Triple SVM Integrated with Enhanced Random Region Segmentation for Classification of Lung Tumors

Sukruth Gowda M A
Dept. of Computer Science and Engineering
School of Engineering
Presidency University
Bengaluru, India

A Jayachandran
Dept. of Computer Science and Engineering
School of Engineering
Presidency University
Bengaluru, India

***Abstract*—The rapid growth of Computer vision and Machine Learning applications, especially in Health care systems, assures a secure, innovative lifestyle for society. The implication of these technologies in the early diagnosis of lung tumors helps in lung cancer detection and promises the survival rate of patients. The existing general diagnosis method of lung radiotherapy, i.e., Computed Tomography imaging (CT), doesn't spot exactly affected parts during injuries on lung malignancy. Herein, we propose a computer vision-based diagnostic method empowered with machine learning algorithms to detect lung tumors. The primary objective of the proposed method is to develop an efficient segmentation method to enhance the classification accuracy of lung tumors by implementing a Triple Support Vector Machine (SVM) for the classification of data samples into normal, malignant, or benign, Random Region Segmentation (RSS) for image segmentation and SIFT and GLCM algorithms are applied for feature extraction technique. The model is trained considering the dataset IQ - OTH or NCCD with 300 epochs, with an accuracy of 96.5% achieved under 200 cluster formations.**

Keywords*—*Benign; computed tomography; malignant; lung cancer; radiation; triple support vector machine

I. Introduction

According to WHO results nearly 10 million death occurred due to cancer. Among 10 million cancer types, about 2.21 million cases are of lung cancer. The most common cancer deaths of 2020 were of lung cancer as per WHO. The deadliest pandemic Covid19 has also affected the Lungs very badly. Hence under these circumstances, it is vital to predict and thereby prevent any onset of cancer in the human body. Early detection of cancer can considerably reduce the mortality rate[1]. In general, cancer is observed with the presence of growth of cells (nodules) in the lungs and if it can spread to the remaining cells of the body is called to be as malignant. The nodules are termed benign if the cells restrain from spreading yet are harmful to the specific organ alone. The development in technology can be utilized in a scalable manner to detect and diagnose the disease [2] [3] [4].

When cells in the lungs grow uncontrollably, lung cancer develops into a tumor and this can cause respiratory problems and spread to other sections of the body. Lung cancer is caused mostly by smoking and drinking alcohol [5]. Around 80% of the people who are observed with lung cancer are habituated to drinking and smoking. People who have the habit of smoking in past are the ones getting lung cancer mostly and only a few of them are never smoked. Lung cancer is mainly caused by the factors such as radiation, smoking from nearby people, living in a polluted environment and other related factors in cases of people who don't smoke [6] [7] [8]. In addition, there is a greater chance of the possibility of spreading into the lungs if any other place of the body is affected with cancer. As an overall result, the detection of cancer and predicting the probability of survival requires timely help in case of malignancy in the early stages. The chance of survival for the detection and progress of treatment at the early stages of cancer is higher when compared to severe stages (Borel 1997). Hence, the proposed research develops a system that detects cancer- free people, and the cancer portions which are benign and do not spread to other portions of the body. The proposed research also determines the malignancy in cancer regions which can spread to further parts of the body to diagnose early and provide appropriate treatment. For diagnosis of lung cancers, a variety of imaging has been used including sputum cytology, CT, chest x-ray, and magnetic resonance imaging (MRI) (http://colah.github.io/posts/). The tumors can be classified as benign or malignant types of cancer and the prognosis of the identified tumor is carried out by an expert like a doctor who can look over the individuals based on the causes. During detection, the number of false positives is reduced significantly which will have a severe impact on the patient's mental or financial well- being.

Manual detection of lung cancer is a very tedious task. To aid in detection, several machine learning approaches to aid in early diagnosis. The input to such approaches can be MRI, CT scan, and X-ray of the affected organ. These inputs are being fed to detect whether the organ is benign, malignant, or normal. Several types of research have already been conducted in the underlying area using different algorithms and techniques. A considerable amount of studies on image processing applications [1] [9] in lung cancer detection paves way for data processing easier for Machine learning classification. With the advent of image segmentation for separating cancerous nodules, the task of detecting the type of cancer and stages of cancer has become quite simpler. In region growth segmentation, watershed segmentation provides an effective result in CT image processing of lung cancer. The SVM and neural networks play a key role in the identification of the type of cancer in the lungs.

The crucial issues are developing a system for automatically identifying lung illness and accurately segmenting the precise region. Additionally, the heterogeneity of the tumor makes segmentation a difficult task [10]. Due to the low feature of photographs, it is difficult to notice impacted areas that are small in size [11]. The aforementioned problems serves as a motivation for effective lung tumor identification using RRS segmentation and supervised machine learning algorithm namely triple SVM.

