An Overview of Lung Cancer Classification Algorithms and their Performances

F. Taher, N. Prakash, A. Shaffie, A. Soliman, A. El-Baz

***Abstract*—In the world, lung cancer is the third most dreadful cancer. Thus, detection of lung cancer cells at early stage is a challenge. The symptoms of lung cancer do not appear in earlier stages which causes high death rates when compared with other types of cancer. In lung cancer detection, image processing algorithms have shown great performance in various high-end tasks. In this paper, different classification methodologies used for the prediction of lung cancer in its early stage are explained. Machine learning techniques are used to identify whether lung tumors are malignant or benign. Machine learning approaches such as: Convolutional neural network (CNN), Support vector machine (SVM), Artificial neural network (ANN), Multi-Layer Perceptron (MLP), K-Nearest Neighbor (KNN), Entropy degradation method (EDM) and Random Forest (RF) are discussed in detail and their performance is evaluated in terms of accuracy, sensitivity and specificity. In this analysis, CNN approach using small dataset shows best result with 96% accuracy compared to other methodologies and EDM shows the worst accuracy of 77.8%.**

***Index Terms*—ANN, benign, cancer, malignant**

I. Introduction

LUNG cancer causes high mortality rates in human population which seriously affect the whole world. Researchers proposed CAD methods which can be applied to computed tomography (CT) images for the classification of pathological objects at an early stage [1]. Most of the tests and procedures can be done very fast, but sometimes it takes days or even months which is common. Such delay can cause serious problems to both patients and care providers and it can affect the survival rate adversely. Therefore, in order to enhance the patient’s condition, the diagnosis-to-treatment process should be made very fast.

Early prediction of lung cancer is possible by the development of the machine learning techniques. In this paper, some of the machine learning techniques for lung cancer prediction are discussed. For creating such CAD system, reference quality dataset is used which will generate the ground truth.

Manuscript received November 10, 2020; revised August 30, 2021.

F. Taher is an Associate Professor and Assistant Dean for Research and Outreach in the College of Technological Innovation, Zayed University, Dubai, U.A.E (phone: +971565257765; e-mail: Fatma.Taher@zu.ac.ae).

N. Prakash is a Research Assistant in the College of Technological Innovation, Zayed University, Dubai, U.A.E. (e-mail: Neema.Prakash@zu.ac.ae).

A. Shaffie is a Postdoctoral Associate of University of Louisville, Louisville, KY, USA. (e-mail: amshaffie.oq@gmail.com).

A. Soliman is a Postdoctoral Associate of University of Louisville, Louisville, KY, USA. (e-mail: ahmedsoli@gmail.com).

A. El-Baz is a Chairperson, Professor and Distinguished Scholar of University of Louisville, Louisville, KY, USA (e-mail: ayman.elbaz@louisville.edu).

Different CAD algorithms are compared using this [2]. Neural network is an effective tool for building an assistive artificial intelligence (AI) based cancer detection system which plays a major role in the classification of the cancer cells as normal or abnormal. An effective cancer treatment can be seen only when the tumor cells are identified from the normal cells. The machine learning based cancer diagnosis [3] mainly focuses on the classification of the tumor cells and training of the neural network which is very important in lung cancer research [4]. This paper discusses different lung cancer classification algorithms such as CNN, SVM, ANN, MLP, KNN, EDM and RF and their performances are evaluated.

The rest of this paper is organized as follows. In Section II, the related works are discussed. In Section III, lung cancer detection methods are described. In Section IV, results are illustrated and finally in Section V, the conclusion is presented.

II. Related works

Main contributions of some of the researchers who tried to develop a lung cancer prediction system analyzed using different classification algorithms are summarized below.

For the classification of lung cancer tumors as benign and malignant, Sasikala et al. [5] proposed a convolutional neural network (CNN) based approach. This system is trained by inputting the lung cancer tissue images of variant shape and size. CNN obtained high accuracy of 96% when compared with other conventional neural network systems which makes this method more efficient. In order to detect the cancer types of different size and shape, CNN will use large datasets for training in the forthcoming years. This paper concludes by suggesting a 3D CNN method that can be used for improving the performance of the system and also by improving the hidden neurons with deep network.

A computer aided lung classification method developed using artificial neural network was presented by Jinsa et al. [6]. The parameters are calculated after the entire lung is segmented from the CT images. The statistical parameters explained in this paper are used as features for classification. Different neural networks are used for the classification process. Thirteen training functions are employed for evaluating the performance of this system. The Traingdx training function gives the highest accuracy rate.

Fang et al. [7] has proposed a lung cancer prediction system based on a new deep learning technique called Google Net which shows better performance such as convergence rate, accuracy, sensitivity, and specificity. Google Net is fine-tuned for the classification of lung cancer cells. This method used less training time and gave better results. Median intensity projection (MIPs) is also discussed in this paper which helps to learn features of cancerous and non-cancerous lung nodules that are compatible with the fine-tuned Google Net. This will increase the accuracy of the system when tested on the validation sets. After 300 epochs, accuracy of 81%, sensitivity of 84%, and specificity of 78% are produced by the trained system which is better than other available programs.

