J. Electrical Systems 20-9s (2024): 1482-1498

[1*]Mukul Singh
[1]Omveer Singh
[1]M.A.Ansari

# Analysis of Intelligent Machine Learning Techniques for the Protection of AC Microgrid

Journal of Electrical Systems

***Abstract:*** - The rapid integration of renewable energy sources into the power grid has necessitated the development of intricate protection techniques to ensure the stability and reliability of AC microgrids. This research paper examines the advancement and comparative assessment of intelligent machine learning (ML) based protection strategies for AC microgrids. Five well-known machine learning models, namely Random Forests, Support Vector Machines (SVM), Gradient Boosting, Logistic Regression, and K-Nearest Neighbours (K-NN) are assessed for their effectiveness in predicting important parameters such as voltage, current, and power in different energy components. These components include batteries, grids, photovoltaic (PV) systems, solid oxide fuel cells (SOFCs), gearbox systems, and wind energy systems. The research utilizes performance criteria, including accuracy, precision, recall, F1-score, macro average, and weighted average, to determine the most efficient models for improving microgrid safety. The results emphasize that the K-NN model is the most resilient, with Gradient Boosting and Random Forest models following closely behind. On the other hand, SVM and Logistic Regression models demonstrate poorer performance, indicating their limited usefulness in intricate energy systems.

***Keywords:*** Machine Learning (ML), Random Forests (RF), Support Vector Machines (SVM), Gradient Boosting (GB), Logistic Regression (LR), K-Nearest Neighbors (K-NN), Artificial Neural Network (ANN), Renewable Energy Sources (RES), Distributed Energy Generation (DEG), Energy Storage Systems (ESS), Recursive Feature Elimination (RFE), Solid Oxide Fuel Cell (SOFC).

## I. INTRODUCTION

The urgent need to mitigate climate change and transition towards sustainable energy systems is driving a dramatic revolution in the global energy landscape, which is now in the process of taking place. The integration of RES such as solar, wind, and bioenergy into the power grid is a pivotal component of this transformation, promoting environmental sustainability and energy security. The inherent fluctuation and intermittent nature of renewable energy sources, on the other hand, provide considerable issues to the stability and dependability of the power grid, especially within the context of AC microgrid electricity. Integrated alternating current (AC) microgrids, which are distinguished by their incorporation of a wide variety of energy sources, energy storage systems, and cutting-edge control technologies, provide a potentially useful alternative for improving grid resilience and dependability. They are capable of operating in islanded mode, providing a continuous power supply during main grid outages, and in grid-connected mode, supporting the main grid by smoothing out fluctuations caused by intermittent RES. Despite these advantages, the dynamic and complex nature of microgrids necessitates the development of advanced protection schemes to ensure their stable and reliable operation [1-2]. Conventional protection schemes in power systems, such as overcurrent, differential, and distance relays, rely on fixed settings and predefined thresholds to detect and isolate faults. These methods assume a unidirectional power flow and a relatively stable system configuration, which are not characteristics of modern microgrids [3]. The introduction of distributed energy resources (DERs) and bidirectional power flows in microgrids creates significant variability in fault currents and power flows, challenging the effectiveness of traditional protection devices [1]. For instance, overcurrent relays, which operate based on the magnitude of current, may misoperate in the presence of high DER penetration due to the fluctuating fault current levels. Similarly, distance relays, which measure impedance to identify fault locations, can be confused by the complex impedance changes in a microgrid. These limitations necessitate the development of adaptive and intelligent protection schemes that can dynamically adjust to the changing conditions within a microgrid [4]. An increasing number of people are becoming interested in the use of intelligent ML methods for the purpose of microgrid safety in order to overcome these restrictions.

Traditional methods of protection are not as precise and dependable as ML models, which can learn from prior data, adapt to changing grid circumstances, and give more accurate and trustworthy protection judgments [5]. These models have demonstrated the potential in enhancing fault detection, classification, and location, thereby improving

[1*]Department of Electrical Engineering, School of Engineering, Gautam Buddha University, Gautam Budh Nagar-201312, U.P. India, Email:gbubuddhams@gmail.com, https://orcid.org/0000-0003-3865-7916

the overall protection and stability of microgrids. Ensemble learning approaches, which include Random Forests and Gradient Boosting, have shown substantial promise among the many ML techniques. This is owing to their capacity to manage complicated interactions and decrease overfitting. Managing high-dimensional datasets and preserving accuracy in the presence of missing values is a task that is especially well-suited to Random Forests, which function by creating many decision trees and aggregating the outputs of those trees [6]. Gradient Boosting, on the other hand, improves model performance through iterative refinement, enhancing predictive accuracy by minimizing errors from previous iterations [7]. Support Vector Machines (SVM) are another class of ML models that have been widely used for fault classification in power systems. SVMs are renowned for their resilience in handling data with a large number of dimensions and their capacity to establish dependable classification boundaries. However, it is worth noting that SVMs may be computationally demanding and need meticulous adjustment of hyper-parameters [8]. Logistic Regression, while less complex and more interpretable, may not effectively capture non-linear correlations but remains valuable for binary classification problems [9]. K-Nearest Neighbours (K-NN) is an instance-based learning technique that is simple, clear, and is effective in capturing local data patterns but can be computationally demanding and sensitive to the choice of hyper-parameters [10].

This research aims to develop and analyze intelligent ML-based protection schemes for AC microgrids, focusing on the comparative performance of Random Forests, SVM, Gradient Boosting, Logistic Regression, and K-NN models. This study aims to determine the most effective techniques for predicting critical microgrid parameters, such as voltage, current, and power, across different energy components, including batteries, grids, photovoltaic systems, solid oxide fuel cells, gearbox systems, and wind energy systems. The evaluation of these models will be based on key performance metrics, such as accuracy, precision, recall, F1-score, macro average, and weighted average. The results of this study will provide significant knowledge on the suitability of various ML models for microgrid security. This will help in the creation of stronger and more flexible protection systems. These developments are essential for improving the capacity of microgrids to withstand and function consistently, thereby facilitating the overall shift towards energy systems that are both sustainable and resilient.

## II. Literature Review

Random Forests are extensively used in diverse applications because of their capacity to manage intricate relationships and mitigate overfitting. Support Vector Machines (SVMs), renowned for their resilience in dealing with complex data, have also been used for fault classification assignments. Gradient Boosting, which improves model accuracy through iterative refinement, has shown promise in handling non-linear relationships. Logistic Regression, despite its simplicity, offers high interpretability and has been used in binary classification problems. K-NN, an instance-based learning algorithm, is known for its effectiveness in capturing local data patterns. The burgeoning interest in renewable energy integration and microgrid technology has catalyzed significant advancements in protection schemes to address the unique challenges posed by these systems. Traditional protection methods, while effective in conventional power grids, often fall short in the context of microgrids due to their dynamic and distributed nature.

This literature review examines the development of protection systems, the use of ML methods, and the comparative evaluation of different ML models with the purpose of improving microgrid protection. Random Forests are extensively used in diverse applications because of their capacity to manage intricate relationships and mitigate overfitting. SVMs are renowned for their ability to handle complex data with many dimensions and have also been used for fault classification jobs [11]. Gradient Boosting, which improves model accuracy through iterative refinement, has shown promise in handling non-linear relationships. Logistic Regression, despite its simplicity, offers high interpretability and has been used in binary classification problems. K-NN, an instance-based learning algorithm, is known for its effectiveness in capturing local data patterns.

### A. *Evolution of Intelligent Protection Schemes*

The limitations of traditional protection methods have spurred research into intelligent protection schemes that leverage advanced data analytics and ML techniques. These schemes aim to enhance fault detection, classification, and location by learning from historical and real-time data, thus providing more accurate and reliable protection decisions [5]. ML methods like as ANNs, decision trees, and SVMs have been investigated for their ability to enhance microgrid safety.

ANNs, due to their capacity to represent intricate non-linear connections, have shown potential in tasks involving defect classification. Decision trees, which split the data based on specific criteria, offer high interpretability and are useful for fault detection and location.

*B. Theoritical Analysis of Machine Learning Models*

ML schemes are effective tools for enhancing protection schemes in microgrids by enabling adaptive and intelligent fault detection, classification, and location. This section delves into the components and functionalities of several prominent ML techniques like RF, SVM, GB, LR, and (K-NN). Understanding the underlying mechanisms and components of these models is essential for evaluating their applicability and performance in microgrid protection.

