BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 70(3), 2022, Article number: e139006
DOI: 10.24425/bpasts.2021.139006

Lung cancer detection using an integration of fuzzy K-means clustering and deep learning techniques for CT lung images

J. Maruthi Nagendra PRASAD[1]*, S. CHAKRAVARTY[1], and M. Vamsi KRISHNA[2]

[1] Centurion University of Technology and Management, Orissa, India
[2] Chaitanya Engineering College, Kakinada, India

Abstract. Computer aided detection systems are used for the provision of second opinion during lung cancer diagnosis. For early-stage detection and treatment false positive reduction stage also plays a vital role. The main motive of this research is to propose a method for lung cancer segmentation. In recent years, lung cancer detection and segmentation of tumors is considered one of the most important steps in the surgical planning and medication preparations. It is very difficult for the researchers to detect the tumor area from the CT (computed tomography) images.

The proposed system segments lungs and classify the images into normal and abnormal and consists of two phases, The first phase will be made up of various stages like pre-processing, feature extraction, feature selection, classification and finally, segmentation of the tumor. Input CT image is sent through the pre-processing phase where noise removal will be taken care of and then texture features are extracted from the pre-processed image, and in the next stage features will be selected by making use of crow search optimization algorithm, later artificial neural network is used for the classification of the normal lung images from abnormal images. Finally, abnormal images will be processed through the fuzzy K-means algorithm for segmenting the tumors separately. In the second phase, SVM classifier is used for the reduction of false positives. The proposed system delivers accuracy of 96%, 100% specificity and sensitivity of 99% and it reduces false positives. Experimental results shows that the system outperforms many other systems in the literature in terms of sensitivity, specificity, and accuracy. There is a great tradeoff between effectiveness and efficiency and the proposed system also saves computation time.

The work shows that the proposed system which is formed by the integration of fuzzy K-means clustering and deep learning technique is simple yet powerful and was effective in reducing false positives and segments tumors and perform classification and delivers better performance when compared to other strategies in the literature, and this system is giving accurate decision when compared to human doctor's decision.

Key words: fuzzy K-means; artificial neural networks; SVM; crow search optimization algorithm.

1. INTRODUCTION

Cancer can be defined as the unrestrained development of cells, and if this development is happening in the lungs, then it is called lung cancer. Among a lot of cancer deaths happening over the world, lung cancer holds the major share. Active smoking is considered as the one of the main causes for the lung carcinoma and passive smoking is another. Lung carcinoma is categorized into 4 stages, which specify the amount of cancer spread in the body parts. Recovery rate of lung carcinoma patients is very mediocre because lung cancer can only be detected by the specialists in the advanced stages. Lung cancers are broadly classified into 2 types, non-small cell lung cancers (NSCLC) and small cell lung cancers (SCLC). For NSCLC 5-year survival rate is 24% and for SCLC it is around 6%. Woman survival rate is different from men, with 23% for women and 16% for men. Computer aided diagnosis system is used for the detection of the lung cancer from image modalities and also it will provide second opinion for the radiologists who will enhance the survival rate of the patients. To decide on the treatment and for the patient's recovery rate enhancement, lung cancer should be diagnosed in the early stages.

Organs and different tissues functioning can be analyzed by the professionals by using various image modalities. Among the many image modalities, due to its ability to show the body parts accurately and its accuracy, CT image modalities are widely used for the lung cancer diagnosis. Even CT images can be used for the detection of lung cancers in the early stages.

The diagnosis of the carcinoma in the CT lung images is the most hectic and demanding assignment and extraction of features from the segmented nodule plays an important role in the diagnosis of the carcinoma in the CT lung images. Features like texture, shape, intensity, and context etc., are various features that can be extracted from the region of interest. Humongous number of features will be extracted and again all these features cannot be used for training the classifier and for testing the classifier, so a subset of features is to be selected with respect to their significance using various techniques like SVM,

*e-mail: maruthiprasad1986@gmail.com

Manuscript submitted 2021-06-16, revised 2021-07-15, initially accepted for publication 2021-08-31, published in June 2022.

kNN, LDA and decision tree etc., and now subset feature vector is used for the classification which will enhance the accuracy of the classifier and reduces cost incurred in the computation. It was observed and identified that artificial neural network classifier can deliver best results in the case of multi-category classification. There exists a lot of literature that shows various segmentation approaches used for the segmentation of tumor from CT lung images, but none was found to be so accurate.

In this work, we aim to develop and evaluate a novel lung cancer detection and segmentation approach that automatically predicts cancer nodules and classify them using a deep learning method which is trained on a public dataset of CT images.

Identified gap:

1. After careful observation it is identified that lots of literature concentrates on **sensitivity** with higher false positives per case.
2. Early detection is not a major issue in most of the state-of-the-art literature.
3. Texture features alone are considered in most of the state-of-the-art literature.
4. False positive reduction is not considered in most of the literature.

Contributions:

The main contribution of this work is:

1) the main objective of the proposed system is to detect lung carcinoma in the earlier stages by considering more features;
2) to use a fusion of texture features and deep learning based features;
3) fuzzy K-means clustering algorithm and deep learning approaches are used for the classification of normal and abnormal CT images;
4) the proposed system uses crow search optimization algorithm (CSOA) for the feature selection for the segmentation of tumor from the abnormal CT images;
5) SVM classification technique is used for the reduction of false positives and
6) finally, all the performance metrics are considered in this work.

Remaining article is presented as follows, we review all the related existing works on Lung Cancer detection and classification techniques in the related work section. Then, we present the methods and materials used, finally we will present the results and discuss the outcomes presented in this research.

2. RELATED WORK

A CAD system is presented by the author in [1], which uses a 3D mass-spring model used in fusion with reconstruction process of spline curves, and a neural network is used for the false positives reduction. A CAD system for the detection and segmentation of juxtavascular and juxtapleural nodules is proposed in [2], where thresholding based on the intensity is used for segmenting the images and K-means clustering is used for the segmentation and identification of nodules and SVM classifier is used for the reduction of the false positives.

In this work [3] an adaptive morphology-based technique is used for the segmentation of the nodule regions in the CT lung images. Adaptive morphological filter uses adaptive structuring elements for improving false positive reduction and to improve nodule detection, texture and intensity features are extracted and given to SVM for detecting the nodules.

A fully automatic system is proposed in [4] for the detection of nodules of any shape and size developed by making use of artificial intelligence and the system provides accuracy of 96.7%. A novel 3D shape based feature descriptor is proposed by [5] for the diagnosis of the CT lung nodules, feature descriptors will be extracted from the nodule candidates and refined using wall discarding technique and at last SVM classifier is used for the classification of nodules from non-nodules. A novel technique is proposed in [6], which uses an integration of fuzzy connection and the evolutionary computation is used for the segmentation of the pulmonary nodules.