Dipanjan et al. [8] aims to develop a 1D CNN model for the classification of non-small cell lung cancer (NSCLC). This model performs better than the conventional CNN methods. This method consumes less time and it detects the NSCLC tumors very accurately. It will thus help the researchers to provide new methods of automated cancer treatments.

A 3D multipath visual geometry group (VGG) evaluated on 3D cubes is proposed by Tekade et al. [9]. The features are extracted from different sources which are available for free access. The proposed approach contains mainly 2 architectures. U-Net architecture is adapted for segmentation of lung nodules from lung CT scan images and 3D multipath VGG-like architecture is proposed for classifying lung nodules and the prediction of their malignancy level. This is useful to predict whether the patient will have the cancer in next two years or not. Combining the two approaches gives a better result for predicting lung nodule detection and also further predicting malignancy level. An accuracy of 95.66% and dice coefficient of 90% is obtained using this approach.

III. Methodology

Lung cancer detection system (Fig. 1) [10] based on chest CT images using machine learning techniques such as CNN, SVM, ANN, MLP, KNN, EDM and RF are discussed as follows.

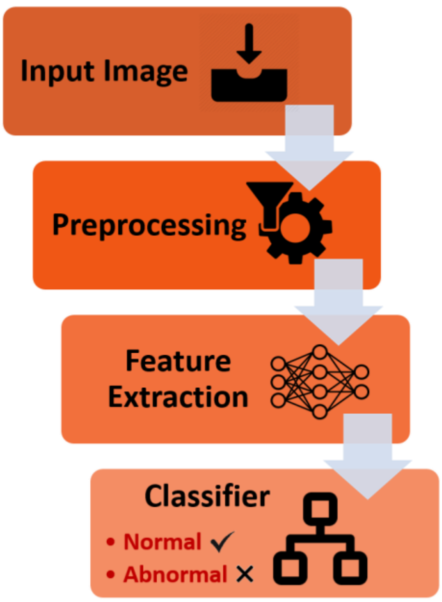
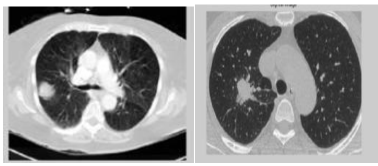


Fig. 1. General block diagram of the lung cancer detection system

In the first stage, lung CT images are preprocessed by applying median filter which minimizes the degradation during acquisition. Then, from the CT image scans, lung regions are extracted. Segmentation of each slice is done to identify tumors. Segmented tumors are then fed as input to the classifier which decides whether the tumor present in a patient's lung is cancerous or non-cancerous [11]. Non-cancerous and cancerous lung images are depicted in Fig. 2. Median filtered images are depicted in Fig. 3.



(a) (b)

Fig. 2. (a) Non-cancerous and (b) cancerous images

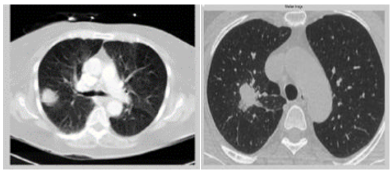


Fig. 3. Median filtered images

A. *Convolutional neural network (CNN)*

A class of deep neural network called convolutional neural network (CNN)/ConvNets can be used in many applications such as image processing, face recognition, object detection etc. This type of neural network is mainly used to identify the cancerous or non-cancerous lung tumors. Among the pattern recognition and computer vision research area, convolutional neural networks (CNNs) models become popular because of their promising outcome on generating high-level image representations.

A CNN is type of neural network [12] composed of several kinds of layers such as convolutional layer, pooling layer and fully connected layers. In order to extract the features from an input image, convolutional layer creates a feature map. The pooling layer keeps the main information only and the other information are cut down. A fully connected input layer flattens the output from the pooling layer. A SoftMax activation function is used by the final layers [5] after passing through the fully connected layer. The final outcome is obtained from the fully connected output layer which helps in the classification of image [5].

Architecture of CNN proposed by Sasikala et al. [5] is shown in the Fig. 4, an image of size b x b x r, where r is the number of the channels given as the input to a convolutional layer. There are k filter kernels of size a x a x q where $a < b$, $q \leq r$ and may vary for each kernel in convolutional layer. In order to produce k feature maps, they are convolved with the input image. Mean or max pooling is used for the sub sampling of each map.

B. *Support vector machine (SVM)*

Vapnik [13] introduced SVM and received considerable recognition due to its high accuracy. An optimal separating hyper plane (OSH) is the basis of this method that separates the training data. The training data is labeled with the output class called maximum margin classifiers by a supervised learning approach, such that the empirical risk can be simultaneously minimized.

