2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS), 8-9 March, Dhaka, Bangladesh

Electrical Power Quality Disturbances Detection in Transmission Lines Using Machine Learning-Enabled Classifier

Arafat Ibne Ikram
Department of Electrical and Electronic Engineering
International Islamic University Chittagong
Chattogram - 4318, Bangladesh
arafatibne.ikram@gmail.com

Md. Shakib Hassan
Institute of Energy Technology
Chittagong University of Engineering & Technology
Chattogram - 4349, Bangladesh
asiksakibhasan4066@gmail.com

Nurjahan Akter
Department of Computer Science & Engineering
Sonargaon University
Dhaka - 1215, Bangladesh
nurjahanalt9456@gmail.com

***Abstract*—There are increasing concerns about power quality disturbances (PQDs) at many phases of energy generation, transformation, distribution, and consumption due to the increasing interconnection of various energy systems. The basis for addressing PQDs is the automatic categorization of voltage or phase angle disturbances. According to a usual standpoint, the three distinct steps of signal analysis, feature selection, and classification should be used to separate the detection of issues with power quality. Nevertheless, signal analysis possesses several inherent deficiencies, mostly stemming from the laborious and inaccurate process of human feature selection. Consequently, this results in diminished classification accuracy when dealing with many disturbances and a compromised ability to withstand disruptive interference. This study focuses on the identification and categorization of PQD using a machine learning-based classifier, taking into account the features of the power quality problems problem, eight different features are taken from the voltage data and used to figure out what caused each PQD. Various types of machine learning models are employed to analyze the dataset, and the effectiveness of the machine learning classifier is assessed by validating its performance using a separate test dataset. Once the machine learning classifier model can classify the disturbances types with 96% accuracy. The proposed classifiers can effectively detect disturbances in the transmission line.**

***Keywords*—Power quality disturbances, Harmonic distortion, voltage disturbances, Machine learning, Classifiers,**

I. Introduction

One of the top objectives for both providers and customers in the contemporary electric network is to get dependable and high-quality electricity [1], [2]. Any disturbances, such as voltage sag, voltage swell, and the presence of harmonics, that might cause failure or malfunction in the customer's equipment are referred to as power quality issues [3]. Power quality issues in the electric network are caused by the integration of renewable energy sources like wind power plants and solar power plants, as well as by the increased use of power electronics converters, the adjustment or switching of large industrial loads, and the increased use of sensitive electronics loads in residential and commercial areas [4], [5], [6], [7]. A method for identifying and categorizing single-stage PQDs utilizing a rule-based decision tree and the Hilbert Transform. Utilizing mathematical relations, the MATLAB software generates PQDs. The Hilbert Transform is used to deconstruct the signals, and the output is then used to extract features for classification.

Numerous studies have been conducted on Power Quality monitoring systems that have become essential during the past 20 years. The increasing prevalence of power electronics–based nonlinear generators and loads in power grids is one factor contributing to the growing need for such devices due to the potential for these components to introduce high-order voltage and current harmonics [8], [9], [10]. The investigated PQDs include pure sine wave, voltage sag, voltage swell, momentary interruption, oscillatory transient, impulsive transient, spike, and notch. The proposed algorithm's effectiveness is demonstrated by calculating its efficiency on testing 50 data sets of each PQD obtained by varying the disturbance parameters [11]. For classifying power quality incidents in distribution networks, a machine learning-based algorithm was proposed. Developing and evaluating a model to categorize the 16 most common power quality events such as Normal, sag, swell, flicker, harmonic, interruption, transient oscillatory, notch, spike, and hybrid fault combination of two, evaluating the best machine learning model using the 'Classification Learners' tool in MATLAB. Implementing the selected machine learning model for testing distribution grid circuits using Simulink. Using simulation data to demonstrate the model's performance in a range of operational scenarios [12].

A hybrid, undecimated wavelet transform-based feature extraction method combining a support vector machine (SVM) and a trous algorithm is recommended for classifying PQDs in distributed generators [13]. Due to its selective nature, the time-frequency-based feature extraction for the classification of PQDs using deep learning algorithms was selected to be

979-8-3503-5028-9/24/$31.00

the average of instantaneous frequency and spectrum entropy. [14]. To identify which machine learning model performs the best, 16 categories of the most common power quality events are characterized using wavelet transform and certain machine learning techniques. The development and testing of a machine learning model for power quality event categorization [15].