A novel automatic system is proposed by [7] for the detection of lung nodules, integration of 2 filters i.e., line structure enhancement and blob like structure enhancement filter are used for the nodule candidates, front surface propagation procedure is used for the detection of the lung nodules. Here [8], a fully automatic system for diagnosing nodules is developed by the authors using region growing algorithm and thresholding algorithm for image segmentation and then ROI is extracted and nodule candidate is detected using self-adaptive template matching technique and the reduction of false positives will be done with the help of FLDA classifier by choosing eleven different features. A novel mechanism is proposed [9] for the discovery of nodules on CT lung images and SVM classifier is used for the reduction of false positives.

In this research [10] author integrates three CAD systems, where one system is responsible for vessel and nodule candidate segmentation and the second system is responsible for false positive reduction and nodule candidate identification and the last one is responsible for the detection of internal and juxtapleural nodules and reduces false positive.

In [11,12], author proposes a system for the diagnosis of nodules using fuzzy clustering models and SVM classifier is used for the classification. In this work [13], ROI is selected and then texture features are extracted from the CT image spectrum and these texture features are used to train the SVM which classify the normal and abnormal images.

The author of this work [14], proposes a fully automatic CAD system for the nodule detection in the CT images. Rule-based algorithms are the initial techniques used for segmenting tumors from the CT lung images; because of their elevated robust nature, these techniques are not appropriate for a massive range of data. Afterwards, a variety of fuzzy algorithms [15] which are adaptive in nature were used for segmenting the tumors from the CT lung images. These algorithms will have a negative impact from human bias and will not be able to handle the differences in the physical world data. Nowadays, with the advancement in the information technology fields, techniques based on artificial intelligence like machine learning or deep learning approaches [16, 17] are utilized for the segmentation of tumors from the CT lung images to overcome some of the

downsides of the specified problems. So, ANN and FKM techniques are used for the segmentation of tumors from CT lung images.

The author in this article [18], evaluates the performance of the CAD system and states that subsolid nodules are diagnosed better with computer-aided detection system when their thickness is 1 mm, rather than 3 or 5 mm section CTs. Author in this article [19] compares, various reading modes of CAD systems powered by artificial intelligence for the detection of lung nodules and it is concluded that the concurrent reading mode and second reading mode will enhance the performance of the CAD system. In this article [20], author evaluates the COMPUTER AIDED DETECTION SYSTEM based on deep learning techniques and also compares the performance of the physician with CAD systems and physician without a CAD system and concludes that physician's performance in the presence of CAD system for the detection of lung nodules is far better when compared to physician's performance in the absence of a CAD system.

In this article [21], author proposed a multimodal deep learning based computer aided detection and diagnosis system for the detection of lung nodules. This system achieves better performance when compared to state-of-the-art literature. Author in this article [22], made a comparison between the two systems targeting the solid pulmonary nodules, first one is a machine learning based CAD system and second one is CAD system based on 3D CNN. Author finally concluded that CAD system based on machine learning delivers lower specificity when compared to CAD system based on 3D CNN and they achieve a similar sensitivity.

In this work [23], author uses NASNet for the development of computer-aided detection system for the early stage detection of COVID-19 using CT images, and this system can classify non-COVID-19 CT images from COVID-19 IMAGES which achieves a decent performance when compared to state of the art literature and this system provided great support for the radiologists during this pandemic.

In this article [24], author proposed a CAD system for the detection of lung nodules and now evaluated the interpretation of the CT images using conventional system and with a CT interpretation system implemented on a cloud and concluded that this interpretation enhanced the nodule detection rate. In this article [25], author proposed a CAD system for the early stage detection of COVID-19 using image processing and machine learning tools. In this approach author prepared two datasets on with X-rays and other with CT images and the proposed system achieves the accuracy of 99.02% over the dataset with CT images and for the dataset with X-ray images the proposed system achieves 85.96% accuracy.

Author in this article [26], proposed a computer-aided detection system based on deep CNNs which are stacked for the detection of lung nodules. A CT image is given as input to the system and uses segmentation to extract region of interest and deliver extracted region of interest to the deep CNN stacked layers for lung nodule presence probability generations. Finally, the proposed system achieves the accuracy of 96.23% and 8 fp's/scan and 96.81% sensitivity. In this article [27], author proposed a system to enhance the reader's performance by using artificial intelligence for the diagnosis of the lung carcinoma on chest radiography. Finally, the authors concluded that the reader's performance can be enhanced using artificial intelligence algorithms.

In this article [28], author presents a review and comparison on various techniques proposed in the literature used for the nodule detection in the CT lung images. In this article [29], the author proposed a computer-aided detection system based on deep CNN for the detection of COVID-19, mini-sets are generated from the imbalanced datasets and apply transfer learning over the mini-sets and this system achieves 0.94 sensitivity.

Author in this article [30], proposed a computer-aided detection system for the detection of lymph nodes in the CT lung images. In this system, convolution network is used for the detection of lymph nodules and then detected lymph nodules are classified using 3D CNN and finally, the author concludes that the system achieves sensitivity of 65% with fp's/patient rate is at 5. In this article [31], authors propose a CAD system based on deep learning to find out the input matrix size which is optimal for the detection of the lung nodules and finally, concludes that optimal matrix size of 896 will make CAD system to deliver good performance when compared with state of the art literature. In this article [32], author proposed a computer-aided detection system based on deep learning algorithm for the detection of periapical lesions in the CT images. Finally, the author concluded that the system achieved good accuracy.

Author in this article [33], proposed a computer-aided detection system which uses pre-trained CNN combination for the generation of multi-CNN for the diagnosis of COVID-19 from the X-ray Images. Multi CNN is used for the feature extraction and extracted features will be given as input to correlation-based technique for the selection of optimal features and then Bayesnet classifier is used for the classification. Finally, the system achieves accuracy of 97.44%.

In this article [34], author presents a comparison between radiologists without CAD and radiologists with CAD for the detection of pulmonary nodules in the chest radiography. In this article [35], author validates the earlier developed computer-aided detection system based on deep learning approach for the detection of lung nodules by using this system in the screening program in china. Finally, the author concludes that the proposed system achieves 89.3% sensitivity while manually radiologists achieved 76% sensitivity.

