

Intelligent PD Classification via SOM with Optimized Correlation

Z H Bohari, M Isa, A Z Abdullah, AAbRahman, M F Sulaima

Abstract—The Malaysian power industry was a regulated, organized electricity supply wise inherited from British colonization and showed to be robust and reliable throughout the millennium. The electricity sector is always committed to ensuring that supply reliability is preserved on a continual basis without any major disturbance. The most prevalent downtime causes are closely linked to high voltage (HV) equipment malfunctions or defects, especially transformers, switchgears and cables. The consequences of these occurrences are catastrophic and lead to the loss of millions of dollars in terms of the refurbishment or replacement of HV infrastructure. Power utility company has mostly attempted to stop this event from happening through HV equipment monitoring to classify the irregularity before it is more severe. Partial discharge (PD) event is among the phenomena that is being measured and evaluated. PD is commonly found in solid, gaseous or fluid form correlated with void. In this research, an unsupervised approach to the neural network is recommended for PD classification. This journal introduces the PD classification use an analysis based on the combo of self-organizing maps (SOMs) and correlation analysis. This project established that PD data was successfully correlated and clustered using MATLAB software's SOM Toolbox and correlation tool to identify the type of PD in a HV apparatus.

Keywords: HV Equipment, PD classification, Partial Discharges, neural network, MATLAB, Self-Organizing Maps, correlation analysis

I. INTRODUCTION

Basically, in Malaysia, electricity structures are based on the old monopoly regulated structure inherited since TNB (Tenaga Nasional Berhad) was known as CEB (Central Electricity Board) in 1949[1]–[3]. Malaysia once before independent was a colonized country by British Empire that per say help to facilitate the nation in built the first energy generation through electrical generator and small electrical distribution system in a mining area in Selangor recorded back in the last millennium. The electrical structure was designed to power up installed pumps use for mining operation. Since then, there are huge turnovers whereas CEB own about 30 energy plants with a total capacity of 40 MW[1], [2]. CEB operates both to transmit and distribute the produced energy mostly to heavy industry and few residential. The reliability and robust electricity system in Malaysia promoted via the introduction the biggest National

Revised Manuscript Received on September 14, 2019.

ZH Bohari, CeRIA, Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia.

AAbRahman, School of Electrical System Engineering, Universiti Malaysia Perlis (UniMAP), Perlis, Malaysia (Email: zulhasrizal@utem.edu.my)

M Isa, School of Electrical System Engineering, Universiti Malaysia Perlis (UniMAP), Perlis, Malaysia.(Email: muzamir@unimap.edu.my)

AZ Abdullah, School of Electrical System Engineering, Universiti Malaysia Perlis (UniMAP), Perlis, Malaysia.(Email: zaidiabdullah@unimap.edu.my)

M F Sulaima, School of Electrical System Engineering, Universiti Malaysia Perlis (UniMAP), Perlis, Malaysia.

Grid that brought plenty abroad investors to drive the nation's development and economic growth. All system related to energy is regulated under Malaysia's Energy Commission (ST)[3], [4].

Even though the system commissioned is proven reliable and robust throughout the century but power utility such as TNB can't predict the future abnormality occurrence for equipment installed without an appropriate HV equipment data and analysis. Each refurbishment or restoration of big HV equipment defects will take months to complete and commonly expensive[5][6]. Normal routine maintenance does the trick to prolong the lifespan of HV equipment but there is uncertainty of equipment defects due to manufacturing, overloading, harsh weather conditions, animal encroachment and human error.

Manufacturing defects and overloading is the common culprit that contributing to electrical system downtime and this happened to be the cause of major national blackout in 1992 and 1996 in Peninsular Malaysia that brought in new Independent Power Plants (IPPs) and new initiative to strengthen the system[1], [2], [7]. Millions of dollars losses recorded due to this event and TNB since then formulated a department to cater on research and development the best practices to do early defects prevention and monitoring.

Power utility use condition based monitoring method to do early malfunction detections; either intrusive or non-intrusive. This is adding on from normal routine maintenance that called preventive corrective or predictive maintenance. Among the effective ways to determine HV equipment or cable condition is through Partial Discharge (PD) measurement[8]–[12]. PD measurement is the best to be measured on HV equipment or cable during online. But the problem was, during most of the installation still energized. Each and every online equipment produced a sound or noise that eventually contaminate the whole set of PD data and finally produced a corrupted information and this will impact on the decision making especially for repairing or refurbishment HV equipment or cable. Electrical engineers were traumatized by bad judgement that might affect the stability of the system[5], [13], [14]. Therefore, in this manuscript a new artificial intelligent (AI) with appropriate statistical correlation is proposed to prevent this dire situation.