The contributions of this research are summarized as follows:

1) The median and Weiner filters are used to preprocess the input images. The reason to choose median filter is that it preserves the sharp edges and Weiner filter removes the additive noise.
2) Subsequently, the precise segmentation is done by using RRS followed by SIFT and GLCM that are used to perform the feature extraction. The features from SIFT is invariant to light and viewpoints and GLCM is used to extract texture features.
3) Further, the triple SVM is used to perform an effective classification of lung tumor based on its comprehensible margin of dissociation.

The paper's organization of the paper is given as follows: Section II describes the image processing concept of region growth segmentation of lung tumors and later Section III has the methodology of a proposed model of segmentation. Section IV describes the experimental set up which includes the dataset description and model building. Section V concludes the final research output.

II. Literature Survey

A boosted deep CNN concept for lung tumor classification was introduced by Rani *et al.* [12] For segmentation purposes, the Advance Target Map Superpixel-based Region was suggested. The tumor region was then quantified using the nanoimaging theory. Image recognition was possible using the idea of boosted deep convolutional neural networks, which provides 97.3% accuracy. The current methodology demonstrated the stated efficacy which proved to greatly increase the execution of the recommended technique. Moreover, the data was handled quickly with deep learning.

Xie *et al.* [13] presented a method for epidermal growth factor receptor (EGFR) prediction with the lung cancer-based alterations. The system employs CNN with the layers counts upto 6 to learn the characteristics of the image deeper, and the process is followed by a Support Vector Machine (SVM) classifier for prediction. Yu's system was evaluated on two kinds of datasets, for which Dataset 1 attains an accuracy score of 76.16% and Dataset2 attains accuracy score of 67.55% respectively. Typical machine learning algorithms are performing inferior to the deep learning models by utilizing the learning obtained through the hierarchical data and deep-layered nature (chaudhary, & Sukhraj Singh 2012). Moreover, the data is being handled quickly with deep learning.

The approaches followed in Kareem *et al.* [14] preserve the fundamental structure of the data while using approaches similar to autoencoder and the approaches also uncover the analysis information via the deep learning approach. When it comes to picture categorization, CNN is the most preferred option and to categorize pictures efficiently, the process of convolution and the traditional process of the neural network is combined effectively in the design of the CNN model.

Lung cancer detection process using the deep learning algorithm has been examined for efficacy by the author da Silva *et al.*l. 2021 through the database named Lung Image Database Consortium (LIDC). The findings of the research demonstrated that the potential evaluation performance of models with deep learning possesses a better accuracy value of about 79.40%. The authors wanted to examine if the deeper layers network schemes might be used to diagnose lung cancer and if there were any more efficient techniques to reduce the downsample impact. Using the same dataset, they are evaluated in comparison with the techniques of Stacked Denoising Auto Encoder (SDAE) and the model of Deep Belief Networks (DBNs). On using DBN and SDAE, the models possess the accuracy value of 81.19% and 79.29%, respectively. The findings of the study showed that deeply layered network algorithms and automatic learning image features have a numerous amount of potential in the field of medical imaging.

A deep neural network model with the help of reinforcement learning has been developed for the early- stage detection of lung cancer by the authors of Ali et al. (The IQ-OTHNCCD lung cancer dataset Published: 19-10-2020, Version 1,DOI:10.17632/bhmdr45bh2.1 Contributor:hamdalla alyasriy.) And the performance has achieved the accuracy of 99.1% and 64.44% for training and testing respectively for the employed model on LIDC/IDRI subsets of data.

Another technique was suggested by (The IQ- OTHNCCD lung cancer dataset Published: 19-10- 2020,Version 1,DOI:10.17632/bhmdr45bh2.1 Contributor:hamdalla alyasriy), in which the authors built a CAD system based on a CNN network with multi- view and multi-scale nature for lung nodule categorization. On the datasets of ELCAP and LIDC/IDRI, the system obtained the values of 92.3% and 90.3% accuracy for detection, respectively. According to the aforementioned research work, CNN performance might be enhanced in many cases to improve accuracy performance and also aids in the detection process of early diagnosis of the lung cancer disease with minimal error [15].

Humayun *et al.* [16] presented the transfer learning (TL) with Convolutional Neural Network (CNN) to classify the lung disease. This work comprised three stages: Initially data augmentation was accomplished followed by pretrained CNN which was used to ensure the classification. Further, the localization was completed, once the classification was done. There are three TL methods such as VGG 16, VGG 19, and Xception that were used with fine-tuning hyperparameters to generate the network for training and testing process. The developed TL with CNN achieved higher accuracy in training and low accuracy in texting process.

Kareem *et al.* [14] developed the lung cancer detection using SVM classifier. In this work, the preprocessing techniques of bit plane slicing, Gaussian filtering, erosion technique and outlining operation were used to enhance the CT images. Subsequently, the Otsu's thresholding was used to separate the nodules followed by SVM that was used to perform the classification. This work mainly concentrated on the preprocessing approach for classifying the lung cancer.

