MULTILAYER PERCEPTRON PREDICTION MODEL FOR TRANSFORMER HEALTH INDEX MONITORING

Nazrul Fariq Makmor
Faculty of Engineering, National Defence University of Malaysia

Afzan Zamzamir
Faculty of Engineering, National Defence University of Malaysia

Ja’afar Adnan
Faculty of Engineering, National Defence University of Malaysia

Azharudin Mokhtarruddin
Faculty of Engineering, National Defence University of Malaysia

他

https://doi.org/10.5109/7395740

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 11, pp.1745-1750, 2025-10-30. International Exchange and Innovation Conference on Engineering & Sciences
バージョン：

MULTILAYER PERCEPTRON PREDICTION MODEL FOR TRANSFORMER HEALTH INDEX MONITORING

Nazrul Fariq Makmor[1], Afzan Zamzamir[2], Ja'afar Adnan[3], Azharudin Mokhtarruddin[4], Yasotharan Vusuvanathan[5], Yulni Januar[6]
[1,2,3,4,5]Faculty of Engineering, National Defence University of Malaysia, Sg. Besi Camp, 57000 Kuala Lumpur, Malaysia
[6]Toshiba Transmission & Distribution Systems Asia Sdn Bhd, Petaling Jaya, Selangor, Malaysia
nazrulfariq@upnm.edu.my

Abstract: *Artificial neural networks (ANNs) emerged through computerized models of human biological operations which focus on simulating neurons. Researchers have extensively applied ANNs to different problems through their extensive study in electrical engineering domains. Dissolved gas analysis (DGA) serves transformer applications by distinguishing between operational and maintenance-required transformers. The main objective of DGA consists of identifying transformer faults that stem from various gas components detected inside the equipment. The key gas method (KGM) analysis serves as a common approach in DGA for transformer health index assessment by evaluating gases that emerge inside the equipment. The researcher designed an MLP architecture as the basic structure while implementing Bayesian Regularization (BR) in combination with Lavenberg Marquardt (LM), and Backpropagation (BP). for training techniques. MLP network uses Sigmoid activation to become operational. The MLP network trained with BR reached 94.44% accuracy as its best achievement among all models.*

Keywords: transformer; dissolve gas analysis; key gas method; multilayer perceptron

1. INTRODUCTION

Power transformers function as essential elements of power grids while representing a major financial commitment in overall power delivery systems. Power disruption caused by transformer malfunctions leads to major financial losses for plant owners because they unable to create electricity. To reduce these risks inspections must happen regularly and transformers need periodic maintenance [1-2]. Transformers that fail will create multiple dangerous circumstances that generate noise pollution while causing power outages and possibly result in transformer fires that produce smoke [3]. Transformers require optimal operation for preventing these kinds of problems. The continuous rise in electricity demand has pushed power transformers into maximum operational range so their potential failures or malfunctions increase.

2. LITERATURE REVIEW

A. Transformer Health Index

The transformer health index functions as a numerical evaluation system for measuring transformer operational efficiency and state. Different factors consisting of temperature and oil quality and insulation condition and winding resistance and extra diagnostic tests need to be analyzed as part of this process [4]. Researchers use transformer characteristics to evaluate its total health condition which allows them to determine its estimated operational life. The assessment system allows quick identification of potential faults so organizations can organize maintenance or replacement activities on schedule. The transformer Health Index (HI) functions as an integrated measurement system which combines operational data from field inspections and laboratory tests to help make proper asset management decisions about transformers [5]. Using this decision-making process transformer operators can maintain their equipment in its best operational state as stated in the referenced material.

Many researchers have focused on discovering which characteristics affect the transformer health index the most. Both data source levels and diagnostic procedures act as essential elements to duplicate primary factors that generate uncertainty. The data collection process shows influence from measurement uncertainties that generate multiple errors between inaccurate measurements and quantization errors. The common imperfections found in data exist within the easily accessible data sources. The uncertainty that affects diagnostic models remains independent from the source between knowledge-based errors and random occurrences. The principle applies to every form of uncertainty no matter what type of uncertainty exists. The framework of the transformer health index appears in Fig. 1 which displays the implementation and development stage.

The transformer health index undergoes changes because of multiple sources of uncertainty which operate within each transformer subsystem. The enhanced decision-making process becomes suitable for measurement and process uncertain systems while delivering practical applications. The system proves effective for making decisions which are unclear or ambiguous. Industrial facilities implement DGA as their standard practice to examine power transformers [6]. The evaluation method tracks and analyzes dissolved gases present in transformer insulating oil because these gases form due to internal defects and irregularities found inside transformers. Regular transformer diagnostics use the DGA as an analysis method [7]. Numerous previous implementations have proven DGA to be successful [8].

