Bishun Zhao[1*]

Shaoqing Zuo[2]

Ling Liu[3]

Chen Li[4]

Min Xu[5]

Violation Intelligent Recognition Algorithm Based on Multi-source Data Fusion

Abstract: - The rapid evolution of the power industry, coupled with the increasing integration of intelligent technologies, has spurred a significant demand for advanced monitoring and identification systems for power equipment and their operational status. As power infrastructure becomes more complex and interconnected, the need for precise monitoring and real-time analysis has become paramount to ensure reliability, safety, and efficiency. In this manuscript, Violation Intelligent Recognition Algorithm Based on Multi-source Data Fusion (VIRA-BO-MSDF) is proposed. The multimedia data includes real-time sensor monitoring data, video and image data, and Power Safety Workflow includes the correct process description of power operation. Initially the multimedia data is preprocessed utilizing Master-Slave Adaptive Notch Filter (MSANF) is used to clean the data. Then, feature extraction is done by Dual Tree Complex Discrete Wavelet Transform (DTCDWT) to extract the gray-scale statistical features such as Homogeneity, Entropy, Energy and harmony. Then, the extracted features are given to Relational Bilevel Aggregation Graph Convolutional Network (RBAGCN) is used to identifying violations in the power industry. In general, RBAGCN does not express some adaption of optimization strategies for determining optimal parameters to assure accurate identification of Violation Intelligent Recognition in power industry. Therefore, Circulatory System-based Optimization (CSBO) is proposed to optimize weight parameter of RBAGCN, which accurately identify violations in the power industry. The proposed model is implemented, efficacy is assessed utilizing some performance metrics likes accuracy, precision, specificity, sensitivity, F1-Score, computation time, ROC, Mean Squared Error. The VIRA-BO-MSDF method provides 28.46%, 21.34 and 33.81% higher accuracy, 22.88%, 26.52% and 34.63% higher Precision, 28.46%, 21.34 and 33.81% higher specificity, 22.37%, 27.89%, and 31.37% higher sensitivity is analyzed with existing method such as Deep learning-depend substation remote construction management and AI automatic violation detection system (DLB-RCM-AVDS), Applying Deep Learning-depend concepts for the detection of device misconfigurations in power systems (DLB-DMC-PS-AVD) and Deep OPF: A Feasibility-Optimized Deep Neural Network Method for AC Optimal Power Flow Problems (FODNN-PS-AVD)respectively.

Keywords: Circulatory System-based Optimization, Dual Tree Complex Discrete Wavelet Transform, Master-Slave Adaptive Notch Filter, power industry, Relational Bilevel Aggregation Graph Convolutional Network,

I. INTRODUCTION

As a vital component of the national economy, the power industry plays a crucial role in ensuring a safe and s[power supply [1]. However, the growing number and complexity of power equipment pose challenges to maintaining safe operations within the industry, as potential violations become increasingly prevalent [2]. Traditional methods for identifying violations are hindered by their reliance on single data sources and limited algorithmic capabilities, falling short of the power industry's demand for efficient and accurate violation detection [3]. To address these challenges, this paper proposes an intelligent identification algorithm leveraging multi-source data fusion, integrating deep learning and machine vision technologies [4]. By amalgamating data from various sources, this approach aims to enhance the industry's capability to identify violations effectively [5]. Through the synergy of deep learning and machine vision, the proposed algorithm seeks to enhance accuracy, efficiency of violation detection within power sector, thus contributing to its safe and reliable operation [6-8]. The dynamic evolution of the power sector, accompanied by the rising integration of smart

[1] *[1*]Engineer,Safety Supervision Dept., Foshan Power Supply Bureau of Guangdong Power Grid Co., Ltd., Foshan, Guangdong, 528000, China*

[1]Email: bishunzhaophd@gmail.com*

[2]Engineer,Safety Supervision Dept., Foshan Power Supply Bureau of Guangdong Power Grid Co., Ltd., Foshan, Guangdong, 528000, China

[3]Senior Engineer, Safety Supervision Dept., Foshan Power Supply Bureau of Guangdong Power Grid Co., Ltd., Foshan, Guangdong, 528000, China

[4]Assistant engineer, Department of Power Distribution Technology, Electric Power Research Institute, CSG, Guangzhou, Guangdong, 510663, China

[5]Senior Engineer, Department of Power Distribution Technology, Electric Power Research Institute, CSG, Guangzhou, Guangdong, 510663, China

technologies, highlights the pressing need for proficient monitoring and identification of power equipment and operational conditions [9]. In acknowledgment of this pressing requirement, this paper introduces an intelligent algorithm for violation identification, grounded in multi-source data fusion [10- 12]. Drawing from diverse data streams including power-related documents, video footage, and images, the algorithm seamlessly combines deep learning and machine vision techniques to deliver comprehensive analysis and decision-making capabilities [13-14]. Empirical results affirm the algorithm's remarkable precision in identifying violations, affirming its resilience and dependability [15]. This innovative approach holds immense potential for bolstering the safety and efficiency of power industry operations. Proposed here is a comprehensive system that amalgamates management functionalities with advanced machine learning object detection techniques [16]. This system is designed to enhance the power construction management process by infusing intelligence and real-time violation detection capabilities [17]. By leveraging cutting-edge machine learning algorithms, the system aims to intelligently monitor construction activities and promptly identify any violations that may occur during the process. This innovative approach promises to revolutionize power construction management by ensuring greater efficiency, safety, and compliance with regulatory standards [18].

A. Problem statement and Motivation

Also, owing to massive dimensions, enormous volumes of data, several challenges, including such systems are complex, need more storage and computing resources. They face challenges owing to unrelated, redundant data, making difficult to find Violation Intelligent Recognition Based Multi-source Data Fusion. Another challenge is the low Violation Intelligent Recognition and high false alert rate. To address issues in identifying violations in real-time, data fusion method offers likely result. Such inspired to do us this investigation work.

It utilized feature level fusion to combine data from various sources, provide further inclusive view of data analysis. Meanwhile relational database managing systems fail to handle important data storage problems, the dataset was evaluated with RBAGCN. The RBAGCN is desired because employs multiple weaker learners to construct strong learners.

B. Contribution

Major contribution of this investigation work is summarized below,

- Violation Intelligent Recognition Algorithm depend on Multi-source Data Fusion (VIRA-BO-MSDF) is proposed.
- The multimedia data encompasses real-time sensor monitoring data, video and image data, while the Power Safety Workflow comprises accurate descriptions of power operation processes for more accurate and informative results.
- Applying fused data to RBAGCN to evaluate performance after fusion.
- In this manuscript is arranged as below: section 2, introduce literature review. Section 3, present proposed method. Section 4 displays outcome with discussions. Section 5, concludes this manuscript.

