A LSTM-based Deep Learning Method with Application to Voltage Dip Classification

Ebrahim Balouji [*], Irene Y.H. Gu [*], Math H.J. Bollen [+], Azam Bagheri [+], Mahmood Nazari [*]

[*] Dept. of Electrical Engineering, Chalmers Univ. of Technology, Gothenburg, Sweden
[+] Electrical Power Engineering, Lulea Univ. of Technology, Skellefteå, Sweden
Emails: {balouji, irenegu}@chalmers.se; {azam.bagheri, math.bollen}@ltu.se

***Abstract*--In this paper, a deep learning (DL)-based method for automatic feature extraction and classification of voltage dips is proposed. The method consists of a dedicated architecture of Long Short-Term Memory (LSTM), which is a special type of Recurrent Neural Networks (RNNs). A total of 5982 three-phase one-cycle voltage dip RMS sequences, measured from several countries, has been used in our experiments. Our results have shown that the proposed method is able to classify the voltage dips from learned features in LSTM, with 93.40% classification accuracy on the test data set. The developed architecture is shown to be novel for feature learning and classification of voltage dips. Different from the conventional machine learning methods, the proposed method is able to learn dip features without requiring transition-event segmentation, selecting thresholds, and using expert rules or human expert knowledge, when a large amount of measurement data is available. This opens a new possibility of exploiting deep learning technology for power quality data analytics and classification.**

***Index Terms*—Artificial intelligence, deep learning, LSTM, RNN, power quality, smart grid, voltage dips.**

I. INTRODUCTION

Deep learning (DL) is a class of machine learning algorithms that use techniques inspired by the human brain's ability to learn. Until recently, lack of training data and computational power has rendered deep learning inefficient on big-data pattern recognition problems such as image processing or power-quality analysis. The problem with lack of data for the latter has been partly solved by the wide-scale availability of power quality (PQ) monitoring equipment. In view of this, in the most recent works, all PQ data coming from PQ monitoring devices to the data center where PQ data is stored, are classified based on signal processing tools for electric power systems. Signal processing tools use extracted features based on the strict and exact threshold setting defined by experts for analyzing and classifying of PQ data. Voltage dip is an important subject in PQ analysis and a regular reason for malfunctions of equipment connected to the power system [1]. The work on voltage dip classification started in [2], [3] and [4] resulted in four classes of dips [5], referred to as A, B, C and D type voltage dips. Later the number of classes was extended to seven, as mentioned in [6] and [7]. The new classes are known as E, F and G types. Most recent work on dip classification centered on these types. Existing methods for dip classification (i.e. for obtaining the dip type from the measurement waveform,) are based on three main solutions: advanced feature extraction; RMS voltage; and complex voltages. In the feature extraction methods, signal processing tools like Support Vector Machine (SVM) [8], Fourier transform [9], wavelet transform [10], etc., are employed to extract useful features from the voltage-dip data. Based on the parameters and thresholds defined by experts, classification of voltage dips is achieved. [11]-[12] contain some examples of these types of works. Based on the RMS voltage, some intuitive classification methods to classify voltage dips have been developed. The methods in [14] and [16] are based on phase-to-neutral RMS sequences. Another example of this type of voltage dip analysis is [17] which considers, as one of the algorithms, the RMS-sequences of the six phase-to-neutral and phase-to-phase-voltages to obtain PN factor, dip type and characteristic. Three complex values of voltages are used as input in the third group of classification methods. The first method in this group is the Symmetrical Components algorithm developed in [18], [19] and [20]. The angle difference between positive and negative sequences is used to obtain the dip type. From negative and positive sequence voltages the characteristic voltage and PN factor are obtained, once the dip type is known. However, there are some problems in complex value extraction [20]. The Space Phasor Model (SPM) is also part of the third group. Reference [21] uses principal components of the SPM in the complex plane, for classifying and characterizing voltage dips.

