J Med Syst (2019) 43: 263
https://doi.org/10.1007/s10916-019-1397-z

IMAGE & SIGNAL PROCESSING

Breast Cancer Diagnosis Using Feature Ensemble Learning Based on Stacked Sparse Autoencoders and Softmax Regression

Vinod Jagannath Kadam[1] · Shivajirao Manikrao Jadhav[1] · K. Vijayakumar[2]

Received: 2 April 2019 / Accepted: 19 June 2019 / Published online: 3 July 2019

Abstract
Nowadays, the most frequent cancer in women is breast cancer (malignant tumor). If breast cancer is detected at the beginning stage, it can often be cured. Many researchers proposed numerous methods for early prediction of this Cancer. In this paper, we proposed feature ensemble learning based on Sparse Autoencoders and Softmax Regression for classification of Breast Cancer into benign (non-cancerous) and malignant (cancerous). We used Breast Cancer Wisconsin (Diagnostic) medical data sets from the UCI machine learning repository. The proposed method is assessed using various performance indices like true classification accuracy, specificity, sensitivity, recall, precision, f measure, and MCC. Simulation and result proved that the proposed approach gives better results in terms of different parameters. The prediction results obtained by the proposed approach were very promising (98.60% true accuracy). In addition, the proposed method outperforms the Stacked Sparse Autoencoders and Softmax Regression based (SSAE-SM) model and other State-of-the-art classifiers in terms of various performance indices. Experimental simulations, empirical results, and statistical analyses are also showing that the proposed model is an efficient and beneficial model for classification of Breast Cancer. It is also comparable with the existing machine learning and soft computing approaches present in the related literature.

Keywords Breast Cancer · Stacked Sparse Autoencoders · Softmax classifier · Ensemble learning · Fine needle aspiration biopsy

Introduction

Breast Cancer and fine needle aspiration biopsy

Breast (carcinoma) Cancer originates in breast tissues and then spreads to other body parts. As per the various reports, this kind of cancer is the second principal cause of cancer death among women worldwide. It is the most prevalent cancer as well as life-threatening diseases among women (if not detected in early stage) [1, 2]. Early screening, correct detection and diagnosis of Breast Cancer are very important to improve the survival rates significantly and to increase chances of recovery. Computer-aided intelligent and automated diagnosis systems, developed by machine learning approaches, are important means in the analysis of breast cancer and it can support medical experts (oncologists) in the medical decision-making process. Fine needle aspiration biopsy (FNAB) of the breast is not only widely accepted as a first-line diagnostic method of breast lesions but also minimally invasive yet maximally diagnostic approach [3]. It is an easy, very safe, faster, less traumatic and cost-effective procedure when compared with open surgical biopsy. In this procedure, a thin and hollow needle is inserted into the mass to take a sample of cells (tissue sample) from an organ or lump (a suspicious area). Collected samples of cells are then examined (analyzed) under a microscope [4, 5].

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Vinod Jagannath Kadam
vjkadam@dbatu.ac.in

Shivajirao Manikrao Jadhav
smjadhav@dbatu.ac.in

K. Vijayakumar
mkvijay@msn.com

1 Department of Information Technology, Dr. Babashaeb Ambedkar Technological University, Lonere, India

2 Department of Computer Science & Engineering, St. Joseph's Institute of Technology, Chennai, India

Related work

Several Machine learning (ML) and soft computing approaches have been applied in the analysis and classification of the data acquired from a digitized image of a fine needle aspirate (FNA) of a breast mass. (e.g. Breast Cancer Wisconsin (Diagnostic) Data Set). These approaches include SVMs [5–16], Decision trees [10, 15, 17–19], Random forest [15], Artificial neural network (ANN) [6, 14, 20] (e.g. Probabilistic neural network [21], Multilayer perceptrons network (MLP) [11, 14, 17, 18], Radial basis function neural network (RBFNN) [5, 6], Fuzzy rough neural network [20] etc.), Logistic regression [10, 15], Naive Bayes classifier [17], K-nearest neighbor [6, 14, 15, 17], Minimum Distance Classifier [11], Linear discriminant analysis [18], kernel-based methods, fuzzy classifiers [22], clustering algorithms [9], evolutionary computations (e.g. Firefly Algorithm [23], Genetic algorithm [11, 12, 24], Ant Colony Optimization [12], simulated annealing [25] and Particle swarm optimization [12, 24, 26, 27] etc.), Fusion of multiple classifiers [17], ensemble method, hybrid approach [15, 16] and Deep learning approaches etc. Sangwook Kim et al. [28] proposed an stacked architecture model using Support Vector Machine for Breast Cancer classification. Ahmed M. Abdel-Zaher et al. [29] proposed Deep Belief Networks for Breast Cancer Classification and tested on UCI Breast Cancer Wisconsin (Original) Data Set. J. Xu et al. [30] applied Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. Qi Zhang et al. [31] proposed unified deep learning (DL) architecture (comprised of the point-wise gated Boltzmann machine (PGBM) and the restricted Boltzmann machine (RBM)) for automatically learning features from shear-wave elastography (SWE) images and for classifying breast cancer. Several researchers have attempted to apply convolutional neural networks to analyze mammograms or for breast cancer screening [32–37]. M. Mohsin Jadoon et al. [38] applied convolutional neural network-discrete wavelet (CNN-DW) and convolutional neural network-curvelet transform (CNN-CT) for Three-Class Mammogram Classification. S. Agrawal et al. [39] proposed a Hybrid model of convolutional neural networks and Linear Classification to Detect of Breast Cancer from Mammogram. K. Liu et al. [40] proposed fully-connected layer first convolutional neural network (FCLF-CNN) (the fully-connected layers are embedded before the first convolutional layer). They tested it on both UCI Wisconsin diagnostic breast cancer (WDBC) database and UCI Wisconsin breast cancer database (WBCD). The result of this study proved that the FCLF-CNN outperforms both MLP and CNN on both structured datasets (WDBC and WBCD datasets) [40]. Yawen Xiao et al. [41] simulated deep learning based unsupervised feature extraction approach that combines stacked autoencoders with support vector machine (SAE-SVM) and obtained 98.25 % highest predictive accuracy on Breast Cancer Wisconsin (Diagnostic) Data Set.

