Research Article

Research on TFe Content of Hematite Based on LU-TELM-SOA and Selection of Band

Yanhua Fu,[1] Lushan Wan,[2] Yachun Mao,[3] Tao Ren,[4] and Dong Xiao [2]

[1]*JangHo Architecture College, Northeastern University, Shenyang 110169, China*
[2]*College of Information Science and Engineering, Northeastern University, Shenyang 110819, China*
[3]*School of Resources and Civil Engineering, Northeastern University, Shenyang 110819, China*
[4]*Software College, Northeastern University, Shenyang 110169, China*

Correspondence should be addressed to Dong Xiao; xiaodong@ise.neu.edu.cn

Iron ore is an important raw material for the steel industry, so it is of great economic significance to determine the grade of the iron ore quickly and accurately. And the TFe content is the main indicator that determines the grade of the iron ore and whether the iron ore can be smelted directly. Unlike manual methods and methods for chemical analysis, the paper uses the selection of band for the near-infrared spectrum based on the pruning method and the two-hidden-layer extreme learning machine based on LU decomposition and seagull optimization algorithm (LU-TELM-SOA) to identify the TFe content. First of all, the paper proposes the selection of band based on the pruning method to retain the sensitive band of the near-infrared spectrum. Aiming at the problems of poor stability and low accuracy of a single LU-TELM (the two-hidden-layer extreme learning machine based on LU decomposition) model, the paper proposes LU-TELM-SOA. The experimental results show that LU-TELM-SOA has the advantages of high accuracy and strong stability.

1. Introduction

Steel is an important metal material that is indispensable for the industrial development of the country and determines the industrial level and technological level. Among them, the iron ore is the raw material of steel. At present, the iron ore with high industrial value includes the magnetite ore and the hematite ore. The chemical formula of hematite is Fe_2O_3, and its theoretical TFe content is 70%. Hematite is the most important industrial iron ore in the world, accounting for more than 60% of the total reserves in the iron ore. The iron ore is widely distributed, but the total reserves of Australia, Russia, and Brazil account for more than half of the global reserves. The TFe content of the iron ore is the main indicator that determines the grade of the iron ore and whether it can be directly smelted. The TFe content of the iron ore used in industrial production ranges from approximately 23% to 70%. The iron ore used in steel mills requires an iron grade of 55% or more and low sulfur, phosphorus, and other elements. The traditional methods for determining the TFe content include chemical titration analysis, instrumental analysis, and X-ray fluorescence spectrometry, but they are time-consuming and labor-intensive. Therefore, the paper uses the near-infrared spectrum and extreme learning machine (ELM) to propose a new method that can quickly and accurately detect the TFe content, which is mainly applied to hematite ore. And the near-infrared spectrum is an electromagnetic radiation wave between visible light (Vis) and mid-infrared (MIR), and the spectral analysis is used to identify substances and determine their chemical composition and relative content based on their spectrum.

The neural networks have strong learning ability, the ability to approximate nonlinear functions, and good fault tolerance and parallelism. The feedforward neural network with s simple structure is one of the types of neural networks and can approximate any function with arbitrary precision when the number of neurons in the hidden layer is large enough. The simplest structure of the feedforward neural network is the single-hidden layer feedforward network (SLFN). There are many types of SLFN, including the back-

propagation algorithm (BP) and its improved algorithm. Although SLFN is widely used, the parameters of SLFN need to be adjusted iteratively by using the gradient descent algorithm. When the training data is large, the gradient descent algorithm will lead to slow learning speed and long training time. Lowe [1] pointed out that the center of the radial basis function neural network (RBF) can be randomly selected and proposed the idea of randomly selecting SLFN parameters for the first time. And the random vector functional-link net (RVFL) [2] only needs to calculate the output weights. ELM [3, 4] was proposed based on SLFN randomly which selects the input weights and thresholds of the hidden layer and calculated the output weights by using the least square method and the Moore-Penrose generalized inverse. The parameter setting is simple, the ability of generalization is better than the traditional SLFN algorithm, and the training speed is faster. Therefore, ELM has been used in many fields, and the current directions are as follows:

(1) The dimensionality reduction and multiple hidden layers: Castano et al. [5] proposed the ELM based on PCA (PCA-ELM) by using principal component analysis (PCA) to achieve the dimensionality reduction. PCA-ELM removes the noise of data and simplifies the structure of the model, thereby improving the learning speed and modeling accuracy. An effective filter maximum relevance minimum redundancy (MRMR) [6] has the same function as PCA. The two-hidden-layer ELM (TELM) [7] and the multiple hidden layer ELM (MELM) [8] were proposed by adding hidden layers. TELM and MELM calculate the weights and thresholds between the remaining hidden layers through the least square method.

(2) The solution of the optimal outer weights based on the swarm intelligence algorithm: the swarm intelligence algorithm simulates the group behaviors and the individual behaviors of animals and plants, which can optimize parameters. At present, the swarm intelligence algorithms that have been applied to ELM include the particle swarm optimization algorithm (PSO) [9] and the differential evolution algorithm (DE) [10]. The swarm intelligence algorithm regards weights and thresholds as individuals and finds the best individuals through the global search and local search.

(3) The selection of the optimal number of nodes in the hidden layer: the different numbers of nodes will lead to the different training errors of ELM, so it is important to obtain the number of nodes with the smallest training error. The current improvement methods include the pruning method, the increment method, and the adaptive method. Rong et al. [11] removed the nodes of the hidden layer with low correlation through statistical testing. Miche et al. proposed OP-ELM [12] that uses the fast leave-one-crossing algorithm to gradually solve the optimal nodes of the hidden layer and TROP-ELM [13] that adds the idea of norm regularization to OP-ELM. Huang and Chen proposed I-ELM [14] and EM-ELM [15]. The

network with smaller nodes of the hidden layer is set firstly, and the number of nodes increases until the training error or the number of nodes reaches the preset value. Lan et al. proposed CS-ELM [16] and TS-ELM [17] by using different methods to screen randomly generated nodes of the hidden layer.

