# LUNG CANCER DETECTION BY USING FUZZY CLUSTERING METHODS & MACHINE LEARNING TECHNIQUES

## Mrs.S.Divya[1], P.Lakshmi Sowmya[2], R.Sreelakshmi[2], Sk.Chandini[2], M.Harika[2]

[1]Assistant Professor, Department of ECE, Narayana Engineering College, Gudur, AP, 524101

[2]UG Student, Department of ECE, Narayana Engineering College, Gudur, AP, 524101.

[1]manju.mahi415@gmail.com, [2]plsowmya4@gmail.com,

------------------------------------------------------------------------***-------------------------------------------------------------------------

**Abstract -** Cancer is a disease where cancerous cells grow out of control, taking over normal cells and organs in the body. Lung Cancer is the uncontrollable formation of the malignant cells in the lungs. The early detection of lung cancer is the most effective way to decrease the mortality rate. However, in this paper, to obtain an exact boundary of the regions, every empirical dispersion of the image is computed by Fuzzy C-Means Clustering segmentation. A classification process based on the Support Vector Machine (SVM), a machine learning classifier is accomplished to distinguish the normal tissue and the abnormal tissue. The experimental evaluation is done using the lung database collected from kaggle portal.

Software Tool:

*Key Words***:** Lung cancer, Fuzzy clustering segmentation, Support vector machine, Machine learning.

## 1.INTRODUCTION

Lung Cancer is considered to be as the main cause of cancer death world wide, and it is difficult to detect in its early stages because symptoms appear only at advanced stages causing the mortality rate to be the highest among all other types of cancers. More people die because of lung cancer than any other type of cancer such as: breast, colon, and prostate cancers. There is significant evidence indicating that the early detection of lung cancer will decreases the mortality rate. The most recent estimates according to the latest statistics provided by the World Health Organization indicates that around 7.6 million deaths worldwide each year because of this type of cancer. Furthermore, mortality from cancer are expected to continue raising, to become around 17 million worldwide in 2030. There are many techniques to diagnosis lung cancer, such as Chest Radiograph (X-ray), Computed Tomography (CT) and Magnetic Resonance Imaging (MRI scan).

However, most of these techniques are expensive and time consuming. In other words, most of these techniques are detecting the lung cancer in its advanced stages, where the patients chance of survival is very low Therefore, there is a great need for a new technology to diagnose the lung cancer in its early stages. Image processing techniques provides a good quality tool for improving the manual analysis. In this project we are using fuzzy clustering segmentation and Support Vector Machine a machine learning classifier for the detection of lung cancer.

## 2. AIM AND OBJECTIVE

The main aim of this project is to detect the lung cancer by using Fuzzy Clustering Segmentation and SVM(support vector machine) a machine learning classifer. The main objective of this project is to construct a program by using machine learning classifier for the detecting the tumor and classify whether it is in normal or abnormal condition.

## 3.LITERATURE SURVEY

**W. Wang and S. Wu:** The paper presents that (1) we apply image processing technique into lung tissue information recognition, the key and hardest task is auto-detecting the tiny nodules, which may present the information of early lung cancer; and (2) the newly developed ridge detection algorithm is to diagnose indeterminate nodules correctly, allowing curative resection of early-stage malignant nodules and avoiding the morbidity and mortality of surgery for benign nodules. The algorithm has been compared to some traditional image segmentation algorithms. All the results are satisfactory for diagnosis

**Summary:** Newly developed ridge detection algorithm is to diagnose indeterminate nodules correctly, allowing curative resection of early-stage malignant nodules

**A. Sheila and T. Ried:** This perspective on Varella-Garcia et al. (beginning on p. XX in this issue of the journal) examines the role of interphase fluorescence *in-situ* hybridization (FISH) for the early detection of lung cancer. This work involving interphase FISH is an important step towards identifying and validating a molecular marker in sputum samples for lung-cancer early detection and highlights the value of establishing cohort studies with biorepositories of samples collected from participants followed over time for disease development.

