



Automatic detection of solitary pulmonary nodules using superpixels segmentation based iterative clustering approach

S. Palanikumar¹ · S. Albert Jerome² · J. P. Jayan³

Received: 7 July 2020 / Accepted: 11 March 2021 / Published online: 21 March 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Lung disease is the premier cause of cancer deaths all over the world. The lung cancer diagnosis is mediocre at an early stage since it is impotent by the radiologist. Various investigations conducted so far manifest clearly that the nodule segmentation algorithms are inefficient. Thus, this investigation has centralized superpixels segmentation based iterative clustering (CSSBIC) and a sophisticated optimization approach for detailed segmentation of pulmonary nodules. The supreme intent of this research paper is the enhancement of lung CT images to identify the tumor effectively and small-scale anomalous nodules segmentation in the lung area. The initial phase an-isotropic diffusion with masking (AIDME) enhancement techniques which can eliminate the noise discern in the images. The next step is to apply the centralized superpixels segmentation based iterative clustering (CSSBIC) algorithm by using an improved picture sequence nodule for abnormal lung tissue prediction. The lung nodule is essentially retrieved using a GWO, based on deep learning techniques, with advanced ONN (ONN) and advanced CNN (CNN). The average time for segmenting the Nodule Slice order is 1.06 s. The best classification accuracy is 97% via GWO based advanced one nearest neighbour (AONN) and 97.6% by GWO based advanced convolutional neural network (ACNN) classifier.

Keywords Lung cancer · Centralized superpixels segmentation based iterative clustering (CSSBIC) algorithm · MSE · PSNR · SSIM · GWO based advanced one nearest neighbour (AONN) · GWO based advanced convolutional neural network (ACNN)

1 Introduction

Unforeseen deaths are often caused by cancer. Different international logbooks revealing that lung cancer tops the diagram of the Homo sapiens early for death. Previous diagnoses can de-escalate it so that oncologists may prescribe suitable treatment in a specified period. A fruit burgeon found in a group of cells causes cancer and is formed through tissue invasion by malignant tumors. CT scan images are favored in this article, Lung cancer. The most sensitive and unusual identification mode offers cross-sectional representations for correct fields of the scanned items with the CT scan or computerized axial tomography (CAT) scan (Mathews and Jeyakumar 2020). This study is aimed at developing a system for supplying CT images as inputs and as desired outputs. During the automated recognition of lung nodules, various researchers proposed diverse methods from a literature point of view. Nonetheless, four processes may follow all existing methods for the classification of the

✉ S. Palanikumar
palanikumar@niuniv.com

S. Albert Jerome
albert_s@rediffmail.com

J. P. Jayan
jeyanjp@gmail.com

¹ Department of Information Technology, Noorul Islam Centre for Higher Education, Thuckalay, Kumaracoil, Tamil Nadu 629180, India

² Department of Biomedical Engineering, Noorul Islam Centre for Higher Education, Thuckalay, Kumaracoil, Tamil Nadu 629180, India

³ Department of Software Engineering, Noorul Islam Centre for Higher Education, Thuckalay, Kumaracoil, Tamil Nadu 629180, India

pulmonary nodule, including pre-processing, segmentation, characteristic extraction, and classification.

Proposed the CAD system for radiologist analysis the lung nodule and improve the radiologist performance to detect the lung lesion with the time consuming and error-free process. It is because even an experienced doctor will not be able to perform an accurate diagnosis with a single slice, thus numerous successive slices were taken for precise diagnosis. Disparate the other traditional technique, which necessitates several slices to make an accurate assessment, a new technique has been proposed which demonstrate that an effective detection and diagnosis in just a single slice can be attained and is comparatively less time-consuming. The key target of the proposed method is to enhance the accuracy and diminish the analysis time, false positive reduction and subjectivity to accurate segmentation of anomalous lung tissue with analysis of CT (computed tomography) scans. The recent contribution of our proposed work includes:

- Enhancement purpose: proposed method is an-isotropic diffusion with masking enhancement (AIDME).
- Segmentation purpose: proposed method is centralized superpixels segmentation based iterative clustering (CSSBIC) algorithm.
- Classification purpose: proposed method is a GWO based advanced one nearest neighbour (AONN) and GWO based advanced convolutional neural network (ACNN).

2 Literature survey

As per the research in current biological world, parallelizing and combining multiple algorithms can improve the efficiency of lung image segmentation methods. On the other side, based on Vijila Rani and Jawhar (2018a) review, we are able to develop robust segmentation method capable of performing lung tumor area boundary detection and classification in an efficient way. Haney et al. (2001) have been computed the estimate of Gaussian function using population oriented tissue maps. Gaussian function provides a bell patterned probability distribution that reflects the intensities of particular tissue classes. The tissue contrast is enhanced and the region is differentiated using K-nearest neighbor (KNN) algorithm. Color ray casting (CRC) method was adapted by Jiang et al. (2004) have to separate the region of interest from the lung image background. Image dependency is the major drawback of this method and gray scale images are out of the scope of this method. Rahman et al. (2007) in their work have proposed a novel machine learning based image pre filtering approach with the combination of a statistical similarity matching technique and relevance feedback scheme to update the query parameters dynamically in a medical image database. To reduce the search space they

have used a probabilistic multi-class support vector machine (SVM) and fuzzy C-Mean (FCM) clustering for categorization and pre-filtering of images. In order to evaluate the performance of their system they have used a medical image database that includes images with ground truth corresponding to the modality, examined body region or orientation.

Sankar and Prabhakaran (2013) have investigated superiority and precision in medical images. Image quality assessments as well as enhancement stage were adopted on low pre-processing techniques based on Gabor filter within Gaussian rules. It is extremely important to determine the irregularity problems in objective figures, particularly in a variety of malignancy tumours such as lung cancer, breast cancer, etc. Avinash et al. (2017) have suggest the Gabor filter and watershed segmentation for the detection of lung cancer. Vijila Rani and Jawhar (2018b) suggest Wavelet Shrinkage adaptive histogram equalization (WSAHE) with anisotropic diffusion filtering. Here the contrast of the CT images capture from lung cancer patients is improved and the various types of noise present in the image are reduced. Quick detection of the lung cancers is beneficial to medical equipment manufacturing industries. Palani Kumar et al. (2014) have suggest Palm print enhancement using recursive histogram equalization Image quality assessments as well as enhancement stage were adopted on low pre-processing techniques.

Armato et al. (1999) have proposed an iterative thresholding algorithm for lung segmentation followed by opening and closing operations using morphological operations for refining the segmentation output. Duda et al. (2001) have suggested segmentation based on discriminative approach. Discriminative classification (DC) is another segmentation method similar to Probabilistic Classification except that the classification is done using generic supervised machine learning and not probabilistic distributions. Sluimer et al. (2005) have suggested the registration scheme in which segmentation of the lungs is refined based on an atlas concept. In accordance to this, apply the voxel classification technique to the border volume of the transform probabilistic atlas. Batenburg and Sijbers (2009) have uses a projection minimization approach that employs tomographic projection data in order to decide the threshold value by minimizing the separation between the forward projection of the sectioned image and the measured projection information. Probability density function (PDF) with novel optimal thresholding algorithm concept is used for lung image segmentation to represent effective lung districts from the CT scan.

