ZULFAQAR Journal of Defence Science, Engineering & Technology
e-ISSN: 2773-5281
Vol. 8, Issue 1 (2025)
DOI: https://doi.org/10.58247/jdset-2025-0801-07
Journal homepage: https://zulfaqarjdset.upnm.edu.my

LEVENBERG MARQUARDT AND BAYESIAN REGULARIZATION TRAINING ALGORITHM BASED MLP NETWORK PREDICTION FOR SHAPE AGGREGATE

Ja'afar Adnan[a], Mohd Salman Mohd Sabri[a], Syahrull Hi-Fi Syam Ahmad Jamil[b], Shanusi Ahmad[c], Nazrul Fariq Makmor[a*]

[a] Department of Electrical & Electronic Engineering, Faculty of Engineering, National Defence University of Malaysia, Sg. Besi Camp, 57000 Kuala Lumpur, Malaysia
[b] Department of Mathematics and Science Computer, Politeknik Tuanku Syed Sirajuddin, 02600 Arau, Perlis, Malaysia
[c] School of Education, College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia

ARTICLE INFO

ARTICLE HISTORY
Received : 01-07-2024
Revised : 15-10-2024
Accepted : 20-01-2025
Published : 31-05-2025

KEYWORDS
Aggregate
Multilayer Perceptron
Training Algorithm
Mean Square Error
Regression

ABSTRACT

The assessment of aggregate quality is predicated on a combination of manual grading and mechanical filtering through established traditional methods. Aggregates are mandated to undergo a series of mechanical, physical, and chemical testing protocols to verify their compliance with predefined standards. The manual evaluation processes are intrinsically inefficient and subjective, resulting in significant temporal resource expenditure. This project aims to design an image processing system capable of categorizing aggregates into distinct classifications. The classification system leverages an artificial neural network (ANN) to perform image analysis for the identification of aggregate morphologies. The study systematically evaluates the performance of various training algorithms for the ANN, specifically juxtaposing Levenberg Marquardt (LM) against Bayesian Regularization (BR) as training paradigms. The findings demonstrate that BR training outperforms other methodologies, as evidenced by superior mean square error (MSE) metrics and improved regression results. The integration of the BR training algorithm with a multilayer perceptron (MLP) network achieves optimal performance in terms of regression accuracy and MSE assessment. Through the implementation of BR training, the network attained an MSE of 1.2042 and a regression coefficient of 0.9892, thereby validating its capability to classify aggregates through image analysis effectively. This novel approach provides researchers with a robust and objective solution that supersedes traditional manual classification methodologies.

1.0 INTRODUCTION

A crucial ingredient in the creation of concrete is aggregate. The two most popular types of rocks used to generate aggregates are still granite and limestone. Form, size, and surface roughness of the aggregate are crucial in the production of high-strength concrete. Features including the kind and degree of stratification of the rock deposit, the type of crushing plant used, and the size reduction ratio all have a significant impact on the form of aggregate particles and the quality of freshly poured and curing concrete [1-2]. It has been shown that a key factor in enhancing the shape is reducing the water to cement ratio necessary to produce a concrete mixture. They found that the price of producing and pouring concrete can be decreased by using this high-quality aggregate.

Good aggregates and bad aggregates are frequently used as categorizations for aggregates. The two forms of good aggregate are angular and cubical, while the four categories of poor aggregate are elongated, flaky, flaky & elongated, and irregular [3]. According to British Standards BS812, Part 103.1 [4], the traditional methods for determining the size and shape of coarse aggregates include mechanical sifting and

*Corresponding Author | Makmor, N. F. | nazrulfariq@upnm.edu.my

manual gauging. The sieving process, also known as "grading analysis," is susceptible to errors because different particle morphologies might occur. The traditional categorization approach has been improved by the development of several techniques involving imaging tools and analytical processes to quantify aggregate dimensions.

Machine vision systems for aggregate categorization are used by Murtagh et al. (2005) and Singh and Rao (2005), respectively [5-6]. These systems are made to operate instantly. In general, the two key phases of the systems are classification and image processing. While the classification stage establishes the kind or calibre of aggregate, the image processing stage extracts significant aggregate features. To evaluate aggregate granularity, Murtagh et al. (2005) proposed a machine vision system based on a multiple-scale image entropy generated from a given image [5]. Based on their visual texture, which differs depending on the mineral content [6], Singh and Rao (2005) categorised ore particles. In the created system, an image processing technique in the RGB colour space is used to retrieve the ore particles' visual texture. For classification, the Radial Basis Function (RBF) neural network uses second-order statistical analysis, such as entropy, contrast, energy, and homogeneity, as well as first-order statistical analysis, such as grayscale values. The manganese, iron, alumina, and aggregate zones are distinguished from one another based on differences in the values of grayscale, entropy, contrast, energy, and homogeneity for each region.