The following are some of the issues with the most recent ways to classifying lung diseases: In classification, it is necessary for testing accuracy to exceed training accuracy [16]. Because training uses the same input data for analysis while testing uses a different input data for categorization. When segmentation is inadequate, classification performance suffers [14]. As a result, developing an efficient segmentation and classification is necessary to improve the classification of lung tumors.

III. Proposed Method

The primary objective of the work is to detect and classify lung cancer using image processing and machine learning techniques. In this research, the RRS based segmentation is proposed for separating the tumor portions from the input image. The SIFT and GLCM extracts optimal features from the segmented image. In that, the SIFT features are invariant light an viewpoints whereas GLCM extracts statistical texture descriptors. Further, the triple SVM is used to perform the classification based on extracted features. The methodology can be outlined as shown in Fig. 1.

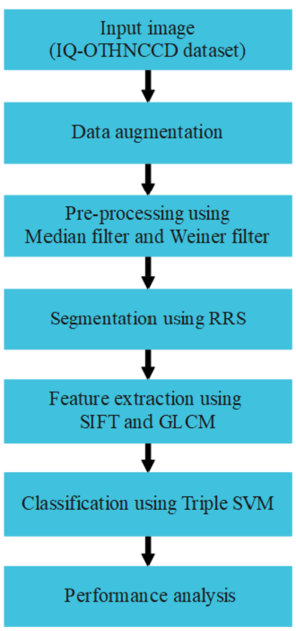

Fig. 1. Architecture of the Proposed Methodology.

Six stages of the proposed methodology have shown in Fig. 1. The procedure of the proposed method begins with collecting the input image data, then the image is augmented, which helps modify the input image to be ready for processing. Using ensembling filters like Weiner and median, the clarity of the vision is enhanced. RRS algorithm is applied to the resultant image, and after the segmentation, we adopt hybrid feature extraction techniques like SIFT and GLCM. The SVM is trained to classify the image into benign or malignant depending on the features. The obtained results are compared with existing techniques in the last phase. The detailed explanation is done in the following sections:

A. Dataset Description

The data for lung cancer has been collected at Iraq- Oncology Teaching Hospital or National Center for Cancer Diseases (IQ-OTH or NCCD) (The IQ- OTH-NCCD lung cancer dataset Published: 19-10-2020, Version 1,DOI:10.17632/bhmdr45bh2.1 Contributor:hamdalla alyasriy). The data collected constitutes three different stages of lung cancer CT scans of patients along with scans of healthy patients over 3 months. For the categories of benign, normal, and malignant groups of lung cancer, 110 patients are providing 1190 CT scans. By analyzing the images of lung cancer, the proposed model can be trained to detect cancer that can spread (malignant) or which cannot spread (benign) or it is a healthy tissue (normal) with the help of this dataset. Among the total data, 55 cases are normal, 40 cases are malignant and 15 cases are benign. The dataset is in DICOM format. The task is to classify the segmented lung nodules into benign, normal, or malignant [17].

B. Data Preprocessing and Morphological Features

The original CT scan images are preprocessed to remove noise to improve the image quality. Before performing the preprocessing, the data augmentation techniques such as cropping, flipping, rotation, translation, contrast, color augmentation are used to increase the number of inputs to meet the real-world requirements in classification. This is achieved by using an image processing technique called filtering. Two types of filters are used in this project to understand the underlying clarity. Fig. 2 and Fig. 3 represent the image after application of median filter and Weiner filter respectively. In the process of lung nodule segmentation, morphology [18][19] is one of the prominent techniques, especially when it comes to handling the cases of tumors that are attached to regions that are not on target for example paranchymal wall or the diaphragm (juxtrapleural) or vessels (juxtavascular). The representation of 0's and 1's as a matrix for a binary view of images on which 1's are called the neighbors.

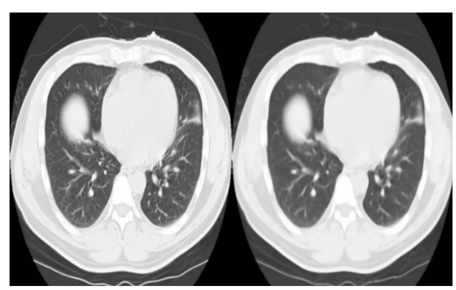

Fig. 2. Original Image and Median Filtered Image.

