

PREDICTION OF LUNG CANCER STAGE WITH ADVANCED TECHNOLOGIES: A COMPARATIVE REVIEW

Prajakta Kale¹, Dr. Santosh K. Yadav

¹*Assistant Professor, Government College of Engineering, Aurangabad, Maharashtra, India*

²*Research Director and Professor, JJT University, Jhunjhunu, Rajasthan, India*

prajakta.kale198@gmail.com , drskyadav@hotmail.com

Abstract: Lung cancer holds the highest position in the list of reasons for deaths in the world. The limitations of most of the Computer Aided Design Systems is that they just predict the whether the given image is cancerous or non-cancerous. It is equally important to study how to detect the stage of the cancer from the extracted nodule. In this paper, we present a comparative analysis of the literature survey in terms of stage classification of lung cancer. The latest research trends are explored presenting their accuracies as well as limitations.

Key Words- lung cancer classification, CAD system, benign or malignant, Neural networks, Deep Learning

1 INTRODUCTION

The survival rate of the lung cancer patients can be dramatically increase if the disease is detected at an early stage. The therapy and medication also depends upon the stage of the cancer. These two facts indicate how important it is to diagnose lung cancer at early disease. The main hindrance in the diagnosis is the overlapping of the symptoms over other lung diseases. The main symptoms of lung cancer include chest pain and infection, cough, loss of appetite and fatigue which are the same as any other normal chronic respiratory disease. This requires a thorough study of the patient's medical history as well as current habits. In case the patient is doubted to have cancer, the further study of his x-rays or scans is needed. These all factors converge upon one thing - the early and accurate diagnosis of lung cancer [1-3].

Considering the anatomical background of lungs [4], [5], the lung is divided into left and right lung as shown in the figure. Both the lungs are covered by a pleural sac with two membranes known as visceral pleura and parietal pleura. The membranes are responsible for protecting the lungs and connecting them to the thoracic cavity. The left lung is smaller in size than the right lung as it contains the heart space. This makes the left lung divided only in two lobes that is upper and lower lobe while the right lung is divided into three lobes, upper, middle and lower lobe. These lobes are separated by the fissure structure with oblique fissures in left lung and oblique and horizontal fissures in right lungs. The windpipe, trachea carries the air from the nostrils to the lungs via the two bronchi's that enter each lung with the part called as carina. The right bronchus gets divided into three lobar bronchi that enter each of three lobes. Similarly, the left bronchus gets divided into two lobar bronchi that enters into two lobes. These lobes are further divided into multiple branches and form an airway tree of bronchi. The deoxygenated blood from the heart is carried via the pulmonary tree structure to the lungs for oxygenation and carried again to the heart after the process. The trees enter and exit the lungs via an anatomical part known as hilum as shown in the figure.