II. LITERATURE REVIEW

Among the frequent investigation works on deep learning based Violation Intelligent Recognition Based Multi-source Data Fusion; some of the latest investigations were assessed in this part.

Yan et al., [19] have presented DL-depend substation remote construction management with automatic violation detection system. In this manuscript, suggested a system aimed at enhancing workplace safety in power construction sites by integrating remote substation construction management with artificial intelligence object detection methods. This system aims to streamline power construction management processes, enable real-time identification of operational violations, thus mitigating the risk of operator injuries. It provides high accuracy and low precision.

Fellner et al., [20] have presented applying DL-depend concepts for the detection of device misconfigurations in power systems. Here, in order to effectively operate in new environments resulting from decentralization of power generation, power systems often rely on grid-supportive functions offered by such devices. Such functions encompass control mechanisms like reactive power dispatch for voltage control otherwise active power reduction based on voltage levels. Through generation and analysis of operational data from power distribution grids, DL-depend method was employed to address detection challenge at hand.It provides high precision and it provides low specificity.

Pan et al., [21] have presented Deep OPF: A Feasibility-Optimized DNN Method for AC Optimal Power Flow Problems. Deep OPF initially trains a DNN method to forecast subset of independent operating variables, subsequently computes remaining ones by solving power flow equations directly. This methodology not only upholds power-flow balance equality constraints but also diminishes quantity of variables to be anticipated by DNN, thereby reducing necessity for numerous neurons, training data. In the training phase, Deep OPF utilizes penalty method alongside zero-order gradient estimation method to ensure compliance with inequality constraints. Additionally, establish criterion for adjusting DNN size based on desired approximation accuracy, evaluating its generalization capability. It attains high specificity and it provides low sensitivity.

Xiao et al., [22] have presented Renewable Energy-Powered Semi-Closed Greenhouse for Sustainable 2 Crop Production utilizing Model Predictive Control and ML for 3 Energy Management. Here, adaptive data fusion method that uses multi-source AIS data to predict vessel trajectories in maritime traffic. Multi-source AIS data is fused together using similar marine mobile service identifier, timestamp. DL techniques were used to improve feature learning, flexibility. It was also uses adjacency distance to improve correlation of dataset trajectory paths. It provides high sensitivity and it provides low F1-Score.

Wang et al., [23] have presented Investigation into Recognition Algorithm of Helmet Violation depend on YOLOv5-CBAM-DCN. To be more precise, employ a graph neural network to acquire a nonlinear representation linking the demanded power to its corresponding allocation. This solution is learned in an unsupervised fashion, with a direct minimization of the associated cost. To accommodate the electrical constraints of the grid, we introduce a novel barrier method that is differentiable and operates effectively on points initially deemed infeasible. It provides high precision and it provides low accuracy.
Nellikkath and Chatzivasileiadis [24] have presented Physics-Informed Neural Networks for AC Optimal Power Flow. Introduce physics-informed neural networks for accurate estimation of AC-Optimal Power Flow outcomes, providing robust assurances regarding their performance. Optimal Power Flow (OPF) algorithms are increasingly employed by power system operators and various stakeholders for diverse applications, encompassing both planning and real-time operations. Though, the original AC OPF problem was frequently intricate to solve owing to its nonlinear, non-convex nature. It provides highF1-Scoreand it provides low precision.

Yuan et al., [25] have presented Multi-Source Data Processing and Fusion Technique for Power Distribution IoT Depend on Edge Intelligence. As the Energy Internet approach advances rapidly, proliferation of sensors within Power Distribution Internet of Things has surged significantly. To address challenges related to convoluted storage, inadequate fusion computing performance of heterogeneous distribution data from multiple sources, presented an edge intelligence-depend PD-IoT multi-source data processing, fusion approach. Initially, a PD-IoT multi-source data processing, fusion architecture is crafted, leveraging edge smart terminals. It attains higher accuracy, lower specificity.

III. PROPOSED METHODOLOGY

The Violation Intelligent Recognition Algorithm Based on Multi-source Data Fusion (VIRA-BO-MSDF) is proposed. Block diagram of proposed VIRA-BO-MSDF approach is represented in Fig 1. This process comprises of five steps such as data acquisition, pre-processing, feature extraction, classification, optimization. Accordingly, detailed description of all step given as below,

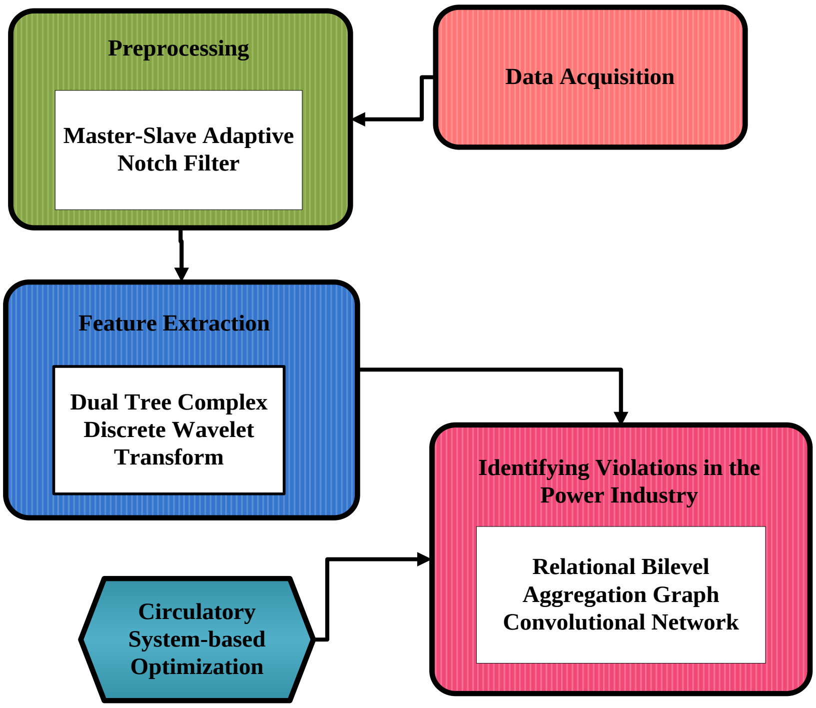

Figure1: Block Diagram for proposed VIRA-BO-MSDF method

A. Data acquisition

The operational procedure incorporates a variety of multi-source heterogeneous data within the power system domain. This data encompasses sample document data and multimedia data, each serving distinct purposes. The sample document data comprises professional text information pertinent to the power system, encompassing fundamental knowledge documents on electrical power and regulatory documents concerning safety supervision. The former entails details regarding equipment, instruments, and operational protocols pertinent to electrical power operation, while the latter delineates violations, their classifications, and corresponding codes, ensuring adherence to safety protocols. On the other hand, multimedia data encompasses real-time sensor monitoring data, video, and image data, augmenting the Power Safety Workflow with accurate descriptions of power operation processes. Then, collected data are given to pre-processing.

