

AN EFFICIENT MODE DETECTION TECHNIQUE OF PULMONARY NODULE IN LUNG CANCER

ARUN B. MATHEWS¹, M.K. JEYAKUMAR²

¹Research Scholar, Department of Computer Science, Noorul Islam Centre for Higher Education, Tamil Nadu, India, e-mail: marthomat@gmail.com, (m): +919447333445.

²Professor, Department of Computer Applications, Noorul Islam Centre for Higher Education, Kumara coil, Tamil Nadu, India, E-mail: jeyakumarmk@yahoo.com, (m): +919486856115

Received: 21.07.18, Revised: 21.08.18, Accepted: 21.09.18

ABSTRACT

Lung cancer is a type of cancer that begins in the lungs. Lung cancer is the leading cause of cancer deaths in many countries, among both men and women. People who smoke have the greatest risk of lung cancer, though lung cancer can also occur in people who have never smoked. The general prognosis of lung cancer is poor because doctors tend not to find the disease until it is at an advanced stage. Tumors can be benign or malignant. Benign tumors usually can be removed and do not spread to other parts of the body. Malignant tumors enter into the bloodstream or lymphatic system and then spread to other sites in the body. Therefore it is very important to find the cancer affected lung region in the earlier stage. By using Linear Iterative Clustering the noise is removed and the color base is provided. In the segmentation process the cancer affected left bottom area is detected and segmented by using Adjustable surface normal overlapping and in classification Advanced Grey Wolf Optimization is used. In this research paper performance output is calculated in the cancer affected area with 100% precision, sensitivity and with best accuracy.

Keywords: Improved Simple Linear Iterative Clustering (ISLIC), Histogram equalization, Adjustable surface normal overlap, A-GWO

INTRODUCTION

Lung cancer is one of the leading causes of death in USA, Europe and India. Surgery, radiation therapy, and chemotherapy are used in the treatment of lung carcinoma. In spite of that, the five-year survival rate for all stages combined is only 14%. However, early detection helps significantly, it is reported that the survival rate for early-stage localized cancer (stage I) is 49%. CT is considered to be the most accurate imaging modality available for early detection and diagnosis of lung cancer. It allows detecting pathological deposits as small as 1mm in diameter. These deposits are called lung nodules. Lymph node involvement from metastasis is particularly a relevant prognostic factor in lung cancer. The same as with tumor size, the greater the number of lymph nodes affected correlates with their spread to other regions and a poorer prognosis. The effort put into studying lymph nodes obtained through surgery using new techniques is to optimize the yield of the specimens and to increase the possibility of a cure by establishing a treatment in line with disease staging. As per the statistics of American Cancer Society, lung cancer is the major cause of the cancer related death in USA. The estimates of new cases and deaths in 2012 are 2, 26,160 and 1, 60,340 respectively. In 2013 the estimated new cases and deaths are 2, 28,190 and 1, 59,480 [1] respectively. In India approximately 63,000 lung cancer cases are reported every year. The image processing

mechanism is widely used in several areas for improving the earlier detection of cancer nodules. Medical images play a very important role in detection and diagnosis of numerous diseases. Here in the preprocessing step the Simple Linear Iterative Clustering method is used which produces a regular lattice and cost low computational complexity, which are the desirable properties for the subsequent applications.

Literature Survey

In cancerous cells the DNA which is present in every cell is not repaired after damage rather these cells grow and form new abnormal cells[1][4][12]. Among many other cancers, Lung cancer is an aggressive and heterogeneous disease and is the leading cause of cancer related deaths resulting in more deaths than other cancers combined. When a person has lung cancer, then they have abnormal cells[3][5] that cluster together to form a tumor (nodule). Not all tumors are cancerous[9][13]. These non-cancerous tumors are called benign nodules. The other cancerous nodules that grow without order, control and obliterate the healthy lung tissues around them are called malignant nodules[6][8]. Lung cancer is classified into two types[10][18] (i) Small Cell Lung Cancer (SCLC) which constitutes 10% to 15% of all lung cancers and (ii) Non-Small Cell Lung Cancer (NSCLC) which shares 80% to 90% of lung cancers. A Computer Aided Detection System (CAD)[12][14][9]

is one of the principal research streams in medical imaging and diagnostic radiology. A well-developed CAD helps in processing image for detection and extraction of abnormalities and also aids in classification of image features between normal and abnormal [15][19]. A CAD system is instrumental in reducing the number of false negative diagnosis [3][22]. The success of a CAD system is measured in terms of accuracy in diagnosis, speed and its degree of automation. Computer aided diagnosis [23][25] based on artificial neural network is used in the classification of the lung cancer [1][6][19]. The features used for the classification are the area, perimeter and the shape. The maximum classification obtained is 90%. For the classification few methods based on the content based Image Retrieval (CBIR) [11][17][21] have been reported. An open source pulmonary nodule image retrieval framework [2] is proposed. Here the system extracts the images of the individual nodules from the LIDC collection and calculates the Haralick co-occurrence [23], Gabor filters and Markov random field features [16][27] of the nodules. The distance measure is used for the retrieval such as the Euclidean, Manhattan and Chebychev. The maximum retrieval rate obtained is to be 88%. A CBIR [3][16][19] system is used in the mammogram images. The shape and margin features are extracted from the images and the retrieval is done by the Euclidean distance. Here the average retrieval rate is about 90%. Few methods based on fuzzy logic have been reported in the classification problems. A classifier has been developed for extracting the fuzzy rules from the texture segmented regions from the HRCT images [4][14][16] of lung cancer patients. A Fuzzy Inference System (FIS) has been constructed starting from the feature extraction procedure applied on the overlapping regions [17][23] from the same organs and deriving simple if-then rules for the classification. Here the proposed method has been tested on 138 regions extracted from the CT scan images acquired from the patients with lung cancer and obtained a sensitivity of 94%. Another type called Fuzzy [21][27] bean based classifier, which is a supervised learning method [8][20][24] and with proper optimization scheme promising result were covered. Differential evolution algorithm [14][18][28] was used in optimizing the required parameters. Classifier was applied to diagnosis of liver dis order with 73.9% classification accuracy. For detection few methods based on the genetic algorithm [5][17] has been reported. A new approach has been introduced by using the genetic algorithm for improving semantic image segmentation system [21][26]. Experiments were conducted and evaluated extensively on benchmark data set for segmentation of 21 objects. For estimating the right number of segments a new method has been used for the automatic segmentation [11][22][30] of human normal and abnormal images. In this method [28] it