Motivation and this work

The deep neural network (DNN) is the superior approach [42] not only in the field of computer vision and signal processing but also in other real-world classification and prediction applications. Feature Engineering is an important step in conventional shallow classifiers [43] but the DNNs automatically learn hierarchies of relevant features directly from the given raw data. Therefore, DNNs may extend a potentially superior classifier for the data acquired from a digitized image of a fine needle aspirate (FNA) of a breast mass. In stacked sparse autoencoders and softmax regression based classification model (SSAE-SM model), the network of stacked sparse autoencoders decreases the dimension of feature space and final Softmax layer is used for classification. The training of the SSAE-SM model involves two phases: the greedy layer-wise pre-training phase and fine-tuning phase. Greedy layer-wise pre-training (Unsupervised) is intended to initialize the DNN and to introduce a useful prior to next fine-tuning training (supervised) phase. The outermost layer of the stacked sparse autoencoders network produces most abstraction representation. This representation is used to train the final Softmax layer. Many researchers have successfully applied this DNN for various medical dataset classifications. Y. Lu, L. Zhang, B. Wang, and J. Yang [44] proposed a modified model using feature ensemble learning based on Sparse Autoencoders for image classification. Unlike SSAE-SM model, this model makes use of all layers of representations. Motivated by these developments, V.J.Kadam et. al. [45] proposed feature ensemble learning based on sparse autoencoders for the automated diagnosis of the Parkinson's disease. In this study, we analyzed the effectiveness of this model for classification of Breast cancer. Simulation and outcomes of this study indicate that the proposed model is an efficient and useful model for classification of Breast Cancer.

The rest of the paper is structured as follows. "Stacked Sparse Autoencoders and Softmax Regression as classifier (SSAE-SM model)" briefly explains Stacked Autoencoders and Softmax layer based Deep Neural Network Classifier (SSAE-SM model). The proposed Feature Ensemble Learning Based on Stacked Sparse Autoencoders and softmax classifier (FE-SSAE-SM model) for Breast Cancer Diagnosis is presented in "Proposed feature ensemble learning based on Stacked Sparse Autoencoders and Softmmax Regression Model (FE-SSAE-SM model) for Breast Cancer diagnosis". "Experiments and results" describes the UCI WDBC dataset, experimental study, and numerical results. "Experiments and results" also includes performance indices, and different comparisons. Conclusions are presented in "Conclusion".

Stacked Sparse Autoencoders and Softmax Regression as classifier (SSAE-SM model)

An Autoencoder (consists of an encoder part and a decoder part) is an Artificial (feed-forward) neural network trained using unsupervised learning (that applies back-propagation approach) to replicate the input representation at the output. This simple learning circuit owns three layers: an input layer, a hidden (encoding) layer, and a decoding (output)layer. It tries to learn approximation to the identity function which forces the hidden (encoding) layer to try to learn good representations of the inputs (or compressed knowledge representations of the original input.) While conceptually simple, Autoencoders play a fundamental role in Deep architecture [46–49].

Sparse Autoencoder (SA) is an Autoencoder imposed with sparseness constraints on the all hidden nodes. It introduces the sparse penalty term.The cost function for training a Sparse Autoencoder (given by Eq. 1) includes three terms. The first term is called mean square error which gives the discrepancy between input x and reconstructed $\hat{x}$ over the whole training data [50].

$$E = MSE + (\lambda \times L2RegularizationTerm) + (\beta \times SparsityRegularizationTerm) \quad (1)$$

λ = The coefficient for the L2 regularization term.
β = The coefficient for the sparsity regularization term.
The Eq. 2 gives L2 regularization term.

$$L2RegularizationTerm = \frac{1}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(w_{ji}^{(l)}\right)^2 \quad (2)$$

Where n_l = Number of layers, l = Layer l, s_l = Number of units in l layer, $w_{ji}^{(l)}$ = the weight value between node i in the layer l and node j in layer $l+1$. The Eq. 3 gives Sparsity Regularization Term.

$$SparsityRegularizationTerm = \sum_{j=1}^{s_2} KL(\rho \| \hat{\rho}_j) \quad (3)$$

Where $\hat{\rho}_j$ = average activation of hidden node j, ρ is a sparsity parameter. $KL(\rho\|\hat{\rho}_j)$ is Kullback-Leibler divergence (between Bernoulli random variable with mean ρ and a Bernoulli random variable with mean $\hat{\rho}_j$ is defined as

$$KL(\rho\|\hat{\rho}_j) = (\rho)log\left(\frac{\rho}{\hat{\rho}_j}\right) + (1-\rho)log\left(\frac{1-\rho}{1-\hat{\rho}_j}\right) \quad (4)$$

An effective deep learning method known as Stacked Sparse autoencoders (SSAE) is a framework in which many such Sparse autoencoder layers are stacked one after another to form a deep learning network. The outputs of each layer are connected to the inputs of each successive layer. Softmax (classifier) layer is used as the last layer of this network. Training of this network is performed in two phases: 1) the unsupervised greedy layer-wise initialization phase 2) the supervised fine-tuning phase. To suitably initialize this Stacked Sparse autoencoder based deep neural network and to introduce a useful prior to next fine-tuning training (supervised) phase, the unsupervised greedy layer-wise algorithm (proposed by Hinton [51, 52]) trains the first sparse autoencoder to minimize the reconstruction error of the raw data, followed by training subsequent sparse autoencoder with the hidden activations of previous sparse autoencoder as input. Then, the last hidden activations are taken as input to train a softmax layer. This greedy layer-wise pre-training sets the stage for a final fine-tune training phase [53]. Finally, the algorithm fine-tunes all parameters of this stacked sparse autoencoder-softmax classifier with the supervised mode for achieving a more specific task. Figure 1 shows SSAE-SM model (Stacked Sparse Autoencoder and Softmax classifer model with two Sparse Autoencoder layers and final softmax layer).

Softmax regression

Softmax (SM) regression classifier, another log-linear model, the superior variant of the logistic regression, is multinomial logistic regression. It is a generalized form (or an extension)of logistic regression to the case where class labels can take more than two possible values and the classes are mutually exclusive. In other words, It is Multi-class Logistic Regression. A softmax classifier (which is trained on provided training data) calculates a separate decimal probability for every

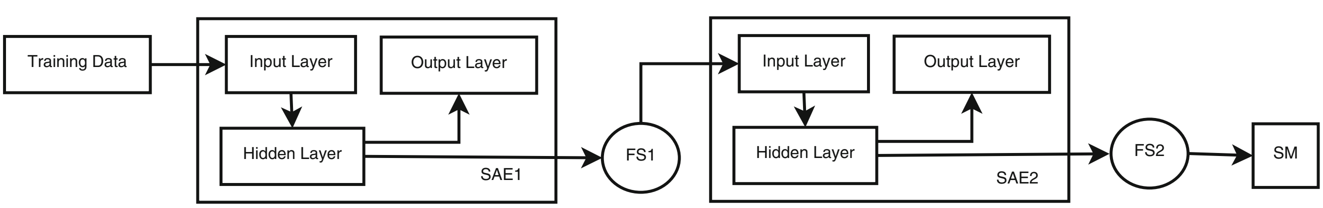

Fig. 1 Stacked sparse Autoencoders - Softmax classifier based model (SSAE-SM model) with two SAEs and Last SM layer [SAE-Sparse Autoencoder, FS-Feature Space, SM-Softmax layer]

possible class with the target class having the highest probability and the decimal probabilities all add up to one. This is the main advantage of using Softmax classifier.

