



# A Survey on Lung Cancer Detection using Image Data Analysis and Machine Learning

Apoorva Mahale<sup>1</sup>, Chetan Rawool<sup>1</sup>, Dinesh Tolani<sup>1</sup>, Deepesh Bathija<sup>1</sup>, Prof. Kajal Jewani<sup>2</sup>

B.E. Students, Dept. of Computer Science, Vivekanand Education Society's Institute of Technology, Mumbai, India<sup>1</sup>  
Asst. Professor, Dept. of Computer Science, Vivekanand Education Society's Institute of Technology, Mumbai, India<sup>2</sup>

**ABSTRACT:** This study aims to highlight the significance of data analytics and machine learning (both burgeoning domains) in prognosis in health sciences, particularly in detecting life threatening and terminal diseases like cancer. Here, we consider lung cancer for our study. For this purpose, preexisting lung cancer patients' data are collected to get the desired results. Data set (in the form of diagnostic images) is run past Matlab for analysis and forecasting. Image processing is employed for this purpose. Medical image segmentation and classification are done to achieve this. The patients' Computed Tomography (CT) lung images are categorized as normal or abnormal. The abnormal images are subjected to segmentation to focus on the tumor portion. Classification depends on features extracted from the images. The emphasis is on the feature extraction stage to yield better classification performance. This information is then fed to machine learning algorithms to discern a pattern that can give some good insights into what combination of features are most likely to result in an abnormality. The ultimate goal is to identify effective and common methods for classification using some well-established machine learning algorithms like FPCM and its improved versions.

**KEYWORDS:** prognosis, Matlab, CT, feature extraction, machine learning

## I. INTRODUCTION

Incorrect and/or late diagnoses are the two leading causes of cancer related deaths. Lung cancer in particular, is a major killer the world over. Poor prognosis means that less than 15% of registered cases survive past five years of diagnosis. The poor prognosis is a consequence of inefficient diagnostic methods for early detection and successful treatment. It has been observed that patients in whom lung cancer has been detected early suffer a lower risk of lung cancer-related mortality than those at an advanced stage of the disease, suggesting that early detection and treatment of lung cancer can go a long way in the successful treatment of cancer. The objective of undertaking this project is to facilitate doctors to provide the best possible treatment by providing useful insights through analysis and diagnosis of lung cancer treatments with the help of predictive models.

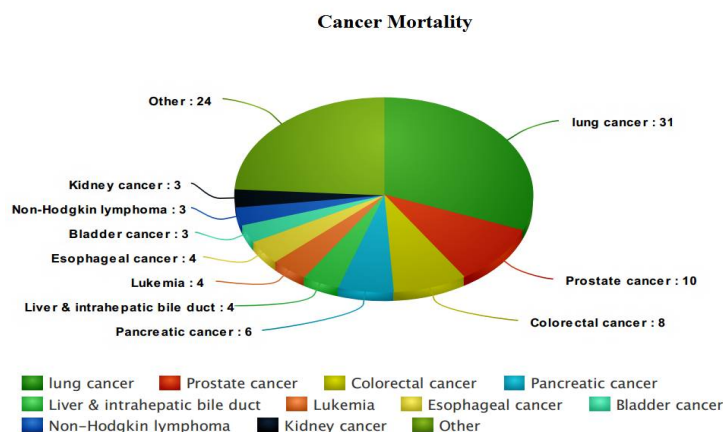


Fig 1: Most widespread forms of cancer worldwide



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 1, January 2017

## II. LITERATURE SURVEY

In this paper [1] a generalized approach is discussed for lung cancer detection using CT scan images of the chest. According to this paper, a computer aided lung cancer detection system involves three main processing stages: enhancement, segmentation and feature extraction. Further, these stages are explained in detail with all possible methods that can be used in each of the stages. Methods used with enhancement are: Median filter, Auto enhancement, Fast Fourier transform. Methods used with segmentation are: Thresholding approach, Watershed Segmentation, Region Growing Segmentation. Feature extraction includes parameters such as Area, Perimeter, Mean, Standard deviation. In the end it compares the efficacy of methods like Neuro-Fuzzy model and Region Growing Method against CT image analysis.

