LUNG CANCER DETECTION USING FUZZY C-MEAN ALGORITHM-BASED ARTIFICIAL NEURAL NETWORK

Affrose[1], Cheruku Sandesh Kumar[2]
[1]Research Scholar, Amity University, Rajasthan, Jaipur India
[2] Amity University, Rajasthan, Jaipur, India

Abstract— Lung cancer is the most dangerous disease. The chest X-ray (CXR) is the most widely used and crucial imaging method for detecting lung cancer. The CXR is chosen due to its accessibility, affordability, non-invasiveness, and ease of acquisition. Machine Learning (ML) is used to automate lung cancer identification. ML in medical imaging has the potential to reduce medical professionals' burden throughout the diagnostics and screening process. Lung segmentation is an important step in lung analysis. Several factors that make lung segmentation difficult are: 1) The size and structure of the lungs change as a result of gender, age, and heart size. 2) Opacity caused by severe pulmonary disease with a high-intensity value. 3) The patient's garments or medical gadgets obscure the entire visibility of the enclosed foreign entity, like the lung field. This study tries to address all of these issues. Graph cut and Fuzzy C-means (FCM) are two segmentation algorithms employed. The segmentation result from both techniques is fed into an ML model for classification followed by feature extraction. The ML model attempts to identify lung images as "normal" or "cancer." The outcome of the ML model followed by both segmented techniques is compared. The comparison results show that the FCM delivers more exact results on screening lung cancer.
Keywords— Lung, Segmentation, Fuzzy, Graph Cut, Support Vector Machine.

Introduction

When examining images for purposes such as object recognition, machine vision, and medical imaging, among others, image segmentation is a crucial first step because of its scope and difficulty. The purpose of an image segment is to separate an image into numerous distinct regions that all share the same characteristics, such as brightness, contrast, saturation, and surface texture. Several methods of partitioning have been described, and citations [1-3] provide in-depth analyses of each. Specifically, edge detection, grouping, boundary identification, and thresholding are the four types of image segmentation procedures discussed in the cited article [1]. The paper deals with lung cancer segmentation and classification. The lung cancer data set was collected from 14 distinct hospitals and it contains 247 CXR. As can be seen in Fig. 1, the collected Gray images all have a resolution of 2048 pixels by 2048 pixels. Among 247 CXRs taken, 93 were considered normal and 154 were diagnosed with cancer.

Submitted: 13/01/2023 **Accepted**: 30/01/2023

Fig. 1. Sample CXRdata

According to their literature review, their SCR (Segmentation in Chest Radiographs) collection includes human-created lung field masks for each CXR in the JSRT database. Figure 2 depicts an abnormal CXR.

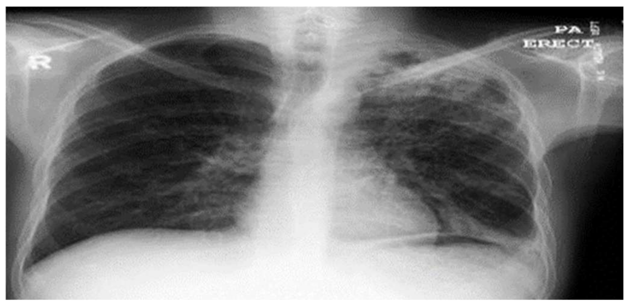

Fig. 2. Abnormal lung image

The previous work on this domain is detailed here. The study [4] proposes a method for recognizing and segmenting lung nodules using a fully convolutional network (FCN), the level set method, and other image processing techniques. To begin, CT scans of the lungs are forwarded to the FCN for segmentation. Second, to detect lung nodules within the lung region, the threshold approach and other image-processing techniques are applied. Then, the level set technique and the threshold strategy, both of which are based on a coordinate system transformation, are used to divide the found lung nodules and speculum. The experimental results demonstrate the efficacy of the suggested method in detecting and segmenting lung nodules, with a 100% accuracy rate and 0.9 dice value in segmentation. Therefore, this methodology can be utilized to better support the clinical detection of lung cancer.