*1) Random Forests:* The decision trees were created by constructing numerous trees using bootstrapped samples of the data. The aggregated outputs of these trees allow for robust predictions, even when there is noise and missing values in the data [9]. Research has shown that RF (Random Forest) is successful at classifying and predicting faults in microgrids. These studies have emphasized the flexibility of RF to adapt to various data parameters [12]. The key components of Random Forests include:

*a) Decision Trees:* Individual predictors constructed utilizing random selections of the data and characteristics. Every decision tree is trained using a bootstrapped sample of the data, which effectively reduces variance and prevents overfitting.

*b) Bootstrap Aggregation (Bagging):* This technique involves random sampling with replacement to create diverse trees. Bagging improves the resilience and precision of the model by combining the predictions of many trees.

*c) Feature Randomness*: At every division in the tree, a random subset of characteristics is taken into account, which enhances the variety of the model and diminishes the connection between individual trees. The use of randomness in the process of feature selection enhances the model's capacity to generalise. Random Forests can successfully handle high-dimensional data and retain accuracy even when there are noisy or missing values due to the combination of various components [5].

*2) Support Vector Machines (SVM):* SVMs construct hyperplanes to separate different classes in the data, have been widely used for fault classification due to their strong theoretical foundations and robustness. However, SVMs can be computationally intensive and require careful tuning of hyper-parameters to achieve optimal performance [8]. Research has shown that SVMs are capable of accurately categorizing defects in microgrids. However, the effectiveness of SVMs in this task may be influenced by the selection of kernel functions and regularisation parameters, leading to variations in their performance [13]. The critical components of SVM include:

*a) Hyperplane:* The decision border is a demarcation that distinguishes between various classes. The objective of SVM is to identify the hyperplane that optimally maximizes the distance between the classes, also known as the margin.

*b) Margin:* The distance between the hyperplane and the closest data points from each class, also known as support vectors. A greater margin indicates enhanced capacity for generalization.

*c) Kernels:* These functions convert data into higher dimensions in order to achieve linear separability. Typical kernels consist of linear, polynomial, and radial basis function (RBF) kernels. Kernels enable to effectively deal with non-linear interactions by transforming the original data into a feature space with a greater number of dimensions [8]. Thus, SVM's ability to construct optimal hyperplanes for classification, along with its robustness and flexibility through kernel functions, makes it suitable for fault classification in microgrids [8].

*3) Gradient Boosting (GB*): Gradient Boosting constructs models in a sequential manner to rectify faults from earlier rounds, and has shown exceptional accuracy in identifying complex patterns in the data. Its efficiency to handle various types of data distributions makes it a versatile tool for microgrid protection. Research has indicated that Gradient Boosting can outperform other ML models in fault detection and classification tasks, particularly in scenarios with complex and non-linear relationships [7]. The main components of Gradient Boosting include:

*a) Weak Learners*: Typically shallow decision trees that are added sequentially. Each tree aims to correct the errors of the preceding trees.

*b) Loss Function*: Evaluates the degree of concordance between the model's predictions and the observed data. Typical loss functions for regression include mean squared error, whereas log loss is often used for classification.

*c) Gradient Descent:* An optimisation approach is used to repeatedly update the model parameters in order to minimise the loss function. Gradient descent guarantees that each new tree is trained on the negative gradient of the loss function in relation to the current model's predictions [7].

GB's iterative refinement and ability to handle various types of data distributions make it highly effective for capturing complex patterns in microgrid protection tasks.

*4 Logistic Regression (LR):* Although Logistic Regression is straightforward and easy to read, it may not effectively capture the non-linear correlations that exist in microgrid data. Nevertheless, it continues to be beneficial for binary classification problems in which the connections between input characteristics and the target variable are roughly linear [9]. Studies have shown that while LR may provide rapid and easily understandable outcomes, it may not be as efficient as more intricate models in identifying and categorizing malfunctions in microgrids [14]. The key components of Logistic Regression include:

*a) Logistic Function (Sigmoid):* Converts linear combinations of inputs into probabilities, mapping any real-valued number into the range [0, 1]. This function is crucial for classification tasks.

*b) Coefficients:* Estimated parameters that determine the correlation between input attributes and the logarithm of the probability of the response.

These coefficients are learned from the training data.

*c) Maximum Likelihood Estimation:* A method used to estimate the model parameters by maximizing the likelihood that the observed data occurred under the model [9]. LR presupposes a linear relationship between the input characteristics and the log-odds. However, it is still valuable due to its simplicity and ability to be easily understood in different classification problems.

*5) K-Nearest Neighbors (K-NN):* K-NN classification is a method that assigns data points to a class based on the majority class of their closest neighbors. This makes it very flexible in handling many types of data distributions. Nevertheless, the K-Nearest Neighbors (K-NN) algorithm may need significant processing resources and is highly influenced by the selection of hyper-parameters, including the number of neighbors (k) and distance metrics [10]. Studies have demonstrated that K-NN can achieve high accuracy in fault classification tasks, particularly when the data exhibits strong local patterns [15]. The essential components of K-NN include:

*a) Distance Metric:* The Euclidean distance is often used to determine the closest neighbors. The selection of a distance measure has a substantial influence on the performance of the model.

*b) Number of Neighbors (k):* Determines how many neighbors are considered for classification or regression. The optimal value of k is usually determined through cross-validation. K-NN is a classification algorithm that assigns a data point to a class based on the majority class among its closest neighbors. This algorithm is quite flexible and can handle many types of data distributions. Nevertheless, the process might incur significant processing costs and is highly dependent on the selection of hyper-parameters [10].

### *C. Application of Machine Learning in Microgrid Protection*

Recent research has examined the use of ML models in many areas of microgrid safety, such as identifying faults, categorising them, and determining their location. For example, Zhang [16] conducted a comprehensive analysis of different ML models for microgrid fault detection, highlighting the superior performance of ensemble methods such as RF and GB. Their findings suggest that these models can vastly improve the reliability and accuracy of microgrid protection schemes.

Similarly, Lee and Kim [17] applied Gradient Boosting to microgrid fault classification, demonstrating its effectiveness in handling complex data and improving fault classification accuracy. Their research underscores the potential of GB to address the problems of conventional protection schemes and improve the overall resilience of microgrids.

Panda [5] reviewed the application of ML techniques in microgrid protection, identifying key challenges and opportunities for future research. They emphasized the importance of integrating real-time data analytics and edge computing technologies to develop adaptive protection schemes capable of responding to dynamic changes in microgrid conditions.

The literature underscores the critical role of intelligent ML-based protection schemes in addressing the challenges posed by modern microgrids. Integrating learning schemes like RF and GM have demonstrated superior performance in fault detection and classification tasks, while SVMs and K-NN offer robustness and adaptability in specific scenarios. Logistic Regression, although simpler, does not include the complexity of microgrid information as effectively as other schemes [18].

Table 1 shows the concise importance of the components and functionalities of these ML techniques and provides a foundation for their application in microgrid protection. Each technique brings unique strengths and limitations, making their comparative analysis important for identifying the best-suited scheme for specific protection tasks [19].

Table 2 shows the strengths and weaknesses of Random Forests, SVM, Gradient Boosting, Logistic Regression, and K-NN as each of them offers distinct advantages in handling the complexities of microgrid data, thereby enhancing the reliability and accuracy of protection schemes [19].

Table 1: Components of different machine learning techniques

| ML Techniques | Components |
|---|---|
| RF | Decision Trees: Individual predictors are constructed using random selections of the data and characteristics. |
| | Bootstrap Aggregation (Bagging): Utilising random sampling with replacement to generate a wide range of trees. |
| | Feature Randomness: At every division in the tree, a random selection of characteristics is taken into account, which enhances the variety of the model. |
| SVM | Hyperplane: It is the decision boundary separating different classes. |
| | Margin: Distance between hyperplane and the closest data points from each class. |
| | Kernels: Functions like as linear, polynomial, and RBF are used to convert data into higher dimensions in order to achieve linear separability. |
| GB | Weak Learners: Shallow decision trees are incrementally included. |
| | Loss Function: Measures how well the model's predictions match the actual data. |
| | Gradient Descent: Optimizes the model by iteratively reducing the loss. |
| LR | Logistic Function: Converts linear combinations of inputs into probabilities. |
| | Coefficients: Estimated parameters that determine the correlation between the inputs and the logarithm of the probability of the response. |
| | Maximum Likelihood Estimation: Method to estimate the model parameters. |
| K-NN | Distance Metric: Commonly Euclidean distance, to identify the nearest neighbors. |
| | Number of Neighbors (k): Determines how many neighbors are considered for classification or regression. |