Proposed feature ensemble learning based on Stacked Sparse Autoencoders and Softmmax Regression Model (FE-SSAE-SM model) for Breast Cancer diagnosis

As discussed above, In stacked sparse Autoencoders-softmax classifier based model (SSAE-SM model), the representation given by the last layer of stacked autoencoders is provided to softmax layer for classification. However, we obtain multiple representations when we train stacked sparse autoencoder based DNN. As discussed by Y. Lu et al. [44] and Kadam et al. [45], to take benefits from representations of all layers, we can integrate them by training different classifiers and apply some combination rule to take the final decision. In this study, As discussed by V.J. Kadam et al. [45], We trained two sparse Autoencoders to produce two representations. Training of first stacked sparse autoencoder producers first representation. The output of the hidden layer (i.e. first representation) of first sparse autoencoder is provided as input to second sparse autoencoder to get the second representation. Again, We concatenated these two representations to form the third representation. We trained three Softmax classifiers corresponding to these three representations. It should be noted that, for the first (even is shallow network) and second classifiers, we require to do fine tuning on the whole network including the input data layer, feature representation layer, and final Softmax layer to enhance the performance. Figures 2 and 3 show proposed FE-SSAE-SM model.

The Training steps are given below (see Fig. 2):

- Step 1:-
  - Train SAE1 with HS1 hidden units on training dataset x. Hidden layer of SAE1 transforms x into features set FS1.
  - Train SAE2 with HS2 hidden units on training feature set FS1. Hidden layer of SAE2 transforms FS1 into features set FS2.
  - Train SM2 on the features set FS2.
  - Fine tune whole network SAE1, SAE2, and SM2.
- Step 2:-
  - Apply input data x to SAE1 to get the feature set FS1.
  - Train SM1 on the features set FS1.
  - Fine tune whole shallow network SAE1 and SM1.
- Step 3:-
  - Apply input data x to SAE1 to get feature set FS1

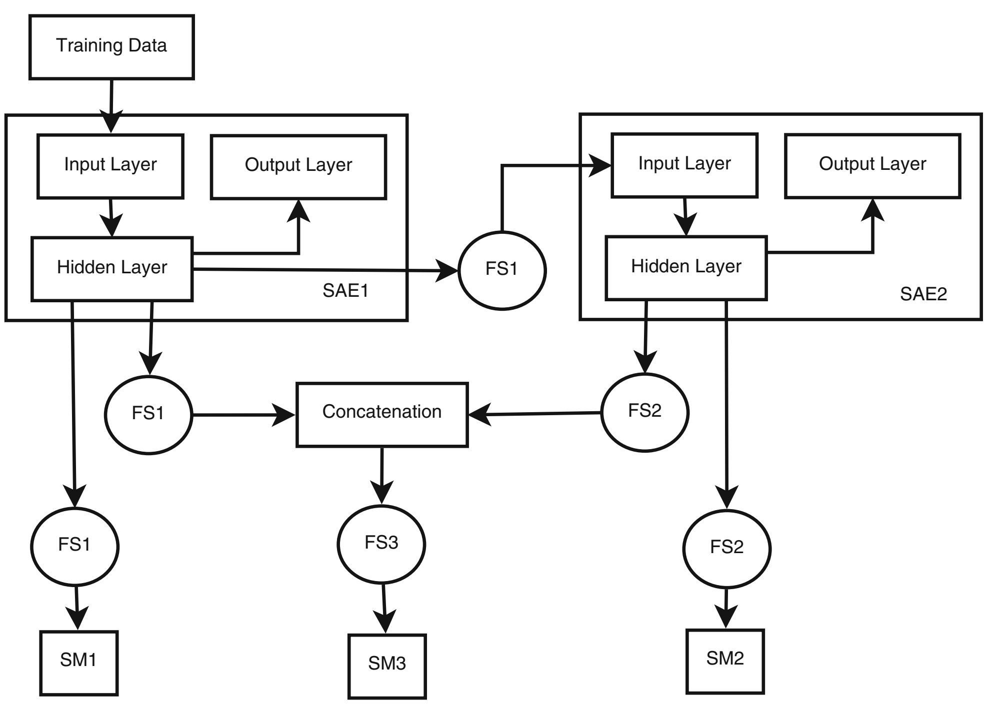

Fig. 2 Overview of training phase of proposed Feature Ensemble Learning Based on Stacked Sparse Autoencoders and Softmax regression model (FE-SSAE-SM model) [SAE-Sparse Autoencoder, FS-Feature Space, SM-Softmax classifier]

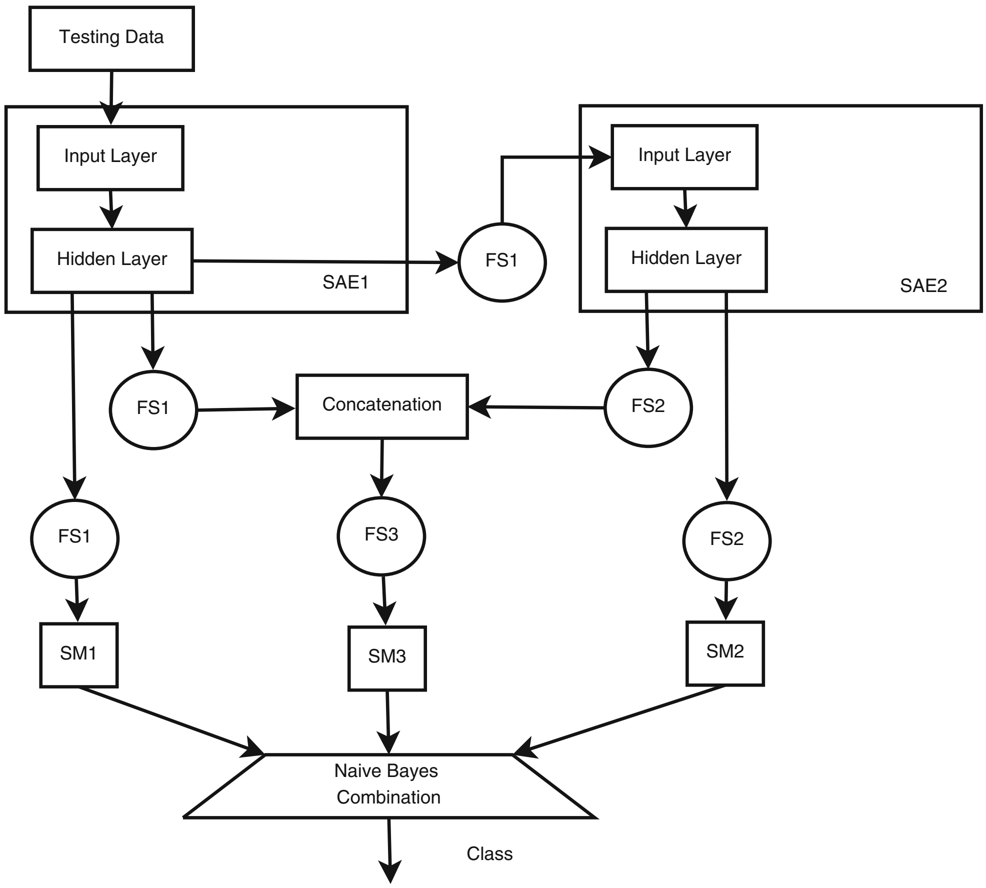

Fig. 3 Overview of testing phase of proposed Feature Ensemble Learning Based on Stacked Sparse Autoencoders and Softmax regression model (FE-SSAE-SM model) [SAE-Sparse Autoencoder, FS-Feature Space, SM-Softmax classifier]

- Apply feature set FS1 to SAE2 to get the feature set FS2.
- concatenate feature set FS1 and feature set FS2 to get the feature set FS3.
- Train SM3 on the features set FS3.

