



LUNG CANCER DIAGNOSIS MODELLING USING BAYESIAN NETWORK

Dr Shribala Nagul¹

1. ECE Department, Matrusri Engineering College, Saidabad, Hyderabad, India

ABSTRACT

With an enormous number of photos, user-assisted segmentation of lung parenchyma pathology-bearing regions becomes perplexing. A new method is proposed for the segmentation of lung tissues from X-ray Computed Tomography (CT) pictures of the chest utilizing a probabilistic graphical model Bayesian Network Model base (BNMB) Image Segmentation. Demonstrating an automated method for seceding the lung parenchyma from the residual chest CT image is the aim of this work. The relationship between the super pixel regions and edge segments is represented by a probabilistic graph constructed from an over-segmentation of the image. We identify regions using an iterative process based on the probabilistic model, and then we merge these regions. Numerous CT image databases are used to evaluate the BNMB, and the results demonstrate improved efficiency and accuracy for both lung CT image segmentation and Region of Interest (ROI) extraction from the affected CT image.

Keywords: Bayesian network, biomarkers, dose-volume metrics, lung cancer, Radiotherapy tumor local control.

1. INTRODUCTION

A deadly illness that affects millions of people worldwide is lung cancer. Most often it is revealed in the later stages, when treatment options may be limited, early detection is essential to achieving better results. A combination of imaging tests, including chest X-rays, CT scans, and PET scans, as well as biopsies to verify the presence of cancerous cells, are usually used to diagnose lung cancer. Globally, lung cancer ranks among the top causes of cancer-related deaths in both men and women. Roughly 80% of cases of lung cancer are categorized as non-small cell lung cancer (NSCLC). When diagnosed, between 25% and 40% of NSCLC patients are in a locally advanced stage III [1]. After radiation therapy, a significant problem in the treatment of patients with locally advanced NSCLC is local failure [2]. Earlier assessment was done using linear and nonlinear models to predict tumor local control, and we have employed a variety of techniques to extract pertinent dose-volume metrics. The objective of this work is to design an effective Bayesian structure learning method for predicting local failure in post-radiation lung cancer patients. Two distinct datasets were used to test our suggested approach. We demonstrate that the suggested approach performs better than traditional Bayesian-based classifiers and that the predictive power can be further increased by adding biological and physical factors to the Bayesian network. A range of methodologies, such as statistical models, machine learning algorithms, and artificial intelligence techniques, can be used to model the

diagnosis of lung cancer. Large patient data sets, such as imaging scans, biopsy results, medical histories, and other pertinent variables that may aid in predicting the propensity for lung cancer, are commonly used to train these models.

2. LITERATURE REVIEW

[3] discusses how the diametric, clinical, biological, and genomic characteristics of patients affect radiation pneumonitis grade 2 (RP2) following non-small-cell lung cancer radiotherapy. RP2 biophysical signalling pathways from a heterogeneous dataset, such as single nucleotide polymorphisms, microRNAs, cytokines, clinical information, and radiation treatment plans before and during radiotherapy, can be interpreted using a Bayesian network (BN) approach. 50 more patients with incomplete data were used for model testing, while 79 patients (21 with RP2) with complete data were used for model building. Relevant predictors were chosen using a large-scale Markov blanket technique that was developed. [4] deals with examining the effectiveness and safety of various treatment options for non-small cell lung cancer and brain metastases in various populations. When it comes to NSCLC patients with brain metastases and EGFR-positive mutations, Osimertinib seems to be the most efficient and secure treatment option. For NSCLC patients with brain metastases who were not selected based on EGFR mutation, the combination of anti-PD1 monoclonal antibody and traditional chemotherapy improves survival; however, this greater benefit is positively correlated with a higher rate of severe adverse events. Given the grave danger that non-small cell lung cancer (NSCLC) poses to human health, a number of clinical trials have shown the advantages of using Chinese herbal injections (CHIs) in conjunction with cisplatin (NP) and vinorelbine [5]. The purpose of this multidimensional network meta-analysis was to determine which CHIs were the best for treating non-small cell lung cancer. For patients with NSCLC, the combination of CHIs and NP is more beneficial than NP by itself. A few of these were noteworthy: Javanica oil emulsion injection plus NP, Shenmai injection plus NP, and Kanglaite injection plus NP. However, additional multicenter RCTs with improved designs are required to corroborate results. A data revolution is occurring in the public health and medical domains. A growing amount of data is available, which has increased interest in machine-learning algorithms. Overview of Bayesian networks (BNs), a machine-learning and knowledge representation tool for risk estimation in medicine is presented in [6]. Three benefits of using BNs for risk modeling over regression-based techniques that are pertinent to health outcomes research are identified: (1) the creation of network structures that facilitate the communication of variable relationships; (2) the application of Bayes's theorem for individual-level risk estimation; and (3) the ease with which they can be converted into decision models. In the era of precision medicine, Bayesian networks are a strong and adaptable tool for analyzing data from outcomes research and health economics. Chemotherapy regimens with adjuvant cytotoxicity for patients whose NSCLC is removed in an early stage. When choosing the best chemotherapy regimen, the clinicians may find some guidance from the analysis [7]. Patients with non-small-cell lung cancer were able to directly compare the safety and effectiveness of various adjuvant chemotherapy regimens thanks to the Bayesian network meta-analysis. The most effective psychosocial intervention for reducing CRF in cancer patients was probably (Mindfulness-based stress reduction) MBSR. Future clinical care for cancer patients should focus on using MBSR, according to medical personnel [8]. This network meta-analysis (NMA) compares the impact of various psychosocial therapies on CRF in a critical manner.

