

# Detection of Lung Cancer by Deep Learning and Machine Learning Techniques

<sup>1</sup>Nazia Fatima, <sup>2</sup>Ayonija Pathre, <sup>3</sup>Mukesh Kumar

<sup>1</sup>Student, Department of Computer Science, Rabindranath Tagore University, Bhopal, MP, India

<sup>2,3</sup>Assistant Professor, Department of Computer Science, Rabindranath Tagore University, Bhopal, MP, India

**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) scan from 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:** Cell Disintegration, Computed Tomography, CNN, Support Vector Machine, Tumor.

## I. INTRODUCTION

Lung cancer is one of the main causes of the death and health issue in many countries with a 5-year survival rate of only 10–16%. 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. 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.

The Computer 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. It facilitates radiologists to assess early risk factors of cancer which is essential in lung cancer research. 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. 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.

Computed Tomography (CT) was developed by the engineer Godfrey N.Hounsfield and the physicist Allan M.Comark in 1972.The two 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 (Lung Cancer Data Consortium), 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.

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. Working on these segmented regions may yield inaccurate features that lead to erroneous outputs. The applications of automatic tumor segmentation are broad, including measuring treatment response, planning of radiation treatment, and to facilitate extraction of robust features for high-throughput radio mics by the MRRN. 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 WORK

Timor Kadir, Fergus Gleeson publishes Lung cancer prediction using machine learning and advanced imaging techniques. In this paper, the authors have provided an overview of the main approaches used for nodule classification and lung cancer prediction from CT imaging data. Here, sufficient training data has been give, the current state-of-the-art is achieved using CNNs trained with Deep Learning achieving a classification performance in the region of low 90s AUC points. When evaluating system performance, it is important to be aware of the limitations or otherwise of the training and validation data sets used, i.e., was the patients' smokers or non-smokers, or were patients with a current or prior history of malignancy included. The 15 features were selected from a palette of over 1,300 classical texture features including Haralick, Gabor, along with simple measures such as mean, standard deviation and volume. They utilized a fully automated feature selection strategy that aimed to select a small subset of features that optimized classification performance over an in-house training dataset. Since it is computationally infeasible to test all combinations of the full palette of features, they utilized a sequential "greedy" algorithm that, starting with the optimal pair of features found by exhaustive search over all pairs of features, selected features one-by-one so as to maximize the performance over the training dataset at each step. Finally, an SVM regression algorithm with a cubic kernel was trained using the libSVM library. The output of this step is a number between 0 and 1 that reflects the likelihood that a particular nodule is malignant. Few advantages of their work are helps to reduce the variability in assessing and reporting the lung cancer risk between interpreting physicians. Improve classification performance by supporting the less experienced or non-specialized clinicians in assessing the risk of a particular nodule being malignant. Few performances are poor.

Jason L. Causey, Yuanfang Guan, Wei Dong, Karl Walker, Jake A. Qualls, Fred Prior, Xiuzhen Huang has proposed Lung cancer screening with low-dose CT scans using a deep learning approach. This papers focus on risk prediction at the patient level, taking into account information from the whole lung. The approach could be combined with others to provide a layered strategy for identifying and diagnosing lung cancer. The approach combines convolutional neural network models and an XG boost classifier to predict the presence of lung cancer at the whole-image level. The authors chose to test the strategy on low-dose CT scan data from the National Lung Cancer Screening Trial (NLST). This

is the first time a whole lung CNN-based classifier has been tested on this large NLST cohort. The algorithm Deep Screener able to make predictions with high accuracy, with an AUC of 0.858, and an AUPRC of 0.788. From the testing results on the NLST cohort, the authors anticipate deep learning algorithms can achieve a performance potentially comparable to human experts and radiologists for lung cancer prediction and detection with low-dose CT scans. Through training and learning from CT images of even a larger population, the approach will yield sensitive, stable, consistent and reliable lung cancer screening with the potential of reducing the human effort and cost of screening. Advantages of this work are improve classification stage accuracy and reliable.

Goran Jakimovski and Danco Davcev publish Using Double Convolution Neural Network for Lung Cancer Stage Detection. Here, the authors propose a lung cancer medical image classifier that is based on a Convolutional Deep Neural Network. To train and test the system, they used CT images of lungs that were previously classified by medical specialists and put into piles of yes/no (yes, the patient is diagnosed with lung cancer; and, no, the patient is cancer-free). In this paper, the authors used the K-means algorithm to pre-classify the images into piles of same slice images, where the DNN can focus on image classification of same slice images. The second novelty is the additional convolution layer with edge sharpening filters, to thoroughly search for cancer. Finally, the main novelty is testing our Deep Neural Network with lung cancer images from Tx stages 2, 3 and 4 and determining at which Tx stage the two algorithms can detect the possibility of cancer. The results were analyzed with medical personnel from the oncology department and were marked as satisfactory to determine cancer. Advantages help in early lung cancer detection and early treatment. The drawback here is that the authors had to decide the minimal value of certainty they would accept as being satisfactory.