To effectively segment an image, the study [5] suggests a Generalized Spatial Fuzzy C-Means (GSFCM) method, which takes into account both the provided pixel and spatial information. This greatly enhances the efficiency of the segmentation.As proved experimentally with a set of MR images, the new GSFCM method beats existing FCM algorithms in a range of cluster validity functions.

The technique provided in the article [6] uses a dynamic and intelligent clustering method called Firefly Search with FCM to automate the process of nodule segmentation. This technique uses the capability of firefly search to discover the best feasible starting points for the FCM's clustering algorithms to improve segmentation results. The performance of cuttingedge algorithms is compared to that of the proposed system. The proposed approach is tested using live-action video.

A computer-aided design (CAD) technique is utilized in the study [7] to automatically separate the lungs from CT scans. The CT data revealed nodules related to the chest wall, and used level-set modeling to isolate the regions of the lungs containing these nodules. This procedure is divided into three stages: The CT scans are binarized initially with an adaptive fuzzy thresholding technique, then the lung with non-isolated nodules is segmented via level set modeling and a convex hull approach. Finally, in step three, the lung is segmented using lobe-specific shape traits. Based on testing results, the solution surpasses currently available methods by 98%.

The goal of the research [8] is to create an automated method for dividing the lung into several regions in the axial plane of the body. Recognizing widespread lung diseases and accurately describing and diagnosing specific anomalies both rely on accurate lung segmentation. In this research, they evaluate the strengths and weaknesses of the two methods side by side. To classify a given lung into its proper area, the method of the anatomic structure utilizes anatomical marker recognition to establish the divisions between the areas, while the ML strategy depends on criteria such as lung shape, volume, and position.

Procedure

We present an overview of the lung segmentation strategies that have been implemented, as well as a comparison of these techniques, feature extraction, and classification, in this part. The architecture of our system is represented in Fig. 3 with the various processing phases, while the following sections go into additional detail on those issues. This system compares the results of Fuzzy-based segmentation to the results of segmenting the input lung CXR using a graph cut optimization technique. It evaluates a set of features using a segmented lung model and a pre-trained binary classifier. Finally, based on, for example, a set of decision rules and thresholds, the classifier's output identifies the CXR input as a TB-positive event.

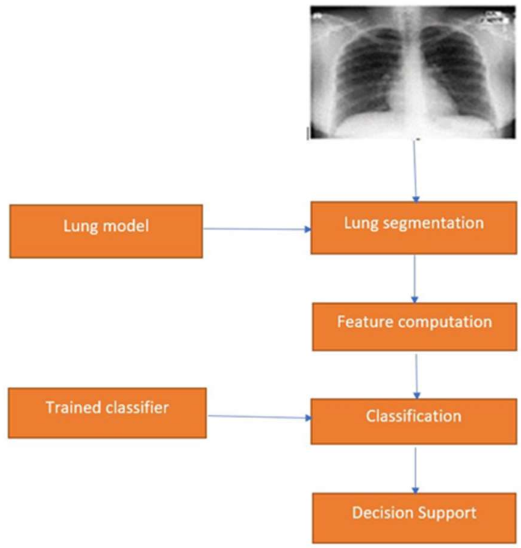

Fig. 3. Review of the system

Segmentation Technique

Graph cut segmentation:

Image segmentation has shown tremendous advancement. With just a few straightforward gathering indicators, people can make a substantial difference in an extreme arrangement of images. A major reason for this evolution is the employment of a diagrammatic method [9]. The schematic graph depicts a fair and adaptable evolution of the approach for dividing images. It includes a simple language for recording simple neighborhood division hints as well as a plethora of powerful computational tools for deriving global division from these local (pairwise) pixel similarities. Some cut processes that have been computationally plotted can be highly convincing. The graph cuts algorithm model leverages an undirected graph G = (V,E)to address computer vision difficulties. The pixel properties, such as brightness, are represented by a collection of vertices V, with edges E connecting them. Many computer vision issues, such as image smoothing and stereo correspondence, can be defined in terms of energy minimization, which graph cuts can easily solve. The formulation that offers the highest a posteriori confidence in a solution is equivalent to the least energy solution for most computer vision problems of this sort. It is necessary to slice a graph. Graph cuts are models that maximize flow while minimizing cuts.Following is a mathematical outline of the possible optimization issue:

