Parallel Ensemble Feature Subset Selection based Deep Learning Approach for Imbalanced High Dimensional Cancer Datasets

Archana Shivdas Sumant[1]

Research Scholar, Compute Engineering Department, MET's Institute of Engineering,
Nashik Affiliated to Savitribai Phule Pune University,
Pune MAHARSHTRA 422003 INDIA
archana.vaidya@ges-coengg.org
https://orcid.org/0000-0002-6213-6216

Dipak V. Patil[2]

Professor and Head, Compute Engineering Department,
GES's R. H. Sapat College of Engineering Management Studies and Research,
Nashik, Maharashtra, 422005 INDIA
dipak.patil@ges-coengg.org
https://orcid.org/0000-0002-9612-8564

Abstract

Cancer is currently one of the most threatening diseases. The analysis of high dimensional microarray data is a method that is frequently used for cancer diagnosis. Multi category imbalance data analysis in the medical sciences plays crucial role. The growing number of cancer patients has made the necessity for cancer classification absolutely necessary. In this study, PE-SU-R-CNN and PE-ChS-R-CNN are employed as parallel ensemble feature selection based deep learning classifiers to address issues for multi-class cancer classification. The deep learning convolutional neural network with soft-max activation function at decision layer is used. The most well known nine cancer microarray gene expression datasets are used in the experiments. The proposed system's empirical findings are compared with deep learning techniques based on bio-inspired feature selection firefly and elephant searches. Finally, a one-way ANOVA statistical significance test with a post hoc Tukey's test is used to draw several inferences about the optimal classifier model to use. The proposed parallel model shows significant improvement over existing methods.

Keywords: High dimensional datasets; imbalance classification; ensemble feature selection; deep learning; parallel computing.

I. INTRODUCTION

The significant challenge that could impact machine learning task like classification performance is dimensionality. Numerous applications in the sciences and engineering, including those in the domains of health, biology, business, sensor network, rely on high dimensional datasets containing hundreds or even thousands of features, some of which are irrelevant, redundant, or noisy [1]. The purpose of feature selection is to reduce irrelevant features from high dimensional datasets. Ensemble feature selections [2-5] are popular due to their accuracy and stability [6, 7]. Deep Learning [8-14] is a recent field of study in machine learning that has had remarkable success in high-level data abstraction and representation. Instead of using a single shallow "fat" structure, it employs numerous levels of non-linear operations to address the relevant machine learning problems.

Most classifier learning techniques, which assume a generally balanced distribution, have found significant difficulties when dealing with data sets with imbalanced class distributions [15-19]. The imbalance data is described as as having much more examples of some classes than others like rare disease. Despite the rarity of these occurrences, classification systems that foresee the small classes tend to be uncommon, unknown, or unnoticed; hence, test samples from the minor classes are misclassified more frequently than those in the majority classes. The proper classification of such samples instance is important in some of the applications like cancer prediction.

In this study, we present a parallel ensemble feature selection for feature reduction for high dimensional imbalanced cancer datasets followed by deep learning frameworks recurrent neural network (RNN) and convolution neural network (CNN). More specifically, we are interested in parallel an asynchronous algorithm for feature selection, which overcomes the restrictions placed by synchronous barriers. Each processing unit is responsible for finding the best features from each vertical partitioned data. Then combining these selected subsets into the final optimal subset. To apply this strategy to the ensemble feature subset selection (EFSS) algorithm, we must first break the general approach's intra-task dependency. The main contribution of this paper is parallel asynchronous approach for EFSS and deep learning RNN and CNN classifier at prediction stage.

The following is how the rest of the paper is organized: We review the current state of the art in deep learning and EFSS parallel approaches in Sect. 2; we present our parallel EFSS method followed by deep learning RNN and CNN approaches for high dimensional cancer datasets in Sect. 3; we test the speedup of our strategies with a sequential approach in Sects. 4 and present and discuss the results; and we present and discuss the conclusions and future work in Sect. 5.

II. LITERATURE REVIEW

A deep learning method stacked denoising autoencoder [20] is used to extract the key correlations between gene expressions of breast cancer. After training, the layer with the lowest validation error and lowest dimension when compared to other encoder stacks is chosen. This validation data set is separate from the training and test sets. ANN , SVM and SVM-RBF are used in training phase.

Kumar, Ansuman, and Anindya Halder.[21] proposed active learning based fuzzy rough set based classifier for microarray cancer data. To enhance performance, unlabeled informative samples are incrementally added to the training set using Gauss-Seidel algorithm. A test pattern's fuzzy membership value for each class is assigned via the fuzzy k-NN algorithm. The algorithm determines which test pattern's class label information has the highest fuzzy membership value based on the final larger training set for a certain class.

PCC-DTCV [22] is a hybrid feature selection process that uses Pearson's correlation coefficients (PCC) for feature reduction and Decision Trees (DT) as classification methods. Grid Search CV is used to optimize the settings for DT (max-depth tuning) and to choose the best feature subset. The best performance is obtained when PCC > 0.4. Accuracy is seen to decrease when PCC value increases to 0.5 and 0.6.

In [23], a two-hybrid technique based on CNN and Relief-F was developed for the classification of cancer microarray datasets. To eliminate noisy features, a cascaded auto encoded deep neural network for feature extraction is used. Softmax function is employed for categorization in the last stage. Auto encoders have the drawback of being slower than other versions. The first method reduces dimensions using ReliefF, whereas the second method reduces features using auto encoders. For classification, SVM and CNN are employed. It has been shown that using the CNN model with reliefF for dimensional reduction produces the greatest results.

Lin et al. [24] proposed two deep learning models: feature based and image based for identification of cancer subtypes. Optimal CNN model with three CNN and three dense layers with ReLU activation function are designed. The dataset has a size of 2q with n samples, where q is the number of genes and w is the filter's width. The first layer's filter has a size of 2w.

Taking everything said above into account, it is clear that a successful gene selection strategy is required, together with unique methodologies and the creation of a promising classifier for more accurate gene prediction that is acceptable. This inspired us to conduct additional research on employing a innovative parallel ensemble feature selection followed by deep neural network-based better classifier for diverse microarray with multi class classification datasets.

Even though many studies have advocated using either filter-based, wrapper-based, or a hybrid of these two to find a subset of most informative genes for better clinical diagnosis, there is still much to be done in terms of performance with ensemble feature selection methods for gaining new insights into the clinical diagnosis,. For predictive accuracy, choosing the best gene from gene expression profiles is particularly difficult since gene selection is NP-hard [25].

We are thus motivated to address this issue by applying a parallel ensemble-based feature selection deep learning classifier. For further testing of the effectiveness of our suggested approach, we compared it to other established state-of-the-art deep learning approaches. A comparison with similar work by other authors concludes by supporting our proposal.