II. BACKGROUND THEORY

Signal Classification

Researchers stated that signal classification is a process that consists of numerous steps such as preparation of load classification, implementing load classification by utilizing clustering method and also

understanding the applications of classification[15]–[17].

Few claimed that the reliability and economic prerequisites can be made by consideration in arranging for the future establishment. Daily and future abnormal signal classification arranging play an exceptionally vital part in determining better decision.

Statistical Correlation

In few literatures shown that statistical correlation is crucial to have better understanding on various interdependences among the variables in a problem. By increasing the knowledge of the set of variables that describe a given problem, the capacity to classify the load data also increases. This opens the possibility of performing better action policies of any related operation linked to the phenomenon under study. This process plays an important figure to increase the consistency and accuracy of the classification process

Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) is a computational technique aiming to replicate human brain functioning. It is referred from the inventor of the first neurocomputer, says a neural network act as a computing framework made up of a number of basic, highly interconnected handling components which process data by their dynamic state reaction to external input [12], [16], [18].

A biological neuron has dendrites to get signals with a cell body to prepare them and an axon to send signals out to the other neurons in the meantime the artificial neuron has a number of input channels, a processing stage and one output that can bring out to multiple other artificial neurons as shown in Figure 1.

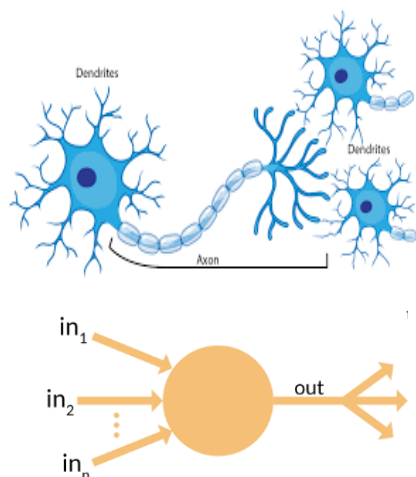


Fig. 1. The Emulation of Human Neurons and ANN

There are two main ANN topologies such as feedforward (FFNN) and feedback. The schematic of a generic feedforward network is shown in Figure 2 which is the most commonly utilized ANN model. The idea behind the use of FFNN models in PD classification is to do classifying without train the network at first.

The application of neural network system for understanding a specific time-series problem that involves determination number of layers and number of nodes in the structure which is done on experimentation basis. It has been set up that single hidden layer with an adequate

number of nodes at the hidden layer and satisfactory data for initialization. Furthermore, the determination number of input nodes which is lagged perceptions of same variable acts as an important part in modelling. Determination of output nodes is generally simple. (M. Najmi Abu Bakar, 2012) and (Mitra & Paul, 2017) recommended that model with small number of nodes at hidden layer will result in a performance of clustering that cannot be processed.

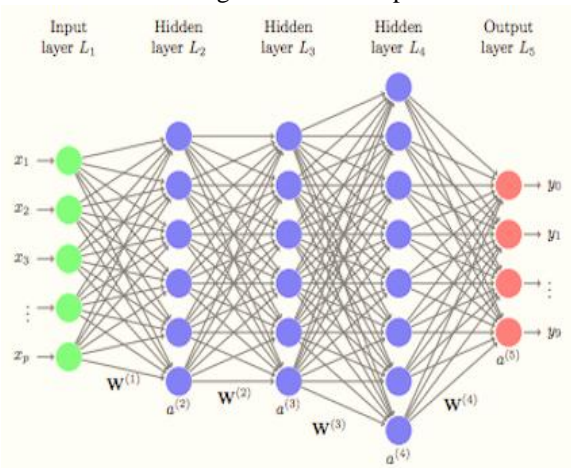


Fig. 2. Feedforward ANNs

Few authors explained that ANNs can be classified in numerous ways; possibility of it is to divide them on the term of learning guideline. A neural network uses either supervised or unsupervised learning. In supervised learning, the network is given with example cases and desired responses[12], [19], [20]. The network weights are then adapted in order to reduce the distinction between network outputs and desired outputs. In unsupervised learning sometimes it appeared in groups of the data to clusters of comparable data that the network is given as it were input signals and also the network weights modification through a predefined mechanism.

Self-Organizing Map (SOM)

The Self-Organizing Maps (SOMs) that known as Kohonen Map was first discovered by Kohonen, 1990 mostly based on unsupervised learning[20]–[22]. Some studies such proved that the SOMs from the ‘Self-Organizing’ is applied to cluster and visualize the information of data that organized on a low-dimensional consists of neurons. Meanwhile for the ‘Maps’ means it can revealed the characteristics to analyse the quality or underlying input data from the obtained mapping

Stefanovicet. al explained about that SOMs could be a collection of nodes that connected to at least one another via a rectangular or hexagonal topology[20]. The foremost topology used is rectangular topology. The rectangular topology of SOMs is displayed in Fig. 2.3 where a circle represents a node.

