IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

Classification of Cancer Datasets using Artificial Bee Colony and Deep Feed Forward Neural Networks

M. Karunyalakshmi[1], Dr. N. Tajunisha[2]

Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, India[1]

Associate Professor, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, India[2]

Abstract: The improvement of microarray technology has facilitated scientists in the direction ofmonitor mRNA expression level of up toward tens of thousands of genes concurrently, inside just one single experiment on a solid surface, made of any glass or silicon. The huge amounts of dataset samples are then exploitedthrough biologists in the direction ofclassify cancer-related genes by means of comparing their expression values in healthy and cancerous tissues. Though, microarray dataset samples generally have thousands of genes or features and simply dozens of samples. Handling of huge amounts of dataset samples are most important challenging task in microarray dataset samples classification .So feature selection or reduction is akey important step in classification of microarray dataset samples with many thousands of features. As a wrapper method, Support Vector Machine-Recursive Feature Elimination (SVM-RFE) is one of the most important feature selection techniques. Though SVM-RFE is able toeliminate irrelevant features successfully, it mightn't deal with the majority of the irrelevant features. In addition SVM-RFE also performs better than other methods by considering threshold valueθ. They mightbe several numbers of ways for choosing of this parameter thusresults higher performance or improved performance. But sometimes they decreases results of the classifier, to solve this problem Artificial Bee Colony (ABC) is introduced in this paper for automatic selection of threshold θ toward enhance classification results in terms of accuracy, precision and recall.This ABC algorithm is performed based on the behavior of colony of three bees: employee, onlooker and scout bees. Then, Deep Feed Forward Neural Network- Recursive Feature Elimination (DFFNN-RFE) is introduced for classification of microarray dataset samples. In DFFNN classifier is the performed based on the calculation of probability distributions that stochastically adjust small parts of the input data for the period of training phase. In addition the proposed work correlation coefficient is introduced in order to filteringof highly redundant features before performing DFFNN-RFE classifier. The experimentation results conclude that the proposed DFFNN-RFE classifier performs better when compared to a nonlinear SVM-RFE, classifier in terms of classification accuracy, precision and recall to cancer dataset.

Keywords: microarray technology, Support Vector Machine(SVM), Recursive Feature Elimination (RFE), Deep Feed Forward Neural Network(DFFNN), Artificial Bee Colony (ABC), feature selection, classification.

1. INTRODUCTION

Breast cancer is a general cancer by means of high death rate in the world wide. Patients in their early on stage frequently have no identifiable symptom and are generallydetectedby means ofcomplexstage cancer. The medium of their continued existence is less than one year. Each and every one these facts highlight the significance of gaining deeper approachingaddicted to gene level of this specific malignant tumor. The development of microarray technology has enabled scientists to screen mRNA expression level of up to tens of thousands of genes simultaneously, within just one single experiment on a solid surface, made of either glass or silicon. The subsequent large amounts of data are then utilized by biologists to identify cancer-related genes by comparing their expression values in healthy and cancerous tissues. However, microarray data usually have thousands of features and only dozens of samples.

The improvement of microarray technology has facilitated scientists in the direction of monitor mRNA expression level of up toward tens of thousands of genes concurrently, inside just one single experiment on a solid surface, made of any glass or silicon. The huge amounts of dataset samples are then exploited through biologists in the direction of classify cancer-related genes by means of comparing their expression values in healthy and cancerous tissues. Though, microarray dataset samples generally have thousands of genes or features and simply dozens of samples. Handling of huge amounts of dataset samples are most important challenging task [1] in

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

microarray dataset samples classification, calls for well-organized feature selection techniques.

In the stage of classification, feature selection techniques are able to be classified into three major types, relying on how they come together the feature selection search by means of the creation of the classifier: filter, wrapper and embedded methods.

In the filter type, feature reduction is carryout without taking considering the performance of the classification algorithm with the purpose ofdeterminationis applied in the direction of the chosen features. Thereforea filter algorithm, usually depending on the relevance measure with the purpose of calculates the significance of each feature designed for the classification model. Some of the Filter methods are Fisher criterion [2], T-statistics [3] and mutual information [4], runningcorrespondingly as the filter approximation the condition of systems [5], choose features simply according in the direction of some statistic which calculates each feature's relevancy by means of the class labels. Filter methods rank each and every one features in terms of their good qualityby means of the relative of each particular gene by means of the class label depending on their univariate scoring metric. The top ranked genes are selected before performing the classification models. The feature score has been computed separately for each feature based on the correlation coefficients among the class and each feature. The major disadvantage of filter method is towardattaining each feature separately at the same time as ignoring the relations among the features.

On the other hand, the wrapper method chooses a subset of features with the purpose is optimized by means of a specified SVM classifier [6]. The SVM classifier is performed based on the black box is run several times on diverse candidate subsets, and whenever, the feature of the candidate subset is calculatedby means of the performance of the classifiermodel trained on this subset. The optimal subset of genes is well-knowndepending on the ranking of classification results obtained from positioning the classifier on each and every one found subsets. For doing this purpose, Genetic Algorithms (GAs) hasbeen proposed and introduced in the recent studies [7-8].Compared with filter methods, the wrapper methods majorly relies on classifier models that construct feature subset with the purpose ofincreasing the accuracy of algorithms .But it becomes more time computation complexity. But both of these feature selection methods provides higher performance than feature extraction [9] methods and Principal Components Analysis (PCA) [10].