Since in PQ analysis, we are dealing with some pattern like voltage drop or interruptions in voltage or current level, DL is capable of learning these patterns in different level of information. Each layer considers all features (input data or output of previous layer) and a result in a large gain in terms of the system's ability to data generalization. In this work, the aim is to go beyond the above-mentioned three categories of

This work was financially supported by the Swedish Energy Agency.

voltage-dip classification and to propose a method that does not require any hand-crafted features Instead the method should automatically learn and detect useful features. From the time when DL was developed, structures such as LeNet [22], GoogLeNet [22] and AlexNet [24] have been found to perform excellently for image and speech signals, e.g., character/digit recognition, traffic sign recognition, and face recognition. Similar methodologies could benefit power engineering research, for example, for automatically learning features of voltage dips without requiring human experts' knowledge, and subsequently perform automatic dip classification.

In this paper, we propose a novel and efficient long short-term memory (LSTM) architecture, a specific type of RNNs (recurrent neural networks), for automatically learning features and classification of three-phase voltage dips. The proposed scheme has been trained on a large measurement dataset and resulted in a classification accuracy of 93% on an independent test set.

II. DEEP LEARNING INTRODUCTION FOR PQ

In this section, we give a brief review on the basic of LSTM, which is employed as a fundamental building block of our feature extraction and classification method, for the sake of convenience in the subsequent sections. On the other hand, each layer categorizes some level of information, refines it and passes along to the next. So we can say that DL, lets the machine to use this process and build a higher article representation. For example, the first layer can look for simple voltage levels, and second layer can look to the collection how much voltage drop we have and the third might extracts which phase have voltage drop and the last one might see whole specialty of voltage dip in once. The neural network can put all of these results together and the machine can recognize the type of voltage dip in this case which is done in Fully Connected (FC) layer. In the subsection the LSTM as the most important element used to develop in this work will be presented.

A. Long short-term Memory (LSTM)

LSTM consists of state boxes receiving the inputs through time. In each time step, an input vector is fed into LSTM, the output is computed according to:

$$h_t = f_w(h_{t-1}, x_t) \tag{1}$$

where $\boldsymbol{x}_t$ is the input vector , $\boldsymbol{h}_t$ and $\boldsymbol{h}_{t-1}$ are the state vectors at time t and $(t-1)$, f_w is a nonlinear activation function where $\boldsymbol{w}$ are the weight parameters. An LSTM is often depicted as in Fig.1.

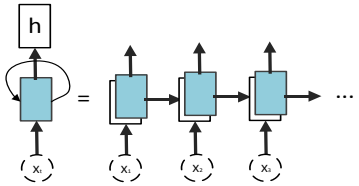

Figure 1. An unrolled LSTM layer.

Mathematically expression of unrolling mentioned in Fig.1.is given in (2) from [24] and [26]

$$\begin{cases} C_t = f \odot C_{t-1} + i \odot g \\ h_t = O \odot \tanh(C_t) \\ \begin{bmatrix} i \\ f \\ o \\ g \end{bmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} \mathrm{W} \begin{pmatrix} h_{t-1} \\ X_t \end{pmatrix} \end{cases} \tag{2}$$

where C_t represents output of cell, f is the forgetting gate controlling the cells on whether or not to erase it; i is input gate on whether to write to cell; g is the gate on how much to write to the cell; o is the output gate deciding how much to reveal cell; σ is the sigmoid function, and $\odot$ is the symbol of functions and can be functions such as *tanh, sigmoid*. In general view to (2), $i \odot g$ is the input from X and state h. $O \odot tanh(C_t)$ represents the state of cell and it can decide how much of cell is going to be revealed to the hidden state. In Fig. 2 the repeating model of unrolled LSTM and (2) is presented.