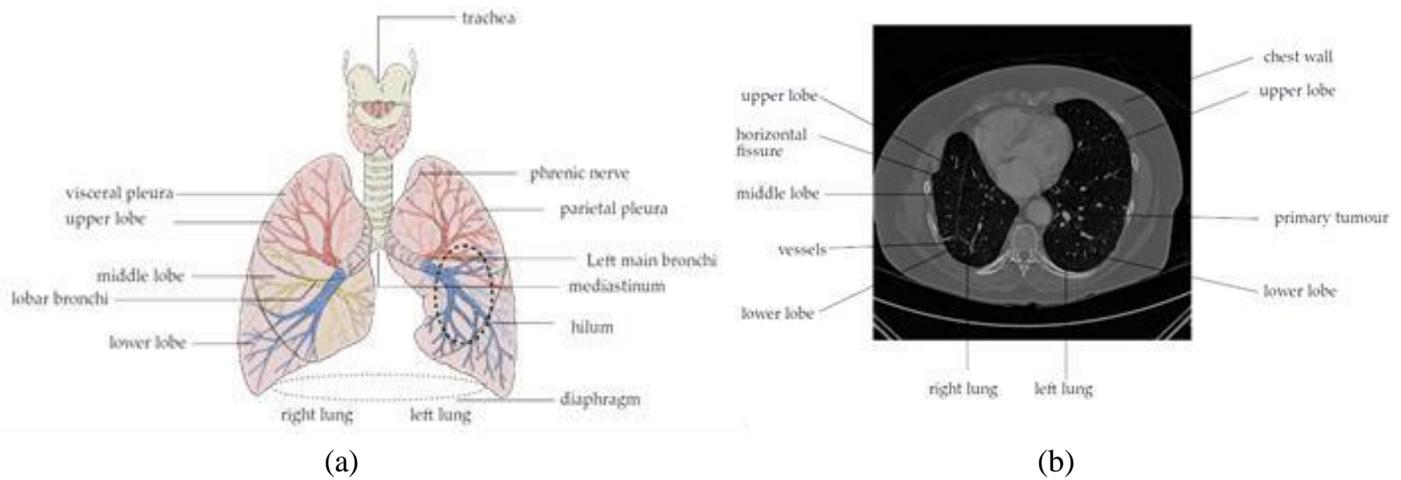


Figure 1. Structure of lungs: (a) Anatomical structure (b) Sample CT Image

Now the next figure illustrates the CT scan with the parts [6]. As we see in the image, we can see that the lung has two different colors – black and grey. The differences in the colors is due to the differences in the densities in terms of radiography. The different parts of the lungs- vessels, tissues, fissures, bones denote various ranges in radiologic densities. These are termed as HU. While studying from a cancer cell perspective, presence of a nodule is considered as the first step which can lead to cancer. But there are more specifications to be considered. An abnormal lung tissue, with circular shape, sharp edges, and 2 to 30 mm in diameter is called a pulmonary nodule. A solitary pulmonary nodule is 30 mm in diameter, with only one side without edges of lung parenchyma. They exist in four different types:

- (a) **Well-circumscribed:** The location of the nodule is at the center without any connection to the neighboring vessels
- (b) **Vascularized:** The location of the nodule is at the center of the lungs with connections to the neighboring blood vessels
- (c) **Juxtapleural:** A good amount of the nodule connects to the periphery of the lungs
- (d) **Pleural tail:** The location of the nodule is at the periphery, and is connected with a thin part called tail. When the nodule size exceeds 30mm in diameter, it is referred to as lung mass or tumor, which is suspicious to lung cancer.

The comprehensive study focuses on the methods proposed in the literature. The structure of the paper is as follows: In the first section we discuss the terminologies and methods for stage predictions of lung cancer. We then discuss about the methods proposed in the literature. We present the comparative analysis of the methods. We then conclude the study with findings of research.

II. METHODS

This section highlights the methods in the literature.

3.1 Anatomical structure of lungs

From the anatomical point of view [7], there are three main classes: T represents the primary tumor's extent, N represents the lymph nodes involvement while the M descriptor represents the distant metastases. Further these three main classes are then divided into sub-categories like T1, T2 depending upon the characteristics. The combinations of these T, N and M descriptors are groups together and the terms as stages of lung cancer.

Summarizing the imaging features of the TNM classification we have the following:

1. **Secondary Primary Lung Cancer:** two or more masses with imaging characteristics
2. **Multifocal Nodules:** multiple nodules with ground glass or solid component
3. **Pneumonic-Type of Adenocarcinoma:** patchy areas of ground glass
4. **Separate Tumor Nodule:** typical lung cancer with separate solid nodule

3.1.1 T component

The T [8] component is divided into five main classes – T0, T1, T2, T3 and T4. These classes are based on the characteristics of sizes and the extent of the nodules into central or mediastina or peripheral structures. If more than one nodule is present, then the position of the second tumors relative to first is determined which further determines the T category. If the primary tumor has invaded the main bronchus, it directly determines the class as T without considering the distance of nodule from carina.

Table 1. T component

T (Primary Tumour)	Description
T0 Tis	No primary tumor Carcinoma in situ
T1 T1a (mi) T1a T1a T1b T1c	Tumour <= 3cm Invasive Adenocarcinoma Superficial tumor spreading in airways Tumor <=1cm Tumor >1 & <=2 Tumor >2 & <=3
T2 T2a T2b	Tumor >3 & <=5 / tumor in visceral pleura/ main bronchus/ atelectasis to hilum Tumor >3 & <=4 Tumor >4 & <=5
T3	Tumor >5 & <=7/ chest wall, pericardium, phrenic nerve/ separate tumors in same lobe
T4	Tumor >7/ mediastinum, diaphragm, heart, carina, trachea, esophagus, spine; tumors in ipsilateral lobe

Processing with the same fashion, if atelectasis has extended to hilum, it is determined as T2a without considering the involvement of lobe or the entire lung. When the diaphragm is involved, it is termed as T4. A pan coast tumor is also termed as T4. When the hilar fat is involved, it is termed as T2a and mediastina fat as T4. When parietal pericardium is involved, is termed as T3. Similarly, when thoracic nerve roots are involved, is termed as T3.

