Indian Journal of Science and Technology, *Vol 9(8), DOI: 10.17485/ijst/2016/v9i8/87962, February 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Effective Lung Cancer Diagnosis: A Survey

K. Prabhavathi* and R. Lakshmi

CSE DEPT, KLNCE, Pottapalayam - 630612, Sivagangai District, Tamil Nadu, India; prabhak.devi@gmail.com, rlaksh_gopi@yahoo.com

Abstract

Background/Objectives: Lung cancer is one of the mostly deadliest cancers across the world. Various approaches have been used for diagnosis of lung cancer. This paper surveys various approaches used for lung cancer diagnosis. **Methods/ Statistical Analysis**: This paper classifies techniques in the following ways, 1) Data mining approach, 2) Medical approach, 3) Biophotonic imaging approach. Also discusses the various pros and cons of these approaches. **Findings:** This paper surveys the different approaches used for lung cancer diagnosis. **Improvements/Applications:** It provides the efficient way for early detection of lung cancer. It reduces the death rate and increases the survival rate.

Keywords: Biophotonics, Biopsy, Optical Surgery, Screening, Tumor

1. Introduction

Lung cancer is the most common cancer in both men and women. Lung cancer accounts for about of all new cancers. Lung cancer accounts for about 27% of all cancer deaths. Lung cancer is the leading cause of cancer death among both men and women[1]. Lung cancer mainly occurs in elderly people. Smoking is the main cause of lung cancer. These include both smokers and nonsmokers. Smokers have high possibility of developing lung cancer when compared to nonsmokers. If lung cancer is detected at an early stage, the possibility of cure is high. Primary cancer can be detected early. If patients are not aware primary cancer can grow into metastasis. In data mining approach various data mining methods such as association rule mining, decision tree classification is used. For association rule mining different algorithm such as Apriori algorithm, hotspot algorithm and k-means and Farthest First Clustering Algorithms[2] are used. In medical approach screening test, biopsy is used. Both the techniques are commonly accepted. But these techniques have their own disadvantages. In Bio photonic imaging approach various probe based imaging techniques such as optical coherence tomography, confocal laser fluorescence endocytoscopy are used. All the three approaches are reviewed. Association rule mining is the best for early detection of lung cancer.

2. Lung Cancer Diagnosis using Different Diagnosis

2.1 Data Mining Approach

2.1.1 Lung Cancer Diagnosis using Apriori Algorithm

Here the data mining techniques are applied for cancer diagnosis. The size and extent of the tumor are described using lung cancer pathologic staging[1]. The patient's stages of cancer are predicted and help the doctor plan appropriate treatment.

The surgery, biopsy is avoided which put a patient's health in danger. The clinical information is only considered and to outplace the pathology report. The association between clinical information and pathology report is studied. The feasibility of applying the clinical information is demonstrated to identify lung cancer using clinical information without surgery[3]. Data mining techniques are used to find an association between clinical and pathological report.

**Author for correspondence*

Association rules are if-then statement that discovers relationship between unrelated data. Apriori algorithm[4] is used to reduce computational effort. Find the frequent item sets: The sets of items that satisfy minimum support, confidence and lift. Use the frequent item sets to generate association rules.

2.2 Methodology used in this Approach

2.2.1 Data Pre-Processing

The clinical and pathological reports are collected. The duplicate data are removed. Provide missing values based on past data.

2.2.2 Decision Tree Classification

Decision tree is used for classification. The attributes with valid values are extracted. Six attributes are identified 'Gender', 'age', 'race', 'reformed smokers', 'length of smoking' and 'pack-year'.

2.2.3 Rules Extraction through Association Mining

The TNM staging[5] is used to categorize the attribute. Interesting rules have been generated. Apriori Algorithm is used for association mining.

2.2.4 Rule Selection

Rules are selected through two approach support-then-confidence-then-lift and confidence-then-lift-then-support.

2.2.5 Rule Evaluation

The rules are evaluated through record matching percentage and accuracy percentage. Using SCL the RM% is higher. This is enumerated as follows (Figure 1).

Using CLS Acc% is higher. This is enumerated as follows (Figure 2).

2.3 Identifying Hotspot in Lung Cancer Data

HotSpot Algorithm is an association rule mining algorithm which is used for identifying hotspots. Identify the average survival time of the patient[6]. The association rule consists of left hand side or antecedent and right hand side or consequent. The consequent is fixed to the target attribute. It can be the average survival time of the patients. The LHS or antecedent defines the segment

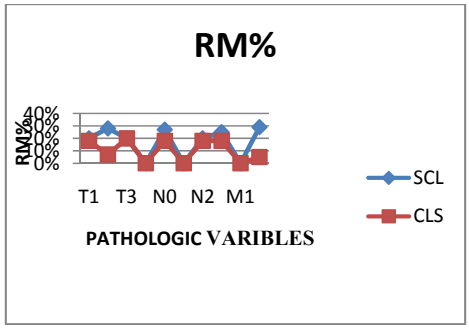

Figure 1. Total RM% of SCL and CLS.