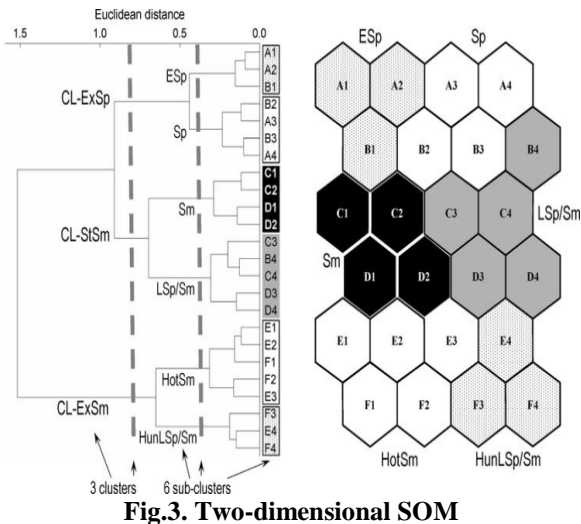


Fig.3. Two-dimensional SOM

The connections between the inputs and the nodes have weights, so a set of weights corresponds to each node. The set of weights forms a vector M_{ij} where $i = 1, 2, 3, \dots, k_x$ and $j = 1, 2, 3, \dots, k_y$ that is often called a neuron or codebook vector. In rectangular SOM, k_x is the number of rows and k_y is the number of column. The dimension of the codebook vector is the same as the number of inputs, $M_{ij} = \{m_{ij}^1, m_{ij}^2, m_{ij}^3, \dots, m_{ij}^n\}$.

III. METHODOLOGY

Before proceed further into this subtopic, below is the research flow that was designed for the PD classification process as shown in Fig. 4.

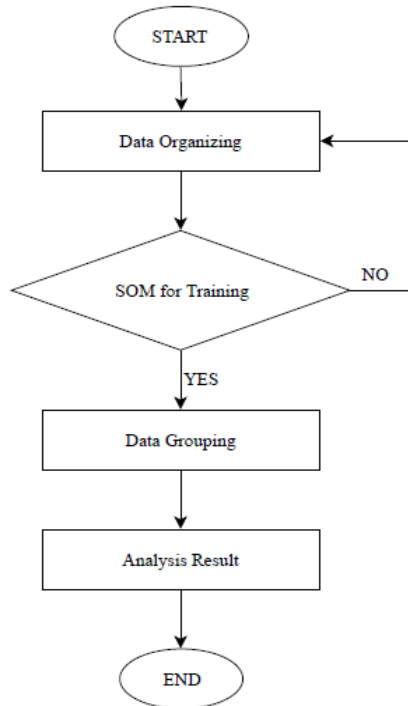


Fig.4. Research Flowchart

*Note: In **Data Organizing** phase, it involved process statistical data transformation and data organization.

3.1 Data Organizing

Data measured by HV Engineer was only one type recorded in voltage (mV). Therefore, it is not sufficient for

MATLAB Self Organizing Maps (SOM) Toolbox to do classification and clustering and usually produced a poor classification result. Value recorded in mille volt was then manipulated by using formula of charge, Q ;

$$Q = C \times V \quad (1)$$

Where Q is charges produced,

C is the capacitance of the cable and

V is the amplitude of voltage recorded.

Capacitance, C of the XLPE cable is calculated by using underground cable capacitance formula;

$$C = \frac{2\pi\epsilon_0\epsilon_r}{\ln\frac{R}{r}} \quad (2)$$

Where ϵ_0 is permittivity of the medium,

ϵ_r is the permittivity of free

space,

r is the radius of the inner conductor and

R is the radius of the sheath.

Finally, both data combine using potential energy, U formula

$$U = \frac{Q \times V}{2} \quad (3)$$

Where, Q is charge produced,

V is the amplitude voltage

recorded.

3.2 Data organizing for SOM

After the input data was selected when performing statistical preprocessing, SOM will be using the selected data which are voltage (V), charges (Q), and potential energy (U) with the suitable data organization and arrangement as shown in Fig. 5 to ensure the classification process are going to smooth and accurate. In the previous paragraph, it was informed that the only one type (millivolt) of original data gathered. Therefore, after the statistical processes were finish, the produced sets of data need to be organized in a correct mode as such it will prepare SOM with more better data arrangement for result improvement. 33 selected PD events and 55 normal situations are mixed up in the data set to determine the ability of SOM to do self-classification without being trained at first.