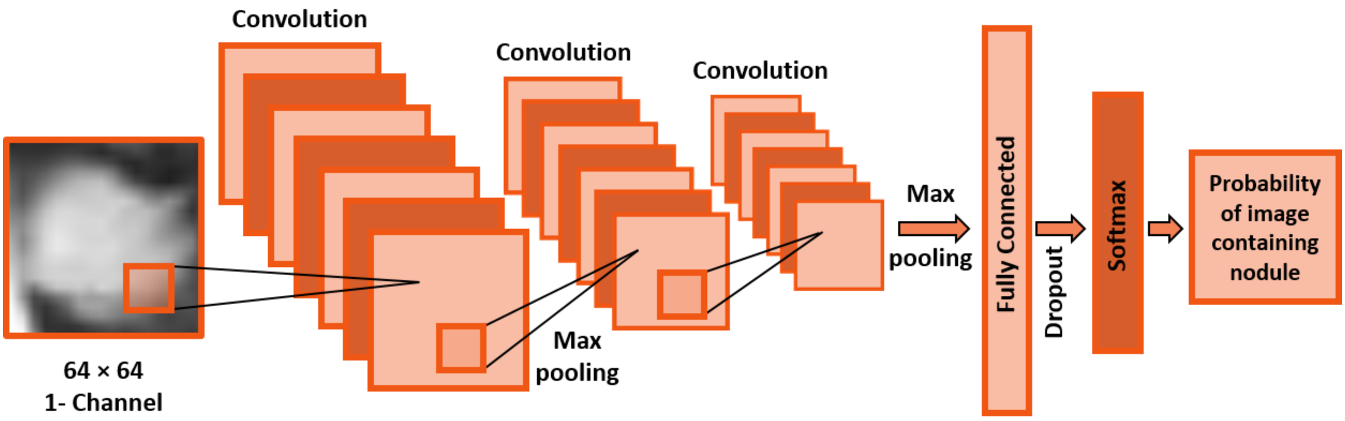


Fig. 4. Architecture of CNN

The optimal hyperplane is determined by:

$$\{h \in \mathbb{H} | \langle w, h\rangle_{\mathbb{H}} + w_0 = 0\} \quad (1)$$

and $\Phi(s_i)$ denotes the mapped data.
Where, the inner product in space $\mathbb{H}$ is indicated as <w, h>, w denotes the quadratic programming problem given as follows:

$$\min_{w,w_0,\xi_{i,\dots,}\xi_N}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i\right) \quad (2)$$

Subject to

$$y_i\,(\langle w, \Phi(s_i)\rangle_{\mathbb{H}} + w_0) - 1 + \xi_i \geq 0 \quad i=1,\dots,N \quad (3)$$
$$\xi_i \geq 0 \quad i=1,\dots,N$$

Where, the number of training samples is denoted as N, ξ_i are slack variables, and C denotes a positive constant. The problem as given in (2) is solved by:

$$\max_{\alpha l_{i,\dots,}\alpha l_N}\left(\sum_{i=1}^{N}\alpha l_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha l_i\,\alpha l_j y_i y_j K(s_i, s_j)\right) \quad (4)$$

Subject to

$$0 \leq \alpha l_i \leq C \quad i = 1\dots, N \quad (5)$$

$$\sum_{i=1}^{N}\alpha l_i y_i = 0 \quad (6)$$

Where, the kernel function is denoted as K (si,sj) = ⟨Φ(si), Φ(sj)⟩, and αli are Lagrange multipliers. Support vector lies near to the OSH in the higher feature space [14].
The SVM learning approach [15] is shown in Fig. 5. In this approach, the support vectors help to maximize the margin of the classifier. Therefore, over-fitting between the classes can be reduced. An SVM classifier with Gaussian kernel is given as follows:

$$K(x_i, x) = e^{-\|x_i - x\|^2/2\sigma^2} \quad (7)$$

where, x_i is the data used for training, x is the support vector and σ is the kernel width, and hyper-parameter of SVM. By applying SVM as in (7) with its specification to data obtained from the feature extraction process, the kernel checks whether the input data is mapped to a feature space of higher dimension. Benign and malignant cells are the two classes of separation. The main strengths of SVM are explained in [16].

The size of the training dataset has a direct impact on the complexity of SVMs.

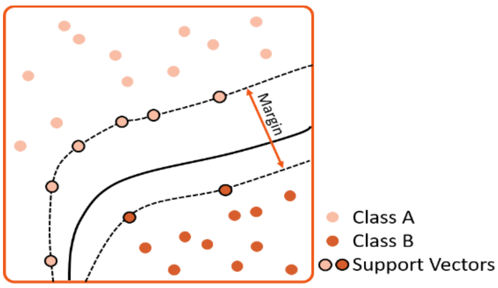


Fig. 5. SVM learning approach

C. *Artificial neural network (ANN)*

In the field of medical image classification, artificial neural network (ANN) [17] are used for classification, pattern recognition, decision-making, dimension reduction etc. which makes it one of the major approaches. Applications where data is not clear, data classification and pattern recognition, ANN can be used [18]. Fig. 6 depicts the ANN architecture which is mainly used in the field of cytology.
The ANN works better in the range 0 to 1, therefore input data is taken in this range. A feed forward neural network [19] is used to determine the unknown function $y= f(x)$ for a given data set $\{x_i, y_i\}_{i=1}^{N}$. The network uses a back-propagation training function and a row vectors of M hidden layer sizes and a feed forward neural network is returned. The equation following determines the relationship connecting the input, output and hidden neurons given by (x_i, $i = 1, 2, \dots, n_1$), (Y_k, $k = 1, 2, \dots, N$) and (h_j, $j = 1, 2, \dots, m_1$) respectively.