In this article [36], author proposed a computer-aided detection system by using MANN for the detection of lung nodules. The proposed system achieves sensitivity of 66.42% without soft tissues and an accuracy of 66.76% and 2.5fp's/image. With soft tissues the proposed system achieves sensitivity of 72.85% and accuracy of 72.96% and 1fp's/image. By using low dose CT images, author in this article [37], proposed a CAD system for the detection of lung nodules in the early stages. An image is given as input to the system and preprocesses the image to remove the additive noise and to enhance the quality of the image. Pre-trained network VGG16, Alex and VGG19 are used for the extraction of features. Genetic algorithm is used to select the optimal features and these features are fed to a variety

of classifiers and finally identified that SVM classifier with the features obtained using pre-trained network VGG19 achieves 96.25% accuracy and 95% specificity and 97.5% sensitivity. In this article [38], author proposed a CAD system which is based on deep learning framework in the multi-scene environment.

In this article [39], author proposed a computer aided detection system for the detection of lung cancer in CT lung images. In the proposed system brightness preserving approach which is applied at multilevel to an image is used for the noise removal and to enhance the image quality. Region of interest is extracted by using deep neural network and then extract features. Later the extracted features are given as input to hybrid spiral optimization approach for selecting optimal features and then ensemble classifier is fed with these selected features for the purpose of classification.

In this article [40], author proposed a computer-aided diagnosis system for the detection of the lung cancer and two algorithms are used for generating the statements providing explanations about the decisions made by the CAD system. In this article [41], author proposed a CAD system based on deep learning and also presented a comparison between the readings of the radiologists without CAD and with CAD for the detection of pulmonary nodules. Finally, author concluded that the radiologists without CAD show less sensitivity when compared to the readings of the radiologists with CAD.

3. MATERIALS AND METHODS

Materials

Like the vast majority of previously mentioned research work, our CAD system will be trained and evaluated on a large open web accessible international dataset for lung cancer detection coordinated by Lung Image Database Consortium Image Collection (LIDC-IDRI) [42], which consists of 1018 cases. Each case consists of CT scan images whose slice thickness range from 0.6 mm to 5 mm and an XML file, annotated by well-trained radiologists. Radiologists followed a 2-phase annotation mechanism to arrive at definite assessment where they marked lesions placed with one of three classes' non-nodules whose size are at least 3 mm and nodules whose size are at least 3 mm and nodules with at most a size of 3 mm.

Methods

Figure 1 shows the complete framework of the proposed lung cancer detection method. The most important step in the framework is the classification of abnormal and normal lung images. Initially the input CT images are sent to pre-processing stage for noise removal, then important features are extracted using feature extraction phase and then CSOA is used for the feature selection and deep learning technique is used for the classification and finally tumor is extracted using fuzzy K-means algorithm. SVM classification is used for the reduction of the false positives.

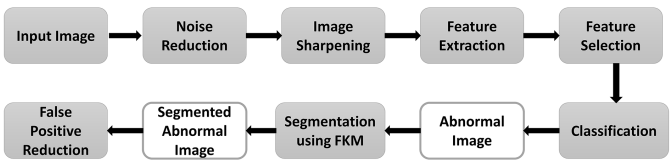

Fig. 1. Overall framework of the proposed system

Pre-processing

Input CT lung images are prone to a lot of noise, to improve the accuracy of the system that detects and classifies abnormality in CT lung images, the noise must be removed. So, in pre-processing, filters are used to enhance the clarity of the CT lung images by evacuating noise from the images. Here in this research, wiener filter is used for the removal of noise.

Based on the stochastic structure wiener filter was developed and is best described as follows:

$$w(f_1,f_2)=\frac{H*(f_1,f_2)\,S_{xx}(f_1,f_2)}{|H(f_1,f_2)|^2\,S_{xx}(f_1,f_2)+S_{\eta\eta}(f_1,f_2)}, \quad (1)$$

where $H(f_1,f_2)$ is the blurring filter and $S_{xx}(f_1,f_2)$, $S_{\eta\eta}(f_1,f_2)$ are the power spectrum of CT image and the noise. Inverse filtering and low pass filtering is used by wiener filter for holding the edges and removing the noise.

Feature extraction

Pre-processed image is next given to feature extraction phase where texture features will be extracted. GLCM properties which have a combination of statistical and morphological information are extracted from an image. Properties or features are standard deviation, sum variance, difference entropy, mean etc.

We do not need to manually extract features from the image in deep learning. By applying weights to its connections, the network automatically extracts information and learns their importance on the output. You give the network the raw image, and it finds patterns within the image as it passes through the network layers, resulting in features. In contrast to standard ML models that require hand-crafted features, neural networks can be regarded of as classifier + feature extractors that are end-to-end trainable.

Neural networks take all the available features and assign random weights. They modify these weights during the training phase to reflect their importance and how they should affect the output prediction. The patterns with the highest frequency of appearance will have higher weights, making them more useful features. Features with the lowest weights, on the other hand, will have relatively little impact on the output.

Deep learning-based feature extraction

CT image is given as input to the VGG-16, for the feature extraction [43, 44], the last max pooling layer is used for the extraction of the image features and this process is shown in Fig. 3.

Feature selection

Once the texture and deep features were extracted and fused [45], features with significance were chosen by making use of crow search optimization algorithm. This feature selec-

tion will reduce the computational cost and enhance the performance of the classification procedure. The framework for feature selection is depicted in Fig. 2.

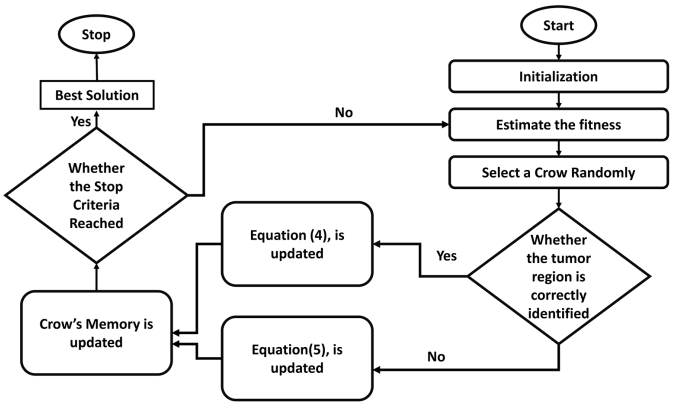

Fig. 2. Frameworks for feature selection using crow search optimization

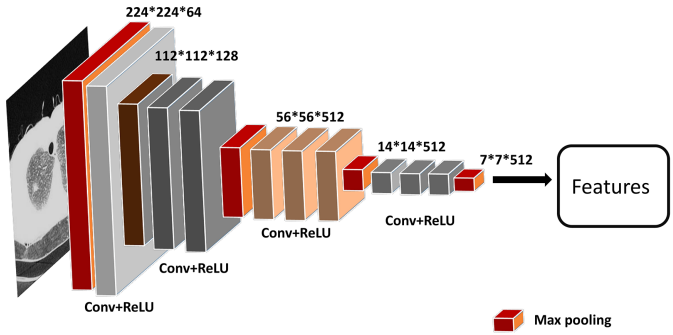

Fig. 3. Deep learning based feature extraction

- Initialization:
  From the feature set, random characteristics are selected and arranged as a matrix represented below, where N represents feature quantity.