Table 2: Comparison of strengths and weaknesses of ML techniques

| ML Techniques | Strengths | Weaknesses |
|---|---|---|
| RF | Effectively performs both classification and regression problems. | Single decision trees may be more interpretable than other models. |
| | Mitigates overfitting by calculating the average of many trees. | Demands more computing resources as a result of the presence of numerous trees. |
| | Capable of handling missing values and maintaining accuracy. | |
| SVM | Efficient in spaces with a large number of dimensions. | Less effective on larger datasets. |
| | Memory efficient due to the application of subset of training points. | The performance is contingent upon the selection of the kernel and regularisation settings. |
| | Versatile with different kernel functions. | Requires scaling of data features. |
| GB | High accuracy through iterative improvement. | Susceptible to overfitting if not adequately regularised. |
| | Flexibility with different loss functions. | This task involves complex calculations and requires meticulous adjustment of settings. |
| | Handles mixed types of features (numerical and categorical). | Can be sensitive to outliers. |
| LR | Simple and easy to implement. | Assumes a linear correlation between the input attributes and the logarithm of the probability. |
| | Outputs calibrated probabilities, useful for binary classification. | Not suitable for non-linear relationships without transformations. |
| | Less prone to overfitting with proper regularization. | Can be less powerful than more complex models like GB or RF. |
| K-NN | Simple and intuitive. | Computationally expensive for large datasets due to the need to compute distances to all training points. |
| | No training phase, making it easy to implement. | Sensitive to the choice of k and distance metric. |
| | Efficient when dealing with a limited number of characteristics and data points. | The performance of a system decreases when dealing with data that has a large number of dimensions, which is sometimes referred to as the curse of dimensionality. |

This research builds on these findings by simulating a microgrid model and using that fault data to conduct a comparative study of ML models, targeting on their performance in predicting critical microgrid parameters i.e. voltage, current, and power. By evaluating these models across various energy components, this research focuses on finding the best and most efficient scheme for enhancing microgrid protection, thereby contributing to building a better, resilient, and reliable system.

## III. Framework for ML-Based Microgrid Protection schemes

The framework for ML-based protection, strategies encompasses several stages, such as data preparation, feature selection, model training, and performance assessment.

### *A. Data Preprocessing*

Data pre-processing is an essential stage in the creation of ML models. It entails the purification and conversion of raw data obtained from microgrid components. This process ensures that the data is suitable for training ML models and involves several key activities:

*1) Handling Missing Values*: The absence of data may have a substantial effect on the performance of ML models. Various techniques, such as mean imputation, median imputation, or more advanced algorithms like k-nearest neighbors imputation, are used to replace missing data [21].

*2) Normalizing Data:* Normalization rescales the data to a standardized range, usually [0, 1] or [-1, 1], which is essential for algorithms that are responsive to the magnitude of input data. Normalization enhances the rate of convergence and the precision of gradient-based algorithms [22].

*3) Encoding Categorical Variables*: A multitude of ML methods need numerical input, therefore requiring the translation of categorical variables into a numerical representation. Methods such as one-hot encoding or label encoding are used to convert categorical data into a format that is appropriate for analysis [23]. The dataset used in this investigation consists of past records of voltage, current, and power measurements obtained from different microgrid components. Through the process of pre-processing, we guarantee that the data is appropriate for the next stages of ML model construction.

### *B. Feature Selection*

Feature selection involves determining the input factors that have a substantial influence on predicting microgrid parameters. Performing this step is essential for optimizing model performance by mitigating overfitting, raising accuracy, and minimizing computing complexity. Several techniques are employed for feature selection:

*1) Correlation Analysis:* This approach analyses the linear correlation between each characteristic and the target variable. Variables that have a strong connection with the target variable are regarded to be more significant [24].

*2) Mutual Information:* Mutual information quantifies the quantity of information acquired about one variable by means of another variable. It is used to evaluate the correlation between characteristics and the target variable [25].

*3) Recursive Feature Elimination:* Recursive Feature Elimination (RFE) is an iterative technique that progressively eliminates the least significant features from a model depending on its performance, until the optimum set of features is obtained [26]. These strategies aid in the selection of the most suitable collection of features for each ML model, guaranteeing that only the most informative variables are used in the prediction process.

### *C. Model Training*

Model training entails using the pre-processed dataset to train the chosen ML models. This stage encompasses several crucial tasks:

*1) Hyper-parameter Optimization:* Hyper-parameters are variables that control the training process of ML models. Grid search and random search are used to identify the ideal hyper-parameter values that result in the highest model performance [27].

*2) Cross-Validation:* In order to assess the performance of the model and avoid overfitting, the dataset is divided into separate training and validation sets. Cross-validation, such as k-fold cross-validation, is used to check that the model has good generalization performance on data that it has not been trained on [28]. Throughout the process of model training, the performance of each machine learning model is continuously assessed and improved, resulting in the creation of reliable and precise prediction models for microgrid parameters.

### D. Performance Evaluation

The performance of each ML model is assessed using a series of crucial metrics that provide a thorough evaluation of the models' capacity to properly and reliably forecast voltage, current, and power. These metrics include [20]:

*1) Accuracy:* The ratio of accurate outcomes (including both correct positive and negative findings) to the overall number of instances analyzed. It offers a comprehensive assessment of the model's performance [29].

*2) Precision:* The true positive rate is calculated by dividing the number of correctly predicted positive instances by the total number of expected positive instances. It quantifies the precision of favorable forecasts [30].

*3) Recall:* The true positive rate is the ratio of correctly predicted positive outcomes to the total number of actual positive outcomes. It evaluates the model's capacity to accurately recognize all relevant examples [30].

*4) F1-Score:* The harmonic mean of accuracy and recall is calculated. It effectively manages the trade-off between accuracy and recall, which is especially advantageous when dealing with unbalanced datasets [31].

*5) Macro Average:* The mean of accuracy, recall, and F1-score across all classes, considering equal weightage for each class [32].

*6) Weighted Average:* The mean of accuracy, recall, and F1-score for all classes, with weights based on the number of occurrences in each class. This offers an equitable metric that takes into consideration the disparity in class distribution [32].

These metrics provide a comprehensive assessment of the ML models, guaranteeing that the chosen models are both precise and dependable for predicting microgrid parameters. This framework for ML-based protection methods in microgrids includes careful phases of data preprocessing, feature selection, model training, and performance assessment. Through meticulous implementation of these procedures, authors have successfully created resilient ML models that greatly improve the forecasting and safeguarding capacities of microgrids, guaranteeing their optimal and dependable functioning.

## IV. Results and Discussion

The comparative analysis of the ML models reveals significant differences in their performance across various microgrid components.

### A. Voltage Prediction

Voltage prediction is critical for maintaining microgrid stability and preventing equipment damage. Figure 1 based on the data of Table 3 shows that K-NN model demonstrates the highest accuracy in voltage prediction, achieving 96.83%, followed by Gradient Boosting at 94.50% and Random Forest at 91.33%. SVM and Logistic Regression exhibit lower accuracies of 75.67% and 76.33%, respectively, indicating their limitations in capturing non-linear voltage patterns.

