International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

COMPARATIVE ANALYSIS OF FAULT DIAGNOSIS IN POWER SYSTEM USING EVOLUTIONARY COMPUTATION AND MACHINE LEARNING

Roshni Rahangdale[1] , Bhanu Pratap Sahu[2] , Aditya Gupta[3] ,Manjari Sahu[4] , Shivani Wahile[5] and Subodh Kant Tiwari[6]

[1]Assistant Professor, Department of Electrical Engineering, Shri Shankaracharya Technical Campus, Bhilai (C.G), India

[2-6]UG Scholar, Department of Electrical Engineering, Shri Shankaracharya Technical Campus, Bhilai (C.G), India

---***---

Abstract - This paper provides a comparative analysis of fault diagnosis in power systems using evolutionary computation and machine learning techniques. Power system faults, including line-to-ground, line-to-line, and three-phase faults, pose significant challenges to the stability and reliability of electrical grids. Efficient and accurate fault diagnosis is crucial for minimizing downtime and maintaining optimal power delivery. This research evaluates the performance of various evolutionary computation algorithms such as genetic algorithms, particle swarm optimization and differential evolution alongside machine learning methods including decision trees, support vector machines, and k-nearest neighbor, in diagnosing power system faults. The study compares the accuracy, speed, and reliability of these methods in detecting and classifying faults under different system conditions. The findings highlight the strengths and limitations of EC and ML-based fault diagnosis techniques in power systems, providing valuable insights for researchers and practitioners in the field.

***Key Words*:** Fault diagnosis, Power system, Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Differential Evolution(DE) , Support Vector Machine (SVM) , Decision Tree (DT), K-Nearest Neighbors (KNN).

1. Introduction

Power system is the set of facilities and parts needed to produce, transfer, distribute, and use electrical energy. Transmission lines, substations, distribution networks, electricity producing facilities, and consumer loads are all included. A power system's main goal is to supply consumers with efficient and dependable electricity while maintaining system stability and safety. The system needs to manage a number of difficulties, including voltage regulation, fault circumstances, and variations in load. Power systems are designed and analyzed by engineers to maximize efficiency, reduce losses, and improve sustainability. Modern power systems strive to integrate renewable energy sources, enhance grid resilience, and expand smart grid features for a sustainable energy future through technological advancements and grid integration.

Fault is an abnormal condition that impedes the system's regular operation—typically causing an interruption in electrical flow—is referred to as a fault in power systems. Numerous factors, including equipment malfunctions, lightning strikes, and human mistake, can result in faults. A malfunction may result in power outages, voltage swings, and possible equipment damage. Predictably identifying and fixing errors early on is essential to keeping the power system stable and dependable.

Different Types of Faults in Power System-

The power system faults are primarily divided into two categories:
1) Open Circuit Fault
2) Short Circuit Fault

1) One or two wires failing is the primary source of the open circuit issue. Because the open circuit fault occurs in series with the line, it is sometimes referred to as a series fault.
The failure in the open circuit is classified as:-
* One open conductor fault;
* Two open conductor faults;
* Three open conductor faults.

2) This kind of failure occurs when the conductors of the various phases come into touch with one another through a power transformer, power line, or other circuit element, which causes a significant amount of current to flow through one or both of the system's phases. There are two types of short-circuit faults: symmetrical and unsymmetrical.

A) Symmetrical faults are those that affect all three phases of the fault system. These kinds of faults maintain their balance even after a fault occurs.
The following subcategories of symmetrical faults exist:-
1) line-to-line-to-line fault
2) Three-phase fault-to-ground connection

B) Unsymmetrical Fault results in unsymmetrical current, which is defined as current that varies in phase and magnitude among the power system's three phases. It is mostly divided into three categories. They are-
1) A single Line-to-Ground fault (L–G)
2) Line-to-Line fault (L–L)
3) Double Line-to-Ground fault (L–L–G)

Fault diagnosis in power systems, fault diagnosis is the process of finding anomalies or problems in an electrical power network and fixing them. To resolve issues that arise in a network for the transmission or distribution of electrical power, fault identification is crucial. Three steps make up the fault diagnosis process:

1) Detection and Identification: This phase entails examining the system for anomalies or flaws. If an anomaly is found, the process advances to the following level. If not, there is no intervention and the system keeps running.

2) Localization and Isolation: The system locates the impacted location within the power system in order to isolate the fault as soon as an anomaly is recognized. In order to aid in the fault's isolation, this step seeks to pinpoint its exact position.

