Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 67 (2018) 69 – 74

www.elsevier.com/locate/procedia

11th CIRP Conference on Intelligent Computation in Manufacturing Engineering, CIRP ICME '17

Machine allocation via pattern recognition in harmonic waves of manufacturing plants

Arnim Reger[a,*], Jonas Dumler[a], Oleg Lobachev[b], Julian Neuberger[b], Rolf Steinhilper[a]

[a] *Fraunhofer Institute for Manufacturing Engineering and Automation, Project Group Process Innovation, Universitaetsstrasse 9, 95447 Bayreuth, Germany*
[b] *Visual Computing of University Bayreuth, 95440 Bayreuth, Germany*

* Corresponding author. Tel.: +49-921-78516-32; fax: +49-921-78516-105. *E-mail address:* a.reger@ipa.fraunhofer.de

Abstract

Non-intrusive load monitoring is currently used to analyze changes in the energy consumption of households. Due to the number of electrical consumers, the associated superpositions and the variety of harmonic waves on the shop floor, current proceedings are not applicable in industrial environment. In this paper, patterns in harmonic waves of four manufacturing plants are analyzed in the time and frequency domain. For machine allocation, features were extracted and classified by k-means and support vector machines with an accuracy of 97.3 and 97.9 %. For comparison, convolutional neural networks were trained with the harmonic profiles in the time domain with an accuracy of 98.7 %.

Peer-review under responsibility of the scientific committee of the 11th CIRP Conference on Intelligent Computation in Manufacturing Engineering

Keywords: Machine allocation; Non-intrusive load monitoring; Production plants; Harmonic waves; Deep learning; Convolutional neural networks

1. Introduction

First introduced by Hart [1], many technologies were developed in over more than 30 years of non-intrusive load monitoring (NILM). There are approaches with sampling rates from 1 Hz up to hundreds of kHz, some of them extracting simple information like difference power signature, others are looking for shapes of transient signatures or complex information like harmonics or electromagnetic interference. [2]

For classification, Khalid et al. [3] generally distinguishes four types of appliances:

- Type I: Two states of operation (on/off)
- Type II: Multi-State appliance with finite operating states
- Type III: Continuously variable devices
- Type IV: Permanent consumer devices

In NILM applications, the appliance signature can be steady state or transient nature. A summary of methods, features and requirements can be found at Zoha et al. [4].

During the long time of NILM research, there are now many real systems in residential applications, but there is a lack of research in industrial environment. Implementing NILM in industrial processes leads to significant difficulties due to a high noise level and a large variety of electrical consumers. In industrial sites, manufacturing plants got many basic electrical appliances, the noise caused by high power motors, fans and invertors can be larger than threshold values in NILM systems. Beyond industrial secrecy, the number of equipment types and temporal patterns would make the research more difficult. [2, 5]

Static power converters are the largest nonlinear loads in the industrial environment. They are used for a variety of purposes e.g. electrochemical power supplies, variable speed drives or arc furnaces. Those nonlinear loads change the sinusoidal nature of the AC power current and cause a harmonic current flow and consequently an AC voltage drop. Each harmonic producing device can have a consistent or load depending variable harmonic current emission characteristic. [6]

For nonlinear device detection, Akbar and Khan [7] and Jonetzko et al. [8] analyzed the harmonic shape of the current signal with high sample rates. The features for classification were extracted by applying the Fast Fourier Transform (FFT). Fuller [9] demonstrated the potential of current spectrum analysis to detect electric machine failures and monitor multiple machines.

Peer-review under responsibility of the scientific committee of the 11th CIRP Conference on Intelligent Computation in Manufacturing Engineering
doi:10.1016/j.procir.2017.12.178

In advanced NILM research the time-frequency-domain was examined for feature extraction with wavelet transforms [10]. Khalid et al. [11] trained neural networks with features from s-Transform.

For very fast detections, Uçar et al. used discrete wavelet transform (DWT) for feature extraction in combination with Extreme Learning Machine [12]. Duarte et al. [13] monitored the voltage for power quality detection using wavelets and support vector machines. Bischa et al. [14] fed the signals to autoencoder for feature extraction and compared the classification of support vector machines (SVM) with convolutional neural networks (CNN).