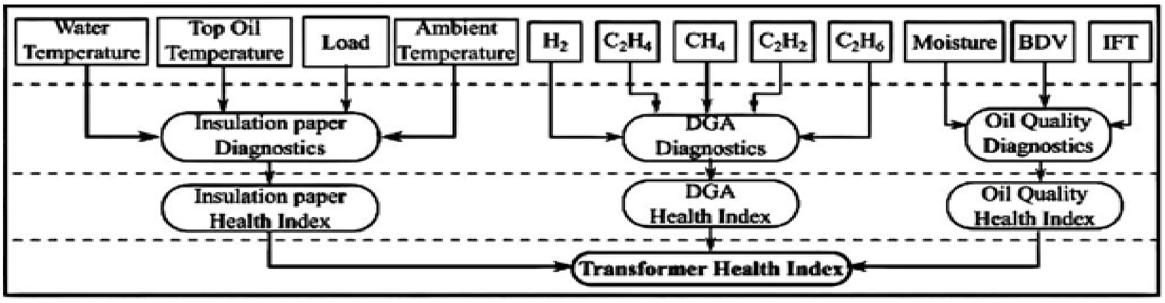


Fig. 1. Transformer health index soft computing framework [8].

B. Multilyer Perceptron Network (MLP)

An ANN functions as a computer network which utilizes neural stem cells extracted from human brain cells and specifically uses neurons as fundamental components [9]. The main structural elements of the human brain consist of tiny cells which we call neurons. The structure and functionality of neurons guide ANNs to re-create the capabilities and creative processing characteristics that the human brain possesses [10]. Mathematical modeling allows ANNs to function in non-parametric grading/ regression and data classification and nonlinear function computation. ANN systems reproduce biological models of human neurons to generate accurate predictive results through training processes. The capability and operational speed of ANNs should match those observed in brain function to be considered alternatives. These systems act as biological process replacements. The MLP network stands among the most commonly employed network architectures for multiple practical uses [11–12]. Rosenblatt invented the Perceptron artificial neuron model while conducting his research in 1958 [7]. The network design included multiple layers which he depicted in Fig. 2.

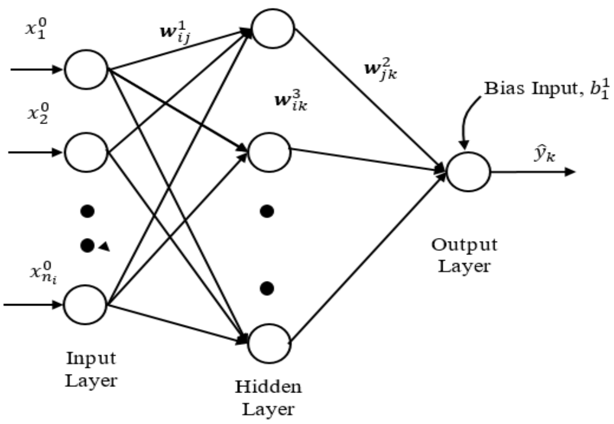


Fig. 2. The structure of the MLP network [13].

Different sections throughout the study use artificial neural networks (ANNs) under the name MLP networks. A multilayer perceptron network comprises three layers which are illustrated in Fig. 1 showing the input layer, hidden layer and the output layer.

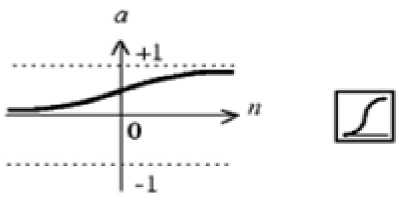
$$\sum_{j=1}^{n_h} w_{jk}^2\, \partial\left(\sum_{i=1}^{n_i} w_{ij}^1 x_i^0(t) + w_{k0}^1 b_j^1\right) \qquad (1)$$

$for\ 1 \leq j \leq n_h$ and $1 \leq k \leq m$

The weight that connects the MLP network's input layer to the hidden layer itself is shown in Eq. 1 as w_{ij}^1. The weights w_{jk}^2 and w_{jk}^2 reflect the connections between the hidden and output layers of the MLP network, respectively. In Eq. 1, the input parameter, indicated as x_i^0, is feeding the MLP network through the input layer, whereas b_j^1 is the threshold of the hidden node. The MLP network, indicated as ∂ (·) in Eq. 1, has been activated in this work using the Tansig activation function. Eq. 1 can be used to find the values of w_{ij}^1, w_{jk}^2 and b_j^1 based on a suitable technique and put at the minimal vector but allow for recurrence at each iteration.

