

# Electric Power Systems Research

## POWER QUALITY DISTURBANCES DIAGNOSIS: A 2D DENSELY CONNECTED CONVOLUTIONAL NETWORK FRAMEWORK

--Manuscript Draft--

<b>Manuscript Number:</b>	EPSR-D-21-03036R1
<b>Article Type:</b>	Research Paper
<b>Keywords:</b>	Machine Learning; Pattern recognition; Power quality disturbances; Electric Distribution Systems.
<b>Corresponding Author:</b>	Raul Vitor Arantes Monteiro, Ph.D. Universidade Federal de Mato Grosso CUIABÁ, BRAZIL
<b>First Author:</b>	Raul Vitor Arantes Monteiro, Ph.D.
<b>Order of Authors:</b>	Raul Vitor Arantes Monteiro, Ph.D. Raoni F. S. Teixeira, PhD. Arturo S. Bretas, PhD.
<b>Abstract:</b>	<p>The fast and accurate diagnosis of power quality disturbances (PQD) aids in avoiding shutdowns and unnecessary procedures, concerning electric energy distribution systems. As such, a number of techniques have been tested and applied in order to reach this objective. Majority of the techniques applied are two-step based. On the first step, power quality disturbances features are extracted. Second step, considering features extracted, disturbance classification is implemented. Recently, relevant literature has presented data-driven signal processing-based approaches, as deep convolutional neural networks (DCNN), which can implement both processing steps while providing automated recognition of patterns and outliers in data. However, not considered by state-of-art, power quality disturbances are evolving in nature, while all possible regularities might not be represented in the dataset. In this work a 2 Dimension Densely Connected Convolutional Network (2D-DenseNet) framework is presented. Case study with synthetic disturbance events are analyzed. Easy-to-implement formulation, built on the 2D-DenseNet, without hard-to-design parameters, highlight potential aspects for real-life implementation.</p>

# POWER QUALITY DISTURBANCES DIAGNOSIS: A 2D DENSELY CONNECTED CONVOLUTIONAL NETWORK FRAMEWORK

**Raul V. A. Monteiro<sup>a,\*</sup>, Raoni F. S. Teixeira<sup>a,b</sup>, Arturo S. Bretas<sup>c</sup>**

<sup>a</sup>Electrical Power Systems Operation and Smart Grid Research Lab., Federal University of Mato Grosso, MT, Brazil

<sup>b</sup>Group for Innovation on Real Information System, Computing Engineering Department, Federal University of Mato Grosso, Campus Várzea Grande, MT, Brazil

<sup>c</sup>Electrical and Computer Engineering Department, University of Florida, Gainesville, FL, USA.

E-mail addresses: raulvitor@ufmt.br; raoni@ufmt.br; arturo@ece.ufl.edu;

\*Correspondence: raulvitor@ufmt.br

February, 15<sup>th</sup>, 2022.

Dear Professor Maria Teresa Nunes Padilha de Castro Correia de Barros,

I am deeply grateful for the opportunity to submit the article of the title **POWER QUALITY DISTURBANCES DIAGNOSIS: A 2D DENSELY CONNECTED CONVOLUTIONAL NETWORK FRAMEWORK** to the journal, **ELECTRIC POWER SYSTEMS RESEARCH**. This is an original work study.

Raul Vitor Arantes Monteiro

## HIGHLIGHTS

# **POWER QUALITY DISTURBANCES DIAGNOSIS: A 2D DENSELY CONNECTED CONVOLUTIONAL NETWORK FRAMEWORK**

- A 2D DenseNet does not need prior extraction of features for PQD signals by some mathematical method, which eliminates the 2-step based model and reduces computational burden;
- Spatio-temporal environment characteristics are considered in a data driven based framework. A 2D DenseNet solution allows for the reduction in the quantity of filters per convolutional layer of DenseNet used to classify PQDs;

# POWER QUALITY DISTURBANCES DIAGNOSIS: A 2D DENSELY CONNECTED CONVOLUTIONAL NETWORK FRAMEWORK

Raul V. A. Monteiro<sup>a</sup>, Raoni F. S. Teixeira<sup>b</sup>, Arturo S. Bretas<sup>c</sup>,

<sup>a</sup> Electrical Power Systems Operation and Smart Grid Research Lab., Federal University of Mato Grosso, MT, Brazil

<sup>b</sup> Group for Innovation on Real Information System, Computing Engineering Department, Federal University of Mato Grosso, Campus Várzea Grande, MT, Brazil

<sup>c</sup> Electrical and Computer Engineering Department, University of Florida, Gainesville, FL, USA.

E-mail addresses: [raulvitor@ufmt.br](mailto:raulvitor@ufmt.br); [raoni@ufmt.br](mailto:raoni@ufmt.br); [arturo@ece.ufl.edu](mailto:arturo@ece.ufl.edu);

\*Correspondence: [raulvitor@ufmt.br](mailto:raulvitor@ufmt.br)

## Abstract

**The fast and accurate diagnosis of power quality disturbances (PQD) aids in avoiding shutdowns and unnecessary procedures, concerning electric energy distribution systems. As such, a number of techniques have been tested and applied in order to reach this objective. Majority of the techniques applied are two-step based. On the first step, power quality disturbances features are extracted. Second step, considering features extracted, disturbance classification is implemented. Recently, relevant literature has presented data-driven signal processing-based approaches, as deep convolutional neural networks (DCNN), which can implement both processing steps while providing automated recognition of patterns and outliers in data. However, not considered by state-of-art, power quality disturbances are evolving in nature, while all possible regularities might not be represented in the dataset. In this work a 2 Dimension Densely Connected Convolutional Network (2D-DenseNet) framework is presented. Case study with synthetic disturbance events are analyzed. Easy-to-implement formulation, built on the 2D-DenseNet, without hard-to-design parameters, highlight potential aspects for real-life implementation.**

***Keywords* – Machine Learning; Pattern Recognition; Power Quality Disturbances; Electric Distribution Systems.**

## 1. Introduction

Industrialization processes aligned with the growth of power electronics based equipment use have increased the impact of power quality disturbances. The intensity of these disturbances and the effects that they cause have led to many efforts towards the detection, identification and location of these events [1].

Considering these efforts, international standards were also elaborated. Among such, the IEEE-1159, IEC-61000-4-30 and EN 50160, which define and characterize power quality disturbances in regards to time and duration, amplitude and form are highlighted [1]. According to [1], power quality disturbances can be classified in the following types: voltage sag, voltage swell, harmonic distortions, notching, voltage fluctuation; interruptions, oscillatory transient and spike. Further, the source of these disturbances are also identified in recent works [2].

Detection, identification and classification of power quality disturbances, play a leading role in the supervision, planning, operation and maintenance of energy distribution systems [3].

Due to its importance, a number of studies have been related to the subject, along with the testing and application of advanced techniques. In [4], the authors use the continuous wavelet transform (CWT) towards the extracting of 5 features from each measured signal for PQD classification, implemented through a ruled-based decision tree. Noise at 40 dB, 30 dB and 20 dB was mixed to measured signals, resulting in a classification accuracy of 99.7%, 98.5% and 93.8%, respectively. The authors in [5] apply the multiresolution S-Transform to produce instantaneous frequency vectors of measured signals, and then use Parseval's theorem to calculate the energy for these vectors. After performing this process of feature extraction, the classification of disturbances is performed. The types of disturbances are classified by region (1, 2 and 3), using the signal frequency characteristics and are identified by the energy contained in these vectors. The authors conclude that the presented algorithm performs well in classifying the disturbances tested. In [6], a technique is presented for detecting and classifying signals using the S-Transform (ST) and a probabilistic artificial neural network. Presented results are most encouraging, with an accuracy in classification of 97.4%. Four features are extracted from the input measured signals. A decision tree (DT)-initialized fuzzy rule base solution for power quality disturbance classification is presented in [7]. These results highlight an average accuracy of 95.22%. In [8], features of the voltage signal are extracted by means of Fourier Transforms (FT) and Wavelet Transforms. A fuzzy system then classifies the disturbances based on extracted features. An accuracy of 96% in the classification is obtained for signal combinations containing disturbances and noise. The authors in [9] present a method that quantifies power quality parameters (PQ) using wavelets and the theory of fuzzy sets. A classification accuracy of 97.65%, 93.54% and 90.32% for signals with noise of 6 dB, 40 dB and 30 dB is reported, respectively. In [10], a chaos synchronization (CS)-based detector is proposed, the results show good accuracy in signal discrimination, fast learning, good robustness and rapid processing in the detection of

disturbances. The ST and the TT transforms are used to extract the features from the measured signals in frequency and time domains in [11], with average general accuracies of 98.1%, 97.2% and 95.2% reached for signals, without noise, at 30 dB noise and 20 dB noise, respectively. In [12], the WT is used for extracting the features from the three-phase signals, and these are then used for classification done through a support vector machines (SVM) based approach. Accuracy of 98.51% in classification is reported. [13] presents an innovating technique using the modified frequency slice WT (MFSWT), which is an extension of the frequency slice WT (FSWT) for extracting the features from signals. Results presented shown to be promising considering further several measured signals with mixed noise. An approach that takes into consideration the combination of two artificial neural networks (ADALINE) is performed in [14], the authors arrived at a classification accuracy of 95.47% when the two signals were merged and 86.23% when two signals were merged with noise. Other studies that use DT, have also reported high accuracy levels [15-18]. In [19], the ensemble empirical mode decomposition (EMD) is used for extracting signal features and a rank wavelet support vector machine (rank-WSVM) is used for classifying the disturbances. The technique reaches average classification accuracy of 92.80%, 92.70%, 90.65% and 90.30% for signals with noise of 20 dB, 30 dB, 40 dB and 50 dB, respectively. In [20], a double resolution ST is used for extracting features and a direct acyclic graph SVM is used for classifying PQDs. [20] reports an accuracy level of 99.3%. The authors in [21] present a method based on Discrete Gabor Transform (DGT) with a Finite Impulse Response Window (FIR-DGT) and a Type-2 Fuzzy Kernel-based Support Vector Machine (T2FK-SVM). The authors report an accuracy of 99.44% in the classification of disturbances. [22] presents a time–frequency-scale transform (TFST) derived from Chirplet transform (CT). Three classifiers are used: SVM, DT and BP, 18 features are extracted from the signals for classification. An accuracy of 99.27% is reported for simulated signals and an average accuracy of 94.26% for real-life signals. The authors in [23] present a modern adaptive signal processing technique called variational mode decomposition (VMD) for power quality (PQ) events detection, where an accuracy of 98.82% was reported. In [24], the authors use deep convolutional neural networks (DCNN) for classifying voltage dips (sags), reaching a classification accuracy of 97.72%. In [25], an integrated multivariate singular spectrum analysis (MSSA) and curvelet transform (CVT) are proposed for feature extraction of PQ disturbance signals. Considering 9 types of disturbances, [25] reported an average accuracy of 99.89%. In [26], the DCNN is applied to PQD classification. A unit construction which consists of 1D convolutional, pooling, and batch-normalization layers is designed to capture multi-scale features and reduce overfitting. Considering 16 types of PQD, [26] reports an accuracy of 99.96%, 99.95%, 99.66% and 98.13%, for pure signals, with noise at 40 dB, 30 dB and 20 dB, respectively. In [27] the authors present a classification model based on compressed sensing and deep convolutional neural network. In [28], authors presented a 1D – 2D convolutional neural network approach.

