A Robust Transform-Domain Deep Convolutional Network for Voltage Dip Classification

Azam Bagheri, Irene Y.H. Gu, *Senior Member, IEEE,* Math H.J. Bollen, *Fellow, IEEE*, Ebrahim Balouji

Abstract-- This paper proposes a novel method for voltage dip classification using deep convolutional neural networks. The main contributions of this paper include: (a) to propose a new effective deep convolutional neural network architecture for automatically learning voltage dip features, rather than extracting hand-crafted features; (b) to employ the deep learning in an effective two-dimensional transform domain, under space-phasor model (SPM), for efficient learning of dip features; (c) to characterize voltage dips by two-dimensional SPM-based deep learning, which leads to voltage dip features independent of the duration and sampling frequency of dip recordings; (d) to develop robust automatically-extracted features that are insensitive to training and test datasets measured from different countries/regions.

Experiments were conducted on datasets containing about 6000 measured voltage dips spread over seven classes measured from several different countries. Results have shown good performance of the proposed method: average classification rate is about 97% and false alarm rate is about 0.50%. The test results from the proposed method are compared with the results from two existing dip classification methods. The proposed method is shown to out-perform these existing methods.

Index Terms-- Power quality, Voltage dip, Machine learning, Deep learning, Convolutional Neural Network.

I. INTRODUCTION

VOLTAGE dips are among the most disruptive power quality events. They regularly cause equipment mal-operation and stoppage of production processes resulting in large economic costs [1]-[3]. Over recent decades, a significant amount of research has been done on voltage dips [4], including propagation of dips, stochastic prediction methods, dip characterization, and mitigation methods. Several methods have also been proposed for voltage dip classification. The methods mainly aim at: quantifying the voltage dip impact on equipment (end-user equipment, but also interface to renewable energy) and getting information on the origin of the voltage dip (e.g. type of electrical fault). Voltage-dip classification is an important challenge in extracting useful information from voltage dips [5]-[8].

An overview of earlier work on dip classification is presented in Section II.A. Earlier work on the use of machine learning for classification of power-quality disturbances [9]-[12] has been based on hand-crafted features. Good performance was obtained, but relatively small amount of data. A classifier trained on a small dataset often shows significant performance degradation when applied to a large dataset. This could be caused by incomplete description of data using hand-crafted features, or not covering sufficient data statistics from the small dataset. Further, human-expert knowledge as well as suitable mathematical tools such as Fourier transforms [9], wavelet transforms [10] or Fuzzy systems [11][12] are required to efficiently extract such features.

Recent work on machine learning, resulting in the new paradigm "deep learning", allows for a completely new approach to handle large amounts of data [13]. Convolutional neural networks (CNNs) are one type of deep learning methods developed specifically to learn the hierarchical representation of input data, as well as to predict the class label of the data. In a CNN, the input data is first transformed through a series of chained convolutional filters with nonlinear activation functions which is equivalent to applying a series of chained multichannel nonlinear filters. By stacking multiple convolution layers with nonlinear activation functions in between, it allows the network to learn sophisticated features of the data. The network is then followed by a series of fully connected (FC) layers with nonlinear activation functions in between. This results in an output vector of class probabilities after the softmax activation function in the last FC layer. The series of chained convolutional filter is used for automatically learning the sophisticated features (in contrast to using hand-crafted features selected by human experts (e.g., rectangular shape, number of segments in a disturbance recording). The series of FC layers is used as a classifier, and can be replaced by conventional classifiers such as support vector machines (SVMs) or AdaBoost (Adaptive boosting). Both parts of the network are usually trained together for obtaining the best network architecture, filter coefficients/weights by adjusting the number of layers, the number of filters and their size and hyper-parameters, such that satisfactory end-to-end performance is obtained under a selected objective criterion or loss function. For a detailed description of the essential steps

involved in a CNN (e.g., convolutional filters, nonlinear activation function, maximum pooling), readers are referred to the excellent set of lecture notes on deep learning and CNNs in [14].

Deep learning enables a machine to learn features by automatically extracting them from the data. In many fields, machine classification performance using features from deep learning has exceeded human performance [13].

Despite a large amount of publications on power quality disturbance classification, only a small number of works have been published on deep-learning methods to the application of power systems: [15]-[20], where only [15] and [17] concern classification of power-quality disturbances. None of the publications takes up voltage-dip classification.

This paper proposes a novel classification method based on deep learning using convolutional neural network (CNN) to automatically extract features and classify voltage dips.

The main contributions of the paper are: (i) developing a specific structure of CNN and extracting robust voltage dip features automatically; (ii) using a 2D representation of space phasor model (SPM) to be independent on duration and sampling frequency of the dip recording; (iii) showing that CNNs are a suitable tool for classifying voltage dips as part of large datasets.

The proposed method was applied to three measured dip datasets, including 5982 dips. The experimental results show robust classification performance in terms of average classification rate (96.78%) and false alarm rate (0.47%).

In the remainder of this paper, Section II reviews related works on voltage dip classification and presents relevant background theory. Section III details the proposed method: the 2D representation of SPM; the proposed 2D-CNN and dip classification. The experimental results and a discussion about obtained results are presented in Section IV. Section V concludes this paper.

II. BACKGROUND

A. Voltage Dip Classification

The research on voltage dip characterization and classification was started in the 1980s [21] and the early 90s [3][22]. The first voltage dip classification for use in three-phase systems, the so-called ABC classification, was proposed in [23]. That classification is based on a number of synthetic dips that fall into four types: A, B, C and D. Type A covers balanced dips, the three remaining types cover voltage dips due to single-phase or phase-to-phase faults [23][24].

This classification is a systematization of the knowledge about dip propagation through different types of three-phase transformers, presented in [3]. Considering two-phase-to-ground faults, the ABC classification is extended to seven different classes, Class A through Class G in [25]. Dips due to motor starting and transformer energizing are included in the classification by [26]. A further extension has been proposed in [27], including two types of voltage swells, referred to as Class H and Class I.

