Network Submission and Monitoring of Power System Operation Safety Monitoring Data using Safe Power Hybrid Classifier (SPHC)

Bo Yan[1], Deming Yu[1*], Yue Qi[2], Meng Liu[1], Kun Cao[1]
[1]Dispatching and Control Centre, State Grid Jibei Electric Power Company Limited
[2]State Grid Jibei Electric Power Co., Ltd
E-mail: demingyu12@outlook.com
[*]Corresponding Author

Keywords: ensemble voting technique, power systems, safety, safe power hybrid classifier (SPHC) algorithm,

Received: September 5, 2024

Ensuring the dependability and security of power systems is critical to guaranteeing uninterrupted electricity supply and avoiding possibly catastrophic outages. Previous power system monitoring methods frequently have restricted incorporation and lack real-time assessment abilities, limiting their ability to detect and tackle safety problems promptly. To tackle these drawbacks, this paper presents the SafePower Hybrid Classifier (SPHC) algorithm, which is intended to improve real-time surveillance and classification of power system security information. The proposed system gathers and sends key security metrics like voltage, current, temperature, humidity, power factor, frequency, and phase imbalance from a network of sensors spread across power plants and transmission lines. The SPHC algorithm uses an ensemble voting method to classify the safety status as Normal, Warning, or Critical, enabling timely intervention using the Power System Operation Safety Monitoring Dataset (PSOSMD). Experimental findings indicate that the SPHC algorithm surpasses conventional classifiers, with performance metrics such as accuracy of 92.4%, precision of 91.3%, recall of 91.8%, F1-score of 91.5%, and MCC of 89.7%, substantially decreasing the possibility of power system failure. Integrating thorough data gathering and sophisticated classification into the proposed system greatly enhances the dependability and stability of power operations.

Povzetek: Predlagan je algoritem SafePower Hybrid Classifier (SPHC) za izboljšano spremljanje varnosti električnega omrežja, ki uporablja senzorjev in metode ansambelskega glasovanja.

1 Introduction

Power system dependability and safety are essential in guaranteeing an uninterrupted and stable supply of electricity, which is required by contemporary society [1]. The need for sophisticated monitoring systems has increased considerably as power grids become more intricate, owing to the incorporation of renewable energy sources and the growth of interconnected networks [2]. These systems should be able to not only identify possible security problems but also react quickly enough to avoid minor anomalies from becoming serious interruptions [3]. As power systems become more dynamic and dispersed, conventional monitoring techniques are increasingly insufficient, requiring the creation of more complex solutions [4].

Numerous methods have been used over time to track power system security, including manual inspections, simple sensor networks, and data logging solutions [5]. While these techniques have yielded useful results, they frequently have major disadvantages. Conventional systems are usually restricted by their incapability to process and assess data in real time, which causes delays in the detection and resolution of security problems. Furthermore, the absence of incorporation among the various elements of the power network outcomes in fragmented data streams that are hard to interpret cohesively. This disjointed strategy impedes the capacity to rapidly detect and tackle growing hazards, affecting the total dependability of the power system.

Recognizing these difficulties, this paper presents an innovative strategy for improving real-time monitoring and safety evaluation of power systems. The proposed solution is based on the SafePower Hybrid Classifier (SPHC) algorithm, which is integrated into an extensive monitoring system. This system is intended to continually collect, transmit, and evaluate data from a network of sensors spread across different power system components, including power plants and transmission lines. The SPHC algorithm uses an ensemble voting method to divide the system's safety status into different groups, allowing operators to react proactively to any identified anomalies. The SPHC algorithm tackles numerous major shortcomings in previous monitoring systems. By combining real-time data analytics and sophisticated classification methods, the SPHC algorithm improves the accuracy and speed of safety evaluations. Unlike conventional approaches, which frequently depend on periodic gathering of data and manual interpretation, the SPHC generates a steady stream of useful knowledge that can be utilized to keep the system stable. Furthermore, the system's capability to integrate data from numerous sources into a unified assessment platform guarantees that security evaluations are thorough and representative of the whole power network.

This research makes significant contributions to the field of power system safety. To begin, the paper describes an innovative classification algorithm developed especially for real-time monitoring of power systems. Second, it shows a detailed experimental assessment to show the algorithm's efficiency, showing that it outperforms previous monitoring methods in terms of accuracy and response. Third, the paper outlines an entire monitoring solution that combines sensor networks, transmission of data, and real-time evaluation, providing an effective framework for guaranteeing power system dependability.

The main aim of this research is to create a system that improves the capability to detect and reduce safety risks in power systems quickly. The objective is to surpass the restrictions of current monitoring techniques by implementing a more incorporated and responsive solution. The SPHC algorithm is unique in that it combines real-time data analysis with ensemble classification methods, resulting in a more precise and timely evaluation of power system safety.

This system is especially helpful in areas where a consistent power supply is critical, like large-scale power plants, renewable energy installations, and urban power grids. The capability to track and react to safety concerns in real time makes this method indispensable for sustaining the stability and effectiveness of contemporary power networks.

The rest of the paper is structured as follows: Section 2 examines related works in power system surveillance and safety evaluation, highlighting the gaps that the SPHC algorithm seeks to fill. Section 3 describes the methodology, which includes the planning and execution of the SPHC algorithm, as well as the entire surveillance framework. Section 4 provides the experimental findings and discusses their implications, emphasizing the benefits of the proposed method. At last, Section 5 concludes the paper and outlines possible fields for future research, such as improvements to the SPHC algorithm and its application to a wider variety of power systems.

2 Related works

This section analyses recent advances in power system tracking, concentrating on the different methods and technologies used to guarantee grid dependability and safety. It investigates the efficacy and constraints of current techniques, such as IoT-based systems, data processing techniques, and safety surveillance enhancements. This section analysis these studies to emphasize the difficulties faced in real-time evaluation and incorporation of safety data, setting the stage for the introduction of the SPHC algorithm as a solution to these problems.

The incorporation of the Internet of Things (IoT) in smart grids has substantially enhanced the capacity to track and regulate grid parameters, resulting in more dependable and effective power delivery. Khan et al. [6] investigated an IoT-based power monitoring system designed specifically for smart grid applications, emphasizing its capability to track and evaluate electrical parameters like voltage, current, and energy usage in real time. The study shows how IoT applications, particularly platforms such as Thing Speak, can help customers and power companies handle the consumption of energy more efficiently, lowering expenses for operations. However, integrating IoT in smart grids poses difficulties, especially in guaranteeing uninterrupted service and data accuracy across different phases of the grid.