Considering the mentioned studies [4]-[25], solutions make use of mathematical methods for extracting measured or synthetic signals features towards event classification. This can increase PQD classification

precision, additive to an implicit computation burden cost. [26, 27, 28], otherwise, do not make use of an two-step approach, once they do not make use of feature extraction methods.

However, in none of the previous reported works, are there found references to the evolving nature of PQDs as well as the limitation of existing data. Thus, presented solutions consider that all patterns classes are present in considered dataset. Further, that events are static. This consideration is based on the hypothesis that all PQDs are not dynamic and evolve in time, which is a strong and somehow unrealistic assumption.

This work presents a data-driven model based framework for PQDs diagnosis. Detection, identification and location of PQDs is realized simultaneously. To avoid the need of a two-step process, a two dimensional (2D) Densely Connected Convolutional Network (DenseNet) based solution is developed. This avoids the implicit computational cost while maintaining accuracy. Towards PQDs location, spatio-temporal characteristics are considered by the framework. The specific contributions of this work towards the state of the art are:

- A 2D DenseNet does not need prior extraction of features for PQD signals by some mathematical method, which eliminates the 2-step based model and reduces computational burden;
- Spatio-temporal environment characteristics are considered in a data driven based framework. A 2D DenseNet solution allows for the reduction in the quantity of filters per convolutional layer of DenseNet used to classify PQDs.

The remainder of the paper is divided as follows: section 2 presents the problem statement. Section 3 presents a case study. Section 4 present the concluding remarks of this work.

## 2. Problem statement

### 2.1 Mathematical models

Sixteen (16) types of PQD are considered in this work. PQDs mathematical models consist of 10, simple types, waveforms as: pure sine, sag, swell, interruption, harmonics, impulsive transients, oscillatory transients, flicker, notch and spike [8]. The 6, complex, types of PQD, consist of sag with harmonics, swell with harmonics, interruption with harmonics, flicker with harmonics, flicker with sag and flicker with swell. The models and model parameters, in line with the standard IEEE-1159, can be found in [8]. They are presented on Table 1.

**Table 1**

Equations and Parameters Variations

Label	PQ disturbances	Mathematical equations	Parameters
-------	-----------------	------------------------	------------

C1	Pure 60 Hz	$v(t) = \sqrt{2} \sin(2\pi \times 60t + \varphi)$	$-\pi \leq \varphi \leq \pi$ $0.1 \leq v_n \leq 0.9$
C2	Sag	$v(t) = \sqrt{2}v_n \sin(2\pi \times 60t + \varphi) [u(t - T_s) - u(t - T_e)]$	$-\frac{\pi}{4} \leq \varphi \leq \frac{\pi}{4}$ $T_s = T, 2T, \dots, 8T$ $T_e - T_s = T, 2T, \dots, 9T$ $1.1 \leq v_n \leq 1.8$
C3	Swell	$v(t) = \sqrt{2}v_n \sin(2\pi \times 60t + \varphi) [u(t - T_s) - u(t - T_e)]$	$-\frac{\pi}{4} \leq \varphi \leq \frac{\pi}{4}$ $T_s = T, 2T, \dots, 8T$ $T_e - T_s = T, 2T, \dots, 9T$ $0 \leq v_n \leq 0.1$
C4	Interruption	$v(t) = \sqrt{2}v_n \sin(2\pi \times 60t + \varphi) [u(t - T_s) - u(t - T_e)]$	$-\frac{\pi}{4} \leq \varphi \leq \frac{\pi}{4}$ $T_s = T, 2T, \dots, 8T$ $T_e - T_s = T, 2T, \dots, 9T$
C5	Flicker	$v(t) = \sqrt{2} \sin(2\pi \times 60t + \varphi) [1 + v_n \sin(2\pi f_n t)]$	$0 \leq \varphi \leq 2\pi$ $0.1 \leq v_n \leq 0.5$ $8 \leq f_n \leq 25$ $0.3 \leq v_n \leq 1$
C6	Oscillatory Transient	$v(t) = \sqrt{2}v_n \sin(2\pi \times 60t) [1 + \sqrt{2}v_n e^{-\alpha(t-T_s)} \sin(2\pi f_n(t - T_s) + \varphi)]$	$150 \leq \alpha \leq 1000$ $T_s = T, 2T, \dots, 7T$ $300 \leq f_n \leq 900$ $0 \leq \varphi \leq 2\pi$
C7	Harmonic	$v(t) = \sqrt{2} [\sin(2\pi \times 60t + \varphi) + \dots + \alpha_{2k} \sin(2\pi(2k) \times 60t + \varphi_{2k}) + \alpha_{2k+1} \sin(2\pi(2k+1) \times 60t + \varphi_{2k+1}) + \dots]$	$-\pi \leq \varphi \leq \pi, THD \geq 5\%$ $0.015 \leq \alpha_{2k} \leq 0.03$ $0.03 \leq \alpha_{2k+1} \leq 0.06$ $-\pi \leq \varphi_{2k} \leq \pi$ $-\pi \leq \varphi_{2k+1} \leq \pi$ $k = 1, 2, \dots, 10$ $0.1 \leq v_n \leq 0.9$ $T_s = T, 2T, \dots, 8T$ $T_e - T_s = T, 2T, \dots, 9T$
C8	Sag with harmonics	$v(t) = \sqrt{2}v_n [\sin(2\pi \times 60t + \varphi) + \dots + \alpha_{2k} \sin(2\pi(2k) \times 60t + \varphi_{2k}) + \alpha_{2k+1} \sin(2\pi(2k+1) \times 60t + \varphi_{2k+1}) + \dots [u(t - T_s) - u(t - T_e)]]$	$-\pi \leq \varphi \leq \pi, THD \geq 5\%$ $0.015 \leq \alpha_{2k} \leq 0.03$ $0.03 \leq \alpha_{2k+1} \leq 0.06$ $-\pi \leq \varphi_{2k} \leq \pi$ $-\pi \leq \varphi_{2k+1} \leq \pi$ $k = 1, 2, \dots, 10$ $1.1 \leq v_n \leq 1.8$ $T_s = T, 2T, \dots, 8T$ $T_e - T_s = T, 2T, \dots, 9T$
C9	Swell with harmonics	$v(t) = \sqrt{2}v_n [\sin(2\pi \times 60t + \varphi) + \dots + \alpha_{2k} \sin(2\pi(2k) \times 60t + \varphi_{2k}) + \alpha_{2k+1} \sin(2\pi(2k+1) \times 60t + \varphi_{2k+1}) + \dots [u(t - T_s) - u(t - T_e)]]$	$-\pi \leq \varphi \leq \pi, THD \geq 5\%$ $0.015 \leq \alpha_{2k} \leq 0.03$ $0.03 \leq \alpha_{2k+1} \leq 0.06$ $-\pi \leq \varphi_{2k} \leq \pi$ $-\pi \leq \varphi_{2k+1} \leq \pi$ $k = 1, 2, \dots, 10$ $0 \leq v_n \leq 0.1$ $T_s = T, 2T, \dots, 8T$ $T_e - T_s = T, 2T, \dots, 9T$
C10	Interruption with harmonics	$v(t) = \sqrt{2}v_n [\sin(2\pi \times 60t + \varphi) + \dots + \alpha_{2k} \sin(2\pi(2k) \times 60t + \varphi_{2k}) + \alpha_{2k+1} \sin(2\pi(2k+1) \times 60t + \varphi_{2k+1}) + \dots [u(t - T_s) - u(t - T_e)]]$	$-\pi \leq \varphi \leq \pi, THD \geq 5\%$ $0.015 \leq \alpha_{2k} \leq 0.03$

			$0.03 \leq \alpha_{2k+1} \leq 0.06$ $-\pi \leq \varphi_{2k} \leq \pi$ $-\pi \leq \varphi_{2k+1} \leq \pi$ $k = 1, 2, \dots, 10$ $0.5 \leq v_{nf} \leq 0.1$ $8 \leq f_n \leq 25$ $0.1 \leq v_{ns} \leq 0.9$ $-\frac{\pi}{4} \leq \varphi \leq \frac{\pi}{4}$ $T_s = T, 2T, \dots, 8T$ $T_e - T_s = T, 2T, \dots, 9T$ $0.5 \leq v_{nf} \leq 0.1$ $8 \leq f_n \leq 25$
C11	Flicker with sag	$v(t) = \sqrt{2}v_n \sin(2\pi \times 60t + \varphi) [v_{nf} \sin(2\pi f_n t)]$ $+ \sqrt{2}v_{ns} \sin(2\pi \times 60t + \varphi) [u(t - T_s) - u(t - T_e)]$	
C12	Flicker with Swell	$v(t) = \sqrt{2}v_n \sin(2\pi \times 60t + \varphi) [v_{nf} \sin(2\pi f_n t)]$ $+ \sqrt{2}v_{ns} \sin(2\pi \times 60t + \varphi) [u(t - T_s) - u(t - T_e)]$	$1.1 \leq v_{ns} \leq 1.8$ $-\frac{\pi}{4} \leq \varphi \leq \frac{\pi}{4}$ $T_s = T, 2T, \dots, 8T$ $T_e - T_s = T, 2T, \dots, 9T$
C13	Impulsive Transient	$v(t) = 1.11v_p (e^{-7.5 \times 10^2(t-T_s)} - e^{-3.44 \times 10^2(t-T_s)}) [u(t - T_s) - u(t - T_e)]$	$0.2 \leq v_p \leq 1$ $T \leq T_s \leq 7T$ $T_e = T_s + 1 \text{ ms}$
C14	Notch	$v(t) = \begin{cases} \sqrt{2} \left[ \sin\left(\omega t + \frac{\pi}{6}\right) + 0.5\rho \sin\left(\omega t + \frac{5\pi}{6}\right) \right] & k = 17 \\ \sqrt{2} \left[ \sin\left(\omega t + \frac{\pi}{6}\right) - 0.5\rho \cos(\omega t) \right] & k = 3.9 \\ \sqrt{2} \left[ \sin\left(\omega t + \frac{\pi}{6}\right) - \rho \sin\left(\omega t + \frac{\pi}{6}\right) \right] & k = 5, 11 \\ \sqrt{2} \left[ \sin\left(\omega t + \frac{\pi}{6}\right) \right] & \text{otherwise} \end{cases}$ $v(t) = \sin(\omega t) + \text{sign}(\sin(\omega t))$	$\omega = 2\pi \times 60, \rho = 0.5$ $0 \leq \alpha \leq \frac{\pi}{2}, \mu = \alpha$ $\mu = \frac{36500 \times 2\pi / 20,000}{\sqrt{2} \times 460\rho \sin(\alpha)}$ $\omega t \in [\alpha + \frac{k\pi}{6}, \alpha + \frac{k\pi}{6} + \mu]$
C15	Spike	$\times \left\{ \sum_{n=0}^9 k [u(t - (t_1 - 0.02n)) - u(t - (t_2 - 0.002n))] \right\}$	$0 \leq t_1 t_2 \leq 0.5T$ $0.01 \leq t_2 - t_1 \leq 0.05T$ $0.1 \leq k \leq 0.4$
C16	Flicker with harmonics	$v(t) = \sqrt{2} \sin(2\pi \times 60t + \varphi) [1 + v_n \sin(2\pi f_n t)]$ $+ \sqrt{2} [\sin(2\pi \times 60t + \varphi) + \dots$ $+ \alpha_{2k} \sin(2\pi(2k) \times 60t + \varphi_{2k})$ $+ \alpha_{2k+1} \sin(2\pi(2k+1) \times 60t + \varphi_{2k+1})$ $+ \dots]$	$0.5 \leq v_n \leq 0.1$ $8 \leq f_n \leq 25$ $-\pi \leq \varphi \leq \pi, THD \geq 5\%$ $0.015 \leq \alpha_{2k} \leq 0.03$ $0.03 \leq \alpha_{2k+1} \leq 0.06$ $-\pi \leq \varphi_{2k} \leq \pi$ $-\pi \leq \varphi_{2k+1} \leq \pi$ $k = 1, 2, \dots, 10$

All PQD signals were generated with 60 Hz frequency.