III. PROPOSED SYSTEMS

The datasets and methodology used in this paper are discussed in this section.

3.1 DATASET USED

For the proposed research in this paper, we make use of publically available microarray datasets [26][27] and are listed in table 1. There are n samples, p features, and C_k classes. These all are high-dimensional cancer micro-array datasets. The n/p ratio is much lower than p, with values ranging from 0.006 to 0.036. The dataset is regarded as high dimensional and statistical analysis of it is difficult when the n/p ratio is smaller than 1 [28]. The binary datasets are Prostate, Leukemia, COLON, DLBCL, Ovarian and CNS. MLL, SRBCT and Lung are multiclass datasets. Table 2 gives class wise distribution of samples for each datasets. The datasets details are as follows:

Prostate: The Prostate dataset has two classes of normal has 50 samples and tumor has 52 samples.
Leukemia: This dataset includes bone marrow samples that were obtained using 7129 markers from 6817 human genes, of which 34 samples (20-ALL and 14 AML) were used for testing and 38 samples (27 ALL and 11 AML) were used for training.

Dataset Name	C_k	Total number of Samples n	Total number of Features p	n/p ratio for high dimensionality
Prostate	2	102	12600	0.008
Leukemia	2	72	7129	0.01
COLON	2	62	2000	0.031
DLBCL	2	47	4027	0.012
Lung	5	203	12600	0.016
Ovarian	2	253	15155	0.017
MLL	3	72	12582	0.006
SRBCT	4	83	2308	0.036
CNS	2	60	7129	0.008

Table 1. Datasets details description used for experimentation with n/p ratio

Dataset Name	C_k	Total number of Samples n	No. of Samples in each class				
			C_1	C_2	C_3	C_4	C_5
Prostate	2	102	52	50	-	-	-
Leukemia	2	72	47	25	-	-	-
COLON	2	62	40	22	-	-	-
DLBCL	2	47	24	23	-	-	-
Lung	5	203	139	21	20	6	17
Ovarian	2	253	162	91	-	-	-
MLL	3	72	24	20	28	-	-
SRBCT	4	83	23	28	12	20	-
CNS	2	60	21	39	-	-	-

Table 2. Datasets with class wise distribution of samples

COLON: There are 62 samples in this dataset, of which 22 are positive and 40 are negative (tumor biopsies are taken from tumor).

DLBCL: There are 47 samples in the diffuse large B-cell lymphoma (DLBCL) data collection, 24 of which are from the "germinal centre B-like" group and 23 from the "activated B-like" group. Each sample is represented by the expression of 4026 genes.

Lung: There are five different classes in the multi-class Lung dataset for categorization. This contains 203 samples of lung tumors, each containing 12,600 genes, including 139 samples of lung adenocarcinomas (labeled as ADEN), 21 samples of squamous cell lung carcinomas, 20 samples of pulmonary carcinoids, 6 samples of small-cell lung carcinomas, and 17 samples of normal lung tissue (labeled as NORMAL).

Ovarian: The ovarian dataset aids in our comprehension of the circumstance in determining whether or not the proteomic patterns in serum indicate the presence of ovarian cancer symptoms. The women with similar family histories are those who are most affected by the cancer. The dataset was created using 253 samples (162 ovarian and 91 normal), which were used to create the proteome spectra.

MLL : The MLL dataset included 28, 20, and 24 samples of each of the three classifications of leukemia: acute myeloid leukemia (AML), myeloid lymphoid leukemia (MLL), and lymphoblastic leukemia (ALL).

SRBCT: The SRBCT sample had four different cancer types. The four classes had samples of 23, 28, 12, and 20 for Ewing sarcoma (EWS), non-Hodgkin lymphoma (NHL), neuroblastoma (NB), and rhabdomyosarcoma (RMS).

CNS: The central nervous system (CNS) dataset shows the prognosis of patients with embryonal tumours. This has a total of 60 samples with 7129 genes, of which 21 are survivors and 39 are failures.

3.2 THE METHODOLOGY

The parallel feature selection and deep learning classification techniques used in this paper are presented in this section.

For dimension reduction in a large dataset, the feature selection is of extreme significance. The use of the fewest effective features from the whole feature set in the dataset may result in a fast solution with appropriate accuracy. For supervised learning tasks, great generalization ability can be attained via ensemble approaches [29] [30]. The challenges of parallelizing computing techniques fall into two categories. (1) The high-dimensional nature of the cancer microarray dataset. (2). Parallel algorithms speed up processing, but at the cost of decreased prediction accuracy [31]. These two challenges are the focus of this study work, which suggests the following solutions. (1) To make the process of data parallelism easier, initial phase is applied on complete data and in the second phase vertically dividing them along the features. (2) Using the Parallel Ensemble Feature Selection (PE-FS) algorithm, the best and most significant features are chosen to improve the categorization of cancer subtypes.

The methods used for high dimensional cancer data classification is shown in Figure 1. Here, Parallel Ensemble Feature selection (PE-FS) SU-R is used in the initial stage of feature selection. Convolutional neural network (CNN) deep learning training and testing are carried out on a reduced feature set X that was chosen in the initial stage. In first stage symmetric uncertainty measure is calculated as per equation 1. Symmetric uncertainty (SU) is a normalized value measure of Mutual Information (MI) obtained using the formula in Eq (2)

$$SU(X,Y) = \frac{2*MI(X,Y)}{H(x)+H(y)} \qquad \textbf{(1)}$$

As per information theory mutual information MI (X; Y) is the level of uncertainty in X resulting from the knowledge of Y [31]. Mutual information in mathematics is given by equation 2.

$$MI(X,Y) = \sum_x \quad \sum_y p(X,Y) log \frac{p(X,Y)}{p(X)*p(Y)} \qquad \textbf{(2)}$$

Here the joint probability distribution function is denoted by P(x, y). P(x) and P(y) are the marginal probabilities of X and Y distributional functions respectively. We can also state MI as per equation 3.

$$MI(X,Y) = H(x) - H(x,y) \qquad \textbf{(3)}$$

Where the conditional entropy H(X|Y) and the marginal entropy H(X) H(X; Y) is the combined entropy of X and Y. If H(X) here denotes the level of ambiguity regarding H(X|Y) quantifies what Y does not speak about X when Y is a random variable. The level of uncertainty in X following understanding Y, which supports the intuitive significance of mutual information as the total sum of knowledge e ach variable tells us something about the other. Marginal entropy H(x) can be defined as equation 4.

$$H(x) = -\int p(X) \log\big(p(X)\big)\, dx \qquad \textbf{(4)}$$

Given that P(X, Y) and P(Y|X) are joint and conditional probability distributions, respectively, and that X and Y are two discrete random variables, the conditional entropy associated with these distributions is defined as per equation 5.

$$H(Y|\, X) = -\sum_{x\in X} \quad \sum_{y\in Y} P(x,y) \quad log_2\, P\,(y|x) \qquad \textbf{(5)}$$

After calculation of SU score as per equation 1 , the non zero positive score features are taken for next step. These features are divided into ℓ vertical partitions. Here the value of ℓ is set to five vertical partitions. Relief score as per equation 6 is computed parallel on each vertical partition.