Figure 2 shows the Training phase of proposed method. In the testing stage, for the ensemble of three softmax classifiers, we examined the Naive Bayes combination method. As explained by Ludmila Kuncheva [54], Naive Bayes combination methods are MAX, MIN and AVG rule etc. [55]. In this study, we used AVG rule of Naive Bayes combination method. Let x be new instance for testing, its class label y takes on k possible values $j = 1, 2, ..k$, we can get corresponding prediction probabilities for the softmax classifier SM_n $(n = 1, 2, 3$ here), denoted as $P_{nj}(x)$. The AVG rule for determining the final value of label y is

$$y = arg \max_{j\in\{1,2,...k\}} \sum_{n=1}^{N} P_{nj}(x)/N, \quad (5)$$

Where N = 3 and k = 2 in this case. Figure 3 shows the testing phase of proposed method.

Experiments and results

Dataset

In this study, we used Breast Cancer Wisconsin (Diagnostic) medical data sets (WDBC) . It was obtained from the UCI Machine Learning dataset repository (https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)) [56]. This dataset was created by Dr. William H. Wolberg, W. Nick Street and Olvi L. Mangasarian and donated by Nick Street. The dataset holds 569 records . There are 357 (62.7%) cases of benign breast changes and 212 (37.3%) malignant breast cancer. Each record consists of ID number, diagnosis (Dataset label: 'B' means benign, 'M' means malignant), and 30 real-valued input features . These 30 features real-valued features are measured from a digitized image of a fine needle aspirate of a breast mass. They represent characteristics of the cell nuclei existing in the image.10 features (1.Radius, 2.texture, 3.Perimeter, 4.Area, 5.Smoothness, 6.Compactness, 7.Concavity, 8.Concave points, 9.Symmetry and 10. Fractal dimension) estimated for each cell nucleus [57–59]. The mean value, The mean of the worst 3 measurements and The standard error of these features were are computed for each image, yielding a database of 30 real-valued input attributes for 569 cases. No missing value is present in the dataset.

Performance indices

Performance indices used for evaluation and comparison are as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\%$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100\%$$

$$Specificity = \frac{TN}{FP+TN} \times 100\%$$

$$Precision(P) = TP/(TP+FP)$$

$$Recall(R) = TP/(TP+FN)$$

$$F-measure = 2 \times (P \times R)/(P+R)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Where TP = true positive, FP = false positive, TN = true negative, and FN = false negative and these values are derived from Confusion Matrix (Table 1).We used here two popular and promising indices F measure and Matthews correlation coefficient (MCC). F measure is defined as the harmonic mean of classification precision and recall [60]. Therefore, unlike other indices, the F measure takes both FP and FN into account [61]. Matthews correlation coefficient is another important measure of the quality of binary classifications.Being more balanced index, MCC estimates a correlation of the classification prediction.

Experimentation

For experimentation purpose, We implemented proposed Feature ensemble based Stacked autoencoder + softmax regression-based model (FE-SSAE-SM) and Stacked autoencoder and softmax regression-based model (SSAE-SM model) using Matlab 2018b environment (we used functions like trainAutoencoder, trainSoftmaxLayer and deepnet etc.). Both networks contain two Autoencoders (SAE1 and SAE2). We tried three combinations on both models:

i) 8 hidden layer neurons in SAE1 and 6 hidden layer neurons in SAE2
ii) 8 hidden layer neurons in SAE1 and 8 hidden layer neurons in SAE2
iii) 8 hidden layer neurons in SAE1 and 10 hidden layer neurons in SAE2

Both models were trained and tested on WDBC dataset. 10-fold cross-validation approach was adopted to evaluate the accuracy and compare the efficiency of FE-SSAE-SM and SSAE-SM model. Training procedures of SSAE-SM and FE-SSAE-SM are given in "Stacked Sparse Autoencoders and Softmax Regression as classifier (SSAE-SM model)" and "Proposed feature ensemble learning based on Stacked Sparse Autoencoders and Softmmax Regression Model (FE-SSAE-SM model) for Breast Cancer diagnosis" respectively. To find the optimal value of parameters λ, β and ρ for both models, we used Grid search. Difference values were assigned to regulation parameters λ, β and ρ to study their effect on accuracy. The coefficient for the L2 regularization term (λ) in the range of 0.005 to 8, Sparsity Proportion (ρ) in the range of 0.01 to 1, Coefficient for the Sparsity regularization term (β) in the range of 1 to 10 and Scaled conjugate gradient with max epoch in the range of 50 to 600 (for all Autoencoders and all softmax classifiers and fine tuning) were considered for experimentation. Performance Comparisons between both model (proposed FE-SSAE-SM and SSAE-SM) in terms of accuracy, sensitivity, and specificity are provided in Table 2. FE-SSAE-SM and SSAE-SM models with the best accuracy are further compared with other state-of-the-art models in terms of accuracy, sensitivity, specificity, recall, precision, f measure, and MCC on WDBC dataset using 10 fold cross

Table 1 Confusion matrix of classification

	Prediction as malignant	Prediction as benign
Actual malignant	TP	FN
Actual benign	FP	TN

Table 2 Performance comparison of FE-SSAE-SM and SSAE-SM (10 fold-cross validation)

Hidden layer size of SAE1 SAE2	Model	Accuracy	Sensivitiy	Specificity
8-6	FE-SSAE-SM	**98.60**	97.16	99.44
	SSAE-SM	98.25	96.69	99.16
8-8	FE-SSAE-SM	98.59	**97.19**	99.436
	SSAE-SM	98.07	96.21	99.16
8-10	FE-SSAE-SM	98.59	96.71	**99.71**
	SSAE-SM	98.07	96.71	98.88

Bold emphasis indicates Best performance

Table 3 Performance Comparison of the proposed FE-SSAE-SM with other classifiers (10 fold-CV)