has been tried to change the FCM method as a fitness function for the combination of it in genetic algorithm to do the image segmentation more accurately. The results shows that the proposed method [24][29] has a significant improvement in the accuracy of image segmentation in comparison to the similar methods.

Image Acquisition

To design a CAD for lung nodule detection system, CT images are well preferred as offers visualization of low contrast or small volume nodules by diminishing the slice thickness. Lung CT images can be acquired from publicly available databases namely Early Lung Cancer Action Program (ELCAP), Lung Image Database Consortium (LIDC) or Medical Image Database. As per the literature there are other imaging techniques and many private databases used by the researchers, which are obtained from private hospitals.

Block Diagram Explanation

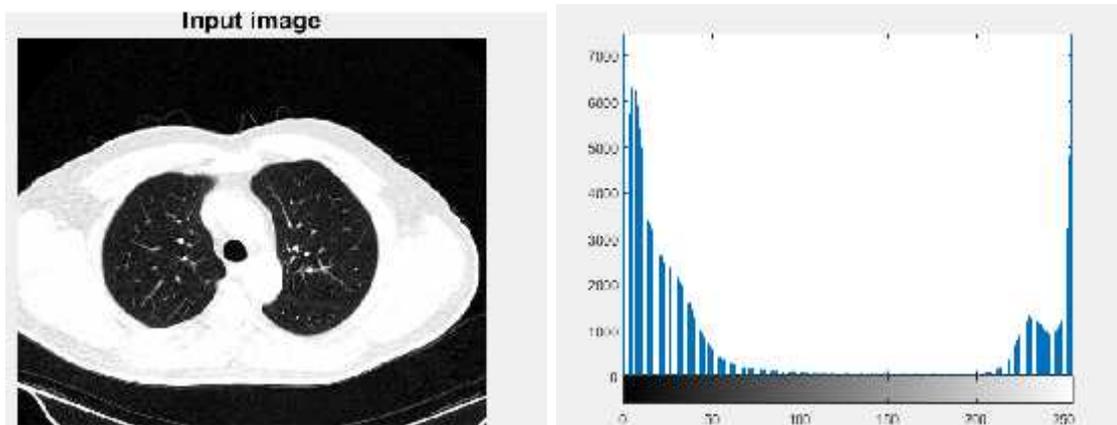
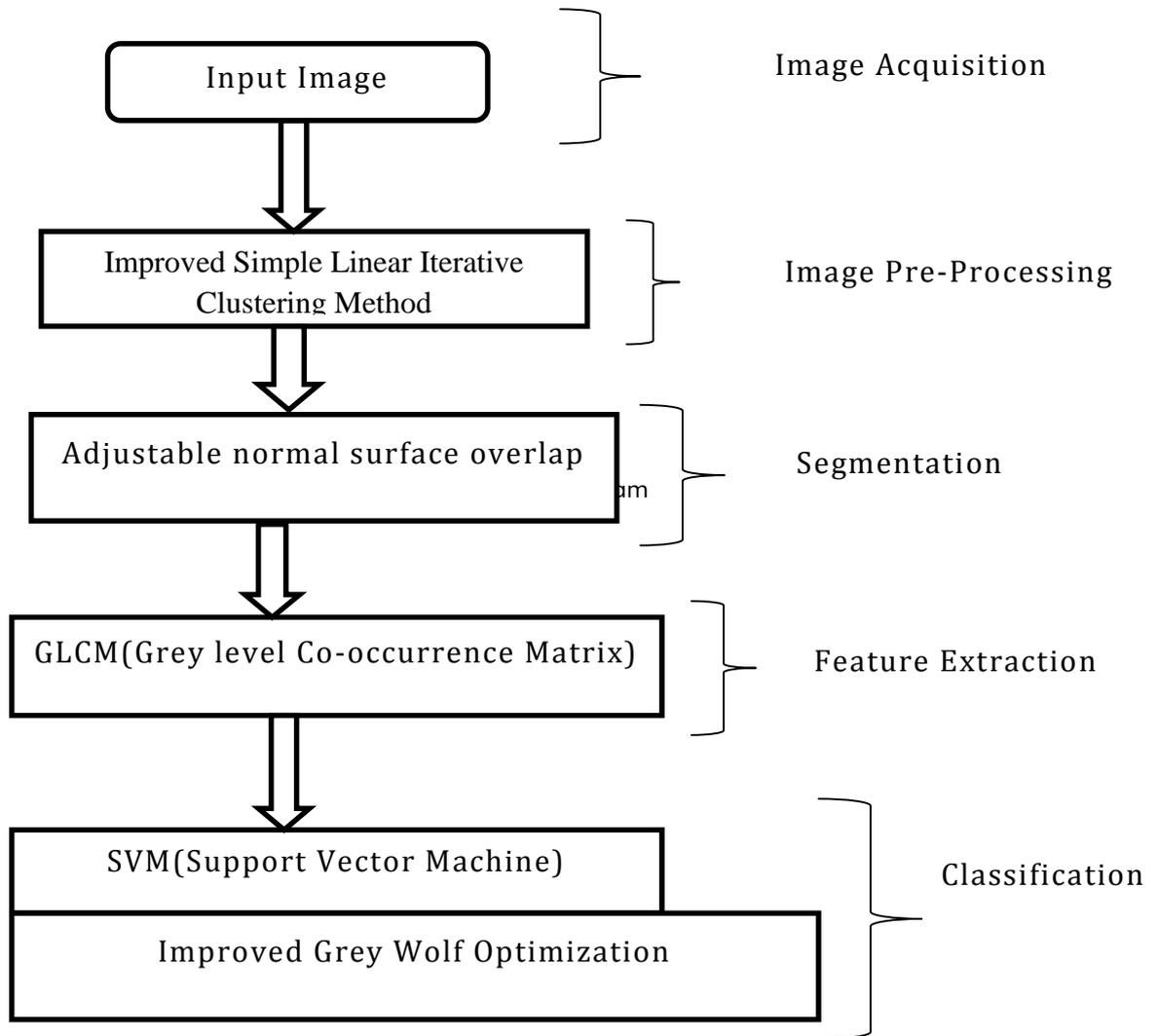
The below given figure.1 describes the block diagram of Pulmonary node Detection of Lung Cancer. The first step consists of the image acquisition. Here the input image is taken as an input from the Dataset. Next step is the image Pre-Processing. Here the noise is removed by using Simple Linear Iterative Clustering method with the help of Histogram Equalization. In the next step Segmentation process is done. Here in the Lung, it detects if there is any cancer affected juxta vascular is there or not is segmented. The next step is the Feature Extraction. Here Adjustable Surface normal overlap is used. Here performance output is calculated. And then the classification is done by Improved Grey Wolf Optimization is done with the help of SVM.