$$Y_k = g\left[\sum_{j=1}^{m1} w_{kj}\, g\left(\sum_{i=1}^{n1} w_{ji}\, x_i + \theta_{in1}\right) + \theta_{hid}\right] \quad (8)$$

Where, $g(z) = 1/\,(1 + e^{-z})$. w_{kj} is the weight from j^{th} hidden neuron to the k^{th} output neuron, w_{ji} is the weight from the i^{th} input neuron to the j^{th} hidden neuron, a bias neuron in the input layer and hidden layer is denoted as θ_{in1} and θ_{hid} respectively. Furthermore, an activation function is used for processing of each neurons in the ANN which is given as follows:

$$O_{pj} = \frac{1}{1+\exp(-\sum_i w_{ji}\, O_{pi} + \theta_j)} \quad (9)$$

where O_{pj} is the the output pattern and O_{pi} is the input pattern. The back-propagation function is used to minimize the weights between pairs of neurons. The adjusted weights are calculated initially as follows:

$$\Delta w_{ji}\ (k_1 + 1) = nl\delta_{pj}O_{pi} + \alpha\Delta w_{ji}\ (k_1) \qquad (10)$$

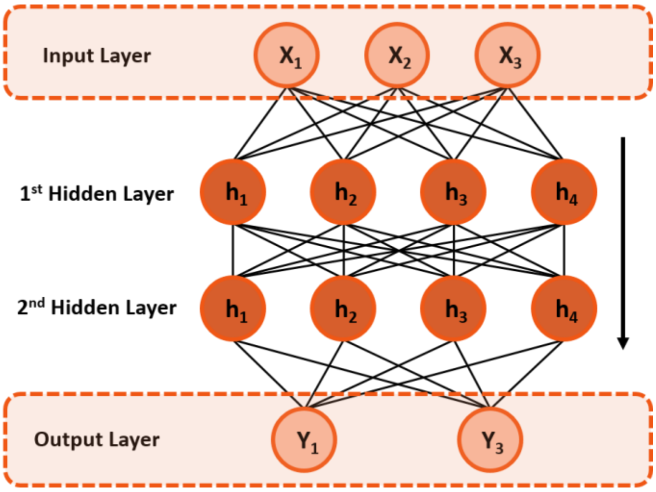


Fig. 6. ANN architecture

Where, the learning rate is denoted as *nl* which is equal to 0.3, α is the momentum term equals to 0.9, k_1 indicates the number of iterations, and δ_{pj} is the error between the desired and actual ANN output values.

In ANN, the final weights can be calculated based on some conditions such as δ_{pj} should become smaller than a threshold value or k_1 has reached another threshold value. Much care is taken when deciding the number of hidden layers. The number of epochs selected varies from 5 to 10 which will decide how much the number of hidden nodes is changed. Finally, for classification of CT images into normal and abnormal, the best trained ANN network [20] is used.

D. Multi-Layer Perceptron (MLP)

The architecture of the MLP classifier is shown in Fig.7 which consists of three layers namely: input, hidden and output layers. There are several neurons present in each layer. Direct learning process is used by the MLP classifier for generating different classes and the optimal weights are calculated by backpropagation training process. The MLP model is trained by the following parameters such as number of hidden layers, alpha, learning rate and solver which is used for optimizing weight [21]. Back propagation neural network is used for enhancing the performance of the MLP.

E. K-Nearest Neighbor (KNN)

Data sets are classified based on their similarity with neighbors using the KNN algorithm. In this classification method, K denotes the quantity of data set. The test sample label determines the similarity among the K nearest neighbors. In this algorithm, the distances between a test sample and database samples is found by using the Euclidean Distances (ED). Between X = (x_1, x_2, x_3, · · · x_n) and Y = (y_1, y_2, y_3, · · · y_n), Euclidean distance is given as follows:

$$ED(x,y) = \sqrt{\sum_{j=1}^{k}(X_i - Y_i)^2} \qquad (11)$$

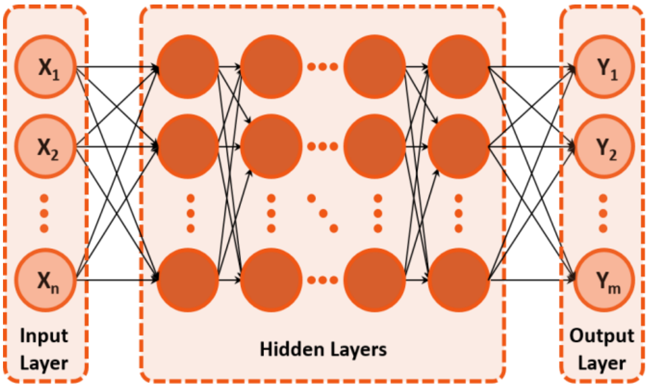


Fig. 7. MLP architecture

F. Entropy degradation method (EDM)

Entropy degradation method (EDM) proposed by Qing Wu et al. [22] is used to diagnose small cell lung cancer (SCLC) from CT images [23]. Data images used for training and testing of good quality are collected from the National Cancer Institute. A number of patient CT scans of high resolution are obtained from open source databases. Pathology diagnosis will provide them with ground truth labels. In this method, from the database, 12 lung CT scans are selected which consists of an equal number of healthy lungs and SCLC patients' lungs.

