ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Website: www.ijareeie.com

**Vol. 8, Issue 2, February 2019**

# Lung Cancer Detection Using Machine Learning Techniques

K.Mohanambal[1], Y.Nirosha[2], E.Oliviya Roshini[3], S.Punitha[4], M.Shamini[5]

Assistant Professor, Dept. of IT, Valliammai Engineering College, Chennai, Tamil Nadu, India[1]

UG Student, Dept. of IT, Valliammai Engineering College, Chennai, Tamil Nadu, India[2]

UG Student, Dept. of IT, Valliammai Engineering College, Chennai, Tamil Nadu, India[3]

UG Student, Dept. of IT, Valliammai Engineering College, Chennai, Tamil Nadu, India[4]

UG Student, Dept. of IT, Valliammai Engineering College, Chennai, Tamil Nadu, India[5]

**ABSTRACT**: Early detection of lung cancer cells can help in a sharp decrease in the lung cancer mortality rate hence it is an aggressive disease which carrying a dismal prognosis with a 5-year survival rate at 18%.Several computer-aided diagnosis systems have been developed to help reduce lung cancer mortality rates.Thus structural co-occurrence matrix (SCM)-based approach is used to extract the feature and to classify nodules into malignant or benign nodules and also into their malignancy level. The computed tomography(CT) scanfrom the lung image database consortium and image database resource initiative datasets provide knowledge concerning nodule positions and their malignancy levels is been deployed here as a model. Support vector machine is been used as a classifier which is (i) to classify the nodule images into malignant or benign nodules and (ii) to classify the lung nodules into malignancy levels (1 to 5).These experimental results reveal that the SCM successfully extracted features of the nodules from the images and, therefore may be considered as a promising tool to support medical specialist to make a more precise diagnosis concerning the malignancy of lung nodules.

**KEYWORDS:** Structural Co-occurrence Matrix(SCM),Malignant nodule, Benign nodule.

## I.INTRODUCTION

Lung cancer is one of the main cause of the death and health issue in many countries with a 5-year survival rate of only 10–16% [1][2].To support radiologists in the identification of early-stage pathological objects, about one decade ago, researchers started the development of CAD methods to be applied to CT examinations [1]. Although most tests and procedures can be done within minutes or hours, inevitable but substantial waiting times can occur between the tests and procedures. In fact, days or even months of delays are not uncommon. Such delays for a potentially life-threatening illness not only lead to unpleasant experience to both the patients and care providers, but also may be linked to adverse survival rate. Therefore, without sacrificing care quality, speeding up the diagnosis-to-treatment process is critically important to improve patient outcome [7].The Compute Tomography (CT) has been shown as the most sensitive imaging modality for the detection Of small pulmonary nodules, particularly since the introduction of the multi-detector-row and helical CT technologies [1][2],It facilitates radiologists to assess early risk factors of cancer which is essential in lung cancer research [11]. The number of deaths caused due to lung cancer is more than prostate, colon and breast cancers combined. Also, most patients detected with lung cancer today are already at an advanced stage as lung cancer is hard to detect in early stages [2].In these features, sizes, shapes and volumetric growing rate have been considered as most reliable features for malignant diagnosis of nodules. However, most of the above features need more accurate segmentation of nodules except texture. Therefore, we take more attention into the texture features for assisting the malignancy diagnosis of nodules in this paper [6]. Computed Tomography (CT) was developed by the engineer Godfrey N.Hounsfield and the physicist Allan M.Comark in 1972.Thetwo scientists were awarded the Nobel Prize for Medicine in 1979 [4], To create such CAD systems, there is always a need for a reference quality dataset that can be used to acquire ground truths and can also act as a basis for comparison of different CAD algorithms. LIDC

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 2, February 2019

(Lung Cancer Data Consortium) [3], Image-based techniques for analyzing lesions are normally performed with detection, segmentation hand-crafted feature engineering and category labelling Zinovev et al. adopted a belief decision tree approach to predict nodule semantic attributes. Related studies on the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset and at simultaneous the CADx system provides a second opinion to help in decision-making [6].However, all these methods rely on nodule segmentation as a prerequisite. Notably, automatic nodule segmentation may influence classification results since methods such as region growing and level set typically depend on initialization [5].Working on these segmented regions may yield inaccurate features that lead to erroneous outputs [5]. The applications of automatic tumour segmentation are broad, including measuring treatment response, planning of radiation treatment, and to facilitate extraction of robust features for high-throughput radiomics by the MRRN [8]. Hence, good results obtained with the proposed method would be a potential to improve medical diagnoses and assist in making more accurate and efficient decisions for these two major health problems.