$$C_i = \begin{bmatrix} F_{11}F_{12}\cdots F_{1m} \\ F_{21}F_{22}\cdots F_{2m} \\ F_{n1}F_{n2}\cdots F_{nm} \end{bmatrix}. \tag{2}$$

- Fitness estimation:
  In this step, for each crow fitness will be determined by making use of the fitness function, the location of each crow is determined and is accessed by making use of predetermined fitness function. Fitness function is given as

$$\text{Fitness function} = \frac{T_P + T_n}{Tp + F_P + F_n + Tn}. \tag{3}$$

  T_p, T_n stands for true positive and true negative and F_p, F_n stands for false positives and false negatives respectively.
- Updating crow's memory and identifying the tumor location:
  Once the fitness estimation is computed then we must find the location of tumor by using crow search optimization algorithm, for doing so we have to consider two scenarios.

Scenario-1: the following equation is used for updating the latest tumor location:

$$S_m^{t+1} = S_m^t + R_m \times FL_m^t \times \left(B_m^t - S_m^t\right), \tag{4}$$

where R_m is an integer value ranging between 0,1 and FL_m^t is the tumor m's flight length at iteration t.

Scenario-2: the following equation is used to update the tumor location randomly when we are dealing with non-tumor locale:

$$S_m^{t+1} = \begin{cases} \text{if } R_m \geq P_n^t, \\ \text{else.} \end{cases} \tag{5}$$

If if $R_m \geq P_n^t$ then use equation (4) to update, otherwise update randomly the position of the tumor.

- Termination criteria:
  The process of feature selection is terminated when the extreme number of iterations are accomplished or when the best solution has been achieved. In this research the highest number of cycles sets used is 40.

Deep learning-based classification

Once we have the significant features extracted then we will use them to train our artificial neural network, our ANN will consist of input, output and concealed layers, the significant features that are chosen by making use of the crow search optimization algorithm are given as input to the input layer and then random weights are distributed to input-hidden and then to hidden-output layer. The ANN classifier is shown in Fig. 4.

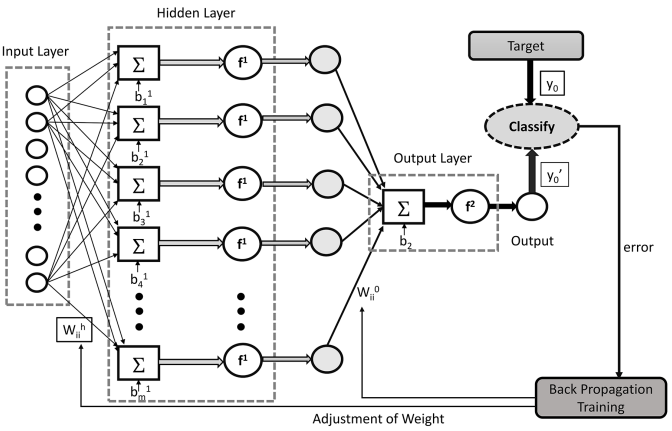

Fig. 4. Artificial neural network for classification

Firstly, the feature given to input layer will be multiplied weights to get H_j, for the hidden layer node j and it's generated by making use of the following equation:

$$H_j = x_j + \sum_{k=1}^{n} U_k w_{ij}^h. \tag{6}$$

Activation function is given as follows:

$$f(H_j) = \frac{1}{1 + e^{-H_j}}. \tag{7}$$

Output function is as follows:

$$O_j = x_i + \sum_{j=1}^{b} w_{i,j}^0 f(H_j). \tag{8}$$

Error value obtained can be given as follows:

$$E = \frac{1}{2n} \sum_{i=0}^{h-1} \sqrt{(T_i - y_i)^2}. \tag{9}$$

Here, n represents parameter number, T_i target value and y_i represents output value. Output layer is assigned with two classes and are represented as follows:

$$\text{Output} = \begin{cases} N1 & \text{for } 0 \le TH < 0.75, \\ N2 & \text{for } 0.75 \le TH. \end{cases} \tag{10}$$

FKM based segmentation

Abnormal images collected after classifications are given to FKM algorithm for segmenting tumors. Fuzzy K-means clustering algorithm arbitrarily chooses a center for a cluster initially, such that not more than 2 clusters have the same centroid. Now update the membership function for computing new centroids, after repeating this process for K number of times form a new group with the same image pixels and nearest new cluster center. For each iteration centers of the K-new clusters change their position until their position becomes stable. Likewise, Euclidean distance between the centroid and pixels will be minimized by FKM clustering algorithm. When Euclidean distance is minimum, distortion is also minimized, hence the distortion in the objective function also reduced.

Objective function is given as

$$J = \sum_{j=1}^{k} \sum_{i=1}^{N} u_{i,j}^m d_{i,j}, \tag{11}$$

where N is the data point's number, squared Eculidean distance between the image pixels and the centroid is $d_{i,j}$.

Membership function $u_{i,j}$ must gratify the below condition:

$$\sum_{j=1}^{N} u_{i,j} = 1, \quad \text{for } i = 1 \text{ to } N. \tag{12}$$

Fuzzy K-means algorithm is as follows:

1. A set of initial clusters is chosen initially and set $p = 1$.
2. Then compute the squared Euclidean distance $d_{i,j}$.
3. And now membership function $u_{i,j}$ is updated:

$$u_{i,j} = \left((d_{i,j})^{1/m-1} \sum_{l=1}^{k} \left(\frac{1}{d_{j,l}} \right)^{\frac{1}{m}-1} \right)^{-1}. \tag{13}$$

Here, $l \neq j$ if $d_{i,j} < \eta$, then set $u_{i,j} = 1$, where η is the small positive number.

4. By taking the initial cluster center calculate the new cluster center set by making use of the following equation:

$$c_j = \frac{\sum_{i=1}^{N} u_{i,j}^m X_i}{\sum_{i=1}^{N} u_{i,j}^m}. \tag{14}$$

5. If $\|c_j - c_{j-1}\| < \varepsilon$ for various values of j ranging from $1-k$, then terminate the iteration, else set $p+1 \to p$ and go to step 2.

