ISSN 2321-743X
International Journal of Research in
Engineering and Bioscience
Volume 7 Issue 3 (Pages 40-48)
IJREB
Journal home page: www.ijreb.org

# DETECTION OF LUNG CANCER USING ADVANCED IMAGE SEGMENTATION ALGORITHM

Sneha S. Nair[1]*, Dr. V. N. Meena Devi[1], Dr. Saju Bhasi[2]

[1]Department of Physics, Noorul Islam Centre for Higher Education, Kumarakovil, Kanyakumari District, Tamilnadu, India.

[2]Department of Radiation Physics, Regional Cancer Centre, Thiruvananthapuram, Kerala, India.
*Corresponding Author
Email address of Author(s): n.sneha85@gmail.com

## ABSTRACT

The predicted outcomes of tumor diagnosis will form the foundation of a Computer assisted Diagnosis framework to identify lung disease early, hence increasing the patient's chance of survival. The extreme variety in the weak level and the relative difference among the pictures divide results less exact, in this way machine learning techniques are used for feature selection, extraction, and in disease prediction. In this research work, a machine learning mechanism with neural networks is used for lung cancer prediction which improves the system performance. The segmentation technique for cancer identification presented here is Modified adaptive threshold segmentation and Support Vector Machines classifier as well as Artificial Neural Network classifier. Here, we conduct tests using the Lung Image Database Consortium datasets, that comprise a database of Computed Tomography pictures, and use these even as input photos to test the efficacy of the suggested strategy. Descriptive and inferential statistical analysis is used to assess the segmentation efficiency of the recommended method. Regarding lung cancer identification, the suggested method scores 96.3% when using an Artificial Neural Network classifier as well as 97% when it employs a Support Vector Machines classifier. This strategy inspires radiologists as well as officials to pay more attention to lung tumors in less time and with greater accuracy.

***Keywords:*** *-Classifier, Lung cancer, Median filter, Tomography, Threshold*

## INTRODUCTION

Throughout the planet, lung tumors are a major cause of mortality. According to the stage of discovery of the abnormal cells within the lungs, tumor development there in the lungs is among the most dangerous and widespread tumor development in the world. So the procedure of early identification of the illness plays a vital and fundamental job to keep away from genuine propelled stages to reduce its level of dissemination (Surendar,, 2021). The point of this exploration is to identify features for precise picture examination as pixels rate and tumor cell marking. Supervised learning, which is used by the vast majority of most CT image-based algorithms, is notoriously inaccurate and necessitates a significant number of manual segmentation training examples (Lakshmanaprabu, et al. 2019; Hu et al. 2020). If an optimum small amount of training image dataset can be constructed, in which each test has pulmonary nodules with comparable size and appearance as that of the target object of the genuine patient, these issues can be resolved (Lokhande, et al. 2022). Acquisition of images, pre-processing, lung delineation, nodule identification, and minimization of false positives are all standard procedures. Acquiring images of the lungs is the first stage (Huidrom, et al. 2022). There are different public databases available for research purposes. Among these is the Lung Image Database Consortium (LIDC), the Early Lung Cancer Action Program (ELCAP), the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) and the Reference Image Database to Evaluate Therapy Response (RIDER) (Ashwini, et al. 2021; Reddy, et al. 2022). There are numerous sorts of arrangement calculations or regularly termed classifiers were utilized for Tumor finding. Artificial Neural Networks, Rough Sets, Genetic Algorithms, Support Vector Machines, and Fuzzy Sets are only a few of the methods that can be used (Bhaskar, et al. 2022). Most useful for analyzing tumors and characterizing diseases have been supporting vector machine (SVM) and artificial neural network (ANN) classifiers. Both have performed exceptionally well in describing tumor development (Rehman, et al. 2022). Specifically, the goal of this study is to validate the use of both SVM and ANN for tumor classification. We test out the threshold segmentation technique also with two classifiers.