The implementation of resilient subspace clustering techniques has tackled advances in processing data and the detection of anomalies for power grid monitoring. Lee et al. [7] proposed a subspace clustering technique for handling large-scale data streams, like synchronized phasor evaluations, which are critical for power grid monitoring. The technique uses a low-rank representation model to reconstruct missing measures and discover anomalies, which addresses the computational and memory difficulties related to uninterrupted data streams in power grids. This method has been proven to surpass previous algorithms, which makes it a useful tool for improving grid dependability.

Wireless sensor networks (WSNs) have also been presented to address data collection problems in power grids, especially those caused by phasor measurement units (PMUs). Manoharan et al. [8] implemented a technique for more efficiently monitoring power grid parameters by combining WSNs and binary logistic regression (BLR). This technique addresses the common issue of data not attaining the main server from PMUs, guaranteeing that vital data is always obtainable for making choices. The suggested technique showed enhanced accuracy and effectiveness in surveillance, highlighting the possibility of WSNs linked with BLR to enhance the stability of the grid.

Fernandez et al. [9] investigated topology change identification in power grids via power line communications (PLC). Their research presented a channel impulse response (CIR)--based method that does not need any changes to the current smart grid architecture. The suggested technique provides a software-only solution for effectively detecting topology changes with low memory and computational demands. The system's resilience in high-noise settings emphasizes its applicability in real-world situations where communication channels are frequently compromised.

The examination of power grid blackouts and the development of countermeasures have also been important areas of research. Zhang et al. [10] examined numerous massive blackout events to determine common causes and risk factors. Their research presented knowledge of power grid vulnerabilities and suggested tactics to improve risk management and functional controls. These recommendations are critical to avoiding future blackouts and guaranteeing the uninterrupted and secure function of power grids.

In line with risk management, Li and Liu [11] proposed a real-time simulation system for assessing power grid vulnerability. Their platform simulates actual grid activities and performs fault scans to assess possible hazards and discover weak points in the grid. This simulation-based method presents technical assistance for the identification and handling of hidden risks, contributing to the total stability and security of power grid functions.

Another area of research has been the safety of power grid personnel during grid activities, with attempts underway to enhance security monitoring systems. Rao et al. [12] created an enhanced YOLOv3-based system for tracking the utilization of safety helmets by power grid workers. By improving detection accuracy for small targets and improving boundary frame regression, their system offers real-time tracking with higher speed and precision, thereby decreasing the risk of accidents during grid maintenance operations.

AI has also been used to tackle security issues in power grids, specifically in the detection of high impedance faults (HIFs). Wang and Dehghanian [13] presented an AI-powered online surveillance system that precisely recognizes HIFs, which are frequently overlooked by traditional monitoring devices. The system's capability to present timely alerts before serious harm happens renders it an indispensable tool for sustaining the security and reliability of power grids, particularly under severe weather conditions.

Hossain et al. [14] highlighted the significance of planning and improving substation grounding grids to protect employees as well as machinery. Their research emphasizes the importance of careful planning in grounding grid design, which is critical for reducing risks and safeguarding infrastructure in power systems. This method is especially important for improving the safety standards of power grid functions.

Finally, Alimi et al. [15] provided an in-depth examination of the most recent machine learning techniques (MLTs) used to improve power system protection and stability. They investigated the use of artificial neural networks (ANN), decision trees (DT), and support vector machines (SVM) in applications such as cyberattack detection, power quality (PQ) disturbances, and dynamic protection evaluations. They also talked about the constraints of previous studies' classifier designs, datasets, and test systems, as well as how to use reinforcement learning (RL) and deep reinforcement learning (DRL) for transient stability assessments. They concluded by outlining important issues and future research directions in this area. Table 1 shows the summary table.

Table 1: Summary table

Study/Method	Key Features/Approach	Accuracy	Recall	Real-Time Capabilities	Integration/Challenges	Limitations
Khan et al. [6] - IoT-based Power Tracking	Real-time tracking of electrical parameters utilizing IoT (ThingSpeak platform).	High	Moderate	Yes	Difficulties in sustaining continuous service and data accuracy	Constrained data accuracy and possible service interruptions
Lee et al. [7] - Subspace Clustering for Data Streams	Low-rank representation model for anomaly discovery in phasor measurements.	High	High	Yes	High computational and memory effectiveness	Memory and computational overhead for continuous data streams
Manoharan et al. [8] - WSN and BLR for Power Grid Tracking	Utilizes WSNs and BLR to resolve data gaps from PMUs for real-time grid parameter monitoring.	High	High	Yes	Enhanced data accuracy and effectiveness	Reliant on wireless sensor network stability
Fernandez et al. [9] - Power Line	Software-only solution for discovering grid topology	Moderate	Moderate	Limited	Minimum computational and memory necessities	Constrained real-time abilities; concentrates

Communications (PLC)	fluctuations using CIR.					on topology fluctuations only
Zhang et al. [10] - Power Grid Outage Evaluation	Blackout incident examination to detect reasons and risk factors; risk management tactics.	High	High	No	Efficient in blackout prevention	No real-time tracking, concentrates on post-incident examination
Li and Liu [11] - Real-Time Grid Vulnerability Simulation	Real-time simulation system for vulnerability and fault examination.	High	High	Yes	Presents technical support for concealed risk discovery	Constrained scalability for massive grids
Rao et al. [12] - YOLOv3-based Helmet Discovery	Enhanced YOLOv3 for real-time monitoring of security helmet use by grid personnel.	High	High	Yes	Precise small-target detection and quicker monitoring	Concentrates on personnel security, not on wider grid tracking
Wang & Dehghanian [13] - AI-Driven HIF Monitoring	AI-based system for detecting high impedance faults (HIFs) in real-time.	High	High	Yes	Timely alerts and high fault detection accuracy	Mainly tackles HIFs, not wider grid problems
Hossain et al. [14] - Substation Grounding Grid Design	Improves substation grounding for personnel and infrastructure safety.	High	High	No	Enhanced safety for substation functions	Concentrates only on grounding grid design, not overall grid security
You et al. [15] - Deep Learning Smart Grid Management	Deep learning for real-time flaw prediction and grid optimization.	High	High	Yes	Enhances effectiveness and security via flaw detection	Constrained incorporation with other grid security systems
SafePower Hybrid Classifier (SPHC)	Ensemble voting method for comprehensive security data incorporation and hazard discovery.	Very High	Very High	Yes	Seamless incorporation of diverse security data from numerous sensors	Surpasses limitations in real-time diagnostics and data incorporation

These previous power systems monitoring techniques are frequently hampered by an absence of incorporation and real-time diagnostic abilities, limiting their capacity to detect and tackle safety problems quickly. Numerous methods struggle with fragmented data sources and fail to efficiently aggregate or evaluate multiple safety metrics, resulting in delays in discovering possible risks and inadequate proactive measures. The SPHC algorithm is proposed to fill these gaps by combining a broad range of safety data gathered by different sensors and using an ensemble voting method for classification. This method guarantees thorough tracking, timely hazard identification, and enhanced intervention tactics, tackling the drawbacks of conventional approaches while improving total system dependability and safety.