3) Restoration and Resolution: Following the isolation of the fault, the system concentrates on fixing the problem and bringing back regular operations. In order to restore the system to its typical operational condition, corrective measures like fixes, replacements, or modifications are carried out during this stage.

2. Methodologies

For fault diagnosis in power system, we have mention two approaches i.e. Evolutionary Computation methods and Machine Learning methods and their comparison which have been discussed in the paper.

2.1. Evolutionary computation (EC)

EC is an important optimization and search tool in Artificial Intelligence(AI). Inspired by natural selection principles, it entails creating a variety of solutions to a problem and allowing them to evolve over subsequent generations via processes like as selection, crossover, and mutation. This approach is especially useful in AI for tackling complex problems, optimizing parameters, and uncovering answers that would be difficult to find with traditional algorithms. In the EC domain, we can mention the following algorithms–
1. Genetic Algorithm
2. Particle Swarm Optimization (PSO)
3. Differential Evolution

2.1.1. Genetic Algorithm (GA)

GA is one of the oldest and best-known natural-based optimization approaches. Fault diagnosis with genetic algorithms is a technology that uses natural selection and evolution concepts to find and correct problems in systems. Genetic algorithms provide a set of alternative solutions that represent various system configurations. These solutions evolve over time based on their effectiveness in diagnosing and correcting issues.

This is the process that it follows:
i) Encoding Solutions: In the context of diagnosing power grid faults, genetic algorithms encode possible solutions into a sequence of genes that may represent various fault parameter configurations, including fault location, kind, magnitude, and others.

ii) Fitness Evaluation: To assess how well a solution diagnoses errors, a fitness function is defined. This function might take into account variables like fault detection robustness, speed, and precision.

iii) Selection: To create offspring solutions, solutions that exhibit better fitness values are chosen to go through genetic processes like crossover and mutation.

iv) Crossover and Mutation: Crossover is the process of integrating genes from two parent solutions to form new solutions, similar to how natural genetic crossover occurs. Mutation causes random changes in genes, allowing us to explore new sections of the solution space.

v) Population Evolution: The population evolves over multiple generations to include better solutions based on the fitness function.

vi)Termination Criteria: The algorithm iterates through generations until it reaches a termination criterion, such as the maximum number of generations or a good result.

Genetic algorithms can aid in the optimization of fault detection parameters, enhance problem identification accuracy, and perhaps minimize the computational resources required for diagnosis. The figure 2.1.1 shows flowchart for GA.

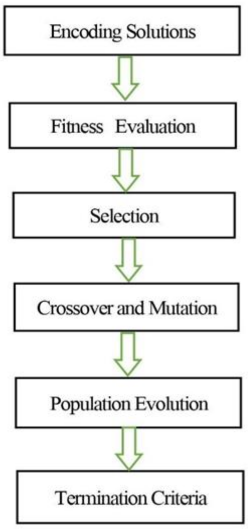

Figure 2.1.1 Flowchart for GA

2.1.2. Particle swarm optimization (PSO)

A computational method called Particle Swarm Optimization (PSO) was motivated by the social behavior of fish schools and flocks of birds. PSO can be used to optimize specific parameters or variables in order to efficiently discover faults in power systems during fault diagnostics.

This is the fundamental process that PSO uses for fault diagnosis:
i) Initialization: PSO initiates the search space by randomly initializing a population of particles. Every particle is a possible fix for the issue at hand.

ii) Evaluation: The degree to which a particle solves the problem determines its fitness or performance. In the context of power system fault diagnosis, this can entail assessing how well, the present parameters identify and pinpoint defects.

iii) Updating Position and Velocity: The particles modify their position and velocity according to the best-known positions of the swarm as well as themselves. This is achieved by combining the prior velocity of the particle, its cognitive component (driven by its best-known position), and its social component (driven by the best-known position of the swarm).

iv) Local and Global Best: The swarm monitors the collective best-known position (global best), while each particle records its individual best-known position (local best).

v) Iterative Optimization: Until a stopping condition is satisfied, such as the convergence of solutions or the maximum number of iterations attained, PSO continues steps 2-4 repeatedly.

vi) Solution Extraction: The optimal solution to the problem is represented by the best-known position of the swarm after all iterations have been completed. This could be the optimal parameter or variable in fault diagnosis that helps locate and identify power system defects.

vii)Validation and Fine-Tuning: In order to make sure the optimized solution works as intended in practical situations, it is next validated and, if needed, adjusted.