3. PROPOSED WORK

Lung cancer continues to be the primary cause of death. Most recent studies rely on quantitative data, such as size, shape, and the ratio of the affected cells. There is strong evidence that early detection of lung cancer will lower the mortality rate when asymptomatic screening methods are used in conjunction with effective treatment. Medical images are processed using computer vision techniques to extract information, such as the identification of cancerous cells. Converting these continuous images to digital format eliminates a lot of diagnostic uncertainty. Nowadays, medical imaging modalities like computed tomography (CT), magnetic resonance imaging (MRI), and sputum cytology generate enormous amounts of data. Numerous methods have been employed in the literature to identify lung cancer; some are capable of identifying the disease at an early stage, while others are only able to identify an advanced stage. The lung cancer diagnosis classifier is shown in Fig.1.

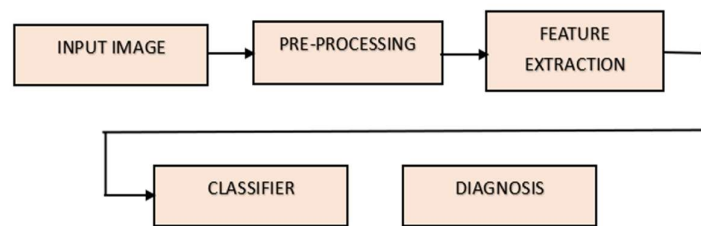


Fig. 1: Lung cancer diagnosis classifier

The four steps of the algorithm are feature extraction, nodule candidate segmentation, image preprocessing, and image acquisition. The segmentation of lung images was done using the active shape model technique, and the texture features were done using the gray level co-occurrence matrix technique. Its invasiveness (exposing patients to radiation) as well as its high false positive and negative rates may be its limitations.

3.1 FLOW CHART

Pre-processing of images is done to cut down on noise and get the images ready for other processes like segmentation. It reduces image distortion and highlights the important details. As a result, an image is corrected. The flow chart of Lung Cancer detection system is shown in Fig.2. It shows the process of how the input image is classified through various stages and output is extracted.