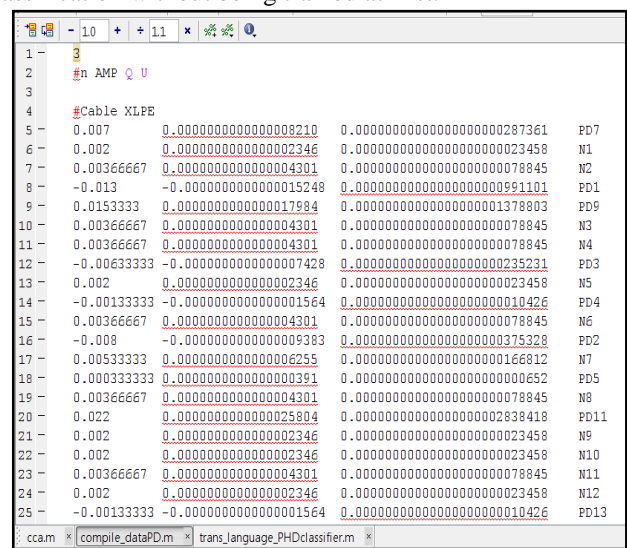


Fig. 5 Data organizing for SOM in MATLAB's m. file

3.3 Training using SOM

During the data training period, it will involve four different normalization methods. The data normalization will have formed the maps to dominate map topology. The four types of normalization methods are *var*, *range*, *log* and *logistic* and below the definition for each method;

Method Description

'var' Variance is normalized to one (linear operation).

'range' Values are normalized between [0,1] (linear operation).

'log' Natural logarithm is applied to the values: $x_{new} = \log(x-m+1)$; where $m = \min(x)$.

'logistic' Logistic or softmax transformation which scales all possible values between [0,1].

For this particular research, only 100 until 320 numbers of neuron are applied to classify small set of measurement data (88 data) for pilot testing.

3.3 U Matrix Mapping Result

After collected all the needed data from unsupervised neural network or self-learning neural network via MATLAB's SOM Toolbox result, these data are compiled in an excel to be further evaluated before the final analysis is conducted to verify the performance of the classifier.

Data collected from SOM are quantization error, topographic error, training time (sec) and map size.

IV. RESULT AND DISCUSSION

Prior to the completion of the data arrangement and organization, in this sub-topic, all results and analyses of each classification with four different normalization methods presented before the final classification is conclusive.

4.1 Var Normalization Results

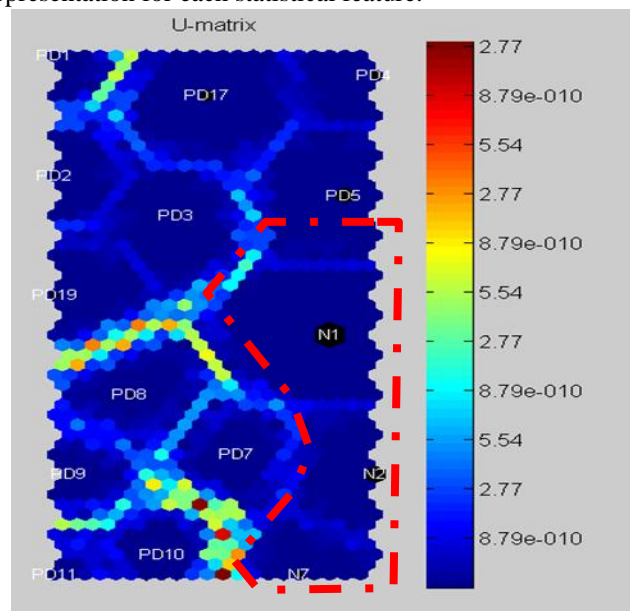
Referring to Table 1, SOM classification has shown the best result using 200, 260 and 300 neurons. For 200 neurons, it's recorded the lowest training time (~1 second) compare with the other two (260 and 300 neurons) but in terms of quantization error it is a bit higher. Hence, either 260 or 300 neurons are preferably to be choosing as the best matching neurons for 'var' normalization.

Next, when 260 and 300 neurons come head to head, both neurons evidence a fast training time at about 4 seconds. But again to produce better classification a good map sizing is needed and this proof that 300 neurons is the best choice for the normalization.

