**Four-Stage Data Science Framework Embedded With DSBNN for Line Trip Fault Prediction in Power Systems**

**1Dr. K. Kalyan Raj,2Dr.B.MadhuKiran, 3Dr.B.Dasu,4Dr.R.Srinivasa Rao**

*1Dr. K. Kalyan Raj, Assistant Professor, Dept of EEE, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru-521356,* *kalyan20.kaniganti@gmail.com*

*2Dr.B.MadhuKiran, Associate Professor, Dept of EEE, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru-521356, A.P*

*3Dr.B.Dasu,Professor, Dept of EEE, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru-521356, A.P*

*4Dr.R.Srinivasa Rao, Professor, Dept of EEE, JNTUK, Kakinada-533003, A. P*

**ABSTRACT**

Data science is a "concept to unify [statistics](https://en.wikipedia.org/wiki/Statistics), [data analysis](https://en.wikipedia.org/wiki/Data_analysis), [informatics](https://en.wikipedia.org/wiki/Informatics), and their related methods" to "understand and analyze actual phenomena" with data. Here we embedded four-stage data science framework for the detection of line trip fault prediction in power systems. In this work, we proposed the novelty of Embedding the DSBNN (Deep Sigmoid based Neural Network) with Data Science for predicting line trip fault in power systems. It involves four stages of the process it undergoes which include pre-processing, Variable Selection, Metrology Prediction, Process Control. Initially, the dataset features are pre-processed and it attains the variable selection stage in the manner of Random support-based AGWO. After those metrology predictions stage, the selected variables are applied to DSBNN is used. Finally, the stage of process control entropy is calculated and the threshold limit is set to predict whether the fault occurs or not. The experimental results revelations the dominance of presented method comparing with the existing methods.

**Keywords:** pre-processing, Adaptive Grey Wolf Optimization (AGWO),DSBNN (Deep Sigmoid based Neural Network), Data science, Power Systems.

**1. INTRODUCTION**

To simply stated, data science is the study of information. It is made up of various tools, methods, algorithms, and procedures. It also includes data storage, management, and analysis in order to extract meaningful information from both structured and unstructured data. It also employs machine learning methods to elicit insights from data. Nowadays, most businesses have an unusually large amount of data. It contributes to the effective use of this massive amount of data.Machine learning (ML), a basic aspect of "Artificial Intelligence" (AI), can play an important role in extracting insights from data, and data science (DS) is driving the transition.Data science is at the forefront of a new scientific paradigm [1], and machine learning has the potential to radically alter the cyber security landscape. The first goal is to ensure the facility system's dependability and stability as power grids grow and demand increases. In power systems, high-accuracy failure prediction enhances operational dependability and stability, potentially saving large losses from power outages. Fault prediction is the process of analyzing and mining prior data in order to predict whether or not there is a fault in the grid, allowing for the implementation of actions to prevent accidents and ensure system recovery [2].

In today's data-driven world, information has become an invaluable asset. Organizations, both big and small, are constantly seeking ways to leverage this vast amount of data to gain insights and make informed decisions. This is where the field of data science comes into play. Data science is a multidisciplinary field that combines techniques from Mathematics, Statistics, Computer Science, and domain knowledge to extract valuable insights and knowledge from data. Fault prediciton has been significantly transformed by the advent of data science. With the abundance of l data available today, data science techniques have emerged as powerful tools for faultl diagnosis. By using advanced algorithms, machine learning, and predictive modelling, data scientists are revolutionizing fault predction by improving accuracy, efficiency, and patient outcomes. In this process, feature selection plays a crucial role in medical applications of data science by identifying the most informative and relevant features from fault datasets. It enables the development of accurate and interpretable models for fault diagnosis, and contingency management planning.

