# Study of Early Detection of Lungs Cancer Using Support Vector Machine and Artificial Neural Network

Faisal Imran, Syed Hassan Tanvir, and Abubakar Yamin

***Abstract*—Lung and bronchial Cancer is one of the world's leading cause of death. It is world's second deadliest disease which is now getting very common among men and women. Due to rapid deformation of climate, excess use of tobacco and working in hazardous waste side i.e. nuclear waste, explosive demolition centers left traces of harmful gases in air, which later can cause lung cancer and other fatal diseases. Cure of this deadliest disease is only possible by regularly examining of individuals, who are working and are exposed to health hazardous environment. X-rays images and Computed Tomography (CT) scans are the sources to detect nodules in lungs, which are the primary source of lungs cancers. Detection, identification and classification of these nodules are always a challenging job for doctors and researchers in medical imaging field. Researchers have developed many methods for nodules detection. Some of these methods include machine learning and artificial neural network. In this paper we have discusses few of these methods both from machine learning and artificial neural network for the early detection of nodules from CT scan.**

***Keywords* – CT scan, nodules, X-ray, support vector machine, artificial neural network.**

## I. INTRODUCTION

LUNGcancer in considered being major fatal disease in whole world. This disease was very common in people who work in hazardous environment or have high consumption of asbestos, certain compounds of nickel, polycyclic aromatic hydrocarbons chromates and arsenic trioxides. People who works in coal mine, manage nuclear waste, demolition of explosive while digging Sulphur has this disease in particular and other common lung diseases. The percentage of lung cancer contribution a lot in overall death rate of the world. Cherry on the top now a day's lung cancer is spreading fast in world due to Cigarettes e-cigarettes, cigars, shisha and e-shisha in youth of world. Utilizing such things are the primary culprits for increasing rate of lung cancer patients these days. Such bad habits directly affect bronchus of lungs that later leads to cancer of bronchus. In this deadly disease, uncontrolled growth of tissues in lung starts. A person chance of survival decreases when the tumor grows up to 3 cm or more. These tissues will not develop in to healthy ones and they further divide in tumors [1]. Lungs cancer always stood first position in morbidity and mortality. According to word health organization [2], approximately 1.3 million deaths recorded annually. It is estimated that 80% of lungs cancer can be cured if and only if it gets attention at right time. In US, the lifetime risk of developing invasive lung cancer is 1 in 7 for men and 1 in 18 in women [3]. Solitary pulmonary nodules have been reported in lung cancer. The early disclosure of Lung Cancer in patients are not more than 15% [4]. Lung cancer is difficult to rectify because it does not show any significant symptoms until the cancer reaches its prior stage. However, periodic screening test of patients can help to reduce the lung cancer and eventually death rate. There are different techniques used by the researchers for the identification of lung cancer nodules, using techniques such as support vector machine and artificial neural network. In this paper we have summarized few of the techniques from SVM and ANN. Fig. 1 shows the different techniques discussed in this paper.

The remaining of the paper is formulated as follows. In section 2, lung cancer identification techniques are introduced. In section 3 and 4, techniques used by support vector machine and artificial neural network are described. In section 5 we have discussed the conclusion.

