



2021

Evaluating the Performance of Extreme Learning Machine Technique for Ore Grade Estimation

Follow this and additional works at: <https://jasm.gig.eu/journal-of-sustainable-mining>



Part of the [Geological Engineering Commons](#), and the [Mining Engineering Commons](#)

Recommended Citation

Abuntori, Clara Akalanya; Al-Hassan, Sulemana; Mireku-Gyimah, Daniel; and Ziggah, Yao Yevenyo (2021) "Evaluating the Performance of Extreme Learning Machine Technique for Ore Grade Estimation," *Journal of Sustainable Mining*: Vol. 20 : Iss. 2 , Article 2.

Available at: <https://doi.org/10.46873/2300-3960.1062>

This Research Article is brought to you for free and open access by Journal of Sustainable Mining. It has been accepted for inclusion in Journal of Sustainable Mining by an authorized editor of Journal of Sustainable Mining.

Evaluating the performance of Extreme Learning Machine technique for ore grade estimation

Clara Akalanya Abuntori ^{a,*}, Sulemana Al-Hassan ^a,
Daniel Mireku-Gyimah ^a, Yao Yevenyo Ziggah ^b

^a Department of Mining Engineering, University of Mines and Technology, Tarkwa, Ghana

^b Department of Geomatic Engineering, University of Mines and Technology, Tarkwa, Ghana

Abstract

Due to the complex geology of vein deposits and their erratic grade distributions, there is the tendency of overestimating or underestimating the ore grade. These estimated grade results determine the profitability of mining the ore deposit or otherwise. In this study, five Extreme Learning Machine (ELM) variants based on hard limit, sigmoid, triangular basis, sine and radial basis activation functions were applied to predict ore grade. The motive is that the activation function has been identified to play a key role in achieving optimum ELM performance. Therefore, assessing the extent of influence the activation functions will have on the final outputs from the ELM has some scientific value worth investigating. This study therefore applied ELM as ore grade estimator which is yet to be explored in the literature. The obtained results from the five ELM variants were analysed and compared with the state-of-the-art benchmark methods of Back-propagation Neural Network (BPNN) and Ordinary Kriging (OK). The statistical test results revealed that the ELM with sigmoid activation function (ELM-Sigmoid) was the best among all the other investigated methods (ELM-Hard limit, ELM-Triangular basis, ELM-Sine, ELM-Radial Basis, BPNN and OK). This is because the ELM-sigmoid produced the lowest MAE (0.0175), MSE (0.0005) and RMSE (0.0229) with highest R^2 (91.93%) and R (95.88%) respectively. It was concluded that ELM-Sigmoid can be used by field practitioners as a reliable alternative ore grade estimation technique.

Keywords: extreme learning machine, artificial intelligence, artificial neural network, grade estimation, kriging

1. Introduction

An important aspect of mining is ore grade estimation, since it determines the viability of actively mining a mineral of interest. This process mainly involves estimating the reserve and grade using statistical procedures with samples obtained during drilling to determine the feasibility of mining the resource. Geostatistics is the conventional ore grade estimation technique which proves to be effective in the grade prediction of relatively uniform and massive deposits [1–4]. However, in extremely heterogeneous data sets, the geostatistical technique tends to perform poorly due to the complicated nature of the variograms obtained which are mostly rendered useless for further analyses and tend to

overestimate or underestimate the resource [3,5]. Also, manual tasking during the geostatistical resource estimation processes encourages bias and may introduce errors in the predicted ore grade values. These practical limitations are found in the most widely used geostatistical technique of Ordinary Kriging (OK). Over the years, in the quest to fix and improve the performance of OK, various kriging techniques such as indicator kriging [6,7], disjunctive kriging [8–10], multigaussian kriging [11,12], probability kriging [13–15], lognormal kriging [16,17] and outlier restricted kriging [18,19] were developed. These modifications resulted in more time consuming, computational complexity and overly expensive resource estimation processes. Due to these shortfalls, alternative resource estimation

Received 11 March 2021; revised 16 April 2021; accepted 18 April 2021.
Available online 25 May 2021.

* Corresponding author.
E-mail address: caabuntori@umat.edu.gh (C.A. Abuntori).

<https://doi.org/10.46873/2300-3960.1062>

2300-3960/© Central Mining Institute, Katowice, Poland. This is an open-access article under the CC-BY 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).

Table 1. Review of AI techniques applied in ore grade estimation.

Author	Technique	Observation
Wu and Zhou [20]	Multilayer Feedforward Neural Network (MLFNN) with Dynamic Quick-Propagation (DQP) variant	The technique overestimated and underestimated low-frequency values such as high-grade values as a result of smoothing. The number of data points employed in the study was 51.
Al-Alawi and Tawo [22]	BPNN	Required more data for prediction. Hence it possessed poor generalisation ability. Number of sample points used was 163.
Kapageridis and Denby [23,47] Kapageridis et al. [48–50] Kapageridis [24]	Radial Basis Function (RBF) and Multilayer Perceptron (MLP)	The neural network's resource estimates gave comparable results to kriging with fewer sample data. The number of drill hole data used in these studies ranged from 50 to 3600.
Matías et al. [51]	MLP, Regularisation Networks (RN) and RBF	Kriging outperformed the MLP, RN and RBF using a total of 1932 samples.
Samanta et al. [52]	Kohonen Neural Network (KNN)	KNN and kriging models performed almost equally well. However, grade values were generally overestimated due to the high nugget effect. The total number of drill holes used in the research was 497.
Samanta et al. [25]	MLFNN and SLFN with the Adaboost algorithm, BPNN	NN generally did not perform well due to the data's low spatial correlation and the high noise of the gold data used. A total of 275 drill hole data was employed in the study.
Chatterjee et al. [35] Samanta et al. [53]	ANN MLFNN with jump network and Genetic Algorithm (GA)	NN outperformed OK using 5149 data points. OK performed slightly better than the NN model. The number of exploratory borehole data used in the research was 181.
Mahmoudabadi et al. [26]	Levenberg–Marquardt Backpropagation (LMBP) with GA	NNs were quite sensitive when the MLP was used with back propagation-based algorithms (LMBP) in generating initial weight values with a limited training dataset. The study applied a total of 65 drill hole data.
Li et al. [21]	Wavelet Neural Network (WNN)	WNN accurately captured the local nonlinearity of the dynamic systems due to its multiscale, multiresolution and localisation ability using 200 drill hole data.
Chatterjee et al. [33]	GA and k-means clustering NN ensemble with SVM and RBF kernel	The estimated results obtained using the SVM and RBF outperformed OK. The number of drill hole data applied in the research was 4745.
Badel et al. [54]	MLFNN with Conjugate Gradient Method (CGM) optimisation and K-means clustering	Results of the Multiple Indicator Kriging (MIK) were more similar to the actual grade values. MIK also had better local precision than the ANN technique using a total number of 1802 data points.
Guo [55]	MLP, X-Ray Diffraction (XRD) and Levenberg–Marquardt (LM)	MLP training preferred correlated data. The dataset used was 82 drill holes.
Dutta et al. [56]	ANN-GA	The results obtained from the hybrid NN models generally did not perform well using a total number of 168 borehole data.
Dutta et al. [36]	Support Vector Regression (SVR) and LM Backpropagation (LMBP) NN algorithm	The SVR model gave the best results out of the NN, and OK methods applied. However, the upgrade was minimal due to the existence of extreme sample values in the 3500 drill hole data employed in the study.
Tahmasebi and Hezarkhani [27]	Adaptive Neuro-Fuzzy Inference System (ANFIS)	ANFIS gave better results than FL, ANN and Kriging. The number of data points used in the study was 258.

(continued on next page)

Table 1. (continued)

