizing leakage current and propagation delay," in *2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI)*, pp. 1–5, IEEE, 2022.

[8] A. I. Ikram, M. S.-U. Islam, M. A. Bin Zafar, M. K. Rocky, Imtiaz, and A. Rahman, "Techno-economic optimization of grid-integrated hybrid storage system using ga," in *2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP)*, pp. 300–305, 2023.

[9] M. A. B. Zafar, M. R. Islam, M. S.-U. Islam, M. Shafiullah, and A. I. Ikram, "Economic analysis and optimal design of micro-grid using pso algorithm," in *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*, pp. 421–424, IEEE, 2022.

[10] A. H. Radwan and A. A. G. Morsi, "Opportunities, challenges and applications of smart grids in housing projects in new settlements in cairo "towards energy efficient urban areas"," in *Urban and Transit Planning: A Culmination of Selected Research Papers from IEREK Conferences on Urban Planning, Architecture and Green Urbanism, Italy and Netherlands (2017)*, pp. 439–456, Springer, 2020.

[11] S. Dorji, A. A. Stonier, G. Peter, R. Kuppusamy, and Y. Teekaraman, "An extensive critique on smart grid technologies: Recent advancements, key challenges, and future directions," *Technologies*, vol. 11, no. 3, p. 81, 2023.

[12] F. Cappa, R. Oriani, E. Peruffo, and I. McCarthy, "Big data for creating and capturing value in the digitalized environment: unpacking the effects of volume, variety, and veracity on firm performance," *Journal of Product Innovation Management*, vol. 38, no. 1, pp. 49–67, 2021.

[13] D. Rangel-Martinez, K. Nigam, and L. A. Ricardez-Sandoval, "Machine learning on sustainable energy: A review and outlook on renewable energy systems, catalysis, smart grid and energy storage," *Chemical Engineering Research and Design*, vol. 174, pp. 414–441, 2021.

[14] M. Alazab, S. Khan, S. S. R. Krishnan, Q.-V. Pham, M. P. K. Reddy, and T. R. Gadekallu, "A multidirectional lstm model for predicting the stability of a smart grid," *IEEE Access*, vol. 8, pp. 85454–85463, 2020.

[15] V. Malbasa, C. Zheng, P.-C. Chen, T. Popovic, and M. Kezunovic, "Voltage stability prediction using active machine learning," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 3117–3124, 2017.

[16] Y. Xu, Z. Dong, K. Meng, R. Zhang, and K. Wong, "Real-time transient stability assessment model using extreme learning machine," *IET generation, transmission & distribution*, vol. 5, no. 3, pp. 314–322, 2011.

[17] V. Arzamasov, K. Böhm, and P. Jochem, "Towards concise models of grid stability," in *2018 IEEE international conference on communications, control, and computing technologies for smart grids (SmartGridComm)*, pp. 1–6, IEEE, 2018.

[18] D. Moldovan and I. Salomie, "Detection of sources of instability in smart grids using machine learning techniques," in *2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 175–182, IEEE, 2019.

[19] N. Shams, P. Wall, and V. Terzija, "Active power imbalance detection, size and location estimation using limited pmu measurements," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1362–1372, 2018.

[20] Y. Zhang, X. Shi, H. Zhang, Y. Cao, and V. Terzija, "Review on deep learning applications in frequency analysis and control of modern power system," *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107744, 2022.

[21] Y. Wang and N. Wang, "Study on an extreme classification of cost-sensitive classification algorithm," in *Big Data Analytics for Cyber-Physical System in Smart City: BDCPS 2019, 28-29 December 2019, Shenyang, China*, pp. 1772–1782, Springer, 2020.

[22] A. Tharwat, "Classification assessment methods," *Applied computing and informatics*, vol. 17, no. 1, pp. 168–192, 2020.