Semiconductor Optoelectronics, *Vol. 41 No. 11 (2022), 538-545*
https://bdtgd.cn/

Semiconductor Optoelectronics ISSN: 1001-5868

PREDICTION OF LUNG CANCER WITH MODIFIED ADAPTIVE THRESHOLD IMAGE SEGMENTATION ALGORITHM BASED ON NEURAL NETWORKS

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. E-mail: n.sneha85@gmail.com*

ABSTRACT

The predicted outcomes of tumor diagnosis will form the foundation of a Computer Aided Diagnosis framework that allows for early detection of lung disease, 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 method proposed here to identify cancer is Modified adaptive threshold segmentation along with Support Vector Machines classifier as well as Artificial Neural Network classifier. Here, we conduct tests using the Lung Image Database Consortium datasets, which 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 suggested 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 [1]. 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 [2,3]. 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 [4]. 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 [5]. There are different public databases available for research purposes. Among these is the Lung Image Database Consortium (LIDC), the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), the Early Lung Cancer Action Program (ELCAP), and the Reference Image Database to Evaluate Therapy Response (RIDER) [6,7]. There are numerous sorts of arrangement calculations or regularly termed classifiers were utilized for Tumor finding. Artificial Neural Networks, Support Vector Machines, Genetic Algorithms, Fuzzy Sets, and Rough Sets are only a few of the methods that can be used [8,9]. 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 [5-8]. 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.

METHODOLOGY

Filtering, fragmentation, as well as feature extraction, are the three main phases of the recognition system. 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). Classifiers are used to improve the accuracy of detection. The proposed methodologies are implemented for cancer detection in MATLAB 2018a software.

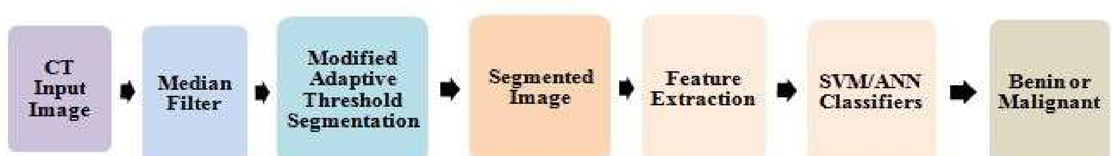

Figure 1: Schematic representation of the new method for identifying lung cancer

The suggested model details the current lung cancer diagnostic process is shown in Figure 1. Medical data are saved, exchanged, and sent utilizing Digital Imaging and Communications in Medicine (DICOM) standards [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, image noise reduction is a must. 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. Lung Segmentation is done using adaptive thresholding and feature extraction is utilized for extracting texture and shape features. Then classification of features is done using ANN and SVM. 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 ADTM, to remove the outer ring structure or parts of the outer ring structure obtained in some cases, in addition to the Adaptive Threshold method, we use a filtering criterion with an assumption that the object's Centroid of Interest will lie in the upper 80% of the image. 13 Shape Features such as 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 Contrast, Entropy Energy, Homogeneity, Cluster Prominence, Cluster Shade, and Dissimilarity, 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. Initially convert all the DICOM images in the LIDC database into JPEG format and obtain a 534-training databases and 150 testing databases containing both benign and malignant types of Lung cancer images. In preprocessing, median filtering is used as a result, edges are maintained whilst noise is suppressed. The preprocessed images are shown in figure 2.

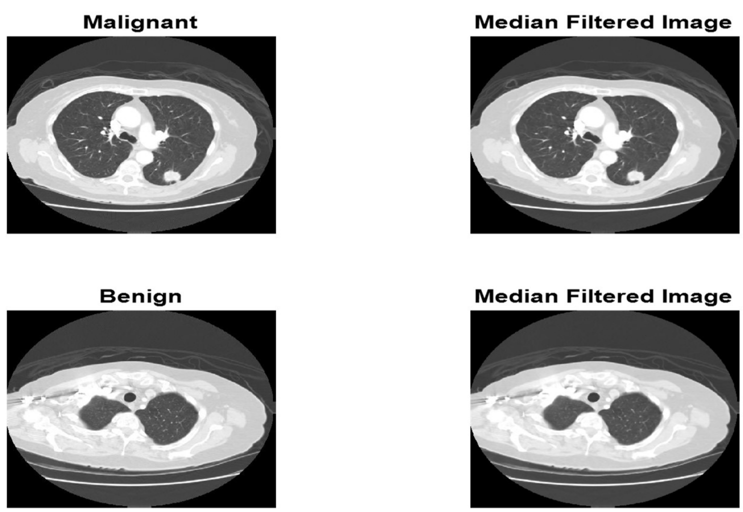

Figure 2. Median filtered image after preprocessing

The segmentation of lung images using adaptive thresholding involves the following steps. Initially, 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. The following figure 3 shows a segmented image using a thresholding approach. It shows important information in image segmentation. The thresholded image has the dual benefits of being easier to analyze and taking down less space for storage.