## 2.2 Deep Convolutional Neural Networks

Deep learning is a class of algorithms that uses various processing units organized in layers to learn features that represent unstructured data. In general, a deep learning algorithm has two main components:

a) a mapping function  $\mathcal{F}$  of the input space for real numbers,  $\mathcal{F} = \{f: \mathcal{X} \rightarrow \mathbb{R}\}$  and b) a classifier

$C = (F, w, b)$  that receives a set of features  $F \subseteq \mathcal{F}$  and provides a label  $y$ . In this study,  $y$  is a number between 1 and 16, representing one type of disturbance, as illustrated on Table 1.

The promise of deep learning is to discover a rich set of features  $F$ , by stacking a sequence of convolutional filters. Training the classifier  $C$  is performed by minimizing a loss function by considering the leaned features  $F$  (via empirical risk minimization) [30, 31].

For the presented work, considering the spatio-temporal characteristic of the environment, a 2D DCNN was used [28].

The 2D DCNN data processing occurs in the kernel filters [28]. These filters capture and extract the signal features into a larger window by means of convolution. The convolution process reduces the network's number of parameters by use of the sliding window and parameter sharing features. After the convolution layer, a pooling layer is usually used and its function is to maintain the essential features of the image while reducing the image dimension. The pooling types which are commonly used are maximum pooling, average pooling and global pooling [28]. The choice of one of the polling types is made according to the kind of information an image brings. For images where a higher resolution is needed for a better feature extraction, maximum pooling layer is the adequate one to be used. An important function that is often used after the pooling layer, is the rectified linear unit (ReLU) activation function. The ReLU function adds non-linearity to the network, thus contributing to the learning process. The last, but not less important, model layer is the fully connected layer (FCL), which is the layer responsible for the classification. When dealing with multi-class output, a final activation function called softmax is used to returns the output probability distribution. Accordingly, the architecture of the 2D DCNN thus becomes smaller, which results in higher processing time at classification. The output of this architecture can be calculated by (1) and (2) [28]:

$$f(l) = pool_{n \times n}(\sigma(k \otimes [D_{in}] + b)) \quad (1)$$

$$f_j(x) = \frac{e^{x_j}}{\sum_{l=1}^k e^{x_l}}, j = 1, 2, \dots, k \quad (2)$$

where, in (1),  $f(l)$  indicates the output of feature extraction part of CNN,  $pool_{n \times n}$  represents max-pooling layer,  $n$  is the pooling window size,  $\sigma$  represents ReLU function,  $k$  indicate convolution kernels,  $D_{in}$ , is the input of the layer (an image for 2D DCNN) and  $b$  is the bias.  $f_j(x)$  represents the output probability and  $x$  represents non-normalized parameters.

The 2D convolution layer can be calculate by (3) [27]:

$$X_i^l = f(\sum_j X_j^{l-1} \otimes k_{jj}^l + b_i^l) \quad (3)$$

where,  $X_i^l$  is 2D matrix output,  $\otimes$  represents 2D convolution operator,  $k_{jj}^l$  is a convolutional kernel with  $j$  by  $j$  dimensions,  $b_i^l$  denotes bias, and  $f$  represents ReLu activation function. Fig. 1 illustrates the idea of a 2D DCNN.

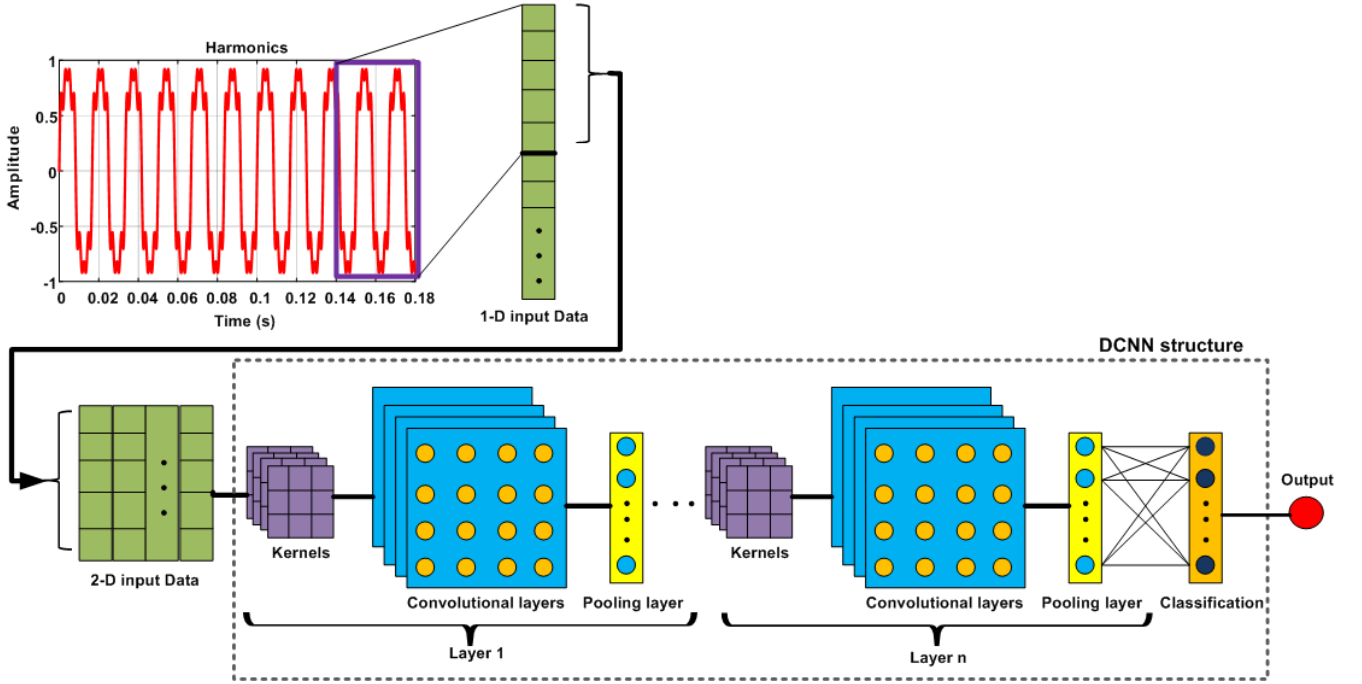


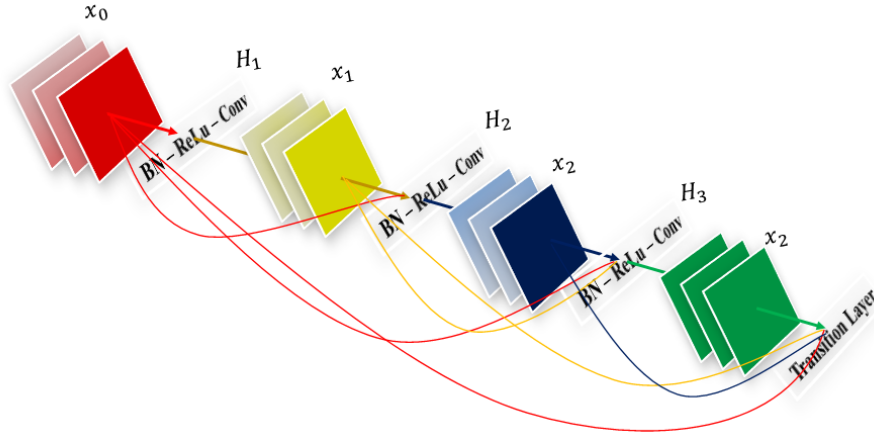
Fig. 1. 2D Convolutional Network structure.

As illustrated in Fig. 1, the original signal is grouped into layers and convolution is performed with a larger data window, which results in a lower number of filters per layer, producing a lower training and classification time. The improvement for real-life applications is two-fold: a) more significant set of features can be continuously learned and b) less computational resources are needed to deploy the model in real-time.

In this work, a 2D Densely Connected Convolutional Neural Network is used to improve the accuracy of PQD classification problem. The proposed solution is compared with three works[26, 27, 28] that are most up-to-date and high performance methods in the literature.

### 2.2.1 Densely Connected Convolutional Network

Firstly introduced by [29], DenseNet is a convolutional network architecture created to ensure maximum information flow between layers in the network by connecting all layers with matching feature-map sizes, directly to each other. In order to preserve the feed-forward nature, each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers. Fig. 2 illustrates this layout schematically.



**Fig. 2.** A 4-layer dense block with growth rate = 3.

To achieve this layout, concatenation is used. Therefore, each layer receives a “collective knowledge” from all preceding layers. This fact allowed the network to be thinner and compact, i.e. the number of channels can be fewer. The resulting of this is that it has higher computational and memory efficiency. Because of its dense connectivity pattern, the authors in [29] refers to their approach as Dense Convolutional Network.

The advantages of DenseNet are [29]:

- The error signal can be easily propagated to previous layers more directly. This can be seen as a implicit deep supervision, as previous layers can get direct supervision from the final classification layer, resulting in a strong gradient flow;
- It has more diversified features and parameters, once each layer in DenseNet receives all preceding layers outputs as inputs;
- Once DenseNet uses features of all complexity levels, it tends to give more smooth decision boundaries, maintaining low complexity features.