Author	Technique	Observation
Tahmasebi and Hezarkhani [57] Tahmasebi and Hezarkhani [31]	Multiple layer Neural network based on GA (ANN-GA) and ANN-based on Fuzzy logic (ANN-FL)	ANN-GA predicted the grade values well with high accuracy than ANN-FL results. The model had a small and individual structure that allowed the user to manage and control the NN training to derive the network's sound performance, especially with small data sets. However, due to the user's interference with the processing, the results could be biased. The number of borehole data used was 65.
Tahmasebi and Hezarkhani [58]	Coactive Neuro-Fuzzy Inference System (CANFIS) with GA and ANFIS-GA	The CANFIS-GA produced the best results due to its high correlation coefficient; however, ANFIS-GA gave best the least error on the testing data set. Data from 156 boreholes was applied in the study.
Maleki et al. [59]	Support Vector Machine (SVM), Backpropagation Neural Networks (BPNN)	The SVM was fast and gave more accurate results than that of the BPNN model using 4000 data samples.
Gholamnejad et al. [32]	MLFNN with Tanh activation function and Levenberg–Marquardt (LM)	The predicted values were deemed acceptable and had a correlation coefficient of 0.8. The number of sample points employed in the study was 2068.
Granek [60]	SVM and Convolutional Neural Network (CNN)	The CNN model was quite complicated, challenging to modify and computationally demanding but had an advantage over SVM by recognising anomalous structures in data. The number of sample points used in the research was 70.
Li et al. [37]	Self-adaptive Learning-based Particle Swarm Optimisation Support Vector Regression (SLPSO-SVR) model	The SLPSO-SVR technique performed better than PSO-SVR, ANN, comprehensive learning PSO-SVR and Grid-SVR. This technique had many advantages which included its rapid training ability and grade estimation using 2000 sample data points.
Jafrasteh and Fathianpour [61]	Local Linear Radial Basis Function (LLRBF) with Skewed activation function (SG), Simultaneous Perturbation Artificial Bee Colony algorithm (SPABC) and BPNN	The standard RBF trained with SPABC-BP algorithm showed higher generalisation ability and better prediction of ore grades for highly skewed data than LLRBF-SG-SPABC-BP and LLRBF-SPABC-BP. The technique was ideal for capturing nonlinear mappings in the 1250 data points used in the study.
Jafrasteh et al. [34]	Random Forest (RF), Gaussian Process (GP), MLP with LM	GP gave the best performance compared to the others since it provided a smoother interpolation and offered a more accurate prediction. IK was the next to provide a better estimation, and MLP gave the worse performance. However, all the techniques were sensitive to sudden variations of the copper concentrations. The number of data points employed in the study was 5647.
Singh et al. [30]	Recurrent Neural Network (RNN)	RNN gave comparable results with kriging. Kriging, however, performed slightly better than RNN. The number of sample points used was 3298.
Jahangiri et al. [62]	Gustafson-Kessel (GK) clustering algorithm with ANN	The accuracy of the results was poor based on the prediction of some elements; however, predictions were more accurate than the mine's estimation techniques. The number of borehole data applied was 1755.

(continued on next page)

Table 1. (continued)

Author	Technique	Observation
Manna et al. [63]	MLP (with 7 Principal Component Analysis (PCA) with ADAM optimiser	The obtained results were promising but required more sample data to ensure the developed model has a good generalisation ability as the samples were few (89 samples).
Zhang et al. [28]	Weighted Least Square Support Vector Regression (WLS-SVR)	The robust weighted WLS-SVR outperformed BPNN, OK and Inverse Distance Weighting (IDW) due to its strong predictive and generalisation ability. The number of samples used the study was 2304.

techniques using Artificial Intelligence (AI) have been applied in ore grade estimation. AI techniques, especially Artificial Neural Networks (ANNs) have effectively been employed in mineral resource estimation using limited data and in highly heterogeneous data sets [20–32]. Most of these techniques applied in literature (Table 1) for ore grade estimation are feedforward neural networks. Their popularity stems from their ability to approximate complex nonlinear mappings between input and output and produce models for a large class of data. From Table 1, the widely used ANN approach in ore grade estimation is the Backpropagation Neural Network (BPNN). Even though majority of the different AI techniques used (Table 1) outperformed kriging [20,22–24,26–28,31,33–37] a few suffered with smoothing in noisy datasets [20,25,30,34,38] which is not ideal in ore grade estimation.

Despite the broad applicability of BPNN, the technique has a wide range of limitations some of which include: overfitting problems, slow convergence and limited reasoning ability. BPNN also requires the model training parameters to be manually tuned in order to obtain optimum results which could lead to local minima with suboptimal solutions [39–41]. The chronological trial and error processes are required in this technique as there are no laid down procedures to ascertain the number of hidden neurons required for the model development [42,43]. Thus, the various issues of feedforward neural networks, including slower learning speed due to the use of gradient-based learning algorithms during the training phase and iteratively tuning all network parameters are addressed with the introduction of the Extreme Learning Machine (ELM) in Huang et al. [39].

In light of the strength and mathematical convenience, the study adopted the ELM approach for the ore grade estimation. The ELM was designed for Single Hidden Layer Feedforward Neural Network (SLFN) that randomly chooses its input

weights and hidden layer biases and can adequately learn on a given data set. It does this by adopting the function approximation in a finite training set resulting in its ability to apply almost any non-linear activation function to produce distinct predictions [39]. However, in practical application of the ELM, different variants exist based on their activation functions. These activation functions have been found to be a key factor in ELM achieving optimum prediction performance.

Therefore, assessing the impact of the activation function on the ELM performance for ore grade estimation has some scientific value worth investigating. In line with that, this study applied the following activation functions: Radial basis, Hard limit, Sigmoid, Triangular basis and Sine. Despite the wide application of ELM for solving diverse science and engineering problems [44–46], there is close to no application of ELM for ore grade estimation. Furthermore, no ore grade estimation study has assessed and compared the ELM variants with the state-of-the-art benchmark methods of BPNN and OK. Therefore, taking into consideration the strength of the ELM, this paper aims at:

- Determining the viability of the variants of ELM as a novel approach for ore grade estimation using exploratory data from a mine in Ghana;
- Determine its generalisation and predictive ability using heterogeneous data set; and
- Perform comparative analyses between the developed variants of ELM (ELM-Sine, ELM-Sigmoid, ELM-Radial Basis, ELM-Triangular Basis and ELM-Hard Limit) and benchmark techniques of BPNN and the OK.

2. Materials and methods

2.1. Study area

This research was conducted in a mine (hereafter Mine X) in Ghana. Mine X deposit is found in the Ashanti belt of the Birimian Supergroup and are

mainly volcanic rocks. The Birimian supergroup consists of northeast-striking belts with significant faults. The deposits found in the Ashanti belt contain mesothermal gold vein-type deposits [64]. Thus, the mineralisation is found in steep NNE–SSW to NE–SW trending shears. The gold is found in two principal ore types: quartz veins with free-milling gold, and sulphide ore containing arsenopyrite, pyrite and rare pyrrhotite and marcasite with refractory gold [65]. A map of the study area is shown in Fig. 1.

2.2. Materials

Secondary data was obtained from Mine X. The data comprised of assay, survey, lithology, and collar files obtained from exploratory drilling programme. The collar data contains the hole ID, the X, Y and Z coordinates of each collar, maximum depth of the drill hole and the type of drilling method. The assay data contains the composite Hole ID, From, To and the Assay value observed in that composited section of the borehole. The survey file consists of composite borehole sections at a particular location with its dip and bearing. The lithology file comprises composite collar data and the area between the “From” and “To” is given the observed rock type. The rock types observed were Schist (SC), Meta Volcanic (MV), Quartz (QU), Greywacke (GK), Phyllite (PH) and Laterite (LAT). The entire data set comprised of 3759 drilled holes.

2.2.1. Statistical description of data

The samples were initially taken at varying depths which made compositing necessary. Compositing involves the averaging of the original assay values to pre-specified lengths. This creates homogenous support of the data to be used for estimation and minimises data variability. Compositing, therefore, produces more robust statistical and structural analyses. The statistics were done on 1 m composite samples to identify if populations within the deposit were significantly different. It was also used to access the effects of the data distribution on the methods used for the grade estimations. The total number of samples obtained after compositing was 301 507 for the Assay and (X, Y, Z) coordinate. Table 2 summarises the descriptive statistics of the entire assay data and (X, Y, Z) coordinate. A cumulative frequency graph was also developed to ascertain the data distribution of the assay values (Fig. 2).

2.3. Methods

This research applied two primary AI techniques for ore grade estimation, the results of which were then compared with those obtained from OK. The AI techniques used include BPNN and ELM. The ELM for ore grade estimation was assessed based on five different activation functions: Radial basis, Sine, Hard limit, Sigmoid and Triangular basis. The OK technique was carried out using Datamine

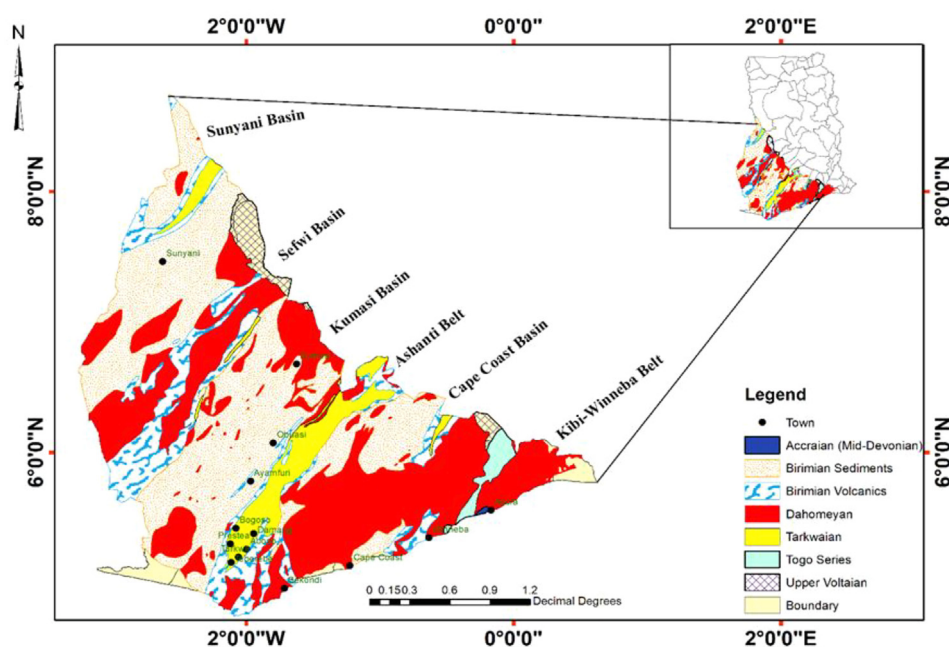


Fig. 1. Geology of the study area.