Preprocessing

Pre-processing is a common name for operations with images at the lowest level of abstraction both input and output are intensity images. These iconic images are of the same kind as the original data captured by the sensor, with an intensity image usually represented by a matrix of image function values (brightness). The aim of pre-processing is an improvement of the image data that suppresses unwilling distortions or enhances some image features important for further processing, although geometric transformations of images are classified among pre-processing methods here since similar techniques are used. Since the Digital Imaging and Communications in Medicine (DICOM) images with noise uses the reconstruction method to enhance resolution denoising these original images is necessary. In the below given figure 2(a) the image is taken as an input. It consists of noise, external noise and so on. The histogram is calculated in the fig

2(b). Here all the intensity is not equal. More variation can be seen in the histogram. Therefore it is important to remove the noise.

Block Diagram



a) Input Image

b) Histogram Graph

Figure 2: Input Images

Preprocessing of an image refers to the procedure of enhancing the quality and interpretability of the input lung image by reducing the noise and unwanted artifact.

Histogram Equalization (HE)

Histogram Equalization (HE) is a very popular technique for improving image contrast. Previously, this image contrast enhancement technique has been applied to the image of man and the natural scenery with good results. HE is a process that changing the distribution of gray scale value in an image so that it becomes uniform. The goal is to obtain a uniform spread of the histogram so that each gray scale value has a relatively equal number of pixels. Flattening histogram obtained by changing the gray scale of a pixel (r) with the new one gray scale (s) with a transformation function T. Mathematically can be written by the equation: $s = T(r)$. r can be recovered from s with inverse transformation is shown in the equation (1)

$$y = T^{-1}(s) \tag{1}$$

Where $0 < s < 1$. The equation used to calculate the HE is shown in the equation (2):

$$k_p = \text{round} \left(\frac{C_i(r-1)}{w, k} \right) \tag{2}$$

Where:

C_i = the cumulative distribution of ith grayscale from original image

round = rounding to the nearest value

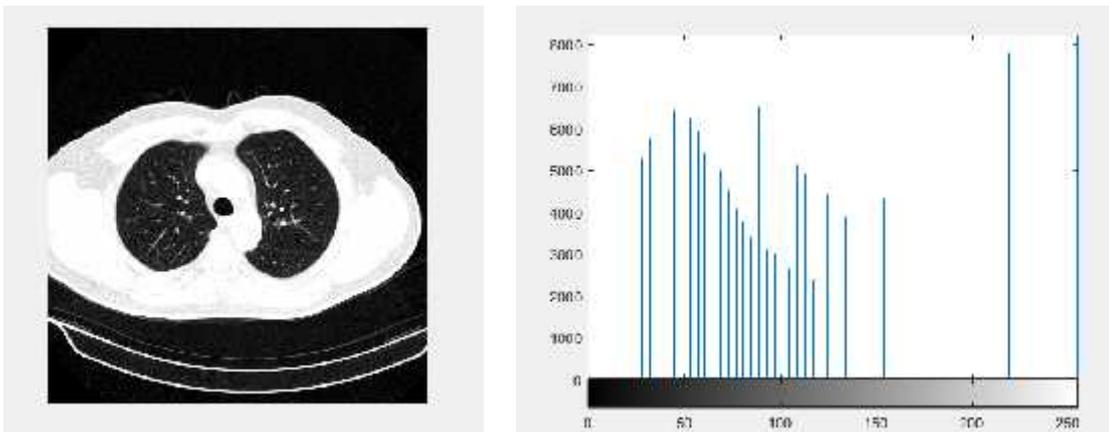
K_0 = gray level value from histogram equalization

w = width of the image

l = height of the image

This method is also useful for images with both background and foreground are bright or both dark. Particularly, this method provides a better view of the x-ray image in the biomedical world, producing clear and detail of the images.

In the below given figure 3(a) the pre-processing is done and the histogram equalization is calculated in the fig 3(b). Here the intensity or the pixel is of mostly



a) Preprocessed Image

b) Histogram Graph

equal.

