Journal of King Saud University – Computer and Information Sciences 32 (2020) 940–948

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Elephant search optimization combined with deep neural network for microarray data analysis

Mrutyunjaya Panda

Department of Computer Science and Applications, Utkal University, Vani Vihar, Bhubaneswar, India

ARTICLE INFO

Article history:
Received 1 August 2017
Revised 2 December 2017
Accepted 3 December 2017
Available online 5 December 2017

Keywords:
Gene expressions
Elephant search
Firefly search
Deep learning
Classification accuracy
Tukey HSD test

ABSTRACT

Even though there is a plethora of research in Microarray gene expression data analysis, still, it poses challenges for researchers to effectively and efficiently analyze the large yet complex expression of genes. The feature (gene) selection method is of paramount importance for understanding the differences in biological and non-biological variation between samples. In order to address this problem, Elephant search (ESA) based optimization is proposed to select best gene expressions from the large volume of microarray data. Firefly search (FFS) is also used to understand the effectiveness of the Elephant search method in feature selection process. Stochastic gradient descent based Deep Neural Network as Deep learning (DL) with softmax activation function is then used on the reduced features (genes) for better classification of different samples according to their gene expression levels. The experiments are carried out on ten most popular Cancer microarray gene selection datasets, obtained from UCI machine learning repository. The empirical results obtained by the proposed elephant search based deep learning approach are compared with the most recent published article for its suitability in future Bioinformatics research. Finally, Statistical significance test by one-way ANOVA with post hoc Tukey's test is conducted to deduce a number of insights on the selection of the best classification model.

1. Introduction

Gene is the basic unit of storage containing hereditary information in living beings. From the technical point of view, it can be treated as a distinct sequence of nucleotides constituting part of a chromosome. Microarrays data analysis is relatively a new technology that aims to help in finding the right treatment for many diseases with the accurate medical diagnosis from the huge number of genes across different experimental conditions. Due to the expensive and complicated nature of the microarray datasets, it is somehow difficult to predict and hence demands careful experimentation with appropriate statistical tools for fruitful analysis. It is well known that the gene expression is a process that maps genes DNA sequence into its corresponding mRNA sequences and then finally to amino acid sequences of proteins. Microarray data analysis is a powerful technology with enormous opportunities presents gene expression profiling to describe the expression levels of hundreds and thousands of genes in cells correlated with corresponding protein, helps one to understand the cellular mechanism of the biological processes. Data Mining helps extract meaningful observations in such a huge and complex microarray gene expressions datasets as a post-genome cancer diagnostics to uncover the details on how the genes are regulated; how genes make an impact on the cancerous mutation of cells and how the performance depends on various medical experimental conditions to name a few. The microarray dataset presents samples or a condition in rows while the respective genes are provided in a column.

The classification data mining is the great impression of dealing with the patients' gene expression profile for a specific disease in a best possible manner, urges of more research in the area for better predictive accuracy. As a large number of genes are present in the Microarray data analysis, it is always suggested to carry out some potential gene (features) selection algorithms in order to find the most informative genes to reduce the curse of dimensionality. This further may be applied with a best possible classifier to predicting the samples correctly to achieve high accuracy reducing the computational cost and more importantly efficient and effective diagnosis and prognosis can then be customized for the treatment for that patient.

E-mail address: mrutyunjaya.cs@utkaluniversity.ac.in
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2017.12.002

The microarray data analysis needs a clear objective to see its successful implementation for a greater cause of the society at large, as cited by Tjaden and Cohen (2006). Clustering is one of the popular technique being used for gene profiling microarray data analysis using K-means clustering and Self-organizing maps (SOM) (Sheng-Bo et al., 2006; Young, 2009). Alshamlan et al. (2013) presents a comprehensive study of objectives and approaches for cancer microarray gene expression analysis and conclude with a detailed investigation on the available approached in this area.

Researchers found that most of the cancer studies with microarray gene expression profiling contain a comparison of various classes of diseases (Simon, 2009; Wang et al., 2007) and their predictions, hence sought for using classification algorithms instead of clustering ones (Dougherty et al., 1995). Support vector machine (SVM) is considered to be one of the most popular and well-established classification methods for microarray data analysis for binary classifications initially (Platt et al., 2000). But, as many cancer datasets are of Multi-class, researchers have proposed to use many variants of SVM such as: DAGSVM (Platt et al., 2000), evolutionary SVM (ESVM) (Huang and Chang, 2007), genetic algorithm based SVM (GASVM) (El Akadi et al., 2009) and Fuzzy SVM (FSVM) (Mao et al., 2005).

Neural network based classifiers are also been proposed by many to effective and efficient microarray data analysis that includes: Functional link Neural network (FNN) (Wang et al., 2007), Extreme learning machines (ELM) (Zhang et al., 2007), improved wavelet neural network (WNN) (Zainuddin and Pauline, 2009), probabilistic neural networks (PNN) (Berrar et al., 2003) and subsequent artificial neural network (SAAN) (Roland et al., 2004).

Apart from single classification or clustering methods for gene classification, ensemble methods are also been used researchers to Multi-class classification problems, but it is noticed that the ensemble methods are not able to improve the performance in comparison the single classifier methods (Ghorai et al., 2010).

The authors (Kothandan and Biswas, 2016) present a comparison between kernel-based methods and decision trees to explore the best predictive model for identifying miRNAs involved in the cancer pathway.

Considering all the above into consideration, one can conclude that an efficient gene selection method is a must with novel approaches followed by the development of a promising fast classifier for better gene prediction with acceptable accuracy.

This motivated us to carry out further experiments on using a novel elephant search-based optimization with a deep neural network classifier for improved performance in diverse microarray dataset with two-class, three-class, and four class classification.

