1D-CNN based real-time fault detection system for power asset diagnostics

Mitiche, Imene; Nesbitt, Alan; Conner, Stephen; Boreham, Philip; Morison, Gordon

Published in:
IET Generation, Transmission & Distribution

DOI:
10.1049/iet-gtd.2020.0773

Publication date:
2020

Document Version
Author accepted manuscript

1D-CNN based Real-Time Fault Detection System for Power Asset Diagnostics

Mitiche, Imene; Nesbitt, Alan; Conner, Steve; Boreham, Philip ; Morison, Gordon

Published in:
IET Generation, Transmission & Distribution

Publication date:
2020

IET Journals
The Institution of Engineering and Technology

1D-CNN based Real-Time Fault Detection System for Power Asset Diagnostics

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

I. Mitiche[1*], *A. Nesbitt* [1], *S. Conner*[2], *P. Boreham*[2], *G. Morison*[1]

[1] *School of Engineering and Built Environment, Glasgow Caledonian University, Glasgow, United Kingdom*
[2] *Innovation Centre for Online Systems, Doble Engineering, Bere Regis, United Kingdom*
* *E-mail: imene.mitiche@gcu.ac.uk*

Abstract: Electromagnetic Interference (EMI) diagnostics aid in identifying insulation and mechanical faults arising in High Voltage (HV) electrical power assets. EMI frequency scans are analysed to detect the frequencies associated with these faults. Time-resolved signals at these key frequencies provide important information for fault type identification and trending. An end-to-end fault classification approach based on real-world EMI time-resolved signals was developed which consists of two classification stages each based on 1D-Convolutional Neural Networks (1D-CNN) trained using transfer learning techniques. The first stage filters the in-distribution signals relevant to faults from out-of-distribution signals that may be collected during the EMI measurement. The fault signals are then passed to the second stage for fault type classification. The proposed analysis exploits the raw measured time-resolved signals directly into the 1D-CNN which eliminates the need for engineered feature extraction and reduces computation time. These results are compared to previously proposed CNN-based classification of EMI data. The results demonstrate high classification performance for a computationally efficient inference model. Furthermore, the inference model is implemented in an industrial instrument for HV condition monitoring and its performance is successfully demonstrated in tested in both a HV laboratory and an operational power generating site.

1 Introduction

Traditional condition monitoring of industrial High-Voltage (HV) power plants often relies on expert diagnosis. This consists of on-site data measurement and analysis to identify potential faults arising within the asset under assessment. Early fault detection allows action to be taken on asset maintenance or replacement. This can be achieved through continuous on-line monitoring, which reduces the risks associated with asset failure including costs, safety and damage [1] leading to power outage [2]. Continuous on-line monitoring generates large quantity of data that needs to be processed and analysed quickly and accurately in order to efficiently determine the health of an asset. Electromagnetic Interference (EMI) diagnostics is one of the asset health monitoring techniques frequently deployed by the power industry for Partial Discharge (PD) and other fault type detection [3]. This method involves measurement of the conducted transient voltages to earth produced by PD and similar phenomena. The measurements are acquired in both the frequency and time domain. Within the EMI diagnostic method the frequency spectrum shows peaks at key frequencies that are associated with faults and time-resolved signals are obtained at these frequencies to reveal the patterns related to fault types. Currently, the fault type identification requires input from highly experienced electrical engineers, called EMI experts, who manually analyse the acquired measurements. As the demand for continuous on-line monitoring of power assets increases, it may be difficult for expert-based diagnostics to meet this demand. Since expert analysis is manual, it is time consuming and not practical for large volumes of data. Hiring more experts is costly and engineers require time and experience to reach the adequate level of expertise. Furthermore, the dependancy on expert judgement alone may introduce human error leading to mis-interpreted measurements.

In order to cope with the data analysis demand challenged by EMI-based continuous on-line monitoring and to meet fast and accurate fault diagnosis, a computationally inexpensive Machine Learning (ML) based solution is proposed in this paper. A two stage ML structure is trained on the EMI time-resolved signals measured on two datasets called in-distribution and out-of-distribution data. The in-distribution data consist of signals measured at multiple operating power stations and examined by experts and labelled according to the interpreted information, where fault patterns were identified in numerous signals. The out-of-distribution data, known to have statistical distributions that differ from that of fault pattern signals, are used to train the first stage ML model in order to filter such signals from the in-distribution signals. The latter are passed to the second stage ML model for a detailed fault type classification. The classification result could be used to support experts in their diagnosis or to interpret the measured signal for a non-expert regarding the health of the diagnosed asset. The focus in this work is to develop of an end-to-end analysis framework which is computationally inexpensive for an edge device implementation and its usage in real-world application. 1D-Convolutional Neural Networks (1D-CNN) architectures are proposed for both ML models to extract features and classify the raw time-resolved signals. 1D-CNN performs one-dimensional convolution and dimensionality reduction operations, and it jointly combines the feature extraction, feature transformation, data fusion and classification steps into one framework. These attributes save computing resources and hence power consumption and processing time. Therefore such architecture is ideally suited for edge implementation. Fig. 1 summarises the process of previously acquired field survey data labelled by EMI experts and how this was used to develop the ML based diagnostic toolset implemented into an edge device for power asset condition monitoring.

1.1 Related work

Intelligent systems which exploit ML and Deep Learning approaches have been successful in asset condition monitoring within the power industry. Various studies in the literature employed 1D-CNN to vibration signals for Structural Health Monitoring (SHM) [4] [5]. The latter were successfully analysed by 1D-CNNs in other papers for bearing [6] and rotating machinery fault diagnosis [7] [8]. The fast computation and straightforward implementation of 1D-CNNs received attention from real-time based applications such as motor fault diagnosis [9], SHM [10], and switch fault detection in HV converters [11] where 1D-CNN was used on raw current and voltage signals. Medical applications of 1DCNN include patient health

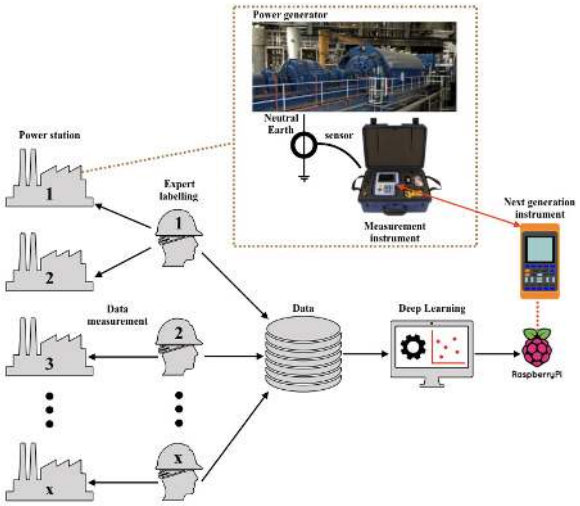


Fig. 1: End-to-end EMI diagnostic with ML-based assistant toolset

monitoring specifically heart beat related anomalies, such as arrhythmia, based on the measured electrocardiogram signals [12] [13]. With regards to Partial Discharge (PD) identification, 1D-CNN was successfully applied to acoustic emission signals for insulation diagnostics in power transformers [14].