Analytical tools like ANNs are incredibly effective in solving difficult and non-linear problems, outperforming other competing techniques (fuzzy logic, evolutionary algorithms, and statistical methods). The ability of ANNs to generalise beyond training data and learn from instances is what has made them so popular. Because they are immune to the "curse of dimensionality" and have a low computing cost by utilising a large amount of data and a lot of dimensions, ANNs are competitive in classification in data mining. A few of the fields where ANNs have been successfully used include pattern recognition and classification, signal and image processing, robot control, weather prediction, financial forecasts, and medical diagnostics. Radial Basis Function (RBF) and MLP are two examples of ANN architectures for pattern categorization problems that have been proposed in the literature [7-8]. The MLP structure is the most well-known and often applied of all of these.

Since to its computational ease, finite parameterization, stability, and smaller structure size for a given problem when compared to other structures, the MLP is well-liked. A straightforward technique that provides a good approximation of any input-output mapping is the MLP [7]. The neural network models are very non-linear with respect to the unknown parameters. The drawback of this characteristic is that it calls for the use of a non-linear optimisation technique, which is frequently associated with problems like slow parameter convergence, demanding computation, and undesirable local minima.

2.0 METHODOLOGY

There was a total of 625 aggregate photos, of which 425 showed good shapes and 200 showed bad shapes. Quality and contrast are enhanced using pre-processing techniques, and a features extraction tool is employed to find key data for categorization. The image is automatically segmented using an iterative thresholding procedure, followed by expanding and shrinking methods, to produce a clearer and better separation between object and backdrop [9]. One of the most challenging problems in this endeavour is the use of geometrical moments for feature extraction in aggregate form classification during the feature extraction stage. The Hu and Zernike moments' invariance property against geometrical changes like scaling, translation, and rotation makes them a promising candidate feature extractor for group recognition. Two sets of seven Hu were obtained, one from the region and the other from the border, based on this reasoning. An artificial intelligence called ANN is modelled after the way the brain works.

The ANN is based on principles found in the human brain and is intended to mimic the way the brain constructs its structures, learns, and operates. The model of nonlinear neurons is shown in Figure 1. Figure 1 depicts the genesis of a neuron as consisting of a network of synapses or connections, a sum, and an activation function. A weighted value is assigned to each neuron's synapse. Assuming the neuron has k synapses, it has k inputs. The model's activation function is represented by $\partial(\cdot)$, the input at each synapse by $(\mathrm{x}_1, \mathrm{x}_2 \,..\, \mathrm{x}_\mathrm{k})$, and the weight at each synapse by $[\mathrm{w}_1, \mathrm{w}_2 \,..\, \mathrm{w}_\mathrm{k}]$. The value of the j^{th} synaptic weights $[\mathrm{w}_\mathrm{j}]$ influences the weight value for processing the synapses to the neuron's output. At the input synapses attached to the neuron, the value of the j^{th} synaptic weights $[\mathrm{w}_\mathrm{j}]$ will be multiplied by the input x_j. The activation function gets a sum process' output and adds up all the multiplied input signals and bias (b). The modelling of neurons can be defined using the two equations below based on Figure 1:

*Corresponding Author | Makmor, N. F. | nazrulfariq@upnm.edu.my

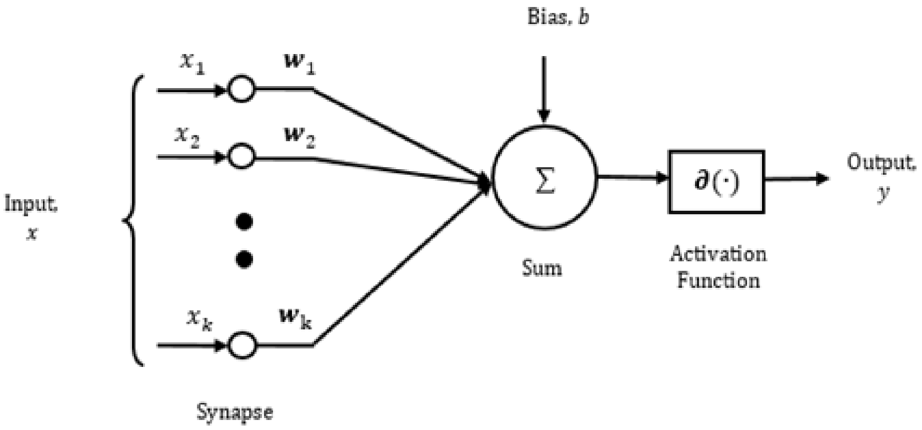

Figure 1. Nonlinear neuron model [10]