The main contribution of this work is to explore the performance of the machine learning classifier determining the PQDs in the transmission line's voltage data of the electrical power system.

II. Methodology

The methodology for the study is shown in Fig. 1, consisting of 3 major sections — 1. Data processing, 2. Features grading using T-NSE, and 3. Classifier model fitting.

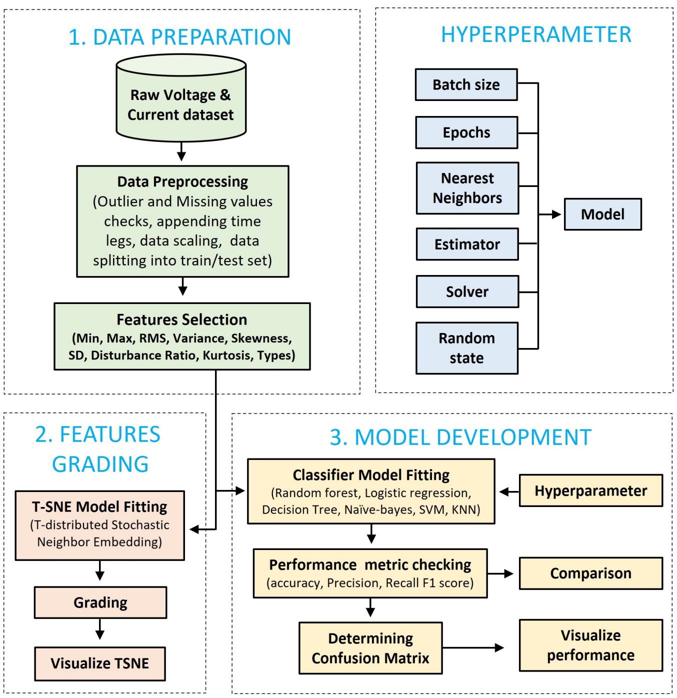


Fig. 1: Methodology

A. Data Processing

Data pre-processing procedures are a crucial stage in the workflow for both statistical analysis and machine learning. It comprises preparing the raw data for modeling and carrying out additional analysis. Here, Voltage data sets of 2.8×10^4 are taken for a case study which is collected from the Sagemaker Database [16], different types of disturbance are defined in columns each containing 20,000 voltage data rows. This raw data of voltage, current, and power needs to be pre-processed before using it to train models. The key stages for data preparation are given as follows —

a) Outliers: Data points known as outliers considerably differ from the remainder of the data sets. The outcomes of analysis and modeling may be distorted because of outliers. In this study, the outliers are replaced with equivalent values, such as mean, median, or tailored estimating approaches for voltage data sets. The signal-to-noise (SNR) of the data set is transformed into the log scale to lessen the outlier impact.

b) Missing Values: Missing data might lead to skewed or inaccurate findings. To remove such an error. Rows and columns having a large percentage of missing values are removed.

c) Appending Time Lags: Time delays are used to highlight variations in time in data sets. This enables the classifier model to take the data's past into account while generating predictions.

d) Data Scaling: Scaling is crucial to prevent the learning process from being dominated by distinct characteristics with varying ranges. Standardization is the process of re-scaling characteristics with a mean of 0 and a standard deviation of 1. The RobustScaler is used for K-Nearest Neighbour (KNN) classifier.

e) Data Sample Split: The data needs to be split into training and testing sets in order to evaluate the model's performance on data that has not been tested. A split ratio of 8:2 is employed where 80% of data is being used for training and 20% of data is being used for testing.

