Volume 6, Issue 8, August 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Evaluation of Staging Classification in Lung Cancer

[1]K. Arutchelvan, [2]Dr. Pon Periasamy
[1] Programmer (SS), Department of Pharmacy, Annamalai University, Tamilnadu, India
[2] Associate Professor, Department of Computer Science, Nehru Memorial College, Tamilnadu, India

Abstract - The most important cause of death for both men and women is Lung cancer. Lung Cancer is a disease of uncontrolled cell growth in tissues of the lung. Detection of Lung Cancer in its early stage is the key of its cure. Data classification is an important task in KDD (Knowledge Discovery in Databases) process. It has several potential applications. The performance of classifiers is strongly dependent on the data set used for learning. It leads to better performance of the classification models in terms of their predictive or descriptive accuracy, diminishing of computing time needed to build models as they learn faster and better understanding of the models. In this paper, a comparative analysis of data classification accuracy using lung cancer data in different scenarios is presented. The predictive performances of popular classifiers are compared quantitatively.

Keywords: Data Mining, Classification, WEKA tool, Clustering and PET/CT.

I. INTRODUCTION

In the modern age of computerized fully automated trend of living, the field of automated diagnostic systems plays an important and vital role. An automated diagnostic system design in Medical Image analysing is one such field where numerous systems are proposed and still many more under conceptual design due explosive growth of the technology today. Lung cancer is considered to be the main cause of cancer death worldwide, and it is difficult to detect in its early stages because symptoms appear only in the advanced stages causing the mortality rate to be the highest among all other types of cancer. More people die because of lung cancer than any other types of cancer such as breast, colon, and prostate cancers.

There is significant evidence indicating that the early detection of lung cancer will decrease mortality rate. There are many techniques to diagnose lung cancer, such as Chest Radiography (x-ray), computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. However, most of these techniques are expensive and time consuming. In other words, most of these techniques are detecting the lung cancer in its advanced stages, where the patients' chance of survival is very low. Therefore, there is a great need for a new technology to diagnose the lung cancer in its early stages. Image processing and data mining techniques provide a good quality tool for improving the manual analysis.

II. NON-INVASIVE STAGING OF LUNG CANCER COMPUTED TOMOGRAPHY

Computed Tomography (CT) is usually the initial imaging modality used to de fine the anatomic extent of lung cancer. It provides not only staging information but also provides information about anatomic relationships that are important for surgical or radiation therapy planning. CT scans for lung cancer staging should be contrast enhanced to define anatomic relationships between the tumor and the mediastinal and pulmonary vessels and also to visualize mediastinal and hilar lymph nodes. CTs should be performed at a minimum of 3.75-mm collimation (slice thickness) and should extend from the supraclavicular region to below the kidneys to allow for visualization of supraclavicular lymph nodes, both adrenal glands, and the majority of the liver.

CT scans can provide rough information about tumor size, but at the stage-determining cut points (e.g., 3, 5 and 7 cm), it is not sufficiently sensitive to determine T-stage. However, at the extremes of tumor size, T-stage may be reliably predicted. Thus, a tumor measuring 1 cm on CT will most likely be a true T1 tumor at pathology, and a 10 cm tumor will likely be a T3 tumor. For tumors that abut the surface of the lung, CT cannot reliably determine visceral pleural invasion, and so the differentiation of T1 from T2 (by visceral pleural invasion) is not possible, though pleural puckering is often suggestive of visceral pleural invasion. Chest wall invasion is often obvious on CT but may also be subtle enough to not be detected except at the time of surgery. Tumors extending into tissue outside of the chest wall or rib erosion are usually reliable predictors of chest wall invasion (T3), and the accuracy of CT for predicting chest wall invasion has been reported to be only 39–87%. Invasion of mediastinal structures such as the esophagus, mediastinal fat, pericardium, and trachea can similarly be imprecise on CT unless there is extensive invasion. Generally, additional con firmatory tests will be required to establish invasion such as bronchoscopy, esophagoscopy, and esophageal (EUS) or endobronchial ultrasound (EBUS), and clinical findings such as phrenic nerve paralysis or intrascapular pain (suggestive of aortic involvement) can aid in establishing T4 status when coupled with CT findings. Invasion of the spine can often be diagnosed if bone destruction is evident.

