



An expert system for EMI data classification based on complex Bispectrum representation and deep learning methods

Imene Mitiche ^{a,*}, Mark D. Jenkins ^a, Philip Boreham ^b, Alan Nesbitt ^a, Gordon Morison ^a

^a Department of Engineering, Glasgow Caledonian University, Glasgow G40BA, United Kingdom

^b Innovation Centre for Online Systems, Doble, Bere Regis BH20 7LA, United Kingdom



ARTICLE INFO

Keywords:

Complex and real-valued Bispectrum
Condition monitoring
Deep learning
EMI diagnostic
Expert system
Insulation faults

ABSTRACT

This paper presents expert system framework based on Machine Learning (ML) for High-Voltage (HV) asset condition monitoring. The work investigates the classification of insulation faults in HV environment, based on real-world time series signals labelled by condition monitoring experts. Extending on our previous work, the proposed approach exploits the Bispectrum analysis and deep learning for feature extraction and classification. The calculated Bispectrum on time series signals can be deployed as the complex-valued Bispectrum, which contains phase information, or as its real-valued magnitude. This can be approached as an image classification problem which can be implemented in various deep networks including Convolutional Neural Network (CNN), Residual Neural Network (ResNet) and their complex-valued version. The employed deep networks performance is compared in terms of their classification accuracy. High classification performance is obtained which produces comparable performance with expert diagnosis. Thus, it can be interpreted as transfer of expert system to an intelligent system.

1. Introduction

The occurring insulation faults within High-Voltage (HV) assets can be either emitted or conducted energy. Electromagnetic Interference (EMI) is the precise frequency and time domain measurement and identification of radio energy that results from electrical activity at defects in HV apparatus. The exertion of EMI method in condition monitoring of HV equipment within the power production industry has been successful in the past decades. This approach is able to detect insulation degradation and conduction faults as well as mechanical defects. Since the first EMI test, in 1980, over 65 different types of defects have been recognised using this approach over 9000 tests (Timperley, 2012). The collected EMI data in both the frequency and the time domain is analysed by experts to identify potential faults. The need for continuous online monitoring of machines is growing due to safety and sparsity of field experts, and training more experts is costly. A dataset containing labelled signals of previous faults, captured using the EMI method, is available to exploit in developing an automatic condition monitoring system. The approach employs classification concept in supervised learning. This can be defined as the prediction of unseen data instances by a model that previously learned from a training dataset.

In this paper, a signal analysis based on Higher Order Statistics (HOS) (Nikias & Petropulu, 1993) is employed to map and extract relevant patterns that are implemented into a range of deep Convolutional Neural Networks (CNNs). Specifically, Bispectrum analysis is employed to obtain two possible mappings. The first is the Bispectrum magnitude which is real-valued component. The second is the complex-valued Bispectrum, with real and imaginary components, which preserves phase information. This analysis has recently received the attention of Biomedical (Sun, Feng, Lu, Wang, & Zhang, 2019; Lanatá, Valenza, Mancuso, & Scilingo, 2011; Li & Liu, 2010) and Audio (Bhalke, Rama Rao, & Bormane, 2016) applications where Bispectrum features were implemented to various Machine Learning (ML) techniques. Bispectrum analysis was also exploited for vibration signals, with non-stationarity characteristics, to identify gearbox fault types (Jinying, Hongxia, & Shihua, 2010). It was demonstrated that such signals contain fault components that cause additional components to modulate the fundamental component. The Bispectrum identifies the degree of phase coupling of a signal while considering the amplitude, unlike the traditional power spectrum obtained by Fast Fourier Transform (FFT) which provides only the power of a signal over frequency. Although the Bispectrum has shown promising results within EEG, audio and vibration

* Corresponding author.

E-mail addresses: imene.mitiche@gcu.ac.uk (I. Mitiche), mark.jenkins@gcu.ac.uk (M.D. Jenkins), pboreham@doble.com (P. Boreham), alan.nesbitt@gcu.ac.uk (A. Nesbitt), gordon.morison@gcu.ac.uk (G. Morison).

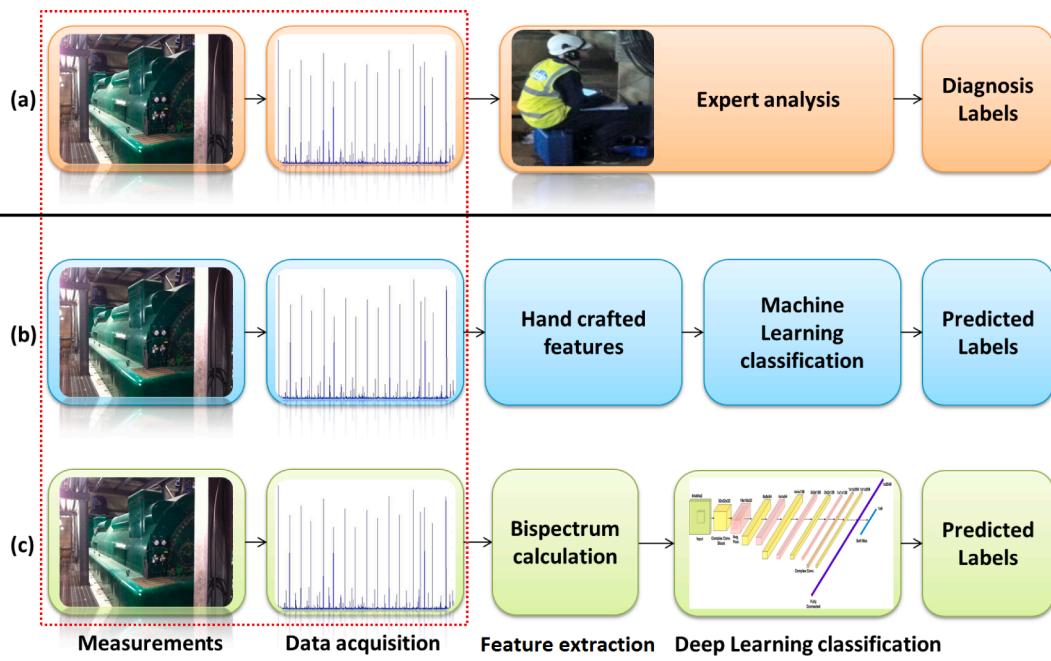


Fig. 1. Condition monitoring methods (a) Expert-based (b) and (c) are Machine Learning-based methods, where in (c) Deep Learning is used for both feature selection and classification.

signals classification, it has not yet been explored in the context of EMI signals classification.

In this paper we propose various ways to utilise the obtained Bispectrum feature into a deep CNN with one channel for the Bispectrum magnitude, and two channels for the complex-valued Bispectrum. Furthermore, a third deep CNN-based framework known as Complex CNN is employed. The performance of these classifiers is evaluated on a large and imbalanced dataset of signals labelled according to common labels practised by the power industry in the United States of America and Canada. The latter were measured using the EMI method at various real-world HV power plant assets and labelled by field experts for diagnosis purposes. Labelled examples include Partial Discharge (PD), Arcing, Exciter pulses etc.