## MATERIALS AND METHODS

Filtering, fragmentation, as well as feature extraction, are the main 3 recognition system phases. To increase the precision as well as reliability of lung cancer diagnosis, the dataset's lung cancer scans can be employed as inputs, and these phases should be preserved. To pinpoint the contaminated area in the input images, we'll use a modified version of adaptive threshold segmentation (ADTM). lassifiers are used to increase detection accuracy. The MATLAB 2018a program is used to implement the suggested approaches for diagnosing cancer.


**Fig 1:**

**Schematic representation of the new method for identifying lung cancer.**

The suggested model details the current lung cancer diagnostic process is shown in Figure 1. Medical data are saved, exchanged, and sent utilizing standards of Digital Imaging and Communications in Medicine (DICOM) [2]. Due to the high-definition images contained within DICOM files, compression is commonly used to reduce their size. As a result, they are noisy. For this reason, it is compulsory to reduce image noise. Since median filtering keeps edges whilst getting rid of noise, it is employed. To determine the extent to which a given pixel is indicative of its environment, using median filter analyses the pixels bordering it. Adaptive thresholding is used in lung segmentation, and texture and form features are extracted by feature extraction. Then, ANN and SVM are used to classify the characteristics. Most commonly, the grayscale/color picture elements are used as inputs for adaptive thresholding (ADT), with the resulting binary image indicating fragmentation in its most basic form. The cutoff is determined for every pixel within the image based upon the local mean intensity within the area immediately surrounding the pixels, with both the sensitivity factor defined by sensitivity [4]. In addition to the Adaptive Threshold approach, we apply a filtering criteria in ADTM with the premise that the object's Centroid of Interest will be in the top 80% of the picture in order to eliminate the outermost ring pattern or portions of the outer ring structure acquired in some circumstances. 13 characteristics of shapes like Equiv Diameter, Area, Perimeter, Solidity Convex Area, Extent, Centroid, Eccentricity, Perimeter, Solidity, Euler Number, Extrema, Major as well as Minor Axis Length, Orientation. 7 GLCM Texture Features such as Cluster Prominence, Contrast, Dissimilarity, Entropy Energy, Homogeneity and Cluster Shade and 8 Intensity Features such as Skewness, Kurtosis, Mean, Smoothness, RMS, Variance, Standard Deviation, Inverse Difference Moment (IDM) are computed.

## RESULTS AND DISCUSSION

To distinguish itself from existing current structures and provide a higher performance of detection of lung cancer this suggested system presents a simultaneous thresholding technique as well as a powerful component extraction approach. This proposed method produces more reliable outcomes than the alternatives. First, convert every DICOM image in the LIDC database to JPEG format. This will yield 150 testing databases and 534 training databases with images of lung cancer, both malignant and benign.In preprocessing, median filtering is used as a result, edges are maintained whilst noise is suppressed. Figure 2 displays the pre-processed pictures.

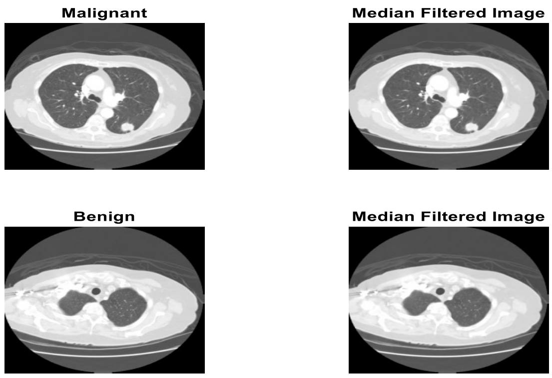


**Fig:2 Median filtered image after preprocessing**

The following procedures are involved in segmenting lung pictures using adaptive thresholding. Firstly, it calculates a global predefined threshold that can transform a grayscale image into a binary one. Image segmentation is typically used to find the location of systems as well as boundaries like lines, and curvatures in imageries. Figure 3 below illustrates a segmented image created with a thresholding technique. It displays significant data related to picture segmentation. The thresholded image has the dual benefits of being easier to analyze and taking down less space for storage.

