



## Novelty Detection on Power Quality Disturbances Monitoring

A. D. Gonzalez-Abreu<sup>1</sup>, M. Delgado-Prieto<sup>2</sup>, J.J. Saucedo-Dorantes<sup>1</sup> and R. A. Osornio-Rios<sup>1</sup>

<sup>1</sup> HSPdigital CA-Mecatronica Engineering Faculty, Autonomous University of Queretaro  
San Juan del Rio 76807, Queretaro, Mexico,

agonzalez63@alumnos.uaq.mx, raosornio@hspdigital.org, jsaucedo@hspdigital.org

<sup>2</sup> MCIA - Electronic Engineering Department, Technical University of Catalonia,  
Terrassa 08222, Barcelona, Spain,  
miguel.delgado@upc.edu

**Abstract.** Complex disturbance patterns take place over the corresponding power supply networks due to the increased complexity of electrical loads at industrial plants. Such complex patterns are the result of a combination of simpler standardized disturbances. However, their detection and identification represent a challenge to current power quality monitoring systems. The detection of disturbances and their identification would allow early and effective decision-making processes towards optimal power grid controls or maintenance and security operations of the grid. In this regard, this paper presents an evaluation of the four main techniques for novelty detection:  $k$ -Nearest Neighbor, Gaussian Mixture Models, One-Class Support Vector Machine, and Stacked Autoencoder. A set of synthetic signals have been considered to evaluate the performance and suitability of each technique as an anomaly detector applied to power quality disturbances. A set of statistical features have been considered to characterize the power line. The evaluation of the techniques is carried out throughout different scenarios considering combined and single disturbances. The obtained results show the complementary performance of the considered techniques in front of different scenarios due to their differences in the knowledge modelization.

**Keywords.** Condition monitoring, fault detection, novelty detection, power quality, power quality disturbances.

### 1. Introduction

Power quality has been critical since the massive introduction of non-linear loads to the power grid, such as protection systems, digital-based equipment, and industrial processors in general. Power Quality (PQ) has become one of the most important issues for both utilities and consumers since it plays a vital role in the proper operation of electrical power systems [1]. Typically, PQ problems are much severe at the utilization level and in the industrial sector. Significant research and development around power quality monitoring schemes are based on fault detection and identification methodologies. Thus, a great effort has been carried out towards high-performance disturbance characterization procedures [2]. Among the most important power quality standards, the IEEE 1159 [3] represents the most comprehensive reference in which the limits of distortions for different levels of the power system are defined. This standard considers the quantitative definition of basic power quality disturbances. However, unlike the

ones caused by the simple combination, complex disturbance patterns are not defined in such international standards. In this regard, the appearance of disturbances does not always correspond to predefined patterns. In this sense, novelty detection represents a current field of investigation applied in multiple machine learning-based applications. Novelty detection refers to identifying abnormal system behavior over which a data-based model of normality has been constructed. Such deviations from the reference patterns are identified as "novelty". The detection of novel patterns during the on-line operation of the corresponding fault detection and identification systems is being considered an essential function for the next generation of monitoring schemes applied to industrial maintenance and supervision. Although a great deal of novelty detection research is available in fields of application like electrical motors or power converters, its application to power quality monitoring schemes has not been clearly explored yet. Most of the few available works dealing with novelty detection and PQ are limited to the detection of stationary changes in voltage and current signals under monitoring [4]. The work presented in [5] brings the concept of novelty detection slightly different from the classical. This study segment the signal in fixed length frames and compares it with a reference one. If a frame surpasses the metric considered as a threshold in the reference frame is a novelty. In [6], an approach for the automatic classification of voltage waveforms is presented. This work classifies the novelties as waveforms not available during the training stage, and its classification method is based on support vector data description. Although these works represent upcoming approaches to the field, they do not consider non-standardized disturbances. Also, neither discuss the suitability of different available novelty detection techniques, limiting their applicability in current industrial scenarios.

Thus, the contributions of this work include the consideration of a novelty detection framework applied to power quality monitoring. The proposed scheme comprises four of the most accepted machine learning techniques applied to novelty detection for their comparison:  $k$ -Nearest Neighbor ( $k$ NN), One-Class Support Vector Machine (OC-SVM), Gaussian Mixture

Model (GMM), and Stacked Auto-Encoder (SAE). The monitoring scheme includes the characterization of the current line through statistical features in the time domain and frequency domain. The performances of the considered techniques are analyzed in front of scenarios of single and combined disturbances. Thus, the analysis is carried out quantitatively through the discussion of the resulting accuracies.