1.1. Motivation and objective

Even though there are lots of research opined of using either filter based or wrapper based or a hybrid of these two, for finding a subset of most informative genes for better clinical diagnosis, still there is lot to achieve in terms of performance with new gene selection (feature selection) methods for obtaining new insights into the clinical diagnosis. Considering gene selection is NP-hard (Patrenahalli and Fukunaga, 1977) and finding optimal gene from gene expression profiles is really a challenge for getting predictive accuracy. There are several recommendations of using classification and clustering methods to address the problem with adding of novel Multi-objective optimization and some sort of suitable classifier for addressing the binary and Multi-class microarray data set as a future scope present in the literature. Hence, we are motivated to solve this issue by using an efficient elephant search-based optimization with stochastic gradient descent based deep learning classifier. Further, we compared with the approach with already established Firefly search optimization for checking the efficacy of our proposed approach. Finally, a comparison with others related work justifies our proposal.

2. Material and methods

This section discusses the datasets and methodology adopted in this paper.

2.1. Datasets used

We use publicly available microarray dataset (Liu et al., 2005; Zhu et al., 2007) for our proposed research in this paper, the details are as provided below:

Prostate Cancer: This dataset contains is used for binary classification to classify the Tumor Vs Normal samples. The training dataset contains 52 prostate tumor sample and 50 normal samples. This dataset contains 12,600 cancer genes.

Leukemia (ALL-AML): This dataset contains bone marrow samples, collected over 7129 probes from 6817 human genes, out of which 38 samples (27 ALL and 11 AML) are for training purposes and 34 samples (20-ALL and 14 AML) for testing purpose. This dataset is used for binary classification.

Colon Tumor: This dataset contains 62 samples out of which 40 samples are negative (tumor biopsies are from tumors) and 22 samples are positive (biopsies are from healthy parts of the colons of the same person. Based on the confidence in the measured confidence level, 2000 out of 6500 genes are selected. This fall under binary classification.

DLBCL-Stanford: Diffuse large B-cell lymphoma (DLBCL) data set contains 47 samples (24 from "germinal center B-like" group and 23 are from "activated B-like" group), where each sample is represented by 4026 genes expressions. This is used for 2-class classification.

Lung-H: The Lung-Harvard dataset contains Multi-class (5-class) for classification. This has 203 samples of lung tumors (139 sample of lung adenocarcinomas (labeled as ADEN) + 21 of squamous cell lung carcinomas (labelled as SQUA) + 20 from pulmonary carcinoids (labeled as COID)+6 from small-cell lung carcinomas (labeled as SCLC) and 17 normal lung samples (labeled as NORMAL) with each sample consisting of 12,600 genes.

Ovarian Cancer: The ovarian dataset makes us understand the situation to distinguish the proteomic patterns in serum to have the symptom of ovarian cancer or not. The cancer is most significant to the women with similar family history. The dataset is obtained from the proteomic spectra generated by mass spectroscopy with 253 samples (162 ovarian samples and 91 normal ones). The raw spectral data contains the relative intensity amplitude for each sample of 15,154 M/Z (molecular mass/charge) identities. This is used for 2-class classification.

Breast Cancer: This dataset contains samples of 78 patients (34 is relapsed category where the distance metastases developed in the patients within 5 years of time and the rest is the non-relapse category for healthy patients within the same period of time). The total number of genes present in the dataset is 24,481. The value of "NaN" symbol in original ratio data is replaced with 100.0. This is used in binary classification.

MLL: This dataset contains 3-class (conventional acute lymphoblastic (ALL), acute myeloid leukemia (AML) and Mixed-lineage leukemias (MLL)). The MLL translocations are basically found in infant leukemias and chemotherapy-induced leukemia with a uniform and distinct pattern to classify all the classes.

SRBCT: SRBCT (small round blue cell tumors) data set contains four different types of childhood cancer tumors such as Ewing's family of tumors (EWS), neuroblastoma (NB), non-Hodgkin lym-

phoma (BL) and rhabdomyosarcoma (RMS). The gene expression values of these tumors based on responses to therapy and prognosis with different treatment options with similar appearance of routine histology makes it's extremely challenging for classification data mining.

CNS: The CNS (central nervous system) dataset presents the outcome prediction of the patients for embryonal tumor. This contains a total of 60 samples (21 are survivors and 39 are failures) with 7129 number of genes. This is used for 2-class classification problem.

2.2. Methodologies used

This section highlights the gene selection and classification methodology adopted in this paper.

2.2.1. Gene selection methods

The gene (or otherwise called as features) selection is of paramount importance for dimension reduction in a huge dataset. The selection of the minimal best genes that represent the original genes in the dataset may either lead to give a faster result with acceptable accuracy.

Bio-inspired search algorithms are such a popular gene selection methods that seem to address the various complex problems with large scale, NP-hard, and multimodal nature; most effectively.

The classical search methods produce local optimum so that they are faster and provide the best accuracy in comparison to the global search optimization methods (such as Genetic algorithm and particle swarm optimization etc.). The local search methods need a good understanding of the initial starting point without which it may not produce an effective result, in contrary, the global search methods do not rely on such initial understanding and are less probable in trapping in a local minimum (Eslami et al., 2013).

The following section discussed the two promising gene selection method for our experimentation.

2.2.1.1. Firefly search. The Firefly search (FFS) was initially coined by Yang (2010) considered to be one among the latest population-based global optimization method which works by mimicking the flashing behavior of the fireflies.

The simplicity in implementation and efficient computation makes FFS makes it an ideal choice, in comparison to artificial bee colony (ABC), Particle Swarm optimization (PSO) and ant colony optimization (ACO) to name a few (Kora and Sri Rama Krishna, 2016). There are around 2000 firefly species which are small insects capable of flashing short and rhythmic light, which in turn attract other fireflies. Since the light intensity attraction decreases with the distances, the fireflies are only visible up-to several hundreds of meters. The objective function used in the algorithm is linked with the fluorescence light behavior of fireflies. The fireflies move randomly if it found no brighter Firefly than it or else it follows the brightest neighboring one.