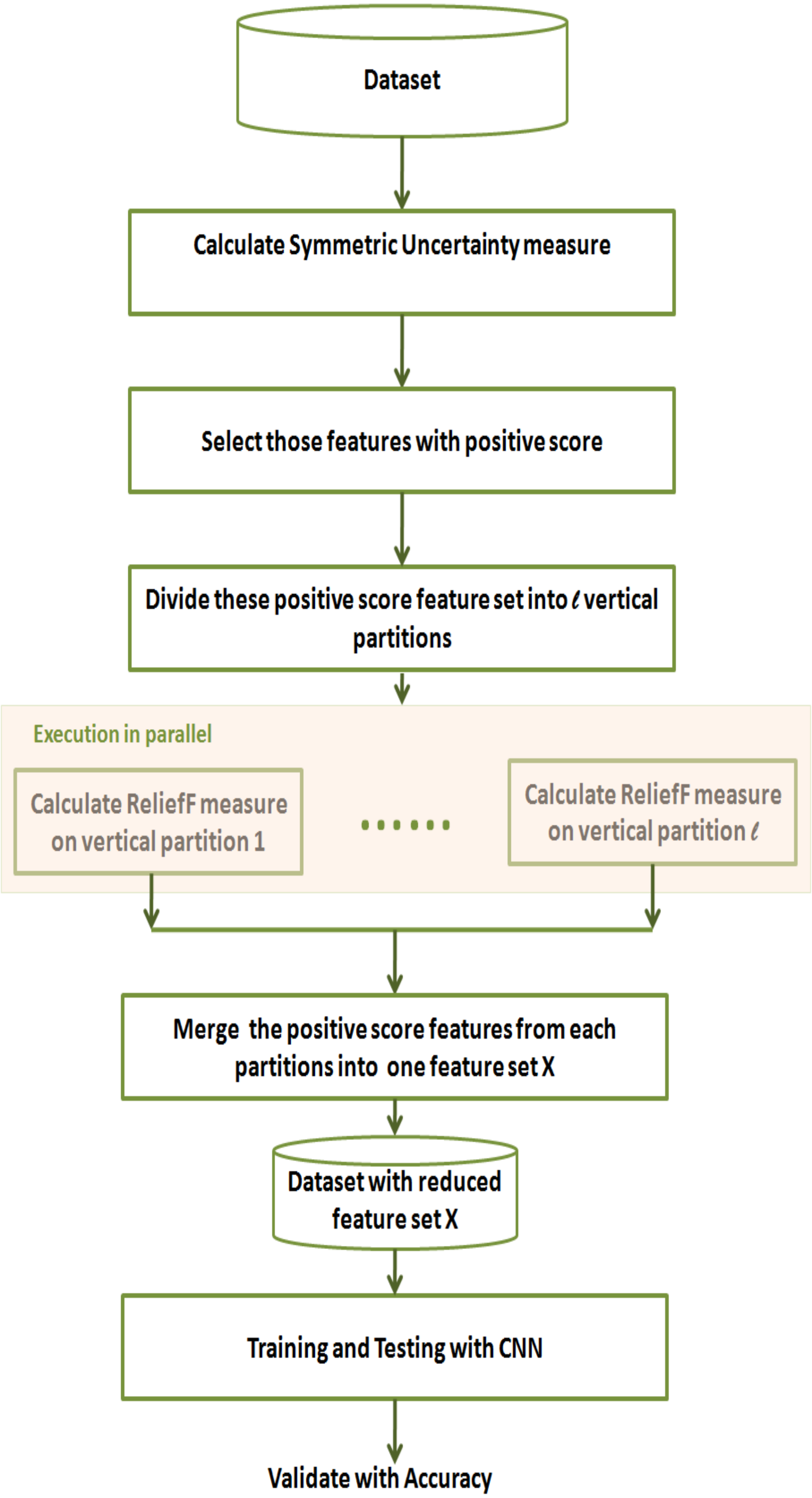

Dataset
Calculate Symmetric Uncertainty measure
Select those features with positive score
Divide these positive score feature set into ℓ vertical partitions
Execution in parallel
Calculate ReliefF measure on vertical partition 1
......
Calculate ReliefF measure on vertical partition ℓ
Merge the positive score features from each partitions into one feature set X
Dataset with reduced feature set X
Training and Testing with CNN
Validate with Accuracy

Fig 1. Parallel Ensemble Feature selection (PE-FS) SU-R followed by deep learning CNN approach for high dimensional cancer datasets

$$W = (W - diff\left(\mathrm{x_{ij}}, \mathrm{near_{hit_{ij}}}\right)^{2} + diff\left(\mathrm{x_{ij}}, \mathrm{near_{miss_{ij}}}\right)^{2})/\mathrm{m} \qquad \mathbf{(6)}$$

The number of sample is j, while the number of features is i and m is neighbor count. The ij samples are used to calculate near hits (NH) and near misses (NM). While determining the relief score, the feature Z is considered. The sequential time complexity of relief score computation is O(i*j*m) [32]. For high dimensional dataset with large number of features relief score computation takes more time. To reduce the time complexity of relief score computation we applied parallel approach with ℓ vertical partitions. The positive relief score feature from each vertical partitions are merged to form final feature set.

In our study [33], we discuss the significance of the SU-R and ChS-R [33] ensemble feature selection methods. For high dimensional data, the two techniques mentioned in this paper perform well. In [34], the stability analysis of the same methodology is covered. When choosing features, stability is a significant factor. The stability of feature selection becomes crucial to provide reliable classification results because even minor variations in data samples might radically alter feature selection. The parallel version of the same is devised here to deal with high dimensional dataset.

One of the most popular deep learning architecture is a convolutional neural networks (CNN) [35,36]. After final feature subset from PE-FS phase training and testing is performed with CNN. CNN uses convolution layer and pooling layer to extract high-level features from 2D input, and integrate them. The biological model of mammalian visual systems is the one by which the design of the CNN is motivated by Hubel, H., et al.[37] where the minute cells that make up the cat's visual cortex are honed region of the visual field known as the receptive field.