1.2 ML for EMI diagnostics literature

Previously the authors of this paper proposed Deep Learning methods based on 2D-CNN for EMI signal classification [15] [16]. This work provided highly accurate results at the expense of high computational complexity, due to signal transformation and feature extraction using the Bispectrum and the 2D-CNN for classification. Despite the successful work proposed in the literature, there is a lack in insulation fault detection using EMI signals. To the best of the authors knowledge the application of 1D-CNNs to the raw EMI time-series signals for fault type classification has not been presented in the literature. The contributions of this work primarily consist of the following:

1. The development of a low computation and highly accurate end-to-end model for automatic fault identification and out-of-distribution signal rejection.
2. Exploited transfer learning to increase the amount of training data for in-distribution classification.
3. Edge implementation of the proposed algorithm to embed ML-based diagnostic toolset in the next generation instruments.

1.3 Paper layout

The remaining structure of this paper is outlined as follows:

Section 2-Describes the measurement of both field and non-field EMI data,
Section 3-Introduces the proposed analysis approach,
Section 4-Details the experimental method of the proposed approach using the labelled data, including the obtained results,
Section 5-Describes the edge implementation of the proposed approach, and results of further testings in HV laboratory and power sites,
Section 6-Discussion and conclusion, providing insight to future work.

2 EMI Measurement

This section details the measurement method used to collect time-resolved signals on which the proposed models are trained and tested. Details on the field signals (in-distribution data) and out-of-distribution data are also provided.

2.1 Field data collection

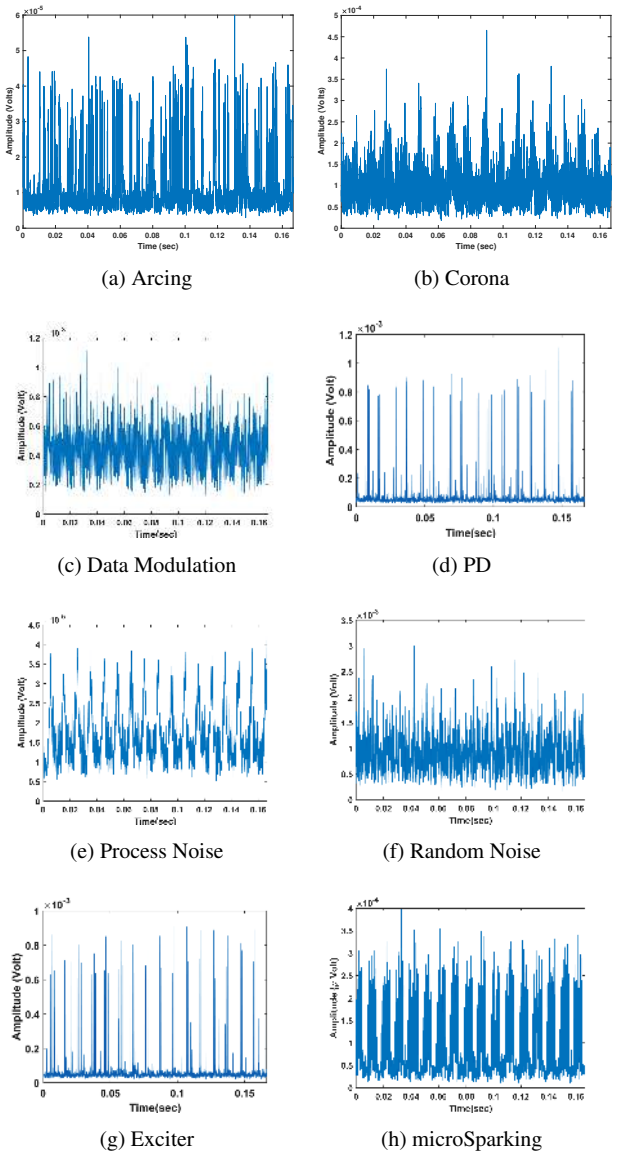
The field data were collected using the EMI method [3] which can be used for both insulation deterioration and mechanical faults detection. Past investigations found that PD was present in a hydrogen cooled generator which indicated looseness in the stator endwindings [17]. Subsequently, the stator was repaired to prevent a winding failure. Similar field investigations and surveys were performed by EMI experts and the reported insulation fault measurements were logged.

The field data were measured from the earth connection of operating motors, generators and transformers. A non-invasive High Frequency Current Transformer (HFCT) sensor was bound around the earth connection in order to collect the conducted or radiated EMI voltage transients that flow due to fault occurrence. This measurement follows the International Special Committee on Radio Interference (CISPR)16 standard [18] for Electromagnetic Compatibility in disturbances measurement and immunity. The frequency and time domain measurements were recorded by a survey device, called PD Survey 200 (PDS200), which has the capability to capture the EMI signature over a frequency spectrum ranging from 50 kHz to 100 MHz. The key frequencies in the recorded EMI spectrum were selected to retrieve the time-resolved signals at a sampling frequency of 24kHz using AM demodulation. The measurements were analysed by an EMI expert to identify the fault type and location

Table 1 Dataset information

Label	# Signal instances
Arcing	546
Corona	59
Data Modulation	261
Partial Discharge	5057
Process Noise	414
Random Noise	449
Exciter	749
microSparking	433
Total	7968

[19] and labels for the time-resolved signals were assigned as presented in Table 1. Example signals are shown in Fig. 2. The data are used in a multi-class classification solution to determine the fault, if any, captured from the EMI measurement for asset health monitoring. However, in real-world application non-fault related signals, referred to as out-of-distribution signals, may be collected in the frequency range at which the PDS200 captures EMI data. Examples of these signals include radio AM, FM and different types of noise. Therefore, such data were collected in order to mitigate the closed-world learning by developing a binary classifier to distinguish data in Table 1 (in-distribution) from the out-of-distribution data.