Classifier	Accuracy	Sensitivity	Specificity	Precision	Recall	f measure	MCC
Decision Trees [Split criterion = Gini's diversity index]							
Coarse Tree [max. number of splits = 4]	93.322	91.038	94.678	0.91	0.91	0.91	0.857
Medium Tree [max. number of splits = 20]	93.322	91.509	94.398	0.907	0.915	0.911	0.857
Fine Tree [max. number of splits = 100]	93.322	91.509	94.398	0.907	0.915	0.911	0.857
Linear Discriminant [Covariance structure = Full]	95.431	88.679	99.44	0.989	0.887	0.935	0.903
Quadratic Discriminant [Covariance structure = Full]	95.255	94.34	95.798	0.93	0.943	0.937	0.899
Logistic Regression	95.606	95.755	95.518	0.927	0.958	0.942	0.907
SVM [Box constraint level = 1]							
Linear SVM [Linear kernel]	97.715	94.811	99.44	0.99	0.948	0.969	0.951
Quadratic SVM [Quadratic kernel]	97.715	95.283	99.16	0.985	0.953	0.969	0.951
Cubic SVM [Cubic kernel]	97.715	95.755	98.88	0.981	0.958	0.969	0.951
Fine Gaussian SVM [Gaussian kernel,Kernel scale = 1.4]	81.371	50.943	99.44	0.982	0.509	0.671	0.617
Medium Gaussian SVM[Gaussian kernel,Kernel scale = 5.5]	97.188	95.755	98.039	0.967	0.958	0.962	0.94
Coarse Gaussian SVM [Gaussian kernel,Kernel scale = 22]	95.431	88.208	99.72	0.995	0.882	0.935	0.904
KNN [Distance metric = Euclidean, distance weight = Equal]							
Fine KNN [Number of neighbors = 1]	95.255	92.453	96.919	0.947	0.925	0.936	0.898
Medium KNN [Number of neighbors = 10]	96.485	91.981	99.16	0.985	0.92	0.951	0.925
Coarse KNN [Number of neighbors = 100]	93.146	82.547	99.44	0.989	0.825	0.9	0.856
Cosine KNN [Distance metric = Cosine, distance weight = Equal,Number of neighbors = 10]	96.661	93.396	98.599	0.975	0.934	0.954	0.928
Cubic KNN [Distance metric = Minkowski, distance weight = Equal,Number of neighbors = 10,]	96.134	91.038	99.16	0.985	0.91	0.946	0.918
Weighted KNN [Distance metric = Euclidean, distance weight = Squared Inverse, Number of neighbors = 10]	96.837	93.396	98.88	0.98	0.934	0.957	0.932
SSAE-SM [Hidden layer size of SAE1- SAE2 = 8-6]	98.243	96.698	99.16	0.986	0.967	0.976	0.962
FE-SSAE-SM [Hidden layer size of SAE1 - SAE2 = 8-6]	98.594	97.17	99.44	0.99	0.972	0.981	0.97

validation method. It is given in Table 3 and Figs. 4, 5, and 6. These classifiers were implemented in Matlab. Parameter settings are also given in the Table 3.

Because there is no standard evaluation protocol for WDBC dataset, studies available in the literatures have adopted different accuracy evaluation protocols, causing

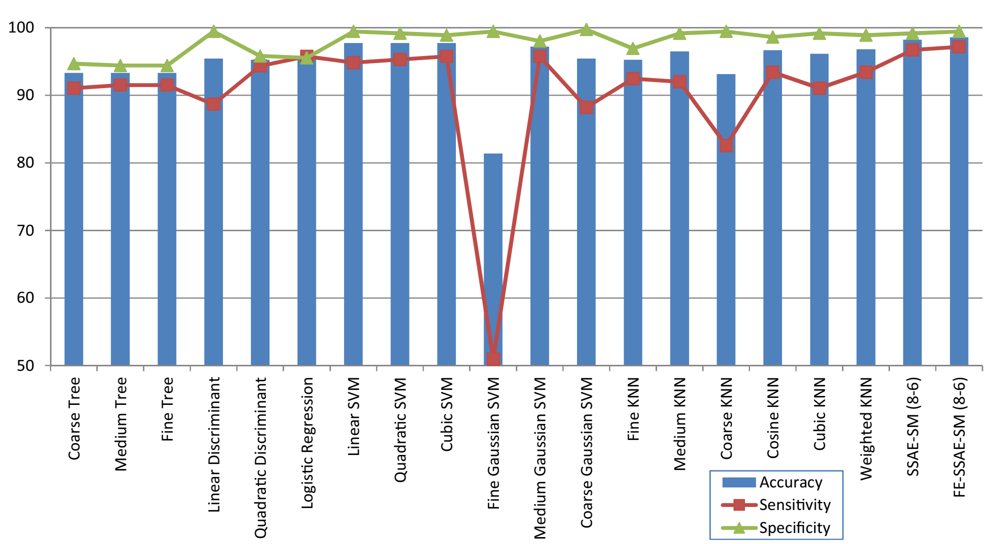

Fig. 4 Comparison of classifiers in terms of Accuracy, Sensitivity, and Specificity

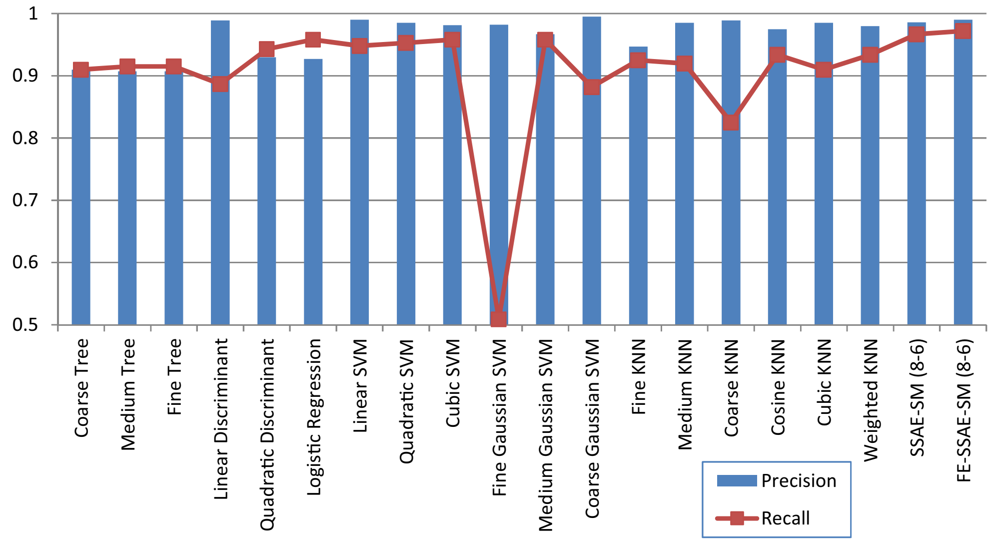

Fig. 5 Comparison of classifiers in terms of Precision and Recall

difficulty in comparing and analyzing performance across these studies. Here, we attempted to compare the performance of FE-SSAE-SM model with other studies (Table 4) whose accuracy evaluation protocols are closest to ours (i.e. 10 fold cross-validation).

