A BINARY BAT INSPIRED ALGORITHM FOR THE CLASSIFICATION OF BREAST CANCER DATA

Doreswamy[1] and Umme Salma M[2]

[1]Department of Computer Science, Mangalore University, Mangalagangothri, Mangalore, India, 574199

[2]Department of Computer Science, Mangalore University, Mangalagangothri, Mangalore, India, 574199

ABSTRACT

Advancement in information and technology has made a major impact on medical science where the researchers come up with new ideas for improving the classification rate of various diseases. Breast cancer is one such disease killing large number of people around the world. Diagnosing the disease at its earliest instance makes a huge impact on its treatment. The authors propose a Binary Bat Algorithm (BBA) based Feedforward Neural Network (FNN) hybrid model, where the advantages of BBA and efficiency of FNN is exploited for the classification of three benchmark breast cancer datasets into malignant and benign cases. Here BBA is used to generate a V-shaped hyperbolic tangent function for training the network and a fitness function is used for error minimization. FNNBBA based classification produces 92.61% accuracy for training data and 89.95% for testing data.

KEYWORDS

Data mining, Classification, Binary Bat, FNN, Breast Cancer

1. INTRODUCTION

Medical data mining is a sub-branch of data mining which deals with extraction, transformation, analysis, interpretation and visualization of medical data stored on a computer. Analysis of medical data is interesting and equally challenging. In medical data mining, classification and prediction of data is not just a matter of accuracy but the matter of life and death. One wrong decision can have a disastrous effect on the life of patients and their families. Thus medical data mining is considered as the decision making frame work which provides assistance for the experts to properly classify and predict the data in a quick time. Classification techniques can be broadly divided into two categories, traditional classification techniques and modern classification techniques.Traditional classification problems are based on the design of classifiers working on the type of structural parameters chosen. If it is a fuzzy classifier then the rules, antecedent, consequent etc acts as the structural parameters, in case of K-Nearest Neighbor (KNN) classifier it is the distance metric and in Artificial Neural Network (ANN) the number of hidden layers, weights and biases serve as the structural parameters. Tuning of these parameters is a chaotic task. The modern classification techniques are the combination of advanced classification techniques such as SVM and ANN with the nature inspired algorithms which are meta-heuristic

DOI :10.5121/ijscai.2016.5301

in nature. Meta heuristic algorithms help us in designing non-parametric classifiers which directly classify the data based upon the updation of optimum decision function also called as cost function or on the basis of rules and conditions. These algorithms provide an optimal solution even in a complex search space. Apart from this, they are capable of escaping from the problem of local minima or maxima [30]. These two characteristics of meta-heuristic algorithms make them capable of producing highly accurate and robust solutions in the shortest time. The selection of nature-inspired algorithm depends upon the problem statement and the solution required for solving it. One such meta-heuristic, nature-inspired algorithm is Bat algorithm [31] which has diverse applications. It can be applied for accomplishment of classification task [16], for optimization problems (both single objective optimization and multi objective optimization) [32], for data prediction and so on. Bat algorithm mimics the way the Bat searches for its prey based upon echolocation technique. Using echolocation the Bat changes its direction and speed based upon the sound that strikes back after reaching the target. It updates its velocity randomly to reach its prey in the shortest span. Earlier studies reveal that Bat algorithm outperforms Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) in providing solution to the unconstrained optimization problems [9]. In this paper the authors are intended to test the performance of Bat algorithm for the classification of breast cancer data into benign and malignant classes. The breast cancer is chosen as a classification problem because it is one of the famous cancers, killing one among every four women [1]. In 2013-14, approximately 64,640 United States women were diagnosed with breast cancer [23]and immediate efforts were made to reduce the death rate by providing proper awareness, analysis and treatment. But, in the developing countries like India, the count is increasing with an alarming rate [2]. A better way to treat a disease is to find its patterns in its early stage.Thus, the authors have specifically chosen the breast cancer data so that through the data mining a betteranalysis can be provided which can help the doctors in decision making. The paper is categorized in the following way, after the introduction part the second section is motivation and related work, followed by preliminary view, proposed model, results and discussions and finally ends with conclusion and future work.

2. MOTIVATION AND RELATED WORK

Nature is both an inspiration and a motivation. Researchers and computer scientists too are inspired by nature and have found solutions for various problems by observing mother nature. The best example is ANN which is built based upon the design and functionality of human brain. Apart from ANN, many nature-inspired algorithms like Particle Swarm Optimization (PSO) algorithm, Bee Colony Optimization (BCO) algorithm, Ant Colony Optimization (ACO) algorithm etc were designed mainly to solve the optimization problems, but now, they are extended to find solutions for diverse problems. One such algorithm is Bat algorithm which has diverse applications [33]. In [6] a simple Bat algorithm was used to solve constrained based optimization problem. A novel hybrid algorithm was designed for global numerical optimization by using Bat algorithm and Harmonic Search (HS) [27] . Another approach used Bat algorithm for solving optimization problems in engineering field [34]. For the first time Bat algorithm was used to select features from various digital data sets [18]. In [22] Optimum-path forest technique and Bat algorithm were combined to select the features using wrapper approach. Gradually from optimization problem and feature selection the researchers focus shifted towards classification and clustering problem. Micro array data was classified using meta-heuristic Bat algorithm [16]. Bat algorithm was also used for clustering by combining traditional K-means clustering with simple Bat algorithm [25].

3. PRELIMINARIES

3.1 Feed-forward neural network

The simplest and the most famous ANN is a feed forward neural network (FNN). In FNN the information flows in unidirectional way i.e. moving forward from an input layer to the output layer. Figure 1 shows a typical feed forward neural network with one hidden layer and one output layer.