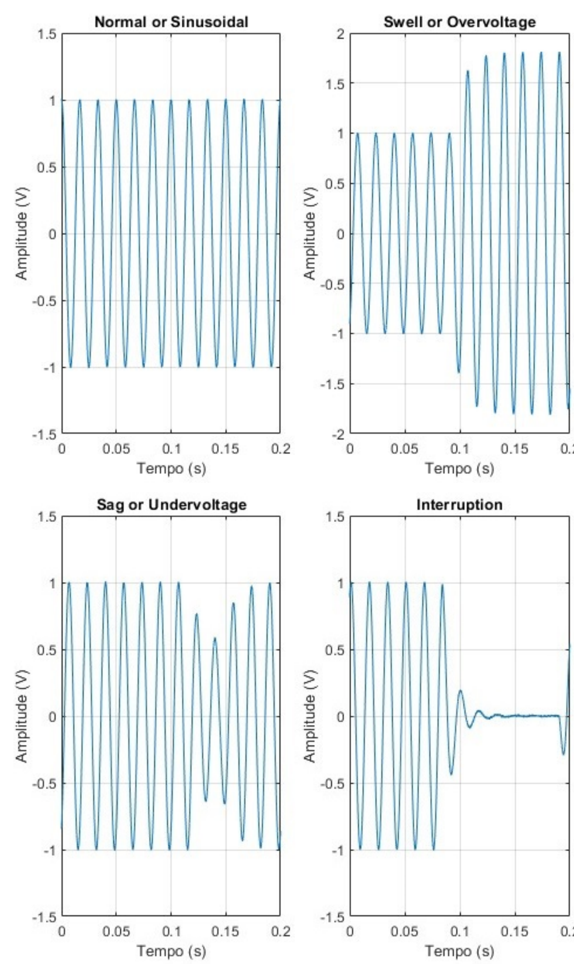


Fig. 2: Normal, Seg, Swell, Interruption Voltage waveform

f) Data Visualization: In power system analysis, PQDs are variations from the typical sinusoidal voltage waveforms. The voltage data of 11 types of PQD is taken as input for determining the best deep-learning model that can classify the types of fault causing the disturbance [1]. Sinusoidal PQD are any abnormalities in a power system's voltage waveforms. Over-voltage, under-voltage, interruption brought on by an external signal, or an asymmetrical line fault, as shown in Fig. 2, can all produce disturbance in power quality.