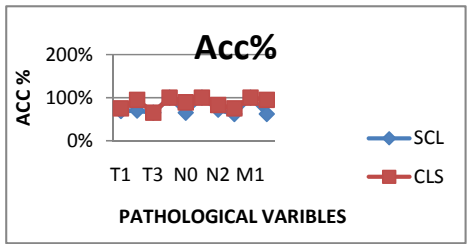

Figure 2. Total Acc% of SCL and CLS.

characteristics for patients. It uses a greedy approach to construct the tree of rules in depth first fashion. The 13 patient attributes are considered, namely 'age', 'birth place', 'diagnosis', 'tumor', 'lymph node', 'surgery performed', 'no surgery', 'radiation therapy', 'lymph node surgery', 'cancer stage', 'past history' and 'lymph node examined'.

2.3.1 Methodology used in this Approach

2.3.1.1 Study Patient Attribute

Study the patient attributes that affects the survival time of the patient.

2.3.1.2 Removal of Redundant Rules

The 2-stage semi-manual procedure is used to remove the redundant rules when using HotSpot algorithm.

2.4 Rule Selection

The lift method is used to compare the average value of a life time of the patient's survival. The cancer grade 1 is well differentiated and grade 2 is moderately differentiated, grade 3 is poorly differentiated.

This is enumerated as follows (Figure 3).

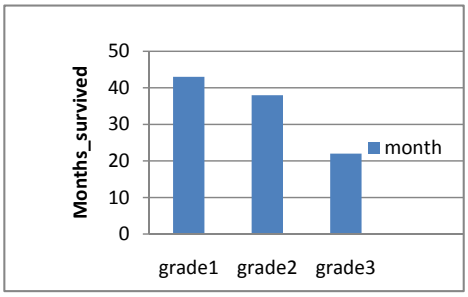

Figure 3. Cancer grade vs. months_survived.

2.5 Medical Approach

2.5.1 Screening

Screening is the most commonly used method to detect lung cancer. Screening is a kind of testing usually done for persons with no symptoms, but likely to develop lung cancer. Low dose computed tomography is the common method for screening. It helps the clinicians in decision making. It describes the benefits, limitations and harms associated with screening for lung cancer[7]. Screening reduces the risk of dying from lung cancer.

2.5.2 Methodology used in this Approach

2.5.2.1 Candidates Suited for Screening

A person who has a history of heavy smoking at least 30-pack-year. Patients who smoke now or quit within past 15 years. Age of the person is between 55 and 74 years. Persons who have a high risk of developing lung cancer is mostly suited for screening.

2.5.2.2 Benefits

Reduce the risk of dying from cancer. It is more cost effective.

2.5.2.3 Limitations

Screening does not detect all lung cancers.

2.5.2.4 Harms

Screening may produce false-positive results. Repeated radiation may even cause cancer in healthy people.

2.5.2.5 Effective Screening Setting

The minimum equipment standard must be met to set up for screening. Standard screening protocol to insure low dose examination. Trained radiologist and technologist are needed to perform the screening.

2.5.2.6 Recommendations

The clinicians must discuss the benefits. Limitation harms to the patient before the screening.

2.6 Biopsy

The biopsy involves the removal of a sample of the tissue from the patient's lung. Patients with lung cancer undergo biopsy under the recommendations of the clinician. The biopsy can be done in four ways.

2.6.1 Methodlogy used in this Approach

2.6.1.1 Bronchoscopic Biopsy

If lung cancer is predicted, this biopsy use bronchoscope to remove the sample of tissue. The bronchoscope is interpolated into the mouth or nose to remove the tissue.

2.6.1.2 Needle Biopsy

The needle biopsy is used when the suspected lung tissue is located near the chest wall. A long needle is inserted into abnormal tissue to remove the lung tissue[8].

2.6.1.3 Open Biopsy

The open biopsy is used when other biopsy cannot detect the lung cancer. Open biopsy is mainly considered when large sample of tissue is needed for lung cancer detection.

2.7 Video-Assisted Thoracoscopic Surgery

It uses thoracoscope that pierce through the chest to remove the tissue.

3. Biophotonic Imaging Approach

Biophotonics is the combination of living being and light. Biophotonics is the development and application of optical techniques, particularly imaging to the study the tissue probe based biophotonic techniques that can image and diffuse lung diseases.

3.1 Optical Surgery

Optical biopsy refers to a technique that uses the properties of light to an instant diagnosis of lung cancer.