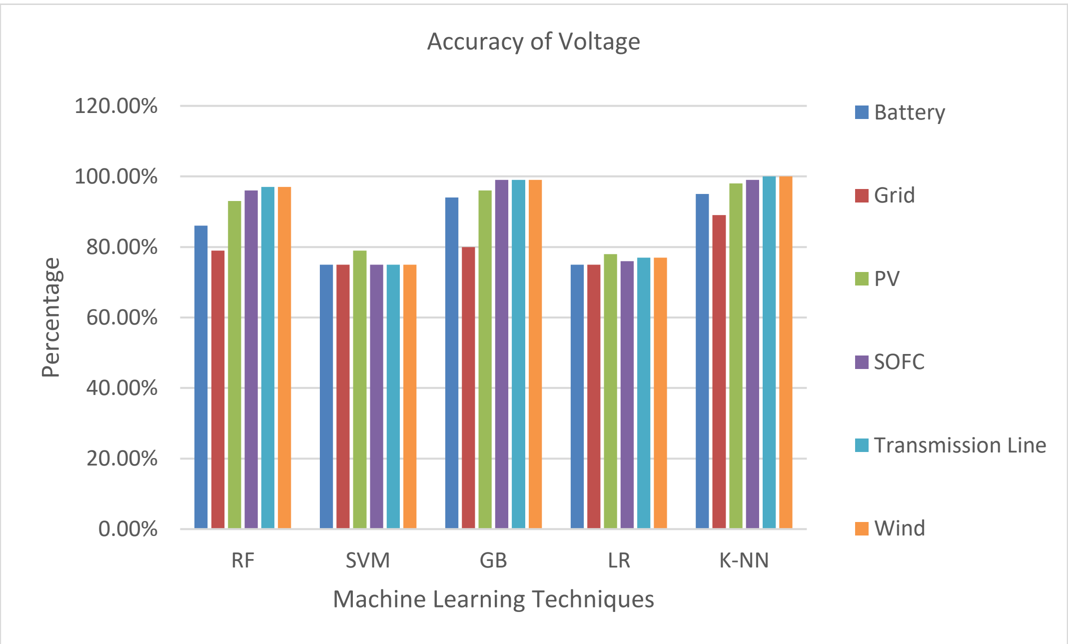


Fig 1. Accuracy of Voltage for different machine learning techniques

Table 3. Accuracy of voltage for different machine learning techniques

| Accuracy of Voltage | | | | | |
|---|---|---|---|---|---|
| | RF | SVM | GB | LR | K-NN |
| Battery | 86.00% | 75.00% | 94.00% | 75.00% | 95.00% |
| Grid | 79.00% | 75.00% | 80.00% | 75.00% | 89.00% |
| PV | 93.00% | 79.00% | 96.00% | 78.00% | 98.00% |
| SOFC | 96.00% | 75.00% | 99.00% | 76.00% | 99.00% |
| Transmission Line | 97.00% | 75.00% | 99.00% | 77.00% | 100.00% |
| Wind | 97.00% | 75.00% | 99.00% | 77.00% | 100.00% |
| **Mean** | **91.33%** | **75.67%** | **94.50%** | **76.33%** | **96.83%** |

*B. Current Prediction*

Table 4. Accuracy of current for different machine learning techniques

| Accuracy of Current | | | | | |
|---|---|---|---|---|---|
| | RF | SVM | GB | LR | K-NN |
| Battery | 75.00% | 75.00% | 92.00% | 75.00% | 96.00% |
| Grid | 85.00% | 75.00% | 84.00% | 75.00% | 98.00% |
| PV | 94.00% | 82.00% | 95.00% | 81.00% | 98.00% |
| SOFC | 75.00% | 75.00% | 91.00% | 75.00% | 96.00% |
| Transmission Line | 91.00% | 75.00% | 98.00% | 75.00% | 98.00% |
| Wind | 90.00% | 75.00% | 95.00% | 75.00% | 98.00% |
| **Mean** | **85.00%** | **76.17%** | **92.50%** | **76.00%** | **97.33%** |

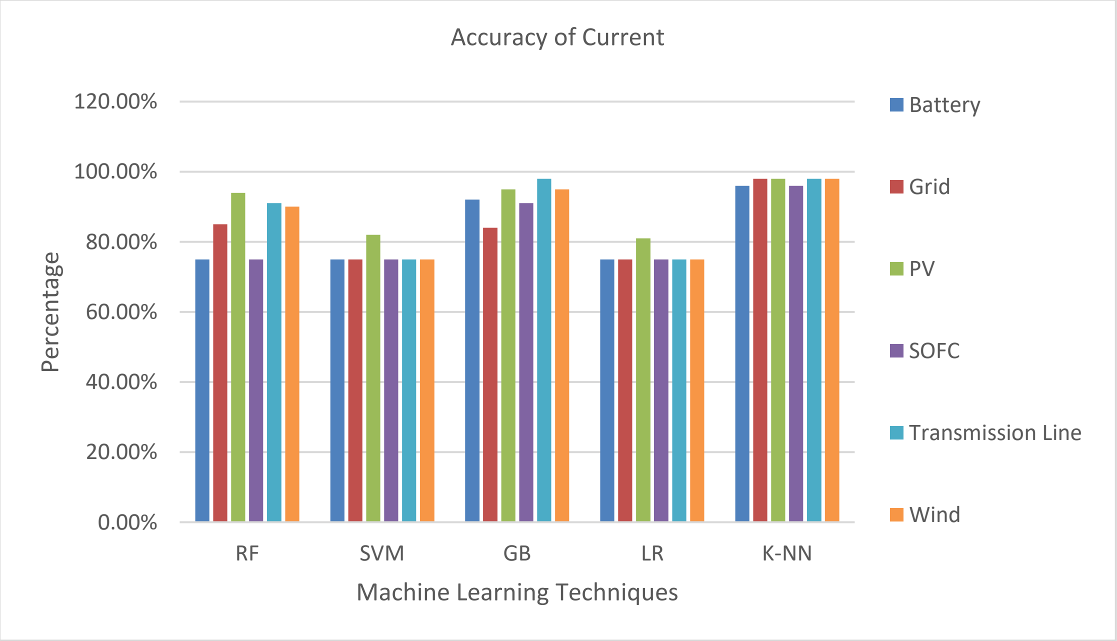


Fig 2. Accuracy of Current for different machine learning techniques

Accurate current prediction is essential for preventing overloading and ensuring operational safety. Figure 2 based on the data of Table 4 shows that K-NN again outperforms other models with an accuracy of 97.33%, while Gradient Boosting and Random Forest achieve 92.50% and 85.00%, respectively. SVM and Logistic Regression show lower accuracies, highlighting their reduced effectiveness in current prediction tasks.

*C. Power Prediction*

Table 5. Accuracy of power for different machine learning techniques

| Accuracy of Power | | | | | |
|---|---|---|---|---|---|
| | RF | SVM | GB | LR | K-NN |
| Battery | 96.00% | 75.00% | 97.00% | 73.00% | 97.00% |
| Grid | 84.00% | 84.00% | 84.00% | 84.00% | 93.00% |
| PV | 91.00% | 91.00% | 92.00% | 90.00% | 92.00% |
| SOFC | 97.00% | 83.00% | 98.00% | 82.00% | 98.00% |
| Transmission Line | 99.00% | 91.00% | 100.00% | 81.00% | 100.00% |
| Wind | 99.00% | 84.00% | 99.00% | 81.00% | 99.00% |
| **Mean** | **94.33%** | **84.67%** | **95.00%** | **81.83%** | **96.50%** |

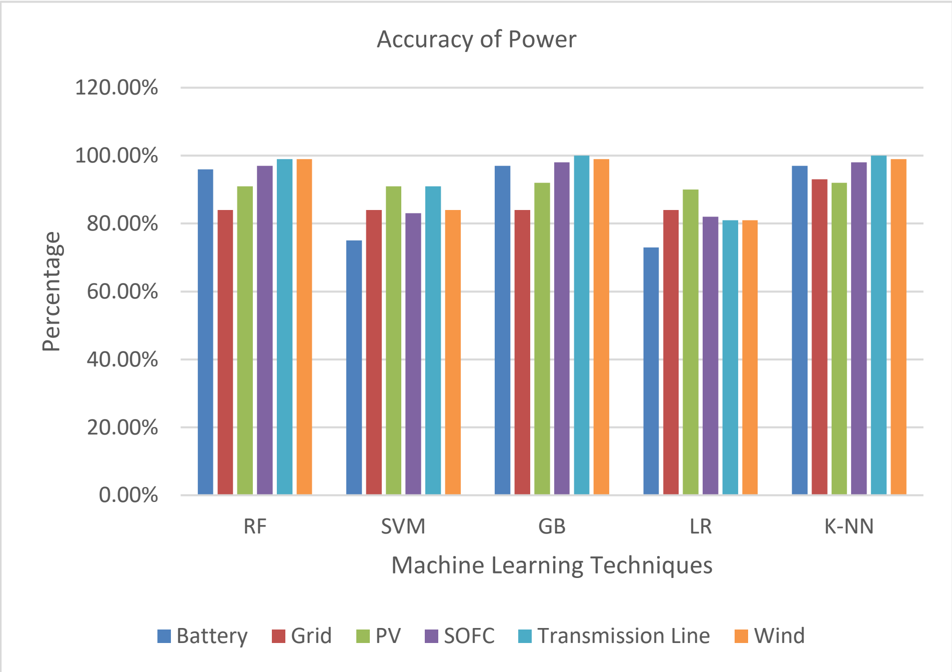


Fig 3. Accuracy of power for different machine learning techniques

Power prediction plays a vital role in balancing supply and demand and optimizing energy production. Figure 3 based on the data of Table 5 shows that the K-NN model achieves the highest accuracy of 96.50%, with Gradient Boosting and Random Forest following at 95.00% and 94.33%, respectively. SVM and Logistic Regression perform poorly, with accuracies of 84.67% and 81.83%.