## 2. Novelty detection techniques

Four novelty detection techniques are considered in this work. The following is a brief and concise explanation of the techniques according to [7]:

- *k-Nearest Neighbours*

The main concept rear this technique is that normal data will be projected near the available neighborhoods, while novelties will be projected far from that neighbors. The  $k$  stands for the number of neighboring points to be considered when evaluating a new datum. Thus, considering assessing a given data sample  $x$ , this point will be accepted as normal if the distance to its nearest neighbor  $y$ ,  $d_{xy}$ , is equal to or less than the distance from  $y$  to its nearest neighbor in the training set. Otherwise,  $x$  will be considered a novelty. Euclidean distance (1) is the most popular choice for univariate and multivariate continuous attributes.

$$\|x - y\| = \sqrt{\sum_{i=1}^D (x_i - y_i)^2} \quad (1)$$

- *One-Class Support Vector Machines*

OC-SVM aims to separate one class of target samples from the rest leading to a one-class characterization problem. Thus, one class is appropriately characterized, that is, the one including all available knowledge, while the rest of behaviors are considered novelties due to measurements are not available. This method finds an optimal hyperplane separating the known class of training data set in a high-dimensional feature space, and so, the test data can be classified [8]. The optimal separating hyperplane can be determined by solving the following constrained optimization problem:

$$\min \left\{ \frac{\|w\|^2}{2} + \frac{1}{N\nu} \sum_{i=1}^N \xi_i - \rho \right\} \quad (2)$$

Subject to

$$w \cdot \Phi(x) \geq \rho - \xi_i \quad \xi_i \geq 0 \quad (3)$$

where  $\nu$  is a penalty factor and  $\xi_i$  is the slack variable for the point  $x_i$ , where  $x_i, i = 1, \dots, N$  denotes training samples. The constants  $w$  and  $\rho$  are the normal vector and offset of the hyperplane, respectively. Despite OC-SVM being a linear function set, it is possible to solve non-linear classification problems using a kernel function. A binary classification indicates the side of the boundary in which the sample encounters states' novelty patterns.

- *Gaussian Mixture Model*

A GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm or the Maximum a Posteriori (MaP) estimation from a well-trained prior model. Its application over a given data sample  $x$  results in an estimation of its membership probability, which can be considered for novelty detection thresholding. A Gaussian mixture model is a weighted sum of  $M$  component Gaussian densities given by equation (4).

$$p(x | \lambda) = \sum_{i=1}^M w_i g(x | \mu_i, \Sigma_i) \quad (4)$$

where  $x$  is a  $D$ -dimensional continuous-valued data vector,  $w_i$  are the mixture weights, and  $g(x | \mu_i, \Sigma_i)$  are the component Gaussian densities.

- *Stacked Autoencoder*

The SAE is a neural network-based structure trained to replicate its input at its output after encoding and decoding stages. The basic autoencoder structure consists of two main components: an encoder and a decoder. The encoder contains the input layer and a hidden layer, which is used to represent the previous layer in a reduced dimension. Otherwise, the decoder takes the compressed information resulting from this hidden layer and returns it to the original dimension through an output layer. The stacking of different autoencoders is usually considered to reduce or compress the input data through several steps, passing from a high dimensionality to a low dimensionality without losing characteristic information relevant to reconstruct the original signal at the output of the structure. An example of a stacked autoencoder is shown in Fig. 1. The hidden layer with the smallest size represents the reduced set of information resulting from the feature reduction process. It is considered to be used for posterior detection and classification of faulty patterns.

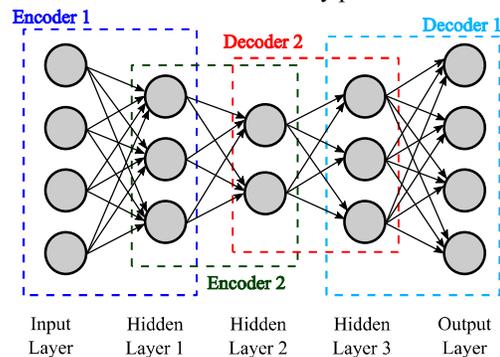


Fig. 1. An example of a stacked autoencoder composed of two autoencoders.

