Extreme Learning Machines with Feature Selection Using GA for Effective Prediction of Fetal Heart Disease: A Novel Approach

Debjani Panda
KIIT University, Bhubaneswar, India
E-mail: pandad@indianoil.in

Divyajyoti Panda
National Institute of Technology, Rourkela, India
E-mail: pandadivya02@gmail.com

Satya Ranjan Dash
KIIT University, Bhubaneswar, India
E-mail: sdashfca@kiit.ac.in

Shantipriya Parida (corresponding author)
Idiap Research Institute, Martigny, Switzerland
E-mail: shantipriya.parida@idiap.ch

Heart disease is considered to be the most life-threatening ailment in the entire world and has been a major concern of developing countries. Heart disease also affects the fetus, which can be detected by cardiotocography tests conducted on the mother during her pregnancy. This paper analyses the presence of heart disease in the foetus by optimizing the Extreme Learning Machine with a novel activation function (roots). The accuracy of predicting the heart condition of the foetus is measured and compared with other activation functions like sigmoid, Fourier, tan hyperbolic, and a user-defined function, called "roots". The best features from the Cardiotocography data set are selected by applying the Genetic Algorithm (GA). ELM with activation functions sigmoid, Fourier, tan hyperbolic, and roots (a novel function), have been measured and compared on accuracy, sensitivity, specificity, precision, F-score, area under the curve (AUC), and computation time metrics. The GA uses three types of regression: linear, lasso, and ridge, for cross-validation of the features. ELM with user-defined activation function shows comparable performance with sigmoid and hyperbolic tangent functions. Features selected from linear and lasso produce better results in ELM than those selected from the ridge. It gives an accuracy of 96.45% as compared to 94.56% and 94.56% respectively with the best features selected from both linear and lasso. The roots activation function also takes 2.50 seconds computation time versus 3.27 seconds and 2.67 seconds for sigmoid and hyperbolic tangent respectively and scores better on all other metrics in designing an efficient model to classify fetal heart disease.

Povzetek: Z metodami strojnega učenja in genetskih algoritmov je analizirana bolezen srca pri fetusih.

1 Introduction

Cardiovascular disease is growing at a very fast rate and as per WHO, 30% of world population deaths occur due to cardiovascular heart diseases, and 23.6 million are expected to be affected by this disease by 2030 [3]. Cardiac disease is not only present in adults but can also be present as a birth anomaly in a newborn child and causes neonatal fatalities. The heart health of the fetus can be monitored to detect abnormal heartbeats and predict diseases affecting the fetus. Thus, predicting the cardiac health of a fetus is the need of the hour. Cardiotocography is one of the most commonly used Nonstress Tests which helps in determining the fetus's well-being in the womb and during labor.

Cardiotocography consists of uterine contractions and fetal heart rate. Fetal heart rate includes attributes like baseline heart rate, variations in baseline heart rate, accelerations, decelerations, and uterine contractions. This test is very useful in studying the base heart rate and uterine contractions pattern and is a vital tool for medical experts to know when a fetus is suffering from an inadequate supply of blood or oxygen to the body or any of its parts. As per the important factors identified by the National Institute of Child Health and Human Development (NICHD), baseline heart rate and its variability, accelerations, deceleration and Nonstress test (NST) are important factors to be considered while examining the well-being of the fetus [24].

The cardiotocography test is carried out by a device

called Electronic Fetal Monitor [27] which gives two signals fetal heart rate (FHR) and uterine contractions (UC). NST and contraction stress test (CST) are two main components of a CTG [8]. The NST determines whether the fetus is distressed and CST determines the placenta's respiratory function.

The normal range of FHR baseline lies between 110 bpm and 160 bpm. If the FHR baseline is higher than 160 bpm for more than 10 minutes, the fetus is considered to be suffering from tachycardia. On the other hand, if the FHR baseline is less than 110 bpm for more than 10 minutes is called bradycardia [6]. Both tachycardia and bradycardia are signs of fetal distress. The conditions are found out from NST which determines the fetal reactivity i.e. the interaction between the sympathetic and parasympathetic autonomous nervous system of the fetus.

Recently machine learning with the use of artificial intelligence has become an important and powerful tool for predicting the heart health of patients. They are effective in both binary and multi-class classification and are effective in predicting cardiac disease. One of the effective tools which are being used for the learning process for single hidden layer feeds forward neural networks (SLFNs) is called extreme learning machine (ELM) [2]. The prime benefit of ELM is that the hidden layer of SLFNs does not require tuning and it also has a fast rate of convergence [13]. The learning speed of ELM is considered to be thousands of times faster than the traditional feed-forward network learning algorithms [11]. Our study mainly focuses on using GA for feature selection and studying the accuracy of ELM using different activation functions.

The following section describes the details of the data set, implementation of ELM as a Classifier that uses the best features identified by the Genetic algorithm. The cross-validation methods used for obtaining the best features are studied thoroughly to study the impact of ELM with four activation functions. The purpose is to study the effectiveness of the novel activation function by comparing it with existing activation functions.

2   Methods

2.1   Workflow diagram

The process flow of our proposed model is as described below in Figure 1. The data set is considered with output class NSP and is pre-processed to remove duplicate entries. Using GA for obtaining the best features, the model is cross-validated with 3 regression models and the performance of ELM is studied before and after feature selection with the existing and novel activation function.

2.2   Dataset details

The Cardiotocography Data Set, obtained from UCI repository [9], has been used for our study and experimentation. The data set originally has 2126 instances with 23

attributes. The CTGs were also classified by three expert obstetricians into 2 types of classes including the class pattern (1-10) and fetal state class (N=Normal, S=Suspect, P= Pathologic). The data set has 21 attributes and two output classes. Our experiment is focused on considering all 21 attributes along with one output class. Similar to other studies conducted on this data set, our experiment also considers 22 attributes where 21 attributes are inputs and the 22nd attribute is the output class "NSP". We have not considered the other output class "CLASS" for our study. 21 attributes with NSP as the output class, described in Table 1.