## II. RELATED WORKS

In [1][2]``Automatic detection of lung nodules in computed tomography images and cancer statistics: Training and validation of algorithms using public research databases," Lung cancer is manifests itself as a non-calcified pulmonary nodules that can be detected reading lung Computed Tomography (CT) images and Computer Aided Detection (CAD) and it is composed of two CAD sub-procedures is presented: CADi which is to analyse the internal parenchymal nodules and CADJP is to identify the nodules attached to the pleura surface.These CAD algorithm will perform the
– lung segmentation;
– candidate detection;
– candidate classification.

The goal of a CADe system is to identify regions of interest in the image that can reveal specific abnormalities and alert physicians to these regions [12]. In [3] ”Lung nodule classification using deep features in CT images," Database that contains CT images of thoracic region for 1010 patients along with the annotation data of suspicious nodules (for both benign and malignant nodules) for a size greater than 3 mm from up-to four radiologists collected over a long period of time. The data that holds diagnostic data for a small number of cases (157 patient) procure from biopsy, surgical resection, progression or reviewing the radiological images to show 2 years of nodule size has been used for the analysis here. An auto-encoder as feature extraction is primarily a two layer network that takes an input f [0;1] and then uses a linear or non-linear transformation to ”encode” the data to a latent space. On the output layer, it uses a decoding transformation to rebuild the data. In “Analysis of human tissue densities: A new approach to extract features from medical images," a new feature extraction method based on human tissue density patterns ,named Analysis of Human Tissue Densities(AHTD) is presented. Machine learning classifiers were register to each feature extractor for two CT image datasets, one to assort lung disease in CT images of the thorax and the other to assortment stroke in CT images of the brain [4]. In ``Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classication,"In is reference paper they applied multi-crop pooling operation which is a specialized pooling strategy for producing multi-scale features to surrogate the conventional max-pooling operation.In this multiple network become the main burden for training CNNs efficiently since they involve more computational costs, especially when dealing with high-resolution images[5]. The automatic classification of malignancy suspiciousness on CT studies is a worthy task, because it would facilitate radiologists to assess early risk factors which is essential in lung cancer research [5].And “A Texture Feature Analysis for Diagnosis of Pulmonary Nodules Using LIDC-IDRI database.The purpose of this paper is to evaluate the performance of texture features on lung CADx and the consistency with expert diagnosis based on visual features from CT images. To obtain more stable results, each nodule in the database are attached by an annotation file which includes the boundary coordinates and some image features recognized by almost four radiologists [6].We calculated the texture features by Haralick method on 2D nodules images to evaluate the performance of texture features on nodule CADx. In [7] “A System-Theoretic Method for Modeling, Analysis, and Improvement of Lung Cancer Diagnosis-to-Surgery Process” here , a system-theoretic method is instigate to analyze the diagnosis-to-treatment process for lung cancer patients who receive surgical resections.Formulas are derived to estimate the mean and co-efficient of variation of waiting time while the diagnosis to-surgery process. This method provides aquantitative tool for the analysis and improvement of lung cancerdiagnosis-to-surgery process. Typically a diagnostic biopsy will be followed by staging tests, which can be noninvasive or invasive. Numerous studies on delays and waiting times during the diagnos is-to-treatment process for lung cancer patients have been carried out from

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 2, February 2019

clinical point of view. The Time to Treat program, which is designed for patients with clinical or radiographic suspicion of lung cancer is introduced.We can decompose such a complex network into a collection of serial processes. In [8] Multiple Resolution Residually Connected Feature Streams For Automatic Lung Tumor Segmentation From CT Images", we developed two multiple resolution residually connected network (MRRN) formulations called incremental-MRRN and dense-MRRN. An algorithm was instructed by using the 377 tumors from the TCIA dataset and validated on the MSKCC and tested on LIDC datasets. The residual stream input consists of feature maps from one of the preceding higher resolution residual streams. Feature integration from multiple resolutions obviates the need for additional un-pooling following RCUBs as used in conventional encoder-decoder like the Unet. CT scan have been used in lung cancer detection, risk assessment, and clinical management. In particular, the increasing quantity of CT image assays has created a unique address for data-driven analysis to capture underlying cancer characteristics at a macroscopic level, allowing identification of prognostic imaging biomarkers [10].