Embedded methods have the advantage with the purpose of they consist of the relationsby means of the classification model, at the same time asreducing less time computation complexity than wrapper methods. Support Vector Machine (SVM) is a machine learning classifierwith the purpose of has been effectivelyfunctional in a variety of fields [11]–[12] because of itsgood simplificationability. Depending on the SVM, a backward elimination processis called as Support Vector Machine- Recursive Feature Elimination (SVM-RFE) [13] is introduced as an effective wrapper method. It starts by means ofeach and every one the feature which removes one feature at a time, the removal of which has the smallest amountcontact on the value of the objective function of SVM, in anticipation of the final feature subset is achieved. SVM-RFE showed its higher classifier results than the other feature selection algorithms [14]-[16]. Some of the work done related to SVM-RFE by merging it with other techniques as new trends [17] with the purpose is able, now and then better classification results. Furlanello et al [18] introduce a new Entropy-based Recursive Feature Elimination (E-RFE) algorithm with the purpose ofremoves chunks of insignificantfeaturesdesigned for each step related to the entropy measure of the SVM weight allocation. Depending on mutual information, [19] SVM-RFE is integrated to additional term with the purpose of make best use of the relevancy with the feature subset and the target class at the same time asreducing timeconsumption of the redundancy amongfeatures. Duan et al [20] develops a new SVM-RFE multiple classifierswith the purpose ofintegrating the ranking functions of multiple SVM classifiers trained on the subsamples formed from k-fold cross validation [21] or bootstrapping [22]. SVM-RFE is proficientin the direction ofremove irrelevant features successfully. On the other hand, the final feature subset generally consist redundant features with the purpose ofmightn't beeliminated in early iterations [23].

In the literature many of the feature selection methods has been introduced in order to remove and select most significant genes designed for microarray data. Those feature reduction methodseliminatethe majority of the unimportant features; it mightn't remove redundant features successfully, thus decreases the accuracy of the classifier. In the recent work some of the classifiers accuracy of the classifier is selected based on the predefined threshold 'θ', howeverthose classifiers reduces performance accuracy as well as increases the execution time of the classifier.They might be several numbers of ways for choosing of this parameter thus results higher performance or improved performance. But sometimes they decreases results of the classifier, to solve this problem Artificial Bee Colony (ABC) is use in this paper for automatic selection of threshold θ toward enhance classification results in terms of accuracy, precision and recall.

2. LITERATURE REVIEW

Hu et al[24] introduces newnonparametricclassifiers such as Pearson correlation coefficient and Significant Analysis of Microarray (SAM) [25]has been describedin the direction based on the top feature selection techniques. But univariate gene rankings have some major drawback. The majority selected features from the feature selection methods becomes redundant genes. So the accuracy of the classifiers is decreases and requires high execution time.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

Yu and Liu [26] introduce a new Redundancy Based Filter (RBF) schema in the direction of solving redundant feature selection problems and the performance results are relativelycapable. The wrapper method combines the procedure of feature selection with classifier results, the chosen features are evaluated based on the classifier's performance. Zhang et al [27] introduces a newwrapper based feature selection algorithm for the selection of genes .This feature selection methods assessing the efficiency of each subset of genes by means of the evaluation of the accuracy percentage of the particular classifier toward be used, training the classifiersimplyby means of the selected genes. But the computational cost of wrapper approach is higher than that of filter based feature selection algorithm.
Liu et al [28] introduces a Genetic Algorithm (GA) based wrapper feature selection approach for microarray dataset samples. In this work the GA is performed based on the mimicking optimization model and natural genetics. On the contrary, wrapper methods integrate the relationsamong genes selection and classifier, which formulate them unique when compared to filter based feature selection techniques.

Ferreira and Figueiredo[29] introduces and develops a newmany Fast Correlation based filter (FCBF) method, follows a relevance-redundancy approach with two major phases are compute the Symmetrical uncertainty evaluation of each gene feature and perform sorting for those gene features, from this remove redundant attributes.Peng et al [30] develops a new 'minimum Redundancy Maximum Relevancy' (mRMR) criterion in which redundancy is calculated by considering Mutual Information (MI) "among pairs of features, while relevance is determined by means of the MI among each gene feature and the class label". This mRMR criterion is well-organized than other criterionfunction because redundancy calculation is restrictedin the direction of a few pairs of the majority relevant features.
Saeys et al [31] introduces a new SVM based feature selection model in order to reduce the dimensionality of feature space with the use of nonlinear kernel function. Here SVM classifier is integrated to Recursive Feature Elimination (SVM-RFE) feature selection in order to reduce computation time of filter methods. It is concluded that this method is more suitable for the classification of microarray dataset samples.Giannakeas et al[32]introduces a new machine learning based SVMs classifier for microarray dataset classification which achieves higher accurateness and robustness. The results of the classifier designed for a known input is the most accepted result betweeneach and every one the random trees.

Statnikov et al[33] introducing a new fundamentalsystem of statistically important sub networks is createdby means of searching a variety of sub networks and handing over scores depending on each sub network's gene expression level. According in the direction of the authors, this fundamental system enhances the stability of the SVM classifier, and moreover permitsdesigned for the extraction of advanced-level biological data which is presented in further databases and formats.
Zhu et al [34] evaluate the accuracy results of SVM classifier with linear regression and Neural Network (NN) classifiers to colon tumor data sets after the completion of feature selection process. Tem and fifty features were chosenby the calculation of t-statistic feature selection and obtain almost 85% of classification accuracy on SVM - RBF kernel. Alladi et al [35] a introduces a Recursive Feature Elimination (RFE) new feature selection method and experiments is performed to colon tumor and leukemia gene expression dataset. Four important genes features were selected using RFE that results 98% of classification accuracy. Symonsand Nieselt [36] develops a new Leave-one-out cross-validation is designed forvaluation and altogether around 220,000 various combination of classifiers and feature selections. From the results it concludes that the linear SVM becomes best classifier than other classification models in the diagnosis of disease based on a variety of test results.