B. Pre-processing utilizing Master-Slave Adaptive Notch Filter

The data pre-processing utilizing MSANF [26] is discussed. The proposed MSANF is used to clean the data. MSANF can be used in a presence of specific frequency components affects data quality. Filters can change their characteristics in response to changes in the quality of data. This process entails converting categorical variables to numerical format, as many data analysis require numerical input. It may be used in the context of data to remove, interference, or unwanted frequency components from the input data. It is recognised that frequency adaptation is crucial MSANF to acquire exact tracking outcomes. Quality is given in equation (1),

$$\dot{\hat{\rho}} = -\frac{\ell}{2\aleph} y_1\left(\ddot{y}_1 + \hat{\rho}^2 y_1\right) \tag{1}$$

Where, $\dot{\hat{\rho}}$ represents the quality, ℓ represents the redundancy in the data, $\aleph$ represents the response speed of filter, y_1 represents the saved data, $\ddot{y}_1$ represents the number of saved data. In data preprocessing, it's essential to ensure data integrity and accuracy through steps like data cleaning, denoising, and feature extraction. The component values is given in equation (2),

$$y_1 = -\frac{C}{\rho_1}\sin\left(\rho_1 f + \phi_1\right) \tag{2}$$

Where, C represents the component values, ρ_1 represents the frequency value convergence, $\rho_1 f$ represents the converting values, ϕ_1 represents the accuracy. Data cleaning involves eliminating abnormal and missing values. Denoising techniques, as MSANF, help remove noise interference. Then, the mathematical calculation of clean data is formulated in equation (3),

$$y_1^2 = \frac{C^2}{2\rho_1^2}\left(1+\sin\left(2\left(\rho_1 f + \phi_1\right)\right)\right) \tag{3}$$

Where, y_1^2 is the samples data calculation, C represents the numeric values, ρ_1 represents the character value, $\rho_1 f$ represents the converting values, ϕ_1 represents the accuracy. Finally, MSANF method cleaned the data. Then, pre-processed data is given to feature extraction.

C. Feature extraction utilizing Dual Tree Complex Discrete Wavelet Transform

The DTCDWT [27] is proposed to extract gray-scale statistical features such as Homogeneity, Entropy, Energy and harmony. The DTCDWT is a data that divides the data into frequency components at various scales. It is a mustier solution technique for feature extraction in data analysis tasks. A powerful instrument for describing a data fusion frame in both the temporal and frequency domains is the wavelet transform. Its resolution is good in the time, frequency domains. It is given in equation (4),

$$g(r) = \sum_{h,d} p_{h,d}\, w_{h,d}(r), \tag{4}$$

Where, $g(r)$ represents the linear input features of basic function, $p_{h,d}$ represents the set of weighting features, d represents the dilation fusion parameter, h represents the scaling factor of the features and $w_{h,d}(r)$ represents the set of basic functions, these functions are obtained from the modification of a scaling function β and mother wavelet function μ it is given in equation (5),

$$g(r) = \sum_{d=-\partial}^{\partial} j_d\, \beta(r-d) + \sum_{d=-\partial}^{\partial}\sum_{h=0}^{\partial} v_{h,d}\, \mu\left(2^h r - d\right) \tag{5}$$

Where, j_d *and* $v_{h,d}$ represents the approximation and detail coefficients respectively, ∂ represents the weight in extraction linear layer, β represents the scaling function modification μ represents the mother wavelet. To create a convolution of y and n, samples are first run by low-pass filter with impulse response n. It is given in equation (6),

$$x(m) = (y \times n)[m] = \sum_{d=-\partial}^{\partial} y[d]\, n[m-d] \tag{6}$$

Where, $x(m)$ signifies feature size, $y[d]$ denotes input features, $n[m-d]$ represents the convolutional extraction process, A high-pass filter (f) is also used to simultaneously deconstruct the object. The outputs provide estimate coefficients from low-pass filter, detail coefficients from high-pass filter. High pass filter (f) is given in equation (7),

$$x_{high}[m] = \sum_{d=-\partial}^{\partial} y[d]\, f[2m-d] \tag{7}$$

Where, $y[d]$ represents input features, is the high filter and represents the $f[2m-d]$ represents the weight of the extracted features. To enhance the feature extraction even more, this decomposition is performed again. The features such as Homogeneity, Entropy, Energy and harmony were extracted. The homogeneity of a value distribution in a data is measured. It can be calculated using a formula that takes into account the proximity of each pixel intensity value to the image's mean intensity. The quality of being homogeneous is referred to as homogeneity. Then the formula for calculating homogeneity is given by the Equation (8).

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \tag{8}$$

Where $p(i,j)$ represents pixel value of data, $i-j$ denotes calculation of value in data. Entropy quantifies the uncertainty of data pixel intensities. It is calculated using the pixel value histogram. Then the formula for calculating Entropy is given by the Equation (9).

$$entropy = \sum_{a,b}^{n} p_{i,j} \log(p_{i,j}) \tag{9}$$

Where $p(i,j)$ denotes pixel value of data. The magnitude of the pixel values in data is represented by energy. Then energy sum of the pixel point in an image and it is given by the Equation (10).

$$energy = \sum\sum p(i,j)^2 \tag{10}$$

Where $\sum\sum p(i,j)$ signifies pixel value of data in energy. Then the formula for calculating harmony is given by the Equation (11).

$$\text{harmony} = 1 - \frac{1}{1+m^2} \tag{11}$$

Where m is the harmony of data. Finally, DTCDWT is extracted features such as Homogeneity, Entropy, Energy and harmony. After completing feature extraction, extracted features are fed to RBAGCN.