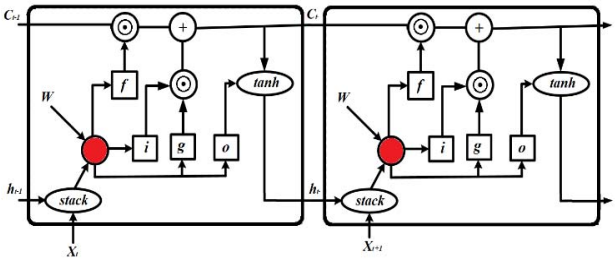

Figure 2. The repeating model of unrolled LSTM in detail. Red circles represents tanh and sigmoid functions.

III. PROPOSED METHOD

A. Overview of the proposed method

The proposed method is shown in the block diagram of Fig. 3. The input three-phase voltage sequence is first pre-processed to RMS sequence and next divided into 30 segments for the computational efficiency in LSTM.

The proposed LSTM architecture for dip feature learning and classification is depicted in Fig. 4, with further details included in Table I. The proposed architecture consists of four LSTM layers, with *tanh* as the activation function. These four layers are used for automatically extracting dip features. Different LSTM layers are used for learning multiscale features from voltage dips. In the final layer, which is a fully connected (FC) layer with a softmax activation function, is used as the classifier. The classifier uses the output from the last layer LSTM as the input, and finds the voltage dip type for a given test data.

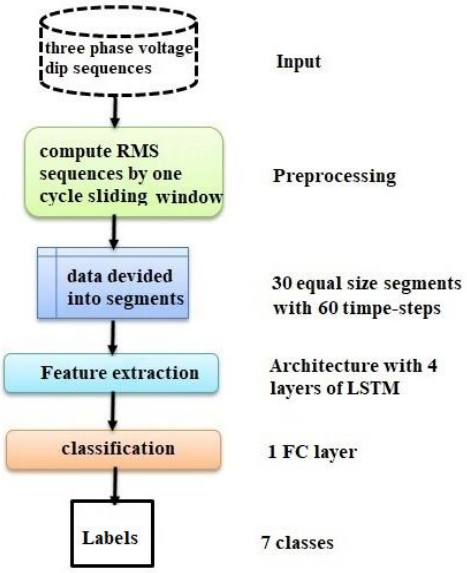

Figure 3. Overview and block diagram of the proposed method.

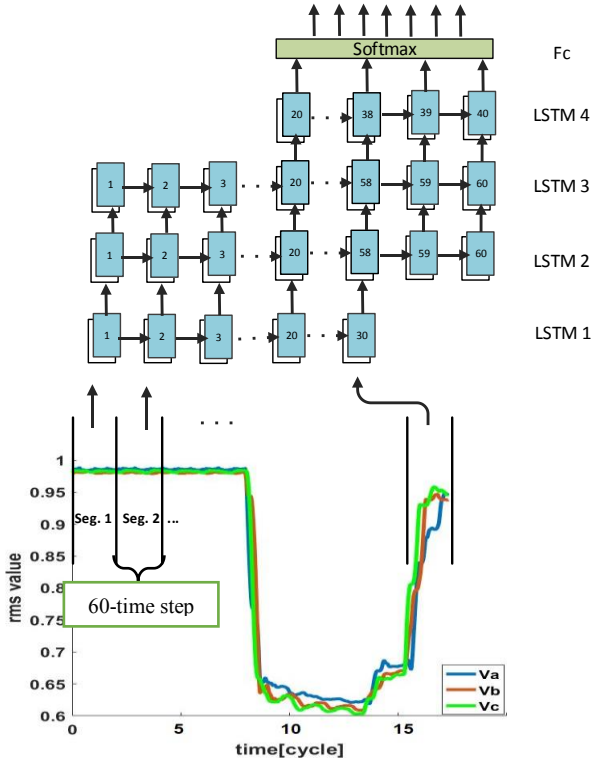

Figure 4. Proposed deep learning method: the LSTM architechture.