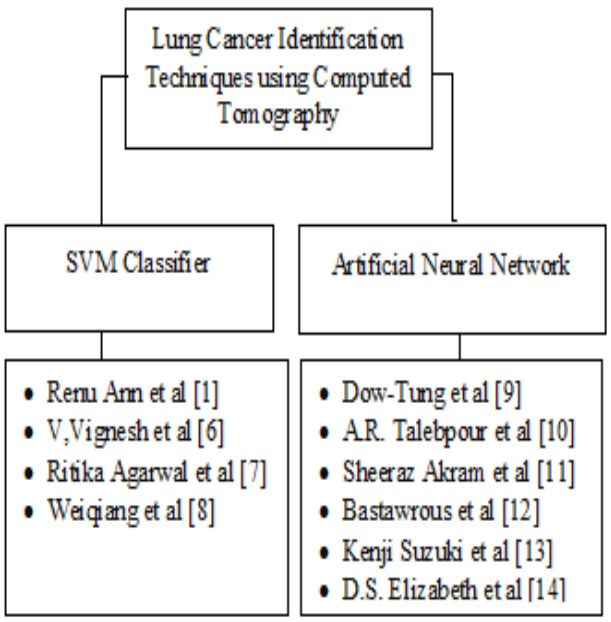


Fig.1: Different techniques discussed in this paper

## II. LUNG CANCER IDENTIFICATION TECHNIQUE

Main Challenge in the medical sciences is to identify lung cancer at its early stage and in cost effective manner. So that third world nations can also afford the rectification and identification of lung cancer. Image processing is one of the cost effective method to identify such kind of diseases. There are many methods which are working effectively to identify lung cancers they are as follow. Computed Tomography (CT), which is a detailed lung screening test. It is painless procedure in which very low dose x-rays are bombarded to screen the lungs in few seconds. This method allows the doctors to see different reformatted image of lungs in different levels of lungs on LCD, Printed film or CD and DVD. Fig. 2 shows the identification of cancer through CT image technique [3].

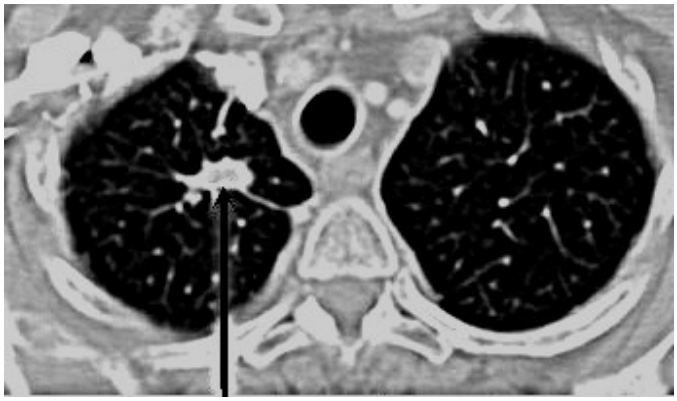

Fig.2 Lung Cancer detection using CT Screening Technique

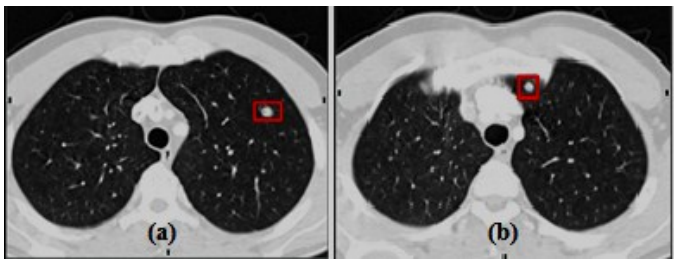


Fig. 3 (a) Normal circular shape nodule (b) Round, hard, bright nodule

Doctors and medical researcher all around the world are spending tremendous amount of efforts in deploying and adopting new technologies and ways for early detection of pulmonary cancer [15]. Computer Tomography scan can display very small nodule which are not possible to be seen on chest radiographs, thus providing a useful way for tumor staging evaluation. Detection on nodules using CT images is very demanding job, reason behind is low contract, location variation and low resolution as shown in the fig 2 and Fig.3 [16]. Many classification technologies have been put forward in detection of nodules using different methodologies of image processing such technologies includes neural-based, feature-based, and template matching classification.

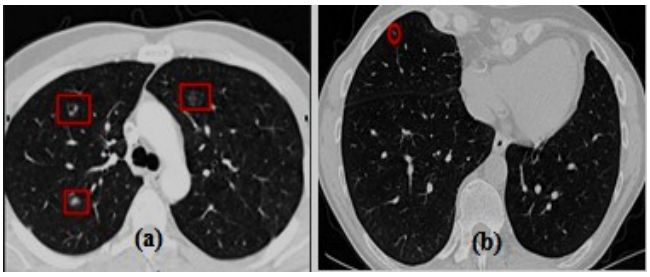


Fig. 4. Lung Cancer sliced nodules/ slices with a darkish, non-solid, low perched or shady appearance (a). A tiny hiding nodule with low brightness (b) [16].