In order to train the model, 5 random scans from each group are selected and the remaining two scans are used for testing. Each CT images contains 100 to 500 axial slices of the chest cavity based on scan parameters. Then, labelling as cluster 1 and cluster 0 will helps to identify the SCLC images and other, respectively. This is done because for SCLC patients, not all CT scans reveal cancer cells. Two additional CT images are used for testing, one for SCLC patient and one without are selected.

Non-cancerous and cancerous images are identified using the SCLC detection [24] of group 0 and group 1 respectively. From the training sets, many features are extracted [22]. The vectorized histogram from cancerous and non- cancerous lungs are fed into the neural network during the training process. Each training set is transformed by EDM into a score [25] which is then converted into probability with the help of a logistic function. For testing, an input without any marking is used for testing which is fed into the neural network. In this case the output is calculated by using the probabilities associated with those scores. The final output will help to find the group to which the testing data belongs to, whether a cancerous or non-cancerous patient [20].

$$a_{m,n} = log(\sum_n \sum_m^N \overrightarrow{point}\ (p{=}j) \times \vec{x_j}\ )+ log\left(\sum \overrightarrow{point}\ (p = n)\right)/(\sum\sum \overrightarrow{point}) \qquad (12)$$

Where n = 0 or 1, (p=n) defines an indicator function. m varies from 1 to the size of input x. N denotes the total size of input x.

The estimated maximum entropy signals' $Y = cd\,f\,(y)$" is used. Its value is calculated by function h, as given below,

$$h = log(det(det(W)) + \frac{1}{N}\sum(log(e + 1 - Y^2)) \qquad (13)$$

Where, W denotes an identity 5x5 matrix.

$$W = W + eta \times (g) \tag{14}$$

Where, eta is used to control convergence speed and the gradient matrix is defined as g.

G. Random Forest (RF)

Random forests are known for their high performance and generalizability. In order to perform the classification, RF model can be used where the dependent variable is categorical. Based on the rules, the data is divided by the tree. The dataset can be split into many regions by using these rules. Variable's influence to the homogeneity or cleanliness of the subsequent child nodes (X2, X3) can be used to compute these rules. The variable X1 becomes a root node because it leads to maximum homogeneity in child nodes. RF model have some other features which helps in the classification process such as Gini Index and Entropy.

IV. Results

A. Dataset

In the CNN approach [26] proposed by Sasikala et al. [5], a dataset consisting of 1000 CT scans are collected. These CT scans are having different nodule sizes. They are, nodule greater than or equal to 3 mm, less than 3 mm, and non-nodule greater than or equal to 3 mm. Among them, training sets consist of 70 images and testing set consists of 30 images.

Lung cancer classification using SVM proposed by Fenwa et al. [27] acquired a total of 80 images which consists of both Chronic Obstructive Pulmonary Disease (COPD) and Idiopathic Pulmonary Fibrosis (IPF). Training and testing are done using 48 and 32 images, respectively.

ANN machine learning algorithm proposed by Naresh et al. [28] uses 111 CT images for stage 1 and 73 samples for stage 2 type of lung cancer. Nodules are described by using the structural and textural features. Among the dataset obtained, 70% are used for training and 30% are used for testing. MLP and KNN algorithms proposed by Sujata et al. [21] uses python programming language for the implementation and the performances are evaluated on DICOM CT images of 1018 cases collected from LIDC-IDRI. In addition to the lung parts, some other parts such as aorta, vena cava, trachea, esophagus are also present in the CT scan images. Morphological opening and local thresholding method are used for extracting Region of Interest (ROI). The features are extracted from the segmented grayscale lung volume. The training set consists of 4877 normal, 36 benign and 53 malignant cases. The testing set consists of 1221 normal, 7 benign and 14 malignant cases.

To evaluate the performance of the EDM algorithm, 100 CT scans are used which contains multiple axial slices (100 to 500 slices) of chest cavity depending on scan parameters. Training set is obtained by randomly selecting 10 of them, where 5 of them will be healthy patients and other 5 SCLC [29] patients. Training input is the extracted vectorized histogram. The remaining samples are taken as testing set. Total of 36 tests are done from these combinations. RF method proposed by Jayaraj et al. [30] also uses a dataset consisting of 1018 images with 512*512-pixel dimensions.