D. Identifying Violations in the Power Industry utilizing Relational Bilevel Aggregation Graph Convolutional Network

In this section, RBAGCN [28] is discussed. RBAGCN is used for identifying violations in the power industry. This model uses RBAGCN as base identifies to efficiently detect Violation Intelligent Recognition in in the power industry complex networks. The data analytics lifecycle for Violation Intelligent Recognition involves aggregating individual identifiers predictions to reduce variance. In the RBAGCN of Violation Intelligent Recognition this could entail combining data from multiple sensors to improve accuracy of multi-source data fusion violations in the power industry. The violation behavior dataset comprises numerous instances, each accompanied by a classification description, image, and data. These violations are curated through manual selection from historical records, such as Work without wearing helmet and No safety check before operation. The process involves cleaning the data to eliminate irrelevant components like functional words and symbols, retaining only the essential keywords and sentences describing the violation behavior. RBAGCN is an application that uses advanced machine learning techniques to improve the performance of multi-source data. Similarity metric function is given in equation (12).

$$k(w,u) = \left(1 - \frac{ir\,(sim(h_w, h_u))}{\pi}\right) \tag{12}$$

Where, ir represents the Intelligent Recognition, h_w, h_u denotes the features of the data, w represents the target data, u denotes the data fusion, π represents the information Recognition. Then, intelligent recognition of the data is calculated in equation (13).

$$\sigma^{(t)}(y) = \left(Q^{(t)} y + c^{(t)}\right) \tag{13}$$

Where, $\sigma^{(t)}(y)$ denotes the linear transformation function, $Q^{(t)}$ represents the quality of data, y denotes the samples data, $c^{(t)}$ denotes the bias vector. Develop an intelligent violation identification model integrated with rule-based systems for precise violation detection. Utilize real time multi source data to assess the presence of any violation behavior using the intelligent identification model. Then, Violation Intelligent Recognition of combining data is calculated in equation (14)

$$M(u) = \left(\left\{D_g(u), S_g(u)\right\}\right) \tag{14}$$

Where, $M(u)$ denotes the maximum value, $D_g(u)$ represents the connected values in the data, $S_g(u)$ represents the disconnected data. Implement intelligent violation identification by constructing a model based on the data from the rule. Enhance the reliability of this model by pre-training it to recognize violation patterns effectively. Then, identifying violations in the power industry is calculated in equation (15).

$$D_g(u) = \{w | w \in N, (w,u) \in Z\} \tag{15}$$

Here, $D_g(u)$ represents the connected detection value in the data, w represents the target data, u denotes the data fusion, N represents the multimodal interactions, $\in$ indicates the Violation, Z denotes the Intelligent Recognition. Finally, RBAGCN is identified violations in the power industry. In this work, CSBO is employed to optimize RBAGCN. Here, CSBO is employed for tuning weight, bias parameter of RBAGCN.

E. Optimization using Circulatory System-based Optimization

In this section, CSBO [29] process is utilized to enhance the weight parameter of RBAGCN. CSBO used to optimize RBAGCN weight parameters which effectively Violation Intelligent Recognition multi-source data Fusion identified violations in the power industry. The optimized weight parameters obtained through the CSBO driven optimization process are then applied within the RBAGCN model to improve identified violations in the power industry performance. The CSBO algorithm principle can be divided into five major stages, which are includes in the following steps.

Step 1: Initialization

In the initialization weight parameter is using the model is expressed in the form of Violation Intelligent Recognition which is used to calculate the modulation detection in weight parameters of RBAGCN is calculated in the equation (16).

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,d} \\ d_{2,1} & d_{2,2} & \dots & d_{2,d} \\ : & : & : & : \\ d_{s,1} & d_{s,2} & \dots & d_{s,d} \end{bmatrix} \tag{16}$$

Where D the parameter weight calculation in the Recognition, $d_{s,d}$ is the number of parameter weight calculation in the Recognition.

Step 2: Random generation

After initialization process, the $M(u), D_g(u)$ weight parameter is created at random through RBAGCN method.

Step 3: Fitness Function

The fitness function for optimizing weight parameters in intrusion detection using machine learning typically involves measuring the model Multi-source Data Fusion align with the actual outcomes truth related to detection. The fitness function measures the performance of the model and guides the optimization process in adjusting the weight parameters. The Fitness function is calculated using equation (17).

$$Fitness\,Function = [optimizing\ M(u), D_g(u)] \tag{17}$$

Where $M(u)$ represent to optimize the data, $D_g(u)$ represent to optimize the improved detection.

Step 4: Exploration phase

Exploration phase is used to evaluate the weight parameter in optimization. The RBAGCN weight parameter can be used to improve the calculation. Then, calculates the evaluation of each optimization vector is given in the equation (18).

$$\rho\left(\frac{\delta u}{\delta s} + u - \Delta u\right) = \Delta q - \Delta \times \gamma \tag{18}$$

Where ρ denotes the character, δu is the value of data fusion, δs is the value of violation, Δu is the detection of data fusion, Δq is the quality of detection. Then the maximum value of Intelligent Recognition is calculated in equation (19).

$$\gamma w_a = \gamma w_{\max} \times rand(1,d) * (\gamma w_{\min} - \gamma w_{\max}) \tag{19}$$

Where γw_a is the Intelligent Recognition weight parameter, $\gamma w_{\max}$ is the Intelligent Recognition maximum weight parameter $rand(1,d)$ is the detection of data, $\gamma w_{\min}$ is the Intelligent Recognition minimum weight parameter.

Step 5: Exploitation phase for optimizing $M(u), D_g(u)$

The term exploitation of weight parameter calculation denotes to process of optimizing method's weights to improve its accuracy in detecting Multi-source Data Fusion using machine learning models proposed method is mathematically expressed equation (20).

$$\gamma w_{a,b}^{new} = \gamma w_{1,b} + \rho_a M(u) * (\gamma w_{3,b} - \gamma w_{2,b}) \tag{20}$$

Where $\gamma w_{a,b}^{new}$ represent the Intelligent Recognition new weight parameter , $\gamma w_{1,b}$ represent the Intelligent Recognition first weight parameter, ρ_a is the character value of data, $M(u)$ is the denotes the maximum value, $\gamma w_{3,b}$ represent the Intelligent Recognition third weight parameter $\gamma w_{2,b}$ represent the Intelligent Recognition second weight parameter proposed method is mathematically expressed equation (21).

$$\rho_b = \frac{D(\gamma w_a) - D_\lambda}{D_v - D_\lambda} \tag{21}$$

Where ρ_b the character value of fusion is, $D(\gamma w_a)$ is the detection of Intelligent Recognition weight parameter, D_λ is the detection of sigmoid, D_v is the detection of variation. Figure 2 shows Flowchart of CSBO for optimizing RBAGCN parameter