(a) Arcing (b) Corona

(c) Data Modulation (d) PD

(e) Process Noise (f) Random Noise

(g) Exciter (h) microSparking

Fig. 2: Labelled EMI signal examples

2.2 *Out-of-distribution data collection*

The out-of-distribution data were collected using a DoblePRIME PD-Guard instrument set up in the ICOS lab at Glasgow Caledonian University. This is an ordinary office environment and there is no HV equipment in or anywhere close to the lab. DoblePRIME PD-Guard operates in the same way as the PDS200 in terms of EMI measurement and time-resolved signals sampling. The data in this experiment were collected through four channels of the PD-Guard, where each was set-up differently as follows.

2.2.1 Channel 1: was connected to a Doble DA100 directional antenna, which has a frequency range of 250MHz-1GHz. The antenna was aimed across the lab towards the window. The DN50-80 preset was selected for this channel, causing the unit to scan a frequency range of 50MHz-1GHz. As the lab is on the 6th floor in a city centre location, it would be expected to collect signals such as mobile phone and digital TV transmissions.

2.2.2 Channel 2: was connected to a Doble TEV probe. This is designed for collecting PD inside metal enclosures and has a useful frequency range of around 100MHz-1.2GHz. It was placed on an Ethernet switch connected to the lab network, with the intention of collecting computer and datacoms related noise. The DN50-80 preset was also used for this channel.

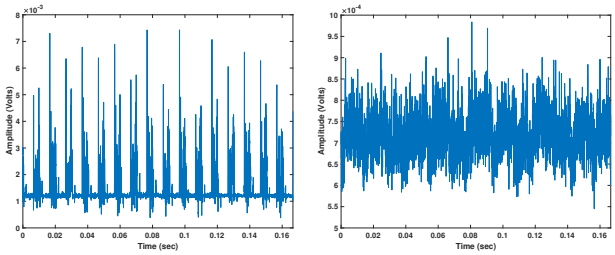
2.2.3 Channel 3: was connected to a Doble HFCT Mini. This sensor is normally used for collecting Radio Frequency (RF) emitted from transformer bushing taps. This was connected to the mains through a small low voltage transformer, with the intention of collecting a representative sample of mains-borne noise. The HFCT Mini preset was set to a frequency range of 100kHz-10MHz.

2.2.4 Channel 4: was connected by a long coaxial cable to a Watson W-881 whip antenna with a frequency range of 25-1900MHz. This was placed on a shelf next to the window. The intention was to collect lower frequency radio signals that would complement those collected by the antenna on Channel 1. So, the HFCT-300 preset was chosen for this channel with a frequency range of 50kHz-100MHz.

The PD-Guard was set up the same as in its normal EMI and PD detection application. Every 2 hours it performed the following actions on each channel:

1. Perform a spectrum analysis recording both the peak and average power.
2. Save the analysis result.
3. For each band in the spectrum, find the frequency where the ratio of the peak to average power is highest.
4. Save the time-resolved signals at each of the frequencies found above.

A total of 13360 signals were collected during one week period. The time-resolved signals were examined by an EMI expert and confirmed the absence of insulation fault signature such as PD. The signals were not labelled individually but they were all grouped under one category (out-of-distribution) with the assumption that they do not contain any of the faults or signal types presented in Table 1. Example of two out-of-distribution collected time-resolved signals are presented in Fig. 3.



(a) Amplitude Modulation (b) Unknown

Fig. 3: Collected signal examples from non-field and non-laboratory environment

3 Analysis Method

This section describes 1D-CNN algorithm and explains how it is used to construct the proposed end-to-end ML-based diagnostic algorithm.

3.1 1D-CNN theory

1D-CNNs were introduced in [4]-[14] to process 1D data and have been proven to have computational benefits among many others. For instance, it was demonstrated that shallow 1D-CNNs are successful in challenging tasks on 1D data, whereas 2D-CNNs often require a larger number of layers to extract meaningful features from 2D data. This allows for easier implementation on CPUs or into real-time and low cost devices. In 1D-CNNs, the convolution operation is performed on data vectors, where input signal vector $\mathbf{x}$ of length N is convolved with a filter vector $\boldsymbol{\omega}$ of length L. This operation can be expressed in (1), and results in a 1D output layer $\mathbf{c}$ of length $(N - L + 1)$, without applying zero-padding.

$$\mathbf{c}(j) = f\Big(\sum_{i=0}^{L-1} \boldsymbol{\omega}(i)\mathbf{x}(j-i) + b\Big), j = 0, 1, ..., N-1 \quad (1)$$

where b is the bias term and $f(\cdot)$ is a non-linear function, which is Rectified Linear Unit (ReLU) [20] in this work. In the proposed architecture, each convolution layer is followed by a max pooling layer where the maximum value in a kernel window function u, with size $m \times 1$ and stride s, is taken over an input vector $\mathbf{c}$ resulting in an output vector $\mathbf{d}$ defined as:

$$\mathbf{d} = \max\Big(u(m \times 1, s)\mathbf{c}\Big) \quad (2)$$

The Fully Connected (FC) layer takes as input the feature map output from the preceding layer and connects all its neuron nodes. Batch Normalisation (BN) is used to normalise and scale the FC layer to adjust its distribution, which allows for faster training [21]. An activation function (e.g. Softmax, Sigmoid etc.) determines the predicted output which is used to compute the cross-entropy loss l, during the network training, as:

$$l = \sum_{k=1}^{C} \mathbf{t}(k) log\Big(\mathbf{p}(k)\Big) \quad (3)$$

where C is the number of classes, $\mathbf{t}(k)$ is the comparison result between the predicted label and the ground truth label such that $\mathbf{t}(k) = 1$ if the input belongs to class k, and $\mathbf{t}(k) = 0$ otherwise, and $\mathbf{p}(k)$ is the predicted probability that the input belongs to class k. Backpropagation is performed using the Adam optimiser [22] and exploits the calculated loss to update the network's weights and biases. The detailed architecture of the implemented 1D-CNNs is presented in Table 2 for the first stage and Table 3 for the second stage. It is observed that both networks have the same depth and parameters, the main difference lies within the final activation layer. The first stage employs Sigmoid activation function for the binary classification problem. The second stage utilises Softmax activation for the multi-class problem. A shallow depth of the networks was chosen for computation purposes. The number of layers, the size and number of convolution kernels in a network significantly influence the feature extraction and classification results [23]. A set of different convolution and max pooling parameters were set in multiple experiments to extract a maximum number of different features at a reasonable computation. The resulting optimum parameters from a grid search method were employed in the proposed model architectures.