The main novelty and contributions in this work are as follows: (a) the combination of complex Bispectrum images and deep CNNs for feature mapping and classification (b) the application of the proposed framework to condition monitoring of HV power plant assets (c) solving a real-world problem as most classification tasks related to machine health monitoring are usually applied to synthetic data.

This paper is organised as follows. Related work to EMI diagnosis is addressed in the next section. Section 3 provides a description of the proposed approach and the employed algorithms. Section 4 describes the data acquisition steps and results of the proposed data analysis approach. The final section draws a conclusion to this paper with future work and recommendations.

2. Related work

This section provides an overview of the different condition monitoring approaches using the EMI diagnosis method and similar techniques in the literature applied to machine health monitoring in general.

Condition monitoring of machinery can be performed using three main methods which are illustrated in Fig. 1. The traditional and current EMI diagnosis method is mostly manual and it involves expert to perform both measurement and data analysis steps. This generates manually a report on the diagnosis along with labelled data to be archived. Various work for this concern was published in (Timperley, Buchanan, & Vallejo, 2018; Timperley & Vallejo, 2017; Timperley,

Vallejo, & Nesbitt, 2014; Timperley, 2008). It is evident that this method is time consuming and relies on expert availability. Modern diagnosis approaches employ ML algorithms, such as Support Vector Machine (Robles, Parrado-Hernandez, Ardila-Rey, & Martinez-Tarifa, 2016), artificial NN (Zhang, Wang, & Chen, 2015; Bin, Gao, Li, & Dhillon, 2012) and random forest (Li et al., 2016). This is popular in vibration and acoustic data analysis for rotary machines diagnosis. Authors in (Mitiche, Morison, Nesbitt, Stewart, & Boreham, 2018; Mitiche et al., 2018) introduced the application of ML to EMI diagnosis by proposing various feature extraction techniques for ML classification. However, the study was limited to a small dataset. The most recent diagnosis method exploits Deep Learning for its ability to learn features from the data, while achieving the state-of-the-art results. Deep Learning algorithms are derived from ML algorithms. In Fig. 1 (c), the Bispectrum could be seen as a mapping from 1D time series to a 2D image or tensor and a feature extraction at the same time. Yet, the Deep Learning algorithm identifies and learns the different features among the classes. Recent work employed ML (Jorkyase et al., 2019) and Deep Learning (Duan et al., 2019) based methods to PD identification. This work was applied to artificially simulated PD faults, however in our paper real-world data, labelled by EMI diagnostic experts, were captured in operating HV stations.

CNN is the popular Deep Learning algorithm applied in the PD field as well as in other fields (Wang, Li, & Zhang, 2019). The first CNN algorithm was introduced by LeCun et al. in 1998 and is known as LeNet-5 which consists of 7 convolutional layers (LeCun, Bottou, Bengio, & Haffner, 1998). However, LeNet was limited to computing resources. Since then, various deep CNN architectures were proposed to reduce the error rate and improve the performance of the prior algorithms including AlexNet (Krizhevsky, Sutskever, & Hinton, 2017), GoogleNet (Szegedy et al., 2015), VGGNet (Simonyan & Zisserman, 2015) and one of the current state-of-the-art Residual Neural Network (ResNet) (He, Zhang, Ren, & Sun, 2016). ResNet was implemented successfully in various applications, for instance in video processing (Huang, Ding, & Chen, 2018) and rotating machinery fault identification (Li, Zhang, & Ding, 2019). Research on the application of deep networks to EMI-based condition monitoring is scarce in the literature, which was a motivation to carry out the work presented in this paper.

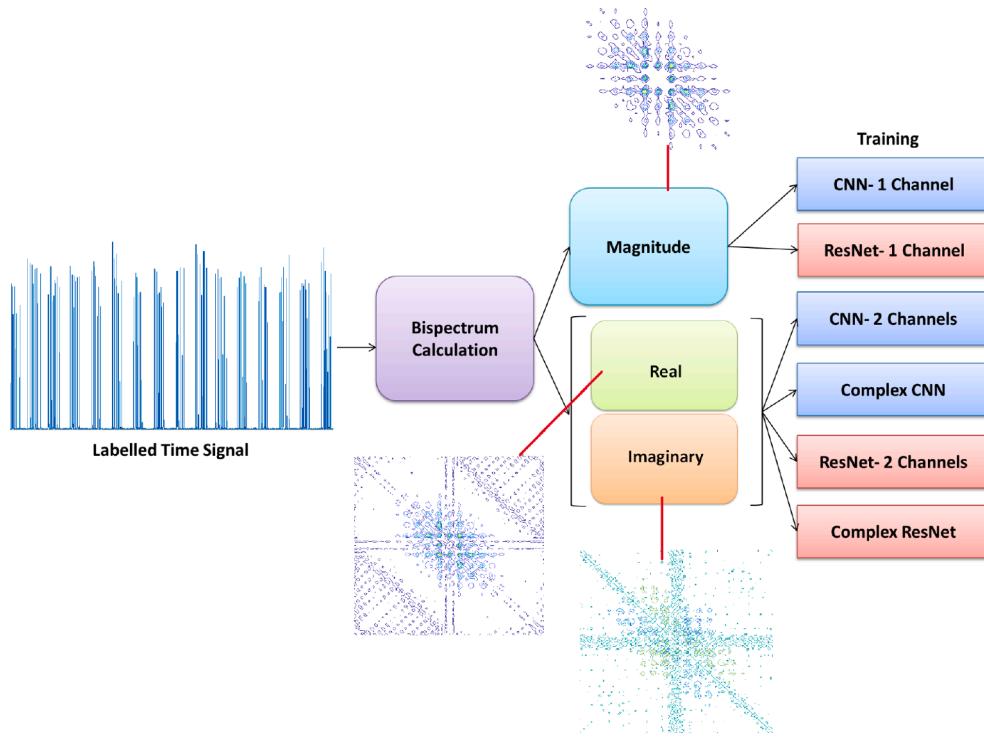


Fig. 2. The proposed analysis for insulation fault detection, using six different deep networks for performance comparison.

3. Pattern recognition methodology

In this section, the combination of both the real and the complex-valued Bispectrum features with various CNN models is described. Details of the analysis undertaken within this paper is summarised in Fig. 2, where first the Bispectrum of the labelled time signals is calculated to obtain a 2D-matrix for a 1D-vector. As the Bispectrum is derived from Fourier Transform (FT), the resulting output is complex-valued. The output could be deployed either as Bispectrum magnitude, by taking the absolute value of the complex Bispectrum, or directly use the complex-valued Bispectrum. This opens various possible ways of implementation into a 2D deep network for classification. The complex network blocks such as convolution and activations are derived from (Trabelsi et al., 2018). The motivation to use the complex form of the Bispectrum over its magnitude lies behind the additional information that can be observed in Fig. 2, which could be useful in classification.