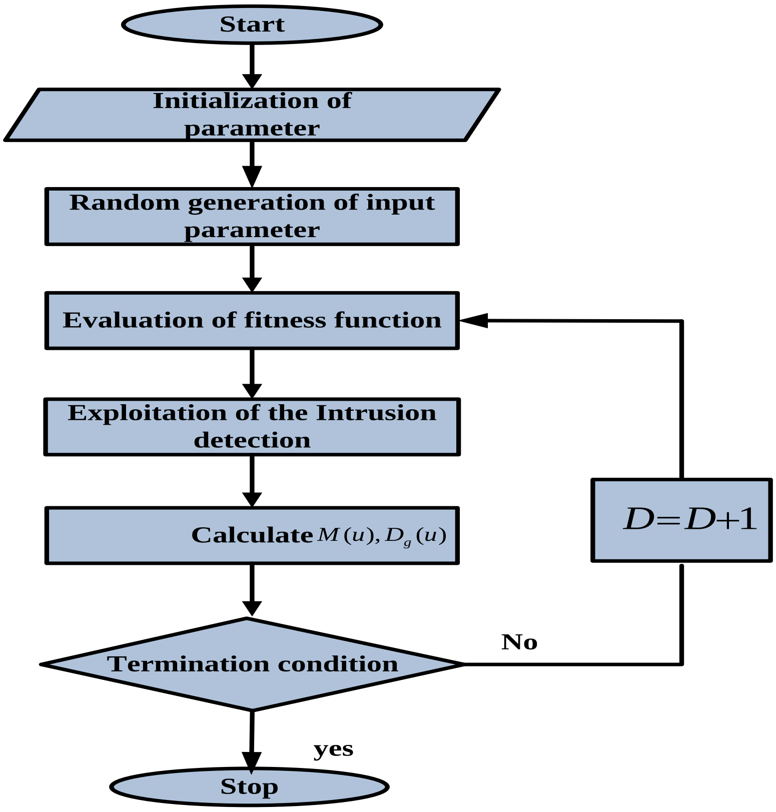

Figure 2: Flowchart of CSBO for optimizing RBAGCN parameter

Step 6: Termination

Finally, the factor $M(u), D_g(u)$ is optimized by CSBO; repeat step 3 till reaches halting criteria $D = D+1$ .RBAGCN is enhanced by CSBO efficiently for detection review with better accuracy. Thus the proposed VIRA-BO-MSDF method effectively classifies the Violation Intelligent Recognition with higher accuracy and low computation time.

IV. RESULT AND DISCUSSION

The simulation results of VIRA-BO-MSDF are discussed. The proposed method is implemented in python using several performance metrics likes accuracy, precision, specificity, sensitivity, F1-score, computation time, ROC and Mean Squared Error are estimated. The proposed method is analyzed with existing technique such as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD

A. Performance metrics

This is evaluated to measure proposed technique efficiency. For that, confusion matrix is essential.

1) Accuracy

The Accuracy is a metric used to evaluate a detection system's performance in correctly identifying and classifying instances within a given dataset, categorized as Violation Intelligent Recognition. Then the calculation of Accuracy is given in equation (22).

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{22}$$

Where, TP denotes true positive, TN signifies true negative, FP denotes false positive, FN signifies false negative.

2) Precision

Precision scales system's ability to detect positive cases correctly out of all predicted positive cases. It is proportion of true positives to total false positives, true positives. It is determined by equation (23),

$$\Pr ecision = \frac{TP}{(TP + FP)} \tag{23}$$

3) Sensitivity

The proportion of correctly forecast true positives among each actual positive instances is calculated as sensitivity. It is assesses ability to detect positive instances correctly which is given in equation (24).

$$sensitivity = \frac{TP}{TP + FN} \tag{24}$$

4) Specificity

Specificity weighs efficiency of method on one class by similar to probability that expression is given in equation (25).

$$Specificity = \frac{(TN)}{(FP + TN)} \tag{25}$$

5) F1-score

F1-score signifies harmonic mean of precision, sensitivity. It is widely utilized as evaluation metric in binary, multi-class classification, combines precision, sensitivity into single metric to gain better understanding of method performance. It computed by equation (26).

$$TPR = \frac{TP}{TP + FN} \tag{26}$$

6) Computation time

The computation time for violation detection can be influenced by numerous factors, comprising complexity of detection, the size of the dataset, computational resources

7) ROC

The ROC curve compares true positive rate to false positive rate at different threshold settings. ROC measures the model's ability to distinguish between attacks, with higher values indicating better performance. This is computed by equation (27).

$$ROC = 0.5 \times \left(\frac{TP}{TP \ + FN} + \frac{TN}{TN + TP} \right) \tag{27}$$

8) Mean Squared Error

The average squared difference among results obtained from an analysis of statistics and the numbers expected from a model is known as the mean squared error is given in equation (28),

$$MSE = \frac{1}{m}\sum_{j=1}^{m}\left(W_j - \hat{W}_j\right)^2 \tag{28}$$

Where, MSE denotes mean squared error, m signifies total data points, W_j represent the observed values, and $\hat{W}_j$ represents the predicted values

B. Performance analysis

Figure 3 to 7 depicts the simulation result of accuracy, precision, specificity, sensitivity, F1-Score, computation time, ROC and Mean Squared Error are analyzed for VIRA-BO-MSDF technique is analyzed with existing technique likes DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD respectively.

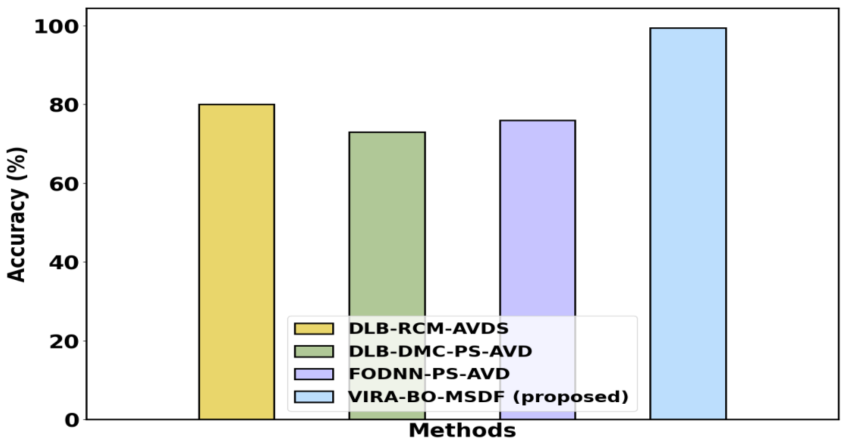

Figure: 3 Accuracy analysis

Figure 3 portrays accuracy analysis. Accuracy, depicted graphically, illustrates the model's efficacy in identifying violations within the power industry. This visualization enables stakeholders to assess performance, pinpoint areas for enhancement, and make informed decisions on deploying and optimizing detection systems. The VIRA-BO-MSDF method provides the 28.46%, 21.34 and 33.81% higher accuracy which is analyzed with existing as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD methods respectively.