III. CT AND MEDIASTINAL NODAL STAGING

By convention, a cut-off point of 1 cm or greater is used to differentiate a potentially met astatically involved node from a non-involved node. Unfortunately, the accuracy of CT for predicting metastases to the mediastinal nodes is low, the false-positive rate of CT in the diagnosis of nodal metastases is between 15 and 80 %, and the false-negative rate is as high as 12 %. A recent meta-analysis by Toloza et al. included 3,438 patients from 20 studies and found that pooled sensitivity and specificity of CT scanning was 57 and 82 %, respectively, with a positive predictive value of only 56 %. An inherent limitation of CT prediction of nodal metastases is that lymph node size is not a reliable indicator of nodal metastatic disease. In a recent analysis of 256 patients (2,891 nodes), Prenzel and colleagues reported that at least 77 % of patients without nodal metastases had at least one lymph node greater than 1 cm in diameter. Furthermore, 12 % of patients with N2 metastases had no enlarged nodes on CT scanning.

TNM Stage Grouping			
Anatomic stage / Prognostic groups			
Occult carcinoma	Tx	No	Mo
Stage 0	Tis	No	Mo
Stage IA	T1a	No	Mo
	Tib	No	Mo
Stage IB	T2a	No	Mo
Stage IIA	T2b	No	Mo
	T1a	N1	Mo
	Tib	No	Mo
	T2a	No	Mo
Stage IIB	T2b	No	Mo
	T3	No	Mo
Stage IIIA	T1a	N2	Mo
	T1b	N2	Mo
	T2a	N2	Mo
	T2b	N2	Mo
	T3	N1	Mo
	T3	N2	Mo
	T4	No	Mo
	T4	N1	Mo
Stage IIIB	T1a	N3	Mo
	T1b	N3	Mo
	T2a	N3	Mo
	T2b	N3	Mo
	T3	N3	Mo
	T4	N2	Mo
	T4	N3	Mo
Stage IV	Any T	Any N	M1a
	Any T	Any N	M1b

IV. T3 TUMOR MORE THAN 7 CM OR ONE THAT DIRECTLY INVADES ANY OF THE FOLLOWING

Parietal pleural (PL3), chest wall (including superior sulcus tumors), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium; or tumor in the main bronchus less than 2 cm distal to the carina1 but without involvement of the carina; or associated atelectasis or obstructive pneumonitis of the entire lung or separate tumor nodule(s) in the same lobe T4 Tumor of any size that invades any of the following: mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina, separate tumor nodule(s) in a different ipsilateral lobeDistant Metastasis (M) M0 No distant metastasis M1 Distant metastasis M1a Separate tumor nodule(s) in a contralateral lobe, tumor with pleural nodules or malignant pleural (or pericardial) effusion2 M1b Distant metastasis (in extrathoracic organs). Regional Lymph Nodes (N) NX Regional lymph nodes cannot be assessed N0 No regional lymph node metastases N1 Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension N2 Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s) N3 Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s).

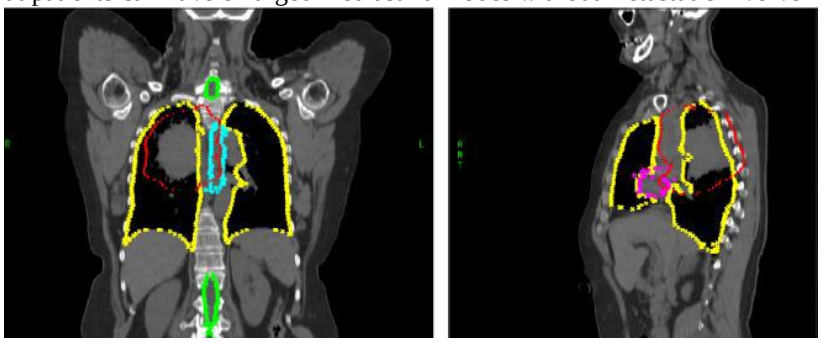
V. POSITRON EMISSION TOMOGRAPHY (PET)