Ricciardi et al. (2011) have proposed the level set active contour model the major drawback of this method is that the initial point selection is highly sensitive and failure to adapt the boundaries because of the in homogeneities in the lung region. Couprie et al. (2011) have suggested image segmentation technique based on graph. Different segmentation

algorithms such as graph cuts, random walker, and shortest path optimization are applied. Here image is viewed as a weighted graph. The common energy function using alternatives of constraint q performing as an exponent on the differentiation amid neighbouring nodes are expressed by these algorithms. The power of the edge weights is put back into working order using a fresh parameter p . This permits including the optimal spanning forest algorithm for watershed. Elizabeth et al. (2012) in their work have proposed an automated method for segmentation of lung parenchyma from chest CT. They have achieved it in two steps. First the lungs have been segmented using optimal thresholding and then the lung border has been reconstructed using the properties of the lung region, namely, convex hull and centroid. They have tested their method by using it in a CAD system for diagnosis of lung cancer and have achieved an accuracy of 97% as compared to 88.5% achieved using segmentation based on optimal thresholding.

Song and Fan (2016) have proposed an automatic lung lesion segmentation approach based on toboggan and back propagation growing algorithm. In this method does not require any training dataset and the result gives high detection sensitivity with little false positive detection. Unsupervised method few lesions give 60% accuracy. Zhang et al. (2017) have developed an advance segmentation method for image sequence of lung nodule based on super pixels and density based spatial clustering algorithm. Here precision (PRI), mean square error (MSE) and voxel index (VOI) values are analyzed which gives the results of PRI, MSE and VOI values 0.9190, 0.0736 and 2.2933 respectively. Vijila Rani and Jawhar (2020) suggest novel superpixel with nanoscale imaging and boosted deep convolutional neural network concept for lung tumor classification method Trials were established and estimated extensively on benchmark data set for segmentation.

Georgiadis et al. (2008) have introduced a novel algorithm based on emphasis of short run and non-uniformity of run length matrix. The features thus extracted shows poor classification performance and high computation complexity, while dealing with large amount of data. By combining 2D wavelets and 3D wavelets, the texture features thus obtained shows better accuracy. Kuruvilla and Gunavathi (2014) have reviewed most of the papers for classification of the nodule candidate based on feature value. Standard feature are geometric, gray level, statistical and gradient features. Vijila Rani et al. (2020) have proposed triple novelty block detection and classification approach for lung tumor analysis. The segmented nodule data was submitted only based on texture during the feature extraction process. Clinical attention was given to the essential texture in separating malignant from benign nodules at the edge of lung nodules. The extraction of features provides certain criteria from which the computer system decides.

Manian et al. (2000) in their work have proposed logical operators for texture classification. They have developed logical operators for feature extraction and extracted features using the logical operators. Feature selection is then carried out using distance and standard deviation. They have performed classification based on Bayes classifier, Euclidean, Mahalanobis distance measures, k-nearest neighbor (KNN) and leave one out classifier. They have achieved a percentage correct classification (PCC) of 78% with Bayes classifier, 96% with Euclidean distance measure, 85% with Mahalanobis distance measure, 94% with KNN and 94% with leave one out classifier. Siva Kumar and Chandrasekar (2013) have been suggested an effective lung nodule detection system for CT images. Nodule segmentation process is done through weighted fuzzy possibility based clustering approach. They demonstrated that radial bias function (RBF) kernel based SVM classifier outperformed the linear and polynomial kernel based SVM classifier. Achieved detection sensitivity and accuracy is 82.05% and 80.36%, respectively.

Choi and Cho (2014) have validated a lung nodule detection scheme for feature extraction and classification to reduce number of FPs. 3-D shape based feature vectors is extracted and the classifier was trained to classify malignant nodules and non-nodules by using supervised learning based support vector machine achieved 6.76 FP per scan and 97.5% detection sensitivity respectively. Teramoto et al. (2017) have been analysis an lung cancers in microscopic images by using a foremost deep learning technique called deep convolutional neural network (DCNN) classifier and results shows that 70% of images were classified correctly. Vijila Rani and Jawhar (2019a) have suggested lung lesion classification scheme using optimization techniques and hybrid (KNN-SVM) classifier that neural networks can greatly help in providing second opinion to healthcare professionals and greatly reduce the time of treatment for the patients.

3 Proposed method

Firstly, the proposed approach aims to strengthen the pulmonary nodule and its segmentation. The segmented nodules are ultimately graded according to two specialized techniques. The next parts provide a chronicle of the proposed method.

3.1 Block diagram

The suggested approach for pulmonary images is shown in Fig. 1. The image acquisition process is initially completed to be accessed by itself or from the Free Access database (Mathews and Jeyakumar 2020). Photographs are taken from the open-access dataset of the Lung Image Collection Consortium (LIDC). The freely open LIDC CT scan involves

