Lung Tumor Staging and Classification with Machine Learning and Deep Learning Models

Jagadeesh K

jagadeeshjpsmk@gmail.com

SNS College of Engineering

Rajendran A

Karpagam College of Engineering

Research Article

Keywords: Lung Tumor, Volumetric Analysis, FCM Clustering, Segmentation, Neural Networks, Classification, Tumor Stage

Posted Date: March 8th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3940572/v1

Additional Declarations: No competing interests reported.

Lung Tumor Staging and Classification with Machine Learning and Deep Learning Models

Jagadeesh. K[1,*] and Rajendran. A[2]

[1]Assistant Professor, Department of Electronics and Communication Engineering, SNS College of Engineering, Coimbatore, 641107, India

[2]Professor, Department of Electronics and Communication Engineering, Karpagam College of Engineering, Coimbatore, 641032, India

[*]Corresponding Author: Jagadeesh.K. Email: jagadeeshjpsmk@gmail.com

Abstract:

In today's worldwide health scenario, Lung Cancer has the highest rates of mortality and morbidity. The accurate and clinical staging of lung cancer category can effectively reduce the death rate, since the treatment phase requires the specific stage of cancer. However, the staging of lung cancer still remains challenging, requires more efforts. The Computed Tomography images (CT) images are utilized for the Computer Aided Diagnosis based cancer diagnosis. With that note, this paper develops a Volumetric Analysis for Lung Tumor Staging and Classification (VA-LTSC), in which the stages are classified based on Tumor Nodule Metastasis (TNM) with Machine Learning and Deep Learning Models Moreover, the proposed model comprises different phases**.** The results are measured using inputs from LIDC-IDRI and LUNA 16, based on classification accuracy, model effectiveness and time complexities and in all, the proposed model outperforms the existing results.

Keywords: Lung Tumor, Volumetric Analysis, FCM Clustering, Segmentation, Neural Networks, Classification, Tumor Stage.

1. Introduction:

In present scenario of human lives, cancer is a remarkable health problem worldwide with higher rates of mortality. Among all cancer types, Lung Cancer is the most deadly in both genders. Lung Cancer can also be stated as Carcinoma, which is the formation of cancerous nodules, with uncontrolled tissue growth. The rate of lung cancer survivors is very minimal approximately 12% with the life span of 5 to 6 years. However, the major problem is the late diagnosis of cancer may cause the treatment process very complicated and mainly minimizes the rate of survival. Hence, the earlier detection of lung cancer can effectively increase the survival rate of patients by providing treatments on time, limits the mortality rate. Based on the tumor cell features, the lung cancer can be classified into two categories, as, small cell lung cancer and non-small cell lung cancer. Among two, the most common cancer type is the non-small cell lung cancer and the rare one is the small cell lung cancer. Additionally, lung tumor staging can be processed with the growth rate of cancer cells and tumor size. The general operations involved in process of lung tumor diagnosis are portrayed in Figure 1.

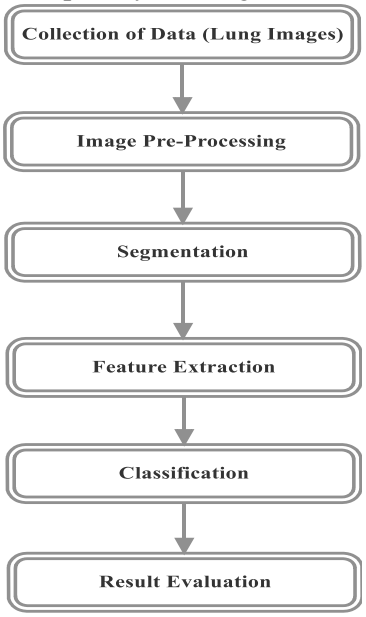


Figure 1: General Framework of Lung Cancer Diagnosis

In clinical practice, the brutality of lung cancer can be described as different levels based on the Tumor, Node and Metastasis (TNM) model [1]. In that Tumor (T) measures the features of the primary tumor like location, size and nodule presence. The Tumor (T) based categorization can be effectively from the input CT scan images. Another descriptor (N) considers the regional lymph nodes, in which the CT cannot be effectively used for the detection [2], [3], [22], [23]. And, 'M' evaluates the growth rate of tumor to the extra-thoracic organ, through the brain, liver, bones and so on [4], [24], [25]. Moreover, staging can be accurately done with some advanced and additional computations. Based on

anatomic features, the T-Stages can be further categorized, which is presented in Table 1, as in 4 stages.

i. Stage 1: the cancer cells are impounded to the lung area
ii. Stage 2 and Stage 3: Cancer growth is within the chest area
iii. Stage 4: Spreading of cancer to other body parts

Table 1: T-based Lung Cancer Staging

Cancer Stage	Tumor Sizes (TS)	Invasion
T_1 T_{1a} T_{1b}	$TS \leq 3cm$ $TS \leq 2cm$ $TS > 2cm \parallel TS \leq 3cm$	Area is covered by visceral pleura attacked in the lobar bronchus
T_2 T_{2a} T_{2b}	$TS > 3cm \parallel TS \leq 5cm$ $TS > 3cm \parallel TS \leq 4cm$ $TS > 4cm \parallel TS \leq 5cm$	Attacks the main bronchus and presented in the hilar region
T_3	$TS > 5cm \parallel TS \leq 7cm$	Attacks the chest wall, pericardium and phrenic nerves.
T_4	$TS > 7cm$	Diaphragm and vessels, trachea are attacked

Based on the Lung Cancer staging criteria, the proposed work concentrates on developing **Accurate Multi-Level Classification Model for Lung Cancer Diagnosis and Staging,** using the CT lung images as input, which is more reliable with minimal distortion and better clarity. Furthermore, the contributions of the proposed model are provided below.