The weight between the input layer nodes of MLP networks and hidden layer nodes is represented by w_{ij}^1 in Eq. 1. The weights w_{jk}^2 and w_{jk}^2represent the hidden layer to output layer connections in the MLP network. The MLP network receives values from x_i^0 , through b_j^1 according to Eq. 1 which includes the hidden node threshold value. This study employs Tansig activation during MLP network operations while using the designation ∂(·) as per the description in Eq. 1. Eq. 1 serves as a method to calculate the w_{ij}^1 , w_{jk}^2 and b_j^1 values using proper techniques which enable recurrences during each iteration to minimize vector values.

The MLP network receives training through three different algorithms which include BP, LM and BR. BR possesses stochastic tabulation as its method while BP and LM use deterministic tabulation for their operations. Logsig transfer function exhibits bipolar sigmoid behavior by producing values between 0 and +1. Activation functions needed for input data or signal selection appear as illustrated in Fig. 3. The data threshold should be expressed anywhere between 0 and +1 in the incoming signal.



a = logsig(n)

Fig. 3. Log-sigmoid activation function [14].

C. Bayesian Regularization (BR) Algortihm

The training and learning technique created by Thomas Bayes is known as D. Bayes' rule. The dataset D enables the application of Bayes' rule to compute θ which medical practitioners call posterior probability [8]. The result of posterior probabilities from Bayes rule contains every possible value within its range. The development of Bayes' rule started when:

$$p(\theta|D) = \frac{p(D|\theta)}{p(D)} \tag{2}$$

Before taking data probability into account, likelihood of data serves as $p(\theta|D)$, then the initial parameter probability emerges as $p(\theta)$. The general idea led to modifications for the probability of distributions in MLP network weight parameters. During training the MLP network acquires its weights from the sample data $p(w|D)$ and represents these weights as w. The depiction of the posterior distribution appears through:

$$p(w|D) = \frac{p(D|w)p(D)}{p(D)} = \frac{p(D|w)}{\int p(D|w)\,p(w)dw} \tag{3}$$

Fig. 3 represents the process through which the BR training algorithm adjusted weight values to modify belief systems according to the Bayesian equation from $p(w)$ through $p(D|w)$.

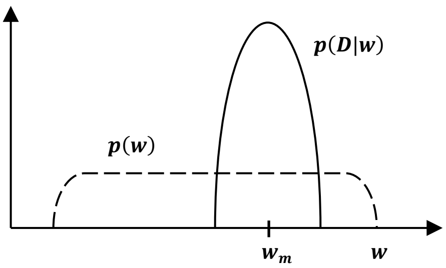


Fig. 4. Change pre weights to posterior/post weights [12].

D. Levenberg Marquardt (LM) Algorithm

A specific gradient table serves as the fundamental operational component of the LM training algorithm. The Levenberg-Marquardt algorithm functions to improve the performance of BP algorithm. The MLP network receives LM algorithm training evaluation through comparison against standard BP training. The predictive performance of the LM training algorithm goes beyond the traditional BP training algorithm while also providing slightly faster convergence results. The relative stability remains intact [15]. The LM training algorithm uses the quasi-Newton approach to develop a second-order training process without the necessity to calculate Hessian matrix components. The LM function results from summing squares while its estimated Hessian matrix contains the following definition:

$$H = J^T J \tag{4}$$

and the gradient of the matrix is calculated as:

$$g = J^T \rho \tag{5}$$

The Jacobian matrix J calculates network error through biases and weights of the network. The array can be computed by applying conventional BP methods since these methods require less computational complexity than Hessian matrix calculations [15]. The LM algorithm uses an equation estimator based on Newton-like methods to modify equations contained within the Hessian matrix according to the following formula:

$$\Delta w = -[J^T J + \mu I]^{-1} J^T \rho \tag{6}$$

The updated weight of the estimator Δw depends on the value of μ parameter. The Hessian matrix undergoes updates following the Newton-like equation to prevent the zero value of this parameter. The Newton estimator-based BP training algorithm improves results when it enhances BP algorithm momentum. The BP training algorithm becomes more effective at finding minimal error local minima due to the increased momentum. The algorithm performs an effective move to reduce the error at every step. As the step function develops the error continues to escalate. The minimum error value needs to decrease during every processing cycle [15].

