Detection and Analysis of lung cancer Via K-means and C-means

P.Govindhan¹, S. Sathiya Devi, S. Gomathi², A. Amirbee³, A. Camilinfanta⁴
Assistant Professor¹, Student²,³,⁴,⁵
Department of ECE
Acharya College of Engineering Technology, Pondicherry, India

Abstract:
Lung cancer is the most prevalent cancer and it is the leading cause of cancer deaths worldwide. According to the overall survival rate of lung cancer patients is only 14%. The lives of cancer patients can be saved only if it is detected in the initial stages. The detection of cancer can be detected by the some of the techniques such as Positron emission tomography/computed tomodraphy. This improves diagnostic accuracy due to the integration of functional (PET) and anatomical information into a single scan. The collection of samples is difficult task for diagnosis. Therefore, computational systems are essential to assist radiologists to elucidate images in accurate diagnosis. The pre-processing methods such as contrast limited adaptive histogram equalization and wiener filter. This performs accuracy in detecting abnormal and normal images. The performance of the proposed methodology was evaluated using receiver operating characteristics curve and via k-means and C-means algorithm. The Proposed Method provides better classification and cancer detection with an overall accuracy of 92.67%. This paper revolves an analysis using different types of algorithms in order to diagnose lung cancer, thereby a survey of different types of algorithm or implementations is given to analyze easily the shapes of cancer cells.

Keywords: CT-computed tomography, PET-positron emission Tomography, FCM, ROC, K-means, C-means

I. INTRODUCTION

Lung cancer is a disease of abnormal cells multiplying and growing into a tumor. It is the leading cause of cancer deaths in both men and women and the mortality rates have been raised for the past 25 years. Lung cancer accounts for more deaths compared to other malignancies in the world. A low five year survival rate of 16% percentage is due to primarily late stage diagnosis. Surgery is often curative if the cancer is diagnosed in early stages of 1 and 2 but only 30% of non small cell lung cancer is made at this early stage. Lung cancer is not simple to diagnose and several types of tests are available in this treatment. Treatments include surgery, chemotherapy and radiation. Radiology aids in the diagnosis of cancer by using various imaging procedure such as CT (computed tomography) and MRI (magnetic resonance imaging). Typically, physician suspecting lung cancer in a patient does other confirmatory tests refer an expensive tomography scan and followed by biopsy for definitive confirmation. On the assumption that it is lung cancers have an exponential pattern of growth and it has been estimated that a tumor of 1 cm in diameter which is the smallest lesion that can be detected in a normal chest radiograph, will have undergone 30 volume doublings. A rapidly growing tumor such as small cell carcinoma may take 2-4 years to achieve this size rather a non small cell carcinoma may take 7-13 years. Therefore plenty of time would have been elapsed for the metastatic spread to develop before the tumor is diagnosed and treated. An obvious approach to these problems is to develop various methods in detecting lung cancer at much earlier stages when treatment becomes more likely to be curative.

II. RELATED WORKS

There are a number of researches which has been carried out in detecting lung cancer in preclinical stages. Studies using non specific gaseous chemical sensing devices-Mass Spectrometry (GC-MS). This involves breath testing analysis. The collected breath is chemically analyzed (GC-6890N, MS5975) which is combined with the SPME fiber. This SPME fiber is used for pre-concentration VOCs in breath sample. Studies using exhaled breadth screening of lung cancer (SCLC) (NSCLC). This method determines the small cell lung carcinoma and non small cell carcinoma which is further divided in the expectation of different types of phenotypes for patients suffering from lung cancer translates into different types of responses to chemo or radiotherapy and different long term prognosis. Studies related to breath samples of cancer patients. The breath samples collected from 96 volunteers results 30 primary lung cancer patients, 26 primary colon cancer patients, 18 primary prostate cancer patients and 22 healthy controls. Nonanol is an aldehyde compound (CH180), this is an aldehyde with high carbon content. It is found in the breath of persons suffering from lung cancer. The selectivity toward ammonia is very high. The sensitivity differential with acetone often found in the breath is 2500 times. Studies related to VOC’s include. A trace concentration of toluene (30ppb) in exhaled breadth is regarded to be the distinctive early symptom of lung cancer which been able to detect using a breath tester. The present standard technique for detection and staging of lung cancer is computed tomography (CT). CT screening programmes lead to the detection of many benign lung nodules which expose patients to radiation and even if proven to reduce the mortality rate. This technique proves not be cost-effective. The present technique describes the techniques of applying various algorithms and implementations to diagnose lung cancer in earlier stages his proves about involving K-means and C-means algorithm which detects and gives obvious research about the shapes of cancer cells.

III PROPOSED SYSTEM

ANALYSIS OF DIFFERENT ALGORITHMS

K-means algorithm: K-means clustering defines a method of vector quantization which is used for cluster analysis. K-means clustering is one of the simplest learning algorithms which is
relatively efficient and gives a better result when all the data sets are distinct and well separated from each other. K-means clustering defines a method of vector quantization which is used for cluster analysis. K-means clustering is one of the simplest learning algorithms which is relatively efficient and gives a better result when all the data sets are distinct and well separated from each other. Idx=k-means(X,K),performs means clustering to partition te observations of the n-by-p data matrix X into K clusters, and returns an n-by-1 vector containing cluster indices of each observation .Rows of X correspond to pints and columns correspond to variables. For example, specify the cosine distance, the number of times to repeat the clustering using new initial values or to use parallel computing.

C-means algorithm:
It is a method of clustering which allows one piece of data into two or more clusters. Clustering or cluster analysis involves assigning data point to clusters which also requires homogenous classes. Clustering can be identified by similarity measures. The similarity measures include distance, connectivity and intensity. Based on the data and application different similarities can be chosen.