The training process is unsupervised and based on a cost function minimization. This cost function measures the error between the input and its corresponding reconstruction at the output. Thus, reconstruction errors higher than those obtained during the training process indicate novel patterns. Adding penalties to the cost function, an autoencoder could include sparsity of the representation, smallness into the derivative of the

representation, and robustness to noise or missing inputs. These coefficients are the L2W regularization term, the sparsity regularization term, and the sparsity proportion term. Thus, the cost function,  $J(x)$ , related to a regularized autoencoder is presented in (5).

$$J(x) = L(x, \hat{x}) + \lambda * \Omega_{weights} + \beta * \Omega_{sparsity} \quad (5)$$

where  $L(x, \hat{x})$  is a loss function (e.g., Mean Squared Error, MSE), that measures the error between the input  $x$  and the reconstruction  $\hat{x}$ .  $\lambda$  is the coefficient for the L2W regularization term  $\Omega_{weights}$ , and  $\beta$  is the coefficient for the sparsity regularization term  $\Omega_{sparsity}$  [9]. Thus, such  $J(x)$  cost function is approached as an optimization problem during the autoencoder training process.

### 3. Experimental methodology

The proposed methodology applied to PQ disturbances novelty detection is composed of three different stages, as shown in Fig. 2. Thus, in the first stage, a complete database of synthetic signals is generated. The synthetic database has been designed following the corresponding definitions of disturbances established by the IEEE Std 1159, leading to the normal signal of operation and different single and combined disturbance conditions. A meaningful set of 16 statistical time and frequency domain features are estimated to characterize all synthetic signals in the second stage. These numerical features are extracted from the related literature and represent a significant characterization process of the signals. Finally, the third stage corresponds to the training of the four novelty detection methods considered for their assessment. Each of the considered novelty detection techniques has distinct parameters that need to be determined. Such parameter selection has been carried out following the corresponding procedures available in the related literature.

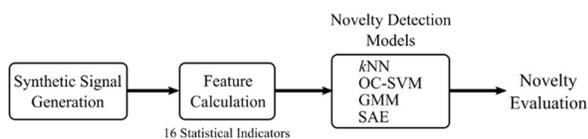


Fig. 2. Scheme for novelty detection techniques evaluation in PQ disturbances.

A complete set of test data, including known disturbances (i.e., single disturbances scenarios), and unknown disturbances (i.e., combined disturbances scenarios), is considered. The performance for each of the four techniques is quantitatively analyzed in front of multiple novelty detection scenarios.

*A. Database synthetic signals and disturbances scenarios*  
Various challenging scenarios of PQ disturbances are considered to validate the proposed methodology, including multiple classes and combinatorial patterns considering single disturbances. It should be noted that most of the studies related to PQ monitoring deal with single disturbances. However, just a few of them consider PQ scenarios, including the combination of two disturbances, which represent a challenging and required framework. Thus, the complete set of conditions or classes considered

are: C1: Normal; C2: Sag; C3: Swell; C4: Interruption; C5: Flicker; C6: Harmonics; C7: Oscillatory Transient; C8: Sag with Harmonics; C9: Sag with Oscillatory Transient; C10: Swell with Harmonics; C11: Harmonics with Oscillatory Transient; C12: Harmonics with Flicker.

A set of 1000 signals for each class is generated to face the training stage properly for the four techniques. The common parameters of the signals are 60 Hz for his fundamental frequency and amplitude in per unit (pu). The signals generation and the posterior procedure of evaluating the techniques have been carried out in Matlab 2020b. Specifically, according to the models developed by parametric equations, the signals have been obtained considering all distinct representations in severity for each class. The signals are created with a window time of 0.1666 ms, equivalent to ten cycles of the signal. This window size is chosen because the voltage is usually measured on a cycle-by-cycle basis, according to [3]. The sampling frequency is 15.36 kHz. Fig. 3 depicts some of the resulting disturbances considered.

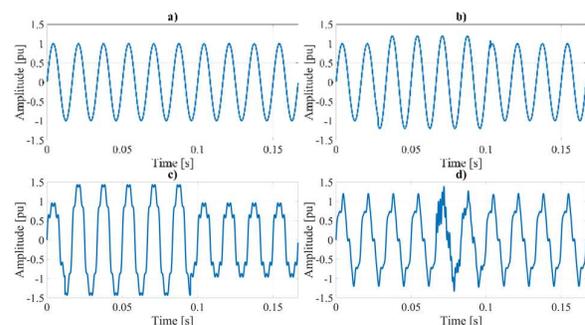


Fig. 3. Electrical signals synthetically generated. a) Normal. b) Swell. c) Swell with Harmonics. d) Harmonics with Oscillatory Transient.