**Summary:** It is used for the early detection of lung cancer

**D. Kim, C. Chung and K. Barnard:** Research has been devoted in recent years to relevance feedback as an effective solution to improve performance of image similarity search. However, few methods using the relevance feedback are currently available to perform relatively complex queries on large image databases. In the case of complex image queries, images with relevant concepts are often scattered across

several visual regions in the feature space. This leads to adapting multiple regions to represent a query in the feature space. Therefore, it is necessary to handle disjunctive queries in the feature space. In this paper, we propose a new adaptive classification and cluster-merging method to find multiple regions and their arbitrary shapes of a complex image query. Our method achieves the same high retrieval quality regardless of the shapes of query regions since the measures used in our method are invariant under linear transformations. Extensive experiments show that the result of our method converges to the users true information need fast, and the retrieval quality of our method is about 22% in recall and 20% in precision better than that of the query expansion approach, and about 35% in recall and about 31% in precision better than that of the query point movement approach, in MARS. 2005 Elsevier Inc. All rights reserved.

**Summary:** Achieves the same high retrieval quality regardless of the shapes of query regions since the measures used in our method are invariant under linear transformations.

**S. Saleh, N. Kalyankar, and S. Khamitkar:** In this paper, we present methods for edge segmentation of satellite images; we used seven techniques for this category; Sobel operator technique, Prewitt technique, Kiresh technique, Laplacian technique, Canny technique, Roberts technique and Edge Maximization Technique (EMT) and they are compared with one another so as to choose the best technique for edge detection segment image. These techniques applied on one satellite images to choose base guesses for segmentation or edgedetection image.

**Summary:** The comparative studies applied by using seven techniques of edge detection and experiments are carried out for different techniques Kiresh, EMT and Perwitt techniques respectively are the best techniques for edge detection.

**L. Lucchese and S. K. Mitra:** Segmentation is the low-level operation concerned with partitioning images by determining disjoint and homogeneous regions or, equivalently, by finding edges or boundaries. The homogeneous regions, or the edges, are supposed to correspond to actual objects, or parts of them, within the images. Thus, in a large number of applications in image processing and computer vision, segmentation plays a fundamental role as the first step before applying to images higher-level operations such as recognition, semantic interpretation, and representation. Until very recently, attention has been focused on segmentation of gray-level images since these have been the only kind of visual information that acquisition devices were able to take and computer resources to hand le. Nowadays, color imagery has definitely supplanted monochromatic information and computation power is no longer a limitation in processing large volumes of data. The attention has accordingly been focused in recent years on algorithms for segmentation of color images and various techniques, ofted borrowed from the background of gray-level image segmentation, have been proposed. This paper provides a review of methods advanced in the past few years for segmentation of color images.

**Summary:** Color layer discrimination in cartographic documents of low graphical quality.

**F. Taher and R. Sammouda:** The analysis of sputum color images can be used to detect the lung cancer in its early stages. However, the analysis of sputum is time consuming and requires highly trained personnel to avoid high errors. Image processing techniques provide a good tool for improving the manual screening of sputum samples. In this paper two basic techniques have been applied: a region detection technique and a feature extraction technique with the aim to achieve a high specificity rate and reduce the time consumed to analyze such sputum samples. These techniques are based on determining the shape of the nuclei inside the sputum cells. After that we extract some features from the nuclei shape to build our diagnostic rule. The final results will be used for a computer aided diagnosis (CAD) system for early detection of lung cancer.

**Summary:** Achieves the same high retrieval quality regardless of the shapes of query regions since the measures used in our method are invariant under linear transformations.

## 4.EXISTING METHOD

### Edge Based Segmentations:

Edge Detection is one of the most commonly used operations in image analysis, and there are many techniques used in detecting edges in images. Edge detection refers to the process of identifying and locating sharp discontinuities in an image. The discontinuities are abrupt changes in pixel intensity which characterize boundaries of objects in a scene. Classical methods of edge detection involve convolving the image with an operator (a 2-D filter), which is constructed to be sensitive to large gradients in the image while returning values of zero in uniform regions. There are large numbers of edge detection operators available, each designed to be sensitive to certain type of edges. Variables involved in the selection of an edge detection operator include Edge orientation, Noise environment and Edge structure. The geometry of the operator determines a characteristic direction in which it is most sensitive to edges. Operators can be optimized to look for horizontal, vertical, or diagonal edges. Edge detection is difficult in noisy images, since both the noise and the edges contain high frequency content. Operators used on noisy images are typically larger in scope, so they can average enough data to discount localized noisy pixels. The edge detection techniques considered here include Sobel, Robert, Canny, Prewitt and LoG (Laplacian of Gaussian).

