Journal of research in engineering and its applications
Vol. 1, Issue. 1, (2018), pp. 34-44

A Lung Cancer Segmentation Using Threshold Based Stochastic Regression Model With Gabor Filter

Ezhil Prakash. M[1], Pawan Patel[2]
Research Scholar,Yeshwantrao Chavan College of Engineering, Nagar Yuwak Shikshan Sanstha, Nagpur[1]
Assistant professor,Maharaja Surajmal Institute of Technology,New Delhi [2]

Abstract

Recently, image processing techniques are widely used in several medical areas for image improvement in earlier detection and treatment stages, where the time factor is very important to discover the abnormality issues in target images, especially in various cancer tumors such as lung cancer. Image quality and accuracy is the core factors of this research. This paper proposes the classification model must be primed or trained on a sample of units with measured features and known classes. Regression model with independent components in which the feature values are treated as covariates. Modified Gabor filter is applied here for filtration process whereas the Laplace is applied to make the filtration process more effective. The Gabor filter is a linear filter which impulse response is defined by a harmonic function multiplied by a Gaussian function. Threshold based stochastic regression model for segmentation is a direct search technique widely used for solving optimization problems based on the values of the objective function when the derivative information is unknown. This technique is mainly used to extract the abnormal portion from the classified input lungs image. The experimental results show the extracted abnormal portion of the input images.

1.INTRODUCTION

Lung cancer has become the most lethal threat to human health, which mainly caused by smoking and air pollution, and has taken 1.6 million people lives. Lung cancer is remediable at the early stage which only occupies less than 25% of the total diagnosed cases. Most of patients have incurable locally advanced or metastatic disease, their 5-year survival approximates one fourteenth of the early stage ones. Thus, early pulmonary malignant nodule detection is crucial for prolonging the patient life even than recovery. Because of the large number of affected individuals, improvements in diagnosis at an early, potentially curable stage would have a major impact on human health.

The National Lung Screening Trial (NLST) compared low-dose computed tomography (CT) to standard chest radiography (CXR) across three annual screens. There were 309 deaths per 100,000 person-years in the CXR group and 247 deaths per 100,000 person-years in the low dose CT group representing a 20% relative reduction in lung cancer mortality within the CT arm compared to CXR3. Of the CT-detected lung cancers, 58% had a prior nodule-positive screen that was not determined to be lung cancer (i.e., nodule positive/cancer negative). An important issue that arose from these studies was the high detection of 4 to 12 mm diameter indeterminate pulmonary nodules (IPNs) that were “suspicious”, but not diagnosed as cancer. Of the IPNs, 96.4% were not diagnosed as, or did not develop into, cancers during the screening period or follow-up. Hence, only 3.6% of IPNs were nascent cancers5. Over diagnosis as "suspicious" is harmful because of patient anxiety, and a subsequent work-up or treatment of these cancers can incur unnecessary costs and morbidity for a condition that may pose no threat if not otherwise treated.

Computed Tomography (CT) is one of the best imaging techniques for soft tissue imaging behind bone structures. A CT image has high spatial resolution, minimizes artifacts and provides excellent visualization of anatomical features for analysis. In earlier stages, lung cancer is most commonly noticeable in CT images radio logically as a non-calcified solitary pulmonary nodule. Nodules are visible as low contrast white, approximately spherical objects within the lung fields.

Image segmentation facilitates delineation of anatomical structures and other regions of interest. Segmentation is one of the most difficult tasks in image processing and determines the outcome of analysis and evaluation of pathological regions. Neural networks are used widely for classification of image analysis in several biomedical systems. The aim of the proposed system is to predict lung tumor through efficient segmentation and neural network classification. Nodules are caused by a variety of disorders, including neoplastic, infectious, inammatory, vascular, and congenital abnormalities. Features on CT which aid in differentiating benign and malignant nodules include size, morphology, and internal characteristics.

Most of the automated nodule detection methods are based on the thoracic CT scan images, neural network based approach, sequential CT images, genetic cellular neural network and 3D template matching technique, mass screening with mobile spiral CT scanner, low dose CT scan images, model based detection in CT scan images, lung nodule classification approach in CT scan images, lung nodule detection using profile matching and back propagation neural network techniques, accuracy of positron emission tomography for nodule detection, local density algorithm and template matching technique on helical CT scan images. These methods correctly detect the nodules in images but some are robust against noise and small nodules in an image can be detected by few and if detected then blood vessels are also detected by them along with the nodules. The evaluation of an incidental nodule to determine whether it reects malignant disease can lead to a long and costly workup. The efforts to detect early lung cancer have led to lung cancer screening with CT in at-risk populations, which is associated with the discovery of even larger numbers of nodules.