1. The input CT images are pre-processed with Median and Adaptive Median Filter.
2. Segmentation is processed with GC-PSO for exactly locating the cancer area
3. First Level Classification is determined with SVM for classifying Normal and Abnormal classes.
4. Second level classification is processed based on the tumor size measurement for classifying the lung images based on the Table 1 criteria, using Adaptive Neuro Fuzzy Model.
5. The obtained results are evaluated based on metrics such as, sensitivity, specificity, error rate, accuracy and time effectiveness.

The remainder sections are framed as follows: Section 2 describes about the image processing and machine learning techniques involved

Among various tumor types, lung tumor is life threatening and common, can be classified as follows,

i. Small Cell Lung Cancer (SCLC)
ii. Non-Small Cell Lung Cancer (NSCLC)

Further, the NSCLC types are classified into the following sub-groups based on the Tumor (T), Lymph Node (N), and Metastasis (M)- TNM based staging, defined as, American Joint Committee on Cancer Classifiers. The classification of lung tumor is presented in Figure 2.

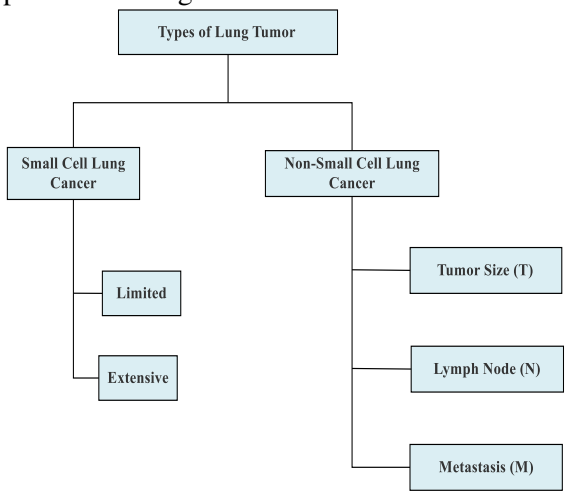


Figure 2: Types of Lung Tumor

Further, the lung tumor staging can be processed with the following steps [7].

1. Stage I- tumor cells are detained only to the lung
2. Stage II- tumor cells are extended to the chest region
3. Stage III- Large tumor cells those are constant to the chest region

4. Stage IV-Tumor cells that are extending to the remaining parts of the body

The proposed model uses the input CT image for processing and classification. The contributions of this proposed work is provided below.

1. The background separation for lung images is processed with global thresholding.
2. Image pre-processing is done with morphological operations.
3. The lung image segmentation is processed with FCM based clustering
4. Deep Neural Networks (TNM) are used for lung tumor staging and classification
5. Performance evaluations of the classification based on metrics such as, precision, accuracy rate, error rate, recall and time complexities.
6. For evidencing model efficiency, the comparison evaluations are carried out with the existing model.

The remainder of this paper is organized as follows, Section 2 describes about the methods and materials of existing work. The working procedure of the proposed model is clearly explained in Section 3. The Section 4 presents the implementation process and result evaluations of the proposed work. Further, the conclusion and future work is provided in the final section 5.

2. Related Works:

An automated Tumor Diagnosis model from CT Lung images has been proposed in [26], using Marker-Controlled Watershed algorithm. For classification, Support Vector Machine (SVM) has been used. The input CT lung images are pre-processed using Gabor Filter. The significant features such as area, perimeter and eccentricity, and given for training and testing. However, the accuracy is not as expected in cancer staging. In [5], [27], sharpening methods are used for results enhancement in addition to the marker-controlled watershed based segmentation.

The work presented in [6], [28], uses the Multi Class SVM, for cancer detection. For pre-processing, median filter has been adopted, in which the detection process has been determined in 3 stages. Deep learning methods are incorporated in [29] for providing better rate of accuracy in cancer image detection. Convolution Neural Networks (CNN) is the commonly used deep learning model for effective lung tumor detection. In [8], [30], Double Convolutional Deep Neural Network (CDNN) has been used for tumor staging. The dataset images were obtained from the Laboratory of Neuro Imaging (LONI) dataset, in which the sample images are categorized into pile 1 and pile 2, as, cancerous and non-cancerous, respectively. Moreover, the work utilized k-Means clustering for classification. And, the model has been combined with convolutional layer and additional pooling layer for accurate results.

In [9], [31], Level set-Active Contour Modeling has been used for lung tumor detection, used histogram based feature extraction. Another work in [10], [32], the input images were enhanced with the effective incorporation of Gabor Filter, and Fast Fourier Transform (FFT). And, segmentation has been processed with thresholding and watershed model. The works [11], [12], [13], [14], [15], [33], [34], [35], [36] and [37] used Grey-Level Co-occurence Matrix (GLCM) for feature extraction. Moreover, the works [16], [17], [38], [39], used Local Binary Pattern (LBP) based segmentation. The paper [40] utilized wavelet transform based feature extraction to extract the texture features of lung images. In a different manner, Genetic Algorithm (GA) has been enforced in [19], [41] for cancer diagnosis. However, the feature detection process is complicated with the number of CT images creates time complexity.

The survey works presented in [18], [42], [43], [44] described various techniques used for lung tumor diagnosis. And, the review work in presented the evolutionary algorithms incorporated in caner image diagnosis.

3. Proposed Work:

The main objective of the proposed model is lung cancer staging, for that, the proposed model comprises of the following sections,

1. Pre-Processing with Adaptive Median Filter
2. Tumor Segmentation
3. Feature Extraction
4. First Level Classification
5. Second Level Classification
6. Result Analysis

The diagrammatic representation of the proposed model work flow is clearly depicted in Figure 2. For that, the input images are obtained from benchmark lung image

datasets. The input CT image contains some noises, which can be processed with filter.

3.1. Image Pre-Processing:

In this process, Adaptive Median Filtering, which are nonlinear is used for reducing the noise from input image. The function includes the edge preserving smoothing operations, in which the sharp edges are smoothened, with the effective determination of median value as presented in (1).

$$Median\,[M(x)+N(x) \neq median\,[M(x)]+Median\,[N(x)]] \quad (1)$$

where, $M(x) and N(x)$ are the two images.