2.3   Data pre-processing and splitting of data sets for model training

Other than the aforementioned 21 features and the output columns, 'CLASS' and 'NSP', the original database has 23 other columns, which were removed. Thereafter, the data set, named 'DT', were split into two subset data sets 'DT_CLASS' and 'DT_NSP' containing 'CLASS' and 'NSP' respectively. 12 duplicate rows were deleted, and the last four rows containing null values were also removed.

The data set of DT_NSP was split to an 80:20 ratio to train the classifiers on 80% of the data and perform the testing on the remaining 20% of the data.

2.4   Feature Selection and classification

Feature Selection is an important part of designing a predictive model to reduce unwanted features and also to reduce the training time of classifiers. In this paper, the important features are identified by using the Genetic Algorithm.

The training data set were given as input to ELM with different activation functions and their accuracy was studied. Linear, lasso, and ridge regression models have been used for cross-validation of candidate feature subsets generated by GA. The attributes selected are considered as best features and the classification algorithms performance has been tabulated.

2.4.1   Genetic Algorithm (GA)

The genetic algorithm is a simple Evolutionary search heuristic algorithm that randomly generates a new population. Its basic objective is to find the attributes with maximum fitness value in the population [14]. Based on the Darwinian Principle, it tries to find the fittest individuals. The entire set of candidate solutions is called a population and each solution is called an individual. Our Genetic algorithm searches for the solution which gives the minimum cross-validation error through linear, lasso, and ridge regression models. The chromosomes are generated with fitness values as true or false for each attribute and after iterating for the total number of generations the features are determined which are best fit to predict the outcome. GA
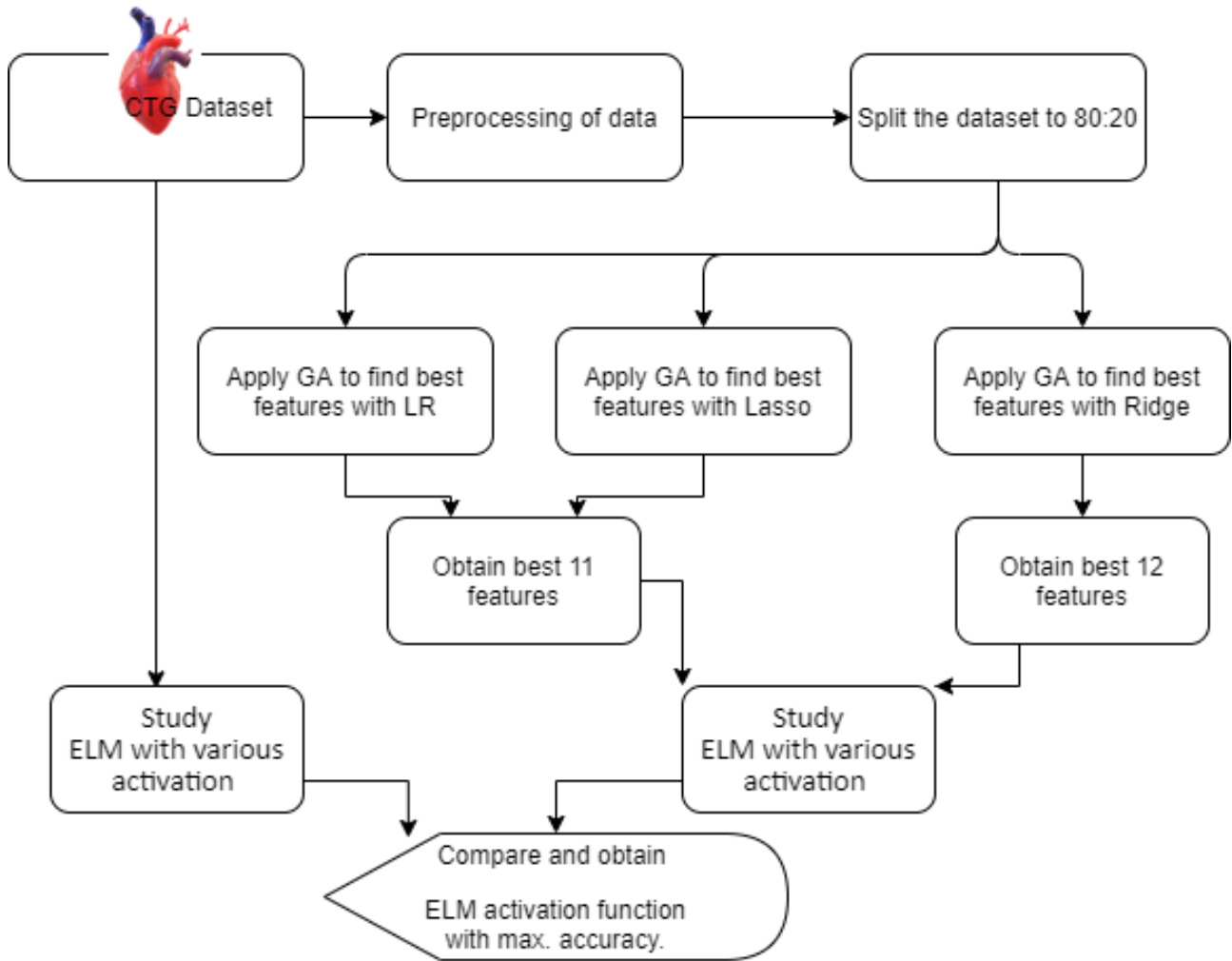
Figure 1: Process flow diagram to study impact of feature selection on ELM with various activation functions.

depends upon the number of generations, number of chromosomes, number of children created during the crossover, and best chromosomes. Depending upon the best fitness values, parents are selected for mating [1]. Crossover has been carried out with 2 parents and mutated to generate the new population and the process was repeated for 20 generations after which the fitness value of features remained constant. Finally, the features with the best fitness values are obtained.

