IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 July 2024, accepted 20 August 2024, date of publication 23 August 2024, date of current version 2 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3449096

RESEARCH ARTICLE

A New Approach Based on Deep Features of Convolutional Neural Networks for Partial Discharge Detection in Power Systems

BELKIS ERISTI, (Member, IEEE)
Department of Electrical and Energy, Vocational School of Technical Sciences, Mersin University, 33110 Mersin, Türkiye
e-mail: beristi@mersin.edu.tr

ABSTRACT Partial discharge (PD) faults often occur as a result of breakdowns in the insulation layer of insulated overhead conductors. PD faults can cause serious problems such as power outages or electrical fire accidents. In this paper, a new PD detection system based on spectral analysis, spectrogram analysis, deep learning algorithms, minimum redundancy–maximum relevance (mRMR) and ensemble machine learning (EML) is presented. In the process of extracting distinctive features in the frequency dimension, 1D spectral data and 2D spectrum data based on frequency-time are obtained from a raw PD signal. Fourier transform-based three-power spectral density analyzes and one spectrogram analysis are performed. Deep features are obtained by using pre-trained 1D convolutional neural network models for 1D spectral data and the pre-trained ResNet-50 model for 2D spectrogram data. The most effective features are determined by applying mRMR feature selection analysis to the obtained deep features. In the last stage, PD detection is performed by applying the selected deep features to the EML classifier. The performance of the proposed PD detection system are evaluated with the VSB common data set. According to the experimental results, the proposed deep feature approach based PD detection system has very high performance.

INDEX TERMS Convolutional neural network, ensemble machine learning, minimum redundancy–maximum relevance, partial discharge, spectral analysis, spectrogram analysis.

I. INTRODUCTION

In recent years, because power line faults in electrical power systems cause serious problems to the power system and end-of-line users, tremendous research has been seen on power line fault detection and solutions have been developed in the industry based on this research. Partial discharge (PD) which is a result of partial dielectric breakdown of an insulator is a common type of fault in transmission lines. PDs are also very challenging to detect because they typically appear as pulses of much shorter duration than 1 microsecond [1], [2]. In electrical power systems, covered conductors are widely used due to their higher operational reliability in forested or dissected terrain areas [3], [4]. However, when tree branches come into contact with covered conductors, PD can occur with a localized electrical discharge that only partially bridges the insulation [5]. In this case, the insulation material may damage and cause a short circuit. As a result of PD, high-frequency transient pulses occur in the voltage signal. This may damage devices and equipment, and the entire power system becomes less reliable and maintenance costs increase [6]. Therefore, reliable and fast PD detection in power systems is very important both to prevent equipment damage and service interruptions and to increase the performance, stability and reliability of the power system.

A. RELATED WORKS

Based on various theoretical studies, faulty signals in power systems are only 5% of normal signals [7]. One of the most effective approaches to detecting faulty signals is to compare the normal signal with the faulty signal. It is very common to use artificial intelligence (AI) approaches in this field.

The associate editor coordinating the review of this manuscript and approving it for publication was Binit Lukose.

TABLE 1. Features, challenges and future works of conventional studies on PD detection.

Study	Methodology	Features	Challenges (Ch) Future works (Fw)
[6]	Ensemble Deep Learning	- Using three deep learning methods - Analyzing the method used with different scenarios.	**Ch:** - The calculations of method are numerous and complex. - The resources used and time spent in the method are quite high. **Fw:** - Parallel computing methods can be developed to reduce computational complexity. - Fault localization can be detected .
[7]	PCA, SVM	- Chunk based feature extraction.	**Ch:** Optimal number of chunks. **Fw:** Deep learning based classifier algorithm.
[9]	CNN-LSTM	- High accuracy. - Training and testing time is short.	**Ch:** - The algorithm used is complex. - Resource usage is high.
[11]	Random forest	- Improved noise removal method	**Ch:** - The data obtained has a low sampling rate - SOMA optimization may cause overfitting problems. **Fw:** The overfitting problem can be eliminated.
[14]	LSTM	- Time decomposition techniques are used to remove the noise portion from the raw signals.	**Ch:** Low accuracy. **Fw:** The accuracy rate can be increased with different algorithms and techniques.
[15]	LSTM	- High accuracy. - Low computational complexity. - Noise filtering. - High generalization ability.	**Ch:** High resource usage
[17]	CNN	- Multiple noise levels are used. - More than one training rate is used. - A hybrid optimization approach is used.	**Ch:** There is transaction complexity.
[18]	CNN-LSTM	- Noise reduced. - LSTM network combined with attention mechanism is used. - Convergence speed is high	**Fw:** - The issue of imbalance between different data set types. - The detection method of multiple algorithm combinations.

In AI techniques, detection processes are carried out using the binary classifier structure, including normal signal and faulty signal. Although faulty signal detection using AI has been widely researched in recent years, it has also been reported that there is a research gap regarding validation and automation of the proposed system [6], [8]. In order to eliminate this gap, it has been seen that there have been many effective studies on AI-based detection of PD faults. Some challenges are encountered in the detection of PDs using AI, such as the dataset imbalance, the presence of background noise, and the large data size [9]. AI-based methods for PD detection are mainly divided into shallow machine learning methods and deep learning methods. When detecting PDs with either of these two methods, a preprocessing stage is generally needed before the classifier. In the preprocessing stages, analyzes such as noise filtering, peak extraction, feature extraction and data size reduction are generally performed to increase the performance of the classifier. Some shallow machine learning models such as support vector machine (SVM) [7], ensemble machine learning (EML) [10], random forest [11], artificial neural network (ANN) [12] have been used in PD detection. Because shallow learning methods cannot capture high-dimensional nonlinearity in classification approaches, deep learning approaches, which have features such as multiple non-linear layers, capturing high-dimensional non-linearity and complex correlation, are used more effectively [13]. It has been seen that deep learning approaches such as recurrent neural network (RNN) [1], long short-term memory network (LSTM) [14], [15], convolutional neural network (CNN) [16], [17], [18] have been widely used in PD detection recently. In general, the Technical University of Ostrava (VSB) dataset [19] consisting of real data is used to evaluate the performance of these modern and innovative PD detection systems.

In [1], a series of preprocessing steps are introduced to PD signals, including phase alignment, signal smoothing, and noise estimation. Then, the obtained global-scale and local-scale features are applied to the RNN classifier and the detection process is carried out. In [5], a CNN-based PD detection system that uses data obtained by filtering low frequencies and extracting pulses is presented. In [6], optimal features and hyperparameters are selected by using the double particle swarm optimization in the pre-training stage. In the classification process, a bagging ensemble system is deployed on three different deep learning architectures: deep neural networks, LSTM-RNN and deep belief networks. In [10], a PD detection approach based on wavelet packet transform, ReliefF feature selection approach and EML classifier is proposed. In [14], after applying a time series decomposition technique to each signal, PD detection is performed via an LSTM classifier. In [15], first, discrete wavelet transform is applied to signals for noise filtering. Then, statistical and entropy feature vectors of each signal are created. In the final stage, an LSTM network is proposed for PD detection. In [17], principal component analysis (PCA)-based dimensionality reduction and four technical indicators-based

feature extraction are performed. Then, all features are provided as input to an optimized CNN. In [18], using the 1D CNN structure, multidimensional features that reflect the complex dynamic changes of PD signals are extracted. Then, the obtained features are used as the input of the LSTM network. In [20], an approach combining autoencoder with data corrections using a specially designed loss function is applied to PD data. Table 1 shows the advantages and disadvantages of different above studies focusing on PD detection strategy.

B. CONTRIBUTIONS

In the literature section discussed in subsection I-A, it is clearly seen that the proposed algorithms in deep learning-based PD detection still have some shortcomings in terms of performance, validation, reliability and automation. There is a serious need to develop more effective deep learning-based PD detection systems due to the efficient, reliable and high-performance features of the deep learning algorithm. Therefore, this paper focuses on developing an innovative algorithm based on deep features for PD detection. For this purpose, four frequency-time data are obtained from a raw PD signal, since the most important information that distinguishes the PD signal is frequency information. In these processes, three separate 1D data are obtained for a PD data by Fourier transform-based power spectral density (PSD) analyzes that Welch, periodogram and multitaper. In addition, 2D time-frequency (TF) spectrum data is obtained with Fourier-based spectrogram analysis for the same PD data. Deep features are obtained by applying the spectral data to three pre-trained 1D CNN models, and the spectrogram data to a pre-trained ResNet-50 model. The most effective features among these deep features are determined as a result of the mRMR feature selection analysis. Thus, in the last stage, PD detection is performed by applying the selected deep features EML-based ensemble subspace k-nearest neighbor (kNN), SVM and decision tree (DT) algorithms. The aim of this paper is to develop an effective PD detection system based on deep feature. The developed deep feature-based system includes both 1-D and 2-D deep learning algorithms. According to the results obtained, it is clearly seen that the proposed PD detection system has a high-performance, fast and reliable structure that does not require noise filtering. The main contributions of this paper are summarized as follows.