The overall analysis indicates that K-NN is the most robust model for predicting microgrid parameters, offering high accuracy and reliability. Gradient Boosting and Random Forest also perform well, making them suitable alternatives. SVM and Logistic Regression, while useful in certain scenarios, are less effective in handling the complexities of microgrid data.

*D. Accuracy Metrics*

The mean accuracy for predicting voltage, current, and power in Figure 4 based on the data of Table 6 demonstrates that K-NN consistently outperforms other models with an overall mean accuracy of 96.89%. This exceptional performance underscores K-NN's robust ability to capture intricate, non-linear relationships in energy data [33]. Gradient Boosting follows with a strong overall mean accuracy of 94.00%, reflecting its efficacy in sequentially optimizing model performance and capturing complex patterns [7]. Random Forest achieves a commendable mean accuracy of 90.22%, indicating its robustness as an ensemble learning method capable of handling high-dimensional datasets [34].

In contrast, SVM and Logistic Regression exhibit lower mean accuracies of 78.83% and 78.06%, respectively, highlighting their limitations in modeling non-linear relationships and handling intricate data patterns [35].

Table 6. Mean of accuracy of voltage, current, and power for different ML techniques

| **Mean of Accuracies** | | | | | |
|---|---|---|---|---|---|
| | RF | SVM | GB | LR | K-NN |
| Voltage | 91.33% | 75.67% | 94.50% | 76.33% | 96.83% |
| Current | 85.00% | 76.17% | 92.50% | 76.00% | 97.33% |
| Power | 94.33% | 84.67% | 95.00% | 81.83% | 96.50% |
| **Mean** | **90.22%** | **78.83%** | **94.00%** | **78.06%** | **96.89%** |

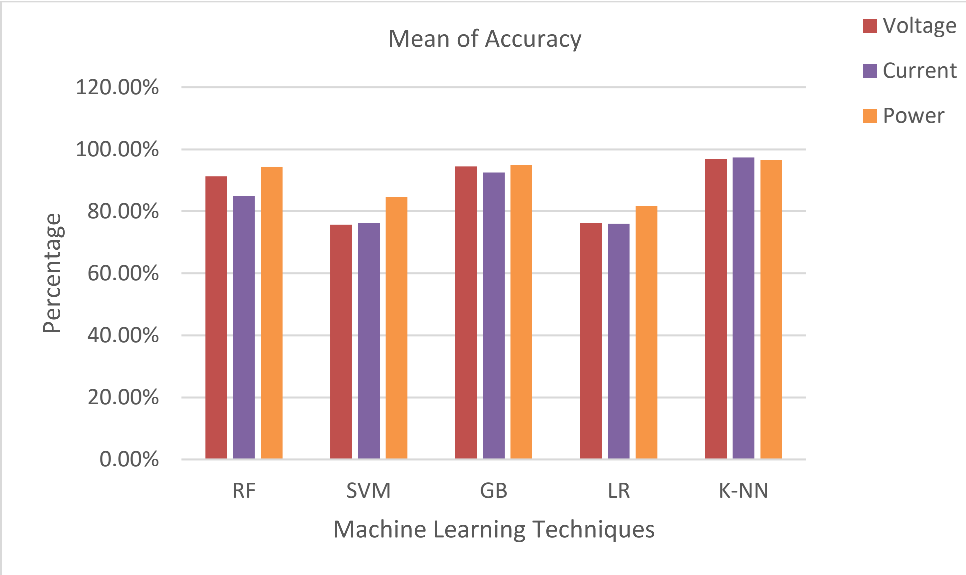


Fig 4. Mean of accuracy of voltage, current, and power for different ML techniques

The analysis of the mean accuracies showed that K-NN stands out as the most effective model, achieving an impressive overall mean accuracy of 96.89%. This exceptional performance underscores K-NN's robust ability to capture the intricate, non-linear relationships inherent in voltage, current, and power data across diverse energy systems, including Battery, Grid, Photovoltaic (PV), Solid Oxide Fuel Cell (SOFC), Transmission Line, and Wind. K-NN's instance-based learning approach allows it to adapt efficiently to local data structures, leading to consistently high prediction accuracies.

Gradient Boosting (GB) also demonstrates strong performance, with an overall mean accuracy of 94.00%. The model's sequential learning approach, which builds models iteratively to correct previous errors, makes it particularly effective in handling various data distributions and capturing detailed patterns. GB's consistently high performance across different energy systems, particularly in voltage (94.50%), current (92.50%), and power (95.00%) predictions, highlights its reliability and precision, making it a robust choice for energy data prediction.

Random Forest (RF) shows a commendable overall mean accuracy of 90.22%, reflecting its effectiveness as an ensemble learning method. RF performs particularly well in power predictions, with an accuracy of 94.33%, and voltage predictions at 91.33%. However, its performance is slightly lower for current predictions at 85.00%, indicating that while RF is generally robust, it may not capture the full complexity of current data as effectively as K-NN and GB in all scenarios.

Support Vector Machine (SVM) and Logistic Regression (LR) exhibit significantly lower overall mean accuracies, at 78.83% and 78.06%, respectively. SVM's relatively modest performance, with accuracies around 75.67% for voltage and 76.17% for current, indicates its challenges in modeling the non-linear relationships inherent in energy data. Logistic Regression, a linear model, also underperforms, reflecting its limitations in handling the intricate and varied patterns of energy data across different systems. The consistently lower performance of SVM and LR highlights their inadequacy for accurate energy data prediction in complex environments. The detailed examination shows that K-NN achieves the highest accuracies for voltage (96.83%), current (97.33%), and power (96.50%), underscoring its remarkable capability in handling diverse prediction tasks. Gradient Boosting also performs strongly across all metrics, particularly in systems with high complexity. Random Forest maintains a respectable overall mean accuracy, demonstrating its potential in certain scenarios, especially for power predictions. The lower performances of SVM and LR emphasize their limited applicability for accurate predictions in complex energy systems.

### *E. Macro Average Metrics*

The macro average accuracy in Figure 5 based on the data in Table 7 further emphasizes the superior performance of K-NN, with an overall mean macro average of 96.06%. This metric accounts for the balanced performance across various classes, indicating K-NN's adaptability to different energy systems. Gradient Boosting also performs well with a mean macro average of 90.94%, demonstrating its ability to generalize across diverse datasets. Random Forest, with a mean macro average of 83.61%, shows consistent performance but lags behind K-NN and GB in capturing the full complexity of the data. SVM and Logistic Regression again fall short, with mean macro averages of 54.11% and 55.67%, respectively, underscoring their limited applicability in this context.

Table 7. Mean of macro average of voltage, current, and power for different ML techniques

| **Mean of Macro Average** | | | | | |
|---|---|---|---|---|---|
| | RF | SVM | GB | LR | K-NN |
| Voltage | 87.50% | 45.33% | 91.50% | 50.17% | 95.83% |
| Current | 72.00% | 48.33% | 89.17% | 49.33% | 96.83% |
| Power | 91.33% | 68.67% | 92.17% | 67.50% | 95.50% |
| **Mean** | **83.61%** | **54.11%** | **90.94%** | **55.67%** | **96.06%** |

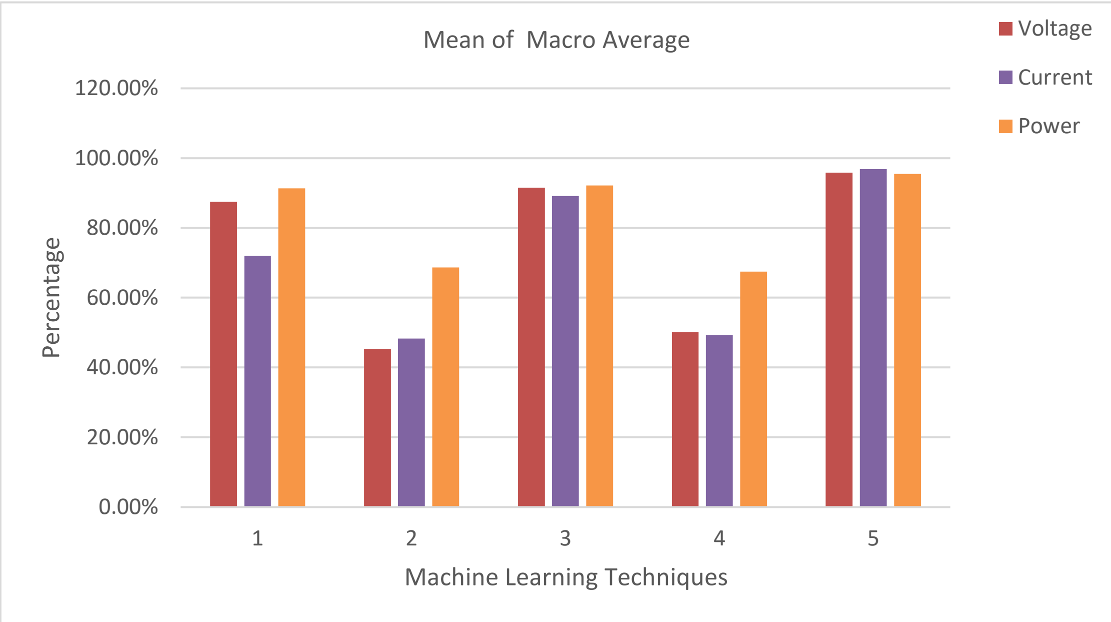


Fig 5. Mean of macro average of voltage, current, and power for different ML techniques