Regression models: It is a supervised method in machine learning to find the correlation of dependant variables in terms of the independent variables. It is effectively used for dimensionality reduction of collinear or multi-collinear variables. The following regression models are used in GA:

Linear Regression: The equation can be written as shown in Equation 1.

$$y = \beta_0 + \sum_{k=1}^{p} of \beta_{ik} x_{ik} \tag{1}$$

Ridge: This method uses L2 regularization, where L2

is the penalty equivalent [23] to the sum of the magnitude of coefficients. This type of regression [22] helps in dealing with a variance that is resultant of the multi-collinearity of variables. It helps in reducing the variance which is a resultant of non-linear relationships between two independent variables.

Lasso: This model is based on L1 regularization in which the least related variables are treated as zero. So, it helps minimize irrelevant features. It adds a penalty to minimize the loss of a model. L1 is the penalty added to the sum of the absolute value of coefficients. For the objective function (Equation 2),

$$\frac{\sum_{I=1}^{N} of f(x_i, \, y_i, \, \alpha, \, \beta)}{N} \tag{2}$$

the lasso regularized version of the estimator will be the solution to the Equation 3.

$$\min_{\alpha, \, \beta} of \frac{\sum_{I=1}^{N} of f(x_i, \, y_i, \, \alpha, \, \beta)}{N}, \text{subject to } \|\beta\|_1 < t \tag{3}$$

Attributes	Description
LB	Fetal base line heart rate
AC	Accelerations per second
FC	Fetal movements per second
UC	Uterine contractions per second
ASTV	percentage of time with abnormal short-term variability
mSTV	mean value of short-term variability
ALTV	percentage of time with abnormal long-term variability
mLTV	mean value of long-term variability
DL	mean light decelerations per second
DS	mean severe decelerations per second
DP	mean prolonged decelerations per second
Width	mean histogram width
Min	low frequency of the histogram
Max	high frequency of the histogram
NMax	number of histogram peaks
Nzeros	number of histogram zeros
Mode	histogram mode
Mean	histogram mean
Median	histogram median
Variance	histogram variance
Tendency	histogram tendency: -1=left asymmetric; 0=symmetric; 1=right asymmetric

Table 1: Cardiotocography (CTG) Data set with detail description of attributes.

where only β is penalized while α is free to take any allowed value, just as β_0 was not penalized in the basic case, and t is a pre-specified free parameter that determines the amount of regularisation.

The basic algorithm used for feature selection through GA is as follows:

1. The initial population was randomly initialized by creating individuals that included/excluded certain features. One particular individual may have the chromosome, as shown in Figure 2, where each box represents a gene, or feature in the data set, green indicates "True" (the feature is included in the chromosome) and red indicates "False" (the feature is excluded from the chromosome).

2. For each generation:

   (a) The fitness score is calculated for an individual as follows:

      i. Using regression models, the presence or absence of the feature is determined. The target is modeled with the features with values like 1 for being present and 0 for being absent.
      For example, if the individual illustrated in Figure 2 is taken, then all the features, excluding the 2nd, 8th, 12th, and 17th features, are taken for modeling.

      ii. The cross-validation scores were determined using negative mean square error (NMSE), calculated as shown in Equation 4.

$$NMSE = -\frac{\sum_{i=1}^{n} of\left(x_i - \frac{\sum_{i=1}^{n} of x_i}{n}\right)^2}{n}$$
$$= \left(\frac{\sum_{i=1}^{n} of x_i}{n}\right)^2 - \frac{\sum_{i=1}^{n} of x_i^2}{n} \quad (4)$$

      iii. The mean of the cross-validation scores was assigned to the fitness value.

   (b) The individuals were sorted in the increasing order of their fitness values.

   (c) The last n individuals (which are the best n individuals of the population) were selected out of the population.

   (d) In the selected individuals, for number i in the range of $\lfloor \frac{n}{2} \rfloor$, the ith and $(n - i)^{th}$ individuals were crossed as shown in Figure 3.

   (e) The daughter chromosome was mutated as shown in Figure 4 to generate new population:

3. The fittest individual was selected and its genes were recorded.

2.5 ELM for multi class classification

Extreme Learning Machines are effective single-layer feedforward networks (SLFNs) with hidden neurons that do
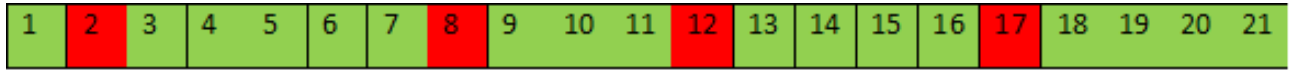
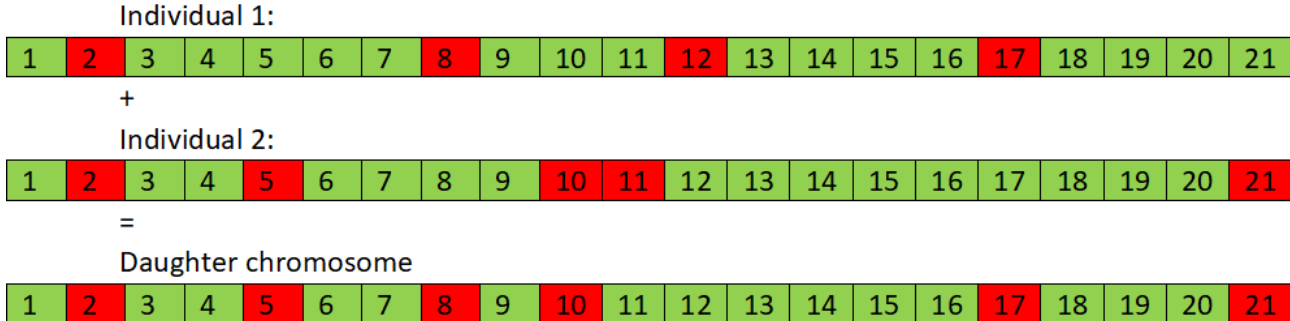Figure 2: A typical chromosome with 21 features