Hereinafter, a NILM application for machine or component allocation in the industrial environment is discussed. Features were extracted with FFT and classified by k-means and support vector machines. In time domain, CNNs were trained with the labeled data sets as well.

2. Methods

2.1. Initial Situation

The research subject was a HAUNI multifilter machine with the main unit A and three auxiliary units B, C and D. Fig. 1 shows the four machines with their related current waveforms connected to the main supply.

2.2. Data Acquisition

For data acquisition, a NI cDAQ-9171 USB-Chassis with a simultaneous analog input module NI 9215 (4 channels, 100 kS/s, 16 bit, ±10 V) and a bandwidth of 420 kHz (-3 dB) was connected to a laptop with a LabView program. For fast measurement, Fluke i200s Rogowski Coils with BNC connector were used. They provide a measurement range up to 200 A at 10 mV/A.

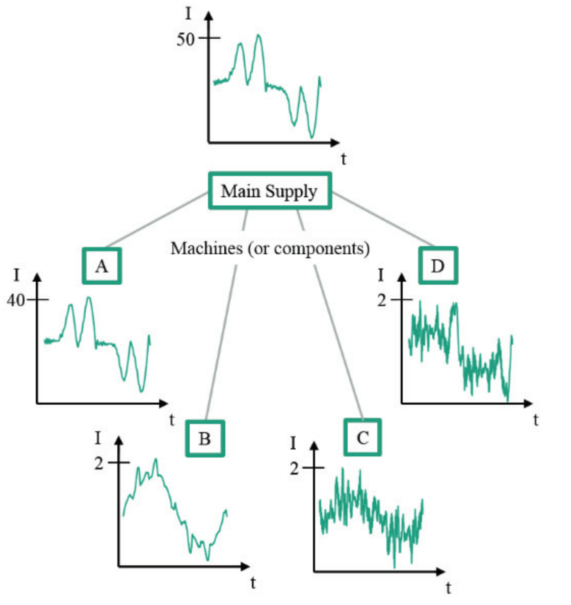

Fig. 1. Schematic of the electric power supply with the related waveforms.

2.3. Generating Labeled Data Sets

A moving average filter followed by a zero-crossing detection algorithm was used in order to cut the signal into single periods of a predefined length of 20 ms (2000 data points) and 1 s (100000 data points) respectively, dependent on the applied feature extraction and learning methods. As shown in Fig. 2, the measured signals from the four single machines were manually overlaid in all 15 possible combinations. The permutation of k out of n machines without repeating can be calculated by the binomial coefficient [15]. For k being a number between 1 and n the number of all possible combinations n_c is given as the sum of these coefficients, see equation 1. In this case 15 different data sets were generated.

$$n_c = \sum_{k=1}^{n} \binom{n}{k} = \sum_{k=1}^{n} \frac{n!}{k! \cdot (n-k)!} \tag{1}$$

2.4. Data Analysis in Time and Frequency Domain

To gain an overview over the statistical distribution, the data sets were analyzed by median, 0.1-quantile and 0.9-quantile in the time domain. In the frequency, first a FFT was used. For advanced analysis, a DWT and a CWT were applied. More