3. PROPOSED METHODOLOGY

Feature selection or reduction is a key important step in classification of microarray dataset samples with many thousands of features. In the literature many of the feature selection methods has been introduced in order to remove and select most significant genes designed for microarray data. As a wrapper method, Support Vector Machine-Recursive Feature Elimination (SVM-RFE) is one of the most important feature selection techniques. Though SVM-RFE is able to eliminate irrelevant features successfully, it mightn't deal with the majority of the redundant features.

In this work has been experimented and evaluated using the cancer dataset samples. Designed for following investigations in favor of this research paper, the learning used well-known datasets which is publicly available. This research work use a new feature selection method, which removes redundant features successfully depending on the correlation between features .Then Deep Feed Forward Neural Network Recursive feature elimination (DFFNN-RFE) classification model is proposed for classification. Artificial Bee Colony (ABC) is use in this paper for automatic selection of threshold θ toward enhances classification results in terms of accuracy, precision and recall. DFFNN-RFE provides higher classification performance for both lung and breast cancer dataset samples. In order to increase the accuracy of DFFNN-RFE classifier first redundant features in the dataset are removed. Introduce the correlation coefficient in the direction ofdetermining the similarity among features. The correlation coefficient among features f_i and f_j is determined as follows,

$$r_{ij} = \frac{\sum_{k=1}^{n}(f_{ik} - \bar{f}_i)(f_{ik} - \bar{f}_j)}{\sqrt{\sum_{k=1}^{n}(f_{ik} - \bar{f}_i)^2}\sqrt{\sum_{k=1}^{n}(f_{ik} - \bar{f}_j)^2}}$$

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

Where $f_{ik}$is denoted as the feature value $f_i$according to the k^{th}sample, and f_i is denoted as the average value of feature f_i of each and every one the n samples. The value of r_{ij}iscomputed insidethe range $[-1, 1]$ and the higher correlation $|r_{ij}|$, denotes those features f_i and $f_j$are most relevant to each other.

STEP 1: Initialize the predefined threshold a θ, the number of the current feature as $i = 1$, and $j = 2$.

STEP 2: Computer_{ij}, if$|r_{ij}| > \theta$, then remove feature *fi* and go to step 4.

STEP 3: Update as $j = j + 1$ and go to step 2 until $j = l$ (l is the number of the original features).

STEP 4: Update as $i = i + 1, j = i + 1$ and go to step 2 until $i = l - 1$.