Traditional convolutional feed-forward networks connect the output  $\ell^{th}$  layer input to the  $(\ell + 1)^{th}$  layer [29], resulting the following layer transition (4):

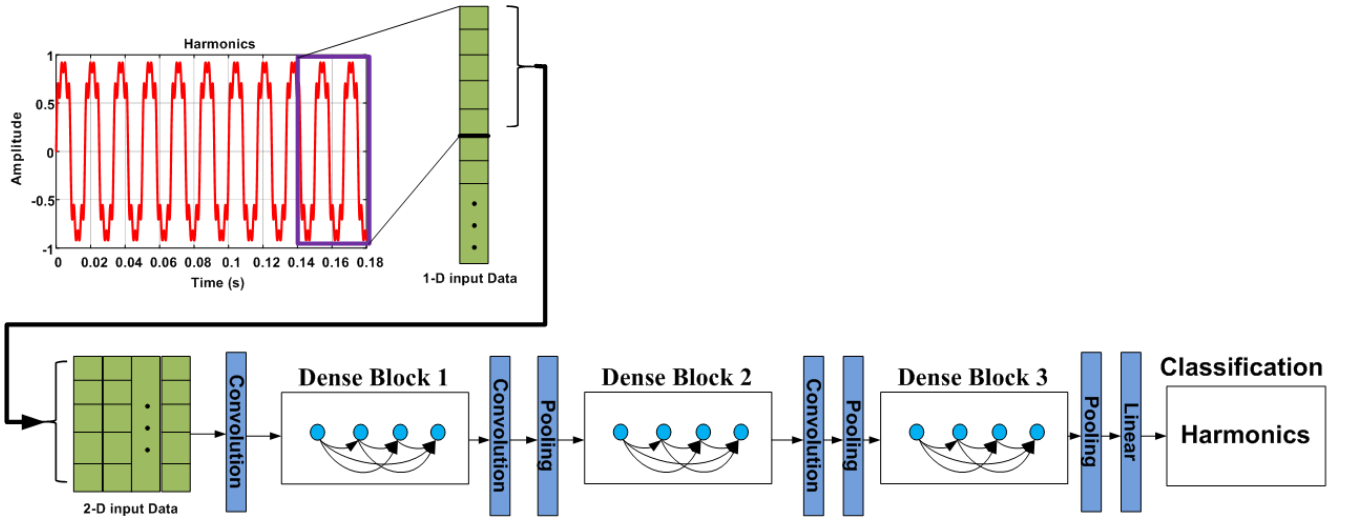
$$X_\ell = H_\ell(X_{\ell-1}) + X_{\ell-1} \quad (4)$$

A DenseNet, however, has a different connectivity pattern. Consequently, the direct connections from any layer to all subsequent layers gives that, the  $\ell^{th}$  layers receives the feature-maps of all preceding layers,  $X_0, \dots, X_{\ell-1}$ , as input [29] (5):

$$X_\ell = H_\ell([X_0, X_1, \dots, X_{\ell-1}]) \quad (5)$$

where  $[X_0, X_1, \dots, X_{\ell-1}]$  refers to the concatenation of the feature-maps produced in layers  $0, \dots, \ell - 1$ . In (5),  $H_\ell(\cdot)$  is defined as a composite function of three consecutive operations: batch normalization (BN), followed by a rectified linear unit (ReLU) and a  $3 \times 3$  convolution (Conv) [29].

The needed concatenation used to build the architecture is not viable when the size of feature-maps changes in time. To facilitate down-sampling the presented network is divided into multiple densely connected dense blocks, as illustrated in Fig. 3.



**Fig. 3.** A deep 2D DenseNet with three dense blocks.

The layers between the blocks are known as transition layers, which do convolution ( $1 \times 1$ ) and pooling ( $2 \times 2$ ). A growth rate can be used to increase the number of channels in each input layer [29], however, in the architecture used in this work it was not used. This function was not used because it would increase the number of feature-maps at the inputs of each layer and the available computer memory could not be enough.

Finally, one can improve the model compactness by reducing the number of feature-maps at transition layers. If a dense block contains  $m$  feature-map, one can let the following transition layer generate  $[\theta m]$  output feature maps, where  $0 < \theta \leq 1$  is referred to as the compression factor. If  $\theta = 1$ , the number of feature-maps across transition layers remain unchanged compared to the pre-set architecture [29]. Once the growth rate was not used in this work, the compression factor was not necessary to be used.

The tested 2D DenseNet structure is presented in Table 2.

**Table 2.**

2D DenseNet architecture for PQD classification. Note that each “conv” layer shown in the table corresponds the sequence BN-ReLu-Conv.

Layers	Output Size	2D DenseNet
Convolution	12 x 38 x 32	3 x 3 conv, stride 2

Dense Block (1)	12 x 38 x 128	[3 x 3 conv] x 3
Transition Layer (1)	12 x 38 x 12	1 x 1 conv
	6 x 19 x 12	2 x 2 average pool, stride 2
Dense Block (2)	6 x 19 x 108	[3 x 3 conv] x 3
Transition Layer (2)	6 x 19 x 6	1 x 1 conv
	3 x 9 x 6	2 x 2 average pool, stride 2
Dense Block (3)	3 x 9 x 198	[3 x 3 conv] x 3
Transition Layer (3)	3 x 9 x 3	1 x 1 conv
	1 x 4 x 3	2 x 2 average pool, stride 2
Classification Layer	3	3 x 3 Global average pool
	16	Fully-connected, softmax

In this work, a 2D DenseNet structure with 3 dense blocks on 14 x 40 input images is used. The initial convolution layer comprises 768k convolutions of size 3x3 with stride 2. For training, sgd algorithm was used. The learning rate is used as 0.002. The minibatch size is selected as 256. The number of epochs chosen was 200 with patience of 50 epochs.

### 2.3 Implementation of the framework

In order that the 2D DenseNet is to be tested on a real system, an embedded system was developed that is constituted of an electronic circuit for acquiring the signal in real-time, along with a mini-microcomputer for performing the classification by means of the 2D DenseNet.

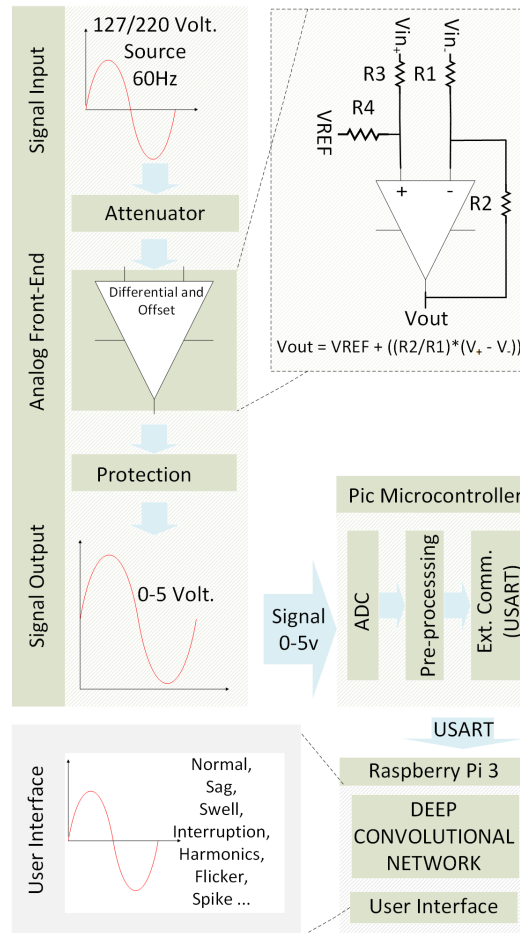
In this stage an Analog Front-End circuit was developed. The circuit is constituted of resistors for performing signal attenuation, an operational amplifier in subtractor mode for performing signal offset and an amplifier subtractor for the zero crossing signal.

Zero crossing was implemented to allow for the identification of the moment at which the AC electric network voltage crosses the line of zero volts. This enables easy synchronism for performing data acquisition.

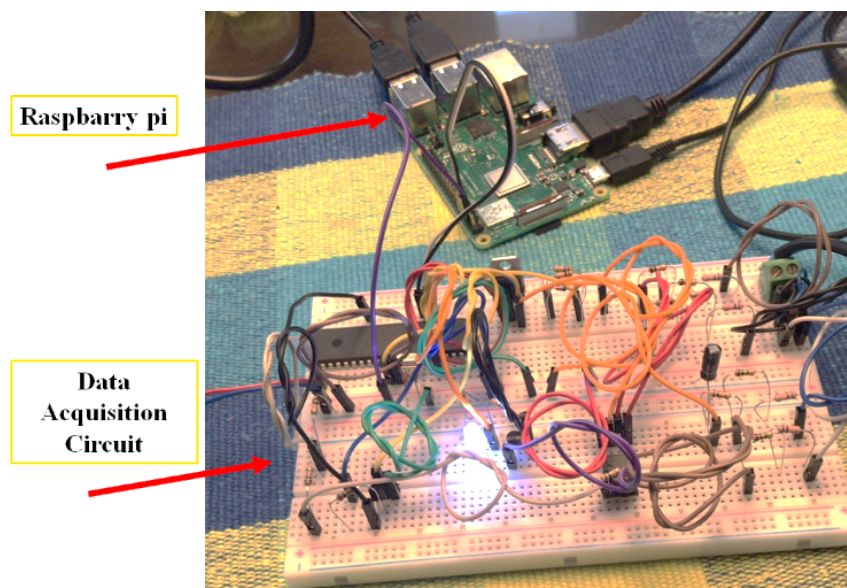
The already attenuated signal coupled with offset was attributed to the ADC channel (Analog to Digital Converter), and the zero crossing signal attributed to an external pin on the microcontroller (PIC 18F4550-1/P). On the microcontroller, external interruption was used per rising edge, and through such, it is possible to capture zero pass-through.

A sampling frequency of 3.36 kHz was used in this work, i.e., 56 samples per cycle. This frequency was chosen, as it is the same as that used for generating the training data for the DCNN.

With the signal sampled in the microcontroller, it was then sent to the mini-microcomputer (Raspberry Pi 3) via UART (Universal Asynchronous Receiver/Transmitter). Fig. 4 presents the block diagram of the complete system. The developed system is shown in Fig. 5.



**Fig. 4.** Block diagram of the embedded system.



**Fig. 5.** The developed real system.

## 2.4 Metrics for the PQDs classificatio

In order to carry out the evaluation of the presented strategy, in this work, metrics used are Recall (6), Precision (7), Accuracy (8) and F1-Score (9) [32]:

$$Recall(R) = \frac{TP}{TP+FN} \quad (6)$$

$$Precision(P) = \frac{TP}{TP+FP} \quad (7)$$

$$Accuracy(ACC.) = \frac{TP+FN}{P+N} \quad (8)$$

$$F1 - Score = 2 \times \frac{P \times R}{P+R} \quad (9)$$

Where, Recall is the rate of true positives or TP sensitivity (correctly predicted that the equipment was switched on), FP are the false positives (predicted device switched on, but was switched off), FN are the false negatives (device switched on, but predicted as switched off). Precision refers to the predicted positive values. Accuracy is the proportion of real results in all cases. F1-Score is the harmonic mean between Precision and Recall.