Methods for automatic classification of recorded voltage dips were first introduced in [28]: the complex positive and negative-sequence voltages were used for determining the dip type: a distinction was originally made between six types of "three-phase unbalanced voltage dip" and one type of "balanced voltage dip" as follows:

- Types C_a, C_b, C_c: known as the C Type family. These three types include unbalanced voltage dips with significant drop in two of the three voltages and a lower or no drop in the third one. The subscript of C indicates the phase with small or no voltage drop.
- Types D_a, D_b, D_c: known as the D Type family. These three types include unbalanced dips with a large drop in one voltage and a lower or no drop for the two other voltages. The subscript of D indicates the phase with a large voltage drop.
- Type A: This type consists of balanced voltage dips.

There exist many variant methods, for classification of measured voltage dips; we use this 7-type ACD classification.

The expert system in [26] aims at voltage dip classification according to underlying cause of voltage dips. Reference [29] uses a support vector machine (SVM) method for voltage dip classification. The work in [30] used wavelet transform to classify power quality events. Although these methods have obtained good performance they require the setting of several threshold levels [26], human expert knowledge [29] or extracting the fundamental frequency component [30].

A CIGRE working group proposed a voltage dip classification based on the number of dropped phase-to-neutral voltages [31]. Its main recommendation was that one should use the notation I-II-III (instead of D-C-A, respectively), to avoid confusion.

The method proposed in [32] distinguishes between balanced dips and two types of unbalanced dips, roughly corresponding to the C and D Type families in the 7-type ACD classification. However, only rms voltages are used, so that information about individual phasors is lost. The methods proposed in [27], [33]-[35] use ellipse parameters for voltage dip classification. The ellipse is corresponding to either space phasor model or the polarization ellipse, in both cases calculated from the three phase-to-neutral voltages. The methods proposed in [27] and [33] are mainly developed based on synthetic dips; the method proposed in [33] needs to extract the fundamental frequency component and requires setting of several thresholds.

B. Space Phasor Model (SPM)

The space phasor model (SPM) of three phase-to-neutral voltages, $V_a(t), V_b(t), V_c(t)$, is given by (1), [27][34].

$$s_t = \frac{2}{3}[V_a(t) + \alpha V_b(t) + \alpha^2 V_c(t)], \quad t = 1, \cdots, n \quad (1)$$

where $\alpha = e^{j2\pi/3}$ and s_t has a complex value.

In the ideal case of sinusoidal balanced voltages, the SPM in the complex plane is in the form of a circle centred around the origin. Voltage unbalance, including unbalanced voltage dips, results in an ellipse. Voltage distortion results in a "distorted ellipse".

An ellipse in the complex plane is defined by just three real parameters: semi-minor axis, semi-major axis and the rotating angle of the ellipse. These three parameters are highly

correlated with some voltage dip characteristics [34][36]. Among these, the rotating angle of the ellipse distinguishes between different types of voltage dips in the 7-type ACD classification [34][36], see Section III-C.

Fig. 1 shows the waveform of a typical dip in time domain and the corresponding SPM in the complex plane.
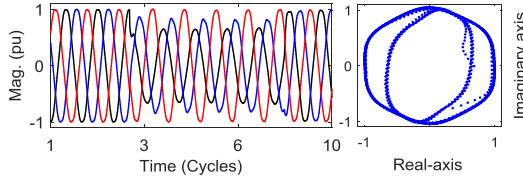

Fig. 1. Waveform (left) and SPM (right) of one typical voltage dip.

C. Deep convolutional neural networks (CNNs)

The architecture of convolutional neural networks typically consists of several layers, including convolutional layers with ReLU (Rectified Linear Unit) as the activation function, pooling layers, and finally the fully connected (FC) layers with softmax as the activation function. A convolutional layer with ReLU is often followed by a pooling layer (e.g., maxpool) to obtain nonlinear down-sampling. This process (i.e., a convolutional layer with a ReLU and a pooling layer) is usually repeated several times to obtain deep network architecture [14]-[37]. For example, for two-dimensional (2D) input signals, a set of 2D spatial filters (sized k by k) is applied. This is similar to the idea of applying a set of 2D FIR filters. However, these filter kernels are obtained through deep learning. The outputs from the filters are corresponding to the features of the signal in the corresponding layer. The ReLU is a nonlinear function $f(x) = max(0, x)$, which is operated element-wise by simply replacing all negative feature values with zero. Using ReLU is found to have greatly accelerated the convergence of the stochastic gradient descent [37]. Max-pooling layer is often adopted in order to introduce the nonlinearity as well as feature down-sampling. For example, a maxpool on a 2 by 2 matrix of features results in one feature value, which is the maximum among these four feature values. The output from each convolutional and pooling layer represents features of the input data in that corresponding scale level.

The fully connected (FC) layers are used for forming the classifier, whose input is the feature map from the last convolutional layer and whose output is the class label. The term "fully connected" implies that every neuron in the previous layer is connected to every neuron in the next layer [37], similar to that in classical artificial neural networks. The last FC layer is usually followed by a softmax function (or, normalized exponential function). It is a generalization of the logistic function. The output of softmax is associated with the probability distribution over L classes of different possible outcomes from the classifier. This part (FC layers with activation functions) can be in principle replaced by other conventional classifiers (e.g., SVMs [38][39], AdaBoost [40][41]). The typical architecture of CNNs is shown in Fig. 2.