3.1. Bispectral analysis

First, the mathematical notation used in this paper denotes scalars by lower case, vectors by bold lower case and matrices by bold upper case. In time-domain HOS analysis, the first-order moment of a discrete stationary signal $\mathbf{x}[n]; n = 1, \dots, N$ is the mean μ and its second-order moment is the variance σ^2 , both are defined in Eq. (1) and Eq. (2) respectively.

$$\mu = E(\mathbf{x}[n]) \quad (1)$$

$$\sigma^2 = E(\mathbf{x}[n]\mathbf{x}[n + \tau_1]) \quad (2)$$

The third-order moment of the signal is defined as:

$$\gamma = E(\mathbf{x}[n]\mathbf{x}[n + \tau_1]\mathbf{x}[n + \tau_2]) \quad (3)$$

where τ_1 and τ_2 are time lags. Assuming that the signal $\mathbf{x}[n]$ is zero mean, then the second and third-order cumulants are equal to the respective moments (Chua, Chandran, Acharya, & Min Lim, 2010). In frequency-domain definition, the third-order cumulant spectrum, known as the

Bispectrum, is the FT of the third-order moment which can be defined as:

$$\mathbf{B}(f_1, f_2) = \sum_{\tau_1=1}^N \sum_{\tau_2=1}^N \sum_{n=1}^N \mathbf{x}[n]\mathbf{x}[n + \tau_1]\mathbf{x}[n + \tau_2] \cdot e^{-j2\pi \cdot (f_1\tau_1 + f_2\tau_2)} \quad (4)$$

Let $n + \tau_1 = m$ and $n + \tau_2 = k$, by rewriting Eq. (4) and splitting the exponent, Bispectrum can be reformulated as:

$$\begin{aligned} \mathbf{B}(f_1, f_2) &= \left\{ \sum_{m=1}^N \mathbf{x}[m] e^{-j2\pi f_1 m} \right\} \cdot \left\{ \sum_{k=1}^N \mathbf{x}[k] e^{-j2\pi f_2 k} \right\} \cdot \left\{ \sum_{n=1}^N \mathbf{x}[n] e^{+j2\pi (f_1 + f_2)n} \right\} \\ &= X(f_1)X(f_2)X^*(f_1 + f_2) \end{aligned} \quad (5)$$

where $X(f)$ is the FT and $X^*(f)$ is the complex conjugate of $X(f)$. This equation allows a faster computation of the Bispectrum, which is a complex measure that contains both magnitude and phase information (Nikias & Petropulu, 1993). By applying the FT of to the higher order correlation function, the Bispectrum provides further information on phase relationships between frequency components, through detecting whether the phase of signal components at frequencies f_1, f_2 and the harmonic $f_1 + f_2$ are interdependent. In other words, it identifies the existence of quadratic (and cubic) non-linear coupling information (Kicinski & Szczepanski, 2004).

3.2. Real network building blocks

The standard building blocks of a CNN are convolution, activation function, and Batch Normalization (BN) which are explained as follows.

3.2.1. Convolution

Convolution is the most important operation in a CNN as it is responsible for information retrieval. A 2D convolution for a layer \mathbf{X} of size $N \times N$ convolved with a filter \mathbf{W} of size $m \times m$ can be expressed in Eq. (6) which results in an output layer of size $(N - m + 1) \times (N - m + 1)$. Here, zero padding is only applied at the edges.

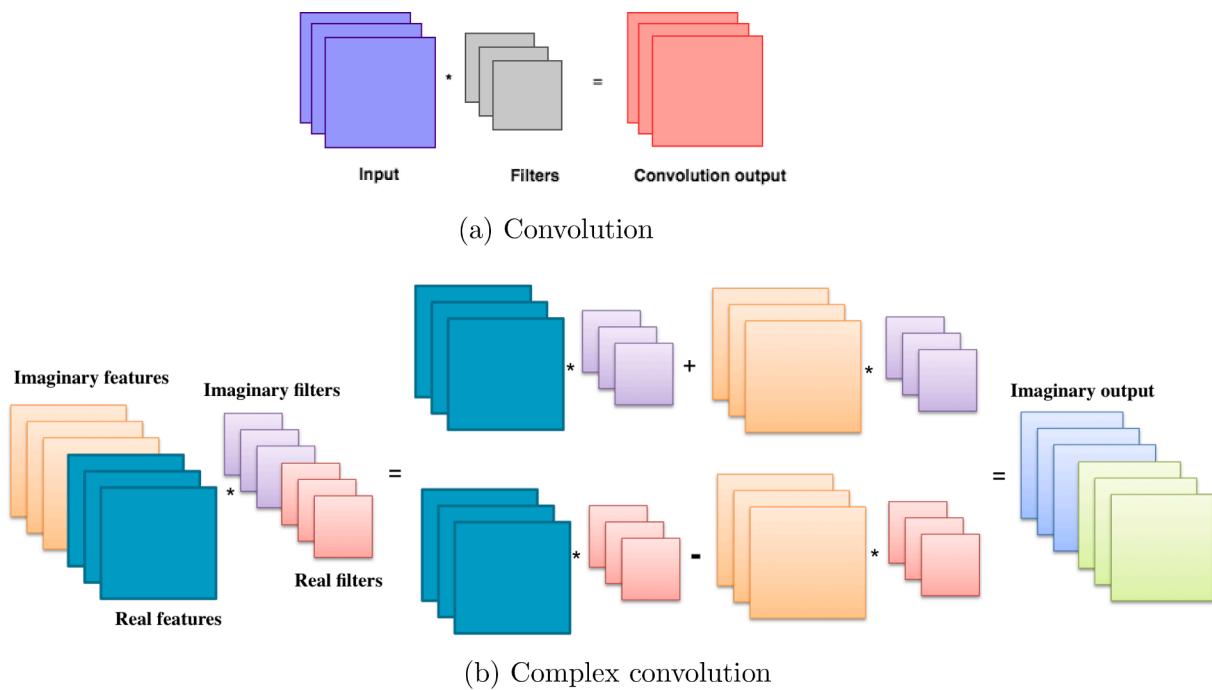


Fig. 3. Complex convolution in (b) based on real convolution in (a), where two channel input is convolved with real and imaginary kernels.

$$\mathbf{W} * \mathbf{X} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \mathbf{W}(a, b) \mathbf{X}(i-a, j-b) \quad (6)$$

From Eq. (6) the convolution is denoted by $*$ which is the sum of the point-wise multiplication between the input layer and the filter matrices.