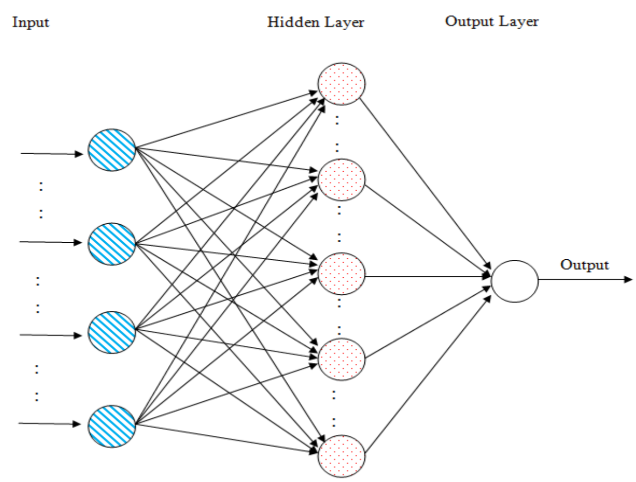

Figure 1: A typical feed forward neural network

Consider an input I which contains several data elements, a simple FNN is designed to generate an output O, satisfying a threshold Θ with the help of weighted summation of inputs and biases. A learning function also called as activation function is used to train the network. The default activation function is sigmoid; however hyperbolic tangent function, Gaussian function and many others can also be used to activate the neural network. The input fed to the FNN is represented by Equation 1.

$$I = \{X_1, X_2, X_3 \ldots X_n\} \tag{1}$$

Where I indicate set of inputs and Xi represents the data samples where i is an integer always i> 0. Hidden layer always uses weighted summation of inputs ($\sum W_i X_i$) and bias (B_i) to carry out learning.

σ, the functionality of the hidden layer is given by the Equation 2.

$$\sigma = \sum_{1}^{J} w_j x_j + b_j \tag{2}$$

The output of the feed forward neural network is given by the Equation (3)

$$Y(\theta)=\sum_{i=1}^{nh} w_i^2 \sigma\left\{\sum_{j=1}^{n} w_{,i,j}^1 x_j + b_{j,i}^1\right\}+b_{j,i} \tag{3}$$

Here y(θ) represent the output following a desired threshold, h indicates the number of neurons in the hidden layer and n indicates the number of input fed to the network. The activation function used is sigmoid and is represented by the Equation 4 and is as shown in Figure 2.

$$Sigmoid\quad(x) = \frac{1}{1+e^{-x}} \tag{4}$$

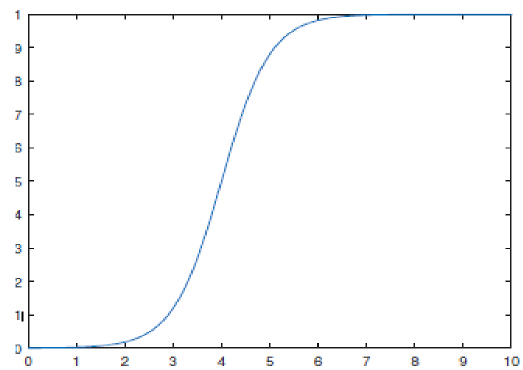

Figure 2: A typical sigmoid function

3.2 Bat algorithm

After Odontocetes, Bats are the mammals which possess a very strong sound propagation technique called Bio−Sonar. Bio−sonar also called as echolocation is used to identify their prey and/or to identify the obstacles while traveling. Echolocation can be defined as phenomenon where the animals use their sound for ranging. The delay in time between the emission of sound from the animals and the echoes reached back after striking the obstacles gives us the actual range of the target from its position. Bats are very clever mammals, they adjust their velocity and change their direction based upon the distance of availability of the prey. This peculiarity of the Bats has inspired the researchers and they came up with the concept of Digital Bats. Xin-She Yang for the first time proposed a novel Bat algorithm which was used to solve many optimization problems [31]. Later, many variations of Bat algorithms were put forth by many researchers. A Binary Bat Algorithm [15], Multi Objective Bat Algorithm (MOBA) [32], Fuzzy logic Bat algorithm [12] etc. Even though there are many variants of Bat algorithm designed to solve various applications the basic working principle remains the same.

It is known that through echolocation the Bat keeps on updating its velocity, position and frequency to catch its prey as soon as possible. The same principle is mimicked by the researchers. A digital Bat is initialized with a fixed population of size *n*, a fixed frequency (F_{min}), initial loudness (A_0), pulse emission rate (*r*) and a wavelength (λ). It is also defined by velocity vector (or velocity matrix) V_i, position vector (or position matrix) X_i and frequency vector (or frequency matrix) F_i which are updated on demand.

Before dynamic updation the authors have some of the predefined assumptions.

1. The pulse rate *r* is inversely proportional to the distance of the Bat from its target (prey) i.e. pulse rate increases as the Bat reaches near to its target.
2. The loudness *A* is directly proportional to the distance of the Bat from its target i.e. as the distance between the Bat and target decreases the loudness also decreases.
3. It is also assumed that the loudness value decreases from a large positive value A_0 to a fixed minimum A_{min}.
4. The initial frequency F_0 is assigned to each Bat randomly which lies in the range of [F_{min} , F_{max}].

The F_{min} and F_{max} are fixed based on the domain size of the given problem. The velocity and position vectors are iteratively updated as follows;

$$V_i(t+1) = V_i(t) + X_i(t) - Gbest)F_i \tag{5}$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{6}$$

$$F_i = F_{\min} + (F\max - F\min)\ \beta \tag{7}$$

In the Equations 7, i is a positive integer ranging from 1 to *n* indicating the i^{th}Bat. t indicates the iteration. *Gbest*is the best solution obtained so far. β is a randomly generated number which is in the range of [0, 1].

From Equations5-6, it can be noticed that difference in frequencies leads the Bats to have different tendency over obtaining the best solutions.

The best solution is given by the formula

$$X_{new} = X_{old} + \varepsilon A^t \tag{8}$$

In Equation 8A^t represents the average loudness of all Bats at the iteration *t* and ε is the random number ranging between [−1, 1].

In order to find the best solution the Bat algorithm explores the dimension space by using updated values of pulse rate and loudness as follows.

$$r_i(t+1) = r_i(0)[1 = \exp(-\lambda t)] \tag{9}$$

$$A_i(t+1) = \alpha A_i(t) \quad (10)$$

In Equations9-10,γ and α are constants and $r_i(0)$ is the final value of r_i. It is to be noted that when $r_i$reaches $r_i(0)$ the value of A_i becomes zero which means the Bat has reached its target.
The pseudo code for Bat algorithm is given in Algorithm 1.

The entire procedure of working of Bat algorithm is represented using flow chart in Figure 3.