- A deep learning-based system with fast, high accuracy and effective performance is proposed for the detection of PD signals. This method consists of combining Fourier-based transforms, CNN architectures, mRMR feature selection and shallow machine learning models.
- Deep feature technology is proposed to precisely reveal PD features from high frequency components. This technology can increase model performance in classifying signals containing high frequency components with artificial intelligence methods.
- PSD and spectrogram analyzes are used to solve the problem of high noise and high data size and to obtain effective features. Using these analyzes together will be an ideal approach to solve classification and detection problems.
- mRMR feature selection approach is proposed to both increase classification performance and obtain lower data size.

C. PAPER ORGANIZATION

The remainder of the paper is organized as follows. Section II describes the VSB dataset used to test the performance of the proposed detection system in this paper. Section III presents all the concepts and methods required in the design of the proposed detection system. The details of the proposed deep feature-based PD detection system are described in Section IV. Section V presents the experiments, results and discussion. Section VI concludes the paper. Table 2 presents nomenclature.

TABLE 2. Nomenclature.

ACC	Accuracy	SEN	Sensitivity
AI	Artificial intelligence	SPE	Specificity
ANN	Artificial neural network	SVM	Support vector machine
CNN	Convolutional neural network	TF	Time-frequency
DT	Decision tree	TN	True negative
EML	Ensemble machine learning	TP	True positive
ENET	Energy and Environmental Technology	VSB	Technical university of Ostrava
FN	False negative	A	Average of v
FP	False Positive	I	Mutual information
kNN	k-nearest neighbor	K	Number of filters
LSTM	long short-term memory network	M	Number of signal samples
MCC	Matthews correlation coefficient	P	Periodogram PSD
mRMR	Minimum redundancy–maximum relevance	$Pwelch$	Welch PSD
PCA	Principal component analysis	S	Number of segments
PD	Partial discharge	Sf	Subset of features
PRE	Precision	v	Data-window
PSD	Power spectral density	w	Frequency of PSD
RNN	Recurrent neural network	y	Spectrum of signal

II. VSB DATASET

Validation of proposed detection systems in machine learning applications is a cornerstone for evaluating system performance. Evaluation of proposed PD detection systems using simulated data is ineffective and inadequate. Thus, there is a need to use real-time data [7]. To overcome this drawback, VSB published a dataset of PD data obtained from medium voltage overhead power lines on Kaggle, the world's largest data science collaboration platform, in 2018 [19]. In the Energy and Environmental Technology (ENET) center at VSB, the PD detection system was designed with a new measuring device developed on three-phase 50 Hz overhead power lines. By installing this meter device on different power lines located in forested areas, a PD dataset called the VSB dataset or ENET dataset was created [21].

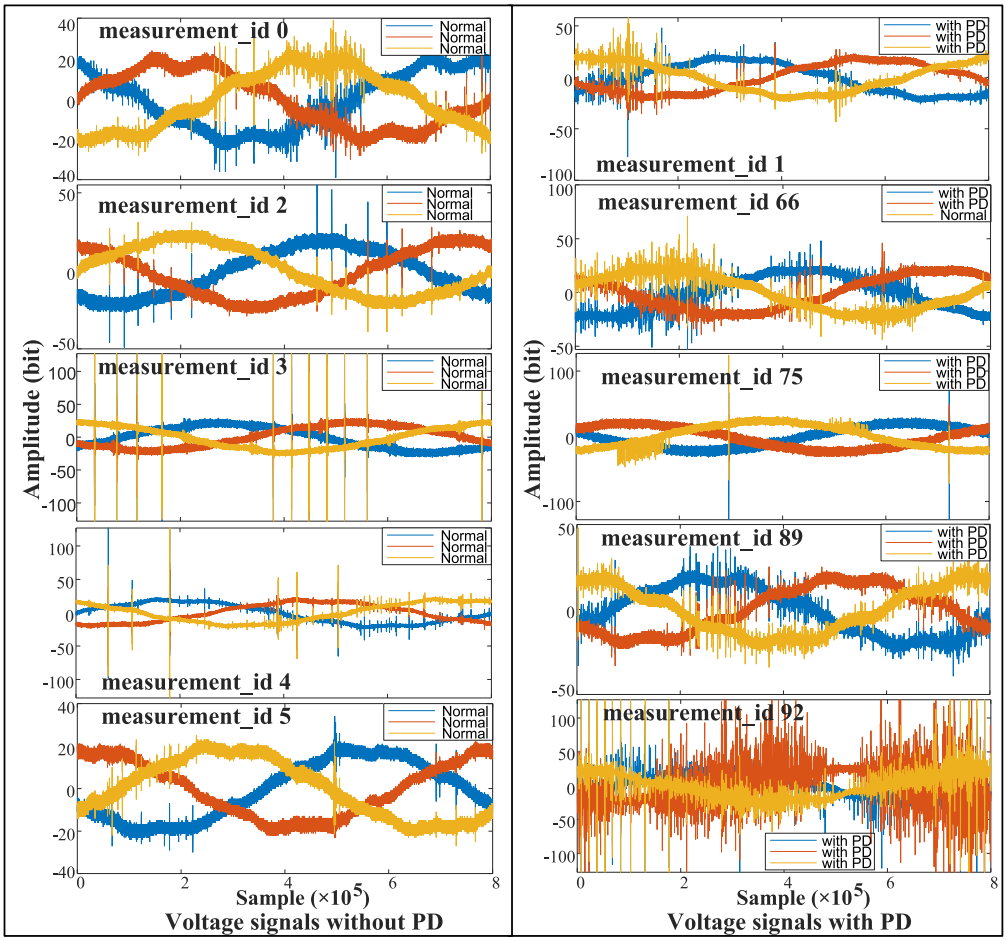

FIGURE 1. Sample signals from VSB dataset.

Each data in the VSB dataset has a sampling frequency of 40 MHz, with 800000 sample points. The total number of samples with normal (0) and faulty (1) class labels in the VSB data set is 8712. In the VSB dataset consisting of real data, there are 8712 labeled voltage signals, of which 8187 are labelled as Non-PD and 525 are labelled as PD. In the measuring device designed to obtain PD data, a single-layer inductor is used as a voltage sensor. This inductor is mounted on the surface of separately covered conductor for each phase. Also the inductors are connected to a capacitance divider. There is a data acquisition card at the capacitance divider output. A detailed description of this measurement system is given in [22]. Table 3 shows the label information of the data in the VSB dataset. In this table, signal_id represents the number of each voltage signal. measurement_id represents the number of a group of three-phase signals. Phase indicates which phase the signal belongs to (0-A, 1-B, 2-C). Target is label of the signal. Fault type depicts whether the three-phase signal is normal or which phases are PD. For example, all three-phase voltage signals in the measurement_id 0 are normal (non-PD), while all three-phase voltage signals in the measurement_id 1 contain PD. Figure 1 presents five randomly selected three-phase signals with PD and non-PD from the VSB dataset. As can be seen in Figure 1, VSB dataset signals have high noise values. Besides, it is very difficult to distinguish PD signals visually.

III. MATERIAL AND METHODS

This section introduces all materials and methods involved in the proposed PD detection system before proceeding on the design of the proposed system in section IV. These materials and methods are: spectral analysis, spectrogram analysis, CNN, mRMR feature selection and shallow machine learning methods.

A. SPECTRAL ANALYSIS

Spectral analysis is a method that displays the frequency distributions of a random and finite-length signal and reveals the repetitive and hidden behavior of the signal [23], [24]. PSD is defined as the energy change in a signal distributed over the measured frequencies [25]. In this paper, periodogram, Welch and multitaper spectral analyzes are used to extract the PSD of raw VSB signals.

1) PERIODOGRAM METHOD

The periodogram method used to find periodicities in a signal is proportional to the square magnitude of the Fourier transform of the product of a windowed time series [26].

TABLE 3. VSB dataset labels.

signal_id	measurement_id	Phase	Target	Fault type
0	0	0 (A)	0	no fault (normal)
1	0	1 (B)	0	
2	0	2 (C)	0	
3	1	0 (A)	1	ABC
4	1	1 (B)	1	
5	1	2 (C)	1	
6	2	0 (A)	0	no fault (normal)
7	2	1 (B)	0	
8	2	2 (C)	0	
...	...	...	...	...
201	66	0 (A)	1	AB
202	66	1 (B)	1	
203	66	2 (C)	0	
...	...	...	...	...
8709	2903	0 (A)	0	no fault (normal)
8710	2903	1 (B)	0	
8711	2903	2 (C)	0	

The equation of periodogram, the first spectrum analysis technique based on Fourier analysis, is as follows,

$$P_{(w)} = \frac{1}{M}\left|\sum_{n=1}^{M} y\,(n)\,e^{-jwn}\right|^2 \tag{1}$$

where P is the periodogram PSD, w denotes the frequency of PSD, M refers to the number of signal samples and y is spectrum of signal [24].