## III. SVM CLASSIFIER

Hond. et. al. has proposed an algorithm which identifies solitary pulmonary nodules. Their proposed algorithm was divided into two stages. In stage 1 the algorithm will process the original CT images given by hospital. DICOM CT images having resolution of 512*512 which will take up adaptive iteration threshold twice to compute complete pulmonary parenchyma segmentation. In second stage histogram analysis were performed to obtain candidate nodules and different features were extracted from ROI. The experimental material was provided by hospital with DICOM CT images. Total numbers of samples were 30 CT images. SVM classifier was trained on the basis of the features which were extracted in the process and were compared with true nodules. This method divide…in to 10 even parts, out of those, nine are named as training sets, remaining one is called test dataset. There proposed algorithm achieved an accuracy of 86% [5].

Renu Ann et al performed the comparison of three classifiers for identification of lung cancer. Morphological operations were used for pre-processing of the images. In image segmentation they have used Grey level Region Growing Algorithm (GLRGA) to add the skipped pixel at its neighboring by using this method the image segmentation process was neatly done. Features were extracted with help of co-occurrence matrix. It describes second order calculation. In first group statistics were calculated from the original image which compensates values like mode and median. In the second group it calculates the relationship of group of two neighbouring pixels. They have applied two classification techniques i.e. supervised classification and un-supervised classification. They have used 3 classifiers namely SVM, KNN and Minimum Distance classifier models. There were total 98 Images with morphological measurements were used as input data. The cross validation framework was used to assess the ability of the classifier. They have calculated Matthews's correlation coefficient for quality of binary classification. They calculated various performance measures too. The results were depicts, that SVM shows better results [1].

V. Vignesh et al. suggested Virtual Dual detection method (VDE). One of the major challenges is to identify the nodules in the chest radiography. Detection of the nodules is one of the important tasks which should be done wisely and accurately if radiologist misses the sub nodules the patient will get expired. To reduce the false positive rate

they suppressed the rib cage or clavicles by VDE method. Grey level based features were extracted from original and VDE CXRs. They have used Japanese Society of Radiological Technology. Lungs tumor was identified at early stage and detection was carried out by CT. V. Vignesh et al. using SVM classifier for classification of the dataset. They have calculated the performance measure. By adopting this method they have achieved the reduction in FP and overall sensitivity of the system was increased to 89% [6].

Ritika Agarwal et al wrote a review paper in which they retrieved medical content based images of lung cancer. They proposed a Computer Aided Diagnosis (CAD) system. CAD system detects lung cancer at early stage. There are several steps involved in the CAD system such as extraction of different region of lungs. Features were extracted by using GLCM. SVM classifier has been used for classification and identification of cancerous nodules. SVM showed 80% efficiency [7]. Weiqiang et al [8] has projected a new method in which they have combined SVM with knowledge of Pathology. They have achieved sequential minimal optimization by deploying game theory to their system. They have compared there results with LCDS system. It proved that validity of their system is authentic. The processing speed of SMO was raised by deploying game theory.

## IV. ARTIFICIAL NEURAL NETWORK

Dow-Tung et al uses a neural network for diagnosis of lungs tumour. He submitted a two stage model for lungs tumour detection on CT images. In first stage segmentations of CT is done to increase the lung nodules features detection by removing the background information and extracting the lungs field area. Blood vessels and nodules are extracted from the segmented image. This field area will serve as region of interest for obtaining the identification rule and to discriminate true or false nodules from ROI using neural fuzzy model. Dow-Tung applied neural fuzzy model on 29 clinical cases including 583 sliced CT images. This method yield approximately 0.3 per image of false positive with a detection rate of 89.3% [9] as shown in the Table 1. A.R. Talebpour et al designed an automatic method for detection of lungs nodules in high resolution CT (HRCT) images. They applied 3D adaptive fuzzy filtering and false positive reduction methods for nodules filtering and detection. Features were extracted from these 3D nodules using binary mask for geometric features and grey level value for texture features. For classification and detection of nodules from non-nodules objects, three layer structures, feed forward neural network with supervised back-propagation method is used. The proposed method is applied on 10 clinical cases with 144 slices per case. The accuracy of the recommended method was reached up to 90% reducing the false positive rate up to 0.2 [10].