4. FALSE POSITIVE REDUCTION

In this phase, by leaving out true positives as is, we try to reduce false positives. In the proposed system vessels, scarring and stains are considered as FPs. FPs are normally eliminated by making use of classification algorithms which is also responsible for identifying the distinctions between non-nodules and nodules. Here the features extracted from the image using histogram of oriented [46] which uses intensity distribution and pixels slopes to identify the objects look and feel.

Advantages of histogram of oriented gradient:

- for image recognition and object identification, histogram of oriented gradients method is utilized;
- HOG is responsible for extracting relevant information and ignore the irrelevant information;
- every individual pixel's direction, magnitude of the vertical component gradients and horizontal component gradients are calculated by the histogram of oriented gradients method and then generate a histogram of 9-bin is generated by arranging the information calculated to find out the data shifts;
- a less biased and optimal model can be made by utilizing the block normalization further;
- application areas of histogram of oriented gradients method are AR and VR and autonomous vehicles field.

For extracting features, the image is first divided into minor areas, in each area a pixel is selected, and its slope is calculated with the help of the equation (15) and this is continued till all the pixels in the region are visited.

$$\nabla f = \frac{\partial f}{\partial x} X + \frac{\partial f}{\partial y} Y, \tag{15}$$

where $\frac{\partial f}{\partial x}$ slope in the direction of X, $\frac{\partial f}{\partial y}$ slope in the direction of Y.

Once all the pixels in the area are completed and all the areas are completed, for each area slope histogram is computed, where the slope indicates the direction of some important textures [47].

These features are given to train the support vector machine classifier for reducing the false positives. Table 1 shows the state-of-the-art classifiers used for the reduction of false positives.

Table 1
State-of-the-art classifiers used for the reduction of false positives

S. no.	State-of-the-art classifiers used for the reduction of false positives	
1.	Support vector machine	[44, 45, 47, 48]
2.	ANN	[4]
3.	Bayesian supervised	[48]
4.	K-nearest neighbours	[49]
5.	Invariant moments	[41]
6.	Rule based	[8]
7.	Fuzzy k-NN classifier	[5]
8.	Feed forward neural networks	[10]
9.	Linear discriminant analysis	[11]
10.	Naive Bayesian and logistic regression	[45]
11.	Optimum path forest	[49, 50]
12.	Massive training neural network	[35]
13.	Lung-RADS^{TM}	[51]
14.	Multivariable logistic regression VGG-16 as feature extractor and support vector machine (SVM) as classifier	[52]
15.	A rule-based classifier followed by multi-view CNN	[53]
16.	Multi-scale gradual integration convolutional neural network (MGI-CNN)	[54]
17.	Multi-scale heterogeneous three-dimensional (3D) convolutional neural network (MSH-CNN)	[55]
18.	A new end-to-end 3D deep convolutional neural net (DCNN), called NoduleNet	[56]
19.	Multi-ringed (MR) – forest framework, against the resource-consuming neural networks (NN) – based architectures	[57]
20.	Combined multiple machine learning-based methods are used	[58]
21.	3D probabilistic deep learning system	[59]

5. RESULTS AND DISCUSSIONS

The proposed method has been tested on various CT images and it has been simulated using MATLAB R2017b. From LIDC-IDRI a dataset is considered for this testing.

The results obtained from this simulation are shown in the figures, first the original image is taken from the dataset which is shown in Fig. 5a and then image is pre-processed as shown in Fig. 5b where we will be improving the image and next we will optimize the image using CSOA algorithm and the resultant image is shown in Fig. 5c and later FKM clustered image is shown in Fig. 5d and finally segmented tumor is shown in Fig. 5e.

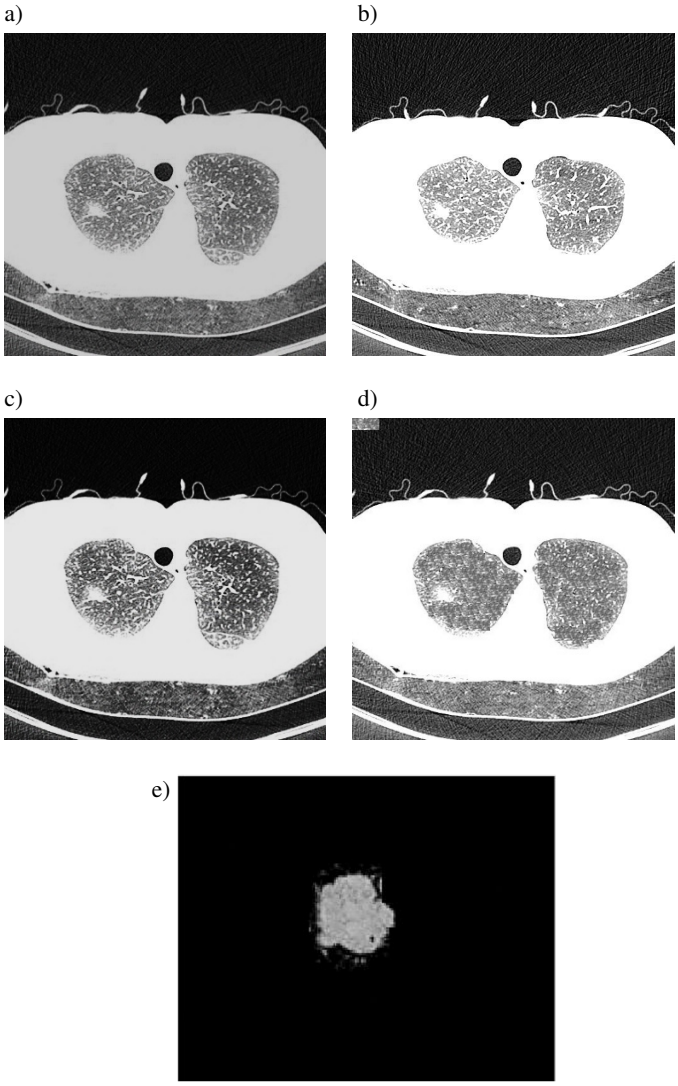

Fig. 5. Results obtained using MATLAB R2017b simulation: a) original CT image, b) pre-processed image, c) CSOA optimized image, d) clustered image, e) segmented tumor

Performance analysis

Performance comparison of state-of-the-art systems in the literature is shown in Table 2.

It is evident from the above table, that the proposed method has achieved a sensitivity of 99% and accuracy of 96% and specificity of 100%. It is also capable of reducing false positives to a greater extent when compared to state-of-the-art literature. Table 3 and Fig. 6 show the comparison of various pre-trained networks.