3.2 Transfer learning

The aim behind transfer learning is to re-use knowledge derived from a particular learning problem to enhance the learning process on a related problem [24]. The concept of transfer learning in Deep Learning refers to setting the starting point of a model as a previously learned model [25]. For EMI signal type classification with 1D-CNN, the early convolutional layers of the trained first stage model are set as the starting weights for the second stage model. The idea is that the second stage model uses the general and low-level features, extracted by the first model [26], that can be found across the EMI signals and the last layers identify unique features within each of the 8 groups. In this paper the strategy of fine tuning the whole network is employed.

3.3 End-to-end approach

The proposed data processing approach performs two learning and classification stages. In the first stage, a binary 1D-CNN classifier is trained on both the field survey data and the out-of-distribution data. This allows for a decision to whether proceed with the fault type classification or to report the measured signals that do not contain a previously seen event as unknown. Here unknown is the label which represents the out-of-distribution data during the inference. In the second stage, a multi-class 1D-CNN is trained on the field survey data. The training and validation data in this stage is similar to the training and validation set used in the first stage. To further reduce computation and improve the second stage performance, the learned weights of the first four layers (two convolution and two max pooling layers) are transferred from stage 1 to stage 2. This means that the 1D-CNN training in stage 2 is initialised with pre-trained weights and biases. The latter from the remaining layers of 1D-CNN is trained to extract features related to the 8 classes for a better learning and pattern recognition, and to discard irrelevant features extracted and learned from the out-of-distribution data. An additional experiment is conducted by training the 1D-CNN of stage 2 without transfer learning and performances comparison is provided in the next section.

Fig. 4 illustrates the inference model of the proposed approach and Algorithm 1 details the procedure steps of this inference. F_1 is the learned function of the first stage ($S1$) with output $\hat{\mathbf{y}}_{S1}$, which decides whether the input signal $\mathbf{x}$ belongs to the in or out-of-distribution data. F_2 is the learned function of the second stage ($S2$) and its output $\hat{\mathbf{y}}_{S2}$ indicates the class label among the classes shown in Table 1.

Algorithm 1 Classification inference

Input: $\mathbf{x} = \{x_1, x_2, ..., x_N\}$
$\hat{\mathbf{y}}_{S1} = F_1(\mathbf{x})$
if $argmax(\hat{\mathbf{y}}_{S1}) = 1$ **then**
  return "Unknown"
else $\hat{\mathbf{y}}_{S2} = F_2(\mathbf{x})$
  return Class: $argmax(\hat{\mathbf{y}}_{S2})$
end if

4 Training and Testing Experiment of the 1D-CNNs

This section details the ML models training procedure and presents the testing and computation results of this experiment.

4.1 Models training on EMI data

Prior to training and testing the models, two pre-processing steps were performed on the time-resolved signals: normalise the signals amplitude between 0 and 1, and split each signal into segments of 4000 sample for data augmentation resulting in the number of signal instances provided in Table 1. The 1D-CNNs were implemented in Tensorflow [27] and Keras [28] application program interface. The training experiment was performed using the ten-fold cross validation method, in each fold 20% of the training data was held for validation to fine tune the model weights.

The models have learnable parameters during their training. The weights in stage 1 and stage 2 (new learning) are initialised using the Xavier initialisation method [29] to avoid exploding or vanishing outputs from the activation layers in the forward

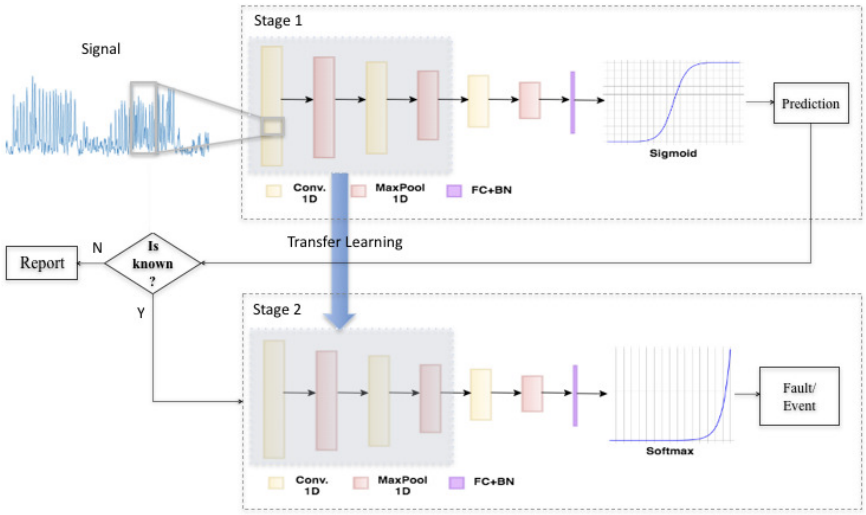


Fig. 4: 1D-CNN based classification inference approach

Table 2 1D-CNN architecture for stage 1

Layer	Kernel size	Stride	# Filters	Data shape
Input				(4000,1)
Conv. 1D_1	3×1	1	64	(3998, 64)
MaxPool 1D_1	3×1	3×1		(1332, 64)
Conv. 1D_2	3×1	1	64	(1330, 64)
MaxPool 1D_2	3×1	3×1		(443, 64)
Conv. 1D_3	3×1	1	64	(441, 64)
MaxPool 1D_3	2×1	2×1		(220, 64)
FC + BN				(100)
Activation (Sigmoid)				(2)

Table 3 1D-CNN architecture for stage 2

Layer	Kernel size	Stride	# Filters	Data shape
Input				(4000,1)
Conv. 1D_1	3×1	1	64	(3998, 64)
MaxPool 1D_1	3×1	3×1		(1332, 64)
Conv. 1D_2	3×1	1	64	(1330, 64)
MaxPool 1D_2	3×1	3×1		(443, 64)
Conv. 1D_4	3×1	1	64	(441, 64)
MaxPool 1D_4	2×1	2×1		(220, 64)
FC + BN				(100)
Activation (Softmax)				(8)

pass. In this initialisation method, the layer?s weights are set to random values from a uniform distribution defined in the range $[-\frac{\sqrt{6}}{\sqrt{n_j+n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j+n_{j+1}}}]$, where n_j is the number of the incoming connections from the previous layer and n_{j+1} is the number of outgoing connections to the next layer. For transfer learning, the first four layers are initialised with weights of the trained first stage model, whereas the last remaining layers are initialised with the Xavier initialisation. The weights are updated during training through the back-propagation using the Adam optimiser with a training scheduler set to start at a learning rate of 0.001 then decreases to 0.0001 after the first 10 epochs to achieve convergence.

