International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Development of a Novel Neural Network Model for Brain Image Classification

Pranati Satapathy, Sateesh Kumar Pradhan, Sarbeswara Hota

Abstract*: *The analysis of brain MRI images is highly beneficial for the medical practitioners. Since the manual study of these images are time consuming and tedious, the automated process using software based system have been developed. The machine learning techniques are applied in developing brain MR image classification process. The classification process consists of dataset preparation, feature extraction, feature reduction and the use of classifier. In this paper, 2D DWT is used for feature extraction and PCA is used for feature reduction. ELM model is used as a classifier. The input weights and biases in ELM are randomly assigned. So EHO algorithm, a newly developed bio inspired algorithm is used to optimally determine the input weights and biases of ELM model. The classification performance of the EHO-ELM model is compared with basic ELM model for three of the brain MR image datasets. From the simulation study, it is found that the proposed EHO-ELM model outperformed the basic ELM model.

Keywords: Elephant Herding Optimization, Machine Learning, Magnetic Resonance Imaging, Principal Component Analysis, Sensitivity.

I. INTRODUCTION

Human brain can be affected with several diseases as the age increases. The early detection of diseased brain help the doctors and radiologists in taking the appropriate clinical decisions [1]. There are various techniques used in getting the brain images. Magnetic Resonance Imaging (MRI) is one of the advanced neuroimaging technique used for the detection of pathological brain. This technique provides brain tissue images of better resolution and it is also radiation free [2]. But the manual detection of pathological brain from brain MRI is a tedious task and may be erroneous because of its vast image contents. So research has been conducted to automate the brain MR image analysis using different machine learning tools. The basic purpose of the automated process is to classify the brain images as normal or pathological. This tool based automated process reduces the diagnostic errors and speeds up the clinical decisions [3].

Manuscript published on 30 September 2019
* Correspondence Author
Pranati Satapathy, Dept. of Computer Science and Applications, Utkal University, Bhubaneswar, India. satapathy.pranati@gmail.com
Sateesh Kumar Pradhan, Computer Science and Applications, Utkal University, Bhubaneswar, India. Email: sateesh.cs@utkaluniversity.ac.in
Sarbeswara Hota*, Computer Application, Siksha O Anusandhan Deemed to be University, Bhubaneswar, India. sarbeswarahota@gmail.com

The general brain MR images classification model is shown in figure 1. The major phases involved in this model are feature extraction, feature reduction and classification. Discrete Wavelet Transformation (DWT) and Gray Level Co-occurrence Matrix (GLCM) are found to be the two popular techniques used for feature extraction from brain MR images [4]. All the extracted features are not relevant in the classification task. So different feature reduction techniques i.e. Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA) methods have been used for reducing the features of brain image datasets. In the next phase, classifiers are designed to classify the images as normal or pathological.

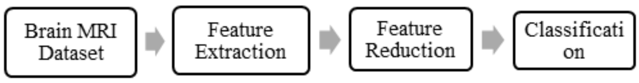

Fig.1. General classification model

The literature study shows that machine learning techniques have been applied by various researchers for classification of brain MR images [5, 6]. M. Sarita *et al.* [7] proposed a probabilistic neural network (PNN) for brain MRI classification into normal and pathological. The authors used wavelet entropy based spider web plots for the feature extraction from the images. 75 number of brain MR images were considered for this work. E. El-Dahshan *et al.* [8] applied two different classification models i.e. k-nearest neighbor (k-NN) and Feed forward Backpropagation Neural Network (BPNN) classifiers for the classification of brain images. 2D DWT was used for feature extraction and PCA was used for dimensionality reduction. From the experimental study, it was found that k-NN classifier outperformed the BPNN classifier. G. B. Huang *et al.* proposed a new learning mechanism known as Extreme Learning Machine (ELM) that overcomes the problems generated by conventional gradient based approaches [9, 10].The connection weights between the input layer and hidden layer are given randomly and the output weights are mathematically calculated using Moore-Penrose generalized pseudo inverse. D. Nayak *et al*. [11] applied the ELM model for the classification of normal and pathological brains from the MR images. The PCA+LDA method was used for feature reduction from the extracted features of the images. The modified sine cosine (MCSA) algorithm was used with the ELM model for the classification purpose. The experimental results concluded that the proposed MCSA-ELM model outperformed other models.