III. PROPOSED METHOD

In this section, we describe the proposed scheme for

automatic feature extraction and classification of voltage dips. The main motivation of this work is to exploit deep learning for automatic extraction of dip features instead of using conventional methods that extract hand-crafted features using 'human experts' knowledge. It has been shown before that SPM domain dip classification is more robust than directly using the waveform or RMS voltage versus time [27][33]-[35]. We therefore propose to employ deep learning in the SPM domain. The proposed method consists of the following steps: (i) Transform three-phase voltages into SPM domain; (ii) Employ 2D CNN (two Dimensional Convolutional Neural Network) for *automatic feature extraction* for different dip classes using supervised training; (iii) Use FC (Full Connected) layers in the CNN as the classifier where the training and classification processes are performed. It is worth mentioning that the classifier (with the FC layers) can be replaced by other types of classifiers, e.g. SVM [38][39], AdaBoost [40][41], and many others. Fig. 3 shows the block diagram of the proposed scheme. In the following subsections, the proposed scheme will be described in details.

A. SPM Transform of Three-phase Voltage Sequences

This subsection describes the approach of forming a 2D representation in the SPM domain from a three-phase voltages. We use SPM in (1) with a per-unit system. It is worth mentioning that SPM in (1) consists of real and imaginary parts with the dynamic range of [-1.1, 1.1] pu (Taking 10% margin for harmonic distortion and voltage magnitude variations). For mathematical convenience, the SPM time series in (1) is written in the vector form,

$$s_t = [s_1, s_2 \ldots s_n]^T \qquad (2)$$

A new matrix S is defined by splitting s_t into its real and imaginary parts:

$$S = [s_{t,R} \quad s_{t,I}]_{n \times 2} \qquad (3)$$

where $s_{t,R} = [s_{1,R} \, s_{2,R} \cdots s_{n,R}]^T \in [-1.1, 1.1]$, and $s_{t,I} = [s_{1,I} \, s_{2,I} \cdots s_{n,I}]^T \in [-1.1, 1.1]$. We then quantize the dynamic range [-1.1, 1.1] of S into 22 equal length segments, where the data sequence in the matrix S is described by a new matrix M (size 22×22), using the following relation:

$$M(k_1, k_2) = \sum_{i=1}^{n} \sum_{(k_1, k_2)} \sqrt{(s_{i,R})^2 + (s_{i,I})^2} \qquad (4)$$

where k_1 and k_2 are determined by:

$$(k_1, k_2) = \begin{cases} (\lceil 10(1.1 - s_{i,I}) \rceil, \ \lceil 10(1.1 - |s_{i,R}|) \rceil) & if \ s_{i,R} < 0 \ \& \ s_{i,I} > 0 \\ (\lceil 10(1.1 - s_{i,I}) \rceil, \ \lceil 10|s_{i,R}| \rceil) & if \ s_{i,R} > 0 \ \& \ s_{i,I} > 0 \\ (\lceil 10|s_{i,I}| \rceil + 11, \ \lceil 10(1.1 - |s_{i,R}|) \rceil) & if \ s_{i,R} < 0 \ \& \ s_{i,I} < 0 \\ (\lceil 10|s_{i,I}| \rceil + 11, \lceil 10 s_{i,R} \rceil + 11) & if \ s_{i,R} > 0 \ \& \ s_{i,I} < 0 \end{cases} \qquad (5)$$

where $\lceil \cdot \rceil$ is the ceiling function, and $|\cdot|$ is the absolute value.

In such a way, all input voltage waveform sequences are converted into a 2D representation of SPM by M matrices with 22×22 components, which are used for the inputs of 2D CNN for automatic feature extraction in SPM domain and subsequently for dip classification.
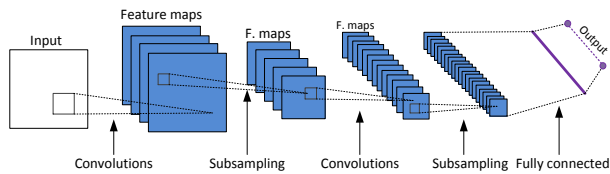
A typical 2D representation of the SPM, as matrix $\boldsymbol{M}$, is shown in Fig. 4.

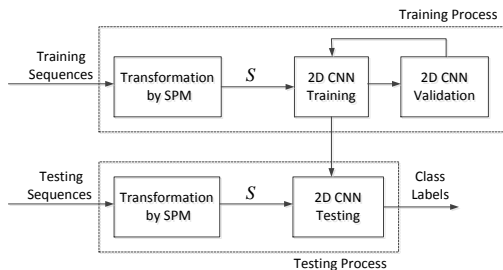Fig. 2. Typical architecture of a Convolutional Neural Networks [42].

Fig. 3. Block diagram of the proposed scheme ($\boldsymbol{S}$ is a 22 by 22 matrix).

B. 2D-CNN for Automatic Feature Extraction in SPM Domain and for Dip Classification.

In this subsection, we describe the proposed 2D-CNN architecture for automatic feature extraction and dip classification. Since the input data in the SPM domain is two dimensional without temporal information, we select 2D-CNN as the deep learning method. Numerous empirical tests have been performed to find out the most appropriate CNN architecture associated with the best end-to-end performance for the dip feature learning and dip classification task. This is usually done by end-to-end training and performance optimization which minimizes some loss function, e.g., the categorical cross-entropy error (see (7)). Currently, there is no general rule or theoretical guide to deep learning methods on how to reach the best architectures and their associated hyper-parameters. This is instead usually done through extensive empirical tests with some test strategies chosen by individual researchers. For example, empirical tests could be first conducted on the number of layers by observing the performance of training, and by observing the overfitting from the performance difference between the training and validation processes. More detailed tuning could then be conducted by changing the number of filters, the size of filter kernels, and other hyper-parameters. This selection process can be repeated until good performance is obtained. We have adopted the above strategy in our experiments to seek an appropriate CNN architecture suited with the associated hyper-parameters for the SPM domain.