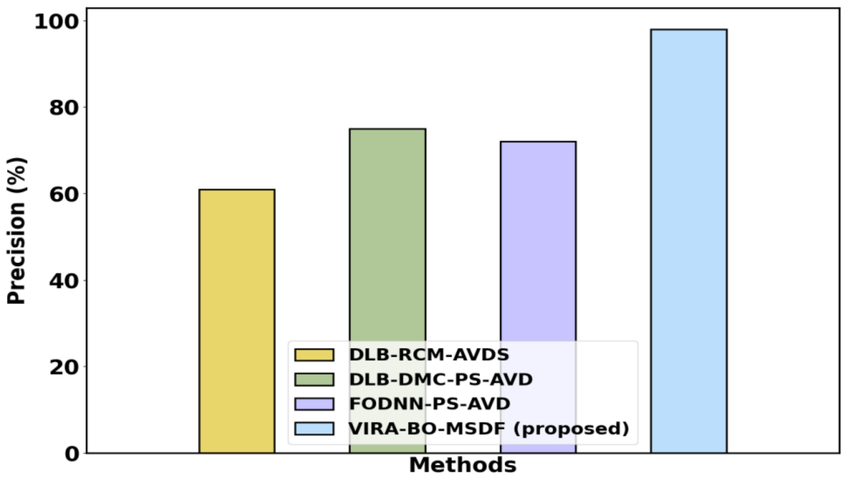

Figure: 4 Precision Analysis

Figure 4 portrays precision analysis. Precision quantifies ratio of accurately identified violations among all instances flagged as violations by the model. A high precision suggests that instances flagged as violations are more likely to be accurate, indicating the model's reliability in identifying true violations. The VIRA-BO-MSDF method provides 22.88%, 26.52% and 34.63% higher Precision which is analyzed with existing method such as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD respectively.

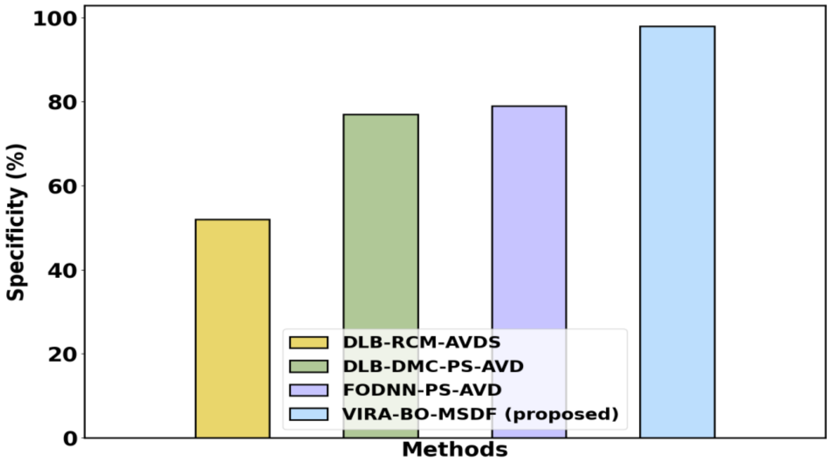

Figure: 5 Specificity Analysis

Figure 5 portrays specificity analysis. Specificity quantifies accuracy of identifying non-violations among instances predicted as such by the model. A high specificity suggests that when the model classifies an instance as not a violation, it is likely accurate, reflecting the model's reliability in discerning non-violative instances. The VIRA-BO-MSDF method provides the 28.46%, 21.34 and 33.81% higher specificity which is analyzed with existing as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD methods respectively.

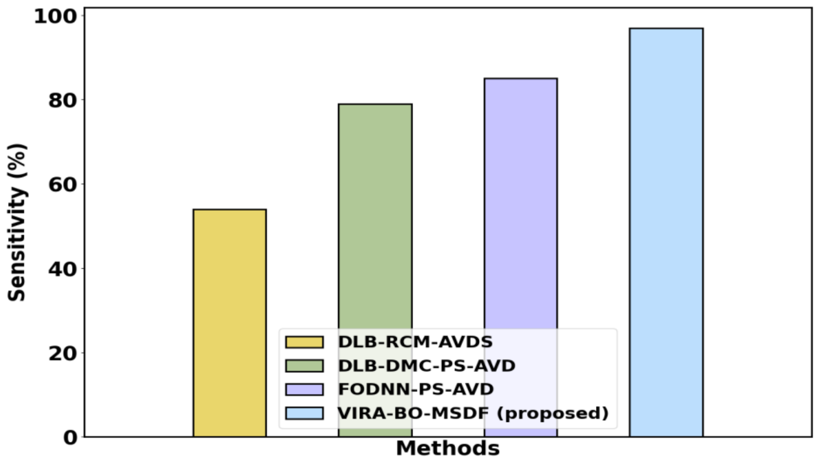

Figure: 6 Sensitivity Analysis

Figure 6 portrays sensitivity analysis. Sensitivity quantifies the accuracy of detecting true violations among all actual violations in the dataset. A high sensitivity signifies the model's effectiveness in identifying genuine violations, reducing the likelihood of missing instances of non-compliance. The VIRA-BO-MSDF method provides achieves an improvement of 22.37%, 27.89%, and 31.37% higher sensitivity which is analyzed with existing method such as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD respectively.

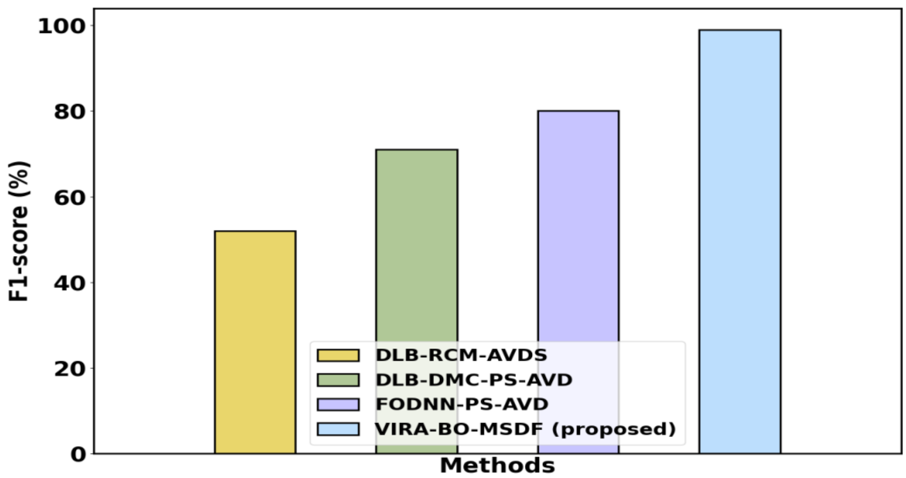

Figure: 7 F1-score Analysis

Figure 7 depicts F1-score analysis. It is derived from harmonic mean of precision, recall, offering balanced evaluation of methods performance. It accounts for the model's capacity to accurately pinpoint violations and its effectiveness in capturing all actual violations. A high F1-score reflects the methods ability to attain both precision, recall simultaneously, ensuring a favorable equilibrium between reducing false positives and false negatives. The VIRA-BO-MSDF method provides achieves an improvement of 19.52%, 25.65%, and 29.82% higher F1-Score are analyzed with existing method such as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD respectively.