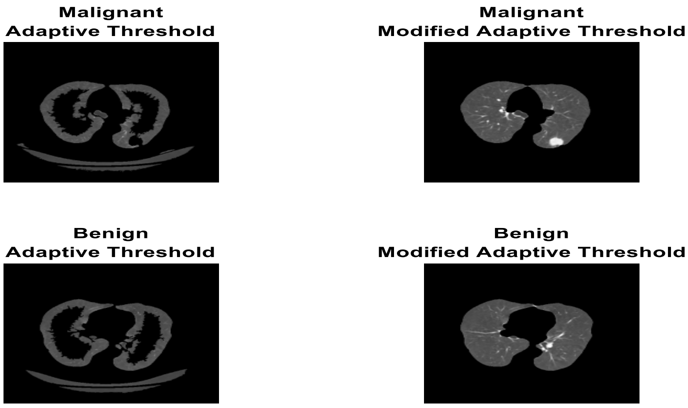


**Fig:3 The resultant images for malignant and benign images from adaptive and modified adaptive thresholds.**

Image processing and pattern recognition are the two main applications of feature extraction. The image's recurring patterns are the feature. Lung cancer is capable of detection using the binarization method, and the interesting area of the image can be extracted. Several steps are involved in identifying and separating particular shapes and pieces, one of which is feature extraction. Keeping track of grayscale and binary pixel data is the basis for the binarization process. For binarization, this is demonstrated that lung tissue photos have a significantly higher proportion of black pixel resolution compared to white pixels, and yet this proportion can serve as a cutoff point to determine whether or not a given image appears to be regular; if somehow the proportion of black pixels within the given image would exceed the threshold, then perhaps the given image seems to be normal, and vice versa. The above binary images are then subjected to morphological processes, which erode the binary image ie., removes small objects from a binary image. The next step was to stretch the images to cover in a certain bounding box that wouldn't otherwise be accessible from the picture's borders. To create a segmented binary image, first determine the area of the segmented section and mark it with a disk shape.

Calculate characteristics such as 13 form features, 7 GLCM texture features, and 8 intensity features from the segmented grayscale image. The single-level discrete 2-D wavelet transform's principal component coefficients are used to extract intensity characteristics. Area, centroid, convex area, eccentricity, equivDiameter, perimeter, euler number, range, extrema, minor, major axis length, orientation, and solidity are the 13 form attributes that were extracted.

Classification accuracy can be measured by the percentage of samples that are correctly labeled. 9 performance metrics like, Mathews Correlation Coefficient, Precision, Error, Accuracy, F1 Score, Specificity, Sensitivity, False Positive Rate and Kappa-Cohen's Kappa, are utilized to systematically analyse the classification effectiveness of the proposed technique values of ADT and ADTM with ANN. Using both the predict class labels and the actual class labels, the confusion matrix is created for several classes [55]. true positive (TP), false-positive (FP), false negative (FN) and TrueNegative(TN) values all exist within the Two-Class of Confusion Matrix. Accuracy, FPR-False positive percentage, errors, Selectivity,Sensitive (True positive rate/Recall), Precise, F1 score, Matthews correlation coefficient, as well as Cohen's kappa is measured to assess the results of ADT as well as ADTM utilizing ANN, as demonstrated in Figure 4.

