International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 8 Issue 3 March - 2024, Pages: 1-9

Enhanced Partial Discharge Detection in High-Voltage Transmission Lines Through Hybrid Neural Networks

Mohammed Eishorafa

Electrical and Electronic Engineering Department, Faculty of Technology,
Sakarya University of Applied Sciences, Sakarya, Turkey

Abstract: *High and medium-voltage transmission lines play a critical role in supplying electricity to urban areas. Manual inspection of these extensive networks is impractical due to their vast reach, leading to potential damage and inefficiency. Additionally, partial discharges (PD), resulting from insulator malfunctions and external impacts, pose a significant threat to the integrity of power lines. Detecting PD promptly is essential to prevent equipment damage and ensure reliable power transmission. In this study, we propose a Bidirectional Long Short-Term Memory (BiLSTM) deep neural network approach to detect PDs. Article BiLSTM model is pre-trained using a unique dataset collected from real power lines with a custom-designed meter by the VSB-Technical University of Ostrava. Notably, the dataset includes real-world signal data with inherent background noise, necessitating preprocessing techniques such as central tendency, statistical dispersion, entropy, and fractality analysis. To enhance classification accuracy, we introduce a novel approach that merges two neural network models, Model160 and Model_320, each utilizing distinct input data. The hybrid model, created by combining these models, achieves remarkable performance, with an accuracy of 97.5% and a Matthews Correlation Coefficient (MCC) of 0.79. Importantly, our hybrid model surpasses Kaggle-winning solutions and other state-of-the-art models trained on the same data, demonstrating its superior capability for PD detection. This research showcases the effectiveness of combining two distinct neural network systems to achieve a synergistic enhancement in performance, setting a new standard for PD detection in high-voltage transmission lines.*

Keywords: Partial Discharge Detection, MCC, Power Line Inspections, BiLSTM, Hybrid Model.

1. INTRODUCTION

Monitoring electrical connections is of utmost importance to maintain a stable energy supply to homes, establishments, and institutions, particularly critical facilities like hospitals. Malfunctions in power transmission lines not only result in power outages but can also lead to fires and disruptions in high or medium voltage installations, power distribution stations, and electrical transformers [1]. PD is a critical issue in power transmission systems, typically caused by the concentration of electrical stress within insulation or on its surface [2]. This phenomenon can be conceptually modeled, as shown in Figure 1, where defects are represented as capacitive insulators [3]. When the electric field across a capacitor exceeds its capacity, partial discharge occurs, generating a current pulse along the capacitor.

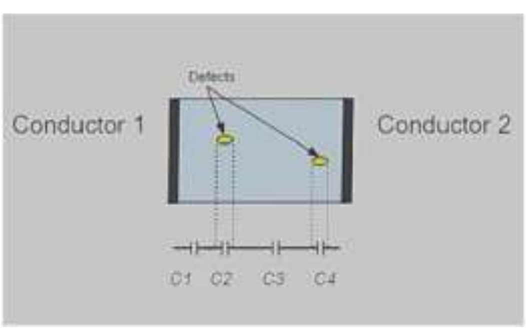

Figure 1: Scheme of electrical model of defects present inside a dielectric [3].

The failure of high voltage equipment is primarily attributed to insulator breakdown in transmission (overhead) lines. PD occurs in about 84\% and 89\% of transformers and insulating cables, respectively. PD can manifest in various mediums, including gases, liquids, or solid insulators, when an appropriate electrical stress level is reached [4]. Although partial discharge does not immediately cause insulator failure, it is often the underlying cause that leads to eventual equipment breakdown [5]. PD serves as an early warning sign of deteriorating insulating materials and their susceptibility to breakdown. If left unaddressed, this deterioration can escalate to a point where the insulating materials can no longer withstand electrical stress, resulting in flashes and catastrophic equipment failure. These failures can be sudden and devastating, particularly in high-voltage situations, causing significant damage and power supply interruptions. Therefore, detecting partial discharge without interrupting power supply has become a significant challenge for electrical engineers. Online monitoring is one method that allows for the detection of partial discharge while maintaining a continuous power supply. Maintenance measures, based on examination results, are then undertaken, ranging from the replacement of insulators to other electrical components on the verge of failure, to prevent future faults. PD can be categorized into three types: Internal Partial Discharge, occurring inside insulation Figure 2, Surface Partial Discharge, tracking across insulation, and Corona Partial Discharge [6].

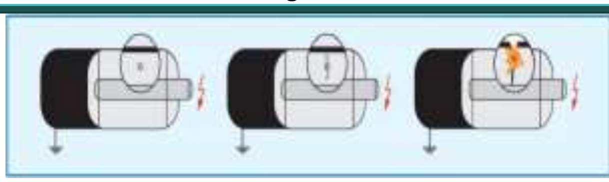

Figure 2: Partial discharge (PD) is a result of concentrations of electrical stress within the insulation or on the surface [2].