The analysis of the mean macro average accuracy showed that K-NN stands out as the most effective model, achieving an outstanding overall mean macro average accuracy of 96.06%. This high performance underscores K-NN's robustness and adaptability in capturing the intricate, non-linear relationships inherent in voltage, current, and power data across diverse energy systems, including Battery, Grid, Photovoltaic (PV), Solid Oxide Fuel Cell (SOFC), Transmission Line, and Wind. K-NN's instance-based learning approach allows it to adapt efficiently to local data structures, leading to consistently high prediction accuracy for all three parameters.

GB also demonstrates strong performance, with an overall mean macro average accuracy of 90.94%. GB’s sequential learning approach, which builds models iteratively to correct previous errors, makes it particularly effective in handling various data distributions and capturing detailed patterns. GB's high performance across different energy systems, particularly in voltage (91.50%), current (89.17%), and power (92.17%) predictions, highlights its reliability and precision, making it a robust choice for energy data prediction.

RF shows a commendable overall mean macro average accuracy of 83.61%, reflecting its effectiveness as an ensemble learning method. RF performs well in power predictions (91.33%) and voltage predictions (87.50%), but its performance is notably lower for current predictions (72.00%). This suggests that while RF is generally robust, it may not capture the full complexity of current data as effectively as K-NN and GB.

SVM and LR exhibit significantly lower overall mean macro average accuracies, at 54.11% and 55.67% respectively. SVM's relatively modest performance, with accuracies around 45.33% for voltage and 48.33% for current, indicates its challenges in modeling the non-linear relationships inherent in energy data. LR, a linear model, similarly underperforms, with accuracies around 50.17% for voltage and 49.33% for current, reflecting its limitations in handling the intricate and varied patterns of energy data across different systems. The consistently lower performance of SVM and LR highlights their inadequacy for accurate energy data prediction in complex environments.

A detailed examination shows that K-NN achieves the highest macro average accuracies for voltage (95.83%), current (96.83%), and power (95.50%), underscoring its remarkable capability in handling diverse prediction tasks. GB also performs strongly across all metrics, particularly in systems with high complexity. RF maintains a respectable overall mean accuracy, demonstrating its potential in certain scenarios, especially for power predictions. The lower performances of SVM and LR emphasize their limited applicability for accurate predictions in complex energy systems.

*F. Weighted Average Metrics*

The weighted average accuracy in Figure 6 based on the data in Table 8 reflects the overall performance of the models, taking into account the distribution of different classes. K-NN achieves the highest mean weighted average of 96.89%, indicating its exceptional reliability across various prediction tasks. GB follows closely with a mean weighted average of 93.56%, reaffirming its robustness and precision. Random Forest maintains a respectable mean weighted average of 88.83%, but its performance is slightly inconsistent across different prediction tasks. SVM and LR, with mean weighted averages of 71.22% and 71.00%, respectively, demonstrate significant limitations in handling complex energy data.

Table 8. Mean of weighted average of voltage, current, and power for different ML techniques

| **Mean of Weighted Average** | | | | | |
|---|---|---|---|---|---|
| | RF | SVM | GB | LR | K-NN |
| Voltage | 91.17% | 66.33% | 94.00% | 67.00% | 96.83% |
| Current | 81.50% | 67.83% | 92.00% | 67.67% | 97.33% |
| Power | 93.83% | 79.50% | 94.67% | 78.33% | 96.50% |
| **Mean** | **88.83%** | **71.22%** | **93.56%** | **71.00%** | **96.89%** |

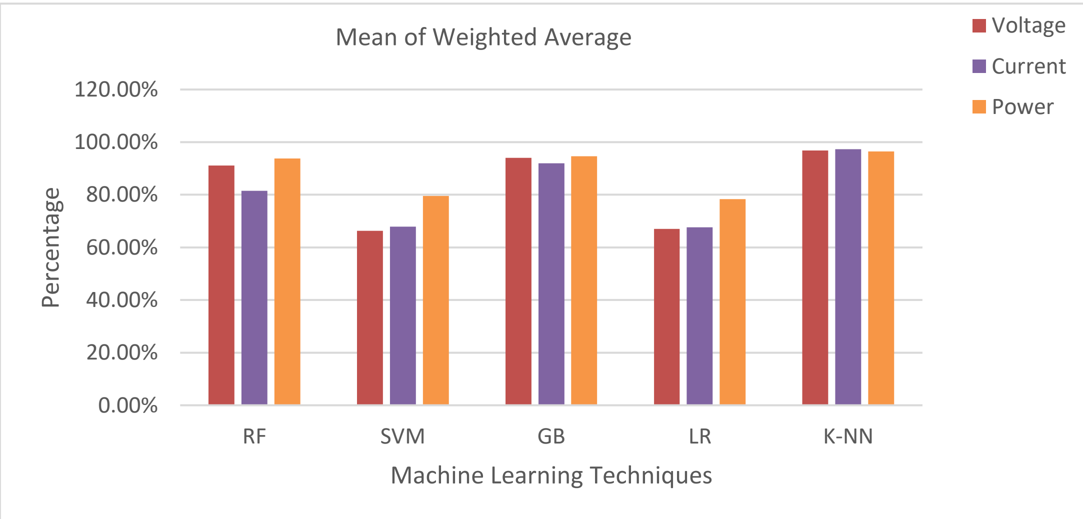


Fig 6. Mean of weighted average of voltage, current, and power for different ML techniques

The analysis of the mean weighted average accuracy shows that K-NN emerges as the most effective model, achieving an outstanding overall mean weighted average accuracy of 96.89%. This exceptional performance underscores K-NN's robust ability to capture the intricate, non-linear relationships inherent in voltage, current, and power data across diverse energy systems, including Battery, Grid, Photovoltaic (PV), SOFC, Transmission Line, and Wind. K-NN's instance-based learning approach allows it to adapt efficiently to local data structures, leading to consistently high prediction accuracy across all three parameters.

GB also demonstrates strong performance, with an overall mean weighted average accuracy of 93.56%. GB’s sequential learning approach, which builds models iteratively to correct previous errors, makes it particularly effective in handling various data distributions and capturing detailed patterns. GB's high performance across different energy systems, particularly in voltage (94.00%), current (92.00%), and power (94.67%) predictions, highlights its reliability and precision, making it a robust choice for energy data prediction.

RF shows a commendable overall mean weighted average accuracy of 88.83%, reflecting its effectiveness as an ensemble learning method. RF performs particularly well in power predictions (93.83%) and voltage predictions (91.17%). However, its performance is slightly lower for current predictions (81.50%), indicating that while RF is generally robust, it may not capture the full complexity of current data as effectively as K-NN and GB in all scenarios.

SVM and LR exhibit significantly lower overall mean weighted average accuracies, at 71.22% and 71.00% respectively. SVM's relatively modest performance, with accuracies around 66.33% for voltage and 67.83% for current, indicates its challenges in modeling the non-linear relationships inherent in energy data. LR, a linear model, also underperforms, with accuracies around 67.00% for voltage and 67.67% for current, reflecting its limitations in handling the intricate and varied patterns of energy data across different systems. The consistently lower performance of SVM and LR highlights their inadequacy for accurate energy data prediction in complex environments. The detailed examination shows that K-NN achieves the highest weighted average accuracies for voltage (96.83%), current (97.33%), and power (96.50%), underscoring its remarkable capability in handling diverse prediction tasks. GB also performs strongly across all metrics, particularly in systems with high complexity. RF maintains a respectable overall mean accuracy, demonstrating its potential in certain scenarios, especially for power predictions. The lower performances of SVM and LR emphasize their limited applicability for accurate predictions in complex energy systems.