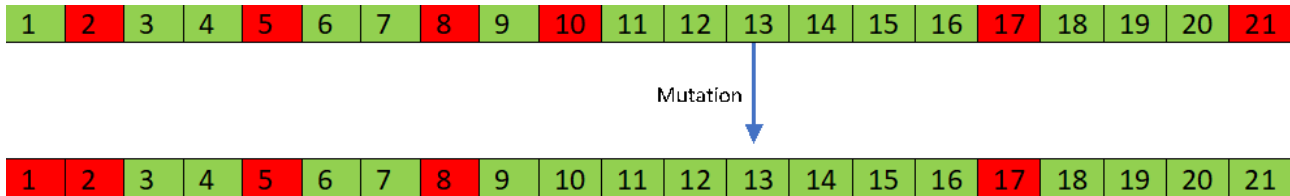Figure 3: Crossing over of two chromosomes

Figure 4: Daughter Chromosome with 21 features undergoing mutation

not require further tuning [17] and can very effectively be trained with minimum time for classification, regression, and feature selection. ELM randomly assigns connections between the input layer and the hidden neurons and they do not change further during the learning process. The output connections are then adjusted to obtain the solution with minimum cost [12]. There are various types of ELM like Simple ELM, ELM of ensembles, Pruned ELM, and incremental ELM [17][12].

In our study, a simple ELM is studied with [5][19] sigmoid, hyperbolic tangent, Fourier and roots activation functions, and their performances have been compared based on training time, accuracy, specificity, F measure score, sensitivity, precision, and AUC.

Basic ELM can be represented as:

For N arbitrary distinct samples $(x_i, \ t_i) \in R^d \times R^m$, SLFNs with L hidden nodes having parameters $(a_i, \ b_i), \ i \in \{1, \ 2, \ \ldots, \ L\}$ are mathematically modelled as in Equation 5.

$$\sum_{i=1}^{L} of\beta_i \ g_i(x_j) = \sum_{i=1}^{L} of\beta_i G(a_i, b_i, x_j) \tag{5}$$
$$= o_j, \ j \in \{1, 2, \ldots, N\}$$

Where β_i, is the output weight of the ith hidden node and $g(x)$ is an activation function.

SLFNs approximates these N samples with zero error. Mathematically, it can be represented as in Equations 2.5 and 2.5

$$\sum_{j=1}^{L} of \|o_j - t_j\| = 0 \tag{6}$$

$$\exists \ (a_i, \ b_i), \ \beta_i \ |$$
$$\sum_{i=1}^{L} of\beta_i \ G(a_i, \ b_i, \ x_j \ ) = t_j, j \in \{1, 2, \ldots, N\} \tag{7}$$

equations and can be written compactly as in Equations 8

$$H\beta = T \tag{8}$$

where,

$$H = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}$$
$$= \begin{bmatrix} G(a_1,b_1,x_1) & \ldots & G(a_L,b_L,x_1) \\ \vdots & \ddots & \vdots \\ G(a_1,b_1,x_N) & \vdots & G(a_L,b_L,x_N) \end{bmatrix}_{N \times L} \tag{9}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix}_{L \times m} \tag{10}$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{11}$$

If X and Y denote the input and output of the function, W_1 and W_2 denote the weight and bias matrices, and G denotes the activation function, then for an ELM learning a model of the form given in Equation 12, W_1 is initialized randomly and W_2 is estimated as shown in Equation 13

$$Y = W_2 G(W_1 X) \tag{12}$$

$$W_2 = G(W_1 X)^+ Y \tag{13}$$

where + denotes Moore-Penrose inverse.

Four different non-linear activation functions have been used in our experiment out of which one function is user-defined. The list of functions used are mentioned in Equations 14-17.

Sigmoid Function:

$$G(a, b, x) = \frac{1}{1 + e^{-(ax+b)}} \tag{14}$$

Fourier Function:

$$G(a, b, x) = \sin(ax + b) \tag{15}$$

Hyperbolic Tangent Function:

$$G(a, b, x) = \tanh(ax + b) \tag{16}$$

Roots Function (User-defined):

$$G(a, b, x) = \begin{cases} 0, x = \frac{-b}{a} \\ \frac{|ax+b|^{n+1}}{ax+b}, x \neq \frac{-b}{a} \end{cases} \tag{17}$$

where $n \in R$ is a parameter, which can take any value between 0 and 1. If the value of n is given as 1, then it becomes a linear function.

The various activation function graph is attached in Fig.5.

3   Results

3.1   Experimental setup

All computations are performed on Intel (R) Core (TM) i5-10210U CPU @2.11GHz with 64bit Windows 10 operating system. Moreover, Python 3.6.5 software package is used to simulate the experiments.

3.2   Metrics and analysis

In the CTG dataset, the output class with value 3 is pathological cases, and our experiment focuses on finding out pathological cases so that they can be used to predict the heart disease of the fetus.

The features selected from GA by applying linear regression and lasso for cross-validation, yield the same set of features and the best 11 features are considered, which are LB, UC, DS, DP, ASTV, ALTV, MLTV, Width, Max, Median, and Variance. The model performance is also measured by applying ridge regression, which considers the best 12 features, which are LB, UC, DS, DP, ASTV, ALTV, MLTV, Min, Max, Nmax, Median, and Variance.

The roots function has been tested with values for n as 0.25, 0.4, and 0.5. The number of hidden units considered for the study is 200. The various metrics used for comparison include confusion matrix, precision, accuracy, F measure, and AUC.

The confusion matrix taken for the classification of pathological cases in the CTG data set is shown in Table 2.