Algorithm 1: Bat Algorithm [15]

1. Initialize Bat population: X_i *(i = 1, 2, ..., n)*
2. Define frequency $F_i$and velocity V_i
3. Initialize pulse rates $r_i$and the loudness A_i
4. **while***t*< Maximumiterations **do**
5.   update frequency and velocity
6.   Calculate transfer function values using Equation (4)
7.   Update Vi, Xi, and Fi using Equations 5 to 7
8.   **if** (*rand* >r_i) then
9.   Select the global best solution (*Gbest*) among the available best solutionsand with the available *Gbest* dimensions modify the dimensions of $X_i$randomly.
10.   **end**
11.   Generate new solution randomly Equation (8)
12.   **if** ((rand <A_i) and ($F(X_i) < F(Gbest)$)) then
13.   Accept the new solutions Increase r_i and reduce $A_i$using Equations (9 to 10)
14.   **end**
15.   Find the current *Gbest* and Rank the Bat
16. **end**

In this paper the authors have chosen a strong variant of bat algorithm called binary bat algorithm (BBA) to build a hybrid model for the classification of breast cancer data using simple feed forward neural network.

Even though there are many Meta-heuristic algorithms available, we choose BBA because of the following reasons;

1. Better Convergence: Compared to other algorithms Bat and its variants have better convergence rate.
2. Auto zooming: Auto zooming is the ability of the Bat to reach the region of promising solution at a sooner rate.
3. Auto switching: Zooming is always accompanied by auto switching where the digital Bats switch from explorative moves to local intensive exploitation of search space. The more efficient auto switching, better is the convergence rate.

4. Parameter control: Parameter control is a mechanism where the parameters are not static; instead they keep on varying with iteration count. In Bat algorithm *A* and *r* are the controlled parameters which vary in each iteration. But, usually in other meta-heuristic algorithms the parameters are fixed. Parameter control aids in auto switching.
5. Frequency tuning: Echolocation behavior of Bats is mimicked by frequency tuning. The frequency tuning property can also be found in PSO, Cuckoo Search, Harmony Search etc. Thus frequency tuning can be exploited to provide some functionality similar to the key features of other swarm intelligence based algorithms.

Apart from these advantages it has been proved by preliminary theoretical analysis that Bat algorithms under right conditions assured global convergence [9].

The complete working details of the BBA are provided in the next sub section.

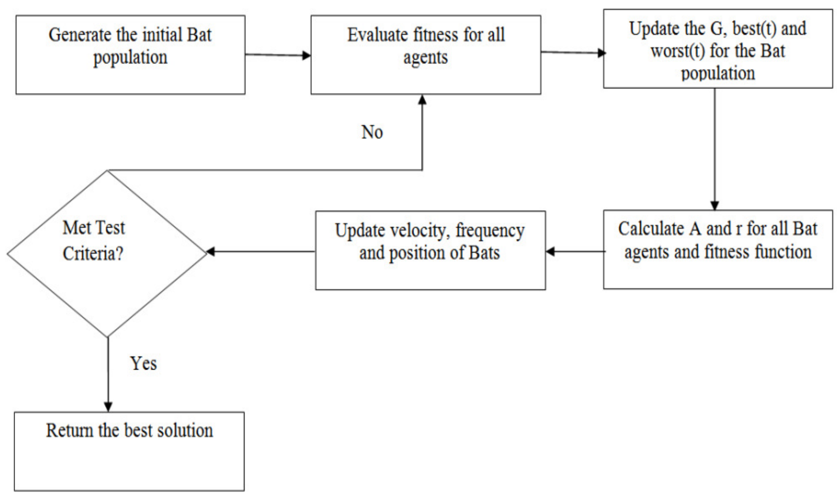

Figure 3: Flow chart representing general procedure of Bat algorithm

3.3 Binary bat algorithm

BBA is conceptually similar to the general Bat algorithm, the difference lies in the search space. General Bat executes in continuous search space, whereas, BBA executes in binary space. Since the binary space is restricted to 0's and 1's the change of velocity and position cannot be performed using Equations5-6 thus, a mechanism to use velocities for changing agent's position is required. In order to update the position of a binary Bat, mapping the velocity values to the probability values is required. This can be done by deploying a transfer function. Care must be taken to select a transfer function that is bound in the interval of [0,1] and the return value of transfer function must be directly proportional to the change in the velocity. Keeping this points in mind the authors have chosen V-shaped transfer function called hyperbolic tangent function which is given by Equation 11 and its mapping from continuous domain to binary domain is given by Figure 5.

A typical V-shaped transfer function looks like Figure 4

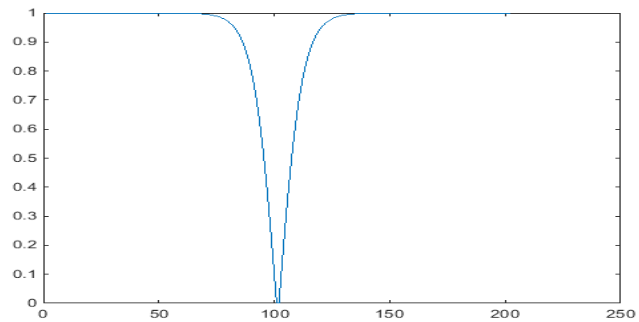

Figure 4: A typical V-shaped transfer function

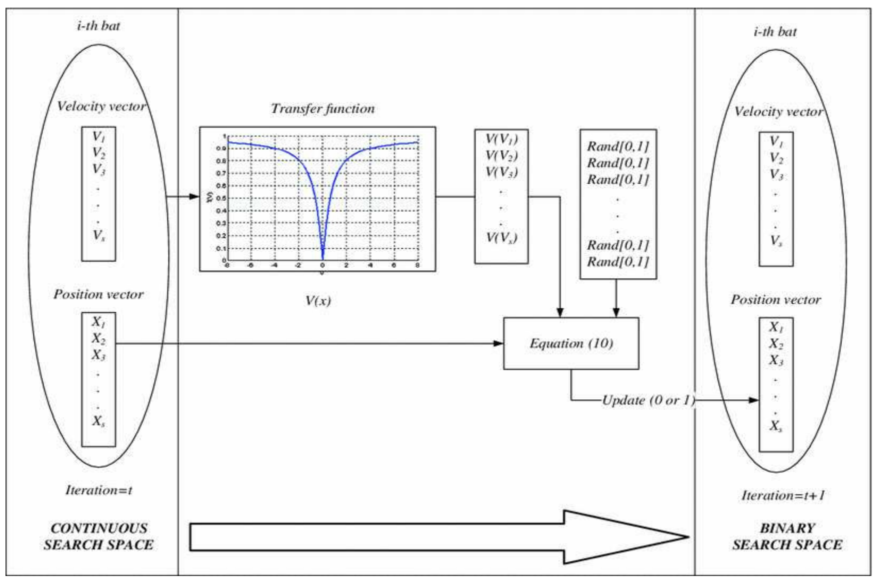

Figure 5: Mapping from continuous domain to binary domain using transfer function[15].

$$V(v_i^k(t)) = \tanh(v_i^k(t)) \tag{11}$$

Using *tanh()* transfer function the probability based change in position of an agent (Binary Bat) is given by Equation 12.

$$\boxed{\begin{array}{l} if\ (rand < V(V_i^k(t+1))) \\ X_i^k(t+1) = X_i^k(t)' \\ else X_i^k(t) \end{array}} \tag{12}$$

Where $V_i^k\ (t)$ and $X_i^k\ (t)$ are the velocity and position of i^{th} agent in k^{th} dimension at the iteration t. Similarly $X_i^k\ (t)'$ is the complement of $X_i^k(t)$.