Metrics used for assessment

Assessment of the proposed system has been done by making use of following measures,

Accuracy will tell you how precise the proposed method is and it is determined by the following equation:

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + F_p + F_n + T_n}. \tag{16}$$

Table 2
Performance comparison of state-of-the-art systems in the literature

Systems	Sensitivity	FP's/Case	Accuracy	Specificity
[9]	80.00%	4.2	–	–
[8]	90.24%	4.54	94.02%	94.20%
[3]	93.45%	1.8	94.88%	94.27%
[14]	82.81%	–	–	–
[11]	82.05%	–	80.36%	76.47%
[4]	92.10%	–	96.7%	94.30%
[2]	95.70%	–	–	94%
[5]	97.90%	6.76	97.4%	97.7%
[6]	83.03% and 91.12%	0.49% and 9.15%	–	–
[47]	82.3	9.2	–	–
[1]	97.00% and 88.00%	6.1 and 2.5	–	–
[7]	95.00% and 91.50%	9.8 and 10.5	–	–
[26]	86.00%	4.9	90%	92%
Proposed method	99%	1.0	96%	100%

Table 3
Comparison of pre-trained network

Pre-trained	Accuracy	Specificity	Sensitivity
VGG-16	78.39	81.66	80.84
VGG-19	81.56	84.96	84.11
ResNet-18	82.42	85.85	85.00
Proposed System with ResNet-50	96.00	100.00	99.00
ResNet-101	87.10	90.73	89.82
GoogleNet	83.84	87.33	86.46
Incep.ResNet-V2	81.16	84.54	83.70
Inception-V3	84.05	87.55	86.68

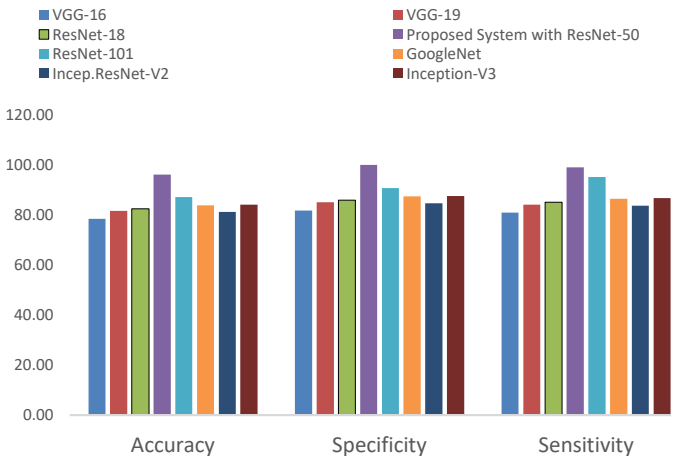

Fig. 6. Comparison of pre-trained network

Sensitivity is the capacity of the system to accurately classify the actual positives and is given by the following equation:

$$\text{Sensitivity} = \frac{T_p}{T_p + F_n}. \tag{17}$$

Specificity is the ability of the system to identify accurately negatives. It is given by the following equation:

$$\text{Specificity} = \frac{T_n}{T_n + F_n}. \tag{18}$$

ROC curve of the proposed system is shown in Fig. 7, orange line in the figure represents the ROC curve when reduction of false positive is disabled and blue line in the figure represents the ROC curve when false positive reduction is enabled where FPs are reduced by 21%. The proposed system can identify almost 90% of all nodules in the data set with 2.1 FPs per case.

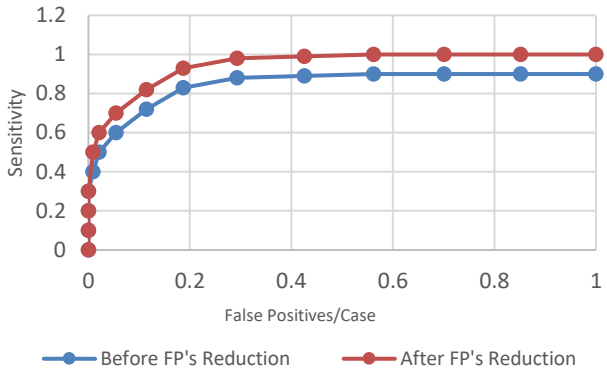

Fig. 7. ROC curve of the proposed system

6. CONCLUSION

In this research, an integration of deep learning and FKM approaches are used for segmenting tumor from CT lung images. The most important step in the framework is the classification of abnormal and normal lung images. Initially the input CT images are sent to a pre-processing stage for noise removal, then important features are extracted using feature extraction phase and then CSOA is used for the feature selection and deep learning technique is used for the classification and finally tumor is extracted using fuzzy K-means algorithm. SVM classification is used for the reduction of the false positives. The proposed system delivers the accuracy of 96% and it reduces false positives, delivering better performance when compared to state-of-the-art literature, and this system delivers an accurate decision when compared to human doctor's decision. Finally, it concludes that, this system can be used by the radiologists in the real time scenario for taking a second opinion before confirming the disease.

One of the limitations of the proposed system is the fact that juxta-pleural, well-circumscribed, vascularized and pleural tail nodules are not identified and staging of cancer is not done in the proposed system.

In the future, we plan to use various features selection mechanisms and identify which feature selection mechanism delivers

better accuracy and efficiency in the classification and tumor segmentation. Finally various features extraction methods will be used to evaluate the performance of the system.

REFERENCES

[1] D. Cascio, R. Magro, F. Fauci, M. Iacomi, and G. Raso, "Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models," *Comput. Biol. Med.*, vol. 42, no. 11, pp. 1098–1109, 2012, doi: 10.1016/j.compbiomed.2012.09.002.

[2] M. Javaid, M. Javid, M.Z.U. Rehman, and S.I.A. Shah, "A novel approach to CAD system for the detection of lung nodules in CT images," *Comput. Methods Programs Biomed.*, vol. 135, pp. 125–139, 2016, doi: 10.1016/j.cmpb.2016.07.031.

[3] A. Halder, S. Chatterjee, D. Dey, S. Kole, and S. Munshi, "An adaptive morphology based segmentation technique for lung nodule detection in thoracic CT image," *Comput. Methods Programs Biomed.*, vol. 197, p. 105720, 2020, doi: 10.1016/j.cmpb.2020.105720.

[4] T. Yanagihara and H. Takizawa, "Pulmonary nodule detection from X-ray CT images based on region shape analysis and appearance-based clustering," *Algorithms*, vol. 8, no. 2, pp. 209–223, 2015, doi: 10.3390/a8020209.

[5] W.J. Choi and T.S. Choi, "Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 37–54, 2014, doi: 10.1016/j.cmpb.2013.08.015.

[6] P. Badura and E. Pietka, "Soft computing approach to 3D lung nodule segmentation in CT," *Comput. Biol. Med.*, vol. 53, pp. 230–243, 2014, doi: 10.1016/j.compbiomed.2014.08.005.