$$u = \sum_{j=1} W_j x_j + b \tag{1}$$

and

$$y = \partial(u) \tag{2}$$

In Eq. (1) and (2), W_j stands for the neuron's weights to the j^{th} synapse, $\partial(\cdot)$ is the activation function, and y is the output product. u is the summing output. x_j represents the j^{th} data or synapse's input signal. Common activation functions include the linear function, piecewise linear function, Logsig function, and fixed limiter function [10]. The ability of ANNs to make accurate predictions is significantly influenced by the training methods used and the architecture of the structure. Hence, improved training techniques were investigated to increase performance even more. A nonlinear functional structure called an MLP neural network can be trained to provide a certain input-output mapping [11]. They added that a forecast cannot be correct when a linear system is modelled using a nonlinear network, such as MLP. Figure 2 is composed of an input layer, a single hidden layer, and an output layer. A single hidden layer in an MLP network is adequate to give accurate prediction results, according to Funashashi (1989) and Cybenko (1989] [12-13]. The remainder of this study will therefore only cover neural networks with a single hidden layer.

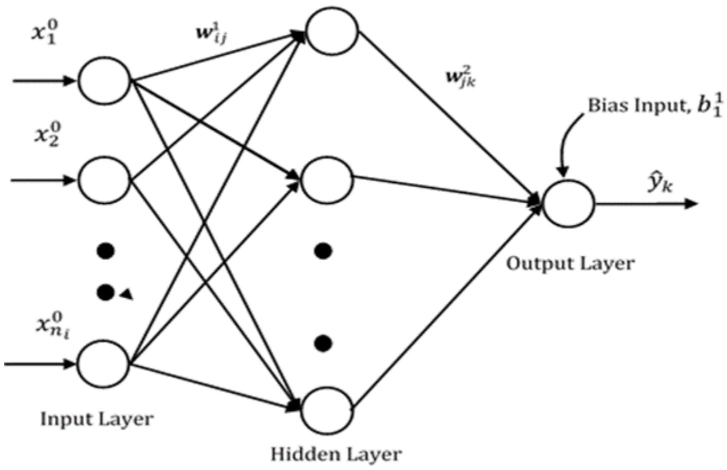

Figure 2. MLP architecture with one hidden layer

*Corresponding Author | Makmor, N. F. | nazrulfariq@upnm.edu.my

The output of the network is given by:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 \, \partial\left(\sum_{i=1}^{n_i} w_{ij}^1 x_i^0(t) + b_j^1\right) \tag{3}$$

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

where m represents the number of network outputs and n_h stands for hidden nodes. The activation function used in this instance with the Logsig activation function to activate the MLP network is $\partial(\cdot)$. The prediction error is determined by minimising the unknown variables w_{ij}^1, w_{jk}^2, w_{ik}^3 and threshold b_j^1, which must converge to optimal values as follows:

$$e_k(t) = y_k(t) - \hat{y}_k(t) \tag{4}$$

The system's actual output is $y_k(t)$, but the expected output is $\hat{y}_k(t)$. The learning phase is a critical stage in a neural network. The process makes sure the neural network can function according to its design requirements. The two most popular learning paradigms are supervised learning and unsupervised learning. Using supervised learning, a global model that links the input and intended result can be developed. On the other hand, supervised learning techniques don't need to estimate utilising tested training models. The learning procedure is distinct from supervised learning because there is no output target. The gathering of a set of input data, which is thought to be a set of random variables, is necessary for unsupervised learning. The datasets will be used to create a density model, and unsupervised learning will be based on prior knowledge. Or, to put it another way, learning is simply dependent on prior experience and is not directed by any specific goals [14]. Unsupervised learning facilitates data compression. For the study, an experimental procedure was followed by a modelling procedure utilising a neural network approach. The additional dataset is collected in addition to the target. Thus, guided training is the better option. BP, LM, Scaled Conjugate Gradient (SCG), and Bayesian Regularization (BR) are examples of supervised training algorithms that are used to model the Blast Pressure Prediction system [15-16].