The algorithm for Binary Bat is given by Algorithm 2.

Algorithm 2: Binary Bat Algorithm [15]

1. Initialize Bat population: X_i $(i = 1, 2, ..., n)$ rand(0 or 1) and $V_i = 0$
2. Define pulse frequency F_i
3. Initialize pulse rates r_i and the loudness A_i
4. **while** t < *Maximum iterations* do
5.   update velocities and adjust frequencies
6.   Using Equation (11) Calculate transfer function value
7.   Using Equation (12) update X_i
8.   **if** (rand $> r_i$) then
9.   Select the global best solution (*Gbest*) among the available best solutions and with the available *Gbest* dimensions modify the dimensions of X_i randomly
10.   **end**
11.   Generate new solution randomly
12.   **if** ((rand $< A_i$)and ($F(X_i) < F(Gbest)$) then
13.   Accept the new solutions Increase r_i and reduce A_i
14.   **end**
15.   Find the current *Gbest* and Rank the Bat
16. **end**

Since the BBA is similar to general Bat algorithm we have continued with the same flowchart and no separate flow chart is given for BBA.

4. PROPOSED WORK

Our problem statement is to classify the given breast cancer data into its two constituent classes (Benign and malignant) using the proposed hybrid model which is a fusion of Bat algorithm and FNN. Here the authors have used Bat algorithm to train the FNN. The authors follow an incremental training approach where the network is trained for a fixed number of iterations and then tested for its performance. The Binary Bat Algorithm is used to minimize the classification error calculated for randomly generated combination of biases and weights. The main aim of the proposed model is to improve the rate of accuracy.

The classification of breast cancer data is carried out in two simple steps:

- Step 1: Representation strategy
- Step 2: Defining fitness function and learning function

4.1 Representation strategy

Binary representation, matrix representation and vector representation are the three widely used methods for representing (encoding)weights and biases in a neural network [36] each having their own set of advantages and disadvantages [35]. The choice of the method depends upon the application. For our binary classification problem the authors have chosen matrix representation method to train feed forward neural network as it is more suitable for training neural network because of its easy decoding phase.

A simple illustrative form of matrix encoding strategy is represented diagrammatically in Figure 6 and correspondingly the dimensions are represented in Equation 13.

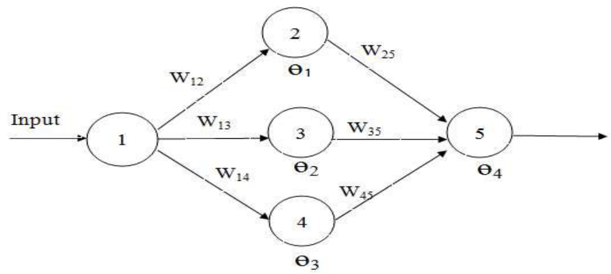

Figure 6: Structure of Feed forward Neural Network

$$i = [W_1, B_1, W_2, B_2] \quad (13)$$

Here W_1 indicates the weight matrix at hidden layer B_1 indicates the bias matrixat hidden layer, W_2 indicates the weight matrix at output layer and B_2 indicates the bias matrix at output layer i.e,

$$W_1 = \begin{bmatrix} w_{12} \\ w_{13} \\ w_{14} \end{bmatrix} \qquad W'_2 = \left[w_{25} w_{35} w_{45} \right]$$

$$B_1 = \left[\theta_1 \theta_2 \theta_3 \right] \text{and} \qquad B_2 = \left[\theta_4 \right]$$

4.2 Defining fitness function and learning function

A learning function is a function used to make the neural network learn. Whereas, the fitness function is a function whose main aim is to minimize the error rate of the output generated from the proposed model as much as possible. So that, obtained result is near to the required result.

For a typical feed forward neural network containing n input nodes h hidden and one output node the learning function is calculated as in Equation 14.

$$f(s_j) = \frac{1}{(1 + \exp(-(\sum_{i=1}^{n} W_{ij} - b_j)))} \quad (14)$$

Where j=1,2,….h

S_j is the output calculated for every hidden node in each iteration. W_{ij} represents the connected weight from the j^{th} node of the hidden layer and i^{th} node of the input layer and b_j is the corresponding bias [35].

The final output is the sum of output derived from all hidden nodes and is given by Equation15.

$$O_k = \sum_{k=1}^{h} W_{kj}.f(S_j) - b_j \tag{15}$$

Where j=1,2,…m

Where W_{kj} represents the connection of weights from the k^{th} output node and j^{th} input node.

$$if(O_k) >= b_j \begin{Bmatrix} class = 1 \\ else \\ class = 0 \end{Bmatrix} \quad 16)$$

$$F_k = \sum_{i=1}^{m} (O_i^k - d_i^k)^2 \tag{17}$$

$$F = \sum_{k=1}^{c} \frac{F_k}{c} \tag{18}$$

Finally the fitness function for the proposed method is given by the equation 18. Here c is the number of training samples used and d_i^k is the desired output of the i^{th} input unit with reference to the k^{th} training sample [13].

5. EXPERIMENTATION AND EVALUATION

The proposed model has been implemented on matlab 2014a platform and tested on a bench mark dataset called Wisconsin Breast Cancer Diagnostic (WBCD) [28]. The Binary Bat Algorithm is used to deploy a fitness function for error minimization and to generate weights and biases required for learning. The feed-forward neural network contains 15 input nodes, 15 hidden nodes and 1 output node. Hyperbolic tangent function is chosen as the learning function to train the network. The network is trained up to 100 iterations to produce the classification rate.

5.1 Data set used

The bench mark data set has been used to check out the performance of the proposed model on the classification of breast cancer data. The breast cancer data set used is Wisconsin Breast Cancer data Diagnostic (WBCD). The dataset is collected from UCI repository [26] and it contains 569 rows and 32 columns (1 class attribute and 31 independent attributes), the same dataset can also be found in [17].