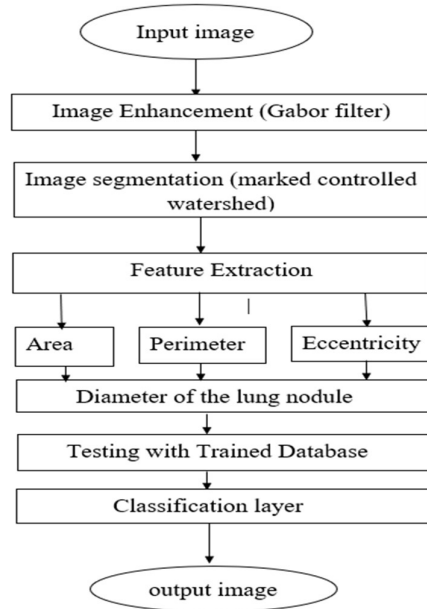


Fig. 2: Flowchart of Lung Cancer detection system

- a) Image Enhancement: The different stages of image processing include gray level conversion, green layer separation, image enhancement, and smoothing. Figure 3 shows the input image. Enhancement techniques come in two varieties: frequency domain and special domain. Because of enhancement, we can either better feed information into image processing techniques or improve the quality of images for human viewers. For improvement, we employ the histogram equalization technique as shown in Fig. 4. In image processing, texture analysis, feature extraction, and image enhancement are all done with the help of linear filters like the gabor filter. Its foundation is the idea of the Gabor wavelet, a mathematical function that bears similarities to the visual cortex's simple cells' receptive fields.

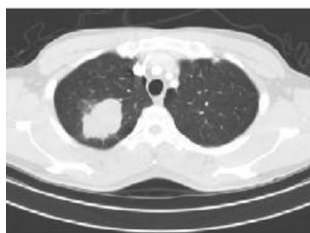


Fig 3: Input image

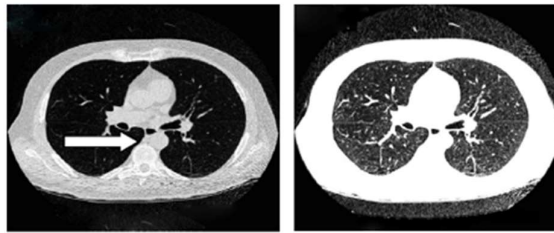


Fig 4(a): Original image Fig 4(b): Enhanced

by Gabor

- b) Image Segmentation: Segmentation is nothing more than the division of an image, and it is commonly used to identify boundaries and objects in an image. We apply the technique of watershed segmentation. Seeds from a watershed segmentation extract show whether there are any objects or backgrounds in the CT scan image. After that, the watershed algorithm is used and the marker location is set to be the regional minimum, usually gradient, of the original input image as shown in Fig.5.

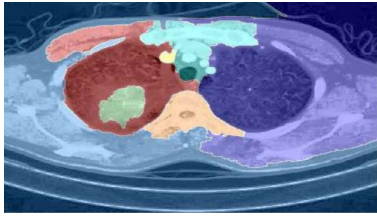


Fig. 5: Image segmentation

AREA	940
LENGTH	347.376
WIDTH	286.912
Eccentricity	1.21074

Fig 6: Feature Extraction

- c) Feature extraction: Following the identification of the cell's nucleus and cytoplasm, we extracted various characteristics that will be applied to the diagnostic procedure to identify cancerous cells. The ability of CAD systems to distinguish between normal and abnormal cells (cancer cells) is the primary challenge facing them in the early diagnosis of lung cancer. Therefore, we can lower or completely eliminate the number of misclassifications by utilizing the relevant features. Various features have been proposed in the literature based on the decision method that has been chosen. We used the following features in our proposed CAD system: perimeter, density, curvature, circularity, eigen ratio, and Nucleus to Cytoplasm (NC) ratio. The first feature is the NC ratio, which is computed by dividing the nucleus area (total number of the pixels in the nucleus region) over the cytoplasm area (total number of pixels in the cytoplasmic region), as follows in the Fig. 6.

$$\text{NC ratio} = \text{Area (Nucleus)} / \text{Area (Cytoplasm)} \times 10$$

$$\text{Eccentricity} = \text{Length of Major Axis} / \text{Length of Minor Axis}$$

- d) Classification: In medical imaging, classification technique is important, particularly for tumor detection and classification. Because lung cancer classification is the last stage of a system where the best results are obtained based on the features available, it is therefore a crucial task for a CAD system. Achieving a high level of true positive rates—even for small cells—and a low number of false positives is the primary goal of the suggested CAD system. In this study, we applied and contrasted support vector machines (SVM) and artificial neural networks (ANNs), two distinct approaches to classification. The use of artificial neural networks (ANNs) in medical image classification is becoming increasingly popular due to their natural ability to approximate functions and make decisions. The input data sets in this work were obtained from features extracted from the sputum color images. We then applied neural network-supervised learning with its specification (multilayer perceptron, MLP) at three layers (input, hidden, and output layers). In the study of cytology, this architecture is frequently employed. Since the ANN performs best in this range, as demonstrated by, the input data has been normalized to fall between 0 and 1.