In the cases, when there are multiple categories of T applicable to the tumor, the highest category of T is considered as the category. This means that if we have small tumor with a high T category, it should be classified according to the invasion. On the other hand, if the large tumor of lower degree is classified according to the size of the tumor.

The size of the tumor is the size of the invasive component or the solid component as we say, that determines the size. Now, considering some special and exceptional cases, if we have a superficial tumor that is spreading around central airways, and it is directly determined as T1a, without considering the location. If we have carcinoma in situ, is Tis. T1a is minimally invasive adenocarcinoma.

3.1.2 N Component

The N [9] component is divided into four main categories in the eighth edition: N0, N1, N2, and N3. The category division is based upon the location of the tumor, with the term nodal involvement. Nodal involvement is the extension of primary tumor into the adjacent nodes. There are further divisions of the N component depending upon the detailed locations. However, the sub-groups are not considered into staging of lung cancer as it cannot detect the small tumors.

Table 2. N component

N (Regional Lymph Node)	Description
N0	No regional node metastasis
N1	Metastasis in hilar nodes
N2	Metastasis in submarine nodes
N3	Metastasis in mediastinal/hilar nodes/supraclavicular nodes

3.1.3 M Component

The M [10] component is divided into four categories as: M0, M1a, M1b, and M1c. Most of the M components lack details for classification. However, contralateral or bilateral pulmonary nodules, pleural or pericardial nodules, pleural or pericardial effusion are termed as M1a category. The M1b category involves the single distant metastasis. M1c is the category with multiple metastasis in a single or multiple organs.

Table 3. M Component

M (Distant Metastasis)	Description
M0	No distant metastasis
M1a	Malignant pleural/ pericardial effusion/ nodules/ separate nodule in collateral lobe
M1b	Single extra thoracic metastasis
M1c	Multiple extra thoracic metastases (in a single or multiple organs)

3.2 Staging in Lung cancer

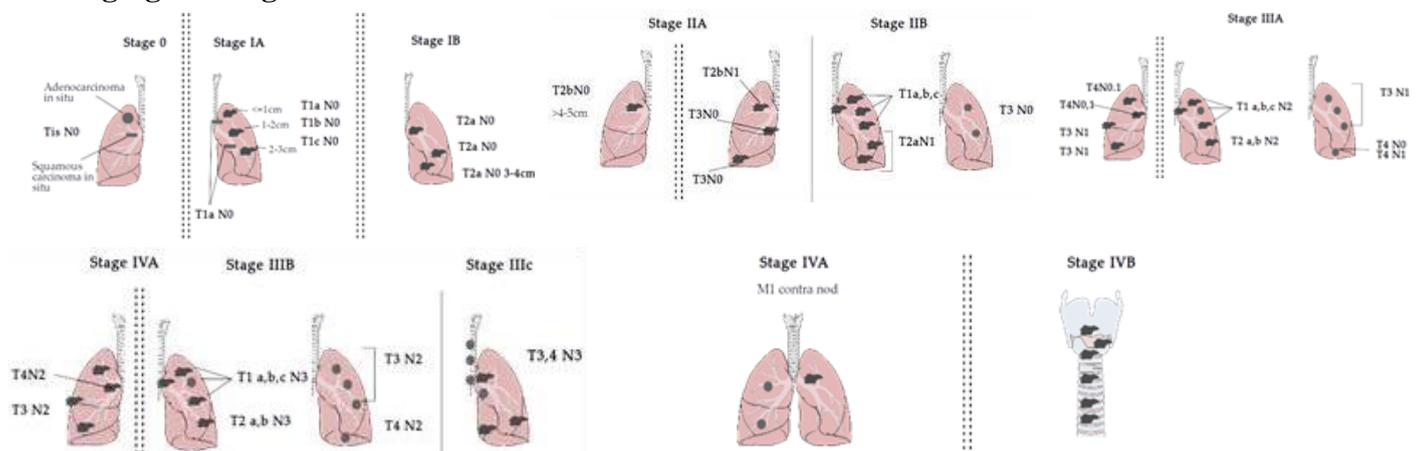


Figure 2. Stages of lung cancer

As already discussed in Introduction chapter, we have stages defined according to the sub-groups of the T, N and M descriptors. The involvement of T1/T2a, N0, and M0 is termed as Stage I. The involvement of T2b/T3, N0, and M0 are termed as Stage II. The stage III is divided into three-subgroups: The involvement of T4, N0, M0; T3/T4, N1, M0; T1/T2, N2, M0 tumors is termed as Stage IIIA. The involvement of T3/T4, N2, M0; T1/T2, N3, M0 is termed as Stage IIIB. The involvement of T3/T4, N3, and M0 tumor is known as Stage IIIB. The Stage IV is also divided into two sub-groups: The involvement of M1a, M1b tumors is known as IVA without considering the T and N categories. The involvement of all

M1c tumors is termed as IVB. Fig 2. depicts the positions of the nodules and accordingly how stages are defined.