Table 2. Descriptive statistics of samples.

Field	Assay (g/t)	X Coordinate (m)	Y Coordinate (m)	Z Coordinate (m)
Minimum	0	11 914.702	10 510.265	−1604.298
Maximum	2276	13 704.413	11 540.677	235.266
Range	2276	1789.711	1030.412	1839.564
Mean	4.008	12 876.910	10 991.710	−463.652
Variance	458.731	209 383.708	10 213.987	51 581.702
Standard Deviation	21.418	457.585	101.064	227.116
Standard Error	0.057	0.833	0.184	0.414
Skewness	49.706	−0.257	0.616	−0.237
Kurtosis	3586.004	−0.871	4.209	0.810

software, whereas the AI techniques were carried out using MATLAB and Python programs.

2.3.1. Extreme Learning Machine

ELM is a learning algorithm for a SLFN which was developed by Huang et al. [39]. ELMs work by iteratively tuning parameters within the network and based on the gaussian probability, it randomly chooses its hidden neurons while the Moore–Penrose generalised pseudo inverse is used to analytically determine the output weights of the SLFN [66]. In order to train a SLFN, Eq. (1) [39] is used.

$$\min_{\gamma} \|R(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}})\gamma - Q\| \tag{1}$$

where $\gamma_i = [\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{im}]^Q$ is the output weight vector linking the i th hidden node with the output node, b_i is the threshold of the i th hidden neuron, $w_i = [w_{i1}, w_{i2}, \dots, w_{im}]^Q$ is the weight vector connecting the i th hidden node and the input neuron. In training a SLFN, the least-squares solution is found using Eq. (2) [67].

$$Q = R\gamma \tag{2}$$

The smallest norm of output weight is achieved when $\gamma = R^\dagger Q$, where Q is the least-squares and R is the output matrix of the hidden layer.

Based on the theory behind ELMs, given N random samples (y_i, k_i) where $y_i = [y_{i1}, y_{i2}, \dots, y_{im}]^T \in R^n$ and $k_i = [k_{i1}, k_{i2}, \dots, k_{im}]^T \in R^m$, ($i = 1, \dots, \tilde{N}$) a standard SLFNs o_j with \tilde{N} hidden neurons and activation function $g(y)$ applied in the training of the samples are modelled mathematically using Eq. (3) [44].

$$\sum_{j=1}^{\tilde{N}} \gamma_j g(w_j \cdot y_i + b_j) = o_j \quad j = 1, 2, \dots, \tilde{N} \tag{3}$$

$w_i \cdot y_j$ represents the inner product of the w_i and y_j whereby the output weight w_i is chosen randomly. Eq. (4) represents the output matrix of the hidden layer [44].

$$R = \begin{bmatrix} g(w_1 \cdot y_1 + b_1) & \dots & g(w_{\tilde{N}} \cdot y_1 + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ g(w_1 \cdot y_N + b_1) & \dots & g(w_{\tilde{N}} \cdot y_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \tag{4}$$

Suppose the number of hidden nodes \tilde{N} is equivalent to the number N of various training data points, $\tilde{N} = N$, the matrix R becomes square and invertible after the input weight vectors w_i and the

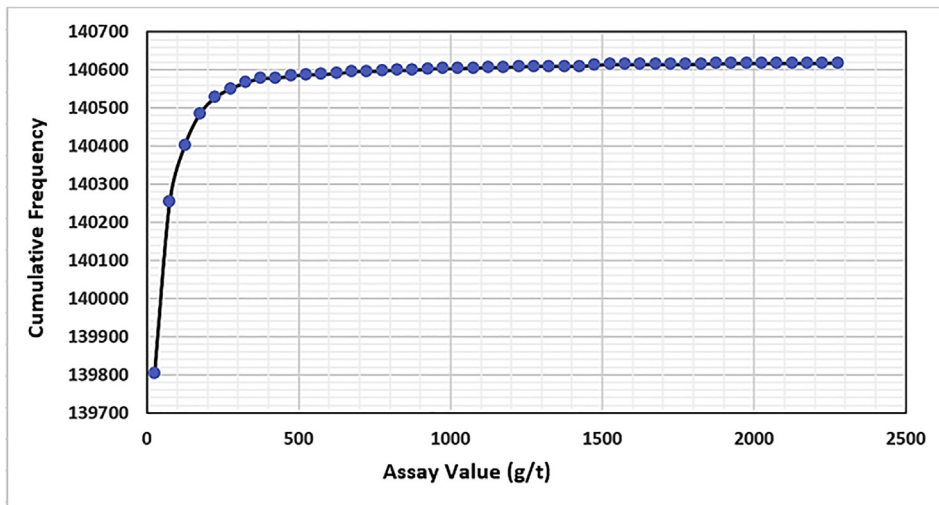


Fig. 2. Cumulative frequency of assay distribution.

biases b_i are selected randomly. Once this happens, SLFNs can estimate the training samples with zero error. None the less N is mostly more than the number of \tilde{N} , different training sample data in practical terms. Detailed mathematical background of ELM is produced in [39]. It is important to note that, the ELM prediction performance is based on the type of activation function used.

The main purpose of an activation function is to determine if a neuron should be activated or not, and is achieved by calculating the sum of weights and the addition of a bias. Thus, the non-linearity of the output node is achieved. If an activation function is not applied to a neural network system, the output will act as a simple linear regression function with limited learning ability [68–70]. The selection of the right activation function in a neural network is crucial since an unsuitable activation function can result in the loss of information from the input parameters during forward propagation and consequently, exponential vanishing gradients during backpropagation [71]. Several types of activation functions can be found in literature, however, the most commonly used such as triangular basis, sigmoid, hard limit, sine and radial basis are applied in this study.

The sigmoid activation function also referred to as the logistic function is non-linear and is widely applied in feedforward neural networks [72,73]. The major advantages of the sigmoid activation function is highlighted by Neal [74], some of which include: easiness to understand and its use in shallow networks. The equation for the sigmoid activation function is shown in Eq. (5) [72]:

$$g(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

The sine activation function is sinusoidal in nature. Hence, it varies from the common activation functions as it rises and falls. The study done by Sopena et al. [75] showed that the sinusoids improve accuracy and shortens training time. Although the sine activation function has been applied, this is rarely used as they are difficult to train [76,77]. It is also saturated as its output converges to zero and flattens as x approaches infinity, it also has numerical problems and converges to local minima [78]. The sine activation function is governed by Eq. (6) [78,79]:

$$f(x) = \sin(x) \quad (6)$$

The Triangular Basis Function (TBF) is a function whose graph is shaped like a triangle, more like an isosceles triangle. It is quite useful in signal processing and when used as an integral transform

function produces more realistic signals. The signals from the function fall within the range of -1 to 1 . The triangular basis activation function is expressed in Eq. (7) [80]:

$$\hat{\theta}_i(y) = \hat{r}\left(\frac{y}{y_i + 1}\right) \quad (7)$$

where:

$$\hat{r}(y) = \begin{cases} 1 - y & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

$\hat{\theta}_i(y)$ is the TBF and y is an independent variable.

Based on the gaussian curve, the Radial Basis Function (RBF) is achieved. RBF applies a parameter which calculates the mean value of a function. RBF is a real-valued function whose observation is solely dependent on the distance from the origin, thus (Eq. (8)) [72]:

$$g(u) = g(\|u\|) \quad (8)$$

otherwise, a distance from another fixed-point, i.e. center c (Eq. (9)) [72] results in:

$$g(u, c) = g(\|u - c\|) \quad (9)$$

Therefore, any function g that satisfies $g(u) = g(\|u\|)$ is a radial function. The Euclidean distance norm and the radial basis function which is commonly taken to be gaussian are merged to obtain an output. The sum of the equations will give Eq. (10) [72] where $y(u)$ is the output and w_i is the weight.

$$y(u) = \sum_{i=1}^N w_i g(\|u - c_i\|) \quad (10)$$

The hard limit function is essentially a transfer function that allows the output neuron to produce a 1 if the input attains a threshold, otherwise, it outputs a 0. It is often used in the perceptron learning rule and as a transfer function, it calculates the output of a layers based on its input. Hard limit activation function is governed by Eq. (11) [81]:

$$\text{Hardlimit}(x) = 0_{x \leq 0} + 1_{x > 0} \quad (11)$$

2.3.2. Backpropagation Neural Network

The AI technique that is commonly used in ore grade estimation is BPNN since it serves as the primary form of neural network [22,26,28,51,55,59]. The basic BPNN structure (Fig. 3) has three (3) layers which is made up of the input, hidden and the output; however, multiple hidden layers are accepted in the BPNN architecture. External input