(4) The improvement of stability based on the ensemble learning: the main idea of ELM based on ensemble learning is to construct different ELM models and get the final output by use different methods. One is to obtain multiple ELM models on the same training set and combine them by weighted linear combination [18, 19] and voting method [20]. Another way is to divide the whole training set into several small training sets and construct multiple ELM models based on different training sets [21, 22]. The genetic ensemble of ELM proposed by Xue et al. [23] generates different ELM models by optimizing the parameters of the hidden layer and integrates the optimal models by sorting.

The paper studies the stability of ELM. Because of the random value of ELM, the stability of a single ELM model is not good, and the results and accuracy of different ELM models are slightly different. And the accuracy of ELM needs to be improved. Thus, the weight is used to deal with different ELM models to improve the stability and accuracy of ELM. Specifically, the ELM model with high accuracy has a larger weight, while the ELM model with low accuracy has a smaller weight. The weights of different ELM models need to get the optimal value, so that the combined model can get the best performance. The weight optimization generally uses the swarm intelligence algorithm, and the monarch butterfly optimization (MBO) [24], the moth search algorithm (MSA) [25], the Harris hawks optimization (HHO) [26], and the seagull optimization algorithm (SOA) [27] have appeared in recent years. The paper selects SOA proposed in 2019. The improvements of SOA include the evolutionary multiobjective seagull optimization algorithm (EMoSOA) [28], the whale optimization with seagull algorithm (WSOA) [29], the improved SOA (ISOA) [30, 31], and the opposition-based seagull optimization algorithm (OSOA) [32], which have changed a little.

Because of the high dimension of the near-infrared spectrum, the selection of band based on the pruning method is proposed to select sensitive bands, so as to achieve the purpose of dimension reduction. Based on the idea of the LU triangulation extreme learning machine (LU-ELM) [33] that uses LU decomposition instead of the traditional generalized inverse methods, the paper takes the two hidden layer extreme learning machines based on LU decomposition (LU-TELM) as the modeling method of a single ELM model. Aiming at the problems of poor stability and low accuracy of a single LU-TELM model, the paper proposes the two-hidden-layer extreme learning machine based on LU decomposition and seagull optimization algorithm (LU-TELM-SOA). In a word, LU-TELM-SOA and the selection of band based on the pruning method are proposed in the paper. The second section explains the principles of LU-

TELM and LU-TELM-SOA. The third section describes the prediction of TFe content of hematite, including the selection of band based on pruning method, LU-TELM, and LU-TELM-SOA. The fourth section concludes the paper.

2. Materials and Methods

2.1. Two-Hidden-Layer Extreme Learning Machine Based on LU Decomposition.
TELM [7] calculates the weights and thresholds between the remaining hidden layers through the least square method. The input parameters and the structure parameters are explained in Figure 1.

2.1.1. First Stage of TELM.
TELM contains a single hidden layer, which is ELM. For the hidden layer, the output and weight are, respectively,

$$H = g(IW1 \times X + B1),$$
$$\beta = H^+ T'. \tag{1}$$

2.1.2. Second Stage of TELM.
TELM contains two hidden layers. When another hidden layer is appended between the hidden layer of the first stage and the output layer, the prediction output is

$$H1 = g(IW2 \times H + B2). \tag{2}$$

According to the weight β, the expected output is

$$H1_* = T'\beta^+. \tag{3}$$

According to $H1_* = H1$, the structural parameters between the two hidden layers are

$$[B2, IW2] = g^{-1}(H1_*)\begin{bmatrix} 1 \\ H \end{bmatrix}^+. \tag{4}$$

Then, the final prediction output is

$$H2 = g(IW2 \times H + B2). \tag{5}$$

The updated weight is

$$\beta_{new} = H2^+ T. \tag{6}$$

Therefore, the predicted output of TELM is

$$Y = H2 \times \beta_{new}. \tag{7}$$

2.1.3. LU-TELM.
TELM uses generalized inverse to solve β, and LU-ELM introduces LU decomposition to solve β. LU decomposition transforms the solution into solving two linear equations; that is, it transforms the complex inversion operation into simple four arithmetic operations. Thus, the paper makes use of LU-TELM based on LU decomposition and TELM.
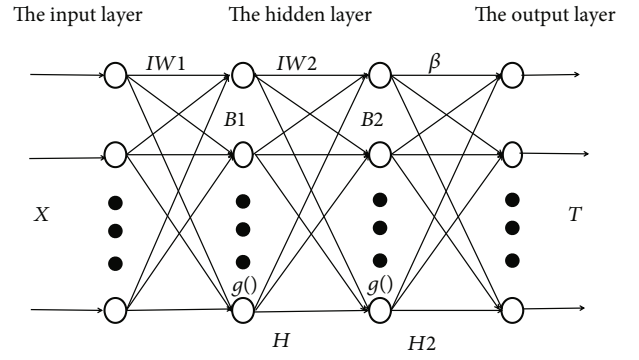

FIGURE 1: The formation and parameters of TELM.

Unlike formula (6) using the generalized inverse to solve β, $H_2^+ = (H_2^T H_2)^{-1} H_2^T$ when $H_2^T H_2$ is a nonsingular matrix. So

$$M\beta = b, \tag{8}$$

where $M = H_2^T H_2$ and $b = H_2^T T$. Therefore, the process of solving β can be replaced by solving a system of the linear equations in the form of formula (8). According to the matrix operation of the sequential Gauss elimination method, the coefficient matrix M can be decomposed into a unit lower triangular matrix L and an upper triangular matrix U, namely,

$$M = LU, \tag{9}$$

which is called the LU decomposition of the coefficient matrix M.