Table 1. The Compilation of 'Var' Results

No. of neuron	Classification Result			
	Map size	Quantization Error	Topographic Error	Training Time (sec)
100	[17, 6]	0.021	0.068	0
120	[17, 7]	0.014	0.034	1
140	[20, 7]	0.009	0.034	1
160	[20, 8]	0.003	0.011	1
180	[23, 8]	0.002	0.011	1
200	[22, 9]	0.001	0.000	1
220	[24, 9]	0.001	0.011	7
240	[27, 9]	0.000	0.045	3
260	[26, 10]	0.000	0.000	4
280	[28, 10]	0.000	0.080	7
300	[27, 11]	0.000	0.000	4
320	[29, 11]	0.000	0.000	3

According to Figure 6, all PD activities are separate from the normal activities. Referring to the red dotted box in the figure, group with normal conditions is group in three different characteristic but closed to each other based on the thickness of the black hexagon formed (N1, N2 and N7). This proved that using 'var' normalization method the classification process is successful. Below figure is the plane representation for each statistical feature.



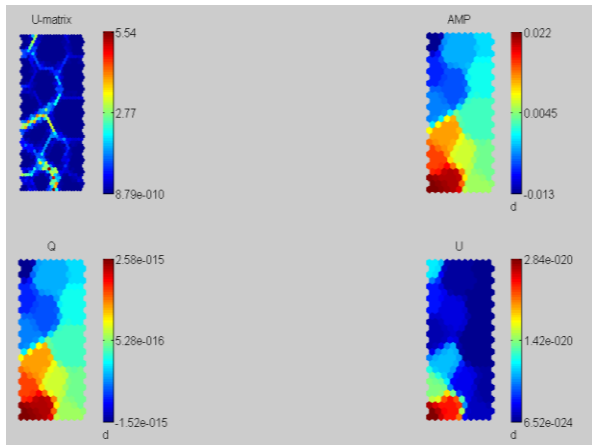


Fig. 6. U Matrix Mapping Result for 300 Numbers of Neurons with Its Plane Representation

4.2 'Range' Normalization Results

For this particular normalization method, 5 neurons shows zero for both quantization error and topographic error. To eliminate, first, the training time must be equal or less than 5 second especially for small data classification. Therefore, neither 300 nor 320 neurons qualify in this criteria and left with only three values of neurons to segregate.

Again, this situation will have to do advance comparison where map sizing will compare between these 3 neurons and 280 neurons is the biggest map size with 23 x 8 size with 4 seconds training time. This value of neurons is selected to be mapped through U Matrix to continue the classification process.

Table 2. The Compilation of 'range' Results

No. of neuron	Classification Result			
	Map size	Quantization Error	Topographic Error	Training Time (sec)
100	[17, 6]	0.002	0.057	0
120	[17, 7]	0.002	0.011	1
140	[20, 7]	0.002	0.159	1
160	[20, 8]	0.001	0.534	1
180	[23, 8]	0.000	0.000	2
200	[22, 9]	0.000	0.000	2
220	[24, 9]	0.000	0.205	2
240	[24, 10]	0.000	0.000	3
260	[26, 10]	0.000	0.534	4
280	[25, 11]	0.000	0.000	4
300	[27, 11]	0.000	0.000	5
320	[29, 11]	0.000	0.000	5

Below in Figure 7, the U Matrix result of 'range' method with 260 neurons. The same good result produced with the previous method. All normal data separated from the PD situation. But via this method, the separation is less prominent compare to 'var' method and its shows that the boundary between PD and normal are almost inseparable and this might be the least favorable method to do PD classification. Again, the plane representation of each input types shown below the 'range' overall U Matrix.

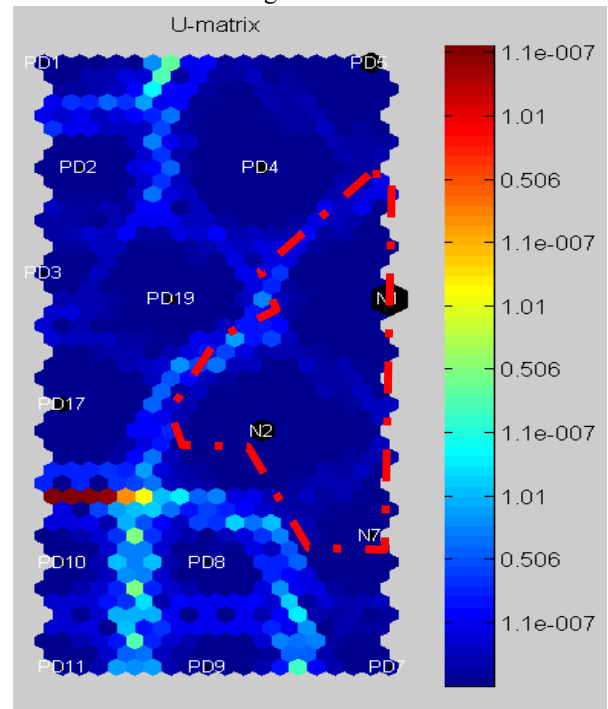


Fig. 7. U Matrix Mapping Result for 260 Neurons with Its Plane Representation

4.3 'Log' Normalization Results

For log normalization method as compiled in Table 3, the result of SOM is easily selected. For overall, all number of neurons produced near zero for both error recorded. Since, the final duo neurons happen to be shared the same value of training time (3 seconds) and quantization and topographic error (0).