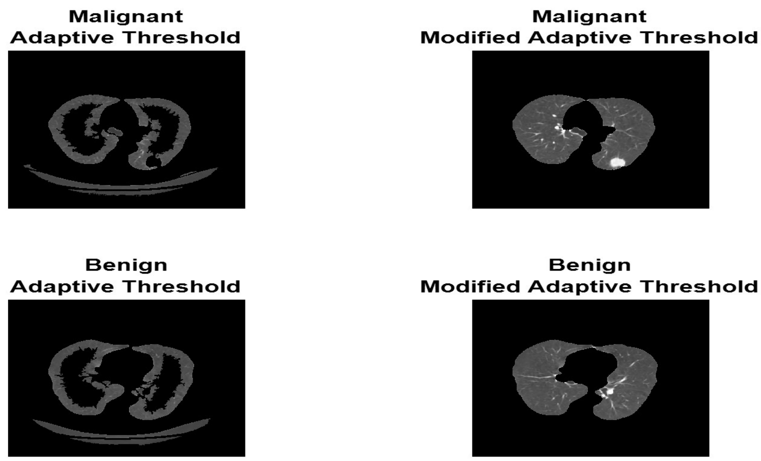

Figure 3. The output of Adaptive and Modified Adaptive Thresholds for Malignant and Benign images

Feature extraction is mainly used in pattern recognition and image processing. The feature is repeated patterns of the image. The binarization method can be used in identifying the presence of lung cancer and extracting the interesting region from that image. Detecting and isolating specific shapes as well as parts requires multiple phases, among them 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 seems 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. Morphological operations are then done on the above binary images which erode 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. Then obtain the area of the segmented region and marked it using a disk shape to get a segmented binary image.

DISCUSSION

From the segmented grayscale image, compute the features such as 13 shape features, 7 GLCM texture features, and 8 intensity features. Intensity features are extracted from the Principal Component Coefficients of Single-level discrete 2-D wavelet transform. These 13 shape features extracted are Area, Centroid, ConvexArea, Eccentricity, EquivDiameter, Perimeter, Euler Number, Range, Extrema, Minor as well as major Axis Length, Orientation, and Solidness.

Classification accuracy can be measured by the percentage of samples that are correctly labeled. Nine performance metrics, such as Error, F1 Score, Accuracy, Specificity, Precision, Sensitivity, False Positive Rate, Mathews Correlation Coefficient, and Kappa-Cohen's Kappa, are utilized to systematically assess the classification effectiveness of the proposed technique values of ADT and ADTM with ANN. The confusion matrix is done for multiple classes from the actual Class labels and the Predict Class Labels [55]. true positive (TP), false negative (FN), false-positive (FP), and TrueNegative(TN) values all exist within the Two-

Class of Confusion Matrix. Accuracy, errors, Sensitive (True positive rate/Recall), Selectivity, Precise, FPR-False positive percentage, F1 score, Matthews correlation coefficient, as well as Cohen's kappa is calculated to assess the results of ADT as well as ADTM using ANN, as demonstrated in Figure 4.