Discussion

In this paper, we applied a Feature ensemble learning method based on stacked Sparse Autoencoders and softmax regression model on UCI WDBC dataset to classify Breast Cancer into benign (non-cancerous) and malignant (cancerous). True accuracy was estimated using 10 fold Cross-validation method. The proposed model gave 98.60% true accuracy with hidden layer size 8 and 6 of Autoencoder 1 and Autoencoder 2 respectively and using proper tuning hyperparameter settings. Highest Sensitivity and Specificity obtained by the proposed model are 97.19% and 99.71% respectively. F measure and MCC obtained by the proposed model are 0.981 and 0.97 respectively. The outcomes of the study show proposed FE-SSAE-SM model outperforms SSAE-SM model and many states-of-the-art classifiers. Comparison of the proposed method and other methods (Available in the literature) is also presented in the paper. Proposed FE-SSAE-SM model outperforms many other existing approaches.

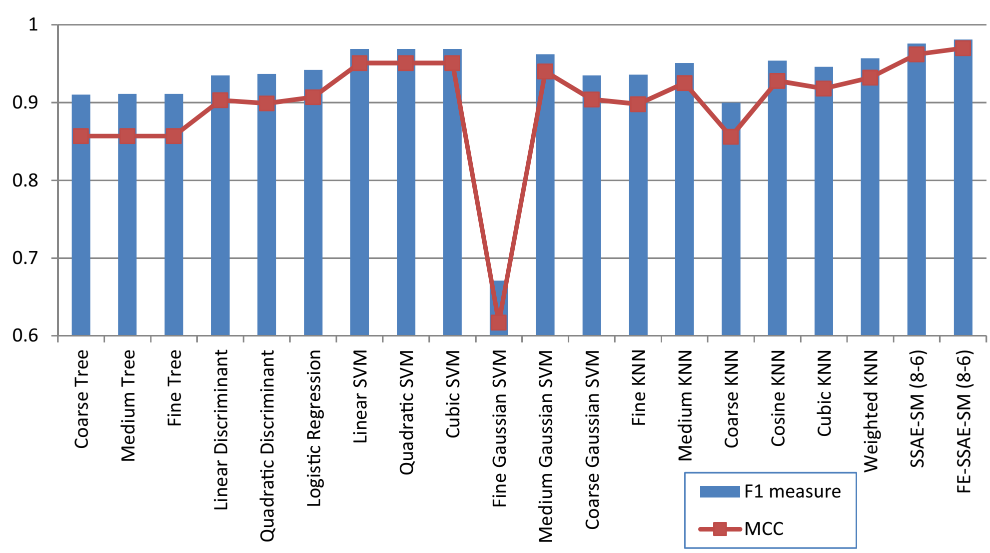

Fig. 6 Comparison of classifiers in terms of F measure and MCC

Table 4 Comparison with other methods in the literature on Breast Cancer Wisconsin (Diagnostic) medical dataset (Results based on 10-fold CV)

Year	Method [Reference]	Accuracy
2010	ACO-SVM [12]	95.96
2010	GA-SVM [12]	97.19
2010	PSO-SVM [12]	97.37
2011	Self-training [62]	85.12
2011	Random co-training [62]	83.54
2011	Rough co-training [62]	88.63
2012	Naive Bayes(NB) [17]	92.97
2012	Multi-Layer Perception [17]	96.66
2012	Decision tree (J48) [17]	93.14
2012	Instance Based for K-Nearest neighbor [17]	95.95
2012	Sequential Minimal Optimization (SMO) [17]	97.71
2012	Fusion of NB and SMO [17]	97.53
2012	Fusion of MLP and SMO [17]	97.71
2012	Fusion of J48 and SMO [17]	94.90
2012	Fusion of IBK and SMO [17]	97.71
2012	Fusion of SMO, IBK and NB [17]	97.36
2012	Fusion of SMO, IBK and MLP [17]	97.18
2012	Fusion of SMO, IBK, and J48 [17]	97.36
2012	Fusion of SMO, IBK, NB and MLP [17]	97.53
2012	Fusion of SMO, IBK, NB and J48 [17]	97.01
2014	Hybrid of K-means and SVM [9]	97.38
2014	Probabilistic neural network [21]	96.31
2015	Independent component analysis + k-NN [6]	91.03
2015	Independent component analysis + RBFNN [6]	90.49
2015	independent component analysis and ANN [6]	90.50
2015	Independent component analysis + SVM (linear) [6]	90.33
2015	Independent component analysis + SVM (quadratic) [6]	89.98
2015	Independent component analysis + SVM (RBF Kernel) [6]	90.86
2016	PSO-Kernel density estimation [24]	98.45
2016	GA-Kernel density estimation [24]	98.45
2016	Artificial immune with semi-supervised learning [63]	98.00
2018	AMBFA [23]	98.21
2018	Binary Firefly Algorithm (BFA) [23]	98.17
2018	Affinity Propagation (AP) clustering + BFA [23]	98.54
2018	AP + AMBFA [23]	**98.60**
2018	Self-Organizing Error Drive (SOED) ANN [14]	96.19
2018	Ensemble of LR + KNN (with SMOTE) [15]	98.32
2018	WAUCE model [16]	97.68
2019	BP neural network [25]	93.9
2019	IGSAGAW + BP neural network [25]	97.5
2019	GAW + BP neural network [25]	95.3
2019	3-NN [25]	92.6
2019	GAW + 3-NN [25]	94.0
2019	IGSAGAW + 3-NN [25]	95.4
2019	Cost sensitive SVM [25]	92.6
2019	GAW + cost sensitive SVM [25]	94.5
2019	IGSAGAW + cost sensitive SVM [25]	95.7
2019	This Study **SSAE-SM**	98.25
2019	This Study **Proposed FE-SSAE-SM**	**98.60**

Bold emphasis indicates Best performance

Conclusion

Early screening, accurate prediction, and diagnosis of Breast Cancers are very important. Computer-aided intelligent and automated medical decision support systems based on machine learning and soft computing methods plays an important role in the early prediction of breast cancer. This paper presents a robust and sound classifier using Feature ensemble learning based on stacked sparse autoencoders and softmax regression to classify widely adopted WDBC UCI data sets. Principal goal of the proposed study was to enhance the accuracy of breast cancer classification. The experimental outcomes and statistical analyses point out that introduced ensemble technique performs better than 'Stacked Sparse Autoencoders + Softmax classifier model'. It also outperforms many machine learning and soft computing classifer (like SVM, KNN and Decision tree etc.) Additionally, this approach is also comparable to the existing techniques available in the related literature. The experimental results and statistical analyses are pointed out that this classifier is really beneficial and efficient model for breast cancer classification.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflict of interest, financial or otherwise. No funding was received for this study.

Research involving human participants and/or animals This research paper does not contain any studies with human participants or animals performed by any of the authors.

Informed consent No humans are involved in this research paper.