Grid operators (PSOs) face a difficult task in providing end-users with uninterrupted power. Although human control over fault infiltration is impossible, it is critical to reliably identify, categories, and localize the fault location. The methods for detecting, classifying, and placing faults in power transmission systems have been thoroughly explored [3, 4].CPPS is more intelligent and stable than a traditional physical electrical power grid. However, because of cyberspace's weaknesses, particularly the complicated interacting process between electrical power and data flow, information security has become a critical issue affecting the facility system's safe and stable operation. As a result, the facility system is constantly vulnerable to network attacks. Unlike network attacks on the Internet, cyber-attacks against CPPS are more focused on disrupting the cyber layer's stability control over the physical layer, even paralyzing the facility system's functioning [5].

The concept of operating reliability was established in order to develop reliability models and assess failure rates and outage consequences in the data processing scale while accounting for various characteristics. Operating reliability theory, rather than classical planning reliability theory, is commonly utilised in real-time power system dispatching and can give a timely assessment to avert potential power grid outages or blackouts [6]. During the recent decade, electrical power systems have moved away from traditional energy systems and toward next-generation smart grid technology [7].

Researchers have discovered that small changes in operational data can help find leakage defects. Several leakage fault detection (LFD) approaches, such as the pressure gradient method, the negative pressure wave method, the inverse transient analysis method, and others, are presented to support this theory. The pressure gradient method works on the idea that when there is a leak, the leakage's upstream flow increases, generating an excessive pressure decrease. The downstream flow is unaffected, resulting in a steady pressure gradient. By comparing simulated pressures and flow rates to measured values at upstream, downstream, and various places along the pipe, the leak will be identified [8]. Many automated fault detection and diagnosis (FDD) approaches have been developed to detect these issues before they have an impact on the long-term functionality of the system [9, 10, 11].Determining the location of the fault is critical for both fault clearance and power restoration transmission. Single line to ground faults, line to line faults, triple line faults, transform faults, and multi-location faults are among the fault types that must be identified initially [12]. Feature engineering is another crucial aspect of prediction analysis. Data scientists identify and transform relevant variables or features that contribute to the predictive accuracy of the models. This involves techniques such as data normalization, dimensionality reduction, and feature selection. By extracting meaningful insights from raw data, data scientists can enhance the performance and interpretability of the predictive models[10].

Validation and evaluation of predictive models are essential to ensure their accuracy and reliability. Data scientists split the available data into training and testing sets, enabling them to assess the model's performance on unseen data. Various evaluation metrics, such as accuracy, precision, recall, and F1 score are employed to measure the model's effectiveness in making predictions. Iterative refinement and optimization of models are performed to enhance their performance over time.

Many studies based on SCADA data have suggested that fault diagnosis and prognosis can be appreciated. To appreciate defect detection, machine learning models with domain knowledge have been presented. The detection of WT gearbox problems was suggested using a deep neural network (DNN)-based architecture [13, 14]. Many data-mining algorithms for alternative energy prediction were reviewed, and some of them might be used to detect faults [15]. In addition to the current situation, the majority of strategies fail to take into account all of the given factors. The unique method is suggested in light of these restrictions. The following are the primary contributions of this paper:

• The variable selection stage combines random support and AGWO (Adaptive grey wolf optimizer) to reduce redundant and unnecessary elements and choose the important variables to increase prediction accuracy and avoid the dimensionality curse.

• Using Data Science and the DSBNN (Deep Sigmoid based Neural Network) to predict power system line trip faults.

The manuscript's structure is organised in the following way: The second section examines the existing literature on the proposed technique. The suggested system is briefly discussed in section 3, the exploratory results are examined in section 4, and the paper is concluded in section 5.

**2. LITERATURE REVIEW**

N. Murugesan et.al. [16] Machine learning approaches are provided for fault detection and supported PV arrays parameters such as, current, radiation and temperature to ensure its safety. Current approaches, on the other hand, typically use supervised learning methods based on different labeled data (so-called error types) and thus have several disadvantages, such as expensive implementation, model simulation, and short time. To improve visualization, this paper proposes an N-semi-regular fuzzy semi-supervised learning system supported by an N-semigraph and some labeled, normalized training data. To support network recovery, the proposed model describes not only the loss, but also the likely failure structure. After building a model, solar systems can learn to independently monitor and detect PV disturbances as the environment changes over time.