2) WELCH METHOD

The Welch method, a modified version of the periodogram, calculates PSD by averaging small windows of the periodogram at each frequency of the signal [27]. Thus, the noise effect is reduced. The welch's PSD equation is given as follows,

$$P_{Welch}\,(w) = \frac{1}{S}\sum_{i=1}^{S} P_i\,(w) \tag{2}$$

and

$$P_i\,(w) = \frac{1}{M}\frac{1}{A}\left|\sum_{n=1}^{M} v\,(n)\,y\,(n)\,e^{-jwn}\right|^2 \tag{3}$$

where P_{welch} is the Welch PSD, P_i is the periodogram PSD of ith segment, S refers the number of segments, $v(n)$ denotes the data-window and A is average of $v(n)$.

3) MULTITAPER METHOD

The multitaper method is a weighted periodogram average by a collection approach of mutually orthogonal windows. The variance and deviation of PSD in the multitaper method are less compared with other spectral methods [28]. The multitaper PSD is shown as follow,

$$\hat{\vartheta}\,(w) = \left|\sum_{n=1}^{K} h_{K-n} y\,(n)\,e^{-jwn}\right|^2 \tag{4}$$

where K is the number of filters and hK-n is the filter impulse.

B. SPECTROGRAM ANALYSIS

A spectrogram is a 2D visual representation of a signal that shows frequency spectrum changes over time [29]. In spectrogram analysis, the signal is first divided into equal overlapping sections of length n. Secondly, windowing is performed for each equal segment. Then, sequential Fast Fourier transform analysis is performed for each segment. Finally, the power of each spectrum is displayed as a segment-by-segment image [30]. In this paper, 2D TF spectrum data is obtained by performing spectrogram analysis for each VSB data. Thus, frequency distribution information in each image provides important distinctive information for the PD detection algorithm.

C. CONVOLUTIONAL NEURAL NETWORK

The CNN is a deep learning architecture that has been used in recent years for classification, segmentation, object detection and localization of image or signal types [31]. Besides, CNN architectures are commonly used as feature extractor to obtain distinctive features of these types. The basic layers in a CNN architecture are the convolution layer, pooling layer and fully connected layer. The convolution layer calculates the receptive field of the image/feature map in the last layer by doing the dot product of the kernels called filters for feature extraction. In the pooling layer, a pooling windowing process is applied to the receptive regions of the feature maps and the mean/maximum values are extracted to reduce the parameters. The fully connected layer is used to connect neurons in one layer to neurons in the next layer. There are different CNN architectures with more layers in the literature. The proposed CNN architectures are designed to meet expectations such as input data type, working structure, results, runtime and performance. This work attempts to extract deep PD features using 1D CNN architecture and ResNet-50 model.

1) CONVOLUTIONAL NEURAL NETWORK-1D

Spectral data obtained from VSB dataset is one-dimensional. In this work, a nine-layer 1D deep CNN architecture created according to the experimental situation is used to obtain deep features of the spectral data. A graphical description of the 1D deep CNN architecture is shown in Figure 2. The input layer of the 1D CNN architecture is a vector of size 16385. The hidden layer consists of 2 convolutional blocks, 1 pooling block, 2 normalization blocks, 2 ReLU blocks and 1 fully connected block. The last output layer is a softmax block for a 2-class output vector.

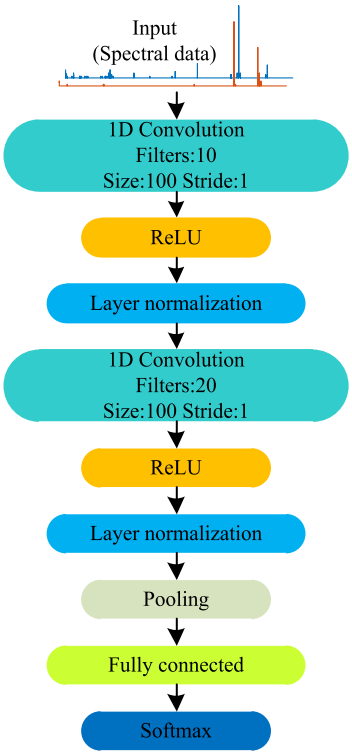

FIGURE 2. 1D deep CNN architecture for spectral data.

2) ResNet-50 MODEL

Although deep learning networks have been an effective intelligent learning algorithm in recent years, they contain some fundamental problems such as complexity in training, significant training fault and vanishing gradients. Residual networks are a type of deep learning algorithm with a network-in-network architecture structure that overcomes the problem of vanishing gradients [32]. In this work, ResNet-50 architecture, a type of residual network, is used to obtain deep features from spectrogram image data. ResNet-50 is a 50-layer CNN algorithm consisting of 48 convolution blocks, a MaxPool block and an average pool block. The general structure of ResNet-50 model is shown in Figure 3.

D. MINIMUM REDUNDANCY – MAXIMUM RELEVANCE

In classification problems, input data set features may have quite large data sizes. These high-dimensional features increase the computational costs of the classifier and may also reduce classification accuracy [33]. The mRMR feature selection approach, which aims to reduce the correlation between each feature, is an effective approach that minimizes the computational cost in classification problems while also increasing accuracy [34]. In the mRMR method, while searching the features that have the most correlation with the final output target variable within the feature set, it is also aimed to find the feature set that has the least correlation with each other. Max-Relevance is expressed as,

$$\max D\left(Sf, c\right), D = \frac{1}{|Sf|}\sum_{x_i \in Sf} I\left(x_i, c\right) \tag{5}$$

where Sf is the subset of features x_i and $I(xi,c)$ is the mutual information between features and class. Min-Redundancy is defined as,

$$\min R\left(Sf\right), R = \frac{1}{|Sf|^2}\sum_{x_i, x_j \in Sf} I\left(x_i, x_j\right) \tag{6}$$

where $I(x_i, x_j)$ is the mutual information between features and features. The mRMR score information can be defined as,

$$\max \Phi\left(D, R\right), \Phi = D - R \tag{7}$$

where $\Phi(D,R)$ is an operator that combines D and R by optimizing them simultaneously [33].

E. MACHINE LEARNING METHODS

In this work, the classification process is carried out using EML based subspace kNN. Besides, detection results are obtained using SVM and DT classifiers for comparison and evaluation. In this subsection, brief explanations and definitions of these classifiers are given.

Ensemble learning is a popular machine learning technique that improves prediction performance by combining multiple models [35]. The various EML classifiers include bagged trees, ensemble bagged trees, ensemble boosted trees, ensemble subspace kNN, ensemble subspace discriminant, and ensemble RUSBoosted trees. Among these models in this paper, the EML classifier based on the ensemble subspace kNN algorithm is used, which achieved successful results. Ensemble subspace kNN algorithm is an EML model that fits different feature subsets of the same dataset. The final step of the ensemble subspace kNN is the mode for classification and mean for regression of the class labels predicted by the model applied to different subsets of data features [36].

SVM is a popular learning algorithm proposed by Cortes and Vapnik that separates multidimensional data classes using a hyperplane [37]. In SVM, the extreme cases of each data class, called support vectors, determine the hyperplane. SVM also finds the hyperplane that maximizes the distance between support vectors of two different classes. Kernel functions are needed for nonlinear applications in SVM. Thus, the data in the nonlinear input space is linearized in the high-dimensional feature space via the kernel function [38]. Although SVM has a binary classification structure, it is successfully used in multi-class problems by adopting a series of binary SVMs. Methods such as one-to-one, one-to-one and directed acyclic graph SVM are the most commonly used methods in multi-class SVM [39].

DT can be defined as a tree-shaped modeling of a series of hierarchical decisions with a simple structure in solving classification and prediction problems. In DT, trees have their root nodes at the top where they are divided into branches according to each specified condition. Each branch may also have sub-branches depending on more specific sub-feature conditions. The splitting of branches is terminated when the final classification decision is reached, and this final stage is defined as the leaf node [40].