The solution process based on LU decomposition is as follows: first, the result of the LU decomposition of M is calculated; that is, the matrices $L = (l_{ij})$ and $U = (u_{ij})$ are calculated from the matrix $M = (m_{ij})$. Let $M = LU$; then

$$\begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ l_{21} & 1 & & & \\ \vdots & l_{32} & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn-1} & 1 \end{bmatrix}$$
$$\cdot \begin{bmatrix} u_{11} & u_{21} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{bmatrix}. \tag{10}$$

According to the rules of matrix multiplication, by comparing the first row elements at both ends of formula (10), the paper can get

$$u_{1j} = m_{1j}, \quad j = 1, 2, \cdots, n. \tag{11}$$

The paper continues to compare the first column of elements at both ends of the equation to get

$$l_{i1} = \frac{m_{i1}}{u_{11}}, \quad i = 2, 3, \cdots, n. \tag{12}$$

Then, the paper compares the remaining elements in the second row at both ends of the equation to get

$$u_{2j} = m_{2j} - l_{21}u_{1j}, \quad i = 2, 3, \cdots, n. \tag{13}$$

The paper continues to compare the remaining elements in the second column at both ends of the equation to get

$$l_{i2} = \frac{m_{i2} - l_{i1}u_{12}}{u_{22}}, \quad i = 3, \cdots, n. \tag{14}$$

Then, the paper continues like this. After calculating the elements in the first $i-1$ row of U and the elements in the first $i-1$ column of L, the paper can continue to calculate the elements in the i row of U and the elements in the i column of L. The calculation formula is as follows:

$$\begin{cases} u_{1j} = m_{1j}, \quad j = 1, 2, \cdots, n, \\ l_{k1} = \dfrac{m_{k1}}{u_{11}}, \quad k = 2, 3, \cdots, n, \\ u_{ij} = m_{ij} - \displaystyle\sum_{m=1}^{i-1} l_{im}u_{mj}, \quad i = 2, 3, \cdots, n, j = i, i+1, \cdots, n, \\ l_{ki} = \dfrac{1}{u_{ii}}\left(m_{ki} - \displaystyle\sum_{m=1}^{i-1} l_{km}u_{mi}\right), \quad i = 2, 3, \cdots, n, l = i+1, \cdots, n. \end{cases} \tag{15}$$

After solving for L and U, the paper can see from formulas (8) and (9)

$$\begin{cases} Ly = b, \\ U\beta = y. \end{cases} \tag{16}$$

The paper solves the first linear equations. By comparing the elements at both ends of the equation, the calculation formula for the element y_k of y is

$$\begin{cases} y_1 = b_1, \\ y_k = b_k - \displaystyle\sum_{i=1}^{k-1} l_{ki}y_i, \quad k = 2, 3, \cdots, n, \end{cases} \tag{17}$$

where b_k is the element at the corresponding position of b.

Then, the paper continues to solve the second linear equation system. By comparing the elements at both ends of the equation,

$$\begin{cases} \beta_n = \dfrac{y_n}{u_{nn}}, \\ \beta_i = \dfrac{y_i - \sum_{j=i+1}^{n} u_{ij}\beta_j}{u_{ii}}, \quad i = n-1, n-2, \cdots, 1. \end{cases} \tag{18}$$

Finally, the output weight β of LU-TELM is solved.

2.2. Seagull Optimization Algorithm. SOA [27] simulates the migration and attack behavior of seagulls. The migration is the global search, and the attack is the local search.

2.2.1. Migration. To prevent collisions, SOA uses an additional variable A to represent the motion behavior of seagulls.

$$A = f_C - \left(t \times \left(\frac{f_C}{\max}\right)\right). \tag{19}$$

In formula (19), f_C makes A decreases linearly, and the range of A is [2,0]. And t is the current number of iteration and $t = 0, 1, 2, \cdots$, max. The current position is $P_S(t)$, so the new location is $C_S(t) = A * P_S(t)$. For seagulls, the direction of the best position is

$$M_S(t) = B \times (P_{bS}(t) - P_S(t)). \tag{20}$$

In formula (20), $B = 2 * A^2 * r_d$ balances the global search and local search. And r_d is a random value and $r_d \in [0, 1]$. The best location for the population is $P_{bS}(t)$. According to $M_S(t)$, the new position of the seagull after moving to the best position is

$$D_S(t) = |C_S(t) + M_S(t)|. \tag{21}$$

2.2.2. Attack. Seagulls prey through spiral motion, and the motion in the XYZ plane is described as

$$\begin{aligned} x &= r \times \cos(\theta), \\ y &= r \times \sin(\theta), \\ z &= r \times \theta, \\ r &= u \times e^{\theta v}. \end{aligned} \tag{22}$$

In formula (22), r is the spiral radius. And θ is a random angle value and the range is $[0, 2\pi]$. And u and v are the correlation constant. Therefore, the attack position of the seagull is

$$P_S(t) = D_S(t) \times x \times y \times z + P_{bS}(t). \tag{23}$$

2.3. Two-Hidden-Layer Extreme Learning Machine Based on LU Decomposition and Seagull Optimization Algorithm. Since the hidden layer thresholds and the input weights of LU-TELM are given at random, the model is different each time and the accuracy of prediction is also different. In response to this problem, the idea of optimization is adopted to assign different weights to the different LU-TELM models. The model with a good prediction effect has a greater weight. Thus, the soft-sensing model with higher forecast accuracy and higher stability is established.
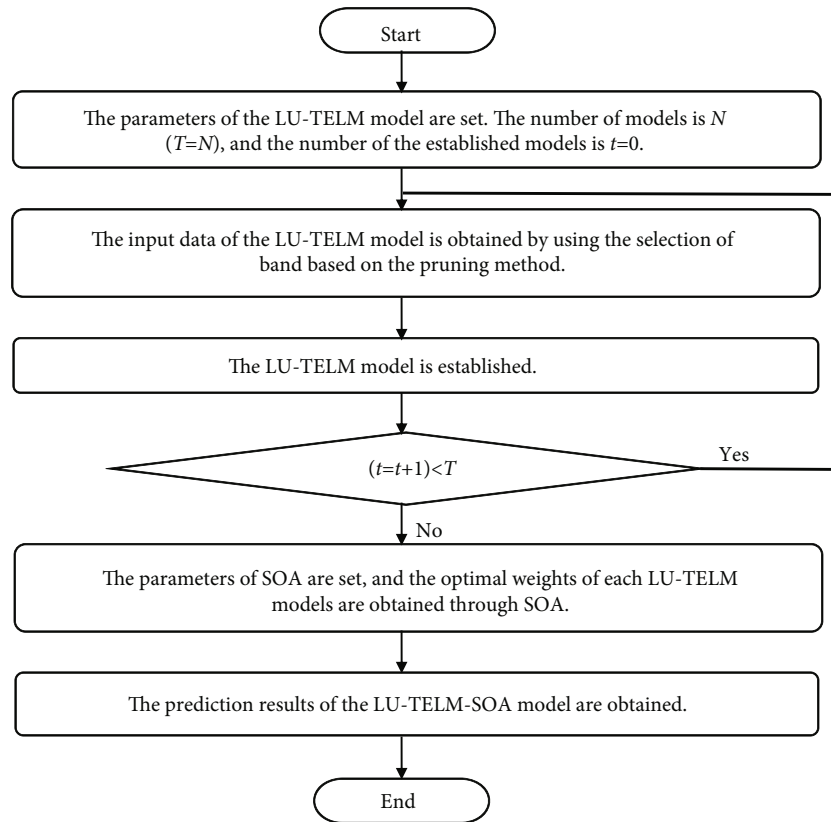
FIGURE 2: The flow chart of LU-TELM-SOA.