References

1. Breast Cancer Awareness Month. National Health Portal Of India, 10 Apr. 2015, (www.nhp.gov.in/breast-cancer-awareness-month_pg).
2. Hashemi, S. H. B., Karimi, S., and Mahboobi, H., Lifestyle changes for prevention of breast cancer. *Electron. Physician* 6(3):894–905, 2014.

3. Yu, Y. H., Wei, W., and Liu, J. L., Diagnostic value of fine-needle aspiration biopsy for breast mass: A systematic review and meta-analysis. *BMC Cancer* 12:41, 2012. https://doi.org/10.1186/1471-2407-12-41.
4. Madubogwu, C. I., Ukah, C. O., Onyiaorah, I. V., Anyiam, D. C. D., Anyanwu, S. N. C., and Chianakwana, G. U., Cost effectiveness of fine needle aspiration cytology for breast masses. *Orient J. Med.* 27(1–2), 2015.
5. Iranpour, M., Almassi, S., and Analoui, M., Breast cancer detection from FNA using SVM and RBF classifier. In: *First Joint Congress on Fuzzy and Intelligent Systems, Ferdowsi University of Mashhad, Iran, 29–31 Aug 2007*, 2007.
6. Mert, A. et al., Breast cancer detection with reduced feature set computational and mathematical methods in medicine Volume 2015. https://doi.org/10.1155/2015/265138.
7. Luo, Z., Wu, X., Guo, S., Ye, B., Guo, S., and Ye, B., Diagnosis of breast cancer tumor based on manifold learning and support vector machine. In: *IEEE International Conference on Information and Automation June 20–23, 2008, Zhangjiajie, China*, 2008.
8. Muthu Rama Krishnan, M., Banerjee, S., Chakraborty, C., Chakraborty, C., and Ray, A. K., Statistical analysis of mammographic features and its classification using support vector machine. *Expert Syst. Appl.* 37:470–478, 2010.
9. Zheng, B., Yoon, S. W., and Lam, S. S., Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst. Appl.* 41(4):1476–1482, 2014. https://doi.org/10.1016/j.eswa.2013.08.044.
10. Bamakan, S. M. H., and Gholami, P., A novel feature selection method based on an integrated data envelopment analysis and entropy model. *Prog. Comput. Sci.* 31:632–638, 2014.
11. Guo, H., and Nandi, A. K., Breast cancer diagnosis using genetic programming generated feature. In: *2005 IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT*, pp. 215–220, 2005. https://doi.org/10.1109/MLSP.2005.1532902.
12. Prasad, Y., Biswas, K. K., and Jain, C. K., SVM classifier based feature selection using GA, ACO and PSO for siRNA design. In: Tan, Y., Shi, Y., and Tan, K. C (Eds.), *Advances in Swarm Intelligence. ICSI 2010. Lecture Notes in Computer Science*, Vol. 6146. Berlin: Springer, 2010.
13. Maldonado, S., Weber, R., and Basak, J., Simultaneous feature selection and classification using kernel-penalized support vector machines. *Inf. Sci.* 181:115–128, 2011.
14. Jafari-Marandi, R., Davarzani, S., Gharibdousti, M. S., and Smith, B. K., An optimum ANN-based breast cancer diagnosis: Bridging gaps between ANN learning and decision-making goals. *Appl. Soft Comput.* 72:108–120, 2018.
15. Naga RamaDevi, G., Usha Rani, K., and Lavanya, D., Ensemble-based hybrid approach for breast cancer data. In: *International Conference on Communications and Cyber Physical Engineering 2018*. Singapore: Springer, 2018.
16. Wang, H., Zheng, B., Yoon, S. W., and Ko, H. S., A support vector machine-based ensemble algorithm for breast cancer diagnosis. *Eur. J. Oper. Res.* 267(2):687–699, 2018.
17. Salama, G. I., Abdelhalim, M. B., and Abd-elghany Zeid, M., Breast cancer diagnosis on three different datasets using multi-classifiers. *Int. J. Comput. Inf. Technol.* 1(Issue 01):2277–0764, 2012.
18. Luukka, P., and Leppalampi, T., Similarity classifier with generalized mean applied to medical data. *Comput. Biol. Med.* 36:1026–1040, 2006.
19. Lavanya, D., and Usha Rani, K., Analysis of feature selection with classification: Breast cancer datasets. *Indian J. Comput. Sci. Eng.* 2(5), 2011. ISSN : 0976-5166.
20. Zhao, J. Y., and Zhang, Z. L., Fuzzy rough neural network and its application to feature selection. In: *The Fourth International Workshop on Advanced Computational Intelligence, Wuhan*. pp 684–687, 2011. https://doi.org/10.1109/IWACI.2011.6160094.
21. Mert, A., Kılıç, N., and Akan, A., An improved hybrid feature reduction for increased breast cancer diagnostic performance. *Biomed. Eng. Lett.* 4(3):285–291, 2014.
22. Lim, C. K., and Chan, C. S., A weighted inference engine based on interval-valued fuzzy relational theory. *Expert Syst. Appl.* 42:3410–3419, 2015.
23. Emami, N., and Pakzad, A., A new knowledge-based system for diagnosis of breast cancer by a combination of affinity propagation clustering and firefly algorithm. *J. AI Data Min.* 7:59–68, 2018.
24. Sheikhpour, R., Sarram, M. A., and Sheikhpour, R., Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer. *Appl. Soft Comput.* 40:113–131, 2016.
25. Liu, N., Qi, E.-S., Xu, M., Gao, B., and Liu, G.-Q., A novel intelligent classification model for breast cancer diagnosis. *Inf. Process. Manag.* 56(3):609–623, 2019.
26. Xue, B., Zhang, M., and Browne, W. N., New fitness functions in binary particle swarm optimisation for feature selection. In: *WCCI 2012 IEEE World Congress on Computational Intelligence June, 10–15, 2012 - Brisbane, Australia*, 2012.
27. Xue, B., Zhang, M., and Browne, W. N., Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Appl. Soft Comput.* 18:261–276, 2014.
28. Kim, S., Kavuri, S., and Lee, M., Deep network with support vector machines. In: Lee, M., Hirose, A., Hou, Z. G., and Kil, R. M. (Eds.), *Neural Information Processing. ICONIP 2013. Lecture Notes in Computer Science*, Vol. 8226. Berlin: Springer, 2013.
29. Abdel-Zaher, A. M., and Eldeib, A. M., Breast cancer classification using deep belief networks. *Expert Syst. Appl.*, 2015. https://doi.org/10.1016/j.eswa.2015.10.015.
30. Xu, J. et al., Stacked Sparse Autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE Trans. Med. Imaging* 35(1):119–130, 2016. https://doi.org/10.1109/TMI.2015.2458702.
31. Zhang, Q. et al., Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics* 72:150–157, 2016.
32. Cano, F., Madabhushi, A., and Cruz-Roa, A., A comparative analysis of sensitivity of convolutional neural networks for histopathology image classification in breast cancer. In: *Proceedings Volume 10975, 14th International Symposium on Medical Information Processing and Analysis; 109750W. Mazatlán, Mexico*, 2018. https://doi.org/10.1117/12.2511647.
33. Ragab, D. A. et al., Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ* 7:e6201, 2018.
34. Kooi, T. et al., Large scale deep learning for computer aided detection of mammographic lesions. *Med. Image Anal.* 35:303–312, 2017.
35. Chougrad, H., Zouaki, H., and Alheyane, O., Deep convolutional neural networks for breast cancer screening. *Comput Methods Programs Biomed.* 157:19–30, 2018. https://doi.org/10.1016/j.cmpb.2018.01.011. Epub 2018 Jan 11.
36. Xiao, T., Liu, L., Li, K., Qin, W., Yu, S., and Li, Z., Comparison of transferred deep neural networks in ultrasonic breast masses discrimination. *Biomed. Res. Int.* 2018:Article ID 4605191, 2018. https://doi.org/10.1155/2018/4605191.
37. Cruz-Roa, A. et al., Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Sci. Rep.* 7:46450, 2017.
38. Jadoon, M. M., Zhang, Q., Ul Haq, I., Butt, S., and Jadoon, A., Three-class mammogram classification based on descriptive CNN features. Biomed. Res. Int. **2017**, Article ID 3640901, 11 pages, 2017. https://doi.org/10.1155/2017/3640901.
39. Agrawal, S., Rangnekar, R., Gala, D., Paul, S., and Kalbande, D, Detection of breast cancer from mammograms using a hybrid