According to a comparison of the detected region with the corresponding pixel on the input image with respect to neighbor pixels of the output image as equals the value of every pixel. In the process of dilation for the given image pixels are added at the boundaries of the object, whereas it's the reverse process in erosion where in which the pixels are removed. The removed or added amount of pixels has been determining the shape and size of the element structure. Neighboring pixels

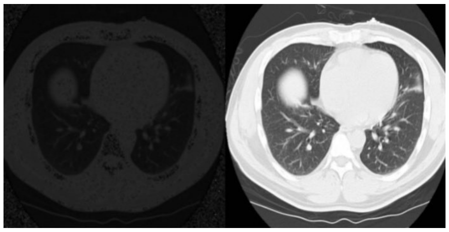

Fig. 3. Original Image and Weiner Filtered Image.

in the input image and the determined pixel through applying rules of morphological operations for the output image. Fig. 4 represents the identification of tumors by using morphological operations.

Cancerous nodules are recognized from the image segmentation in the upcoming stage of analysis and it is being depicted in Fig. 5 as left and right lungs. By using the proposed approach, the extraction of lungs from the images is done with the help of the region growing segmentation approach. The output of the nodule segmentation for one of the lung CT scans is shown in Fig. 6.

Fig. 5. Left and Right Lung Segmentation.

Fig. 6. Segmentation of Cancer Nodules.

Fig. 6 illustrate the segmentation output of the entire dataset. Fig. 7 shows the original CT scan images and Fig. 8 shows the segmented output using region growth algorithm.

C. Segmentation

Image segmentation [6] [2] [3] is the process to partition the image into multiple segments used to locate objects and boundaries. The region growth segmentation examines the neighboring pixels to determine whether it is to be added to the region or not. This is continued until no more regions can be added. The region is grown iteratively by considering all the neighboring pixels to the region.

The difference between the region means and pixel intensity is used as the similarity measure. A certain threshold value is set for the image. The region growth commences from the dark region of the image taken for segmentation. From Fig. 6 one can see that the cancerous nodules are segmented from the original image. Hence this segmentation helps in understanding the shape and size of the cancer nodules. Therefore, by using region growth segmentation, the tumor cells are identified. The following equations depict the operation of region growth segmentation. For every pixel in a region, the neighboring pixels are compared and if the difference is less than the threshold, then the neighboring pixel is added to the region.

Segmenting the image $\boldsymbol{R}$ into $\boldsymbol{n}$ non-empty subsets (R_1, R_2, R_3,......, R_n) in this segmentation. The following conditions given in equation (1) and (2) are required to be met during the segmentation.

- R_i, i = 1,2, . . . , n is digitally connected that is the regions has the contiguous lattice points.

$$U_{i=1}^{n} R_i = R ----> (1)$$

$$R_i \cap R_j = \phi, i \neq j ----> (2)$$

Where, ϕ defines the null set. Equation (1) defines the segmentation is complete i.e., each pixel must be in the desired region whereas the equation (2) states that the points in a region is connected in the predefined sense. The RRS gives the capability to segment an image into any number of objects, where is the user-specified arbitrary number of objects. Also treats the image as a graph, with the pixels acting as its nodes or vertices. Hence, the proposed RRS is used to segment lung tumor from the input image.

The main goal is to calculate the similarity of images in different regions

D. Feature Extraction

After performing the RRS based segmentation, the segmented lung tumor *(SI)* is given as input to the feature extraction. Image feature extraction is the most important stage in any classification problem.

SIFT determines the local features of an image. These are often termed as key points and constitute of scale and rotation