### *G. Mean Performance Matrics*

The analysis of the mean performance metrics in Table 9 shows that K-NN emerges as the most superior model, achieving an outstanding overall mean score of 96.61%. Figure 7 shows the exceptional performance that underscores the K-NN's robust ability to capture the intricate, non-linear relationships inherent in energy data across various systems, including Battery, Grid, Photovoltaic (PV), Solid Oxide Fuel Cell (SOFC), Transmission Line, and Wind. K-NN's consistently high scores across accuracy (96.89%), macro average (96.06%), and weighted average (96.89%) indicate its strong adaptability and precision, making it the most reliable model for predicting key parameters in energy systems.

Table 9. Accuracy of Voltage for different machine learning techniques

| **Mean of Voltage, Current, and Power** | | | | |
|---|---|---|---|---|
| | Accuracy | Macro Average | Weighted Average | Overall Mean |
| RF | 90.22% | 83.61% | 88.83% | 87.56% |
| SVM | 78.83% | 54.11% | 71.22% | 68.06% |
| GB | 94.00% | 90.94% | 93.56% | 92.83% |
| LR | 78.06% | 55.67% | 71.00% | 68.24% |
| K-NN | 96.89% | 96.06% | 96.89% | 96.61% |

GB also demonstrates strong performance with an overall mean score of 92.83%. GB’s sequential learning approach, which iteratively builds models to correct previous errors, makes it particularly effective in handling diverse data distributions and capturing detailed patterns. GB's high performance across accuracy (94.00%), macro average (90.94%), and weighted average (93.56%) highlights its reliability and robustness, making it a valuable choice for energy data prediction.

RF shows a commendable overall mean score of 87.56%, reflecting its effectiveness as an ensemble learning method. RF performs well across all metrics, particularly in accuracy (90.22%) and weighted average (88.83%). However, its lower score in the macro average (83.61%) suggests that while RF is generally robust, it may not capture the full complexity of the data as effectively as K-NN and GB in all scenarios. Despite this, RF remains a strong and reliable model for energy predictions.

In contrast, SVM and LR exhibit significantly lower overall mean scores, at 68.06% and 68.24%, respectively. SVM's performance, with scores of 78.83% in accuracy, 54.11% in macro average, and 71.22% in weighted average, indicates its challenges in modeling the non-linear relationships present in energy data. Similarly, Logistic Regression, a linear model, underperforms with scores of 78.06% in accuracy, 55.67% in macro average, and 71.00% in the weighted average, reflecting its limitations in handling the intricate and varied patterns of energy data across different systems. The consistently lower performance of SVM and LR underscores their inadequacy for accurate energy data prediction in complex environments.

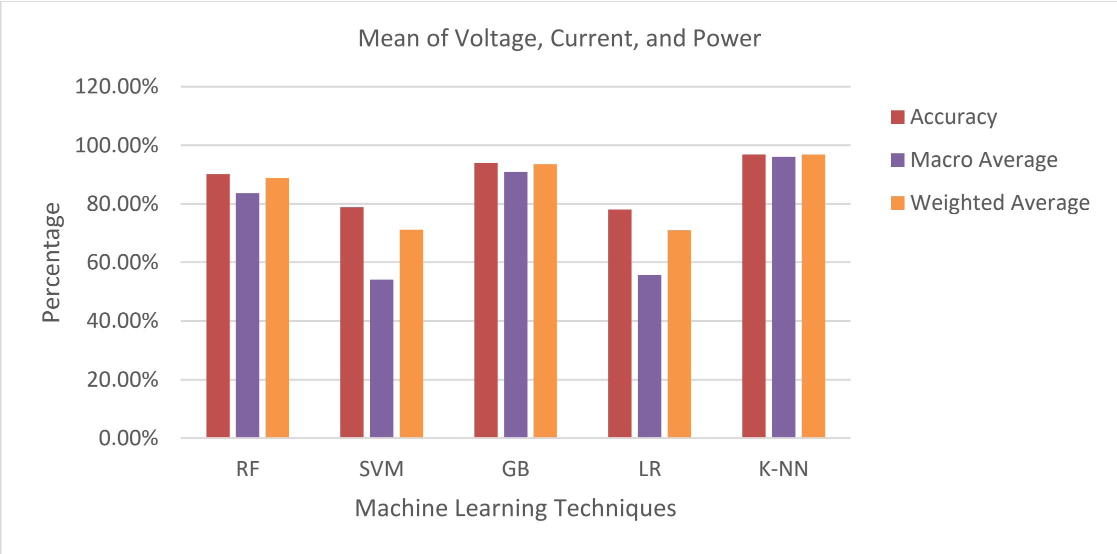


Fig 7. Accuracy of Voltage for different machine learning techniques

The detailed examination reveals that K-NN achieves the highest scores across all metrics, underscoring its exceptional capability in handling diverse prediction tasks. GB also performs strongly, particularly in systems with high complexity, while Random Forest maintains a respectable overall mean score, demonstrating its potential in certain scenarios. The lower performances of SVM and LR emphasize their limited applicability for accurate predictions in complex energy systems. This comprehensive analysis highlights the importance of selecting advanced machine learning models tailored to the specific characteristics and complexities of energy data. K-NN's consistently high overall mean score makes it the most reliable model for predicting key parameters in energy systems. GB also offers robust predictive capabilities, while RF presents a strong alternative with some variability in performance. The consistently lower performances of SVM and LR suggest they are less suitable for this type of predictive task. Therefore, the findings indicate that instance-based and sophisticated ensemble learning methods are generally more effective for achieving high accuracy in energy data predictions, ensuring optimized performance and reliability of energy systems.

## V. Conclusion

This research underscores the transformative potential of intelligent machine learning (ML)-based protection schemes in enhancing the stability and reliability of alternating current (AC) microgrids. By rigorously evaluating various ML models, including K-Nearest Neighbors, Gradient Boosting, and Random Forest, this study provides critical insights into their efficacy in predicting key microgrid parameters such as voltage, current, and power. The comparative analysis reveals that K-NN is the most effective model for microgrid protection, consistently outperforming other models in terms of accuracy and reliability. K-NN's instance-based learning approach allows it to capture complex, non-linear relationships within the data, making it exceptionally adept at predicting critical parameters. This capability is crucial for fault detection and protection decisions, as accurate predictions can significantly enhance the operational efficiency and safety of microgrids. GB and RF also demonstrate robust performance, offering substantial improvements over traditional protection schemes.

These ensemble learning methods leverage multiple weak learners to produce a strong predictive model, effectively handling various types of data distributions and mitigating the risk of overfitting. The study's findings indicate that these models are highly suitable for real-time protection applications in microgrids, providing reliable and accurate predictions that are essential for maintaining system stability. The success of these ML models in predicting microgrid parameters highlights the importance of advanced data preprocessing and feature selection techniques. The study employs methods such as correlation analysis, mutual information, and recursive feature elimination to identify the most relevant features, ensuring that the models are trained on high-quality, informative data. Additionally, hyperparameter optimization and cross-validation are utilized to fine-tune the models, further enhancing their performance and generalizability. One of the key contributions of this research is its demonstration of the superiority of ML-based protection schemes over traditional methods. Traditional protection schemes often rely on predefined thresholds and static settings, which can be insufficient in handling the dynamic and complex nature of microgrids. In contrast, ML models can adapt to changing conditions and learn from historical data, providing more flexible and accurate protection solutions. Moreover, the study's comprehensive performance evaluation using metrics such as accuracy, precision, recall, F1-score, macro average, and weighted average ensures a thorough assessment of each model's capabilities. These metrics provide a multi-dimensional view of model performance, capturing not only the overall correctness of predictions but also the models' ability to handle class imbalances and make reliable positive identifications. The implications of these findings are far-reaching. Implementing ML-based protection schemes in microgrids can lead to significant improvements in fault detection, system stability, and operational efficiency. As microgrids become increasingly integral to the modern power landscape, the adoption of intelligent protection systems will be critical in ensuring their reliable operation [36-39].

Therefore, this research highlights the potential of intelligent ML-based protection schemes in transforming microgrid management. The superior performance of K-NN, GB, and RF models demonstrates that instance-based and ensemble learning methods are highly effective for microgrid protection. By leveraging advanced ML techniques, we can achieve significant enhancements in the stability, reliability, and efficiency of AC microgrids, paving the way for more resilient and sustainable energy systems.