Once the numerical feature database is generated, a training set and test set are defined. The training set includes 90% of the known samples, and the test set the remaining 10% in addition to the same number of novelty condition samples. A 10-fold-cross validation approach is taken into account throughout the training process.

In this sense, three scenarios are defined in order to test each one of the four techniques. Table I depicts the scenarios with the classes considered as known and novelty for the test. For the three scenarios, two classes are considered novelties. In this regard, for each test, only one class is tested, so for each one of the three scenarios, two different tests -one for each class unknown- have been performed.

Table I – Scenarios selected and the classes considered in each one as known and novelties.

Scenario	Know Classes	Novelties	
		Test 1	Test 2
1	C1, C2, C3, C4, C6, C7	C5	C8
2	C1, C2, C3, C4, C5, C6, C7, C8	C9	C11
3	C1, C2, C3, C4, C5, C6, C7, C8, C9, C11	C10	C12

### B. Feature extraction calculation

The signals have been characterized in the time domain and frequency domain for the database, calculating eight statistical indicators for each one of these domains. FFT is calculated for the frequency domain. As a result, a feature characterization array of 16 statistical indicators is obtained in this stage. Table II depicts the equations for the statistical indicators considered.

Table II – Set of statistical features

Mean	$\bar{x} = \frac{1}{n} \cdot \sum_{k=1}^n  x_k $	(6)
Deviation Standard	$\sigma = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^n (x_k - \bar{x})^2}$	(7)
Variance	$\sigma^2 = \frac{1}{n} \cdot \sum_{k=1}^n (x_k - \bar{x})^2$	(8)
Skewness	$S_k = \frac{E[(x_k - \bar{x})^3]}{\sigma^3}$	(9)
Kurtosis	$k_u = \frac{E[(x_k - \bar{x})^4]}{\sigma^4}$	(10)
Energy	$E = \sum_{k=1}^n ( x_k ^2)$	(11)
Entropy	$ET = - \sum_{k=1}^n x_k^2 \log(x_k^2)$	(12)
Log-Entropy	$LE = - \sum_{k=1}^n \log(x_k^2)$	(13)

### C. Configuration of techniques' hyperparameters

A proper configuration of each technique is needed in order to reach a significant level of performance. In this regard, for each technique, its configuration has been based on an empirical approach following related literature. Next, for each technique considered, a description of its key hyperparameters -and their configuration- is presented.

- *k-Nearest Neighbours*

The parameter  $k$  of the  $k$ NN method is a specific parameter defined by the user. Often  $k$  is set to a natural number closer to the square of training samples [10]. Low values for  $k$  like 1 or 2 can be noisy and subject to the effects of outliers. Large values for  $k$  smooth over things.  $k$  should not be so large that other categories will always outvote a category with only a few samples in it. Nonetheless, not always increasing the value of  $k$  makes the prediction accurate [11].

In this work, the value of  $k$  is settled in a value of 3, resulting in the best performance. The values that have been tested are 3,5,10,20, and 30. One established the value of  $k$ , the threshold for novelty detection with  $k$ NN has been defined. With the average Euclidean distance of the  $k$  nearest neighbors in the training data, the knowledge data, a histogram with the adequate resolution, is created, and a 95th percentile is selected. So, the calculated distance of test samples over this threshold is considered a novelty.

- *One-Class Support Vector Machine*

In OC-SVM, the penalty factor  $\nu$  is settled in 0.01. This parameter leads to a less or more number of support vectors and established the decision boundary. An optimal value of this parameter should capture the data complexity either to avoid overfitting. The larger the penalty factor, the more overfitted the hyperplane. The smaller the penalty factor, the more adapted the hyperplane. In general, a value that detects 90-95 % of the available data is correct. The novelty metric is binary, on one side or the other of the hyperplane.

- *Gaussian Mixture Models*

To select the number of components of the GMM, a range of component numbers is evaluated. Then, using the Bayesian information criteria (BIC), the GMM with the number of components resulting in the lowest BIC is selected. The range of the number of components is between the number of classes known and 1.5 times that number. The threshold is selected as a 95th percentile of the distribution from the sample distance to each mean of each component in the GMM.