**Inference:** The lung cancer detection on CT image using image processing achieves a higher detection rate of 90.1% while Neuro-Fuzzy Model achieves 89.3% and Region Growing Method achieves 88.5% detection rate.

In this paper [2], Fuzzy C-Means (FCM) and different extensions of FCM algorithm are discussed. The exclusive FCM algorithm yields better results for segmenting noise free images, but it fails to segment images downgraded by noise, outliers and other imaging artifacts. This paper presents an image segmentation approach using Modified Fuzzy Possibilistic C-Means algorithm (MFPCM), which is a generalized adaptation of the standard Fuzzy C-Means Clustering algorithm and Fuzzy Possibilistic C-Means algorithm. The flaw of the conventional FCM technique is eliminated by modifying the standard technique, the Modified FCM algorithm is defined by modifying the distance measurement of the standard FCM algorithm to permit the labeling of a pixel to be influenced by other pixels and to restrain the noise effect during segmentation. Rather than having one term in the objective function, a second term is included, forcing the membership to be as high as possible without a maximum frontier restraint of one. Experiments are carried out on real images to observe the performance of the proposed modified Fuzzy Possibilistic FCM technique in segmenting the medical

images. Standard FCM, Modified FCM, Possibilistic C-Means algorithm, Fuzzy Possibilistic C-Means algorithm are compared with Modified FPCM to explore the accuracy. The three most important parameters used to determine the accuracy of the Modified FPCM are similarity, false positive and the false negative ratio.

**Inference:** Modified FPCM achieves higher accuracy of 94.25% while Fuzzy Possibilistic C-Means Clustering (FPCM) achieves 92.50% accuracy, Possibilistic C-Means Clustering Algorithm (PCM) achieves 91.00% accuracy, Modified FCM achieves 89.50% accuracy and standard FCM shows least accuracy of 86.03%.

In this paper [3], SVM classifier is used for classifying the tumor as Benign or Malignant once the CAD (Computer Aided Diagnosis) system for finding the lung tumor using the lung CT images is in place.

Support vector machines are supervised learning models with corresponding learning algorithms that perform data analysis and pattern recognition, used for classification. The basic SVM takes a collection of input data and for each input given, predicts which of the two classes forms the input, making it a non-probabilistic binary linear classifier. From the given set of training example data, each noted as belonging to one of two categories, an SVM algorithm builds a model that assigns new examples into one category or the other. In the proposed system, we choose the linear classifier technique.

Best hyper plane is the one that exhibits the largest separation or margin between the two classes. So we choose the hyper plane such that its distance from the nearest data point on each side is maximized. Such a hyper plane is known as the maximum margin hyper plane and the linear classifier it defines is known as a maximum classifier. In case of SVM classifier out of 9 features at a time only two features are selected for classification, which produces result as either benign or malignant.

**Inference:** SVM provides the accuracy of 92.5%. The accuracy of a system can be improved if training is performed using a very large image database.

This paper [4] discusses the difficulties encountered while detecting lung nodules in radiographs. These difficulties are: Variation in nodule size, variation in the density of the nodule and difficulty of appearance of nodule anywhere in the

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Issue 1, January 2017

lung field, where they are likely to be obscured by ribs, the mediastinum and similar structures beneath the diaphragm, causing a large variation of contrast to the background. To overcome these issues, the paper proposes a Computer Aided Diagnosis (CAD) system for detection of lung nodules. This paper initially applies different image processing techniques for lung region extraction. Further, the Fuzzy Possibilistic C Means (FPCM) algorithm is used for segmentation. For learning and classification, Extreme Learning Machine (ELM) is used. The experimentation has been performed on 1000 images obtained from reputed hospitals.