Lee Chia-Yen et al. [17] TFT-LCD panel processes are based on experimental design and technical experience in process monitoring and internal control throughout the production line. This study proposes a three-step data science framework embedded with multiple data processing and machine learning techniques to identify performance influencing variables, predict photo-spacing process metrology results, and propose method control in the color filter manufacturing process. To confirm the proposed paradigm, an empirical study was conducted with a leading TFT-LCD manufacturer in Taiwan.

Maren David Dangut et.al [18] propose a hybrid machine learning approach that combines linguistic communication processing approaches with ensemble learning to predict ultra-rare aerospace component failures. A log-based dataset from a real aviation central maintenance system is used to validate the proposed approach. In this data set, unplanned component replacement is extremely rare. The results show that the proposed method outperforms existing unbalanced and ensemble learning methods in terms of precision, recall and f1 score. It also found that focusing exclusively on minority class models could help address class imbalances.

Cheng He et al. [19] A new roller bearing fault diagnosis method using a combined extreme point symmetric shape decomposition (ESMD), multi-step weighted permutation entropy (CMWPE) and gravitational search algorithm with multiple adaptive constraint strategy (MACGSA) optimized statistical process support vector was proposed. machine (LSSVM). ESMD and CMWPE can obtain a more sensitive set of high-dimensional feature vectors to deal with the problem of modal nuance of the internal state function (IMF) and small changes in error characteristics. Finally by tuning the LSSVM, MACGSA was able to improve the fault detection accuracy. The ESMD model is used to generate a series of IMFs based on rolling data; The IMF CMWPE values ​​are then extracted to form a set of high-dimensional feature vectors; and finally the MACGSA-LSSVM model is used to classify the faults. This method has a higher diagnostic accuracy than previous methods.

Pengfei Liang et.al [20] semi-supervised and high-precision adversarial learning system is designed for one-time and simultaneous transmission error detection using generative adversarial networks and time-frequency imaging. The proposed strategy consists of two parts. In the first section, a continuous wavelet transform is used to convert one-dimensional raw vibration data into two-dimensional time-frequency images. Labeled and unlabeled time-frequency images are fed into the built competitive learning model in the second step to diagnose a single and simultaneous gearbox failure. Finally, the technique is tested with two case studies. The results show that it achieves the best accuracy with higher accuracy and fewer training steps than other intelligent fault diagnosis methods currently available in the literature.

**3. PROPOSED SYSTEM**

This paper proposes a four-stage data science framework incorporating machine learning approaches for predicting data-based line trip faults in power systems. Pre-processing, variable selection, metrology prediction, and process management are all part of a four-stage data science framework. Remove factors having a large number of non-available (N/A) or identical values during the pre-processing stage. Also, see if the missing values are the result of parallel computers.The variable selection stage, which employs random support and the AGWO (Adaptive grey wolf optimizer), eliminates redundant and unnecessary variables while focusing on the important ones to improve prediction accuracy and avoid the dimensionality curse. Feature selection is a crucial step in data science that involves selecting a subset of relevant features or variables from a larger set of available features. The goal is to reduce dimensionality, improve model performance, enhance interpretability, and eliminate redundant or irrelevant information. In fault diagnosis, feature selection methods are essential for identifying the most relevant and informative features from datasets. The machine learning technique is used in the metrology prediction step to collect the chosen variables. Current, voltage, active power, and user reactive power are all electrical measures taken during the DSBNN (Deep Sigmoid based Neural Network) approach. The technique control stage tweaks the process for Entropy-based line trip fault prediction; the most important duty is determining whether or not a problem exists when the power system is operating. Figure 1 depicts a diagram of the methodology being described.Bottom of Form



**Figure1:** Overall Proposed Trip Fault Prediction in Power Systems Methodology

**3.1 Pre-Processing**

In a stage-1 data science framework the Pre-processing is the important step to attain the dataset effectively. Let us consider is the Testbeddataset, where *n* is the number of records in the Testbeddataset. In the Set of features represented by. In Some places where the value is empty that case removes the factors with many non-available (N/A) and alsoeliminates the identical values.