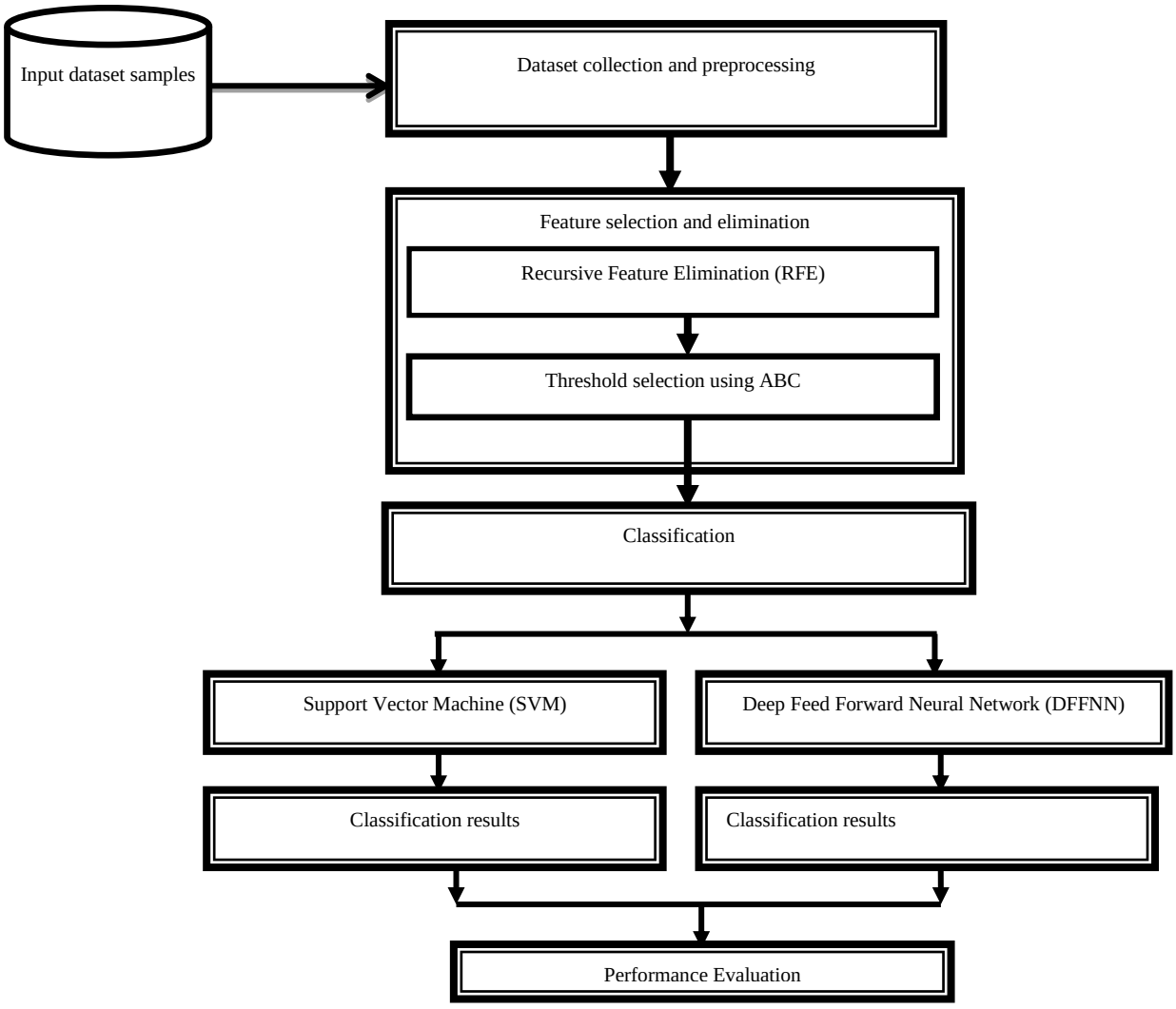

Fig1. Overall Process Flow Diagram

Overall Process of Algorithm

STEP 1: Select the cancer dataset.

STEP 2: Recursive Feature Elimination (RFE) using correlation coefficient

- Initialize the predefined threshold as θ, the number of the current feature as $i = 1$, and $j = 2$.
- Computerr_{ij}, if$|r_{ij}| > \theta$, then remove feature *fi* and go to step 4.
- Update as $j = j + 1$ and go to step 2 until $j = l$ (l is the number of the original features).
- Update as $i = i + 1, j = i + 1$ and go to step 2 until$i = l - 1$.

STEP3: Automatic selection of threshold θ using Artificial Bee Colony algorithm

STEP 4: Classification algorithm

- Initialize the original feature set as *F*.
- Define the parameters (θ) consisting of W, b_v, and b_h, the parameters are

$$W = \begin{pmatrix} w_{0,0} & w_{0,1} \\ w_{1,0} & w_{1,1} \\ w_{2,0} & w_{2,1} \end{pmatrix}, b_h = (b_{h0}, b_{h1}), h_v = (b_{v0}, b_{v1}, b_{v2})$$

- Train a linear DFFNN with feature set *F*.
- For any arbitrary hidden node h_k
- Compute auto encoder is a combination of a weight matrix W, a hidden bias vector b_h along with a visible vias

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

vector bv, and a saturating nonlinearity function weight vector $\boldsymbol{\omega} = \sum_i y_i x_i \alpha_i$

➢ Probabilistically modified input data point z_i given input x, with corruption level of the input.

➢ Compute cross entropy

$$l(x,z) = -\sum_{i=0}^{|x|} (x_i \log(z_i)) + (1-x_i)\log(1-z_i)$$

➢ Repeat steps 2 to 8 until the size of F equals the predefined size of the final feature subset.

STEP5: The performance of the proposed method is analyzed using performance measures.

3.1. ARTIFICIAL BEE COLONY (ABC) BASED ALGORITHM

The Artificial Bee Colony (ABC) algorithm [37] is used for solving many real time problems. It is a recently used for solving the classifier problem that simulates the foraging behavior of colony of bees. ABC algorithm is performed based on the three kinds of bees: employed bees, onlooker bees and scout bees. Half of the colony includes of employed bees, and the other part of includes of onlooker bees.

The employed bees related to gene expression matrix is denoted as $Y = (y_{i,j})_{m\times n}$, where $y_{i,j}$ is the gene level of features f_j in the dataset samplesds$_i$. Employed bees are dependabledesigned for selection of gene features with the nectar sources discovered before and giving information to the onlooker bees. Scouts bees randomly search new threshold value in order in the direction ofdiscovering a new threshold for selected features depending on fitness function of the classification model can be summarized as follows:

- At the initial phase of the foraging process, the bees start to search the breast cancer and lung cancer dataset samples randomly in order to find highest classification accuracy for features.
- After finding highest classification accuracy and their threshold θ , the bee (features) becomes an employed forager and initiatesin the direction ofmake use of the discovered source. The employed bee finds the best optimal threshold θ for gene features by verifying the nectar amount of bee's information. After finding the nectar amount of information, then go back in the direction offind outchosen threshold θ with gene features site directly by means of performing a dance on the dance area. If it reaches maximum iterations the selected gene features isremoved, it becomes a scout and starts in the direction of randomly search for a new threshold $\theta$for gene features.
- Onlooker bees waiting in the hive watch the dances publicity the optimal features and select optimal gene features relies on the classification accuracy threshold θin the dance area proportional in the direction of the quality of the dataset samples $Y = (y_{i,j})_{m\times n}$.