The working procedure of Firefly algorithm is detailed in Fig. 1.

The advantages of using Firefly algorithms lay down with the following points.

- FFA deals with natural way of multimodal optimization by dividing the whole population into subgroups, then each subgroup into local modes and in each local mode, there is existence of global optimum,
- The FFA converges faster through nonlinear attractiveness behaviors among its multiple agents.
- The simplicity makes it popular in use for diverse optimization problems.

2.2.1.2. Elephant search. The Elephant Search Algorithm (ESA) (Deb et al., 2015) is a highly nonlinear, multimodal global optimization technique inspired by the biological habits of elephant herds.

The Elephant search is considered to be a good optimization technique for its intensification in local search space for a better solution. It provides global optimal solutions with covering reasonable search space without falling into local optima.

Firefly search (FFS)

Step-1: begin
Step-2: Define objective function $f(x_i)$ and initialize a population of fireflies X_i , $i=1,2,\ldots,n$
Step-3: Light intensity value (I_i) is calculated at X_i by using objective function $f(X_i)$
Step-4: initialize Light absorption coefficient (γ)
Step-5: (t< Maximum generation of fireflies)
and i=1 to n for all n fireflies
and j=1 to n, for all n fireflies
Step-6: Use Cartesian distance measure to obtain the distance between X_i and X_j
Setp-7: If intensity at J is more than the intensity at I ($I_j > I_i$)
Step-8: Move fireflies from I to J in all dimension (D)
Step-9: end if
Step-10: Attractiveness varies with distance r via $e^{-\gamma r^2}$
Step-11: New solutions are evaluated and then light intensity are updated
Step-12: end of j
Step-13: end for i
Step-14: rank the fireflies in order to find the current best
Step-15: end while
Step-16: Post-processing and then end of procedure

Fig. 1. Firefly Search Algorithm working procedure.

Further, in this process, when two male elephant collides, they move opposite to each other depending on their unequal mass where their visual range depicts their radius of the moving bodies. The mass represents the fitness value and the elephant having the higher mass pulls out the other elephant and becomes the winner and then scouts for better position further, in search of the food. It is worth noting here that the male elephants are roaming near to the female elephants rather they move independently to explore the best position on their own in search of the food.

The Pseudo code for Elephant search optimization process is outlined in Fig.2.

This search algorithm does the following tasks for its implementation.

- The best possible solution is achieved through an iterative process that is represented by the lifetimes of the searching elephants.
- The local searches are led by some chief female elephants to a higher likelihood of obtaining best results.
- The male elephants are Rangers to lead the elephant clan so that the whole elephant can go out of local optimum.

While elephant search is being implemented, the following are addressed for their effectiveness.

- The visual range of each elephant is fixed and can be calculated using Euclidean distance. It can be observed that the visual range is better in male elephants rather than the female ones.
- Secondly, in cases when two or more elephants are searching for visual range, current fitness values are taken for comparison. The elephant with the higher fitness value shall remain and the others can randomly be removed.
- In this basic ESA (elephant search algorithm), it is observed that only a single female elephant group exists and there is no separation of the group.
- In case an elephant dies, the new baby elephant of the same sex is born for group's gender balancing and fixed group size.

2.3. Deep learning

Deep learning (Min et al., 2016) is termed as universal approximator because of its mapping from input to output as $y = f(x)$ to find out correlation among attributes x and y present in the dataset. Neural networks are modeled based on the working of the human brain for pattern recognition. Deep learning (Deep neural network or DNN) differs from the conventional neural network in terms of depth, consisting of more than one hidden layers apart from input and output layer. This is why deep learning is also called as "stacked neural network". Minimum of three hidden layers can be thought of as deep learning. Deep learning further can be seen to have feature hierarchy since they combine and aggregate the features from one layer to the next. This way, it increases complexity and level of abstraction and makes it a good choice for handling the very large and high dimensional complex data set. The performance comparison of deep Learning with conventional learning methods are shown in Fig. 3.

The deep neural network needs many hyper-parameter to be set for implementation and at the same time, it is to be noted that finding optimal set of values for that hyper-parameter may not be feasible using gradient descent algorithm due to several constraints like the dataset is a mix of both real and discrete; each

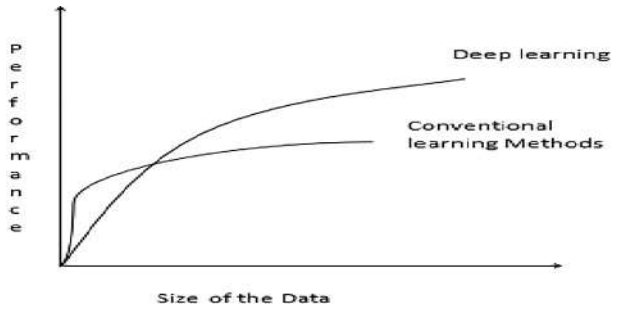

Fig. 3. Deep learning with data size.

Elephant search Algorithm (ESA)

Initialization of parameters: parameters initialize ( )

Initialize the Elephant Groups ( )

While stopping criteria or the number of iteration to be used are satisfied,

    move the male elephant ( )

If two different groups elephants are engaged in visual range,

    then avoid local optimum ( )

        and then in that condition, move by group leader ( )

Female elephant deep Search ( ) is then performed

Elephants die and new elephant born ( )

Finally, update each group's fitness value ( )

end while

Post-processing and end of procedure

Fig. 2. Elephant Search Optimization process.

hyper-parameter is difficult to be optimized alone and finding a local minima involves a great deal of time. Initially, the weights of a deep neural network are small enough so that the activation function (softmax activation function is used here) operates linearly with large gradient value. The learning rate of the deep neural network should be chosen efficiently so that the validation error is kept to a minimum. Further, looking at the input, more network capacity is needed and hence sought for a large number of hidden layers. The L1 or L2 regularization scheme is needed to check whether the deep neural network can provide better solutions. In this process, three hidden layers are considered with ReLu (Rectified Linear Unit), whereas at the output layer, softmax activation functions combined with MCXENT (multi-class cross entropy) are considered. No hidden layer should be less than a quarter of the input layer's nodes. For larger data size, more hidden layers are advised. At the same time, if one chose a number of hidden layers as same as that of input nodes, then there is a chance of identity loss and at the same time, too many hidden layers may result in noise and overfitting. In order to avoid overfitting, L1 and L2 regularization may be employed.