Let

$$f = \{f_1, f_2, \dots, f_p, \dots f_N\}$$

be a binary vector whose elements represent the label assignments for pixel pϵP (where P - pixels collection in the CXR and N – pixels count) in the lung region and background, respectively. We use a minimization technique to get the best possible setting for f, which we call the objective function.

$$E(f) = E_d(f) + E_s(f) + E_m(f)$$

where E_d, E_sand E_m stand for the CXR's region, boundary, and lung model characteristics. Lung border pixels p and q are restricted by the following limitations on their boundaries:

$$E_x(f) = \sum_{(p,q)\epsilon O} \exp\left(-(I_p - I_q)^2\right)$$

This phrase incorporates the exponential intensity fluctuations across pixels used to characterize the cut. When the differences in intensity are largest, just a little quantity is present. A standard lung model is a 2-D array that accounts for the likelihood that a given pixel, p, is a part of the lung field. Figure 4 gives the outcome of lung segmentation by the graph cut method.

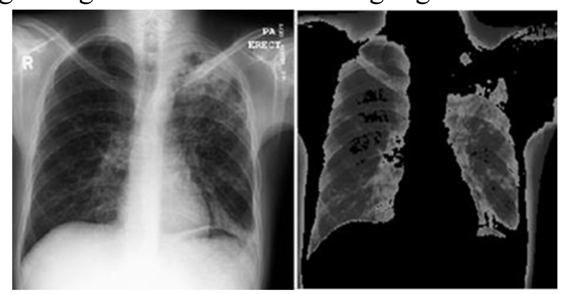

Fig. 4. Outcome of lung segmentationby Graph cut method

Fuzzy C-Means segmentation:

Clustering is a technique for categorizing data into comparable groups based on the relationships between individual bits of information. The feature vectors are divided into N groups using a clustering approach. Cn is the location of the core of each succeeding cluster. Fuzzy detection and pattern recognition are two applications of Fuzzy Clustering. FCM is the most often used fuzzy clustering algorithm, however, there are others. To assign uncertain values, fuzzy cost modeling uses reciprocal distance. The algorithm takes a set number of clusters, N, as input. We determine where people in a cluster fall on average. As a result, an object class is divided into N different clusters. Under all weighting settings, the FCM cluster strives to minimize the Mean Squared Error (MSE). Each feature vector can be allocated a wide range of fuzzy membership values using the FCM. The appropriate weight of the feature vector across all clusters determines the final segmentation. For good reason, the FCM algorithm is the most extensively used and highly acknowledged method of image segmentation: it has superior uncertainty characteristics and preserves significantly more information than hard segmentation approaches [10,11]. Figure 5 depicts the fuzzy method for segmenting the lungs.

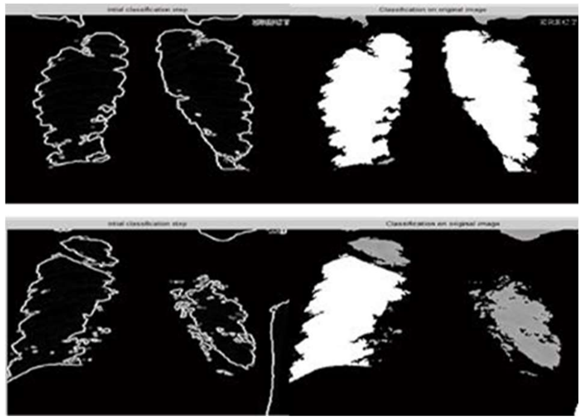

Fig. 5. Fuzzy-based segmentation (Normal &CancerImage)

Machine Learning Model

In this post, we will go over the principles of the Support Vector Machine, a ML model used for categorization (SVM). Since Vapnik's initial suggestion of the SVM [12], the ML research community has paid close attention. SVMs can surpass competing for data classification algorithms in terms of classification accuracy, according to current research. Simulations have helped with data classification to solve real-world difficulties. Extensive testing and comparisons with other supervised learning systems have shown that Sims is the winner. However, for some datasets, the effectiveness of SVM is significantly reliant on the cost parameter and kernel parameters. This usually necessitates a lot of cross-validation on the side of the user before they can figure out which values produce the greatest outcomes. The phrase "model selection" is commonly used to describe this process. Because of the time investment required, model selection is a practical concern. We experimented with a few SVM algorithm options that could affect the output. Such factors include the number of training instances, the standard deviation of the Gaussian kernel, the weights associated with slack variables to account for the unequal distribution of labels, and the choice of kernel functions.