The basic idea of LU-TELM-SOA is to optimize the weights of multiple LU-TELM models through SOA, so that the final error of the model is minimized. If the weight of a single model is too large, the final error will increase and the stability of the model will decrease. Therefore, the paper specifies the range of the weight parameters as $[0, m](m \in [0.15, 0.50])$. The number of LU-TELM models is $N(N \geq 2)$. Finally, the different weights of each model are accumulated based on the optimal weights obtained by SOA and summed to obtain the LU-TELM-SOA model. The ultimate objective of LU-TELM-SOA is to reduce the error between the predicted output and the actual output.

$$\min (x - (a_1 \times x_1 + a_2 \times x_2 + \cdots + a_N \times x_N)). \quad (24)$$

In formula (24), x is the expected output of LU-TELM-SOA, and $x_1, x_2, \cdots, x_N$ are the predicted output of LU-TELM. And $a_1, a_2, \cdots, a_N$ are weights of LU-TELM.

The flow chart of LU-TELM-SOA is shown in Figure 2.

3. Results and Discussion

3.1. Acquisition and Processing of Sample. The collection area of the iron ore sample in the paper is the Anshan Iron Mine in China, which is widely distributed and has large reserves. First, the core was obtained along the orientation of the band created by the siliceous and iron, and the samples were obtained by cutting in the direction perpendicular to the core column. Then, the sample was ground, and the spectral data was obtained by using the SVC HR-1024 portable spectrometer. At the same time, the chemical methods were used to obtain the TFe content of the sample. Finally, the number of hematite samples obtained was 91.

As shown in Figure 3, the spectral curves of hematite are parallel, but there are serious data fluctuations around 1900 nm and 2500 nm. To remove the interference of data fluctuations, the paper constructs the fluctuating residual to eliminate the bands with severe data fluctuations. The fluctuating residual is defined as the absolute error between adjacent bands, namely,

$$E_C(i, j) = GP(i, j) - GP(i - 1, j). \quad (25)$$

In formula (25), $i = 2, 3, \cdots, N$ represents the i-th band in the sample, $j = 1, 2, \cdots, M$ represents the j-th sample, $GP(i, j)$ represents the reflectance of the j-th sample in the i-th band, $GP(i\text{-}1, j)$ represents the reflectance of the j-th sample in the i-1-th band, and $E_C(i, j)$ represents The fluctuation residual of the i-th band of the j-th sample.

When the data fluctuates normally, the fluctuating residual fluctuates around 0 and the floating range is small or changes regularly. On the contrary, it floats around 0 and the floating range is large, with irregular changes. Due to a large number of samples and the abnormal samples have been deleted, either the mean value of fluctuating residual or the median value of fluctuating residual can be used. If no abnormal samples are deleted, the median value of fluctuating residual is used.
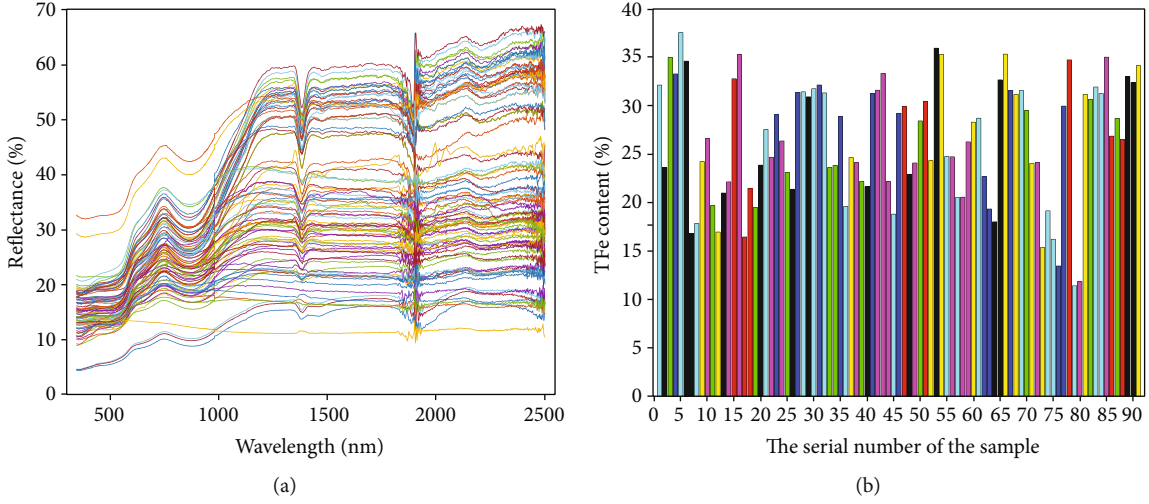
(a)

(b)

FIGURE 3: The spectral data and the TFe content of hematite. (a) The spectral data. (b) The TFe content.