The number of epochs neither to be very less for better parameter learning nor to be very large, to avoid overfitting on training data. Iteration defines the number of mini-batch wise parameter updates in a row. Mini-batch refers to the number of examples considered at a time for computing gradients and parameter updates. Even though the choice of mini batch size largely depends on the applications, a size of 1 will not provide the benefits of parallelism; a size of 10 will be too small for GPU but acceptable for CPU; but, a size of more than 10 to 100 may provide expected results.

For faster training becomes possible, stochastic gradient descent (SGD) optimization combined with a Nesterov momentum updater is used in this paper.

The parameter setting used in this gene selection and classification strategy is as follows:

- Firefly search/Elephant search: Absorption type-0.001 (setting of the absorption coefficient of the firefly population members), Betamin-0.33 (set the zero distance attractiveness of the Firefly members), Accelerator type-normal, Chaotic coefficient-4.0, Chaotic mapping type-logistic map, Number of iterations-20, Objective type-merits/multiobjective, Population size-20 (particles in the swarm). Report frequency-20 (set how frequently reports are generated)
- Deep learning (Deep Neural Network): Activation function-Softmax, Weight initialization method-XAVIER, Bias initialization-1.0, Distribution function- Normal Distribution, Learning Rate-0.1, Bias Learning rate-0.01, Momentum-0.9, Updater for stochastic gradient descent- NESTEROVS, Gradient normalization threshold-1.0, Loss function-Loss MCXENT, ADADELTA's rho parameter-0.0, ADADELATA Epsilon parameter-1.0E-6, RMSPROP's RMS delay parameter-0.95, ADAM's Mean decay parameter-0.9, ADAM's variance parameter-0.999, Number of Epochs-10, Optimization algorithm- SGD (stochastic gradient descent), Batch size-100, Seed-1, Number of decimal places-2

The Pseudo code of the Deep Neural Network classifier is outlined in Fig. 4.

Here, S = selected set, C = Candidate set and $F = S \cup C$. Input weights = W_F, selected input weights, as selected with attributes in S = Ws, Wc as candidate weights, G_F indicates corresponding gradient to W_F, G_{Fj} indicates G_F to select one feature such as jth one from C,W_{Fj} indicates W_F with newly selected input weights representing the weights associated with jth input node in W_F. $S = S \cup F_j$ and $C = C \backslash F_j$ indicates that S and C are to be updated by adding or removal of j respectively.

ADADELTA (Zeiler, 2012) is a per dimension learning rate method basically used for gradient descent method in deep learning classifier, that takes minimum computation overhead, no manual tuning of learning rate (hence dynamic adaptation) and robust to noisy data in selection of hyper-parameters.

ADAM (Kingma and Ba, 2015), is a straightforward and simple stochastic gradient descent optimization method used to adaptive estimate of the lower order moments efficiently. This way, it takes less memory for computation, which is interesting.

3. Experimental findings and discussion

In this, we present our experimental results and discuss its effectiveness in microarray gene expression profiling. Fig. 5 shows

Deep Neural Network Algorithm:

Step-1: Input: X=(X_1,X_2,…,Xn) denotes data matrix of n samples, Y=(Y_1,Y_2,…,Y_n)^T as their corresponding output labels, the maximum number of selected attributes K.

Step-2: Initialize: S={bias}, C=F and W_c=0

Step-3: While $|S| \le k+1$ do

Step-4: Assign W_c (candidate weights)=0;

Step-5: Update weight of hidden layers as well as input weight W_s;

Step-6: Multiple times Drop out to be used and then obtain average G_{F_C}

Step-7:calculate $j = \arg\max_{c \in C} \left\| G_{F_C} \right\|_q$

Step-8: update Learning rates using AdaDelta

Step-9: Initialize W_{Fj} with Xavier initialization

Step-10: Perform $S = S \cup F_j$ and $C = C \backslash F_j$

Fig. 4. Pseudo code for Deep Neural network algorithm.

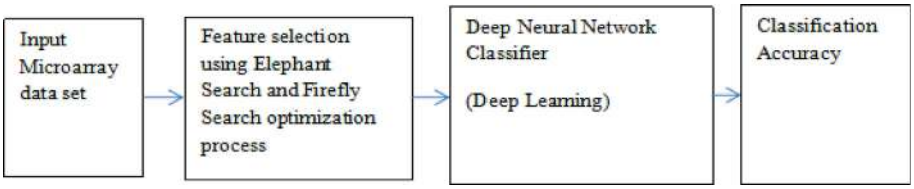

Fig. 5. Experimental process.

Table 1
Firefly search-based optimization with Deep learning classifier.

Sl.No.	Dataset	Original Attribute (genes)	Instances	Number of classes	Reduced Attributes (genes)	Time in seconds	Accuracy in%
1	Prostate Cancer	12,601	102	2	5189	1.24	87.26
2	Leukemia (ALL-AML)	7130	38	2	2463	1.84	100
3	Colon Tumor	2001	62	2	562	0.43	77.43
4	DLBCL-Stanford	4027	47	2	1805	0.79	89.36
5	Lung-Harvard	12,601	203	5	5304	15.53	93.11
6	Ovarian Cancer	15,155	253	2	35	17.23	97.24
7	Breast Cancer	10	286	2	6	0.75	65.39
8	MLL	12,582	72	3	190	14.59	80.56
9	SRBCT	2308	83	4	768	0.63	93.98
10	CNS	7129	72	2	1526	0.45	56.67

Table 2
Elephant search-based optimization with Deep learning classifier.