5.2 Parameter setup

A neural network based classification model goes through three phases viz., training, validation and testing. Training phase is a phase where the data is trained so that the network can learn about the patterns which can help in classification.Validation phase is a phase where we check the model with different parameters and come up with the finest set of parameters required for proper classification.Testing is a phase where the unknown data is given to the network to check its ability, how better does it classifythe unknown data based upon the previously learnt knowledge.Since learning plays a vital role in building a neural network based classification model, first and foremostwe check the performance of various learning functions on validating data. The training function which provides highest accuracy will be the chosen function to carry out the entire experimentation.A set of four different V-shaped transfer functions are analyzed by executing them in 10 independent trials, for 100 iterations (which is the usual standard) andtheir corresponding accuracies are tabulated in Table 1.

Table 1 Impact of various V-shaped transfer functions on the proposed model

Function	Formula	Maximum accuracy in %	Maximumtime taken in secs
F1: hyperbolic tangent function	$\lvert\tanh(x)\rvert$	89.95	101.57
F2: erf function	$\left\lvert \mathrm{erf}\left(\sqrt{\frac{\pi}{2x}}\right)\right\rvert$	85.23	744.30
F3: arctan	$\left\lvert \frac{2}{\pi} * \mathrm{atan}\left(\frac{\pi}{2x}\right)\right\rvert$	83.83	114.99
F4: inverse of square root of x2	$\left\lvert \frac{x}{1+\sqrt{x^2}}\right\rvert$	73.28	359.36

Figure 1 to 4 shows the MSE transfer curve of the four different transfer functions with maximum accuracy as given in Table 1.

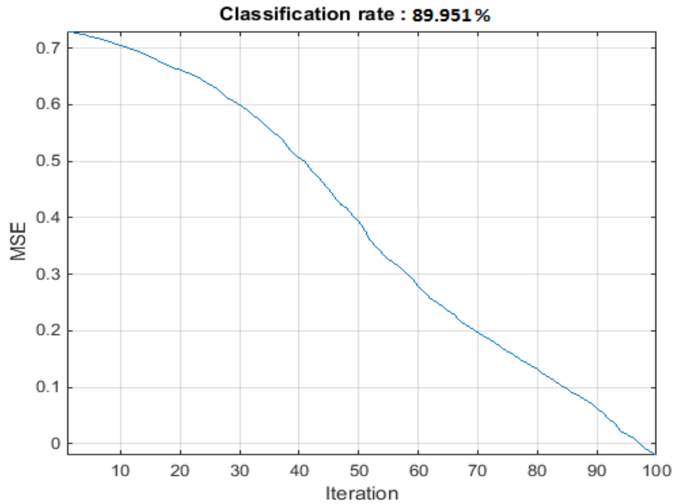

Figure 1: MSE transfer curve for F1 V-shaped transfer functions.

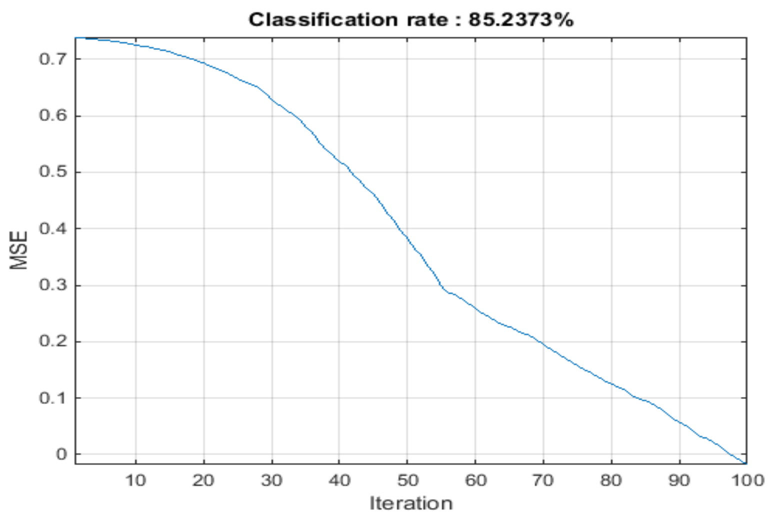

Figure 2: MSE transfer curve for F2 V-shaped transfer functions.

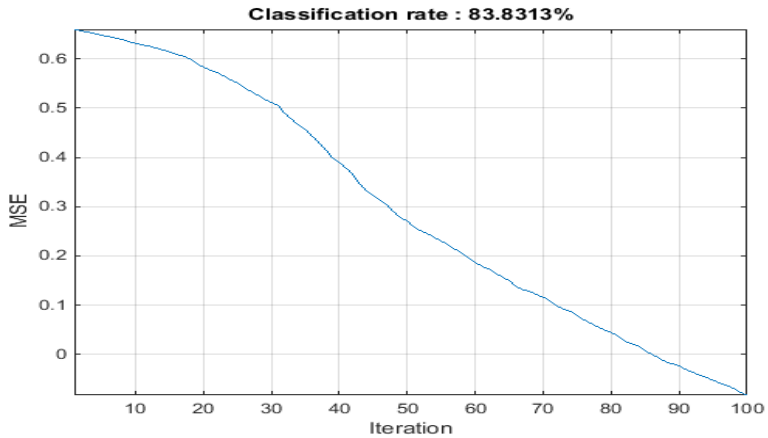

Figure 3:MSE transfer curve for F3 V-shaped transfer functions.

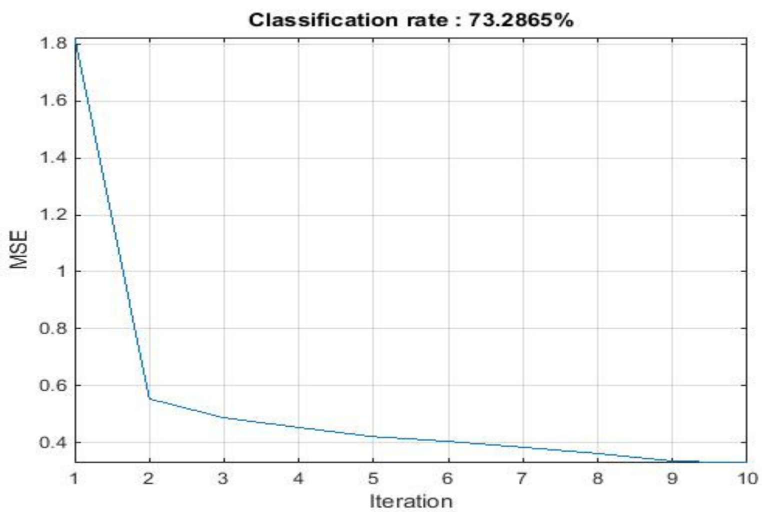

Figure 4: MSE transfer curve for F4 V-shaped transfer functions.