A convolutional neural network has three different types of layers: convolutional layers, pooling layers, and fully connected layers. Feature maps and filters are the components of convolutional layers. The neurons of the layer are essentially the filters. They provide an output value like a neuron and have weighted inputs. A patch or receptive field is a fixed-size square that serves as the input size. The result of one filter applied to the preceding layer is the feature map. The pooling layers feature map is down sampled from the earlier layers. Following one or more convolutional layers, pooling layers are used to consolidate the learned and represented features in the feature map from the preceding layers. As a result, pooling may be thought of as a technique, to reduce the model's over fitting of the training data and generally compress or generalized feature representations. The typical flat feed-forward neural network layer is made up of fully connected layers, the layers could either a softmax activation or a nonlinear activation function to produce class prediction probabilities. After the convolutional and pooling layers have completed their work of feature extraction and consolidation, fully connected layers are employed to complete the network. They are combined to produce final nonlinear features, which the network then uses to make predictions.

Hyper parameters	Model
Epoch	200
Momentum	0.9
Learn-Rate	0.001
Batch-size	64
Optimizer	SGD
Activation Function	ReLU
Output Function	Softmax

Table 3 Experimental hyper parameters of CNN model

Table 3 shows experimental hyper parameters used for setting up our CNN model. Our CNN model has two fully connected layers, two pooling layers, and two convolutional layers. The input layer of the network includes neurons according to the amount of the input data. To reduce the size of the feature map by 85%, down sampling was accomplished using a concatenated average and maximum pooling method. The activation function in a neural network is in control of converting the node's summed weighted input into the activation of the node or output for that input. The rectified linear (ReLU) activation function is a piecewise linear function that outputs zero if input is 0.0 or neagtive otherwise the input directly. The vanishing gradient issue is solved by the rectified linear activation function, which enables models to learn more quickly and perform better.

The rectified linear unit (ReLU), which facilitates training of deep neural networks by stabilizing gradients on backpropagation, is used for all nonlinear functions [38]. In order to improve network training by stabilizing the loss landscape, batch normalization was also applied between the convolutional and ReLU layers [39, 40].

$$Loss = -\sum_{i=1}^{Output\ Size} x_i \log \hat{x}_i \qquad \textbf{(7)}$$

The error rate must be calculated in order to calculate the model performance. We utilized the categorical cross-entropy cost function as a loss function generated by equation 7 because this is a multiclass classification problem [41].

Here output size is the number of scalar values in the model output, $\hat{x}_i$ is the i-th scalar value in the model output, x_i is the associated target value. This loss is an excellent indicator of how easily two discrete probability distributions may be distinguished from one another. In this situation, x_i denotes the likelihood that event i will take place, and the total of all x_i equals 1, indicating that precisely one event might happen.

The minus sign makes sure that when the distributions approach one another, the loss decreases.

An iterative technique for optimizing an objective function with acceptable smoothness qualities is stochastic gradient descent, or SGD [42]. Since it uses an estimate of the gradient instead of the actual gradient (derived from the whole data set), it can be thought of as a stochastic approximation of gradient descent optimization (calculated from a randomly selected subset of the data). This minimizes the extremely high computational burden, especially in high-dimensional optimization problems, allowing for faster iterations at the cost of a reduced convergence rate.

On the output layer, a softmax activation function σ as per equation 8 is employed to transform the outputs into probabilistic values and enable the model's prediction output to be chosen from among several classes.

$$\sigma(\vec{Y})_i = \frac{e^{Y_i}}{\sum_{j=1}^{k} e^{Y_j}} \qquad \textbf{(8)}$$

The input vector is represented as$(\vec{Y})$. Here e^{Y_i} is standard exponential function for input vector, e^{Y_j} is standard exponential function for output vector and k is number of classes.

The second proposed method PE-FS ChS-R followed by deep learning classifier CNN works same as shown in Figure 1. Here at first step chi-squared score with equation 9 is computed to select positive score features from entire feature set.

$$\chi 2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - \mathrm{E_{ij}} \right)^2}{\mathrm{Eij}} \qquad \textbf{(9)}$$

Here E_{ij} is the observed value, while O_{ij} is the predicted value. It establishes a meaningful connection between two feature vectors. This technique evaluates a feature that predicts class relationships and computes a score. Chi-squared attribute assessment determines a feature's value by computing the Chi-squared statistic's value in relation to the class. H_0 is the initial hypotheses, and presumption that there is no connection between the two features and it is tested Chi-squared test. Then after selecting positive scored features from the first step with chi-squared score are divided into ℓ vertical partitions. A parallel relief with equation 6 is computed on each ℓ vertical partitions. A positive score features are merged into single set for deep learning training and testing with CNN.

IV. RESULT AND DISCUSSION

The two devised systems parallel ensemble feature selection SU-R followed by deep learning CNN approach (PE-SU-R-CNN) and parallel ensemble feature selection ChS-R followed by deep learning CNN approach (PE-ChS-R-CNN) results are discussed in this section.

The two methods used for comparison are Firefly search and elephant search for feature selection, followed by a deep neural network built on a stochastic gradient descent technique with a soft max activation function [43].

The Elephant Search Algorithm is a highly nonlinear, multimodal global optimization method was developed in response to the biological behaviors of elephant herds. The intensification of the Elephant search in the local search space is regarded as an effective optimization approach for improved solutions. It offers universally optimal solutions while adequately spanning the search space and avoiding local optima.

Firefly search is a recent population-based global optimization technique that functions by imitating the flashing pattern of fireflies. There are about 2000 different types of firefly species, which are tiny insects with the ability to flash brief, rhythmic lights that draw other fireflies. The firefly can only be seen for a few hundred meters or so due to the fact that the light intensity attraction diminishes with increasing distance. The fluorescence light behavior of fireflies is related to the objective function utilized in the algorithm. If there are no fireflies that are brighter than it or the one it is following, the fireflies move randomly.

To measure system performance accuracy is calculated as equation 9. Confusion matrix for classification task is calculated as in table 4.

$$\%\boldsymbol{Accuracy} = 100 * (\mathrm{TN} + \mathrm{TP})/(\mathrm{TN} + \mathrm{TP} + \mathrm{FN} + \mathrm{FP}) \qquad \textbf{(10)}$$

Class	Actual :No	Actual :Yes
Predicted :No	True Negative **(TN)**	False Positive **(FP)**
Predicted :Yes	False Negative **(FN)**	True Positive **(TP)**

Table 4 Confusion matrix for accuracy calculation

.