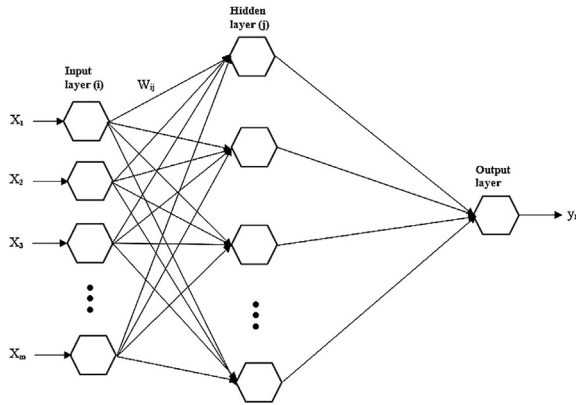


Fig. 3. Basic architecture of BPNN.

parameter are received into the network via the input layer, i.e. X, Y and Z coordinates to each input neuron $X_j = (X_1, X_2, X_3, \dots, X_m)^T$ which are assigned specific weights w_{ij} and a bias b_i (Eq. (12)) [43]. The input values are then transformed into weighted inputs and are transferred to the hidden layer. A mathematical nonlinear activation function is then used to decide if the data in the input neuron should be activated or not after which the transformed data is given out through the output neuron. As shown in Eq. (12), the input of the output layer is obtained from the output of the hidden layer. The linear activation function is used to transform the input of the hidden layer to the output layer which produces the final network output y_i .

$$y_i = f\left(\sum_{j=1}^m (w_{ij}X_j + b_i)\right) \tag{12}$$

Designing a BPNN model involves a critical process of, selecting a suitable number of hidden layers, hidden neurons, training algorithm, and the transfer function. Studies led by several scholars have shown that for solving complicated problems, a BPNN having a single hidden layer is enough as a universal approximator [43,82,83]. Hence, one hidden layer was used in this research. The hyperbolic tangent transfer function was employed in the hidden layer whereas the linear transfer function was applied in the output layer to give out the ore grade value for the BPNN model. The training algorithm applied was the Levenberg–Marquardt optimisation method which is primarily used for solving nonlinear least-squares problems. The Levenberg–Marquardt algorithm works by combining the gradient descent and the gauss-newton methods. The gradient descent technique works by updating the parameters in the steepest direction to reduce the sum of squared errors. On

the contrary, the Gauss-Newton method is applied, by summing the squared errors and reducing it by assuming the least-squares function to be locally quadratic in parameters, thereby finding the minimum quadratic value [84]. The detailed mathematical background of the Levenberg–Marquardt can be found in [84].

2.3.3. Ordinary kriging

Geostatisticians use the variogram as a fundamental tool to measure the spatial continuity of the ore grade data. The experimental variogram is the average variability between samples versus the distance between samples [85]. This variogram model is computed using Eq. (13) [86]:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [Z(u_\alpha + h) - Z(u_\alpha)]^2 \tag{13}$$

where: u is a vector of coordinates; $z(u)$ is variable under consideration as a function of spatial location; h is the distance between the two points and expressed as a vector; $N(h)$ is the number of pairs found at distance h apart; and $Z(u + h)$ is the value of a second variable at location h units from u . The spherical model is widely applied in most orebodies (Fig. 4). This model is characterised by Eqs. (14)–(16) [87].

$$\gamma(h) = 0 \text{ for } h = 0 \tag{14}$$

$$\gamma(h) = C_0 + C \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \text{ for } h < a \tag{15}$$

$$\gamma(h) = C_0 + C \text{ for } h \geq a \tag{16}$$

where: a is the range and corresponds to the intuitive idea of the range of influence of the regionalised variables. Beyond this value, samples are no longer

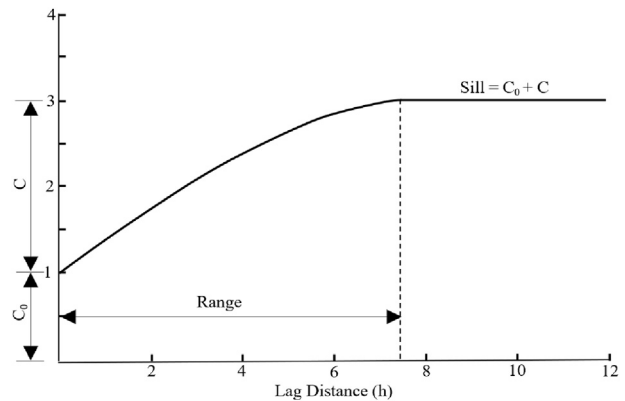


Fig. 4. Spherical semi-variogram model.

auto-correlated; C_0 is the nugget variance and represents the random portion of variations of the regionalised variables; C is the spatial variance and is the predictable/structural part of the spatial variance, and $C + C_0$ is the Sill.

The grade estimation model can be expressed in the matrix form, as shown in Eq. (17) [86].

$$[C][W] = [D] \quad (17)$$

where:

$$[C] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdot & \cdot & \cdot & \sigma_{1n} & 1 \\ \sigma_{21} & \sigma_{22} & \cdot & \cdot & \cdot & \sigma_{2n} & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{n1} & \sigma_{n2} & \cdot & \cdot & \cdot & \sigma_{nn} & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}, \quad [W] = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \\ \lambda \end{bmatrix},$$

$$[D] = \begin{bmatrix} \sigma V v_1 \\ \sigma V v_2 \\ \cdot \\ \cdot \\ \cdot \\ \sigma V v_n \\ 1 \end{bmatrix}$$

In matrix C , σ_{ij} is the covariance between any two data points i and j ; from matrix D , $\sigma V v_i$ is the average covariance between a sample and the block to be estimated, while matrix W , shows a_i which is the weight to be assigned to sample i .

2.4. Statistical evaluation tools

The efficiency of the predicted results of the various models were compared based on their Correlation Coefficient (R), Mean Square Error (MSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2) and Root Mean Squared Error (RMSE) which are shown in Eqs (18)–(22) [28,43,55,88,89].

$$MSE = \frac{1}{N} \left(\sum_{i=1}^n (Z_i - Z_i^*)^2 \right) \quad (18)$$

$$RMSE = \sqrt{\frac{1}{N} \left(\sum_{i=1}^n (Z_i - Z_i^*)^2 \right)} \quad (19)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Z_i - Z_i^*)^2}{\sum_{i=1}^n \left(Z_i - \frac{1}{n} \sum_{i=1}^n Z_i \right)^2} \quad (20)$$

$$R = \frac{\sum_{i=1}^n (Z_i - \bar{Z}_i) (Z_i^* - \bar{Z}_i^*)}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z}_i)^2} \times \sqrt{\sum_{i=1}^n (Z_i^* - \bar{Z}_i^*)^2}} \quad (21)$$

$$MAE = \frac{\sum_{i=1}^n |Z_i - Z_i^*|}{N} \quad (22)$$

where: Z is the actual assay value obtained from drilling; Z^* is the predicted assay value; \bar{Z} is the mean of the actual grade; \bar{Z}^* is the mean of predicted grade; n is the sample; and N is the total number of samples.

2.5. Data preprocessing for model development

For good predictions using ANN, enough data is required. This study adopted the popular hold-out cross-validation technique commonly employed in ANN modelling to divide the data [43]. As such, the training data set should be larger than the testing data set. Hence, 80% of the data representing 241 206 data points were used for training the network, whereas the remaining 20% representing 60 301 data points were used for testing. The data division was done randomly to prevent bias. However, both divisions possessed similar statistical characteristics to guarantee the models generalisation ability to predict ore grade accurately.

During the data preparation, the data sets were first normalised. This was necessary because, the grade values range between 0 and 2000 while the coordinates range between -1600 to 13700 , which fall in different ranges. Moreover, without the data normalisation the large values (coordinates) essentially influence the results due to its more significant value but may not be more important as a predictor. Therefore, the aim was to change the values of the dataset to a standard scale without distorting the differences in range values and improving the model's validation accuracy. In this research, the scaler applied essentially scaled the data to range from -1 to 1 . The normalisation formula used is expressed in Eq. (23) [90].

$$p_i = p_{\min} + \frac{(p_{\max} - p_{\min}) \times (q_i - q_{\min})}{q_{\max} - q_{\min}} \quad (23)$$

where: p_i is the normalised data, q_i is the actual drill hole data, q_{\max} and q_{\min} are the maximum and minimum values of the actual drill hole data with p_{\min} and p_{\max} values set at -1 and 1 , respectively.

3. Results and discussion

3.1. Models developed

3.1.1. ELM model for ore grade estimation

Based on the experimental results, the optimum number of neurons for the developed ELM model was 50. The five activation functions; sigmoid, hard limit, sine, triangular basis and radial basis were applied separately. The optimum ELM model had three input nodes with a single hidden layer made up of 50 hidden neurons and one output with structure [3-50-1]. The three input variables were the X, Y, and Z coordinates while the output was the ore grade.