However, when one has a multiclass configuration on the classifier, it becomes necessary to calculate the average for the F1 – score, for Precision and for Recall. For such, one can use the following metrics:

- Weighted;
- Micro;
- Macro.

The examples used herein for the F1 - Score are applied to the remaining metrics.

The first of these ‘weighted’ calculates the F1 – score for each class in an independent fashion, but when it sums these together, it uses a weight that depends on the number of real labels from each class. As such, it favors the majority class (10):

$$F1 - Score_{weighted} = (F1_{class1} \times W_1 + F1_{class2} \times W_2 + \dots + F1_{classN} \times W_N) / N_{samples} \quad (10)$$

Where,

$N_{samples}$  – is the number of samples.

Here 'micro' uses the global number of TP, FN, and FP and calculates F1 directly, which results in a higher penalty when its model does not have a good performance in the minority classes (11):

$$F1 - Score_{micro} = F1_{class1+class2+class3} \quad (11)$$

Finally, 'macro' calculates F1 separated by class, but not using aggregation weights, which results in a higher penalty when its model does not have a good performance in the minority classes (12):

$$F1 - Score_{macro} = (F1_{class1} + F1_{class2} + \dots + F1_{classN})/N \quad (12)$$

Where,

N – is the number of classes.

In this way, when one has F1 – micro, Precision - micro and Recall – micro, and since Precision = Recall in the case of micro – average, these too are equal to their harmonic average, i.e., for F1 – micro (13):

$$F1 - micro = Precision - micro = Recall - micro \quad (13)$$

Finally, (9) is also the general accuracy of the classifier, i.e., it is the proportion of examples that are classified correctly from among all the examples analyzed. As such, we have (14):

$$Accuracy = F1 - micro = Precision - Micro = Recall - micro \quad (14)$$

### 3. Case study

#### 3.1 Preprocessing of Data

A 768,000 PQDs data set was used for 2D DenseNet training, from those 10% of the data were random selected and sent to validation. A 1,000 samples are generated for each disturbance as a test set, to test the performance of trained models.. Different to other studies, the option was made here to train the DenseNet with all the PQDs all at once instead of separately. Gaussian white noise of 20 dB is added to the synthetic PQDs data. Table 3 summarizes the data structures used.

**Table 3**

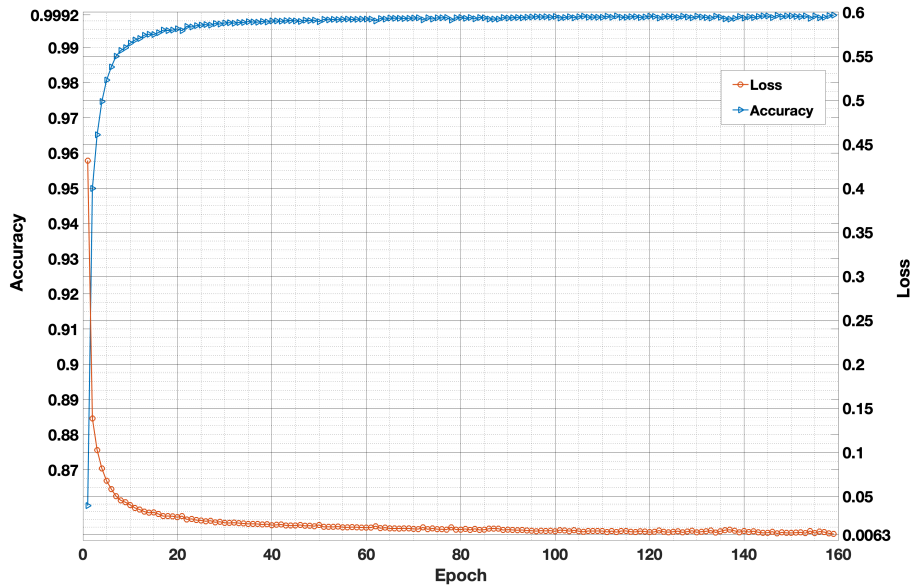
Data Used

	2D DenseNet	Noise grade
Train set	768,000	Pure to 50 dB
Validation set	76,800	Pure to 50 dB
Test set	16,000	Pure to 50 dB

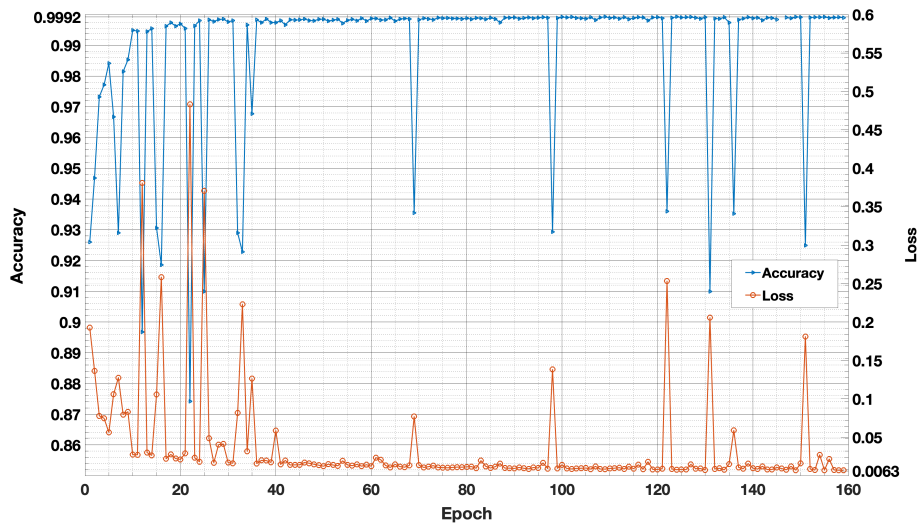
A microcomputer was used with a processor Intel ®CoreTMi7, 3.40 GHz X 4, 16GB of RAM and a video card ®NVIDIA GEFORCE GTX1080 8G GPU with a CUDA processor.

### 3.2 Performance and Comparative Analysis

The accuracy and loss values for training and validation are shown in Fig. 6 and Fig. 7, respectively.



**Fig. 6.** Accuracy and loss values during training



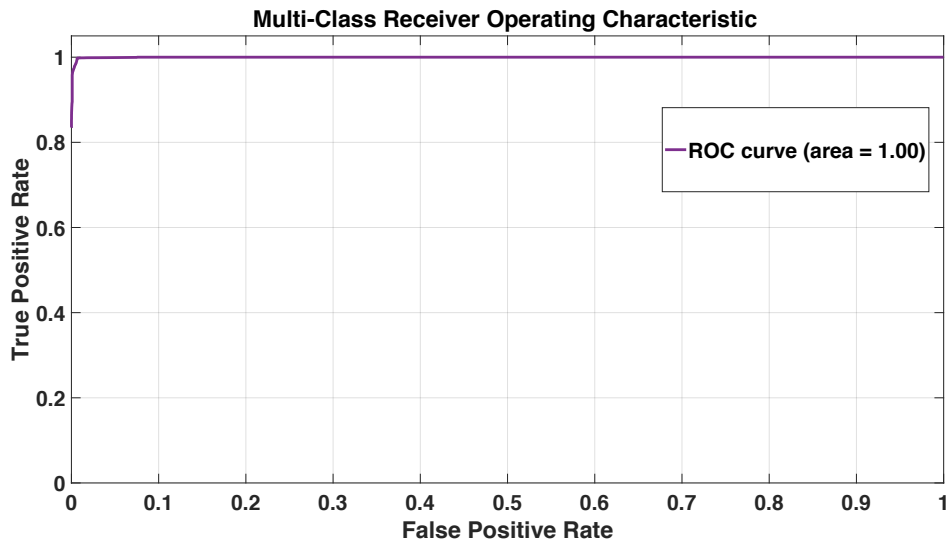
**Fig. 7.** Accuracy and loss values during validation.

At each epoch, the accuracy of the models in the validation set was detected when it was improved. As such, the model with the best performance in the training and validation set is selected as the final model. The final model and its properties are shown on Table 4.

**Table 4****Final Model 1**

Type of ANN	Best loss in validation set	Best accuracy in validation set	Elapsed epoch	Consuming time for each epoch	Training time	Number of parameters	Model size
2D DenseNet	0.0094	0.9992	145	699s	1,689 min	231,530	1.1 MB

As seen from Table 4, the accuracy and validation of 2D DenseNet was 99.92%, with a training time of 1,689 minutes, the best loss in validation was 0.0094 and the model size resulted 1,1 MB. A multi-class receiver operating characteristic (ROC) curve is illustrated in Fig. 8.

**Fig. 8.** Multi-class ROC curve.

According to [29], ROC curves can be used to analyse classifiers performances. Further, the AUC (area under curve) shows the average performance value of a classifier and can be used to compare different classifiers. As bigger is the AUC, better is a classifier performance.

As one can see in Fig. 8, the ROC curve for the 2D DenseNet, showed that almost all positives and negatives examples were classified correctly. The classifier performance can be analyzed by its AUC as well. Fig. 8 shows that the 2D DenseNet AUC curve area is equal to 1.00, resulting that the 2D DenseNet performed well for the case presented herein.

An important property of the presented solution is that the original model is reused as the starting point for update both features and classifier. Note that the most important motivation for representation learning is to learn features that can transfer well to different downstream tasks. This continuously strategy takes into account new signals captured in real time.

Once one has the trained model in hand, it is possible to test it and compare with other DNNs approaches. The detailed performance of the methods for each type of disturbance is shown in Tables 5-6. In order to fully appreciate the results, a Discrete Wavelet Transform (DWT) + Neural Network had been trained over the same sample and the results are shown in Table 5. The Neural Network architecture is shown in Table 7.