TABLE I. LSTM ARCHITECTURE, WHERE FOUR LSTM LAYERS ARE USED FOR FEATURE EXTRACTION, ONE FC LAYER IS USED FOR CLASSIFICATION.

Layers	Input size	Output size	Cells
LSTM1 +tanh +BN	30	60	30
LSTM2 +tanh+BN	60	60	60
LSTM3 +tanh+BN	60	60	60
LSTM3 +tanh+BN	40	40	40
FC1 + softmax	40	1x7	7

B. Dataset and labels

The dataset used in this work consists of 5982 voltage dip events. The classification method presented [21] has been used for labelling which consists of seven classes; class A and 6 subclasses of C and D. Subclasses are represented by C_a, C_b, C_c, D_a, D_b, D_c. Table II shows the distribution of measurement data over the seven classes, as defined in [18] and [21].

The three-phase voltage definition based on classes is related to the voltage drops in the different phases. In voltage dip class A, a serious voltage drop in all phases is observed. In the case of having serious voltage drop in two phases, the voltage dip is typically labeled as C_a, C_b and C_c. Finally, in the case where a serious voltage drop is observed in one phase, the three-phase voltage dip is represented in one of the classes called: D_a, D_b and D_c. See [18] and [19]for more details.

TABLE II. DATASET DESCRIPTION: THE NUMBER OF DIPS IN EACH CLASS

Class	A	C_a	C_b	C_c	D_a	D_b	D_c	Total
# data	2232	693	742	631	464	618	602	5982

The dataset is then randomly split into three subsets: 60% is used for training, 20% for validation, and the remaining 20% for testing. An example of a Type A voltage dip and the corresponding RMS sequence shape is shown in Fig. 5. A serious voltage drop is observed in all three phases.

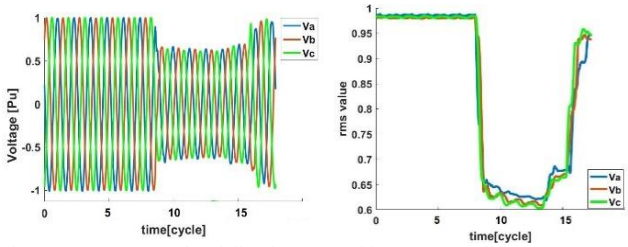

Figure 5. An example of dip class A, and its corresponding RMS curves.

C. Forming the input vector

For forming the input vector:First, we compute the RMS sequences from each three-phase measurement data using sliding over one sample with one cycle window length. The original waveform is first normalized in range (-1,1). The resulting three-phase RMS sequences are normalized in range (0,1) and concatenated as follows:

$$\mathbf{x}_{dip} = [RMS_A, RMS_B, RMS_C] \quad (3)$$

where $V_{Dip} = \begin{bmatrix} V_A[0] & V_A[1] \dots V_A[599] \\ V_B[0] & V_B[1] \dots V_B[599] \\ V_C[0] & V_C[1] \dots V_C[599] \end{bmatrix}$ is the measurement three-phase voltage data for each dip event.

To efficiently employ the LSTM, we further split the RMS sequence in (3) into 30 equal sized segments. The reason behind this segmentation is based on the ability of LSTM cells to memorize the number of input time step. As mentioned in [24], the optimized value for the time step of data fed into each cell is 60 samples. Segments and time steps are depicted in Fig. 4.

IV. EXPERIMENTAL RESULTS

In this section, experimental results from using the proposed voltage dip classification scheme will be presented.

A. Setup

The computer used for our experiments is a workstation with an Intel i7 3.40 GHz CPU, 48 GB RAM and an NVIDIA Titan Xp 128 GB GPU. The program is implemented in python, using Keras library with TensorFlow backend. The dataset is partitioned with 60% for training, 20% for validation, and 20% for the testing. The hyper-parameters used to train are carefully tuned and are given in Table III.