3.1.2. BPNN model for ore grade estimation

The input and output data sets used in the ELM model was the same used to develop the BPNN model. Three layers comprising input, hidden and output layers made up the developed BPNN model. As demonstrated in literature, a single hidden layer is capable of approximating any complex problem [82], hence, one hidden layer was employed in this study. The hyperbolic tangent and linear transfer functions were employed in the hidden and output layers to capture both non-linearity and linearity between the input–output data. In training the

BPNN model, the Levenberg–Marquardt algorithm was applied [91]. The best BPNN model obtained in this research has three input nodes, ten neurons in the hidden layer and a single output node, with the structure [3-10-1].

3.1.3. OK model for ore grade estimation

As shown in Table 2 the grade distribution shows positive skewness with possible outliers. Therefore, there was the need to apply a top cut to minimise the influence of outliers on the mean and the skewness. The top cut value was based on the log probability curves' (Fig. 5) analysis showing inflexion at specific points, which indicate subpopulation. Hence, the bottom cut value is shown to be at 0.01 g/t, but due to the high-grade values in the data, the top cut value is at 15 g/t. Since the deposit is located in the Birimian, and due to the high-grade records in that structure, geologists usually consider the top cut value to be 12 g/t. However, in this research, 15 g/t is considered the top cut value based on the observed inflexion (Fig. 5).

Structural analyses were first performed on the data. The experimental variogram was generated using the exploratory data with a top cut value of 15 g/t. Based on the experimental points at various lag distances, the variogram parameters were then obtained. The nugget variance $C_0 = 13.4$, spatial variance $C = 9.4$ and range a in the X, Y and Z directions were 21.2, 15.03 and 72.54 respectively. These were obtained as the model parameters from the experimental variogram. The spherical model was then superimposed onto the experimental variogram based on the experimental variogram parameters (Fig. 6).

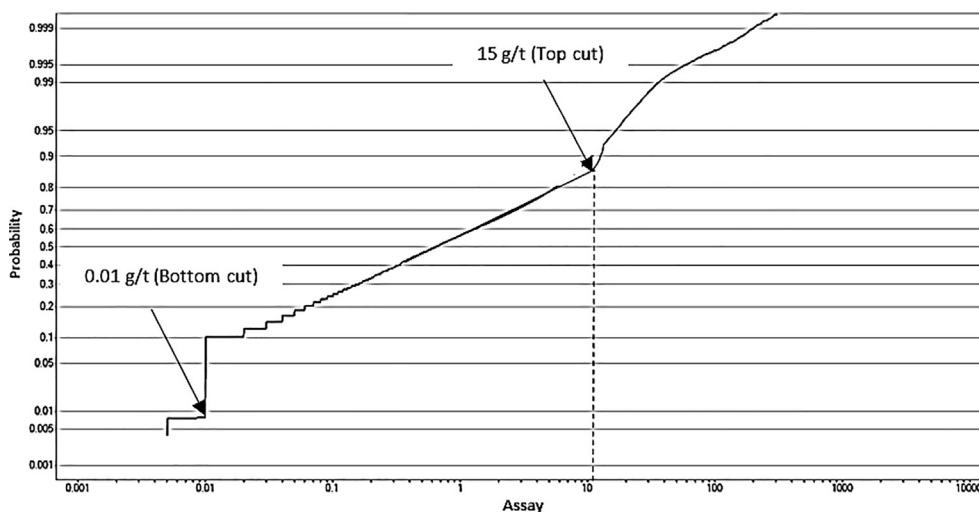


Fig. 5. Log probability plot for assay data.

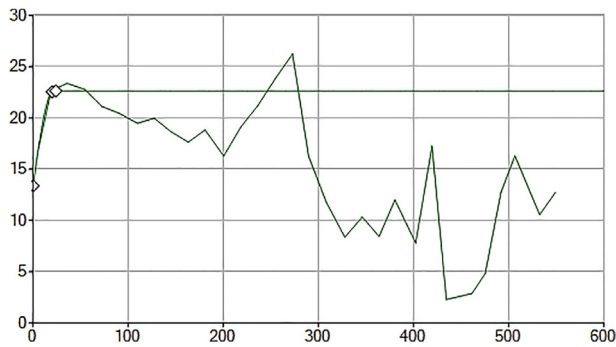


Fig. 6. Directional variogram across strike at 60° .

Based on the results obtained from the structural analyses, ore grade estimation using OK was then conducted. The performance of OK in the ore grade prediction is summarised in Table 4 based on the statistical evaluation tools.

3.2. Comparative analysis of ELM variants

In considering the dimensioned error statistic indicators (MSE, MAE and RMSE) as shown in Table 3, it was observed that the estimation technique, ELM-Sigmoid model obtained the lowest MAE, MSE and RMSE values of 0.0175, 0.000524 and 0.022900 correspondingly. These results (MSE, MAE and RMSE) were interpreted based on the rule of thumb that states that, for a model to approximate closely to the actual data, the error values are closer to zero. The next technique that performed nearly as well as ELM-Sigmoid was ELM-Sine which was

closely followed by ELM-Radial basis. The ELM-Triangular basis performed fairly while the ELM-Hard limit gave the poorest results.

The R^2 (Table 3) indicates that the nearer the value is to 1 or 100% the better the predicted results. In effect this shows the extent the model applied could explain the prediction variation level from the model compared with actual data. It is observed in Table 3 that the ELM-Sigmoid had the highest R^2 value of 91.93%, followed by ELM-Sine, 87.91%. The subsequent models: ELM-Radial basis; and ELM-Triangular basis performed fairly with R^2 values measuring 68.86% and 55.69% while the ELM-Hard limit performed poorly with R^2 of 33.37%.

The statistical tool that measures the strength between the relationship of two variables, i.e., actual and estimated values is known as the correlation coefficient R , and they fall within the range of -1 to 1 . In effect it explains the level of prediction accuracy of the model. In Table 3, it is noticed that the ELM-Sigmoid and ELM-Sine had R values exceeding 0.9, making them perform better than the other models, however, ELM-Sigmoid had the highest R of 0.96. On the contrary, ELM-Radial basis had R of 0.83 while that of ELM-Triangular basis and ELM-Hard limit was 0.75 and below 0.6 respectively. The presented ELM test results are further illustrated in Fig. 7 ((a)-(e)) for visual observation. To this end, it can be stated that the ELM-Sigmoid has demonstrated strong calibration power and best generalisation on the training and testing data with great adaptability as compared to the other models.

Table 3. Optimal train and test results for ELM variants.

Technique	Train					Test				
	Performance Criteria					Performance Criteria				
	MAE	MSE	RMSE	R^2	R	MAE	MSE	RMSE	R^2	R
ELM-Hard limit	0.0447	0.0043	0.0657	0.3432	0.5858	0.0448	0.0043	0.0657	0.3337	0.5777
ELM-Triangular basis	0.0176	0.0029	0.0539	0.5573	0.7465	0.0360	0.0029	0.0536	0.5569	0.7463
ELM-Radial basis	0.0320	0.0021	0.0454	0.6868	0.8287	0.0320	0.0020	0.0449	0.6886	0.8298
ELM-Sine	0.0209	0.0008	0.0285	0.8767	0.9363	0.0212	0.0008	0.0280	0.8791	0.9376
ELM-Sigmoid	0.0172	0.0005	0.0229	0.9201	0.9592	0.0175	0.0005	0.0229	0.9193	0.9588

Table 4. Optimal train and test results.

Technique	Train					Test				
	Performance Criteria					Performance Criteria				
	MAE	MSE	RMSE	R^2	R	MAE	MSE	RMSE	R^2	R
OK	1.6473	7.0565	2.6564	0.5392	0.7343	1.6473	7.0565	2.6564	0.5392	0.7343
BPNN	0.0272	0.0022	0.0464	0.5725	0.7566	0.0555	0.0054	0.0735	0.2866	0.5354
ELM-Sigmoid	0.0172	0.0005	0.0229	0.9201	0.9592	0.0175	0.0005	0.0229	0.9193	0.9588