E. Back propagation (BP) Algorithm

Applications of BP training method extend throughout various control systems together with ANNs [16]. Measurements of brain signals through electrical activity tend to be linked with heart signals or electrical activity detected in ECGs. The placement of electrodes takes place across hand, foot and chest areas to detect such signals. The acquired electrical signals along with their measured impulses become the training base for the BP algorithm. The training process of the algorithm conducts operations through the steepest descent method. The BP approach utilizes learning phase data either as weights or protocols to obtain derivative values for performing effective weight adjustments [8]. Through this process BP modifies the weight between the jth hidden layer neuron and the *i*-output layer neuron directly.

$$w_{ji}(t) = w_{ji}(t-1) + \Delta w_{ji}(t) \tag{7}$$

and

$$b_j(t) = b_j(t-1) + \Delta b_j(t) \tag{8}$$

The updated weight, $\Delta w_{ji}(t)$ and $\Delta b_j(t)$ given by:

$$\Delta w_{ji}(t) = \eta_w \gamma_j(t) x_i(t) + \alpha_w \Delta w_{ji}(t-1) \tag{9}$$

and

$$b_j(t) = \eta_b \gamma_j(t) + \alpha_b \Delta b_j(t-1) \tag{10}$$

The subscripts w and b in Eq. 7 and Eq. 8 respectively denote the weight and threshold of the steepest decent estimator. The constants α_w and α_b regulate the likelihood of the previous parameter to the present benchmark. The likelihood momentum must be carefully chosen. BP won't be able to draw lessons from the past if the momentum is set too high. If it is adjusted too low, however, the BP will continue to behave as it did previously. The BP training algorithm's η_w and η_b learning rates are noted, whereas $\gamma_j(t)$ represents an error that occurs at the jth neuron. The error afterwards changed into an input and returned to the weight. Eq. 11 indicates error at the output node on the assumption that the activation function is functioned in a linear parameter:

$$\gamma(t) = y_k(t) - \hat{y}_k(t) \tag{11}$$

where $y_k(t)$ is the predicted output, and $\hat{y}_k(t)$ is the desired output. The production at the hidden layer is

$$y_j(t) = F'\big(x_i(t)\big) \sum_j \rho_j^k(t)\, w_{jk}^2(t-1) \quad (12)$$

where $F'\big(x_i(t)\big)$ is derived from $F\big(x_i(t)\big)$ by respecting to $x_i(t)$. As a steepest descent approach, the BP algorithm suffers from a poor convergence speed or rate. The pursuit of a global minimum may lead to local minima. It responds similarly to user-selected parameters [8].

3. METHODOLOGY

A. Data Collection

This research method integrates KGM within a system that analyzes transformer data obtained from DGA. The data organization follows the conditions specified in Table 1. The Total Dissolved Gas Concentration (TDCG) calculation determines the conditions by adding gas concentrations specified in Table 2. The TDCG selects appropriate criteria from its collected data to structure information. The data sorting process uses the values derived from the TDCG stage. Through data from dissolved gas analysis and furan testing and oil testing the ANN determined the health condition of transformers which operated optimally. Malaysia's power provider company uses a complete approach to record transformer oil gas concentrations in great detail. The current research utilized ANN to establish health index determination for power transformers. The Klang Valley contains fifty transformers that serve as the basis.

The presented methodology demonstrates both accurate gathering techniques and powerful analytical tools as essential elements for transformer management enhancement. The methodology should be expanded through future investigations by using combined diagnostic instruments such as vibration analysis and thermal imaging for comprehensive transformer health assessment. Predictive analytics integration enables the system to predict upcoming failures in advance which helps maintain transformers proactively. The development of these new technologies will make power distribution systems more reliable and efficient thus resulting in better service quality for consumers.

B. Pre Processing Dataset

The research methodology of this study integrates the KGM into a DGA data analysis system that was developed specifically for transformer data assessment. The data organization followed the parameters listed in Table 2. The TDCG calculation performs the parameter identification process by adding together the gas concentrations listed in Table 3. The TDCG value functions as an essential data arrangement criterion that permits the data to direct this process. The TDCG results guide the data sorting process to achieve accurate classification of information.