3.3 Proposed methods in the literature

One of the main drawbacks in the lung cancer Computer aided diagnostic (CAD) systems is that there is very less research in the direction of staging of lung cancer. In the study of state-of-art, many proposed models exist only up to the detection of nodules in the cancer along with detecting whether it is cancerous or non-cancerous. We have found several research methods in the literature with diverse techniques ranging from image processing till latest technologies like convolutional neural networks (CNN) [11], [12].

In [13], Ignatious et. al has proposed a CAD model for detection of stage in lung cancer. They have used the watershed segmentation technique for segmenting the nodules while enhancing them with the sharpening filters. The detected nodules undergo feature extraction with five main features – area, convex area, perimeter, eccentricity and mean intensity. The extracted features are then provided as input to four classifiers – Support Vector Machines (SVM), Naïve Bayes Multinomial Classifier (NVBM), Random Trees and Naïve Bayes Trees. The methodology is applied to 200 CT images in JPEG format obtained from Regional Cancer Center. The lung nodule detected is classified into one of the stages as T1, T2, T3 and T4. The model with random trees has outperformed other classifiers with an accuracy of 94%.

In [14], Kulkarni and Panditrao have proposed a CAD system model with watershed marker controlled algorithm with Gabor filter for enhancing the images. The nodules segmented then have been used to extract three features – area, perimeter and eccentricity. The extracted features are then fed to SVM classifier for classifying the nodules into four stages like Ignatious et al. i.e. T1, T2, T3 and T4. The research has been carried out with the Computed Tomography (CT) scans obtained from NIH/NCI Lung Image database Consortium (LIDC) database.

Besides traditional classifiers, CNN models are also explored for detection of stages. In [15], Kirienko et. al have proposed a CNN model for stages classification. The research has been carried out on fluorodeoxyglucose positron emission tomography (FDG-PET)/ CT scans obtained from an institutional database. The images are resized as 512 by 512 and clipped between -1000 and 400 Hounsfield units. The images are further normalized between 0 and 1. The lesions, 128X128 in size are cropped around the center and the resulting dataset undergoes data augmentation by rotating the patches around 10 degrees. The CNN model has been divided into two parts. The first extracts the features from the nodules segmented and the second part classifies the nodules into stages. However, the model classifies the nodules into binary classes. The stages T1 and T2 are classified as 0 while T3 and T4 are classified as 1. The testing accuracy has reached 90%.

Moitra and Mandal [16] have proposed a 1-D CNN model for staging of lung cancer with an accuracy of 96%. They have used data from NSCLC Radiogenomics Collection having the PET/CT scan images. The images were resized to 256 by 256 in size followed by Gaussian blurs. The images are then segmented using Otsu's thresholding images. Features with second order derivative of Gaussian are then extracted and fed as an input to 1-D CNN model. The model is used to classify the nodules into T, N and M stage with accuracies as 96%, 94% and 99% respectively.

III. DISCUSSIONS

Table 4 summarizes all the models discussed above with their specifications.

Table 4. Summary of the proposed models in literature

Method	Proposed Model	Dataset used	Staging method	Accuracy
[13]	Random trees	RCC	T1, T2, T3, T4	94%

[14]	SVM	LIDC	T1, T2, T3, T4	-
[15]	CNN	Institutional database	T1, T2 as 0 and T3, T4 as 1	90%
[16]	1-d CNN	NSCLC Radiogenomics	T, N and M	96%, 94% and 99%

The models proposed in [13] and [14] have used watershed and marker controlled watershed algorithm. However, these techniques are tedious to automate. In [14], they have specifically mentioned the numerical results of the model, they have got satisfactory accuracy. In [15] and [16], they have used CNN model for classification of lung cancer staging. The accuracy obtained is quite high. However, [15] have only used binary classification technique. Specific stages have not been mentioned. The [16] model is good enough provided the features are extracted accurately. These learnings will be useful while approaching for a model for stage classification.