Early stopping criteria was implemented, to prevent overfitting the data, over 100 epochs. This method involves stopping the training if the validation loss does not improve with a patience number of epochs of 10. The models were compiled with different loss functions, such that 1D-CNN in stage 1 employed the binary cross-entropy loss for the binary classification task, and 1D-CNN in stage 2 employed the categorical cross-entropy loss for the multi-class task. It is observed that the data classes are unbalanced, thus the Keras embedded attribute of weighing the classes in calculating the loss during training was utilised.

Figure 5 illustrates the loss and accuracy during training and validation of each model. It is shown that the models has converged in the early epochs by the early stopping mechanism before the 100 epochs were reached to avoid over-training and overfit the data.

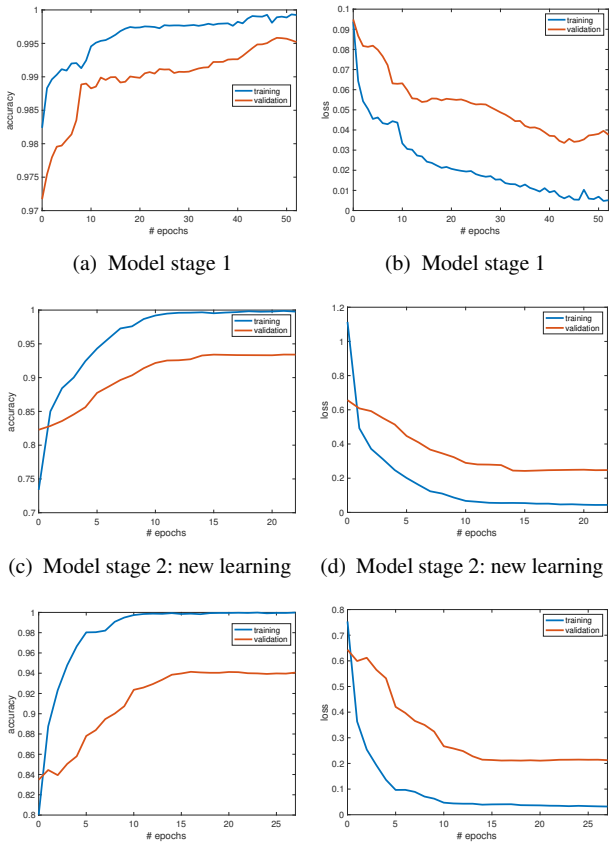
4.2 Results & Discussion

The average performance resulting from the ten folds for each 1D-CNN testing as well as the previous methods proposed in the literature, are presented in Table 4. Stage 1 achieved the top performance in all metrics and solved efficiently the binary classification problem, where only one boundary is required to separate the two data categories. A 3D visualisation of the learned features from this stage is provided in Fig. 7 to illustrate this separation in the feature space. This was obtained by using the t-Stochastic Neighbour Embedded (t-SNE) clustering algorithm [31] on the FC layer output features of the 1D-CNN. The approach involves a projection of the 100 dimension features obtained at the FC layer to a 3D feature space. The second stage solved a multi-class problem of 8 categories which resulted in a light performance drop when compared to stage 1. However, this performance is considered high and sufficient for a real-world multi-class problem and is in line with the results obtained in [15] and [16]. Note that the method in [16] used a sub-set of data that is used in this paper, this sub-set consists of balanced data classes. Furthermore, the 1D-CNN based approach has computation advantage with more realistic dataset, larger and unbalanced. The normalised confusion matrices for each of the second stage models are presented in Fig. 6 to show the performance quality in the multi-class problem. Further classification metrics for each class, called specificity which represent the true positive rate and sensitivity representing the true negative rate, are presented in Table 5. It is observed that transfer learning outperforms the new learning approach across the majority of the presented metrics with a enhanced classification on classes with lower number of examples such as corona and random noise. These results affirm the previous claims on improved performance when employing transfer learning [24] [32] [33]. The t-SNE of the FC layer features in stage 2 (transfer learning) is illustrated in Fig. 8. It is seen that features of the different classes are well clustered and separated in overall. The t-SNE graphs reflect the obtained test accuracy and indicate that the extracted features from the 1D-CNNs are substantial and discriminative.

The computation time of each model training during one fold was recorded as well as the end-to-end inference and results are presented in Table 6. In order to provide a fair comparison, each model was trained under the same processing resources (one GPU) and conditions. It is worthwhile to point out that the training time of the model with transfer learning is lower than the new learning. The end-to-end inference time, which involves the steps in Algorithm 1, is calculated for the classification of one raw signal and is only 0.25 seconds. The latter achieves a significant reduction in inference time from the previously proposed methods based on 2D-CNN. This is due to

Table 4 Average classification results over the ten folds. Bold font indicates the top performance in the multi-class problem

Metrics	test acc. %	test loss	precision	recall	specificity	F1-score
Stage 1	99.4	0.037	0.99	0.99	0.99	0.99
Stage 2: new learning	90.89	0.338	0.90	0.89	0.97	0.89
Stage 2: transfer learning	93.8	**0.216**	0.93	0.93	**0.98**	**0.93**
Method in [15]	92.24	-	-	-	-	-
Method in [16] (RV-CNN)	86.94	-	0.91	0.87	-	-
Method in [16] (CV-CNN)	**95.37**	-	**0.95**	**0.95**	-	-
Method in [30] SVM based	87.8	-	0.68	0.87	0.97	0.76