- *Stacked Autoencoder*

One of the main difficulties when considering an SAE is related to the design of its structure, that is, selecting optimal values for each configuration parameter. Indeed, the SAE requires the definition of specific hyperparameters that must be selected to obtain a proper performance. These hyperparameters include: (i) the hidden layers size, (ii) the L2W regularization, and (iii) the sparsity regularization. For the hidden size of the layers, a deep network is proposed based on reducing the features. In this approach to obtain the maximum optimal representation in each hidden layer, three encoder layers are considered. The reduction ratio is between one-third and one-tenth regarding the number of features as input considered [12]. In this approach, the values selected are one-eighth for the first layer concerning the input size and then one quarter for the following two layers. That means that the final representation is of size three eighth from the input size. This reduction ratio is selected according to [13] that indicates that a fewer hidden size can lead to a poor generalization ability of the SAE. Also, too many hidden layers cause invalid representations to be extracted. An empirical procedure based on a coarse grid search has been considered regarding the regularization terms in the cost function. The values result for the hyperparameters selection are summarized in Table III.

Table III - Hyperparameters values for the autoencoder configuration

Hyperparameter	Value Selected	
Number of Autoencoder	3	
	Layer	Size
Hidden layers size	1	14
	2	10
	3	6
L2W Regularization	0.000001	
Sparsity Regularization	0.0001	

A histogram with the MSE of all training samples is performed and based on the Cumulative Distribution Function to select the threshold to define a novelty. A 90% percentile value is defined as a threshold, so the MSE obtained test samples higher than the threshold is considered a novelty.

#### 4. Results and discussion

The results that are presented next include the analysis of the techniques' performance in front of the different scenarios of known and unknown disturbances. The performance of the considered techniques in front of the tests presents significant variations depending on the configuration parameters and complexity of the novelty detection scenario. In this regard, the resulting novelty detection accuracy for the confusion matrix in each of the considered tests and techniques is summarized in Table IV.

Table IV – Accuracy of each technique as novelty detector

	Scenario #1		Scenario #2		Scenario #3	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
<b>KNN</b>	84.72%	91.14%	87.14%	88.74%	89.7%	86.36%
<b>GMM</b>	84.78%	83.38%	83.92%	83.26%	83.72%	82.96%
<b>OC-SVM</b>	89.14%	90.04%	88.44%	88.06%	90.28%	87.6%
<b>SAE</b>	78.88%	85.18%	81.68%	81.9%	85.3%	82.6%

The accuracy results depict that in scenario 1, test 1 is more complicated to describe the pattern distribution, and only the OC-SVM is near to reach 90%. Exclusively in test 1, the novelty class is a simple disturbance, so the characterization with the statistical indicators considered is not sufficient to assign the proper class to the test samples. All other techniques achieve a performance of less than 85%. In test 2, *k*NN presents the best performance. In scenario 2, *k*NN and OC-SVM are maintained as the best techniques. However, GMM maintains a lower level of performance as in scenario 1. In scenario 3 the same trend is maintained similar to scenario 2, with only slightly increased SAE performance. Again GMM presents the lowest accuracy. The results of each technique related to known and novelty classes are presented next. A more descriptive performance is depicted through Table V to Table VIII.

Table V - Detail of *k*NN performance

	Scenario #1		Scenario #2		Scenario #3	
	K	N	K	N	K	N
<b>Test 1</b>	91.2%	45.8%	91.6%	51.4%	91.08%	75.8%
<b>Test 2</b>	91.08%	91.4%	91.58%	66%	91.04%	39.4%

K-Known set; N-Novelty set

In the *k*NN technique, the three scenarios for the known classes present a good performance of around 91%. However, detecting novelties is lower than 80% in almost all the scenarios and tests.

Table VI - Detail of GMM performance

	Scenario #1		Scenario #2		Scenario #3	
	K	N	K	N	K	N
<b>Test 1</b>	92.1%	40.8%	90.86%	28.4%	88.32%	37.6%
<b>Test 2</b>	90.34%	41.6%	90.52%	25.2%	88.98%	22.2%

K-Known set; N-Novelty set

In the GMM technique, the performance for the known classes is around 90% decreasing 2% through more complicated data distribution as in scenario 3. The local performance related to detecting novelty patterns is under 40%, decreasing in half when more data is added to the scenario.

