Jurnal Kejuruteraan 36(x) 2024: xxx-xxx
https://doi.org/10.17576/jkukm-2024-36(X)-XX

Activation Function based MLP Network for Shape Aggregate Classification

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

[a]*Faculty of Engineering, Universiti Pertahanan Nasional Malaysia, 57000 Sungai Besi, Kuala Lumpur, Malaysia*
[b]*Royal Service Corp Directorates, Army Head Quarters, Ministry of Defense, Jalan Padang Tembak, Kuala Lumpur, Malaysia*
[c]*Department of Mathematics and Science Computer, Politeknik Tuanku Syed Sirajuddin, 02600 Arau, Perlis, Malaysia*
[d]*Department of Computer Science & Information Technology, Maulana Azad National Urdu University, India*

**Corresponding author: nazrulfariq@upnm.edu.my*

Received XX month 20XX, Received in revised form XX Month 20XX
Accepted XX Month 20XX, Available online Month 20XX

ABSTRACT

The assessment of aggregate quality depends on manual grading together with mechanical filtering through traditional methods. Aggregates need to pass through multiple mechanical and physical as well as chemical tests to verify their compliance with established standards. The evaluation procedures performed by hand prove to be inherently inefficient and subjective and take up too much time. A project seeks to create an image processing system which will classify aggregates into various categories. The classification system employs an artificial neural network (ANN) to analyse images for determining aggregate shapes. The study compares the performance of different training algorithms for the ANN. The study compares the performance of Levenberg Marquardt (LM) against Bayesian Regularization (BR) as training algorithms. The results show that BR training outperforms other methods since it provides better mean square error (MSE) values and enhanced regression outcomes. The combination between BR training method and MLP network delivers optimal performance levels regarding regression accuracy and MSE measurement. Through BR training the network obtained an MSE of 1.2042 and a regression of 0.9892 which confirms its successful ability to classify aggregates through image analysis. Through this alternative method researchers gain an efficient and objective solution to replace traditional manual classification approaches.

Keywords: Aggregate classification, MLP network, Training algorithm, MSE, Regression

INTRODUCTION

Aggregate quality is one of the most important factors which affect the production quality of concrete since these materials are key in defining the strength, durability, and performance of the concrete being produced. Of all the rock types used for aggregate, granite and limestone are the most common because of the abundance of these two and their mechanical properties. High-strength concrete is defined and depends on three parameters namely: The properties include the shape of the aggregate particles, the distribution and surface texture. These characteristics have the effect of increasing the packing density as well as workability of the concrete mix thus leading to better bonding results. Also, the characteristics of aggregate particles are also changed by the nature of the rock deposits from where they are derived, choice of crusher and degree of reduction. These and many other factors, therefore, determine the quality of concrete to be produced, pointing out why it is necessary to use quality aggregates for better performance (Ray et al. 2021; Amin et al. 2020).

The improvement of aggregate shape and size can be crucial to the overall quality of concrete, and this is an area that should be given priority. This brings about enhanced concrete blending that makes the product stronger, less permeable to water, thus making them more durable and resistant to weather and natural conditions (Sosa et al., 2021). The utilization of high-quality aggregates can also improve the performance of concrete and reduce costs at the time of production and application of concrete.

There are two general categories of aggregates based on their quality, good aggregates and poor aggregates. Aggregates used in concrete must be of good quality and should have angular and cubical shapes so that they interlock with one another when mixed. This interlocking is important since it increases the matrix strength of the concrete and, therefore, the performance. On the other hand, the poor-quality aggregate, for instance, elongated, flaky, or irregular shape prove to be detrimental to the workability and strength properties of concrete. These undesirable shapes pose some difficulties in attaining a homogenous blend, which leads to a weaker concrete structure (Zhang et al., 2020).

The previous practices in size and shape determination of coarse aggregates have used mechanical sieving and manual measurements as expounded by the British Standard BS812 Section 103.1 (British Standard BS1881, 1983). However, this process is not without its flaws because the morphologies of the particles that are being formed can cause variations and inaccuracies in grading analysis. In this regard, there has been an attempt to address the above challenges through the development of new methods of assessment that incorporates modern imaging systems and analytical software. These techniques enhance the measurement of the aggregate dimensions and shapes which enhances the quality of the aggregate assessment and overall quality of concrete produced (Subramaniam et al., 2024).