Predicted → Actual ↓	1	2	3
1	TN	TN	FP
2	TN	TN	FP
3	FN	FN	TP

Table 2: Confusion matrix

where,

- TP: True positive, where output class 3 is predicted as pathological case

- TN: True negative, where output classes 1 and 2 are predicted as non-pathological (normal or suspect) case

- FP: False positive, where output classes 1 and 2 are predicted as pathological case, and

- FN: False-negative, where output class 3 is predicted as non-pathological case.

The metrics used for measuring classification success are mentioned in Equations 18-23.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{19}$$

$$Specificity = \frac{TN}{TN + FP} \tag{20}$$

$$Precision = \frac{TP}{TP + FP} \tag{21}$$

$$F\ measure = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \tag{22}$$

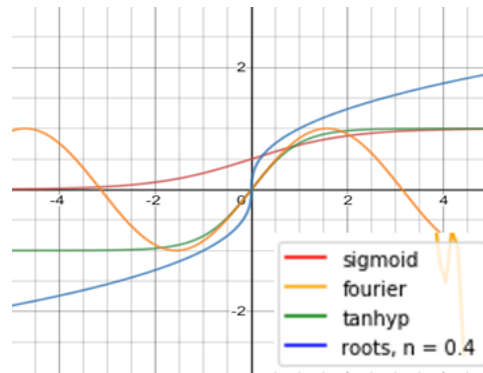$$AUC = \frac{Sensitivity + Specificity}{2} \tag{23}$$

Figure 5: Graph of various activation functions in ELM

The inbuilt ELM module python has been compared with ELM models with sigmoid, Fourier, hyperbolic tangent, and roots (n = 0.25, 0.4, 0.5) based on their accuracy for predicting heart disease before and after feature selection. The results are tabulated as shown in Table 3.

The built-in ELM function in python suffers from the problem of under-fitting for which our study focuses on an alternate set of activation functions to be used for building the model. The graph to measure the accuracy of different activation functions with varying hidden nodes before feature selection is shown in Figure 6. The user-defined activation function, which has been named as "roots" is plotted against the other available functions of ELM. Our model outperforms other in-built functions in terms of accuracy in many instances when hidden inputs are varied from 0 to 1000.

When hidden inputs cross 200 units, the graphs of a sigmoid, hyperbolic tangent and roots activation functions are almost consistent. The hidden inputs are fixed at 200, to study other metrics for evaluating the roots function performance.

It is also observed from Table 3, the function of that root with n=0.4 has given optimum results. The graphs have been plotted with n=0.4 while using the function of the root for ELM. Figure 7 shows the graph after using selected features from linear regression and Lasso in GA. Figure 8 shows the results of ELM with features selected from Ridge regression in GA.

The results showed that ELM with Sigmoid, Roots and Hyperbolic tangent activation functions performed better than the Fourier activation function. The Fourier activation function is not considered for further study as it is not sensitive to feature selection. The three functions were then analyzed based on their computation time and other metrics for classifying pathological cases. It depicts that the function of the roots takes lesser time than sigmoid and tan hyp to compute the results for the testing samples. The activation functions of ELM were studied with the original feature set and a reduced feature set by applying GA and results have been tabulated in Table 4 and Table 5.

When compared to selected features using GA with ridge, GA with Linear and Lasso regression yielded better performance on all metrics. The graph in Figure 6 shows the performance of the 4 activation functions before applying feature selection. The performance of ELM models improved after feature selection using the three activation functions sigmoid, hyperbolic tan, and roots(n=0.4). The roots activation function performed better than the other 2 activation functions. Figure 7 depicts the graph for measuring the performance of ELM after feature selection. The best features derived from Linear and Lasso cross-validation in GA have been used to plot the graph. Another graph shows the improved performance of ELM after using feature selection from best features obtained from Ridge with GA and is shown in Figure 8. The results are dependent upon the number of hidden inputs taken and can change. It has also been observed that by varying the hidden inputs from 0 to 1000, ELM with roots activation function, has outperformed in classification in the majority of the cases.

The standard inbuilt ELM remains unaffected and even the performance degrades after feature selection. Its performance reduced from 11.11% to 10.17% after using feature selection. However, the customized ELMs show improvement in terms of accuracy. The ELM with sigmoid function has improved from 92.67 to 94.56%, with hyperbolic tangent function the performance improved from 93.14 to 94.56% and with roots activation function (n=0.4) the accuracy improved from 94.33 to 96.45%. The ELM with Fourier activation function remains unchanged before and after feature selection and has given 90.07% accuracy throughout the experiment.

4 Discussion

Various fetal disease prediction systems have been proposed for diagnosing the fetus's health. The works have been carried out on Cardiotocography data set using either all the features or a subset of them. One of the study [30] focuses on using various types of ANN like MLPNN, PNN, and GRNN models using the entire data set to identify the fetal state and have reported the overall classification accuracy's for MLPNN, PNN, and GRNN as 90.35, 92.15, and

Classifier	Before Feature Selection	After Feature Selection		
		Linear	Lasso ($\alpha = 0.0001$)	Ridge ($\alpha = 0.0001$)
ELM inbuilt in Python	11.11	10.17	10.17	10.16
ELM with new activation function(n=0.25)	94.56	95.98	95.98	95.74
ELM with new activation function(n=0.4)	94.33	96.45	96.45	95.04
ELM with new activation function(n=0.5)	94.8	96.21	96.21	95.04
ELM with sigmoid activation	92.67	94.56	94.56	94.33
ELM with Fourier activation	90.07	90.07	90.07	90.07
ELM with hyperbolic tangent activation	93.14	94.56	94.56	93.85

Table 3: Classification Performance of ELM (Accuracy%) with best features obtained by applying Genetic Algorithm (GA) using Linear, Lasso and Ridge regression models.