Depending on the particular needs, features, and fault diagnosis issue at hand of the power system, PSO can be modified and used in a variety of ways. The figure 2.1.2. shows flowchart for PSO. The figure 2.1.2 shows flowchart for PSO.

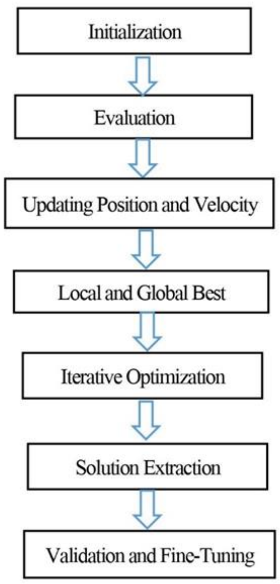

Figure 2.1.2 Flowchart for PSO

2.1.3. Differential evolution (DE)

DE is a numerical optimization approach that solves optimization problems using a continuous search space. Storn and Price launched DE in 1997, and it has since acquired popularity for its simplicity, effectiveness, and capacity to solve optimization problems with non-linear and non-differentiable objective functions. DE method can be effective in fault diagnosis in power systems due to its capacity to optimize and seek for optimal solutions in complex and high-dimensional domains. This is the process that it follows:

i) Feature Extraction: The initial step is to extract useful characteristics from sensor data. These characteristics could include voltage and current measurements, fault initiation times, and impedance values. The choice of features has a considerable impact on the effectiveness of fault diagnosis.

ii) DE Algorithm: DE operates on a set of candidate solutions, each indicating a potential fault kind and location. The method iteratively improves these solutions using the techniques below:

a) Mutation: New candidate solutions are created by introducing variations between current solutions in the population. This adds diversity and enables the algorithm to examine various sections of the search space.

b) Crossover: Trial solutions are created by combining elements from both mutant and basic solutions. This helps incorporate Benefits of diverse solutions.

c) Selection: The original solution and the trial solution compete using a fitness function. The solution with the highest fitness value forms part of the next generation population.

iii) Fitness Function: The fitness function is crucial for guiding the DE algorithm. The fitness function in fault diagnosis can be constructed to minimize the discrepancy between actual and anticipated measurements for a certain fault type and location or to maximize the separation of features associated with different fault kinds.

iv) Fault Classification: Once the DE algorithm has converged, the solution with the highest fitness value is deemed the most likely fault scenario. This solution identifies the kind and location of the fault in the power distribution system.

Overall, DE is a viable approach to fault diagnostics in power distribution systems. Its ability to manage complicated interactions and identify global solutions makes it an effective tool for guaranteeing system reliability. The figure 2.1.3 shows flowchart for DE.

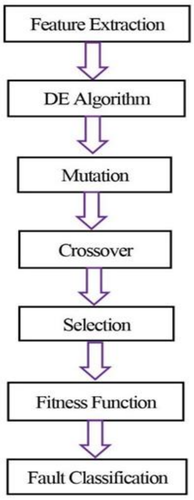

Figure 2.1.3 Flowchart for DE

2.2 Machine learning (ML)

ML is a subset of artificial intelligence (AI) that involves creating systems or models that can learn from data and improve their performance on a job over time without being explicitly programmed for it. It's about creating algorithms that let computers to learn and make predictions or judgments based on patterns and information in the data they are exposed to. In ML, we mention the following methods -
1. Support Vector Machine
2. Decision Tree
3. K-Nearest Neighbor

2.2.1. Support vector machine (SVM)

SVM is a powerful machine learning technique that can effectively identify and classify faults in power systems. SVMs work by finding an optimal hyper plane that separates different fault types in a high-dimensional feature space, enabling accurate fault detection and classification.

This is the process that it follows:

1. Data collection: Compile information on voltage, current, frequency, and other pertinent characteristics from a variety of sensors installed in the power system.

2. Feature Extraction: Take pertinent features out of the data that has been gathered. Phase angles, harmonics, RMS values, and other properties might be included in this list.

3. Data preprocessing: To guarantee consistency and boost SVM performance, clean up the data, deal with missing values, and normalize the features.
4. Training Phase: The SVM model is trained using historical data that has been tagged with fault circumstances. Based on the features that are extracted, the SVM learns to classify the data into fault and non-fault circumstances
.
5. Testing Phase: on determine whether a power system problem exists, apply the trained SVM model on fresh, untainted data.