Figure 3:Preprocessed image

Improved simple Linear Iterative Clustering Method

Simple Linear Iterative Clustering (SLIC) algorithm is increasingly applied to different kinds of image processing because of its excellent perceptually meaningful characteristics. For superpixels to be useful they must be fast, easy to use, and produce high quality segmentations. Among all the superpixel algorithms, the Simple Linear Iterative Clustering (SLIC) method is widely adopted due to its practicality. However, the resultant superpixels sometimes do not well adhere to the edges. Here an Improved Simple Linear Iterative Clustering method is used to solve the problem. For many vision tasks, compact and highly uniform superpixels that respect image boundaries, such as those generated by SLIC are desirable. Our approach generates superpixels by clustering pixels based on their color similarity and proximity in the image plane. This is done in the five-dimensional [labxy] space, where [lab] is the pixel color vector in

CIELAB color space, which is widely considered as perceptually uniform for small color distances, and xy is the pixel position. While the maximum possible distance between two colors in the CIELAB space is limited, the spatial distance in the xy plane depends on the image size. It is not possible to simply use the Euclidean distance in this 5D space without normalizing the spatial distances. In order to cluster pixels in this 5D space, we therefore introduce a new distance measure that considers superpixel size. Using it, we enforce color similarity as well as pixel proximity in this 5D space such that the expected cluster sizes and their spatial extent are approximately equal.

Distance Measure

SLIC takes a desired number of approximately equally-sized superpixels K as input. So each superpixels will have approximately N/K pixels. Hence, for equally sized superpixels, there would be

a superpixel center at every grid interval $S = \sqrt{(N/K)K}$ superpixel cluster centers $C_k = [l_k, a_k, b_k, x_k, y_k]$ with $k = [1, K]$ at regular grid intervals S are chosen. Since the spatial extent of any cluster is approximately S^2 , it can be assumed that pixels associated with this cluster lie within $2S \times 2S$ area around the superpixel center in the xy plane. Euclidean distances in CIELAB colorspace are meaningful for small distances. If spatial pixel distances exceed this perceptual color distance limit, then they begin to outweigh pixel color similarities. Distance measure D_s is shown in the equation (4).

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \tag{3}$$

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \tag{4}$$

$$D_s = d_{lab} + (m / S) * d_{xy} \tag{5}$$

where D_s is the sum of the lab distance and the xy plane distance normalized by the grid interval S . A variable m is introduced in D_s allowing us to control the compactness of superpixel. The greater the value of m , the more spatial proximity is emphasized and the more compact the cluster. This value can be in the range $[1, 20]$. Authors of the algorithm have chosen $m=10$.

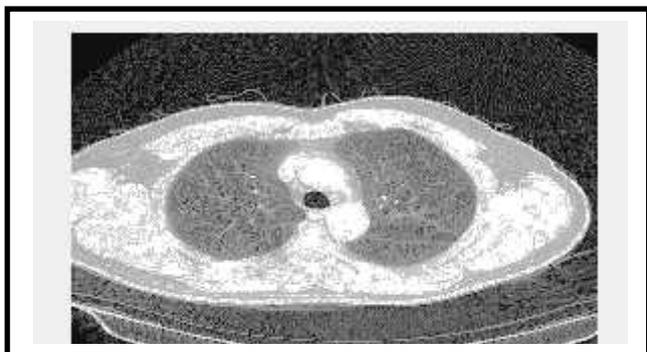
Algorithm

Simple Linear Iterative Clustering is the state of the art algorithm to segment superpixels which doesn't require much computational power. It begins by sampling K regularly spaced cluster centers and

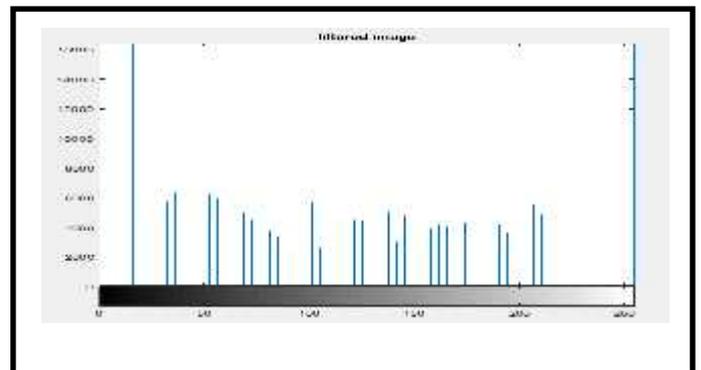
moving them to seed locations corresponding to the lowest gradient position in a 3×3 neighborhood. This is done to avoid placing them at an edge and to reduce the chances of choosing a noisy pixel. Image gradients are computed is shown in the equation (6)

$$G(x, y) = \sqrt{|I(x + 1, y) - I(x - 1, y)|^2 + |I(x, y + 1) - I(x, y - 1)|^2} \tag{6}$$

Where $I(x, y)$ is the lab vector corresponding to the pixel at position (x, y) . This takes into account both color and intensity information. Each pixel in the image is associated with the nearest cluster center whose search area overlaps this pixel. After all the pixels are associated with the nearest cluster center, a new center is computed as the average labxy vector of all the pixels belonging to the cluster. At the end of this process, a few stray labels may remain, that is, a few pixels in the vicinity of a larger segment having the same label but not connected to it. It enforces connectivity in the last step of the algorithm by relabeling disjoint segments with the labels of the largest neighbouring cluster. In the below given figure 4(a)(b) the Advanced Simple Linear Iterative Technique is applied to the pre-processed image. Here the color base is of White, Black and Grey. Based on the intensity value or the pixel (0-255) range is calculated. If the pixel value is of 0, it represents black. If the pixel value is 255, it is considered as white. In between pixel value is of grey in color. After calculating the pixel value, the mean value is calculated.