3.0 RESULTS AND DISCUSSION

Prediction performance study is required to show that the MLP neural network can predict explosive pressure. The three steps of analysis in the MATLAB neural network tools (nntool) consist of 70% training and 30% testing. Examples include checking the MSE for errors and using regression to get the best fit [17]. The performance of the training method is assessed using the lowest MSE and highest regression performance. The relative error during the prediction phase should be as low as possible, according to the lowest MSE. The worst-case scenario for regression performance happens when the measurement is closest to 0, and the best performance happens when the measurement is closest to 1. Using MATLAB's neural network tool, the MSE and regression values for the difference training procedure were determined [18-19]. In Table 1, the performance of the MLP network using three distinct training algorithms is presented in descending order of highest to lowest MSE performance.

Table 1. MSE and Regression performance of MLP network

Training Algorithm	MSE Performance	Regression Performance	Number of Epoch
BR with Tansig	1.2042	0.9892	287
BR with Logsig	1.4235	0.9760	316
LM with Tansig	1.5825	0.9672	27
LM with Logsig	1.7473	0.9521	31

The performance of the MLP network is significantly influenced by the activation functions Tansig and Logsig, as these functions directly impact the network's outputs. The effectiveness of the training process was evaluated through simulation, yielding data summarized in Table 1. The combination of the BR training algorithm with the Tansig activation function achieved a minimum MSE of 1.2042. Conversely, the LM approach using the Tansig activation function resulted in a higher MSE of 1.5825. The study also investigated the Logsig activation function during both BR and LM training experiments, yielding MSE

*Corresponding Author | Makmor, N. F. | nazrulfariq@upnm.edu.my

results of 1.4235 and 1.7473, respectively, which indicated suboptimal performance compared to the Tansig activation function. Table 1 provides a comprehensive summary of the MLP network's regression outcomes associated with various training methodologies and activation functions. The regression performance score reached 0.9892 when utilizing the Tansig activation function in an MLP network trained with BR. The MLP network employing the LM training algorithm with the Tansig activation function secured the second-best performance, achieving a regression score of 0.9760. The regression scores for the MLP network trained with BR and LM using the Logsig activation function were recorded at 0.9672 and 0.9521, respectively, further evidencing the influence of activation functions on overall network performance.

Although the MLP network trained with the BR algorithm and Tansig activation function exhibited excellent MSE and regression results, it did not surpass the basic MLP architecture employing the LM training algorithm with Tansig during the constrained training period of 27 epochs. The experimental findings suggest that while BR facilitates superior long-term accuracy, LM demonstrates a faster initial convergence rate. The BR training algorithm operates under a stochastic model characterized by random variables, whereas the LM training algorithm is based on a deterministic model with predictable attributes. The distinctions between these underlying models contribute to the performance discrepancies illustrated in Table 1. Research on deterministic models has been extensive, aimed at identifying an algorithm that garners consensus among researchers. Many LM-based algorithms encounter challenges in achieving optimal performance levels, often becoming trapped in local minima during the training process. Although the BR technique requires an extended convergence period of 287 epochs, it ultimately attains greater accuracy than other combinations. In contrast, the LM training approach reaches convergence in merely 27 epochs, albeit with a lower accuracy compared to that achieved by BR.

4.0 CONCLUSIONS

The efficacy of MLP networks can be deduced from the results obtained across various domains. The BR training algorithm distinguishes itself among the methodologies discussed in this study, yielding the most favourable regression outcomes alongside the lowest MSE values. In contrast, the LM training algorithm, while executing its functions more swiftly, is associated with comparatively subpar regression predictions and elevated MSE values, indicating diminished reliability in forecasting capabilities. It is essential to recognize that there remains considerable potential for further enhancement of the approach in the future, despite the BR method demonstrating superior accuracy and reduced error rates compared to the LM technique. Among the strategies examined in the prior section, one promising avenue for performance enhancement lies in the modification of the MLP network architecture. For instance, the implementation of deep learning methodologies could prove advantageous, as it facilitates the definition and exploration of diverse input perspectives, potentially resulting in more intricate models. The superior accuracy achieved by the BR training algorithm suggests that it is well-suited for future training endeavours, even though its training duration exceeds that of the LM method.

5.0 CONFLICT OF INTEREST

The authors declare no conflicts of interest.

6.0 AUTHORS CONTRIBUTION

Adnan, J. (Project administration; Supervision)
Mohd Sabri, M. S. (Validation; Writing - review & editing)
Ahmad Jamil, S. H. F. S. (Resources; Software; Writing - original draft)
Ahmad, S. (Data curation; Formal analysis)
Makmor, N. F. (Conceptualization; Writing - original draft)

7.0 ACKNOWLEDGEMENTS

This research is entirely supported by the UPNM/2023/GPJP/TK/5 grant from the GPJP. The authors express their gratitude to the Ministry of Higher Education (MOHE) and the National Defence University of Malaysia (UPNM) for providing the approved funding that made this research feasible.