Next, comparison between map sizings will determine the better option for classification. The different is on the map size and the better map sizing is on 300 neurons. This value of neurons is proceeding with U-Matrix analysis for performance verification.

Table 1. The Compilation of 'log' Results

No. of neuron	Classification Result			
	Map size	Quantization Error	Topographic Error	Training Time (sec)
100	[11, 9]	0.000	0.205	0
120	[12, 10]	0.000	0.057	1
140	[13, 11]	0.000	0.057	2
160	[13, 12]	0.000	0.057	2
180	[15, 12]	0.000	0.034	2
200	[15, 13]	0.000	0.000	3
220	[16, 14]	0.000	0.091	4
240	[17, 14]	0.000	0.000	2
260	[17, 15]	0.000	0.148	3
280	[18, 16]	0.000	0.386	4
300	[19, 16]	0.000	0.000	3
320	[19, 17]	0.000	0.466	3

As depicted in Figure 8, the mapping result exposed a bit of unsatisfactory mapping. Whereas, there are certain normal (label N7) situation data is drift away (green dotted line) to the PD activities region. But generally most of the normal data still properly divided into their activity type and it can consider almost victorious where the border formed clearly at the right data position. One additional bonus from this map is the ability to zoom in the most critical PD activity as yellow dotted box.

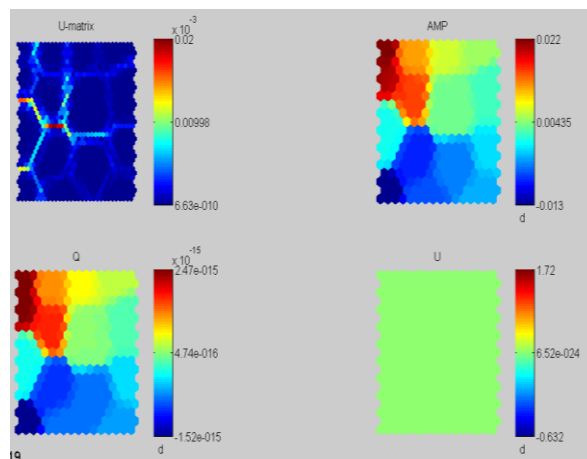
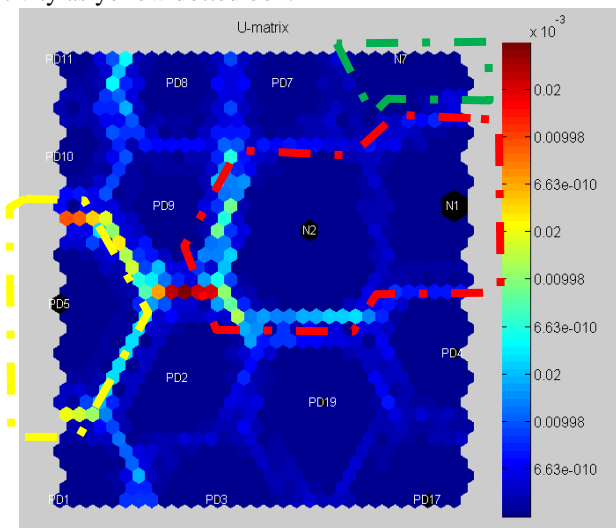


Fig. 8. U Matrix Mapping Result for 300 Neurons with Its Plane Representation

4.4 'Logistic' Normalization Results

Finally, using the 'logistic' normalization method, generally most of the number of neurons achieved zero error (quantization and topographic) except for certain value (180, 200, 260 and 300 neurons).

The fastest neurons to accomplish zero for both error is 160 neurons but with a decent small map size. Follow by 220 and 280 neurons that getting zero error but the earlier neurons produced lower training time. To single out the better, a good map sizing will determine the winner and 280 neurons selected for this method for U Matrix mapping.