In recent years, deep neural networks have gained prominence in pattern recognition [7]. Multilayer Perceptrons (MLPs) are a type of neural network commonly used for various problems [8]. Other popular deep neural network architectures include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks [9]. Additionally, machine learning algorithms such as support vector machines, Nearest Neighbor, Linear Regression, K-Means, and others have been employed in various applications [10]. In this article, we explore the application of artificial intelligence and deep learning techniques to monitor partial discharge online without the need for power interruptions. Our research encompasses four stages: data acquisition for training and testing, preprocessing of the data to prepare it for training, binary classification of the presence or absence of the partial discharge phenomenon, model design and training, and model testing using Kaggle-provided data, which undergoes the same preprocessing steps as the training dataset. Based on what was mentioned above, this problem is big and can be a difficult job to manually monitor it periodically using sensors and specialized devices. So the science of artificial intelligence will be used to contribute to solving this dilemma and discovering the phenomenon of PD in the supply lines of high or medium voltage for the three phases by designing an AI model to predict and classify if the phenomenon was formed or not. Accordingly, the appropriate maintenance procedure should be taken to address the problem without loss of energy supply.

2. Literature review

The interest in the measurement and localization of PD phenomena has garnered significant attention both locally and globally. Various entities, including academic institutions' electrical and electronic departments, electricity distribution companies, and the military, have been actively involved in this field since at least the early 1940s [11-14]. Early research has indicated that factors such as size, cavity shape, temperature, and electrical pressure can influence the occurrence of PD [15]. PD remains a prominent cause of insulator damage, although it does not typically lead to immediate failure. The maintenance of insulators in optimal condition is crucial for both Direct Current (DC) and Alternating Current (AC) voltage applications. Consequently, numerous methods have been devised to assess insulator condition, with the detection of PD emerging as a prominent indicator [16]. While industrialists can employ specialized sensors to identify PD in transformers, the precise location of the phenomenon remains challenging through electrical measurements alone. Early detection of PD is vital to prevent complete insulator failure and mitigate equipment damage, thereby reducing overall costs [17]. The detection of PD using electrical measurements can be complicated by electromagnetic interference and environmental noise, including that from human activity and wildlife, such as birds. Chemical approaches have proven effective for PD detection by identifying combustible gases [18]. Insulation equipment failures can be attributed to two primary mechanisms: chemical dissolution and nitrogen ion bombardment [19].

The application of artificial intelligence (AI) in various domains, including engineering, has shown remarkable efficacy in addressing electrical faults. AI methods have been employed to replace chemical detection approaches effectively. For example, Akbal et al. employed hybrid Artificial Neural Network (ANN) methods to predict sheath currents, aiding in the prevention of cable faults during underground cable line installations [20]. Yaw et al. developed an expert system to detect short-circuits in stator windings of permanent-magnet synchronous machines (PMSMs) [21]. Wang et al. proposed the use of Convolutional Neural Networks (CNNs) for fault detection in power systems, such as transmission lines [22]. High-impedance faults (HIFs) on overhead power lines pose significant risks, including insulator breakdown leading to power line contact with trees or the ground, potentially causing wildfires [23]. To address this, an Online Monitoring AI system was proposed to achieve accurate and rapid HIF detection [24]. Tehrani et al. developed an AI model integrating a Long Short-Term Memory (LSTM) architecture with an attention layer and one-dimensional CNN to extract feature and frequency information [25].

In this article, we introduce a novel deep learning model for PD prediction without interrupting the power supply. The model comprises four inputs representing processed three-phase voltage signals and one output for prediction. Our discussion comprises three main parts: data processing and feature engineering, AI model design, and testing and comparison. To facilitate our research, two distinct working areas are presented in Figure 3.

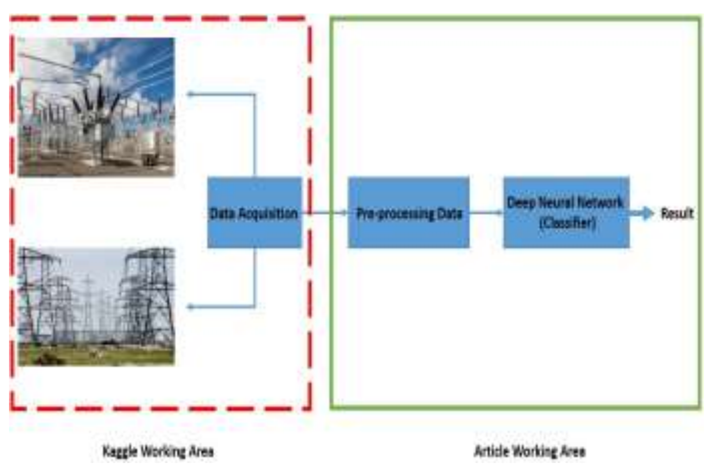

Figure 3: working areas in the research process.

First, the Kaggle working area is responsible for providing training and testing datasets collected from power lines using a new meter designed at the ENET Centre at VŠB. Second, the article working area involves utilizing this data to train an AI model following data preprocessing.

3. DATA PROCESSING AND FEATURE ENGINEERING, EXTRACTION

The training and testing datasets were obtained from the Kaggle competition website which gives the competitors the ability or even anyone to publish and find trained AI models and even datasets for free. It is necessary to put the dataset in the form to be understandable before designing the AI model by preprocessing the raw data and extracting the features from them because the use of raw data as it is and its effect on the neural network in many cases leads to the inability of the model to learn. The signal data comes from the real environment, not a lab, and they contain a lot of background noise. These signals are measured by the patented device of ENET with a lower sampling rate (cost efficiency purpose). In addition to this, metering devices were deployed in more than 20 different locations (8,712 phases). This implies that the spectrum of noise and quality of PDs are so different from each other, that the correct and robust classification is still an ongoing problem.