Sl. No.	Dataset	Original Attributes (genes)	Instances	Number of classes	Reduced Attributes (genes)	CPU Time in seconds	Accuracy in %
1	Prostate Cancer	12,601	102	2	4267	1.33	88.24
2	Leukemia (ALL-AML)	7130	38	2	1044	0.31	92.11
3	Colon Tumor	2001	62	2	572	0.41	79.03
4	DLBCL-Stanford	4027	47	2	1717	0.44	91.49
5	Lung-Harvard	12,601	203	5	4545	2.92	94.10
6	Ovarian Cancer	15,155	253	2	384	1.67	99.21
7	Breast Cancer	10	286	2	6	1.09	73.43
8	MLL	12,582	72	3	190	15.13	80.56
9	SRBCT	2308	83	4	306	0.66	83.14
10	Central Nervous System (CNS)	7129	72	2	1621	0.47	53.34

the experiment process carried out in this research. The Input Microarray dataset is applied to Elephant search (Fig. 2) as well as to Firefly search optimization process (Fig. 1) separately to obtain the reduced features (genes), as a part of feature selection process. The reduced features are then applied to Deep Neural network classifier (Fig. 4) as a deep learning process to obtain the better gene expression classifications.

To address the possible overfitting of the data during training of neural network model, 10-fold cross validation method is adopted, where the whole training data is divided into 10 subsets and then running the classification algorithm 10 times. Here, at first, 9/10 part of the training data is used to train the model and the rest 1/10 is used for the testing purpose. The process continues till each 1/10 subset is used exactly once for validation of the model. At last, the average accuracy is calculated to understand the effectiveness of the proposed classification algorithm.

Finally, the results are validated with the obtained classification accuracy and CPU model build time. Predictive accuracy or classification accuracy can be obtained as the ratio between the numbers of samples correctly classified to the total number of samples.

Tables 1and 2 demonstrate the proposed approach with FFS and ESA based Deep learning respectively.

It can be observed from Tables 1 and 2 that Deep learning works well for almost in all datasets except for CNS dataset. This may be due to the less number of instances available for classification.

Further, a comparative study is provided in Table 3, for the understanding the efficacy of the proposed approach. The comparison opines that the proposed approach achieves comparable accuracy for all the microarray datasets. While verifying the suitability of others work as given in Table 3, it is observed that Vural et al. (2015) have used singular value decomposition along with information gain to reduce the number of attributes till they obtain less than that of the number of samples and Mukkamala et al. (2005) presented their work with different number of reduced attributes.

A step forward in this direction, Tables 4 and 5 presents a comparative analysis is made with others work in terms of classification accuracy and the execution time in seconds for building the classification model. From the results in Tables 4 and 5, one can see the proposed ESA-DL is fastest among all other algorithms for all datasets; provides best accuracy in Lymphoma and Prostate cancer dataset and acceptable accuracy for other dataset. It can also be observed that FFS-DL is the best in leukemia dataset, while ELM (Yashoda and Ponmuthurammalingam, 2015) is best in MLL and SRBCT datasets.

Understanding the reduced number of genes may play a vital role in achieving high accuracy for microarray data analysis, extensive comparison are carried out with the most recent works being proposed by the researchers which is shown in Table 6.

It is quite evident from the Table 6 that FFS + DL is 100% accurate for Leukemia dataset, ABC-SVM is best for all other datasets.

Table 3
Accuracy (%) Comparison with some existing research (Part-1).

Method/Dataset	ALL-AML	Colon Tumor	SRBCT	Lung-H	DLBCL	Prostate Cancer
SVM Vural et al. (2015)	97.14	83.87	95.18	93.6	98.7	–
ANN Vural et al. (2015)	91.43	83.87	95.18	92.12	94.81	–
Random Forest Vural et al. (2015)	91.43	87.1	86.75	90.64	90.91	–
PSA (Glinsky et al., 2004)	–	–	–	–	–	77
MARS with 6 attributes Mukkamala et al. (2005)	–	–	–	–	–	68.2
Random forest with 6 attributes Mukkamala et al. (2005)	–	–	–	–	–	80.2
LGP with 6 attributes Mukkamala et al. (2005)	–	–	–	–	–	92.1
Ours (FFS + DL)	100	77.42	93.98	93.11	89.36	87.26
Ours (ESA + DL)	92.11	79.03	83.14	94.1	91.49	88.24

Table 4
Accuracy (%) Comparison with some existing research (Part-2).

Methods	Dataset	% accuracy	Execution time in seconds
Pillar NN (Susmi et al., 2015)	ALL-AML (Leukemia)	85.6	–
NN (Susmi et al., 2015)		82.35	–
ELM (Yashoda and Ponmuthurammalingam, 2015)		79	25
BPN (Yashoda and Ponmuthurammalingam, 2015)		55	49
ESA + DL (ours)		92.11	0.44
FFS + DL (ours)		100	0.79
ELM (Yashoda and Ponmuthurammalingam, 2015)	Lymphoma (DLBCL)	65	30
BPN (Yashoda and Ponmuthurammalingam, 2015)		47	53
ESA + DL (ours)		91.49	0.44
FFS + DL (ours)		89.36	0.79

Table 5
Accuracy (%) Comparison with some existing research (Part-3).