**Inference:** The experimental result shows that the proposed Computer Aided Diagnosing system is capable of detecting the false positive nodules correctly. Also, the use of Extreme Learning Machine increases the efficiency of detecting cancer nodules.

### III. PROPOSED SYSTEM

Five phases included in the proposed computer aided diagnosis system for lung cancer detection are as follows:

1. Extraction of lung region from chest computer tomography (CT) images
2. Segmentation of lung region using Modified Fuzzy Possibilistic C-Means
3. Feature extraction from the segmented region
4. Formation of diagnosis rules from the extracted features
5. Classification of occurrence and nonoccurrence of cancer in the lung

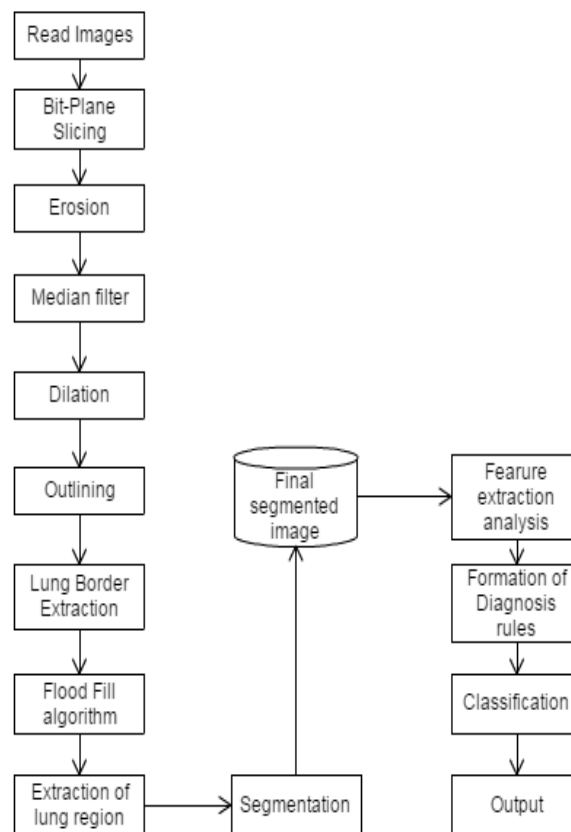


Fig 2: Proposed System Design

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 1, January 2017

In our proposed method, we preprocess the given test images for eliminating the noise and enhancing the contrast. Afterwards, texture features will be extracted from them. Among the several features extracted, distinct features are selected that will be utilized later on for classification. Support Vector Machines (SVM) is used to classify whether the test image is normal or abnormal.

## Computer Aided Diagnosis Process:

### Phase 1: Extraction of Lung Region from (CT) Images

The first phase of the proposed Computer Aided Diagnosing system is the extraction of lung region from the (CT) scan images. This phase uses basic image processing methods. The procedure for performing this phase using the image processing methods is provided in Fig 3.

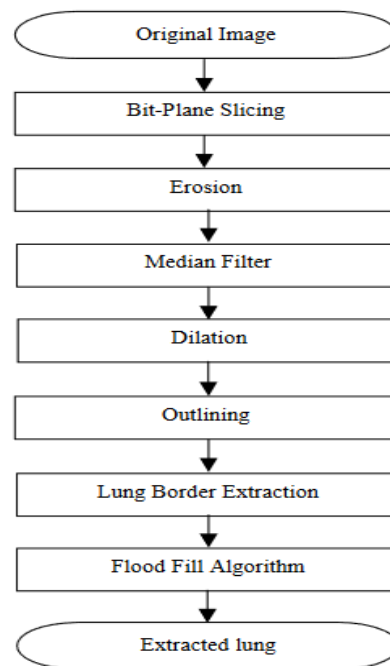


Fig 3: Extraction of Lung Region from (CT) Images

The image processing methods used in this phase are Bit-Plane Slicing, Erosion, Median Filtering, Dilation, Outlining, Lung Border Extraction and Flood-Fill algorithm. Generally, the CT chest image contains not just the lung region, but also the background, heart, liver and other organ areas. The chief objective of the lung region extraction process is to detect the lung region and pinpoint regions of interest (ROIs) in the CT scan image.