3.2.2. Batch normalization

BN is used to normalise and scale an input layer to adjust its distribution due to the parameters change in the previous layer. This allows for faster training (Ioffe & Szegedy, 2015). The BN is applied independently to each dimension of a multidimensional layer $\mathfrak{B} = \{\mathfrak{B}^1, \dots, \mathfrak{B}^d\}$ with d dimension, through the following equation:

$$BN(\mathfrak{B}) = \gamma \frac{\mathfrak{B} - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (7)$$

where μ, σ^2 are the mean and variance of the batch \mathfrak{B} , respectively and ϵ is a constant introduced for numerical stability. The parameters γ and β are learnable to ensure each layer has zero mean and variance of 1 (Ioffe & Szegedy, 2015).

3.2.3. Rectified Linear Unit (ReLU)

Deep neural networks are trained using the Stochastic Gradient Descent approach, which back-propagates the prediction error through the network to update the weights. Networks with a large number of layers causes the gradient error to diminish significantly as it is propagates closer to the input layer and as a result it has an insignificant effect on updating the weights. This issue is called vanishing gradient (Goodfellow, Bengio, & Courville, 2016). ReLU is known for solving the vanishing gradient problem, compared to sigmoid and hyperbolic tangent activation functions, while allowing the model to train faster (Xu, Wang, Chen, & Li, 2015). Thus it employed by default in CNNs (Wu, Xu, Zhu, & Guo, 2018; Wang et al., 2019). ReLU function is defined in Eq. (8) such that it is zero when the input is negative and it has a gradient of 1 when the input is positive.

$$f(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (8)$$

3.3. Complex network building blocks

The mathematical definition of the main complex blocks including convolution, ReLU and BN are described as follows.

3.3.1. Complex convolution

The 2D complex convolution is performed by convolving a complex filter matrix $\mathbf{W} = \mathbf{A} + j\mathbf{B}$ by a complex input $\mathbf{Z} = \mathbf{Z}_R + j\mathbf{Z}_I$, where the matrices \mathbf{A}, \mathbf{B} , and vectors $\mathbf{Z}_R, \mathbf{Z}_I$ consist of real-valued entities corresponding to the real and imaginary parts respectively. Note that in Electronic and Electrical Engineering j is used to denote the imaginary unit to avoid confusion with electric current i (Boas, 2006). The convolution in frequency domain can be formulated as follows:

$$\mathbf{W} * \mathbf{Z} = (\mathbf{A} * \mathbf{Z}_R - \mathbf{B} * \mathbf{Z}_I) + (j(\mathbf{B} * \mathbf{Z}_R + \mathbf{A} * \mathbf{Z}_I)) \quad (9)$$

The resulting convolution has also real and imaginary parts which can be written as:

$$\begin{bmatrix} Re(\mathbf{W} * \mathbf{Z}) \\ Im(\mathbf{W} * \mathbf{Z}) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} * \begin{bmatrix} \mathbf{Z}_R \\ \mathbf{Z}_I \end{bmatrix} = \begin{bmatrix} \mathbf{A} * \mathbf{Z}_R - \mathbf{B} * \mathbf{Z}_I \\ \mathbf{B} * \mathbf{Z}_R + \mathbf{A} * \mathbf{Z}_I \end{bmatrix} \quad (10)$$

It can be seen that the complex convolution is a particular case of real convolution with double the number of filters and parameters to account for the imaginary part of the input. The difference between both operations is illustrated in Fig. 3.

3.3.2. Complex BN

A BN method for Complex CNN is utilised to fit within the complex model, which is calculated in the following steps. First, the centered data is multiplied by the inverse square root of the 2×2 covariance matrix defined in Eq. (12) as:

$$\hat{\mathfrak{B}} = \mathbf{V}^{-1/2}(\mathfrak{B} - \mu) \quad (11)$$

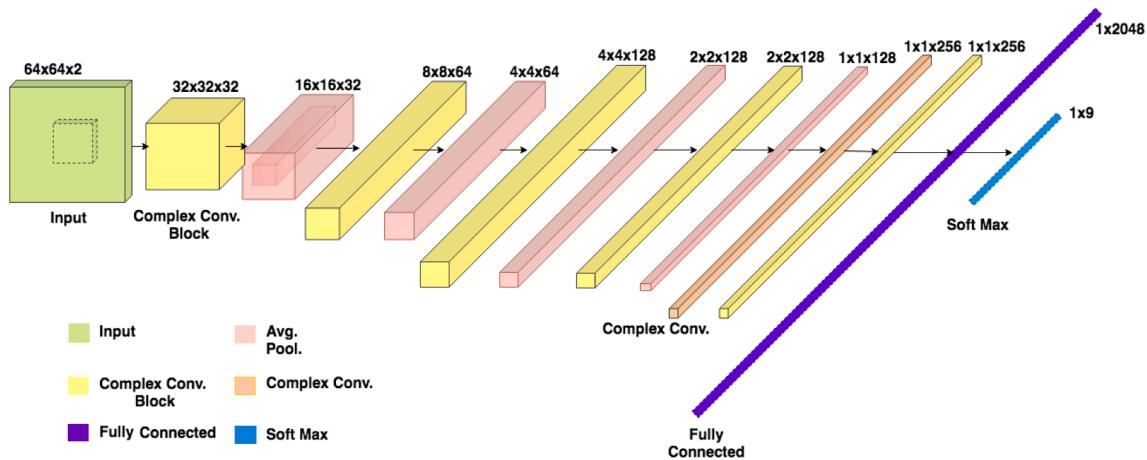


Fig. 4. CNN architecture used in both real and complex cases.

Table 1

Structure of the real CV-CNN architecture. The convolution block consists of convolution, BN and ReLU.

Layer	Kernel	No. filters	Stride	Operation
1	6×6	16	2	Conv. block
2	2×2	-	2	MaxPool
3	3×3	32	2	Conv. Block
4	2×2	-	2	MaxPool
5	3×3	64	1	Conv. Block
6	2×2	-	2	MaxPool
7	3×3	64	1	Conv. Block
8	2×2	-	2	MaxPool
9	3×3	128	1	Conv.+ReLU
10	3v3	128	1	Conv. Block
11	2×2	-	2	MaxPool
12	Fully Connected, ReLU, SoftMax			

$$\mathbf{V} = \begin{bmatrix} Cov(Re(\mathfrak{B}), Re(\mathfrak{B})) & Cov(Re(\mathfrak{B}), Im(\mathfrak{B})) \\ Cov(Im(\mathfrak{B}), Re(\mathfrak{B})) & Cov(Im(\mathfrak{B}), Im(\mathfrak{B})) \end{bmatrix} \quad (12)$$

The BN is then calculated for $\hat{\mathfrak{B}}$ as:

$$BN(\hat{\mathfrak{B}}) = \Gamma \hat{\mathfrak{B}} + \beta \quad (13)$$

where β is a complex parameter with real and imaginary learnable components, Γ is a 2×2 scaling parameter defined in Eq. 14.

$$\Gamma = \begin{bmatrix} \gamma_{rr} & \gamma_{ri} \\ \gamma_{ri} & \gamma_{ii} \end{bmatrix} \quad (14)$$

γ_{rr} , γ_{ii} and γ_{ri} are set to $1/\sqrt{2}$, $1/\sqrt{2}$ and 0 respectively at the start of the operation. Moving averages are implemented to estimate the complex BN parameters, where they are initialised to zero for β , $Cov(Re(\mathfrak{B})$, $Im(\mathfrak{B})$), and to $1/\sqrt{2}$ for $Cov(Re(\mathfrak{B}), Re(\mathfrak{B}))$ and $Cov(Im(\mathfrak{B}), Im(\mathfrak{B}))$.