A major advance in the preoperative staging of lung cancer has been the development of positron emission tomography (PET). PET works by taking advantage of the fact that cancer cells and cells with high metabolic activity have increased cellular uptake of glucose and increased rates of glycolysis. Radio labelled fluoro deoxy glucose (FDG), a glucose analog, undergoes uptake by cancer cells but following phosphorylation is unable to undergo further meta bolization and is retained intracellularly. The intensityOf FDG uptake can be quantified (standardized uptake value or SUV) and is considered abnormal if greater than 2.5 or if it is above the normal background uptake in the mediastinum. PET provides information about the metabolic activity of a lesion, but its spatial resolution is poor compared to the precise anatomic information that CT scan provided. The development of integrated PET/CT scanners where 2D axial

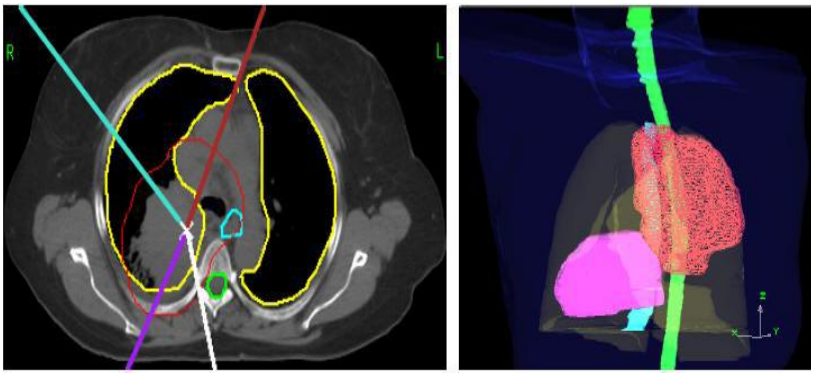
CT images are co-localized with colorized PET data has enabled more accurate assessment of location of metabolic activity than either modality when performed separately. It should be noted that the CT component of most integrated PET/CT scans is performed without intravenous contrast, making detailed information about the relationships of pulmonary vasculature to tumors or mediastinal and hilar nodes not possible. Furthermore, as the CT is performed without breath holding, the lung parenchyma is under inflated and subject to motion artefact which obscures detail. This is particularly relevant when evaluating small (<5 mm) intraparenchymal nodules. PET/CT rarely provides additional information regarding T-stage over what is available from a contrast-enhanced CT, though there is some evidence to suggest that the level of FDG uptake may offer prognostic information regarding survival. Its main role in the staging and pre-treatment assessment of lung cancer lies in its ability to predict mediastinal nodal and distant metastases.

VI. PET/CT AND MEDIASTINAL STAGING

PET/CT has significantly improved the ability to predict the presence of metastatic disease in mediastinal nodes. CT is limited by the fact that patients can have enlarged mediastinal nodes without metastatic involvement, and conversely.



The figure shows a single CT image slice of a representative lung cancer patient, with contours for the various OARs, together with a 3D representation the CT images showing the tumor wrapping around the esophagus. The dimension of the space of possible 4.