Receiver operating characteristics:
Receiver operating characteristics (ROC) is a graphical plot which illustrates the performance of binary system. The curve is created by plotting the true positive rate against the false positive rate. True positive rate (TPR) is also called as sensitivity. The false positive rate (FPR) is known as fallout. ROC is also known as relative operating characteristics curve because it is a comparison of two operating characteristics (TPR, FPR) as the criteria changes.

Pre-Processing
The quality of the images (PET/CT) images is affected by the contrast variation and noise. The processing techniques such as clahe and wiener filtering are applied to reduce the disturbances without affecting the image details.

ROI Extraction
It is necessary to extract lung regions from the surrounding anatomical parts. This implements and removes all the external structures and the internal parts of lungs like blood vessels, bronchi are eliminated except the lung lobes. ROI extraction is considered for accurate extraction of lung ROI. Feature extraction determines the intensity values between normal and abnormal images. The presence of lung cancer drastically changes the appearance of the texture of the lung. These texture features play a vital role in medical imaging for automatic classification between normal and abnormal images.

Classification
Classification of the PET/CT images is mainly require to identify the presence of lung cancer. This method is preferred for classifying the regions of normal and abnormal

\[ \text{Mean} = \mu = \sum_{i=0}^{G-1} i \cdot P(i) \]

\[ \text{Variance} = \sigma^2 = \sum_{i=0}^{G-1} (i - \mu)^2 \cdot P(i) \]

\[ \text{standard deviation}, \sigma = \sqrt{\sum_{i=0}^{G-1} (i - \mu)^2 \cdot P(i)} \]

skewness, \(\gamma_1 = \sigma^{-3} = \sum_{i=0}^{G-1} (i - \mu)^3 \cdot P(i)\)

kurtosis, \(\gamma_2 = \sigma^{-4} = \sum_{i=0}^{G-1} (i - \mu)^4 \cdot P(i)\)

\[ \text{Entropy}, h = -\sum_{i=0}^{G-1} P(\hat{i}) \cdot \log_2[ P(\hat{i}) ] \]

Fig1.1 is an image with affected cancer cells, now that image is undergoing to various analysis to predict the stage of the cancer using several algorithm such as c means k means and roc curve. First c means followed by k means and roc will be applied on the input image for different analysis.

IV. RESULTS

C-Means

C-Means Output

This image determines about the small cell lung carcinoma with the initial stage which is detected using computed tomography (CT), the cells with white spot are affected cells this image will be used to predict stage of the cancer using different algorithm. K means has clustering advantage but difficult to predict K value and different initial partitions can result in different final clusters and its does not work well with clusters (in the original data) of different size and different density. C means algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance cluster center and data point.

C-Means Output Com
Fig. 1.3 shows the output of the c mean algorithm of cancer affected image in this figure the cancer affected cell may be viewed clearly when compared to the previous image, using mat lab tools the complementary of the last image fig1.2 have be taken to predict the cells clearly, the black spot region are the cancer affected cells.

Figure 4. C-means
Fig 1.4 shows the analysis of C means algorithm which is used to predict the cancer affected cells clearly, with in the lung cells the black spot regions are the cells which are affected by the cancer, with fuzzy c means algorithm, clustering allows one piece of data belong to two or more cluster and the centroid of the cluster is computed as being the mean of all the points.

Figure 5. K-Means analysis
Fig. 1.5 show the output of K means algorithm, K means clustering advantage, if variables are huge then k means most of the times computationally faster than hierarchical clustering anf K means produce tighter cluster than hierarchical clustering especially if the cluster are globular.

Figure 6. K-means output image 2
Fig1.6 shows the analysis of the k means algorithm in which huge dataset will be solved easily, with pre clustering reducing the space into disjoint smaller sub spaces. The white spot can be identified as the cancer affected cells with the lung encircled region.

Figure 7. K-means analysis
Fig. 1.7 shows the analysis of the cancer image

Figure 8. Receiver operating characteristics

Figure 9. (Tumor seed, back ground seed, tumor probability map)
Level of 250 iteration
Fig. 3
Level set with alpha value 0.5
Fig. 4

Figure 10. Region of operating Characteristics

V. CONCLUSION & FUTURE SCOPE

Developing a system is feasible if and only if it is beneficial and removes all the drawbacks of the existing system and also enhances the way of operation to make the operation perform easier. Lung cancer diagnosis is a well-fitting health-related application of this technology, which is of utmost importance in the health sector, because lung cancer has the highest death rate among all cancer types, and it brings a high yearly global burden. Our basic aim is to develop this system to improve its functionality and removes all the drawback of the existing system. Cancer is termed as the unwanted growth of tissues in the cells. Where the treatment to cure any type of cancer takes a long process and the symptoms may or may not occur in a short period. It depends upon each individual therefore it is difficult to diagnose or to cure after the disease has reached its extent. The process involved in treating cancer involves Radiation, Chemotherapy, and Drugs. The person has to undergo radiation and continuous drugs. To overcome these diagnostic properties such as Time Consuming, Risk of radiation, Pain bearing and high cost this system is developed. Cancer is a leading cause of death worldwide. In 2012, cancer accounted for 8.2 million deaths, and number of deaths is projected to increase to over 13 million in 2030. Early detection is essential to improve successful treatment and
reduce cancer mortality and cancer screening in the asymptomatic general population might be a particularly promising approach to achieve this goal. However, only few cancer screening programs are widely used. For most deadly cancers, such as pancreatic or gastric cancer, no reliable population-based screening exists, and for other common malignancies, like breast or colorectal cancer, there is large potential for improving currently used screening methods.

VI. REFERENCES