B. Performance evaluation

These above-mentioned classifiers can be compared with the help of confusion matrix [31]. True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) [32] are used in representing the confusion matrix (Table I). Sensitivity, specificity and accuracy can be calculated using TP, FN, FP and TN. The correctly classified lung cancer images are given as true positive and the wrongly predicted non- cancerous images are given as false positive.

In CNN method [5], an input sample image is fed to the trained model which is followed by preprocessing, feature extraction and finally identifying the cancer spot.

Implementation of this neural network is performed in MATLAB. Some parameters such as weight, learning rate, gradient moment and hidden neurons are also used for training. In order to improve the accuracy of this method, two convolution and subsampling layers are also used.

This method shows an accuracy of 96%, sensitivity of 87.5 % and specificity of 100%. SVM method [27] uses a total training time of 146.86s and total recognition time of 0.284s. This method is able to obtain an accuracy of 96%, sensitivity of 90% and specificity of 90%. While considering ANN approach [28], an accuracy of 92.68%, sensitivity of 55.26% and specificity of 100% can be seen. In the MLP and KNN methods [21], accuracy, sensitivity and specificity are almost same which is given in Table II. EDM algorithm [22] makes 10 FP predictions and it also misses 6 cases when the patients actually diagnosed with SCLC.

It shows an accuracy of 77.8%, sensitivity of 83.33% and specificity of 72.22%. By using RF method, accuracy of 89.9%, sensitivity of 90.85% and specificity of 88.32% is obtained. Evaluation metrics showing the performances of all these methods is presented in Table II. Fig.8 shows the performance criteria of all methods. Accuracy of CNN, SVM, ANN, MLP, RF and KNN are high when compared with EDM. Sensitivity of CNN, SVM and RF is high. ANN, MLP and KNN are showing low sensitivity. CNN, ANN, MLP and KNN are having 100% specificity.

From these studies, CNN is proved to be a good classifier showing better performance by using a smaller dataset compared to other methods. SVM, MLP, RF and KNN are better compared to ANN and EDM [33]. EDM algorithm needs improvement because of many false positive predictions [34]. This can be rectified by using large datasets.

TABLE I
CONFUSION METRIX

Parameters	CNN	SVM	ANN	MLP	KNN	EDM	RF
TP	7	19	17	1200	1200	30	600
TN	22	19	17	21	21	26	400
FP	0	6	0	1	1	10	8
FN	1	6	3	20	20	6	10

TABLE II
EVALUATION METRICS

Author	Algorithms	Sensitivity	Specificity	Accuracy
Sasikala et al. [5]	CNN	87.5 %	100%	96%
Fenwa et al. [27]	SVM	90%	90%	96%
Naresh et al. [28]	ANN	55.26%	100%	92.68%
Sujata et al. [21]	MLP	51.2%	100%	98.31%
Sujata et al. [21]	KNN	51.3%	100%	98.30%
Qing et al. [22]	EDM	83.33%	72.22%	77.8%
Jayaraj et al. [30]	RF	90.85%	88.32%	89.9%