6. Evaluation and Validation: Use measures like accuracy, precision, recall, and F1-score to evaluate the SVM model's performance. To make sure the model is reliable, check its predictions against recognized fault scenarios.

The optimal hyperplane for classifying the data points into distinct groups is identified via SVM. In order to identify the decision boundary that best generalizes to yet-to-be-found data, it optimizes the margin between the classes. The data points that are closest to the hyperplane, or support vectors, are used to determine this decision boundary. The figure 2.2.1 shows flowchart for SVM.

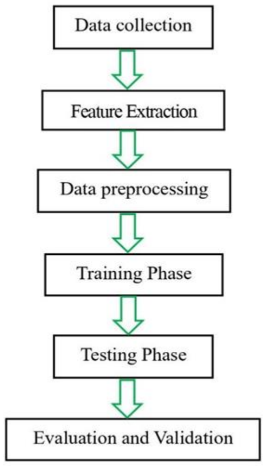

Figure 2.2.1 Flowchart for SVM

2.2.2. Decision Tree (DT)

Decision Tree is a supervised learning technique that may be applied to both classification and regression issues, however it is most commonly employed to solve classification problems. It is a tree-structured classifier, with core nodes representing dataset attributes, branches representing decision rules, and leaf nodes representing outcomes. Fault detection in power systems with decision trees involves developing a tree-like model to examine system parameters such as voltage, current, and frequency. The program evaluates these factors to find potential errors and their sources. The following are the stages involved in diagnosing faults in power distribution systems using a decision tree:

1. Data Acquisition: The first step is to collect data from the electricity distribution system. This data typically contains voltage and current measurements taken at various places along the network. Fault data can be retrieved as follows: Fault recorders collect high-frequency data during fault incidents. Protection relays provide information on fault currents and voltage sags. Smart meters can provide voltage and current data in high detail.

2. Feature extraction: Relevant features are retrieved from the raw data to help determine the type of issue. These features could include: Magnitudes are the root mean square (RMS) values of voltage and current during the fault.
Ratios: Phase current ratios, zero sequence current to positive sequence current, and other ratios.
Rates of Change: The voltage and current change at the beginning of the problem.

3. Training using Decision trees: A decision tree is constructed from a dataset of past failure occurrences. Every data point depicts a distinct fault situation, along with associated attributes and the fault type (such as single-phase or three-phase). The data is separated iteratively by the decision tree method according to the qualities that are most useful in distinguishing between various fault types.

4. Fault Diagnosis: Once trained, the decision tree can be used to identify future problem events. The measured features of a new fault event are input into the decision tree, which traverses the tree using the decision rules at each node. The decision tree eventually classifies the failure type according to the terminal node it reaches.

The figure 2.2.2 shows flowchart for DT.

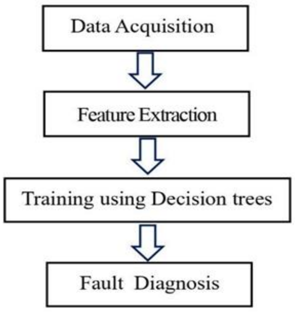

Figure 2.2.2 Flowchart for DT

2.2.3 K-Nearest Neighbor (KNN)

The K-Nearest Neighbor (KNN) method is a prominent machine learning tool for classification and regression problems. It is based on the assumption that similar data points would have similar labels or values.

This is an explanation of how it functions:
1. Data Acquisition and Preprocessing: At various locations across the distribution network, the system tracks voltage and current signals. These signals diverge from typical operational values during failures. The KNN uses the recorded voltage and current waveforms as its data, along with any other pertinent properties. Preprocessing may include methods to guarantee consistent data representation, such as normalization or noise reduction.

2. Training Phase: Labeled data from a historical dataset must be used. Measurements of voltage and current under typical circumstances as well as different fault scenarios (such as single-phase and line-to-ground faults) are included in this data. The KNN algorithm gains knowledge by examining the connections between the training data's characteristics (voltage/current) and the associated fault labels. In essence, it learns by heart the traits of various fault kinds found in the training set.

3. Fault Detection and Classification: In real-time operation, the system records voltage and current measurements in the event that a defect is suspected. These measurements combine to create a new data point that the trained K-NN model must classify. Within the training data, the K-NN algorithm determines the separation between the new data point and its k-nearest neighbors, or k closest data points. For each new data point, the projected fault is determined by taking the majority class (fault type) among the k-nearest neighbors.