— Median value of fluctuating residual
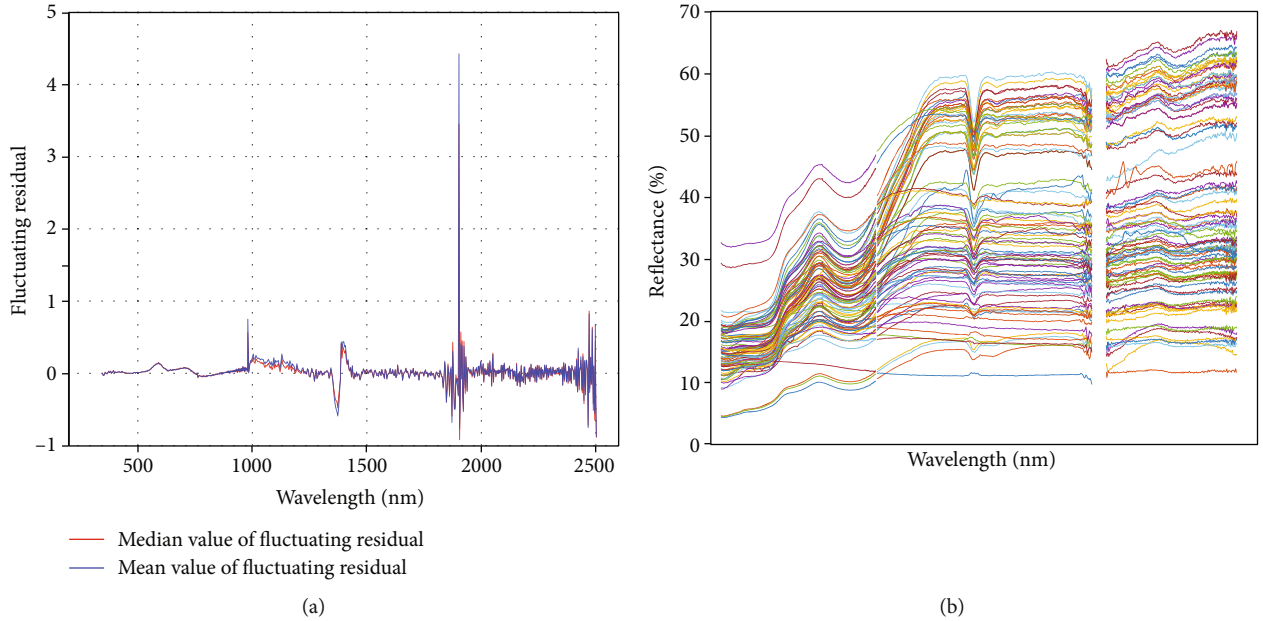— Mean value of fluctuating residual

(a)

(b)

FIGURE 4: The fluctuating residual of each band and the spectral data after deleting the abnormal wave band. (a) The fluctuating residual. (b) The spectral data after deleting the abnormal wave band.

$$E_{\text{mean}}(i) = \frac{1}{M} \sum_{j=1}^{M} E_C(i, j),$$

$$E_{\text{median}}(i) = \text{median}(E_C(i, 1 : M)).$$

(26)

It can be seen from Figure 4 that the bands with severe data fluctuations are 980.3 nm, [1871.6, 1924.2] nm, and [2466.2, 2502.1] nm, and there are 937 bands without severe data fluctuations.

3.2. Selection of Band Based on the Pruning Method. In the paper, RMSEc and RMSEx are the root mean square errors of the test set and the training set, respectively, and R2c and

TABLE 1: The conditions for discrimination.

Condition	Mark
RMSEc > RMSE and R2c < R	1
RMSEc < RMSE and R2c > R	2
RMSEc > RMSE and R2c > R	3
RMSEc < RMSE and R2c < R	4

R2x are the coefficient of determination of the test set and the training set, respectively.

The spectral data has a total of 937 bands without serious data fluctuations, and each band has a different effect on the

TABLE 2: The result of the selection of band based on the pruning method and the comparison of TELM and PCA-TELM.

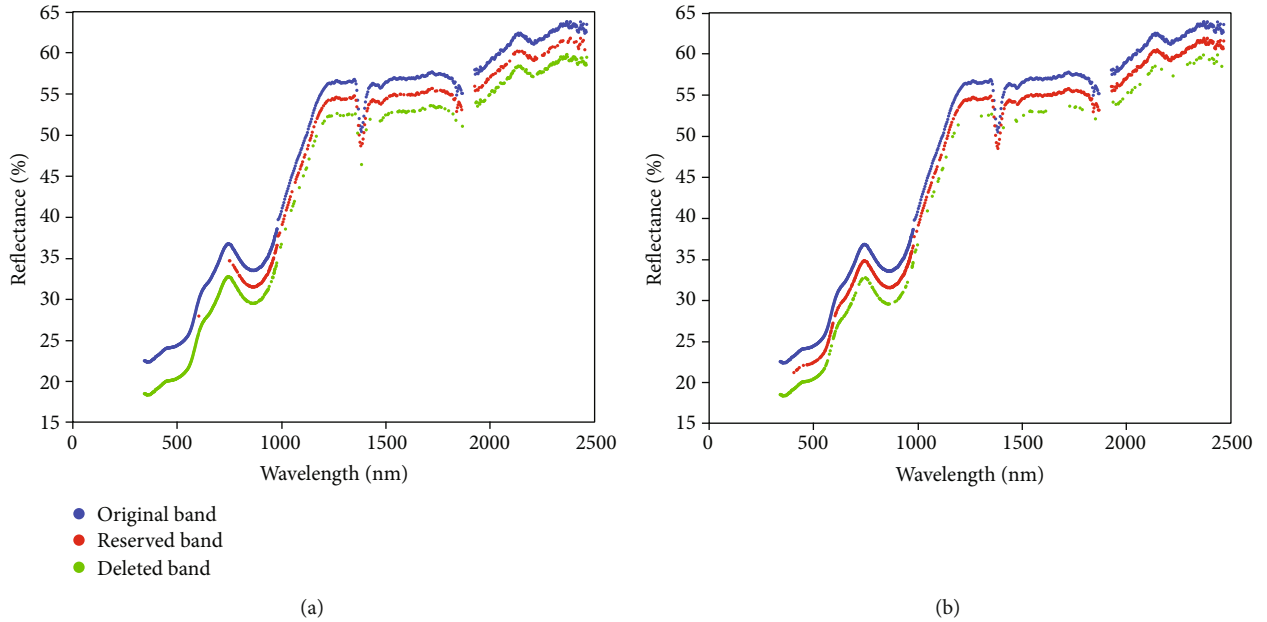| Method	Input band	Selected condition	Mark	RMSEc	R2c
TELM	973	——	——	3.2350	0.8035
TELM	937	——	——	3.1648	0.7913
PCA-TELM	937	——	——	3.4235	0.7923
TELM	351	RMSEc > RMSE and R2c < R	1	2.4602	0.8856
PCA-TELM	351	——	——	3.1150	0.8123
TELM	656	RMSEc > RMSE	1,3	2.3882	0.8849
PCA-TELM	656	——	——	2.6501	0.8583