3.2. Tumor Region Segmentation:

The lung images comprises of large region of air, in which the radiographic density ranges from -600 HU to -400 HU in the images. The significant part is to separate the air-fillled regions from the lung tissues. Moreover, each pixel of the tumor region denotes the gray rate based on the HU value of the anatomical structure of lungs. The connectivity of pixel grey rate and the corresponding density is given as,

$$c(P_{(a,b)}) = \frac{HU-y}{n} \quad (3)$$

Where, y, denotes the rescale intercept and the slope rate determined by 'n'. Here, the segmentation is processed with Guaranteed Convergence Particle Swarm Optimization (GCPSO).

3.3. Feature Extraction Operation:

After segmentation, feature extraction operations are employed for obtaining significant features of input CT images for performing efficient training and testing in first and second level of classifications. Moreover, the probability mass function is determined from the input histogram images, which are further processed with the following steps.

i. Determine the Cumulative Distributive Function (CDF) based on the gray levels

ii. Gray levels are determined by,

$$CDF_{new} = CDF \times (No.of\ gray\ levels - 1) \quad (4)$$

iii. The new gray levels are mapped with the number of pixels and the equivalent histogram is determined.

The process of feature extraction extracts the most important features of segmented lung tumor, which is provided as the input for the classification model. The size and shape of the tumor image is accurately measured with the determinations of three following geometric features, which are, area, perimeter and eccentricity.

i. Area of the segmented cancer is computed based on the number of pixels that are in the lung nodule. The computations are carried out with the binary image by the summation of pixel rates in the image.

ii. Perimeter is the factor defines the scalar quantity provides the total pixel denotes the tumor border. It is computed with adding the registered value with 1, at the boundary.

iii. Eccentricity (EC) is the factor can be defined as irregularity index or the roundness. The computation is given as,

$$EC = \frac{length(major\ axis)}{length\ (minor\ axis)} \quad (5)$$

3.4. First Level Classification:

For performing first level classification, the cancerous and non-cancerous images are separated from the input CT images. For that, Support Vector Machine based classification technique is incorporated, which is efficient in performing binary classifications. The technique defines the input vector, which is having the high-dimensional space. With that, the optimal hyper-plane is determined. Moreover, by the effective determination of optimal hyper-plane, the appropriate classification results for cancerous and non-cancerous images are determined. And, the equation is given as,

$$MX + N = 0 \quad (7)$$

Where, X, is the number of CT image samples, M, perpendicular vector that separates the hyper-plane and 'N' is the offset factor. And, the pictorial representation in presented in Figure 3. In SVM, the extracted features are used for training the model and perform appropriate classification during the process of training. The result classifies the input images under cancerous and non-cancerous images. And, the non-cancerous images are further given for second level classification. In second level classification, the images are further processed for perform cancer staging.

3.5. Second Level Classification:

The main objective of the second level classification is to determine the cancer staging appropriately, as, Stage I, Stage II, Stage III and Stage IV. Based on the results, the earlier treatment can be effectively provided to the patients and thereby, mortality rate can be reduced to the whole. For that, Neuro Fuzzy Interface Model is used in this phase of work for cancer staging. Furthermore, the classifier is built with Sugeno Fuzzy Interference Model, effectively performs with conditional cases, IF-THEN. And, the classifier model with Neuro Fuzzy Interface comprises of five layers for performing accurate results.

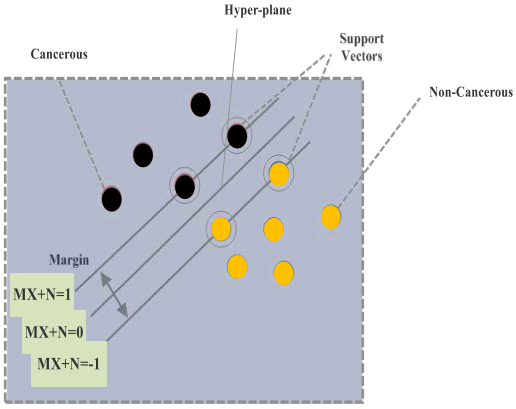


Figure 3: SVM in Binary Classification

1. In the first layer, the extracted feature vectors are converted into levels as, low, medium and high, ranges between 0 and 1. And, the output of this layer is determined as the fuzzy facto's grade point. Further, the factors with comprehensive bell function (bf) is given as,

$$bf(i,e) = \frac{1}{1+\{(\frac{m-C_c}{W_c})^2\}^{S_c}} \quad (8)$$

Where, C_c, is the center of curve, W_c, is the its width and 'S_c' is the slope of curve. Based on the computations, the curve parameters are changed with respect to the fuzzy factors and defined with the fuzzy-rule set using IF class.

2. Second layer for framing the fuzzy rule set for performing training operations. The execution of this layer is processed with the formation of fuzzy rule sets based on IF-Then Classes.

3. The third layer also defines the rule set with fuzzy logics with respect to the normalized derivations based on the previous level.

4. The resultant parameter is provided at the fourth layer, represents THEN condition for processing.

5. The collective results are determined in the fifth layer by combining the previous layer results based on the feature vectors (F).

The classifier results four classes, as, Stage I, Stage II, Stage III and Stage IV based on the extracted features. Moreover, the Rule-set (RS) based fuzzy interference is computed in the following equation.

$$RS_K = IF\,(F_1\ is\ bf(1,1))\,AND(F_1\ is\ bf(i,e)\ THEN\ (C\ is\ C_{out})) \quad (9)$$

3.6. Metrics for Result Analysis:

For evaluating the performance of the proposed model for cancer staging, different evaluation metrics are used. The statistical measures such as, classification accuracy (CA), Precision Value (PV), Sensitivity, Specificity, F1 Score, Area under the ROC (AUC) are used. Here, the CA is calculated based on the cross-validation of the processing with training and testing operations with input CT images. For determining the metric results, the results of True Positive, True Negative, False Positive and False Negative rates from the classification results.