3.3.3. Complex ReLU

In this work the activation is applied to the real and imaginary parts of the complex value independently. Here, a complex ReLU (cReLU) is utilised as follows:

$$cReLU(\mathbf{Z}) = ReLU(\mathbf{Z}_R) + jReLU(\mathbf{Z}_I) \quad (15)$$

This results in preserved phase and magnitude information when the angle $\theta \in]0, \pi/2[$. The real part is discarded when $\theta \in [\pi/2, \pi]$, the imaginary part when $\theta \in [3\pi/2, 2\pi]$ and both parts are discarded when $\theta \in [\pi, 3\pi/2[$.

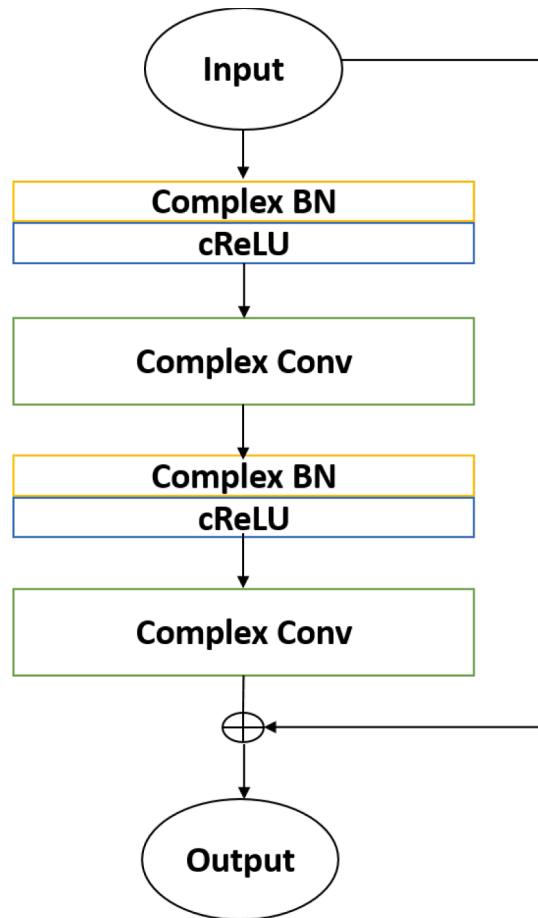


Fig. 5. Complex ResNet residual block.

3.4. CNN architecture

Three different CNN models are employed, which have similar architecture as presented in Fig. 4. The real-valued CNN model was constructed using the 2D convolution, BN and ReLU described above, in addition to the standard Keras max pooling. The overall architecture is summarised in Table 1 where a convolution block consists of convolution, BN and ReLU in order. The Complex CNN follows similar architecture in terms of the number of layers, kernel, filters and stride, however complex convolution, complex BN and cReLU are employed

Table 2
Structure of the real ResNet architecture.

Layer	Kernel	Depth	Stride	Operation
1	7×7	32	2	Conv
2	3×3	-	1	MaxPool
3	3×3	32	1	Residual Block $\times 2$
4	3×3	64	1	Residual Block $\times 2$
5	3×3	128	1	Residual Block $\times 2$
6	3×3	256	1	Residual Block $\times 2$
7	Average Pool, Fully Connected, SoftMax			

instead of the real-valued function.

3.5. Deep ResNet architecture

ResNet is simply a CNN with skip connections between the residual blocks to avoid the vanishing/exploding gradients (Szegedy et al., 2015). In this work, two architectures of ResNet are employed: real and complex-valued. Here the ResNet architecture proposed by (He et al., 2016) was followed. The Real ResNet employs the real-valued 2D convolution, BN and ReLU, whereas the Complex ResNet model uses the complex 2D convolution, cReLU and complex BN, in addition to the standard Keras average/max pooling. ResNet model consists of a number of stacked residual blocks joined by skip connection between the

input and output of the block. The content of each block is illustrated in Fig. 5. The overall ResNet architecture is summarised in Table 2.

4. Experimental Set-up

In this section, the process of EMI data acquisition is described. Details on the employed data set are also provided and results on the proposed analysis are presented.

4.1. Data acquisition

EMI data collection is performed in frequency domain and online during normal operation of the asset. It is totally passive and non-invasive, thus no design modifications are required. The emitted EMI signals can be captured across a wide frequency range between 10's of kHz up to 100 MHz. EMI can both be radiated and conducted from the discharge site and the resulting EMI spectrum is unique for each physical location and type of defect within the electrical system (Timperley et al., 2014). An EMI measurement of an HV asset is carried out as follows. Using a split-core High-Frequency-Current-Transformer (HFCT), signals are measured on the neutral earth cable of a generator stator winding, at the grounding transformer, or, a motor cable supply conduit. These signals are usually recorded with a PD Surveyor 200 (PDS200) instrument that functions under The Comité International Spécial des