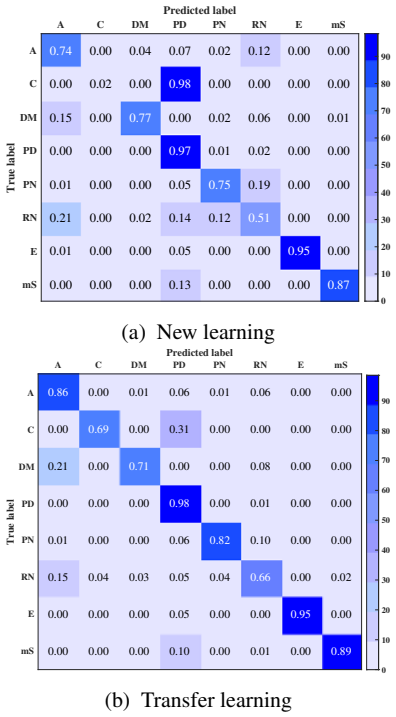


(a) Model stage 1 (b) Model stage 1

(c) Model stage 2: new learning (d) Model stage 2: new learning

(e) Model stage 2: transfer learning (f) Model stage 2: new learning

Fig. 5: Training and validation accuracy and loss curves for each model



(a) New learning

(b) Transfer learning

Fig. 6: Confusion matrix resulting from each model in stage 2. A: arcing, C: corona, D: Data Modulation, PD: partial discharge, PN: process noise, RN: random noise, E: exciter, mS: microSparking

Table 5 Specificity and sensitivity results from stage 2 classification for each class

Metrics	New learning		Transfer learning	
Model	Specificity	Sensitivity	Specificity	Sensitivity
Arcing	**0.97**	0.74	**0.97**	**0.83**
Corona	**0.99**	0	**0.99**	**0.66**
Data Modulation	**0.99**	**0.72**	**0.99**	**0.72**
Partial Discharge	0.89	**0.97**	0.93	**0.97**
Process Noise	0.98	0.73	**0.99**	**0.80**
Random Noise	**0.97**	0.50	**0.97**	**0.64**
Exciter	0.99	0.92	**1**	**0.93**
microSparking	0.99	0.86	**1**	**0.88**
Average	0.97	0.68	**0.98**	**0.80**

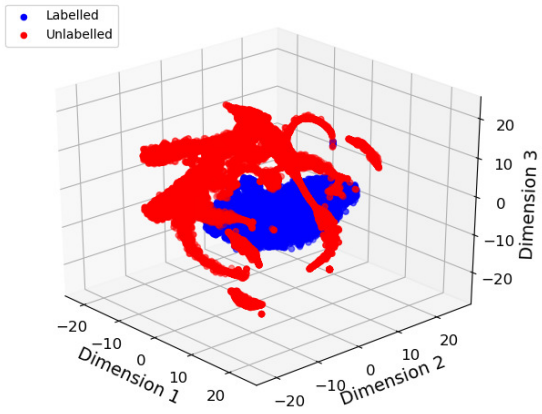


Fig. 7: t-SNE clustering of stage 1 feature space

the complex computation of the Bispectrum mapping of one signal, its feature extraction and classification using the 2D networks; ResNet in [15], the Real Valued-CNN (RV-CNN) and the Complex Valued-CNN (CV-CNN) in [16].

5 Edge Implementation

The proposed data processing approach was implemented into a Raspberry Pi 3b+ which is part of an EMI survey instrument that has

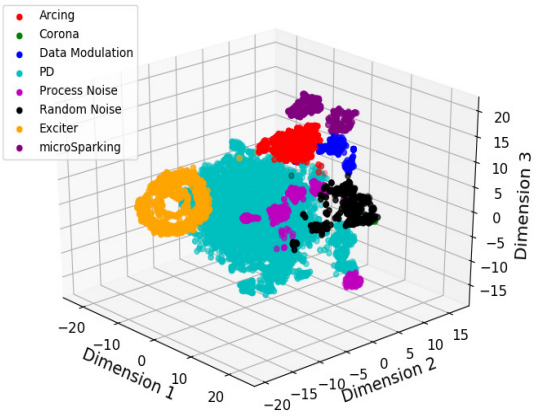


Fig. 8: t-SNE clustering of stage 2 feature space

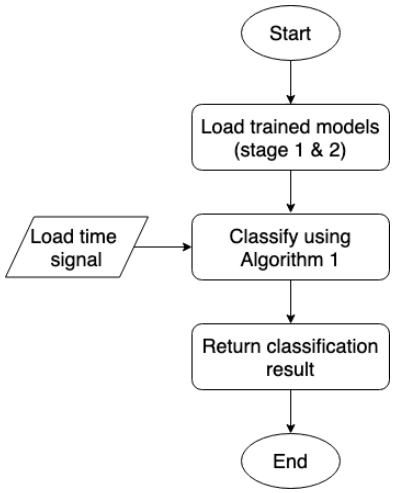


Fig. 9: Flow diagram of the inference on Raspberry Pi

Table 6 Computation time results comparison. Top performance in Bold font indicates top performance in each comparison

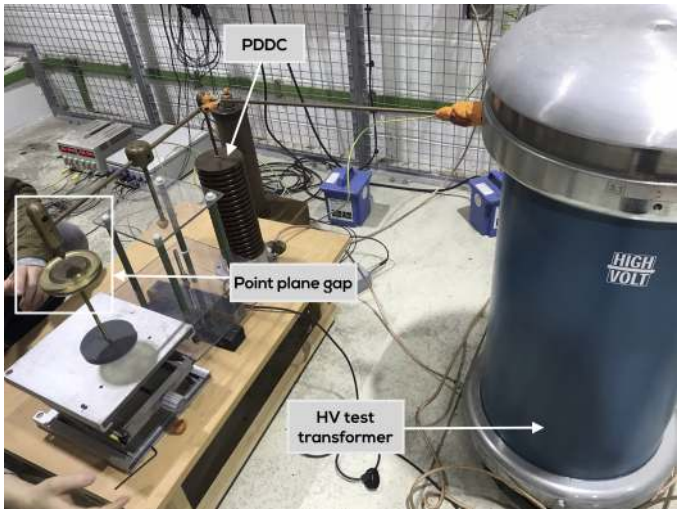
Model	Time
Stage 1 training	47min52s
Stage 2 training: new learning	36min16s
Stage 2 training: transfer learning	**28min35s**
Inference end-to-end	**0.25s**
Inference in [15]	1.59s
Inference in [16] (RV-CNN)	2.74s
Inference in [16] (CV-CNN)	26s

the capability to acquire data in the form of EMI frequency spectrum, phase resolved pattern and time-resolved signal. The adapted Raspberry Pi consists of the Broadcom BCM2835 system on a chip, operating on the ARM1176JZF-S Core (ARM V6K) 700 MHz CPU processor, with a RAM of 1GB. The chip requires a power input of 5V/2.5A DC and has a Micro SD card reader for system and data storage, in which the proposed algorithm was implemented. Further technical details on the device can be found in [34]. The Raspberry Pi is one part of the data acquisition instrument, which is responsible for displaying the graphical user interface and performing the inference of the proposed approach in Algorithm 1. The following packages were installed on the device in order to run the inference: python3.5, numpy, scipy, keras, tensorflow and pillow. A flow diagram of this process is presented in Fig. 9, which was completely executed in half a second. This implementation was tested both in laboratory experiment and at an operating power site, which will be described in detail as follows.