**3.2 Random Support Based Adaptive Grey Wolf Optimization (AGWO)**

The stage-2 data science framework sorts the dataset input data based on the support value of features in the data. In this circumstance, the Random Support value evaluation is based on certain features.

****(1)

Where denotes the support value, denotes the chosen random optimal features set. We proceed to the Adaptive GWO, which is a swarm intelligence technique in a Four-stage data science framework it mimics wolves' control structure, which is known for its collective hunting. Grey wolves mostly value boarding a group more. They need a rigorous, progressive structure that is socially dominant wolf , with an Alpha (α) who can be either a male wolf or a female wolf. The alpha often takes on the majority of the duties associated with leadership. Most wolves expect the pack to comply with their orders. The alpha's subordinate wolves, known as the Betas (β), help him in his basic leadership responsibilities. The pack's enforcer and advisor to the alpha is the beta. Delta (δ) wolves are in charge of the omega wolves and answer to the alpha and beta wolves. Alpha (α) is the mathematical word for the best-fitting solution and modelled as Random Support-based AGWO.The second and third best solutions are given by  Beta and Delta group respectively. so the β,δ wolfs in coordination will lead the hunt along with the alpha.In this the hunting pattern od the wolfs is mathematically modeled to develop a algorithm to handle the optimization problem.The pseudo-code of Random Support-based Adaptive GWO method is given in algorithm1.



Bottom of Form

**Algorithm 1:** Pseudo Code of the Random Support AGWO

**Step 1:**Search agents are used to initialise the GWO parameter (), vectors*,* and the maximumgreat number of iteration ().

** (2)

**(3)

The assessmentslinearly decrease from 2 to 0 over the time of iterations and*,*  are random vectors in [0, 1]. The parameter is linearly updated in each cycle to go from [2-0] as indicated by the condition (4),

 (4)

Where represents the number of iterations and is the total number of iterations taken into consideration for the optimization of fault prediction problem.

**Step 2:** Depending on the size of the pack, we randomly create wolves.

**Step 3:** Considering the Random Support value we evaluate each hunt agent's fitness esteem (5)

**(5)

Where is ascharacterized in (6) and🡪iteration amount,🡪coefficient matrix,🡪targetlocation, and 🡪 grey wolf location.

 (6)

**Step 4:** The best hunt mediator (), second-dominant hunt negotiator () and also the third hunt mediator () by utilizing the circumstance are determined using (7),

, and  (7)

Where, and  (8)

**Step 5:** current hunt agent's location is updated using the following condition (9)

**** (9)

**Step 6:**The fitness value is determined for all the hunts.

**Step 7:**Update the values of *,* and*.*

**Step 8:** check for a termination condition until the maximum number of iterations are reached ; otherwise, proceed to step 5.

The AGWO optimization algorithm gives the most influential set of features as output. Here, features are randomly chosen for the dimensionality lessening of the dataset by utilizing aRandom Support-based AGWO algorithm that can reduce the computational burden as well as increase the accuracy in classifying the features thar are actually necessary for perfect interpretation of the data.

**3.3 Deep Sigmoid Based Neural Network(DSBNN)**

A mathematical function with a distinctive "S"-shaped curve, also called a sigmoid curve, is called a sigmoid function. The following is a representation of the sigmoid function: (10)

The input is (sig), and the output is (f). he neural network, which is the result of a sigmoid function, must continuously learn to solve problems with more expertise or to use a range of techniques to get better results. When it receives fresh knowledge from the system, it learns how to react to a replacement scenario.A deep neural network is a sort of machine learning system that uses numerous layers of nodes to extract high-level functions from input data. It involves converting data into a more imaginative and abstract form.A connection layer, pooling, Sigmoid-based normalization, and convolution are just a few of the layers of the proposed DSBNN that address the issues with CNN. Figure 2 illustrates the construction of the Deep Sigmoid-based Neural Network (DSBNN).