2.LITERATURE SURVEY

Hongbo Zhua, Chun-Hyok Paka;b, Chunhe Song [1] 2017, proposes a novel lung cancer detection method for CT images based on the super-pixels and the level set segmentation methods. In the proposed methods, the super-pixels method is used to segment the lung region and the suspected lung cancer lesion region in the CT image. The super-pixels method and a level set method are used to segment the suspected lung cancer lesion region simultaneously. Finally, the cancer is determined by the difference between results of the two segmentation methods.

Samuel Hawkins1, Hua Wang2, 3, Ying Liu [2] 2016, determines if quantitative analyses ("radiomics") of low dose CT lung cancer screening images at baseline can predict subsequent emergence of cancer.

Public data from the National Lung Screening Trial (ACRIN 6684) were assembled into two cohorts of 104 and 92 patients with screen detected lung cancer (SDLC), then matched to cohorts of 208 and 196 screening subjects with benign pulmonary nodules (bPN). Image features were extracted from each nodule and used to predict the subsequent emergence of cancer.

S.K. Vijai Anand [3] 2010, proposed system efficiently predicts lung tumor from Computed Tomography (CT) images through image processing techniques coupled with neural network classification as either benign or malignant. The lung CT image is denoised using non-linear total variation algorithm to remove random noise prevalent in CT images. Optimal thresholding is applied to the denoised image to segregate lung regions from surrounding anatomy. Lung nodules, approximately spherical regions of relatively high density found within the lung regions are segmented using region growing method. Textural and geometric features extracted from the lung nodules using gray level co-occurrence matrix (GLCM) is fed as input to a back propagation neural network that classifies lung tumor as cancerous or non-cancerous.

Anam Tariq1, M. Usman Akram [4] 2013, proposes a computerized system for lung nodule detection in CT scan images. The automated system consists of two stages i.e. lung segmentation and enhancement, feature extraction and classification. The segmentation process will result in separating lung tissue from rest of the image, and only the lung tissues under examination are considered as candidate regions for detecting malignant nodules in lung portion. A feature vector for possible abnormal regions is calculated and regions are classified using neuro fuzzy classifier.

Devinder Kumar Alexander Wong David A. Clausi [5] 2015, describes a CAD system which uses deep features extracted from an auto encoder to classify lung nodules as either malignant or benign. We use 4303 instances containing 4323 nodules from the National Cancer Institute (NCI) Lung Image Database Consortium (LIDC) dataset to obtain an overall accuracy.

Thomas A. Lampert, Andr´e Stumpf, and Pierre Ganc¸arski [6] 2016 found that: the rank of a detector is highly dependent upon the method used to form the ground truth; and that although STAPLE and LSML appear to represent the mean of the performance measured using individual annotations, when there are few annotations, or there is a large variance in them, these estimates tend to degrade. Furthermore, one of the most commonly adopted combination methods consensus voting accentuates more obvious features, resulting in an overestimation of performance.

Dianhui Wang and Ming Li [7] 2017, contribute to the development of randomized methods for neural networks. The proposed learner model is generated incrementally by stochastic configuration (SC) algorithms, termed SC networks (SCNs). In contrast to the existing randomized learning algorithms for single layer feed-forward networks, we randomly assign the input weights and biases of the hidden nodes in the light of a supervisory mechanism, and the output weights are analytically evaluated in either a constructive or selective manner. As fundamentals of SCN-based data modeling techniques, we establish some theoretical results on the universal approximation property.

Journal of research in engineering and its applications
Vol. 1, Issue. 1, (2018), pp. 34-44