*Corresponding Author | Makmor, N. F. | nazrulfariq@upnm.edu.my

List of Reference

[1] Anderson, D. J., Smith, S. T., & Au, F. T. (2016). Mechanical properties of concrete utilising waste ceramic as coarse aggregate. *Construction and Building Materials, 117,* 20-28.

[2] Pitarch, A. M., Reig, L., Tomás, A. E., & López, F. J. (2019). Effect of tiles, bricks and ceramic sanitary-ware recycled aggregates on structural concrete properties. *Waste and Biomass Valorization, 10*, 1779-1793.

[3] Fu, R., Hu, X., & Zhou, B. (2017). Discrete element modeling of crushable sands considering realistic particle shape effect. *Computers and Geotechnics, 91*, 179-191.

[4] Method for determination of slump, BS 1881, Part 102, British Standard BS1881 (1983).

[5] Murtagh, F., Qiao, X., Crookes, D., Walsh, P., Basheer, P. M., Long, A., & Starck, J. L. (2005). A machine vision approach to the grading of crushed aggregate. *Machine Vision and Applications, 16*, 229-235.

[6] Singh, V., & Rao, S. M. (2005). Application of image processing and radial basis neural network techniques for ore sorting and ore classification. *Minerals Engineering, 18*(15), 1412-1420.

[7] Norizan, Y. N. A., Din, M. F. M., Zamri, W. F. H. W., Hashim, F. R., Jusoh, M. T., & Rahman, M. R. A. (2018, February). The effects of different heat treatment annealing on structural properties of LaFe11. 5Si1. 5 compounds. In *AIP Conference Proceedings (Vol. 1930, No. 1)*. AIP Publishing.

[8] Jamil, S. H. F. S. A., Kadir, J. A., Hashim, F. R., Mustapha, B., Hasan, N. S., & Januar, Y. (2020, August). Optimization of ecg peaks for cardiac abnormality detection using multilayer perceptron. In *2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (pp. 169-173). IEEE.

[9] Al-Batah, M. S., Isa, N. A. M., Zamli, K. Z., Sani, Z. M., & Azizli, K. A. (2009). A novel aggregate classification technique using moment invariants and cascaded multilayered perceptron network. *International Journal of Mineral Processing, 92*(1-2), 92-102.

[10] Haykin, S. Neural Networks & Learning Machines, third ed., Prentice Hall, New York, 2011.

[11] Hashim, F. R., Nagappan, P., Ishak, M. T., Joini, N. F., Makmor, N. F., Saleh, M. S., & Zolkiply, N. (2021). Solar location estimation using logsig based activation function using artificial neural network approach. *Zulfaqar Journal of Defence Science, Engineering & Technology, 4*(1).

[12] Funahashi, K. I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural networks, 2*(3), 183-192.

[13] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems, 2*(4), 303-314.

[14] Neagoe, V. E., & Stoica, R. M. (2018, July). A new neural approach of supervised change detection in SAR images using training data generation with Concurrent Self-Organizing Maps. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 4792-4795). IEEE.

[15] Ahmad, S., Ahmad, K. A., Hashim, F. R., & Syafuan, W. M. (2019). Terrain masking and radar exposure modelling based on raster cells for pre-flight planning for low flying helicopters. *Defence S&T Technical Bulletin, 12*(2).

[16] Ahmad, K. A., Ahmad, S., & Hashim, F. R. (2018, February). MATLAB implementation of satellite positioning error overbounding by generalized Pareto distribution. In *AIP Conference Proceedings (Vol. 1930, No. 1).* AIP Publishing.

[17] Hashim, F. R., Jamil, S. H. F. S. A., Kadir, J. A., Hasan, N. S., Mustapha, B., & Januar, Y. (2019, November). Tansig Based MLP Network Cardiac Abnormality. In *2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (pp. 199-203). IEEE.

[18] Nadia, A. R., Mat Isa, N. A., & Desa, M. K. M. 2020. Efficient single and dual axis solar tracking system controllers based on adaptive neural fuzzy inference system. *Journal of King Saud University-Engineering Sciences 3*2(7): 459-469.

[19] Tuan Zizi, T. K., Ramli, S., Wook, M., & Mohd Shukran, M. A. 2023. A review of classification techniques for the prediction of human emotion through heart rate and eye movement. *Zulfaqar Journal of Defence Science, Engineering & Technology, 6*(2).

*Corresponding Author | Makmor, N. F. | nazrulfariq@upnm.edu.my