The figure 2.2.3 shows flowchart for KNN.

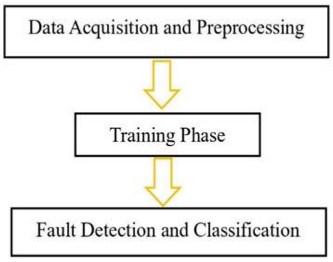

Figure 2.2.3 Flowchart for KNN

3. Comparative Results

3.1 Comparison between EC and ML

The Table 3.1 shows a comparison between EC and ML based on several parameters.

Parameters	Evolutionary Computation (EC)	Machine Learning (ML)
1.Nature	Evolutionary algorithms inspired by natural selection and swarm behavior	Supervised learning algorithms focused on classification.
2.Applicability	Primarily used for optimization problems, feature selection, and evolving solutions	Applied to a wide range of classification and regression problems
3.Convergence speed	Numerous factors, including population size and mutation rates, might affect convergence.	Faster to train on average than evolutionary algorithms
4.Parameter Sensitivity	Sensitive to adjustment of the parameters, and different problem domains may require different optimal parameter choices	Though frequently simpler, parameters can also affect its sensitivity.
5.Handling Noisy Data	Because they are population-based, they are robust against noise.	Sensitive to noise; tuning or preprocessing can be necessary

Table 3.1 Comparison between EC and ML

3.2. Comparison between GA and KNN

The Table 3.2 shows a comparison between GA and k-NN based on several parameters.

Parameters	GA	KNN
1 Nature	Genetic algorithms are population-based algorithms for optimization that draw inspiration from genetics and natural selection. They entail a population of viable solutions developing over several generations.	K-Nearest Neighbors is an instance-based, non-parametric learning method that categorizes new cases according to a similarity metric after storing all of the cases that are available.
2 Purpose	The main applications of genetic algorithms are in search and optimization, where the goal is to find the best answer across a wide search space.	K-Nearest Neighbors is primarily utilized in regression and classification tasks, where the result is a continuous value or a class membership depending on the input data.
3 Complexity	Genetic algorithms entail population initialization, selection, crossover, mutation, and evaluation phases, they can have a high computing complexity, particularly for vast search spaces or complex fitness functions	K-Nearest Neighbors just stores the training data, it has a comparatively low computing complexity when being trained. However, because it involves calculating the distances between each training instance and the test instance, the prediction complexity can be considerable, particularly for big datasets.
4. Performance	Genetic algorithms can be used to solve complicated optimization issues with broad search areas, but they may take longer and more processing power to reach the best answer.	K-Nearest Neighbors is a good choice for regression and classification problems, particularly if the decision boundaries are non-linear.

Table 3.2 Comparison between GA and KNN

3.3 Comparison between PSO and SVM

The Table 3.3 below shows a comparison between PSO and SVM based on several parameters.

Parameters	PSO	SVM
1. Nature	Swarm intelligence optimization method, modeled after fish or bird social behavior	For problems involving regression and classification, supervised learning is used.
2. Purpose	Applies to optimization issues such as determining the best course of action given an objective function.	Especially useful in high-dimensional environments, this algorithm is used for classification and regression problems.
3. Complexity	Less hyper parameters to adjust, straightforward to apply.	Requires fine-tuning of parameters such as kernel type and regularization parameter, and can be computationally demanding,
4. Performance	Does well with discrete or multimodal functions, but may have trouble with continuous optimization situations.	Offers strong performance and excellent accuracy in general, particularly for binary classification tasks.

Table 3.3 Comparison between PSO and SVM

3.4 Comparison between DE and DT

The Table 3.4 below shows a comparison between DE and DT based on several parameters.

Parameters	DE	DT
1 Nature	DE is a stochastic optimization approach based on population and inspired by natural selection.	The model learns decision rules from the data in decision trees, a supervised learning technique used for regression and classification tasks
2 Purpose	DE looks for the global optimum of a given objective function and is mainly utilized in optimization problems.	Decision trees are used for both classification and regression tasks in machine learning, providing interpretable models and helps in decision-making.
3 Complexity	DE can be computationally costly, particularly when dealing with complex objective functions or high-dimensional optimization issues	Although decision trees are quick and easy to train, they can get complicated and prone to overfitting when used with large datasets or deeply rooted trees.