The proposed CNN architecture consists of four convolutional layers (Conv1-Conv4) and three fully connected layers (FC1-FC3). The convolutional layers are designed for automatic extracting dip features, while the fully connected layers serve as a classifier. Table I shows the detailed architecture of the proposed 2D-CNN architecture. As can be seen in Table I, the first convolutional layer *Conv1* contains 16 filter bands, each being a 5×5 spatial domain filter that performs, $x_1 = f_1(s * w + b)$, the spatial convolution

followed by ReLU nonlinear activation function $f(x) = max(x, 0)$. In the second and fourth convolutional layers, a 2×2 maximum pooling is applied, with its output as the largest value from the 2×2 area. The outputs from the four convolution layers (after ReLU and maxpool) contain the multi-resolution features of the input voltage dip. After four convolutional layers, three FC layers are employed as the classifier. We used FC layers for simplicity in fulfilling the classification task. Each FC layer consists of several cells (neurons), and the last FC layer consists of seven cells equal to the number of dip classes in our designed scheme. Table I shows the complete architecture of the convolutional network used for the feature extraction (rows 2-5) and for the classification (rows 6-8). Using the parameters given in Table I, one can build up the complete convolutional network using the TensorFlow library.
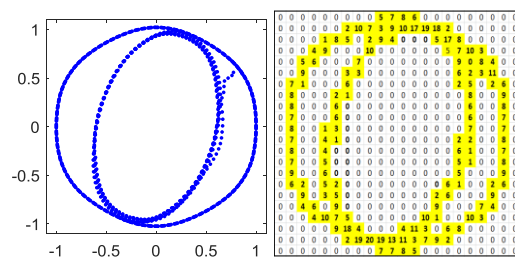

Fig. 4. A typical example of 2D representation of SPM; left: SPM in complex plane, right: corresponding matrix $\boldsymbol{M}$.

C. Voltage Dip Classes and Labelling

In our scheme, we follow the 7-type ACD classification [28]: C_a, C_b, C_c, D_a, D_b, D_c and A, as explained in Section II.A. The before-mentioned strong correlation between the rotating angle of the ellipse and the dip class is shown in Table II [34]. The parameter T, is obtained by dividing the rotating angle by 30°, is directly related to the dip class. The voltage dip classes, determined from this approach, are used as labels for supervised learning process of 2D-CNN. Class A is determined once the difference between semi-minor and semi-major axes is less than a threshold (0.05 pu).

TABLE I
Proposed Deep Learning Method: 2D CNN Architecture for Automatic Feature Extraction and Classification of 7 Class Dips

Layers	Filter Size / # cells	Input size	Output Size
2D Conv1+ReLU	$(5,5) \times 16$	22×22	22×22×16
2D Conv2+ReLU+Maxpool	$(3,3) \times 32$	22×22×16	11×11×32
2D Conv3+ReLU	$(3,3) \times 64$	11×11×32	11×11×64
2D Conv4+ReLU+Maxpool	$(3,3) \times 128$	11×11×64	6×6×128
FC1+ReLU	1024	4608×1	1024 × 1
FC2+ReLU	128	1024 × 1	128 × 1
FC3+Softmax	1x7	128×1	7×1

TABLE II
Relation Between Angle φ , Parameter T, and Dip Class

Ranges of angles (φ)	Parameter T	Dip Class
(0°, 30°]	1	D_b
(30°, 60°]	2	C_c
(60°, 90°]	3	D_a
(90°,120°]	4	C_b
(120°,150°]	5	D_c
(150°,180°]	6	C_a

IV. Experimental Results

The proposed deep learning-based scheme has been applied to three measurement datasets under a number of experiments. In the following, we describe the criteria for evaluation, experimental setup, test results and performance evaluation.

A. Criteria for Evaluating the Performance

To verify the performance of the proposed method, we used two sets of criteria: performance of deep learning in CNN; performance of classifier on each individual class.

For deep learning in CNN, we adopt the conventional used definitions of accuracy and loss by the categorical cross-entropy, as follows [43]:

$$Accuracy = \frac{\#\ correctly\ predicted\ dips\ in\ all\ classes}{total\ number\ of\ dips} \quad (6)$$

$$Loss = -\sum_x p(x) \ln q(x) \quad (7)$$

where $p(x)$ is the correct label for the input x, and $q(x)$ is the predicted label. It is worth noting that accuracy and loss are the average values over all classes.

For assessing the performance of individual dip classes, we use a commonly adopted way in the machine learning area: the "confusion matrix". In this paper, the confusion matrices are shown only for the test datasets. Each row of the confusion matrix represents the instances in an actual class while each column represents the instances in a predicted class. The diagonal entries show the number of correctly predicted classes where the non-diagonal entries show the incorrect classifications.

We use the classification rate (CR) and false alarm rate (FAR) for each class as a measure of the performance. For class i, the CR and FAR are defined as follows [40]:

$$CR_i = \frac{TP_i}{TP_i + FN_i} \quad (8)$$

$$FAR_i = \frac{FP_i}{FP_i + TN_i} \quad (9)$$

where $i = 1,2,...,7$ indicates the class number, TP_i and TN_i are the true positive and true negative values for i^{th} class, FP_i and FN_i are the false positive and false negative values for i^{th} class, respectively.

B. Setup

1) Datasets

Three voltage dip datasets are used in this study, details are given in Tables III and IV. The Sweden dataset is measured in one location in a low voltage network. The global dataset includes voltage dips measured in different locations and voltage levels around the world by PQube power quality monitors [44]. The UK dataset consists of voltage dips obtained at many medium voltage locations during approximately one month of recording. The dips are roughly uniformly distributed over the different classes in UK and Sweden datasets. In global dataset the number of balanced voltage dips dominates, but the unbalanced dips are again roughly uniformly distributed over the different classes.

2) CNN hyper-parameters

The hyper-parameters of CNN, learning rate (lr) and

dropout (dr), were selected optimally in the learning process through 2D grid search. The number of epochs and batch size were set as 250. The optimizer 'adam' was used, also 'softmax' activation was used in the FC3 layer, while in all other layers 'ReLU' was used.