Figure 1 to 4 and Table 1 clearly suggest that *tanh()* function performs better than its counter parts.

The V-shaped transfer functions are much better in updating the position than the S-shaped functions because, in V-shaped transfer functions the search agents are assigned the values either 0 or 1.

V-shaped transfer functions tend more often to form the complement of variables using Equation (12). This mechanism promotes and guarantees changing the position of search agents proportional to their velocities. This is the main reason for the superiority of v-shaped test functions.

However, it is also important to analyze the effect of change in number of iterations in a neural network. Thus we check the performance of tanh() on different iterations.

The Table 2 provides the details about the change in accuracy of the model with change in number of iterations. Here the minimum, maximum and mean accuracy of the tanh() for different iterations is given.

Table 2: Impact of number of iterations on the proposed model.

Sl. No	No .of Iterations	Maximum accuracy in %	Mean accuracy in %	Minimum accuracy in %
1.	10	48.33	37.59	40.97
2.	30	85.41	83.04	65.71
3.	50	92.64	87.88	84.2
4.	75	78.73	68.10	50
5.	100	89.95	76.71	67.3
6.	150	72.23	69.66	65.47
7.	200	74.34	72.2	68.26

After deciding to go with which transfer function and fixing up the iterations, we move on to check the impact of other parameters.The hidden nodes play a major role in deciding the complexity of the ANN structure. More the number of hidden nodes more complex will be the structure and lesser the number of hidden nodes, simple will be the structure. Our concern is to keep the structure as simple as possible and at the same time not compromising with the accuracy.Thus, knowing about an optimal number of hidden nodes that can serve our purpose is must.

As the maximum number of inputs taken in our model is 15, we restrict the maximum number of hidden nodes to 15.Since there is a thumb rule that the hidden nodes can't exceed the number of input nodes[8].

Table 3 provides the impact of various number of hidden nodes on the accuracy of the model. Table 3 clearly specify that 15 number of hidden nodes are more suitable for carrying out the classification task.

Table 3: Impact of number of hidden nodes on the proposed model.

Sl. No	No .of Hidden nodes	Maximum accuracy in %	Mean accuracy in %	Minimum accuracy in %
1.	3	72.4	61.86	50.48
2.	5	75.21	66.06	55.70
3.	10	83.83	68.96	54.30
4.	15	92.61	88.7	84.40

In nature inspired algorithms the role of Number of Particles (NoP) chosen and how many times these particles aretrained has a major impact on deciding the accuracy of the model. Table 4 provides the information regarding the change in performance of the model withchange in NoP values.

Table4: Impact of number of particles (NoP) on the proposed model

Sl.No	NoP count	Maximum accuracy in %	Mean accuracy in %	Minimum accuracy in %
1.	10	56.23	41.62	24.90
2.	20	64.85	54.70	33.74
3.	30	92.61	88.70	84.4
4.	40	80.31	62.38	47.75
5.	50	82.77	60.93	47.95
6.	60	68.80	56.01	37.00

From Table 4 it is clear that NoP at 30 produces good classification results.Similar to NoP, the NoV parameter was checked with different values and its details are tabulated in Table 5.

Table5: Impact of number of dimensions (NoV) on the proposed model

Sl.No	NoV count	Maximum accuracy in %	Mean accuracy in %	Minimum accuracy in %
1.	25	75.75	60.68	45.51
2.	50	80.47	61.09	33.74
3.	75	81.63	62.29	36.55
4.	100	82.95	57.29	40.77
5.	125	86.46	59.76	42.16
6.	150	92.61	88.7	84.4
7.	175	87.22	65.35	75.21
8.	200	77.85	62.05	49.05

Finally the pulse rate (r) and loudness (A) are confirmed by checking different values and the corresponding accuracies are tabulated in Table 6 loudness values are checked from 0.99 to 0.1pulse rate is checked from 0.2 to 1.

Table6: Impact of A and r on the proposed model

Sl. No	A and r values	Maximum accuracy in %	Mean accuracy in %	Minimum accuracy in %
1.	0.9 and 0.2	84.60	69.18	46
2.	0.7 and 0.4	84.69	70.51	54.1
3.	0.5 and 0.6	92.61	88.70	84.4
4.	0.3 and 0.8	81.01	67.50	60
5.	0.1 and1.0	65.90	46.50	34.6

The A=0.5 and r=0.6 produces good classification results when compared to other values.Apart from these, the two constants Q_{min} and Q_{max} are assigned values of 1 and 5 respectively. These values are fixed by referring the literature [13-14].

The final set of parameters for BBA algorithm confirmed after an intense preliminary study, careful examination and repeated experimentation are tabulated in Table 7.

Table7:Finalizedset of parameters for BBA

Parameter	Value
Maximum Iterations	50
Number of Particles (NoP)	30
Number of Dimensions (NoV)	150
Loudness (A)	0.5
Pulse Rate (r)	0.6
Minimum frequency (Qmin)	1
Maximum frequency (Qmax)	5
Initial Frequency (for each particle)	0
Initial Velocity (for each particle)	0
Initial Position (for each particle)	0

5.3 Results

In order to get the better results the algorithm was executed 10 times for the given dataset under 10 fold cross validation scheme and the highest accuracy was selected as the best accuracy of the model.The results include Confusion matrices for training and testing phase. Receiver Operating Characteristic (ROC) curve for the testing data and MSE performance plot for testing data.

From Table 8 and 9, the confusion matrices for the WBCD datasets for training and testing phase is given which provide us the information regarding the number of true positives (TP), true negatives (TN) false positives (FP) and false negatives (FN) obtained for the WBCD dataset.

Table8: Confusion matrix for WBCD dataset used for training

WBCD Dataset	Malignant	Benign
Malignant	200	33
Benign	9	327

Table 9: Confusion matrix for WBCD dataset used for testing

WBCD Dataset	Malignant	Benign
Malignant	195	40
Benign	17	317

The Receiver Operating Characteristic (ROC) curve obtained for the WBCD dataset during testing is as shown in Figure 5.It provides the details of the area covered by the proper classification.