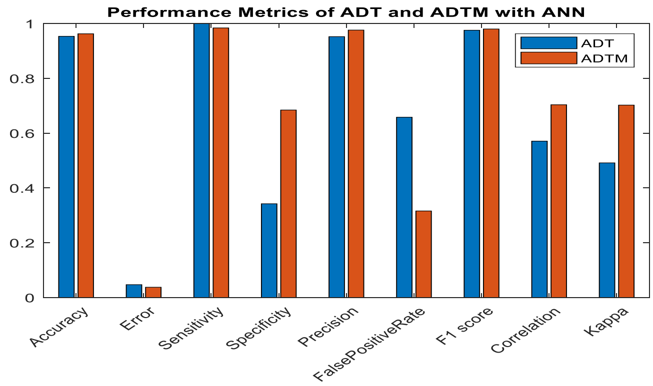


**Fig:4 Evaluation of ADT and ADTM with ANN Performance**

A target class and an output class confusion matrix are shown in Figure 5. Within the first two diagonal cells, we can see the total number of exact classifications made by the network which is trained and the percentage of those classifications. 488 of the 534 photos have been correctly marked as benign (the true positive). It is the same percentage (91.4%) as 534 out of 534 pictures. The malignancy of 12 instances is also confirmed (true negative). That's 2.2% of all photos, by the way. Only eight (1.5% of all cases) malignant pictures are mistakenly labeled as benign. Equally concerning is the fact that 4.9% of all data points are accounted for by the incorrect classification of benign images as malignant (26 cases). Overall, ADTM with ANN classifier obtain a 96.3% success rate with an error margin of 3.7%. Figure 6 shows an SVM classifier performance evaluation. The study found that the overall accuracy of 97.00% achieved is higher than that of the other classifiers examined. The SVM is least dependent on the sample size because it only uses support vectors to construct the segregating hyperplane, so increasing the number of training samples did not affect the accuracy.

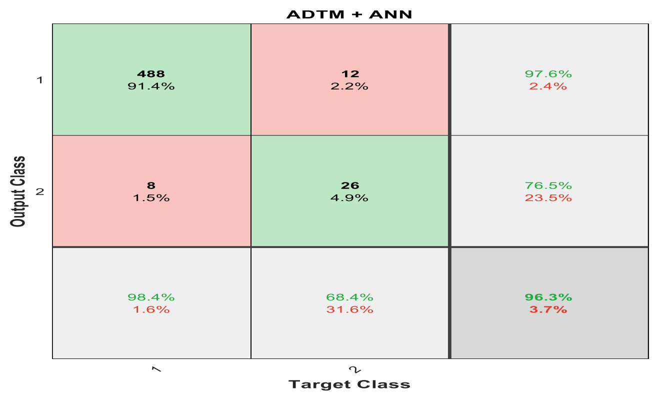


**Fig:5 Confusion Matrix for ADTM with ANN classifier**

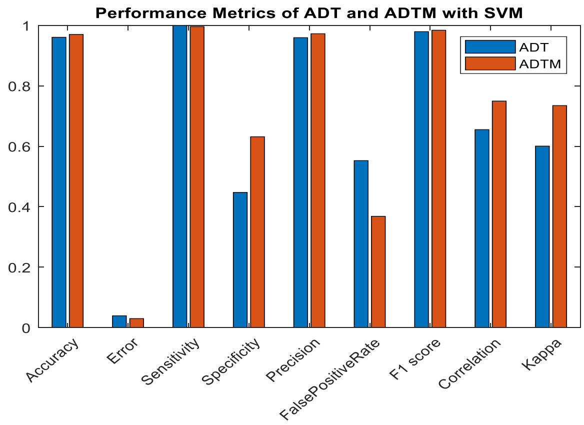


**Fig:6 Performance evaluation with SVM classifier**

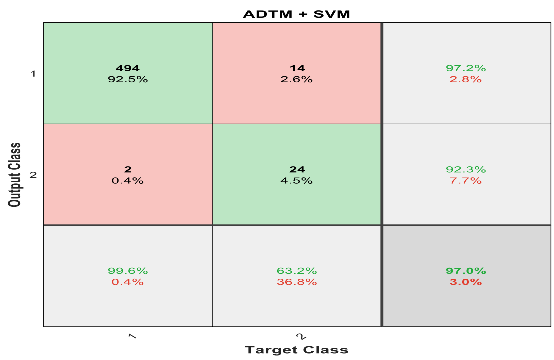


**Fig:7 ADTM with SVM's Confusion Matrix**

Figure 7 shows the target class and output class confusion matrix for ADTM with SVM. Of the 534 pictures, 494 are benign (the true positive). That's equivalent to 534 photos, or 92.5% of the total. Conversely, 14 instances are appropriately identified as malignant (true negative) (true negative). This translates to 2.6% among complete images. 2 out of each malignant picture are wrongly categorized that benign (false positive) but this equals 0.4%. Another 4.5% of said data set consists of 24 benign images that were labelled as cancer mistakenly (False negative). ADTM utilizing the SVM classifier has an error margin of 3.0% overall.