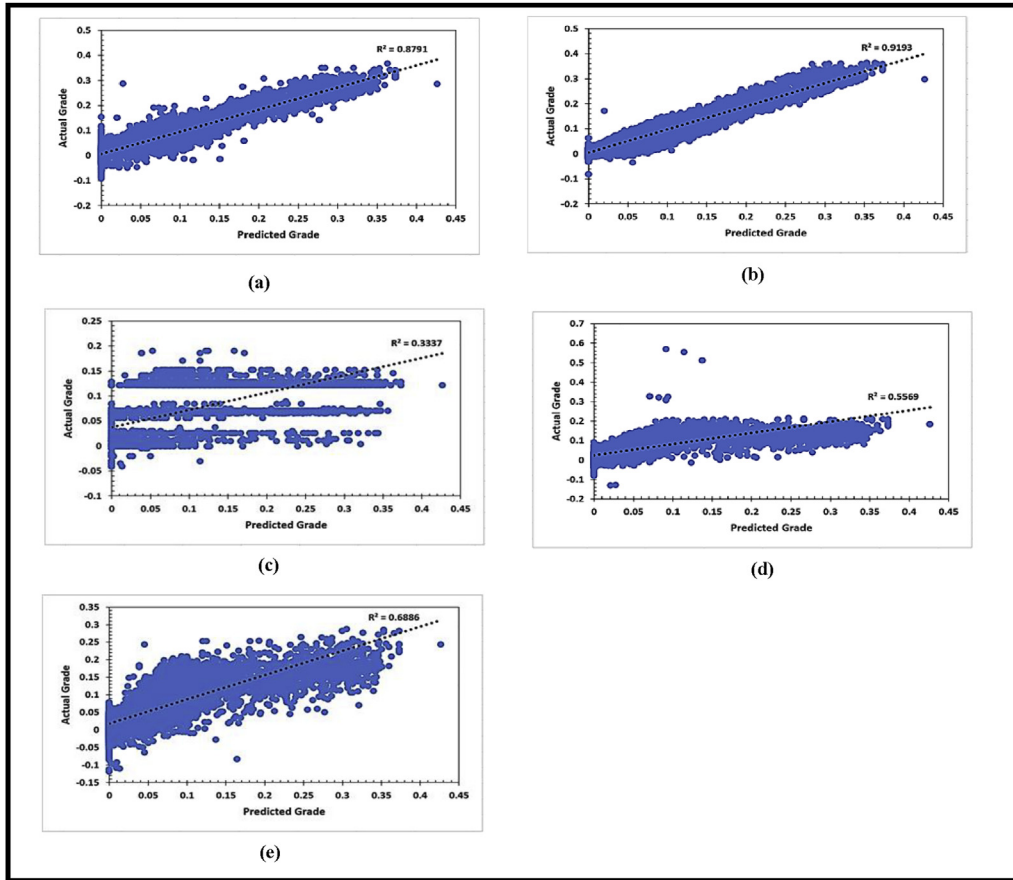


Fig. 7. Test results: (a) ELM-sine; (b) ELM-Sigmoid; (c) ELM-Hard limit; (d) ELM-Triangular basis; (e) ELM-Radial basis.

3.3. Performance evaluation of ELM-sigmoid with other investigated techniques

The ELM-Sigmoid was the best performing model out of the various ELM methods applied in this

study. Hence the ELM-Sigmoid is evaluated against state-of-the-art methods of OK and BPNN. The comparison (Table 4) aimed to determine if the proposed ELM-Sigmoid technique could produce comparable or superior results to those obtained

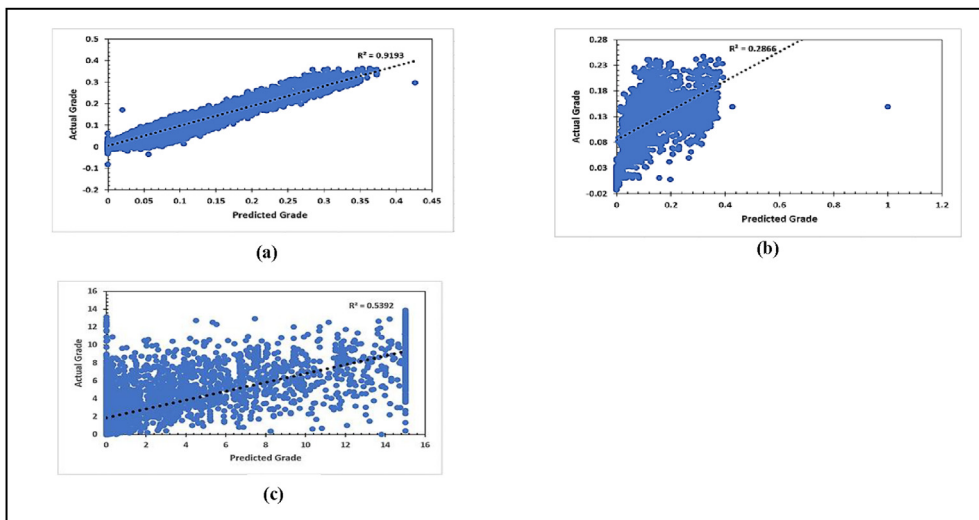


Fig. 8. Test results: (a) ELM-sigmoid, (b) BPNN and (c) OK.

from the state-of-the-art established BPNN and OK methods.

Comparatively, ELM-Sigmoid was superior over BPNN and OK as it recorded the lowest errors and had the best performing R^2 and R values. BPNN had MAE, MSE and RMSE of 0.0555, 0.0054 and 0.0735 while OK recorded 1.4127, 6.1599, and 2.4819 respectively. The ELM-Sigmoid had 0.0175, 0.0005, 0.0229, 0.9193, 0.9588 for MAE, MSE, RMSE, R^2 and R . A visual comparison of the various models based on the actual and predicted grade values are shown in Fig. 8. The interpretation is that the ELM-Sigmoid model produced a better fitted values to the actual ore grade data than the other methods. The strength of the ELM-Sigmoid comes from the fact that there is less manual tasking (human interference) and minimum fine-tuning adjustable parameters in the model development process. Moreover, optimum predictions were achieved because the ELM technique is not gradient descent based type of algorithm which can trap in local minima and therefore produces global best solutions.

4. Conclusions

In this study, ELM has successfully been applied in ore grade prediction of heterogeneous data sets. Five variants of the ELM based on triangular basis, radial basis, hard limit, sine and sigmoid activation functions were developed and tested with data from a mine in Ghana. The proposed ELM techniques were then compared with state-of-the-art established methods of BPNN and the conventional OK. ELM-Sigmoid model generated comparable grade predictions similar to the actual grade values than the other models applied. Thus, the ELM-Sigmoid gave the lowest MAE, MSE and RMSE values of 0.0175, 0.0005 and 0.0229 and highest R^2 and R of 0.9193 and 0.9588 respectively. Based on the results obtained, it was decided that the proposed ELM-Sigmoid model has shown encouraging application potential in ore grade estimation and therefore serve as a suitable alternative to the established BPNN and OK models employed in this research. The proposed ELM-Sigmoid model's efficiency was attributed to its inherent ability to randomly select its weights and biases, faster computational speed, less manual tasking in the model development and producing global minimum solutions because it is not a gradient descent type of algorithm where it can trap in local minima.

The developed model for the studied Mine could be adopted as an alternative ore grade estimation tool since OK method which is usually employed produced poor estimates due to the heterogeneous

nature of the deposit. Furthermore, the proposed methodology can definitely be replicated for other deposits. This is because it has proven to possess excellent generalisation ability and self-adaptive characteristic feature where it can automatically learn on any given dataset from any Mine.

Ethical statement

The authors state that the research was conducted according to ethical standards.

Funding body

This research was funded by Ghana National Petroleum Corporation (GNPC) Professorial Chair in Mining Engineering at the University of Mines and Technology (UMaT), Ghana.

Conflicts of interest

None declared.

References

- [1] Sinclair AJ, Deraisme J. Geostatistical study of the eagle copper vein, northern British Columbia. *Cim Bull* 1974;746: 131–42.
- [2] Dominy SC, Camm GS, Phelps RFG. Narrow vein mining-A challenge to the operator. *Mine Plan Equip Select* 1997;97: 125–32.
- [3] Roy WD. Using geostatistics to estimate the resources of a narrow vein gold deposit61. Nova Scotia: DalTech; 2000. PhD Dissertation.
- [4] Dominy SC, Noppé MA, Annels AE. Errors and uncertainty in mineral resource and ore reserve estimation: the importance of getting it right. *Explor Min Geol* 2004;11(1–4):77–98. <https://doi.org/10.2113/11.1-4.77>.
- [5] Dominy SC, Annels AE, Camm GS, Wheeler P, Barr SP. Geology in the resource and reserve estimation of narrow vein deposits. *Explor Min Geol* 1999;6(4):317–33.
- [6] Journel AG. Nonparametric estimation of spatial distributions. *J Int Assoc Math Geol* 1983;15(3):445–68. <https://doi.org/10.1007/BF01031292>.
- [7] Arslan H. Spatial and temporal mapping of groundwater salinity using ordinary kriging and indicator kriging: the case of Bafra Plain, Turkey. *Agric Water Manag* 2012;113:57–63. <https://doi.org/10.1016/j.agwat.2012.06.015>.
- [8] Yates SR, Warrick AW, Myers DE. Disjunctive Kriging: 1 Overview of estimation and conditional probability. *Water Resour Res* 1986;22(5):623–30. <https://doi.org/10.1029/WR022i005p00623>.
- [9] Daya AA. Application of disjunctive kriging for estimating economic grade distribution in an iron ore deposit: a case study of the Choghart North Anomaly, Iran. *J Geol Soc India* 2014;83(5):567–76. <https://doi.org/10.1007/s12594-014-0085-3>.
- [10] Mendes MP, Ribeiro L. Nitrate probability mapping in the northern aquifer alluvial system of the river Tagus (Portugal) using Disjunctive Kriging. *Sci Total Environ* 2010;408(5): 1021–34. <https://doi.org/10.1016/j.scitotenv.2009.10.069>.
- [11] Verly G. The multigaussian approach and its applications to the estimation of local reserves. *J Int Assoc Math Geol* 1983; 15(2):259–86.
- [12] Afzal P, Madani N, Shahbeik S, Yasrebi AB. Multi-Gaussian kriging: a practice to enhance delineation of mineralised