The CA is defined as the ratio of number of correctly classified samples to the total amount of images taken. Here, the k-fold cross validation is processed with the classification result; average error rate is calculated, based on the following equation.

$$Err = \frac{1}{K}\sum_{i=0}^{K}\frac{FP_i+FN_i}{TP_i+TN_i+FP_i+FN_i} \quad (10)$$

Other computations for Precision Value (PV), Sensitivity, Specificity, F1 Score, Area under the ROC (AUC) are provided below.

$$True\ Positive\ Rate\ (TPR) or\ Sensitivity = \frac{TruePositive}{TruePositive+FalseNegative} \quad (11)$$

$$True\ Negative\ Rate\ (TNR) or\ Specificity = \frac{TrueNegative}{TrueNegative+FalsePositive} \quad (12)$$

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$F_1 - Score = \frac{2*precision*recall}{precision+recall} \quad (14)$$

$$CA = \left(\frac{TP+TN}{Tot.Samples}\right) \times 100\% \quad (15)$$

This section describes the working process of the proposed model for appropriate lung tumor segmentation and staging. As the model concentrates on measuring the volume of tumor cell, the three dimensional reconstruction of CT image is defined with width and height of each pixel. Here, the pixel volume can be computed as 3D voxel. For appropriate tumor segmentation, morphological operations are processed. Then, FCM algorithm is used for grouping up of the pixels into three clusters as, background cells, tumor cells with airways and soft lung tissues, based on their intensity rates. The clusters including tumor cells are further considered for airways removal with respect to the tumor cell properties. Based on the size of the tumor, staging will be performed using the Deep Learning model called, Deep Neural Networks (DNN).

3.7. Fuzzy C Means Clustering for Lung Tumor Segmentation:

FCM based image segmentation is used here for making exact decision making and clinical analysis. The segmentation is processed with contouring based process. Based on that, similar clusters are formed by combining the pixels in the feature space as in Figure 4. And, the steps for processing in FCM for lung tumor segmentation are presented below (given in Figure 5).

i. choosing two centroids with minimal value

ii. Assign values for centroids

iii. Compute Membership Matrix

$$MM_{x,y} = \frac{1}{\sum_{j=1}^{M}\left[\frac{|a_i - M_j|}{|a_j - M_n|}\right]^{\frac{2}{k-1}}} \quad (16)$$

Here, 'M' denotes the cluster number, where 'MM'>1.

iv. Compute centroids for clusters

$$c_{center} = \frac{\sum_{a=1}^{c} M_{i,j}^{k} * a_x}{\sum_{a=}^{c} M_{xy}^{k}} \quad (17)$$

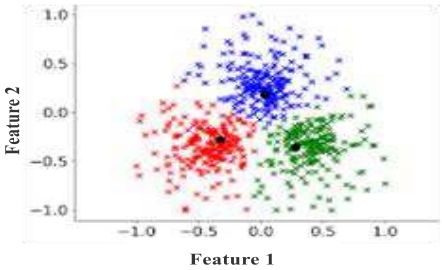


Figure 4: Clusters by combining Pixels

The computation process continues till the complete segmentation of cancer nodule is determined and the numbers of iterations are noted.

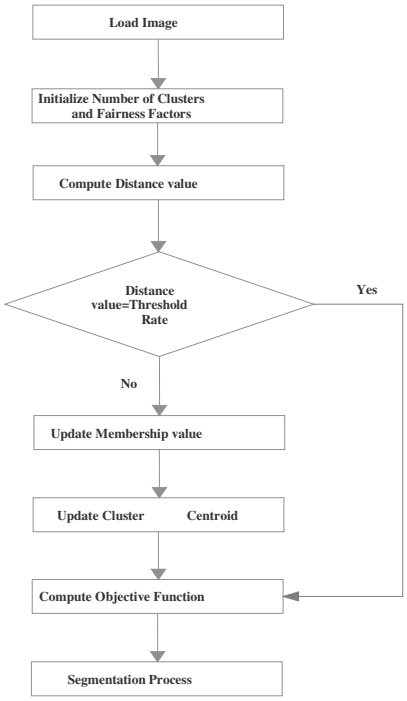


Figure 5: Process of FCM

3.8. Lung Tumor staging using Deep Neural Networks (DNN):

In this section, DNN is used for tumor staging based on TNM based shape evaluations. The framework of deep neural networks comprises of hidden nodes that were enhanced for processing with complex calculations. The framework of DNN model for lung tumor staging is depicted in the following figure 6.

In performing TNM based classification, lung tumor classifies the tumor nodule from benign to malignant, into four stages. Each tumor stage illustrates the survival rate with decisions for treatments based on tumor shape and size. Based on the diameter of the tumor nodule, the primary tumor (T) is classified as, $\{T_X, T_1, T_2, T_3, T_4\}$, the regional lymph nodes (N) $\{N_X, N_0, N_1, N_2, N_3\}$, and, the distant metastasis (M) is categorized as, $\{M_X, M_0, M_1\}$. According to the sub-staging, the final stage of tumor is diagnosed.