4 Performance	DE is well-known for its ability to reliably and effectively identify global optima in a wide range of optimization problems, particularly those involving continuous spaces.	Decision trees are capable of producing good results on datasets of any size, but when faced with correlated or noisy characteristics, they may perform poorly and necessitate the use of ensemble approaches or pruning.

Table 3.4 Comparison between DE and DT

4. Conclusion

The comparison of evolutionary computation with machine learning for fault identification in power systems highlights the unique features of both approaches. Using natural selection as inspiration, evolutionary computation works well for parameter optimization in defect detection. It performs best in situations involving intricate, nonlinear interactions. Conversely, machine learning exhibits robustness in real-world applications due to its data-driven methodology and capacity to adapt to a variety of datasets. When comparing the methods of evolutionary computation with machine learning for fault identification in power distribution systems, each approach has advantages and disadvantages. Complex, nonlinear issues can be handled using evolutionary computation, such as genetic algorithms, but they may take longer and require more computing power. However, machine learning, using methods like SVM or decision trees, can provide accurate and effective solutions, but might also
struggle with complex, evolving problems.

5. Future scopes

1. Enhanced Accuracy: By combining EC and ML techniques, future research can concentrate on raising the accuracy of defect diagnosis algorithms. This might entail creating increasingly complex evolutionary algorithms and using cutting-edge machine learning models.

2. Real-time Monitoring: Using EC and ML approaches, it may be possible to create real-time fault diagnosis systems that continuously monitor power systems. This would enhance system dependability by decreasing downtime and enabling faster failure identification and response.

3. Big Data Integration: Future studies may examine how to combine EC and ML methods with big data analytics, given the abundance of data in power systems. Large amounts of data from sensors and other sources could be processed in this way to improve fault diagnosis accuracy.

4. Multi-objective Optimization: By simultaneously taking accuracy, speed, and cost-effectiveness into account, EC approaches can be further investigated for multi-objective optimization in fault diagnosis. Machine learning algorithms can help discover intricate connections among several goals.

5. Adaptive Systems: Adaptive EC and ML algorithms that can automatically modify and improve their parameters depending on shifting system conditions and fault characteristics may be incorporated into future fault diagnosis systems.

6. References

[1] Ming-Yuan Cho and Thi Thom Hoang, "Feature selection and parameters 0ptimization of SVM using Particle Swarm Optimization for fault classification in power distribution system", Department of Electrical Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan ,11 July 2017.
[2] Davi L. de Souza, Matheus H. Granzotto, Gustavo M. de Almeida, Luís C. Oliveira-Lopes, "Fault detection and diagnosis using Support Vector Machines - a SVC and SVR comparison,2014.
[3] Melanie Mitchell and Charles E. Taylor, "Evolutionary Computation: An Overview", Anta Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico ,March 1999.
[4] Pranay V. Ambade and Sanjeev S. Gokhale, "Fault section estimation by neural networks and genetic algorithm", YCCE, Nagpur,India,March,2018.
[5] Yuhui Shi, "Particle Swarm Optimization: development, applications and resources", February 2001.
[6] TC Srinivasa Rao, SS Tulasi Ram, JBV Subrahmanyam, "Fault recognition and diagnosis based on a Decision Tree for power distribution system", 25 March 2019.
[7] Aida Asadi Majd, Haidar Samet and Teymoor Ghanbari, "K-NN based fault detection and classification methods for power transmission systems", 2017.
[8] Tanweer Alam, Shamimul Qamar, Amit Dixit, Mohamed Benaida, "Genetic Algorithm: reviews, implementations, and applications", June 2020.
[9] Ahmed G. Gad, "Particle Swarm Optimization algorithm and its applications: a systematic review", 19 April 2022.
[10] P. P. Wasnik, Dr. N. J. Phadkule, Dr. K. D. Thakur, "Fault detection and classification in transmission line by using KNN and DT technique", April 2020.
[11] Baskar. D, Dr. Selvam. P, "Machine learning framework for power system fault detection and classification", February 2020.
[12] R. Ramaraj, P.M. Lall, V. S. K. Murthy, "Power system fault diagnosis using wavelet and genetic algorithms", 2005.
[13] A. K. Sinha, H. S. Dhillon, S. S. Murthy, "Electric power distribution system engineering", CRC Press,2018.
[14] A. K. Pradhan. "Power System Reliability: Safety and Management", CRC Press,2009.
[15] K. S. Swarup, P. D. Kulkarni, "Fault detection and classification in power system using support vector machine", 2006.