ELM (Yashoda and Ponmuthurammalingam, 2015)	Prostate	85	20
BPN (Yashoda and Ponmuthurammalingam, 2015)		64	37
ESA + DL (ours)		88.24	1.33
FFS + DL (ours)		87.26	1.24
ELM (Yashoda and Ponmuthurammalingam, 2015)	MLL	91	15
BPN (Yashoda and Ponmuthurammalingam, 2015)		66	30
ESA + DL (ours)		81	15.13
FFS + DL (ours)		80.56	14.59
ELM (Yashoda and Ponmuthurammalingam, 2015)	SRBCT	98	10
BPN (Yashoda and Ponmuthurammalingam, 2015)		78	22
ESA + DL (ours)		93.98	0.66
FFS + DL (ours)		92.77	0.63

Ours proposed ESA + DL have achieved almost near accuracy in comparison to ABC-SVM in most of the cases, but with low accuracy in case of Colon dataset.

In order to have a deeper visualization of the accuracy results obtained for several datasets; box plot for various algorithms is drawn, which is shown in Fig.6. It can be seen from the Fig.6 that a less number of genes selections perform better in terms of accuracy in some cases (Table 3 and Table 6) whereas in case of others (Table 4 and Table 5), our proposed approach with a large number of genes works well in terms of acceptable accuracy. Whereas the range between the upper tails and lower tails of the box plots are small, there is only a single data point fall below the lower tail, as in the case of ESA-DL (or ESA-DNN).

To summarize, a detailed assessment of performances of ESA + DL approach with ABC-SVM as highlighted in Table 6 in terms of% accuracy, a graphical representation using box plot is presented in Fig.6. It seems from Fig.6 that ABC-SVM performs better in comparison to all others.

In order to validate the experimental finding and to a have a right conclusion, we propose to use one-way ANOVA with post hoc Tukey HSD (Honest significant difference) test to have the statistical significance test (Haynes et al., 2013). The results obtained after the statistical significance tests are outlined in Tables 7 to 11.

3.1. Conclusion from ANOVA

The p-value corresponding to the F-statistic of one-way ANOVA is higher than 0.05, suggesting that the treatments are not significantly different for that level of significance. The Tukey HSD test, as well as other multiple comparison tests like Scheffe or Bonferroni, might not narrow down which of the pairs of treatments are significantly different. Even though the accuracy comparison of ours proposed ESA-DNN and ABC-SVM do not suggest the presence of significantly different treatment pairs in one-way ANOVA, still we perform the multiple comparison tests to uncover any newness for the conclusion. In some instances, a Bonferroni test of a small set of pairs might show significance, even though 1-way ANOVA suggests that there are too much noise and randomness in our data used for comparative analysis.

3.2. Tukey HSD test

The p-value corresponding to the F-statistic of one-way ANOVA is lower than 0.05 which strongly suggests that one or more pairs of treatments are significantly different. Here, we used $k = 2$ treatments, for which one shall apply Tukey's HSD test to each of the 1 pairs to pinpoint which of them exhibits statistically significant difference.

We first establish the critical value of the Tukey–Kramer HSD Q statistic based on the $k = 2$ treatments and $\nu = 8$ degrees of freedom for the error term, for significance level $\alpha = 0.01$ and 0.05 (p-values) in the Studentized Range distribution. We obtain these critical values for Q, for α of 0.01 and 0.05, as $Q_{\alpha} = 0.01$, $k = 2$, $\nu = 8$ as 4.7452 and 3.2612, respectively.

3.3. Post hoc Tukey HSD test calculator results

Post-hoc Tukey HSD test calculates by evaluating whether $Q_{i,j}$ is more than the critical $Q_{i,j}$ value for all relevant pairs of treatments.

Table 6
Accuracy (%) Comparison with some existing research (Part-4): with reduced genes shown inside the bracket.

Algorithm	Colon	Leukemia	Lung	SRBCT	Lymphoma
ABC-SVM (Alshaman et al., 2016)	95.61(20)	93.05(14)	97.91(8)	95.36(10)	96.96(5)
GA-SVM (Alshaman et al., 2016)	93.53(83)	91.99(51)	95.83(62)	92.77(74)	93.93(43)
PSO-SVM (Alshaman et al., 2016)	93.55(78)	95.83(53)	94.79(65)	93.97(68)	96.96(8)
PSO-SVM (Qi et al., 2007)	85.48(20)	94.44(23)	–	–	–
PSO-SVM (Javad and Giveki, 2013)	87.01(2000)	93.06(7129)	–	–	–
GA-SVM (Peng et al., 2003)	93.55(12)	–	–	–	–
mRMR-GA (El Akadi et al., 2009)	–	–	–	95(5)	–
ESVM (Huang and Chang, 2007)	–	–	95.75(7)	–	--
ESA + DL (ours)	79.03 (572)	92.11 (1044)	94.1 (4545)	93.98 (306)	91.49 (1717)
FFS + DL (ours)	77.42 (562)	100 (2463)	93.11 (5304)	92.77 (768)	89.36 (1805)

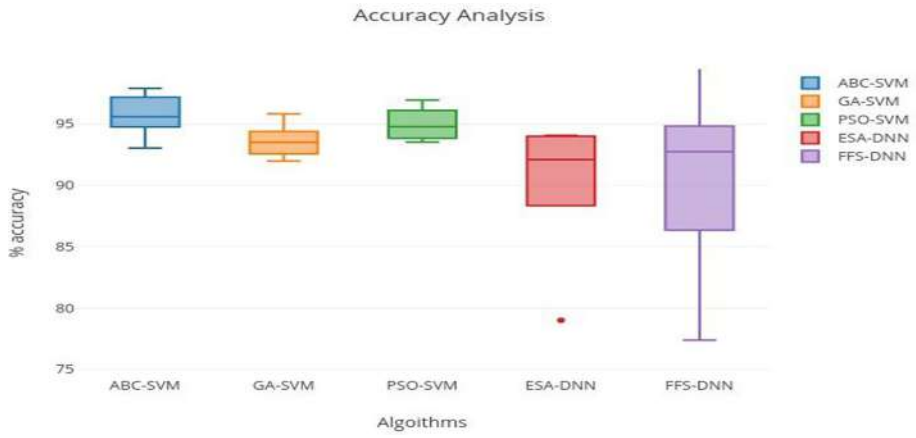

Fig. 6. Box plot for accuracy comparison.