The threshold concentration value appearing in Table 3 functions as TDCG performs as the essential reference standard to segment obtained data into four specific categories. The transformer passing all requirements without any deviations is categorized under "Condition 1" indicating its healthy condition. The second condition involves transformers that possess minor imperfections yet maintain operation capabilities while operators need to track these flaws. A transformer classified as "Condition 3" shows signs of less favorable health since specific problems have been identified which need maintenance before the end of six months. The condition referred to as "Condition 4" requires urgent transformer repairs within three months because it incorporates substantial defects that need speedy intervention to prevent additional deterioration.

The classification system helps both transformer health monitoring and maintenance scheduling through its systematic strategy so that critical matters receive prompt attention for operational reliability and safety needs.

Table 1. Gasses invalved in KGM

Type of gas	Formula
Hydrogen	H_2
Methane	CH_4
Acetylene	C_2H_2
Ethylene	C_2H_4
Ethane	C_2H_6
Carbon Monoxide	CO
Carbon Dioxide	CO_2

Table 2. Limit assigned as input

TDCG limit	Condition
720	1
721-1920	2
1921-4630	3
>4630	4

4. RESULT & DISCUSSION

The research used Matlab software while adopting the same analytical techniques from studies [16-17]. The technique presents two distinct ways to interpret data classification along with prediction tasks. The research first established the best setup for MLP network structures. The analysis established the MLP network would reach its maximum configuration with 10 hidden nodes. The analysis determines the best number of iterations required for the MLP network operation. The time required for calculations expands because of the complex design elements which include the chosen structure together with training algorithms and activation functions [18].

Table 2. MLP optimum structure

Training algorithm	Num. of iteration
BR using Logsig	312
LM using Logsig	27
BP using Logsig	23

Table 3. MLP optimum structure

Training algorithm	Accuracy performance (%)		
	Training	Testing	Overall
BR using Logsig	94.12	94.76	94.44
LM using Logsig	92.48	92.25	92.37
BP using Logsig	87.52	86.27	86.90

BR with Logsig delivered the highest accuracy performance obtained 94.12% and 94.76% of training and testing value respectively where it be the most effective approach.. Followed by LM algorithm using Logsig as the second highest accuracy performance where it yielded 92.48% for training value and reached 92.25% for the testing. The lowest performance of training algorithm is

BP with Logsig where it provided the performance value of 87.52% and 86.27% for both training and testing performance.

Among all models of the MLP network structures demonstrated 94.44% accuracy which stands as the highest prediction overall accuracy rate according to Table 3. The precision value of the MLP model indicated minimal significant prediction errors with 86.90% precision. The LM using Logsig placed in second position among the training algorithm after securing 92.37% accuracy. The BR using Logsig exhibited superior performance than the LM using Logsig which placed second below the BR.

5. CONCLUSION

The main purpose of classifier evaluation is to determine how models predict the transformer health index across different categories and classification groups [19-20]. Each of the examined prediction models showed a capability to deliver strong results in terms of accuracy together with accuracy performance analysis. The BR with Logsig prediction model surpasses all other models by accomplishing an accuracy of 94.44% for overall accuracy value while 94.12% and 94.76% for testing and training value obtained respectively. The BP training algorithm's model achieved the worst results among all prediction models because it returned to 86.90% of overall accuracy, 87.52% training result and testing value of 86.27%. All training algorithm prediction models demonstrate equal performance according to the determined results which establish that high classification performance is possible through any model. Additional prediction models including LM and BP need assessment to establish improved performance capability. A larger number of datasets available for the classifiers would enhance their prediction capability and generalization ability.

The research discoveries establish substantial grounds for creating improved maintenance plans that enhance both power system reliability and system-wide performance results. The research develops an improved diagnostic approach which uses a MLP network classifier inside DGA to enhance transformer fault detection accuracy. The MLP network enables performance assessment of power transformers to determine their health index as part of the current operational methodology. Future studies should research advanced machine learning approaches which extend beyond MLP network usage to diagnose transformers accurately by examining deep learning algorithms and ensemble methods. The analysis of these methods on extensive datasets will demonstrate their capacity as well as their strength for real-world power transformer diagnosis.

Further research should work to create a complete real-time monitoring framework which unites the diagnostic abilities of the MLP classifier with IoT technologies. This system would enable ongoing transformer assessment which enables predictions for maintenance before outages occur to preserve critical infrastructure and minimize operational interruptions.

The development of predictive maintenance strategies would benefit from adding environmental elements to diagnostic models and investigating their performance effects on transformer operations. The study provides researchers with numerous potential directions to explore which will enhance power system reliability as well as efficiency.