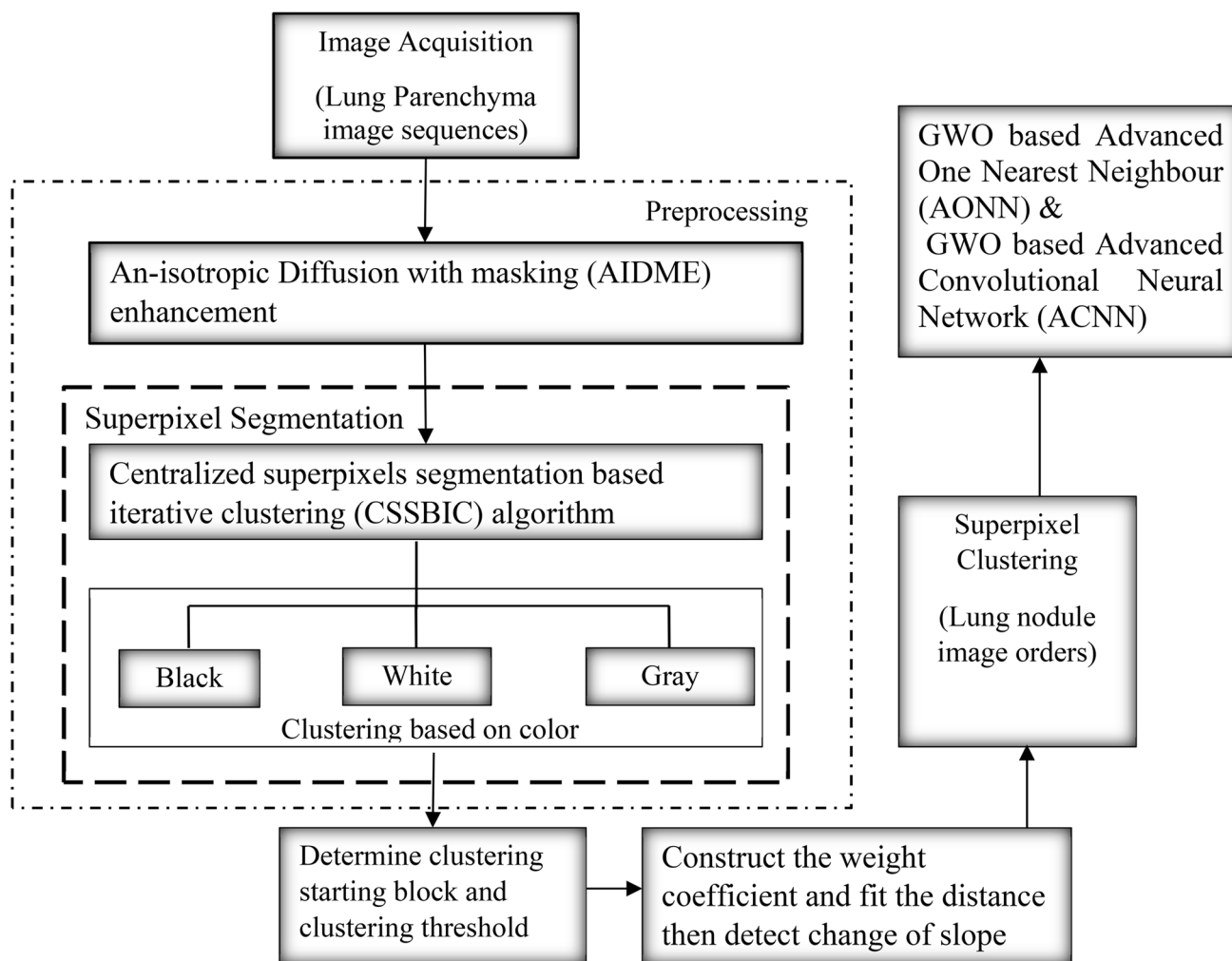


Fig. 1 Block diagram for proposed method of lung nodule image segmentation

80 patients with 505 CT study clinical findings on site. 455 patients with 710 CT nodules have been tested. 700 lung cancer images are nominated to execute the classification process. The first phase of modern innovations is the pre-processing of the lung image using the An-isotropic Diffusion with masking enhancement (AIDME) method. During preparation, a CT input image is eradicated by noise during image development, and then eliminates unwanted signals that may be used for the processing of those mistakes. The CT scan of the pulmonary image is also identified by certain criteria of the Gaussian approaches. In conjunction with the creation of contrasts, noise removal from the CT image is necessary to enhance medical image diagnosis. The An-isotropic Diffusion with masking enhancement (AIDME) favors the Gaussian image model.

The clarity of the medical images low since the images contains such deficiencies, which can impact the speckle and

visual noise. The existence of spatula noise is unpleasant as it undermines the picture quality and impacts human diagnostics and interpretation activities. Noise from speckles is a multiplying noise, so it is not as simple as the extra noise. The conventional techniques do not reduce some noises, particularly speckle noise. A filter is then applied to the goal for the elimination of speckle noise.

The an-isotropic diffusion with masking (AIDME) enhancement is proposed for the removal of annoying noise in the input CT pulmonary images; image improvement technology is used to incorporate an-isotropic diffusing with masking enhancement technique. In this case, the picture is split into three sub-pieces: black, gray, and white. To increase the accuracy of the proposed method, the tumor area detected from the image should be perfectly observed. Finally, the proposed AIDME preprocessed techniques are accomplished.

3.2 Centralized superpixels segmentation based iterative clustering (cssbic) algorithm

The goal of the algorithm is to classify pixels into a small, uniform picture-like area. The segmentation of the superpixels would create an issue with the clusters as each superpixel has its color and shape. We are implementing a new algorithm called Centralized superpixels segmentation based iterative clustering (CSSBIC) algorithm, which blends three-dimensional color pixels in a hybrid form of black and white and green. The suggested method consists of two stages: clustering and fusion. In the initial step, the pixels are fused to produce the new superpixels. Then the initial superpixels are reinforced, along with the iterative clustering phase the final superpixels fuse with a very small superpixel. The flexibility of the chosen method gives a very simple means of using it alone in which the parameter shows the number of superpixels and the algorithm performance is relatively high. The findings of the experiments reveal that the alternative

solution provides greater segmentation than the new superpixel approach. Algorithm 1, describes about the proposed Centralized superpixels segmentation based iterative clustering (CSSBIC) method.

$$dl_{ab} = \sqrt{((l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2)} \quad (1)$$

$$d_{xy} = \sqrt{((x_k - x_i)^2 + (y_k - y_i)^2)} \quad (2)$$

where dl_{ab} is the lab distance, d_{xy} is the plane distance, (x, y) is the pixel at position.

Then the Distance measure D_s is shown in the Eq. (3)

$$D_s = dl_{ab} + (m/S) * d_{xy} \quad (3)$$

where l, a, b values of the CIELAB color space, D_s is the sum of the lab distance and the xy plane distance normalized by the grid interval S . A variable m is introduced in D_s which can control the compactness of super pixel with x, y pixel coordinates.

Algorithm 1: Centralized superpixels segmentation based iterative clustering (CSSBIC) algorithm

Step 1: Input Image Initialization Process

```

//***** Initialize cluster center  $C_k = [L_k, a_k, b_k, x_k, y_k]^T$  in the hexagonal grid, pixels label matrix I,
distance matrix D from pixels to cluster centers, hexagonal grid spacing S and radius r of circular structure element
*****//

```

Step 2: Move Cluster center

```

//*****initial cluster center to the smallest gradient position in the hexagonal region.*****//

```

Step 3: Repeat Process

```

//*****Repeat the same condition*****//

```

While residual error \geq threshold E do

```

//***** If the residual error greater than or equal to threshold Do the for condition*****//

```

For each cluster center C_k do

Obtain sub image cluster center & Distance calculation process

```

//*****Obtain sub images (use  $[x_k, y_k]$  as the center,  $2*S$  as the side) containing cluster centers. Calculate
the distance between each pixel in the sub image and cluster center. If the distance is less than the previous value,
then update its I and D*****//

```

End for // *****End the for condition*****//

Update the cluster center

```

//***** calculate the mean of [equation 1] L, a, b, x and y of each superpixel to update the cluster center.
Recalculate the residual error, go to the third step, and continue the execution*****//

```

End while // *****End the While condition*****//

Step 4: Mask Prediction Process

```

//*****Obtain all non-connected regions of the CT image and then perform the open operation of the
circular structure element radius r. The final result is subtracted from the original CT image denoted as mask
*****//

```

Step 5: Perform distance transform

```

//*****Perform distance transform on mask, and assign each small region to the nearest
superpixel *****//

```

Step 6: Computation Process

```

//*****Compute the superpixel adjacency matrix for subsequent operation. Until all CT image
sequences are segmented*****//

```

Step 7: Obtained Output.

```

//*****Execute connection operations and sequential output*****//

```

3.3 Grey wolf optimization based classification algorithm

In general, image optimization involves in reducing the unwanted features of the image and converting it into an optimized form so that it can undergo efficient processing.

Here the GWO algorithm makes use of Meta heuristic optimization step to optimize the continuous function, the number of grey wolves is represented as N, and the solution search space is represented as dimension d.

In spite of having a good convergence rate, GWO is incapable to perform the detection, global optima with the same convergence rate of the algorithm. Thus, to reduce this effect

and improve its efficiency, GWO with classification algorithm is established. The main purpose of using GWO with classification algorithm is to identify the best wolf values. For the classification purpose two GWO based classifier are proposed. They are,

- GWO based advanced one nearest neighbour (AONN) classifier.
- GWO based advanced convolutional neural network (ACNN) classifier.

Algorithm 2: GWO based Advanced One Nearest Neighbour (AONN) classifier

Begin

Initialize the advance grey wolf population

Initialize Test Data as Number of grey wolves (NumGW)

For each NumGW, calculate Accuracy and grey wolf position

Set condition of i=1 //*****Iteration 1 and then generate $R \leftarrow$ random number (between 1 to NumGW). Calculate the first three best fit values*****//

Check condition R //*****Check condition R from a database (both public and in-house clinical) of numbers so that R should not repeat. Update the value. One by one data points are taken from TestData .Set as an position*****//

Position \leftarrow **Test Data(R)**

//*****Pass Position to ONN as input, assigns a point x to the class of its closest neighbour in the feature space and get ONN Output. (Fitness evaluation)*****//

Calculate Parameters

//*****such as, Accuracy \leftarrow Target Value – ONN Output

Best_Value \leftarrow Accuracy

Best_Position \leftarrow Position*****//

Then do, i=2 Iteration 2 //*****for each grey wolf NumGW.

Replace worst fit gray wolf value with best fit*****//

Check condition if (Best_Value > Accuracy)

Then Calculate //*****alternatively, Best_Value=Accuracy

Best_Position=Position*****//

Get the best solution with best fitness gray wolf value.