The authors in [12] used modified PSO based ELM model for the classification of brain MRI datasets. This study motivates to use ELM model in classification of brain MR images. One limitation in this tool is that the random assignment of weights and biases in the hidden layer affects prediction accuracy and these weights may not be always optimal [13]. To overcome this problem, various researchers have used different bio inspired algorithms with ELM model.

In this work, we have used Elephant Herding Optimization (EHO) algorithm for the optimization of ELM model. The classification performance of this EHO-ELM model is compared with the standard ELM model.

A good amount of work has been performed using EHO algorithm. G. Wang *et al.* [14] have proposed a new type of bio inspired optimization method known as EHO for solving optimization problems. The performance of EHO was compared with other three evolutionary algorithms i.e. BBO, GA and DE by taking fifteen test cases. EHO outperforms the other three methods on most benchmark problems. A. Sahlol *et al.* [15] used EHO to train the ANN that updates the weights and biases of the network. This model was implemented in classification task and it achieved better classification accuracy in comparison with SVM, Naive Bayes and Decision tree. The authors in [16] used EHO algorithm to tune SVM parameters and achieved highest classification accuracy as compared with GA and grid search method. N. Meena *et al.* [17] suggested an improved EHO for solving a multi objective distributed energy resource (DER) accommodation problem of distribution systems. The optimal solutions obtained using the improved EHO method are promising as compared to GA, PSO and Teaching Learning based Optimization method (TLBO). In [18], the authors proposed different techniques to enhance the performance of EHO algorithm and found better performances as compared to other bio inspired algorithms in various application fields. So this study indicates that EHO can be used as an efficient optimization algorithm [19][20].

The basic goal of this work is to design a novel classification model using EHO-ELM model and compare its performance with the standard ELM for three of the brain MR image datasets. The paper layout is as follows. The methodologies are described in section 2 .The datasets description and the experimental setup along with the experimental comparisons are discussed in section 3.The conclusion and future scope of this work are described in section 4.

II. METHODOLOGY

This section discusses the DWT technique for feature extraction, PCA for feature reduction, ELM model for classification and the EHO algorithm. The proposed EHO-ELM model is also depicted in this section.

A. Discrete Wavelet Transformation (DWT)

In this work 2D DWT is employed to generate the wavelet coefficients from the MR images. The wavelet coefficients provide localized frequency information that are beneficial in classification. These wavelets coefficients will be used as feature vectors. The wavelets depict the hierarchical framework of the image information. The cascaded filter banks of both high pass and low pass filters are used to decompose the signal into different sub bands as shown in Figure 2.

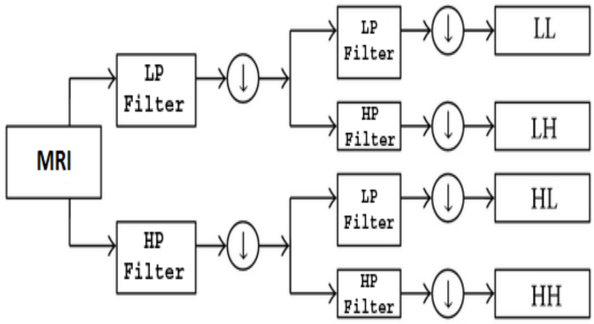

Fig 2. 3-level decomposition of 2D DWT

The image is processed along two dimensions at each level generating four sub bands i.e. LL, LH, HH, HL. The LL sub band is utilized for the next level of transformation. The LL sub band is considered as the approximation component of the image. The other three sub bands are considered as the detailed components of the image. The calculation of wavelet coefficients are performed using Haar wavelet function for the LL sub band. This algorithm uses level-3 decomposition to extract features.

B. Principal Component Analysis (PCA)

PCA is a popular technique for reducing the features of a dataset into a lower-dimension feature space. In this work min-max normalization method is used to normalize the dataset before applying PCA. PCA uses the largest Eigen vectors of the correlation matrix. PCA preserves the variance of the reconstructed data. This technique reduces the set of correlated variables to a subset of linearly uncorrelated variables called principal components. The basic steps of PCA are described as follows:

Let $D\ [X \times Y]$ be the input dataset.