6. REFERENCES

[1] M. V. A. J. Wang, A. J. Vandermaar, K. D. Srivastava, Review of condition assessment of power transformers in service. IEEE Electrical Insulation Magazine, 18(6) (2002) 12-25.

[2] A. E. Abu-Elanien, M. A. A. Salama, Asset management techniques for transformers. Electric Power Systems Research, 80(4) (2010) 456-464.

[3] A. Christina, M. A. Salam, Q. M. Rahman, F. Wen, S. P. Ang, W. Voon, Causes of transformer failures and diagnostic methods–A review. Renewable and Sustainable Energy Reviews, 82 (2018) 1442-1456.

[4] G. C. Jaiswal, M. S. Ballal, P. A. Venikar, D. R. Tutakne, H. M. Suryawanshi, Genetic algorithm–based health index determination of distribution transformer. International Transactions on Electrical Energy Systems, 28(5) (2018) e2529.

[5] A. Jahromi, R. Piercy, S. Cress, J. Service, W. Fan, An approach to power transformer asset management using health index. IEEE Electrical Insulation Magazine, 25(2) (2009) pp.20-34.

[6] S. S. Ghoneim, I. B. Taha, A new approach of DGA interpretation technique for transformer fault diagnosis. International Journal of Electrical Power & Energy Systems, 81 (2016) 265-274.

[7] K. D. Patekar, B. Chaudhry, DGA analysis of transformer using Artificial neutral network to improve reliability in Power Transformers. In 2019 IEEE 4th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), (2019) 1-5.

[8] J. I. Aizpurua, B. G. Stewart, S. D. McArthur, B. Lambert, J, G. Cross, Towards a comprehensive DGA health index. In 2018 IEEE 2nd International Conference on Dielectrics (ICD), (2018) 1-4.

[9] N. H. Mohd Nor, N. I. Mansor, N. A. Hasim, Artificial neural networks: A new frontier in dental tissue regeneration, Tissue Engineering Part B: Reviews, (2024).

[10] A. Shvarts, R. Bos, M. Doorman, P. Drijvers, Reifying actions into artifacts: process–object duality from an embodied perspective on mathematics learning, Educational Studies in Mathematics, 117(2) 2024.

[11] A. K. Al-Mhdawi, H. S. Al-Raweshidy, A smart optimization of fault diagnosis in electrical grid using distributed software-defined IoT system. IEEE Systems Journal, 14(2) (2019) 2780-2790.

[12] A. H. F. S. A. Jamil, J. A. Kadir, J. M. Jamil, F. R. Hashim, S. Shaharuddin, N. F. Makmor, Multilayer Perceptron Optimization of ECG Peaks for Cardiac Abnormality Detection. In 2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE), (2022) 37-41.

[13] S. Tangirala, Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. International Journal of Advanced Computer Science and Applications, 11(2) (2020) 612-619.
[14] S. H. F. S. A. Jamil, A. R. Alias, A. R., M. T. A. Rahman, F. R. Hashim, S. Shaharuddin, M. S. M Sabri, Cardiac abnormality prediction using Logsig-based MLP Network. In 2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE) (2022) 42-46.
[15] L. Sun, Y. Liu, B. Zhang, Y. Shang, H. Yuan, Z. Ma. An integrated decision-making model for transformer condition assessment using game theory and modified evidence combination extended by D numbers. Energies, 9(9) (2016) 697.
[16] T. Mamee, N. Ditnin, P. Sripadungtham, N. Nanthawirojsiri, Short-Term Photovoltaic Power Forecasting Using a Neural Network Dynamic Time Series. International Exchange and Innovation Conference on Engineering & Sciences, (2024).
[17] F. R. Hashim, S. H. F. S. A. Jamil, J. A. Kadir, N. S. Hasan, B. Mustapha, Y. Januar, Tansig Based MLP Network Cardiac Abnormality. In 2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), (2019) 199-203.
[18] J. Galupino, J. Dungca, J, Estimation of Permeability of Soil-Fly Ash Mix using Machine Learning Algorithms. International Exchange and Innovation Conference on Engineering & Sciences, (2023).
[19] F. R. Hashim, Y. Januar, A. Zamzamir, J. Adnan, M. T. Ishak, Health Index of Transformer's Monitoring using Artificial Intelligent. Zulfaqar Journal of Defence Science, Engineering & Technology, 6(2) (2023).
[20] S. Li, X. Li, Y. Cui, H. Li, Review of transformer health index from the perspective of survivability and condition assessment. Electronics, 12(11) (2023) 2407.