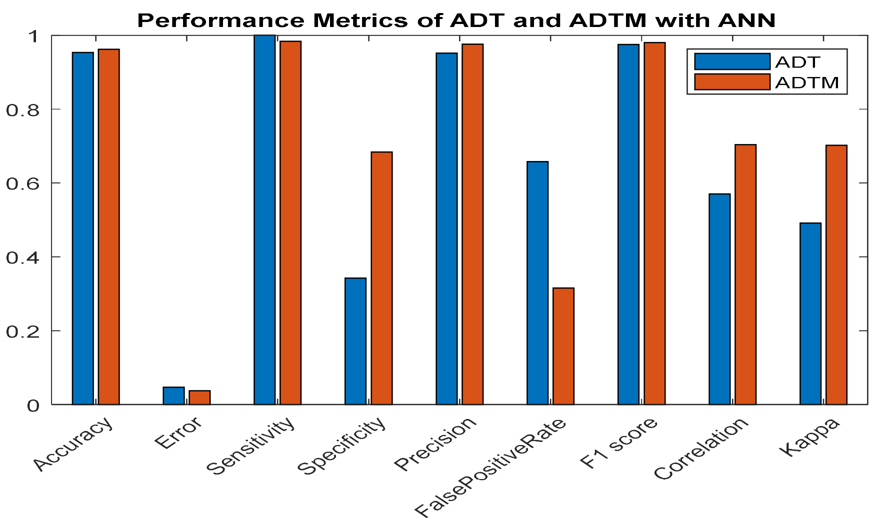

Figure 4. Performance evaluation of ADT and ADTM with ANN

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 correct classifications made by the trained network and the percentage of those classifications. 488 images out of 534 have been correctly labeled 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 the ANN classifier achieves a 96.3% success rate with a margin of error 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.