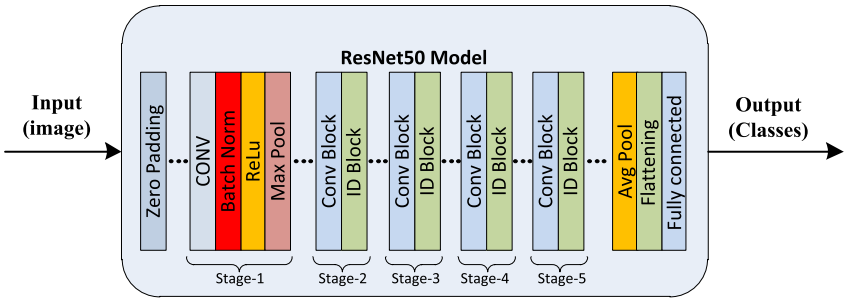

FIGURE 3. The general structure of ResNet-50 model.

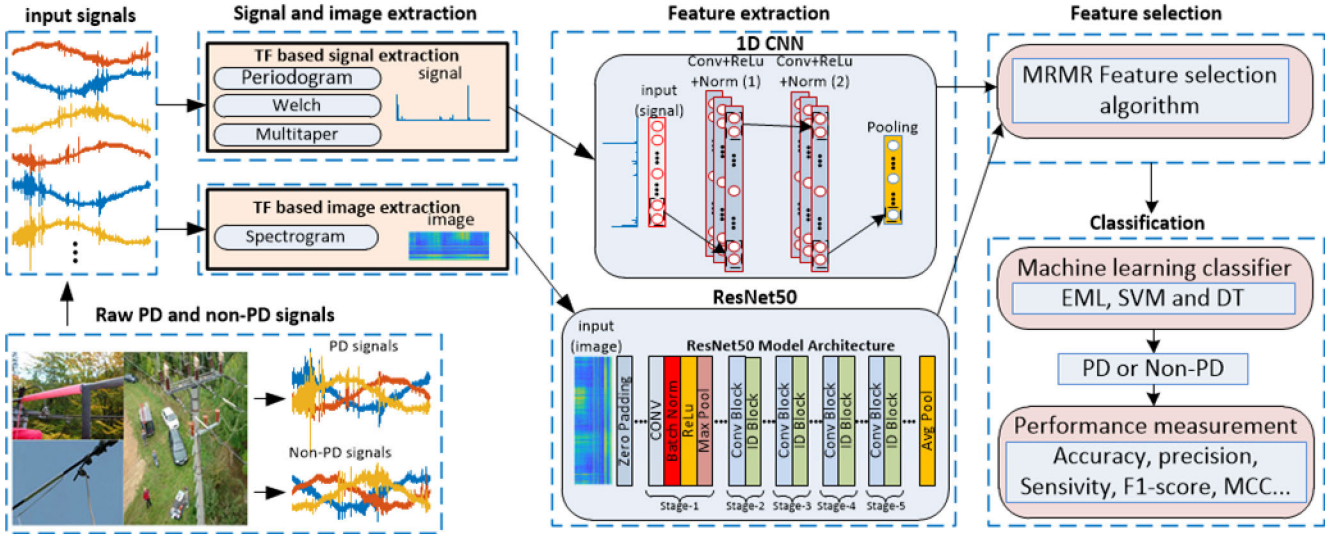

FIGURE 4. The block diagram of the proposed deep feature-based PD detection system.

IV. THE PROPOSED DEEP FEATURE-BASED PD DETECTION SYSTEM

The main task of the proposed deep feature-based detection system in this paper is to identify the PDs occurring in high voltage power cables. PD signals containing high frequency components naturally have a very high raw voltage data. Additionally, PD signals contain different levels of noise components. Thus, the proposed deep feature-based detection system has an effective algorithm that can overcome these difficulties. Two key aspects of the proposed detection system make it significantly superior to similar state-of-the-art works showing an outstanding detection performance. The first key step is to obtain distinctive features by applying deep learning approaches to the time-frequency changes of the raw PD signal. The other step is to find the effective features among the obtained deep features. Initially, the input signal data collected by ENET center from the power cable is applied the proposed algorithm. Deep features are extracted based on time-frequency information about the waveforms of PD pulses by using a series of deep learning algorithms. Thus, the data size is significantly reduced and distinctive features are obtained. mRMR feature selection approach is applied to all features and very effective features are obtained for the final step that classifier step. As a result of mRMR feature selection process, the feature data size is further reduced and a feature set having high classification performance is created. The final step of the proposed system is binary-based classification processes using the features obtained for the data input signals. Figure 4 shows the steps of the working-flow of the proposed deep feature-based detection system. The proposed detection system in this paper consists of four consecutive stages: signal and image extraction, feature extraction, feature selection and classification.

A. SIGNAL AND IMAGE EXTRACTION

In the machine learning-based algorithms, the dataset presented to the machine learning method should be ready for use without the need for further work [7]. When the VSB dataset is directly applied to the machine learning method, some challenges may be encountered. These challenges are generally that the dataset consists of raw data, signals contain noise, data size in each signal is very high, and PD features are present in high frequency components. The first three stages of the proposed detection system are designed to overcome these challenges. In the signal and image extraction stage, each VSB data consisting of time-amplitude information is converted into four different new data containing time-frequency information obtained by Fourier-based

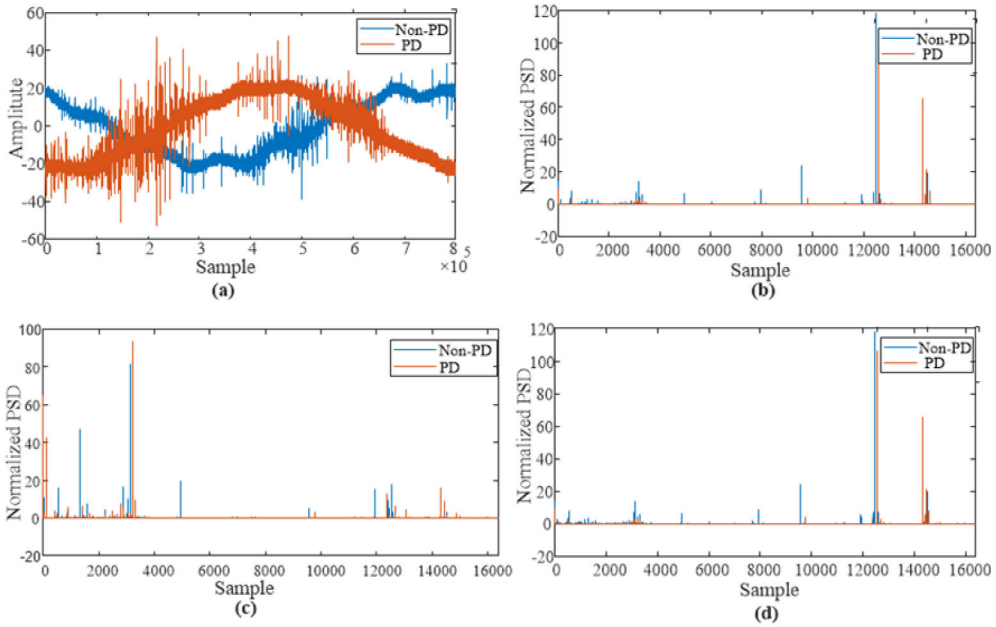

FIGURE 5. Graphics of a non-PD and a PD signal: a) raw data; b) PSD graph with Welch method, c) PSD graph with periodogram method and d) PSD graph with multitaper method.

analyzes. As a result of these analyzes, each VSB signal with high data size and unclear underlying characteristics is converted into more distinctive frequency-based image and signals with reduced data size and more distinct PD features. Thus, in the next deep feature-based feature extraction stage, it is possible to quickly extract effective features from each image and signal data representing the PD signal.

In the signal extraction process, Welch-, periodogram- and multitaper-based PSD analyzes are performed for the raw all data, respectively. Each VSB data is converted into a new data containing 1-D PSD information. Hamming window is used in periodogram and Welch PSD analyses. For the multitaper, windowing is not required. Taper type in the multitaper analysis is specified as 'sine'. In all PSD analyses, DFT points is selected as 2^{15} (32768). There are $2^{15}/2+1$ (16385) points in the each PSD estimation. Thus, the VSB dataset data of 1×800000 size is reduced to 1×16385 data size, which is more suitable for 1-D CNN input. The z-score based normalization approach is implemented to the each PSD estimation. Figure 5 shows the PSD changes obtained for a non-PD and a PD signal.

In the image extraction process, all VSB signals are converted into image-based TF spectrums that represent the power distribution between frequencies as a result of spectrogram analysis processes. Thus, in the obtained spectrum image data, the x-axis represents time and the y-axis represents frequency. The color or intensity of the image pixels in the graph represents the intensity of frequency changes in the signal. In other words, a TF spectrum is a visual representation of how the frequency content of a signal changes with time. Figure 6 shows the TF spectrums of the PD and non-PD signals in the 0-20 MHz range (20 MHz is Nyquist frequency of a VSB signal) given in Figure 5.a. As can be seen from Figure 6, PDs in TF spectrums cannot be distinguished by the human eye due to noise and are not suitable for expert PD identification. However, deep learning algorithms are capable of learning with such fine details.