In all three case studies, we partitioned the dataset(s) into training/validation/test subsets. The data in the training and validation sets was split into 5 parts, 4/5 was used for training, and the remaining 1/5 for validation. For computing the validation error, we used 5-fold cross-validation: the data were partitioned into 5 folds, where 1 fold was used as the validation set, and the remaining 4 folds were used as the training set. This process was repeated five times such that each fold was used once for validation. The validation error was then obtained as the average error over the five different validation sets. The testing process was conducted solely on the test dataset where the data were kept separated from the training and validation data.

TABLE III
Detailed Information of the Three Voltage Dip Datasets

Name	# Dips	Location of measurements
D1	705	Sweden
D2	4361	Global
D3	916	UK

TABLE IV
The Number of Dips for Every Class of Each Dataset

Dataset	A	C_a	C_b	C_c	D_a	D_b	D_c	Total
D1	166	110	100	74	83	77	95	705
D2	1946	448	500	429	265	412	361	4361
D3	120	135	142	128	116	129	146	916
Total	2232	693	742	631	464	618	602	5982

3) Case studies

To test the proposed scheme and its robustness, we have designed three case studies. In the first case study, we mixed all three datasets (with a total of 5982 dip measurements), and then split the whole data into 75% for training and 25% for testing. In the next two case studies, we used training data and testing data from different datasets (from different locations) in order to further study the robustness of the proposed scheme. Table V summarizes the three case studies.

TABLE V
Three Case Studies and the Corresponding Training/Test Datasets.

Case Study	Training set	Test set
C1	0.75 (D1+D2+D3)	0.25 (D1+D2+D3)
C2	D2+D3	D1
C3	D1+D2	D3

C. Results and Performance Evaluation

In this section, we will describe the results of each case study in detail.

1) Case study 1

This case study was applied to the summed datasets D1+D2+D3, with a 75% and 25% partition into training and test sets. Fig. 5 shows the performance of the proposed scheme where the accuracy and loss are shown as a function of the number of epochs. Further, Table VI shows the resulting confusion matrix on the test set, and Table VII shows the performance (classification rate and false alarm rate, as in (8)

and (9)) on the test set for each individual class.

The table shows 4 misclassifications between D_a and C_b, 9 between D_c and C_a and 7 between D_c and C_b.

The other misclassifications are between unbalanced dip classes and balanced dip Class A, for instance, 2 between Class D_c and Class A.

To find the optimal learning rate and dropout, a 2D grid search was performed. Table VIII shows the results of the grid search, where the table contains the training accuracy values.
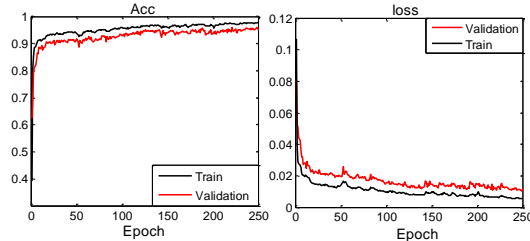

Fig. 5. Case study 1: performance on training and validation, using 5-fold cross-validation. Left: accuracy vs. epochs, Right: loss vs. epochs.

TABLE VI
Case Study 1: Resulting Confusion Matrix on the Test Set.

	$\widetilde{A}$	$\widetilde{C}_a$	$\widetilde{C}_b$	$\widetilde{C}_c$	$\widetilde{D}_a$	$\widetilde{D}_b$	$\widetilde{D}_c$
A	**556**	0	1	0	0	0	0
C_a	0	**170**	0	0	0	0	2
C_b	1	0	**184**	0	0	0	0
C_c	0	0	0	**154**	1	0	0
D_a	0	0	4	0	**112**	0	0
D_b	1	3	0	0	0	**151**	0
D_c	2	9	7	0	0	3	**129**

TABLE VII
Case Study 1: Performance on Each Class: CR and FAR.

Class	A	C_a	C_b	C_c	D_a	D_b	D_c	Ave.
CR(%)	99.8	98.9	99.5	99.4	96.6	97.4	86.0	**96.78**
FAR(%)	0.8	0.9	0.9	0.0	0.2	0.4	0.1	**0.47**

TABLE VIII
Grid Search for Optimal Hyper-Parameter Selection; Case Study 1

$\frac{dr}{lr}$	0.30	0.40	0.50	0.60	0.70
0.0003	95.37	96.04	96.89	96.57	96.79
0.0004	96.38	95.50	96.84	96.51	97.04
0.0005	96.79	95.77	97.72	97.05	96.91
0.0006	97.24	96.91	97.58	97.44	97.45
0.0007	97.25	95.58	96.27	96.98	96.85

2) Case Study 2

This case study was designed for testing the robustness of the proposed method using several countries' data (D2+D3: global+ UK) for training, and then applying the trained system for classification of dips obtained from only one country (D1: Sweden). Fig. 6 shows the training and validation performance as a function of number of epochs.

The resulting confusion matrix on the test set is shown in Table IX, and the performance (classification rate and false alarm rate) on the test set for each individual class in Table X.

The table shows that misclassifications occur between consecutive classes in Table II and between unbalanced voltage dip classes (D and C) and balanced dip class (A).

The results in Tables IX, X and Fig. 6, show a small drop in performance (1.71% decrease of average classification rate, and 0.38% increase of false alarm rate, compared to results from Case study 1). This small drop of performance is anticipated as the training data is from different counties whereas the test data is from one country and in this case actually from one location. From the results, it shows that our proposed scheme is very robust, achieving reasonably good test results (95.1% classification rate, 0.85% false alarm rate).
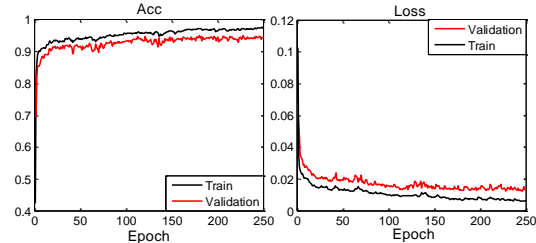

Fig. 6. Case Study 2: performance on training and validation, using 5-fold cross-validation. Left: accuracy vs. epochs, Right: loss vs. epochs.

