Shape of Aggregate Classification by using MLP Network based Training Algorithm and Activation Function

Mohd Salman Mohd Sabri[1, b)], Syahrull Hi-Fi Syam Ahmad Jamil[2, c)], Ja'afar Adnan[1, d)] and Fakroul Ridzuan Hashim[1, a)]

[1]*Faculty of Engineering, Universiti Pertahanan Nasional Malaysia, 57000 Sungai Besi, Kuala Lumpur, Malaysia*
[2]*Department of Mathematics and Science Computer, Politeknik Tuanku Syed Sirajuddin, 02600 Arau, Perlis, Malaysia*

[a)]*Corresponding author: fakroul@upnm.edu.my*
[b)]*salman@upnm.edu.my*
[c)]*syahrull@ptss.edu.my*
[d)]*jaafar@upnm.edu.my*

Abstract. Mechanical sifting and hand grading have long been used to assess the quality of aggregates. It must pass a range of mechanical, chemical, and physical testing to produce superior aggregates; these tests are typically performed manually and are sluggish, arbitrary, and time-consuming. This work aims to develop an image-based classification system that can categorise aggregates. An artificial neural network was used to reprocess the image after it had been taken in order to classify its shapes. In comparison to Backpropagation (BP) training techniques, the Bayesian Regularization (BR) methodology offers better performance with reduced mean square error (MSE) and higher regression. The LM training approach using the Multilayer Perceptron (MLP) network-based MSE offers the maximum regression and the lowest mean square error (MSE). The BR-trained network has 1.4235 MSE and 0.9760 regression capabilities.

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 poor aggregates are two classifications that are frequently used to categorise aggregates. Poor aggregate can be divided into four types: elongated, flaky, flaky & elongated, and irregular, as opposed to good aggregate, which can be divided into two types: angular and cubical [3]. According to British Standards BS812, Section 103.1 [4], mechanical sifting and hand gauging have traditionally been employed to determine the size and shape of coarse aggregates. Due to different particle morphologies, errors may be generated during the sieving process, also known as "grading analysis." Many methods have been developed to enhance the traditional categorization approach, utilising imaging tools and analytical algorithms to assess 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 Artificial Neural Networks (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 Multilayered Perceptron (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.

Because 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. As a result, the neural network model needs a large amount of data and a lengthy training period in order to be properly trained. However, enhancing the learning capacity of the training algorithm might be able to resolve the issue mentioned earlier. The Bayesian Regularization (BR) training algorithm, an improved variant of the Backpropagation (BP) training method, can solve the issue even though BP is known to be stuck in local minima.

Methodology

There were 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 artificial neural networks (ANN) is modelled after the way the brain works. The artificial neural network 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.

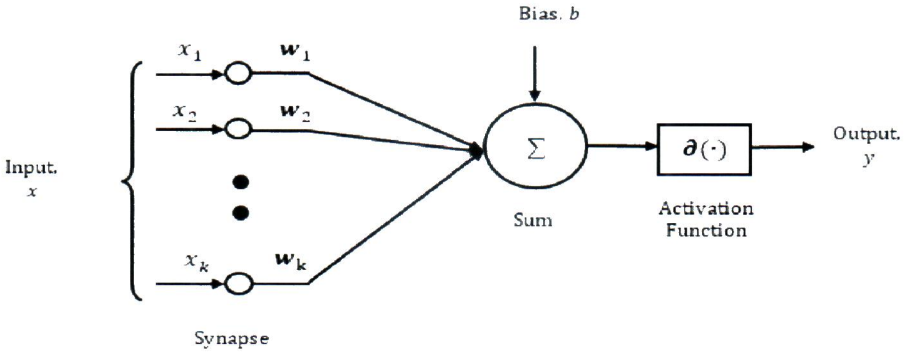

FIGURE 1. Nonlinear neuron model [10]

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 $(x_1, x_2 .. x_k)$, and the weight at each synapse by $[\mathbf{w}_1, \mathbf{w}_2 .. \mathbf{w}_k]$. The value of the j^{th} synaptic weights $[\mathbf{w}_j]$ 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 $[\mathbf{w}_j]$ 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 mathematical modelling of neurons can be defined using the two equations below based on Figure 1:

$$u = \sum_{j=1} \mathbf{W}_j x_j + b \tag{1}$$

and

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

In Equations (1) and (2), $\mathbf{W}_j$ stands for the neuron's weights to the j^{th} synapse, () 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 looked into 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 4 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) [12] and Cybenko (1989] [13]. The remainder of this study will therefore only cover neural networks with a single hidden layer.

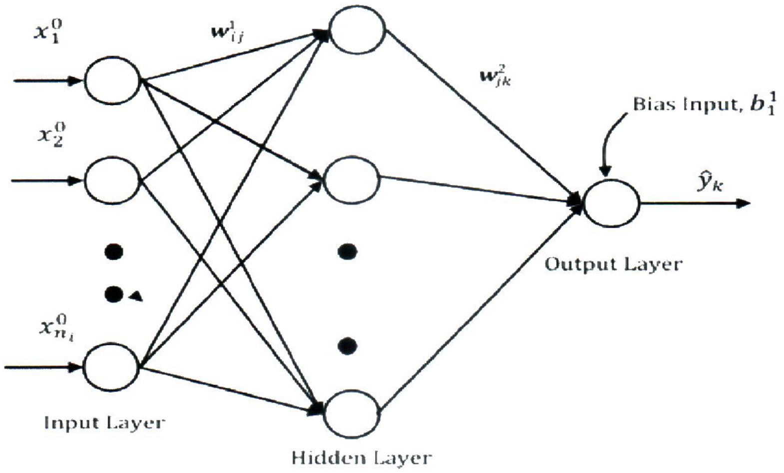

FIGURE 1. MLP architecture with one hidden layer

The output of the network is given by:

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

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

where n_h 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 $\boldsymbol{\partial}(\cdot)$. 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:

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

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. Backpropagation (BP), Scaled Conjugate Gradient (SCG), Lavenberg-Marquardt (LM), and Bayesian Regularization (BR) are examples of supervised training algorithms that are used to model the Blast Pressure Prediction system [15–16].