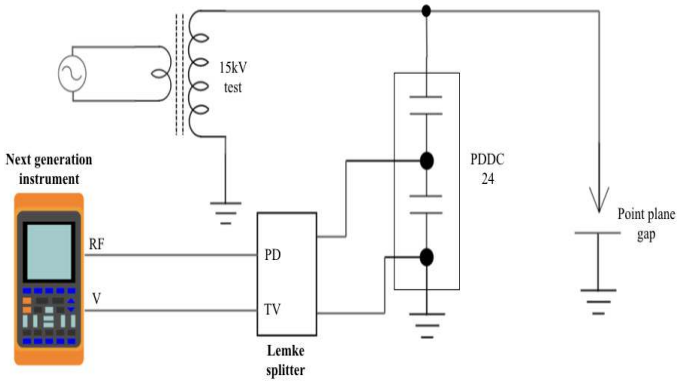
5.1 Laboratory testing

Artificial data was created in a HV laboratory environment (see Fig. 10a), where a point plane gap arrangement was created to generate PD [35] as illustrated in the electrical circuit schematic in Fig. 10b. A voltage of 15kV was generated as input to the circuit and the PD voltage transients were captured using the PD Decoupling (PDDC) unit feeding its output to a Lemke LDS-6 splitter [36]. Frequency and time domain data were recorded in the next generation instrument using the EMI method described in **Section 2**. The time-resolved signal was recorded by the developed EMI survey instrument and it was successfully classified as PD by the implemented end-to-end inference approach. An additional signal consisting of background noise only was recorded by the instrument and classified as Random Noise.

Further out-of-distribution signals were collected, from the office environment and the set-up measurement described in 2.2, and tested



(a) HV lab experiment set-up



(b) Point plane gap experiment schematic

Fig. 10: HV laboratory experiment

on the developed approach for validation. This involves a total of 53275 signals of which 100% were correctly classified as out-of-distribution and hence successfully filtered by the first stage model.

5.2 *Power site testing*

The developed instrument was tested in a major UK power station, the station name remains anonymous for confidentiality reasons. Specifically, tests were performed on two 667MW hydrogen/water cooled synchronous generators producing 24kV at 16kA. The HFCT was clamped around different connections of these generators including the blue, yellow and red phase couplers, and the neutral earth connection. The data measurement and classification was performed by the same instrument used in the HV laboratory. A total of 40 time-resolved signals were recorded between both generators (20 signals per generator). The signals were retrieved at frequencies 1, 5, 10, 15 and 40MHz from the EMI frequency spectrum. The developed end-to-end classifier identified 23 signals as unknown classes and the remaining 17 signals as PD. The measurements were analysed by two EMI experts confirming the presence of PD in signals classified as PD and a mixture of PD with other new phenomena classes (e.g. floating potential, slot discharge) in signals classified as unknown. These results are supported by the confusion matrix without normalisation in Fig. 11.

Fig. 12 shows photographs taken at the power station during the testing of the implemented end-to-end approach in the instrument, where two examples of the recorded signals were classified as PD and unknown.

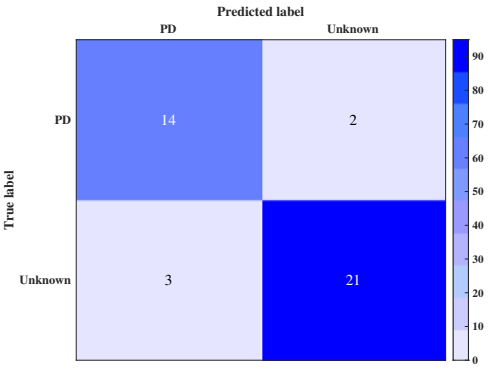


Fig. 11: On-site testing confusion matrix

(a) PD signal

(b) Unknown signal

(c) PD result

(d) Unknown result

Fig. 12: On-site testing results of the end-to-end solution

6 Conclusion

An end-to-end framework for classification of EMI time-resolved signals measured from real-world power stations and from a non-HV environment was developed. The analysis was conducted to achieve highly accurate automatic condition monitoring of HV assets at low computation cost. 1D-CNNs and transfer learning methods were followed to satisfy these performance requirements. Results demonstrate that the proposed approach achieved high classification results at lower computation cost compared to the previously proposed 2D-CNN based approaches. Furthermore, the proposed approach provides robustness against out-of-distribution data, a problem that is not addressed in previous methods. The developed approach was implemented in an edge device developed by ICOS Doble as a demonstration of ML-based condition monitoring to the power industry. The successful tests carried out on various data from the test dataset, HV lab experiments and real-world power station measurements widen the range of possibilities to utilising this approach as an informative diagnostic toolset to clients or a supportive toolset to experts. Future work will be carried out to tackle the multi-label classification problem to enhance the model with the capability to identify two or more phenomena recorded in one data measurement. The model will also be enhanced to build robustness against noise label as data labels rely on expert judgement which may differ within experts.