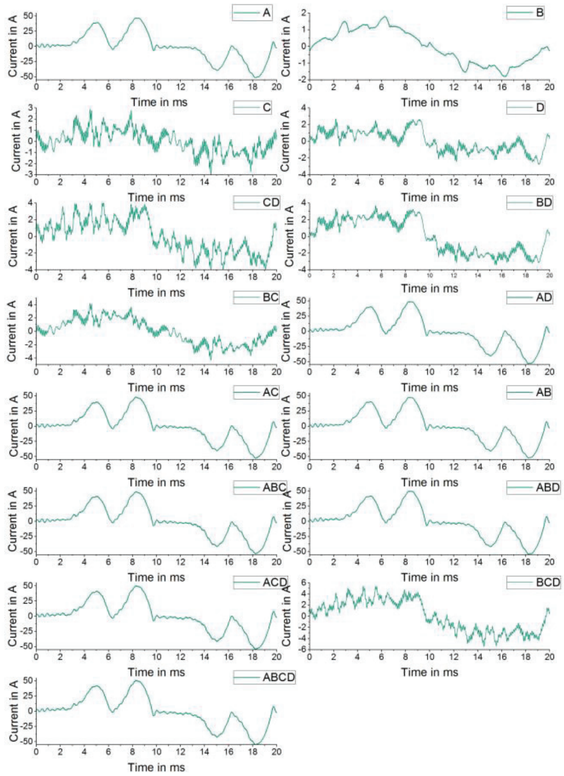

Fig. 2. Distribution of the Signals A, B, C, and D.

details about the wavelet theory can be found in previous research [16].

2.5. Machine Learning

Classical methods of pattern recognition include a process of analysis and classification. The analyzer extracts an array of significant features from the original data set by transformations and mathematical operations, which have to be defined by the user. There are a big amount of classification methods. In this case k-nearest neighbors classifier (k-NNC) and support vector machines (SVM) were used.

K-nearest neighbor classifiers are among the simplest machine learning techniques and belong to the class of lazy learning algorithms. For a given test sample the k nearest training samples are determined by a distance function. The class for the test sample is then assigned with a majority voting among the labels. [17]

A Support Vector Machine (SVM) can be described as a neuron trained by an algorithm that only considers samples of the training data, which are closest to the class limits. Thus the risk of over-fitting is reduced. For non-linear separable problems, a transformation of the vector space, called kernel, is performed. [18] [19]

2.6. Convolutional Neural Networks

In recent years, convolutional neural networks, short CNN, have risen in popularity for recognizing patterns in data. They are widely used for finding patterns in images, but have proven their usefulness in automatic speech recognition (ASR) [20].

To better simulate real world problems, noise was added at the training data set with a normal distribution to all samples. Each training example consisted of the previously mentioned 2000 signal samples in the time domain with a 20 ms timespan.

2.7. Architecture for Modelling Machine Allocation

The initial idea for an architecture capable of modelling machine allocation was based on one for modelling speech in automated speech recognition, described in [21]. Similar to the model described, the core of our architecture is built out of one convolutional layer and one maximum pooling layer, organized in a block. Additionally, a 30 % chance of dropout was applied after each max pooling layer. Three of those blocks were used in succession, followed by a fully connected layer, mapping the features found in the previous blocks to output neurons.

The first block, has an input length equal to the full 20 ms of one whole signal period. Features are found in this input and passed on to the following blocks, sampled down further and weighted, finally flattened and passed into the dense layer, which maps found patterns to machines. This dense layer utilizes the widely used sigmoid function as activation, allowing for a probability-like output.

At training, all convolutional blocks were only trained for one epoch. In successive epochs, the weights of those blocks are frozen, reducing trainable parameters, thus saving computation time.

3. Results

3.1. Data Analysis in Time and Frequency Domain

First, the four basic training data sets A, B, C and D were analyzed for their statistical distribution and spectral harmonics as shown in Fig. 3 and Fig. 4.

In Fig. 4 the selected features are marked by a dot. To show the influence of higher harmonics, the 1st harmonic waveform at 50 Hz is not shown and considered. For classification, the highest four values above the 1st harmonic were selected.

To show the spectral distribution over all 15 combinations, an inverse discrete wavelet transform (IDWT) and a continuous wavelet transform (CWT) were generated as shown in Fig. 5 and Fig. 6. In Table 1 the related signal combinations for both figures are shown.

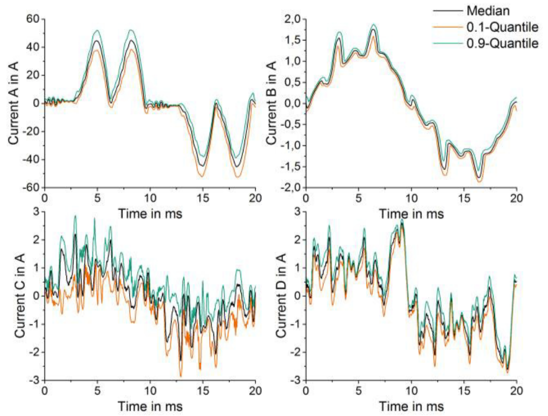

Fig.3. Statistical distribution of the four basic data sets A, B, C and D.