Step 1: Calculate the empirical mean using Equation (1).

$$e[i] = \frac{1}{Y} \sum_{j=1}^{Y} D[i,j] \qquad (1)$$

Step 2: Calculate the deviations of the data from the mean and put in a separate matrix $E\ [X \times Y]$ using Equation (2).

$$E = D - e.h \qquad (2)$$

Here $h\ [1 \times N]$ is a one dimensional array of all 1s.

Step 3: Calculate the covariance matrix C using Equation (3).

$$C = \frac{1}{Y}\ E.E^T \qquad (3)$$

Step 4: Calculate the Eigen vectors and Eigen values of the covariance matrix using Equation (4).

$$V^{-1}CV = D \qquad (4)$$

Here V is the Eigen vector matrix, D is the diagonal matrix of Eigen values of C.

Step 5: Sort the Eigen vectors and Eigen values in the descending order.

Step 6: Select the components and form the new feature vector.

C. Extreme Learning Machine (ELM) model

ELM is a relatively new kind of training method developed for Single Layer Feed forward Network (SLFN) [9].

Convergence speed and better learning rate are the two most important features of ELM that makes it very popular among researchers.

It exhibits good generalization performance, requires simple and fixed network structure, does not need setting up or tuning of network parameters and is also free from converging to local minima . In contrast to traditional gradient-based learning algorithms, ELM adapts to nontraditional training procedure where in the weights between hidden layer and the output layer are found out in a non-iterative way from the randomly chosen weights and biases of the hidden layer using the Moore–Penrose (MP) generalized inverse.

The basic steps of ELM are as follows:

Let $M = \{(X_i, y_i) \mid \in R^d,\ y_i \in R, i = 1,2, \ldots., n\}$ be the dataset.

Step-1: Assign the weight w_i and bias $b_i(i = 1,2, \ldots, N)$ between the input and hidden layer randomly.

Step-2: Compute H as the output of the hidden layer using the Eq.5.

$$H = \begin{pmatrix} g(W_1 .X_1 + b_1) & g(W_2 .X_1 + b_2) & \cdots & g(W_N .X_1 + b_N) \\ g(W_1 .X_2 + b_1) & g(W_2 .X_2 + b_2) & \cdots & g(W_N .X_2 + b_N) \\ \vdots & \vdots & \cdots & \vdots \\ g(W_1 .X_n + b_1) & g(W_2 .X_n + b_2) & \cdots & g(W_N .X_n + b_N) \end{pmatrix}_{n \times N} \quad (5)$$

Step-3: Find out the output weight β, *where* $\beta = H^{\dagger}T$, where $H^{\dagger}$ is the MP generalized inverse of H and $T = (y_1, y_2, y_3, \ldots \ldots, y_n)^T$.

D. Elephant Herding Optimization (EHO) Algorithm

EHO is a new type of swarm based meta-heuristic optimization algorithm proposed by G. Huang *et al.* [14]. Its basic aim is to solve optimization problem. This algorithm is also inspired from the herding behavior of elephant groups. The elephants normally are wild animals and they have some social nature involving female and male elephants. An elephant group consists of several clans. The matriarch is the oldest female elephant and is the leader of the clan. A clan is composed of one female elephant and some other family members of her. Female elephants generally stay in a group, but after certain age, male elephants go away from their family group. EHO is simulated using clan updating operation and clan separating operation. These two operations are described as follows.

i) Clan Updating Operation

The elephants in a clan are updated with reference to the position of the matriarch and the position of the matriarch is updated with reference to the elephant at the center position of the clan.

In a clan C_i, the next position of an elephant $e_{C_i,j}$ is represented using Equation (6).

$$e_{C_i,j}(t+1) = e_{C_i,j}(t) + \propto \times \left(e_{best,C_i} - e_{C_i,j}(t)\right) \times r \quad (6)$$

Here $\propto$ is the scale factor that represents the matriarch influence in the clan, e_{best,C_i} is the matriarch in clan $C_i$and r is a random number between 0 and 1.

The fittest elephant or the matriarch is updated using Equation (7).

$$e_{best,C_i}(t+1) = \beta \times e_{center,C_i}(t) \quad (7)$$

Where $\beta \in [0,1]$ is a factor that represents the influence of the elephant e_{center,C_i} on the matriarch e_{best,C_i}in the clan C_i.