### BIT PLANE SLICING:

Bit plane slicing is used to highlight the contribution made to the total image appearance by specific bits hence it is useful for analyzing importance played by each bit of the image. In bit plane slicing each pixel is represented by 8 bits, the image is composed of eight 1-bit planes. Plane one is composed of least insignificant bits and if all these 1-bit planes are combined we will get the most insignificant image (to the right side of the original image in above example). Plane 8 contains the most significant bits and when all the plane 8 bits are combined we will get the most significant and clear image (the one to the lower right of the above example), thus in bit plane slicing only the higher order bits contain the visually significant data.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 1, January 2017

## EROSION:

Erosion is one of the two basic operators in the area of mathematical morphology. It is also known by the name "shrink". It turns off pixels which were near pixels that were off originally, thereby eating away at the edges of the items. The erosion operator takes two pieces of data as inputs. The first is the image which is to be eroded. The second is a (generally small) collection of coordinate points called a structuring element. This structuring element determines the precise effect of the erosion on the input image.

The opening of set  $X$  by structuring element  $B$  is denoted as  $X \ominus B$ , where:

$$X \ominus B = X - b = \{ z : (B + z) \subseteq X \}$$

If  $X$  is any gray scale shape and  $B$  is a symmetric structuring element, the output of erosion is the set of translation points such that the input set  $X$  contains the translated structuring element. This equation implies that the erosion of  $X$  by  $B$  is the set of all points  $b$  such that  $B$ , translated by  $b$ , contains  $X$ .

## MEDIAN FILTERING:

The Median filter is a nonlinear filter. Unlike linear filters, median filters replace the pixel values with the median value available in the local neighborhood (say, 5x5 or 3x3 pixels around the central pixel value). Also, median filter is edge preserving. It is used to remove salt and pepper noise.

Step 1: Input the image.

Step 2: Select a 3X3 window around the current pixel within the image.

Step 3: Sort the pixels in increasing order and store it to vector.

Step 4: Find the median of the vector.

Step 5: Replace the current pixel with the median value.

Step 6: Repeat the steps 2 to 5 until all the pixels within the image are processed.

Step 7: Result.

## DILATION:

Dilation is a basic mathematical morphology operator. The basic effect of the operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels. Thus areas of front pixels grow in size, while holes within those regions become lesser.

The dilation operator takes as input two pieces of data. Firstly, it takes the image to be dilated. Secondly, it accepts as input a (usually small) set of coordinate points called a structuring element. This structuring element decides the explicit effect of the dilation on the input image. The opening of set  $X$  by structuring element  $B$  is denoted as  $X \oplus B$ , where:

$$X \oplus B = X + b = \{ x + b : x \in X \& b \in B \}$$

If  $X$  is any gray scale shape and  $B$  is balanced structuring element. The output of dilation is the set of translated points such that translate of the returned structuring element has a non-empty intersection with  $X$ . This equation is based on obtaining the reflection of  $B$  about its origin and changing this reflection by  $b$ . this dilation of  $X$  by  $B$  then is the set of all displacements,  $b$ , such that  $x$  and  $b$  lap over by at least one element.

## OUTLINING:

Outlining is basically used to create the outline of the whole image including lungs, ribcage etc. This is performed after all the noise and distortions from the image are removed.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 1, January 2017

## LUNG REGION EXTRACTION:

It is used to extract the lung region after the image outlining is performed, so that the resultant image can be further used for flood fill algorithm.