3.1 DATA DESCRIPTION

Each training and testing dataset signals have the same properties of sample time, several samples, and grid operating frequency. Each three-phase voltage signal is measured in a sample of 25 nanoseconds, which is a great sampling time for a high-quality ADC signal. In addition, the electrical system operates at 50 Hz, meaning that each phase has been converted from analog to digital model with a sample number equal to 800,000. In short, the sample number covers a complete voltage signal. As shown in figure 4, there is almost no difference between the unproblematic (class-0) signal and the problematic (class-1) signal in the raw signal, which is the difficulty of this problem.

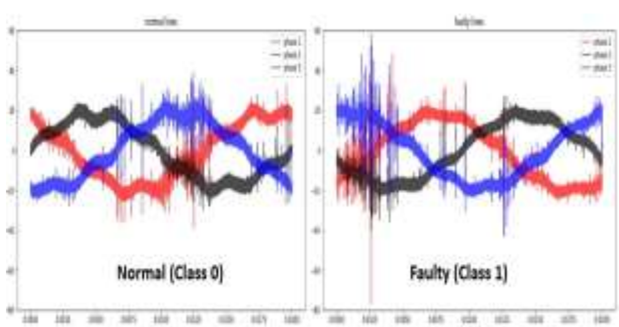

Figure 4: Normal 3-phase power lines on the left and faulty 3-phase power lines on the right.

The training dataset and testing dataset sizes are 8,712 and 20,336, respectively. After observing the distribution of the classes that out of the 8,712 phase signals (in different 20 locations), only 525 signals are class 1, representing only 6\% of the total data (imbalance data) as shown in figure 5. It is expected because partial discharge is not a normal thing to have occurred.

Figure 5: Bar chart input data distribution.

There are many techniques or algorithms to remove noise from signals like Wavelet Denoising, Fourier Denoising, and Moving Average Denoising. In this article, wavelet Denoising was used to remove sinusoidal waves from the signal. Consequently, the result of removing sinusoidal components from both types of three-phase signals is shown in figure 6 for normal lines on the left and on the right for faulty lines.

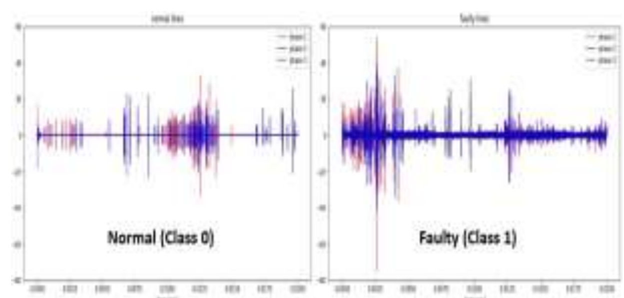

Figure 6: Class-1 denoised 3-phase power lines on the left and Class-0 denoised 3-phase power lines on the right.

As illustrated each phase has a sample size of 800,000 and these data range between -128 to 127 because it is an 8int data type. Additionally, these data regardless of whether it is original or processed are not convenient to input into the machine learning (ML) model. It is not great modeling to design a model with 800,000 inputs in a signal form. It is not correct to train the model with each phase separately because the three phases have correlations between them or even affect each other. If this bad ML perception is represented in reality as shown in figure \ref{fig_7}, it means that there will be three ML models and each model has 800,000 inputs. Therefore the data should be converted from signal shape to vector shape and extract the features and combine all three phases into a single

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 8 Issue 3 March - 2024, Pages: 1-9

vector to treat them together and not separately. This, in turn, contributes to building a robust artificial intelligence system.

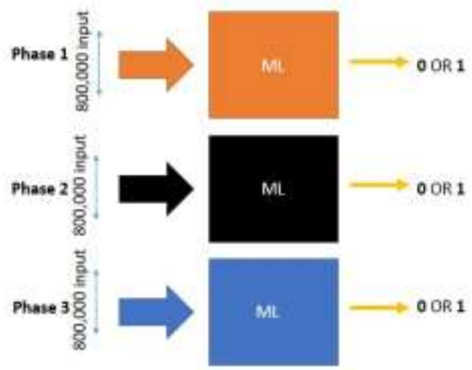

Figure 7: Bad ML visualization design.

3.2 Converting Data Signal to Vector

This process aims to make an AI model or algorithm handle a set of values at a time instead of one by one. As shown in figure 8, the vectorization stages start by taking input data as signals and finish by producing two output data types as vectors. These vectors represent the simple and complex feature extracted data for 160 and 320 windows.

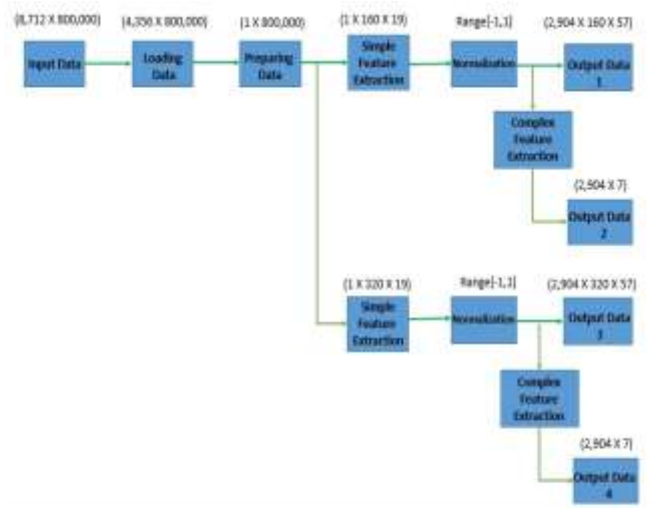

Figure 8: Vectorization stages.