In this paper, in order to increase the learning ability of the ResNet-50 architecture, we focused on frequencies above 100 kHz, where the PD frequency features of each signal are more intense. As a result of some detection analyses, the best detection performance is obtained for TF spectrums obtained in the 500 kHz-20 MHz range. Thus, each VSB signal is converted into TF spectrums in the range of 500 kHz-20 MHz. Figure 7 shows the TF spectrums in the range of 500 kHz-20 MHz for the same signals given in Figure 6. As seen in Figure 7, PDs cannot be distinguished visually, but high frequency changes in TF spectrums can be seen in more detail. As mentioned in the previous sections, the difficulties encountered in detecting PDs in the VSB dataset continue to exist in the TF spectrum graphs. It is not possible to identify a trace representing PD in Figure 6 (b) and Figure 7 (b). The proposed PD detection system has a structure that can overcome this challenge. Each TF spectrum image has provisions of 1361×1075 pixels, resolution of 300 dpi, 24-bit depth and RGB.

B. FEATURE EXTRACTION

In the feature extraction stage, distinctive deep learning features are obtained from spectral and spectrogram data. Frequency-amplitude based image and signals obtained from the previous stage are applied to deep learning algorithms. Thus, this stage overcomes the feature extraction limitations encountered in traditional machine learning-based classification workflows, thanks to deep learning-based feature

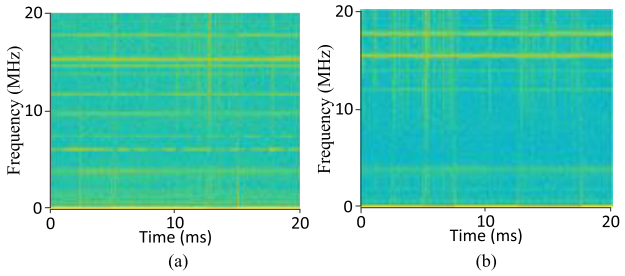

FIGURE 6. TF spectrums of (a) PD signal and (b) non-PD signal in the range 0-20 MHz.

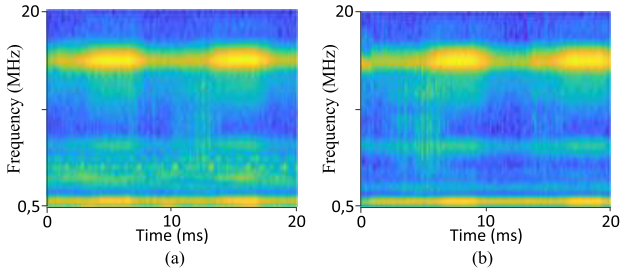

FIGURE 7. TF spectrums of (a) PD signal and (b) non-PD signal in the range 500 kHz-20 MHz.

extractor models that can extract a comprehensive feature set for any PD data. In this paper, 1D CNN architecture for Welch-, periodogram- and multitaper-based TF signals is built with nine layers. ResNet-50 architecture is used to obtain features from 2D spectrogram images. Three pre-trained 1D CNN models are created as a result of training three 1D CNN models using Welch-, periodogram- and multitaper-based PSD data, respectively. After the ResNet-50 model is trained using spectrogram image data, a pre-trained ResNet-50 model is created. Pre-trained models are created using 70% of the data for training and 30% for testing. In all pre-trained models, the learning rate is determined by changing it to 0.001, 0.0001 and 0.00001 during the training phase. The solvent is determined by comparing the results obtained with SGDM or ADAM optimization methods. SGDM optimization method is chosen for ResNet-50 and ADAM optimization method is chosen for all 1D CNN models. Mini-batch size for 1D CNN models and ResNet-50 model is set as 16 and 64, respectively. While the pre-trained ResNet-50 model is trained with 20 epochs and a total of 1900 iterations, the 1D CNN pre-trained models are trained with 15 epochs and a total of 11430 iterations. In the pre-trained 1D CNN, deep features of PD and non PD signals are obtained from the activation data in the output of the pooling layer at level-7. Features of the pre-trained ResNet-50 algorithm are obtained from the activation data in the output of the 2D global average pooling layer at level-173. As a result of this stage, a total of 3×20 and 1×2048 features are obtained from 1D CNN and ResNet-50 architectures, respectively.

C. FEATURE SELECTION

The size of the data obtained after the feature extraction stage is 1×2108. These features may increase computational costs and consequently reduce PD detection accuracy. In the feature selection stage, features with high classification performance are determined and the data size is reduced. Thus, the classification stage exhibits very high performance in terms of speed and efficiency. In the feature selection stage, mRMR is used to remove features that did not increase PD detection performance. After calculating the relevance and redundancy scores for all features, the corresponding mRMR scores are determined. The 1×1 data created with the feature with the best mRMR score is applied to the input of the classification stage. In the classification stage, PD detection performance is determined as a result of training and testing for this 1×1 data. Then, a 1×2 feature dataset is created by adding the feature with the second best mRMR score and the same processes are repeated to obtain PD detection results. These processes are repeated for the last ranked features according to the mRMR score, respectively, and PD detection results are obtained. Finally, the features in the feature set with the highest Matthews correlation coefficient (MCC) coefficient among the detection results are used in the classifier input of the proposed PD detection system. Figure 8 shows the mRMR score values obtained for all features.

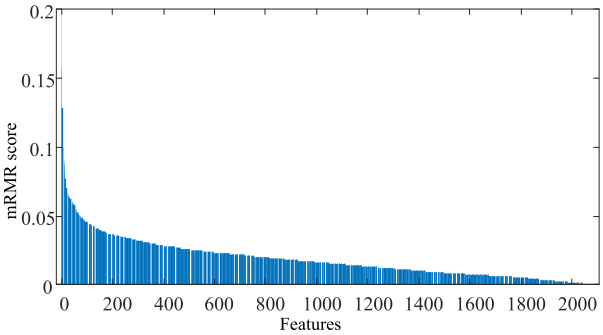

FIGURE 8. mRMR scores of all features.

D. CLASSIFICATION

The final stage of the proposed PD detection system is classification. The output of the classification stage is the decision stage of the proposed PD detection system. The effectiveness of the classification stage can be evaluated with two main situations: the performance of the data in the classifier input and performance of the classifier. As mentioned in the previous stages, features with high classifier performance are obtained. At this stage, classification performances are evaluated using three different classifiers. Thus, as a result of the classification stage, effective PD detection with high performance can be achieved. At this stage, PD detection processes are carried out with the binary classification approach using the ensemble subspace kNN classification method with 5-fold cross-validation. Additionally, PD detection results are obtained using SVM and DT algorithms to make comparisons and evaluations. In the classification stage of the proposed system, classification results are obtained with the ensemble subspace kNN, SVM and DT by using the features selected

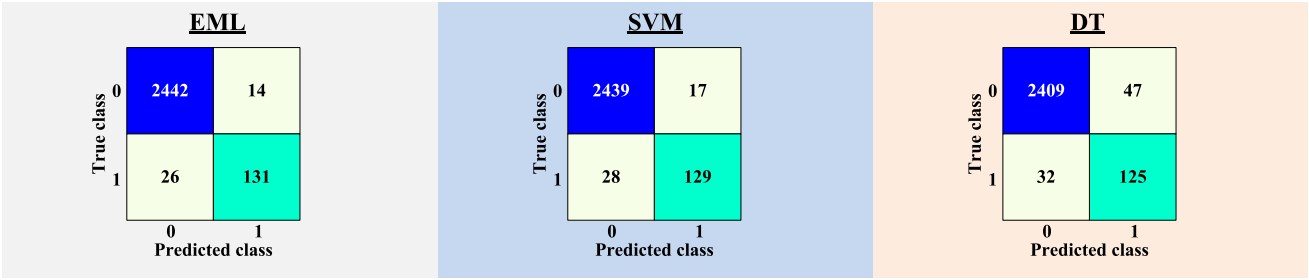

FIGURE 9. Confusion matrices for proposed deep feature-based detection system.

with the mRMR method. MATLAB Classification Learner Toolbox is preferred in all three classification processes.