Table 7
Statistical significance test with One-way ANOVA with post hoc Tukey HSD Test (Descriptive statistics of k = 2 independent treatments; ESA + DNN and ABC-SVM).

Treatment	ABC-SVM	Pooled Total	ESA + DNN
Observations N	5	5	10
Sum $\sum x_i$	478.8900	450.6100	929.5000
Mean $\bar{x}$	95.7780	90.1220	92.9500
Sum of squares $\sum x_i^2$	45,880.7139	40,768.6775	86,649.3914
Sample variance s^2	3.3969	39.7008	28.0407
Sample std. dev. s	1.8431	6.3009	5.2953
Std. dev. of mean	0.8242	2.8178	1.6745

The test is also performed by using studentized range distribution as well as p-values corresponding to an observed value of $Q_{i,j}$.

The statistics in Tables 9–11 concludes that even though the ABC-SVM is somewhat better accuracy than others, still it is not that much significant to outweighs the other, suggesting the effectiveness (equally significant) of our proposed approach.

From all the above discussions, we found the following:

Deep learning fails to perform better since it may over-fit when a complicated model is chosen to learn from an easy problem. Also, the Deep learning depends largely on the depth of the network; it is envisaged of working well in complex problems. It is well understood that direct Multi-class classification results in low accuracy in comparison to two-class classification. At the same time, it's a very challenging job to obtain high accuracy in the microarray dataset because of less number of samples in comparison number of attributes even after reduction. While comparing with others existing research, Statistical significance test along with post hoc tests inhibits deeper understanding to choose the best model out of many.

Table 9
Tukey HSD results for ESA-DNN and ABC-SVM.

Treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inference
ESA-DNN vs. ABC-SVM	2.7245	0.0902088	Insignificant

Table 10
Scheffé multiple comparison.

Treatments pair	Scheffé T-statistic	Scheffé p-value	Scheffé inference
ESA-DNN vs. ABC-SVM	1.9265	0.0902074	Insignificant

Table 8
One-way ANOVA of k = 2 independent treatments (ESA-DNN and ABC-SVM).

Source	Sum of squares SS	Degrees of freedom ν	Mean square MS	F statistic	p-value
Treatment	79.9758	1	79.9758	3.7114	0.0902
Error	172.3906	8	21.5488	–	–
Total	252.3664	9	–	–	–

Table 11
Bonferroni and Holm multiple comparison.

Treatments pair	Bonferroni and Holm T-statistic	Bonferroni p-value	Bonferroni inference	Holm p-value	Holm inference
ESA-DNN vs. ABC-SVM	1.9265	0.0902074	Insignificant	0.0902074	Insignificant

4. Conclusion and future scope

From the literature search and experiments conducted in this paper, it is understood that a reliable and detailed analysis is most essential to the success of the gene expression data analysis for cancer classification. As the gene expression datasets are complex ones, novel gene selection followed by an effective and efficient classification strategy needs utmost care for diagnosis of the disease. The result obtained opined that elephant search method could select most appropriate genes out many redundant genes present in the dataset. The most promising, very recent deep learning classification technique is then found promising ones with good accuracy. One way ANOVA and Post hoc Tukey HSD Test is used to check the suitability of the proposed approach. The result opines that our proposed approach is equally significant with the best of the method available in the literature so far. Further, the effectiveness of our proposed method is to be tested in the big dataset with a large number of samples along with a large number of attributes in future.

References

Tjaden, B., Cohen, J., 2006. A survey of computational methods used in microarray data interpretation. Appl. Mycology Biotechnol., Bioinf. 6, 7–18.

Sheng-Bo, G., Lyu, M.R., Lok, T.-M., 2006. Gene selection based on mutual information for the classification of multi-class cancer. In: Proceedings of the 2006 international conference on Computational Intelligence and Bioinformatics – Volume Part III, Ser. ICIC'06. Springer-Verlag, pp. 454–463.

Young, Y.T., 2009. Efficient multi-class cancer diagnosis algorithm, using a global similarity pattern. Comput. Stat. Data Anal. 53 (3), 756–765. https://doi.org/10.1016/j.csda.2008.08.028.

Alshamlan, Hala M., Badr, Ghada H., Alohali, Yousef, 2013. A study of cancer microarray gene expression profile: objectives and approaches. Proceedings of the World Congress on Engineering 2013 Vol II, WCE 2013, July 3–5, 2013, London, U.K.

Simon, R., 2009. Analysis of DNA microarray expression data. Best Practice Res. Clinical Hematol. 22 (2), 271–282.

Wang, L., Chu, F., Xie, W., 2007. Accurate cancer classification using expressions of very few genes. IEEE/ACM Trans. Comput. Biol. Bioinf. 4 (1), 40–53.

Dougherty, J., Kohavi, R., Sahami, M., 1995. Supervised and Unsupervised Discretization of continuous features. In: Machine Learning: Proceedings of the twelfth international conference. Morgan Kaufmann, pp. 194–202.

Platt, J.C., Cristianini, N., Shawe-Taylor, J., 2000. Large margin dags for multiclass classification. In: Advances in Neural Information Processing Systems. MIT Press, pp. 547–553.

Huang, H.-L., Chang, F.-L., 2007. ESVM: Evolutionary support vector machine for automatic feature selection and classification of microarray data. Biosystems 90 (2), 516–528.

El Akadi, A., Amine, A., El Ouardighi, A., Aboutajdine, D., 2009. A new gene selection approach based on minimum redundancy-maximum relevance (mrmr) and genetic algorithm (GA). In: IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2009, pp. 69–75.