The first stage is input data stage, the raw signal data with shape (8,712 X 800,000) with range [-128,127] and also 800,000 represent one cycle at 50Hz. After then, in loading data stage, since the signal size (8,712 X 800,000) is larger than the RAM that the Kaggle Kernel has, this function will perform a simple task. It is dividing the input signal into two intervals (4,356 X 800,000) and passing it to the preparing data function. Then, in preparing data stage, each phase is sent separately to the feature extraction stage which means (1 X 800,000) will be sent to the next step to extract the desired features. As shown in figure 9, the feature time step was set to be 160 and 320, which means there will be 160 and 320 windows.

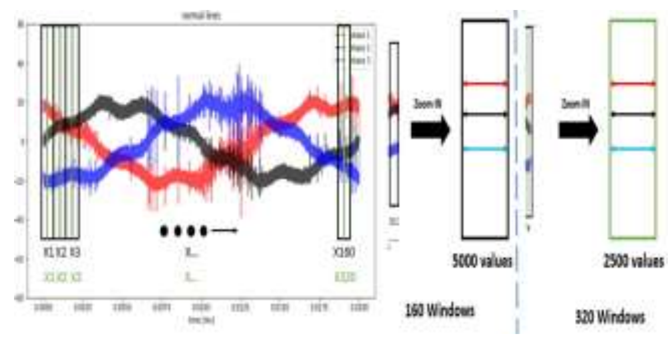

Figure 9: Dividing 3 phase signals to 160,320 windows and zoomed in the window.

Simple extracting features, each phase is processed separately and features are extracted from it. As mentioned previously, the number of windows is 160 and 320, which means the features will be extracted from 5000 voltage values, and 19 features will be extracted from each phase. So after that, the shape of the vector will be (1 X 160 X 19) and (1 X 320 X 19) for each phase but for every 3 phases, the vectors shape will be 3*(1X 160 X 19) = (1 X 160 X 57) and 3*(1X 320 X 19) = (1 X 320 X 57). After converting each phase from (1 X 800,000) as a signal to (1 X 160 X 57) and (1 X 320X 57) as a vector but the voltage values could be in the range [-128,127] because 8int data type. Hence, these vectors should be normalized. Normalization is a data preparation technique that is frequently used in machine learning and means rescales the values into a range of [0, 1] but because the dataset has negative values have been rescaled into a range of [-1, 1]. Min-Max scaling technique was proposed to normalize data in this paper because the range of original data is known in the range [-128,127] and the result of normalizing is putting data in the range [-1, 1].

3.3 Simple Feature Extraction

Sometimes a machine learning model needs preprocessed input data by feature extraction as this would help the model to learn faster and better. That leads to extract features from raw data before giving them to the machine learning model and that process is called feature engineering. It is one of the crucial steps in the process of predictive modeling, it is the process of transforming arbitrary data into a form that is understandable in supervised learning models. Simple extracted features belong to the measures of central tendency and statistical dispersion. Whereas measures of central tendency are numerical measures used to measure the concentration or aggregation of data such as mean, standard deviation, Standard deviation top and bottom, Percentile[0 1th 25th 50th 75th 99th 100th], Relative percentile and max range.

Series1 : [1, 2, 1, 2, 1, 2, 1, 2, 1, 2], mean = 1.5, $\sigma = 0.5$ (1)

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 8 Issue 3 March - 2024, Pages: 1-9

Series2 : [1, 1, 2, 1, 2, 2, 2, 1, 2, 1], mean = 1.5, $\sigma = 0.5$ (2)

When looking at the two series (1) and (2), the first one is regular and it is easy to guess the next value because it alternates 1 and 2 regularly but the second one is irregular which has either a value of 1 or 2, chosen randomly, each with probability 1/2. Even though both series have the same mean and standard deviation but they don't have the same complexity. This means that moment statistics, such as mean, standard deviation and variance, will not distinguish between these two series in terms of complexity. Nor will rank order statistics distinguish between these series. Complexity is one of the tools of time series analysis. And that there is great interest in measuring the complexity of many signals such as bird song, ECG, a protein sequence, or DNA. The complexity measurements for time series data can be divided into three primary groups, i.e., fractality (mono- or multi-fractality) for self-similarity (or system memorability or long-term persistence), methods derived from nonlinear dynamics (via attractor invariants or diagram descriptions) for attractor properties in phase-space, and entropy (structural or dynamical entropy) for the disordered state of a nonlinear system. Shannon’s entropy has been used to measure the complexity of discrete systems [26]. Although the entropy formula was conceived in the thermodynamic area, the entropy concept has spread to different disciplines adapting its meaning regarding the applied area and making tools for many applications [27]. Fractals are infinitely complex patterns that are self-similar across different scales. That means the Fractal dimension is a measure of how "complicated" a self-similar figure is. The terms fractal dimension (FD) and fractal were coined by Mandelbrot in 1975 [28]. Moreover, Humans are fractal. Their circulatory system, lungs, and brains are like trees. Many fractality feature types were extracted from data such as Petrosian FD, Katz FD and Higuchi FD.