E. EVALUATION METRICS

In general, the evaluation metric is a measurement tool that measures the performance of classification-based systems [41]. MCC metric is ideal metrics for evaluating proposed algorithms for prediction problems. All values found in the complexity matrix are used in the calculation to evaluate the model performance with the MCC metric [41]. Besides, detection models in the Kaggle competition on the VSB dataset were evaluated according to the MCC metric value, and the MCC metric value of the winning model was 70.6% [1]. Some evaluation metrics, as well as MCC metric, are used to evaluate the performance of the PD detection approach proposed in this paper. Accuracy metric, which is the ratio of total correct examples to total examples, is used to measure the performance of the model. Precision metric, defined as the ratio of true positive predictions to the total number of positive predictions made by the model, is used. Recall metric is used, which is the ratio of the number of true positive samples to the sum of true positive and false negative samples. Specificity metric, also known as true negative rate, is also used to evaluate the proposed PD detection system.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A series of experiments are carried out to evaluate the performance of the proposed deep feature-based PD detection system. All analyzes in this paper are carried out in the Matlab(R) 2021 software environment and on a computer with Intel(R) Core(TM) i7-3770, 3.90 GHz, 32.0 GB RAM, and GeForce GTX 2060 GPU. Since the proposed system is defined as a PD detection system, the results are in the form of binary classification. In classification process, '0' is labeled as non-PD signal and '1' as PD signal. The training data of the VSB dataset, consisting of 8712 data, are used in the training and testing processes of the proposed detection system. In all experiments, randomly selected 70% of the total 8712 data are used for the training of the proposed detection system. The remaining 30% are used in testing process. In the training dataset, 5731 data of the total 6099 data are non-PD data and the remaining 368 data are PD data. In the test dataset, 2456 of the total 2613 data are non-PD data and the remaining 157 data are PD data.

A. DETECTION PERFORMANCE OF THE PROPOSED SYSTEM

First, the performance of the proposed deep feature-based detection system in predicting PDs is investigated. The VSB dataset is an unbalanced dataset due to the small number of PD data. Since MCC is an effective approach in evaluating unbalanced data, MCC performance measurement provides important information when evaluating the prediction performance of the proposed system. In addition to the MCC measurement, other metrics given in section IV-E are also included in the performance evaluation of the proposed system. Detection results of proposed system are obtained for the EML algorithm used in the classification stage. Additionally, the results of SVM and DT algorithms are also included as a comparison. The performance measurements of proposed system are shown in Table 4. The accuracy of all machine learning algorithms of proposed system is illustrated by confusion matrices shown in Figure 9. According to the obtained results, the proposed system detects approximately 99% of non-PD data. The proposed system with EML algorithm for PD data shows a high detection performance of 83.4%. MCC metrics are 0.8538, 0.8430 and 0.7447 for EML, SVM and DT, respectively. It is seen that the proposed system with EML algorithm has a higher detection performance than the system with the other two algorithms, SVM and DT. While the system with the SVM algorithm exhibits a performance close to the system with EML, the system with the DT algorithm has a lower performance. Thus, it can be seen that classifier selection is very important in the classification stage, which is the final stage of the PD detection system.

B. RESULTS OF FEATURE EXTRACTION ANALYSIS

The most important main contribution of the proposed PD detection system is the feature extraction structure. In this analysis process, performance results of four different feature sets obtained based on signal and image data are obtained. Thus, the results of these analysis processes provide information about the details of the feature extraction structure. EML, SVM and DT based PD detection results are obtained using spectral and spectrogram feature sets. The detection performance of the total feature set obtained from these four feature vectors is also achieved.

TABLE 4. Results of proposed deep feature-based detection system.

Evaluation metrics	Machine learning algorithms		
	EML	SVM	DT
MCC	0.8602	0.8430	0.7447
Accuracy	0.9847	0.9828	0.9698
Precision	0.9034	0.8836	0.7267
Recall	0.8344	0.8217	0.7962
Specificity	0.9943	0.9931	0.9809

Feature extraction analyzes are performed without applying the feature selection process. MCC and accuracy metric results are included in this analysis. The results obtained for all feature sets are given in Table 5. The bold results that are not underlined are the highest performance results obtained for the four feature sets. According to these results, it can be said that feature sets based on 1D CNN have good detection performance, especially non-PD data. Thus, the total feature vector has a very high detection performance, as can be seen from the underlined bold results in Table 4.

TABLE 5. Results of feature extraction analyses.

Feature extraction method	Feature size	Classifier	MCC	Accuracy
F1=Welch+1D CNN	1×20	EML	0.7336	0.9709
		SVM	0.7151	0.9694
		DT	0.6887	0.9659
F2=Periodogram+1D CNN	1×20	EML	0.7130	0.9686
		SVM	0.7136	0.9694
		DT	0.6983	0.9663
F3=Multitaper+1D CNN	1×20	EML	**0.7611**	**0.9747**
		SVM	0.7408	0.9721
		DT	0.7581	0.9740
F4= Spectrogram +ResNet-50	1×2048	EML	0.7352	0.9663
		SVM	0.7405	0.9701
		DT	0.6899	0.9613
F1+F2+F3+F4 (All features)	1×2108	EML	**0.8461**	**0.9832**
		SVM	0.8384	0.9824
		DT	0.7405	0.9701

C. RESULTS OF FEATURE SELECTION

In feature selection stage, mRMR feature selection process is applied to the feature set with 1×2108 data size obtained during the feature extraction stage. Performance results of the features sorted through the mRMR process are determined. Thus, as a result of this analysis, the feature set for the classifier input of the proposed detection system is determined. In this analysis process, firstly, only the first ranked feature is applied to the EML classifier input and the classification performance is obtained. Then, the classification result is obtained for the feature set created using the first ranked feature and the second ranked feature. Then, the second ranked feature is added to this feature and the classification result based on MCC is obtained. Thus, features are added sequentially until the last feature and the detection performance of each feature set is determined.

Figure 10 shows the changes in MCC values obtained from sequential detection processes for the VSB dataset. In the proposed detection system, the highest MCC coefficient is obtained with only 246 features selected from the all dataset consisting of four feature sets. Among the 246 highest mRMR-ranked features, the number of Welch, periodogram, multitaper and spectrogram features are 11, 16, 13 and 206, respectively. As seen in Figure 10, the MCC value obtained for feature set of 20 highest mRMR-ranked features is significantly higher than the MCC value obtained separately for the four feature sets given in Table 4. Thus, by means of the feature selection stage, the detection system has a high detection performance and the data size in the classifier input is reduced by about 1/8.5.

In Table 6, the mRMR score of 60 highest ranked features obtained from mRMR feature selection process and the dataset information to which the feature belongs are given. The first highest ranked feature is Welch feature-5 with 0.1960 mRMR score. The second highest ranked feature is multitaper feature-48 with 0.1584 mRMR score. Among the 60 highest mRMR-ranked features, the number of Welch, periodogram, multitaper and spectrogram features are 6, 5, 6 and 43, respectively. Thus, it can be seen that each feature extraction approach contributes significantly to the high performance of the proposed detection system.

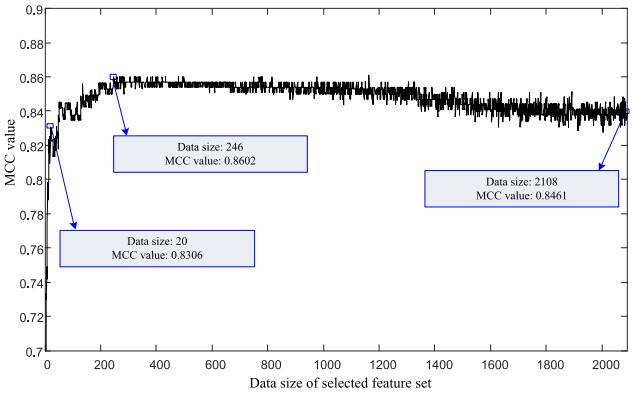

FIGURE 10. MCC values obtained using EML for sequential features sorted through mRMR feature selection.

D. TIME CONSUMPTION OF PROPOSED SYSTEM

This section shows the time complexity of the proposed PD detection system, and Table 7 shows the time consumption results. In this table, the average time values obtained for the processes performed at each stage are given. It can be said that the time consumption value of feature selection and training processes is higher than other processes. The detection time for a VSB data is approximately 4 seconds. This detection time value is a very excellent result for a VSB data with 800000 samples. This value will be even lower on computers with high GPU and processor. As a result, it can be said

TABLE 6. Ranking of best 60 features.