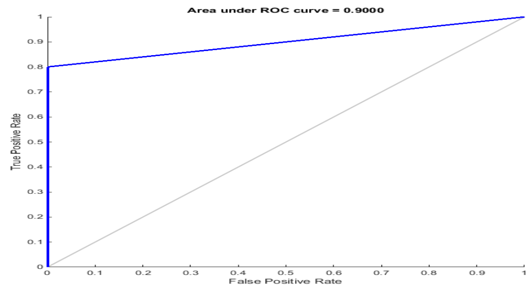

Figure 5: ROC curve for WBCD dataset.

The Area Under Curve (AUC)covered by WBCD dataset is 0.9000 i.e. 90% of the total area. The performance plot provides the information regarding the minimum error of the model.The performance plot for WBCD is as shown in Figure 6.The MSE of the proposed model is found to be 0.34.

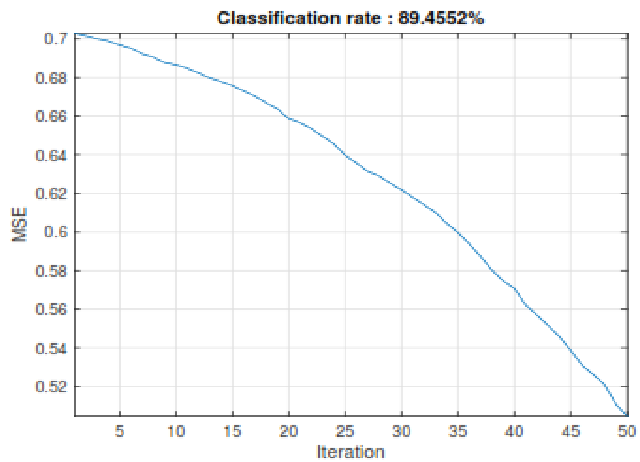

Figure 6: MSE curve for WBCD dataset.

The evaluation measures such as precision, recall, accuracy, Matthews coefficient etc are calculated using true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives of the testing confusion matrix is given in Table 10.

Table 10: Various measures deduced from Confusion matrix for WBCD data set.

Measure	Formula	Value
Sensitivity (TPR)/Recall	TPR = TP / (TP + FN)	0.9198
Specificity (TNR)	SPC = TN / (FP + TN)	0.8880
Precision (PPV)	PPV = TP / (TP + FP)	0.8298
Negative Predictive Value (NPV)	NPV = TN / (TN + FN)	0.9491
Accuracy	ACC = (TP + TN) / (P + N)	0.8991
Matthews Correlation Coefficient	F1 = 2TP / (2TP + FP + FN)	0.7932

5.4 Comparative analysis

We compared our results with nearly 9 other classification techniques which included both nature inspired algorithms and regular techniques. The accuracy values (in %) of various classification techniques and that of proposed model are mentioned in Table 7.

Table 11: Comparing accuracy of various classification techniques

Algorithm/Techniques	**Accuracy in %**
ACO [4]	47.45
BBO [14]	91.1
KNN [29]	80.03
MNN [10]	92.1
NaiveBayes [29]	91.63
PSO [11]	91.16
Random Forest [24]	89.12
ES [19]	91.81
Proposed Method	**92.61**

The classification accuracy obtained for WBCD dataset is 92.61% for training and 89.91% for testing which is higher than other techniques however, MNN give a tough competition to the proposed model. Since the advantages of BBA are superior to PSO in many aspects,we claim that the proposed model is better than all other compared techniques.

1. CONCLUSION AND FUTURE WORK

From the available results and comparative analysis we can strongly conclude that the proposed model - BBA inspired Feed-forward neural network performs very well in classifying the data into benign and malignant classes and giving us the maximum accuracy of 92.61% for training WBCD data and 89.951% accuracy for testing. Even though the accuracy of PSO and GSA is higher than BBA, the time taken and MSE obtained is less. Thus the time efficiency and error minimization makes the proposed model more suitable than other algorithms in solving binary classification problem.

But it is to be noted that, both the dataset and the structure of the FNN is kept simple. In future we are interested in carrying out the classification task on huge dataset and complex network structure.

ACKNOWLEDGEMENT

Wisconsin Hospital and its team: A Wisconsin Original Breast Cancer dataset was obtained from theUniversity of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. A hearty thanks to the institutesand personnel, my teachers and my fellow researchers for their timely support.

Funding: Maulana Azad National Fellowship for Minority Students, (Grant/Award Number: ‘F1-17/2013-14/MANF-2013-14-MUS-KAR-24350’).