3 Methodology

This section describes the methodology used to create and assess the SPHC algorithm. It describes the dataset utilized to train and test the approach, the planning and execution of the SPHC algorithm, and an overview of the entire strategy. The goal is to present a thorough discussion of how the SPHC algorithm improves real-time surveillance and classification of power system safety information. By incorporating sophisticated classification techniques, the methodology intends to enhance the dependability and efficacy of power system safety management.

3.1 Dataset

The dataset namely Power System Operation Safety Monitoring Dataset (PSOSMD) used by the SPHC algorithm includes key measurements gathered from a network of sensors spread across power plants, substations, and transmission lines. This dataset is critical for improving the real-time monitoring and categorization of power system safety data. Each entry in the dataset incorporates multiple features that present a thorough summary of the system's functional status and environmental circumstances.

This PSOSMD dataset contains 10,000 instances, which includes a wide range of functional measurements from sensors distributed across power plants, substations, and transmission lines. Preprocessing steps were performed prior to model training, comprising min-max normalization to guarantee consistent scaling of features between 0 and 1, and IQR (Interquartile Range) outlier removal to boost data quality and model effectiveness. The dataset was then divided into training and testing sets, with 80% (8,000 instances) used to train the SPHC algorithm and 20% (2,000 instances) used to test its efficacy. This division enables a comprehensive assessment of the model's accuracy and generalizability to new data. The SPHC algorithm uses a dataset that includes key measurements collected from a network of sensors spread across power plants, substations, and transmission lines. This dataset is essential for real-time tracking and classification of power system security data. Each entry in the dataset contains numerous features that provide a comprehensive overview of the system's operational status and environmental conditions.

Sensor ID is a distinctive identifier allocated to each sensor, enabling data from different sensors installed throughout the power system to be differentiated and tracked more effectively. This feature is essential for handling sensor efficiency and data honesty, as it ensures that data can be precisely attributed to its source.

The Location feature specifies the exact location where the sensor is installed, like a substation, power plant, or transmission line. Comprehending the location assists in contextualizing the data, as various components of the power system might encounter different functioning circumstances and problems.

The Data Type defines the type of measure being logged by the sensor, which may contain voltage, current, or temperature. This feature is required for evaluating multiple facets of power system health, as each kind of data presents information about particular operational parameters. Voltage and current are essential for assessing electrical efficiency, while temperature and humidity provide data about ambient circumstances that can affect devices.

Voltage (V) and current (A) are basic measures that represent the electrical potential variance and the flow of electric charge, correspondingly. Tracking these parameters is critical for guaranteeing they remain within acceptable functioning limitations to avoid concerns like power surges, overheating, or equipment damage.

Temperature (°C) monitors the heat at the sensor's position, which can identify possible issues such as equipment overheating or external conditions interfering with efficiency. High temperatures may indicate equipment faults or future breakdowns.

Humidity (%) indicates the amount of moisture in the air at the sensor's position. High humidity can impact the efficiency and lifespan of electrical equipment, perhaps resulting in insulation breakdowns or corrosion.

Power Factor is a measurement of how effectively electrical power is used. A power factor near one denotes effective energy consumption, whereas a lower value implies inadequacies that could contribute to increased functional expenses and system failures.

Frequency (Hz) measures the stability of an alternating current. Deviations from the normal frequency can impair the operation of electrical equipment and the entire system stability, thus it is critical for ensuring correct operation.

Phase Imbalance (%) is the voltage variation between stages in a three-stage system. A low percentage suggests a well-balanced system, but a high percentage might produce ineffectiveness or harm to equipment.

Timestamps capture the date and time of each measure, which is critical for monitoring shifts over time, recognizing patterns, and associating data with particular events or situations.

At last, the Status feature evaluates the entire state of the power system using sensor readings. It divides the system into Normal, Warning, and Critical states, which is crucial for prioritizing actions and tackling possible difficulties in a timely fashion. Table 2 displays a sample of the dataset with ten rows.

Table 2: Sample dataset

Sensor ID	Location	Data Type	Voltage (V)	Current (A)	Temperature (°C)	Humidity (%)	Power Factor	Frequency (Hz)	Phase Imbalance (%)	Timestamp	Status
001	Substation A	Voltage	240	600	60	55	0.88	60	2	2024-08-28 10:00:00	Normal
002	Power Plant B	Temperature	230	490	85	60	0.85	48.5	3	2024-08-28 10:05:00	Warning
003	Transmission Line 1	Current	235	485	82	57	0.87	51.2	2.5	2024-08-28 10:10:00	Normal
004	Substation C	Voltage	220	520	78	50	0.89	60	2	2024-08-28 10:15:00	Normal
005	Power Plant A	Temperature	300	570	80	70	0.82	59	3	2024-08-28 10:20:00	Critical
006	Transmission Line 2	Current	315	580	83	62	0.84	59.8	2.7	2024-08-28 10:25:00	Normal
007	Substation B	Voltage	335	605	75	53	0.88	60	2	2024-08-28 10:30:00	Normal
008	Power Plant C	Temperature	310	575	95	68	0.83	59.2	3.5	2024-08-28 10:35:00	Warning
009	Transmission Line 3	Current	320	590	81	56	0.86	60.1	2.2	2024-08-28 10:40:00	Normal
010	Substation D	Voltage	320	600	80	54	0.87	60	2	2024-08-28 10:45:00	Normal

3.2 SPHC algorithm

The SPHC algorithm combines numerous sophisticated machine-learning classifiers to enhance power system safety monitoring. The algorithm follows a set of well-defined stages to guarantee an efficient and precise forecast of system status.

The algorithm starts by initializing five improved classifiers, each with its own set of hyperparameters to improve efficiency. These classifiers include Improved RandomForest (RF), Improved J48, Improved Support Vector Machine (SVM), Improved naive Bayes (NB), and Improved k-Nearest Neighbors (KNN).