S. Sasikala, M. Bharathi, B. R. Sowmiya proposes Lung Cancer Detection and Classification Using Deep CNN. A convolutional neural network based system was implemented to detect the malignancy tissues present in the input lung CT image. Lung image with different shape, size of the cancerous tissues has been fed at the input for training the system. This paper presents lung cancer detection based on chest CT images using CNN. In the first stage, lung regions are extracted from CT image and in that region each slices are segmented to get tumors. The segmented tumor regions are used to train CNN architecture. Then, CNN is used to test the patient images. The main objective of this study is to detect whether the tumor present in a patient's lung is malignant or benign. A sample image has been fed as an input to the trained model and the model at this stage is able to tell the presence of

cancer and locate the cancer spot in the sample image of a lung cancer. The process involves the feeding the input image, preprocessing, feature extraction, identifying the cancer spot and indicate the results to the user. The proposed system is able to detect the presence and absence of cancerous cells with accuracy of about 96%. Here, the specificity obtained is 100% which shows that that there is no false positive detection.

Joseph A. Cruz, David S. Wishart proposes Applications of Machine Learning in Cancer Prediction and Prognosis. This paper compares and assesses the performance of different machine learning that are being applied to cancer prediction and prognosis. Specifically, the authors identified a number of trends with respect to the types of machine learning methods being used, the types of training data being integrated, the kinds of endpoint predictions being made, the types of cancers being studied and the overall performance of these methods in predicting cancer susceptibility or outcomes. While ANNs still predominate it is evident that a growing variety of alternate machine learning strategies are being used and that they are being applied to many types of cancers to predict at least three different kinds of outcomes. It is also clear that machine learning methods generally improve the performance or predictive accuracy of most prognoses, especially when compared to conventional statistical or expert-based systems. Improve the accuracy of predicting cancer susceptibility, recurrence and mortality. Machine learning predictor against other predictors is critical to choosing the optimal tool.

Ch. Shravya, K. Pravalika, Shaik Subhani proposed Prediction of Breast Cancer Using Supervised Machine Learning Techniques. Breast Cancer is the most often identified cancer among women and major reason for increasing mortality rate among women. This paper is a relative study on the implementation of models using Logistic Regression, Support Vector Machine (SVM) and K Nearest Neighbor (KNN) is done on the dataset taken from the UCI repository. Here, the authors obtained the breast cancer dataset from UCI repository and used spyder as the platform for the purpose of coding. The methodology involves use of classification techniques like Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Logistic Regression, with Dimensionality Reduction technique i.e. Principal Component Analysis (PCA). The work mainly focused in the advancement of predictive models to achieve good accuracy in predicting valid disease outcomes using supervised machine learning methods. The analysis of the results signify that the integration of multidimensional data along with different classification, feature selection and dimensionality reduction techniques can provide auspicious tools for inference in this domain. Achieve good accuracy in predicting valid disease outcomes. But performance is poor.

Dr. S. Senthil & B. Ayshwarya introduced Lung Cancer Prediction using Feed Forward Back Propagation Neural Networks with Optimal Features. In this paper an integrated framework for predicting lung cancer is introduced using Neural Network with Particle Swarm Optimization. Neural networks deals with different problem solving methods in which the neurons are being trained and tested by given database. The lung cancer features are extracted for predicting the cancer stage based on certain feature used in the system. Feature selection is used to identify predictive subsets of cancer cells within a database and reduce the number of cancer cells presented to the computation method. Better performance can be achieved by discarding some features. Lung disease is the uncontrolled development of strange cells that affect one or both lungs. The main objective is to predict and early detection of lung cancer by using neural network with optimal features. Initially the Lung database are collected and given as input to the system. Then data preprocessing is applied on the input images, for the enhancing the image to get the high contrast images. The enhanced images are trained and tested by neural network compared with sample training database. Particle Swarm Optimization (PSO) is applied to extract the features of the given input images and further process is proceeded to detect the lung cancer. On this basis, neural network classifier is employed to analyze those input images features are classified as cancerous or non-cancerous. Provides the earlier warning to the users and it is also cost and time saving benefit to the user.