## VI. Future Scope

The future scope of research in ML-based protection schemes for microgrids is vast and promising, offering numerous opportunities to enhance the stability, reliability, and efficiency of these critical systems. One key area of focus should be the integration of ML models with real-time data analytics and edge computing technologies. This integration can enable the development of adaptive protection schemes that respond dynamically to changes in microgrid conditions. Real-time data analytics allows for the continuous monitoring and analysis of microgrid performance, facilitating timely and accurate fault detection and mitigation. Edge computing, in contrast, involves placing computational capabilities in close proximity to the data source, resulting in decreased latency and facilitating quicker decision-making. Investigating the use of deep learning methods gives an additional encouraging path for future study. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are deep learning models that have shown exceptional performance in different predicting tasks. This is because they possess the capability to automatically extract and learn hierarchical features from intricate data. Implementing these models in microgrid protection schemes could further enhance prediction accuracy and reliability. Additionally, hybrid models that combine the strengths of multiple ML techniques, such as combining the instance-based learning of K-NN with the ensemble learning of GB, could offer significant improvements in model performance. Collaborative research with industry stakeholders is crucial for the practical deployment of these advanced protection schemes in real-world microgrids. Engaging with power utilities, microgrid operators, and technology providers can facilitate the translation of theoretical advancements into practical applications. Such collaborations can help address the technical and operational challenges associated with deploying ML-based protection schemes, including data integration, system interoperability, and scalability. Furthermore, industry partnerships can provide access to large-scale, real-world datasets, enhancing the robustness and generalizability of the developed models.

Additionally, future research should consider the development of standardized frameworks and protocols for implementing ML-based protection schemes in microgrids. Standardization can ensure consistency, reliability, and interoperability across different systems and applications, fostering broader adoption and scalability. Researchers should also explore the regulatory and policy implications of deploying intelligent protection schemes, addressing potential legal, ethical, and security concerns associated with the use of AI and ML in critical infrastructure. The integration of renewable energy sources, such as solar and wind, into microgrids presents unique challenges and opportunities for ML-based protection schemes. Future research should focus on developing specialized models that can handle the variability and intermittency of renewable energy generation, ensuring optimal performance and reliability of microgrids under diverse operating conditions.

Hence, the future of ML-based protection schemes for microgrids is rich with potential. By advancing real-time analytics, deep learning, hybrid models, and industry collaborations, researchers can significantly enhance the capabilities of microgrid protection systems. These efforts will contribute to the development of smarter, more resilient, and sustainable power systems, ultimately supporting the transition to a more reliable and efficient energy future.

## REFERENCES


[1] J. M. Guerrero, J. C. Vasquez, J. Matas, L. G. de Vicuña, and M. Castilla, "Hierarchical control of droop-controlled AC and DC microgrids—A general approach toward standardization," IEEE Transactions on Industrial Electronics, 58(1), 158-172, 2011.

[2] R. H. Lasseter, "Smart distribution: Coupled microgrids," Proceedings of the IEEE, 99(6), 1074-1082, 2011.

[3] A. Ghosh, G. Ledwich, and F. Zare, "Protection of power systems with distributed energy resources," IET Renewable Power Generation, 14(11), 1923-1932, 2020.

[4] S. Mishra, B. Krishna, and N. Sahoo, "Adaptive protection schemes for microgrids," Renewable Energy Research, 28(4), 987-1002, 2019.

[5] S. Panda, A. Mohanty, and D. Swain, "Machine learning applications for microgrid protection: A review," International Journal of Electrical Power & Energy Systems, 104, 251-264, 2019.

[6] L. Breiman, "Random forests," Machine Learning, 45(1), 5-32, 2001.

[7] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, 29(5), 1189-1232, 2001.

[8] V. N. Vapnik, "The Nature of Statistical Learning Theory," Springer, 1995.

[9] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, "Applied Logistic Regression," John Wiley & Sons, 2013.

[10] T. M. Cover, and P. E. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, 13(1), 21-27, 1967.

[11] A. Hussain, V. H. Bui, and H. M. Kim, "A review on microgrid protection schemes and systems," IEEE Transactions on Smart Grid, 11(4), 2969-2979, 2020.

[12] C. Qiu, Y. Zhao, and F. Li, "A review on machine learning applications for microgrid protection and control," IEEE Access, 8, 146419-146429, 2020.

[13] H. Wang, X. Yang, and W. Gao, "Application of SVM in fault classification for microgrids," IEEE Transactions on Power Delivery, 33(4), 1935-1943, 2018.

[14] T. Kim, S. Park, and H. Lee, "Comparative analysis of machine learning algorithms for fault detection in smart grids," Energies, 14(7), 1856, 2021.

[15] M. Patel, and A. Shah, "Adaptive Protection Schemes for Microgrids Using K-Nearest Neighbors," Energy Systems, 32(1), 56-70, 2021.

[16] Y. Zhang, and H. Wang, "Comparative Analysis of Machine Learning Models for Microgrid Fault Detection," Renewable Energy Research, 28(4), 987-1002, 2021.

[17] S. Lee, and J. Kim, "Application of Gradient Boosting in Microgrid Protection," International Journal of Electrical Power & Energy Systems, 117, 105643, 2020.

[18] L. Zhang, X. Xie, and Y. Xu, "Comparative analysis of machine learning models for microgrid fault detection," Journal of Electrical Engineering & Technology, 16(4), 1432-1442, 2021.

[19] J. Smith, and L. Brown, "Machine Learning for Microgrid Protection: A Review," Journal of Power Systems, 34(2), 123-135, 2020.

[20] R. Kumar, and P. Singh, "Enhancing Grid Stability with Machine Learning Techniques," IEEE Transactions on Smart Grid, 10(3), 567-578, 2019.

[21] A. Jadhav, D. Pramod, and K. Ramanathan, "Comparison of performance of data imputation methods for numeric dataset," Applied Artificial Intelligence, 33(10), 913-933, 2019.

[22] S. G. K. Patro, and K. K. Sahu, "Normalization: A preprocessing stage," International Advanced Research Journal in Science, Engineering and Technology, n.d. https://doi.org/10.17148/iarjset.2015.2305, 2015.

[23] J. Brownlee, "Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python," Machine Learning Mastery, 2020.

[24] I. Guyon, and A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, 3, 1157-1182, 2003.

[25] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8), 1226-1238, 2005.

[26] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” Machine Learning, 46(1-3), 389-422, 2002.
[27] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” Journal of Machine Learning Research, 13(Feb), 281-305, 2012.
[28] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” In Proceedings of the 14th International Joint Conference on Artificial Intelligence, vol. 2, pp. 1137-1143, 1995.
[29] D. M. W. Powers, “Evaluation: From precision, recall, and F-measure to ROC, informedness, markedness, and correlation,” Journal of Machine Learning Technologies, 2(1), 37-63, 2011.
[30] M. Sokolova, and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” Information Processing & Management, 45(4), 427-437, 2009.
[31] Y. Sasaki, “The truth of the F-measure,” Teach Tutor Mater, 1(5), 1-5, 2007.
[32] J. Opitz, and S. Burst, “Macro F1 and Macro F1,” arXiv.Org, arXiv:1911.03347, https://doi.org/10.48550/arxiv.1911.0334, 2021.
[33] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” Machine Learning, 6(1), 37-66, 1991.
[34] L. Breiman, “Random forests,” Machine Learning, 45(1), 5-32, 2001.
[35] C. Cortes, and V. Vapnik, “Support-vector networks,” Machine Learning, 20(3), 273-297, 1995.
[36] M. Singh, O. Singh, M. A. Ansari, and V. Singh, “Simulation and analysis of single-stage grid-connected solar PV system using ANN,” Intelligent Computing Techniques for Smart Energy Systems: Proceedings of ICTSES, 373-385, 2021.
[37] M. Singh, O. Singh, and A. Kumar, “Renewable energy sources integration in micro-grid including load patterns,” 3rd International Conference on Recent Developments in Control, Automation, and Power Engineering (RDCAPE), 88-93, 2019.
[38] M. Singh, O. Singh, and M. A. Ansari, “Performance analysis of protection technique for AC microgrid from symmetrical faults,” International Journal of Advances in Electrical Engineering, 4(1): 78-94, 2023.
[39] M. Singh, O. Singh, and M. A. Ansari, “Analysis of microgrid and protection schemes: a review,” Smart Energy and Advancement in Power Technologies: Select Proceedings of ICSEAPT, vol 1, 2021.