The hyperparameters were chosen to improve efficiency while minimizing overfitting. The Improved RandomForest employs 100 trees to guarantee diversity and resilience, with a maximum depth of 10 to avoid excessive intricacy and overfitting. To enhance feature selection and decrease the risk of overfitting, each split was assigned a square root of total attributes. The Improved J48 classifier uses a 0.15 confidence factor for aggressive pruning to simplify the model and enhance interpretability, as well as a minimum of 5 instances per leaf to guarantee prediction stability. The Improved SVM utilizes a regularization parameter 'C' of 1.0 to efficiently balance margin maximization and error reduction, whereas the RBF kernel captures the data's non-linear relationships.

The Improved NaiveBayes uses kernel density estimation to represent continuous features more accurately, as well as supervised discretization to better manage numeric data. Finally, the Improved k-Nearest Neighbors is configured with 5 neighbors to balance local influence and generalization. Inverse distance weighting is used to prioritize closer data points, and cross-validation improves the model's resilience and aids prevent overfitting.

During the training phase, each improved classifier is trained on the training dataset ('X_train' and 'Y_train'). This step includes fitting each model to the data, allowing it to learn patterns and relationships that are necessary for creating precise forecasts. After training, each classifier predicts on the test dataset ('X_test'). These forecasts are gathered and saved independently, preparing them for the next ensemble approach. The ensemble approach uses majority voting to make its final forecast. For each instance in the test dataset, forecasts from all five improved classifiers are combined. The majority vote among these forecasts determines the result, guaranteeing that the most common prediction is selected. This method uses the advantages of each classifier to improve overall prediction accuracy. The mathematical expression for the majority voting procedure is as follows:

$$y_{pred}(i) = mode\{y_{RF}(i), y_{J48}(i), y_{SVM}(i), y_{NB}(i), y_{KNN}(i)\} \quad (1)$$

Where:

- $y_{pred}(i)$ is the end predicted label for instance
- $y_{RF}(i), y_{J48}(i), y_{SVM}(i), y_{NB}(i), y_{KNN}(i)$ are the forecasts from the RandomForest, J48, SVM, NaiveBayes, and k-Nearest Neighbors classifiers, correspondingly.
- mode refers to the statistical mode function, which chooses the most commonly recurring prediction among the classifiers.

The algorithm finishes by returning the forecasted labels for the test dataset as 'Y_predictions', resulting in an in-depth evaluation of the power system's safety status using the integrated knowledge from all classes. Algorithm 1 demonstrates the SPHC algorithm.

Algorithm 1: SafePower Hybrid Classifier (SPHC) Algorithm		
Input	:	**X_train, Y_train:** Training attributes and labels **X_test:** Testing attributes
Output	:	**Y_predictions**: Final forecast labels for X_test
Step 1	:	**Initialize classifiers with hyperparameters tuning:** • I_RF (numTrees=100, maxDepth=10, numFeatures=sqrt(total_features)) • I_J48 (confidenceFactor=0.15, minNumObj=5, useUnpruned=False) • I_SVM (C=1.0, kernel=RBFKernel, gamma=auto) • I_NB (useKernelEstimator=True, useSupervisedDiscretization=True) • I_KNN (k=5, distanceWeighting=Inverse, crossValidate=True)
Step 2	:	**Training Phase:** • Train I_RF on X_train, Y_train • Train I_J48 on X_train, Y_train • Train I_SVM on X_train, Y_train • Train I_NB on X_train, Y_train • Train I_KNN on X_train, Y_train
Step 3	:	**Ensemble Tactic:** • Predict with each classifier on X_test • Store predictions: RF_predictions, J48_predictions, SVM_predictions, NB_predictions, KNN_predictions
Step 4	:	**Majority Voting:** • For each instance i in X_test:

		o Gather votes from RF_predictions, J48_predictions, SVM_predictions, NB_predictions, KNN_predictions o Calculate the majority vote for i o Add majority vote to Y_predictions
Step 5	**:**	Return Y_predictions

Improved random forest:

Random Forest is an ensemble learning technique that generates a large number of decision trees during training and returns the mode of the classes for classification issues or the mean prediction for regression tasks. It is well-known for its durability and efficacy in managing complicated datasets, owing to its capacity to decrease overfitting and increase forecast accuracy by averaging numerous decision trees. The Improved RandomForest classifier was created on this basis by making many significant improvements to enhance its efficiency even more. Raising the number of trees ('numTrees') to 100 improves the classifier's stability and minimizes variance, as combined predictions from a larger number of trees result in a more dependable and consistent model.

$$y_{RF}(i) = mode\{T_1(x_i), T_2(x_i), \dots, T_{100}(x_i)\} \quad (2)$$

where:

- $y_{RF}(i)$ is the final prediction for instance i.
- $T_j(x_i)$ denotes the prediction created by the j-th decision tree for instance i.
- mode represents the statistical mode function, which chooses the most common prediction among the trees.

Furthermore, restricting each tree's maximum depth ('maxDepth') to 10 keeps the model from becoming overly complicated lowering the risk of overfitting and guaranteeing greater adaptability to new data. To increase variety among the trees and enhance total predictive efficiency, the number of attributes considered at each split ('numFeatures') is set to the square root of the total number of attributes. These improvements contribute to a more precise and reliable RandomForest model.

Improved J48:

J48 is an execution of the C4.5 algorithm, which generates decision trees for classification tasks. This classifier is well-known for its comprehension and ability to manage both numerical and categorical data sets. J48 creates a decision tree by iteratively splitting the data using the feature with the greatest information gain, resulting in a useful tool for recognizing trends and creating decisions depending on complicated data sets.

The Improved J48 decision tree classifier improves its efficacy by adjusting particular parameter values to enhance generalization. By lowering the confidence factor to 0.15, the algorithm engages in more aggressive pruning, which simplifies and decreases the tree's intricacy, thus avoiding overfitting. Raising the minimum number of instances per leaf ('minNumObj') to 5 prevents the tree from forming excessively particular branches, resulting in a more generalizable model. Furthermore, setting the pruning choice to false ('useUnpruned=False') guarantees that the tree undergoes pruning, simplifying the model and improving its capability to generalize to novel, unseen data.
In terms of prediction, for a given instance i, the Improved J48 classifier creates a prediction $y_{J48}(i)$ using the built decision tree. The prediction is denoted by:

$$y_{J48}(i) = Class(LeafNode(Tree(x_i))) \quad (3)$$

where:

- x_i is the feature vector of the instance i.
- $Tree(x_i)$ denotes the decision tree performed to an instance x_i.
- LeafNode represents the terminal node of the tree where x_i ends up.
- The class represents the class label allocated to the leaf node, which is the end prediction $y_{J48}(i)$.

These improvements finding in a more efficient J48 classifier with enhanced efficiency across various datasets.