TABLE III. DATASET PARTITION, PARAMETERS AND HYPER-PARAMETERS USED IN OUR EXPERIMENTS

Parameters and Hyper Parameters	case study
# train data set	3583 (60%)
# validation data set	1195 (20%)
# test data set	1195 (20%)
Total number of events	5982
batch size	50
# epochs	500
data dimension	30
time steps	60
# classes	7
η(learning rate)	0.001
η_decay (decay of learning rate)	0
FC dropout	0.3
recurrent dropout	0.3
RNN dropout = 0	0.3
estimated processing time	9 hours
optimizer	Adam

In Table III, the data split, hyper parameter ratio and optimizer used are shown with green, blue and yellow color, respectively. The hyper-parameters were obtained based on the grid search where many values were tried, and the combination of parameters that has led to the best possible results are used for the test later. In addition, the optimizer was chosen as "*adam*" (adaptive moment estimation) for parameter updates on each epoch. The "*adam"* optimizer can be described by

$$v_t = \beta_1 m_t - 1 + (1-\beta_1) g_t$$
$$m_t = \beta_1 m_t - 1 + (1-\beta_1) g_t \quad (4)$$

where m_t and v_t are the estimation of first and second gradients respectively. More detail can be found in [27].

B. Results

For evaluate the effectiveness of the proposed LSTM architecture, a range of tests were performed.

Training and validation:

Fig. 6 shows the performance (accuracy and loss) of the proposed scheme from the training and validation as a function of epochs. Further, we summarize the performance from the training, validation and testing in Table IV.

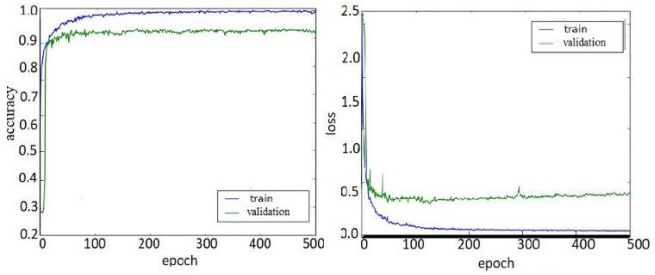

Figure 6. Performance from the proposed LSTM method. Left: accuracy vs. epochs; Right: Loss vs. epochs; Blue curve: from the training; Green curve: from the validation.

TABLE IV. SUMMARY OF THE PERFORMANCE ON THE TRAINING, VALIDATION AND TEST SET. PARAMETERS IN THE CLASSIFIER ARE FIXED FROM THE TRAINING PROCESS AT EPOCH=500.

Dataset	Accuracy (%)	Loss
Training set	99.01	0.0800
Validation set	93.12	0.5102
Test set	**93.40**	**0.5001**

Testing:

The testing performance is shown in the 4th row of Table IV, where the parameters of the classifier were fixed from the training process at epoch=500.

Observing the difference of between the training and validation/testing, one can see that there is a drop of about 5.6%. This is probably caused by the overfitting. This issue may be the subject of further study by fine tuning hyper-parameters (e.g. grid search) and refining the architecture.

Performance:

To further evaluate the performance of the proposed LSTM scheme in each individual class, Table V shows the confusion matrix on the test data set.

TABLE V. CONFUSION MATRIX FOR TEST DATASET: RESULTS OF CLASSIFICATION FOR EACH CLASS INDIVIDUALLY. $\hat{A}$ $\hat{C}$ AND $\hat{D}$ ARE THE LABELS OBTAINED FROM THE CLASSIER, WHILE A, C AND D ARE THE "GROUND TRUTH" LABELS.

Classes	A	C_a	C_b	C_c	D_a	D_b	D_c
$\hat{A}$	**237**	4	0	3	0	0	0
$\hat{C}_a$	3	**167**	0	3	0	2	1
$\hat{C}_b$	0	2	**196**	0	7	0	5
$\hat{C}_c$	0	0	3	**162**	1	7	3
$\hat{D}_a$	6	0	2	9	**98**	0	2
$\hat{D}_b$	2	5	0	0	0	**121**	1
$\hat{D}_c$	2	2	3	1	1	1	**133**

Based on the confusion matrix given in Table V, the classification rate and false alarm rate for each class is calculated and is shown in Table VI.