LIC output image



Histogram graph

Figure 4: Color base image by applying ILIC technique

Segmentation

Image segmentation is a technique of partitioning an image in to multiple regions of the lung refers to the process of extracting the lung region from other anatomical parts of the body in chest CT images. This process plays a vital role in nodule detection by improving accuracy and precision that helps in early diagnosis of lung cancer. An accurate segmentation will reduce the computational cost of detection. In CT image of a lung, the anatomical structures that may require segmentation are lungs themselves, the airways, the vessels, lung lobes. Segmentation is a

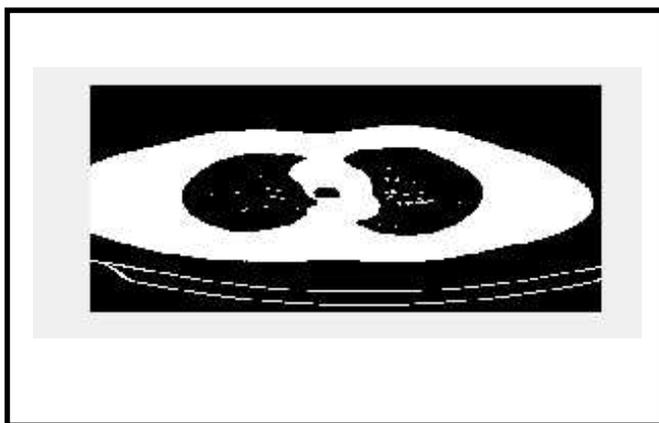
complex activity due to pulmonary structures of similar densities namely arteries, veins, bronchioles and different scanners used. Numerous publications have addressed the issue of segmentation of lungs. The proposed techniques are measures with respect to accuracy, processing time and level of automation.

Adjustable Surface Normal Overlap

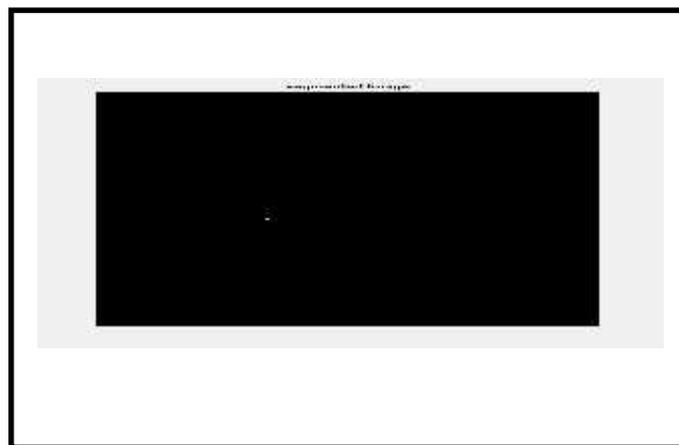
We developed a novel computer-aided detection (CAD) algorithm called the surface normal overlap method that we applied to colonic polyp detection and lung nodule detection in helical computed

tomography (CT) images. It is able to achieve 100% sensitivity. The adjustable surface normal overlap step is critical for detecting lesions. Generally speaking, both colonic polyps and lung nodules tend to have some convex regions on their surfaces and thus, the inward pointing surface normal vectors, near these features tend to intersect or nearly intersect within the tissue. Pulmonary vessels in the lungs and haustral folds in the colon also have convex surfaces, but since they have a dominant curvature along a single direction (as opposed to high curvature in two directions as is common on the surfaces of polyps and nodules), the score for vessels and folds is generally less than that for nodules and polyps. Providing robustness to variations from perfectly spherical objects is critical to the success of

this algorithm in real patient data. Our algorithm provides robustness both in the radial direction (objects with non-constant distance from surface points to center) and in the transverse direction (objects with non-uniform magnitude of curvature). Robustness in the radial direction is provided by the fact that normal vectors can intersect at different distances from the surface, thus allowing many non-spherical but roughly globular objects to have a significant response. In the below given figure 5(a)(b), the left bottom cancer affected area is found in the lung. Clear detection of cancer affected area in the lung is very important to find out in the earlier detection. Here clearly the cancer affected lung area is detected. The segmented image is shown in the figure 5(b).



Segmentation Image



Juxta Vascular region

Figure 5: Segmented image

Feature Extraction

Feature extraction describes the relevant shape information contained in a pattern so that the task of classifying the pattern is made easy by a formal procedure. In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. The main goal of feature extraction is to obtain the most relevant information from the original data and represent that information in a lower dimensionality space. Feature extraction is an important step in the construction of any pattern classification and aims at the extraction of the relevant information that characterizes each class. In this process relevant features are extracted from objects/ alphabets to form feature vectors. These feature vectors are then used by classifiers to recognize the input unit with target output unit.

GLCM (Gray Level Co-Occurrence Matrix)

Gray Level Co-Occurrence Matrix (GLCM) has proved to be a popular statistical method of extracting textural feature from images. Gray Level Co-occurrence Matrix is a tabulation of how often different combinations of pixel brightness values occur in an image. GLCM contains the information about the positions of pixel having similar gray level values. GLCM calculation units receive pairs of gray level values as input. A GLCM is a matrix where the

number of rows and columns is equal to the number of gray levels, G , in the image. The matrix element $P(i, j | \Delta x, \Delta y)$ is the relative frequency with which two pixels, separated by a pixel distance $(\Delta x, \Delta y)$, occur within a given neighborhood, one with intensity 'i' and the other with intensity 'j'. The matrix element $P(i, j | d, \theta)$ contains the second order statistical probability values for changes between gray levels 'i' and 'j' at a particular displacement distance d and at a particular angle (θ) . Using a large number of intensity levels G implies storing a lot of temporary data, i.e. a $G \times G$ matrix for each combination of $(\Delta x, \Delta y)$ or (d, θ) . Due to their large dimensionality, the GLCM's are very sensitive to the size of the texture samples on which they are estimated. Thus, the number of gray levels is often reduced. GLCM matrix formulation can be explained with the example illustrated in fig 2.1 for four different gray levels. Here one pixel offset is used (a reference pixel and its immediate neighbour). If the window is large enough, using a larger offset is possible. The top left cell will be filled with the number of times the combination 0,0 occurs, i.e. how many time within the image area a pixel with grey level 0 (neighbour pixel) falls to the right of another pixel with grey level 0 (reference pixel).