3.RESEARCH METHODOLOGY

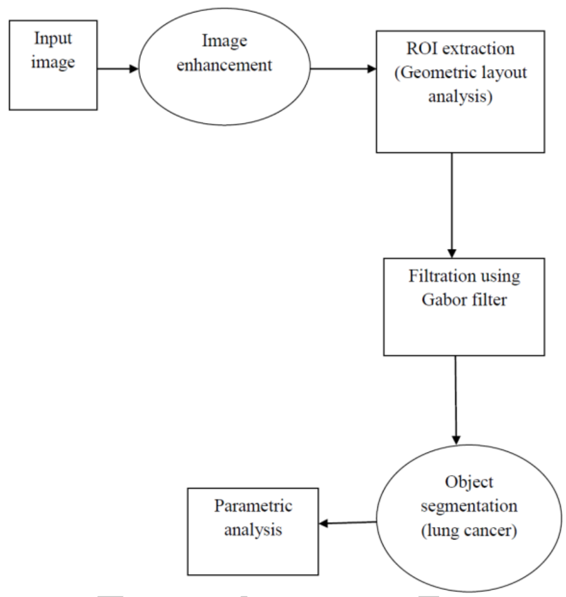


Figure 1. Proposed Architecture

3.1Filtration

Modified Gabor filter is applied here for filtration process whereas the Laplace is applied to make the filtration process more effective. The Gabor filter is a linear filter which impulse response is defined by a harmonic function multiplied by a Gaussian function. This filter can be used to detect line endings and edge borders over multiple scales and with different orientations. While operators that focus on global information are essential to describing a variety of physical systems, operators that focus on local information are essential to analyzing physical signals. For example, linear, time invariant (or shift-invariant) systems are usually analyzed with Fourier or Laplace transforms which are global operations.

$$\mathbf{G(x,y,\theta,\mu,\beta) = exp\ (-\frac{x^{2}+\beta^{\wedge}2y^{\wedge}2}{2\mu^{2}})\ exp(i(2\prod\frac{x}{\beta}+\theta))\ (1)}$$

The above equation indicates Gaussian function for linear filtration process

$$\mathbf{G\ (s) = \int_{0}^{\infty} e^{-X^{2}\beta^{\wedge}2y^{2}} f(t)dt . \int_{0}^{\infty} e^{2\pi x/\beta} f(t)dt} \qquad \mathbf{(2)}$$

The above equation indicates Gabor filtration equation with the impulse response.

To detect the edges of the eyes, the parameters of Gabor wavelet are set as follows: x=1, y=4, β =2π, pmax =π/2 and μ = 2. In this experiment, Gabor wavelet with horizontal orientation is selected because it produces discriminative Gabor features better than other orientation. The parameter r represents the scale of the filter. The greater value of r is set, the size of filter is becomes small. The parameter μ is a standard deviation of the Gaussian function along the x and y-axes. By using this parameter, the width of the filter can be changed and it can increase or decrease the thickness of the edge

3.2 Threshold Based Stochastic Regression Model

Thresholding method is used as an initial approach to segment out lung tissue from the dataset. This model must be primed or trained on a sample of units with measured features and known classes.

Regression model with independent components in which the feature values are treated as covariates we can explore to build the segmentation models by assigning coefficients to the mean topic-assignment of the words in the image z_d, and applying a softmax function in order to obtain a distribution over classes. Alternatively, one could consider more flexible models such as Gaussian processes, however that would considerably increase the complexity of inference.The classification model with stochastic regression models on basis of threshold can be determined by below equation,

$$\boldsymbol{p}(\boldsymbol{c}^{\boldsymbol{d}}/\boldsymbol{z}^{-\boldsymbol{d}}\ \eta)=\boldsymbol{exp}(\eta_c^T \boldsymbol{z}^{-\boldsymbol{d}})/\sum_{l=1}^{C} \boldsymbol{exp}(\eta_l^T \boldsymbol{z}^{-\boldsymbol{d}}) \qquad \textbf{(3)}$$

Where equation (3) shows that the co-efficient η with discrete element z^{-d} which is equals to exponential of co-efficient at time T to the power and with C no. of filtered images in classification model, divisible by the sum of exponential in η from time T to l=1 with the discrete element.

$$\boldsymbol{\zeta}_{i,j}^{(t)} = (\mathbf{1}-\boldsymbol{\rho}_t)\boldsymbol{\zeta}_{i,j}^{(t-1)} + \boldsymbol{\rho}_t(\boldsymbol{\tau} + \boldsymbol{D}\sum_{n=1}^{N_d} \boldsymbol{\omega}_{n,j}^{d} \boldsymbol{\phi}_{n,i}^{d}) \qquad \textbf{(4)}$$

Equation (4) shows the stochastic model for the proposed segmentation model that is represented by $\zeta_{i,j}^{(t)}$ that have the probability ρ_t with time factor and that have summation with number of discrete element N_d that have $\phi_{n,i}^{d}$ as the odd coefficient of the dataset images.

$$\boldsymbol{net}(\boldsymbol{t},\boldsymbol{f}) = (\boldsymbol{x} * \boldsymbol{w})[\boldsymbol{t},\boldsymbol{f}] = \sum_{\boldsymbol{m}} \sum_{\boldsymbol{n}} \boldsymbol{x}[\boldsymbol{m},\boldsymbol{n}]\boldsymbol{W}[\boldsymbol{t}-\boldsymbol{m}_{\boldsymbol{o}}\boldsymbol{f}-\boldsymbol{n}] \qquad \textbf{(5)}$$

Where $net(t,f)$ is the output of proposed model with stochastic models, *x* is input image and *w* is Gabor filter matrix.