REFERENCES

[1] American Society of Breast Cancer Research.Cancer facts and figures, http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2009/index, accessed on 10.04.2015.
[2] Breast Cancer India. Trends of breast cancer in India, http://www.breastcancerindia.net/statistics/trends.html, accessed on 10.04.2015.
[3] Clark, P. & Niblett, T. (1989) "Thecn2induction algorithm", Machine learning, 3(4): pp261–283.
[4] Dorigo, M., Maniezzo, V., Colorni, A., & Maniezzo, V. (1991). "Positive feedback as a search strategy", Citeseer.
[5] Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., & Haussler, D. (2000) "Support vector machine classification and validation of cancer tissue samples using microarray expression data", Bioinformatics, 16(10), pp 906-914.
[6] Gandomi, A. H., Yang, X.-S., Alavi, A. H., & Talatahari, S. (2013) "Bat algorithm for constrained optimization tasks", Neural Computing and Applications, 22(6), pp1239–1255.
[7] Holland, J. H. (1992). Genetic algorithms. Scientific american, 267(1):66–72. Huang, G.-Q., Zhao, W.-J., & Lu, Q.-Q. (2013) "Bat algorithm with globalconvergence for solving large-scale optimization problem", Application Research of Computers, 30(3), pp1–10.
[8] How to select number of hidden neurons//http://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netwAccessed on 27.04.2016
[9] Huang, G.-Q., Zhao, W.J., & Lu, Q.-Q. (2013) "Bat algorithm with globalconvergence for solving large-scale optimization problem", Application Research of Computers, 30(3), pp1-10
[10] Javadi, M., Arani, S. A. A. A., Sajedin, A., & Ebrahimpour, R. (2013) "Classification of ECG arrhythmia by a modular neural network based on mixture of experts and negatively correlated learning". Biomedical Signal Processing and Control, 8(3), pp289-296.
[11] Kenndy, J. & Eberhart, R. (1995) "Particle swarm optimization", In Proceedings of IEEE International Conference on Neural Networks, volume 4, pp1942–1948.
[12] Khan, K., Nikov, A., & Sahai, A. (2011) "A fuzzy bat clustering method for ergonomic screening of office workplaces", In Third International Conference on Software, Services and Semantic Technologies S3T 2011, pp 59–66.
[13] Mirjalili, S., Hashim, S. Z. M., & Sardroudi, H. M. (2012) "Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm", Applied Mathematics and Computation, 218(22), pp11125–11137.
[14] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014) "Let a biogeography-based optimizer train your multi-layer perceptron", Information Sciences, 269,pp188–209.
[15] Mirjalili, S., Mirjalili, S. M., & Yang, X.-S. (2014)"Binary bat algorithm", Neural Computing and Applications, 25(3-4), pp663–681.
[16] Mishra, S., Shaw, K., & Mishra, D. (2012) "A new meta-heuristic bat inspired classification approach for microarray data", Procedia Technology, 4, pp802–806.
[17] mldata (2014) "WBCD benchmark data from mldata repository", http://mldata.org/repository/search/?searchterm=wisconsin&data=Data. Accessed online on 2014-05-10.
[18] Nakamura, R. Y., Pereira, L. A., Costa, K., Rodrigues, D., Papa, J. P., & Yang, X. (2012) "BBA: A binary bat algorithm for feature selection", 25th IEEE Conference on Graphics, Patterns and Images (SIBGRAPI), pp291–297.
[19] Neg, S. C., Chung, C., Leung, S. H., & Luk, A. (1994) "An evolutionary search algorithm for adaptive iir equalizer", IEEE International Symposium on Circuits and Systems, volume 2, pp53-56.
[20] Quinlan, J. R. (1990) "Decision trees and decision-making", Systems, Man and Cybernetics, IEEE Transactions on, 20(2), pp339-346.
[21] Rashedi, E., Nezamabadi-Pour, H., and Saryazdi, S. (2009). Gsa: a gravitational search algorithm. Information sciences, 179(13):2232–2248.
[22] Rodrigues, D., Pereira, L. A., Nakamura, R. Y., Costa, K. A., Yang, X.-S., Souza, A. N., & Papa, J. P. (2014) "A wrapper approach for feature selection based on bat algorithm and optimum-path forest", Expert Systems with Applications,41(5), pp2250–2258.
[23] Siegel, R., Naishadham, D., & Jemal, A. (2013)"Cancer statistics 2013", CA: a cancer journal for clinicians, 63(1), pp11-30.
[24] Strobl, C., Malley, J., & Tutz, G. (2009) "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests", Psychological methods, 14(4), pp323.
[25] Tang, R., Fong, S., Yang, X.-S., & Deb, S. (2012) "Integrating nature-inspired optimization algorithms to k-means clustering", International Conference on Digital Information Management (ICDIM), pp116–123.

[26] UCI (2014) “Wisconsin original benchmark data from uci repository”, https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original).Accessed online on 10-05-2014.

[27] Wang, G. & Guo, L. (2013) “A novel hybrid bat algorithm with harmony search for global numerical optimization”, Journal of Applied Mathematics, 2013.

[28] Wisconsin Breast Cancer Diagnostic (2015) https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic). Accessed online on10-02-2015.

[29] Wu, X., Kumar V., Quinlan J.R., Motoda H., & Zhou Z.H. (2008) “Top 10 algorithms in data mining”, Knowledge and information systems, 14(1), pp1-37.

[30] Yang, X.-S. (2010)“Nature-inspired metaheuristic algorithms”, Luniver press.

[31] Yang, X.-S. (2010).“A new metaheuristic bat-inspired algorithm”, In Nature-inspired cooperative strategies for optimization (NICSO 2010), pp65–74.

[32] Yang, X.-S. (2011) “Bat algorithm for multi-objective optimization”, International Journal of Bio-Inspired Computation, 3(5), pp267–274.

[33] Yang, X.-S. & He, X. (2013) “Bat algorithm: literature review and applications”, International Journal of Bio-Inspired Computation, 5(3), pp141–149.

[34] Yang, X.-S. & Hossein Gandomi, A. (2012) “Bat algorithm: a novel approach for global engineering optimization”, Engineering Computations, 29(5), pp464–483.

[35] Zhang, J.-R., Zhang, J., Lok, T.-M., & Lyu, M. R. (2007) “A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training”, Applied Mathematics and Computation, 185(2), pp1026–1037.

[36] Zhang, N. (2009) “An online gradient method with momentum for two-layer feed- forward neural networks”, Applied Mathematics and Computation, 212(2), pp488–498.

AUTHORS

Doreswamy received B.Sc degree in Computer Science and M.Sc Degree in Computer Science from University of Mysore in 1993 and 1995 respectively. Ph.D degree in Computer Science from Mangalore University in the year 2007. After completion of his Post-Graduation Degree, he subsequently joined and served as Lecturer in Computer Science at St. Joseph’s College, Bangalore from 1996-1999.Then he has elevated to the position Reader in Computer Science at Mangalore University in year 2003. He was the Chairman of the Department of Post-Graduate Studies and research in computer science from 2003-2005 and from 2009-2008 and served at varies capacities in Mangalore University at present he is the Chairman of Board of Studies and Professor in Computer Science of Mangalore University. His areas of Research interests include Data Mining and Knowledge Discovery, Artificial Intelligence and Expert Systems, Bioinformatics, Molecular modelling and simulation, Computational Intelligence, Nanotechnology, Image Processing and Pattern recognition. He has been granted a Major Research project entitled “Scientific Knowledge Discovery Systems (SKDS) for Advanced Engineering Materials Design Applications” from the funding agency University Grant Commission, New Delhi, India. He has been published about 30 contributed peer reviewed Papers at national/International Journal and Conferences. He received SHIKSHA RATTAN PURASKAR for his outstanding achievements in the year 2009 and RASTRIYA VIDYA SARASWATHI AWARD for outstanding achievement in chosen field of activity in the year 2010.This space is for writing your short resume.

Umme Salma M received BSc and MSc (Computer Science) degree from Kuvempu University. She has secured 1st rank in MSc.Cs in 2009 and is a gold medalist. She is an awardee of Maulan Azad National Fellowship. Currently pursuing her PhD in Mangalore university and her research topic is Exploration of advanced datamining techniques for the classification of breast cancer data.