IV. CONCLUSIONS

Before proposing a model for CNN, it is necessary to study and analyze the state-of-art in the domain. This paper focuses on understanding the stages of lung cancer, methods to predict those and technologies used in the past to analyze their merits and demerits. Accuracy, simplicity and using less computational resources are the main objectives to be taken into consideration. In future we aim at proposing a model for stage classification based on these learnings.

V. ACKNOWLEDGMENT

REFERENCES

- [1] Mesleh AM, *Jordan Journal of Biological Sciences*, 2017 10:239
- [2] H. Mahersia, M. Zaroug, L. Gabralla, "Lung Cancer Detection on CT
- [3] Aerts, H.J.W.L.; Velazquez, E.R.; Leijenaar, R.T.H.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 2014, 5, 4006.
- [4] Sylvan, E. CT-Based Measurement of Lung Volume and Attenuation of Deceased. Master's Thesis, Linköpings Universisty, Linköping, Sweden, 23 September 2005.
- [5] Walker, C.M.; Rosado-de-Christenson, M.L.; Martínez-Jiménez, S.; Kunin, J.R.; Wible, B.C. Bronchial Arteries: Anatomy, Function, Hypertrophy, and Anomalies. *RadioGraphics* 2015, 35, 32–49.
- [6] Zhou, X.; Hayashi, T.; Hara, T.; Fujita, H.; Yokoyama, R.; Kiryu, T.; Hoshi, H. Automatic segmentation and recognition of anatomical lung structures from high-resolution chest CT images. *Comput. Med. Imaging Graph.* 2006, 30, 299–313.
- [7] Schneider, F., & Dacic, S. (2017). Histopathologic and molecular approach to staging of multiple lung nodules. *Translational lung cancer research*, 6(5), 540–549. <https://doi.org/10.21037/tlcr.2017.06.11>
- [8] Weissferdt, A.; Moran, C.A. *Diagnostic Pathology of Pleuropulmonary Neoplasia*; Springer: New York, NY, USA, 2013; ISBN 978-1-4419-0786-8.
- [9] Oda, H.; Bhatia, K.K.; Oda, M.; Kitasaka, T.; Iwano, S.; Homma, H.; Takabatake, H.; Mori, M.; Natori, H.; Schnabel, J.A.; et al. Automated mediastinal lymph node detection from CT volumes based on intensity targeted radial structure tensor analysis. *J. Med. Imaging* 2017, 4, 044502.
- [10] Loutfi, S.; Khankan, A.; Ghanim, S.A. Guidelines for multimodality radiological staging of lung cancer. *J. Infect. Public Health* 2012, 5, S14–S21.
- [11] G. Chartrand, P. M. Cheng, E. Vorontsov et al., "Deep learning: a primer for radiologists," *RadioGraphics*, vol. 37, no. 7, pp. 2113–2131, 2017.

- [12] Q. Dou, H. Chen, L. Yu, J. Qin, and P. A. Heng, "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, pp. 1558–1567, 2017.
- [13] Ignatious, S.; Joseph, R.; John, J.; Prahladan, A. Computer Aided Lung Cancer Detection and Tumor Staging in CT image using Image Processing. *Int. J. Comput. Appl.* 2015, 128, 29–33.
- [14] Kulkarni, A.; Panditrao, A. Classification of lung cancer stages on CT scan images using image processing. In *Proceedings of the 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, Ramanathapuram, India, 8–10 May 2014; IEEE: Ramanathapuram, India, 2014; pp. 1384–1388
- [15] Margarita Kirienko, Martina Sollini, Giorgia Silvestri, Serena Mognetti, Emanuele Voulaz, Lidija Antunovic, Alexia Rossi, Luca Antiga, Arturo Chiti, "Convolutional Neural Networks Promising in Lung Cancer T-Parameter Assessment on Baseline FDG-PET/CT", *Contrast Media & Molecular Imaging*, vol. 2018, Article ID 1382309, 6 pages, 2018. <https://doi.org/10.1155/2018/1382309>
- [16] Moitra, Dipanjan & Mandal, Rakesh. (2020). Classification of Non-Small Cell Lung Cancer using One-Dimensional Convolutional Neural Network. *Expert Systems with Applications*. 159. 113564. 10.1016/j.eswa.2020.113564.
- [17] Prajakta Kale, Dr. Santosh Yadav "Lung cancer prognosis by implementing various evolutionary Image processing steps." By the International journal of analytical and experimental modal analysis. ISSN NO:0886-9367, Volume XII, Issue IX, September/2020, page no.1103-1117. <http://www.ijaema.com/gallery/130-ijaema-september-4579.pdf>