**Table 5**

The detailed performances of 2D DenseNet and DWT + NN

Disturbances number	Types of disturbances	Performance of DCNN	
		2D DenseNet	Performance of well-known method DWT + NN
1	Pure 60 Hz	100	90.30
2	Sag	100	90.40
3	Swell	100	98.50
4	Interruption	99.80	96.10
5	Flicker	100	98.60
6	Oscillatory Transient	100	29.30
7	Harmonic	100	99.80
8	Sag with harmonics	99.90	83.90
9	Swell with harmonics	100	97.60
10	Interruption with harmonics	100	82.60
11	Flicker with sag	100	80.30
12	Flicker with Swell	99.80	95.10
13	Impulsive Transient	100	99.50
14	Notch	100	92.30
15	Spike	100	98.70
16	Flicker with harmonics	99.40	00.00
-	<b>AVE</b>	<b>99.93</b>	<b>83.31</b>

**Table 6**

Evaluation metrics

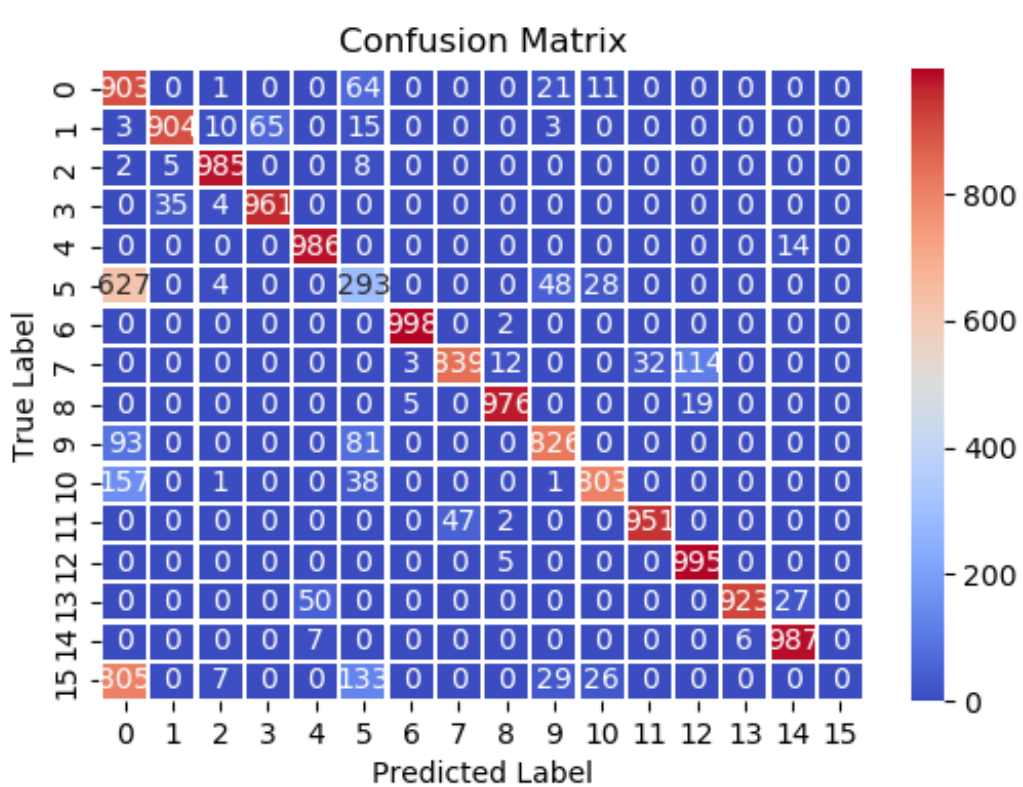
Evaluation metrics	Performances of DCNN
	2D DenseNet
Precision	99.93%
Recall	99.93%
F1-score	99.93%

**Table 7**

**DWT + NN architecture features**

Network Name	Number of layers	Type
DWT + NN	Number of layers: 2; Neurons per layer – unit 1: 30 (Activation function: ReLu); unit 2: 20 (Activation function: ReLu); Dropout rate: 20% for all layers; Classification layer: Activation function Softmax.	Multilayer Perceptron

Fig. 9 shows the DWT + NN technique confusion matrix.



**Fig. 9. DWT + NN confusion matrix.**

As one can see in Table 5, the 2 D DenseNet besides the fact that does not need a prior features extraction technique, performed way better than the DWT + NN technique.

It is not always easy to make comparisons between machine learning models. As reasons for such, one can cite the data quantities used for training, validation and tests; the data shape of input data; the way data is generated for build a dataset; how noise grade was added to the dataset, among others. However, despite all the dataset inconsistencies that one can find within related published research, an attempt is made here to compare the presented model with other state-of-art solutions.

Accordingly, some studies in the literature use mixed noise, while some studies use a single noise level. Average accuracy values are used in studies that conduct experiments according to multiple noise levels.

Table 8 shows a comparison between other state-of-art solutions and the 2D Densenet, while Table 9 shows their architecture features.

**Table 8**

Comparison with state-of-art

Reference	Number of PQDs	Accuracy
[26]	16	0.9942
[27]	15	0.9986
[28]	13	0.9997
2D DenseNet	16	0.9993

**Table 9**

State-of-art solutions and 2D DenseNet architecture features

Network name and reference	Number of layers (not including BN and Dropout layers)	Type	Data set pre-processing technique	Model size
Deep 1D CNN [26]	Conventional 1D layer: 6; Layer depth – 3; Total filters: 448; Activation function: ReLu	DCNN	None	1.97 MB
CS + Deep 1D CNN [27]	Conventional 1D layer - 6; Layer depth – 3; Total filters: 224; Activation function: Softmax	DCNN	Compressed Sensing	Not available
Deep Hybrid (1D and 2D) CNN [28]	Conventional 1D layer – 3; Layer depth – 3 Activation function: Softmax GoogleNet Conventional 2D layer: 110; Layer depth – 11 Activation function: Softmax	DCNN	None	Not available GoogleNet: 47.58 MB
2D DenseNet	Conventional 2D layer – 9; Layer depth – 3; Total filters: 384; Activation function: Softmax	Densely Connected DCNN	None	1.1 MB

When reading the performance in Table 8 and architecture features in Table 9, one must investigate the results deeper. In first place, the only up-to-date relevant research paper that taking the same number of PQDs signals is [26]. When comparing [26] results with 2D DenseNet results, one can see that the presented model perform better.

Another relevant point when comparing the proposed model with [26] is the model size. As one can see in Table 4, the 2D DenseNet size is 1.1 MB while [26] model size is 1.97 MB. The model size

reduction is only possible due the reduced number of filters found in 2D DenseNet. As one can see, the proposed model herein has 384 filters while the model proposed by [26] has a total of 448 filters. This fact allows, the proposed model herein, to achieve a faster processing time what makes the model more suitable for embedded devices, turning those devices more efficiently, as one can see on section 3.4. To make a more accurate comparison regarding the models processing times, [26] model was trained and built. Subsection 3.4 shows the comparisons mentioned.

Another comparison that can be done is with [27] proposal. While the approach proposed herein takes 16 PQDs into account in the used dataset, [27] stroke out the Flicker with harmonics PQD. By analyzing Table 5 and 6, one can see that this is the hardest compound PQD for the model to classify. In this sense, by only not considering this PQD, [27] improved considerably their model performance. Another point that can be highlighted in [27] work, is the use of a compressed sensing method applied to the signals before introduce them to the 1D deep learning network. That said, there is a dataset pre-processing, which detracts from the main advantage of using 1D and 2D networks, which is its power in extracting image features.

The same arguments used in the last comparison can be used when comparing [28] model with the 2D DenseNet model. Moreover, [28] proposed a hybrid 1D – 2D deep learning network approach, using a 2D CNN pre-trained model known as GoogleNet. The presented model concatenate the outputs of one 1D model with a 2D model and used it as the input of a final fully connected layer. The complexity of [28] deep learning network structure needs a much higher computational effort, compared to 2D DenseNet proposed herein, resulting a model with a much higher size and a much longer processing time.

A confusion matrix, Fig. 10, provides information on the classification performance of single and composite signals.

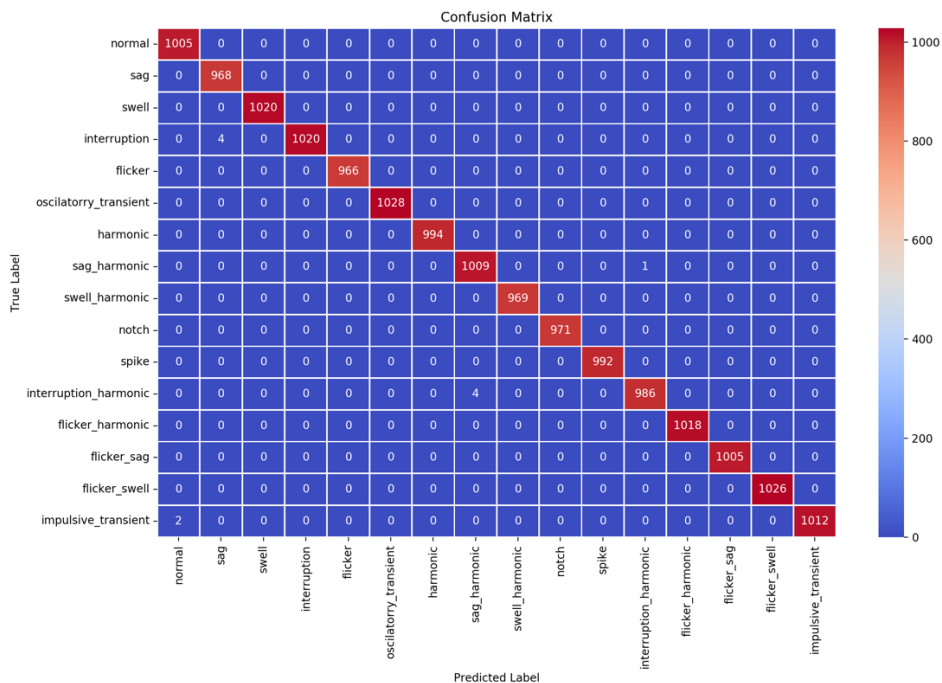
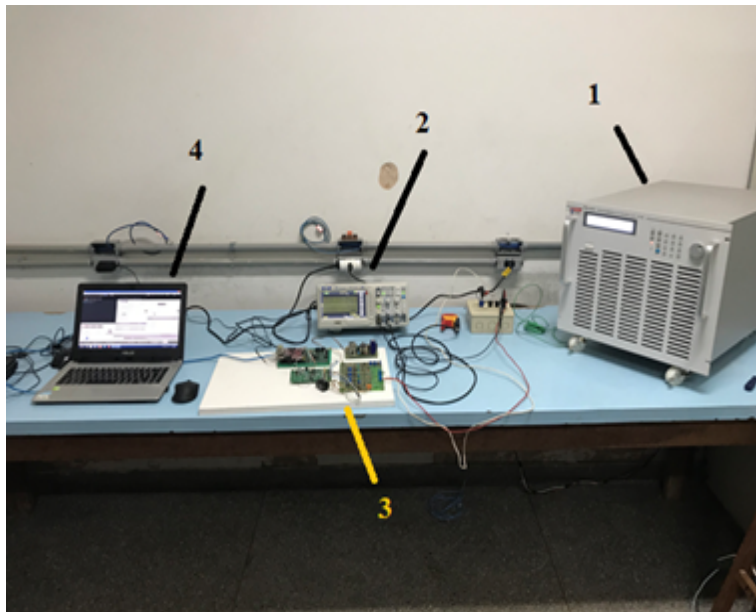


Fig. 10. Confusion matrix.

### 3.3 Laboratory Measurements

All the signals used for training the 2D DenseNet were generated by the software MatLab R2017a, by means of mathematical equations described in [8]. However, in order that the performance of the 2D DenseNet was truly tested and validated, PQDs were generated by means of a PQD generating source, and later classified by the trained network.

To this end, the acquisition of these PQDs was performed in the laboratory, as illustrated in Fig. 11.