4. RESULTS

- a) Sample collection: Samples were gathered from more than 200 patients at four Medical Diagnostic Centers in the state of Khartoum. Each patient's complete medical history, including smoking history, was recorded, as were all investigational records. The results of each investigation were then compared to the results produced by this diagnosis model. The

conditional probability percent to use as input data to the program is shown in table 1.

Table 1: Conditional Probability Table

Variables	CP True	CP False	Quantity in Sample
smoking	0.20	0.8	40
Lung cancer	0.30	0.005	60
X-Ray (LC)	0.60	0.02	36
Bronchitis	0.25	0.5	50

b) Visualizing the Bayesian Network as a Graph: The network structure can be visualized using the biograph object. The properties of nodes and edges can be changed as described in appendix. The probabilities of smoking and its results are shown in the Fig.7 which explains the graph structure. Probabilistic inference is the process of determining the probability distribution of variables given certain evidence. Pearls created a message-passing technique for precise inference in singly-connected networks by taking advantage of local independencies among. By propagating beliefs between neighbouring nodes, the method may calculate the conditional probability of any variable given any piece of data. Graph structure with conditional probability is shown in Fig. 8.

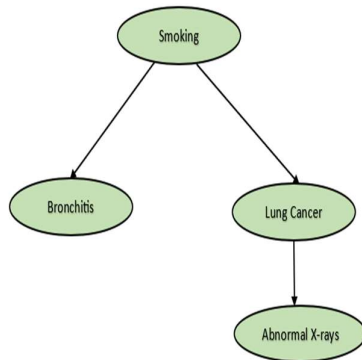


Fig. 7: Bayesian Network – Graph structure

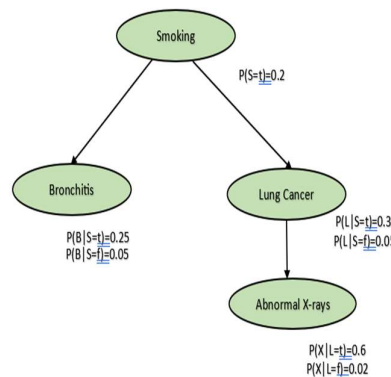


Fig.8: Bayesian Network

Graph -CPT

c) The three situations are

- To evaluate the likelihood that a patient with bronchitis has lung cancer, instantiate $B=t$ (true) and update the network.
- With the observation that the patient has bronchitis ($B = t$), the probability of a true condition for all other nodes has increased. In particular, the probability of smoking history increases because smoking is one leading cause of chronic bronchitis. In turn, because smoking is also associated with lung cancer, the probability of lung cancer increases and so does the probability of an abnormal chest X-ray test. In case the patient has not been evaluated for bronchitis but the chest X-ray shows some

abnormalities, the network instantiated $X = t$ and initialized again with the new evidence.

- Given the observed abnormal X-ray results, the probability of lung cancer has increased significantly because of the direct dependency of node X (X-rays) on node L (lung cancer). Finally, suppose the patient has both been diagnosed with bronchitis and received positive results for his/her chest X-ray.

The three situations can be compared by plotting the probabilities as bar charts as shown in Fig. 9. Evidence of bronchitis and evidence of abnormal X rays increase the probability of lung cancer with smoking history, one directly and the other indirectly. When compare the three situations by plotting the probabilities, it is observed that the evidence of bronchitis and evidence of abnormal X-rays increase the probability of lung cancer with smoking history, one directly and the other indirectly. The effect of a positive versus negative bronchitis diagnosis in presence of abnormal X-ray results can now be compared. Instantiate $B = f$ (false) and compare with previous estimates $B = t$ (true). When bronchitis is ruled out ($B = f$), the probability of smoking history decreases with respect to the case in which the bronchitis is confirmed ($B = t$). The effect is propagated across the network and affects the probability of lung cancer in a similar manner as shown in Fig. 10 below.