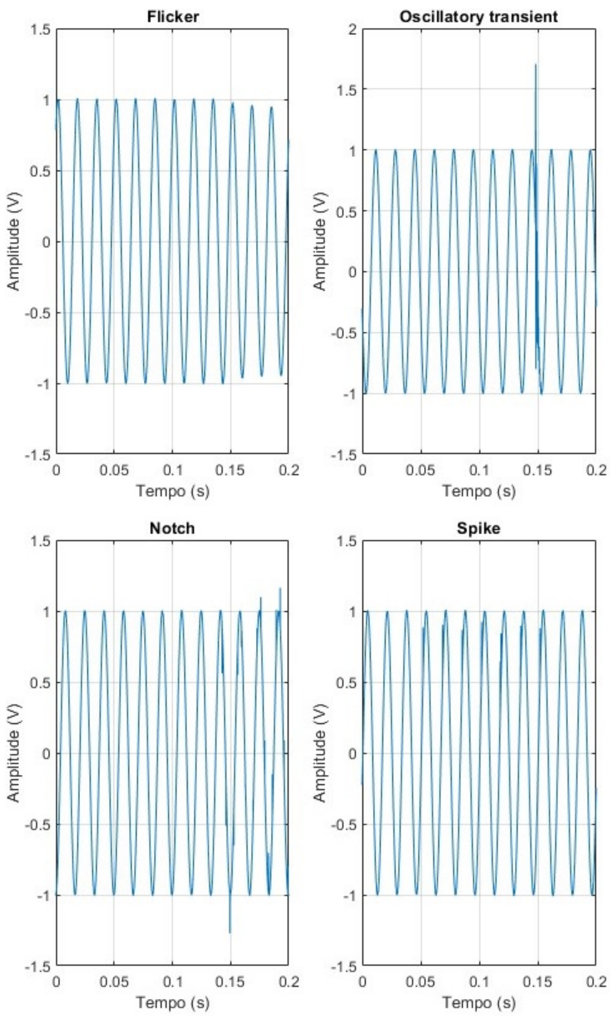


Fig. 3: Flicker, Oscillatory Transient, Notch, and Spike Voltage waveform

Momentary disruption is a lightning strike or switching activity that results in a rapid, large-amplitude, brief voltage shift is called transient disturbance, Rapid and erratic voltage changes that cause apparent light variation in incandescent bulbs are known as flickers and Voltage notching is the process of reducing voltage levels at particular frequencies for a brief period. Equipment switching, operation, or other forms of interference can all contribute to it, And an instance of a PQD known as an oscillatory transient is one in which voltage levels change rapidly and repeatedly, as shown in Fig. 3. Oscillatory transients can result from a variety of things, such as rapid switching processes, malfunctions, and interactions between different pieces of equipment in the power system. These interruptions can also be brought on by lightning strikes, swapping capacitors banks, and other quick changes in the network. The existence of non-sinusoidal waveform parts at multiples of integers of the basic frequency is known as harmonic distortion and these can be caused by nonlinear loads [1]. Harmonic disturbance can simultaneously occur with flicker, sag, and swell events as shown in Fig. 4 of the data set.

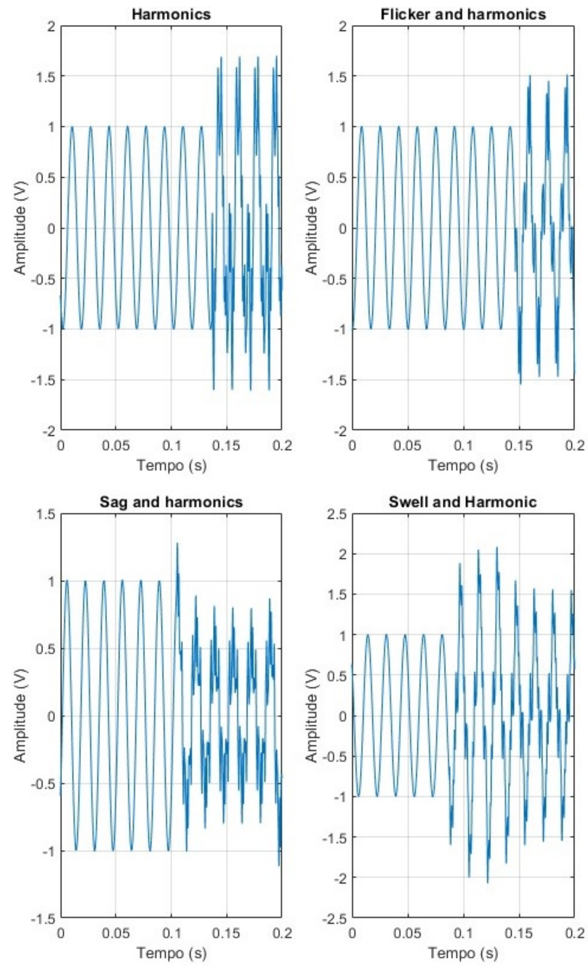
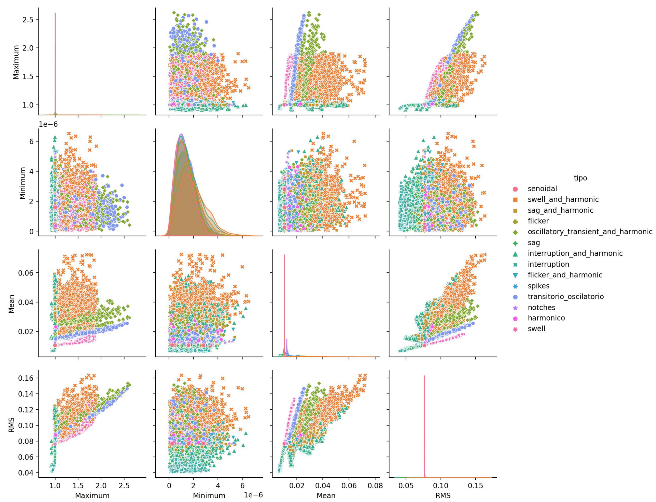