Bottom of Form



**Figure 2:** Architecture of the proposed DSBNN

|  |
| --- |
| **Input:** Features Selected from the Random Support AGWO**Output:** fault occurs or not a prediction |
| 1. **Begin:**
2. Set all weights and biases using (11), (12).
3. **For all** input image If**do**
4. //Convolutional layer
5. **For**to *n* do
6. **For** layers =1 to **do**
7.
8. **End for**
9. **End for**
10. //sigmoid-based normalization layer equation (14),
11. **For**to*n***do**
12. **For** layers =1 to **do**
13. **//** sigmoid

**End for****End for** //Upgrade weights 1. **For**to 1 **do**
2. **For** to **do**
3. **If**!= max-pooling layer **then**
4. Upgrade weights and biases
5. **End if**
6. Upgrade the thresholds
7. **End for**
8. **End for**
9. **End**
 |

**Algorithm 2:**Pseudo code of the proposed DSBNN

The weights and biases of the prior layers in the structure design affect the DSBNN classifier's final result. The model is then improved with conditions (11) and (12) for each layer independently..

**** (11)

 (12)

Where represents the weight, represents the bias,**gives the layer number, **represents the regularization parameter, ** means defines the learning rate, is the total number of training sets,**represent the value of momentum, **is the upgrading phase and finally represents the cost function.

The proposed DSBNN classifier have different layers and are explained below,

**Step 1: Convolutional layer**: This layer finishes the convolution of the input data with the kernel by using a condition (13).

(13)

Wheredefines the reproduced segmented images,  represents the filter, and represents the number of components in& the output vector is.

**Step 2: Sigmoid-based normalization layer**

Linearly transforming data to fit within a specific range is called normalization. Data normalization uses the Z-score normalization method, which linearly transforms the data. Equation (14) describes how Z-scores are normalized**:**

(14)

Here, is the normalized output, f is the value of the sigmoid function, which is the norm of the output image of the convolutional layer and is the variance of the values ​​of the output image of the convolutional layer. The output image of the convolution layer is normalized by a sigmoid function using equation (14). This layer results in a sigmoid-based normalized image and is given as input to the collection layer. This layer creates a normalized image based on the reference value and is expected to contribute to the aggregation layer.

**Step 3: Pooling layer:**Down-sampling is another name for this layer. The pooling method decreases the size of output neurons from the convolution layer to reduce computational intensity and avoid overfitting. The max-pooling algorithm selects only the highest value in each feature map, resulting in fewer output neurons. Pooling layers are often used after convolution layers to simplify the data within the output of the convolution layer.

**Step 4: A completely connected layer:**The activation function computes the probability distribution of the classes. So the result layer uses the softmax function to find the result of the previous layer that matches the previous normal or malignant or benign.

 (15)

Where, represents the resulting image. Here, DSBNN is equipped with sigmoid function-based normalization to control layer atrophy, and the important measurement classifications are current, voltage, active power and reactive power of users during the process.

**3.4 Entropy Calculation**

Entropy is used to recognise the texture of measurements. The condition (16) calculates the entropy of the variables as follows:

   (16)

The main goal of Line Trip Fault Prediction Using Entropy is to determine whether or not there is a fault during the operation of a power system by forecasting whether or not the threshold value will reach a high level, indicating that a fault has occurred, otherwise it will not.

**4. RESULTS**

Experimental results show that proposed strategy is more effective than traditional methods. The conclusions of the application are based on Matlab2018a software running on Intel CoreTM i7 processors running at 1.6 GHz. Extensive experimental studies were conducted to determine the optimal subsets of features selected by AGWO that provide the highest possible accuracy using different classifiers.

**4.1 Dataset Description**

The Advanced Diagnostics and Prognostics Testbed (ADAPT) lab at the NASA Ames center expects to convey some way to judge the adequacy of symptomatic calculations at distinguishing deficiencies in power frameworks. The calculations are assessed utilizing information from the Power System (EPS),which recreates components of a standard aviation vehicle power framework. The EPS takes into consideration the controlled inclusion of flaws in repeatable disappointment situations to check if analytic calculations can recognize and detach these shortcomings. This dataset was created from the EPS within the ADAPT lab. Every information document relates to at least one test run of the work. During a test, an information procurement framework orders the work into various arrangements and records information from sensors that action framework factors like voltages, flows, temperatures and switch positions. Flaws were infused in an exceedingly portion of the exploratory runs.