The entropy in machine learning is a measure of the randomness in the information being processed. The higher the entropy, the harder it is to draw any conclusions from that information. Entropy is a useful tool in machine learning to understand various concepts such as feature selection, building decision trees, and fitting classification models, etc [29]. Many entropy feature types were extracted from data such as Permutation Entropy(PE), Singular Value Decomposition entropy (SVD entropy), Approximate Entropy (ApEn) and Sample Entropy (sampEn).

4. THE PROPOSED AI MODEL DESIGN

In general and mainly, recurrent neural network RNN type was proposed in designing AI model but particularly, Bidirectional Long short-term memory (BiLSTM) preferred to RNN because vanishing gradients was a series problem and which makes the network not learn significantly. In addition to this, the provided data is unbalance and BiLSTM can overcome this problem because it consists of two LSTMs in the forward and the backward direction and also can handle sequence time series data. This in turn gives the strong advantage of increasing the amount of training data because it is processed from both directions. Thus, it will produce a strong model and work better than the LSTM, but takes more time during training.

4.1 Proposed Deep Neural Network Design

Functional API model structure was used to develop and build the AI model because building high performance and accurate partial discharge classifier model needs to be multi-input in order to compromise between the extracted features which are related to central tendency and features related to complexity and entropy of data. And there are two types of data which were produced from pre-processing and feature extraction stages. As shown in figure 10, the proposed model, or can be called model_hybrid combines model_160 and model_320 to compromise the properties of both models. Hence the model_hybrid deals with data in shape (2904 X 160 X 57) with extra complex feature data (2904 X 7) and shape (2904 X 320 X 57) with extra complex feature data (2904 X 7). The model_hybrid consists of many layers as follows:

BiLSTM layer with 128 cells accepts the input data in shape (2904 X 160 X 57) then BiLSTM layer with 64 cells which receives the output of the previous BiLSTM (128) layer, then the attention layer that offers a great advantage in that it can give some information more important than others or even ignore it completely. In general, it simulates cognitive attention. It receives the output of the previous BiLSTM (64) layer.

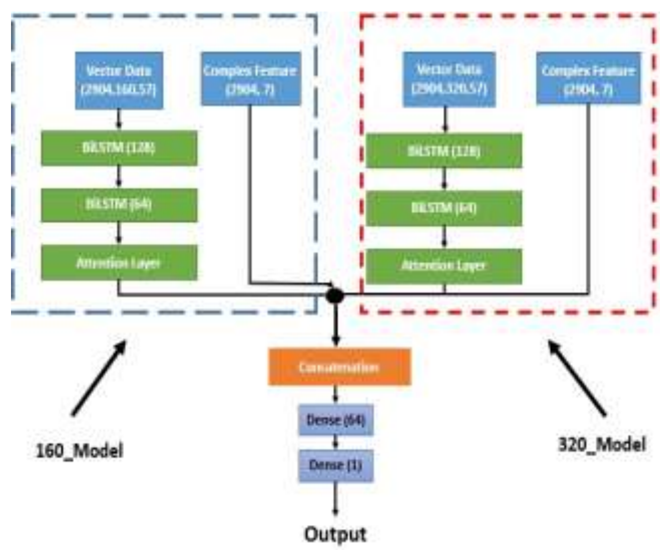

Figure 10: Proposed Deep Neural Network block diagram.

The same data pre-processing, and feature extraction procedures have been done on the model_320 side but the dataset has been divided into 320 windows. Moreover, the 320_model has been trained on the same hyperparameters as 160_model. The first BiLSTM layer which has 128 cells accepts the input data in shape (2904 X 320X 57), then the

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 8 Issue 3 March - 2024, Pages: 1-9

BiLSTM layer with 64 cells which receives the output of the previous BiLSTM (128) layer, and then the attention layer receives the output of the previous BiLSTM (64) layer. The concatenation layer connects the output of two attention layers with the complex feature data in shape (2904X7), then a dense layer with 64 cells using ReLU as an activation function. Since LSTM accepts inputs sequence of vectors, so dense layer is used because it accepts input as a sequence of vectors and converts it into 1D vector outputs. Finally, dense layer with 1 cell using sigmoid as an activation function and this layer is an output layer which predicts whether it is class 0 or 1. The total number of trainable parameters is 727,777.

4.2 K-Fold Validation

The training dataset was divided into training and validation. The validation part was used to examine the performance of the model during training. The testing dataset could not be used to examine the model performance because it doesn't provide the target and it is hidden by Google. The testing dataset was used to evaluate the competitors by Google Kaggle.

K-fold approach provides the ability to use the dataset in training and uses an amount from it for verification purposes. The great importance of this approach appears with a small dataset, as following this extends the data theoretically. For example, if the dataset is divided into five folds as followed in this article and as shown in figure 11. That means the accuracy is calculated five times and the overall accuracy is the average.

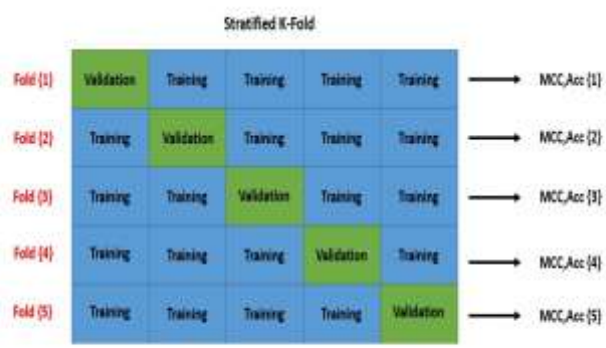

Figure 11: Stratified training data in K-fold.