The following are the steps to perform lung region extraction:

- Step 1. Scan the image from left to right to locate the leftmost pixel of the lung region.
- Step 2. Draw a vertical line along this pixel from top to bottom, representing the boundary.
- Step 3. Scan the edge map from the right to left, from the first row.
- Step 4. Obtain a pixel that is black, indicating an edge path. Traverse the pixel path by regarding all the surrounding pixels in a clockwise priority and taking into consideration the pixel with the highest priority.
- Step 5. The pixels surrounding the edge pixel (but of lower priority) are kept in a stack (to be used only if the traversal process reaches a dead end).
- Step 6. If a dead end is reached, pop out from the stack a lesser priority pixel and continue with the traversal process.
- Step 7. Store the pixels traversed in a plotting list to be used later for drawing the lung boundary.
- Step 8. Continue traversing to the next pixel till the bottom of the image is reached.
- Step 9. If the bottom of the image is not reached, the path is discarded, the plot list is erased and the process is continued from Step 4. Else get pixels from the plotting list at a discrete interval.
- Step 10. Plot the first pixel.
- Step 11. Draw a simple curve between two consecutive pixel position separated by the discrete interval. Continue plotting till the last pixel in the Plotting List is plotted.

## FLOOD FILL ALGORITHM:

After the lung region extraction is performed, flood fill algorithm is used to fill the lung area so that it can be further used for the process of segmentation.

The flood-fill algorithm accepts three parameters: a start node, a target color, and a replacement color. The algorithm looks for all nodes in the array that are connected to the start node through a path of the target color and converts them to the replacement color. There are several ways in which the flood-fill algorithm can be structured, but they all make use of a stack data structure, explicitly or implicitly.

Flood-fill (node, target-color, replacement-color):

- Step 1. If target-color is equal to replacement-color, return.
- Step 2. If the color of node is not equal to target-color, return.
- Step 3. Set the color of node to replacement-color.
- Step 4. Perform Flood-fill (one step south of the node, target-color, replacement-color).
- Perform Flood-fill (one step north of the node, target-color, replacement-color).
- Perform Flood-fill (one step west of the node, target-color, replacement-color).
- Perform Flood-fill (one step east of the node, target-color, replacement-color).
- Step 5. Return.

## MODIFIED POSSIBILISTIC FUZZY C-MEANS:

The second phase of the proposed CAD system is the lung region segmentation. Segmentation is performed for pinpointing the cancer nodules in the lung. This phase will identify the Region of Interest (ROI) that helps in determining the cancerous region. Modified Fuzzy Possibilistic C-Means (MFPCM) is used in the proposed technique for segmentation due to improved accuracy of MFPCM.

MFPCM adjust its membership function according to the weight; the segmentation can be performed with better accuracy. When the boundary regions of the cancer nodules are taken into consideration, the usage of FPCM will sometimes misjudge the edge of nodule due to its fixed membership function. This issue is overcome by the usage of MFPCM.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 1, January 2017

The objective function is composed of two expressions: the first is the fuzzy function and uses a fuzziness weighting exponent, the second is the possibilistic function and uses a typical weighting exponent.

## MODIFIED FPCM ALGORITHM:

1. Set the number of clusters =  $c$  and the index of fuzziness =  $m$ . Also initialize the fuzzy cluster Centroid vector =  $v$  randomly and set  $\epsilon > 0$  to a small value,
2. Set the neighborhood size and the window size includes the calculation of cluster centers and membership matrix,
3. Evaluate the modified distance measurement using the equation mentioned as  $d_{ki}(x_j, v_i)$ ,
4. Update the membership matrix and the distance measurement.

The objective function of the MFPCM can be formulated as follows:

$$J_{MFPCM} = \sum_{i=0}^c \sum_{j=1}^n (\mu_{ij}^m w_{ji}^m d^{2m}(x_j, v_j) + t_{ij}^n w_{ji}^n d^{2n}(x_j, v_j))$$

$U = \{\mu_{ij}\}$  indicates fuzzy partition matrix, and is described as:  $\mu_{ij} = \left[ \sum_{k=1}^c \left( \frac{d(x_j, v_i)}{d(x_j, v_k)} \right)^{\frac{2m}{m-1}} \right]^{-1}$