Bottom of Form

.

**4.2 Performance Analysis**

Accuracy, Sensitivity, Specificity, Precision, Recall, and F-measure are statistical metrics that can be stated in TP, FP, FN, and TN esteem.

**4.2.1** Accuracy is one of the statistical measures used to evaluate the performance of our proposed work. Accuracy is measured by dividing the number of correct predictions by the total number of observations in the dataset. It measures the classification accuracy of power disturbance data. The condition is used to check the accuracy of (17), (17)

Where TN represents the true negative, TP gives the true positive, FP gives the false positive, and FN represents the false negative.

**4.2.2 Sensitivity**

The number of true positives that a classification test can effectively distinguish is called sensitivity. This shows how effective the test is in classifying the data. The conditional sensitivity (18) is calculated.

 (18)

**4.2.3 Specificity**

A classification test's specificity is defined as the number of true negatives accurately detected. It expresses how well the test distinguishes between normal and abnormal data. Utilizing the condition, specificity is calculated (19).

 (19)

**4.2.4 Precision**

Precision is the proportion of the predicted positive instances that were correct text sizedata given by equation (20).

 (20)

**4.2.5 Recall**

Recall is the relative relationship between the conventional information identified and the general information available in the dataset with condition (21).

 (21)

**4.2.6 F-measure**

This is the measure used to determine the accuracy of the test. The best value of F-measure is 1 and the worst value is 0. Equation (22) is used to find it**.**

 (22)

WOSIG (without sigmoid) and existing techniques ANN (Artificial neural network), KNN (K-Nearest Neighbour), and NB (Nave Bayes) are compared in Figure 3 and Table 1. The proposed technique's accuracy is 98 percent, which is significantly higher than the previous techniques' 70, 77, and 61 percent accuracy. In addition, when compared to existing approaches, the other metrics of Sensitivity, Specificity, Precision, Recall, and F-measure have improved.

|  |  |
| --- | --- |
| **C:\mathumadev\praveen doc\Fw__pravin_results\fig1.jpg** | **C:\mathumadev\praveen doc\Fw__pravin_results\fig2.jpg** |

**Figure 3:**Performance measure for proposed and existing techniques.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Techniques**  | **Accuracy** | **Sensitivity** | **Specificity** | **Precision** | **Recall** | **F-measure** |
| Proposed | 0.989362 | 0.974533 | 0.998907 | 0.998261 | 0.974533 | 0.986254 |
| WOSIG | 0.978059 | 0.967742 | 0.984699 | 0.976027 | 0.967742 | 0.971867 |
| ANN | 0.703457 | 0.759036 | 0.69243 | 0.328696 | 0.759036 | 0.458738 |
| KNN | 0.771277 | 0.939163 | 0.735697 | 0.429565 | 0.939163 | 0.589499 |
| NB | 0.611702 | 0.390244 | 0.617908 | 0.027826 | 0.390244 | 0.051948 |

**Table 1:**Performance measure values for proposed and existing techniques.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Techniques** | **Human error** | **Equipment** | **Fire** | **Weather** | **Unknown** |
| Proposed | 276 | 107 | 209 | 302 | 21 |
| WOSIG | 265 | 102 | 198 | 294 | 18 |
| ANN | 210 | 82 | 162 | 243 | 12 |
| KNN | 238 | 90 | 179 | 255 | 16 |
| NB | 186 | 71 | 143 | 219 | 9 |

**Table 2:**Count of disturbances due to different causes values for existing and proposed techniques.

|  |  |  |
| --- | --- | --- |
| **Techniques** | **MW Loss** | **No of Customers Affected** |
| Proposed | 74.5872 | 929 |
| WOSIG | 69.2678 | 915 |
| ANN | 58.6574 | 609 |
| KNN | 61.1549 | 763 |
| NB | 55.3658 | 546 |

**Table 3:**Count of disturbances due to MW lost and affecting a

number of customers.