TABLE IX
Case Study 2: Resulting Confusion Matrix on the Test Set.

	$\widetilde{A}$	$\widetilde{C}_a$	$\widetilde{C}_b$	$\widetilde{C}_c$	$\widetilde{D}_a$	$\widetilde{D}_b$	$\widetilde{D}_c$
A	**161**	0	1	1	1	1	1
C_a	2	**107**	0	0	0	1	0
C_b	1	0	**97**	0	1	0	1
C_c	1	0	0	**72**	1	0	0
D_a	0	0	1	2	**80**	0	0
D_b	1	0	0	1	0	**75**	0
D_c	1	9	4	1	0	1	**79**

TABLE X
Case Study 2: Performance on Each Class: CR and FAR.

Class	A	C_a	C_b	C_c	D_a	D_b	D_c	Ave.
CR (%)	97.0	97.3	97.0	97.3	96.4	**97.4**	83.2	**95.07**
FAR(%)	1.2	1.6	1.1	0.8	0.5	0.5	**0.4**	**0.85**

3) Case Study 3

In this case study the proposed deep network was trained and validated on datasets D1+D2 (Sweden + global) and was tested on the third dataset (D3: UK).

The performance for CNN training and validation is shown in Fig. 7, the confusion matrix is shown in Table XI, and the performance for individual classes is shown in Table XII.

From observing the results in the confusion matrix, the number of misclassifications between unbalanced and balanced voltage dip classes and between consecutive classes (as in Table II) have increased.

In this case study, the accuracy is more impacted when one compares the results with those in Case Study 2. No obvious reason for this is found.

Observing the results in Tables XI, XII and Fig. 7, one can see that there is again a small drop in performance: 3.12% decrease in average classification rate and 0.67% increase in false alarm rate. These results also support the notion that our proposed scheme is robust, achieving 93.66% classification rate, and 1.14% false alarm rate. Table XIII summarizes the performance for the three case studies.
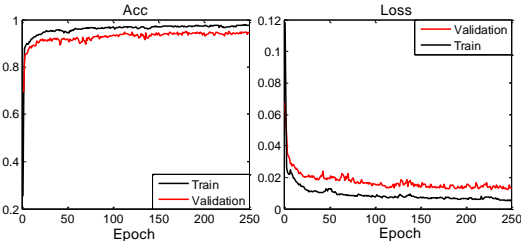
Fig. 7. Case Study 3: performance on training and validation, using 5-fold cross-validation. Left: accuracy vs. epochs, Right: loss vs. epochs.

TABLE XI
Case Study 3: Resulting Confusion Matrix on the Test Set

	$\tilde{A}$	$\tilde{C}_a$	$\tilde{C}_b$	$\tilde{C}_c$	$\tilde{D}_a$	$\tilde{D}_b$	$\tilde{D}_c$
A	120	0	0	0	0	0	0
C_a	2	125	0	0	0	7	1
C_b	0	0	137	0	2	0	3
C_c	2	0	0	118	5	3	0
D_a	4	0	8	4	100	0	0
D_b	1	0	0	1	0	127	0
D_c	2	9	7	0	0	4	124

TABLE XII
Case Study 3: Performance on Each Class: CR and FAR.

Class	A	C_a	C_b	C_c	D_a	D_b	D_c	Ave.
CR (%)	100	94.1	98.6	94.6	88.8	95.4	84.3	93.66
FAR(%)	1.7	1.6	1.5	1.4	0.7	0.7	0.4	1.14

TABLE XIII
Summary of Performance on the Test Set for all Cases Studies

Case Study	Loss	ACC (%)	ACR (%)	FAR (%)
C1	0.0057	97.72	96.78	0.47
C2	0.0104	95.18	95.07	0.85
C3	0.0152	93.59	93.66	1.14

D. Comparison with Other Methods

Ideally, comparisons should be made for systems that are designed for classifying the same types and number of classes, using the same datasets. However, such comparisons are very difficult as early work on voltage dip classification rarely uses large datasets. It is also difficult to find systems with the same type and number of classes as the proposed system. Hence, comparisons selected in this part may only be served as a relative performance indication, rather than absolute comparison. In the following, the proposed method is compared with the voltage dip classification methods proposed in [26] and [29]. Two different approaches were conducted in those studies: (a) Mixing all data from different countries and with a 75% and 25% partition on the training and test sets (such as Case Study 1); (b) Training on data from different countries but test on data from only one country (such as Case studies 2 and 3). The results for the first approach are shown in upper-side, and for second approach in lower-side of Table XIV. The table shows that the proposed method achieves the highest classification accuracy, which is further evidence for its effectiveness.

It should be noted that the accuracy and false alarm rate of the proposed method, for the second approach, are average of obtained results on Case studies 2 and 3. The same method was used for SVM method.

E. Discussion

1) Comparison Between Different Case Studies

As shown in Table XIII, the best performance was obtained for Case study 1. Mixing all datasets, in both training and testing processes, increases the uniformity of the data and decreases its diversity and thus results in a better performance.

The results from Case studies 2 and 3 show that the proposed method is sufficiently robust: training deep network on voltage dips from different countries and then testing it on dips from only one country does not significantly deteriorate the total performance. It means no new training is needed for new locations.

2) Misclassification

There are two types of misclassification: the first type of misclassification is "between different unbalanced classes"; the second type of misclassification is "between balanced and unbalanced classes".

The first type of misclassification are the ones with a significant characteristic phase-angle jump, that are close to the border between two classes when considering the rotational angle of the ellipse (Table II). These can easily be misclassified as was shown in [45] but for a completely different classification algorithm.