- zones by Concentration-Volume fractal model in Dardevey iron ore deposit, SE Iran. *J Geochem Explor* 2015;158:10–21. <https://doi.org/10.1016/j.gexplo.2015.06.011>.
- [13] Journel AG. Recoverable reserves estimation-the geostatistical approach. *Min Eng* 1985;37(6):563–8.
- [14] Verly G, Sullivan J. Multigaussian and probability krigings-application to the jerritt canyon deposit. *Min Eng* 1985;37(6):568–74.
- [15] Adhikary PP, Dash CJ, Bej R, Chandrasekharan H. Indicator and probability kriging methods for delineating Cu, Fe, and Mn contamination in groundwater of Najafgarh Block, Delhi, India. *Environ Monit Assess* 2011;176(1–4):663–76. <https://doi.org/10.1007/s10661-010-1611-4>.
- [16] Armstrong M, Boufassa A. Comparing the robustness of ordinary kriging and lognormal kriging: outlier resistance. *Math Geol* 1988;20(4):447–57. <https://doi.org/10.1007/BF00892988>.
- [17] Yamamoto JK. On unbiased backtransform of lognormal kriging estimates. *Comput Geosci* 2007;11(3):219–34. <https://doi.org/10.1007/s10596-007-9046-x>.
- [18] Arik A. Outlier restricted kriging: a new kriging algorithm for handling of outlier high grade data in ore reserve estimation. In: Proceedings of the 23rd international symposium on the application of computers and operations research in the Minerals industry. Littleton: Society of Mining Engineers; 1992. p. 181–7.
- [19] Leuangthong O, Nowak M. Dealing with high-grade data in resource estimation. *J S Afr Inst Min Metall* 2015;115(1):27–36. <https://doi.org/10.17159/2411-9717/2015/v115n1a3>.
- [20] Wu X, Zhou Y. Reserve estimation using neural network techniques. *Comput Geosci* 1993;19(4):567–75. [https://doi.org/10.1016/0098-3004\(93\)90082-G](https://doi.org/10.1016/0098-3004(93)90082-G).
- [21] Li XL, Xie YL, Guo QJ, Li LH. Adaptive ore grade estimation method for the mineral deposit evaluation. *Math Comput Model* 2010;52(11–12):1947–56. <https://doi.org/10.1016/j.mcm.2010.04.018>.
- [22] Al-Alawi SM, Tawo EE. Application of artificial neural networks in mineral resource evaluation. *J King Saud Univ Eng Sci* 1998;10(1):127–38.
- [23] Kapageridis I, Denby B. In: Panagiotou GN, Michalakopoulos TN, editors. Ore grade estimation with modular neural network systems. A case study. *Information Technology in the Mineral Industry*. vol. 52. Rotterdam: AA Balkema Publ; 1998.
- [24] Kapageridis IK. Grade interpolation using radial basis function networks. *Mine Plan Equip Select* 2002:1–9.
- [25] Samanta B, Bandopadhyay S, Ganguli R, Dutta S. A comparative study of the performance of single neural network vs. Adaboost algorithm based combination of multiple neural networks for mineral resource estimation. *J S Afr Inst Min Metall* 2005;105(4):237–46.
- [26] Mahmoudabadi H, Izadi M, Menhaj MB. A hybrid method for grade estimation using genetic algorithm and neural networks. *Comput Geosci* 2008;13(1):91–101. <https://doi.org/10.1007/s10596-008-9107-9>.
- [27] Tahmasebi P, Hezarkhani A. Application of adaptive neuro-fuzzy inference system for grade estimation; case study, sarcheshmeh porphyry copper deposit, Kerman, Iran. *Aust J Basic Appl Sci* 2010;4(3):408–20.
- [28] Zhang X, Song S, Li J, Wu C. Robust LS-SVM regression for ore grade estimation in a seafloor hydrothermal sulphide deposit. *Acta Oceanol Sin* 2013;32(8):16–25. <https://doi.org/10.1007/s13131-013-0337-x>.
- [29] Badel M, Angorani S, Panahi MS. The application of median indicator kriging and neural network in modeling mixed population in an iron ore deposit. *Comput Geosci* 2010;37(4):530–40. <https://doi.org/10.1016/j.cageo.2010.07.009>.
- [30] Singh RK, Ray D, Sarkar BC. Recurrent neural network approach to mineral deposit modelling. In: 4th int. Conference on recent advances on information Technology. 1–5; 2018.
- [31] Tahmasebi P, Hezarkhani A. Application of a modular feedforward neural network for grade estimation. *Nat Resour Res* 2011;20(1):25–32. <https://doi.org/10.1007/s11053-011-9135-3>.
- [32] Gholamnejad J, Kasmaee S, Kohsary A, Nezamolhosseini. A grade estimation of ore stockpiles by using artificial neural networks: case study on choghart iron mine in Iran. *Int J Min Miner Eng* 2012;4(1):17–25. <https://doi.org/10.1504/IJMME.2012.047997>.
- [33] Chatterjee S, Bandopadhyay S, Machuca D. Ore grade prediction using a genetic algorithm and clustering Based ensemble neural network model. *Math Geosci* 2010;42(3):309–26. <https://doi.org/10.1007/s11004-010-9264-y>.
- [34] Jafrasteh B, Fathianpour N, Suárez A. Comparison of machine learning methods for copper ore grade estimation. *Comput Geosci* 2018;22(5):1371–88. <https://doi.org/10.1007/s10596-018-9758-0>.
- [35] Chatterjee S, Bhattacharjee A A, Samanta B, Pal SK. Ore grade estimation of a limestone deposit in India using an artificial neural network. *Appl GIS* 2006;2(1):1–20. <https://doi.org/10.2104/ag060003>.
- [36] Dutta S, Bandopadhyay S, Ganguli R, Misra D. Machine learning algorithms and their application to ore reserve estimation of sparse and imprecise data. *J Intell Learn Syst Appl* 2010;2(2):86–96. <https://doi.org/10.4236/jilsa.2010.22012>.
- [37] Li XL, Li LH, Zhang BL, Guo QJ. Hybrid self-adaptive learning based particle swarm optimisation and support vector regression model for grade estimation. *Neurocomputing* 2013;118:179–90.
- [38] Samanta S, Bandopadhyay S, Ganguli R. Data segmentation and genetic algorithms for sparse data division in nometallic placer gold grade estimation using neural network and geostatistics. *Explor Min Geol* 2004;11(1–4):69–76.
- [39] Huang GB, Zhu QY, Siew CK. Extreme learning machine: a new learning scheme of feedforward neural networks. *IEEE Int Joint Conf Neural Network* 2004;2:985–90. <https://doi.org/10.1109/IJCNN.2004.1380068>.
- [40] Huang Y. Advances in artificial neural networks - methodological development and application. *Algorithms* 2009;2(3):973–1007. <https://doi.org/10.3390/algorithm2030973>.
- [41] Adeel A, Larijani H, Javed A, Ahmadinia A. Impact of learning algorithms on random neural network based optimisation for LTE-UL systems. *Netw Protoc Algorithm* 2015;7(3):157–78. <https://doi.org/10.5296/npa.v7i3.8295>.
- [42] Liu J, Yan K, Zhao X, Hu Y. Prediction of autogenous shrinkage of concretes by support vector machine. *Int J Pavement Res Technol* 2016;9(3):169–77. <https://doi.org/10.1016/j.ijprt.2016.06.003>.
- [43] Arthur CK, Temeng VA, Ziggah YY. Novel approach to predicting blast-induced ground vibration using Gaussian process regression. *Eng Comput* 2020;36(1):29–42. <https://doi.org/10.1007/s00366-018-0686-3>.
- [44] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. *Neurocomputing* 2006;70(1–3):489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>.
- [45] Poerio DV, Brown SD. Localized and adaptive soft sensor based on an extreme learning machine with automated self-correction strategies. *J Chemometr* 2020;34(7):1–18. <https://doi.org/10.1002/cem.3088>.
- [46] Dash R, Dash PK, Bisoi R. A self adaptive differential harmony search based optimised extreme learning machine for financial time series prediction. *Swarm Evol Comput* 2014;19:25–42. <https://doi.org/10.1016/j.swevo.2014.07.003>.
- [47] Kapageridis IK, Denby B. Neural network modelling of ore grade spatial variability. In: International conference on artificial neural networks; 1998. p. 209–14.
- [48] Kapageridis IK, Denby B, Hunter G. GEMNET II – a neural ore grade estimation system. In: 29th international symposium on the application of computers and operations research in the minerals industries; 1999. 10pp.
- [49] Kapageridis IK, Denby B, Hunter G. Integration of a neural ore grade estimation tool in a 3D resource modeling package. *Proc Int Jt Conf Neural Netw* 1999;6:3908–12.
- [50] Kapageridis IK, Denby B, Schofield D. GEMNET II – an alternative method for grade estimation. 2000. p. 1–8.