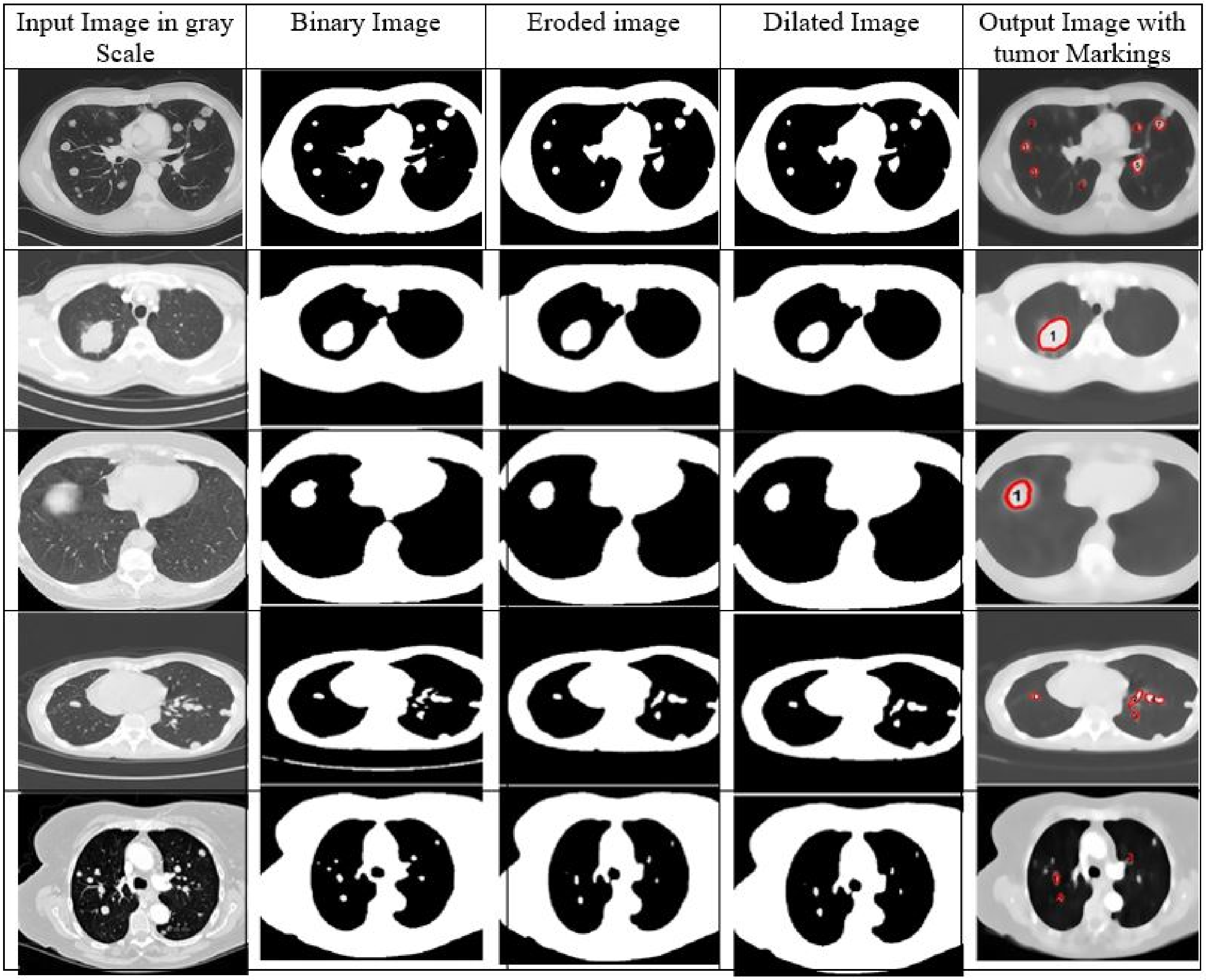

Fig. 4. (a) Represents the Input Image in Gray Scale Format. (b) Represents the Binary Image of the given Input Image. (c) Represents the Image after Applying Erosion Operation. (d) Represents the Image after Applying Dilation Operation. (e) Represents the Output Image along with the Tumor Markings.

invariant. The advantage of using SIFT is that it is independent of the size and orientation of an image. Gaussian blur is used to compute the scale of an image. The idea is to find out the local maxima and local minima for the images and then remove the low contrast key points. Finding the keypoint descriptor is the final stage for SIFT and it computes the orientation and magnitude of a descriptor which are expressed in equation (3) and (4), respectively. According to the local image gradient directions, the keypoints are allocated orientations.

$$m(x,y) = \sqrt{(Diff_x)^2 + (Diff_y)^2} - - - - > (3)$$

$$\theta(x,y) = \tan^{-1} \frac{Diff_x}{Diff_y} - - - - > (4)$$

where, x and y are the dimensions of segmented lung tumor image (*SI*) ; $Diff_{\mathrm{x}}$ and $Diff_{\mathrm{y}}$ are pixel differences; The calculations of magnitude and direction are performed for an each pixel in an adjacent region around the keypoint at the Gaussian-blurred image.

GLCM (Grey Level Co-occurrence Matrix (Mohanaiah et al. 2013) determines the spatial relationship of image pixels. This calculates how often the pixel pairs and spatial relationships occur in an image. The statistical measures are obtained from the GLCM matrix. The most common measures obtained are contrasted (local variation in the matrix), correlation (joint probability of the pixel pairs), energy (sum of squared elements), homogeneity (closeness of the distribution) and entropy. Hence, this GLCM provides the statistical texture descriptors for the segmented lung tumor image *(SI)*. Further, the features from both the SIFT and GLCM are concatenated together as shown in equation (5).

$$F_{ea} = \{m, \theta, GLCM\} - - - - > (5)$$

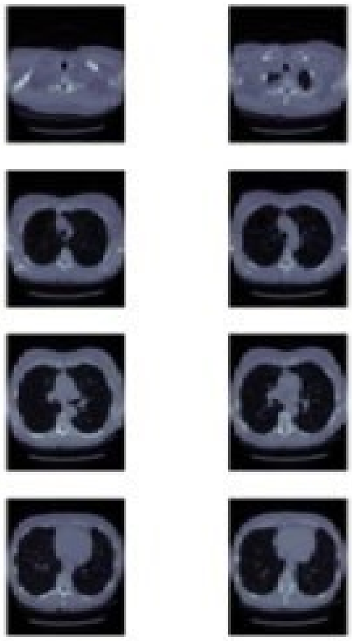

Fig. 7. Original CT Images.