The second type of misclassification contains two sub-types: (i) unbalanced dips are classified as balanced dip (Class A); shallow unbalanced dips result in an ellipse in which the difference between semi-minor and semi-major axes is very small, therefore the 2D representation of the SPM is similar to a circle; (ii) balanced voltage dip is classified as one of the unbalanced voltage dip classes; heavy harmonic distortion in balanced voltage dips results in the polygon instead of circle shape of SPM, that leads to misclassification.

In overall, observing the results in Table XIII, the proposed scheme is shown to be effective for dip classification.

3) Voltage Dip Representation

The 2D representation of voltage dips allows us to cope with challenges as: duration of voltage dip recording, sampling frequency, and the variety among a large amount of dips, since all different voltage dips are represented by the same 22×22 matrix. The results from case studies 2 and 3 show that the 2D representation of SPM provides robust features for voltage dip classification.

TABLE XIV
Performance of Different Methods for Voltage Dip Classification; Accuracy and False Alarm Rate on Test Set

Train and Test on All Countries				
Method	# of dips in dataset	# of class	ACC (%)	FAR(%)
Proposed method	5973	7	97.72	0.47
Expert system [26]	916	7	97.00	0.30
SVM method[29]	1720	5	92.80	0.98
Train on Several Countries, and Test on Different Countries				
Proposed method	5973	7	96.37	0.995
SVM method[29]	1720	5	94.05	1.41

V. CONCLUSION

This paper proposes deep convolutional neural networks (CNNs) for automatic feature extraction and voltage dip

classification. A robust transform-domain method consisting of a space phasor model (SPM) of three-phase voltage dips has been used. The resulting 2D representation of the SPM presents any voltage dip recording independent of its duration or sampling frequency, by the same matrix. The fully-automatically extracted features, by the CNN, are fed to the fully connected neural layers, to classify voltage dips into seven different classes based on the 7-type ACD classification. The proposed method was applied to three-phase voltage dip datasets (about 6000 dips measured in different countries and at different voltage levels) distributed over seven classes.

The proposed method achieved high classification accuracy and small false alarms overall, as well as for each individual class. Comparison with two existing methods provided further evidence for the effectiveness of the proposed method.

REFERENCES

[1] 6th CEER benchmarking report on the quality of electricity and gas supply, Chapter 3-4 , Nov. 2016.

[2] L. E. Weldemariam, V. Cuk, J. F. G. Cobben and J. B. M. Waes, "Regulation and classification of voltage dips", presented at the 22th Int. Conf. on Electricity Distribution (CIRED), Stockholm, Sweden, June 10-13, 2013.

[3] M. F. McGranaghan, D. R. Mueller, and M .J. Samotyj, "Voltage sags in industrial systems," IEEE Trans. Industry Applications, vol. 29, no. 2, pp. 391-399, Mar./Apr. 1993.

[4] A. Bagheri and M. H. J. Bollen, "Developments in voltage dip research and its applications, 2005–2015," in Proc. ICHQP, 2016, pp. 48-54.

[5] A. A. J.-Matta, L. G. Pesquer and C. L. T.-Rodríguez, "Classification of voltage sags according to the severity of the effects on the induction motor," DYNA, vol. 82, no. 190, pp. 96-104, 2015.

[6] A. Băloi, F. Molnar-Matei, A. Pană, F. Băloi and F. Dilertea, "LabVIEW implementation for three-phase voltage dip classification," presented at the 16th Int. Conf. Electric Power Engineering, Kouty nad Desnou, Czech Republic, May 20-22, 2015.

[7] S. R. D. Della, G. P. Ferraria and A. Flammini, "Distributed monitoring system for voltage dip classification over distribution grid," Sustainable Energy, Grids and Networks, vol. 6, pp. 70-80, June 2016.

[8] Y. Wang, A. Bagheri, M. H. J. Bollen and X. Y. Xiao, "single-event characteristics for voltage dips in three-phase systems," IEEE Trans. Power Delivery, vol. 32, no. 2, pp. 832-840, Apr. 2017.

[9] A. K. Ghosh and D. L. Lubkeman, "The classification of power system disturbance waveforms using a neural network approach," IEEE Trans. Power Delivery, vol. 10, pp. 671–683, July 1990.

[10] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier, Part I: Theoretical foundation and Part II: Application," in Proc. IEEE Power Eng. Society Winter Meeting, 1997.

[11] J. S. Huang, M. Negnevitsky, and D. T. Nguyen, "A Neural-Fuzzy Classifier for Recognition of Power Quality Disturbances", IEEE Trans. Power Delivery, vol. 17, no. 2, pp. 609–616, Apr. 2002.

[12] M. Negnevitsky and M. Ringrose, "A Neuro-Fuzzy System for Recognition of Power Quality Disturbances" (Panel Paper), in Proc. IEEE/PES General Meeting, San Francisco, CA, USA, 12-16 June 2005.

[13] M. Najafabadi, F. Villanustre, T. M Khoshgoftaar, N. Seliya, R. Wald and E. Muharemagic, "Deep learning applications and challenges in big data analytics," Journal of Big Data, vol. 2, no. 1, pp. 1-21, Feb. 2015.13

[14] Convolutional Neural Networks, Available online: http://cs231n.github.io/convolutional-networks/

[15] E. Balouji and O. Salor, "Classification of power quality events using deep learning on event images," in Proc. IPRIA, pp. 216-221, 2017.14

[16] A. Gensler, J. Henze, B. Sick and N. Raabe, "Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM neural networks," in Proc. SMC, pp. 9-12, 2016.

[17] J. Ma, J. Zhang, L. Xiao, K. Chen and J. Wu, "Classification of power quality disturbances via deep learning," Journal of IETE, Technical Review, vol. 34, no. 4, pp. 408-415, 2017.