Jelo Salomon Lung Cancer Detection using Deep Learning. In this paper, the author proposes a method of detecting lung cancer in a CT scan using a 2D-UNet model on a web application. The author cropped 2D cancer masks on its reference image using the center of the lung cancer given in the dataset and trained a model with different techniques and hyper parameters. Finally the result is evaluated using a dice coefficient and confusion matrix metrics. The author reaches 65.7% accuracy on the dice coefficient and an average 0.88% true positive rate and 0.71% false positive rate on a test set of positive and negative samples. A web application has been developed to demonstrate a proof of concept. The application requires a user to upload a CT Scan. The application then processes the file and displays the images to the user. The user then chooses which scan he or she wants to predict then the application pre-processes the CT scan and infers the image to the predictive model. The output of the model is then displayed to the user. The user has the choice to view the images via a carousel or a gallery mode.

Andre Dekker, Cary Dehing-Oberije, Dirk De Ruyscher, Philippe Lambin proposes Survival Prediction in Lung Cancer Treated with Radiotherapy. Bayesian Network (BN) models are a class of machine learning models with some unique

characteristics that make them suitable for medical applications. First of all, they offer a natural way to represent and reason under the uncertainties inherent to the medical domain. Second, they can be described in the form of a graph making them relatively easily interpretable for the medical community and providing an effective means to reason about new links and graphs. Third, BN models take into account the dependence relations between the features, revealing the cause-effect relationships included in medical data. The final goals of this research is to use these knowledge based models as a prior and combine them with the available data to improve performance but maintain interpretability and apply them to individualized radiotherapy decision support. They improved the performance. An important limitation of this work is the small size of the validation set, leading to rather coarse ROC curves and making definitive statements on significance of the observed differences questionable.

Animesh Hazra, Nanigopal Bera & Avijit Mandal proposed Predicting Lung Cancer Survivability using SVM and Logistic Regression Algorithms. One of the major and frequent bases of cancer deaths globally in terms of both instance and transience is lung cancer. The main reason behind the increasing of deaths from it is detecting the disease lately and faults in effective treatment. So, the early detection is needed to save lives from this disease. The survivability rate of lung cancer can be predicted with the help of modern machine learning techniques. Accordingly, it would be clever to determine the survival possibilities among the patients. In this study data cleaning, feature selection, splitting and classification techniques have been applied for predicting survivability of lung cancer as accurately as possible. The clinical data taken from CGD (Clinical and Genomic Data) portal is used for further processing to get the desired outcome. The proposed methodology built here containing of three phases. At first pre-processing of data set is done. Then the dataset will be split into two lays, one is used for training phase and another for testing phase. Next, classification on the dataset using SVM and LR algorithms were performed. It provides computationally efficient classifier for medical applications but time and space complexity is more.

### III. CONCLUSION

From the related work we have observed that in early stage lung cancer, treatment options generally include surgery, chemotherapy, radiation therapy, and radiofrequency ablation. Historically, patients with early stage lung cancer considered inoperable were offered conventional external-beam radiation therapy. Conventional radiation usually involves delivering radiation via two-dimensional beams using a linear accelerator (usually from the front and back). In conventional treatment, planning (simulation) is done using a diagnostic X-ray

machine and radiation is delivered based on these simulation (X-ray) images. Detailed anatomical view (such as in a three-dimensional computed tomography [CT]) is not available; therefore accurate information about dose distribution to the target and surrounding normal structures is lacking.

### 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," *Pattern Recognit. Lett.*, vol. 94, pp. 211218, Jul. 2017.
- [5] W. Shen et al., "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *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, Fedoria Rugless, 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.
- [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.



**Ms. Ayonija Pathre**, MTech,  
Asst. Professor, Rabindranath Tagore  
University, Bhopal, MP, India.  
*Research Area:* Network Security, Data  
Mining



**Prof. Mukesh Kumar**, MTech  
Asst. Professor, Rabindranath Tagore  
University, Bhopal, MP, India.  
*Research Area:* Image Processing &  
Info Retrieval & Network Security.

#### AUTHORS BIOGRAPHY



**Nazia Fatima** has completed BE computer Science and now she is submitting the paper for M Tech Computer science from Rabindranath Tagore University, Bhopal, MP, India.

#### Citation of this Article:

Nazia Fatima, Ayonija Pathre, Mukesh Kumar, “Detection of Lung Cancer by Deep Learning and Machine Learning Techniques” Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 5, Issue 2, pp 48-52, February 2021. Article DOI <https://doi.org/10.47001/IRJIET/2021.502008>

\*\*\*\*\*