The concatenated features are given to the Triple SVM to perform the classification.

E. Triple Support Vector Machine Classification

Initially, the features from the SIFT and GLCM are taken as input to the triple SVM to perform the classification. SVM is a supervised machine learning algorithm that uses a hyperplane to classify the data. The task is to find out a best-fitting hyperplane that segregates data based on the given input features and the process is termed a marginal classifier. The SVM has proved to achieve good performance in the data which has no prior knowledge from the related study. The basic idea is to map the input data (i.e., concatenated features of SIFT and GLCM (FFFFFF)) onto a higher dimensional feature space and determine the separate margin between the classes in the feature space. SVM is built using a kernel function that has the best fitting hyperplane [7] [20]. In the current work, Triple SVM is used for multi-class classification, which helps in preventing the loss of information where class data is considered against each of the other classes. Further, Triple SVM achieves better performances in the IQ-OTH or NCCD dataset, because it minimizes the inter-class and maximize the intra-class. The linear classifier is depends on the linear discriminant function

$$f(F_{ea})$$

which is expressed in equation (6).

$$f(F_{ea}) = w^T F_{ea} + b - - - - > (6)$$

Where, w is the weight vector and b is the bias value.

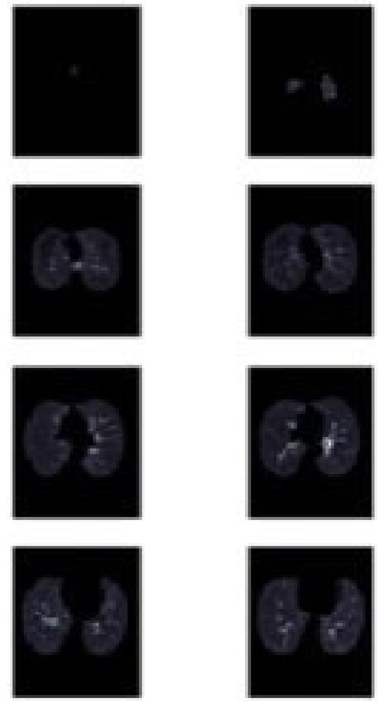

Fig. 8. Segmented Output.

Building the model involves classifying the images into normal, benign, or malignant cases. The classification algorithm named SVM has been employed as the technique to be carried for the current phase [14] [21] [22]. The dataset is split into train and test sets. The classification is achieved by optimal separating hyperplane. The data after feature extraction is fed to the SVM kernel function as input. The model is built using a polynomial SVM kernel.

IV. Results and Discussion

The proposed system is implemented using Google Co-lab software (http://colah.github.io/posts/). The model is trained for 100 iterations and achieved an accuracy of 96.5%. The f1-score for normal, Benign and Malignant cases are 0.92, 1 and 0.93, respectively. It is likely to achieve validation accuracy of over 99% when the number of iterations is increased by 300 epochs. Table I and Fig. 9 shows the classification report and accuracy plot of triple-SVM, respectively.

TABLE I. Classification Report

	Precision	Recall	f1-Score	Support
Normal Cases	1.00	0.86	0.92	7
Benign Cases	1.00	1.00	1.00	7
Malignant Cases	0.88	1.00	0.93	7
Accuracy			0.965	21
Macro Avg	0.96	0.95	0.95	21
Weighted Avg	0.96	0.95	0.95	21

Algorithm 1 Region Random Segmentation Algorithm

Step 1: The lung CT is uploaded as the input image.
Step 2: The coordinate of the lung(left and right) is specified.
Step 3: Contrast enhancement is done so that the lung region can be easily extracted.
Step 4: The threshold value is considered as a 20% gray threshold of the entire image.
Step 5: The seed value to be selected for both lungs and a point value which is the neighboring pixel is considered.
Step 6: The point value is verified against seed value and threshold value and repeatedly moved across the pixel till it reaches edges.
Step 7: A graph represented with edges and vertices is extracted from the resultant lung image on which edges are the relationship between pixels and vertices are the image pixels.
Step 8: To do nodule localization and for the lung region representation two seed points are user defined.
Step 9: Differences in intensity between the pixels which are unlabelled are being computed by utilizing the euclidean method distance as the seed point.
Step 10: Gaussian weight computation and normalization of resultant value.
Step 11: Gaussian weighting function used by seeded pixels of lung region and the reached nodule by the probability of walking across each pixel by computation two probabilities with the help of 2 predefined pixels.
Step 12: The maximum probability (pn>pl) pixel will be selected as nodule pixel and label that pixel from the vector of probabilities.
Step 13: Nodule representation as the region is separately considered as nodule label pixels to segment.