The figure shows the left three panels show CT image slices in the treatment plane for patient #8 in our study; the rightmost panel shows a beam's eye view of the same patient's case. The PTV is shown in red colour, the lungs in yellow color, and the spinal cord in green colour, the heart in pink colour, and the esophagus in cyan. The solid lines in the third panel show the beam directions. Treatments are so larger for these locally-advanced lung cases than for prostate gland cases, because the larger treatment fields contain a greater big total number of beamless. cancer can involve nodes without necessarily causing their enlargement (defined as >1 cm in short axis dimension). PET, however, relies on metabolic activity of tissue rather than on anatomic size criteria and thus has an advantage over CT in this regard. Standardized quantitative criteria for an abnormal PET scan finding in the mediastinum have not been defined; however, an SUV > 2.5 or uptake greater than the background activity of lung or liver is frequently used. False positivity may occur within flammatory tissue. Thus, patients with granulomatous in flammation or infection may have nodes with increased FDG uptake leading to falsely positive scans. Furthermore, there must be a critical mass of FDG-avid tissue to allow its detection, thus nodes with only microscopic involvement with metastases will not appear FDG avid.

The spatial resolution with current generation PET scanners is about 7 mm. accordingly, an FDG-avid node smaller than this is generally highly suspicious for metastatic involvement, whereas a non-avid node may still harbour tumor. A recent analysis of 2,865 patients with lung cancer demonstrated a pooled sensitivity and specificity of 74 and 85 %, respectively. Most studies have reported negative predictive values of 90 % or greater, and in general, a patient with a negative PET scan and a peripheral lung cancer may proceed to definitive therapy without further evaluation of the mediastinum. The American College of Surgeons Oncology Group (ACOSOG) ZD0050 trial prospectively evaluated the use of PET (not integrated PET/ CT) in the preoperative staging and work-up of patients with proven lung cancer, 303 patients were enrolled in the study, of which 76 % were clinical stage I prior to PET scanning. PET identified unsuspected metastatic disease in 6 %; however, PET over diagnosed metastatic disease in 7 % of patients who were subsequently found to be without metastases.

PET identified 82 of 302 (27 %) patients as having N2/3 disease; however, only 46/82 (56 %) were true positives (positive predictive value = 56 %, negative predictive value = 87 %). Overall, PET identified 61 (20 %) patients in whom nontherapeutic thoracotomy could potentially be avoided. There was a relatively high rate of false positives in this study (44 %), which strongly argues in favour of histologic confirmation of PET-positive mediastinal nodes. Indeed, both the American College of Chest Physicians and the European Society of Thoracic Surgeons recommend mediastinal nodal sampling to con firm PET-positive mediastinal nodes. Patients with small peripheral lung cancers that have negative nodes by PET and CT criteria may safely undergo either surgical resection or ablative therapy as their physiologic performance allows. The incidence of occult N2 metastases in this group of patients is between 5 and 11 %. Patients with central tumors, however, are at signi fi cantly higher risk for occult N2 disease. In a study by Lee et al., patients with peripheral tumors smaller than 2 cm and a radiologically negative mediastinum had a 3 % incidence of occult N2 nodal metastases but increased to 22 % if the tumor was centrally located.

The incidence of occult N2 disease increased substantially for tumors that were greater than 2 cm and centrally located and approached 30 %. Thus, even patients who have negative mediastinal nodes by PET and CT should be considered for histologic assessment of the mediastinal nodes if they have central tumors that are greater than 2 cm in size. The same applies to cases where the risks of surgery are great and the marginal bene fi t of an operative approach would be greatly offset by the presence of occult nodal disease, examples might include patients who require pneumonectomy, resection of Pancoast tumors, and patients with severe comorbidities or of extremely advanced age.

VII. INVASIVE STAGING FOR LUNG CANCER

Invasive surgical staging is generally performed either with surgery (mediastinoscopy, anterior mediastinotomy [Chamberlain procedure], or VATS lymph node sampling) or via endoscopy (transbronchial needle aspiration, with or without endobronchial ultrasound [EBUS] guidance, and esophageal ultrasound [EUS] fine-needle aspiration). Occasionally transthoracic CT-guided biopsy may be performed to sample nodes not immediately adjacent to the tracheobronchial or esophageal structures and therefore inaccessible to mediastinoscopy or endoscopic methods.