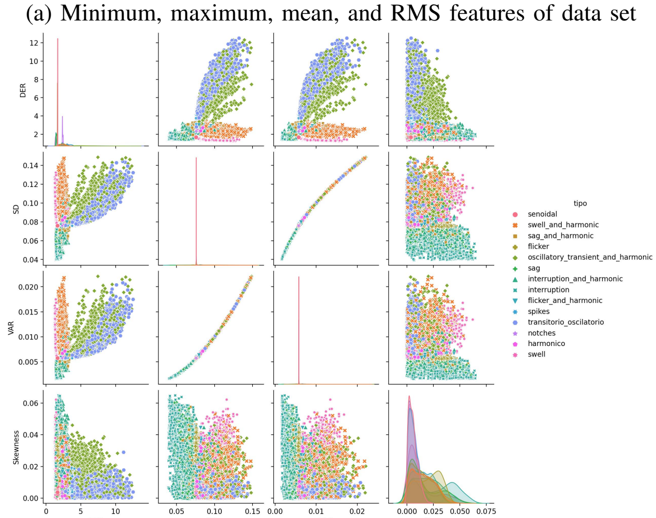
Fig. 4: Harmonics, Flicker-harmonics, Sag-harmonics, and Swell-harmonics voltage waveform

B. Features selection

A key component of getting data ready for machine learning classifier training involves feature selection. To enhance the accuracy of your models, lower the complexity of computation, and avoid over-fitting, which entails locating and selecting the most pertinent and instructive characteristics from the data set. The data set is reduced in dimensions. High-dimensional data might present problems such as potential over-fitting and higher processing complexity. By reducing the number of unnecessary characteristics, feature selection increases the generalization and effectiveness of models. Various methodologies can be utilized to conduct feature selection within the framework of machine learning classifiers [2] such as - Minimum, maximum, Mean, Root-mean-square (RMS), Disturbance-energy-ratio (DER), Standard deviation (σ), Variance (σ^2), Phase Skewness and Kurtosis values as per disturbance type shown in Fig. 5.



(a) Minimum, maximum, mean, and RMS features of data set



(b) Disturbance ratio, standard deviation, variance and skewness of data set

Fig. 5: Features of data set

The utilization of a pair plot shown in Fig. 5 serves as a valuable visualization tool that provides numerous advantages in facilitating a thorough comprehension of given features of the data set. A pair plot offers a comprehensive perspective on the relationships, patterns, and distributions present in the data by generating scatter plots for every possible combination of variables and incorporating diagonal plots to depict the distributions of individual variables. In Fig. 5b, the relation between disturbance ratio and skwnees seems proportional for some types of disturbance, and others are inversely proportional. Even though the raw voltage data may seem symmetrical as s sinusoidal data, looking at the features of the data set gives another broader aspect of different data types and how each of the disturbances can be differentiated from each other.

1) T-SNE model fitting: T-Distributed Stochastic Neighbor Embedding (T-SNE) is a method used for reducing the complexity of data sets. This technique provides numerous advantages in terms of facilitating the understanding and interpretation of intricate data [17]. Here, a total of 8 features — min, max, mean, RMS, disturbance ratio, SD, variance, and skewness are considered for a total of 14 types of disturbance in power quality. The considered disturbance types of the power quality are Sag, Swell, Flicker, Interruption, Oscillatory transient, Notch, Spike, Harmonics, flicker with harmonics, sag with harmonics, swell with harmonics, interruption with harmonics, and Oscillatory transient with harmonics. The t-SNE algorithm is applied for capturing intricate non-linear associations inside data sets characterized by high dimensional. Fig. 6 displays the T-SNE projection, which aids in comprehending the data set and the association among various types of disturbance, specifically concerning the voltage characteristic.

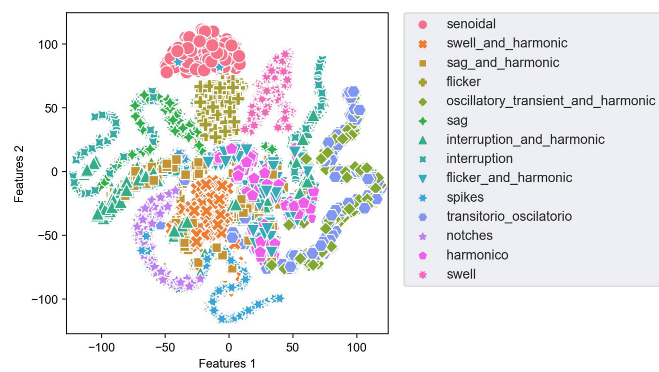


Fig. 6: T-SNE Projection

C. Classifier Model

In the field of machine learning, a classifier model pertains to a certain neural network architecture that is devised to allocate input data points into distinct categories or classes. Classification tasks encompass the prediction of the most suitable class label for a given input. machine learning classifiers possess significant efficacy due to their inherent ability to autonomously acquire and extract hierarchical features from the input data, hence facilitating their capacity to effectively address intricate patterns and variations. The machine learning classifier model from scikit-learn library is used to import the following functions - random forest, Gaussian naive Bayes, decision tree, support vector, logistic regression, K-Nearest Neighbour, and bagging classifier. RobustScaler() function is used to process the input train data for the K-Nearest Neighbour classifier. By fitting the process data set of the training series, the model is established.