****

**Figure 4:**The highest classification accuracy achieved by the proposed classifier.

Figure 4 shows how the proposed and current techniques WOSIG without sigmoid, ANN (Artificial neural network), KNN (K-Nearest Neighbor), and NB (Nave Bayes) perform in terms of classification accuracy. The most important features during a power outage, such as MW loss and the number of customers impacted, are employed by five classifiers to forecast the best accuracy. Following that, the Cause appears as an important informative feature, which is employed by two classifiers to achieve the best accuracy. The “MW lost” is the amount of active power lost in MW units when a particular type of disruption occurs. “No. of Affected Customers” is the second important When a specific type of disturbance happens, the power fault has an undesirable effect on the feature.

|  |  |
| --- | --- |
| **C:\mathumadev\praveen doc\Fw__pravin_results\fig4.jpg** | **C:\mathumadev\praveen doc\Fw__pravin_results\fig5.jpg** |
| **(a)** | **(b)** |

**Figure 5:**Count of disturbances due to different causes and MW lost and affecting a

number of customers.

The examination of the types of disturbances that occur in count of cause is shown in Table 2 and Figure 5(a). Human error, equipment failure, fire, bad weather, and the unknown are all examples of disturbances. It is explained that the main sources of disturbances are Weather and Equipment Failure, with Weather and Equipment Failure causing the greatest amount of disruptions. Table 3 and Figure 6(b) show the number of disturbances that cause MW loss and affect a large number of consumers for each type of disturbance. Customer contacts are found to be the source of the greatest amount of disruptions, resulting in MW loss and affecting a huge number of customers.



**Figure 6:** comparison of convergence graph proposed optimization with existing optimization

Figure 6 shows that the proposed calculation's components impact is better to the next existing PSO and ABC algorithms. In cycle 10, the introduced strategy achieves the global ideal incentive. However, the current methods for dealing with their ideal worth are more advanced than cycle 30.So in our work, we choose the proposed optimization.

**5.CONCLUSION**

In this paper, we precede the four-stage data science framework is embedded with DSBNN for the detection of line trip fault prediction in power systems. Prediction of fault in power system includes data science undergoes the process of Four stages pre-processing, Variable Selection, Metrology Prediction, Process Control. Implementation of proposed optimization methods like AGWO adds more computational efficiency in the feature selection methods for improving their efficiency. However these methods increase the computational burden. The feature selection methods making them more domain specific, powerful by multiple data interaction and real time feature selection. At every stage, the process controls and detects the fault occurs or not by proceeding with enthalpy calculation in the Final stage. The proposed technique results attain better accuracy 98% when comparing with the existing techniques. The novel DSBNN proposed technique reaches far better accuracy and all other performance measures too when comparing with normal DNN.