Sheeraz Akram et al performed an automatic thresholding for detection of nodules for lungs cancer in CT scan using artificial neural network classifier. The proposed method includes segmentation of lungs CT images, detection of nodules for the area of interest, feature extraction and at last neural network classification. Akram accomplished the segmentation using optimized thresholding for the identification of circular overlaps regions. The thresholding value is calculated iteratively for maximum optimization. The Hessian Unit (HU) is set to minimum if the vowels value is less than thresholding value and if the vowels value is greater than thresholding. Extraction and detection of candidate nodules is carried out using lungs lobes extraction, hole filling and contour correction. Potential nodules are separated from vessels using hole filling. Contour correction is performed to include the juxta-pleural nodules in the lungs lobe. The geometric features (2D & 3D), statistical features based on intensity values (2D & 3D) are calculated from the candidate nodules. The ANN with feed forward back propagation is trained using these extracted nodules. Akram proposed ANN classifier achieve the accuracy of 91.87% with a sensitivity of 96.55% and 0.40 FP/scan [11]. Hany Ayad Bastawrous et al proposed a method uses Gabor filter along with neural networks for the discloser of ground glass opacities (GGO) nodules in Chest CT images. This model enhanced the nodule detection process by applying Gabor filter on the CT images. Interested objected having high intensity values are extracted, using morphological operations which includes thresholding and labelling. Region of interest (ROI) is identified using feature extraction analysis. After the features are extracted, the similarities between the potential cancer candidates are determined by applying a template and some Gaussian reference models. Hany proposed algorithm was applied over 715 slices containing 25 ground glass opacities.

| Method | Acc/ Sens | False Positive | Dataset |
|---|---|---|---|
| Dow-Tung[9] | 89.3% | 0.30 | 29 cases, 583 slices |
| A.R. Talebpour [10] | 90.0% | 0.20 | 10 cases, 144 slices |
| SheerazAkram [11] | 91.87%/ 96.55% | 0.40 | 49 CT Scan |
| Bastawrous [12] | 92% without ANN 81% with ANN | 0.75 without ANN 0.25 with ANN | 715 slices, 25 GGO 715 slices, 25 GGO |
| Kenji Suzuki [13] | 97% without multiple MTANN 82% with multiple MTANN | 6.7 without MTANN 1.2 with MTANN | 69 cases |
| D.S. Elizabeth[14] | 94.44% | 5-11 | 150 CT, 1564 slices |

Table 1 Analysis of different artificial neural networks for detection of nodules

This model achieves nodules detection sensitivity of 92% with false positive rate of 0.76 FP/slice. Finally artificial neural network was applied to reduce the FP/slice rate to 0.25, but doing so reduce the detection sensitivity to 84% [12]. Kenji Suzuki et al massive-training ANN for the enhancement of lung nodules. To identify the region of interest, segmentation of lungs regions in a CT images is performed. Suzuki uses otsu's thresholding and rolling ball technique for the segmentation and to include nodules attached to the pleura segmented region. To enhance the nodules in CT images Suzuki train the massive training ANN filter with the entire lung region and corresponding teaching image containing maps for likelihood of being a lung nodule. To increase the accuracy rate the MTANN filter is trained 100,000 times. The proposed technique is applied on dataset containing 69 cases and achieving a performance of 97% with false positive rate of 6.7 FP/slice. To reduce the false positive rate to 1.2 FP/slice Suzuki applied multiple massive training ANN. This process decreases the FP/slice but also decrease the accuracy to 82% [13].

D.S. Elizabeth et al contemplated radial basis function neural network for the recognition of lung cancer in CT images. Greedy snake algorithm is used for the segmentation of lung parenchyma from each slice. This segmentation is performed after the pre-processing of CT images. To identify the region of interest from the extracted parenchyma region growing algorithm is used. Shape and texture feature were derived after the labelling of cancerous or non-cancerous nodules. This labelling is performed using domain experts. Radial basis function neural network (RBFNN) is trained using labelling of corresponding area of interest and features were extracted. For the detection of valid nodules from the CT image this technique selects only those areas of interest more than 9 pixels and it must exist in at least 3 slices. The proposed technique achieve an accuracy of 94.44% in a dataset of 150 chest CT images, 1564 slices and an error rate of 5-11 [14].

## V. CONCLUSION

Lung cancer is one of the most lethal cancers among all other cancers in the world. Early identification is the only cure. Early detection is only possible in the correct identification of nodules in CT scan. It has been always difficult for the medical specialist to recognize and classify cancer nodules due to its varying features. Correct detection of lungs nodules is only possible if the CT scan is analysed properly. Different image processing techniques have been established for the analysis of CT scan. An algorithm will provide more accurate results if it leads to proper analysis of lungs cancer. Among all of these methods neural network proves to be more accurate with an accuracy ranging between 90-97%.