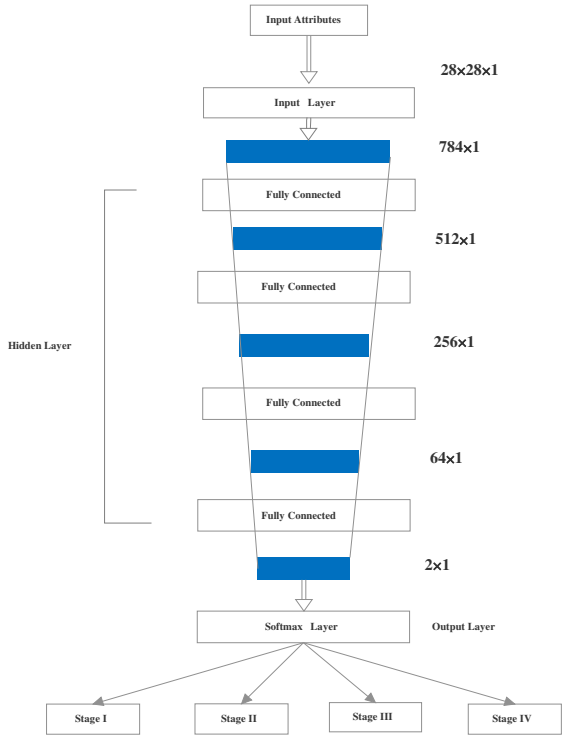


Figure 6: Architecture of DNN for Lung Tumor Staging

Moreover, the tumor spreading is occurred based on tumor diameter, spreading rate of lung nodule and lung metastasis. In the proposed model, the diagnosed tumor cell is categorized into 'T' sub-stage, as the other type of staging requires regional and distant rate of tumor data. The diameter and center of the tumor nodule was calculated based on the connected regions feature. The maximal rate of diameter of two planes (XY) is selected as the larger diameter of the three dimensional rate of the primary tumor and the cented connected to the larger diameter is selected as centroid. The tumor diameter is vital as a minor variation, significantly changes the treatment process and recovery rate of the patient.

In the training process of the model, in each layer of DNN, the random weight is derived with respect to the Gaussian distribution rate. And, the output is evaluated with respect to the operations of forward and backward propagation. The lung tumor staging and classification was determined and computed based on the categorization.

3.9. Result Evaluation Metrics:

For analyzing the model performance of the proposed work, the factors such as Classification Accuracy (CA), Precision Value (PV), Sensitivity, Specificity, Error Rate (ER) and F1-measure. The evaluations are processed based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values.

a. Classification Accuracy:

CA can be defined as the amount of correctly classified CT images to the amount of total images tested. The computation is given as,

$$CA = \frac{Amount\ of\ accurately\ classified\ images}{No.of\ total image\ tested} \times 100\% \quad (18)$$

b. Precision Value:

Precision Value can be calculated based on the classification of appropriate stages. And, the formula is given as,

$$PV = \left(\frac{True\ Positive}{True\ Positive + False\ Positive}\right) \times 100\% \quad (19)$$

c. Sensitivity:

Rate of sensitivity is defined as the True Positive rate obtained from the results and the formula is given as follows,

$$Sensitivity\ Rate = \frac{TruePositive}{TruePositive + FalseNegative} \quad (20)$$

d. Specificity:

Rate of Specificity is computed based on the true negative rate, can also be denoted as Recall and the computation is given as,

$$Specificity\ Rate = \frac{TrueNegative}{TrueNegative + FalsePositive} \quad (21)$$

e. Error Computation:

The average error rate in diagnosing the tumor staging is calculated as in the following equation.

$$Error\ Rate = \frac{1}{n}\sum_{i=0}^{n}\frac{FP_i + FN_i}{TP_i + TN_i + FP_i + FN_i} \quad (22)$$

e. F1-Measure:

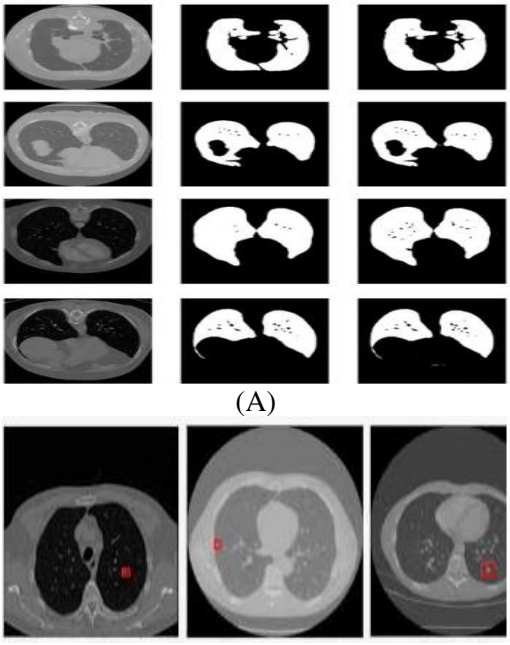
F1 measure is calculated from the precision value and recall rates of previous results.

$$F_1 - Score = \frac{2 * PV * recall}{PV + recall} \quad (23)$$

4. Dataset Description:

The input lung CT images are obtained from two datasets, Lung Image Database

Consortium (LIDC-IDRI) [20] and LUNA 16 dataset [21]. The LIDC-IDRI dataset comprises of 1018 patients that are obtained from seven academic centers and medical imaging companies. The data contains an XML file with explanations of CT scan. There are 244, 527 CT image from 1010 patients in the data set. Based on that, the detection process is done at two levels, as, patient level and nodule level. Here, the DICOM images are with resolution 512×512, in which the width ranges from 65 to 764 slices. Further, another dataset LUNA 16 is the subset of LIDC-IDRI dataset, comprises of heterogeneous scans. Since, the CT lung images are very thin slices, the slice thickness with more than 2.5 mm are avoided. There are 36,378 CT images, which are classified as nodules greater than 3 mm and non-nodules. The lung nodules with different headers that is nearer than the summation of the radii that is merged. The data set results in framing 2290, 1602, 1186 and 777 nodules. The sample images from the aforementioned two data sets are presented in Figure 7.



(A)

(B)

Figure 7: Sample Lung Input Images from LIDC-IDRI (A) and LUNA 16 (B) datasets

The proposed model is trained with input samples from Lung Image Database Consortium (LIDC-IDRI) [29] [30]. The dataset includes images from medical imaging companies and academic centers. Additionally, the samples are categorized into three groups, as follows.

$$\begin{cases} Size\_of\_nodule \geq 3 \\ Size\_of\_nodule < 3 \\ non-nodule \geq 3 \end{cases}$$

All the images in the dataset are with 512×512 resolution rate and width is given as, 65-760 slices. The different images from the dataset, used for training and testing, are presented in Figure 8.