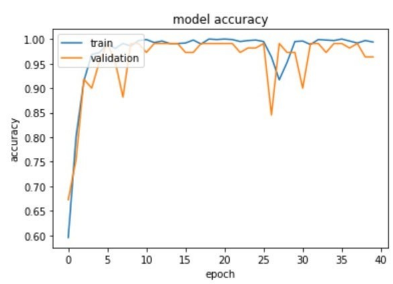

Fig. 9. Accuracy Plot of Triple-SVM Model (300 Epochs).

For classification, 80% of the data is used for training, while the remaining 20% of the data is used for testing. The correlation matrix for the classification accuracy is shown in below Fig. 10.

Table II displays the study of the proposed RRS and triple SVM model's results in relation to the suggested segmentation methodology, variation with various feature extraction methods, and region-based segmentation. Utilizing triple SVMs, the suggested RRS together with selected features from SIFT and GLCM provided higher accuracy of 96.5%. However, when the background has a comparable texture, the segmentation using standard region growth is compromised. Next, little modifications to the input data have an impact on the classification performed using decision trees.

Fig. 10. Correlation Matrix.

TABLE II. RESULT ANALYSIS

Methods	Result Obtained
Particle Swarm Optimization, Genetic Algorithm, SVM	$Accuracy$: 89.50%
K-NN Classification using Genetic Algorithm	$Accuracy$: 90%
Artificial Neural Network Approach	$Accuracy$: 90%
Region Growing Approach with GLCM and Decision Trees	$Accuracy$: 69% $Specificity$: 68% $Sensitivity$: 70%
Region Growing Approach with GLCM and Triple SVM classifier	$Accuracy$: 89% $Specificity$: 90% $Sensitivity$: 93%
Region Growing Approach with SIFT, GLCM and Decision Trees	$Accuracy$: 89% $Specificity$: 85% $Sensitivity$: 84%
Region Growing Approach with SIFT, GLCM and Triple SVM classifier	$Accuracy$: 93% $Specificity$: 91% $Sensitivity$: 87%
Proposed Segmentation RR algorithm with SIFT, GLCM and Triple SVM classifier	$Accuracy$: 96.5% $Specificity$: 94% $Sensitivity$: 89%

The comparative analysis of RRS with Triple SVM with existing researches such as TL-VGG 16 (Humayun *et al.* [16]), TL-VGG 19, TL- Xception and SVM (Kareem *et al.* [14]) are shown in the Table III. From the Table III, it is concluded that the RRS with Triple SVM achieves better results than the TL-VGG 16 (Humayun et al. 2022), TL- VGG 19, TL-Xception and SVM. But, the proposed RRS is used to perform precise segmentation of lung tumor portions which leads to improve the classification using triple SVM. The capacity of handling higher dimensional spaces of triple SVM is used to increase the classification accuracy.

TABLE III. COMPARATIVE ANALYSIS

Methods	Accuracy
TL-VGG 16[16]	83.39 %
TL-VGG 19 [16]	80.97 %
TL-Xception[16]	89.68 %
SVM [14]	89.88 %
RRS with Triple SVM	96.5 %

Table III shows the comparative analysis of various methods. RRS with Triple SVM has been compared to previous researches such as, TL-VGG 16, TL-VGG 19 and TL-Xception [16], and SVM [14]. This comparison is given in Table III. According to Table III, the RRS with Triple SVM outperforms the existing TL-VGG 16, TL-VGG 19, TL- Xception and SVM in terms of accuracy (96.5%). Also, the exact segmentation of lung tumor sections were carried out using the proposed RRS which enhances the classification using triple SVM. Triple SVM's ability to handle greater dimensional spaces is employed to improve classification accuracy.

From the overall analysis, the classification is improved by the precise segmentation of lung tumor sections which is carried out utilizing the proposed RRS. To increase classification accuracy, Triple SVM is used because of its capacity for handling higher dimensional spaces. The recommended RRS and a few SIFT and GLCM features combined with triple SVMs produced a greater accuracy of 96.5%. Further, the proposed RRS with Triple SVM was evaluated with existing models such as TL-VGG 16, TL-VGG 19, TL-Xception and SVM in terms of accuracy. From that study, it clearly hows that proposed RRS with Triple SVM has achieved higher accuracy of 96.5%.

V. Conclusion

A lung cancer detection and classification model is built using an effective segmentation and triple SVM classifier. The CT scan images are preprocessed and segmented to obtain the cancerous nodules in the lungs. The features are extracted using SIFT and GLCM. The input is then fed to triple SVM classifier with SIFT and GLCM extracted features. The classification accuracy of lung tumor is enhanced based on precise segmentation achieved by RRS. The triple SVM is trained for 300 epochs initially and achieved an accuracy of 96.5% with 200 clusters. If the cluster size is increased to 500, then the model achieves more accuracy. The technique can be enhanced by using multiple algorithm comparisons as well as incorporating deep learning algorithms into this dataset. The proposed methodology can also be implemented using different lung cancer datasets to study the efficiency of the proposed methodology. The prediction accuracy of the proposed model can be varied according to the number of epochs, clusters, and varying preprocessing feature extraction used in building the SVM model. However, the proposed RRS with triple SVM doesn't predict the size of affected portion and classify the supplementary abnormality of tumor. Therefore, the future work involves identifying the volume of the lung and classifying it with any other abnormality and also identifying the thoracic tumor.