## REFERENCES


[1] Thomas, Renu Ann, and Sahoo Subhendu Kumar. "Automatic detection of lung nodules using classifiers." Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on. IEEE, 2014.

[2] http://www.who.int/mediacentre/factsheets/fs297/en/; (Assessed on September 23-09-2015 at 9:39 am)

[3] https://www.cedars-sinai.edu/Patients/Programs-and-Services/Imaging-Center/For-Patients/Exams-by-Procedure/CT-Scans/CT-Lung-Screening.aspx (Assessed on Sep. 23-09-2015 at 9:39 am)

[4] Chunhui Wei. The progress of lung cancer diagnosis in early stage[J]. Journal of Clinical Pulmonary Medicine.2010,Vol.15,no.8,1136-1138

[5] Shao, Hong, Li Cao, and Yang Liu. "A detection approach for solitary pulmonary nodules based on ct images." Computer Science and Network Tech. (ICCSNT), 2012 2nd Int. Conf. on. IEEE, 2012.

[6] Vignesh, V., and K. Kothavari. "Classification and detection of lung nodules using virtual dual energy in CXR images." Green Computing Comm. and Electrical Eng. (ICGCCEE), . Int. Conf.on. IEEE, 2014.

[7] Agarwal, Ritika, Ankit Shankhadhar, and Raj Kumar Sagar. "Detection of Lung Cancer Using Content Based Medical Image Retrieval." Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on. IEEE, 2015.

[8] Liu, Weiqiang, et al. "Fast algorithm of support vector machines in lung cancer diagnosis." Medical Imaging and Augmented Reality, 2001. Proceedings. International Workshop on. IEEE, 2001.

[9] Daw-Tung Lin and Chong-Ren Yan "Lung nodules identification rules extraction with Fuzzy Network" Proceedings of the 9th int. conf. on Neural Info. Processing (ICONIP02), Vol. 4 Pg. 2049-2053

[10] A.R. Talebpour, H.R. Hemmati, M. Zarif Hosseinian, "Automatic Lung Nodules Detection In Computed Tomography Images Using Nodule Filtering And Neural Networks" The 22nd Iranian Conference on Electrical Engineering (ICEE 2014)

[11] Sheeraz Akram, Muhammad Younus Javed, Ayyaz Hussain, "Automated Thresholding of Lung CT Scan for Artificial Neural Network Based Classification of Nodules", ICIS 2015, June 28-July 1 2015, Las Vegas, USA

[12] Hany Ayad Bastawrous, Takuya Fukumoto, Norihisa Nitta, Masaru Tsudagawa. "Detection of Ground Glass Opacities in Lung CT Images Using Gabor Filters and Neural Networks", IMTC 2005 – Instrumentation and Measurement Technology Conference Ottawa, Canada, 17-19 May 2005

[13] Kenji Suzuki, Zhenghao Shi, and Jun Zhang. "Supervised Enhancement of Lung Nodules by Use of a Massive-Training Artificial Neural Network (MTANN) in Computer-Aided Diagnosis (CAD)"

[14] D.S. Elizabeth, H.K. Nehemiah, C.S. Retmin Raj, A. Kannan., "Computer-aided diagnosis of lung cancer based on analysis of the significant slice of chest computed tomography image", IET Image Process., 2012, Vol. 6, Iss. 6, pp. 697–705.

[15] BramV. Ginneken, Bart M. Haar Romeny, and Max A. Viergever, "Computer-Aided Diagonsis in Chest Radiography: A survey", IEEE Trans. on Medical Image. Vol, 20, pp. 1228-1241, Dec (2001)

[16] http://www.website-go.com/acc/; (Accessed on September 27-09-2015 at 9:50pm)


## AUTHORS


**Faisal Imran**, Faisalali129@gmail.com,
Bahria University, Islamabad, Pakistan.

**Syed Hassan Tanvir**, hassantanvir2008@yahoo.com
Bahria University, Islamabad, Pakistan.

**Abubakar Yami**, abubakar.yamin@gmail.com
Bahria University, Islamabad, Pakistan