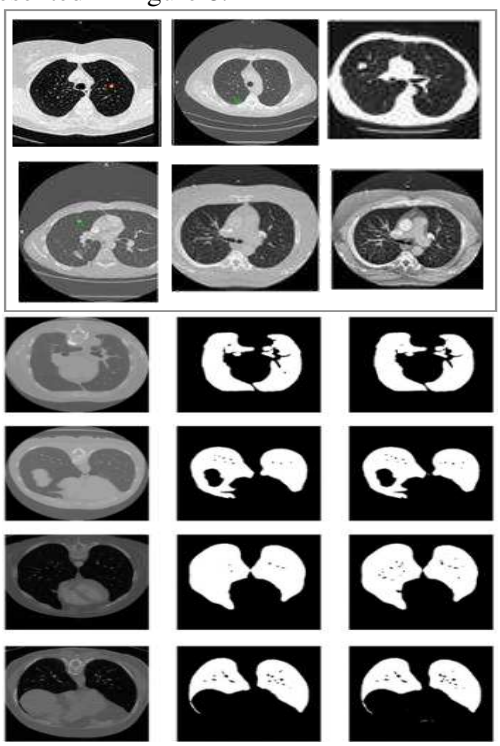


Figure 8: Samples from Dataset

5. Results and Discussions:

The Genetic algorithm, Machine learning algorithm and Deep learning techniques are employed in this analyzing Lung cancer CT images and results of classification process are discussed here. The performance is analysed with the classification of images on the following values,

i. True positive
ii. True negative
iii. False positive
iv. False negative

By determining the above values, the sensitivity (True Positive Rate), Specificity (Recall), Precision and Accuracy Rate are calculated.

5.1 Filters in Image Pre- processing

The Lung cancer CT images are preprocessed using average filter, Median filter and adaptive bilateral filter. The results of SSI and SMPI as explained before, which are discussed for the average filter, median filter and adaptive bilateral filter. Table 2

shows the SSI and SMPI value for the three filters.

Table 2: SSI and SMPI measurement for Average filter, Median filter and Adaptive bilateral filter

Sample Images	SSI			SMPI		
	Mean	Median	Adaptive	Mean	Median	Adaptive
Image 1	0.8086	0.8208	0.9621	0.9638	0.9788	0.9857
Image 2	0.8087	0.8232	0.9658	0.9452	0.959	0.9895
Image 3	0.8091	0.8209	0.9588	0.9696	0.9799	0.9883
Image 4	0.7937	0.808	0.9671	0.9703	0.9836	0.9958
Image 5	0.8078	0.822	0.9705	0.9706	0.9833	0.9851

5.2 Segmentation results

The Lung cancer CT images are segmented using particle swarm optimization (PSO), Guaranteed Convergence particle swarm optimization (GPSO), and Modified GPSO and Fuzzy C-means techniques. The figure 9 and 10 depicts the true positive and true negative comparison and false positive and false negative comparisons respectively for the PSO, GPSO, MGPSO and FCM methods that have been implemented and described.

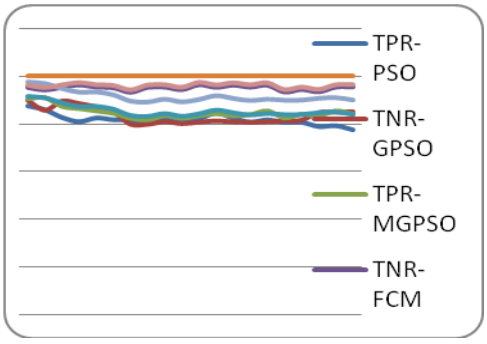


Figure 9: True positive and True negative rate (PSO, GPSO, M-GPSO, FCM)

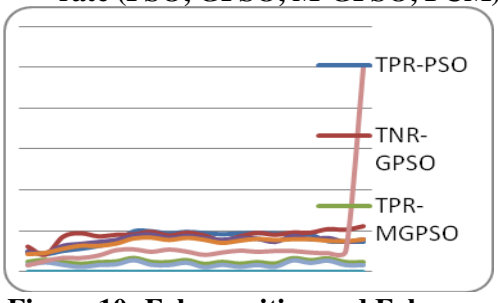


Figure 10: False positive and False negative rate (PSO, GPSO, M-GPSO, FCM)

5.3 Classification Results

The results that obtained from segmentation are classified by PNN, Multi level classification and DNN method. The classification performances of PNN for 20 sample images are depicted in table 3 In MGCPSO, a new particle (φ) identical to GCPSO is created, which searches the region surrounding the current personal best position.

Table 3: Overall classification performance

Metric	PNN	Multi-Level	DNN
TPR	95.35	97.013	98.17
TNR	94.6	95.3	97.4
FPR	0.026617	0.053312	0.063955
FNR	6.9724	4.0074	3.8283
Precision	93.0276	95.9926	96.1717
Accuracy	94.4736	95.7814	96.99625

Table 4 depicts the comparison of classification accuracy of the existing and proposed methods. Chaunzwa, T. L(2021) have implemented KNN model for cancer image classification and obtained 71% of accuracy; also implemented Linear SVM and obtained 68% of classification accuracy; and by CNN model obtained 72% of classification accuracy. Qin R(2020) have implemented CNN (multi grained method) and obtained 72% of classification accuracy.

Table 4: Comparison of existing and proposed method

	Methods	Accuracy %
Proposed	PNN	94.4736
	Multi-level Classification	95.7814
	DNN(Volumetric Analysis)	96.99625
Chaunzwa, T. L(2021)	KNN	71
	linear SVM	68
	CNN	72
Qin,R(2020)	CNN(multi trained)	72

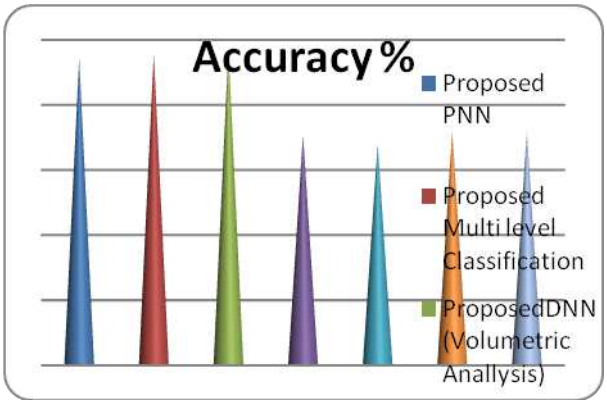


Figure 11: Comparison of existing and proposed method

The proposed methodologies as described, the classification accuracy that have obtained is better while comparing with the above-mentioned methodologies. Figure 11 shows the comparison of existing and proposed methodologies. The proposed DNN technique along with the Fuzzy C- Means segmentation obtains the highest classification accuracy as 96.99625%.