Khorram et al. (2017) have proposed techniques for the design of the machine vision system for classification of aggregates which can be considered as a breakthrough in material engineering. These are real-time systems and are made up of two major parts: The classification module and the image processing module were identified for the system. Classification for the aggregates is based on the appearance of the aggregates and this ability helps in classifying the aggregates as per type or calibration. At the same time, the image processing module provides the means for obtaining the relevant features from the images of the aggregates captured by the camera, and thus, enhances the analysis.

In their research, Khorram et al. proposed a new machine vision method to analyze aggregate sizes with the help of a method called multi-scale image entropy analysis. This makes it possible to analyze the characteristics of the aggregate based on specific images and, consequently, improve the accuracy of size classification. This work is complemented by Duboselard et al. (2015) that proposed classification methods for ores based on the fluidized bed visual texture analysis that depends on the evaluated concentrations of minerals. Their new RGB color space image processing method allows to isolate the picture of visual texture of ore particles, which makes classification more accurate.

These image processing techniques are incorporated with Artificial Neural Networks (ANNs) to improve the classification of aggregates to the next level. The ANNs categorize aggregates using the second-order statistical measures of entropy, contrast, energy, and homogeneity together with the first-order statistical measures based on gray scale values. This set of statistical descriptors enables the analysis of the aggregates 'shape at the level of the picture. Additionally, the depiction of elements like manganese, iron, alumina, and specific area of the aggregate in the various regions is feasible with the help of the greyscale variations and necessary calculated metrics of entropy, contrast, energy, and uniformity (Tripathy & Guru, 2017).

The application of ANNs in analyses makes them superior to the fuzzy logic systems, evolutionary algorithms, and conventional statistical approaches especially in classifying aggregates where the relationships are non-linear. Deep learning methods such as ANNs have become very popular due to the model's ability to generalize from the training examples and learn from a variety of inputs. This ability to learn from examples is especially valuable in the data mining categorization problem, and ANNs exhibit very little sensitivity to the 'curse of dimensionality.' This, coupled with their ability to process large data sets that may include multiple variables, makes ANNs popular in many analytic applications.

The field of application of ANN is vast and can be applied in areas such as pattern recognition, classification, signal and image processing, control of robots, meteorological and financial prediction, and diagnostics. They have also been widely used in these domains due to their flexibility and ability to provide a strong performance. In the field of pattern classification problems, two primary network architectures are widely used in artificial neural networks: Among these they are the Radial Basis Function (RBF) networks, and the Multi-Layer Perceptrons (MLPs). These architectures are designed to improve the capability of learning and classifying intricate data patterns in the network (Norizan et al., 2018; Ahmad Jamil et al., 2020).

The MLP architecture is known to be the most significant and the most applied type of neural network among all the known ones. It is explained by several unique features, such as computational ease, limited number of parameters, stability, and compactness. These aspects render the MLP as the ideal solution to solve problem such as (Sabri et al., 2024). The MLP architecture performs very well in providing direct solutions hence its capability of determining the correlation between the input and output with a lot of precision.

Neural network models including the MLP are used to estimate non-linear relationship that may have unknown parameters involved. This means that a non-linear optimization technique is necessary for this purpose. However, optimization process can be proved to be cumbersome in the sense that it may slow down the convergence of parameters, it may consume a lot of computational power, and it may give sub-optimal solutions which are trapped in local optima. Thus, training a neural network model requires large data sets and long train times to properly learn and generalize the data.

To address these challenges, research activities have directed towards improving the learning capability of the training algorithms. Other development is the use of Bayesian Regularization (BR) training algorithm which is seen as an

https://doi.org/10.17576/jkukm-2024-36(X)-XX

enhancement of the Backpropagation (BP) training algorithm. BP is, however, not without its challenges, which include getting stuck in local minima that may hinder the performance of the model (Hang et al., 2023). The problem of overfitting is handled by the BR algorithm by adding a regularization term to the error function. This term punishes complexity, which helps to reduce overfitting tendency. Therefore, over iterations, the network able to come up with better and more accurate features that can be used to map between the input features and the output classes. This also helps in increasing the accuracy of the prediction as well as the speed at which other macro-level classification processes are conducted.