D. Hyper-parameter

Hyper-parameters refer to the parameters of a machine learning algorithm that are predetermined before the commencement of the learning process. The hyperparameters utilized in the aforementioned machine learning classifier are presented in Table I.

TABLE I: Hyper-parameter of Classifier [18]

Parameter	Value	Classifier
Data Sample	20000	All
Split ratio	8:2	All
Epochs	3	All
Iteration	1000	Random Forest, Decision-Tree, GaussianNB, KNN, SVM classifier
	3000	Logistic Regression
Number of neighbors	100	KNN
Solver	L-BFGS	Logistic Regression
Estimator	SVC (n=10)	Bagging Classifier
Random state	1	All
	0	Bagging Classifier

III. Results and Analysis

A weighted average considers each type of disturbance in power quality that is tested with validation with each classifier. The weighted average accuracy for determining the PQD using different classifiers is shown in Fig. 7. The Random Forest classifier shows the best results for determining the different types of PQD with around 95.5% weighted average validation score, followed by the decision tree classifier 92.2% weighted accuracy score. This means these two machine learning classifiers can determine the types of disturbance in the power system with 92.2 to 95.5% accuracy.

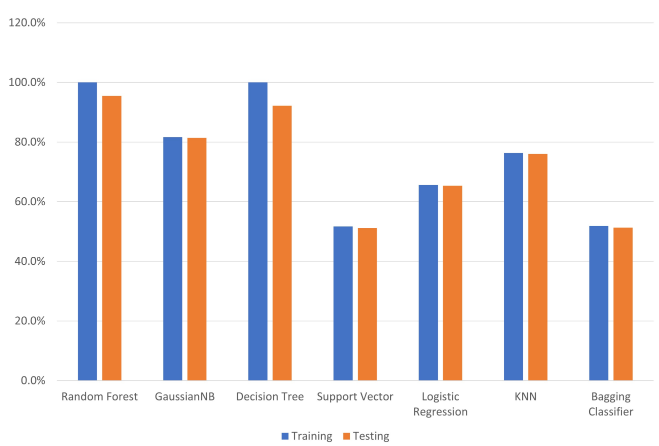


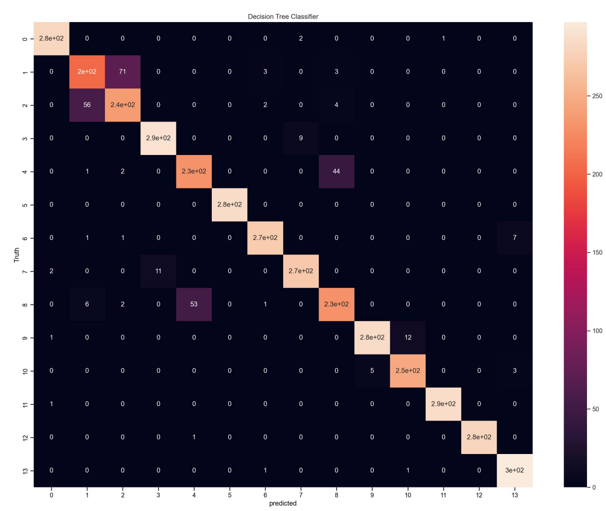
Fig. 7: Validation score of different classifier

A total of 14 different types of disturbance in power are tested using affirmation models and the classification report is provided in Table II. The random forest classifier and the decision tree classifier model performed best compared to other machine learning classifier models with an accuracy of around 92.2% to 95.5%. The overall accuracy of the model is the weighted average on the individual performance detection of the given problem as shown in Table II.