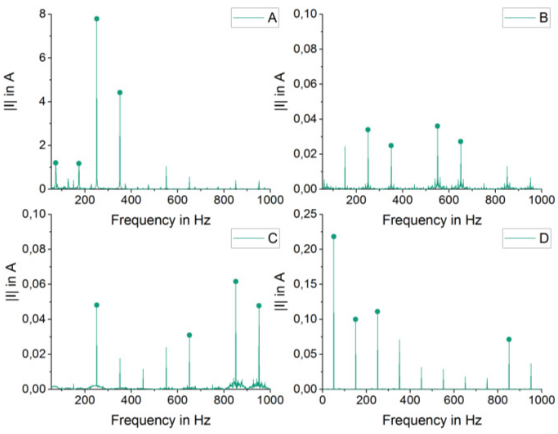

Fig.4. Spectral distribution of the four basic data sets with training points.

Table 1. Timesteps and signal combinations for CWT and IDWT

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A							A	A	A	A	A	A		A
	B				B	B			B	B	B		B	B
		C		C		C		C		C		C	C	C
			D	D	D		D				D	D	D	D

In Fig. 5 the IDWT of the signal combinations is shown. In order to fit the frequencies of interest (multiples of 50 Hz) into the bandpass ranges, the signal was down sampled by the factor 3 to a new sample rate of 33333 Hz.

The calculated 10-level IDWT leads to a frequency range up to ~32.5 Hz in the highest level. The main amplitude of 50 Hz is part of level 9 with a frequency range from 32.5 Hz to 65.1 Hz. For the CWT the signal first was down sampled by

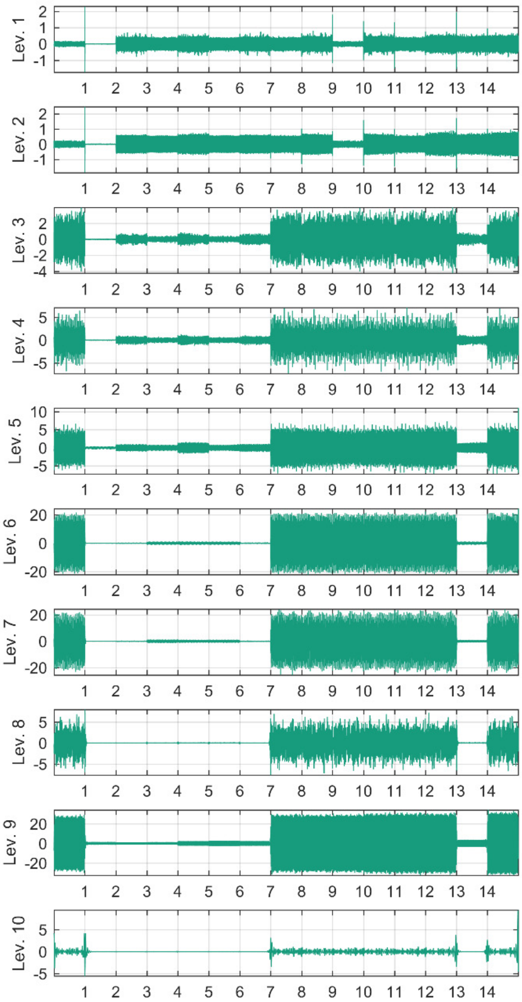

Fig. 5. IDWT of the signal combinations with a Meyer wavelet with a center frequency of 0.6902 Hz.