The e_{center,C_i} of clan C_i is determined using Equation (8).

$$e_{center,C_i} = \frac{1}{N}\sum_{j=1}^{N} e_{C_i,j} \quad (8)$$

Where N represents number of elephants in clan C_i.

ii) Clan Separating Behavior

Since the male elephants will go away from the clan after certain age, this behavior can be implemented assuming the elephants with lowest fitness values.

Let e_{worst,C_i} be the elephant having lowest fitness values. Its new position can be determined after separation using Equation (9).

$$e_{worst,C_i} = e_{min} + (e_{max} - e_{min} + 1) \times rand \quad (9)$$

Where e_{max} and e_{min} are the upper bound and lower bound of the elephant position respectively. $rand$ is a random number that follows stochastic distribution in the range$[0, 1]$.

E. Proposed EHO-ELM model

The proposed EHO-ELM model is shown in Fig. 3. The feature extraction is performed using 2D DWT technique. The PCA algorithm is used for feature reduction. Then the ELM model is taken as the classifier. The input weights and biases between the input layer and hidden layer are optimized using EHO algorithm. The MSE is taken as the objective function. The EHO-ELM algorithm is described as follows.

Algorithm: EHO-ELM

1. Initialize the population of elephants as the input weights and biases of ELM model
2. Calculate the output weights using MP generalized inverse.
3. For the input training samples, calculate the actual outputs.
4. Determine the MSE by finding the differences.
5. Consider the MSE as the fitness function of each elephant.
6. Sort the elephants according to the fitness values.
7. Update the elephants using Equation (6) and Equation (7).
8. Separate the worst male elephant and update its position using Equation (9).
9. Repeat the steps 4 to 8 until stopping criterion is met.

III. SIMULATION STUDY

This section consists of dataset preparation, data preprocessing, training and testing the model and result analysis processes.

A. Datasets Preparation

For the simulation study, three brain MR image datasets are downloaded from the Harvard medical school website. i.e. Alzheimer, Glioma and Multiple Sclerosis and used in this work. These image datasets refer to three different types of brain diseases. The datasets are preprocessed before classification.

B. Data Preprocessing

This phase consists of feature extraction and feature reduction steps. In this work, 2D DWT technique is used to extract the features from the images. There are 1296 features extracted for each brain image dataset with the 3- level decomposition using Haar wavelet function.

PCA is used for feature reduction as most of the features are not relevant for the classification. Before applying