End for

End

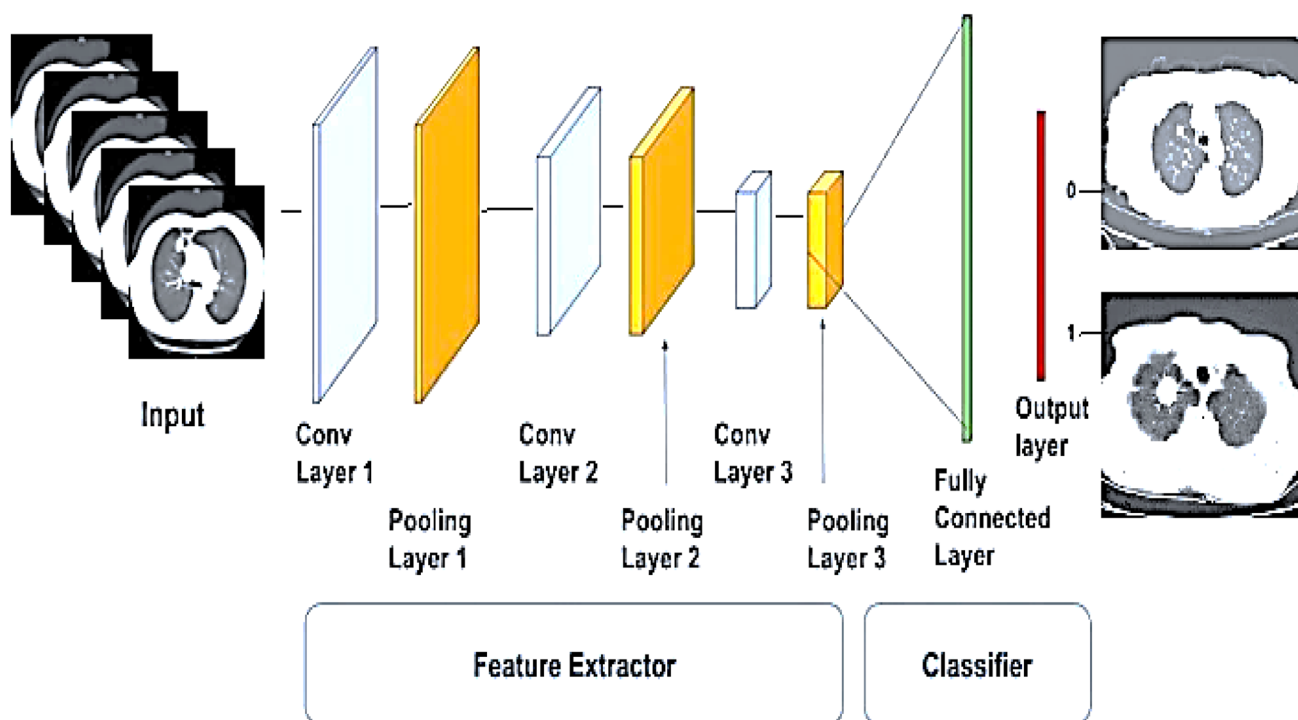


Fig. 2 Steps in deep learning based ACNN algorithm

In the GWO algorithm, it is necessary to determine the distance between the individual and the prey as shown in the Eq. (4),

$$D = |C \cdot X_p(t) - X(t)|. \quad (4)$$

Algorithm 2 explains the details about the GWO based advanced one nearest neighbour (AONN) classifier. Initially the distances between data points are calculated where one neighbor is considered. The classification is accomplished using AONN and deep learning based GWO based advanced convolutional neural network (ACNN) classifier. This research involves an artificial classification scheme for CT image lung cancer, using a GWO based advanced convolutional neural network (ACNN) classifier, a major profound learning technique. The key objective of the study is to simplify the classification of the ACNN.

The ResNet 50 alternative is to pre-train the ACNN and can extract functionality from the network fed input picture. The ResNet 50 is primarily used to train more than one million pictures. ResNet 50 has the advantage of providing

remaining contacts, which prevents the lack of information after the exercise. Thus, ResNet can be used for ACNN training to increase processing speed. Figure 2 contains three convolutional layers, three pooling layers, and two fully connected layers of ACNN for grouping. The (Advance Convolutional Neural Network) ACNN was trained with our initial database as tests were performed. In the first convolution layer, the picture still joins. Start the input matrix at the top left of the image.

After this, the computer selects a smaller filter matrix. For convolution, the filter is used, i.e. it goes through the input image. The filter uses its values to multiply the original pixel values. The bought items shall be summarized to the final amount. Since the filter reads only the top left corner of the image, one-piece pushes to the right and a similar operation is done. But as the filter is transferred throughout, the obtained matrix forms a matrix compared to the input matrix. The development of classifications may be benign or malignant. Algorithm 3 explains the details of the proposed GWO based advanced convolutional neural network (ACNN) classifier.

Algorithm 3: GWO based Advanced Convolutional Neural Network (ACNN)

classifier

Begin**Initialize** the advance grey wolf population**Initialize** Test Data as Number of grey wolves (NumGW)**For** each NumGW, calculate Accuracy and grey wolf position

Set condition of i=1 //*****Iteration 1 and then generate $R \leftarrow$ random number (between 1 to NumGW). Calculate the first three best fit values*****//

Check condition R //*****Check condition R from a database (both public and in-house clinical) of numbers so that R should not repeat. Update the value. One by one data points are taken from TestData .Set as an position*****//

Position \leftarrow Test Data(R)

//*****Pass Position to CNN as input, assigns a point x to the class of its closest neighbour in the feature space and get CNN Output. (Fitness evaluation)*****//

Calculate Parameters//*****such as, Accuracy \leftarrow Target Value – CNN OutputBest_Value \leftarrow AccuracyBest_Position \leftarrow Position*****//

Then do, i=2 Iteration 2 //*****for each grey wolf NumGW. Replace worst fit gray wolf value with best fit*****//

Check condition if (Best_Value > Accuracy)**Then Calculate** //*****alternatively, Best_Value=Accuracy