TABLE VI. CLASSIFICATION RATEAND FALSE ALARM ON THE TEST SET.

Class	A	C_a	C_b	C_c	D_a	D_b	D_c
False alarm (%)	5.2	7.78	4.0	9.8	9.18	7.6	8.27
Classification rate (%)	94.8	92.22	96	90.2	90.82	93.4	91.73

Observing Table VI, one can see that the classification accuracies are relatively uniform, with the highest classification rate 96% for the class C_b , and the lowest 90.2% for the class C_c. Furthermore, the variance of false alarm rates is also relatively small, with the lowest false alarm 4% and highest 9.8%. Observing the results of Tables IV – VI, one can conclude that the proposed DL method has effectively learned features from RMS voltage dip sequences, which has led to good classification performance.

V. CONCLUSION

In this paper, a deep learning (DL)-based solution is developed for voltage dip feature extraction and classification. The proposed method learns three-phase voltage dip patterns from time-dependent RMS sequences. The dataset used in this work consists of 5982 voltage dips obtained from several countries, partitioned into training, validation and test subsets according to 60%, 20% and 20%. The proposed method is shown to be effective, with an average classification accuracy of 93.4% on the test set. In addition, analysis of the results on individual classes is shown that relatively uniform classification rates and small false alarms were achieved over all classes (the range of classification rates is [90.2%, 96.0%], false alarm is [4.0%, 9.8%]). For dip feature learning, the proposed 4-layer LSTM structure is employed for extracting voltage dip features from RMS sequences. For classification, one fully connect (FC) layer with *softmax* activation function is employed for classifying 7 types of voltage dips using the features provided by the LSTM outputs. The combination of feature extraction layers (i.e., LSTMs) and classification layer (i.e., FC) is shown to be effective. Observing the results from the proposed method implies that DL-based methods may be a good choice for future PQ analysis without requiring human experts, if large data is available. Furthermore, it could provide opportunities for developing automatic intelligent counter-measures for some parts of the electricity grid.

REFERENCES

[1] M. Bollen and I. Y. H. Gu, "Signal Processing of Power Quality Disturbances," Wiley-IEEE Press, Hoboken, USA, 2006.

[2] M. F. McGranaghan, D. R. Mueller, and M. J. Samotyj, "Voltage Sags in Industrial Systems," *IEEE Trans. Ind. Appl.*, vol. 29, no. 2, pp. 397–403, 1993.

[3] C. L. Becnel, "Maintaining Process Continuity During Voltage Dips," *IEEE Trans. Ind. Appl.*, vol. IA-18, no. 4, pp. 324–328, 1982.

[4] K. Yao, D. Koval, W. Xu, and J. Salmon, "An Investigation of Voltage Dips by A Phasor Methodology," Proc. of IEEE Canadian Cong. Elect. Computer. Eng. conf., Ottawa, Canada, 1998.

[5] M. H. J. Bollen, "Characterisation of Voltage Sags Experienced by Three Phase Adjustable Speed Drives," *IEEE Trans. Power Del.*, vol. 12, no. 4, pp. 1666–1671, 1997.

[6] 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," Proc., IEEE Power Eng. Conf. Soc., New York, 1999.

[7] M. H. J. Bollen, "Understanding Power Quality Problem: Voltage Sags and Interruptions," Piscataway, NJ, USA: IEEE, 2000.

[8] L. Lovisolo , J.A. Moor Neto , K. Figueiredo , L. M. D. Laporte, and J. C. D. S. Roch, "Location of Faults Generating Short-Duration Voltage Variations in Distribution Systems Regions from Records Captured at One Point and Decomposed into Damped Sinusoids," *IET Generation, Transmission and Distribution,* vol. 6, no. 12, pp. 1225 –1234, 2012.