References

[1] N. S. Nadkarni and S. Borkar, "Detection of lung cancer in ct images using image processing," pp. 863–866, 2019.

[2] R. S. Kabade and M. Gaikwad, "Segmentation of brain tumour and its area calculation in brain mr images using k-mean clustering and fuzzy c-mean algorithm," *Int. J. Comput. Sci. Eng. Technol*, vol. 4, no. 05, pp. 524–531, 2013.

[3] B. A. Skourt, A. El Hassani, and A. Majda, "Lung ct image segmentation using deep neural networks," *Procedia Computer Science*, vol. 127, pp. 109–113, 2018.

[4] B. Pellini and A. A. Chaudhuri, "Circulating tumor dna minimal residual disease detection of non–small-cell lung cancer treated with curative intent," *Journal of Clinical Oncology*, vol. 40, no. 6, pp. 567–575, 2022.

[5] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using glcm approach," *International journal of scientific and research publications*, vol. 3, no. 5, pp. 1–5, 2013.

[6] U. Kamal, A. M. Rafi, R. Hoque, J. Wu, M. Hasan *et al.*, "Lung cancer tumor region segmentation using recurrent 3d-denseunet," pp. 36–47, 2020.

[7] P. Sangamithraa and S. Govindaraju, "Lung tumour detection and classification using ek-mean clustering," pp. 2201–2206, 2016.

[8] M. P. Arakeri *et al.*, "Real time multi-object tracking based on faster rcnn and improved deep appearance metric," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 12, 2021.

[9] P. R. Katre and A. Thakare, "Detection of lung cancer stages using image processing and data classification techniques," pp. 402–404, 2017.

[10] S. P. Sankar and D. E. George, "Regression neural network segmentation approach with lidc-idri for lung lesion," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 5571–5580, 2021.

[11] P. Nanglia, S. Kumar, A. N. Mahajan, P. Singh, and D. Rathee, "A hybrid algorithm for lung cancer classification using svm and neural networks," *ICT Express*, vol. 7, no. 3, pp. 335–341, 2021.

[12] K. V. Rani and S. J. Jawhar, "Superpixel with nanoscale imaging and boosted deep convolutional neural network concept for lung tumor classification," *International Journal of Imaging Systems and Technology*, vol. 30, no. 4, pp. 899–915, 2020.

[13] Y. Xie, W.-Y. Meng, R.-Z. Li, Y.-W. Wang, X. Qian, C. Chan, Z.-F. Yu, X.-X. Fan, H.-D. Pan, C. Xie *et al.*, "Early lung cancer diagnostic biomarker discovery by machine learning methods," *Translational oncology*, vol. 14, no. 1, p. 100907, 2021.

[14] H. F. Kareem, M. S. Al-Huseiny, F. Y. Mohsen, and K. Al-Yasriy, "Evaluation of svm performance in the detection of lung cancer in marked ct scan dataset," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3, p. 1731, 2021.

[15] J. Yang, B. Wu, L. Li, P. Cao, and O. Zaiane, "Msds-unet: A multi-scale deeply supervised 3d u-net for automatic segmentation of lung tumor in ct," *Computerized Medical Imaging and Graphics*, vol. 92, p. 101957, 2021.

[16] M. Humayun, R. Sujatha, S. N. Almuayqil, and N. Jhanjhi, "A transfer learning approach with a convolutional neural network for the classification of lung carcinoma," vol. 10, no. 6, p. 1058, 2022.

[17] H. Alyasriy, "The iq-othnccd lung cancer dataset," 2020.

[18] A. Chaudhary and S. S. Singh, "Lung cancer detection on ct images by using image processing," pp. 142–146, 2012.

[19] S. HARIPRASAD, H. Gururaj *et al.*, "Artifact removal techniques for lung ct images in lung cancer detection," *International Journal of Data Informatics and Intelligent Computing*, vol. 1, no. 1, pp. 21–29, 2022.

[20] H. Hu, Q. Li, Y. Zhao, and Y. Zhang, "Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2880–2889, 2020.

[21] K. Sun and D. Yang, "Human health big data evaluation based on fpga processor and big data decision algorithm," *Microprocessors and Microsystems*, vol. 81, p. 103793, 2021.

[22] A. Naik and D. R. Edla, "Lung tumor classification using cnn-and glcm-based features," in *ICT Systems and Sustainability*. Springer, 2021, pp. 157–163.