ELM activation functions →/	sigmoid		roots (n=0.4)		Hyperbolic tangent	
	Original DT	Best features (11 attributes)	Original DT	Best features (11 attributes)	Original DT	Best features (11 attributes)
Confusion Matrix	294 29 2 26 28 2 4 23 15	299 25 1 24 30 2 2 18 22	306 19 0 21 32 3 4 17 21	301 24 0 24 32 0 2 13 27	295 30 0 24 30 2 3 24 15	299 26 0 24 29 3 1 19 22
Accuracy	92.67	94.56	94.33	96.45	93.14	94.56
Sensitivity	35.71	52.38	50.00	64.29	35.71	52.38
Specificity	98.95	99.21	99.21	100.00	99.48	99.21
Precision	78.95	88.00	87.50	100.00	88.24	88.00
F-measure	49.18	65.67	63.64	78.26	50.85	65.67
AUC	67.33	75.80	74.61	82.14	67.59	75.80
Computation Time in secs	3.05	3.27	2.31	2.50	2.67	2.67

Table 4: Measurement Metrics for Sigmoid, Roots, and Hyperbolic Tangent activation functions before and after feature selection by GA through linear/lasso regression with 200 hidden inputs (in %).

ELM activation functions →/	sigmoid		roots (n=0.4)		Hyperbolic tangent	
	Original DT	Best features (12 attributes)	Original DT	Best features (12 attributes)	Original DT	Best features (12 attributes)
Confusion matrix	294 29 2 26 28 2 4 23 15	295 25 1 25 30 1 6 16 20	306 19 0 21 32 3 4 17 21	297 27 1 21 33 2 2 16 24	295 30 0 24 30 2 3 24 15	294 30 1 25 30 1 6 18 18
Accuracy	92.67	94.33	94.33	95.04	93.14	93.85
Sensitivity	35.71	47.62	50.00	57.14	35.71	42.86
Specificity	98.95	99.48	99.21	99.21	99.48	99.48
Precision	78.95	90.91	87.50	88.89	88.24	90.00
F-measure	49.18	62.50	63.64	69.57	50.85	58.06
AUC	67.33	73.55	74.61	78.18	67.59	71.17
Computation Time in secs	3.05	3.17	2.31	2.55	2.67	2.71

Table 5: Measurement Metrics for Sigmoid, Roots, and Hyperbolic Tangent activation functions before and after feature selection by GA through ridge regression with 200 hidden inputs (in %).
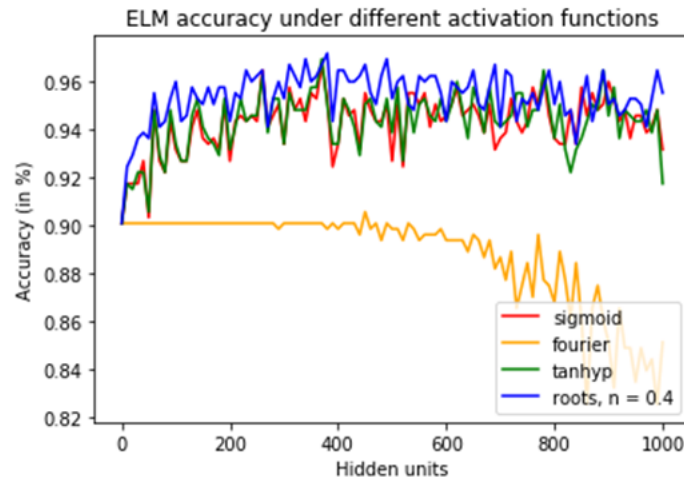
Figure 6: ELM accuracy before feature selection with Sigmoid, Fourier, Hyperbolic tangent(tanhyp) and Roots(user-defined) activation functions.
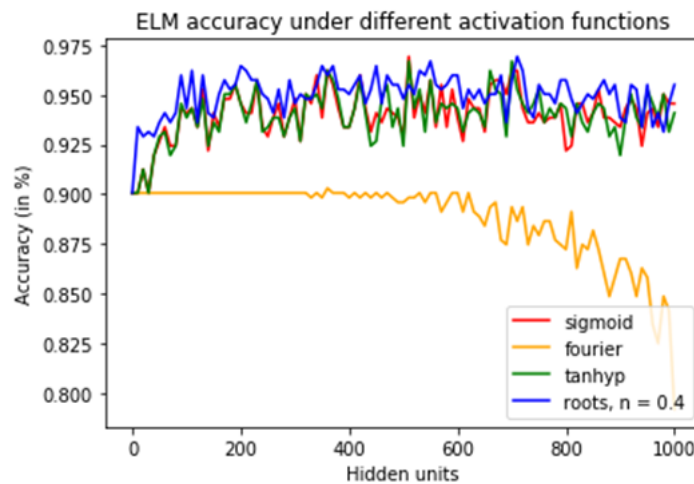

Figure 7: ELM accuracy with sigmoid, Fourier, hyperbolic tangent (tanhyp) and roots(user-defined) activation functions, after feature selection with best 11 features selected through GA with linear/lasso regression for cross-validation.

91.86%, respectively. Another work proposes using Discriminant Analysis, Decision Trees, and Artificial Neural network for identifying the fetal status [15] using all features of the CTG data set and have reported 82% accuracy for DA, 86.4% for DT, and 97.8% accuracy for ANN. The work also establishes the fact that giving rules for identification is always better i.e DT even with lower accuracy is better interpretative for results rather than an Artificial neural network which resembles a black box where processes involved are unknown. Another work including all the features [31], focuses on studying fetal well-being using The Least Square SVM method with Particle Swarm Optimization and Decision Trees. This method yielded 91.62% accuracy with all 2162 instances and had been validated using 10-fold cross-validation. The PSO played a major role in optimizing the penalty factor of LS-SVM. A similar work proposed using Adaptive Neuro-Fuzzy inference Systems (ANFIS) [21] to differentiate pathological cases from nor-

mal ones and reported accuracy of 97.2% for normal cases and 96.6% accuracy for pathological states. Rough Neural Networks suggested in another study for fetal risk assessment [4] was provided with upper and lower boundaries in input layer as well as hidden layers and gave an accuracy of 92.95% for pathological cases using the entire set of features.