Feature order	Feature number	mRMR score	Dataset	Feature order	Feature number	mRMR score	Dataset	Feature order	Feature number	mRMR score	Dataset
1	5	0.1960	Welch	21	305	0.0701	Spectrogram	41	817	0.0600	Spectrogram
2	48	0.1584	Multitaper	22	16	0.0693	Welch	42	578	0.0598	Spectrogram
3	37	0.1292	Periodogram	23	1670	0.0676	Spectrogram	43	17	0.0598	Welch
4	1359	0.1281	Spectrogram	24	54	0.0675	Multitaper	44	286	0.0598	Spectrogram
5	185	0.1214	Spectrogram	25	1536	0.0662	Spectrogram	45	694	0.0597	Spectrogram
6	1908	0.1172	Spectrogram	26	1504	0.0657	Spectrogram	46	959	0.0594	Spectrogram
7	980	0.1029	Spectrogram	27	361	0.0654	Spectrogram	47	2030	0.0587	Spectrogram
8	1525	0.1019	Spectrogram	28	1692	0.0651	Spectrogram	48	1002	0.0587	Spectrogram
9	58	0.1004	Multitaper	29	1132	0.0643	Spectrogram	49	21	0.0584	Periodogram
10	2100	0.0903	Spectrogram	30	1933	0.0637	Spectrogram	50	1038	0.0582	Spectrogram
11	1487	0.0882	Spectrogram	31	763	0.0634	Spectrogram	51	20	0.0573	Welch
12	3	0.0870	Welch	32	55	0.0633	Multitaper	52	934	0.0571	Spectrogram
13	49	0.0846	Multitaper	33	1108	0.0631	Spectrogram	53	950	0.0565	Spectrogram
14	1295	0.0800	Spectrogram	34	822	0.0625	Spectrogram	54	169	0.0564	Spectrogram
15	1730	0.0769	Spectrogram	35	749	0.0623	Spectrogram	55	1917	0.0557	Spectrogram
16	84	0.0763	Spectrogram	36	24	0.0622	Periodogram	56	607	0.0549	Spectrogram
17	730	0.0761	Spectrogram	37	1743	0.0606	Spectrogram	57	2080	0.0549	Spectrogram
18	59	0.0760	Multitaper	38	1119	0.0603	Spectrogram	58	1706	0.0546	Spectrogram
19	40	0.0727	Periodogram	39	1663	0.0602	Spectrogram	59	27	0.0544	Periodogram
20	2054	0.0704	Spectrogram	40	4	0.0602	Welch	60	615	0.0534	Spectrogram

that the proposed PD detection system is quite fast and its processing complexity is low.

E. COMPARISON WITH RELATED WORKS

As explained in section I, previous studies have presented significant work on PD detection using artificial intelligence method-based approaches. The PD detection algorithms presented in some of these studies are evaluated by using the VSB dataset. Our proposed PD detection system has been compared with eight studies [1], [5], [7], [15], [16], [17], [21], [43] using the same real-time VSB dataset. Thus, readers can easily compare the resulting metric values. Table 8 shows the comparison results based on accuracy, MCC and recall metric values. As can be seen from the results, the proposed work has a very high performance compared to comparison studies. It can be said that the reason why the proposed system has a high performance compared to recent papers is due to two main contributions. The first is the spectral and spectrogram analyzes performed for PD signals. The second is that the presented algorithm includes two different deep feature extraction approaches. It also shows that the proposed feature selection approach significantly increases the performance of the model. It can be seen that only [7] has a higher MCC and recall metrics. However, the accuracy value is lower than the accuracy value of the proposed algorithm. The high MCC and recall values is due to the fact that all PD data is used only in the test dataset.

Computational time analysis is performed in only two studies [17], [42] for the detection systems proposed in these comparison papers. In [17], computation time values between 381.09 s and 630.38 s are obtained for test processes. In [42], speed values between 197.7 s and 314.9 s are obtained for test processes. In both papers, detection time values for only one PD data are not encountered. In [27], the detection time of a PD data is given by four different GPU devices. These detection times are between 0.426 s and 1.494 s. In our paper, according to the time consumption results given in Table 5, the fastest detection system is the Periodogram-based detection system (Periodogram + 1D CNN + SVM). This system has a detection time of approximately 0.120 s and the MCC value of this system is 0.7136. This computational time is quite short compared to other paper times. The detection time of our proposed system, which has an MCC value of 0.8461, is approximately 3.4 s, and it can be said that it is a good detection time compared to other studies. In this detection period, especially the spectrogram analysis time is approximately 2 s. Therefore, it can be said that extra computational time may be needed to achieve high detection performance. In future works, detection systems with a detection time of less than 1 s and at least the MCC value obtained in this paper can be developed.

F. CHALLENGES AND LIMITATIONS OF PROPOSED SYSTEM

In the previous subsections, the results obtained about the performance of the proposed PD detection system and its advantages compared to related works are included. In this subsection, the limitations of the proposed system and the

TABLE 7. Time consumption results of the proposed system.

Stage	Process	Average time (s)
Signal extraction (per data)	Welch	0.041
	Periodogram	0.035
	Multitaper	0.761
Image extraction (per data)	Spectrogram	1.946
Feature extraction	1D CNN (Pre-training)	1.252 (s/per iteration)
	1D CNN (Feature extraction from activation data)	0.088
	ResNet-50 (Pre-training)	3.644 (s/per iteration)
	ResNet-50 (Feature extraction from activation data)	0.092
Feature selection	mRMR	195.220
Classification	EML (Training)	0.763
	EML (Test)	0.438
	SVM (Training)	0.066
	SVM (Test)	0.002
	DT (Training)	0.086
	DT (Test)	0.008

TABLE 8. Comparison with similar study.

Study	Technique	Training/Test samples	MCC	Accuracy	Recall
[1]	RNN	6972/1740	0.7300	0.8820	0.7900
[5]	Extract pulses+CNN	6972/1740	0.8170	0.9670	-
[7]	SVM+PCA	6140/2572	0.9143	0.9687	0.9130
[15]	Wavelet Transform +LSTM	6972/1740	0.7400	0.9700	0.7570
[16]	Resnet18 + VggNet11	6972/1740	0.7570	0.9730	-
[17]	CNN	2323/581	-	0.9667	-
[21]	Wavelet Transform +LSTM	-	0.8500	0.8600	0.810
[43]	LSTM	6972/1740	0.8140	0.9785	0.7700
	The proposed system	**6099/2613**	**0.8602**	**0.9847**	**0.8344**

challenges encountered during the experiments are discussed. The main challenge in experimental analysis lies in the structure of the PD signals. Visual detection and interpretation of PD data is a challenging task for the responsible expert and underlying characteristics of PD data might be highly uncertain. Additionally, real-world PD signals may contain a significant amount of background noise and they have high data size. Due to these problems of PD signals, the proposed detection system encountered two different challenges. The first is to perform challenging analyzes for signals with very high data size in the signal and image extraction stages. Among the PD data with 800000 samples, only a few data contain the underlying characteristics of the PD. The remaining data represents the normal signal with noise. The second is to achieve high PD detection performance in these challenging data. However, regardless of the complexity challenges of PD data, the proposed system proved its robustness, stability, consistency and efficiency. A possible limitation of this paper is that, as mentioned above, all signal samples are analyzed. Detection performance can be improved by focusing only on patches containing possible PD ranges. Another limitation is that no noise filtering is performed on the signals. Filtered PD signals may be used instead of raw PD signals at the input of the proposed PD detection system. Thus, even though the filtering process creates a computational burden, it has the potential to increase detection performance.

VI. CONCLUSION

In this paper, a novel PD detection system based on the deep features of CNN algorithms is proposed. In the proposed detection system, frequency-based features are obtained as a result of applying spectral analysis and spectrogram analysis to PD data. These frequency features are applied to 1D CNN models and ResNet-50 models to obtain deep features. Among these deep features, the most effective features are obtained with the mRMR feature selection approach. Thus, before the final classification stage, a PD dataset is converted into a frequency feature set, the frequency feature set is converted into a deep feature set, and the deep feature set is reduced to the selected deep feature set. In the last stage of proposed PD detection system, the PD detection process is carried out by applying the final deep feature set to EML, SVM and DT classifiers. Some experimental results are obtained by applying the ENET dataset to the proposed PD detection system. According to the obtained results, the proposed deep feature-based PD detection system has high detection accuracy and the detection speed is quite high. It also offers higher performance than similar PD detection algorithms in the literature. Moreover, the proposed detection system overcomes the limitations of traditional approaches and has a significant improvement in PD detection performance (MCC: 86.02%) compared to the winning models in the Kaggle competition (MCC: 70.6%) over the VSB dataset.

Experimental results show that the deep feature-based system proposed in this paper can accurately detect PD signals in real-world environments with high performance. Spectral and spectrogram analyzes for PD signals and the use of two different deep feature extraction approaches play a key role in this performance. In addition, the feature selection approach is also very effective for the detection performance of the system. In future works, faster and higher-performance algorithms based on spectral and spectrogram analyzes can be developed at the signal and image extraction stage of the proposed detection system. Additionally, an effective noise filtering process can be integrated into the system.

REFERENCES

[1] C. Huang, S. Ding, S. Li, and R. Liu, "LMFE: Learning-based multiscale feature engineering in partial discharge detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 5, pp. 5848–5856, May 2024.