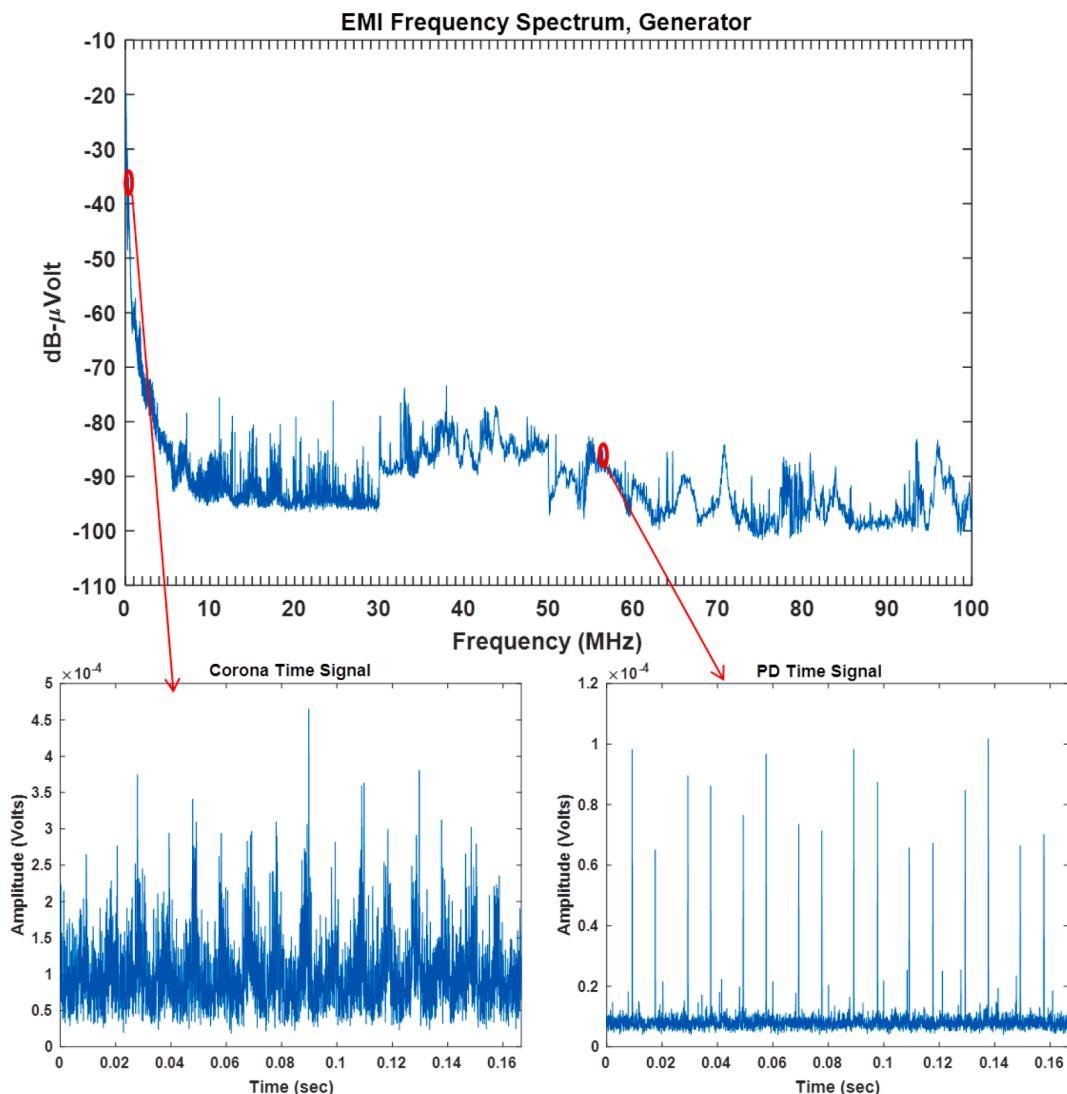


Fig. 6. EMI frequency spectrum example and time resolved Partial Discharge (PD) and Corona signals extracted at 1 MHz and 56 MHz.

Table 3

Data set information. The number of train/val/test instances is obtained after signal segmentation (refer to 4.2).

Class	No. Training	No. Validation	No. Testing
Arcing	772	166	166
Corona	83	18	18
Data Modulation	370	79	79
PD	415	1588	1588
Process Noise	585	125	125
Random Noise	712	152	152
Exciter	1061	228	228
microSparking	613	132	132
Total	4611	2488	2488

Perturbations Radioélectriques (CISPR) 16 standard (33.100.10-Emission, 2015) for EMI filter types, to ensure compatibility of measurements and results interpretation across EMI measurement instruments. The measuring device provides a frequency spectrum from which time signals can be retrieved at a sampling rate of 24 kHz, at a frequency of interest, using AM demodulation. The PDS200 looks at signals based on the CISPR16 bandwidths and moves through the frequency spectrum making appropriate power filtered response measurements at each frequency. Quasi-peak measurements are made over a 1s time-period at each selected frequency. EMI measured frequency spectrum is explained further in (Timperley & Vallejo, 2017; Timperley et al., 2014). For time resolved signals the instrument permits any frequency between 50 kHz to 100 MHz to be selected, for e.g. where the maximum envelope energy exists. It then makes the slower demodulated response measurement which is sampled at 24 kHz. Fig. 6 illustrates an example of EMI frequency spectrum and two time-resolved signals retrieved at 1 MHz and 56 MHz from the spectrum obtained through EMI measurement. The captured time signals are labelled by EMI experts according to the occurring fault such as PD, corona, arcing etc. or non-faulty signals containing noise. In this paper, a total of 8 different signal types are treated as different classes in a multi-class classification problem. Details on the data are presented in Table 3, from which a large and imbalanced data set is observed. Example time series signal of Arcing and Corona faults are presented in Fig. 7 along with their respective Bispectrum representations as images since these are treated as image classification problem. This analysis yields a 2D mapping of the non-linear interaction between frequency pairs in the signal. It is observed that Arcing and Corona have unique patterns, and more importantly preserving the complex nature of the Bispectrum provides more information.

4.2. Data analysis and results

The end-to-end steps for signals analysis are described in this section followed by classification results. First, the captured time series signals from all classes are segmented in non-overlapping windows with length of 4000 samples. This approach is performed for data augmentation and the window size is chosen based on the smallest measured signal as the signals length vary between 4 k-240 k samples. Next, all data segments are normalised between 0 and 1, then the zero mean signal is obtained prior to Bispectrum calculation as described in Section 3.1. The obtained Bispectrum images are of size 64x64x2, for the complex and real networks with 2 channels. For real networks with 1 channel, the Bispectrum images are 64x64x1.

Finally, the labelled data is randomly split into 70% to train the classifiers, 15% for validation to tune the weights and the remaining 15% for testing to evaluate the learned model. It is observed that the number of training instances in PD does not match the split ratio, this is because a previously trained model with large PD data (10587 instances) compared to the remaining classes resulted in a bias towards PD. This is a common problem in ML known as class imbalance, one method of minimising the effect of this issue is to drop data instances, in this case from PD class, and hence reduce the bias. The six deep networks are implemented in Keras with a learning rate starting at 0.001 then set to 0.0001 after 10 epochs. The models are trained over 100 epochs with a batch size of 32 and ADAM optimization. It is observed that the dataset is imbalanced, thus automatic class weighting approach is followed during training. The utilised loss function is categorical cross-entropy for multi-class classification. The final step is performed 10 times for reliability. In order to distribute similar computation across all networks, the models with 1 channels are implemented with a double number of convolutional filters. Performance results are presented in Table 4 which includes the best obtained classification accuracy, average accuracy and the standard deviation (std) from the mean of the obtained accuracies across the 10 iterations.

Table 4

Testing results. Best performance is underlined.