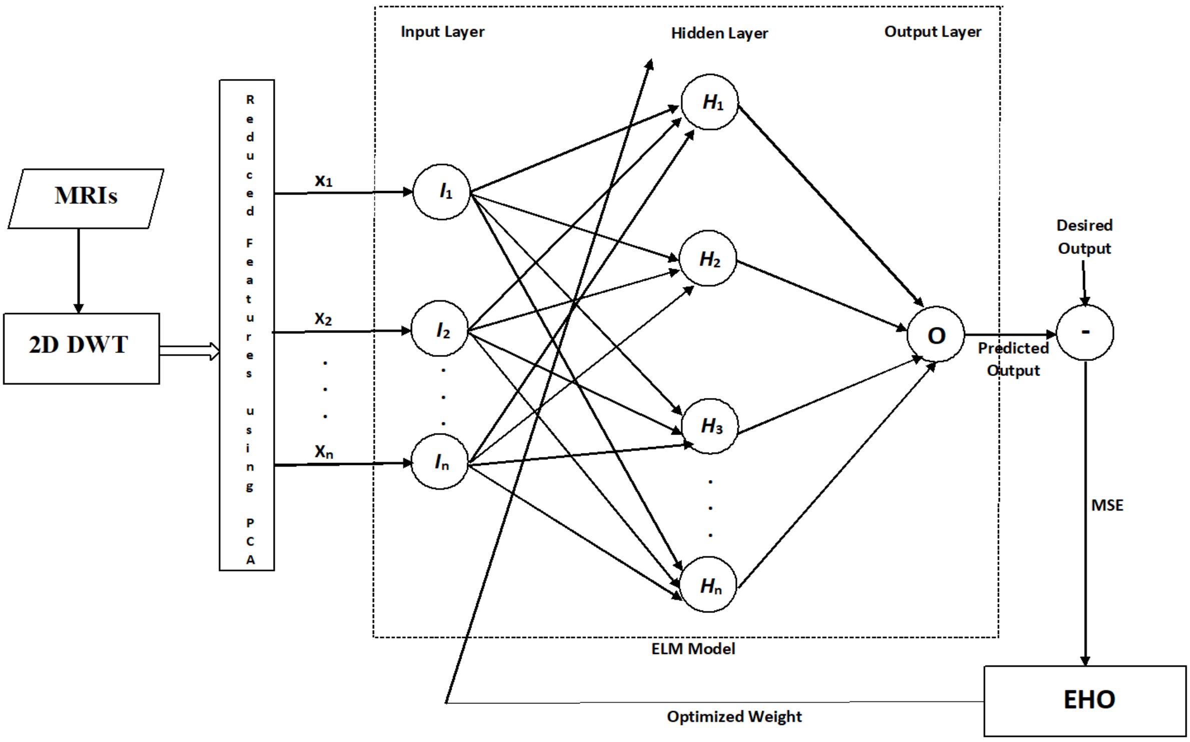

Fig. 3. Proposed EHO-ELM model

PCA, the data are normalized using min-max normalization for each dataset. The variance of 95% is taken as the threshold value for choosing the principal components. The features are reduced to 34, 24 and 39 for Alzheimer, Glioma and Multiple Sclerosis dataset respectively.

C. Training and testing the model

The training and the testing of the ELM model is performed with 5-fold cross validation of the dataset. In this process, the entire dataset is divided into 5 separate subsets. In each run, one set is used for testing and the rest are used for the training purpose. So 20 runs are performed for each dataset and the average performance measures are evaluated. The EHO algorithm is used for determining the weights and biases associated with the hidden layer. After the weights and biases are frozen, the model is validated with the test dataset. The classification performance of the proposed model is measured with different metrics. The number of hidden nodes are taken as 60 and sigmoid activation function is used in the ELM. The hidden nodes are determined experimentally with varying from 5 to 80. For implementing the EHO algorithm, the population size is 50, number of clans is 5, the value of $\propto$ is 0.5 and β is 0.1.

D. Performance metrics

Different metrics have been used as the performance measures for the classification problem. The following performance measures are determined in this paper.

i) Accuracy
ii) Specificity
iii) Recall
iv) Precision
v) F-score

The above measures are defined using Equation (10) to Equation (14).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$Specificity = \frac{TN}{TN+FP} \quad (12)$$

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$F-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (14)$$

Here *TP*= Count of the samples that belong to class *A* and predicted as *A*, *FP*= Count of the samples that belong to class B and predicted as A, *TN*= Count of the samples that belong to class B and predicted as *B*, *FN*= Count of the samples that belong to class *A* and predicted as *B*. Table 1 to Table 5 shows the above measures for the three datasets for different models.

Table 1. Accuracy (in %) of different models for three datasets

Models	Alzheimer	Glioma	Multiple Sclerosis
ELM	88.57	79.07	77.27
PCA+ELM	91.43	81.4	86.36

EHO-ELM	91.43	83.72	81.82
PCA+EHO-ELM	94.29	86.05	88.64

Table 2. Recall of different models for three datasets

Models	Alzheimer	Glioma	Multiple Sclerosis
ELM	0.8666	0.6956	1
PCA+ELM	0.9333	0.9565	1
EHO-ELM	0.9333	0.7826	1
PCA+EHO-ELM	0.9333	0.8268	1

Table 3. Specificity of different models for three datasets

Models	Alzheimer	Glioma	Multiple Sclerosis
ELM	0.9	0.9	0.6667
PCA+ELM	0.9	0.65	0.8
EHO-ELM	0.9	0.9	0.7333
PCA+EHO-ELM	0.95	0.9	0.7333

Table 4. Precision of different models for three datasets

Models	Alzheimer	Glioma	Multiple Sclerosis
ELM	0.8666	0.8889	0.5833
PCA+ELM	0.875	0.7586	0.7
EHO-ELM	0.875	0.9	0.6333
PCA+EHO-ELM	0.9333	0.9047	0.7368

Table 5. F-score of different models for three datasets

Models	Alzheimer	Glioma	Multiple Sclerosis
ELM	0.8666	0.7804	0.7368
PCA+ELM	0.9032	0.8461	0.8235
EHO-ELM	0.9032	0.8372	0.7777
PCA+EHO-ELM	0.9333	0.8636	0.8484