Angular Second Moment

Angular Second Moment is also known as Uniformity or Energy. It is the sum of squares of entries in the GLCM. Angular Second Moment measures the image homogeneity. Angular Second Moment is high when image has very good homogeneity or when pixels are very similar is shown in the equation (7)

$$ASM = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_{i,j}^2 \tag{7}$$

Where i, j are the spatial coordinates of the function p(i, j), Ng is gray tone.

Entropy

Entropy shows the amount of information of the image that is needed for the image compression. Entropy measures the loss of information or message in a transmitted signal and also measures the image information, is shown in the equation (8)

$$Entropy = - \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_{i,j} \log_2 p_{i,j} \tag{8}$$

Correlation

Correlation measures the linear dependency of grey levels of neighboring pixels. Digital Image Correlation is an optical method that employs tracking & image registration techniques for accurate 2D and 3D measurements of changes in images. This is often used to measure deformation, displacement, strain and optical flow, but it is widely applied in many areas of science and engineering. One very common application is for measuring the motion of an optical mouse, is shown in the equation (9)

$$Correlation = \frac{\sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} (I_{i,j})P(I_{i,j}) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{9}$$

Classification

After the candidate nodules are detected and false positives are reduced, the resultant set of nodules must be classified to be benign or malignant ones. Most of the pulmonary nodules are benign but they may represent an early stage of lung cancer. Early detection of a malignant (cancerous) nodule increases the survival rate of the diseased. Many Computer Aided Diagnosis (CADx) systems have been developed that differentiate malignant lesions from benign ones, and also gives the insight into the proximity of detected nodule to be malignant.

Svm (Support Vector Machine)

Support Vector Machines (SVM) are supervised learning models with associated learning algorithms that analyze data used for classification. A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. The main contributions are as follows:

(1) In order to improve the generalization ability of the SVM model, the GWO algorithm is used to optimize the key parameters of the SVM.

(2) The initial population of the original GWO algorithm is generated randomly, thereby improving the convergence speed and better solution ability of the algorithm.

(3) The proposed Improved GWO-SVM model can get better classification accuracy, and the recognition accuracy is better than support vector machine model based on the original GWO, SVM model based on grid search and ELM or the other methods. And the proposed method is achieved precisely classify to the color difference classification problems of printing and dyeing products.

The problem of data classification in SVM can be described as the system generates a hyperplane and continuously adjusts the plane until the point belonging to the two types in data set is located on the two side of the hyperplane, and the distance about the point of sample set to the classification plane is as large as possible. In the sample space, the classification of hyperplanes is shown in the equation (10)

$$\omega^T x + b = 0 \tag{10}$$

where, (; ; = ... 12 d) is a normal vector, which determines the direction of the hyperplane; b is a displacement term, which determines the distance between the hyperplane and the origin. Suppose the hyperplane can correctly classify the training samples.

That is, for (x, y) D, ii€ if y 1 i = + , is there x b 0 T i + > ; If, y 1 i = - then there is x b 0 T i + < . Order is shown in the equation (11)

$$\begin{cases} \omega^T x_i + b \geq 1, y_i = +1 \\ \omega^T x_i + b \leq -1, y_i = -1 \end{cases} \tag{11}$$

These training samples are nearest to the hyperplane that are called "support vectors", and the sum of the distance between the two dissimilar support vectors to the hyperplane can be expressed as:

$$M = \frac{2}{\|\omega\|} \tag{12}$$

And it is called "margin". In order to find the partition hyperplane with the maximum margin, namely, it is to find the constraint parameters and b, only to make -1 maximum, which is equivalent to make the 2 minimal. For support vector machines in linear and non-separable cases, is shown in the equation (13)

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=0}^m \mu_i \tag{13}$$

$$s.t \ y_i(\omega^T x_i + b) \geq 1 - \mu_i, \ \mu_i \geq 0, \ i=1,2,..m \tag{14}$$

Where C is a predefined balance parameter to balance empirical risk and structural risk.

Improved grey wolf optimization algorithm

GWO algorithm was simulation about social hierarchy and group hunting behavior of grey wolf population, through tracking, encircling, hunting, attacking of wolf population and other processes to achieve the optimization of intelligent algorithm. The Improved GWO algorithm has the advantages of simple principle, few parameters to be adjusted, easily to implement and strong capability to global search. The GWO algorithm is used to optimize the continuous function, the number of grey wolves is set to N, and the solution search space is set to d dimension. The location of the i wolf is defined in the d dimension space, and is the current best individual in the grey wolf population, and are individual grey wolves whose fitness value rank in second and third, respectively. The global optimal solution of the optimization problem is the location of the final prey.

In the GWO algorithm, it is necessary to determine the distance between the individual and the prey is shown in the equation (15)

$$D = |C \cdot X_p(t) - X(t)| \tag{15}$$

Where X (t p) represents the location of the prey at the time of t-th generation;

X(t) represents the location of the grey wolf individual at the time of t-th generation; Constant C is the swing factor is shown in the equation (16)

$$C = 2r_1 \tag{16}$$

Where, r1 is a random number from 0 to 1. The position updating formula of grey wolf is shown in the equation (17)

$$X(t+1) = X_p(t) - A.D \tag{17}$$

Where, A is the convergence factor is shown in the equation (18)

$$A = 2a_2 - a \tag{18}$$

Where r2 is a random number from 0 to 1, while, a linearly decreases from 2 to 0 as the number of iterations increases.