Classifier	Best accuracy %	Avg. accuracy %	\pm std
CNN:1channel	72	68	0.029
CNN:2channels	88	81	0.033
Complex CNN	82	76	0.031
ResNet:1channel	89	80	0.069
ResNet:2channels	<u>98</u>	<u>97</u>	<u>0.010</u>
Complex ResNet	96	93	0.029

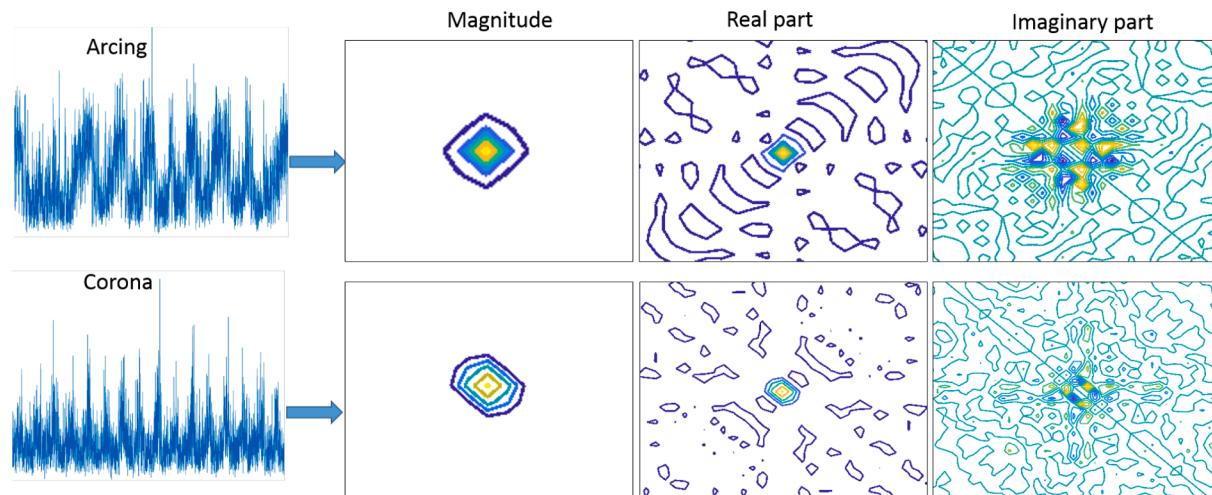


Fig. 7. Examples of Arcing and Corona signals and their different Bispectrum representations.

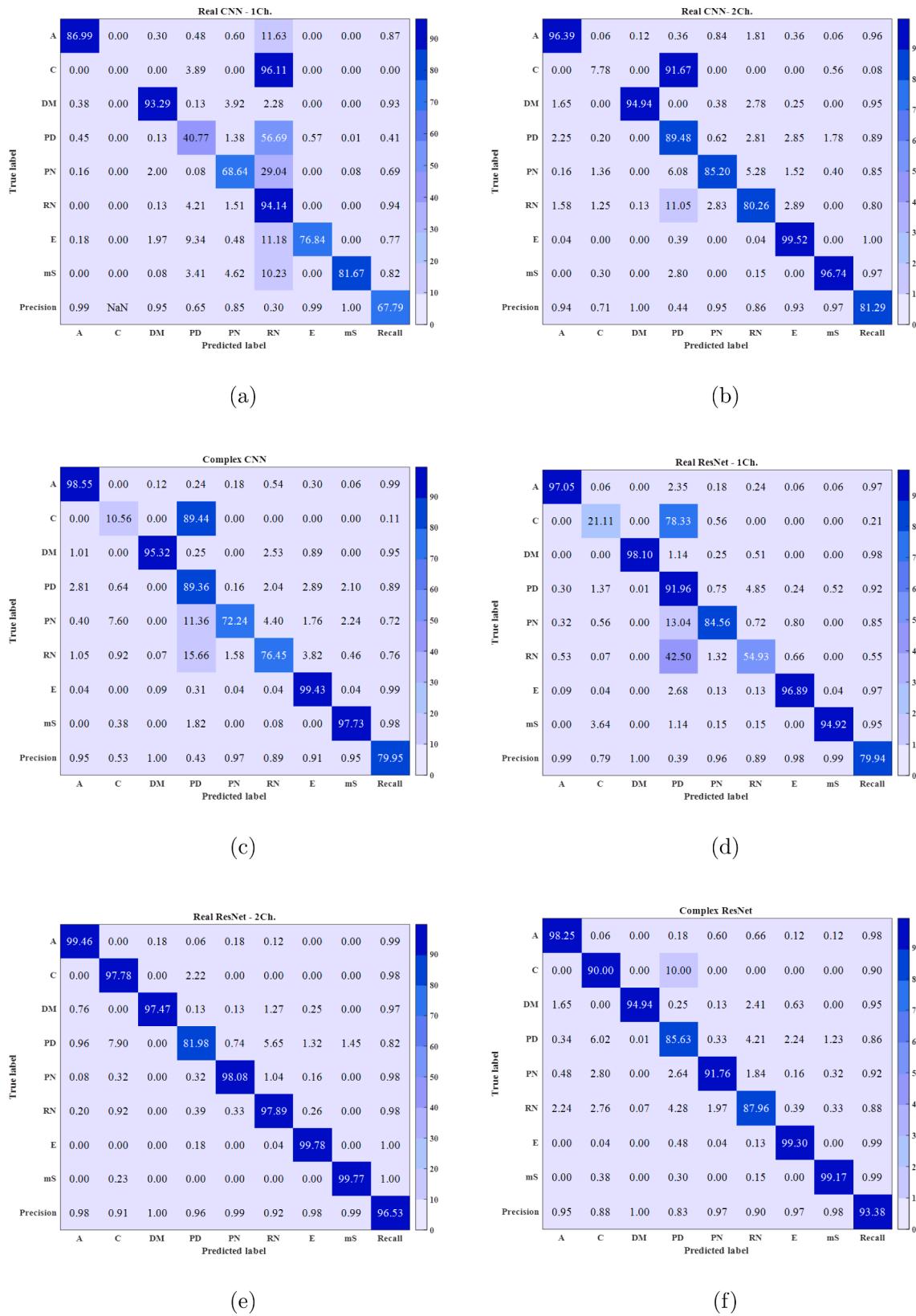


Fig. 8. Average confusion matrix for (a) CNN-1channel (b) CNN-2channels (c) Complex CNN (d) ResNet-1channel (e) ResNet-2channels (f) Complex ResNet.

It is seen that the ResNet with 2 channels achieves the best performance with regards to all measures, while real CNN has the lowest performance, in particular the 1 channel model. The average confusion matrices, across the 10 iterations, for each model are illustrated in Fig. 8

along with precision and recall measures. A lower classification is performed by the real CNNs which is improved by Complex CNN, however a high confusion between the classes, particularly with PD, is still existent. The different ResNet models improve upon this confusion, where a

practical and reasonable performance which satisfies both the inter and intra class accuracies is achieved by the real ResNet with 2 channels. It is observed that most models exploiting both real and imaginary parts of the Bispectrum, including ResNet with 2 channels, Complex ResNet and Complex CNN, result in an improved performance as ranked respectively. This could be due to the advantage of utilising the imaginary part of the Bispectrum which reveals phase information that aids to distinguish between the classes more efficiently.