METHODOLOGY

Out of the 625 compiled pictures, 425 of the images represented angular and cubic shapes while the other 200 of the images depicted undesirable flaky and elongated and flaky and elongated and irregular shapes. The images should be pre-processed to improve the resolution of the images for better quality of the images. The feature extraction tool selects features necessary for classification in the next step out of all the features of the input data. A thresholding method is used to segment the images first before the contraction and expansion process takes place. Both detail amplification and background aggregate discrimination are beneficial based on the studies conducted in (Wang et al. 2022). The main issue in extracting features through geometrical moments is the way the moments can be used to classify the shape of the aggregate. Hu and Zernike moments turn out to be especially suitable for this purpose. Hu and Zernike moments remain invariant to geometric transformations such as resizing and movement of the objects and rotation of the axes and therefore make them suitable selection tools for identifying groups correctly. Seven Hu moments are views from two different sites: One of them is the grouping of moments from the region to provide comprehensive shape information while the other is the grouping of moments from the border.

This is effectuated through the ANN's imitation of human brain architecture, in addition to its learning and working principles. The biological structure of the ANN enables it to learn about relationships and multiple patterns of image data within a large aggregate data set. Figure 1 represents the model of nonlinear neurons to illustrate the basic architecture of ANN and their elements that make it possible to perform complex a classification task.

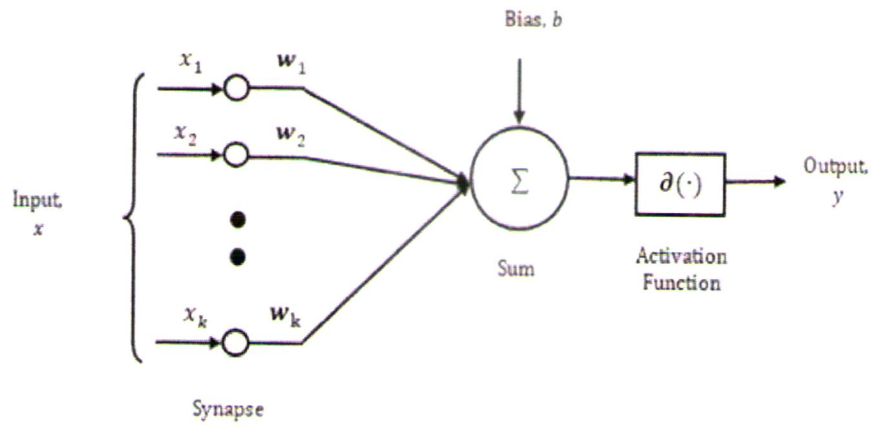

FIGURE 1. Nonlinear neuron model (Elbrächter et al. 2021)

Figure. 1 visually represents the architecture of a neuron, highlighting its key components: a network of interconnected synapses (or connections), a summation function, and an activation function. Each synapse in the neuronal network is assigned a weighted value, reflecting its importance in the processing of input signals. If we assume a neuron has 'k' synapses, it consequently possesses 'k' inputs. The activation function of the model is denoted as $\partial(\cdot)$, with the input at each synapse represented by $(x_1, x_2 \ldots x_k)$, and the corresponding synaptic weight by $(w_1, w_2 \ldots w_k)$. The value of the jth synaptic weight, denoted as $[w_j]$, plays a crucial role in determining the influence of that particular synapse on the neuron's output. Specifically, at the input synapses connected to the neuron, the value of $[w_j]$ is multiplied by the input signal x_j. The activation function then takes the result of a summation process, combining all the weighted input signals along with a bias term (b). This entire process, representing the mathematical modeling of neurons, is visually depicted and explained in Fig. 1, illustrating how individual inputs are processed and transformed to produce an output signal

https://doi.org/10.17576/jkukm-2024-36(X)-XX

$$u = \sum_{j=1} W_j x_j + b \qquad (1)$$

and

$$y = \partial(u) \qquad (2)$$

In the context of neural networks, specifically within equations (1) and (2), the term W_j represents the weights associated with a neuron's connection to the jth synapse. The symbol $\partial(\cdot)$ denotes the activation function, which introduces non-linearity into the network, and y signifies the output produced by the neuron. The variable u represents the summed output before the activation function is applied. The input signal or data received by the j^{th} synapse is represented by x_j. Common activation functions include the linear function, the piecewise linear function, the Logsig function, and the fixed limiter function, as noted by Makmor et al. (2024).