Table 1. The Compilation of 'logistic' Results

No. of neuron	Classification Result			
	Map size	Quantization Error	Topographic Error	Training Time (sec)
100	[17, 6]	0.003	0.023	0
120	[17, 7]	0.002	0.011	1
140	[20, 7]	0.001	0.034	0
160	[20, 8]	0.000	0.000	2
180	[23, 8]	0.000	0.182	6
200	[22, 9]	0.000	0.386	3
220	[24, 9]	0.000	0.000	2
240	[27, 9]	0.000	0.034	4
260	[26, 10]	0.000	0.568	3
280	[28, 10]	0.000	0.000	3
300	[27, 11]	0.000	0.386	3
320	[29, 11]	0.000	0.011	7

Referring to Figure 9, the segregation of normal from PD data is quite peculiar and indistinguishable. Even though in terms of the recorded value (training time, quant and topo error) is the best but the mapping is not fit enough to discriminate the characteristic. It is proven that 'logistic' method is not suitable for PD classification.

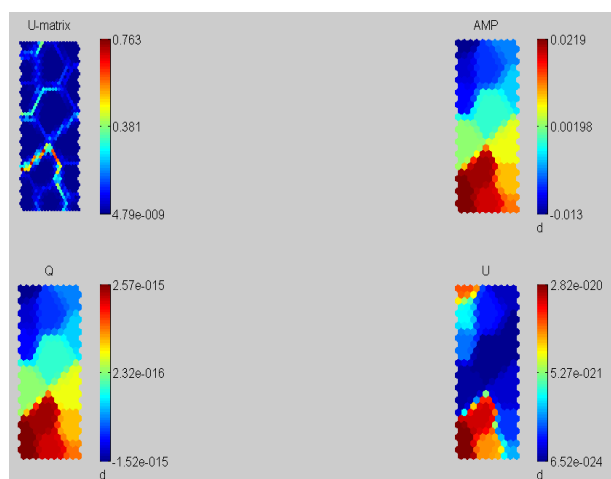
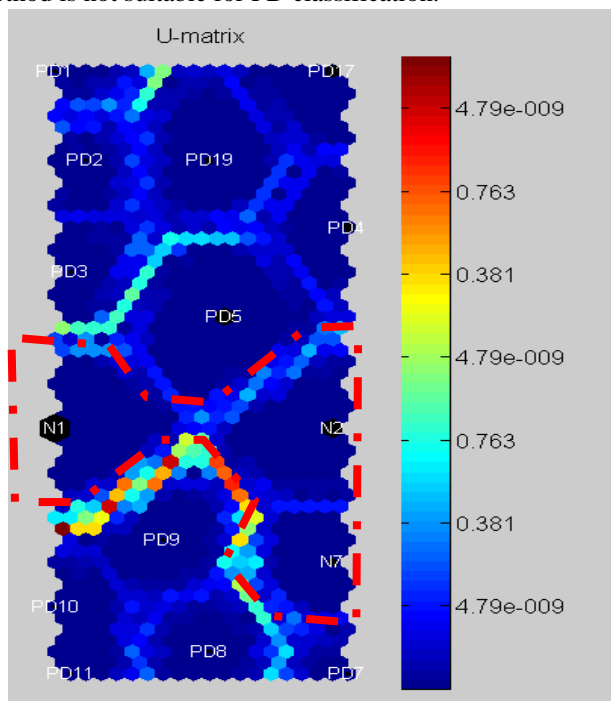


Fig. 9. U Matrix Mapping Result for 280 Neurons with Its Plane Representation

V. CONCLUSION

Overall, to conclude this research paper, it is proven that SOM has the ability to do intelligent segregation and classification without any doubt. All types of normalization method successfully determine the different characteristic of PD events even some of the method not producing concrete classification where some of normal data being group with PD data. The best classification result achieved by 'range' normalization with the boundary generated really differentiates between normal and PD data.

This SOM combine with statistical correlation method help to facilitate the future HV engineer to be effortlessly discriminating the PD data without being subjected to human erroneous decision by considering internal or

external interference as PD activity. This will ultimately save the power utility in millions of dollars in the defect correction activity and ensure the healthiness of HV equipment or cable can be sustained in a long period of time.

For future research, more data and varying statistical variation needed to be tested before conclude on the robustness of SOM with statistical correlation in classifying PD activity from raw mV measurement data gathered by power utility HV engineer.

VI. ACKNOWLEDGMENT

The authors would like to acknowledge UniversitiTeknikal Malaysia Melaka (UTeM), Universiti Malaysia Perlis (UniMAP) and Ministry of Education Malaysia for the financial support and providing the resources through for the research..