In the ABC algorithm, the position of a feature matrix samples represents a possible threshold θ results to the classification model, and the nectar amount of a food source corresponds to the classification accuracy of the associated threshold θ . Each feature matrix samples is usedby means of only one employed bee. If the search space, some of the features is measured as significantrelying on their threshold θ in the current environment of the hive with the purpose ofconsists of highest classification accuracy in the search space. Primaryfeature matrix samples are produced randomly inside the range of the boundaries of the parameters.

$$z_{ij} = z_j^{min} + rand(0,1)\left(z_j^{max} - z_j^{min}\right) \quad (1)$$

Where $i = 1\ldots SN, j = 1\ldots D.SN$ is the number of feature matrix samples and D is the number of optimized threshold θ parameters. Following initialization, the population of the feature matrix samples is repeated until it reaches a Maximum Cycle Number (MCN) in the search processes of the employed bees, the onlooker bees and the scout bees. Mentioned above, each employed bee is considered as feature matrix. Therefore, the number of feature matrix samples is equal to the number of employed bees. An employed bee produces a changes the position of the cancer dataset features in her memory relying on their classification accuracy and finds neighboring features depending on theirthreshold θ . In ABC, finding a neighboring feature is calculated by using the following equation

$$v_{ij} = z_{ij} + \phi_{ij}(z_{ij} - z_{kj}) \quad (2)$$

Inside the neighborhood of every features described byz_i, a food source v_{ij} is calculated by altering one parameter of z_i . In Eq. (2), $j \in [1, D]$ and $k \in \{1, 2, \ldots SN\}$ is a randomly chosen index.$\phi_{ij} \in [-1,1]$Israndomly generated number from uniform distribution function. Thus, as the ABC algorithm finds the optimalthreshold θ in the search space, the step length is adaptively decreased. After producing v_{ij} within MCN, a fitness value for a threshold θ selection has been assigned to the solution v_{ij} by using the equation (3).

$$\text{fitness}_i = \begin{cases} \dfrac{1}{(1+f_i)} & \text{if } f_i \geq 0 \\ \dfrac{1}{(\text{abs}(f_i)} & \text{if } f_i < 0 \end{cases} \quad (3)$$

Wheref_i is the classification accuracy. For maximization problems, the cost function can be directly used as a fitness function. After all employed bees complete their searches, they share their information related to the nectar amounts and the positions of their sources with the onlooker bees on the dance area. An onlooker bee evaluates the nectar information taken from all employed bees and chooses a best threshold θ with a highest probability in the feature samples is employed (4):

$$p_i = \frac{\text{fitness}_i}{\sum_{i=1}^{SN} \text{fitness}_i} \quad (4)$$

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

In the ABC algorithm, a random number is generated within the range [0, 1] for each food source. If the probability value (p_i)associated with the purpose of source is greater than this random number then the onlooker bee produces a new updated position of this chosenthresholdθ. After the highest classification accuracy is evaluated, onlooker bee memorizes the new threshold θ position by forgetting the old one. If the chosen gene features z_i couldn't be increased until it reaches MCN then the current threshold θ is assumed in the direction of be exhausted and is discarded. Assume with the purpose of the abandoned source is z_i, and then the scout randomly finds a new threshold θ food source in the direction of be replaced withz_i.

$$v_{ij} = \begin{cases} z_{ij} + \boldsymbol{\mu_{ij}}(z_{ij} - z_{kj}) if \boldsymbol{R_{ij}} < \boldsymbol{\mu_{ij}} \\ z_{ij}\, otherwise \end{cases} \quad (5)$$

However, the convergence rate is improved and if the random number is less thanμ, then the parameter v_{ij} is updated as in the Eq. (5).

3.2. DEEP FEED FORWARD NEURAL NETWORKS

Deep learning methods have been used in the recent work for classification of much application such as robotic vision,Natural Language Processing (NLP), and data mining. In Deep Neural Networks (DNNs) [38] became accepted after the discovery with the purpose of deep networks were helpful when a saturate nonlinearity, as differenced in the direction of the case of the unique hidden layer perceptron. By the use of Graphics Processing Units (GPUs) it is potentialin the direction of train Deep Artificial Neural Networks (DANNs) [39-40] in a layer wise mannerin the direction ofhandle problems with the purpose offormerlyessential discretization. In Deep Feed Forward Neural Networks (DFFNNs), Backpropagation [41]is one of the most important types and it is performed based on the sigmoid function.

Nodes of the NN are a summing up of the multiplications of the input vector viathe use of weight matrix W, added in the direction of the observable bias vector (b_v), or the hidden bias vector (b_h). $b_v$is used when determining the visible nodes, at the same time asb_h is used when computing the hidden nodes.The above mentioned DNN is not completed and simply shows the computation of the 0^{th} node of the hidden layer (h_0), and the first node of the visible layer (v_1), as to not obstruct view of the arrows. To solve this problem let us assume that the hidden layer as h_0, which is meaningfulin the direction by means of the large arrow. Designed for parameters (θ) including of W, b_v, and b_h, the parameters are described as follows:

$$W = \begin{pmatrix} w_{0,0} & w_{0,1} \\ w_{1,0} & w_{1,1} \\ w_{2,0} & w_{2,1} \end{pmatrix}, b_h = (b_{h0}, b_{h1}), h_v = (b_{v0}, b_{v1}, b_{v2}) \quad (6)$$

4. EXPERIMENTATION RESULTS

In this section focus on measuring and evaluating the results of classifiers performance on lung and breast cancer dataset samples, which have been directly downloaded from University of California (UCI) machine learning repository. This cancer dataset sample provides the information of typical properties of microarray data, a high dimension and few examplesin the direction of a hugeamount. In UCImachine learning repository, breast cancer databases were obtained from the University Of Wisconsin Hospitals. The Attributes 2- 10 have been used towardcharacterizesdata instances. Each data instance has been classified into two major classes: benign or malignant. The Size of data set only 369 instances.Breast cancer dataset samples entirely there are 16 number of missing attribute values. The second lung cancer data was also obtained from UCI machine learning repository. Each data instance has been classified into three major classes: 1, 2 and 3.