SVMs, or supervised learning models, are a class of related algorithms for doing these tasks. This family is connected to linear classes in general. SVM offers the unique ability to maximize the geometric margin while minimizing the empirical classification error at the same time. As a result, Maximum Margin Classifiers, or SVMs, are used. SVM is based on the Structural Risk Minimization (SRM) hypothesis. Support vector machines project the input vector onto a higher-dimensional space to generate a maximal separation hyperplane. A hyperplane divides the data in half, and two parallel hyperplanes are formed on either side of the dividing line. The separating hyperplane is defined by maximizing the distance between two parallel hyperplanes. The classifier's generalization error is thought to decrease as the distance between the parallel hyperplanes increases. In this case, we consider data in the form of

$$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \ldots\ldots (x_n, y_n)\}$$

Training data is divided (or separated) using a hyperplane, and provides

$$w.x + b = 0$$

The scalar b is multiplied by the p-dimensional vector w in this formula. The vector w is perpendicular to the dividing hyperplane. The offset option b can be used to increase the margin. When b is not present, the hyperplane is compelled to pass through the origin, limiting the search space. SVM and parallel hyperplanes both attract our curiosity because we are interested in maximization. A pair of parallel hyperplanes can be described using an equation.

$$w.x + b = 1$$

$$w.x + b = -1$$

If the training set can be divided into linearly separable subsets, we pick such hyperplanes that do not have points among them and then maximize their distance. The distance to the hyperplane is calculated as 2/(|w|). As a result, lowering |w| is a goal.

Results and Discussion

Coding tasks related to segmentation, feature extraction, and classification are performed in MATLAB. Results from the developed method can distinguish between healthy and cancerous CXRs that have been uploaded. In this research, we suggest two methods, graph cut and fuzzy, for the automatic segmentation of lungs from CXR. The findings of segmentation help with the quantitative analysis of lung characteristics because segmentation is necessary to calculate lung parameter. Classified images are compared based on the segmentation results and feature extraction performed using each technique. Table I displays a comparison of the SVM output obtained by applying both segmentation outputs.

The accuracy, precision, and recall are compared to identify the best technique for segmentation. The SVM result followed by the graph cut method gives an accuracy of 93.33%, precision of 91.39%, and recall of 95.17%. Similarly, the FCM gives the result of 96%-accuracy, 97.32%-precision, and 94.77%-recall. Its shows that the FCM method increases the accuracy by2.67% when compared withthe Graph cut technique. Table 1 is converted to a bar chart and it will display in figure 6.

TableI: Segmentation Technique Evaluation

Segmentation	Graph cut	Fuzzy C-means
ML Model	SVM	SVM
Accuracy	93.33%	96%
Precision	91.39%	97.32%
Recall	95.17%	94.77%