In order to simulate the hunting behavior of the grey wolf, suppose wolf, wolf and wolf have a better understanding to the location of the prey, so the grey wolf population can use the position of the three to decide the location of the prey. When the wolves get the prey location information, leader wolf command and to hunt. The grey wolf population to update its position according to the location information of , and wolf are is shown in the equation

$$D_{\alpha} = |C_1 \cdot X_{\alpha}(t) - X(t)| \tag{19}$$

$$D_{\beta} = |C_2 \cdot X_{\beta}(t) - X(t)| \tag{20}$$

$$D_{\delta} = |C_3 \cdot X_{\delta}(t) - X(t)| \tag{21}$$

$$X_1 = X_{\alpha} - A_1 \cdot D_{\alpha} \tag{22}$$

$$X_2 = X_{\beta} - A_2 \cdot D_{\beta} \tag{23}$$

$$X_3 = X_{\delta} - A_3 \cdot D_{\delta} \tag{24}$$

$$X_p(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{25}$$

Where, according to the formula (19)–(224), the distance between the individual in the population and the wolf, wolf and wolf is calculated, and the distance and direction of the individual moving to the prey are determined by the formula (25). The optimization process of GWO algorithm is that a group of gray wolves in the search space randomly is generated, in the process of evolution, the position of the prey is evaluated and located by , and wolf, and the rest of individual in the group to calculate the distance between themselves and their prey, they complete the omni-directional proximity, encirclement, attack, and so on, and finally capture the prey. Here a dataset consisting of 40 images were taken into consideration. Each image is made to run and the performance output is calculated. These values are linked with the SVM classifier and the confusion matrix is got as a result. The confusion matrix is shown in the below given figure In the below given figure 6, the performance is calculated and the confusion matrix is calculated by using GLCM. The values are linked with the SVM classifier. The values such as the True Positive (TP) is of 20, True Negative (TN) is of 0, False positive (FP) is of 1 and the False Negative (FN) is of 19.

Conclusion

As earlier the cancer affected area is detected, the treatment can be taken and cured. Now a days, more techniques were found out in the lung cancer detection, in this research the cancer affected area is clearly detected. Best performance is calculated in the output with 97.5% accuracy, 100% sensitivity and precision, 95% specificity. Till now, in the before research 95.4% accuracy is only calculated. But in this research 97.5% accuracy is got in the output. Therefore this research will be very usefull in the earlier detection of cancer in the lung nodule with best accuracy. Future research aims to develop hybrid artificial intelligence techniques employing the following genetic algorithms; hybrid of fuzzy inference system (FIS), Ant Colony Optimization (ACO). Looking forward to conduct further experiments and then to test this algorithm on a large dataset having different pathologies. In-depth technical analysis, researches, investigations and implementations carried out in the area of Detection of pulmonary nodule in lung cancer have promoted the scope for further research. Although the obtained results demonstrated here provide the effectiveness of the proposed methods, there is still a lot of scope for improvement in the earlier detection of Lung Cancer and their types. Future work will propose the implementation of this approach and at the end, we

can better give global diagnostics for any given method.



Figure 6: Confusion Matrix

References

1. Abdulla, AA and shaharum, "Lung cancer cell classification methods using Artificial Neural Network", Information Engineering Letters, SM 2012, Voi. 2, pp. 50-58
2. Lei Liu , Guo-qing Liao , Pei He , Hong Zhu, Peng-hui Liu, Yi-mei Qu , Xiao-ming Song, "Detection of circulating cancer cells in lung cancer patients with a panel of marker genes"Elsevier 2008
3. Hiroaki Nomori, MD, "Sentinel Node Mapping in Lung Cancer: The Japanese Experience" Elsevier 2009
4. Jose M. GalbisCaravajal, Antonio Cremades Mira Angel Zuniga Cabrera, "Sentinel Lymph Node in Pulmonary Carcinoma", Elsevier 2014
5. Weixing Wang, Shuguang Wu, "A Study on Lung Cancer Detection by Image Processing", IEEE publication 2008
6. JunjieZhanga, Yong Xiaa, Hengfei Cui, Yanning Zhang, "Pulmonary nodule detection in medical images", Elsevier 2018
7. Abhinao A Somnath, Devendra G Ingale, "Lung Nodule Detection Methods", International Journal of Computer Sciences and Engineering, 2006
8. Junjie Bai , Xiaojie Huang , Shubao Liu, Qi Song, and RoshniBhagalia, "Genetic Programming-based Feature Transform and Classification for the Automatic Detection of Pulmonary Nodules on Computed Tomography Images", Research Gate Publication, 2015
9. Neelima Singh and A. Asuntha, "Image Processing used for Lung Cancer Detection in Medical Images", International Science Press, 2016
10. MunimandaPremChander , M. Venkateshwara Rao, T. V. Rajinikanth, "Detection of Lung Cancer Using Digital Image Processing Techniques",International Journal of Medical Imaging publication, 2017
11. Deshpande , Dhanesh D. Lokhande , Rahul P. Mundhe, JuileeM.Ghatole, "Lung Cancer Detection with fusion of CT and MRI Images Using Image Processing", International Journal of Advanced Research in Computer Engineering, 2015
12. PreetiKatiyar, Dr. Krishna Singh, "Lung Tumor Detection and Segmentation in CT Images" International Journal of Innovations & Advancement in Computer Science, 2017.
13. K.-T. Bae, J.-S. Kim,, Y.-H. Na, "Pulmonary nodules: automated detection on CT images with morphologic matching algorithm", preliminary results, Radiology, Vol. 236, pp. 286294, 2005.
14. C.-W. Bong, H.-Y. Lam and H. Kamarulzaman, "A Novel Image Segmentation Technique for Lung Computed Tomography Images", Communications in Computer and Information Science, Vol. 295, pp. 103-112, 2012.
15. T. Chhabra, G.Dua, T. Malhotra, "Comparative Analysis of Methods to Denoise CT Scan Images", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, No. 7, pp. 3363-3369, 2013.
16. M. Ceylan, Y. Ozbay, O.-N. Ucan and E. Yildirim, "Novel method for lung segmentation on chest CT images: complex-valued arti_cial neural network with complex wavelet transform", Turk J. Elec. Eng. and Comp. Sci., Vol.18, No.4, pp. 613-623, 2010.
17. K. Devaki, V. MuraliBhaskaran and M. Mohan, "Segment Segmentation in Lung CT Images-Preliminary Results", Special Issue of International Journal on Advanced Computer Theory and Engineering (IJACTE), Vol. 2, No. 1, pp. 84-89, 2013.
18. A. El-Baz, G. Gimelfarb, R. Falk and M. Abo El-Ghar, "3D MGRF-based appearance modelling for robust segmentation of pulmonary nodules in 3D LDCT chest images, in Lung Imaging and Computer Aided