7 References

1 Dukic, G., Čukarić, A.: 'New algorithm for detecting power transformer faults based on m-robust estimation of sound signals', *IET Generation, Transmission Distribution*, 2014, 8, pp. 1117–1126

2 Ali, A., Khan, A.Q., Hussain, B., Raza, M.T., Arif, M.: 'Fault modelling and detection in power generation, transmission and distribution systems', *IET Generation, Transmission Distribution*, 2015, 9, pp. 2782–2791

3 Timperley, J.E., Vallejo, J.M.: 'Condition assessment of electrical apparatus with emi diagnostics', *IEEE Transactions on Industrial Applications*, 2017, 53, pp. 693–699

4 Abdeljaber, O., Avci, O., Kiranyaz, M.S., Boashash, B., Sodano, H., Inman, D.J.: '1d-cnns for structural damage detection: Verification on a structural health monitoring benchmark data', *Neurocomputing*, 2018, 275, pp. 1308–1317

5 Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., Inman, D.J.: 'Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks', *Sound and Vibration*, 2017, 388, pp. 154–170

6 Eren, L., Ince, T., Kiranyaz, S.: 'A generic intelligent bearing fault diagnosis system using compact adaptive 1d cnn classifier', *Signal Processing Systems*, 2019, 91, pp. 179–189

7 Ma, S., Cai, W., Liu, W., Shang, Z., Liu, G.: 'A lighted deep convolutional neural network based fault diagnosis of rotating machinery', *Sensors*, 2019, 19, pp. 2381

8 Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., et al.: 'Convolutional neural network based fault detection for rotating machinery', *Sound and Vibration*, 2016, 377, pp. 331–345

9 Ince, T., Kiranyaz, S., Eren, L., Askar, M., Gabbouj, M.: 'Real-time motor fault detection by 1-d convolutional neural networks', *IEEE Transactions on Industrial Electronics*, 2016, 63, pp. 7067–7075

10 Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Inman, D.J.: 'Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks', *Sound and Vibration*, 2018, 424, pp. 158–172

11 Kiranyaz, S., Gastli, A., Ben-Brahim, L., Al-Emadi, N., Gabbouj, M.: 'Real-time fault detection and identification for mmc using 1-d convolutional neural networks', *IEEE Transactions on Industrial Electronics*, 2019, 66, pp. 8760–8771

12 Kiranyaz, S., Ince, T., Gabbouj, M.: 'Real-time patient-specific ecg classification by 1-d convolutional neural networks', *IEEE Transactions on Biomedical Engineering*, 2016, 63, pp. 664–675

13 Kiranyaz, S., Ince, T., Hamila, R., Gabbouj, M.: 'Convolutional neural networks for patient-specific ecg classification'. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, August 2015, pp. 2608–2611

14 Woon, W.L., Aung, Z., El.Hag, A.: 'Intelligent monitoring of transformer insulation using convolutional neural networks', In: 6th Data Analytics for Renewable Energy Integration. Technologies, Systems and Society (DARE), Dublin, Ireland, September 2018, pp. 127–136

15 Mitiche, I., D. Jenkins, M., Boreham, P., Nesbitt, A., Stewart, B.G., Morison, G.: 'Deep residual neural network for emi event classification using bispectrum representations', In: 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, September 2018, pp. 186–190

16 Mitiche, I., D. Jenkins, M., Boreham, P., Nesbitt, A., Morison, G.: 'Deep complex neural network learning for high-voltage insulation fault classification from complex bispectrum representation', In: 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, September 2019, pp. 1–5

17 Timperley, J.: 'Emi diagnostics: Steam turbine generators', Doble, 2016, https://www.doble.com/emi-diagnostics-steam-turbine-generators/ , accessed: October 2018

18 CISPR16: 'Cispr/cis/a-radio-interference measurements and statistical methods', 2015

19 Timperley, J.E., Longo, P., McKim, D., Vallejo, J.M.: 'Electromagnetic interference data collection from bus coupler capacitors', In: 36th Electrical Insulation Conference (EIC), San Antonio, USA, June 2018, pp. 178–181

20 Nair, V., E. Hinton, G.: 'Rectified linear units improve restricted boltzmann machines', In: 27th International Conference on Machine Learning (ICML), Haifa, Israel, June 2010, pp. 807–814
21 Ioffe, S., Szegedy, C.: 'Batch normalization: accelerating deep network training by reducing internal covariate shift', In: 32nd International Conference on Machine Learning (ICML), Lille, France, July 2015, pp. 448–456
22 Kingma, D.P., Ba, J.: 'Adam: A method for stochastic optimization', In: 3rd International Conference for Learning Representations (ICLR), San Diego, USA, December 2015
23 Huang, S., Tang, J., Dai, J., Wang, Y.: 'Signal status recognition based on 1dcnn and its feature extraction mechanism analysis', *Sensors*, 2019, 19
24 Pan, S.J., Yang, Q.: 'A survey on transfer learning', *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22, pp. 1345–1359
25 Olivas, E.S., Guerrero, J.D.M., Sober, M.M., Benedito, J.R.M., Lopez, A.J.S.: In: 'Transfer learning', 'Handbook Of Research On Machine Learning Applications and Trends: Algorithms, Methods and Techniques - 2 Volumes'(Information Science Reference - Imprint of: IGI Publishing, 2009, pp. 242–264
26 Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: 'How transferable are features in deep neural networks?'. In: Advances in Neural Information Processing Systems 27, Montreal, Canada, December 2014, pp. 3320–3328
27 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al.: 'TensorFlow: Large-scale machine learning on heterogeneous systems', 2015, http://tensorflow.org/, online; accessed: February 2020
28 Chollet, F., et al.: 'Keras'. GitHub, 2015 https://github.com/fchollet/keras
29 Glorot, X., Bengio, Y.: 'Understanding the difficulty of training deep feedforward neural networks', In: In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS10), Society for Artificial Intelligence and Statistics, Sardinia, Italy, May 2010, pp. 249–256
30 Mitiche, I., Morison, G., Nesbitt, A., Hughes.Narborough, M., Stewart, B.G., Boreham, P.: 'Classification of partial discharge signals by combining adaptive local iterative filtering and entropy features', *sensors*, 2018, 18, p. 406
31 van der Maaten, L., Hinton, G.: 'Visualizing data using t-sne', *Machine Learning Research*, 2008, 9, pp. 2579–2605
32 Fawaz, H.I., Forestier, G., Weber, J., Idoumghar, L., Muller, P.: 'Transfer learning for time series classification', In: IEEE International Conference on Big Data, Seattle, USA, December 2018, pp. 1367–1376
33 Gabriela, C.: In: 'A comprehensive survey on domain adaptation for visual applications', 'Domain Adaptation in Computer Vision Applications' (Springer International Publishing, 2017), pp. 1–35
34 'Raspberry pi 3 model b+', https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/, online; accessed: February 2020
35 Danikas, M.G.: 'Detection and recording of partial discharges below the inception voltage with a point-plane electrode arrangement in air: Experimental data and definitions', *Journal of Electrical Engineering*, 2010, 61, pp. 177–182
36 Reid, A., Judd, M., Duncan, G.: 'Simultaneous measurement of partial discharge using tev, iec60270 and uhf techniques', In: IEEE International Symposium on Electrical Insulation, San Juan, USA, July 2012, pp. 439–442