Whereas, accuracy reflects model performance and predictability. There are many statistical tools that can measure performance in the field of deep learning such as accuracy, F1 score, and recall but the most common one is accuracy. Accuracy doesn't work well every time, especially when dealing with unbalanced data and in this article the data in there is unbalanced, so depending on the accuracy in measuring how the model works would be wrong. In 1975, a new tool proposed by Brian W. Matthews is called the Matthews Correlation Coefficient (MCC) for evaluating performance. Moreover, MCC is not sensitive to unbalanced data because it takes into account four parameters as follows: true negatives (TN), true positives (TP), false negatives (FN), and false positives (FP). So the MCC expression is as (3) [30].

$$MCC = \frac{TN \ X \ TP - FP \ X \ FN}{\sqrt{(TN+FN)(FP+TP)(TN+FP)(FN+TP)}} \tag{3}$$

The closer the value is to one, the more reliable the model works, but if the value is zero, this means that the model does not learn and gives random results, nothing more. In addition, if the value approaches one from the negative side, this also means that the system is working fine, but in reverse, and this can be fixed by reversing the output.

4.3 AI Model Training

The model has been trained 30 times in each fold and during the training, the Matthews Correlation Coefficient (MCC) was continually being checked and at best MCC the model weights were saved in the .h file. After that, the weights that achieve the best MCC will be relied on. As shown in figures \ref{fig_12} to \ref{fig_15}, training and validation accuracy and loss in each fold at each epoch. The figures \ref{fig_16} to \ref{fig_18} show the MCC in each fold at each epoch as well.

After the designing and training deep learning models have been finished at the same hyperparameters such as a number of epochs, learning rate, etc. The comparison between three designed AI models is illustrated in table 1, the model_hybrid is superior to other models in MCC, training and validation accuracy, and loss but the other two models are superior to model_hybrid in having less model training time and number of trainable parameters.

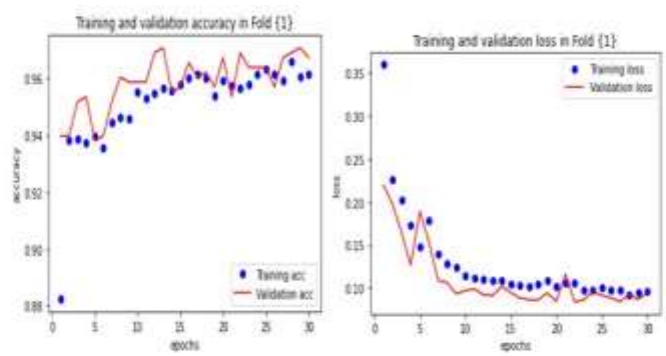

Figure 12: Training and validation accuracy and loss in fold 1.

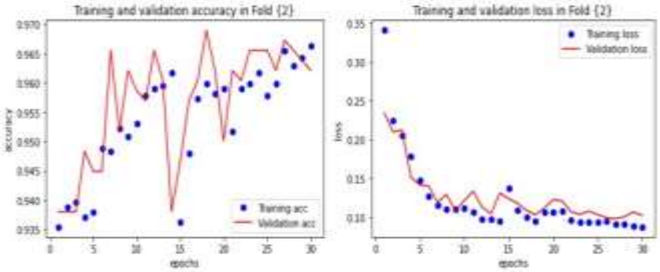

Figure 13: Training and validation accuracy and loss in fold 2.

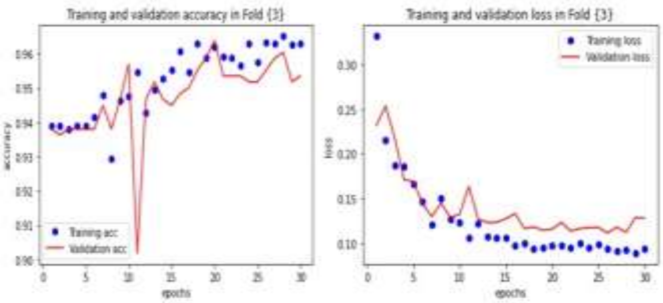

Figure 14: Training and validation accuracy and loss in fold 3.

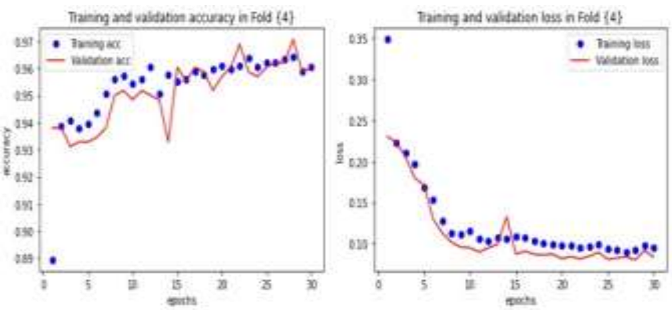

Figure 15: Training and validation accuracy and loss in fold 4.