E. Result Analysis

The Table 1 to Table 5 reflect the different performance measures of the classification models used in this paper for the three brain MRI datasets. From the accuracy values given in Table 1, it is found that the classification accuracy of the proposed PCA+ EHO-ELM model is 94.29%, 86.05% and 88.64% for the Alzheimer, Glioma and Multiple Sclerosis datasets respectively. For Multiple Sclerosis dataset, the PCA+ELM model gives accuracy of 86.36% which is better than EHO-ELM model without feature reduction. The graphical representation of the accuracy I the form of a bar chart is shown in Fig. 4. The other measures i.e. recall, specificity, precision values of the proposed PCA+ EHO-ELM model are better than the other models for these three datasets.

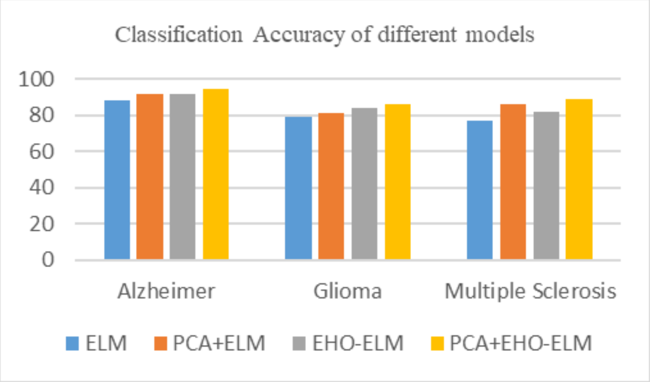

Fig. 4. Bar chart of Classification accuracy of different models

The F-score value of the proposed PCA+EHO-ELM model is 0.9333, 0.8636 and 0.8484 for the Alzheimer, Glioma and Multiple Sclerosis datasets respectively. The Table 5 depicts the F-score values of all the models for the three datasets and Fig. 5 shows the bar chart representation of the F-score values of different models for the three datasets. It is found from Table-5 that the F-score value of the proposed PCA+EHO-ELM is better as compared to the other models.

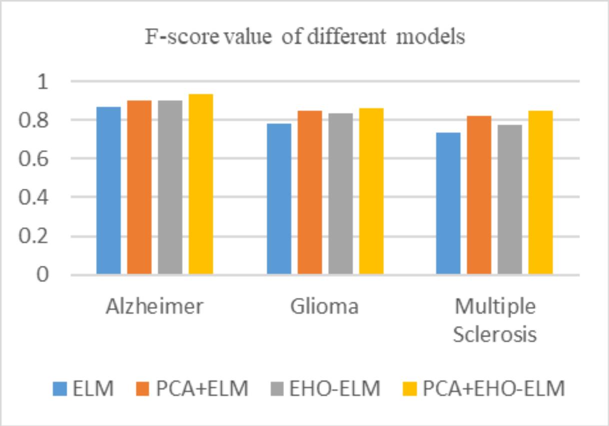

Fig. 5. Bar chart of F-score values of different models

So this simulation study concludes that the proposed PCA+ EHO-ELM model outperformed the basic ELM, PCA+ ELM and the EHO-ELM model for the Alzheimer, Glioma and Multiple sclerosis brain MR image datasets.

IV. CONCLUSION

This paper proposes the ELM model hybridized with EHO algorithm, a newly developed bio inspired algorithm for the classification of three brain MR image datasets i.e. Alzheimer, Glioma and Multiple Sclerosis. As the weights and biases of the hidden layer in ELM are taken randomly, the EHO algorithm is used to optimally determine these weights and biases. This work also uses 2-D DWT technique to extract the features from the brain MR images and PCA technique is used to reduce features. The classification performance of the proposed PCA+EHO-ELM model is compared with other three models. The results obtained from the simulation study demonstrates that the proposed model outperformed the other models in classification of Alzheimer, Glioma and Multiple Sclerosis datasets.