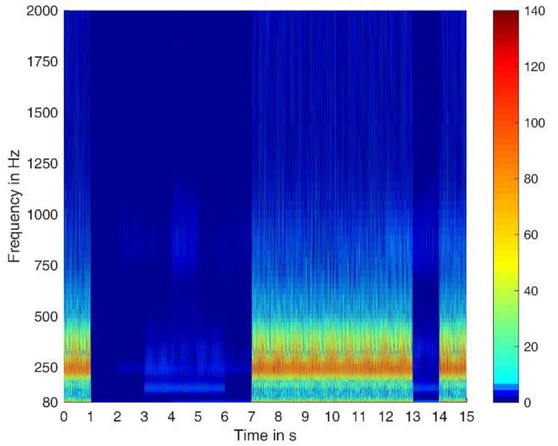

Fig. 6. CWT of the signal combinations analysed with a morlet wavelet with a center frequency of 0.8125 Hz.

taking every 8th data point into account in order to reduce the computing time. This led to a new sample frequency of 12.5 kHz. Fig. 6 shows the absolute values of the computed wavelet coefficients over a frequency range from 80 up to 2000 Hz.

Such spectral analyses can help to identify the relevant features in harmonics. From timestamp 2 till 7, the IDWT shows between level 3 and 5 significant harmonics, in the CWT in the frequency range up to 1000 Hz.

3.2. K-Nearest Neighbors

As described in the previous chapter, for classification the highest four values above the 1st harmonic were selected as features. The result for k-NNC is shown in Fig. 7.

In the k-NNC shown in Fig. 7 the distance metric was set to Euclidean and distance weight to equal.

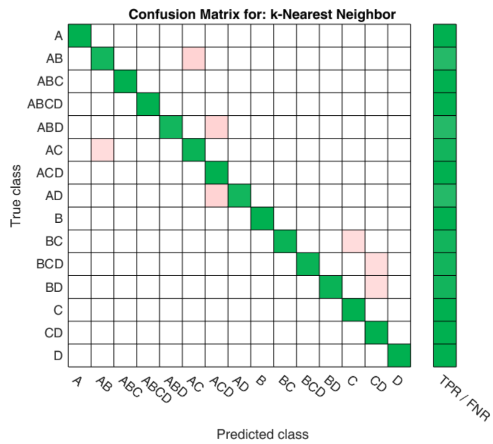

Fig.7. Confusion matrix for KNN with an accuracy of 97.3 %.

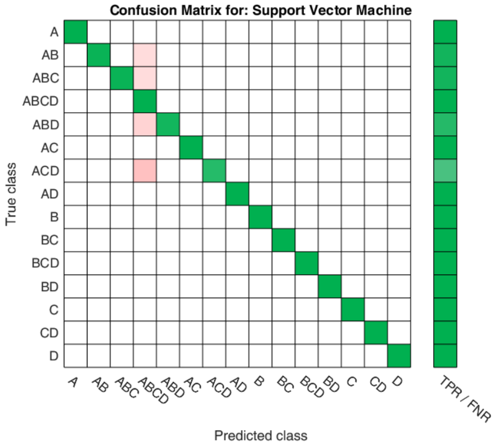

Fig. 8. Fine-Gaussian SVM with 97.9 % accuracy.

3.3. Support Vector Machine

Beside KNN a SVM was fed with the four harmonics and their amplitude from FFT as well. In addition, 25 % of the 2000 datasets for each of the 15 different labels were holdout for validation. The results are shown in Fig. 8.

For the analysis, the manual kernel scale was set to 0.7, the kernel scale mode to manual, the box constraint level to 1.0 and the multiclass method to one-vs-one.

3.4. Convolutional Neural Networks

Impact of number of convolutional blocks on loss

We tested several configurations of our model with differing numbers of used convolutional blocks. As shown in Fig. 9, by using three blocks, we achieved best accuracy, which replicates the findings of Abdel-Hamid et al. [20]. By using less than three blocks, we could not model the problem.

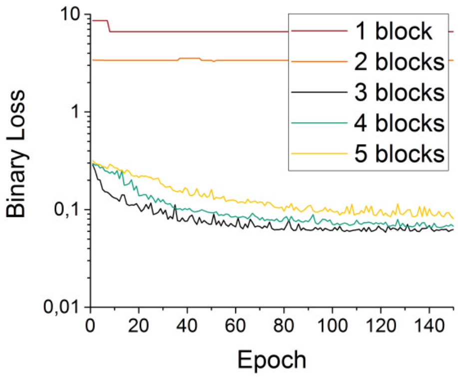

Fig. 9. Impact of number of convolutional blocks on loss.