ANN predictive accuracy depends heavily on training methodologies together with the overall network architecture design. Scientists have investigated better training methods because they want to enhance performance results. An MLP network functions through its non-linear functional structure to create specific input-output mappings according to Hashim et al. (2021). The researchers argued that employing non-linear networks like MLP to depict linear systems would produce incorrect output results. Fig. 2 shows the standard MLP design with its input layer followed by one hidden layer and ending with an output layer. The MLP network requires only one hidden layer to produce accurate prediction outputs according to independent research by Funashashi (1989) and Cybenko (1989). The research will concentrate solely on neural networks with one hidden layer because this proven architecture requires a simplified analysis.

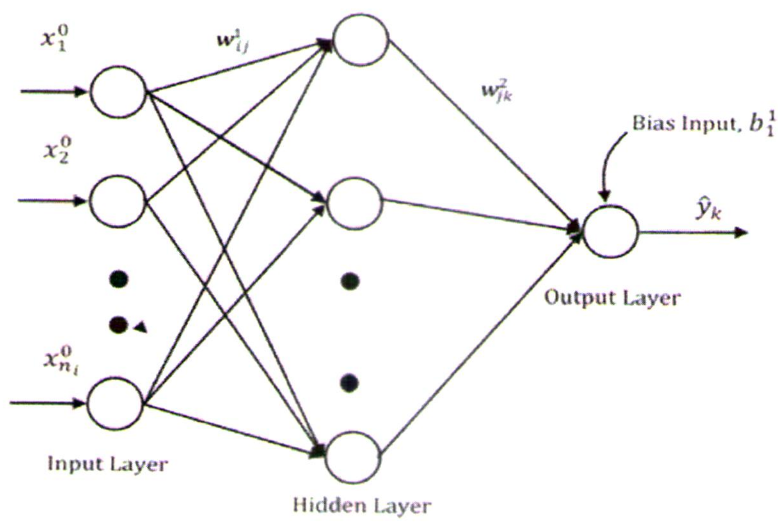

FIGURE 2. MLP architecture with one hidden layer (Makmor et. al., 2024)

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

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

In the realm of system analysis, the actual output is denoted as $\hat{y}_k(t)$, while its corresponding predicted output is represented as $\hat{y}_k(t)$. The training phase takes the center stage as the most crucial aspect in the development and deployment of neural networks. The basic procedure ensures that the neural network is functioning well and responding to the specifications and performance standards that it was meant to meet. Machine learning comprises of two classes of learning which include the supervised learning and the unsupervised learning. Supervised learning is an efficient technique of establishing a global model to map between the input data and expected results. Training involves the use of labelled data to build the mapping between the input and the output. A Supervised learning systems works by identifying the mapping between the input and the output, unlike the training models. The only major difference that can be drawn between supervised learning and the other type of learning is that in the latter case there are no targets for the output and no labelled data. As for the approach of unsupervised learning, the input information must be received in the form of a set of random variables. It is possible to generate the density models due to unsupervised learning to extend previous knowledge about the identification of patterns and structures in data.

The learning process in this paradigm relies on experience because it does not have a goal based on Prudencio et al. (2023). Data compression can be considered as one of the main applications of unsupervised learning since the algorithm creates abstractions of data. The study used an experimental research design before adopting the neural network modelling as the analytical tool. The additional dataset is gathered by the research team simultaneously with the target information. The selection is for guided training as this training model performs better than the other one. The trained methods BP with SCG LM, and BR are used in the prediction output as posted by Ahmad et al. (2019) and Ling & Mat Isa (2023).