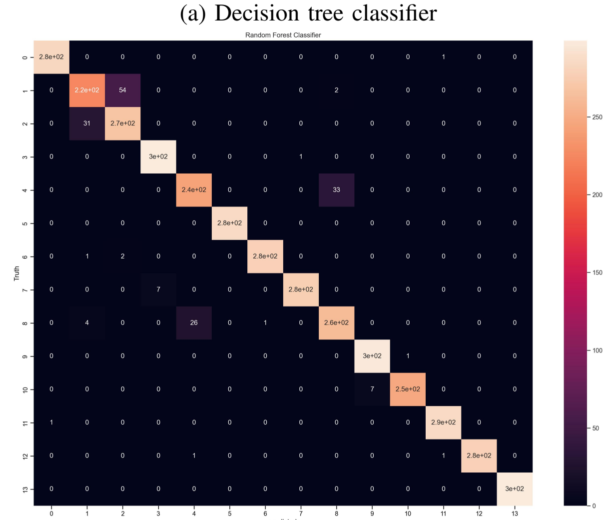
TABLE II: Classification Report

Types of Disturbance	Random Forest			Decision Tree		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Flicker	0.99	0.99	0.99	0.99	0.99	0.99
Flicker with harmonic	0.85	0.79	0.82	0.76	0.72	0.73
Harmonic	0.83	0.89	0.85	0.76	0.79	0.77
Interruption	0.98	1.00	0.99	0.96	0.96	0.96
Interruption with harmonic	0.90	0.87	0.89	0.80	0.82	0.81
Notches	1.00	1.00	1.00	1.00	1.00	1.00
Normal	0.98	1.00	0.99	0.98	0.95	0.97
Oscillatory transients	1.00	1.00	1.00	0.96	0.99	0.98
Oscillatory transient with harmonic	0.99	1.00	0.99	0.96	0.96	0.96
Sag	0.99	0.98	0.98	0.96	0.96	0.96
Sag with harmonic	0.87	0.90	0.89	0.81	0.78	0.79
Spikes	1.00	0.97	0.98	0.94	0.96	0.95
Swell	0.99	1.00	0.99	0.99	1.00	0.99
Swell with harmonic	1.00	0.99	1.00	1.00	1.00	1.00
Weighted avg. accuracy		0.95			0.92	

The confusion matrix of the best-performing classifier for the PQD is shown in Fig 8.



(a) Decision tree classifier



(b) Random forest classifier

Fig. 8: Confusion matrix

The confusion matrix is a tabular representation of a comprehensive assessment of the performance of a machine learning model in a classification problem. The evaluation process aids in comprehending the efficacy of the model in accurately categorizing situations into distinct classes by juxtaposing its predictions with the verifiable ground truth. The utilization of a confusion matrix proves to be highly advantageous in obtaining valuable insights regarding the specific categories of errors that the model is generating. The overall data for all the classifier models used in the proposed study is given in Table III.

TABLE III: Performance between different classifiers

Classifier Models	Accuracy Score	Precision Score	Recall Score	F1 Score
Random Forest	0.955	0.955	0.955	0.955
Gaussian Naive Bayes	0.814	0.837	0.814	0.818
Decision Tree	0.922	0.922	0.922	0.922
Support Vector	0.512	0.554	0.512	0.467
Logistic Regression	0.654	0.662	0.654	0.632
K-Nearest Neighbors	0.760	0.762	0.760	0.757
Bagging Classifier	0.513	0.576	0.513	0.473

The evaluation of a model's performance should begin with accuracy, but it is crucial to also take into account additional metrics such as precision, recall, and F1-score. This comprehensive approach is particularly important in situations where there is a disparity in class or varying costs related to false positives and false negatives.

IV. Conclusions

The utilization of power quality monitoring technologies has become imperative, prompting numerous recent research to concentrate on the identification and categorization of power quality issues. Currently, a significant impediment to the direct comparison of PQD classification techniques is the absence of a standardized database that may serve as a benchmark. In this study, we put out a proposition for an open-source transmission line voltage dataset, that was taken as a case study to train different types of machine learning classifier models to detect the PQDs within the dataset. These disturbances are characterized by different levels of repetition and random characteristics. Additionally, the application has two reference classifiers that perform best for classifying such problems. Given the impressive performance exhibited by these classifiers, we propose that they be employed as standard benchmarks within the research community to advance the development of novel and enhanced methods for PQD classification.