Best_Position=Position*****//

Get the best solution with best fitness gray wolf value.**End for****End**

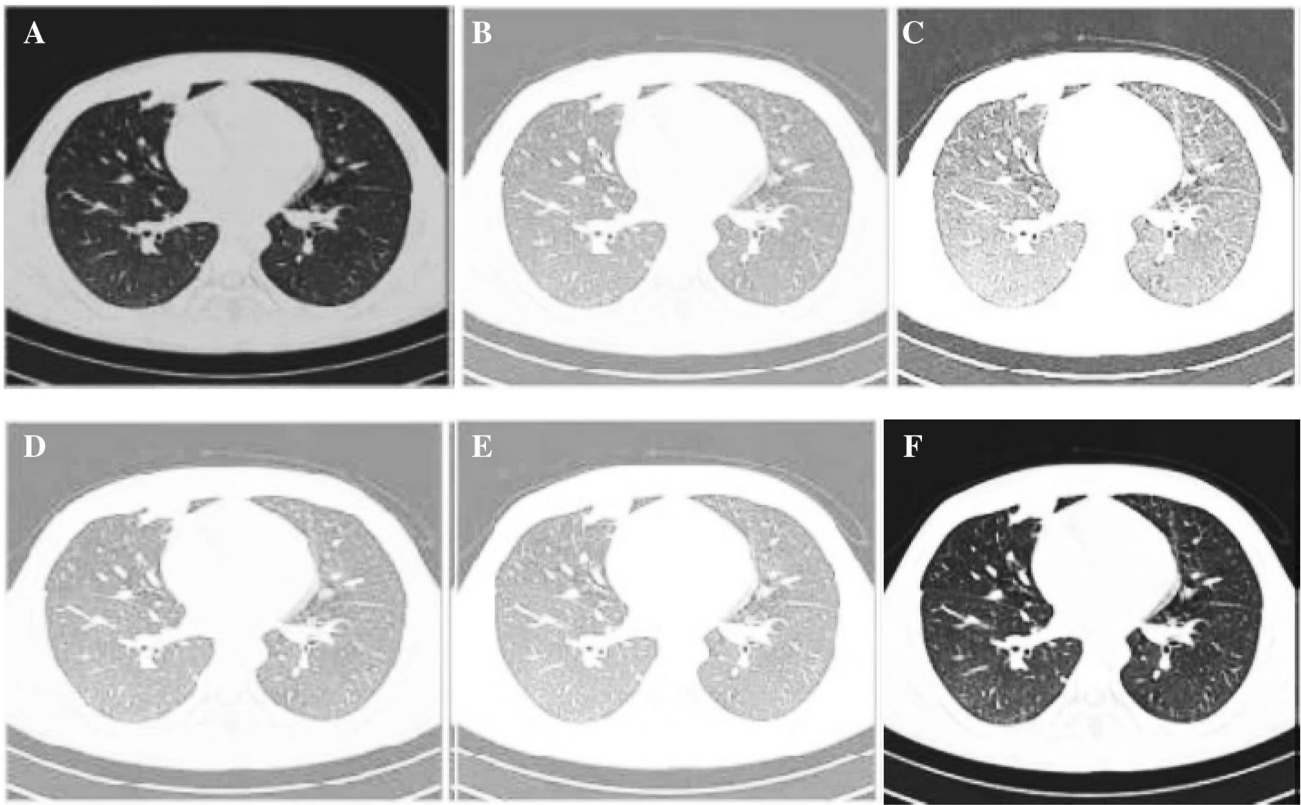


Fig. 3 SAMPLE TEST IMAGE 1 image enhancement results

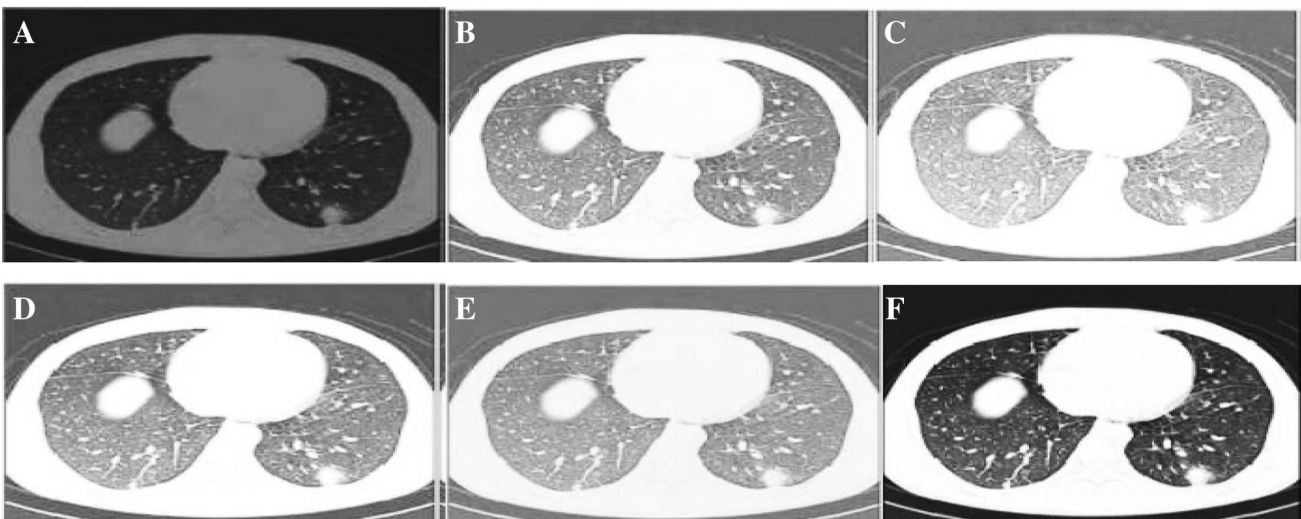


Fig. 4 SAMPLE TEST IMAGE 2 Image enhancement results

4 Performance measures

The enactment of the proposed algorithm has been estimated in terms of peak signal-to-noise ratio, mean square error and

structural similarity index methods (Mathews and Jeyakumar 2020).

Equation (5) estimates the mean-squared error to compute PSNR

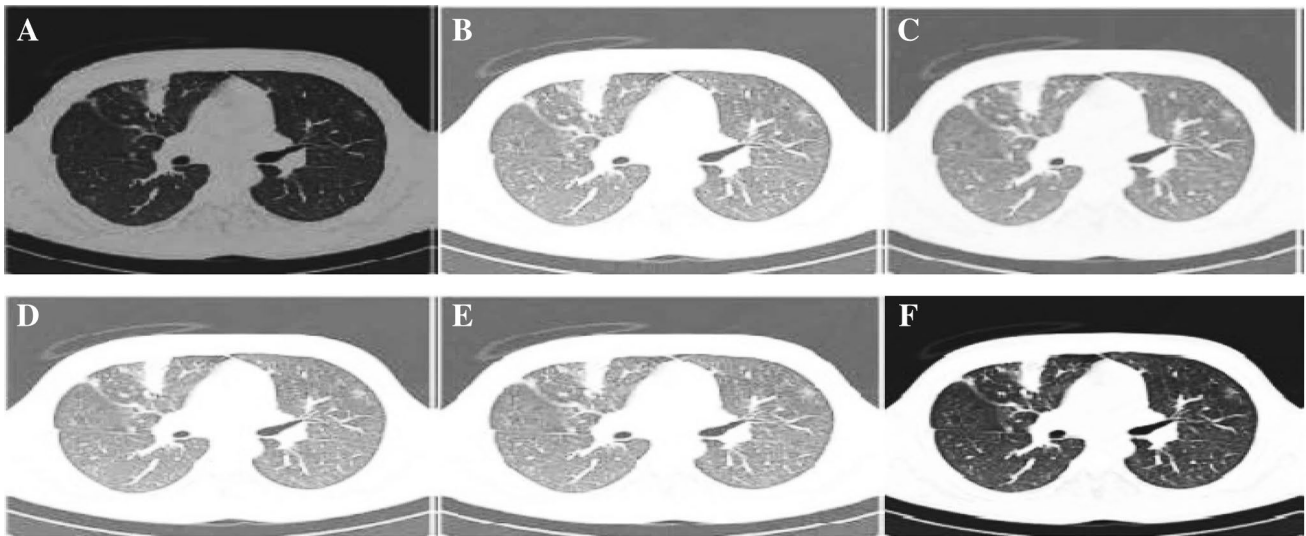


Fig. 5 SAMPLE TEST IMAGE 3 Image enhancement results

Table 1 Quality parameters performance measure of SAMPLE TEST IMAGE 1

	PSNR (dB)	Entropy value	MSE
Sample test input image 1 (ST1)	30.18	9.85	5.50E+3
Contrast limiting adaptive histogram equalization of ST1	25.92	7.68	4.62E+3
Modified adaptive histogram equalization of ST1	31.17	10.65	5.91E+3
Trilateral filter of ST1	24.95	8.65	6.16E+3
Histogram equalization of ST1	25.21	9.22	6.22E+3
An-isotropic diffusion with masking (AIDME) enhancement of ST1(proposed)	69.12	10.36	4.12E+3