[18] X. Tang, B. Li, H. Wang, J. Wei, G. Tang, Z. Zhang and K.J. Chen, "Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism," in Proc. ISPSD, pp. 31–34, 2016.

[19] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu and Y. Zhang, "Short-Term residential load forecasting based on LSTM recurrent neural network," IEEE Trans. Smart Grid, in print.

[20] Y. He, J. Deng and H. Li, "Short-term power load forecasting with deep belief network and copula models," in Proc. IHMSC, pp. 191-194, 2017.

[21] C. L. Becnel, "Maintaining process continuity during voltage dips," IEEE Trans. Ind. Appl., vol. 8, no. 4, pp. 324-328, July/Aug. 1982.

[22] L. Conrad, K. Little, and C. Grigg, "Predicting and preventing problems associated with remote fault-clearing voltage dips," IEEE Trans. Ind. Appl., vol. 27, no. 1, pp. 167-, Jan./Feb. 1991.

[23] M. H. J. Bollen, "Characterization of voltage sags experienced by three-phase adjustable-speed drives", IEEE Trans. Power Delivery, vol. 12, no. 4, pp. 1666-1671, Oct. 1997.

[24] M. H. J. Bollen and L.D. Zhang, "Different methods for classification of three-phase unbalanced voltage dips due to faults," Electric Power Systems Research, vol. 66, no. 1, pp. 59-69, July 2003.

[25] M. H. J. Bollen, and R.A.A. De Graaff, "Behavior of AC and DC drives during voltage sags with phase-angle jump and three-phase unbalance," In Power Engineering Society, Winter Meeting, IEEE, vol. 2, pp. 1225-30, 1999.

[26] E. Styvaktakis, M. H. J. Bollen and I. Y. H. Gu, "Expert system for classification and analysis of power system events," IEEE Trans. Power Delivery, vol. 17, no. 2, pp. 423-428, Apr. 2002.

[27] V. Ignatova, P. Granjon and S. Bacha, "Space vector method for voltage dips and swells analysis," IEEE Trans. Power Delivery, vol. 24, no. 4, pp. 2054-2061, Sep. 2009.

[28] L. D. Zhang and M. H. J. Bollen, "Characteristic of voltage dips (sags) in power systems," IEEE Trans. Power Delivery, vol. 15, no. 2, pp. 827–832, Apr. 2000.

[29] P. G. V. Axelberg, I. Y. H. Gu and M. H. J. Bollen, "Support Vector Machine for classification of voltage disturbances," IEEE Trans. Power Delivery, vol. 22, no. 3, pp. 1297-1303, July 2007.

[30] S. Santoso, E. J. Powers, W. M. Grady and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier. I. Theoretical foundation," IEEE Trans. Power Delivery, vol.15, no.1, pp. 222-228, Jan. 2000.

[31] Voltage dip immunity of equipment and installations, CIGRE/CIRED/UIE Working Group C4.110, Apr 2010. www.uie.org; www.e-cigre.org.

[32] S. Z. Djokic, J. V. Milanovic, D. J. Chapman, M. F. McGranaghan, and D. S. Kirschen, "A new method for classification and presentation of voltage reduction events," IEEE Trans. Power Delivery, vol. 20, no. 4, pp. 2576-2584, Oct. 2005.

[33] V. Ignatova, P. Granjon, S. Bacha and F. Dumas, "Classification and characterization of three phase voltage dips by space vector methodology," presented at Int. Conf. Future Power Systems, Nov. 2005.

[34] A. Bagheri, M. H. J. Bollen and I. Y. H. Gu. "Improved characterization of multi-stage voltage dips based on the space phasor model," Electric Power Systems Research, vol. 154, pp. 319-328, Jan. 2018.

[35] M. R. Alam, K. M. Muttaqi and A. Bouzerdoum, "A new approach for classification and characterization of voltage dips and swells using 3-d polarization ellipse parameters," IEEE Trans. Power Delivery, vol. 30, no. 3, pp. 1344-1353, June 2015.

[36] A. Bagheri and M. H. J Bollen, "Characterizing three-phase unbalanced dips through the ellipse parameters of the space phasor model", in Proc. ISGT-europ, pp. 1–6, 2017.

[37] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Advances in neural information processing systems, pp.1097-1105, 2012.

[38] J. Wolfshaar, M. F. Karaaba and M. A. Wiering, "Deep convolutional neural networks and support vector machines for gender recognition," in Proc. IEEE SSCI, 2015, pp. 218-230. Y. Cao, R. Xu and T. Chen, "Combining convolutional neural network and support vector machine for sentiment classification," in Proc. CNCSMP, 2015, pp. 144-155.

[39] Y. Gao, W. Rong, Y. Shen and Z. Xiong, "Convolutional neural network based sentiment analysis using Adaboost combination," in proc. IJCNN, 2016, pp. 1333-1338.

[40] S. Yang, L. F. Chen, T. Yan, Y. H. Zhao and Y.-J. Fan, "An ensemble classification algorithm for convolutional neural network based on AdaBoost," in proc. ICIS, 2017, pp. 401-406.

[41] Convolutional Neural Networks (LeNet), available: http://deeplearning.net/tutorial/lenet.html.

[42] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning", MIT Press. [online]. Available: http://www.deeplearningbook.org/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2018.2854677, IEEE Transactions on Power Delivery

9

[43] D. Burke, J. Brundage and R. Redfield, "Measurement of the false positive rate in a screening program for human immunodeficiency virus infections", The New England Journal of Medicine, vol. 319, pp. 961–964, Oct. 1988.

[44] PQube live world map of power quality, [online]. Available: http:www.powerstandards.com/product/pqube-classic/highlights/

[45] M.H.J. Bollen, "Algorithms for characterizing measured three-phase unbalanced voltage dips," IEEE Trans. Power Delivery, vol. 18, no. 3, pp. 937-944, July 2003.