REFERENCES

1. S. M. Nor., "Privatisation and changes in organisation - a case study of a Malaysian privatised utility (BL)," 1999.
2. M. Azman, "Brief Outlook on Malaysian Electricity Supply Industry, TNB & MNC-CIGRE," 2013.
3. Suruhanjaya Tenaga Malaysia, "Peninsular Malaysia Electricity Supply," *Issn:2289-7666, St:(P)14/09/2014*, 2014.
4. Suruhanjaya Tenaga, "Energy Malaysia," *Suruhanjaya Tenaga (Energy Comm.*, vol. 6, 2015.
5. K. T. Muthanna, A. Sarkar, K. Das, and K. Waldner, "Transformer insulation life assessment," *IEEE Trans. Power Deliv.*, 2006.
6. D. H. Zhu, P. A. Crossley, Z. D. Wang, and K. J. Cornick, "Partial discharge location in power transformers," *IEE Proc. - Sci. Meas. Technol.*, vol. 147, no. 5, pp. 249–255, 2002.
7. Z. H. Bohari, H. S. Azemy, M. N. M. Nasir, M. F. Baharom, M. F. Sulaima, and M. H. Jali, "Reliable short term load forecasting using self organizing map (SOM) in deregulated electricity market," *J. Theor. Appl. Inf. Technol.*, vol. 79, no. 3, 2015.
8. D. B. Go, "Gaseous Ionization and Ion Transport: An Introduction to Gas Discharges," p. 6, 2012.
9. M. F. Rahman and P. Nigude, "A New Approach to Study Partial Discharge Inception due to Particles in Transformer Oil using Electric Field Analysis," *2019 Int. Conf. High Volt. Eng. Technol.*, pp. 1–6, 2019.
10. J. Maki *et al.*, "Construction and Verification of Novel Insulation for Next Generation Power Module," *2019 Int. Conf. Electron. Packag.*, pp. 325–328, 2019.
11. NXP semiconductors, "Application note: An 1629 UHF RFID label antenna design," pp. 1–56, 2008.
12. W. J. K. Raymond, H. A. Illias, and A. H. A. Bakar, "Classification of partial discharge measured under different levels of noise contamination," *PLoS One*, 2017.
13. Z. Ngadiron, N. H. Radzi, Z. Yassin, and I. Amin, "Review on Restructuring of Malaysia Electricity Supply Industry," *Appl. Mech. Mater.*, 2015.
14. Z. H. Bohari, M. A. M. Yusof, M. H. Jali, M. F. Sulaima, and M. N. M. Nasir, "Feature combination analysis in smart grid based using SOM for Sudan national grid," in *IOP Conference Series: Materials Science and Engineering*, 2015, vol. 100, no. 1.

15. T. Okamoto, N. Hozumi, and T. Imajo, "Partial discharge pattern recognition for four kinds of electrode systems," *Conf. Rec. IEEE Int. Symp. Electr. Insul.*, vol. 1992-June, pp. 375–378, 1992.
16. M. H. Jali, T. A. Izzuddin, Z. H. Bohari, H. I. Jaafar, and M. N. M. Nasir, "Pattern recognition of EMG signal during load lifting using Artificial Neural Network (ANN)," in *Proceedings - 5th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2015*, 2016.
17. J. Do Nascimento Cruz *et al.*, "Bio-inspired printed monopole antenna applied to partial discharge detection," *Sensors (Switzerland)*, vol. 19, no. 3, 2019.
18. M. H. Jali *et al.*, "Joint torque estimation model of SEMG signal for arm rehabilitation device using artificial neural network techniques," *ARNP J. Eng. Appl. Sci.*, vol. 9, no. 7, 2014.
19. H. P. Burgener and K. Fröhlich, "Probability of partial discharge inception in small voids," *Conf. Electr. Insul. Dielectr. Phenom. (CEIDP), Annu. Rep.*, 2001.
20. J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Trans. Neural Networks*, 2000.
21. F. M. Riese and S. Keller, "SUSI: Supervised Self-Organizing Maps for Regression and Classification in Python," 2019.
22. S. Coenen, S. Tenbohlen, and S. M. Markalous, "Detection of Partial Discharges in Power Transformers using UHF PD Measurements and Acoustic Measurements."

AUTHORS PROFILE



Zul Hasrizal Bohari was graduated with Master of Electrical Engineering from UNITEN in 2013 and Bachelor Degree in Electrical and Electronic Engineering from UKM back in 2007. The author has published more than 20 journals in various publisher and started to do research works related to electrical engineering and artificial intelligence (AI) since 2011. Among the achievement is being awarded few local and international awards on new innovation related to energy and artificial intelligence (AI).



Muzamir Isa was graduated with D. Sc of HV Engineering from Aalto University in 2013 and Bachelor Degree in Electrical and Electronic Engineering from UKM back in 2002. The author has published more than 100 journals in various publisher and started to do research works related to high voltage engineering since 2013. Among the achievement is being awarded few local and international awards on new innovation related to HV engineering.