Table 2 Quality parameters performance measure of SAMPLE TEST IMAGE 2

	PSNR (dB)	Entropy value	MSE
Sample test input image 2 (ST2)	29.07	9.25	5.62E+3
Contrast limiting adaptive histogram equalization of ST2	28.14	7.14	6.62E+3
Modified adaptive histogram equalization of ST2	30.47	10.32	5.18E+3
Trilateral filter of ST2	25.47	8.24	6.69E+3
Histogram equalization of ST2	25.21	9.22	6.22E+3
An-isotropic diffusion with masking (AIDME) enhancement of ST2(proposed)	68.42	10.07	4.62E+3

Table 3 Quality parameters performance measure of SAMPLE TEST IMAGE 3

	PSNR (dB)	Entropy value	MSE
Sample test input image 3(T3)	31.49	9.92	5.18E+3
Contrast limiting adaptive histogram equalization of ST3	21.32	7.48	6.09E+3
Modified adaptive histogram equalization of ST3	31.68	10.47	6.72E+3
Trilateral filter of ST3	28.35	8.45	5.62E+3
Histogram equalization of ST3	22.52	9.20	4.70E+3
An-isotropic diffusion with masking (AIDME) enhancement of ST3 (proposed)	69.47	10.24	3.69E+3

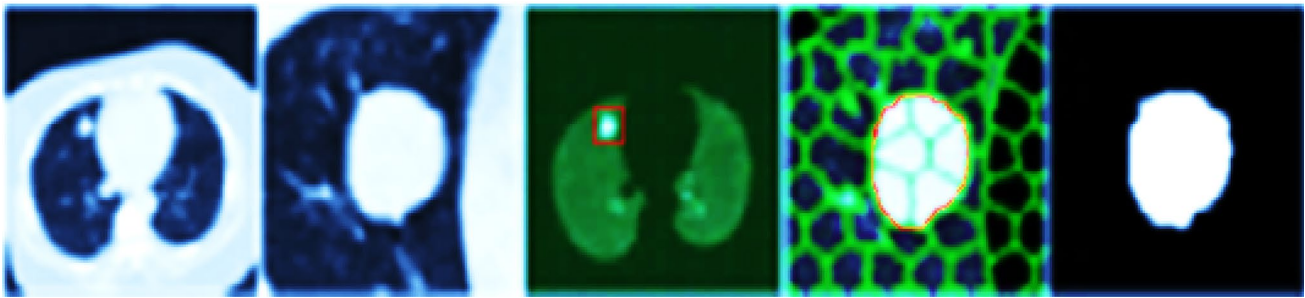


Fig. 6 Sample 1 segmentation result of our suggested CSSBIC method for solitary pulmonary nodule detection

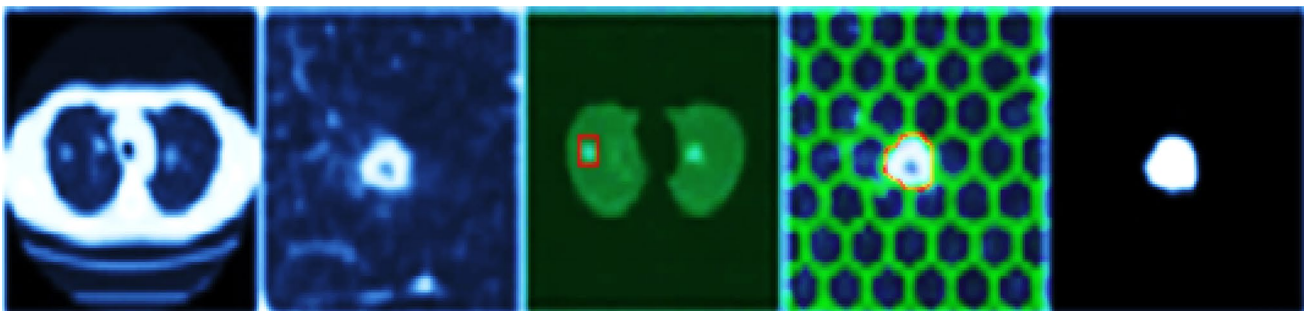


Fig. 7 Sample 2 segmentation result of our suggested CSSBIC method for solitary pulmonary nodule detection

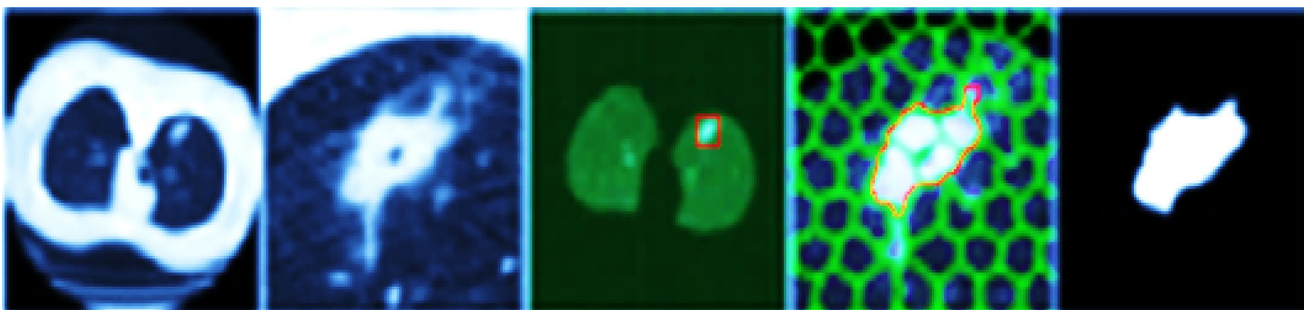


Fig. 8 Sample 3 segmentation result of our suggested CSSBIC method for solitary pulmonary nodule detection

Table 4 Summarizes the effect of varying threshold and ‘i’ values on the accuracy of solitary nodules detected.

Dataset	Threshold	i = 10		i = 10	
		AGWO-ONN		AGWO-CNN	
		Accuracy	N	Accuracy	N
Publically available LIDC dataset	0.65	97.0	44	97.6	44
	0.75	85	38	85	38
	0.85	66.67	20	66.7	20
In-house clinical dataset	0.65	90.7	34	92.5	34
	0.75	82	27	85.33	30
	0.85	60.76	21	62.15	24

‘N’ is the number of solitary nodules detected

$$MSE = \sum_{M_1, N_1} \frac{[I_1(m_1, n_1) - I_1(m_2, n_2)]^2}{M_1 \times N_1} \tag{5}$$

M1 is the number of rows in the input images N1 number of columns in the input images, respectively. The Eq. (6) computes PSNR,

$$PSNR = 10 \log_{10} \left(\frac{R1^2}{MSE} \right) \tag{6}$$

The SSIM index is considered on various windows of a lung image. The measure between two window x and y common size N × N. Equation (7) shows the computation of the SSIM index.