Result 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. 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 Logsig	1.4235	0.9760	316
BP with Logsig	1.9752	0.9368	17
BR with Purelin	2.6956	0.9228	402
BP with Purelin	3.0275	0.8257	21

Using the activation functions of Logsig and Purelin, the MLP network is turned on. Based on simulation, Table 1's BR training algorithm with Logsig activation function might offer the MSE performance with the lowest value of 1.4235. MLP network trained by BP training algorithm with Logsig activation function, MSE of 1.9752. The BR and BP training algorithms with MSE values of 2.6956 and 3.0275, respectively, were activated by Purelin activation function to produce the simulation results. The MLP network's regression performance after being trained using those training techniques and activation functions is also shown in Table 1. Once more, an MLP network trained using BR with Logsig can provide a regression performance score of 0.9570. With a regression score of 0.9368, the MLP network using the BP training technique and Logsig activation function comes in second. Regression results from the MLP network trained by BR and BP and activated by the Purelin activation function, respectively, reveal values of 0.9228 and 0.8257. MSE and regression performance is shown by MLP network trained by BR training algorithm and Logsig activation function, however it is unable to surpass the simplest structure provided by MLP trained by BP training method activated by Logsig with 17 number of epoch.

The BR training algorithm is based on a stochastic model, whereas the BP training algorithm was built from a deterministic model, clearly demonstrating the performance differences between the two training methods in Table 1. The deterministic model is by far the most thoroughly studied when seeking for a recognisable algorithm, whereas a stochastic model is made up of a set of random variables. As a result, most BP-based algorithms struggle to perform effectively because they frequently get trapped in local minima during training. Although the BR method takes a while to converge with 316, the accuracy it achieves is higher than that of other combinations. The BP training algorithm, on the other hand, can converge quickly with only 17 epochs, but it is unable to deliver acceptable accuracy.

Conclusion

The effectiveness and capacity of the MLP network in forecasting aggregate form are shown by the prediction results. The accuracy shown by the BR training method is the best, with the smallest MSE and maximum regression performances, according to the data. Because of this, even though the BP training strategy just needs a few epochs and has a short processing time, it can only produce worse regression results and a larger MSE. Although the BR performs better than the BP, it cannot match the BR training algorithm's performance.

Acknowledgments

This research is fully supported by GPJP grant, UPNM/2022GPJP/TK/2. The authors fully acknowledged the Ministry of Higher Education (MOHE) and National Defence University of Malaysia (UPNM) for the approved fund, making this important research viable and effective.

References

[1] D. J. Anderson, S. T. Smith and F. T. K. Au, Mechanical properties of concrete utilising waste ceramic as coarse aggregate, Construction and Building Materials, 117 (2016) 20-28.
[2] A. M. Pitarch, L. Reig, A. E. Tomás and F. J. López, Effect of tiles, bricks and ceramic sanitary-ware recycled aggregates on structural concrete properties, Waste and Biomass Valorization, 10(6) (2019) 1779-1793.
[3] R. Fu, X. Hu and B. Zhou, Discrete element modeling of crushable sands considering realistic particle shape effect, Computers & Geotechnics, 91 (2017) 179-191.
[4] Method for determination of slump, BS 1881, Part 102, British Standard BS1881 (1983).
[5] F. Murtagh, X. Qiao and D. Crookes, A machine vision approach to the grading of crushed aggregate. Machine Vision & Applications, 16 (2005) 229-235.
[6] V. Singh and S. M. Rao, Application of image processing and radial basis neural network techniques for ore sorting and ore classification, Minerals Engineering, 18(15) (2005) 1120-1132.
[7] Y. N. A. Norizan, M. F. M. Din, W. F. H. W. Zamri, F. R. Hashim, M. T. Jusoh and M. R. A. Rahman, The effects of different heat treatment annealing on structural properties of LaFe11.5Si1.5 compound, AIP Conference Proceedings (2018) 1930.
[8] S. H. F. S. Ahmad Jamil, J. A. Kadir, F. R. Hashim, B. Mustapha, N. S. Hasan and Y. Januar, Optimization of ECG peaks for cardiac abnormality detection using multilayer

perceptron, 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), (2020) 169-173.

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

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

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

[12] K. Funashashi, On the approximate realization on continuous mapping by neural networks, Neural Networks. 2 (1989) 182-192.

[13] G. Cybenko, Approximation by superposition of the sigmoid function. Mathematics of Control, Signals & Systems. 2 (1989) 303-314.

[14] V. Neagoe and R. Stoica, A new neural approach of supervised change detection in SAR images using training data generation with concurrent self-organizing maps, IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium. (2018) 4792-4795, Valencia.

[15] S. Ahmad, K. A. Ahmad, F. R. Hashim and W. M. Syafuan, 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) (2019) 318-329.

[16] K. A. Ahmad, S. Ahmad and F. R. Hashim, MATLAB implementation of satellite positioning error overbounding by generalized Pareto distribution, AIP Conference Proceedings (2018) 1930.

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