## III.PROPOSED SYSTEM

The image processing techniques are mostly used for prediction of lung cancer and also for early detection and treatment to intercept the lung cancer. To prognosis the lung cancer various features are extracted from the images therefore, pattern recognition based approaches are useful to predict the lung cancer. In the proposed system, we use Python software for analysis. In image processing procedures, the methods involved are image pre-processing, segmentation and feature extraction techniques have been discussed in detail. We are line up to get the more accurate results by using various enhancement and segmentation techniques. The structural co-occurrence matrix (SCM) proficiency was appeal to pluck out features from images of nodules and categorize them into malignant or benign nodules and their malignancy levels. The CT test from the lung image database consortium and image database resource induce datasets provide information referring to nodule positions and their malignancy levels. The classification stage utilise **Support Vector Machine algorithm (SVM)** and applied them to two tasks:

(i) to classify the nodule images into malignant or benign nodules and
(ii) to classify the lung cells into malignancy levels (1 to 5).

These experimental result reveals that the SCM successfully extracted features of the lung nodules from the images and, therefore may be considered as a promising tool to support medical specialist to make a more pin point diagnosis concerning the malignancy of lung nodules.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Website: www.ijareeie.com

Vol. 8, Issue 2, February 2019

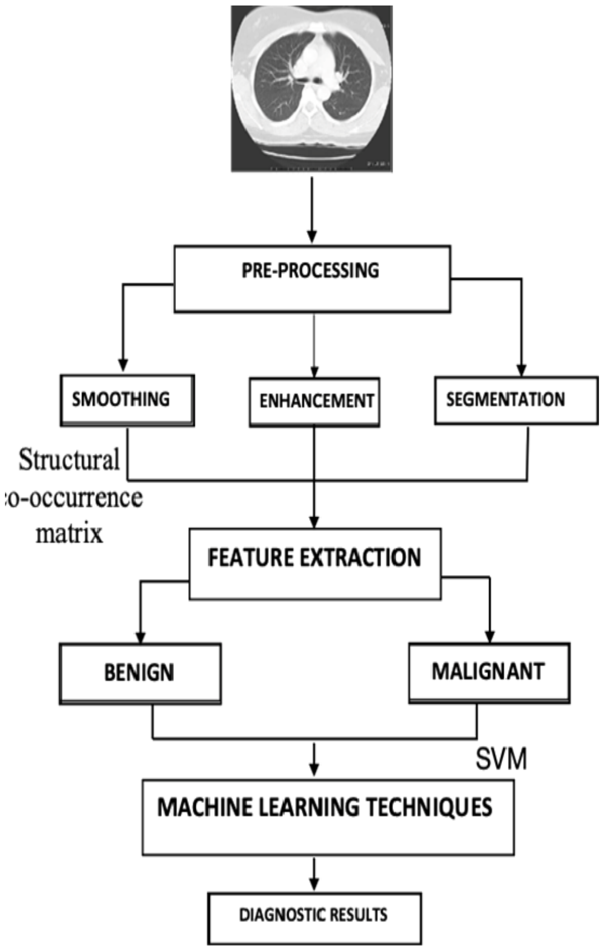


**Fig.1 Architecture of the system**

## IV.METHODOLOGY

The lung cancer detection system using image processing is used to classify the present of lung cancer in a CT- images. In this study, we use Python software for analysis. In image processing procedures, methods such as image pre-processing, segmentation and feature extraction have been discussed in detail. We are line up to get the more accurate results by using various enhancement and segmentation techniques. The structural co-occurrence matrix (SCM) proficiency was appeal to pluck out features from images of nodules and categorize them into malignant or benign nodules and their malignancy levels.The classification stage utilize **support vector machine algorithm** and applied them to two tasks: (i) to classify the nodule images into malignant or benign nodules and (ii) to classify the lung nodules into malignancy levels (1 to 5). These experimental results reveal that the SCM successfully extracted features of the nodules from the images and, therefore may be considered as a promising tool to support medical expert to make a more sparse detecting concerning the malignancy of lung nodules. Thus, computer-aided diagnosis (CAD) systems have arisen to overcome such situations. CAD systems are categorized into two groups[9], to wit, the

(i) detection systems (CADe) and,
(ii) diagnostic systems (CADx).

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

**International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering**

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijareeie.com*

**Vol. 8, Issue 2, February 2019**

CADx systems perform an automatic diagnosis based on features extracted from the system input images [9].The automatic classification of the nodules into malignant or benign using CT images supports the medical specialist when assessing nodules and at the same time the CADx system provides a second opinion to help in decision-making.