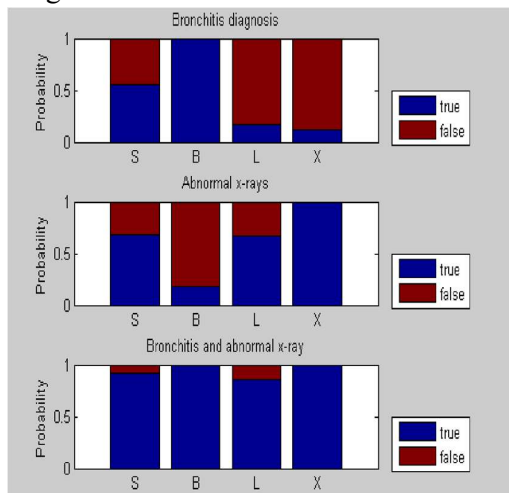


Fig. 9: Comparison of 3 situations by plotting with evidence of probabilities abnormal x-ray Result

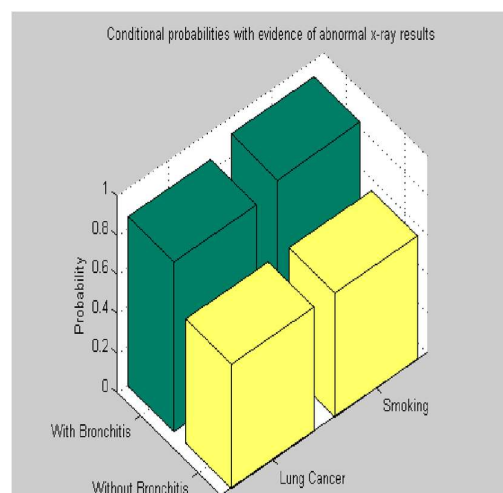


Fig. 10 Conditional probability

5. CONCLUSION

The likelihood of the patient actually having bronchitis has increased, based on the observation that they do. Specifically, as smoking is one of the main causes of chronic bronchitis, the likelihood of having smoked in the past rises. Consequently, the likelihood of both lung cancer and an abnormal chest X-ray test rises because smoking is linked to lung cancer as well. The likelihood of lung cancer with a smoking history is increased by indications of bronchitis and abnormal X-rays, one directly and the other indirectly. In comparison to the instance where bronchitis is verified to be real, the likelihood of smoking history falls when bronchitis is ruled out (bronchitis is false). The influence spreads throughout the network and has a comparable impact on the likelihood of developing lung cancer. In conclusion, research that employs Bayesian networks to model lung cancer has the potential to yield important insights into the

intricate interactions between the variables that affect the onset and course of the illness. Bayesian network models have the potential to enhance diagnosis and therapy by forecasting the chance of acquiring lung cancer, detecting risk factors, and creating customized treatment plans. These efforts can ultimately result in improved patient outcomes.

REFERENCES

- [1] Ahmedin Jemal, Freddie Bray, *Global Cancer Statistics- A Cancer Journal for Clinicians*, 61, 2021, 69-90.
- [2] J.S. Leena Jasmine, A. Govardhan, S. Baskaran, *Classification of Micro classification in Mammogram using Nonsampled Contourlet Transform and Neural Network*, *European Journal of Scientific Research*, 46(4), 2020, 531-539.
- [3] Yi Luo et. al., *Unravelling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis*, *Radiotherapy and Oncology-Journal of European Society*, 123(1), 2017, 85-92.
- [4] Lihui Liu et. al., *Efficacy and safety of treatment modalities across EGFR selected/unselected populations with non-small cell lung cancer and brain metastases: A systematic review and Bayesian network meta-analysis*, *Research article on Lung Cancer*, 158, 2021, 74-84.
- [6] Paul Arora et.al. *Value in health Journal*,22(4), 2019, 439-445.
- [7] Lan-Lan Pang et.al., *Investigation of the optimal platinum-based regimen in the postoperative adjuvant chemotherapy setting for early-stage resected non-small lung cancer: a Bayesian network meta-analysis*, *BMJ Open*, 12(6), 2022
- [8] Yang Yuan et. al., *Effectiveness comparisons of various psychosocial therapies for cancer-related fatigue: A Bayesian network meta-analysis*, *Journal of Affective Disorders*, 2022, 471-481