Improved SMO (SVM):

The Support Vector Machine (SVM) is a potent classification method known for its ability to identify the best hyperplane that divides data into different classes. The SMO (Sequential Minimal Optimization) algorithm is widely utilized to effectively address the SVM optimization issue, which makes it ideal for massive data sets and complicated classification tasks. To enhance classification accuracy, the Improved SMO (SVM) classifier is carefully tuned in terms of hyperparameters. The regularization parameter 'C' is set to 1.0, balancing the trade-off between increasing the margin between classes and reducing classification failures. This setting allows the model to generalize well to novel data by regulating the intricacy of the decision limit. The Radial Basis Function

(RBF) kernel was chosen because of its capability to map data into higher dimensions, enabling the SVM to better manage non-linear relationships. The gamma parameter is set to auto, which allows the model to adapt to changing data distributions and improve the division between classes.
In terms of prediction, for a given instance x, the Improved SMO (SVM) classifier creates a prediction $y_{SVM}(x)$ using the decision function:

$$y_{SVM}(x) = sign(\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b) \quad (4)$$

where:

- α_i are the Lagrange multipliers attained from the SMO algorithm.
- y_i are the class labels of the support vectors.
- $K(x, x_i)$ is the kernel function (Radial Basis Function in this case).
- b is the bias term.
- sign function computes the class label by assessing the sign of the decision function.

Improved Naive Bayes:

Naive Bayes is a probabilistic classifier that relies on Bayes' theorem and presumes feature independence. It is well-known for its ease and effectiveness, especially in situations involving massive data sets and categorical variables. Despite its simplicity, Naive Bayes can be extremely successful when properly tuned for a particular task. The Improved Naive Bayes classifier uses sophisticated methods to enhance its handling of numerical data. By using kernel density estimation ('useKernelEstimator=True'), the classifier abandons the assumption of a normal distribution for numerical attributes in Favor of a more flexible estimation technique. This enhancement enables more precise modelling of feature distributions, particularly when the actual data deviates from normal. Furthermore, supervised discretization ('useSupervisedDiscretization=True') is used to convert continuous attributes into categorical bins, taking advantage of their relationship with the target variable.
For the Improved Naive Bayes classifier, the predicted class label $y_{NB}(i)$ for a given instance i is computed by discovering the class with the highest posterior probability. Given a feature vector x_i for instance i, the predicted class $y_{NB}(i)$ is calculated as:

$$y_{NB}(i) = \arg max_y \, (P(y).P(x_i|y)) \quad (5)$$

where:

- $P(y)$ is the prior probability of class y.
- $P(x_i|y)$ is the possibility of the feature vector x_i given class y, assessed utilizing kernel density estimation or supervised discretization.

Improved k-Nearest Neighbours (IBk):

The k-nearest Neighbours (KNN) algorithm is an easy but effective classification technique based on similarity. It categorizes data points using the majority class of their k-nearest neighbours, making it simple and understandable. KNN is extremely adaptable and effective in a variety of applications, especially when the decision boundary between classes is complicated. To enhance the performance of the Improved k-Nearest Neighbours (IBk) classifier, various strategic parameter adjustments are made. Setting the number of neighbours ('k') to 5 strikes a balance between bias and variance, making the classifier more resistant to noise while maintaining accuracy. The employing of distance weighting with an inverse technique gives closer neighbours more impact, which enhances classification accuracy, particularly in areas with changing data density. Additionally, using cross-validation ('crossValidate=True') during training assists in autonomously choosing the optimum value of 'k', guaranteeing that the classifier operates optimally. These refinements result in a more precise and dependable k-Nearest Neighbours classifier, which makes it suitable for complicated and varied datasets.
For the Improved k-Nearest Neighbours (IBk) classifier, the predicted class label $y_{KNN}(i)$ for a given instance i is computed by considering the class labels of its k-nearest neighbours, with a distance-weighted voting method.

Given an instance x_i, the predicted class $y_{KNN}(i)$ is calculated as follows:

$$y_{KNN}(i) = \arg max_y \, (\sum_{j=1}^{k} w_j . [\![(y_j = y)) \quad (6)$$

where:

- w_j is the weight allocated to the j-th nearest neighbor, typically calculated as $w_j = \frac{1}{d_j}$ with d_j being the distance from the j-th neighbor to the instance i.
- $[\![(y_j = y)$ is an indicator function that is 1 if the class label of the j-th neighbor y_j is equal to y, and 0 otherwise.
- k is the number of neighbors considered.

This section describes the procedure of using an extensive data set and a sophisticated classification algorithm to improve power system safety monitoring. This method tackles the shortcomings of previous monitoring systems by using the SPHC algorithm, which uses ensemble voting to incorporate numerous classification models. The capability of the SPHC algorithm to manage real-time data, precisely classify safety statuses and present timely alerts greatly enhances power operations' dependability and stability. The proposed system's strong classification guarantees that it can handle safety problems and reduce possible risks in power systems. Figure 1 displays the system architecture of the SPHC algorithm.