RESULT AND DISCUSSION

A specific investigation must be performed to establish the forecasting power of MLP networks when predicting explosive pressure values. MATLAB's neural network toolbox (nntool) hosts the analytical process which implements a 70:30 dataset separation for training and testing purposes while using a network structure featuring 10 hidden layers. The MSE assessment for measuring inaccuracies is followed by regression analysis to reach optimal model fitting according to Nadia et al. (2020) and Tuan Zizi et al. (2023). The MSE minimization and achievement of maximum regression performance serve as indicators to evaluate the training process effectiveness. The predictive phase requires minimal relative error because it directly affects the resulting MSE value. The regression performance shows its worst condition when the value approaches zero, but its best state occurs when the value reaches one. The MSE and regression performance of the differential training method were calculated using MATLAB neural network software. The three different training procedures applied to MLP network yielded performance results shown in Table 1. The list of procedures shows descending MSE performance which starts from inaccurate methods at the top and includes more accurate procedures at the bottom.

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
BR with Purelin	2.6956	0.9228	402
LM with Purelin	2.8752	0.8572	36

Performance of the MLP network depends heavily on the activation methods Tansig, Logsig and Purelin because they affect its outcomes. The evaluation of training process effectiveness occurred through simulation which produced the data in Table 1. The combination of BR training method with Tansig activation function reached a minimum MSE value of 1.2042. The LM approach with Tansig activated function resulted in an MLP network with a higher MSE value of 1.5825. The research examined the Logsig activation function during BR and LM training experiments. The measured MSE results were 1.4235 and 1.7473 which showed inferior performance than when using Logsig activation function. Purelin activation function unable to outperform Tansig and Logsig since the MSE value are higher than others with 2.6956 and 2.8752 both for BR and LM training algorithm, respectively. Table 1 presents an extensive summary of the MLP network's regression results which resulted from different training approaches and activation functions.

The regression performance score reached 0.9892 when using Tansig activation function in MLP network trained with BR. The MLP network using LM training algorithm and Tansig activation function obtained the second-best position by achieving 0.9760 in regression performance. The regression scores from the MLP network trained with BR and LM with Logsig activation function evaluation reached 0.9672 and 0.9.521 which demonstrated additional evidence of activation function effects on overall performance. On the other hand, BR and LM training algorithm activated by Purelin capable to give 0.9228 and 0.8572 on regression, respectively.

The MLP network trained through BR

ISSN:0128-0198 E-ISSN:2289-7526

training algorithm and Tansig activation function showed excellent MSE and regression results but failed to outperform the basic MLP structure using LM training algorithm with Tansig activation function during restricted training of 27 epochs. The experimental outcome indicates that BR leads to superior long-term precision although LM displays faster initial convergence speed. The BR training algorithm operates under stochastic modelling with random variables, but the LM training algorithm derives from a deterministic model with predictable features. The differences between the underlying models result in the performance variations demonstrated in Table 1. Research about the deterministic model has been extensive throughout the search for an algorithm that all parties can agree upon. Most LM-based algorithms have trouble reaching their best performance level because they easily get stuck in training process local minima. The BR technique takes a long time to converge at 287 epochs, yet it achieves better accuracy than other possible combinations. The LM training approach completes its convergence in only 27 epochs although its accuracy level remains lower than what BR achieves.

CONCLUSION

The performance of Multilayer Perceptron (MLP) networks can be inferred from the outcomes obtained in different fields. The BR training algorithm stands out among the methods discussed in the paper and provides the best regression results as well as the lowest value of MSE. The Levenberg-Marquardt training method, on the other hand, performs its tasks faster but is followed by relatively inferior regression predictions and high MSE values that signify less reliability in the forecast.

It is important to understand that there is still some potential for further development of the approach in the future, even though the BR approach has been proven to be more accurate and have fewer errors than the LM method. Of all the approaches examined in the previous section, one of the promising areas of improving performance could be the modification of the structure of the MLP network. For example, the application of deep learning techniques may be beneficial because it defines and searches for various input perspectives which may result in more complex models.

The fact that BR training algorithm has given superior accuracy results means that the algorithm is suitable for future training even if the training time was longer than that of the LM method. Future research studies may help advance the development of deep architectures which can consider of various parameters to improve the predictive performance. Continuing this research has the potential to enhance the functioning of MLP networks in every field, thus bringing about a better capability of predictive models.

ACKNOWLEDGEMENT

This investigation is entirely supported by the UPNM/2022GPJP/TK/2 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 vital research possible and productive.

DECLARATION OF COMPETING INTEREST

None

REFERENCES

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): 318-329.