$T = \{t_{ij}\}$  indicates a typical partition matrix, and is represented as:  $t_{ij} = \left[ \sum_{k=1}^n \left( \frac{d(x_j, v_i)}{d(x_j, v_k)} \right)^{\frac{2n}{n-1}} \right]^{-1}$

$V = \{v_i\}$  indicates  $c$  centers of the clusters, and is represented as:

$$V_i = \frac{\sum_{j=1}^n (\mu_{ij}^m w_{ji}^m + t_{ij}^n w_{ji}^n) * x_j}{\sum_{j=1}^n (\mu_{ij}^m w_{ji}^m + t_{ij}^n w_{ji}^n)}$$

## Abbreviations:

$m$  - Index of Fuzziness

$c$  - No. of clusters

$v$  - Fuzzy cluster centroid vector

$J$  - Objective function

$v_i$  -  $P$  dimensional center of cluster

$d$ - Distance between  $i^{\text{th}}$  data and  $j^{\text{th}}$  cluster

$n$ - No. of patterns in data set

$w_{ji}$ - Weight of point  $j$  in relation to class  $i$

$x_j$ -  $j^{\text{th}}$  pattern

$t_{ij}$ - Possibility degree of pattern

$U$ - Membership matrix  $U$  is initialized with random values between 0 to 1.

## Phase 3: Feature Extraction from the Segmented Region

After the segmentation is implemented on the lung region, the features can be gained from it for Determining the diagnosis rule for identifying the cancer nodules in the lung region perfectly. The features that are used in this paper in order to create diagnosis rules are:

1. Area of the candidate region
2. The maximum drawable circle (MDC) inside the candidate region
3. Mean intensity value of the candidate region.

## Phase 4: Formation of Diagnosis Rules from the Extracted Features

After the necessary features are derived, the following diagnosis rules can be applied to detect the occurrence of cancer nodule. The three rules involved for diagnosis are as follows:

Rule 1: At the beginning the threshold value  $T1$  is set for area of region. If the area of candidate region exceeds the



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 1, January 2017

threshold value, then it is eliminated for further consideration. This rule will help in contracting the steps and time necessary for the upcoming steps.

Rule 2: In this rule maximum drawable circle (MDC) is considered. The threshold T2 is defined for value of maximum drawable circle. If the radius of the drawable circle for the candidate Region is less than the threshold T2, then that region is considered as non-cancerous nodule and is excluded for further consideration. Implementing this rule has the effect of rejecting large number of vessels, which in general have a thin oblong, or line shape.

Rule 3: In this, the range of value T3 and T4 are set as threshold for the mean intensity value. Then the mean intensity values for the candidate regions are calculated. If the mean intensity value of candidate region goes down minimum threshold or goes above maximum threshold, then that region is pretended as non-cancerous region.

By implementing all the above rules, the maximum of regions which does not examined as cancerous nodules are eliminated. The remaining candidate regions are considered as cancerous regions. This CAD system helps in neglecting all the false positive cancer regions and helps in identifying the cancer regions more accurately. These rules can be passed to classifier in order to detect the cancer nodules for the supplied lung image.

## Phase 5: Classification of Occurrence and Nonoccurrence of Cancer in the Lung

The final phase in the proposed CAD system is the classification of occurrence and nonoccurrence of cancer nodule for the supplied lung image. The classifier used in this paper is Support Vector Machine.

## IV. MACHINE LEARNING ALGORITHM

### Support Vector Machines (SVMs)

Support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification. Consider the pattern classifier, which uses a hyper plane to separate two classes of patterns based on given examples  $\{x(i), y(i)\} i=1..n$ . Where  $(i)$  is a vector in the input space  $I=R^k$  and  $y(i)$  denotes the class index taking value 1 or 0.

A support vector machine is a machine learning method that classifies binary classes by finding and using a class boundary the hyper plane maximizing the margin in the given training data. The training data samples along the hyper planes near the class boundary are called support vectors, and the margin is the distance between the support vectors and the class boundary hyper planes. The SVM are based on the concept of decision planes that define decision boundaries.