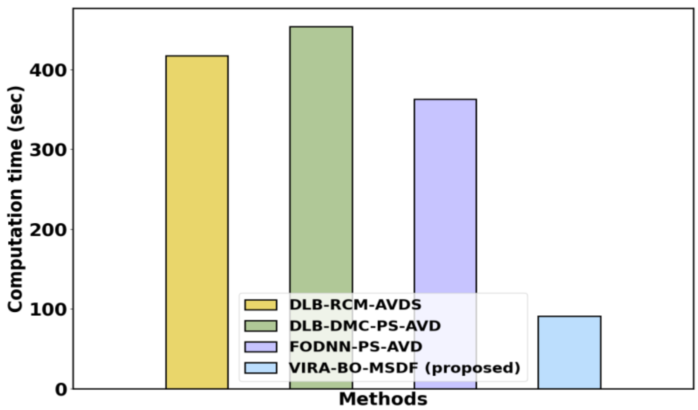

Figure 8: Computation time analysis

Figure 8depicts computation time analysis. The computation time relates to the duration needed for the underlying algorithms to execute tasks critical to the analysis of violation. This encompasses various stages including preprocessing and the execution of additional processing steps essential for accurate identification of violation in power industry. The VIRA-BO-MSDF method provides achieves an improvement of 29.88%, 26.65%, and 19.78% lower computational time are analyzed with existing method such as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD respectively.

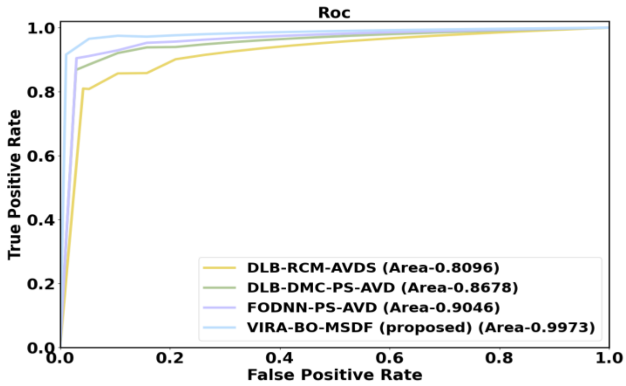

Figure: 9 ROC Analysis

Figure 9 depicts ROC analysis. The ROC curve evaluates a methods ability to differentiate among true positive, false positive rates across various identification thresholds. It illustrates the balance between sensitivity and specificity offering insights into methods performance. In this context, the proposed VIRA-BO-MSDF method achieves improvements of 0.93%, 0.94%, and 0.97% higher ROC when analyzed with existing methods such as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD respectively.

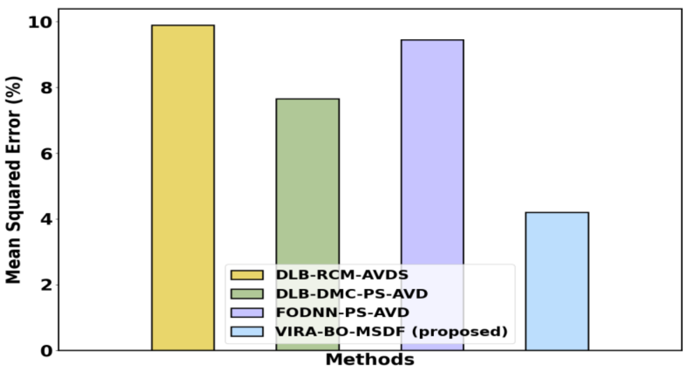

Figure 10: Mean Squared Error analysis

Figure 10 depicts Mean Squared Error analysis. MSE calculates the mean of squared differences among actual, predicted values produced by a model. Applied to identifying violations, MSE can gauge the precision of predictions concerning metrics like power consumption and voltage levels. The VIRA-BO-MSDF method provides achieves an improvement of 32.58%, 26.77%, and 30.98% lower Mean Squared Error are analyzed with existing method such as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD respectively.

C. Discussion

The various RBAGCN models are trained and tested using word and character and numerical value approaches. During the experiment, changed the hyper parameters that shape the RBAGCN Change a single parameter and the others will be fixed. It also successfully developed RBAGCN models that identifying violation in power industry. The Violation Intelligent Recognition Based Multi-source Data Fusion accuracy of VIRA-BO-MSDF is 22.55%, 24.72%, 29.63% greater than existing techniques likes DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD respectively. Similar to this, the precision of proposed method is 96.94% analyzed with sensitivity of comparison techniques of 82.54%. The proposed method VIRA-BO-MSDF has high F1-score and specificity evaluation metrics than existing methods. Therefore, the comparative methods are expensive than the proposed technique. As a result, the proposed technique is identifying violation in power industry effectively and efficiently.

V. CONCLUSION

The proposed, Violation Intelligent Recognition Algorithm Based on Multi-source Data Fusion (VIRA-BO-MSDF) has successfully implemented. An intelligent violation identification algorithm leveraging multi-source data fusion is proposed. By harnessing RBAGCN and CSBO, this algorithm accurately detects potential violations within the power industry. Experimental findings validate its effectiveness and performance superiority, offering robust support for ensuring the safe operation and management of power facilities. Future activities will focus on DTCDWT performance and efficiency while validating and implementing it across various practical application scenarios. The VIRA-BO-MSDF technique is implemented in python. The Performance metrics likes accuracy, precision, specificity, sensitivity,F1-score are analyzed. TheVIRA-BO-MSDF technique provides 19.52%, 25.65%, and 29.82% higher F1-Score, 29.88%, 26.65%, and 19.78% lower computational time, 0.93%, 0.94%, and 0.97% higher ROC, 32.58%, 26.77%, and 30.98% lower Mean Squared Errorare analyzed with existing method such as DLB-RCM-AVDS, DLB-DMC-PS-AVD and FODNN-PS-AVD methods respectively.

REFERENCES

[1] R.W. Liu, Y. Guo, J. Nie, Q. Hu, Z. Xiong, H. Yu, H, and M. Guizani, “Intelligent edge-enabled efficient multi-source data fusion for autonomous surface vehicles in maritime internet of things”, IEEE Transactions on Green Communicationsand Networking, 6(3), 1574-1587, 2022.

[2] P. Zhang, T. Li, G. Wang, D. Wang, P. Lai and F. Zhang, “A multi-source information fusion model for outlier detection”, Information Fusion, 93, 192-208, 2023.