- [51] Matías JM, Vaamonde A, Taboada J, González-Manteiga W. Comparison of kriging and neural networks with application to the exploitation of a slate mine. *Math Geol* 2004;36(4): 463–86.
- [52] Samanta B, Bandopadhyay S, Ganguli R, Dutta S. Sparse data division using data segmentation and kohonen network for neural network and geostatistical ore grade modeling in nome offshore placer deposit. *Nat Resour Res* 2004;13(3): 189–200.
- [53] Samanta B, Ganguli R, Bandopadhyay S. Comparing the predictive performance of neural networks with ordinary kriging in a bauxite deposit Comparing the predictive performance of neural networks with ordinary kriging in a bauxite deposit. *Min Technol* 2005;114(3):A129–39.
- [54] Badel M, Angorani S, Panahi MS. The application of median indicator kriging and neural network in modeling mixed population in an iron ore deposit. *Comput Geosci* 2011;37(4): 530–40. <https://doi.org/10.1016/j.cageo.2010.07.009>.
- [55] Guo WW. A novel application of neural networks for instant iron-ore grade estimation. *Expert Syst Appl* 2010;37(12): 8729–35. <https://doi.org/10.1016/j.eswa.2010.06.043>.
- [56] Dutta S, Misra D, Ganguli R, Samanta B, Bandopadhyay S. A hybrid ensemble model of kriging and neural network for ore grade estimation. *Int J Surf Min Reclam Environ* 2006; 20(1):33–45. <https://doi.org/10.1080/13895260500322236>.
- [57] Tahmasebi P, Hezarkhani A. Comparison of optimised neural network with fuzzy logic for ore grade estimation. *Austr J Basic Appl Sci* 2010;4(5):764–72.
- [58] Tahmasebi P, Hezarkhani A. A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. *Comput Geosci* 2012;42:18–27. <https://doi.org/10.1016/j.cageo.2012.02.004>.
- [59] Maleki S, Ramazia HR, Moradi S. Estimation of Iron concentration by using a support vector machine and an artificial neural network-the case study of the Choghart deposit southeast of Yazd, Yazd, Iran. *Geopersia* 2014;4(2):201–12.
- [60] Granek J. Application of machine learning algorithms to mineral prospectivity mapping. Vancouver: University of British Columbia; 2016. p. 105. <https://doi.org/10.14288/1.0340340>. PhD Dissertation.
- [61] Jafarsteh B, Fathianpour N. A hybrid simultaneous perturbation artificial bee colony and back-propagation algorithm for training a local linear radial basis neural network on ore grade estimation. *Neurocomputing* 2017;235:217–27. <https://doi.org/10.1016/j.neucom.2017.01.016>.
- [62] Jahangiri M, Ghavami Riabi SR, Tokhmechi B. Estimation of geochemical elements using a hybrid neural network-Gustafson-Kessel algorithm. *J Min Environ* 2018;9(2):499–511. <https://doi.org/10.22044/jme.2017.5513.1363>.
- [63] Manna B, Samanta B, Chakravarty D, Dutta D, Chowdhury A, Santra A, et al. Hyperspectral signature analysis using neural network for grade estimation of copper ore. *IOP Conf Ser Earth Environ Sci* 2018;169(1):1–12.
- [64] Smith AJ, Henry G, Frost-Killian S. A review of the birimian supergroup-and tarkwaian group-hosted gold deposits of Ghana. *Episodes* 2016;39(2):177–97. <https://doi.org/10.18814/epiiugs/2016/v39i2/95775>.
- [65] Milési JP, Ledru P, Ankrah P, Johan V, Marcoux E, Vinchon C. The metallogenic relationship between Birimian and Tarkwaian gold deposits in Ghana. *Miner Deposita* 1991; 26(3):228–38. <https://doi.org/10.1007/BF00209263>.
- [66] Cao J, Lin Z, Huang GB. Self-adaptive evolutionary extreme learning machine. *Neural Process Lett* 2012;36(3):285–305. <https://doi.org/10.1007/s11063-012-9236-y>.
- [67] Liang NY, Huang GB, Saratchandran P, Sundararajan NA. A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Trans Neural Network* 2006; 17(6):1411–23. <https://doi.org/10.1109/TNN.2006.880583>.
- [68] Sakthivadivel DA. Formalising the use of the activation function in neural inference. arXiv; 2021. p. 1–7. preprint arXiv: 2102.04896.
- [69] Dushkoff M, Ptucha R. Adaptive activation functions for deep networks. *Electron Imag* 2016;19:1–5. <https://doi.org/10.2352/ISSN.2470-1173.2016.19.COIMG-149>.
- [70] Wang Y, Li Y, Song Y, Rong X. The influence of the activation function in a convolution neural network model of facial expression recognition. *Appl Sci* 2020;10(5):1–20. <https://doi.org/10.3390/app10051897>.
- [71] Hayou S, Doucet A, Rousseau J. On the impact of the activation function on deep neural networks training. In: International conference on machine learning; 2019. p. 2672–80.
- [72] Karlik B, Olgac AV. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Int J Artif Intell Expet Syst* 2011;1(4):111–22.
- [73] Nwankpa C, Ijomah W, Gachagan A, Marshall S S. Activation Functions: comparison of trends in practice and research for deep learning1–20. arXiv; 2018. Prepr. arXiv1811.03378.
- [74] Neal RM. Connectionist learning of belief networks. *Artif Intell* 1992;56:71–113.
- [75] Sopena JM, Romero E, Alquezar R. Neural networks with periodic and monotonic activation functions: a comparative study in classification problems. *Inst Elec Eng Conf Pub* 1999; 1(470):323–8. <https://doi.org/10.1049/cp:19991129>.
- [76] Wong KW, Leung CS, Chang SJ. Handwritten digit recognition using multi-layer feedforward neural networks with periodic and monotonic activation functions. In: IEEE: In Object recognition supported by user interaction for service robots. 16; 2002. p. 106–9. <https://doi.org/10.1109/icpr.2002.1047806> (3).
- [77] McCaughan DB. On the properties of a periodic perceptrons. In: Proceeding of international conference on neural networks. vol. 1; 1997. p. 188–93. <https://doi.org/10.1007/BF01023055>.
- [78] Lapedes A, Farber R. Nonlinear signal processing using neural networks: prediction and system modelling. Los Alamos National Laboratory: Los Alamos; 1987. p. 1–50.
- [79] Gashler MS, Ashmore SC. Modeling time series data with deep Fourier neural networks. *Neurocomputing* 2016;188: 3–11. <https://doi.org/10.1016/j.neucom.2015.01.108>.
- [80] Melkonian D. Similar basis function algorithm for numerical estimation of Fourier integrals. *Numer Algorithm* 2010;54(1): 73–100. <https://doi.org/10.1007/s11075-009-9324-x>.
- [81] Huang G, Mao K, Siew CK, Huang DS. Fast modular network implementation for support vector machines. *IEEE Trans Neural Network* 2005;16(6):1651–63. <https://doi.org/10.1109/TNN.2005.857952>.
- [82] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Network* 1989; 2(5):359–66.
- [83] Arthur CK, Temeng VA, Ziggah YY. Soft computing-based technique as a predictive tool to estimate blast-induced ground vibration. *J Sustain Min* 2019;18(4):287–96. <https://doi.org/10.1016/j.jsm.2019.10.001>.
- [84] Gavin HP. The levenburg-marquardt algorithm for nonlinear least squares curve-fitting problems. Durham: Department of Civil and Environmental Engineering, Duke University; 2020. p. 19.
- [85] Bhunia GS, Shit PK, Chattopadhyay R. Assessment of spatial variability of soil properties using geostatistical approach of lateritic soil (West Bengal, India). *Annal Agra Sci* 2018;16(4): 436–43. <https://doi.org/10.1016/j.aasci.2018.06.003>.
- [86] Mert BA, Dag A. A computer program for practical semi-variogram modeling and ordinary kriging: a case study of porosity distribution in an oil field. *Open Geosci* 2017;9(1): 663–74. <https://doi.org/10.1515/geo-2017-0050>.
- [87] Sinclair AJ, Blackwell GH. Applied mineral inventory estimation. Cambridge: Cambridge University Press; 2006. p. 381.
- [88] Nezamolhosseini SA, Mojtahedzadeh SH, Gholamnejad J. The application of artificial neural networks to ore reserve estimation at choghart iron ore deposit. *Anal Numeric Meth Min Eng* 2017;6:73–83.
- [89] Asante-Okyere S, Shen C, Ziggah YY, Rulegeya MM, Zhu X. Investigating the predictive performance of Gaussian process regression in evaluating reservoir porosity and

- permeability. *Energies* 2018;11(12):1–13. <https://doi.org/10.3390/en11123261>.
- [90] Mueller AV, Hemond HF. Extended artificial neural networks: incorporation of a priori chemical knowledge enables use of ion selective electrodes for in-situ measurement of ions at environmentally relevant levels. *Talanta* 2013;117: 112–8. <https://doi.org/10.1016/j.talanta.2013.08.045>.
- [91] Moré JJ. The Levenberg-Marquardt algorithm: implementation and theory. In: *Numerical analysis*; 1978. p. 105–16. <https://doi.org/10.1007/bfb0067700>.