FIGURE 5: The result of the selection of band based on the pruning method. (a) 351 bands. (b) 656 bands.

TFe content. If it is the sensitive band, the evaluation index of the model will be significantly reduced after the band is deleted. Otherwise, it will be significantly improved, which the evaluation index is better than the limiting condition. Based on the idea, the RMSEc of the TELM model becomes larger (and R2c becomes smaller), that is, the limiting condition is better than the evaluation index of the model, after the band is deleted. And the sensitive bands are exhaustively selected. RMSE and R are the limiting conditions, and RMSEc and R2c are the evaluation index of the model. As shown in Table 1, the bands meeting the selected condition need to be retained.

The paper uses the TELM and sets the ratio of the training set to the test set to 3 : 1. The input is all sensitive bands, and the output is the TFe content. The number of neurons in the hidden layer is 1000, and the activation function is the dsig function ($g(x) = (1 − e^{−x})/(1 + e^{−x})$). For the screening of a single wave band, the optimal result is selected through 1000 cycles. Let RMSE = 3.1648 and R = 0.7913. The paper takes PCA-TELM as a comparison. The number of principal components input by PCA-TELM is 10 because the cumula-

TABLE 3: The comparison of TELM and LU-TELM.

Method	RMSEc	R2c	RMSEx	R2x
TELM	2.4602	0.8856	1.1820	0.9596
LU-TELM	2.0716	0.9147	1.1820	0.9596

tive contribution rate is greater than 99%. And the remaining parameters are the same as TELM. The results show that TELM with the band as input is better than PCA-TELM. As shown in Table 2, the TELM with 351 bands is selected based on the principle of simplifying the network structure. And the result of the selection of band based on the pruning method is shown in Figure 5.

3.3. LU-TELM. LU-TELM uses LU decomposition to solve β, the input is 351 bands, and the other settings are the same as the TELM with 351 bands input. As shown in Table 3, the experimental results show that LU-TELM is better than TELM.

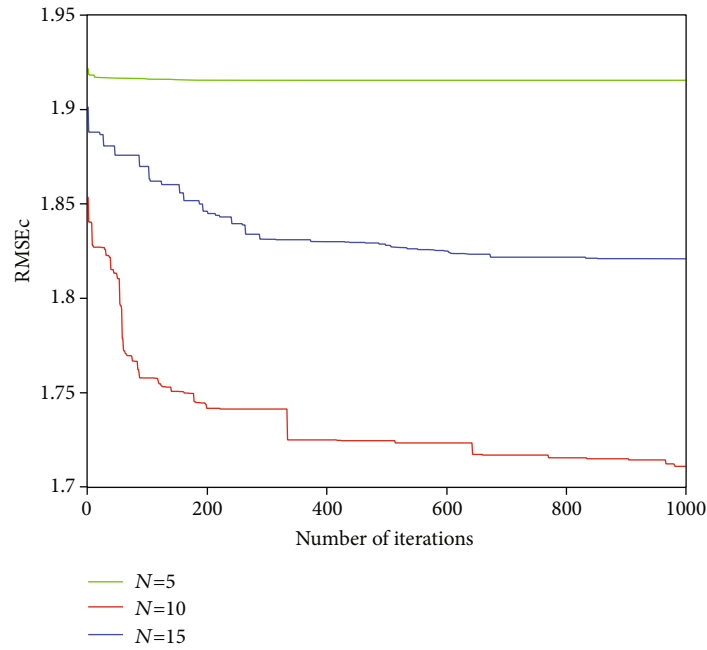*3.4. LU-TELM-SOA.* Since the weights and thresholds are randomly selected, LU-TELM is unstable. Because the

Figure 6: The iterative process of LU-TELM-SOA.

Table 4: The comparison of LU-TELM and LU-TELM-SOA.

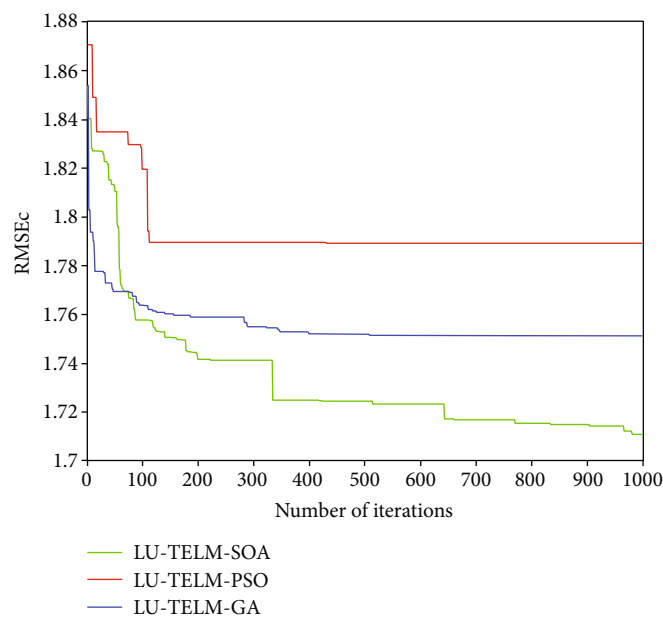| Model	N	m	RMSEc	R2c	RMSEx	R2x
LU-TELM	—	—	2.0716	0.9147	1.1820	0.9596
	5	0.4	1.9154	0.9262	1.1997	0.9596
LU-TELM-SOA	10	0.4	1.7109	0.9396	1.1879	0.9596
	15	0.2	1.8208	0.9332	1.8334	0.9596
LU-TELM-PSO	10	0.4	1.7892	0.9347	1.1935	0.9596
LU-TELM-GA	10	0.4	1.7513	0.9373	1.1920	0.9596