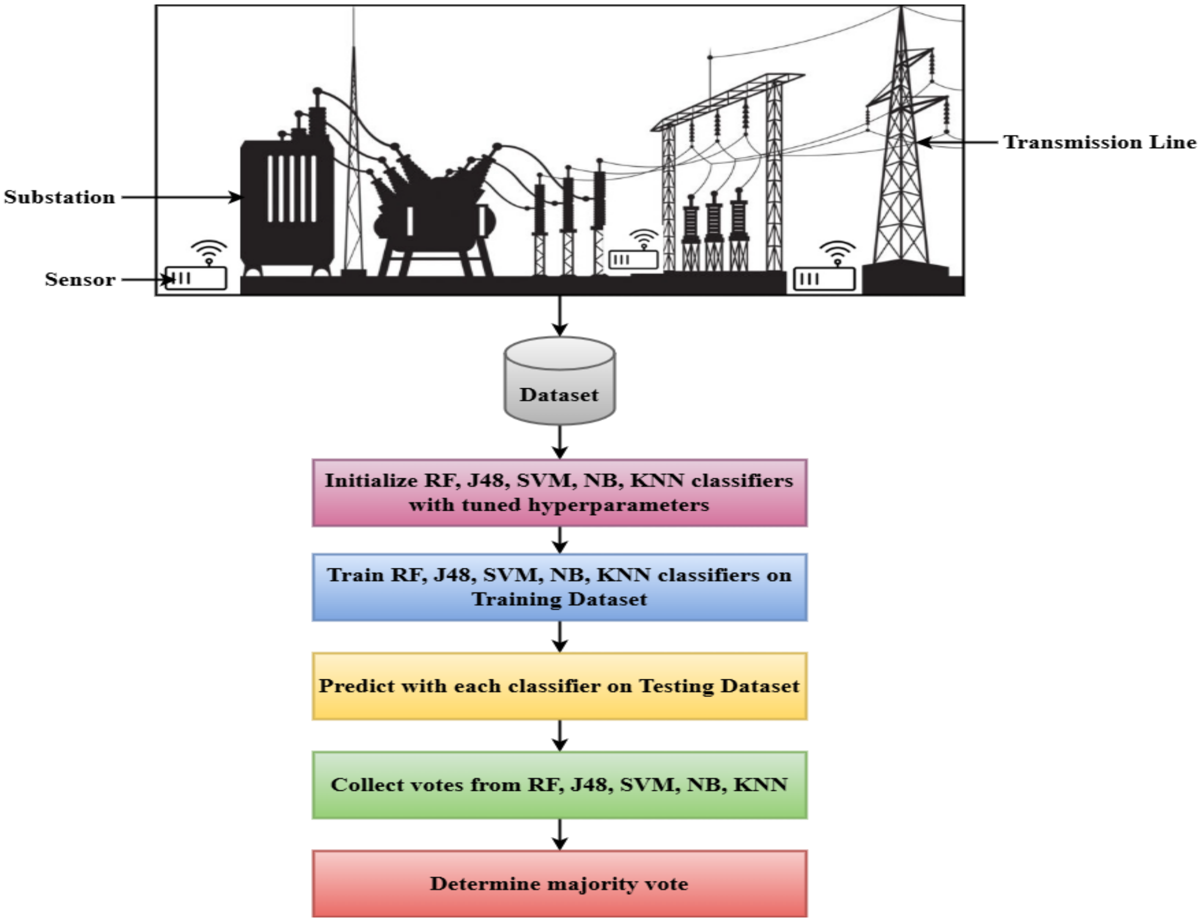

Figure 1: System architecture of SPHC algorithm

4 Experimental results and discussions

This section provides an in-depth evaluation of the experimental findings attained by comparing the SPHC algorithm to numerous well-known classifiers: RandomForest, J48, SMO (SVM), NaiveBayes, and k-Nearest Neighbors (IBk). The experiments were executed in Java utilizing the Weka tool, with a software configuration that included Windows 11 64-bit OS, Apache NetBeans IDE 15, and JDK 8. The experiments' hardware specifications include an Intel Core i7 processor, 16 GB of RAM, and a 512 GB SSD, which enabled effective data processing and model training. Each classifier's effectiveness was evaluated using important metrics such as accuracy, precision, recall, F1-score, and MCC. These metrics present an overall assessment of each classifier's ability to handle the dataset and make precise forecasts.

4.1 Performance metrics

To examine the efficiency of the classifiers, the following metrics were utilized:

Accuracy: This metric calculates the percentage of accurate outcomes among all cases assessed. It is a basic indicator of a classifier's overall efficacy, including both true positives and true negatives. The equation for accuracy is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

where TP stands for True Positives, TN for True Negatives, FP for False Positives, and FN for False Negatives. A higher accuracy denotes a greater overall accuracy of the classifier.

Precision: Precision measures the proportion of true positives among those predicted as positive. It is especially essential in cases where the expense of false positives is high. Precision is calculated as:

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

High precision means that when the classifier forecasts a positive class, it is likely to be correct.

Recall: Recall, also referred to as sensitivity or true positive rate, is an indicator of how many actual positives the classifier accurately identifies. It is critical in circumstances where missing a positive case is expensive. The recall is given by:

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

High recall indicates that the classifier accurately recognizes the majority of the actual positive cases.

F1-score: The F1-score strikes a balance between precision and recall, which is especially helpful when the dataset is imbalanced. It is the harmonic mean of precision and recall and is computed as:

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (10)$$

A higher F1 score denotes that the classifier executes well in terms of both precision and recall.

Matthews Correlation Coefficient (MCC): MCC is a more thorough metric that considers all four categories (TP, TN, FP, and FN) and is particularly helpful when dealing with imbalanced datasets. It's computed utilizing the formula:

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (11)$$

An MCC value near 1 implies a high positive correlation between the predicted and actual classifications.

4.2 Experimental setup

The classifiers were evaluated on a standard dataset with the Weka tool, which offers a reliable setting for executing and comparing machine learning techniques. The SPHC algorithm was tested against RandomForest, J48, SMO (SVM), NaiveBayes, and k-Nearest Neighbors (IBk). The dataset was divided into two sets: training and testing. Each classifier was trained on the training set and assessed on the testing set. The performance metrics for each classifier were computed and tabulated. Table 3, compares the efficiency of each classifier across the different metrics. The SPHC algorithm consistently surpassed the others, with the greatest values across all metrics.

Table 3: Performance metrics comparison

Classifier	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**MCC (%)**
RandomForest	88.8	85.4	86.2	85.8	82.3
J48	85.3	83.2	84.1	83.6	79.6
SMO	89.2	87.0	87.6	87.3	84.8
NaiveBayes	81.6	80.0	80.4	80.2	75.2
k-Nearest Neighbors	86.1	84.6	85.1	84.8	81.0
SPHC	92.4	91.3	91.8	91.5	89.7

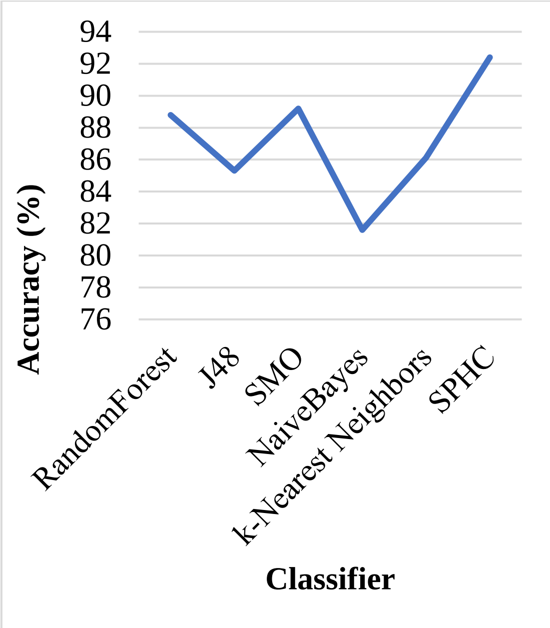

Figure 2: Accuracy comparison

Figure 2 compares the accuracy of each classifier, revealing that the SPHC algorithm outperforms all others with an accuracy of 92.4%. This higher accuracy is most likely owing to the SPHC algorithm's hyperparameter-tuned ensemble methods, which enable it to better model the data and decrease error rates. The integration of various base classifiers within SPHC algorithm improves its capacity to capture different trends in the data, resulting in more precise predictions.