**Fig. 11.** The collection scheme for PQDs in the laboratory.

The equipment used for data collection is described in the following, in accordance with Fig. 11:

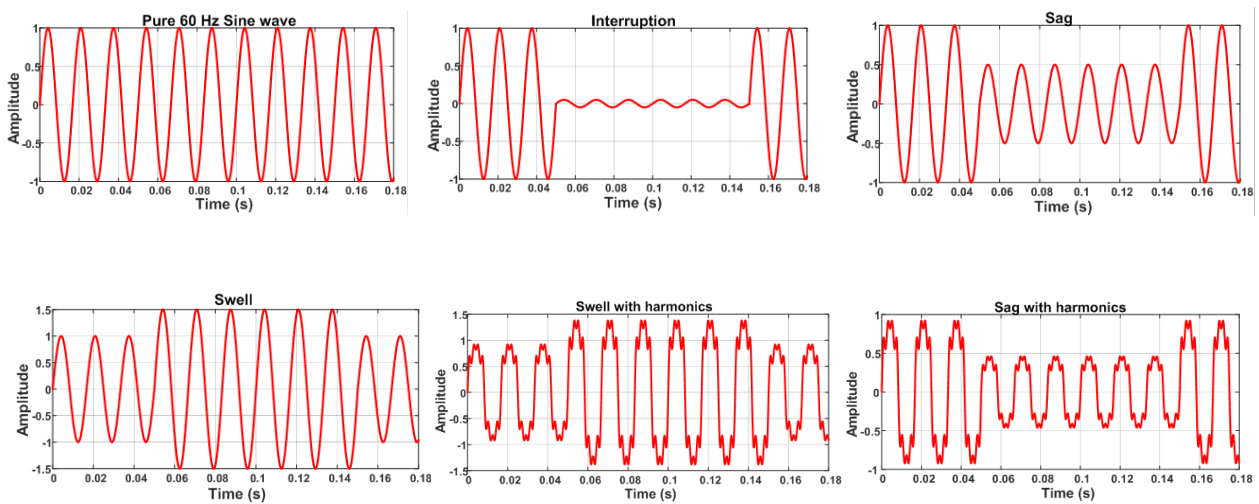
1 – The Chroma programmable supply, model 61702. This programmable supply generates pure voltage signals. Among its functions, emphasis is placed on the generation of more than 30 types of harmonic distortion, interruption, sag and swell. In order to test the trained convolutional neural network, the signals were generated at a frequency of 60 Hz;

2 – An oscilloscope from MIT 1052DL. This oscilloscope was used to monitor the PQD signals generated by the above-mentioned programmable supply.

3 and 4 – A data acquisition system. System on Chip (SoC) DM3730 supplied by Texas Instruments and composed of a general digital processing core with an ARM Cortex-A8 de 1 GHz architecture and a digital processing core for signals from the C64+ of 600 MHz family. The SoC counts on 512 MB of RAM memory, where the data transference rate between the two cores is performed through a shared memory region. The DSP performs the acquisition of analogical signals through an analog signal-conditioning interface connected to the DSP serial port. The signals acquired and processed by the DSP are thus transferred to the region of shared memory and later accessed by the ARM processor. In turn, the ARM core performs the storage of data received in a text file of the CSV type, at a rate of 3300

samples per second. This device was used to perform the acquisition of data from the generated signal. The sampling frequency for the acquisition was 3.3 kHz.

Sags, swells and interruptions can be caused by load switching on electrical power systems. The compound PQDs can be generated through simultaneous line faults, loads switching among others. Hence, despite other works whose PQDs deep learning approaches are done only by signals simulations, we generated six real signals in laboratory. In this way, noises generated by wire connections and by the signal generator itself, gives more real signals so one can test the proposed deep learning approach. The generation and acquisition was performed for signals with interruptions, sags, swells and pure sine. For each type of PQD, 30 samples were collected for the tests. The acquisitioned signals are shown in Fig. 12.



**Fig 12.** Real-life tested PQDs.

The detection results of the proposed method are shown in table 10.

**Table 10**

Detection result of real-life tested PQDs

PQDs type	Pure 60 Hz	Interruption	Sag	Swell	Swell with harmonics	Sag with harmonics
Test set	30	30	30	30	30	30
Correct samples	30	30	30	30	30	30

As one can see on table 10, the proposed 2D DenseNet architecture approach is suitable to detect simple and compound PQDs. Therefore, in a real-life system, disturbances can be easily speculated according the disturbance type, and the targeted measures can be taken to solve it, so as to improve the power quality of some microgrid, for example.

Since the model effectively predicted new data from results not used in training, the experiments

performed do not indicate overfitting. In addition to testing the model with computer-simulated data, we also used data generated in real-time by a programmable power supply, and the model's performance remained unchanged.

### 3.4 Comparison among Real Detection Times

Since the structure of the 2D DenseNet is smaller, its detection time in real time should be lower than the detection time of [26]. Fig. 12 illustrates the real-life tested PQDs generated with the equipment showed in Fig. 11.

For a signal without disturbance, i.e., pure, the detection time for [26] was 1.1730 seconds, while for 2D DenseNet was 1.0555 seconds. Pure signal detection is important. If the classifier has not been trained to detect and classify pure signals, it could identify it as any of the PQDs it has been trained to detect, except the pure one, leading the classification process to fail.

For an interruption, the detection time for [26] was 1.9910 seconds, while for 2D DenseNet was 1.5328 seconds. The detection time for sag with [26] model was 1.7022 seconds, while for 2D DenseNet the time was 1.5042 seconds, for a signal with sag. For a signal with swell, the detection time for [26] was 1.9833 seconds, while for 2D DenseNet the time was 1.6024 seconds. Swell with harmonics signal result was 1.7708 and 1.4998, for [26] and 2D, respectively. Finally, for a signal with sag with harmonics, the result was 1.9400 and 1.5540 for [26] and 2D, respectively. This information is summarized on Table 11.

**Table 11**

Detection Time for both DCNN

DCNN	[26]	2D DenseNet
	Time (s)	Time (s)
Pure 60 Hz	1.1730	1.0555
Interruption	1.9910	1.5328
Sag	1.7022	1.5042
Swell	1.9833	1.6024
Swell with harmonics	1.7708	1.4998
Sag with harmonics	1.9400	1.5640

## 4. Conclusion

This paper presented 2D densely connected convolutional network based approach for power quality disturbances classification.

The 2D convolutional network architecture is compared with the state-of-art solutions. The simulation test results showed that the proposed method has higher accuracy, less complex structure and no need for signal pre-processing.

A series of advantages brought by the 2D DenseNet, such as strong gradient flow, diversified features, richer parameters and more smooth decision boundaries maintaining low complexity features, can be highlighted as the characteristics that make this approach more accurate for dealing with the spatio-temporal environment characteristics of PQDs.

The presented deep network architecture performed well for mathematical built and real-life power quality disturbances signals. This shows the approach proposed herein is suitable for single and compound PQDs, what shows that it can be a useful tool for real-life applications.

A comparative analysis of the 2D DenseNet model processing time was done. The results proved that the proposed DCNN has a smaller processing time, turning this model proposal more suitable for embedded devices and allowing the increase of their efficiency.

Future works about power quality disturbances classification should focus on the model size of the proposed deep network, so it can improve its performance on embedded devices making them more efficient. The application of new deep neural networks to PQDs classification problem, as the inception neural network, is also of interest for future works.

## **Acknowledgment**

This study was realized with support in part from the National Council for Scientific and Technological Development – CNPq Project 409687/2018-9 and in part from Mato Grosso Support Foundation - FAPEMAT Project 204690/2017.

## **References**

- [1]R. Raja Singh, Yash S. M., Shubham S. C., Indragandhi V., Vijayakumar V., SARavan P., Subramaniaswaqmy V. “IoT Embedded Cloud-Based Intelligent Power Quality Monitoring System for Industrial Drive Application”. *Fut. Gen. Comp. Systems.*, vol. 112, pp. 884-898, 2020. <https://doi.org/10.1016/j.future.2020.06.032>.
- [2]Dugan, R. C.; McGranaghan, M. F.; Santoso, S.; Beaty, W. H., “Electrical Power Systems Quality,” McGraw-Hill, Second Edition, 2004.
- [3]M. Ghiasi, S. Esmailnamazi, R. Ghiasi, M. Fathi. “Role of Renewable Energy Sources in Evaluating Technical and Economic Efficiency of Power Quality”. *Technol. Econ. Of Smart Grids Sustain. Energy*, vol. 5, no. 1, pp. 1-13, 2020. <https://doi.org/10.1007/s40866-019-0073-1>

- [4] F. Zhao, S. Member, and R. Yang, "Power-Quality Disturbance Recognition Using S-Transform," *IEEE Trans. on Power Deliv.*, vol. 22, no. 2, pp. 944–950, 2007. <https://doi.org/10.1109/TPWRD.2006.881575>
- [5] M. Gargoom, S. Member, N. Ertugrul, and W. L. Soong, "Automatic Classification and Characterization of Power Quality Events," *IEEE Trans. on Power Deliv.*, vol. 23, no. 4, pp. 2417–2425, 2008. <https://doi.org/10.1109/TPWRD.2008.923998>
- [6] S. Mishra, S. Member, C. N. Bhende, B. K. Panigrahi, and S. Member, "Detection and Classification of Power Quality Disturbances Using S-Transform and Probabilistic Neural Network," *IEEE Trans. on Power Deliv.*, vol. 23, no. 1, pp. 280–287, 2008. <https://doi.org/10.1109/TPWRD.2007.911125>
- [7] S. R. Samantaray, "Decision tree-initialised fuzzy rule-based approach for power quality events classification," *IET Gener. Transm. & Dist.* vol. 23, Iss. 4, pp. 538–551, 2010. <https://doi.org/10.1049/iet-gtd.2009.0508>
- [8] R. Hooshmand and A. Enshae, "Detection and classification of single and combined power quality disturbances using fuzzy systems oriented by particle swarm optimization algorithm," *Electr. Power Syst. Res.*, vol. 80, no. 12, pp. 1552–1561, 2010. <https://doi.org/10.1016/j.epsr.2010.07.001>
- [9] S. K. Meher and A. K. Pradhan, "Fuzzy classifiers for power quality events analysis," *Electr. Power Syst. Res.*, vol. 80, pp. 71–76, 2010. <https://doi.org/10.1016/j.epsr.2009.08.014>
- [10] C. Huang, C. Lin, and C. Kuo, "Chaos Synchronization-Based Detector for Power-Quality Disturbances Classification in a Power System," *IEEE Trans. Power Deliv.*, vol. 26, no. 2, pp. 944–953, 2011. <https://doi.org/10.1109/TPWRD.2010.2090176>
- [11] C. Lee and Y. Shen, "Optimal Feature Selection for Power-Quality Disturbances Classification," *IEEE Trans. Power Deliv.*, vol. 26, no. 4, pp. 2342–2351, 2011. <https://doi.org/10.1109/TPWRD.2011.2149547>
- [12] H. Eris and Y. Demir, "Automatic classification of power quality events and disturbances using wavelet transform and support vector machines," *IET Gener., Transm., & Dist.*, vol. 6, Iss. 10, pp. 968–976, 2012. <https://doi.org/10.1049/iet-gtd.2011.0733>
- [13] B. Biswal and S. Mishra, "Power signal disturbance identification and classification using a modified frequency slice wavelet transform," *IET Gener., Transm., & Dist.*, vol. 8, Iss. 2, pp. 353–362, 2014. <https://doi.org/10.1049/iet-gtd.2013.0171>
- [14] M. Valtierra-rodriguez, S. Member, R. D. J. Romero-troncoso, S. Member, R. A. Osornio-rios, and A. Garcia-perez, "Detection and Classification of Single and Combined Power Quality Disturbances Using Neural Networks," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2473–2482, 2014. <https://doi.org/10.1109/TIE.2013.2272276>
- [15] N. Kishor, S. Member, J. P. S. Catalão, and S. Member, "Optimal Feature and Decision Tree-Based Classification of Power Quality Disturbances in Distributed Generation Systems," *IEEE Trans. Sust. Ener.* vol. 5, no. 1, pp. 200–208, 2014. <https://doi.org/10.1109/TSTE.2013.2278865>
- [16] M. S. Manikandan, S. R. Samantaray, and S. Member, "Detection and Classification of Power