Table 5 shows proposed parallel ensemble approaches followed by deep neural CNN model performance obtained. The proposed system performance is compared with existing two methods Firefly with deep learning and elephant search with deep learning. To compare proposed system accuracy with existing system accuracy we have calculated Δ as per equation 11.

$$\Delta_{\mathbf{i}} = \ Accuracy\ of\ proposed\ system - \%Accuracy\ of\ existing\ system \qquad \textbf{(11)}$$

Δ_1 is computed to show improvement of proposed PE-SU-R-CNN with existing firefly with deep learning approach. The average improvement is 8.65 percentages.

Δ_2 is computed to show improvement of proposed PE-SU-R-CNN with existing elephant search with deep learning approach. 10.25 percentages are improved on average by PE-SU-R-CNN.

Δ_3 is computed to show improvement of proposed PE-ChS-R-CNN with existing firefly with deep learning approach. It is observed that 7.97 average percentage improvement in performance by PE-ChS-R-CNN over firefly + deep learning approach.

Δ_4 is computed to show improvement of proposed PE-ChS-R-CNN with existing elephant search with deep learning approach. We observe improvements of 9.57 average percentages here.

This analysis reveals that the suggested system outperforms the current elephant search + deep learning strategy and that PE-SU-R-CNN outperforms PE-ChS-R-CNN.

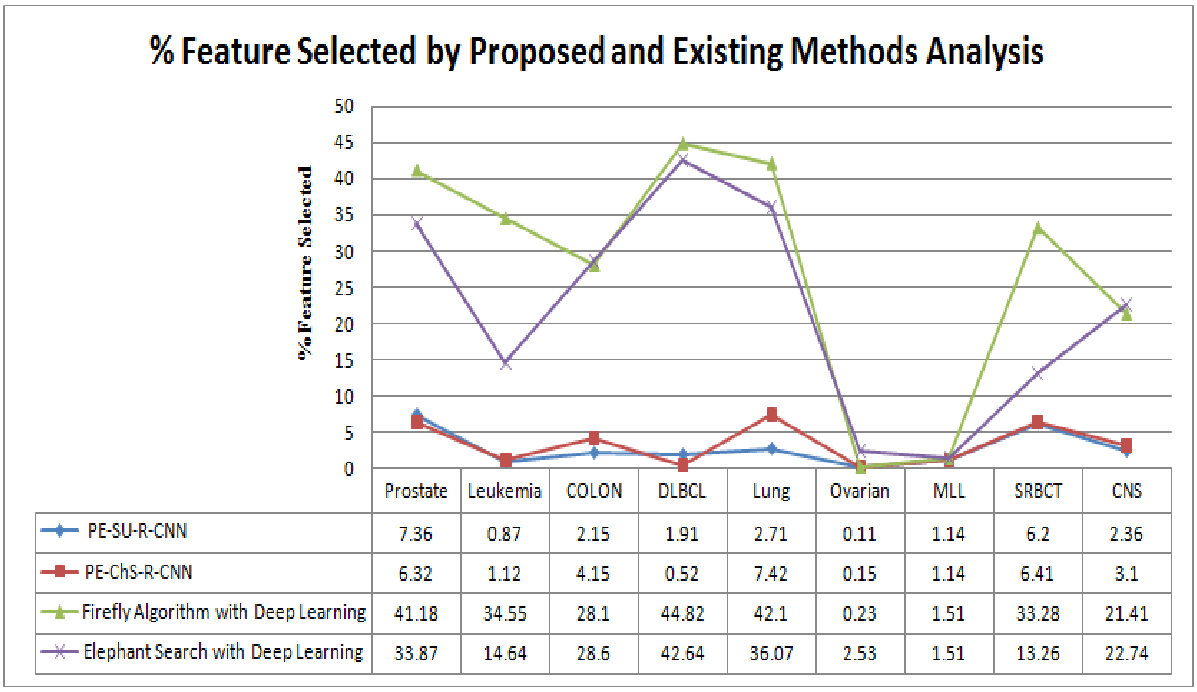
It is observed that practically all datasets benefit from deep learning. The proposed system improvement over existing firefly+deep learning and elephant search + deep learning approach shows that parallel ensemble feature selection outer perform in feature selection than existing bio inspired feature selection firefly and elephant search for high dimensional data. It is also observed that less number of features is selected by parallel ensemble feature selection methods for all datasets. When compared system accuracy it is observed that proposed systems outer perform in almost all cases except elephant search+ deep learning achieves 0.3 percent improvement over PE-ChS-R-CNN approach. The highest improvement observed is with CNS dataset. The lowest improvement is observed with ovarian dataset.

According to Tables 5, the proposed PE-SU-R-CNN is the fastest algorithm across all datasets and offers the best accuracy in the datasets for ovarian, leukemia, colon, and DLBCL cancer. It also offers adequate accuracy for other datasets. For each dataset, the top-performing approach is indicated with bold. The only exception is the leukemia dataset, where PE-SU-R-CNN and Firefly+ deep learning are tied for accuracy.

					Proposed methods				Existing Methods				Improvement of PE-SU-R-CNN		Improvement of PE-ChS-R-CNN	
Sr. No.	Dataset Name	C_k	Total number of Samples n	Total number of features p	PE-SU-R-CNN		PE-ChS-R-CNN		Firefly Algorithm with Deep Learning (A)		Elephant Search Based Algorithm with Deep Learning (B)		Compared with approach (A)	Compared with approach (B)	Compared with approach (A)	Compared with approach (B)
					NoF Selected	Accuracy	NoF Selected	Accuracy	NoF Selected	Accuracy	NoF Selected	Accuracy	Δ_1	Δ_2	Δ_3	Δ_4
1	Prostate	2	102	12600	927	88.42	796	100	5189	87.26	4267	88.24	1.16	0.18	12.74	11.76
2	Leukemia	2	72	7129	**62**	**100**	80	99.7	2463	**100**	1044	92.11	0	7.89	-0.3	7.59
3	COLON	2	62	2000	**43**	**100**	83	**85.79**	562	77.43	572	79.03	22.57	20.97	8.36	6.76
4	DLBCL	2	47	4027	77	**100**	**21**	**100**	1805	89.36	1717	91.49	10.64	8.51	10.64	8.51
5	Lung	5	203	12600	**342**	98.7	935	**99.8**	5304	93.11	4545	94.1	5.59	4.6	6.69	5.7
6	Ovarian	2	253	15155	**16**	**100**	23	**100**	35	97.24	384	99.21	2.76	0.79	2.76	0.79
7	MLL	3	72	12582	**144**	**86.5**	**144**	**83.33**	190	80.56	190	80.56	5.94	5.94	2.77	2.77
8	SRBCT	4	83	2308	**143**	**96.5**	148	95.4	768	93.98	306	83.14	2.52	13.36	1.42	12.26
9	CNS	2	60	7129	**168**	**83.33**	221	**83.33**	1526	56.67	1621	53.34	26.66	29.99	26.66	29.99
	Average				**214**	**94.83**	**272**	**94.15**	**1982**	**86.18**	**1627**	**84.58**	**8.65**	**10.25**	**7.97**	**9.57**

Table 5. Proposed Parallel Ensemble Feature Selection approaches with deep learning classifier compared with existing Firefly and Elephant Search Based Algorithm with Deep Learning Classifier. Number of features selected by each algorithm is NoF.