6. Conclusion and Future work

By comparing classification techniques along with the work methodologies the Deep learning technique that implemented to identify the sub stages of the tumour obtains the better accuracy results. This volumetric analysis identifies the tumour size, shape and location based on the tumour nodule metastasis. Whereas the PNN classification along with MGPSO segmentation identifies the tumour and Multilevel classification using SVM classification along with PSO segmentation identifies the tumour stage. In future, the model can be further enhanced for processing with real-time medical data for appropriate diagnosis and staging.

Declarations:
Ethical Approval - Not applicable
Funding – No funding received for this work
Availability of data and materials – The data analyzed can be accessed from the manuscript for the current study

References:

[1] Weissferdt, A.; Moran, C.A. "Diagnostic Pathology of Pleuropulmonary Neoplasia"; Springer: New York, NY, USA, 2013; ISBN 978-1-4419-0786-8.

[2] Oda, H.; Bhatia, K.K.; Oda, M.; Kitasaka, T.; Iwano, S.; Homma, H.; Takabatake, H.; Mori, M.; Natori, H.; Schnabel, J.A.; et al. "Automated mediastinal lymph node detection from CT volumes based on intensity targeted radial structure tensor analysis". J. Med. Imaging 2017, Vol. 4, 044502.

[3] Loutfi, S.; Khankan, A.; Ghanim, S.A. "Guidelines for multimodality radiological staging of lung cancer" J. Infect. Public Health 2012, Vol. 5, pp. 14–21.

[4] Kulkarni, A.; Panditrao, A. "Classification of lung cancer stages on CT scan images using image processing". In Proceedings of the 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, India, 8–10 May 2014; IEEE: Ramanathapuram, India, 2014; pp. 1384–1388.

[5] Ignatious, S.; Joseph, R.; John, J.; Prahladan, A. "Computer Aided Lung Cancer Detection and Tumor Staging in CT image using Image Processing". Int. J. Comput. Appl. 2015, Vol. 128, pp. 29–33.

[6] Jiao, Z.; Gao, X.; Wang, Y.; Li, J.; Xu, H. "Deep Convolutional Neural Networks for mental load classification based on EEG data". Pattern Recognition. 2018, Vol. 76, pp. 582–595.

[7] Jakimovski, G.; Davcev, D. "Using Double Convolution Neural Network for Lung Cancer Stage Detection". Appl. Sci. 2019, Vol. 9, 427.

[8] Dhara, A. K., Mukhopadhyay, S., Dutta, A., Garg, M., & Khandelwal, N. "A combination of shape and texture features for classification of pulmonary nodules in lung CT images". Journal of Digital Imaging, 29, 2016.

[9] Han, F.,Wang, H., Zhang, G., Han,H., Song, B., Li, L.,Moore,W., Lu, H., Zhao, H., & Liang, Z. "Texture feature analysis for computer-aided diagnosis on pulmonary nodules". Journal of Digital Imaging, 28, 2014, pp. 99–115.

[10] Wei, G., Cao, H., Ma, H., Qi, S., Qian, W., & Ma, Z. "Content-based image retrieval for lung nodule classification using texture features and learned distance metric". Journal of Medical Systems, 42, 13, 2017.

[11] Wei, G., Ma, H., Qian, W., Han, F., Jiang, H., Qi, S., & Qiu, M. "Lung nodule classification using local kernel regression models with out-of-sample extension. Biomedical Signal Processing and Control, Vol. 40, 2018, pp. 1–9.

[12] Sergeeva, M., Ryabchikov, I.,Glaznev,M.,&Gusarova, N. "Classification of pulmonary nodules on computed tomography scans. Evaluation

of the effectiveness of application of textural features extracted using wavelet transform of image," in 2016 18th Conference of Open Innovations Association and Seminar on Information Security and Protection of Information Technology (FRUCT-ISPIT), 2016, pp 291–299.

[13] Shen,W., Zhou, M., Yang, F., Yang, C., & Tian, J. " Multiscale convolutional neural networks for lung nodule classification BT" information processing in medical imaging, 2015, pp. 588–599.

[14] Shen, W., Zhou, M., Yang, F., Yu, D., Dong, D., Yang, C., Zang, Y., & Tian, J. " Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification". Pattern Recognition, 61, 2016.

[15] Li, X., Shen, L., & Luo, S. " A solitary feature-based lung nodule detection approach for chest X-ray radiographs". IEEE Journal of Biomedical and Health Informatics, Vol. 22, No. 2, 2018, pp. 516–524.

[16] Boroczky, L., Zhao, L., & Lee, K. P. "Feature subset selection for improving the performance of false positive reduction in lung nodule CAD". IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 3, 2006, pp. 504–511.

[17] Armato Samuel, G., III; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Zhao, B. Aberle, D.R.; Henschke, C.I.; Hoffman, E.A, "Data From LIDC-IDRI". 2015.

[18] Setio, A.A.A.; Traverso, A.; de Bel, T.; Berens, M.S.N.; van den Bogaard, C.; Cerello, P.; Chen, H.; Dou, Q.; Fantacci, M.E.; Geurts, B.; "Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images": The LUNA16 Challenge. Med. Image Anal. **2017**, Vol. 42, pp. 1–13.

[19] Hesamoddin Hosseini, Reza Monsefi, Shabnam Shadroo, "Deep Learning Applications for Lung Cancer Diagnosis: A systematic review" Image and Video Processing (eess.IV); Computer Vision and Pattern Recognition (cs.CV); Machine Learning, 2022.