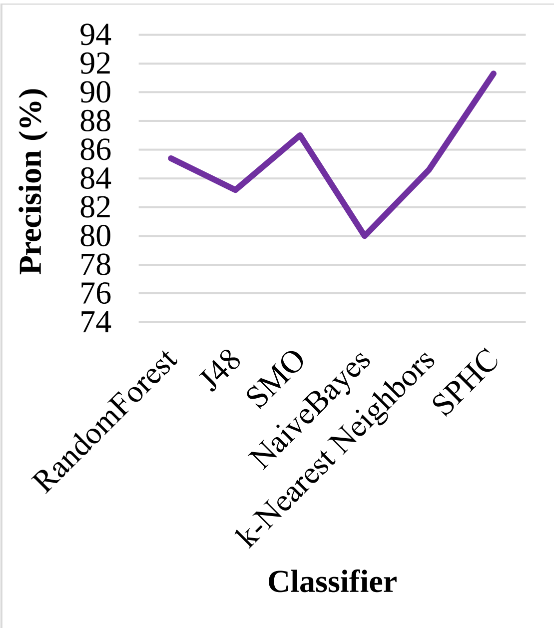

Figure 3: Precision comparison

Figure 3 depicts the precision comparison, with the SPHC algorithm also leading at 91.3%. The high precision suggests that SPHC algorithm efficiently reduces false positives, which makes it especially useful in situations where erroneous positive predictions could have serious effects. The SPHC algorithm's strong management of various attribute types and its ensemble method most likely assist in its high precision, guaranteeing that it correctly recognizes true positives while preventing incorrect classifications.

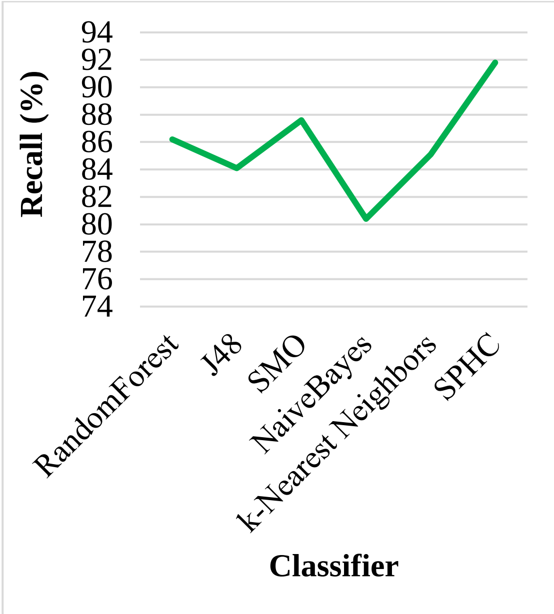

Figure 4: Recall comparison

Figure 4 compares the classifiers' recall rates, with the SPHC algorithm attaining the highest recall in this assessment at 91.8%. High recall indicates that SPHC algorithm is efficient at detecting true positives, which is critical in applications where missing a positive instance could be costly. SPHC algorithm's high recall could be attributed to its thorough learning strategy, which incorporates numerous models and improves their parameters to guarantee that the most positive instances are identified.

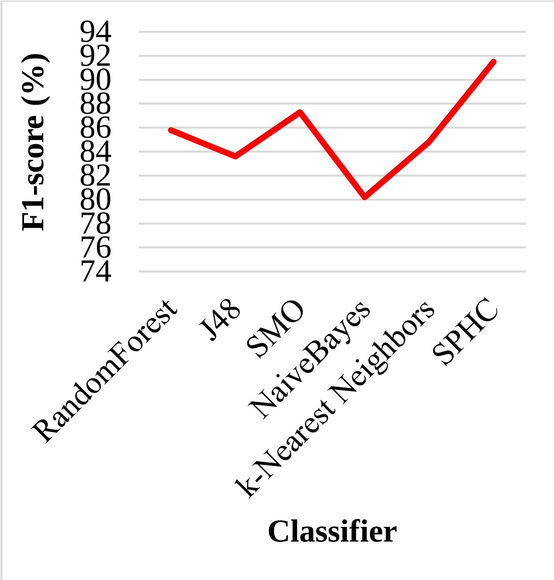

Figure 5: F1-score comparison

The F1-score comparison in Figure 5 indicates that the SPHC algorithm achieves a balanced efficiency with an F1-score of 91.5 percent. This high F1 score shows that SPHC algorithm strikes a good balance between precision and recall, which makes it an excellent selection for purposes that require both metrics. SPHC algorithm's capability to retain high precision and recall concurrently demonstrates its advanced design, which efficiently manages various levels of data intricacy.

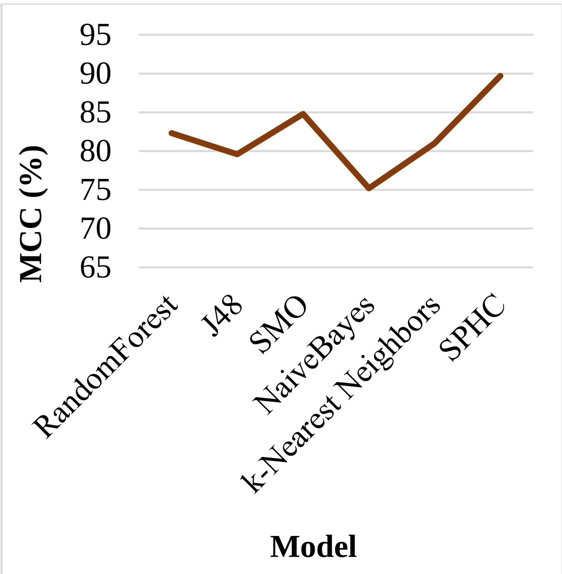

Figure 6: MCC comparison

Figure 6 compares the MCC values of the classifiers and shows that the SPHC algorithm has the highest MCC of 89.7%. This suggests a strong correlation between predicted and actual classifications, even in the existence of class imbalances. The high MCC value demonstrates SPHC algorithm's efficiency in making accurate predictions across a broad spectrum of data situations. The algorithm's usage of ensemble learning is critical to attaining this high MCC, guaranteeing that it works well even when the dataset provides difficulties like imbalanced classes.

The experimental findings explicitly show the SPHC algorithm's advantage in all performance metrics. SPHC algorithm regularly surpasses conventional classifiers such as RandomForest, J48, SMO, NaiveBayes, and k-Nearest Neighbors by incorporating sophisticated methods like ensemble learning and model improvement. The SPHC algorithm's capability to attain high accuracy, precision, recall, F1-score, and MCC demonstrates its resilience and dependability when dealing with complicated and varied datasets. These findings indicate that SPHC algorithm is an effective tool for classification tasks, providing substantial advantages over traditional approaches which renders it appropriate for practical uses where high predictive efficiency is critical.