*** NoF is Number of Features**



	Prostate	Leukemia	COLON	DLBCL	Lung	Ovarian	MLL	SRBCT	CNS
PE-SU-R-CNN	7.36	0.87	2.15	1.91	2.71	0.11	1.14	6.2	2.36
PE-ChS-R-CNN	6.32	1.12	4.15	0.52	7.42	0.15	1.14	6.41	3.1
Firefly Algorithm with Deep Learning	41.18	34.55	28.1	44.82	42.1	0.23	1.51	33.28	21.41
Elephant Search with Deep Learning	33.87	14.64	28.6	42.64	36.07	2.53	1.51	13.26	22.74

Fig 2. Number of Features selected percentage of proposed PE-SU-R-CNN, PE-ChS-R-CNN and Existing Firefly and Elephant search with deep learning approach.

Figure 2 shows the percentage of selected features by proposed PE-SU-R-CNN, PE-ChS-R-CNN, and existing firefly and elephant searches using deep learning. It can be seen from this analysis and the accuracy in table 5 that our suggested method produces better accuracy with fewer features than existing systems.

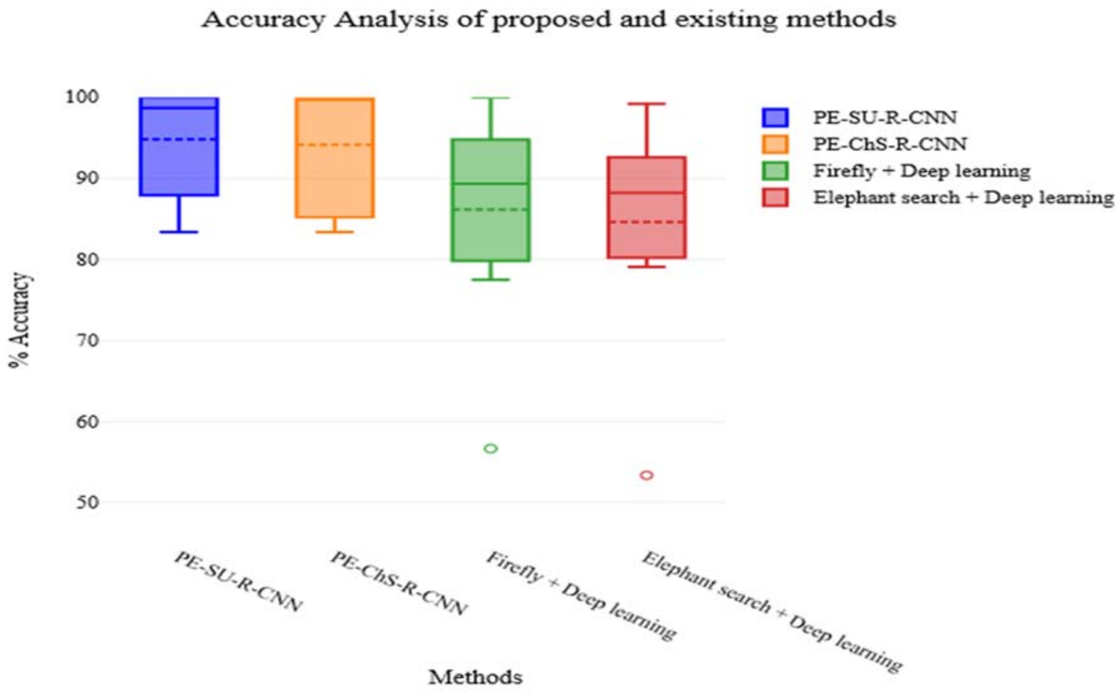


Fig 3. Accuracy analysis using box plot of proposed PE-SU-R-CNN, PE-ChS-R-CNN and Existing Firefly, Elephant search with deep learning approach.

In summary, Fig. 3 shows a graphical presentation using a box plot of the results of the Proposed PE-SU-R-CNN and PE-ChS-R-CNN technique with existing firefly +deep learning and elephant search+ deep learning as emphasized in Table 5 in terms of accuracy. According to Fig. 3, proposed PE-SU-R-CNN and PE-ChS-R-CNN appears to perform better than existing systems.

It is common knowledge that the feature selection procedure affects a deep learning model's accuracy. Compared to two-class classification, accuracy for multi-class problems is low. Because there are fewer samples in the microarray dataset than there are attributes, even after reduction, it can be exceedingly difficult to achieve high accuracy. The statistical significance test and post hoc tests prevent a deeper understanding and the selection of the best model out of several options when compared with other previous studies.
We validated our system using one-way ANOVA with post hoc Tukey HSD (Honest Significant Difference) test to have the statistical significance test in order to confirm the experimental data and to come to the correct conclusion. The next section discusses the findings from the statistical significance tests.

4.1 CONCLUSIONS FROM SIGNIFICANCE TEST

The importance of our proposed approaches is demonstrated by the results of a one-way ANOVA with post-hoc Tukey HSD Calculator, Scheffé, Bonferroni, and Holm multiple comparisons[44][45], which are presented in tables 6 through 10.

Treatment	Proposed Method	Existing Method	Pooled Sum
Observations N	16	16	32
Sum ΣX_i	1,500.80	1,339.59	2,840.39
Mean	93.8	83.7244	88.7622
$\sum X^2$	141,547.17	114,629.51	256,176.68
sample variance s^2	51.4751	164.8783	130.8854
sample std. dev. s	7.1746	12.8405	11.4405
std. dev. of mean	1.7937	3.2101	2.0224

Table 6 Statistical significance test calculations

source	sum of squares SS	degrees of freedom ν	mean square MS	F- statistic	p-value
Treatment	812.1458	1	812.1458	7.5076	0.0102
Error	3,245.30	30	108.1767		
Total	4,057.45	31			

Table 7 Statistical significance test with One-way ANOVA k=2

The p-value corresponding to the F-statistic of one-way ANOVA is lower than 0.05 as shown in table 7, suggesting that the one or more treatments are significantly different.