[7] B. Chen *et al.*, "Automatic segmentation of pulmonary blood vessels and nodules based on local intensity structure analysis and surface propagation in 3D chest CT images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 7, no. 3, pp. 465–482, 2012, doi: 10.1007/s11548-011-0638-5.

[8] J. Gong, J. Yu Liu, L. Jia Wang, B. Zheng, and S. Dong Nie, "Computer-aided detection of pulmonary nodules using dynamic self-adaptive template matching and a FLDA classifier," *Phys. Medica*, vol. 32, no. 12, pp. 1502–1509, 2016, doi: 10.1016/j.ejmp.2016.11.001.

[9] A. Teramoto and H. Fujita, "Fast lung nodule detection in chest CT images using cylindrical nodule-enhancement filter," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 8, no. 2, pp. 193–205, 2013, doi: 10.1007/s11548-012-0767-5.

[10] N. Camarlinghi *et al.*, "Combination of computer-aided detection algorithms for automatic lung nodule identification," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 7, no. 3, pp. 455–464, 2012, doi: 10.1007/s11548-011-0637-6.

[11] S. Sivakumar and C. Chandrasekar, "Lung nodule detection using fuzzy clustering and support vector machines," *Int. J. Eng. Technol.*, vol. 5, no. 1, pp. 179–185, 2013.

[12] G.T. Reddy, M.P.K. Reddy, K. Lakshmanna, D.S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis," *Evol. Intell.*, vol. 13, no. 2, pp. 185–196, 2020, doi: 10.1007/s12065-019-00327-1.

[13] H.M. Orozco, O.O.V. Villegas, L.O. Maynez, V.G.C. Sanchez, and H.D.J.O. Dominguez, "Lung nodule classification in frequency domain using support vector machines," *2012 11th Int. Conf. Inf. Sci. Signal Process. their Appl. ISSPA 2012*, no. July, pp. 870–875, 2012, doi: 10.1109/ISSPA.2012.6310676.

[14] T. Messay, R.C. Hardie, and S.K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Med. Image Anal.*, vol. 14, no. 3, pp. 390–406, 2010, doi: 10.1016/j.media.2010.02.004.

[15] G.T. Reddy and N. Khare, "Hybrid Firefly-Bat optimized fuzzy artificial neural network based classifier for diabetes diagnosis," *Int. J. Intell. Eng. Syst.*, vol. 10, no. 4, pp. 18–27, 2017, doi: 10.22266/ijies2017.0831.03.

[16] B. Chandrasekaran and S. Fernandes, "Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus. The COVID-19 resource centre is hosted on Elsevier Connect, the company ' s public news and information website," *Diabetes Metab Syndr.*, vol. 14(4), no. January, pp. 337–339, 2020.

[17] N. Deepa *et al.*, "An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier," *J. Supercomput.*, vol. 77, no. 2, pp. 1998–2017, 2021, doi: 10.1007/s11227-020-03347-2.

[18] S. Park *et al.*, "Computer-aided detection of subsolid nodules at chest CT: improved performance with deep learning–based CT section thickness reduction," *Radiology*, vol. 299, no. 1, pp. 211–219, 2021.

[19] H.-H. Hsu *et al.*, "Performance and reading time of lung nodule identification on multidetector CT with or without an artificial intelligence-powered computer-aided detection system," *Clin. Radiol.*, vol. 76, pp. 626.e23-626.e32, 2021.

[20] S.Y. Choi, S. Park, M. Kim, J. Park, Y.R. Choi, and K.N. Jin, "Evaluation of a deep learning-based computer-aided detection algorithm on chest radiographs: Case–control study," *Medicine (Baltimore).*, vol. 100, no. 16, p. e25663, 2021.

[21] A.B. Menegotto and S.C. Cazella, "Multimodal Deep Learning for Computer-Aided Detection and Diagnosis of Cancer: Theory and Applications," *Enhanc. Telemed. e-Health Adv. IoT Enabled Soft Comput. Framew.*, pp. 267–287, 2021.

[22] R.M. Perl, R. Grimmer, T. Hepp, and M.S. Horger, "Can a Novel Deep Neural Network Improve the Computer-Aided Detection of Solid Pulmonary Nodules and the Rate of False-Positive Findings in Comparison to an Established Machine Learning Computer-Aided Detection?," *Invest. Radiol.*, vol. 56, no. 2, pp. 103–108, 2021.

[23] M. Ghaderzadeh, F. Asadi, R. Jafari, D. Bashash, H. Abolghasemi, and M. Aria, "Deep Convolutional Neural Network–Based Computer-Aided Detection System for COVID-19 Using Multiple Lung Scans: Design and Implementation Study," *J. Med. Internet Res.*, vol. 23, no. 4, p. e27468, 2021.

[24] E.J. Hwang, J.M. Goo, H.Y. Kim, J.Yi, S.H. Yoon, and Y. Kim, "Implementation of the cloud-based computerized interpretation system in a nationwide lung cancer screening with low-dose CT: comparison with the conventional reading system," *Eur. Radiol.*, vol. 31, no. 1, pp. 475–485, 2021.

[25] A. Saygılı, "A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods," *Appl. Soft Comput.*, vol. 105, p. 107323, 2021.

[26] M.G. Nakrani, G.S. Sable, and U.B. Shinde, "Lung Nodule Detection from Computed Tomography Images Using Stacked Deep Convolutional Neural Network," in *Intelligent Systems, Technologies and Applications*, Springer, 2021, pp. 237–246.

[27] H. Yoo *et al.*, "AI-based improvement in lung cancer detection on chest radiographs: results of a multi-reader study in NLST dataset," *Eur. Radiol.*, pp. 1–11, 2021.

[28] S.R. Ziyad, V. Radha, and T. Vayyapuri, "Overview of computer aided detection and computer aided diagnosis systems for lung nodule detection in computed tomography," *Curr. Med. Imaging*, vol. 16, no. 1, pp. 16–26, 2020.

[29] B. Narayanan, R. Hardie, V. Krishnaraja, C. Karam, and V. Davuluru, "Transfer-to-Transfer Learning Approach for Computer Aided Detection of COVID-19 in Chest Radiographs," *AI*, vol. 1, no. 4, pp. 539–557, 2020, doi: 10.3390/ai1040032.

[30] E. Kurniawan, P. Prajitno, and D.S. Soejoko, "Computer-Aided Detection of Mediastinal Lymph Nodes using Simple Architectural Convolutional Neural Network," *J. Phys. Conf. Ser.*, vol. 1505, no. 1, p. 12018, 2020.