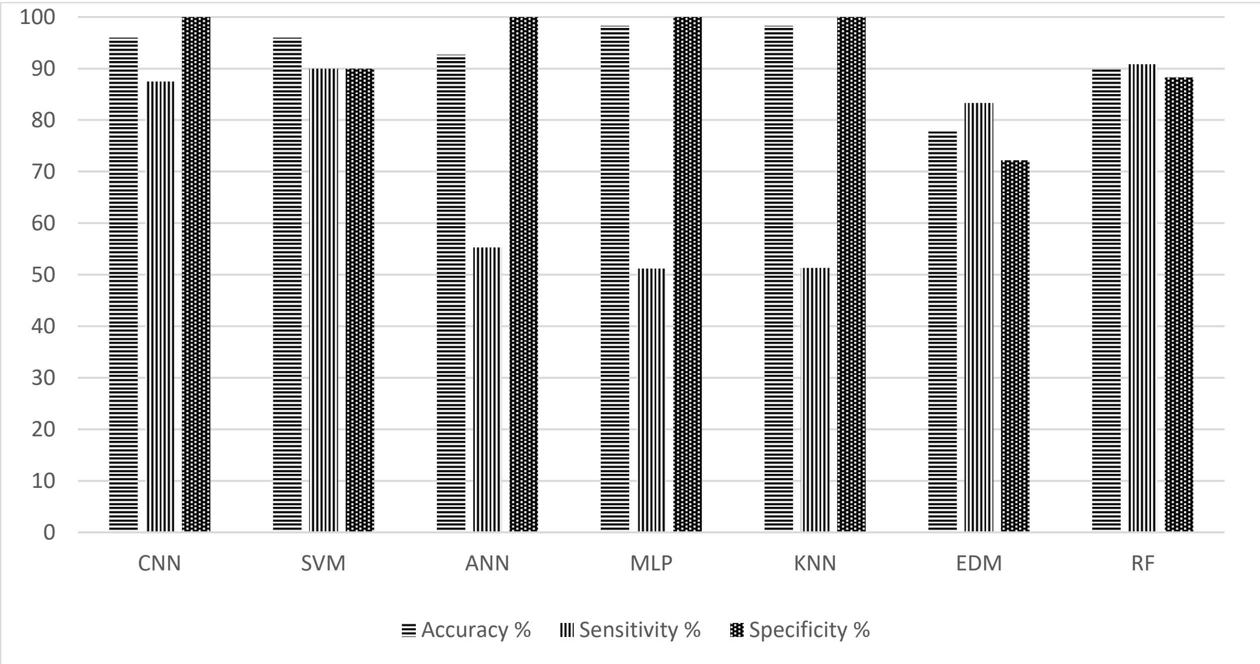


Fig. 8. Bar chart for performance criteria

V. CONCLUSION

In this paper, different machine learning techniques used in the classification of lung tumor such as CNN, SVM, ANN, MLP, KNN, EDM and RF are explained and their performances are evaluated in terms of accuracy, sensitivity and specificity. CNN system is able to classify the benign and malignant cells with high accuracy of 96% with a test data of 30 images. Sensitivity and specificity of the CNN based system are also compared to other techniques. SVM classifier also shows better accuracy of 96% with test data of 32 images, hence used for differentiating the pulmonary lung nodules into malignant and benign to assist the radiologist and for the future enhancement. ANN, MLP, KNN and RF are able to achieve high accuracy of 92.68%, 98.31%, 98.30% and 89.90% respectively with large datasets. The EDM method shows the least accuracy of 77.8%. EDM requires a large space for improvement. In order to enhance the performance, EDM method is combined with CNN, where large datasets and deeper network are used for training. Thus, in CT lung imaging, this method can be used for various applications.

REFERENCES

[1] N. Camarlinghi, "Automatic detection of lung nodules in computed tomography images: Training and validation of algorithms using public research databases", Eur. Phys. J. Plus, vol. 128, no. 9, p. 110, Sep. 2013.

[2] K. Mohanambal, Y. Nirosha, E. Oliviya Roshini, "Lung Cancer Detection Using Machine Learning Techniques", IJAREEIE, vol.8, no 2, pp.266-271, February 2019.

[3] N. Panpaliya, N. Tadas, S. Bobade, R. Aglawe, A. Gudadhe, "A survey on early detection and prediction of lung cancer", *International Journal of Computer Science and Mobile Computing*, pp.175-184, 2015.

[4] Q. Song, L. Zhao, X. Luo, and X. Dou, "Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images", *Journal of Healthcare Engineering*, vol.2017, pp.1-7, Article ID 8314740.

[5] S. Sasikala, M. Bharathi, B.R. Sowmiya, "Lung Cancer Detection and Classification using Deep CNN", *IJITEE*, vol.8, no. 2S, pp. 259-262, December 2018.

[6] J. Kuruvilla, K. Gunavathi, "Lung cancer classification using neural networks for CT images", *Computer Methods and Programs in Biomedicine*, pp. 202-209, 2013.

[7] T. Fang, "A novel computer aided lung cancer detection method based on transfer learning from GoogLe Net and median intensity projections", *IEEE International Conference on Computer and Communication Engineering Technology*, pp.286-290, 2018.

[8] D. Moitra, R.Kr. Mandal, "Classification of non-small cell lung cancer using one-dimensional convolutional neural network", *Expert Systems with Applications*, pp.1-10, May 2020.

[9] R. Tekade, K. Rajeswari, "Lung cancer detection and classification using deep learning", *4th International Conference on Computing Communication, Control and Automation*, pp.1-5, 2018.

[10] A. Asuntha, A. Brindha, S. Indirani, A. Srinivasan, "Lung cancer detection using SVM algorithm and optimization techniques", *Journal of Chemical and Pharmaceutical Sciences*, pp.3198-3203, December 2016.

[11] N.S. Reddy, V. Khanaa, "Detection and Prediction of Lung cancer using Different algorithms", *IJEAT*, pp.2088-2093, September 2019.

[12] Q. Song, L. Zhao, X. Luo. X. Dou, "Using Deep Learning for Classification of lung Nodules on Computed tomography Images", *Journal of Healthcare Engineering*", pp.1-7, 2017.

[13] Vapnik, V.N. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998

[14] S. Khan, S. Hussain, S. Yang, K. Iqbal, "Effective and Reliable Framework for Lung Nodules Detection from CT Scan Images", *Nature*, pp.1-14,2019.

[15] F. Taher, N. Werghi and H. Al-Ahmad, "Computer Aided Diagnosis System for Early Lung Cancer Detection", *Algorithms*, vol. 8, no. 4, pp. 1088-1110, Nov. 2015, ISSN 1999-4893.

[16] T.A. Keziah, P. Haseena, "Lung Cancer Detection Using SVM Classifier and MFPCM Segmentation", *IRJET*, pp.3114-3118, 2018.

[17] L. Bertolaccini, P. Solli , A. Pardolesi , A. Pasini, "An overview of the use of artificial neural networks in lung cancer research", *Journal of Thoracic Disease*, pp.924-931, 2017.