**Declarations**

**Funding                      :**No funding has been provided.

**Conflict of interest    :**There is no conflict of interest for authors.

**Data availability        :**Not applicable.

**REFERENCES**

1. Sarker, Iqbal H., A. S. M. Kayes, ShahriarBadsha, HamedAlqahtani, Paul Watters, and Alex Ng. "Cybersecurity data science: an overview from a machine learning perspective." *Journal of Big Data* 7, no. 1 (2020): 1-29.
2. Zhang, Senlin, Yixing Wang, Meiqin Liu, and ZhejingBao. "Data-based line trip fault prediction in power systems using LSTM networks and SVM." *Ieee Access* 6 (2017): 7675-7686.
3. Raza, Ali, AbdeldjabarBenrabah, ThamerAlquthami, and Muhammad Akmal. "A review of fault diagnosing methods in power transmission systems." *Applied Sciences* 10, no. 4 (2020): 1312.
4. Abdelgayed, Tamer S., Walid G. Morsi, and Tarlochan S. Sidhu. "Fault detection and classification based on co-training of semisupervised machine learning." *IEEE Transactions on Industrial Electronics* 65, no. 2 (2017): 1595-1605.
5. Cao, J., Wang, D., Qu, Z., Cui, M., Xu, P., Xue, K. and Hu, K., 2020. A novel false data injection attack detection model of the cyber-physical power system. *IEEE Access*, *8*, pp.95109-95125.
6. Zhang, Guanglun, HaiwangZhong, Qiang Wang, and Xin Cao. "A Machine-Learning Based Method to Analyze the Correlation between Meteorological Data and Component Outages of Power System." In *2020 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, pp. 352-357. IEEE, 2020.
7. Ibrahim, Muhammad Sohail, Wei Dong, and Qiang Yang. "Machine learning driven smart electric power systems: Current trends and new perspectives." *Applied Energy* 272 (2020): 115237.
8. Xue, Puning, Yi Jiang, Zhigang Zhou, Xin Chen, Xiumu Fang, and Jing Liu. "Machine learning-based leakage fault detection for district heating networks." *Energy and Buildings* 223 (2020): 110161.
9. Bode, Gerrit, Simon Thul, Marc Baranski, and Dirk Müller. "Real-world application of machine-learning-based fault detection trained with experimental data." *Energy* 198 (2020): 117323.
10. Fahim, Shahriar Rahman, YeahiaSarker, Subrata K. Sarker, MdRafiqul Islam Sheikh, and Sajal K. Das. "Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification." *Electric Power Systems Research* 187 (2020): 106437.
11. Gharsellaoui, Sondes, MajdiMansouri, Mohamed Trabelsi, Shady S. Refaat, and HassaniMessaoud. "Fault diagnosis of heating systems using multivariate feature extraction based machine learning classifiers." *Journal of Building Engineering* 30 (2020): 101221.
12. Rao, TC Srinivasa, SS Tulasi Ram, and J. B. V. Subrahmanyam. "Neural network with adaptive evolutionary learning and cascaded support vector machine for fault localization and diagnosis in power distribution system." *Evolutionary Intelligence* (2020): 1-12.
13. Dunderdale, Christopher, Warren Brettenny, Chantelle Clohessy, and E. Ernest van Dyk. "Photovoltaic defect classification through thermal infrared imaging using a machine learning approach." *Progress in Photovoltaics: Research and Applications* 28, no. 3 (2020): 177-188.
14. Lara-Benítez, Pedro, Manuel Carranza-García, Jorge García-Gutiérrez, and José C. Riquelme. "Asynchronous dual-pipeline deep learning framework for online data stream classification." *Integrated Computer-Aided Engineering* Preprint (2020): 1-19.
15. Qu, Fuming, Jinhai Liu, Hongfei Zhu, and Bowen Zhou. "Wind turbine fault detection based on expanded linguistic terms and rules using non-singleton fuzzy logic." *Applied Energy* 262 (2020): 114469.
16. Murugesan, Dr N., R. Anitha, and M. Ganesan. "N-Semi Regular Graph-Based Fuzzy Semi-supervised Learning Approach for Fault Detection and Classification in Photovoltaic Arrays." *International Journal of Advanced Research in Engineering and Technology* 11, no. 9 (2020).
17. Lee, Chia-Yen, and Tsung-Lun Tsai. "Data science framework for variable selection, metrology prediction, and process control in TFT-LCD manufacturing." *Robotics and Computer-Integrated Manufacturing* 55 (2019): 76-87.
18. Dangut, Maren David, ZakwanSkaf, and Ian K. Jennions. "An integrated machine learning model for aircraft components rare failure prognostics with log-based dataset." *ISA transactions* (2020).
19. He, Cheng, Tao Wu, Changchun Liu, and Tong Chen. "A novel method of composite multiscale weighted permutation entropy and machine learning for fault complex system fault diagnosis." *Measurement* 158 (2020): 107748.
20. Liang, Pengfei, Chao Deng, Jun Wu, Zhixin Yang, Jinxuan Zhu, and Zihan Zhang. "Single and simultaneous fault diagnosis of gearbox via a semi-supervised and high-accuracy adversarial learning framework." *Knowledge-Based Systems* 198 (2020): 105895.