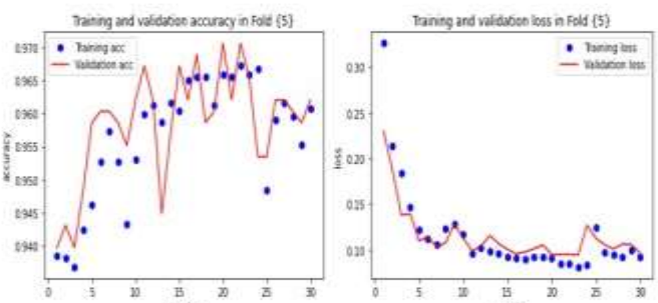

Figure 16: Training and validation accuracy and loss in fold 5.

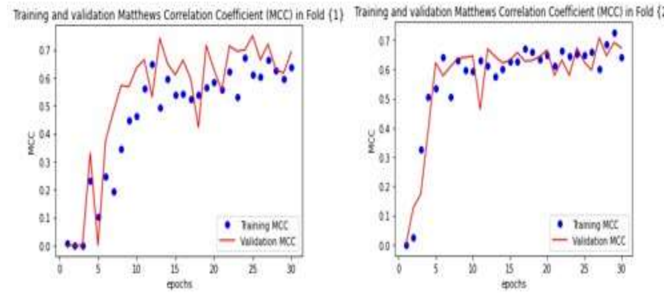

Figure 17: Training and validation of Matthews Correlation Coefficient (MCC) in Fold 1 on the left and fold 2 on the right.

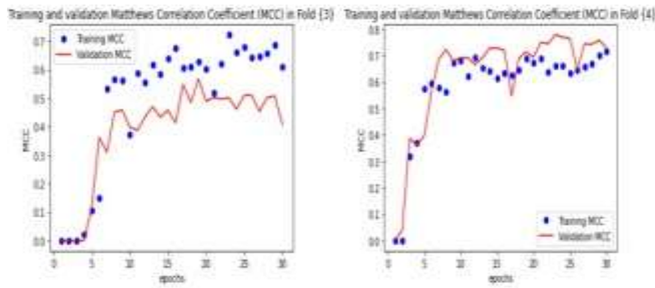

Figure 18: Training and validation of Matthews Correlation Coefficient (MCC) in Fold 3 on the left and fold 4 on the right.

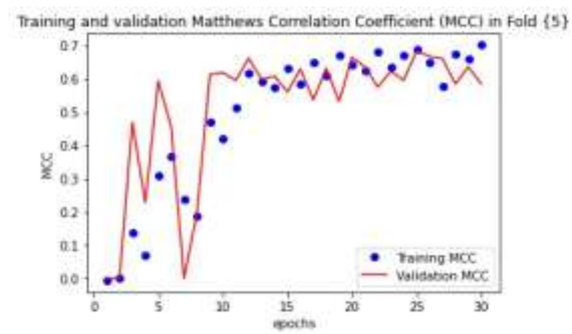

Figure 19: Training and validation of Matthews Correlation Coefficient (MCC) in Fold 5.

5. Testing and Comparison

After the model_hybird has been trained and in each fold model weights have been saved, the result of training is five files in the .h extension. They have been produced to use in testing the model. Model_hybrid has been classified 990 of 20,336 faulty lines. In figure 20 a bar chart of the distribution of the classifier is shown as a percentage.

Table 1: The comparison between three designed AI models

Comparison parameters	***Model_160***	***Model_320***	***Model_hybird (Proposed Model)***
Max training accuracy	0.96	0.96	0.975
Max training Validation	0.96	0.97	0.975
Max training Loss	0.1	0.1	0.1
Max Validation Loss	0.1	0.1	0.1
MCC	**0.75**	**0.71**	**0.79**
No.of detected faulty Lines	1023	1047	990
Model training Time	7 Hours	10 Hours	20 Hours
No.of trainable parameters	363,873	364,033	727,777

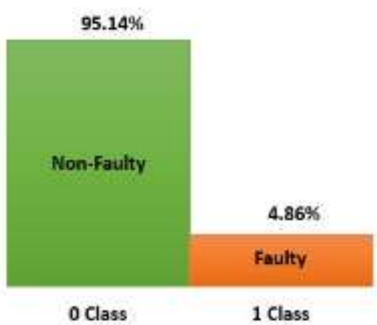

Figure 20: The bar chart of the output distribution of the model_hybrid.

Table 2: Performance comparison between different AI methods [25]

Method with contributions	***MCC***	***Accuracy***
Kaggle Winner	0.450	0.922
LSTM-Attention	0.387	0.913
DNN with Peak features	0.258	0.873
CNN with frequency features	0.293	0.917
Frequency-based Multi-Task learning with Attention Mechanism	0.433	0.961
Article proposed mode	**0.79**	**0.975**

6. Conclusion

In this article, the problem of partial discharge, the importance of studying this phenomenon, and the benefits of designing the AI model to detect this problem were discussed.

In addition to this, a Deep Neural network-based system design has been made for a power line fault detection system. The novelty of this article is combining 160_model and 320_model which have been created using BiLSTM network architecture for the detection of partial discharges before any permanent damage occurs. These three models were trained at the same hyper-parameters. The Model_hybrid is used as the accountable model to detect the PD because it has the best MCC value although it took much time in training. Hence, Relying on the transfer learning technique, this trained model can be used to solve a problem related to PD with the form and nature of the datasets being different, for instance. MCC of model_160 and model_320 reached 0.71 and 0.75, respectively. In addition to this, the MCC of model_hybrid reached 0.79.