## CONCLUSION

A diagnosis is achieved in this study by employing picture preprocessing as well as image analysis. To detect lung cancer, binarization technology is employed to transform the image into a binary format, which is then compared with something like a threshold level. After the lung CT scan has been segmented, Important data is extracted using a strong feature extraction technique. These procedures allow for the detection of nodules and the extraction of relevant features. Extracted features are used to determine how to categorize diseases at various stages. We employ a feed-forward neural network in combination with an ANN and a support vector machine classifier to make predictions about the stages at which lung cancer might develop. Once the neural network has been trained using the extracted features, it is put to the test on both malignant and benign photos. Our proposed approach demonstrates enhanced accuracy in classifying lung cancer diagnoses. The proposed method's accuracy for lung cancer detection utilizing the ANN classifier is 96.3% and the SVM classifier is 97%, which indicates the SVM classifier is giving high accuracy with ADTM for the identification of lung cancer.

## REFERENCE


1. Surendar, P. "Diagnosis of lung cancer using hybrid deep neural network with adaptive sine cosine crow search algorithm." *Journal of Computational Science* 53 (2021): 101374.
2. Lakshmanaprabu, S. K., Sachi Nandan Mohanty, K. Shankar, N. Arunkumar, and Gustavo Ramirez. "Optimal deep learning model for classification of lung cancer on CT images." *Future Generation Computer Systems* 92 (2019): 374-382.
3. Hu, Qinhua, Luis Fabricio de F. Souza, Gabriel Bandeira Holanda, Shara SA Alves, Francisco Hercules dos S. Silva, Tao Han, and Pedro P. Reboucas Filho. "An effective approach for CT lung segmentation using mask region-based convolutional neural networks." *Artificial intelligence in medicine* 103 (2020): 101792.
4. Lokhande, Narendra Lalchand, and Tushar Hrishikesh Jaware. "Lung CT image segmentation: a convolutional neural network approach." In *Information and Communication Technology for Competitive Strategies (ICTCS 2020) ICT: Applications and Social Interfaces*, pp. 385-392. Springer Singapore, 2022.
5. Huidrom, Ratishchandra, Yambem Jina Chanu, and Khumanthem Manglem Singh. "Neuro-evolutional based computer aided detection system on computed tomography for the early detection of lung cancer." *Multimedia Tools and Applications* 81, no. 22 (2022): 32661-32673.
6. Ashwini, S. S., M. Z. Kurian, and M. Nagaraja. "Lung cancer detection and prediction using customized selective segmentation technique with SVM classifier." In *Emerging Research in Computing, Information, Communication and Applications: ERCICA 2020, Volume 2*, pp. 37-44. Singapore: Springer Singapore, 2021.

7. Reddy, N. Sudhir, and V. Khanaa. "LDDC-Net: Deep Learning Convolutional Neural Network-based lung disease detection and classification." *Journal of Algebraic Statistics* 13, no. 1 (2022): 526-542.

8. Bhaskar, Nuthanakanti, and T. S. Ganashree. "Pulmonary Nodule Classification and Lung Cancer Stage Identification by Deep Learning Approach." *NeuroQuantology* 20, no. 11 (2022): 6297.

9. Rehman, Mujeeb Ur, Arslan Shafique, Yazeed Yasin Ghadi, Wadii Boulila, Sana Ullah Jan, Thippa Reddy Gadekallu, Maha Driss, and Jawad Ahmad. "A novel chaos-based privacy - preserving deep learning model for cancer diagnosis." *IEEE Transactions on Network Science and Engineering* 9,no.6 (2022):4322-4337.