Mao, Y., Zhou, X., Pi, D., Sun, Y., Wong, S.T.C., 2005. Multiclass cancer classification by using fuzzy support vector machine and binary decision tree with gene selection. J. Biomed. Biotechnol. 2 (8), 160–171.

Zhang, R., Huang, G.-B., Sundararajan, N., Saratchandran, P., 2007. Multicategory classification using an extreme learning machine for microarray gene expression cancer diagnosis. IEEE/ACM Trans. Comput. Biol. Bioinf. 4 (3), 485–495.

Zainuddin, Z., Pauline, O., 2009. Improved wavelet neural network for early diagnosis of cancer patients using microarray gene expression data. In: IJCNN 2009 International Joint Conference on Neural Networks, pp. 3485–3492.

Berrar, D., Downes, S., Dubitzky, W., 2003. Multiclass cancer classification using gene expression profiling and probabilistic neural networks. In: Pacific Symposium on Biocomputing, pp. 5–16.

Roland, L., Dawn, D., Holger, S., Dirk, T., Klaus, R., Siegfried, P., Mathias, W., 2004. The subsequent artificial neural network (sann) approach might bring more classificatory power to ann-based DNA microarray analyses. Bioinformatics 20 (18), 3544–3552.

Ghorai, S., Mukherjee, A., Sengupta, S., Dutta, P., 2010. "Multicategory cancer classification from gene expression data by multiclass nppc ensemble", in Systems in Medicine and Biology (ICSMB). International Conference on 2010, 4–48.

Kothandan, Ram, Biswas, Sumit, 2016. Comparison of Kernel and Decision Tree-Based Algorithms for Prediction of microRNAs Associated with Cancer. Curr. Bioinform. 11, 143–151.

Patrenahalli, N., Fukunaga, K., 1977. A branch and bound algorithm for feature subset selection. IEEE Trans. Comput. 26 (9), 917–922.

Liu, Huiqing, Li, Jinyan, Wong, Limsoon, 2005. Use of extreme patient samples for outcome prediction from gene expression data. Bioinformatics 21 (16), 3377–3384. Oxford University Press.

Zhu, Zexuan, Ong, Y.S., Dash, M., 2007. Markov blanket-Embedded Genetic Algorithm for gene selection. Pattern Recognition 49 (11), 3236–3248.

Eslami, Mahdiyeh, Shareef, Hussain, Khajehzadeh, Mohammad, 2013. Firefly algorithm and pattern search hybridized for global optimization. In: Huang, D.-S. (Ed.), ICIC 2013, LNAI 7996. Springer, pp. 172–178.

Yang, X.-S., 2010. Firefly algorithm, stochastic test functions and design optimisation. Int. J. Bio-Inspired Comput. 2 (2), 78–84.

Kora, Padmavathi, Sri Rama Krishna, K., 2016. Hybrid firefly and Particle Swarm Optimization algorithm for the detection of Bundle Branch Block. Int. J. Cardiovascular Acad. 2, 44–48. Elsevier.

Deb, Suash, Fong, Simon, Tian, Zhonghuan, 2015. Elephant search algorithm for optimization problems. In: The Tenth International Conference on Digital Information Management (ICDIM 2015). IEEE, pp. 249–255.

Seon-woo Min, Byunghan Lee, and Sungroh Yoon, Deep Learning in Bioinformatics, https://arxiv.org/ftp/arxiv/papers/1603/1603.06430.pdf

Matthew D. Zeiler, ADADELTA: An adaptive learning rate method, arXiv:1212.5701v1 [cs.LG] 22 Dec 2012

Diederik P. Kingma and Jimmy Lei Ba, ADAM: A method for stochastic optimization, arXiv:1412.6980v8 [cs.LG] 23 Jul 2015.

Vural, H., Subasi, A., 2015. Data Mining techniques to classify Microarray gene expression data using selection by SVD and information gain. Modeling Artificial Intelligence 6 (2), 171–182.

Glinsky, G.V., Glinski, A.B., Stephenson, A.J., Hoffman, R.M., Gerald, W.L., 2004. gene expression profiling predict clinical outcomes of prostate cancer. J. Clinical Investigation 113 (6), 913–923.

Mukkamala, S., Liu, Q., Veeraghattam, R., Sung, A.H., 2005. Computational Intelligent techniques for tumor classifications (using microarray gene expression data). International Journal of lateral computing 2 (1), 38–45.

Susmi, S.J., Khanna Nehemiah, H., Kannan, A., Christopher, J., 2015. Relevant gene selections and differentiation of Leukemia gene expression data. In: Sethy, N.R. (Ed.), Proc. of Emerging research in computing, information, communication and applications. Springer, pp. 503–510.

Yashoda, M., Ponmuthurammalingam, P., 2015. A microarray gene expressions with classifications using extreme learning machines. Genetica 47 (2), 523–534.

Alshaman, H.M., Badr, G.H., Aohali, Y.A., 2016. ABC-SVM: Artificial Bee Colony and SVM method for microarray gene selection and multiclass cancer classification. Int. J. Mach. Learning Comput. 6 (3), 184–190.

Qi, S., Shi, W.-M., Wei, K., Ye, B.-X., 2007. A combination of modified particle swarm optimization algorithm and support vector machine for gene selection and tumor classification. Adv. Comput. Sci. 71 (4), 157–162.

Javad, M., Giveki, D., 2013. Automatic detection of erythematosquamous diseases using pso–svm based on association rules. Eng. Appl. Artif. Intell. 26 (1), 603–608.

Peng, S., Xu, Q., Ling, X.B., Peng, X., Du, W., Chen, L., 2003. Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. FEBS Lett. 555 (2), 358–362.

Haynes, Winston, 2013. Tukey's Test. In: Dubitzky, Werner, Wolkenhauer, Olaf, Cho, Kwang-Hyun, Yokota, Hiroki (Eds.), Encyclopedia of Systems Biology. Springer, pp. 2303–2304.