Table VII - Detail of OC-SVM performance

	Scenario #1		Scenario #2		Scenario #3	
	K	N	K	N	K	N
<b>Test 1</b>	89.34%	88%	88.1%	91.2%	89.88%	94.2%
<b>Test 2</b>	88.52%	99.2%	87.54%	92.4%	90.38%	59.6%

K-Known set; N-Novelty set

The OC-SVM shows an equilibrated performance. The performance to detect novelties is slightly more outstanding on some occasions, almost reaching 100% on one occasion. Nevertheless, the lowest performance is near to 60% related to detect novelties.

Table VIII- Detail of SAE performance

	Scenario #1		Scenario #2		Scenario #3	
	K	N	K	N	K	N
<b>Test 1</b>	90.28%	10.4%	89.64%	18%	90.4%	34.4%
<b>Test 2</b>	89.34%	60.2%	90.08%	16.6%	89.98%	9%

K-Known set; N-Novelty set

The SAE is a technique where unexpected values have been obtained. In this sense, the known class performance is suitable on an average of 90% in all scenarios and number tests. However, the most high performance in novelty patterns is 60%, and on some occasions, lower than 10% is observed. In Fig. 4, a signal from the known class and a signal for the novelty class are depicted to show the work made by the SAE encoding and decoding the information.

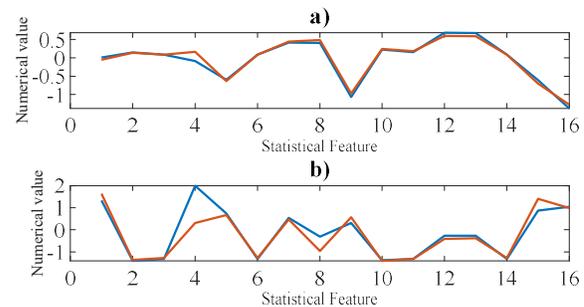


Fig. 4. Input feature vector and its reconstruction by the SAE. a) Known class sample. b) Novelty class sample.

To summarize, Fig. 5 shows the average performance obtained for each technique through the three scenarios and the two tests in each one. In this figure is depicted the performance for known classes and novelties. It must be noted that the four techniques are suitable to identify the known class, but in novelties, OC-SVM is superior to the others. This superior yield could be because selecting a

hyperplane adjusts the data patterns or distribution better, leading to a better separation of the classes. The four techniques have a global performance of around 90%, although it is due to the excellent performance obtained in identified known classes.

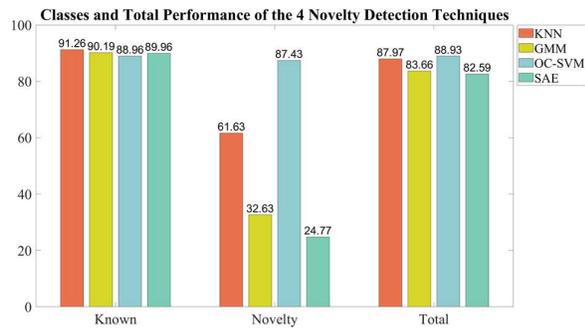


Fig. 5. Overall performance of the four techniques as novelty detectors.

In this sense, suppose that different classes' configurations adding to the complex knowledge model could bring different performances. Also, the considerations of an exhaustive grid search or stochastic search or algorithm to looking for the best hyperparameters of each technique could considerably increase the four techniques' global performance.

## 5. Conclusion

A novelty framework of four techniques is presented and evaluated in various PQ disturbances scenarios as novelty detectors in this work. There are three important aspects in this work. First, the consideration of novelty detection in the PQ area. Although a few works present Novelty Detection in PQ, the increase of the loads attached at the grid leads to complex patterns, and detecting them is a current challenge to improve condition monitoring power systems. Second, the consideration of different scenarios with simple standardized and combined disturbances patterns. However, combined disturbances are not described or defined, and reported them is a need that must face it. Third, the four techniques evaluated are among the most typical in the Novelty Detection area, so considering these techniques involves the different approaches, based in domain, based on distance, based on probabilistic and new trend like deep learning approaches. With the evaluation, advantages, and disadvantages of each of the techniques exposed, the performance results bring an interesting perspective. It showed that some techniques work excellently in detecting novelty patterns, even not present a good performance in the known classes. Also, the configuration of parameters in each technique plays a vital role in achieving a suitable performance. The results achieved for the techniques could be misinterpreted if the global performance is observed. For this reason, detail of performances in detecting the classes known and novelties are depicted and discussed, resulting in some techniques best detecting the known classes. In this sense, a hybrid scheme could provide the potential of detecting adequately known and novelty patterns. Also, the next step in this research is to find a qualitative way to show the approximation of the novelty detection techniques presented to visualize in 2-Dimensional the representation

achieve for each technique without losing information due to the compression technique considered.