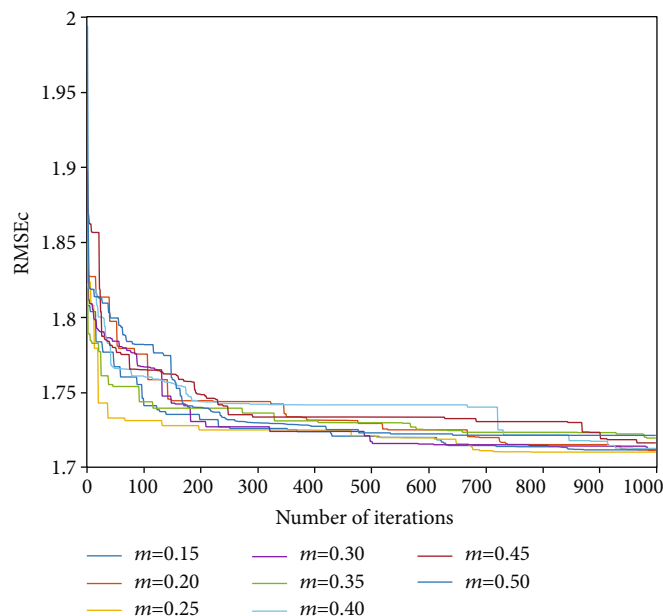Figure 7: The comparison of LU-TELM-SOA, LU-TELM-PSO, and LU-TELM-GA.

FIGURE 8: The results of LU-TELM-SOA with different ranges of weight parameter ($N = 10$).

TABLE 5: The results of LU-TELM-SOA with different ranges of weight parameter ($N = 10$).

m	RMSEc	R2c	Sum of weights
0.15	1.7118	1.1923	0.9763
0.20	1.7114	1.1945	1.0025
0.25	1.7103	1.1900	0.9955
0.30	1.7125	1.1977	1.0048
0.35	1.7197	1.1870	1.0094
0.40	1.7128	1.1881	0.9895
0.45	1.7164	1.1942	0.9847
0.50	1.7214	1.1911	0.9475

performance of each model is different, they are assigned to different weights. In the paper [27], the seagull optimization algorithm (SOA) was proposed, and the SOA is better than the particle swarm optimization algorithm (PSO), the genetic algorithm (GA), and so on. Then, the SOA is used to find the best weight distribution. The paper sets the parameters of SOA based on [27], and $f_c = 2$, $u = 1$, and $v = 1$. The population of seagulls is better at 100 [27], and the paper sets $n = 1000$ to improve the optimization performance. The larger the maximal number of iterations is, the better the model is, so the paper sets max $= 1000$. Let $P = xyzC1$ and $C1$ as the adjustment factor, and $C1 = 0.000001$. For point C beyond the search space, if $C > m$ then $C_{new} = |C - m|$. If $C \leq 0$, then $C_{new} = -C$. The paper stipulates that the range of $N(N \geq 2)$ weight parameters is $[0, m](m \in [0.15, 0.50])$. In SOA, RMSEc is used as the fitness function.

When $m = 0.4$, the performance of the model with different number of LU-TELM models is compared. When $N = 15$, $RMSE_C$ will be greater than 10 if $m = 0.4$, which is abnormal. Thus, the paper chooses the model with $m = 0.2$ and the best

performance. When $N = 10$, the performance of LU-TELM-SOA is the best (Figure 6 and Table 4).

In the paper, PSO and GA are used to replace SOA, and LU-TELM-PSO and LU-TELM-GA algorithms are proposed. The paper compares the performance of LU-TELM-SOA, LU-TELM-PSO, and LU-TELM-GA under the same conditions, where the same conditions refer to $N = 10$ and $m = 0.4$. The inertia coefficient of PSO is 1.0 ($\omega = 1.0$), and the acceleration constant of PSO is 1.0 ($c1 = c2 = 1.0$). The crossover probability of GA is 0.4, and the mutation probability of GA is 0.2. The other parameters of PSO and GA are the same as SOA. The experimental results show that RMSEc of LU-TELM-SOA is better than that of LU-TELM-PSO and LU-TELM-GA, and LU-TELM-SOA can avoid falling into local minimum too early (Figure 7). According to the performance of the model from good to bad, the order is LU-TELM-SOA, LU-TELM-PSO and LU-TELM-GA in the paper (Table 4).

For the range of weight parameters, there is no obvious performance advantage or disadvantage in $m \in [0.15, 0.50]$ (Figure 8 and Table 5).

4. Conclusions

In the paper, the selection of band based on the pruning method and LU-TELM-SOA is used to establish the model of TFe content of hematite with higher accuracy and better stability. The 351 sensitive bands are screened out by the selection of band based on the pruning method, and TELM with 351 sensitive bands input is better than PCA-TELM. And the LU-TELM is superior to TELM, and LU-TELM-SOA proposed in the paper solves the stability problem of LU-TELM. Finally, the RMSEc of the TFe content of hematite obtained was 1.7109, achieving high accuracy. The SOA algorithm used in the paper has the inherent problems of

swarm intelligence algorithm, and it needs to be improved in the future. For example, the swarm intelligence algorithm has some problems in the balance of global search and local search. And it is also a research direction for choosing swarm intelligence algorithms with better performance than SOA.

Data Availability

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. Lowe, "Adaptive radial basis function nonlinearities, and the problem of generalisation," in *First IEE International Conference on Artificial Neural Networks, (Conf. Publ. No. 313)*, pp. 171–175, London, UK, 1989.

[2] Y. H. Pao, G. H. Park, and D. J. Sobajic, "Learning and generalization characteristics of the random vector functional-link net," *Neurocomputing*, vol. 6, no. 2, pp. 163–180, 1994.