7. Acknowledgments

The authors cordially thank the Sakarya University of Applied Sciences in Turkey for supporting this work.

8. References

[1] Modawy, Duaa Elgassim Abbas. Fault Levels in Transmission Substations Level 110 KV. Diss. Sudan University of Science and Technology, 2018.

[2] Kessler, Ole. "The importance of partial discharge testing: PD testing has proven to be a very reliable method for detecting defects in the insulation system of electrical equipment and for assessing the risk of failure." IEEE Power and Energy Magazine 18.2 (2020): 62-65.

[3] Url-1<https://eepower.com/technical-articles/what-is-partial-ischarge/#>, date retrieved 29.07.2022.

[4] Thakor,K.and shah,P.(2020),Partial Discharge Detection Techniques: A Review.'International Research Journal of Engineering and Technology (IRJET),7,(10)PP 885 - 890.

[5] Densley, Chapter 11 in Engineering Dielectrics, Vol. 1, “Corona measurement and interpretation”, R. Bartnikas and E. J. McMahon, Editors, STP 669, ASTM, Philadelphia/West Conshohocken, 1979.

[6] Url-2 < https://www.fluke.com/en-us/learn/blog/leak-detection/partial-discharge>, date retrieved 29.07.2022.

[7] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." Neural networks 61 (2015): 85-117.

[8] Sarle, Warren S. "Neural networks and statistical models." (1994).

[9] Url-3 < https://viso.ai/deep-learning/deep-neural-network-three-popular-types/>, date retrieved 29.07.2022.

[10] ATALI, G., PehlIvan, İ., Gürevin, B., \ ŞEKER, H. İ. (2021). Chaos in metaheuristic based artificial intelligence algorithms: a short review. Turkish Journal of Electrical Engineering and Computer Sciences, 29(3), 1354-1367.

[11] A. Kellen, ”Investigation in Scandinavian Countries of Partial Discharges,” ERA, IB/2185,1963.

[12] R. LaKoste, “Review, Progress, and Trends in Partial Discharges in France,” ERA, IB/2187,1963.

[13] A. Austen and W. Hackett, “Internal Discharges in Dielectrics - Observation and Analysis,” IEE, Vol. 91, No. 44,1944.

[14] C. Boyles and R. Hinton, ”Seven Years of Corona Testing,”.

[15] Nattrass, David A. "Partial discharge. XVII. The early history of partial discharge research." IEEE Electrical Insulation Magazine 9.4 (1993): 27-31.

[16] Wenrong, Si, et al. "Digital detection, grouping and classification of partial discharge signals at DC voltage." IEEE Transactions on Dielectrics and Electrical Insulation 15.6 (2008): 1663-1674.

[17] Acoustic Emission Detection of Partial Discharges in Power Transformers, EPRI Final Report no. EL-4009, 1985.

[18] Eleftherion, Peter M. "Partial discharge. XXI. Acoustic emission based PD source location in transformers." IEEE Electrical Insulation Magazine 11.6 (1995): 22-26.

[19] Morshuis, Peter HF. "Degradation of solid dielectrics due to internal partial discharge: some thoughts on progress made and where to go now." IEEE Transactions on Dielectrics and Electrical Insulation 12.5 (2005): 905-913.

[20] Akbal, Bahadır. "Applications of artificial intelligence and hybrid neural network methods with new bonding method to prevent electroshock risk and insulation faults in high-voltage underground cable lines." Neural Computing and Applications 29.8 (2018): 97-105.

[21] Nyanteh, Yaw, et al. "Application of artificial intelligence to real-time fault detection in permanent-magnet synchronous machines." IEEE transactions on industry applications 49.3 (2013): 1205-1214.

[22] Wang, Miaoyiquan, Weiguo Tong, and Shibo Liu. "Fault detection for power line based on convolution neural network." Proceedings of the 2017 International Conference on Deep Learning Technologies. 2017.

[23] Ghaderi, Amin, Herbert L. Ginn III, and Hossein Ali Mohammadpour. "High impedance fault detection: A review." Electric Power Systems Research 143 (2017): 376-388.

[24] Wang, Shiyuan, and Payman Dehghanian. "On the use of artificial intelligence for high impedance fault detection and electrical safety." IEEE Transactions on Industry Applications 56.6 (2020): 7208-7216.

[25] Tehrani, P., \& Levorato, M. (2020, November). Frequency-based multi task learning with attention mechanism for fault detection in power systems. In 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) , Tempe, AZ, USA ,(pp. 1-6).

[26] Ur1\5<https://www.sciencedirect.com/science/article/abs/pii/ S0960077915002817>, date retrieved 29.07.2022.

[27] Url\-6 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7516527/>, date retrieved 29.07.2022.

[28] Albers; Alexanderson (2008). "Benoit Mandelbrot: In his own words". Mathematical people : profiles and interviews. AK Peters. p. 214. ISBN 978-1-56881-340-0.

[29] Goh, C., Hamadicharef, B., Henderson, G. T., \& Ifeachor, E. C. (2005, June). Comparison of fractal dimension algorithms for the computation of EEG biomarkers for dementia. In 2nd international conference on computational intelligence in medicine and healthcare (CIMED2005), Portugal.

[30] Jurman, Giuseppe, Samantha Riccadonna, and Cesare Furlanello. "A comparison of MCC and CEN error measures in multi-class prediction." (2012): e41882.