4.PERFORMANCE ANALYSIS

A. Image enhancement using Gabor Filtration method

The main objective of image enhancement is to provide better input for other automated image processing techniques. Image enhancement techniques can be divided into two broad categories: Spatial and frequency domain methods. Unfortunately, there is no general theory for determining what "good" image enhancement is when it comes to human perception. If it looks good, it is good. However, when image enhancement techniques are used as pre-processing tools for other image processing techniques, the quantitative measures can determine which techniques are most appropriate. Gabor filter was used to enhance lung images according to the comparison results of FFT and Gabor filtration. Image presentation based on Gabor function constitutes an excellent local and multi-scale decomposition in terms of logons that are simultaneously (and optimally) localization in space and frequency domains. A Gabor filter is a linear filter whose impulse response is defined by a harmonic function multiplied by a Gaussian function. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function. Figure 2 shows (a) the original image and (b) the enhanced image using Gabor Filter.

Figure 2 . (A) Original image showing lung effected

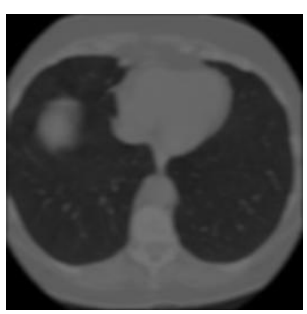


Figure 2. (B) Gabor Enhancement with cancer

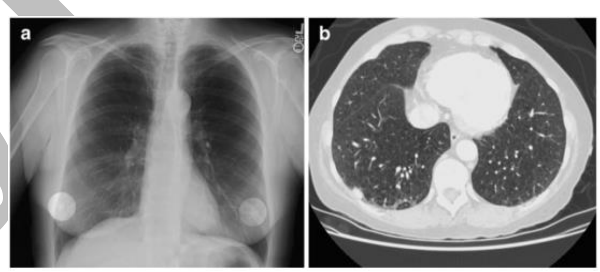
Journal of research in engineering and its applications
Vol. 1, Issue. 1, (2018), pp. 34-44

B. Region Of Interest Extraction

The geometric layout analysis aims at producing a hierarchical representation of the image, which embeds its geometric structure, i.e. classified blocks, each representing a homogeneous region of the image, and their spatial relationships. This structure allows us to describe the image layout at different levels of detail, e.g. a body of text can be viewed as a single coherent element, as well as, at a higher detail level, a set of lines. Skew estimation detects the deviation of the image orientation angle from the horizontal or vertical direction. Early image reading systems assumed that images were printed with a single direction of the text and that the acquisition process did not introduce a relevant skew. The advent of flatbed scanners and the need to process large amounts of images at high rates, made the above assumption unreliable and the introduction of the skew estimation phase became mandatory. In fact, a little skewing of the image is often introduced during processes such as copying or scanning. Moreover, today images are ever more free styled and text aligned along different directions is a common feature.

The image decomposition segments the document image into homogeneous blocks of maximum size (image segmentation), and to classify them into a set of predefined data types (classification). Image segmentation takes into consideration only the geometric layout of the image, e.g. the spacing among different regions, while blocks classification employs specific knowledge about the data types to be discriminated, e.g. features can be devised to distinguish among text, pictures, or drawings.

Here our concentration was on the lung region as region of interest to simplify the representation of an object into something that is more meaningful and easier to analyze. It is important to separate regions of interest from other parts of the image to be account as white background to concentrate on object itself for partitioning into normal and abnormal parts. Lung regions segmentation from the background based on morphological operations as following diagram Figure 3.



A.LUNG IMAGE **B.ROI EXTRACTED IMAGE**

Figure 3. Morphological Operations for Lung Region Extraction

The PA chest x-ray demonstrates two anomalies or regions of interest (ROI) within the right and left lower lobes. The right lower lobe abnormality was scored a 4 on the likert scale, whilst the left was scored a 1 as it was felt the opacity was most likely to represent an overlying rib. Subsequent CTA confirms the presence of the nodule on the posterior aspect of the patient's right lower lobe.