Theory of SVM is described below:

Given training data  $(x_i, y_i)$  for  $i = 1 \dots N$ , with  $x_i \in R^d$  and  $y_i \in \{-1, 1\}$

Every input  $x$  is initially mapped into a higher dimension feature space  $F$ , by  $z = \phi(x)$  through a nonlinear mapping  $\phi: R^n \rightarrow F$ .

$$f(x) = W^T x + b$$

$W$  is the normal to the line, and  $b$  the bias.  $W$  is known as the weight vector.

Instead, the SVM can be formulated to learn a linear classifier,  $f(x) = \sum ( ) +$

A decision plane is one that separates between assets of objects having different class memberships. SVM is a useful technique for data classification. A classification task usually involves with training and testing data which consists of some data instances. Each instance in the training set contains one "target value" (class labels) and several "attributes" (features).



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 1, January 2017

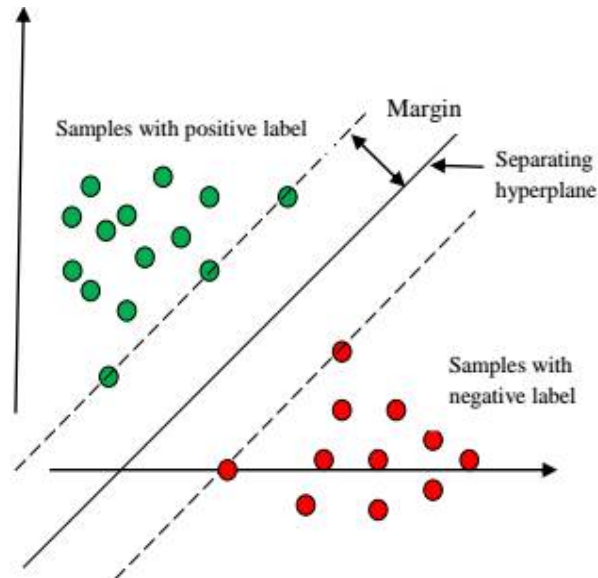


Fig 4: Classification using SVM Classifier

## V. CONCLUSION

The issues encountered in early detection of lung cancer nodules have been overcome in this paper. This paper comes up with an efficient computer aided diagnosis system for early lung cancer detection. Chest computer tomography images have been used over here as input datasets. In the first phase of the proposed technique, the lung region is extracted from the chest tomography images. Different basic image processing techniques are used for this purpose. In the second phase, the extracted lung region is segmented with the aid of modified fuzzy possibilistic c-means algorithm. The next phase is derivation of features for diagnosis from the segmented image. Next, diagnosis rules are generated from the extracted features. Finally with the obtained diagnosis rules, classification is performed to detect the occurrence of cancer nodules. For evaluation purpose, different real time chest computer tomography images are used. The empirical results demonstrate the advantage of the proposed CAD system for detecting lung cancer. The experiment shows that using SVM will result in higher classification accuracy.

## VI. FUTURE SCOPE

This project mainly deals with Computer Aided Diagnosis System (CAD) for early detection of lung cancer nodules from the Chest Computer Tomography (CT) image using MFPCM, approach. This technique is highly successful in detecting lung cancer. The proposed CAD system with SVM approach provides the best results. For improving the performance of the proposed approaches, some future enhancements would be essential to improve upon the current research work. The main aim of the future enhancements would be to increase the sensitivity and specificity of the system.

The future scope for this research is to use Extreme Learning Machine in the CAD system. The use of ELM would provide even better results. Moreover the training time and the accuracy of the system would also be bettered by using ELM. The second enhancement in this research would be to use new clustering algorithms which can provide better results than the proposed clustering algorithms.