[18] F. Taher, N. Werghi, H. Al-Ahmad and C. Donner, "Extraction and Segmentation of Sputum Cells for Lung Cancer Early Diagnosis", *Algorithms Journal of Machine Learning for Medical Imaging*, pp. 512-531, vol. 6, August 2013.

[19] N.M Numan, S.M. Abuelenin, M.Z. Rashad, "Prediction of Lung Cancer Using Artificial Neural Network", *IJICIS*, pp.1-19,2018.

[20] I.M. Nasser, S.S. Abu-Naser, "Lung Cancer Detection Using Artificial Neural Network", *IJEAIS*, pp. 17-23, 2019.

[21] S.R. Patale, R. Borhade, "Classification and Detection of Lung Cancer using Machine Learning Approach", *International Journal of Pure and Applied Mathematics*, vol.118, no. 24, pp. 1-9, 2018.

[22] Q. Wu, W. Zhao, "Small cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm", *International Symposium on Computer Science and Intelligent Controls*, October 2017, proceedings, 2017, pp.88-91.

[23] M. Buscema, "Back propagation Neural Networks", *Substance Use & Misuse*, pp.233-270, 1998.

[24] "*Data Science Bowl 2017 data description*," https://www.kaggle.com/c/data-science-bowl-2017, accessed:2017-01-13.

[25] G. Wang, "A perspective on deep imaging", *IEEE Access*, vol. 4, pp.8914–8924, 2016.

[26] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[27] O.D. Fenwa, F.A. Ajala, A.A. Adigun, "Classification of cancer of the lungs using SVM and ANN", *International Journal of Computers & Technology*, pp.6418-6426, October 2015.

[28] P. Naresh, R. Shettar, "Early Detection of Lung Cancer Using Neural Network Techniques", *International Journal of Engineering Research and Application*, pp.78-83, August 2014.

[29] https://www.cancer.org/cancer/lung-cancer/if-you-have-small-cell-lung-cancer-sclc.html.

[30] D. Jayaraj, S.Sathiamoorthy, "Random Forest based Classification Model for Lung Cancer Prediction on Computer Tomography Images", *ICSSIT 2019*, pp. 100-104, 2019.

[31] O. Gunaydin, M. Gunay, O. Sengel, "Comparison of Lung Cancer Detection Algorithms", *Proceedings of IEEE*, pp. 1-4, 2019.

[32] R. Sathishkumar, K. Kalaiarasan, A. Prabhakaran, M. Aravind, "Detection of Lung Cancer using SVM Classifier and KNN algorithm", *Proceedings of International Conference on System, Computation, Automation and Networking*, pp.1-7, 2019.

[33] N. S. Reddy, V Khanaa, "Detection and prediction of lung cancer using different algorithms", *International Journal of Engineering and Advanced Technology (IJEAT)*, pp.2088-2093, vol. 8, September 2019.

[34] M. Glatzer, A. Rittmeyer, J. Müller, I. Opitz, "Treatment of limited disease small cell lung cancer: the multidisciplinary team", *Thoracic Oncology*, pp.1-10, 2017.

Fatma Taher is an associate professor and assistant dean for research and outreach in the College of Technological Innovation at Zayed University, Dubai, UAE. Her research interests are in the areas of signal and image processing, pattern recognition, deep learning, Machine learning, artificial intelligence, medical image analysis, especially in detecting of the cancerous cells, kidney transplant and autism. In addition to that, watermarking, remote sensing and satellite images researches. Dr. Fatma Taher has published more than 70 papers in international Journals and conferences. She served as a member of the steering, organizing and technical program committees of many international conferences, and has served on many editorial and reviewing boards of international journals and conferences. Dr. Fatma is the chair of IEEE UAE section and the chair of the Education Committee in British Society in UAE. Dr. Fatma has received many distinguished awards such as: Best paper award of the first prize in the PhD Forum of the 20th IEEE International Conference on Electronics, Circuits, and Systems (ICECS), PhD Forum, Dec. 8-11, 2013. And recently, she was awarded the UAE Pioneers award as the first UAE to create a computer aided diagnosis system for early lung cancer detection based on the sputum color image analysis, awarded by H.H Sheik Mohammed Bin Rashed Al Maktoum, 15th Nov. 2015. In addition to that she was awarded an innovation award at the 2016 Emirati Women Awards by H. H. Sheik Ahmed Bin Saeed Al-Maktoum. Chairman of Civil Aviation Authority, and LOreal-UNESCO For Women in Science Middle East Fellowship 2017.

Neema Prakash is currently working as research assistant in the College of Technological Innovation at Zayed University, Dubai, UAE. She received M.tech degree in Optoelectronics and Optical communication from University of Kerala, India (2016), B.tech degree in Electronics and Communication from Mahatma Gandhi University, Kerala, India (2012). Her research interests are in the areas of image processing, designing electronic circuits and telecommunication. She has gained experience in image processing working in Raman Research Institute, India and has good experience in testing electronic circuits and components, working in Indian Space Research Organization, India.