4.3 Discussion

The experimental findings show that the SPHC algorithm surpassed conventional classifiers like RandomForest, J48, SMO, NaiveBayes, and k-Nearest Neighbors in important performance metrics like accuracy, precision, recall, F1-score, and MCC. Particularly, SPHC algorithm outperformed all other classifiers in terms of accuracy (92.4%), precision (91.3%), recall (91.8%), F1-score (91.5%), and MCC (89.7%). SPHC algorithm's outstanding results can be attributed to its ensemble learning strategy, which combines numerous classifications models and uses a voting method to improve decision-making. Furthermore, fine-tuning of model parameters increased the classifier's capacity to precisely identify intricate trends in power system security information, resulting in better efficiency than standalone classifiers such as RandomForest and SMO. However, while SPHC algorithm performed admirably in this setting, its scalability and generalizability across different power grid circumstances and configurations remain possible drawbacks. Further study is required to evaluate its adaptability to larger, more intricate networks and differing functional situations, which may have an influence on its efficacy in real-world applications.

5 Conclusion

This study tackled the vital problem of guaranteeing power system safety and dependability by implementing the SPHC algorithm, which was developed to improve real-time monitoring and classification of power system safety data. The study aimed to improve the identification and avoidance of power system breakdowns with a more comprehensive and real-time assessment technique. The results show that the SPHC algorithm, which utilizes an ensemble voting method, surpasses conventional approaches in correctly identifying safety statuses as Normal, Warning, or Critical. This innovation benefits the area by providing a resilient solution that improves the dependability and stability of power activities, lowering the risk of possible catastrophic breakdowns. However, the study recognizes the constraints of sensor data coverage and the requirement for additional testing in a variety of functional settings. Future research could fill these gaps by investigating more safety metrics, adapting the SPHC algorithm to other kinds of vital infrastructure, and integrating sophisticated methods of machine learning to improve classification accuracy. Extending the use of SPHC algorithm beyond power systems and combining it with other multidisciplinary safety monitoring systems could greatly expand its influence and usefulness across different industries.

References

[1] Peyghami, S., Palensky, P., & Blaabjerg, F., 2020. An overview of the reliability of modern power electronic-based power systems. *IEEE Open Journal of Power Electronics*, 1, 34-50. https://doi.org/10.1109/ojpel.2020.2973926

[2] Rivas, A. E. L., & Abrao, T., 2020. Faults in smart grid systems: Monitoring, detection and classification. *Electric Power Systems Research*, 189, 106602. https://doi.org/10.1016/j.epsr.2020.106602

[3] Siniosoglou, I., Radoglou-Grammatikis, P., Efstathopoulos, G., Fouliras, P., & Sarigiannidis, P., 2021. A unified deep learning anomaly detection and classification approach for smart grid environments. *IEEE Transactions on Network and Service Management*, 18(2), 1137-1151. https://doi.org/10.1109/tnsm.2021.3078381

[4] Gams, M., & Kolenik, T., 2021. Relations between electronics, artificial intelligence and information society through information society rules. *Electronics,* 10(4), 514. https://doi.org/10.3390/electronics10040514

[5] Xie, B., Zhao, D., & Hong, T., 2020. Transformer monitoring and protection in dynamic power systems–a review. *Frontiers in Energy Research*, 8, 150. https://doi.org/10.3389/fenrg.2020.00150

[6] Khan, F., Siddiqui, M. A. B., Rehman, A. U., Khan, J., Asad, M. T. S. A., & Asad, A., 2020, February. IoT-based power monitoring system for smart grid applications. In 2020 *International Conference on Engineering and Emerging Technologies* (ICEET) (pp. 1-5). IEEE. https://doi.org/10.1109/iceet48479.2020.9048229

[7] Lee, Y. H., Kim, S. J., Lee, K. Y., & Nam, T., 2023. Online Robust Subspace Clustering With Application to Power Grid Monitoring. *IEEE Access*, 11, 27816-27828. https://doi.org/10.1109/access.2023.3257357

[8] Manoharan, H., Teekaraman, Y., Kirpichnikova, I., Kuppusamy, R., Nikolovski, S., & Baghaee, H. R., 2020. Smart grid monitoring by wireless sensors using binary logistic regression. *Energies,*13(15),3974. https://doi.org/10.3390/en13153974

[9] Fernandez, J. H., Omri, A., & Di Pietro, R., 2023. Power grid surveillance: Topology change detection system using power line communications. *International Journal of Electrical Power & Energy Systems*, 145, 108634. https://doi.org/10.1016/j.ijepes.2022.108634

[10] Yi, Z. H. A. N. G., Feng, Z. H. A. N. G., Youchun, L. I., Lv, T. A. N. G., Xuehu, P. E. N. G., & Jianguo, M. O., 2021, February. Analysis Of Typical Power Grid Blackout Accidents And Suggestions For Countermeasures. In *Journal of Physics: Conference Series* (Vol. 1754, No. 1, p. 012023). IOP Publishing. https://doi.org/10.1088/17426596/1754/1/012023

[11] Li, T., & Liu, J., 2021, August. Research on risk assessment of power grid based on real-time simulation system. *In Journal of Physics: Conference Series* (Vol. 2005, No. 1, p. 012127). IOP Publishing. https://doi.org/10.1088/1742-6596/2005/1/012127

[12] Rao, F., Xi, X., Wu, S., Sheng, J., & Zhang, Y., 2022. Safety Helmet Monitoring of Power Grid Staff Based on Improved YOLOv3. *In Mechatronics and Automation Technology* (pp. 58-66). IOS Press. https://doi.org/10.3233/atde221150

[13] Wang, S., & Dehghanian, P., 2020. On the use of artificial intelligence for high impedance fault detection and electrical safety. *IEEE Transactions on Industry Applications,* 56(6), 7208-7216. https://doi.org/10.1109/tia.2020.3017698

[14] Hossain, M. S., Ahmed, R., & Hossain, S., 2021. Design and optimization of substation grounding grid for ensuring the safety of personnel and equipment. *Journal of Electrical Power & Energy Systems,* 5(1), 71-80. https://doi.org/10.26855/jepes.2021.08.001

[15] Alimi, O. A., Ouahada, K., & Abu-Mahfouz, A. M., 2020. A review of machine learning approaches to power system security and stability. *IEEE Access*, 8, 113512-113531. https://doi.org/10.1109/access.2020.3003568