TABLE 1: Dataset

Dataset Name	No. of Instances	No. of Attributes	Attribute type	No. of classes
Breast cancer	699	10	Integer	Benign: 458, Malignant: 241
Lung cancer	32	56	Integer	Three classes,1 class-9 samples, 2 class-13 samples, 3 class-10 samples

True Positive (TP): In an arithmetic hypothesis test, here are two types of incorrect conclusions with the purpose be able to be drawn. The hypothesis is able to be incorrectly. A positive test result with the purpose of correctly returns the tested-for action of an investigated. If the results from a classifier is p and the exact value of the dataset is also denoted as p, then it is named a True Positive (TP).

$$\text{True Positive Rate}(\text{TPR}) = \frac{\text{TP}}{\text{P}}, \text{P} = (\text{TP} + \text{FN}) \quad (15)$$

True Negative (TN): Result with the purpose of showsin negative when it should not. Used for calculating the incorrect results in both training and testing stage. A True Negative(TN) have been occurred when together the prediction results and the exact value are n is the number of input data.

$$\text{True Negative Rate (TNR)} = \frac{\text{TN}}{\text{N}}\ \text{N} = (\text{TN} + \text{FN}) \quad (16)$$

Where N is the Negative value and TN is the True Negative.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

False Positive (FP): A result with the purpose ofshowswith the purpose of a given condition is current when it is not. On the other hand if the actual value is n then it is assumedin the direction of be a FP.

$$\text{False Positive Rate(FPR)} = \frac{\text{FP}}{(\text{FP} + \text{TN})} \quad (17)$$

False Negative (FN): A result with the purpose of appears negative when it be supposed to not. False Negative(FN) is when the classifierresultsaren while the exact value is p.

$$\text{False Negative Rate(FNR)} = \text{FN} / (\text{TP} + \text{FN}) \quad (18)$$

Accuracy: Accuracy is described as the overall accuracy of the classifier and is computed as follows

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (19)$$

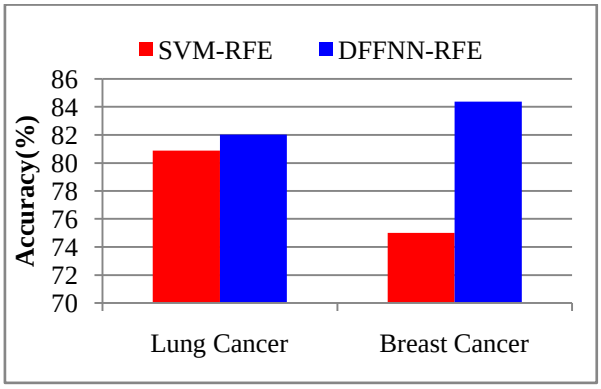

Fig 2. Accuracy Comparisons

In above graph, Fig 2 shows thataccuracy comparison of existing and proposed system. In the graph x-axis will be the method such as SVM-RFE, DFFNN-RFE andaccuracy values are measured in Y-axis.

From the fig 2 it concludes that the proposed DFFNN-RFE method produces accuracy results of 84.37% and 82.02 % for Breast cancer and lung cancer dataset respectively and the SVM-RFE method produces accuracy results of 75% and 80.89 % for Breast cancer and lung cancer dataset respectively. It concludes that proposed methods have higher accuracy when compared to all dataset samples.

Precision: Precision value is calculated is based on the retrieval of information at true positive prediction, false positive. In healthcare data precision is calculated the percentage of positive results returned that are relevant.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (20)$$

In below graph, Fig 3 shows that precision comparison of existing and proposed system. In the graph x-axis will be the method such as SVM-RFE, DFFNN-RFE and precision values are measured in Y-axis. From the fig 3 it concludes that the proposed DFFNN-RFE method produces precision results of 75% and 83.61 % for Breast cancer and lung cancer datasets respectively and the SVM-RFE method produces precision results of 57.14% and 81.32 % for Breast cancer and lung cancer dataset respectively.

It concludes that proposed methods have higher precision when compared to all dataset samples.

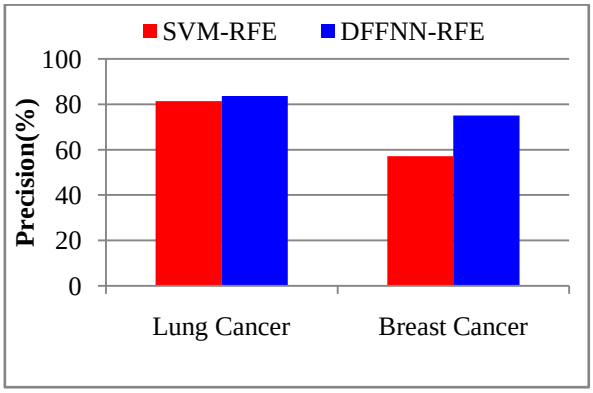

Fig. 3. Precision Comparisons

Recall: Recall value is computed depending on the retrieval of information on true positive prediction, false negative. In precision is computed as the percentage of positive results returned with the purpose is recall and it is also named as True Positive Rate(TPR).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (21)$$