Impact of dropout

Dropout can be used as an easy way to prevent neural networks from overfitting [21]. In our experiments, we could observe this, too, and achieved minimal loss, when using a dropout chance of 10 % to 30 %. Fig. 10 shows the influence of the dropout.

Impact of alternative training methods on loss

When training the CNN, we decided to freeze all weights of the convolutional blocks after the first epoch and only fine-tune the weights for the fully connected layer in all following epochs. Training the convolutional blocks for only one epoch instead of more made no impact on the final loss as well as the resulting prediction accuracy.

Impact of hyperparameters for the convolution blocks

Choosing hyperparameters for our convolutional blocks had a big impact on the modeling capabilities of our architecture. Most notably was the correct choice of activation function for the convolution layers. We tested the commonly used sigmoid, hyperbolic tangent and rectified linear unit activation functions [22]. As shown in Fig. 11, using a rectified linear unit as activation resulted in the lowest loss.

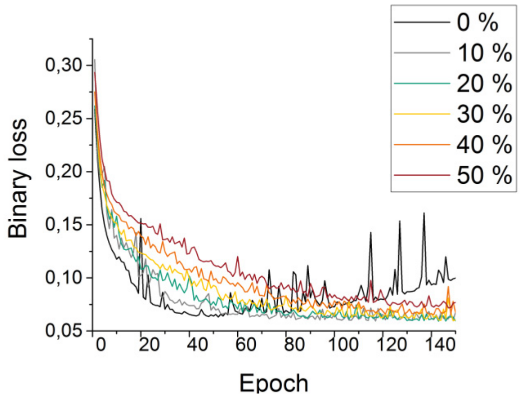

Fig. 10. Impact of dropout.

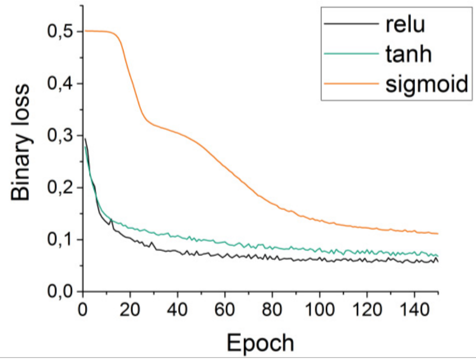

Fig. 11. Impact of activation.

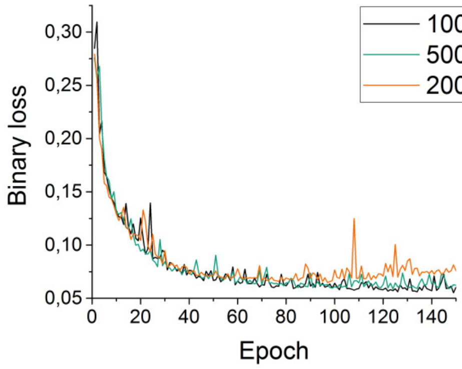

Fig.12. Impact of number of neurons in the fully connected layer.

Impact of number of neurons in the fully connected layer

In our tests, we found that a layer of at least 100 neurons activated by a sigmoid function works best. Using more than 500 neurons led to higher losses. In Fig. 12 the impact of the number of neurons is shown.

Overall, we achieved a prediction accuracy of 98.7 % with a binary loss value of 0.61 on our test set. To show the impact hyperparameters on the produced loss and accuracy of our architecture, we only changed one parameter at a time. We only showed the ones with the biggest impact and omitted the ones that provided only a marginal improvement.

4. Conclusion

With the presented NILM application for machine or component allocation in the industrial environment high accuracy rates were gained. With a three phase measurement, 150 training data sets per second can be acquired at a net frequency of 50 Hz. With the feature extraction of the four highest harmonics above the first, the harmonic current emission characteristic shows sufficient accuracy by classification with k-NN (97.3 %) and SVM (97.9 %). The trained CNN got the highest accuracy with 98.7 %.

5. Further Work

The conducted tests made clear, that CNNs can model machine allocation very well. The next steps will be the testing of their capabilities for larger numbers of connected machines, to prove their suitability in larger networks with more manufacturing plants.