Future studies should be focused on integrating the best-performing machine learning classifier to detect synthesized real-time disturbance data

References

[1] K. Thirumala, S. Pal, T. Jain, and A. C. Umarikar, "A classification method for multiple power quality disturbances using ewt based adaptive filtering and multiclass svm," *Neurocomputing*, vol. 334, pp. 265–274, 2019.

[2] S. Wang and H. Chen, "A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network," *Applied energy*, vol. 235, pp. 1126–1140, 2019.

[3] Y. Xi, Z. Li, X. Zeng, X. Tang, Q. Liu, and H. Xiao, "Detection of power quality disturbances using an adaptive process noise covariance kalman filter," *Digital Signal Processing*, vol. 76, pp. 34–49, 2018.

[4] U. Singh and S. N. Singh, "A new optimal feature selection scheme for classification of power quality disturbances based on ant colony framework," *Applied Soft Computing*, vol. 74, pp. 216–225, 2019.

[5] A. I. Ikram, A. Ullah, D. Datta, A. Islam, and T. Ahmed, "Optimizing energy consumption in smart homes: Load scheduling approaches," *IET Power Electronics*, Jan 2024.

[6] A. I. Ikram, M. Shafiullah, M. R. Islam, and M. K. Rocky, "Techno-economic assessment and environmental impact analysis of hybrid storage system integrated microgrid," *Arabian Journal for Science and Engineering*, Feb 2024.

[7] H. Liu, F. Hussain, Y. Shen, S. Arif, A. Nazir, and M. Abubakar, "Complex power quality disturbances classification via curvelet transform and deep learning," *Electric Power Systems Research*, vol. 163, pp. 1–9, 2018.

[8] M. H. Bollen and I. Y. Gu, *Signal processing of power quality disturbances*. John Wiley & Sons, 2006.

[9] E. Hossain, M. R. Tür, S. Padmanaban, S. Ay, and I. Khan, "Analysis and mitigation of power quality issues in distributed generation systems using custom power devices," *Ieee Access*, vol. 6, pp. 16 816–16 833, 2018.

[10] S. K. Khadem, M. Basu, and M. Conlon, "Power quality in grid connected renewable energy systems: Role of custom power devices," in *International Conference on Renewable Energy and Power Quality (ICREPQ)*. IEEE, 2018.

[11] R. Saini, O. P. Mahela, and D. Sharma, "An algorithm based on hilbert transform and rule based decision tree classification of power quality disturbances," in *2018 IEEE 8th Power India International Conference (PIICON)*. IEEE, 2018, pp. 1–6.

[12] T. P. Tun and G. Pillai, "Power quality event classification in distribution grids using machine learning," in *2021 56th International Universities Power Engineering Conference (UPEC)*. IEEE, 2021, pp. 1–6.

[13] A. Yılmaz, A. Küçüker, G. Bayrak, D. Ertekin, M. Shafie-Khah, and J. M. Guerrero, "An improved automated pqd classification method for distributed generators with hybrid svm-based approach using un-decimated wavelet transform," *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107763, 2022.

[14] T. S. Gunawan, B. Y. Husodo, E. Ihsanto, and K. Ramli, "Power quality disturbance classification using deep bilstm architectures with exponentially decayed number of nodes in the hidden layers," in *Recent Trends in Mechatronics Towards Industry 4.0: Selected Articles from iM3F 2020, Malaysia*. Springer, 2022, pp. 725–736.

[15] T. P. Tun and G. Pillai, "Power quality event classification in distribution grids using machine learning," in *2021 56th International Universities Power Engineering Conference (UPEC)*. IEEE, 2021, pp. 1–6.

[16] Amazon, "Amazon sagemaker," 2023. [Online]. Available: https://aws.amazon.com/sagemaker/

[17] H. Perez and J. H. Tah, "Improving the accuracy of convolutional neural networks by identifying and removing outlier images in datasets using t-sne," *Mathematics*, vol. 8, no. 5, p. 662, 2020.

[18] R. Machlev, A. Chachkes, J. Belikov, Y. Beck, and Y. Levron, "Open source dataset generator for power quality disturbances with deep-learning reference classifiers," *Electric Power Systems Research*, vol. 195, p. 107152, 2021.