Table 1. Comparison between different approaches

Approach	Data Mining Approach		Medical Approach		Biophotonic Imaging Approach		
Criteria	Apriori	HotSpot	Screening	Biopsy	OCT	LCE	Endocytoscopy
Advantages	Support diagnosis of lung cancer	Identify the factors affecting the survival	Reduce the risk of dying from cancer	Provides accurate results	Increases the flexibility of visualizing images	Also used in skin inspection	Avoid the need for biopsy
Disadvantages	No medical evidences	Increased processing time	Does not detect all the cancer	Needs surgery	Low resolution	Low penetration	Requires the contrast agent
Risks	Need support from doctor	Need medical evidences	Radiation risk,false positive results	Risk patient life	Costly	costly	Costly
Need for surgery	No	No	No	Yes	Optical Biopsy	Optical Biopsy	Optical Biopsy
Basis for detection	Patient's life style	Patient's life style	Possibility of developing cancer	When cancer is suspected through X-ray	Uses light property to image the infected area	Uses light property to detect scaffold of lung	Uses light property to differentiate between normal and tumor cells

Previously surgical biopsy is used where a sample of tissue from a patient is taken to complete the report. It takes more time. The life of the patient is at risk. The commonly used biophotonic imaging techniques are Optical Coherence Tomography, Confocal Laser Fluorescence Endomicroscopy and Endocytoscopy. These techniques are discussed below in detail.

3.1.1 Optical Coherence Tomography (OCT)

OCT uses IR light, 15-µm resolution. It can differentiate carcinoma. It has better penetration power.Current technology provides 2-3 mm of penetration; this is acceptable for imaging early bronchogenic carcinomas[9]. It requires no contrast agent for differentiating the images of tissue. Only a short time is required to image makes large-organ.

3.1.2 Laser Fluorescence Endomicroscopy (LFE)

Laser Fluorescence Endomicroscopy is a probe-based imaging technique has been fitted for bronchoscopy. Laser Florescence Endomicroscopy has higher resolution but less penetration than OCT[9]. It uses blue argon laser light for imaging. It requires a contrast agent for better resolution. It is a cable of detecting the acinar scaffold of the lung and the basement membrane under bronchial mucosa.

3.1.2 Endocytoscopy

Endocytoscope is simply a contact microscope that requires a contrast agent to provide resolution of bronchial mucosa[9]. With Endocytoscopy the detailed images of the nuclei is investigated that provides the fine discrimination between normal and malignant cells.

4. Discussion

Table 1 provides the comparison between different approaches. The various criteria such advantages, disadvantages, risk associated, whether there is the need for surgery and the basis for detection are discussed for all the approaches.

5. Conclusion

A survey of several approaches used for early detection of lung cancer is presented and each approach has been discussed in detail. The comparison between different approaches has also been made. Thus this work provides required information on lung cancer detection.

6. References

1. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population

in 2008. International Journal of Cancer. 2013 Mar; 132(5):1133–45.
2. Dharmarajan A, Velmurugan T. Lung cancer data analysis by k-means and farthest first clustering algorithms. Indian Journal of Science and Technology. 2015; 8(15):1–8.
3. Yang H, Chen YP. Data mining in lung cancer pathologic staging diagnosis: Correlation between clinical and pathology information. Expert Systems with Applications. 2015 Sep; 42(15):6168–76.
4. Chai S, Yang J, Cheng Y. The research of improved Apriori Algorithm for mining association rules. IEEE Conference on Service System and Service Management; Chengdu. 2007 Jun 9-11. p. 1–4.
5. Edge SB, Compton CC. The American Joint Committee on Cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. Annals of Surgical Oncology. 2010 Jun; 17(6):1471–4.
6. Agrawal A, Choudhary A. Identifying HotSpots in lung cancer data using association rule mining. 11th IEEE International Conference on Data Mining Workshops (ICDMW); Vancouver, BC. 2011 Dec 11. p. 995–1002.
7. Wender R, Fontham ET, Barrera E, Colditz GA, Church TR, Ettinger DS, Etzioni R, Flowers CR, Gazelle GS, Kelsey DK, LaMonte SJ, Michaelso JS, Sullivan DC, Wolf A, Brawley OW, Smith RA. American cancer society lung cancer screening guidelines. CA: A Cancer Journal for Clinicians. 2013 Mar-Apr; 63(2):106–17.
8. Greif J, Marmor S, Schwarz Y, Staroselsky AN. Percutaneous core needle biopsy vs. fine needle aspiration in diagnosing benign lung lesions. Acta Cytologica.1999 Sep-Oct; 43(5):756–60.
9. Newton RC, Kemp SV, Shah PL, Elson D, Darzi A, Shibuya K, Mulgrew S, Yan GZ. Progress toward optical biopsy: Bringing the microscope to the patient. Lung, 2011 Apr; 189(2):111–9.