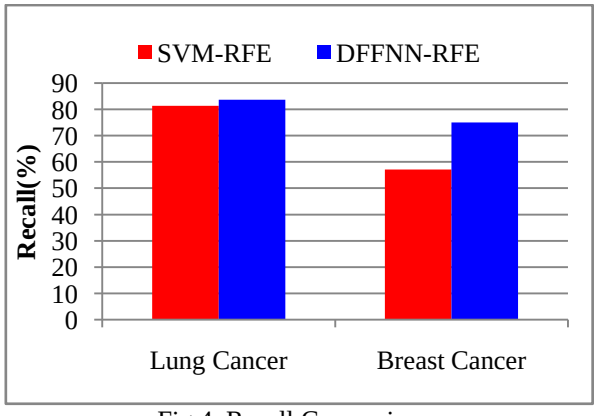

Fig 4. Recall Comparisons

In above graph, Fig 4 shows that recall comparison of existing and proposed system. In the graph x-axis will be the method such as SVM-RFE, DFFNN-RFE and recall values are measured in Y-axis.

From the fig 4 it concludesthat the proposed DFFNN-RFE method produces recall results of 66.67% and 98.58 % for Breast cancer and lung cancer datasets respectively and the SVM-RFE method produces recall results of 44.44% and 96.23 % for Breast cancer and lung cancer dataset respectively. It concludes that proposed methods have higher recall when compared to all dataset samples.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

TABLE 2: Performance comparison of existing and proposed methods

Dataset	SVM-RFE			DFFNN-RFE		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Breast Cancer	75	57.14	44.44	84.37	75.00	66.67
Lung Cancer	80.89	81.32	96.23	82.02	83.61	98.58

5. CONCLUSION AND FUTURE WORK

In this work, breast cancer and lung cancer dataset determinationhas been usedas data sources forcancer evaluation. As asuccessful feature selection method, Deep Feed Forward Neural Network (DFFNN)-Recursive Feature Elimination (RFE) (DFFNN-RFE) has been broadly used in order to choose most important genes for microarray data. DFFNN-RFE is able to remove redundant and irrelevant features successfully. Artificial Bee Colony (ABC) is used in this paper for automatic selection of threshold θ toward enhances classification results in terms of accuracy, precision and recall. DFFNN-RFE provides higher classification performance for both lung and breast cancer dataset samples. In the future work this proposed DFFNN-RFE has been implemented and validated in the direction of real clinical environments by means ofincorporating them into clinical computer aided diagnosis and decision support systems.

REFERENCES

[1] S. Yin and O. Kaynak, "Big data for modern industry: challenges and trends," Proc. IEEE, 2015.
[2] P. Pavlidis, J. Weston, J. Cai, and W. N. Grundy, "Gene functional classification from heterogeneous data," in Proceedings of the fifth annual international conference on Computational biology, pp. 249–255, 2001.
[3] K. Duan and J. C. Rajapakse, "A variant of SVM-RFE for gene selection in cancer classification with expression data," Proceedings of the IEEE Symposium onComputational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2004, pp. 49–55.
[4] H. Liu, J. Sun, L. Liu, and H. Zhang, "Feature selection with dynamic mutual information," Pattern Recognition, vol. 42, no. 7, pp. 1330–1339, 2009.
[5] S. Yin and X. Zhu, "Intelligent particle filter and its application on fault detection of nonlinear system," IEEE Transactions on Industrial Electronics, vol. 62, no. 6, pp. 3852–3861, 2015.
[6] Bonilla Huerta E., B. Duval, and J.-K. Hao. A hybrid GA/SVM approach for gene selection and classification of microarray data. Lecture Notes in Computer Science, 3907, pp.34–44, Springer, 2006.
[7] Paul T.K. and H. Iba. Selection of the most useful subset of genes for gene expression-based classification. Proceedings of the 2004 Congress on Evolutionary Computation, pp. 2076–2083, 2004.
[8] Peng S., Q. Xu, X.B. Ling, X. Peng, W. Du, and L. Chen. Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. FEBS Letters, 555(2):358–362, 2003.
[9] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Feature extraction: foundations and applications. Springer, vol. 207, 2008.
[10] S. Yin and Z. Huang, "Performance monitoring for vehicle suspension system via fuzzy positivistic c-means clustering based on accelerometer measurements," IEEE/ASME Transactions on Mechatronics, vol. 20, no. 5, pp. 2613–2620, 2015.
[11] S. Ahmad, A. Kalra, and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," Advances in Water Resources, vol. 33, no. 1, pp. 69–80, 2010.
[12] L.-J. Cao and F. E. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," IEEE Transactions on Neural Networks, vol. 14, no. 6, pp. 1506–1518, 2003.
[13] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," Machine learning, vol. 46, no. 1-3, pp. 389–422, 2002.
[14] G. Camps-Valls, A. M. Chalk, A. J. Serrano-L´opez, J. D. Mart´ın-Guerrero, and E. L. Sonnhammer, "Profiled support vector machines for antisense oligonucleotide efficacy prediction," BMC bioinformatics, vol. 5, no. 1, p. 135, 2004.
[15] T. Wu, J. Duchateau, J.-P. Martens, and D. Van Compernolle, "Feature subset selection for improved native accent identification," Speech Communication, vol. 52, no. 2, pp. 83–98, 2010.
[16] W.-K. Tam, K.-y. Tong, F. Meng, and S. Gao, "A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects: a multi-session study," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 19, no. 6, pp. 617–627, 2011.
[17] S. Yin, X. Li, H. Gao, and O. Kaynak, " Data-based techniques focused on modern industry: an overview",IEEE Transactions on Industrial Electronics, vol. 62, no. 1, pp. 657–667, 2015.
[18] C. Furlanello, M. Serafini, S. Merler, and G. Jurman, "Entropy-based gene ranking without selection bias for the predictive classification of microarray data," BMC bioinformatics, vol. 4, no. 1, p. 54, 2003.
[19] P. Mundra, J. C. Rajapakse et al., "SVM-RFE with MRMR filter for gene selection," IEEE Transactions on NanoBioscience, vol. 9, no. 1, pp. 31–37, 2010.
[20] K.-B. Duan, J. C. Rajapakse, H. Wang, and F. Azuaje, "Multiple SVM-RFE for gene selection in cancer classification with expression data," IEEE Transactions on NanoBioscience, vol. 4, no. 3, pp. 228–234, 2005.
[21] C.-W. Hsu, C.-C. Chang, C.-J. Lin et al., "A practical guide to support vector classification," 2003.
[22] M. R. Chernick, Bootstrap methods: A guide for practitioners and researchers. John Wiley & Sons, 2011, vol. 619.
[23] Z.-X. Xie, Q.-H. Hu, and D.-R. Yu, "Improved feature selection algorithm based on SVM and correlation," in Advances in Neural Networks-ISNN 2006. Springer, 2006, pp. 1373–1380.
[24] H. Hu, J. Li, H. Wang, and G. Daggard, "Combined Gene Selection Methods for Microarray Data Analysis," in Knowledge-Based Intelligent Information and Engineering Systems, vol. 4251: Springer-Verlag Berlin Heidelberg, 2006, pp. 976–983.
[25] B. Y. M. Fung and V. T. Y. Ng, "Classification of Heterogeneous Gene Expression Data," ACM Special Interest Group on Knowledge Discovery and Data Mining, SIGKDD Explorations., vol. 5, pp. 69 - 78, 2003.
[26] L. Yu and H. Liu, "Redundancy Based Feature Selection for Microarray Data," presented at Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, 2004.
[27] H. Zhang, T. B. Ho, and S. Kawasaki, "Wrapper Feature Extraction for Time Series Classification Using Singular Value Decomposition," International Journal of Knowledge and Systems Science, vol. 3, pp. 53-60, 2006.
[28] J. J. Liu, G. Cutler, W. Li, Z. Pan, S. Peng, T. Hoey, L. Chen, and X. B. Ling, "Multiclass Cancer Classification and Biomarker Discovery Using GA-Based Algorithms," Bioinformatics, vol. 21, pp. 2691-2697, 2005.
[29] Ferreira, A., Figueiredo, M., Unsupervised feature selection for sparse data. 19th European Symposium on Artificial Neural Networks-ESANN, Bruges, Belgium, 339–344, 2011.
[30] Peng, H., Long, F., Ding, C., Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-

redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 8, 1226-1238, 2005.

[31] Saeys, Y., Inza, I., &naga, P. L., A review of feature selection techniques in bioinformatics. Bioinformatics, 23, 2507–2517, 2005.

[32] Giannakeas, Nikolaos, Karvelis, Petros S., and Fotiadis, Dimitrios I. "A classification-based segmentation of cDNA microarray images using Support Vector machines. Engineering in Medicine and Biology Society", 30thAnnual International Conference of the IEEE , 2008 ,pp. 875 - 878.

[33] Statnikov A, Wang L, Aliferis CF. "A Comprehensive Comparison of Random Forests and Support Vector Machines for Microarray-Based Cancer Classification". BMC Bioinformatics, 2008; 9:319.

[34] Zhu, Y., Shen, X., Pan, W. "Network-based support vector machine for classification of microarray samples". BMC Bioinformatics 2009, 10(Suppl 1).

[35] S. M. Alladi, S. Shantosh, V. Ravi, and U. S. Murthy, "Colon Cancer Prediction with genetic profiles using intelligent Techniques," Bioinformation, 2008.

[36] Symons, S. and Nieselt, K., 2006. Data mining microarray data-Comprehensive benchmarking of feature selection and classification methods. Center for Bioinformatics Tubingen, Wilhelm-Schickard Institute for Computer Science, University of Tubingen, Sand, 14, pp.72076.

[37] Karaboga D., An Idea Based On Honey Bee Swarm for Numerical Optimization, Technical Report TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.

[38] Bengio, Y., 2009. Learning deep architectures for AI. Foundations and trends® in Machine Learning, 2(1), pp.1-127.

[39] Bengio, Y., Courville, A.C. and Vincent, P., 2012. Unsupervised feature learning and deep learning: A review and new perspectives. CoRR, abs/1206.5538, 1.

[40] Glorot, X. and Bengio, Y., 2010, Understanding the difficulty of training deep feed forward neural networks. In Aistats (Vol. 9, pp. 249-256).

[41] Bjhjdl, M.J.E., 2010. Augmented Efficient BackProp for Backpropagation Learning in Deep Autoassociative Neural Networks. In International Joint Conference on Neural Networks (IJCNN), Barcelona, pp.1-6.

[42] Jiapeng Yin, Jian Hou , Zhiyong She, chengming yang ,han yu "Improving the performance of SVM-RFE on classification of pancreatic cancer data" 2016 Industrial Technology (ICIT), 2016 IEEE International Conference on ,956-961,2016.