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inference
Proposed Vs Existing	3.8749	0.010246	**Significant**

Table 8 Tukey HSD results

treatments pair	Scheffé TT-statistic	Scheffé p-value	Scheffé inferfence
Proposed Vs Existing	2.74	0.0102454	**Significant**

Table 9 Scheffé test results

treatments pair	Bonferroni and Holm TT-statistic	Bonferroni p-value	Bonferroni inference	Holm p-value	Holm inference
Proposed Vs Existing	2.74	0.0102454	**Significant**	0.0102454	**Significant**

Table 10 Bonferroni and Holm results: all pairs simultaneously compared

The test shows the results are significant and from this we can prove significance of our proposed system.

V. CONCLUSIONS AND FUTURE SCOPE

If relevant features were chosen from a high-dimensional dataset during the feature selection phase, deep learning would perform better on the cancer microarray dataset. Additionally, because deep learning is expected to perform well with complicated issues, the network's depth plays a significant role in how it works. According to the results, PE-SU-R and PE-ChS-R ensemble feature selection were able to choose the most suitable genes from among the many redundant genes included in the dataset. The most promising, most recent deep learning classification method, CNN, is then identified as one of the most accurate and promising. The appropriateness of the proposed approaches is examined using an ANOVA and a post hoc Tukey HSD test, among other methods. The findings suggest that our suggested strategy is on par with the most effective approaches currently described in the literature. In the future, proposed approaches will be used to test the usefulness of our suggested strategy for high sample and also for less features datasets.

Declarations

Conflicts of interest The author's declare that they have no conflict of interest.

References

[1] Miao, Jianyu, and Lingfeng Niu. "A survey on feature selection." *Procedia Computer Science* 91 (2016): 919-926.

[2] Chandralekha, M., and N. Shebagavadivu. "An improved tree model based on ensemble feature selection for classification." *Turkish Journal of Electrical Engineering and Computer Science* 27.2 (2019): 1290-1307. https://doi.org/10.3906/elk-1808-85

[3] Hoque, Nazrul, Mihir Singh, and Dhruba K. Bhattacharyya. "EFS-MI: an ensemble feature selection method for classification." *Complex & Intelligent Systems* 4.2 (2018): 105-118. https://doi.org/10.1007/s40747-017-0060-x

[4] Tsymbal, Alexey, Seppo Puuronen, and David W. Patterson. "Ensemble feature selection with the simple Bayesian classification." *Information fusion* 4.2 (2003): 87-100. https://doi.org/10.1016/S1566-2535(03)00004-6

[5] Ben Brahim, Afef, and Mohamed Limam. "Ensemble feature selection for high dimensional data: a new method and a comparative study." *Advances in Data Analysis and Classification* 12.4 (2018): 937-952. https://doi.org/10.1007/s11634-017-0285-y.

[6] Salman, Reem, Ayman Alzaatreh, and Hana Sulieman. "The stability of different aggregation techniques in ensemble feature selection." *Journal of Big Data* 9.1 (2022): 1-23. https://doi.org/10.1186/s40537-022-00607-1

[7] Yang, Pengyi, et al. "Stability of feature selection algorithms and ensemble feature selection methods in bioinformatics." *Biological Knowledge Discovery Handbook* (2013): 333-352. https://doi.org/10.1002/9781118617151.ch14

[8] Novitasari, D. C. R., P. Wulandari, and D. Z. Haq. "Cervical Cancer Diagnosis System Using Convolutional Neural Network ResidualNet". *International Journal of Computing*, vol. 21, no. 1, Mar. 2022, pp. 61-68, doi:10.47839/ijc.21.1.2518.

[9] Wu, Zebin, et al. "Deep learning for classification of pediatric otitis media." *The Laryngoscope* 131.7 (2021): E2344-E2351. https://doi.org/10.1002/lary.29302.

[10] Hamolia, V., V. Melnyk, P. Zhezhnych, and A. Shilinh. "Intrusion Detection In Computer Networks Using Latent Space Representation And Machine Learning". *International Journal of Computing*, vol. 19, no. 3, Sept. 2020, pp. 442-8, doi:10.47839/ijc.19.3.1893.

[11] Ngo, D., L. Pham, A. Nguyen, T. Ly, K. Pham, and T. Ngo. "Sound Context Classification Based on Joint Learning Model and Multi-Spectrogram Features". *International Journal of Computing*, vol. 21, no. 2, June 2022, pp. 258-70, doi:10.47839/ijc.21.2.2595.

[12] Sulam, Jeremias, Rami Ben-Ari, and Pavel Kisilev. "Maximizing AUC with Deep Learning for Classification of Imbalanced Mammogram Datasets." In *VCBM*, pp. 131-135. 2017. http://dx.doi.org/10.2312/vcbm.20171246

[13] Asuntha, A., and Andy Srinivasan. "Deep learning for lung Cancer detection and classification." *Multimedia Tools and Applications* 79, no. 11 (2020): 7731-7762. https://doi.org/10.1007/s11042-019-08394-3.

[14] H. Chen, H. Zhao, J. Shen, R. Zhou and Q. Zhou, "Supervised Machine Learning Model for High Dimensional Gene Data in Colon Cancer Detection," *2015 IEEE International Congress on Big Data*, 2015, pp. 134-141, doi: 10.1109/BigDataCongress.2015.28.

[15] Chawla, Nitesh V., Nathalie Japkowicz, and Aleksander Kotcz. "Special issue on learning from imbalanced data sets." *ACM SIGKDD explorations newsletter* 6, no. 1 (2004): 1-6.

[16] Fawcett, Tom, and Foster Provost. "Adaptive fraud detection." *Data mining and knowledge discovery* 1, no. 3 (1997): 291-316.

[17] Gong, Joonho, and Hyunjoong Kim. "RHSBoost: Improving classification performance in imbalance data." *Computational Statistics & Data Analysis* 111 (2017): 1-13.

[18] Alqatawna, Jafar, Hossam Faris, Khalid Jaradat, Malek Al-Zewairi, and Omar Adwan. "Improving knowledge based spam detection methods: The effect of malicious related features in imbalance data distribution." *International Journal of Communications, Network and System Sciences* 8, no. 05 (2015): 118.

[19] Athitya Kumaraguru, M., Viji Vinod, N. Rajkumar, and S. Karthikeyan. "Parallel selective sampling for imbalance data sports activities." In *Soft Computing: Theories and Applications*, pp. 879-886. Springer, Singapore, 2020.

[20] Danaee, Padideh, Reza Ghaeini, and David A. Hendrix. "A deep learning approach for cancer detection and relevant gene identification." In *Pacific symposium on biocomputing 2017*, pp. 219-229. 2017.