Can't do work properly on the smooth transition imageNoise sensitivity is another big disadvantage.

## 5.PROPOSED METHOD

### Fuzzy C-Means Clustering Segmentation:

Clustering is the process of dividing the data into groups based on similarity of objects; information that is logically similar physically is stored together. Fuzzy Clustering has been used in many fields like pattern recognition and Fuzzy identification. A variety of Fuzzy clustering methods has been proposed and most of them are based upon distance criteria. The most widely used algorithm is the Fuzzy C-Mean algorithm (FCM). The FCM was introduced by J.C.Bezdek.

**SVM:**

An SVM model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane(MMH).
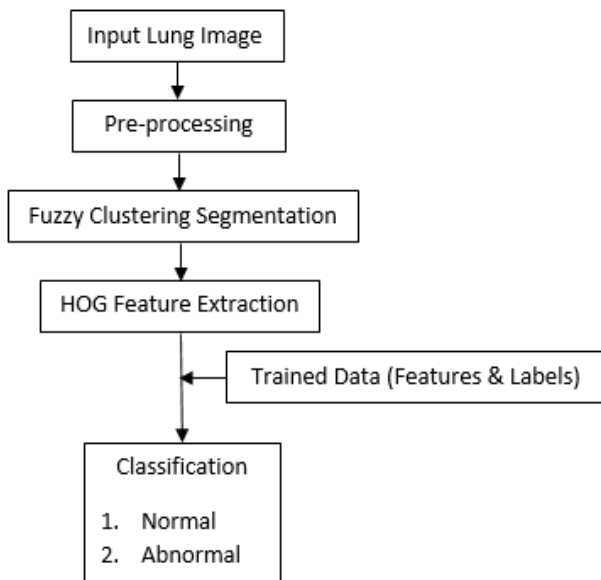


Fig: Block Diagram of Proposed Method.

FCM gives best result for overlapped data set and comparatively better than k-means algorithm and edge based methods. SVM works relatively well when there is a clear margin of separation between classes. SVM is more effective in high dimensional spaces. SVM is effective in cases where the number of dimensions is greater than the number of samples. SVM is relatively memory efficient.

Fuzzy c-means (FCM) clustering algorithm has been widely used in many medical image segmentations. SVM is used in Image-based analysis and classification tasks, Security-based applications, Speech recognition.

## 6.NUMBER OF MODULES

**1.Lung Cancer:** *Cancer* is a disease in which cells in the body grow out of control. When cancer starts in the lungs, it is called *lung cancer.*Lung cancer begins in the lungs and may spread to lymph nodes or other organs in the body, such as the brain. Cancer from other organs also may spread to the lungs. When cancer cells spread from one organ to another, they are called *metastases.*Lung cancers usually are grouped into two main types called small cell and non-small cell (including adenocarcinoma and squamous cell carcinoma). These types of lung cancer grow differently and are treated differently. Non-small cell lung cancer is more common than small cell lung cancer.

**2.Fuzzy Clustering Segmentation:** Fuzzy c-means (FCM) is a data clustering technique in which a data set is grouped into *N* clusters with every data point in the dataset belonging to every cluster to a certain degree. For example, a data point

that lies close to the center of a cluster will have a high degree of membership in that cluster, and another data point that lies far away from the center of a cluster will have a low degree of membership to that cluster.The fcm function performs FCM clustering.

**3.Support Vector machine (SVM):** Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:
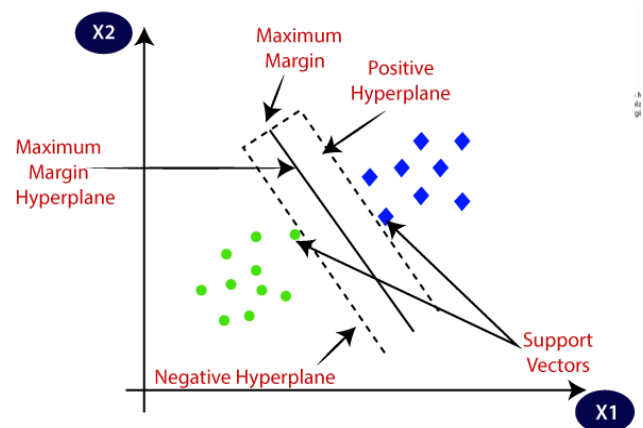