approach of deep learning and linear classification. In: *2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai*, pp. 1–6, 2018. https://doi.org/10.1109/ICSCET.2018.8537250.
40. Liu, K., Kang, G., Zhang, N., and Hou, B., Breast cancer classification based on fully-connected layer first convolutional neural networks. *IEEE Access* 6:23722–23732, 2018. https://doi.org/10.1109/ACCESS.2018.2817593.
41. Xiao, Y., Wu, J., Lin, Z., and Zhao, X., Breast cancer diagnosis using an unsupervised feature extraction algorithm based on deep learning. In: *Proceedings of the 37th Chinese Control Conference July 25–27. Wuhan*, 2018.
42. Vijayakumar, K., and Arun, C., Automated risk identification using NLP in cloud based development environments. *J. Ambient Intell. Human Comput.*, 2017. https://doi.org/10.1007/s12652-017-0503-7.
43. Kadam, V. J., Yadav, S. S., and Jadhav, S. M., Soft-margin SVM incorporating feature selection using improved elitist GA for arrhythmia classification. In: Abraham, A., Cherukuri, A., Melin, P., and Gandhi, N. (Eds.), *Intelligent Systems Design and Applications. ISDA 2018. Advances in Intelligent Systems and Computing*, Vol. 941. Cham: Springer, 2018.
44. Lu, Y., Zhang, L., Wang, B., and Yang, J., Feature ensemble learning based on sparse autoencoders for image classification. In: *2014 International Joint Conference on Neural Networks (IJCNN), Beijing*, pp. 1739–1745, 2014. https://doi.org/10.1109/IJCNN.2014.6889415.
45. Kadam, V. J., and Jadhav, S. M., Feature Ensemble Learning Based on Sparse Autoencoders for Diagnosis of Parkinson's Disease. In: Iyer, B., Nalbalwar, S., and Pathak, N. (Eds.) , *Computing, Communication and Signal Processing. Advances in Intelligent Systems and Computing*, Vol. 810. Singapore: Springer, 2019.
46. Hinton, G. E., and Salakhutdinov, R. R., Reducing the dimensionality of data with neural networks. *Science* 28(5786):504–507, 2006.
47. Bengio, Y., and LeCun, Y., Scaling learning algorithms towards AI. In: Bottou, L., Chapelle, O., DeCoste, D., and Weston, J. (Eds.) *Large-Scale Kernel Machines*: MIT Press, 2007.
48. Ranzato, M. A., Poultney, C., Chopra, S., LeCun, Y., Chopra, S., and LeCun, Y., Efficient learning of sparse representations with an energy-based model. In: *Advances in Neural Information Processing Systems 19 (NIPS'06)*, pp. 1137–1144. MIT Press, 2007.
49. Baldi, P., Autoencoders, unsupervised learning, and deep architectures. In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012.
50. Ng, A., CS294A Lecture notes. Sparse autoencoder. pp. 72, 2011. (https://web.stanford.edu/class/cs294a/sparseAutoencoder.pdf).
51. Hinton, G. E., Osindero, S., and Teh, Y.-W., A fast learning algorithm for deep belief nets. *Neural Comput.* 18(7):1527–54, 2006.
52. Hinton, G. E., Learning multiple layers of representation. *Trends Cogn. Sci.* 11(10):428–34, 2007.
53. Erhan, D., Bengio, Y., Courville, A., Manzagol, P.-A., Vincent, P., and Bengio, S., Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.* 11:625–660, 2010.
54. Kuncheva, L. I., *Combining Pattern Classifiers, Methods and Algorithms*, p. 544. New York: Wiley, 2004.
55. Kittler, J., Hatef, M., Duin, R. P. W., and Matas, J., On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* 20(3):226–239, 1998. https://doi.org/10.1109/34.667881.
56. Dua, D., and Graff, C., *UCI Machine Learning Repository (http://archive.ics.uci.edu/ml)*. Irvine: University of California, School of Information and Computer Science, 2019.
57. Wolberg, W. H., Street, W. N., and Mangasarian, O. L., Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer Lett.* 77(2–3):163–171, 1994.
58. Nick Street, W., Wolberg, W. H., and Mangasarian, O. L., Nuclear feature extraction for breast tumor diagnosis. *Proc. SPIE* 1905:861–871, 1993.
59. Street, W. N., and Wolberg, W. H., Breast cancer diagnosis and prognosis via linear programming. *Oper. Res.* 43(4):570–577, 1995.
60. Pradeep Mohan Kumar, K., Saravanan, M., Thenmozhi, M., and Vijayakumar, K., Intrusion detection system based on GA-fuzzy classifier for detecting malicious attacks. *Concurr. Comput. Pract. Exp.* e5242, 2019. https://doi.org/10.1002/cpe.5242.
61. Joseph Manoj, R., Anto Praveena, M. D., and Vijayakumar, K., An ACO–ANN based feature selection algorithm for big data. *Cluster Comput.*, 2018. https://doi.org/10.1007/s10586-018-2550-z.
62. Miao, D., Gao, C., Zhang, N., and Zhang, Z., Diverse reduct subspaces based co-training for partially labeled data. *Int. J. Approx. Reason.* 52:1103–1117, 2011.
63. Peng, L., Chen, W., Zhou, W., Li, F., Yang, J., and Zhang, J, An immune-inspired semi-supervised algorithm for breast cancer diagnosis. *Comput. Methods Prog. Biomed.*, 2016. https://doi.org/10.1016/j.cmpb.2016.07.020.