[3] Wu, Y., Wang, Q., Guo, N., Tian, Y., Li, F., & Su, X. (2023). Efficient Multi-Source Self-Attention Data Fusion for FDIA Detection in Smart Grid. Symmetry, 15(5), 1019.

[4] X. Li, and Y. Zhong, “Exploration of a network security situational awareness model based on multisource data fusion”, Neural Computing and Applications, pp.1-13, 2023.

[5] A. Abid, F. Jemili, O. Korbaa, “Real-time data fusion for intrusion detection in industrial control systems based on cloud computing and big data techniques”, Cluster Computing, 1-22, 2023.

[6] Y. Ma, Z. Xie, S. Chen, Y. Wu and F. Qiao, “Real-time driving behavior identification based on multi-source data fusion”, International journal ofenvironmental research and public health, 19(1), 348, 2021.

[7] Y. Chen, Z. Hua, Y. Tang and B. Li, “Multi-Source Information Fusion Based on Negation of Reconstructed Basic Probability Assignment with Padded Gaussian Distribution and Belief Entropy”, Entropy, 24(8), p.1164, 2021.

[8] Q. Yuan, Y. Pi, L. Kou, F. Zhang, Y. Li and Z. Zhang, “ Multi-source data processing and fusion method for power distribution internet of things based on edge intelligence”, Frontiers in Energy Research, 10, 891867, 2022.

[9] R.R Shrivastwa, S. Guilley and J.L. Danger, “Multi-source fault injection detection using machine learning and sensor fusion”, In InternationalConference on Security and Privacy, 2021 (pp. 93-107). Cham: Springer International Publishing.

[10] J. Wei, F. Zhang, J. Lu, X. Yang and Y. Yan, “Fault diagnosis method for machinery based on multi-source conflict information fusion”, MeasurementScience and Technology, 33(11), 115007, 2021.

[11] J. Xing, W. Wu, Q. Cheng and R. Liu, “Traffic state estimation of urban road networks by multi-source data fusion: Review and new insights”, PhysicaA: Statistical Mechanics and its Applications, 595, 127079, 2022.

[12] H. Huang, Z. Liu, X. Han, X. Yang and L. Liu, “A belief logarithmic similarity measure based on dempster-shafer theory and its application in multi-source data fusion”, Journal of Intelligent & Fuzzy Systems, (Preprint), 1-13, 2023

[13] X. Mi, T. , Lv, Y. Tian and B. Kang, “Multi-sensor data fusion based on soft likelihood functions and OWA aggregation and its application in target recognition system”, ISA transactions, 112, 137-149, 2021.

[14] J. Zhou and Y. Lei, “Multi-source Heterogeneous Data Fusion Algorithm Based on Federated Learning”, In International Conference on SoftComputing in Data Science, 2023 (pp. 46-60). Singapore: Springer Nature Singapore.

[15] J. Xiong, H. Shi, G. Guo, X. Liu, C. Xiao and S. Zhu, “ Research on intelligent transportation and shared parking spaces inurban areas based on multi-source data integration”, Computers and Electrical Engineering, 110, 108835, 2023

[16] Y. Shang, C. Ma, K. Yang and D. Tan, “ Regenerative Braking Control Strategy Based on Multi-source Information Fusion under Environment Perception”, International Journal of Automotive Technology, 23(3), 805-815, 2022.

[17] X.U.E. Ke and B. Han, “ Cold Chain Logistics Path Planning and Design Method based on Multi-source Visual Information Fusion Technology”, InternationalJournal of Advanced Computer Science and Applications, 14(10), 2023.

[18] K. Preethi, “AN INTELLIGENT ONLINE DRUNK DRIVING DETECTION SYSTEM BASED ON MULTI-SENSOR FUSION TECHNOLOGY”, International Journal of Management Research and Reviews, 13(2), pp.7-11,2023.

[19] K. Yan, Q. Li, H. Li, H. Wang, Y. Fang, L. Xing, Y. Yang, H. Bai and C. Zhou, “Deep learning-based substation remote construction management and AI automatic violation detection system”, IET Generation, Transmission & Distribution, 16(9), pp.1714-1726, 2022.

[20] D. Fellner, T.I. Strasser and W. Kastner, “Applying deep learning-based concepts for the detection of device misconfigurations in power systems”, Sustainable Energy, Grids and Networks, 32, p.100851, 2022.
[21] X. Pan, M. Chen, T. Zhao and S.H. Low, “ DeepOPF: A feasibility-optimized deep neural network approach for AC optimal power flow problems”, IEEE Systems Journal, 17(1), pp.673-683, 2022.
[22] Y. Xiao, X. Li, J. Yin, W., Liang and Y. Hu, Adaptive multi-source data fusion vessel trajectory prediction model for intelligent maritime traffic. Knowledge-Based Systems, 277, 110799, 2023.
[23] L. Wang, Y. Cao, S. Wang, X. Song, S. Zhang, J. Zhang and J. Niu, “Investigation into recognition algorithm of helmet violation based on YOLOv5-CBAM-DCN”, IEEE Access, 10, pp.60622-60632, 2022.
[24] R. Nellikkath and S. Chatzivasileiadis, “Physics-informed neural networks for ac optimal power flow”, Electric Power Systems Research, 212, p.108412, 2022.
[25] Q. Yuan, Y. Pi, L. Kou, F. Zhang, Y. Li and Z. Zhang, “ Multi-source data processing and fusion method for power distribution internet of things based on edge intelligence”, Frontiers in Energy Research, 10, p.891867, 2022.
[26] Y. Ge, L. Yang and X. Ma, “A harmonic compensation method for SPMSM sensorless control based on the orthogonal master-slave adaptive notch filter”, IEEE Transactions on Power Electronics, 36(10), 11701-11711, 2021.
[27] H.L. Minh, S. Khatir, R.V. Rao, M. Abdel Wahab and T. Cuong-Le, “A variable velocity strategy particle swarm optimization algorithm (VVS-PSO) for damage assessment in structures”, Engineering with Computers, 39(2), 1055-1084, 2023.
[28] L. Yuan, G. Huang, F. Li, X. Yuan, C.M. Pun and G. Zhong, “RBA-GCN: Relational Bilevel Aggregation Graph Convolutional Network for Emotion Recognition”, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023.
[29] M. Ghasemi, M.A. Akbari, C. Jun, S.M. Bateni, M. Zare, A. Zahedi, H.T. Pai and S.S. Band, M. Moslehpour and K.W. Chau, “Circulatory System Based Optimization (CSBO): An expert multilevel biologically inspired meta-heuristic algorithm”, Engineering Applications of Computational Fluid Mechanics, 16(1), 1483-1525, 2022.