## Acknowledgments

This research work has been partially supported by FOFIUAQ-2018 FIN 201812 and CONACyT doctoral scholarship number 735042. The authors would like to thank the support provided by the Catalan Agency for Management of University under the grant 2017 SGR 967.

## References

- [1] Y. Naderi, S. H. Hosseini, S. Ghassem Zadeh, B. Mohammadi-Ivatloo, J. C. Vasquez, and J. M. Guerrero, "An overview of power quality enhancement techniques applied to distributed generation in electrical distribution networks," *Renew. Sustain. Energy Rev.*, vol. 93, no. March, pp. 201–214, 2018.
- [2] P. Khetarpal and M. M. Tripathi, "A critical and comprehensive review on power quality disturbance detection and classification," *Sustain. Comput. Informatics Syst.*, vol. 28, p. 100417, 2020.
- [3] IEEE Power and Energy Society, *IEEE Recommended Practice for Monitoring Electric Power Quality*, vol. 2019. New York City, NY, USA: IEEE Standard 1159-2009 (Revision IEEE Std 1159-2009), 2019.
- [4] E. B. Kapisch, L. M. A. Filho, L. R. M. Silva, and C. A. Duque, "Novelty Detection in Power Quality Signals with Surrogates: A Time-Frequency Technique," *Int. Conf. Syst. Signals, Image Process.*, vol. 2020-July, pp. 373–378, 2020.
- [5] L. R. M. Silva, E. B. Kapisch, C. A. Duque, L. M. De Andrade Filho, A. S. Cerqueira, and P. F. Ribeiro, "The concept of novelty detection applied to Power Quality," *IEEE Power Energy Soc. Gen. Meet.*, vol. 2016-Novem, 2016.
- [6] A. E. Lazzaretti, D. M. J. Tax, H. Vieira Neto, and V. H. Ferreira, "Novelty detection and multi-class classification in power distribution voltage waveforms," *Expert Syst. Appl.*, vol. 45, pp. 322–330, 2016.
- [7] J. A. Cariño Corrales, "Fault detection and identification methodology under an incremental learning framework applied to industrial electromechanical systems," Universitat Politècnica de Catalunya, 2017.
- [8] J. C. Bravo-Rodríguez, F. J. Torres, and M. D. Borrás, "Hybrid Machine Learning Models for Classifying Power Quality Disturbances: A Comparative Study," *Energies*, vol. 13, no. 11, p. 2761, Jun. 2020.
- [9] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Res.*, vol. 37, no. 23, pp. 3311–3325, Dec. 1997.
- [10] J. C. Bravo-Rodríguez, F. J. Torres, and M. D. Borrás, "Hybrid Machine Learning Models for Classifying Power Quality Disturbances: A Comparative Study," *Energies*, vol. 13, no. 11, p. 2761, Jun. 2020.
- [11] J. A. Cariño, M. Delgado-Prieto, D. Zurita, A. Picot, J. A. Ortega, and R. J. Romero-Troncoso, "Incremental novelty detection and fault identification scheme applied to a kinematic chain under non-stationary operation," *ISA Trans.*, vol. 97, pp. 76–85, 2020.
- [12] J. He, M. Ouyang, C. Yong, D. Chen, J. Guo, and Y. Zhou, "A Novel Intelligent Fault Diagnosis Method for Rolling Bearing Based on Integrated Weight Strategy Features Learning," *Sensors*, vol. 20, no. 6, p. 1774, Mar. 2020.
- [13] W. Qiu, Q. Tang, J. Liu, Z. Teng, and W. Yao, "Power Quality Disturbances Recognition Using Modified S Transform and Parallel Stack Sparse Auto-encoder," *Electr. Power Syst. Res.*, vol. 174, no. February, p. 105876, Sep. 2019.