Table 5 Dice parameter—segmentation performance measures of proposed CSSBIC method

Images	Dice			
	Otsu segmentation	Fuzzy C means clustering	Watershed segmentation	Proposed CSSBIC method
Sample Test Image1	0.0036	0.1472	0.0045	0.1520
Sample Test Image2	0.0052	0.2295	0.0030	0.2459
Sample Test Image3	0.0026	0.2086	0.0052	0.1965
Sample Test Image4	0.0036	0.2426	0.0123	0.2993
Sample Test Image5	0.0014	0.1611	0.0053	0.1526
Sample Test Image6	0.0060	0.1639	0.0040	0.2337
Average	0.0037	0.1922	0.0057	0.2133

Table 6 SSIM parameter—segmentation performance measures of proposed CSSBIC method

Images	SSIM			
	Otsu segmentation	Fuzzy C means clustering	Watershed segmentation	Proposed CSSBIC method
Sample Test Image1	0.0001	0.9653	0.9087	0.9681
Sample Test Image2	0.0001	0.9842	0.9154	0.9853
Sample Test Image3	0.0001	0.9730	0.9405	0.9767
Sample Test Image4	0.0001	0.9641	0.9353	0.9774
Sample Test Image5	0.0001	0.9422	0.9083	0.9423
Sample Test Image6	0.0001	0.9391	0.9448	0.9726
Average	0.0001	0.9613	0.9255	0.9704

Table 7 Demonstrate an Evaluation with other publications with respect to lung tumor detection

Study	Method	Dataset	Performance
Hua et al. (2011)	Graph-search	15 cases	Sensitivity = 98.6% Specificity = 99.5%
Cascio et al. (2012)	Double threshold cut and neural network	LIDC 84 scans and 148 nodules	88% sensitivity with 2.5 FPs/scan
Alakwaa et al. (2017)	3D convolutional neural networks (CNN)	Kaggle Data Science Bowl, 2017	Accuracy 86.6%
Zhang et al. (2017)	Superpixels and density-based spatial clustering	1458 CT sequence images	Accuracy = 91.90%
Vijila Rani and Jawhar (2019b)	Nanoscale imaging technique based lung cancer detection	LIDC CT scan of 455 patients and in-house clinical dataset ICCN of 80 patients CT exams	Accuracy = 97 (LIDC Dataset)
Vijila Rani and Jawhar (2020)	Superpixel with nanoscale imaging and boosted deep convolutional neural network concept	450 LIDC database and in-house clinical data set from International Cancer Center Neyyoor (ICCN, Tamil Nadu)	Accuracy = 97.3%
Proposed method 1	Centralized superpixels segmentation based iterative clustering (CSSBIC) algorithm and GWO based advanced one nearest neighbour (AONN)	LIDC CT scan of 455 patients and in-house clinical dataset of 80 patients CT exams	Accuracy = 97 (LIDC Dataset) Accuracy = 90.7 (In-house Clinical Dataset)
Proposed method 2	Centralized superpixels segmentation based iterative clustering (CSSBIC) algorithm and GWO based advanced convolutional neural network (ACNN)	LIDC CT scan of 455 patients and in-house clinical dataset of 80 patients CT exams	Accuracy = 97.6 (LIDC Dataset) Accuracy = 92.5 (In-house Clinical Dataset)

All Confusion Matrix

Output Class	1	444 63.5%	7 1.0%	98.4% 1.6%
	2	14 2.0%	234 33.5%	94.4% 5.6%
		96.9% 3.1%	97.1% 2.9%	97.0% 3.0%
		1	2	3
		Target Class		

Fig. 9 Show that the confusion matrix output from the GWO based advanced one nearest neighbour (AONN) classifiers

Confusion Matrix

Output Class	1	446 63.8%	5 0.7%	98.9% 1.1%
	2	12 1.7%	236 33.8%	95.2% 4.8%
		97.4% 2.6%	97.9% 2.1%	97.6% 2.4%
		1	2	3
		Target Class		

Fig. 10 Shows that the confusion matrix output from the GWO based advanced convolutional neural network (ACNN) classifiers

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_2)(2\sigma_{xy} + c_1)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (7)$$

where, μ_x the average of x_i
 μ_y the average of y_i
 σ_x^2 the average of x_i^2
 σ_y^2 the average of y_i^2
 σ_{xy} the covariance of x and y .

$$C_1 = (K_1L)^2, C_2 = (K_2L)^2$$

Dice: The relationship involving ground truth ‘G’ and segmented image ‘F’ such that, it computes the ratio of the common to the sum of number of elements between them.

$$Dice = \frac{2|G \cap F|}{|G| + |F|} \quad (8)$$

5 Results and discussion

This section estimates the results obtained by applying the centralized superpixels segmentation-based iterative clustering (CSSBIC) algorithm. The photos are collected from LIDC or the clinical archive in-house. Initially, the image processing is calculated on the image using the simulation results performed through MATLAB 2018a. The image obtained must be pre-processed in order to heat up the image for additional steps. The initial aim of the unused noise in the lung images of the input CT is to eliminate anisotropic diffusion filtering. And sharp masking is integrated in order to sharpen the image and increase the segmentation quality of the image.

In comparison with the proposed method and Mathews and Jeyakumar (2020), four prevailing approaches were compared: (i) trilateral filter, (ii) histogram equalization (HE) (Couprie et al. 2011), (iii) contrast limiting adaptive histogram equalization (CLAHE) (Zuiderveld 1994; Pizer et al. 1987) and (iv) modified adaptive histogram equalization (MAHE). The output images of the suggested and prevalent approaches display Figs. 3, 4 and 5. In Fig. 3 shows row 1: sample test input image 1 (ST1), trilateral filtered image of ST1, histogram equalized image of ST1. Row 2: contrast limiting adaptive histogram equalized image of ST1, modified adaptive histogram equalized image of ST1, an-isotropic diffusion with masking (AIDME) enhancement image of ST1 (proposed). In Fig. 4 shows the row 1: sample test input image 2(ST2), trilateral filtered Image of ST2, Histogram equalized image of ST2. Row 2: contrast limiting adaptive histogram equalized image of ST2, Modified adaptive histogram equalized image of ST2, an-isotropic diffusion with masking (AIDME) enhancement image of ST2 (proposed). Finally in Fig. 5 shows, the row 1: sample test input image 3 (ST3), trilateral filtered image of ST3, histogram equalized image of ST3. Row 2: contrast limiting adaptive histogram equalized image of ST3, modified adaptive histogram equalized image of ST3, An-isotropic Diffusion with masking (AIDME) enhancement image of ST3 (proposed).

Tables 1, 2 and 3 can be seen that the suggested algorithm in terms of the MSE, PSNR and ENTROPY estimation indicators enriched than other algorithms, and has an improved enhancement effect.

In Figs. 6, 7 and 8 Column (i) shows the original lung CT images, Column (ii) shows the results of the local enlargement of (i), Column (iii) shows the results of lung nodule image masks Column (iv) shows the result of super pixel

identification in the lung nodule (vii) present the final segmentation results of solitary pulmonary nodule using our proposed CSSBIC method.

As shown in Table 4, the number of solitary nodules detected in the widely available LIDC data set and clinical data set is affected by the variation of the threshold value. The threshold is commensurate with the exactness (Tables 5, 6). Table 7 shows a summary of the current study publication subject to successful assessment criteria with other lung tumor identification publications.

Figure 9 show that the confusion matrix output from the GWO based advanced one nearest neighbour (AONN) classifiers. In this figure, the first two diagonal cells illustrate the number and percentage of correct classifications by the trained network. For example, 444 lung cancer images are unerringly classified as benign. This corresponds to 63.5% of all 699 lung cancer images. Similarly, 234 cases are unerringly classified as malignant. Overall, 97.0% of the predictions are correct and 3.0% are erroneous.

700 lung cancer images are nominated to execute classification process. A confusion matrix of 3×3 presents the output of GWO based advanced convolutional neural network (ACNN) classifier which estimates the accuracy of the image. Figure 10 show that the confusion matrix output from the GWO based advanced convolutional neural network (ACNN) classifiers. Benign category includes 444 lung cancer images which are being are classified precisely. This corresponds to 63.5% of all 699 lung cancer images. On the contrary, malignant category includes 238 cases that are classified accurately. Thus 34.0% of all lung cancer images have been categorized efficaciously. Overall, 97.6% of the predictions are precise and 2.4% are erroneous. The average time consumption for each CT slice order with the proposed methodology is 12.32. Then the average segmentation time of a CT image is 1.06 s.