Fig: SVM Classification

**4.Machine Learning:** Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

## 7.RESULTS



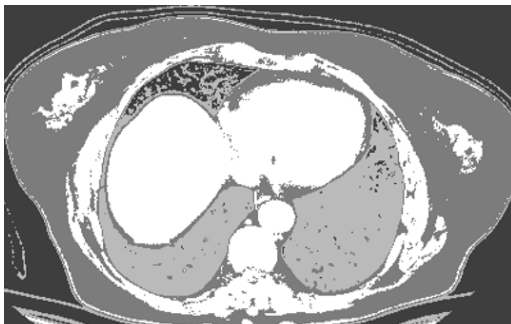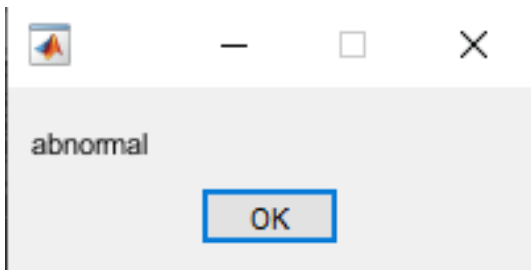Fig: Input Image



Fig: Segmented Image



Fig: Message Box Output
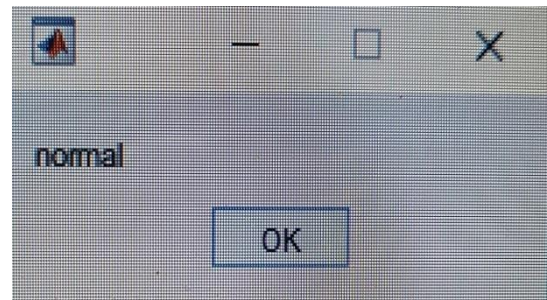


Fig: Input Image



Fig: Segmented Image



Fig: Message Box Output

## 8.CONCLUSIONS

In this paper, Fuzzy C-Means clustering segmentation technique for segmenting lung cancer images. Then classificatin of lung cancer through images as normal/abnormal is done by using SVM. The most effective strategy to reduce lung cancer mortality is early detection. Every empirical dispersion of the image is computed by Fuzzy C-Means Clustering segmentation in this proposed method to provide an exact boundary of the regions. A machine learning classifier is used to distinguish between normal and pathological tissue using a classification approach based on the Support Vector Machine (SVM).

## FUTURE SCOPE

FCM gives best result for overlapped data set and comparatively better than k-means algorithm and edge based methods. SVM works relatively well when there is a clear margin of separation between classes. SVM is more effective in high dimensional spaces. SVM is effective in cases where the number of dimensions is greater than the number of samples. SVM is relatively memory efficient.Fuzzy c-means (FCM) clustering algorithm has been widely used in many medical image segmentations. SVM is used in Image-based analysis and classification tasks, Security-based applications, Speech recognition.

## REFERENCES

[1] W. Wang and S. Wu, "A Study on Lung Cancer Detection by Image Processing", proceeding of the IEEE conference on Communications, Circuits and Systems, pp. 371-374, 2006.

[2] A. Sheila and T. Ried "Interphase Cytogenetics of Sputum Cells for the Early Detection of Lung Carcinogenesis", Journal of Cancer Prevention Research, vol. 3, no. 4, pp. 416-419, March, 2010.

[3] D. Kim, C. Chung and K. Barnard, "Relevance Feedback using Adaptive Clustering for Image Similarity Retrieval," Journal of Systems and Software, vol. 78, pp. 9-23, Oct. 2005.

[4] S. Saleh, N. Kalyankar, and S. Khamitkar," Image Segmentation by using Edge Detection", International Journal

on Computer Science and Engineering(IJCSE), vol. 2, no. 3, pp. 804-807, 2010.

[5] L. Lucchese and S. K. Mitra, "Color Image Segmentation: A State of the Art Survey," Proceeding of the Indian National Science Academy (INSA-A), New Delhi, India, vol. 67, no. 2, pp. 207-221, 2001.

[6] F. Taher and R. Sammouda," Identification of Lung Cancer based on shape and Color", Proceeding of the 4th International Conference on Innovation in Information Technology, pp.481-485, Dubai, UAE, Nov. 2007.