	POSITIVE CTA	**NEGATIVE CTA**
POSITIVE CAD	49	116
NEGATIVE CAD	20	121

TABLE 1. CT ANGIOGRAPHY SENSITIVITY

4.1 CTA:CT Angiography Sensitivity

Sensitivity analysis is defined as the study of how uncertainty in the output of a model can be attributed to different sources of uncertainty in the model input. In the context of using Simulink Design Optimization™ software, sensitivity analysis refers to understanding how the parameters and states (optimization design variables) of a Simulink model influence the optimization cost function. Examples of using sensitivity analysis include:
Before optimization — Determine the influence of the parameters of a Simulink model on the output. Use sensitivity analysis to rank parameters in order of influence, and obtain initial guesses for parameters for estimation or optimization.

After optimization — Test how robust the cost function is to small changes in the values of optimized parameters.

If a person has a disease, how often will the test be positive (true positive rate)? Put another way, if the test is highly sensitive and the test result is negative you can be nearly certain that they don't have disease. A Sensitive test helps rule out disease (when the result is negative). Sensitivity rule out or "Snout"

Sensitivity= true positives/(true positive + false negative)

Accuracy:

If a person does not have the disease how often will the test be negative (true negative rate)? In other terms, if the test result for a highly specific test is positive you can be nearly certain that they actually have the disease. A very specific test rules in disease with a high degree of confidence Accuracy rule in or "Spin".

Accuracy=true negatives/(true negative + false positives)

Predictive value for a positive result (PV+): PV+ asks "If the test result is positive what is the probability that the patient actually has the disease?"

PV+= true positive/(true positive + false positive)

Predictive value for a negative result (PV-): PV- aks “If f the test result is negative what is the probability that the patient does not have disease?"

PV-= true negatives/(true negatives +false negatives)

TP 49

TP + FN

49 + 20 = 71%

Accuracy

TN 121

TN + FP

121 + 116 = 51%

Positive Predictive Value

TP 49

TP + FP

49 + 116 = 29.6%

Negative Predictive Value

TN 121

TN + FN

121 + 20 = 85.8% Accuracy obtained: 92%

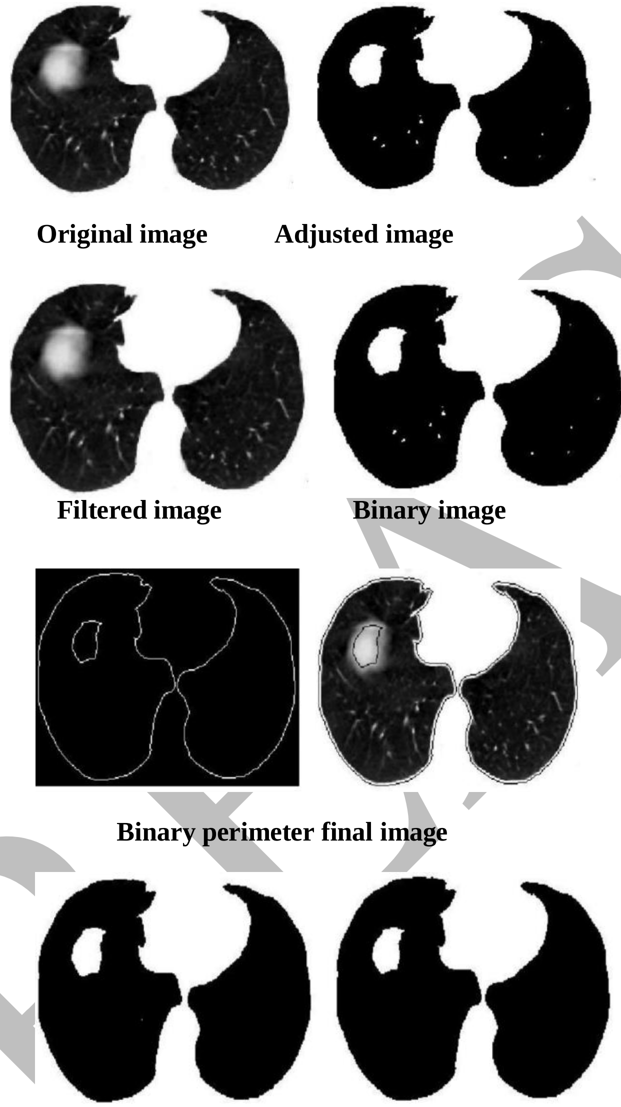
(After overall analysis)

Geometric features of probability densities of interest apart from the standard location (mean) and variability (variance) parameters include values describing the decay (towards remote parts of the support), the curvature (reflecting the correlation structure) and the modality (number of local extrema).