[31] Y.-G. Kim, S.M. Lee, K.H. Lee, R. Jang, J.B. Seo, and N. Kim, "Optimal matrix size of chest radiographs for computer-aided detection on lung nodule or mass with deep learning," *Eur. Radiol.*, vol. 30, no. 9, pp. 4943–4951, 2020.

[32] F.C. Setzer *et al.*, "Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images," *J. Endod.*, vol. 46, no. 7, pp. 987–993, 2020.

[33] B. Abraham and M.S. Nair, "Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier," *Biocybern. Biomed. Eng.*, vol. 40, no. 4, pp. 1436–1445, 2020.

[34] M. Haber, A. Drake, and J. Nightingale, "Is there an advantage to using computer aided detection for the early detection of pulmonary nodules within chest X-Ray imaging?," *Radiography*, vol. 26, no. 3, pp. e170–e178, 2020.

[35] X. Cui *et al.*, "Validation of a deep learning-based computer-aided system for lung nodule detection in a Chinese lung cancer screening program." *Eur. Respiratory Soc.*, p. 4168, 2020, doi: 10.1183/13993003.congress-2020.4168.

[36] K. Rajagopalan and S. Babu, "The detection of lung cancer using massive artificial neural network based on soft tissue technique," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1–13, 2020.

[37] A. Elnakib, H.M. Amer, and F.E.Z. Abou-Chadi, "Early lung cancer detection using deep learning optimization," *Int. J. online Biomed. Eng.*, vol. 16, no. 6, pp. 82–94, 2020, doi: 10.3991/ijoe.v16i06.13657.

[38] Q. Zhang and X. Kong, "Design of Automatic Lung Nodule Detection System Based on Multi-Scene Deep Learning Framework," *IEEE Access*, vol. 8, pp. 90380–90389, 2020.

[39] P.M. Shakeel, M.A. Burhanuddin, and M.I. Desa, "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier," *Neural Comput. Appl.*, pp. 1–14, 2020.

[40] A. Meldo, L. Utkin, M. Kovalev, and E. Kasimov, "The natural language explanation algorithms for the lung cancer computer-aided diagnosis system," *Artif. Intell. Med.*, vol. 108, p. 101952, 2020.

[41] T. Kozuka *et al.*, "Efficiency of a computer-aided diagnosis (CAD) system with deep learning in detection of pulmonary nodules on 1-mm-thick images of computed tomography," *Jpn. J. Radiol.*, vol. 38, no. 11, pp. 1052–1061, 2020.

[42] S.G. Armato *et al.*, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, 2011, doi: 10.1118/1.3528204.

[43] D. Vasan *et al.*, "IMCFN: Image-based malware classification using fine-tuned convolutional neural network architecture," *Comput. Networks*, vol. 171, p. 107138, 2020.

[44] E. Kot, Z. Krawczyk, K. Siwek, L. Krolicki, and P. Czwarnowski, "Deep learning based framework for tumour detection and semantic segmentation," *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 69, p. e136750, 2021, doi: 10.24425/bpasts.2021.136750.

[45] F. Gil and S. Osowski, "Fusion of feature selection methods in gene recognition," *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 69, p. e136748, 2021, doi: 10.24425/bpasts.2021.136748.

[46] M. Firmino, G. Angelo, H. Morais, M.R. Dantas, and R. Valentim, "Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy," *Biomed. Eng. Online*, vol. 15, no. 1, pp. 1–17, 2016, doi: 10.1186/s12938-015-0120-7.

[47] R. Kruse and M. Steinbrecher, "Visual data analysis with computational intelligence methods," *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 58, no. 3, pp. 393–401, 2010, doi: 10.2478/v10175-010-0037-z.

[48] A. El-Baz, A. Elnakib, M. Abou El-Ghar, G. Gimel'Farb, R. Falk, and A. Farag, "Automatic detection of 2D and 3D lung nodules in chest spiral CT scans," *Int. J. Biomed. Imaging*, vol. 2013, p. 517632, 2013, doi: 10.1155/2013/517632.

[49] A.S. Iwashita *et al.*, "A path- and label-cost propagation approach to speedup the training of the optimum-path forest classifier," *Pattern Recognit. Lett.*, vol. 40, no. 1, pp. 121–127, 2014, doi: 10.1016/j.patrec.2013.12.018.

[50] T.M. Nunes, A.L.V. Coelho, C.A.M. Lima, J.P. Papa, and V.H.C. De Albuquerque, "EEG signal classification for epilepsy diagnosis via optimum path forest – A systematic assessment," *Neurocomputing*, vol. 136, pp. 103–123, 2014, doi: 10.1016/j.neucom.2014.01.020.

[51] M. Kaminetzky *et al.*, "Effectiveness of Lung-RADS in reducing false-positive results in a diverse, underserved, urban lung cancer screening cohort," *J. Am. Coll. Radiol.*, vol. 16, no. 4, pp. 419–426, 2019.

[52] Z. Shi *et al.*, "A deep CNN based transfer learning method for false positive reduction," *Multimed. Tools Appl.*, vol. 78, no. 1, pp. 1017–1033, 2019.

[53] S.A. El-Regaily, M.A.M. Salem, M.H. Abdel Aziz, and M.I. Roushdy, "Multi-view Convolutional Neural Network for lung nodule false positive reduction," *Expert Syst. Appl.*, vol. 162, p. 113017, 2020, doi: 10.1016/j.eswa.2019.113017.

[54] B.-C. Kim, J.S. Yoon, J.-S. Choi, and H.-I. Suk, "Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection," *Neural Networks*, vol. 115, pp. 1–10, 2019.

[55] Z. Xiao, N. Du, L. Geng, F. Zhang, J. Wu, and Y. Liu, "Multi-scale heterogeneous 3D CNN for false-positive reduction in pulmonary nodule detection, based on chest CT images," *Appl. Sci.*, vol. 9, no. 16, p. 3261, 2019.

[56] H. Tang, C. Zhang, and X. Xie, "Nodulenet: Decoupled false positive reduction for pulmonary nodule detection and segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 266–274.

[57] H. Zhu *et al.*, "MR-forest: A deep decision framework for false positive reduction in pulmonary nodule detection," *IEEE J. Biomed. Heal. informatics*, vol. 24, no. 6, pp. 1652–1663, 2019.

[58] P. Petousis, A. Winter, W. Speier, D.R. Aberle, W. Hsu, and A.A.T. Bui, "Using sequential decision making to improve lung cancer screening performance," *IEEE Access*, vol. 7, pp. 119403–119419, 2019.

[59] O. Ozdemir, R.L. Russell, and A.A. Berlin, "A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose CT scans," *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1419–1429, 2019.