- Quality Disturbances Using Sparse Signal Decomposition on Hybrid Dictionaries,” *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 27–38, 2015. <https://doi.org/10.1109/TIM.2014.2330493>
- [17] R. Kumar, B. Singh, D. T. Shahani, A. Chandra, and K. Al-haddad, “Recognition of Power-Quality Disturbances Using S-Transform-Based ANN Classifier and Rule-Based Decision Tree,” *IEEE Trans. Ind. Appl.*, vol. 51, no. 2, pp. 1249–1258, 2015. <https://doi.org/10.1109/TIA.2014.2356639>
- [18] U. Singh and S. N. Singh, “Application of fractional Fourier transform for classification of power quality disturbances,” *IET Sci. Meas. Technol.*, vol. 11, Iss. 1, pp. 67–76, 2016. <https://doi.org/10.1049/iet-smt.2016.0194>
- [19] Z. Liu, Y. Cui, and W. Li, “A Classification Method for Complex Power Quality Disturbances Using EEMD and Rank Wavelet SVM,” *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1678–1685, 2015. <https://doi.org/10.1109/TSG.2015.2397431>
- [20] J. Li, Z. Teng, Q. Tang, and J. Song, “Detection and Classification of Power Quality Disturbances Using Double Resolution,” *IEEE Trans. Instrum. Meas.*, vol. 65, no. 10, pp. 2302–2312, 2016. <https://doi.org/10.1109/TIM.2016.2578518>
- [21] S. Naderian and A. Salemnia, “Method for classification of PQ events based on discrete Gabor transform with FIR window and T2FK-based SVM and its experimental verification,” *IET Gener., Transm., & Dist.*, vol. 11, Is. 1, pp. 133–141, 2017.
- [22] U. Singh and S. N. Singh, “Detection and classification of power quality disturbances based on time – frequency-scale transform,” *IET Sci. Meas. Technol.*, vol. 11, Iss.6, pp. 802-810, 2017. <https://doi.org/10.1049/iet-smt.2016.0395>
- [23] T. Chakravorti and P. K. Dash, “Multiclass power quality events classification using variational mode decomposition with fast reduced kernel extreme learning machine-based feature selection,” *IET Sci. Meas. Technol.*, vol. 12, Iss.1 pp. 106–117, 2018. <https://doi.org/10.1049/iet-smt.2017.0123>
- [24] Bagheri, I. Y. H. Gu, S. Member, and M. H. J. Bollen, “A Robust Transform-Domain Deep Convolutional Network for Voltage Dip Classification,” *IEEE Trans. Power Deliv.*, vol. 33, no. 6, pp. 2794–2802, 2018. <https://doi.org/10.1109/TPWRD.2018.2854677>
- [25] M. Sahani and P. K. Dash, “Automatic Power Quality Events Recognition Based on Hilbert Huang Transform and,” *IEEE Trans. Ind. Informatics*, vol. 14, no. 9, pp. 3849–3858, 2018. <https://doi.org/10.1109/TII.2018.2803042>
- [26] S. Wang and H. Chen, “A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network,” *Appl. Energy*, vol. 235, pp. 1126–1140, 2019. <https://doi.org/10.1016/j.apenergy.2018.09.160>
- [27] Wang, J., Xu, Z., & Che, Y. (2019). Power Quality Disturbance Classification Based on Compressed Sensing and Deep Convolution Neural Networks. *IEEE Access*, 7, 78336–78346. <https://doi.org/10.1109/ACCESS.2019.2922367>

- [28] Sindi, H., Nour, M., Rawa, M., Öztürk, Ş., & Polat, K. (2021). A novel hybrid deep learning approach including combination of 1D power signals and 2D signal images for power quality disturbance classification. *Expert Systems with Applications*, 174. <https://doi.org/10.1016/j.eswa.2021.114785>
- [29] Huang, G., & Weinberger, K. Q. (n.d.). *Densely Connected Convolutional Networks*. arXiv.org. 2018.
- [30] T. Hastie, R. Tibshirani, J. Friedman. "The Elements of Statistical Learning". Springer-Verlag, 2009. ISBN 978-0-387-84857-0.
- [31] I. Goodfellow, Y. Bengio, Aaron Courville. "Deep Learning". The MIT Press, 2006.
- [32] Liu, H., Wu, H., & Yu, C. (2019). A hybrid model for appliance classification based on time series features. *Energy and Buildings*, 196, 112–123. <https://doi.org/10.1016/j.enbuild.2019.05.028>

Raul V. A. Monteiro<sup>a</sup>, Raoni F. S. Teixeira<sup>b</sup>, Arturo S. Bretas<sup>c</sup>,

<sup>a</sup> Electrical Power Systems Operation and Smart Grid Research Lab., Federal University of Mato Grosso, MT, Brazil

<sup>b</sup> Group for Innovation on Real Information System, Computing Engineering Department, Federal University of Mato Grosso, Campus Várzea Grande, MT, Brazil

<sup>c</sup> Electrical and Computer Engineering Department, University of Florida, Gainesville, FL, USA.

E-mail addresses: [raulvitor@ufmt.br](mailto:raulvitor@ufmt.br); [raoni@ufmt.br](mailto:raoni@ufmt.br); [arturo@ece.ufl.edu](mailto:arturo@ece.ufl.edu);

\*Correspondence: [raulvitor@ufmt.br](mailto:raulvitor@ufmt.br)

## RESPONSE TO REVIEWERS

**We thank the reviewers for their questions and comments. They have allowed us to much improve the paper. In the following additions are the answers to specific questions and comments.**

**Reviewer #1: The paper is about power quality disturbances classification using deep learning**

### **1. Correct the vn inequality in c5 flicker**

*Answer: Dear Sir / Madam. The inequality has been corrected.*

**2. You have presented a good introduction. Various two stages methods and their success rates are presented. For example, it is written" the authors use the continuous wavelet transform (CWT)towards the extracting of 5 features from each measured signal for PQD classification, implemented through a ruled-based decision tree. Noise at 40 dB, 30 dB and 20 dB was mixed to measured signals, resulting in a classification accuracy of 99.7%, 98.5% and 93.8%, respectively". It is required to compare the results with the easily implementable DWT plus NN or SVM over the same sample tests or another well developed methods. Otherwise the results would not be fully appreciated.**

*Answer: Dear Sir / Madam. The last column, from the left to the right, of table 5 has been added to the text. DWT + NN results has been added to the text as requested. Table 7 and figure 9 have also been added to the text as a complement of the requested.*

**Reviewer #2: The authors have presented 2 Dimension Densely Connected Convolutional Network (2D-DenseNet) framework. Case study with synthetic disturbance events were analyzed. The reviewer has the following comments.**

### **1) Need more clarification on the new contribution of the paper.**

*Answer: Dear Sir / Madam. The new contributions of the paper relies on the unprecedented use of a Densely Connected 2D Convolutional Neural Network (2D DenseNet) for PQD classification problem. Due to its Densely characteristics, this network architecture allows the use of less filters than others convolutional neural networks that can be found on literature, as shown throughout the paper.*

*The text below can be found at the end of section 1, 4<sup>th</sup> page:*

*“This work presents a data-driven model based framework for PQDs diagnosis. Detection, identification and location of PQDs is realized simultaneously. To avoid the need of a two-step process, a two dimensional (2D) Densely Connected Convolutional Network (DenseNet) based solution is developed. This avoids the implicit computational cost while maintaining accuracy. Towards PQDs location, spatio-temporal characteristics are considered by the framework. The specific contributions of this work towards the state of the art are:*

- A 2D DenseNet does not need prior extraction of features for PQD signals by some mathematical method, which eliminates the 2-step based model and reduces computational burden;*
- Spatio-temporal environment characteristics are considered in a data driven based framework. A 2D DenseNet solution allows for the reduction in the quantity of filters per convolutional layer of DenseNet used to classify PQDs.”*

**2) The state of the art comparison of the current results with the papers [26,27,28] listed in the table 7 are too close in the accuracy and relevant methods. Therefore, the authors need to extensively compare the proposed work with the referred state of the art papers [26-28] in all aspects (methods/deep learning approach/layers/performance etc) and distinguish in the considerable amount to establish the new contribution.**

*Answer: Dear Sir / Madam. Table 9 has been added to address the reviewer question.*

**3) Whether the model is overfitting?**

*Answer: Dear Sir / Madam. Thank you for this question. The information below has been added to the text. Last paragraph, page 22.*

*“Since the model effectively predicted new data from results not used in training, the experiments performed do not indicate overfitting. In addition to testing the model with computer-simulated data, we also used data generated in real-time by a programmable power supply, and the model's performance remained unchanged.”*

**4) Considering the achieved results of 99.99 percentage, the authors need to illustrate the real challenge in this research work, limitations and research gap need to be addressed for future work.**

*Answer: Dear Sir / Madam. The text below can be found at the last paragraph in conclusion section:*

*“Future works about power quality disturbances classification should focus on the model size of the proposed deep network, so it can improve its performance on embedded devices making them more efficient. The application of new deep neural networks to PQDs classification problem, as the inception neural network, is also of interest for future works.”*

**Dear revisers, we would like to thank you each of your time and patience in revising this paper.**

**Best regards,**

**Corresponding author: Raul Vitor Arantes Monteiro**

**Declaration of interests**

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

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

**CRediT author statement**

**A POWER QUALITY DISTURBANCES  
DIAGNOSIS: A 2D DENSELY CONNECTED  
CONVOLUTIONAL NETWORK  
FRAMEWORK**

**Raul V. A. Monteiro:** Software, Data Curation, Term, Conceptualization, Writing – Original Draft, Project administration. **Raoni F. S. Teixeira:** Supervision, Software, Data Curation, Resources. **Arturo S. Bretas:** Writing – review & editing; Formal analysis.