In order to access them before feeding them into a deep learning network, for example, a data-preprocessing has to be carried out, which often involves a sophisticated analysis of the data. The reason for ‘strapping’ the input data for predictive learning algorithms to its essentials is the resulting ability to increase the efficiency of the training and cross validation process.

5.SEGMENTATION OF LUNG CANCER

In general image segmentation is the process of dividing an image into multiple parts. This is typically used to identify objects or other relevant information in digital images. Our concentration was on lung nodules segmentation, global thresholding and morphological operations had been the core of this process. The segmented nodules are used for feature extraction as future step in the detection system. In this section we used extracted region of interest image to locate objects and boundaries in given image.



Original image **Adjusted image**

Filtered image **Binary image**

Binary perimeter final image

Eroded Binary **Eroded Final**

Figure 4. Segmentation of Lung cancers

6.CONCLUSION

According to the stage of discovery of the cancel cell in the lungs, lung cancer is the most dangerous and widespread cancer in the world, so detection process of the disease plays a very important and essential role to avoid serious stages and distribution percentages. The main idea of this work is to detect and segment lung nodule for the future segmentation as cancerous and non-cancerous cells. Thus the lung CT images were subjected to various processing steps to get extracted features for classified systematic future usage. Our future work will be concentrated on identifying the effective features for fuzzy means segmentation with classification.

REFERENCE

[1]. Hongbo Zhua, Chun-Hyok Paka;b, Chunhe Songc, Shengchang Doua; Hai Zhaoa, “A novel lung cancer detection algorithm for CADs based on SSP and Level Set”, 0928-7329/17/$35.00c 2017.

Samuel Hawkins, Hua Wang, Ying Liu, Alberto Garcia, Olya Stringfield, “Predicting malignant nodules from screening CTs”, 10.1016/j.jtho.2016.07.002.

[2]. S.K. Vijai Anand, “Segmentation coupled Textural Feature Classification for Lung Tumor Prediction”, 978-1-4244-7770-8/10/$26.00 ©20 1 0 IEEE.

[3]. Devinder Kumar, Alexander Wong, David A. Clausi, “Lung Nodule Classification Using Deep Features in CT Images”, 2015 12th Conference on Computer and Robot Vision.

[4]. Thomas A. Lampert_, Andr´e Stumpf, and Pierre Ganc¸arski, “An Empirical Study into Annotator Agreement, Ground Truth Estimation, and Algorithm Evaluation”, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. XX, NO. Y, MARCH 2016.

[5]. Macedo Firmino1*, Antônio H Morais1, Roberto M Mendoça2, Marcel R Dantas3, Helio R Hekis3 and Ricardo Valentim3, “Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects”, Firmino *et al*. *BioMedical Engineering OnLine* 2014, **13**:41.

[6]. Dianhui Wang, *Senior Member, IEEE*, and Ming Li, *Student Member, IEEE*, “Stochastic Configuration Networks: Fundamentals and Algorithms”, IEEE TRANSACTIONS ON CYBERNETICS, VOL. 47, NO. 10, OCTOBER 2017.

[7]. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, COMMUNICATIONS OF THE ACM | JUNE 2017 | VOL. 60 | NO. 6.

[8]. Anita chaudhary, Sonit Sukhraj Singh, “LUNG CANCER DETECTION ON CT IMAGES BY USING IMAGE PROCESSING”, 2012 International Conference on Computing Sciences.

[9]. Filipe Rodrigues, Mariana Lourenc¸o, Bernardete Ribeiro, Senior Member, IEEE, and Francisco C. Pereira, Member, IEEE, “Learning Supervised Topic Models for Classification and Regression from Crowds,” DOI 10.1109/TPAMI.2017.2648786, IEEE Transactions on Pattern Analysis and Machine Intelligence.

[10]. Gian Domenico Licciardo, *Senior Member, IEEE*, Carmine Cappetta , *Student Member, IEEE*, and Luigi Di Benedetto, *Member, IEEE*, “Design of a Gabor Filter HW Accelerator for Applications in Medical Imaging”, 2156-3950 © 2018 IEEE