Another step will be the reduction of the network for an application on embedded systems. Furthermore, we want to train the network on data, sampled while only a single machine is running and let it predict which machines are running when multiple are operating. For that task, we need to further improve the architecture.

References

[1] Hart GW. Nonintrusive appliance load monitoring. Proceedings of the IEEE 1992; Volume 80: 1870-1891.

[2] Trung KN, Zammit O, Dekneuvel E, Nicolle B, Van CN, Jacquemod G. An innovative non-intrusive load monitoring system for commercial and industrial application. International Conference on Advanced Technologies for Communications 2012: 23-27.

[3] Khalid SN, Abubakar I, Mustafa MW, Shareef H, Mustapha M. An Overview of Non-Intrusive load monitoring Methodologies. IEEE Conference on Energy Conversion 2015: 54-59.

[4] Zoha A, Gluhak A, Imran MA, Rajasegarar S. Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. Sensors 2012; Volume 12: 16838-16866.

[5] Adabi A, Mantey P, Holmegaard E, Kjaergaard MB. Status and challenges of Residential and Industrial Non-Intrusive Load Monitoring. IEEE conference on Technologies for Sustainability 2015: 181-188.

[6] IEEE Std 519-2014. IEEE Recommended Practice and Requirements for Harmonic Control in Electric Power Systems; 2014.

[7] Akbar M, Khan ZA. Modified Nonintrusive Appliance Load Monitoring for Nonlinear Devices. International Multitopic Conference 2007: 1-5.

[8] Jonetzko R, Detzler M, Gollmer KU, Guldner A, Huber M, Michels R, Naumann S. High frequency non-intrusive electric device detections and diagnosis. IEEE International Conference on Smart Cities and Green ICT Systems 2015: 1-8.

[9] Fuller AE. Harmonic Approaches to Non-Intrusive Load Diagnostics. Master Thesis. MIT; 2008.

[10] Su YC, Lian KL, Chang HH. Feature Selection of Non-intrusive Load Monitoring Systems using STFT and Wavelet Transform. IEEE 8th International Conference on e-Business Engineering 2011: 293-298

[11] Khalid K, Mohamed A, Shareef H, Sabeeh M. Event-based S-transform approach for nonintrusive load monitoring. Przegląd Elektrotechniczny 2016; Volume 5: 194-198.

[12] Uçar F, Alçin ÖF, Dandil B, Ata F. Machine learning based power quality event classification using wavelet – Entropy and basic statistical features. IEEE 21st International Conference on Methods and Models in Automation and Robotics 2016: 414-419.

[13] Duarte C, Delmar P, Goossen KW, Barner K, Gomez-Luna E. Non-intrusive load monitoring based on switching voltage transients and wavelet transforms. IEEE Future of Instrumentation International Workshop 2012.

[14] Binsha P, Sachin Kumar S, Athira S, Soman KP. Power Quality Signal Classification using Convolutional Neural Network. International Journal of Computer Technology and Applications 2016; Volume 9: 8033-8042.

[15] Flajolet P, Sedgewick R. Analytic Combinatorics. 1st Edition. Cambridge: Cambridge University Press; 2009.

[16] Reger A, Oette C, Aires AP, Steinhilper R. Pattern recognition in load profiles of electric drives in manufacturing plants. IEEE 5th International Electric Drives Production Conference 2015.

[17] Hitendra Sarma T,Viswanath P, Sai Koti Reddy D, Sri Raghava S. An improvement to k-nearest neighbor classifier. 3rd International Conference on Data Management 2010.

[18] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd Edition. Wiley-Interscience; 2000.

[19] Abe S. Support Vector Machines for Pattern Classification. London: Springer London; 2010.

[20] Abdel-Hamid O, Mohamed A, Jiang H, Deng L, Penn G, Yu D. Convolutional Neural Networks for Speech Recognition. IEEE/ACM transactions on audio, speech and language processing 2014; Volume 22: 1533-1545.

[21] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 2014; Volume 15: 1929-1958.

[22] Olgac AV, Karlik B. Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks. International Journal of Artificial Intelligence And Expert Systems (IJAE) 2011; Volume 1: 111-122.