The above works have reported a maximum accuracy of 97.8% and have used all the features for the experiment. In comparison, our work has yielded 96.45% accuracy with only 11 features, thus reducing the computation cost and time.

Our paper focuses on studying the efficiency of ELM with novel activation function for the effective classification of fetal heart disease. The accuracy of our model is compared before and after feature selection using GA. GA uses regression models for cross-validation of the best features and linear as well lasso have yielded the same 11
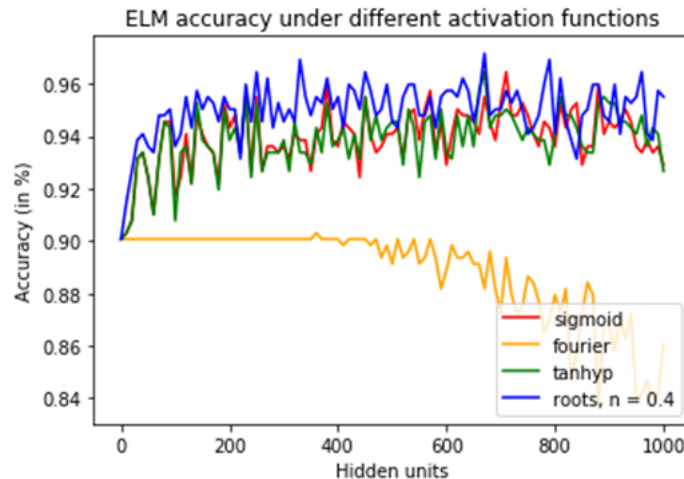
Figure 8: ELM accuracy with sigmoid, Fourier, hyperbolic tangent (tanhyp) and roots(user-defined) activation functions, after feature selection with best 12 features selected through GA with ridge regression for cross-validation.

best features in our experiment and have given better accuracy than the 12 features selected with ridge regression. As compared to a study [29], the features selected by convolution neural networks, MKNet and MKRNN which resulted in classification accuracy of 90%, our feature selection has given better performance and accuracy has improved by almost 6%.

For classification of heart disease, extraction of important features plays an important role as evident from [28], [20], [32]. Generalized discriminant analysis has been used with the Radial basis kernel function or Gaussian function of ELM to analyze heart rate signals and the process has achieved 100% accuracy. The impact of feature selection was therefore explored in our study using GA as GA in the study [20]. gave improved results for the classification of heart disease. Our accuracy also improved from 94.33% to 96.45% after using the best features obtained from GA with a lasso and linear regression models, similar to [32] where accuracy improved by 5.6% using PCA.

Our model has given improved results of 96.45% accuracy as compared to another work [7] which used ANN with ELM for classifying fetal heart disease and have given 93.42% and 91.84% using ELM and ANN respectively.

ELM with our novel activation function roots (n=0.4), also outperformed the results of classification using various other classifiers given in [10], where XGBoost gave the best results with (>92%) and was comparable with other optimized ELM models used for classification of various other diseases.[18]

The number of hidden inputs in our study has been considered to be 200 as compared to 2 to 3 input units suggested in the work [16] and the inputs have been selected by varying the units from 0 to 1000 and the optimized value has been considered for the ELM.

Our novel approach gave 100% specificity and 100% sensitivity as compared to other classification models [26]. The best features selected by other studies [25] are also the common features that have been selected by using GA with cross-validation using linear and lasso and have obtained accuracy >2% as compared to classification and regression decision trees and Self-organizing maps.

5 Conclusion

ELM with sigmoid and roots activation functions produced accuracy above 95%. ELM takes less time than other neural networks to get trained as their input weights and biases do not need to be tuned further, but it depends on the activation function used. In this experiment, the function of the roots was faster than other functions when hidden units were set to 200. The Genetic Algorithm has played an important factor in improving the accuracy of ELMs through feature selection. Other activation functions can also be used to see the effect on various parameters for classifying pathological cases in Cardiotocography data sets. The models can be used as an effective tool to aid medical experts in detecting cardiological abnormalities in the fetus.

Future work can be carried out on the optimization of various other activation functions of ELM to analyze the impact of the selection of hidden units on computation time and accuracy. Depending upon the dataset, the number of hidden units, and the number of generations, the activation function can be optimized to find the value of n in the function of the user-defined roots, which will determine the best results. The current study has used regression techniques for cross-validation in GA and in the future other techniques can be used to examine the model.

References

[1] Aalaei, S., Shahraki, H., Rowhanimanesh, A., Eslami, S., 2016. Feature selection using genetic algorithm

for breast cancer diagnosis: experiment on three different datasets. Iranian journal of basic medical sciences 19, 476.

[2] Albadra, M.A.A., Tiuna, S., 2017. Extreme learning machine: a review. International Journal of Applied Engineering Research 12, 4610–4623.

[3] Alwan, A., et al., 2011. Global status report on noncommunicable diseases 2010. World Health Organization. `https://doi.org/10.2471/blt.11.091074`.

[4] Amin, B., Gamal, M., Salama, A., Mahfouz, K., El-Henawy, I., 2019. Classifying cardiotocography data based on rough neural network. machine learning 10. `https://doi.org/10.14569/ijacsa.2019.0100846`.

[5] Cao, J., Lin, Z., 2015. Extreme learning machines on high dimensional and large data applications: a survey. Mathematical Problems in Engineering 2015. `https://doi.org/10.1155/2015/103796`.

[6] Chen, C.Y., Chen, J.C., Yu, C., Lin, C.W., 2009. A comparative study of a new cardiotocography analysis program, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE. pp. 2567–2570. `https://doi.org/10.1109/iembs.2009.5335287`.

[7] Cömert, Z., Kocamaz, A.F., Güngör, S., . Classification and comparison of cardiotocography signals with artificial neural network and extreme learning machine `https://doi.org/10.17678/beuscitech.338085`.