6 Conclusion

This research work is based on an iterative clustering algorithm that identifies lung nodules, which in turn improves image consistency. Iterative clustering is used for noise reduction scanning images and various validation tests are correlated with the chosen noise removal approaches. Three separate output measurements (including PSNR, MSE, and entropy) are measured for the pre-processed lung CT images. GWO based advanced one nearest neighbour (AONN) GWO based advanced convolutional neural network (ACNN) is used to procure the solitary nodes.

Classification accuracy assesses the efficiency of the data collection. Confusion matrices show the best classification precision. 97.6% of the accuracy of the proposed GWO based advanced convolutional neural network (ACNN) has

been auspiciously accomplished. The technique chosen indicates increased sensitivity and shortened estimation times. Early diagnosis will greatly increase patients' chances of survival. The suggested technique, therefore, shows the excellent capacity for automated detection of the lung tumor in the advanced CAD scheme.

7 Future work

The analysis is not limited to this work or to the findings. Because of the difference in the form and scale of this nut, the key restriction of the proposed lung lesion sectioning system does not specifically segment cavitory and Juxta vascular Nodules. The potential future scope of the work highlights the revolutionary approach needed to identify the precise position of the lung nodules.

Acknowledgements First the authors thank the National Cancer Institute and then acknowledge for free public available online LIDC-IDRI database & in-house clinical ICCN database used in this study. Finally, we wish to thank the anonymous reviewers for helping to strengthen this paper.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alakwaa W, Nassef M, Badr A (2017) Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). *Int J Adv Comput Sci Appl* 8(8):409–417
- Armato SG, Giger CJ, Moran JT, Blackburn K, MacMahon H (1999) Computerized detection of pulmonary nodules on CT scans. *Radiographics* 19(5):1303–1311
- Avinash S, Manjunath K, Senthil Kumar S (2017) An improved image processing analysis for the detection of lung cancer using Gabor filters and watershed segmentation technique. *IEEE Trans Image Process* 74:178–193
- Batenburg KJ, Sijbers J (2009) Optimal threshold selection for tomogram segmentation by projection distance minimization. *IEEE Trans Med Imaging* 28(5):676–686
- Cascio D, Magro R, Fauci F, Lacomini M, Raso G (2012) Automatic detection of lung nodules in CT datasets based on stable 3D mass spring models. *Comput Biol Med* 42(11):1098–1109
- Choi WJ, Cho TS (2014) Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Comput Methods Progr Biomed* 113:37–54
- Coupric C, Grady L, Najman L, Talbot H (2011) Power watershed: a unifying graph-based optimization framework. *IEEE Trans Pattern Anal Mach Intell* 33(7):1384–1399
- Duda RO, Hart PE, Stork DG (2001) *Pattern classification*, 2nd edn. Wiley Inter Science, New Jersey
- Elizabeth DS, Nehemiah HK, Raj CSR, Kannan A (2012) A novel segmentation approach for improving diagnostic accuracy of CAD systems for detecting lung cancer from chest computed tomography images. *J Data Inf Qual* 3(2):1–14

- Georgiadis P, Cavouras D, Kalatzis I, Daskalakis A, Kagadis GC, Sifaki K, Malamas M, Nikiforidis G, Solomou E (2008) Improving brain tumor characterization on MRI by probabilistic neural networks and non-linear transformation of textural features. *Comput Methods Progr Biomed* 89(1):24–32
- Haney SM, Thompson PM, Cloughesy TF, Alger JR, Toga AW (2001) Tracking tumor growth rates in patients with malignant gliomas: A test of two algorithms. *Am J Neuroradiol* 22(1):73–82
- Hua P, Song Q, Sonka M, Hoffman EA, Reinhardt JM JM (2011) Segmentation of pathological and diseased lung tissue in CT images using a graph-search algorithm. *Biomed Imaging*, pp 2072–2075
- Jiang C, Zhang X, Huang W, Meinel C (2004) Segmentation and quantification of brain tumor. In: *Virtual environments, human-computer interfaces and measurement systems (VECIMS)*, pp 61–66
- Kuruville J, Gunavathi K (2014) Content based image retrieval for CT images of lungs. *Int J Soft Comput* 9(6):386–390
- Manian V, Vasquez R, Katiyar P (2000) Texture classification using logical operators. *IEEE Trans Image Process* 9(10):1693–1703
- Mathews AB, Jeyakumar MK (2020) Automatic detection of segmentation and advanced classification algorithm. In: *2020 fourth international conference on computing methodologies and communication (ICCMC)*, Erode, India, 2020, pp 358–362. <https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00067>
- Palani Kumar S, Kumar MS, Rajeesh J (2014) Palmprint enhancement using recursive histogram equalization. *Imaging Sci J* 61:447–457
- Pizer SM, Amburn EP, Austin JD (1987) Adaptive histogram equalization and its variations. *Comput Vis Graph Image Process* 39:355–368
- Rahman MM, Bhattacharya P, Desai BC (2007) A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *IEEE Trans Inf Technol Biomed* 11(1):58–69
- Ricciardi S, Tomao S, de Marinis F (2011) Efficacy and safety of erlotinib in the treatment of metastatic non-small-cell lung cancer. *Lung Cancer Target Ther* 2:1–9
- Sankar K, Prabhakaran M (2013) An improved architecture for lung cancer cell identification using Gabor filter and intelligence system. *Int J Eng Sci* 2:38–43
- Siva Kumar S, Chandrasekar C (2013) Lung nodule detection using fuzzy clustering and support vector machines. *Int J Eng Res Technol* 5(1):179–185
- Sluimer I, Prokop M, Ginneken BV (2005) Toward automated segmentation of the pathological lung in CT. *IEEE Trans Med Imaging* 24(8):1025–1038
- Song J, Fan L (2016) Lung lesion extraction using a toboggan based growing automatic segmentation approach. *IEEE Trans Med Imaging* 35(1):337–353
- Teramoto A, Tsukamoto T, Kiriya Y, Fujita H (2017) Automated classification of lung cancer types from cytological images using deep convolutional neural networks. *Bio Med Res Int* 1–6
- Vijila Rani K, Jawhar SJ (2018a) Emerging trends in lung cancer detection scheme—a review. *Int J Res Analyt Rev* 5(3):530–542
- Vijila Rani K, Jawhar SJ (2018b) Novel method for lung tumour detection using wavelet shrinkage-based double classifier analysis. *IETE J Res*. <https://doi.org/10.1080/03772063.2018.1557086>
- Vijila Rani K, Jawhar J (2019a) Lung lesion classification scheme using optimization techniques and hybrid (KNN-SVM) classifier. *IETE J Res*. <https://doi.org/10.1080/03772063.2019.1654935>
- Vijila Rani K, Jawhar SJ (2019b) Novel Technology for lung tumor detection using nanoimage. *IETE J Res*. <https://doi.org/10.1080/03772063.2019.1565955>
- Vijila Rani K, Jawhar SJ (2020) Superpixel with nanoscale imaging and boosted deep convolutional neural network concept for lung tumor classification. *Int J Imaging Syst Technol*. <https://doi.org/10.1002/ima.22422>
- Vijila Rani K, Thinkal Dayana C, Sujatha Therese P, Eugene Prince M (2020) Triple novelty block detection and classification approach for lung tumor analysis. *Int J Imaging Syst Technol*. <https://doi.org/10.1002/ima.22509>
- Zhang W, Zhang X, Zhao J, Qiang Y, Tian Q, Tang X (2017) A segmentation method for lung nodule image sequences based on superpixels and density-based spatial clustering of applications with noise. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0184290>
- Zuiderveld K (1994) Contrast limited adaptive histogram equalization. In: *Graphic gems IV*. Academic Press Professional, San Diego, pp 474–485