- Diagnosis”, chapter 3, pp. 5163, Taylor and Francis edition, 2011.
19. K.-Z. Faizal and V. Kavitha, “An Effective Segmentation Approach for Lung CT Images Using Histogram Thresholding with EMD Refinement, Proceedings of International Conference on Internet Computing and Information Communications”, *Advances in Intelligent Systems and Computing*, Vol. 216, pp. 483-489, 2014.
 20. A. Farag, A. Ali, J. Graham, S. Elhabian, A. Farag and R. Falk, “Feature Based Lung Nodule Classification”, *ISVC 2010, Part III*, Vol. 6455, pp. 7988, 2010.
 21. A. Farag, H. Abdelmunim, J. Graham, “Variational approach for segmentation of lung nodules”, in Proc. of the IEEE International Conference on Image Processing (ICIP 11), pp. 21572160, 2011.
 22. T. Gao, X. Sun, Y. Wang and S. Nie, “A Pulmonary Nodules Detection Method Using 3D Template Matching, *Foundations of Intelligent Systems*”, AISC, Vol. 122, pp. 625633, 2012.
 23. J. Gong, T. Gao, R.-R. Bu, X.-F. Wang and S.-D. Nie, “An Automatic Pulmonary Nodules Detection Method Using 3D Adaptive Template Matching”, *Communications in Computer and Information Science*, Vol. 461, pp. 3949, 2014.
 24. Y. Guo, C. Zhou, H.-P. Chan, A. Chughtai, J. Wei, L.-M. Hadjiiski and E.-A. Kazerooni, “Automated iterative neutrosophic lung segmentation for image analysis in thoracic computed tomography”, *Med. Phys.*, Vol.40, No. 8, pp. 081912/1-081912/11, 2013.
 25. U. Javed, M.-M. Riaz, T.-A. Cheema and H.-F. Zafar, “Detection of Lung Tumor in CE CT Images by using Weighted Support Vector Machines”, In. proc. of the 10th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 113-116, 2013.
 26. M. Mabrouk, A. Karrar and A. Sharawy, “Computer Aided Detection of Large Lung Nodules using Chest Computer Tomography Images”, *International Journal of Applied Information Systems (IJ AIS)*, Vol. 3, No. 9, pp. 12-18, 2012.
 27. H.-M. Orozco and O.-O. Villegas, “Lung Nodule Classification in CT Thorax Images using Support Vector Machines”, In proc. of the 12th Mexican International Conference on Artificial Intelligence, pp. 277- 283, 2013.
 28. H. Shao, L. Cao and Y. Liu, “A Detection Approach for Solitary Pulmonary Nodules Based on CT Images”, in Proc. of the 2nd International Conference on Computer Science and Network Technology, pp. 1253-1257, 2012.
 29. Y. Zheng, K. Steiner, T. Bauer, J. Yu, D. Shen and C. Kambhmettu, “Lung nodule growth analysis from 3D CT data with a coupled segmentation and registration framework”, in Proc. of the IEEE 11th International Conference on Computer Vision (ICCV 07), 2007.
 30. G. Niranjana and M. Ponnaivaikko “Statistical analysis of features and classifiers in identifying nodules and its staging in lung CT images “, *ARPJN Journal of Engineering and Applied Sciences*, 2016.
 31. J. C.L. Jiang, S.W. He, Y.D. Zhang, H.X. Duan, T. Huang, Y.C. Huang, G.F. Li, L.J. Ma, G.B. Zhou, Y. Cao, “Air pollution and DNA methylation alterations in lung cancer”, *Oncotarget*, vol.8, pp. 1369-1391, 2017.
 32. Theresa M.M. and Bharathi V.S. “Computer aided diagnostic (CAD) for feature extraction of lungs in chest radiograph using different transform features”, *Biomedical Research*, pp. 1–8, 2017.
 33. Shen W., Zhou M., Yang F., Yu D., Dong D., Yang C., Zang Y., Tian J. “Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition*”, vol.61, pp. 663–673, 2017.
 34. Liu K., Kang G. “Multiview convolutional neural networks for lung nodule classification”, *International Journal of Imaging Systems and Technology*, vol.27, pp. 12–22, 2017.
 35. Wu L., Cheng J.Z., Li S., Lei B., Wang T., Ni D. FUIQA: “Fetal ultrasound image quality assessment with deep convolutional networks”, *IEEE Transactions on Cybernetics* vol.47, pp. 1336–1349, 2017.
 36. Koichi Sato, Hiroaki Akamatsu et.al “Correlation between immune-related adverse events and efficacy in nonsmall cell lung cancer treated with nivolumab” *Elsevier, Lung Cancer* vol.115 pp.71–74, 2018.
 37. Shaohua Cui, Zhuoan Cheng, Wenxin Qin, Liyan Jiang “Exosomes as a liquid biopsy for lung cancer” *Elsevier*, pp. 46-51, 2018.
 38. Daichi Fujimoto et.al, “Efficacy and safety of nivolumab in previously treated patients with nonsmall cell lung cancer: A multicenter retrospective cohort study”, *Elsevier*, pp.14-20, 2018.