[3] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[4] G. B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels," *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.

[5] A. Castano, F. Fernández-Navarro, and C. Hervás-Martínez, "PCA-ELM: a robust and pruned extreme learning machine approach based on principal component analysis," *Neural Processing Letters*, vol. 37, no. 3, pp. 377–392, 2013.

[6] T. Liu, L. Hu, C. Ma, H. L. Chen, and Z. Y. Wang, "A fast approach for detection of erythemato-squamous diseases based on extreme learning machine with maximum relevance minimum redundancy feature selection," *International Journal of Systems Science*, vol. 46, no. 5, pp. 919–931, 2015.

[7] B. Y. Qu, B. F. Lang, J. J. Liang, O. D. Crisalle, and A. K. Qin, "Two-hidden-layer extreme learning machine for regression and classification," *Neurocomputing*, vol. 175, pp. 826–834, 2016.

[8] D. Xiao, B. Li, and Y. Mao, "A multiple hidden layers extreme learning machine method and its application," *Mathematical Problems in Engineering*, vol. 2017, Article ID 4670187, 10 pages, 2017.

[9] Y. Xu and Y. Shu, "Evolutionary extreme learning machine–based on particle swarm optimization," *Lecture Notes in Computer Science*, vol. 3971, pp. 644–652, 2006.

[10] Q. Y. Zhu, A. K. Qin, P. N. Suganthan, and G. B. Huang, "Evolutionary extreme learning machine," *Pattern Recognition*, vol. 38, no. 10, pp. 1759–1763, 2005.

[11] H. J. Rong, Y. S. Ong, A. H. Tan, and Z. Zhu, "A fast pruned-extreme learning machine for classification problem," *Neurocomputing*, vol. 72, no. 1–3, pp. 359–366, 2008.

[12] Yoan Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, and A. Lendasse, "OP-ELM: optimally pruned extreme learning machine," *IEEE Transactions on Neural Networks*, vol. 21, no. 1, pp. 158–162, 2010.

[13] Y. Miche, M. van Heeswijk, P. Bas, O. Simula, and A. Lendasse, "TROP-ELM: A double-regularized ELM using LARS and Tikhonov regularization," *Neurocomputing*, vol. 74, no. 16, pp. 2413–2421, 2011.

[14] G. B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," *Neurocomputing*, vol. 71, no. 16–18, pp. 3460–3468, 2008.

[15] Guorui Feng, Guang-Bin Huang, Qingping Lin, and R. Gay, "Error minimized extreme learning machine with growth of hidden nodes and incremental learning," *IEEE Transactions on Neural Networks*, vol. 20, no. 8, pp. 1352–1357, 2009.

[16] Y. Lan, Y. C. Soh, and G. B. Huang, "Constructive hidden nodes selection of extreme learning machine for regression," *Neurocomputing*, vol. 73, no. 16–18, pp. 3191–3199, 2010.

[17] Y. Lan, Y. C. Soh, and G. B. Huang, "Two-stage extreme learning machine for regression," *Neurocomputing*, vol. 73, no. 16–18, pp. 3028–3038, 2010.

[18] M. Van Heeswijk, Y. Miche, T. Lindh-Knuutila et al., "Adaptive ensemble models of extreme learning machines for time series prediction," in *the 19th International Conference on Artificial Neural Networks*, pp. 305–314, Limassol, Cyprus, 2009.

[19] Hui-Xin Tian and Zhi-Zhong Mao, "An ensemble ELM based on modified AdaBoost.RT algorithm for predicting the temperature of molten steel in ladle furnace," *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 1, pp. 73–80, 2010.

[20] J. Cao, Z. Lin, G. B. Huang, and N. Liu, "Voting based extreme learning machine," *Information Sciences*, vol. 185, no. 1, pp. 66–77, 2012.

[21] N. Liu and H. Wang, "Ensemble based extreme learning machine," *IEEE Signal Processing Letters*, vol. 17, no. 8, pp. 754–757, 2010.

[22] A. Samat, P. du, S. Liu, J. Li, and L. Cheng, "E2LMs: ensemble extreme learning machines for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1060–1069, 2014.

[23] X. Xue, M. Yao, Z. Wu, and J. Yang, "Genetic ensemble of extreme learning machine," *Neurocomputing*, vol. 129, pp. 175–184, 2014.

[24] G. G. Wang, S. Deb, and Z. Cui, "Monarch butterfly optimization," *Neural Computing and Applications*, vol. 31, no. 7, pp. 1995–2014, 2019.

[25] G. G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Computing*, vol. 10, no. 2, pp. 151–164, 2018.

[26] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and

applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.

[27] G. Dhiman and V. Kumar, "Seagull optimization algorithm: theory and its applications for large-scale industrial engineering problems," *Knowledge-Based Systems*, vol. 165, pp. 169–196, 2019.

[28] G. Dhiman, K. K. Singh, A. Slowik et al., "EMoSOA: a new evolutionary multi-objective seagull optimization algorithm for global optimization," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 2, pp. 571–596, 2021.

[29] Y. Che and D. X. He, "A hybrid whale optimization with seagull algorithm for global optimization problems," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6639671, 31 pages, 2021.

[30] X. Ji, Y. Pan, G. Jia, and W. Fang, "A neural network-based prediction model in water monitoring networks," *Water Supply*, vol. ws2021046, 2021.

[31] X. Chen, Y. Li, Y. Zhang, X. Ye, X. Xiong, and F. Zhang, "A novel hybrid model based on an improved seagull optimization algorithm for short-term wind speed forecasting," *Processes*, vol. 9, no. 2, p. 387, 2021.

[32] H. Jiang, Y. Yang, W. Ping, and Y. Dong, "A novel hybrid classification method based on the opposition-based seagull optimization algorithm," *IEEE Access*, vol. 8, pp. 100778–100790, 2020.

[33] Y. Kutlu, A. Yayık, E. Yildirim, and S. Yildirim, "LU triangularization extreme learning machine in EEG cognitive task classification," *Neural Computing and Applications*, vol. 31, no. 4, pp. 1117–1126, 2019.