Amin, M., Tayeh, B. A., & Agwa, I. S. 2020. Effect of using mineral admixtures and ceramic wastes as coarse aggregates on properties of ultrahigh performance concrete. *Journal of Cleaner Production*. 273: 123073.

Duboscland, P., Larnier, S., Konik, H., Herbulot, A., & Devy, M. 2015. Automated visual grading of grain kernels by machine vision. In *Twelfth International Conference on Quality Control by Artificial Vision 2015*, 9534: 112-119. SPIE.

Elbrächter, D., Perekrestenko, D., Grohs, P., & Bölcskei, H. 2021. Deep neural network approximation theory. *IEEE Transactions on Information Theory*. 67(5): 2581-2623.

Hang, L., Liu, D. and Xie, F., 2023. A hybrid model using PCA and BP neural network for time series prediction in chinese stock market with TOPSIS analysis. *Scientific Programming*, *2023*(1), p.9963940.

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).

Jamil, S. H. F. S. A., Kadir, J. A., Hashim, F. R., Mustapha, B., Hasan, N. S., & Januar, Y. 2020. Optimization of ECG peaks for cardiac abnormality detection using multilayer perceptron. In *2020 10th IEEE International*

Conference on Control System, Computing and Engineering (ICCSCE): 169-173. IEEE.

Khorram, F., Morshedy, A. H., Memarian, H., Tokhmechi, B., & Zadeh, H. S. 2017. Lithological classification and chemical component estimation based on the visual features of crushed rock samples. *Arabian Journal of Geosciences*, 10: 1-9.

Ling, Q., & Mat Isa, N. A. 2023. Printed circuit board defect detection methods based on image processing, machine learning and deep learning: A survey. *IEEE Access* 11: 15921-15944.

Makmor, N. F., Hashim, F. R., Jamil, S. H. F. S. A., & Din, M. F. M. 2024. Classification of shape aggregate by using MLP network-based Bayesian regularization training algorithm. In *AIP Conference Proceedings* 2925(1). AIP Publishing

Method for determination of slump, BS 1881, Part 102, British Standard BS1881, 1983.

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* 32(7): 459-469.

Norizan, Y. N. A., Din, M. F. M., Zamri, W. F. H. W., Hashim, F. R., Jusoh, M. T., & Rahman, M. R. A. 2018. The effects of different heat treatment annealing on structural properties of LaFe11. 5Si1. 5 compounds. In *AIP Conference Proceedings* 1930(1). AIP Publishing.

Prudencio, R. F., Maximo, M. R. & Colombini, E. L., 2023. A survey on offline reinforcement learning: Taxonomy, review, and open problems. *IEEE Transactions on Neural Networks and Learning Systems*.

Ray, S., Haque, M., Sakib, M. N., Mita, A. F., Rahman, M. M., & Tanmoy, B. B. 2021. Use of ceramic wastes as aggregates in concrete production: A review. *Journal of Building Engineering*, 43: 102567.

Sabri, M. S. M., Hashim, F. R., Jamil, S. H. F. S. A., & Din, M. F. M. 2024. Shape aggregate classification using MLP based activation function. In *AIP Conference Proceedings* 2925(1). AIP Publishing.

Sosa, M. E., Zaccardi, Y. A. V. & Zega, C. J., 2021. A critical review of the resulting effective water-to-cement ratio of fine recycled aggregate concrete. Construction and Building Materials, 313, p.125536.

Subramaniam, D. N., Sajeevan, M., Pratheeba, J., Wijekoon, S. H. B. & Sathiparan, N., 2024. Characterization of the shape of aggregates using image analysis and machine learning classification tools. *Geomechanics and Geoengineering*, *19*(4), pp.421-443.

Tripathy, D. P., & Guru Raghavendra Reddy, K. 2017. Novel methods for separation of gangue from limestone and coal using multispectral and joint color-texture features. *Journal of The Institution of Engineers (India): Series D*, 98: 109-117.

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).

Wang, J., Lu, S., Wang, S. H., & Zhang, Y. D. 2022. A review on extreme learning machine. *Multimedia Tools and Applications*, 81(29): 41611-41660.

Zhang, T., Zhang, C., Zou, J., Wang, B., Song, F., & Yang, W. 2020. DEM exploration of the effect of particle shape on particle breakage in granular assemblies. *Computers and Geotechnics*, 122: 103542.