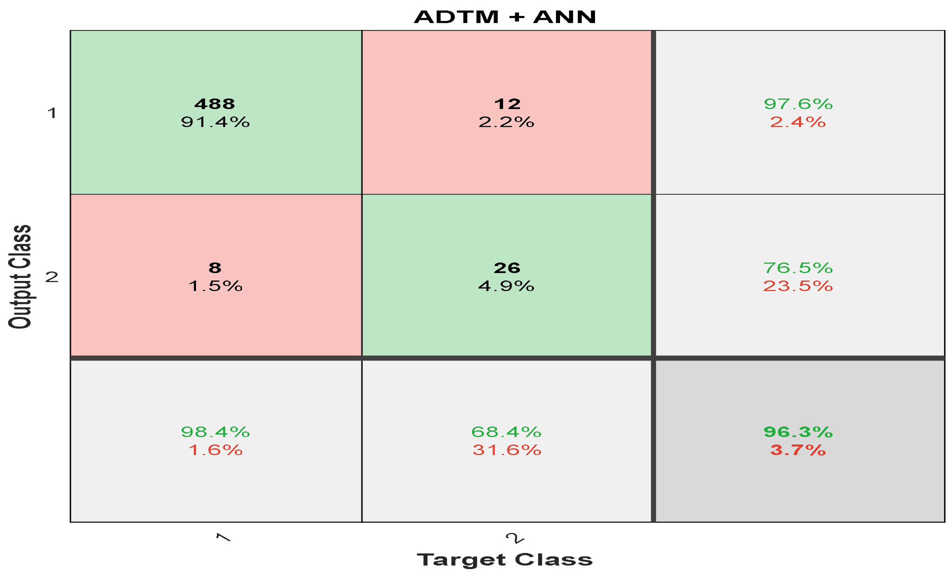

Figure 5: Confusion Matrix for ADTM with ANN classifier

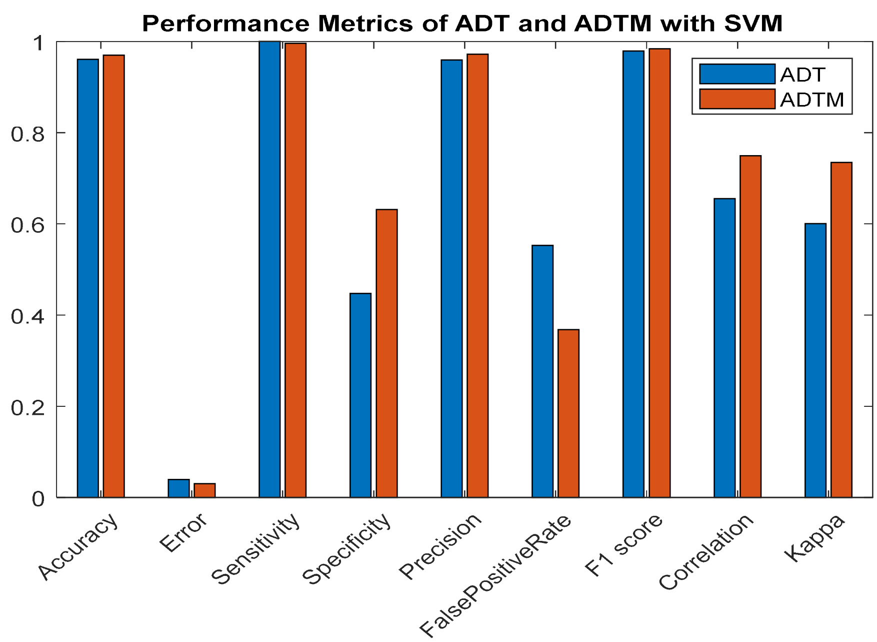

Figure 6. Performance evaluation with SVM classifier

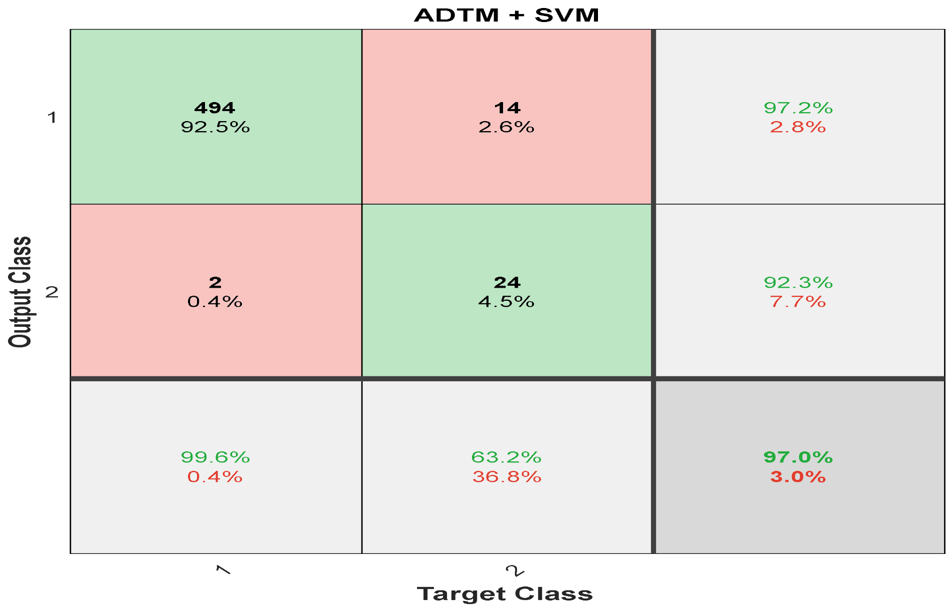

Figure 7. Confusion Matrix for ADTM with SVM

In Figure 7, the confusion matrix for ADTM with SVM with target class and output class is indicated. Out of 534 images 494 images belongs to 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 all 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 photos that were wrongly labeled as cancer (False negative). ADTM using 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 CT scan of the lungs has been segmented, a robust feature extraction method is used to pull out key information. 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 accuracy of the proposed method for lung cancer detection using the ANN classifier is 96.3% and the SVM classifier is 97%, which indicates the SVM classifier is giving high accuracy with ADTM for lung cancer detection.

REFERENCES

1. Surendar, P. (2021). Diagnosis of lung cancer using a hybrid deep neural network with adaptive sine cosine crow search algorithm. *Journal of Computational Science*, *53*, 101374.

2. Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). The optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems*, *92*, 374-382.

3. Hu, Q., Souza, L. F. D. F., Holanda, G. B., Alves, S. S., Silva, F. H. D. S., Han, T., & Reboucas Filho, P. P. (2020). An effective approach for CT lung segmentation using mask region-based convolutional neural networks. *Artificial intelligence in medicine*, *103*, 101792.

4. Lokhande, N. L., & Jaware, T. H. (2022). Lung CT Image Segmentation: A Convolutional Neural Network Approach. In *Information and Communication Technology for Competitive Strategies (ICTCS 2020)* (pp. 385-392). Springer, Singapore.

5. Huidrom, R., Chanu, Y. J., & Singh, K. M. (2022). Neuro-evolutional based computer-aided detection system on computed tomography for the early detection of lung cancer. *Multimedia Tools and Applications*, 1-13.

6. Ashwini, S. S., Kurian, M. Z., & Nagaraja, M. (2022). Lung Cancer Detection and Prediction Using Customized Selective Segmentation Technique with SVM Classifier. In *Emerging Research in Computing, Information, Communication, and Applications* (pp. 37-44). Springer, Singapore.

7. Reddy, N. S., & Khanna, V. (2022). LDDC-Net: Deep Learning Convolutional Neural Network-based lung disease detection and classification. *JOURNAL OF ALGEBRAIC STATISTICS*, *13*(1), 526-542.

8. Bhaskar, N., & Ganashree, T. S. (2022). Pulmonary Nodule Classification and Lung Cancer Stage Identification by Deep Learning Approach. *NEUROQUANTOLOGY*, *20*(11), 6297-6309.

9. Rehman, M. U., Shafique, A., Ghadi, Y. Y., Boulila, W., Jan, S. U., Gadekallu, T. R., ... & Ahmad, J. (2022). A Novel Chaos-Based Privacy-Preserving Deep Learning Model for Cancer Diagnosis. *IEEE Transactions on Network Science and Engineering*.