[9] M. H. Chia and A. M. Khambadkone, "Subcycle Voltage Dip Classification Using Matrix Pencil Method With Ellipse Fitting Algorithm," *IEEE Trans. Industry Applications*. vol. 51, no. 2, pp. 660–1668, 2015.

[10] H. Eris and Y. Demir "Automatic Classification of Power Quality Events And Disturbances Using Wavelet Transform And Support Vector Machines," *IET Generation, Transmission and Distribution* vol. 6, no. 10, pp. 968 – 976, 2012.

[11] T. C. Oliveira, J. M. C. Filho and R. C. Leborgne, "Voltage Sags: Validating Short-Term Monitoring by Using Long-Term Stochastic Simulation," *IEEE Trans. on Power Delivery*, vol. 24, no. 3, 2009.

[12] F. B. Costa and J. Driesen, "Assessment of Voltage Sag Indices Based on Scaling and Wavelet Coefficient Energy Analysis," *IEEE Transactions on Power Delivery*, Vol. 28, no. 1, pp. 336–346, 2013.

[13] K. Chen, J. Hu, J. He, "A Framework for Automatically Extracting Overvoltage Features Based on Sparse Auto-encoder," *IEEE Transactions on Smart Grid*, no. 99, pp. 1-11, April. 2016.

[14] P. G. V. Axelberg, I. Y. H. Gu and M. H. J. Bollen "Support Vector Machine for Classification of Voltage Disturbances," *IEEE Transactions on Power Delivery*, vol. 22, no. 3, pp. 1297 – 1303, 2007.

[15] S. Z. Djokic, J. V. Milanovic, D. J. Chapman and M. F. McGranaghan, "Shortfalls of Existing Methods For Classification And Presentation Of Voltage Reduction Events," *IEEE Trans. Power Del*., vol. 20, no. 2, pp. 1640–1649, 2005.

[16] M. Guder, O. Salor, I. Çadirci, B. Ozkan and E. Altintas, "Data Mining Framework for Power Quality Event Characterization of Iron and Steel Plants," *IEEE Trans on Industry Applications*, vol. 51, no. 4, pp. 3521 – 3531, 2015.

[17] M. H. J. Bollen, "Algorithm for Characterizing Measured Three-Phase Unbalanced Voltage Dips," IEEE Trans. Power Del., vol 20, no 4, pp. 2576 - 2584, 2005.

[18] Y. Wang, M. H. J. Bollen, X. Y. Xiao and M. Olofsson, "Single-Event Characteristics for Voltage Dips in Three-Phase Systems," *IEEE Transactions on Power Delivery*, vol.32, no. 2, pp. 832-840, 2017.

[19] L. Zhang and M. H. J. Bollen, "Characteristic of Voltage Dips (Sags) In Power Systems," *IEEE Trans. Power Del.*, vol. 15, no. 2, pp. 827–832, 2000.

[20] M. H. J. Bollen, "Algorithms for Characterizing Measured Three-Phase Unbalanced Voltage Dips," *IEEE Trans. Power Del.*, vol. 18, no. 3, pp. 937–944, 2003.

[21] 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, 2018.

[22] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-Based Learning Applied," Proceedings of The IEEE, vol. 86, no. 11, pp. 2278 – 2324, 1998.

[23] S. M. A. Eslami, S. Mohamed, P. Battaglia, M. Jaderberg and N. Heess, "Unsupervised Learning of 3D Structure from Images," eprint arXiv:1607.00662, 07/2016.

[24] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[25] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory. Neural Computation," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[26] T. N. Sainath, O. Vinyals and A. H. Sak, "Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks," proc. of IEEE Int. conf. on Acoustics, Speech and Signal Processing (ICASSP) 2015.

[27] J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *Journal of Machine Learning Research*, no. 12, pp. 2121–2159, 2011.