[20] Xinrong Lu, Y. A. Nanehkaran, and Maryam Karimi Fard, "A Method for Optimal Detection of Lung Cancer Based on Deep Learning Optimized by Marine Predators Algorithm" Swarm Intelligence and Neural Network Schemes for Biomedical Data Evaluation, Computational Intelligence and Neuroscience, 2021

[21] Saba T, Khan MA, Rehman A, "Region extraction and classification ofskin Cancer: a heterogeneous framework of deep CNN features fusion and reduction". J Med Syst 2019, Vol. 43:289.

[22] Abbas N, Saba T, Mehmood Z, Rehman A, Islam N, Ahmed KT. "An automatednuclei segmentation of leukocytes from microscopic digital images". Pak JPharm Sci 2019, Vol. 32, No. 5, pp. 2123–38.

[23] Abbas N, Saba T, Rehman A, Mehmood Z, Javaid N, Tahir M, "Plas-modium species aware based quantification of malaria, parasitemia in light microscopy thin blood smear". Microsc Res Tech 2019, Vol. 82, pp. 1198–214.

[24] Fahad HM, Khan MUG, Saba T, Rehman A, Iqbal S. "Microscopic abnormality classification of cardiac murmurs using ANFIS and HMM". Microsc Res Tech, 2018, Vol. 81, No. 5, pp. 449–457.

[25] Kurihara Y, Matsuoka S, Yamashiro T, Fujikawa A, Matsushita S, Yagihashi K, "MRI of pulmonary nodules". Am J Roentgenol 2014, Vol. 202, pp. 10–6.

[26] Iftikhar S, Fatima K, Rehman A, Almazyad AS, Saba T. "An evolution basedhybrid approach for heart diseases classification and associated risk factors identification". Biomed Res, 2017, Vol. 28, No. 8, pp. 3451–5.

[27] Anita Chaudhary, Sonit Sukhraj Singh "Lung cancer detection on CT images using image processing", computing sciences 2012 international conference, IEEE, 2012.

[28] A. Mansoor,"Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends," Radio Graphics , vol. 35 no. 4 pp. 1056-1076, 2015

[29] J. Wang, "Prediction of Malignant and Benign of Lung Tumor using a Quantitative Radiomic Method," Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1272-1275, 2016.

[30] E. Dandil, "Artificial neural network-based classification system for lung nodules on computed tomography scans," 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), pp.382-386, 2014.

[31] Huang, C. H., Zeng, C., Wang, Y. C., Peng, H. Y., Lin, C. S., Chang, C. J., &

Yang, H. Y. “A study of diagnostic accuracy using a chemical sensor array and a machine learning technique to detect lung cancer”. Sensors, Vol. 18, No. 9, pp. 28-45, 2018.

[32] Nasser, I. M., & Abu-Naser, S. S. “Lung Cancer Detection Using Artificial Neural Network”. International Journal of Engineering and Information Systems (IJEAIS), Vol. 3, No. 3, pp. 17-23, 2019.

[33] Manikandan, T., & Bharathi, N. “Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier”. Journal of medical systems, Vol. 40, No. 7, 2016.

[34] Mokhled S. Al-Tarawneh, “Lung Cancer Detection Using Image Processing Techniques”, Leonardo Electronic Journal of Practices and Technologies, Issue 20, January- June 2012, pp.147-158. ISSN 1583-1078.

[35] Boroczky, L., Zhao, L., & Lee, K. P. “Feature subset selection for improving the performance of false positive reduction in lung nodule CAD”. IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 3, 2006, pp. 504–511.

[36] Bhatia, S., Sinha, Y., & Goel, L. “Lung cancer detection: A deep learning approach”. In Soft Computing for Problem Solving, pp. 699-705, Springer, Singapore, 2019.

[37] Ignatious, S.; Joseph, R.; John, J.; Prahladan, A. “Computer Aided Lung Cancer Detection and Tumor Staging in CT image using Image Processing”. Int. J. Comput. Appl. 2015, Vol. 128, pp. 29–33.

[38] Suzuki K, Abe H, MacMahon H, Doi K, “Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN)”. IEEE Trans Med Imaging, 2006, 25(4):406–416.

[39] Kavitha MS, Shanthini J, Sabitha R, “ECM-CSD: an efficient classification model for cancer stage diagnosis in CT lung images using FCM and SVM techniques. J Med Syst, 2019.

[40] Thiyagarajan M, “A survey on computer-aided diagnosis systems for lung cancer detection”. Int Res J Eng Technol IRJET, 03(05), 2016, 1562–1570.

[41] Kavitha MS, Shanthini J, Bhavadharini RM, “ECIDS-enhanced cancer image diagnosis and segmentation using artificial neural networks and active contour modelling”. J Med Imaging Health Inform 10(2):428–434, 2020.

[42] Senthil Kumar K, Venkatalakshmi K, Karthikeyan K “Lung cancer detection using image segmentation by means of various evolutionary algorithms” Comput Math Methods Med, 2019.

[43] https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI

[44] Armato Samuel, G., III; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Zhao, B. Aberle, D.R.; Henschke, C.I.; Hoffman, E.A, “Data From LIDC-IDRI”. 2015.

Dr. K. Jagadeesh is presently Assistant Professor, Department of Electronics and Communication Engineering, SNS College of Engineering, Coimbatore – 641107, affiliated to Anna University, Chennai, Tamilnadu, India. He received the B.E, M.E and Ph.D. degree from Anna University, Chennai. His Research interests include Medical Image Processing, Communication Networks, Machine Learning and Neural Networks.

Professor Dr. A. Rajendran is presently Professor, Department of Electronics and Communication Engineering, Karpagam College of Engineering, Coimbatore – 641032, affiliated to Anna University, Chennai, Tamilnadu, India. He received the B.E, M.S (By Research) and Ph.D. degree from Anna University, Chennai. His Research interests include Medical Imaging, Pattern Recognition, Image Processing, VLSI Signal Processing and Wireless Communication. He is also reviewer in top most journals like Informatics in Medicine - Elsevier Publication, IEEE Journal of Biomedical and Health Informatics and etc.,