5. Conclusion

This paper leveraged the Bispectrum method proposed in previous work for the classification of real-world insulation faults, which were captured at various HV power plants and labelled by EMI experts. Results show that using the complex-valued Bispectrum enhances the performance in both inter and intra-class classification. ResNet with 2 channels achieved the best performance while the performance of Complex ResNet and Complex CNN could be improved in the future. Findings in this preliminary study are promising and researchers are encouraged to further explore deep complex networks and exploit the real and complex-valued Bispectrum. Future work will investigate the implementation of this work in condition monitoring of HV power plants.

CRediT authorship contribution statement

Imene Mitiche: Methodology, Investigation, Software, Writing - original draft. **Mark D. Jenkins:** Writing - review & editing, Software. **Philip Boreham:** Funding acquisition, Resources. **Alan Nesbitt:** Data curation, Project administration. **Gordon Morison:** Conceptualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the funder of this project Doble Engineering and their EMI experts for providing labelled data.

References

- 33.100.10-Emission (2015). CISPR/CIS/A - radio-interference measurements and statistical methods. International Electrotechnical Commission (IEC).
- Bhalke, D., Rama Rao, C., & Bormane, D. (2016). Hybridisation of mel frequency cepstral coefficient and higher order spectral features for musical instruments classification. *Archives of Acoustics*, 41, 427–436.
- Bin, G., Gao, J., Li, X., & Dhillon, B. (2012). Early fault diagnosis of rotating machinery based on wavelet packets-empirical mode decomposition feature extraction and neural network. *Mechanical Systems and Signal Processing*, 27, 696–711.
- Boas, M. L. (2006). *Mathematical methods in the physical sciences*. Hoboken, NJ: Wiley.
- Chua, K. C., Chandran, V., Acharya, U. R., & Min Lim, C. (2010). Application of higher order statistics/spectra in biomedical signals-a review. *Medical Engineering and Physics*, 32, 679–689.
- Duan, L., Hu, J., Zhao, G., Chen, K., He, J., & Wang, S. X. (2019). Identification of partial discharge defects based on deep learning method. *IEEE Transactions Power Delivery*, 34, 1557–1568.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas*.
- Huang, W., Ding, H., & Chen, G. (2018). A novel deep multi-channel residual networks-based metric learning method for moving human localization in video surveillance. *Signal Processing*, 142, 104–113.
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning, Lille*.
- Iorkyase, E. T., Tachtatzis, C., Glover, I. A., Lazaridis, P., Upton, D., Saeed, B., & Atkinson, R. C. (2019). Improving rf-based partial discharge localization via machine learning ensemble method. *IEEE Transactions Power Delivery*, 34, 1478–1489.
- Jinying, F. A. H., Hongxia, S. B. P., & Shihua, T. C. B. (2010). Bispectrum entropy feature extraction and its application for fault diagnosis of gearbox. In *International Conference on Fuzzy Systems, Barcelona*.
- Kicinski, W., & Szczepanski, A. (2004). Quadratic phase coupling phenomenon and its properties. *Hydroacoustics*, 7, 97–106.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of ACM*, 60, 84–90.
- Lanata, A., Valenza, G., Mancuso, C., & Scilingo, E. (2011). Robust multiple cardiac arrhythmia detection through bispectrum analysis. *Expert Systems with Applications*, 38, 6798–6804.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings IEEE*, 86, 2278–2324.
- Li, C., Sanchez, R. V., Zurita, G., Cerrada, M., Cabrera, D., & Vásquez, R. E. (2016). Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mechanical Systems and Signal Processing*, 76–77, 283–293.
- Li, S., & Liu, Y. (2010). Feature extraction of lung sounds based on bispectrum analysis. In *Proceedings of the third International Symposium on Information Processing, Qingdao*.
- Li, X., Zhang, W., & Ding, Q. (2019). Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal Processing*, 161, 136–154.
- Mitiche, I., Morison, G., Nesbitt, A., Hughes-Narborough, M., Stewart, B. G., & Boreham, P. (2018). Classification of emi discharge sources using time-frequency features and multi class support vector machine. *Electric Power Systems Research*, 163, 261–269.
- Mitiche, I., Morison, G., Nesbitt, A., Stewart, B. G., & Boreham, P. (2018). Entropy-based feature extraction for electromagnetic discharges classification in high-voltage power generation. *Entropy*, 20, 549.
- Nikias, C. L., & Petropulu, A. P. (1993). *Higher-order spectral analysis: a nonlinear signal processing framework*. Prentice Hall.
- Robles, G., Parrado-Hernandez, E., Ardila-Rey, J., & Martinez-Tarifa, J. M. (2016). Multiple partial discharge source discrimination with multiclass support vector machine. *Expert Systems with Applications*, 55, 417–428.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations, San Diego*.
- Sun, L., Feng, Z., Lu, N., Wang, B., & Zhang, W. (2019). An advanced bispectrum features for eeg-based motor imagery classification. *Expert Systems with Applications*, 131, 9–19.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition, Boston*.
- Timperley, J. E. (2008). Generator condition assessment through emi diagnostics. In *ASME Power Conference Florida*.
- Timperley, J. E. (2012). Application of emi diagnostics to hydro generators. In *Presentation At The Northwest Hydroelectric Association Annual Conference, Portland*.
- Timperley, J. E., Buchanan, D., & Vallejo, J. M. (2018). Electric generation condition assessment with electromagnetic interference analysis. *IEEE Transactions on Industrial Applications*, 54, 1921–1929.
- Timperley, J. E., & Vallejo, J. M. (2017). Condition assessment of electrical apparatus with emi diagnostics. *IEEE Transactions on Industrial Applications*, 53, 693–699.
- Timperley, J. E., Vallejo, J. M., & Nesbitt, A. (2014). Trending of emi data over years and overnight. In *Proceedings of IEEE Electrical Insulation Conference Philadelphia*.
- Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J. F., Mehri, S., Rostamzadeh, N., Bengio, Y., & Pal, C. J. (2018). Deep complex networks. In *International Conference on Machine Learning, Stockholm*.
- Wang, R., Li, W., & Zhang, L. (2019). Blur image identification with ensemble convolution neural networks. *Signal Processing*, 155, 73–82.
- Wu, S., Xu, J., Zhu, S., & Guo, H. (2018). A deep residual convolutional neural network for facial keypoint detection with missing labels. *Signal Processing*, 144, 384–391.
- Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolution network. In *Deep Learning Workshop in International Conference on Machine Learning, Lille*.
- Zhang, X., Wang, B., & Chen, X. (2015). Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine. *Knowledge-Based Systems*, 89, 56–85.