[21] Kumar, Ansuman, and Anindya Halder. "Active Learning Using Fuzzy-Rough Nearest Neighbor Classifier for Cancer Prediction from Microarray Gene Expression Data." *International Journal of Pattern Recognition and Artificial Intelligence* 34, no. 01 (2020): 2057001.

[22] Fathi, Hanaa, Hussain AlSalman, Abdu Gumaei, Ibrahim IM Manhrawy, Abdelazim G. Hussien, and Passent El-Kafrawy. "An efficient cancer classification model using microarray and high-dimensional data." *Computational Intelligence and Neuroscience* 2021 (2021).

[23] Kilicarslan, Serhat, Kemal Adem, and Mete Celik. "Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network." *Medical hypotheses* 137 (2020): 109577.

[24] Lin, Chun-Yu, Peiying Ruan, Ruiming Li, Jinn-Moon Yang, Simon See, Jiangning Song, and Tatsuya Akutsu. "Deep learning with evolutionary and genomic profiles for identifying cancer subtypes." *Journal of Bioinformatics and Computational Biology* 17, no. 03 (2019): 1940005, doi: 10.1142/S0219720019400055.

[25] Motieghader, Habib, Ali Najafi, Balal Sadeghi, and Ali Masoudi-Nejad. "A hybrid gene selection algorithm for microarray cancer classification using genetic algorithm and learning automata." *Informatics in Medicine Unlocked* 9 (2017): 246-254.

[26] Liu, Huiqing, Jinyan Li, and Limsoon Wong. "Use of extreme patient samples for outcome prediction from gene expression data." *Bioinformatics* 21, no. 16 (2005): 3377-3384.

[27] Zhu, Zexuan, Yew-Soon Ong, and Manoranjan Dash. "Markov blanket-embedded genetic algorithm for gene selection." *Pattern Recognition* 40, no. 11 (2007): 3236-3248.

[28] Burnham, K. P., & Anderson, D. R, "Multimodel Inference: Understanding AIC and BIC in Model Selection,". *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644 (2004).

[29] Sharma, Suvita Rani, Birmohan Singh, and Manpreet Kaur. "A Novel Approach of Ensemble Methods Using the Stacked Generalization for High-dimensional Datasets." *IETE Journal of Research* (2022): 1-16.

[30] Kshatri, Sapna Singh, Deepak Singh, Bhavana Narain, Surbhi Bhatia, Mohammad Tabrez Quasim, and Ganesh Ram Sinha. "An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach." *IEEE Access* 9 (2021): 67488-67500.

[31] Venkataramana, Lokeswari, Shomona Gracia Jacob, and Rajavel Ramadoss. "A parallel multilevel feature selection algorithm for improved cancer classification." *Journal of Parallel and Distributed Computing* 138 (2020): 78-98.

[32] Robnik-Šikonja, Marko, and Igor Kononenko. "Theoretical and empirical analysis of ReliefF and RReliefF." *Machine learning* 53, no. 1 (2003): 23-69. Robnik-Šikonja, Marko, and Igor Kononenko. "Theoretical and empirical analysis of ReliefF and RReliefF." *Machine learning* 53, no. 1 (2003): 23-69.

[33] Sumant, Archana Shivdas, and Dipak Patil. "Ensemble feature subset selection: integration of symmetric uncertainty and chi-square techniques with RReliefF." *Journal of The Institution of Engineers (India): Series B* (2022): 1-13.

[34] Sumant, Archana Shivdas, and Dipak Patil. "Stability Investigation of Ensemble Feature Selection for High Dimensional Data Analytics." In *International Conference on Image Processing and Capsule Networks*, pp. 801-815. Springer, Cham, 2022.

[35] Khan, Salman, Hossein Rahmani, Syed Afaq Ali Shah, and Mohammed Bennamoun. "A guide to convolutional neural networks for computer vision." *Synthesis lectures on computer vision* 8, no. 1 (2018): 1-207.

[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[37] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," The Journal of physiology, vol. 160, pp. 106-154, 1962.

[38] Nair, Vinod, and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines." In *Icml*.

[39] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." In *International conference on machine learning*, pp. 448-456. PMLR, 2015.

[40] Banerjee, Chaity, Tathagata Mukherjee, and Eduardo Pasiliao Jr. "An empirical study on generalizations of the ReLU activation function." In *Proceedings of the 2019 ACM Southeast Conference*, pp. 164-167. 2019.

[41] Chen, Chien-Hua, Po-Hsiang Lin, Jer-Guang Hsieh, Shu-Ling Cheng, and Jyh-Horng Jeng. "Robust multi-class classification using linearly scored categorical cross-entropy." In *2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, pp. 200-203. IEEE, 2020.

[42] Yang, Jing, and Guanci Yang. "Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer." *Algorithms* 11, no. 3 (2018): 28.

[43] Panda, Mrutyunjaya. "Elephant search optimization combined with deep neural network for microarray data analysis." *Journal of King Saud University-Computer and Information Sciences* 32, no. 8 (2020): 940-948.

[44] Vasavada, N. "One-way ANOVA (ANalysis Of VAriance) with post-hoc Tukey HSD (Honestly Significant Difference) test calculator for comparing multiple treatments." OneWay_Anova_with_TukeyHSD/_get_data (2016).

[45] Kim, Hae-Young. "Analysis of variance (ANOVA) comparing means of more than two groups." *Restorative dentistry & endodontics* 39, no. 1 (2014): 74-77. https://doi.org/10.5395/rde.2014.39.1.74

Author Profile

ARCHANA S. SUMANT is currently pursuing the Ph.D. degree in Computer Engineering with MET's Institute of Engineering, Nashik, Maharashtra, India. She received her Masters' Degree in Computer Engineering from V. J. T. I., Mumbai University (INDIA) in 2010 & Bachelors' Degree in Computer Engineering from Walchand College of Engineering Sangli Shivaji University (INDIA) in 2002. Her areas of interests are parallel computing and machine learning. She has teaching experience of 20 years. She is life member of ISTE and IAENG.

DIPAK V PATIL received B.E. degree in Computer Engineering in 1998 from the University of North Maharashtra India and M.Tech. Degree in computer engineering in 2004 from Dr. B. A. Technological University, Lonere India. He has done Ph.D. degree from S.R. T. M. University, Nanded. Currently, he is a Professor and Head in Computer Engineering Department at GES's R. H. Sapat College of Engineering Management Studies and Research, Nashik, India affiliated to the Savitribai Phule Pune University, Pune, Maharashtra, India. He has many publications in peer-reviewed journals of international repute with good indexing like SCIE, SCOPUS, etc. His research interests include soft computing, data mining, system simulation and modelling.