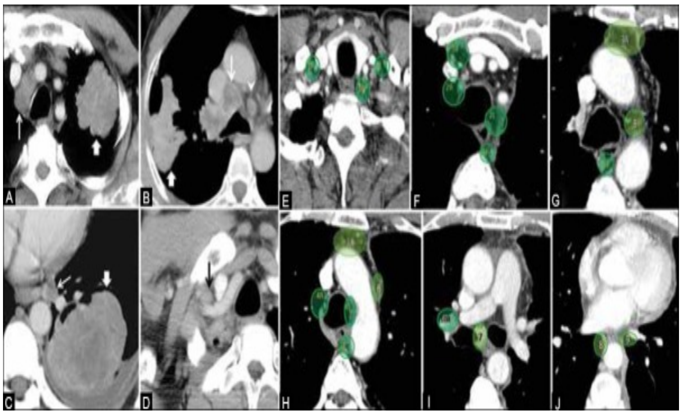


**Fig 2.Lung cancer nodules.**

## V.FUTURE ENHANCEMENT

The lung cancer detection system using the machine learning technique is much efficient and gives the betterment result to the radiologist and assist them. This enhances with the additional features for upgrading in the future. On this image processing system to support the radiologist to detect the malignancy nodules and the level of malignancy of the lung.

## VI.RESULT AND CONCLUSION

Here we processed the CT scan images to differentiate the benign and the malignant nodule and its level of the growth of the cancer cells by the machine learning system and detecting the growth of the cancerous cell in the initial stage can be made in our project.Here it presented an approach to differentiate the pulmonary nodules into malignant and benign nodules to assist the radiologist and for the future enhancement. Further loads ought be directed at improving the SVM for classifying malignancy levels of nodules through experiments with various alternatives.

## REFERENCES

[1].N.Camarlinghi, ``Automatic detection of lung nodules in computed tomography images: Training and validation of algorithms using public research databases," Eur. Phys. J. Plus, vol. 128, no. 9, p. 110,Sep. 2013.
[2]. R. L. Siegel, K. D. Miller, and A. Jemal, ``Cancer statistics, 2016," CA,Cancer J. Clin., vol. 66, no. 1, pp. 730, Jan. 2016
[3].D. Kumar, A. Wong, and D. A. Clausi, ``Lung nodule classification using deep features in CT images," in Proc. 12th Conf. Comput. Robot Vis. (CRV), Jun. 2015, pp. 133138.
[4]. P. P. RebouçasFilho, E. D. S. Rebouças, L. B. Marinho, R. M. Sarmento,J. M. R. Tavares, and V. H. C. de Albuquerque, ``Analysis of human tissue densities: A new approach to extract features from medical images,"PatternRecognit. Lett., vol. 94, pp. 211218, Jul. 2017.
[5]. W. Shen et al., ``Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classication," Pattern Recognit., vol. 61, pp. 663673, Jan. 2017.
[6]. FangfangHan ,Guopeng Zhang, Huafeng Wang, Bowen Song, Hongbing Lu, Dazhe Zhao, Hong Zhao and Zhengrong Liang "A Texture Feature Analysis for Diagnosis of Pulmonary Nodules Using LIDC-IDRI Database" 2013.
[7]. Hyo Kyung Lee,Student Member, IEEE,FengJu,Member, IEEE, Raymond U. Osarogiagbon, Nicholas Faris, Xinhua Yu, FedoriaRugless, Shan Jiang, and JingshanLi,Fellow," A System-Theoretic Method for Modeling, Analysis, and Improvement of Lung Cancer Diagnosis-to-Surgery Process",2017.
[8]. JueJiang , Yu-chi Hu, Chia-Ju Liu, DarraghHalpenny, Matthew D. Hellmann, Joseph O. Deasy, Gig Mageras and HariniVeeraraghavan" Multiple Resolution Residually Connected Feature Streams For Automatic Lung Tumor Segmentation From CT Images",2018.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijareeie.com*

**Vol. 8, Issue 2, February 2019**

[9]. I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H.C.de Albuquerque, and J. M. R. Tavares,``Automati3D pulmonary nodule detection in CT images: A survey," Comput. Methods Programs Biomed., vol. 124, pp. 91107, Feb. 2016.
[10]. P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G. van Stiphout, P. Granton, C.M. Zegers, R. Gillies, R. Boellard, A. Dekker, et al., Radiomics: extracting more information from medical images using advanced feature analysis, Eur. J. Cancer 48 (4) (2012) 441–446.
[11]. M.M. Wahidi, J.A. Govert, R.K. Goudar, M.K. Gould, D.C. Mc Crory, Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer?:Accp evidence-based clinical practice guidelines, CHEST J 132 (3_suppl) (2007) 94S–107S.
[12]. Edson Cavalcanti Neto, Edson Cavalcanti Neto, V .H.C. Albuquerque, Joao Manuel R. S. T avares," Automatic 3D pulmonary nodule detection in CT images",February 2016.