Further future enhancements would be incorporating cutting edge technologies like nanotechnology, genetic algorithm etc., into the CAD system for better performance of the system.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 1, January 2017

## REFERENCES

- [1] Sruthi Ignatious, Robin Joseph, "Computer Aided Lung Cancer Detection System", Global Conference on Communication Technologies (GCCT 2015).
- [2] M.Gomathi, "A Parameter Based Modified Fuzzy Possibilistic C-Means Clustering Algorithm for Lung Image Segmentation", Global Journal of Computer Science and Technology Vol. 10 Issue 4 Ver. 1.0 June 2010.
- [3] Swati P. Tidke, Vrishali A. Chakkarwar, "Classification of Lung Tumor Using SVM", International Journal Of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 5 September 2012.
- [4] M.Gomathi, Dr.P.Thangaraj, "A Computer Aided Diagnosis System for Detection of Lung Cancer Nodules Using Extreme Learning Machine", International Journal of Engineering Science and Technology, ISSN: 0975-5462, Vol. 2(10), 2010.
- [5] Ayman El-Baz et al, "Computer-Aided Diagnosis Systems for Lung Cancer: Challenges and Methodologies", International Journal of Biomedical Imaging Volume 2013, Article ID 942353, 46 pages
- [6] C.Kavitha et al, "Classification of Lung Cancer Nodules using SVM Kernels", International Journal of Computer Applications (0975 – 8887), Volume 95– No.25, June 2014
- [7] FatmaTaher, NaoufelWergghi, Hussain Al-Ahmad, "Computer Aided Diagnosis System for Early Lung Cancer Detection", Algorithms 2015, doi:10.3390/a8041088, ISSN 1999-4893
- [8] Armato, S.G, Giger, M.L. and MacMahon, H., "Automated Detection of Lung Nodules in CT Scans: Preliminary Results", Medical Physics, Vol. 28, Pp. 1552–1561, 2001.
- [9] El-Baz, A., Gimel'farb, G., Falk, R. and El-Ghar, M.A., "A New CAD System for Early Diagnosis of Detected Lung Nodules", IEEE International Conference on Image Processing, Vol. 2, Pp. II-461-II-464, 2007. A Computer Aided Diagnosis System for Lung Cancer Detection using Machine Learning Technique 275
- [10] Ginneken, B.V., Romeny, B.M. and Viergever, M.A., "Computer-Aided Diagnosis in Chest Radiography: A Survey", IEEE Transactions on Medical Imaging, Vol. 20, No. 12, 2001.
- [11] Gurcan, M.N, Sahiner, B., Petrick, N., Chan, H., Kazerooni, E.A., Cascade, P.N. and Hadjiiski, L., "Lung Nodule Detection on Thoracic Computed Tomography Images: Preliminary Evaluation of a Computer-Aided Diagnosis System", Medical Physics, Vol. 29, No. 11, Pp. 2552- 2558, 2002.
- [12] Fiebich, M., Wormanns, D. and Heindel, W., "Improvement of Method for Computer-Assisted Detection of Pulmonary Nodules in CT of the Chest", Proceedings of SPIE Medical Imaging Conference, Vol. 4322, Pp. 702–709, 2001.

## BIOGRAPHY

**Kajal Jewani** is an Assistant Professor at the Department of Computer Science, Vivekanand Education Society's Institute of Technology, Mumbai (India). Her interests include Image Processing and Theory of Computation.

**Apoorva Mahale** is a final year Computer Science undergraduate student at Vivekanand Education Society's Institute of Technology, Mumbai (India). Her interests include Data Analytics and Web Development.

**Chetan Rawool** is a final year Computer Science undergraduate student at Vivekanand Education Society's Institute of Technology, Mumbai (India). His interests include exploring new Open Source tools.

**Dinesh Tolani** is a final year Computer Science undergraduate student at Vivekanand Education Society's Institute of Technology, Mumbai (India).

**Deepesh Bathija** is a final year Computer Science undergraduate student at Vivekanand Education Society's Institute of Technology, Mumbai (India).