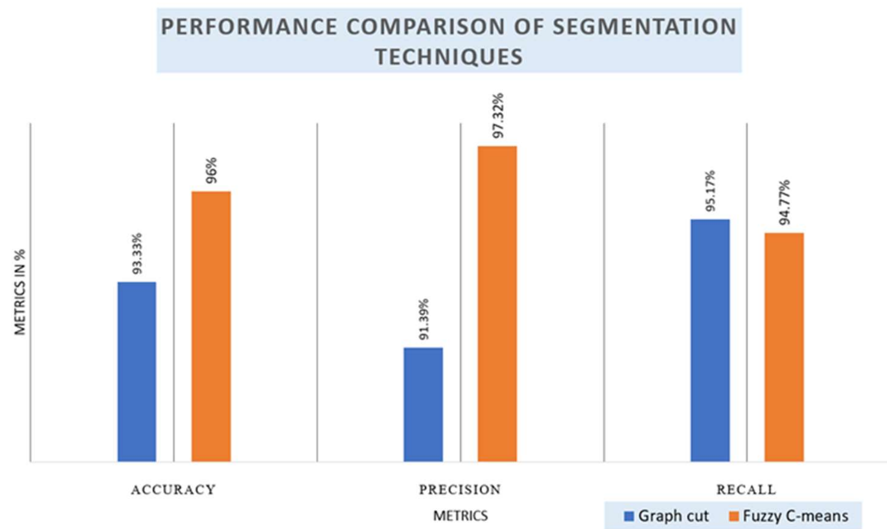

Fig. 6. Performance comparison of segmentation techniques

Conclusion

The detection and treatment of lung disorders rely heavily on computerized lung segmentation from CXR images. Patients' CXRs often show opacities in the lungs, making it challenging to segment the lungs. A segmentation method is proposed in this study as a means of addressing the issue. We employed both graph's cut-based segmentation and fuzzy clustering-based segmentation to automatically extract lung tissue from CT data. Considering that all lung parameters necessitate segmentation to be computed, the results of the segmentation research are useful for the numerical analysis of lung parameters. The segmentation output is then used to draw out the most crucial elements. The SVM model is used to categorize images into normal and pathological categories based on the retrieved attributes. The SVM classification accuracy is improved by 2.7% thanks to the FCM segmentation. This allows us to evaluate the relative merits of several segmentation techniques for distinguishing between normal and cancerous lung tissue.

REFERENCES

Haralick, R. M.—Shapiro, L. G, “Image Segmentation Techniques”,Comput. Vision Graphics Image Process., vol. 29, pp. 100–132,1985.

Pal, N.—Pal, S, “A Review on Image Segmentation Techniques”, Pattern Recognition, vol. 26, pp. 1277–1294,1993.

Fu, S. K.Mui, J. K,“A Survey on Image Segmentation”, Pattern Recognition, vol. 13, pp. 3–16,1981.

Y. Chunran, W. Yuanvuan and G. Yi, "Automatic Detection and Segmentation of Lung Nodule on CT Images," 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 2018, pp. 1-6, doi: 10.1109/CISP-BMEI.2018.8633101.

H. Van Lung and J. -M. Kim, "A generalized spatial fuzzy c-means algorithm for medical image segmentation," 2009 IEEE International Conference on Fuzzy Systems, Jeju, Korea (South), 2009, pp. 409-414, doi: 10.1109/FUZZY.2009.5276878.

S. Shaik Parveen and C. Kavitha, "Segmentation of CT lung nodules using FCM with firefly search algorithm," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2015, pp. 1-6, doi: 10.1109/ICIIECS.2015.7193181.

S. Azimi and H. Rabbani, "A level set-based method for lung segmentation in CT images," 2014 22nd Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 2014, pp. 1917-1920, doi: 10.1109/IranianCEE.2014.6999854.

A. Misra, M. Rudrapatna and A. Sowmya, "Automatic lung segmentation: a comparison of anatomical and machine learning approaches," Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004., Melbourne, VIC, Australia, 2004, pp. 451-456, doi: 10.1109/ISSNIP.2004.1417503.

F. Yi and I. Moon, "Image segmentation: A survey of graph-cut methods," 2012 International Conference on Systems and Informatics (ICSAI2012), Yantai, China, 2012, pp. 1936-1941, doi: 10.1109/ICSAI.2012.6223428.

Bezdek, James & Ehrlich, Robert & Full, William, “FCM—the Fuzzy C-Means clustering-algorithm”, Computers & Geosciences, vol. 10, pp. 191-203, 1984, doi: 10.1016/0098-3004(84)90020-7.

Ming-Chuan Hung and Don-Lin Yang, "An efficient Fuzzy C-Means clustering algorithm," Proceedings 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 2001, pp. 225-232, doi: 10.1109/ICDM.2001.989523.

V. Vapnik. “The Nature of Statistical Learning Theory”, NY: Springer-Verlag, 1995.