REFERENCES

1. G. Mohan, M. M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification", *Biomedical Signal Processing and Control*, 39, 139-161, 2018.
2. P. V. Prasad, Magnetic resonance imaging: methods and biologic applications, 124, *Springer Science & Business Media*, 2006.
3. J. Jiang, P. Trundle, J. Ren, "Medical image analysis with artificial neural networks", *Computerized Medical Imaging and Graphics*, 34(8), 617-631,2010.
4. N. V. Shree, T. N. R. Kumar, "Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network", *Brain Informatics*, 5, 23–30, 2018.
5. S. Chaplot, L.M.Patnaik, N. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical signal processing and control*, 1(1), 86-92,2006.
6. J. Jiang, P. Trundle, J. Ren, "Medical image analysis with artificial neural networks", *Computerized Medical Imaging and Graphics*, 34(8), 617-631,2010.
7. M. Saritha, K. Paul Joseph, Abraham T. Mathew, "Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network", *Pattern Recognition Letters*, 34, 2151–2156, 2013.
8. E. Dahshan, T. Hosny, A. Salem, "Hybrid intelligent techniques for MRI brain images classification", *Digital Signal Processing*, 20(2), 433-441, 2010.
9. G.B. Huang, Q. Zhu, C. Siew, "Extreme Learning Machine: Theory and applications", *Neurocomputing*, 70, 489-501, 2006.
10. G.B. Huang, H. Zhou, X. Ding, R. Zhang, " Extreme Learning Machine for Regression and Multiclass Classification", *IEEE transaction on System, Man and Cybernetics*, 42(2), 513-529, 2012.
11. D. R. Nayak, R. Dash, B. Majhi, S. Wang, "Combining extreme learning machine with modified sine cosine algorithm for detection of pathological brain", *Computers & Electrical Engineering*, 68, 366-380, 2018.
12. D. R. Nayak, R. Dash, B. Majhi, "Discrete ripplet-II transform and modified PSO based improved evolutionary extreme learning machine for pathological brain detection", *Neurocomputing,282*, 232-247,2018.
13. G. B.Huang, Q. Y. Zhu, C. K. Siew,"Extreme learning machine: a new learning scheme of feedforward neural networks", *Neural networks*, *2*, 985-990, 2004.
14. G. G. Wang, S. Deb, X. Gao, L. Coelho, "A new metaheuristic optimization algorithm motivated by elephant herding behavior, *International Journal of Bio-Inspired Computation*, 8(6), 394–409, 2017.
15. A. Sahlol, F. Ismail, A. Abdeldaim, A. Hassanien, "Elephant Herd Optimization with Neural Network: A case study on Acute Lymphoblastic Leukemia Diagnosis", *in: 12th IEEE International Conference on Computer Engineering and System (ICCES)*, 657-662, 2017.
16. E. Tuba, Z. Stanimirovic, " Elephant herding optimization algorithm for support vector machine parameters tuning", *In: Proceedings of the 2017 International Conference on Electronics, Computers and Artificial Intelligence (ECAI),* 1–5, 2017.
17. N. Meena, A. Swarnkar, N. Gupta, K. Niari, "Improved Elephant Herding Optimization for MultiObjective DER accommodation in Distributed systems", *IEEE Transaction on Industrial Informatics*, 14(3), 1029-1039, 2018.
18. M. Elhosseini, R. Sehiemy, Y. Rashwan, X. Gao, "On the performance improvement of elephant herding optimization", *Knowledge-Based System*, 166, 58-70, 2019.
19. A.Hassanien, M. Kilany, E. Houssein, H. Alqaheri, " Intelligent Human emotion recognition based on elephant herding optimization tuned support vector regression", *Biomedical Signal Processing and Control*, 45, 182-191, 2018.
20. G. Wang, S. Deb, L. Coelho, "Elephant herding Optimization", *In: IEEE International Symposium on Computational and Business Intelligence*, 1-5, 2015.

AUTHORS PROFILE

Pranati Satapathy: She is working as a Lecturer in Integrated MCA department, Utkal University and continuing her research work in Utkal University. Her research area includes medical image processing and soft computing**.** She has 8 years of teaching experience.

Sateesh Kumar Pradhan: He is working as a Professor in the department of Computer Science and Applications in Utkal University. His research area includes computer networking, parallel computing, soft computing etc. He has published more than 50 research papers in various journals and conferences.

Sarbeswara Hota: He is working as an Assistant Professor in the department of Computer application in SOA deemed to be university. He has 17 years of teaching experience. His research area includes financial computing, soft computing and machine learning.