[2] B. Bajwa, C. Butani, and C. Patel, "A novel approach towards predicting faults in power systems using machine learning," *Electr. Eng.*, vol. 104, no. 1, pp. 363–368, Feb. 2022.

[3] K. Chen, T. Vantuch, Y. Zhang, J. Hu, and J. He, "Fault detection for covered conductors with high-frequency voltage signals: From local patterns to global features," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1602–1614, Mar. 2021.

[4] S. Misak, M. Kratky, and L. Prokop, "A novel method for detection and classification of covered conductor faults," *Adv. Electr. Electron. Eng.*, vol. 14, no. 5, pp. 481–489, Dec. 2016.

[5] G. Michau, C.-C. Hsu, and O. Fink, "Interpretable detection of partial discharge in power lines with deep learning," *Sensors*, vol. 21, no. 6, p. 2154, Mar. 2021.

[6] W. Elmasry and M. Wadi, "EDLA-EFDS: A novel ensemble deep learning approach for electrical fault detection systems," *Electric Power Syst. Res.*, vol. 207, Jun. 2022, Art. no. 107834.

[7] W. Elmasry and M. Wadi, "Detection of faults in electrical power grids using an enhanced anomaly-based method," *Arabian J. Sci. Eng.*, vol. 47, no. 11, pp. 14899–14914, Nov. 2022.

[8] A. Prasad, J. Belwin Edward, and K. Ravi, "A review on fault classification methodologies in power transmission systems: Part—I," *J. Electr. Syst. Inf. Technol.*, vol. 5, no. 1, pp. 48–60, May 2018.

[9] Y. Xi, X. Tang, Z. Li, Y. Shen, and X. Zeng, "Fault detection and classification on insulated overhead conductors based on MCNN-LSTM," *IET Renew. Power Gener.*, vol. 16, no. 7, pp. 1425–1433, May 2022.

[10] B. Erişti, "A partial discharge fault detection method based on wavelet packet transform, ReliefF feature selection and ensemble learning algorithm," *Firat Univ. J. Eng. Sci.*, vol. 35, no. 2, pp. 505–516, 2023, doi: 10.35234/fumbd.1284537.

[11] S. Misák, J. Fulnecek, T. Vantuch, T. Buriánek, and T. Jezowicz, "A complex classification approach of partial discharges from covered conductors in real environment," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 2, pp. 1097–1104, Apr. 2017.

[12] L. Li, J. Tang, and Y. Liu, "Partial discharge recognition in gas insulated switchgear based on multi-information fusion," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 2, pp. 1080–1087, Apr. 2015.

[13] S.-H. Bae and K.-J. Yoon, "Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 595–610, Mar. 2018.

[14] M. Dong and J. Sun, "Partial discharge detection on aerial covered conductors using time-series decomposition and long short-term memory network," *Electric Power Syst. Res.*, vol. 184, Jul. 2020, Art. no. 106318.

[15] Y. Xi, F. Zhou, and W. Zhang, "Partial discharge detection and recognition in insulated overhead conductor based on bi-LSTM with attention mechanism," *Electronics*, vol. 12, no. 11, p. 2373, May 2023.

[16] W. Wang and N. Yu, "Partial discharge detection with convolutional neural networks," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, Aug. 2020, pp. 1–6.

[17] R. Srivastava and V. Avasthi, "Deep convolutional neural network for partial discharge monitoring system," *Adv. Eng. Softw.*, vol. 180, Jun. 2023, Art. no. 103407.

[18] Z. Li, N. Qu, X. Li, J. Zuo, and Y. Yin, "Partial discharge detection of insulated conductors based on CNN-LSTM of attention mechanisms," *J. Power Electron.*, vol. 21, no. 7, pp. 1030–1040, Jul. 2021.

[19] (2018). *VSB Power Line Fault Detection*. Kaggle. [Online]. Available: https://www.kaggle.com/c/vsb-power-line-fault-detection/data

[20] L. Klein, J. Dvorský, D. Seidl, and L. Prokop, "Novel lossy compression method of noisy time series data with anomalies: Application to partial discharge monitoring in overhead power lines," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108267.

[21] N. Qu, Z. Li, J. Zuo, and J. Chen, "Fault detection on insulated overhead conductors based on DWT-LSTM and partial discharge," *IEEE Access*, vol. 8, pp. 87060–87070, 2020.

[22] S. Mišák and V. Pokorný, "Testing of a covered conductor's fault detectors," *IEEE Trans. Power Del.*, vol. 30, no. 3, pp. 1096–1103, Jun. 2015.

[23] A. Said and H. Göker, "Spectral analysis and bi-LSTM deep network-based approach in detection of mild cognitive impairment from electroencephalography signals," *Cognit. Neurodynamics*, vol. 18, no. 2, pp. 597–614, Apr. 2024.

[24] Z. Zhang, "Spectral and time-frequency analysis," in *EEG Signal Processing and Feature Extraction*. Singapore: Springer, 2019, pp. 89–116.

[25] E. Buss, T. Aust, M. Wahby, T.-L. Rabbel, S. Kernbach, and H. Hamann, "Stimulus classification with electrical potential and impedance of living plants: Comparing discriminant analysis and deep-learning methods," *Bioinspiration Biomimetics*, vol. 18, no. 2, Mar. 2023, Art. no. 025003.

[26] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. Hoboken, NJ, USA: Wiley, 1996.

[27] A. Meziani, K. Djouani, T. Medkour, and A. Chibani, "A lasso quantile periodogram based feature extraction for EEG-based motor imagery," *J. Neurosci. Methods*, vol. 328, Dec. 2019, Art. no. 108434.

[28] M. A. Wieczorek and F. J. Simons, "Minimum-variance multitaper spectral estimation on the sphere," *J. Fourier Anal. Appl.*, vol. 13, no. 6, pp. 665–692, Dec. 2007.

[29] L. Klein, P. Žmij, and P. Krömer, "Partial discharge detection by edge computing," *IEEE Access*, vol. 11, pp. 44192–44204, 2023.

[30] C. Ardito, Y. Deldjoo, T. D. Noia, E. D. Sciascio, and F. Nazary, "Visual inspection of fault type and zone prediction in electrical grids using interpretable spectrogram-based CNN modeling," *Exp. Syst. Appl.*, vol. 210, Dec. 2022, Art. no. 118368.

[31] S. Duran, K. Üreten, Y. Maraş, H. H. Maraş, K. Gök, E. Atalar, and V. Çayhan, "Automatic detection of spina bifida occulta with deep learning methods from plain pelvic radiographs," *Res. Biomed. Eng.*, vol. 39, no. 3, pp. 655–661, Jul. 2023.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[33] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Feature selection for high-dimensional data," *Prog. Artif. Intell.*, vol. 5, pp. 65–75, May 2016.

[34] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[35] O. Sagi and L. Rokach, "Ensemble learning: A survey," *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1249, Jul. 2018.

[36] R. Vempati and L. D. Sharma, "EEG rhythm based emotion recognition using multivariate decomposition and ensemble machine learning classifier," *J. Neurosci. Methods*, vol. 393, Jun. 2023, Art. no. 109879.

[37] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[38] B. Eristi and H. Eristi, "Classification of power quality disturbances in solar PV integrated power system based on a hybrid deep learning approach," *Int. Trans. Electr. Energy Syst.*, vol. 2022, pp. 1–13, Jun. 2022.

[39] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[40] R. Sekhar, N. Solke, and P. Shah, "Lean manufacturing soft sensors for automotive industries," *Appl. Syst. Innov.*, vol. 6, no. 1, p. 22, Feb. 2023.

[41] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process*, vol. 5, no. 2, pp. 1–11, Mar. 2015.

[42] L. Lavazza and S. Morasca, "Comparing ϕ and the F-measure as performance metrics for software-related classifications," *Empirical Softw. Eng.*, vol. 27, no. 7, p. 185, Dec. 2022.

[43] C. Zhang, M. Chen, Y. Zhang, W. Deng, Y. Gong, and D. Zhang, "Partial discharge pattern recognition algorithm of overhead covered conductors based on feature optimization and bidirectional LSTM-GRU," *IET Gener., Transmiss. Distrib.*, vol. 18, pp. 680–693, Feb. 2024.

BELKIS ERISTI (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical education and the Ph.D. degree in electrical and electronics engineering from Firat University, Türkiye, in 2000, 2002, and 2015, respectively. She was a Lecturer with the Department of Electrical and Electronics Engineering, Tunceli University, in March 2016. She is currently working as an Assistant Professor with the Department of Electrical and Energy, Vocational School of Technical Sciences, Mersin University. Her research interests include signal processing, pattern recognition, power quality, power systems, and computer vision.