[8] Cömert, Z., Kocamaz, A.F., Güngör, S., 2016. Cardiotocography signals with artificial neural network and extreme learning machine, in: 2016 24th Signal Processing and Communication Application Conference (SIU), IEEE. pp. 1493–1496. `https://doi.org/10.1109/siu.2016.7496034`.

[9] Dua, D., Graff, C., 2017. UCI machine learning repository. URL: `http://archive.ics.uci.edu/ml`.

[10] Hoodbhoy, Z., Noman, M., Shafique, A., Nasim, A., Chowdhury, D., Hasan, B., 2019. Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data. International Journal of Applied and Basic Medical Research 9, 226. `https://doi.org/10.4103/ijabmr.ijabmr_370_18`.

[11] Huang, G., Huang, G.B., Song, S., You, K., 2015. Trends in extreme learning machines: A review. Neural Networks 61, 32–48. `https://doi.org/10.1016/j.neunet.2014.10.001`.

[12] Huang, G.B., Wang, D.H., Lan, Y., 2011. Extreme learning machines: a survey. International journal of machine learning and cybernetics 2, 107–122. `https://doi.org/10.1007/s13042-011-0019-y`.

[13] Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70, 489–501. `https://doi.org/10.1016/j.neucom.2005.12.126`.

[14] Huang, J., Cai, Y., Xu, X., 2007. A hybrid genetic algorithm for feature selection wrapper based on mutual information. Pattern Recognition Letters 28, 1825–1844. `https://doi.org/10.1016/j.patrec.2007.05.011`.

[15] Huang, M.L., Hsu, Y.Y., 2012. Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network `https://doi.org/10.4236/jbise.2012.59065`.

[16] Jadhav, S., Nalbalwar, S., Ghatol, A., 2011. Modular neural network model based foetal state classification, in: 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW), IEEE. pp. 915–917. `https://doi.org/10.1109/bibmw.2011.6112501`.

[17] Li, B., Li, Y., Rong, X., 2013. The extreme learning machine learning algorithm with tunable activation function. Neural Computing and Applications 22, 531–539. `https://doi.org/10.1007/s00521-012-0858-9`.

[18] Li, Q., Chen, H., Huang, H., Zhao, X., Cai, Z., Tong, C., Liu, W., Tian, X., 2017. An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis. Computational and mathematical methods in medicine 2017. `https://doi.org/10.1155/2017/9512741`.

[19] Miche, Y., Sorjamaa, A., Bas, P., Simula, O., Jutten, C., Lendasse, A., 2009. Op-elm: optimally pruned extreme learning machine. IEEE transactions on neural networks 21, 158–162. `https://doi.org/10.1109/tnn.2009.2036259`.

[20] Nikam, S., Shukla, P., Shah, M., . Cardiovascular disease prediction using genetic algorithm and neuro-fuzzy system `https://doi.org/10.21172/1.82.016`.

[21] Ocak, H., Ertunc, H.M., 2013. Prediction of fetal state from the cardiotocogram recordings using adaptive neuro-fuzzy inference systems. Neural Computing and Applications 23, 1583–1589. `https://doi.org/10.1007/s00521-012-1110-3`.

[22] Panda, D., Ray, R., Abdullah, A.A., Dash, S.R., 2019. Predictive systems: Role of feature selection in prediction of heart disease, in: Journal of Physics: Conference Series, IOP Publishing. p. 012074. `https://doi.org/10.1088/1742-6596/1372/1/012074`.

[23] Panda, D., Ray, R., Dash, S.R., 2020. Feature selection: Role in designing smart healthcare models, in: Smart Healthcare Analytics in IoT Enabled Environment. Springer, pp. 143–162. `https://doi.org/10.1007/978-3-030-37551-5_9`.

[24] Parer, J., Quilligan, E., Boehm, F., Depp, R., Devoe, L.D., Divon, M., Greene, K., Harvey, C., Hauth, J., Huddleston, J., et al., 1997. Electronic fetal heart rate monitoring: research guidelines for interpretation. American Journal of Obstetrics and Gynecology 177, 1385–1390. `https://doi.org/10.1016/s0002-9378(97)70079-6`.

[25] Peterek, T., Gajdoš, P., Dohnálek, P., Krohová, J., 2014. Human fetus health classification on cardiotocographic data using random forests, in: Intelligent Data analysis and its Applications, Volume II. Springer, pp. 189–198. `https://doi.org/10.1007/978-3-319-07773-4_19`.

[26] Sahin, H., Subasi, A., 2015. Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. Applied Soft Computing 33, 231–238. `https://doi.org/10.1016/j.asoc.2015.04.038`.

[27] Schmidt, J.V., McCartney, P.R., 2000. History and development of fetal heart assessment: a composite. Journal of Obstetric, Gynecologic, & Neonatal Nursing 29, 295–305. `https://doi.org/10.1111/j.1552-6909.2000.tb02051.x`.

[28] Singh, R.S., Saini, B.S., Sunkaria, R.K., 2018. Detection of coronary artery disease by reduced features and extreme learning machine. Clujul Medical 91, 166. `https://doi.org/10.15386/cjmed-882`.

[29] Tang, H., Wang, T., Li, M., Yang, X., 2018. The design and implementation of cardiotocography signals classification algorithm based on neural network. Computational and mathematical methods in medicine 2018. `https://doi.org/10.1155/2018/8568617`.

[30] Yılmaz, E., 2016. Fetal state assessment from cardiotocogram data using artificial neural networks. Journal of Medical and Biological Engineering 36, 820–832. `https://doi.org/10.1007/s40846-016-0191-3`.

[31] Yılmaz, E., Kılıkçıer, Ç., 2013. Determination of fetal state from cardiotocogram using ls-svm with particle swarm optimization and binary decision tree. Computational and mathematical methods in medicine 2013. `https://doi.org/10.1155/2013/487179`.

[32] Zhang, Y., Zhao, Z., 2017. Fetal state assessment based on cardiotocography parameters using pca and adaboost, in: 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE. pp. 1–6. `https://doi.org/10.1109/cisp-bmei.2017.8302314`.