VIII. CERVICAL MEDIASTINOSCOPY

Cervical mediastinoscopy (CM) was fi rst described by Carlens at the Karolinska Institute, Sweden in 1959. It has remained the gold standard method of histologic sampling of mediastinal nodes since then but will likely be replaced by less invasive endoscopic methods, as a growing body of literature suggests similar efficacy of EBUS and EUS. Media stinoscopy is performed as an outpatient surgery under general anaesthesia and is now more commonly performed with a video media stinoscope, which greatly improves visualization. The patient is placed in a supine position, and a 2-cm transverse incision is placed 1 cm above the suprasternal notch. The platysma is divided and the strap muscles separated in the midline, exposing the underlying anterior surface of the trachea. Incision of a thin layer of investing fascia allows entry into the avascular pretracheal space. The mediastinoscope is then directed into this space under direct vision. Further dissection is performed bluntly with the aid of a suction/cautery device. The development of the video media stinoscope significantly enhanced visualization and now allows for extremely accurate dissection and nodal sampling (Fig.). The media stinoscope can readily access stations 2R, 2L, 4R, 4L, and 7, and attempts to sample nodes from these stations should be made. It is extremely unusual not to be able to sample nodal tissue from the 4R, 4L, and 7 stations in nearly every patient. Nodes in the 2L region are not infrequently absent or diminutive. Furthermore, excessive dissection in the left paratracheal region does place the left recurrent laryngeal nerve at potential risk for traction injury, so judgment is required.

IX. CLASSIFICATION

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fraudulent and valid activities determined on a record-by-record basis. The classifier-training algorithm uses these preclassified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier. Some well known classification models are: a) Classification by decision tree induction b) Bayesian Classification c) Neural Networks d) Support Vector Machines (SVM).

X. EXPERIMENTAL STUDY AND ANALYSIS

WEKA Tool

The system use WEKA (www.cs.waikato.ac.nz/ml/weka/), an open source data mining tool for our experiment. WEKA is developed by the University of Waikato in New Zealand that implements data mining algorithms using the JAVA language. WEKA is a state-of-the-art tool for developing machine learning (ML) techniques and their application to real-world data mining problems. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. WEKA implements algorithms for data pre-processing, feature reduction, classification, regression, clustering, and association rules. It also includes visualization tools. The new machine learning algorithms can be used with it and existing algorithms can also be extended with this tool.

Dataset Description: Performance of computer simulation on a lung cancer dataset available UCI Machine Learning Repository. The features describe different factor for cancer reoccurrence. The dataset contains 768 instances shown as no reoccurrence-events while 85 instances as reoccurrence-events.

Classifier	Confusion Matrix	Precision Avg.	Recall Avg.	F-Measure
ZEROR	a b <-- classified as 0 9 | a = 1 0 23 | b = 2	0.719	0.517	0.719
NAIVE BAYES	a b <-- classified as 5 4 | a = 1 3 20 | b = 2	0.356	0.775	0.781
LOGISTIC	a b <-- classified as 6 3 | a = 1 3 20 | b = 2	0.813	0.813	0.813
1BK	a b <-- classified as 3 6 | a = 1 4 19 | b = 2	0.667	0.688	0.674
BAGGING	a b <-- classified as 3 6 | a = 1 0 23 | b = 2	0.851	0.813	0.776

XI. CONCLUSION

An experiment Conducted to find the impact of lung cancer data on the predictive performance of different classifiers. The system selects five popular classifiers considering their qualitative performance for the experiment. After analysing the quantitative data generated from the computer simulations and find that the general concept of improved predictive performance of all above classifiers but 1BK performance is not significant.

REFERENCE

[1] Wang X And Gotoh O, "Cancer Classification Using Single Genes‖, Genome Informatics, Vol. 23, Pp.179-188, 2009.

[2] Dechang Chen “Developing Prognostic Systems of Cancer Patients by Ensemble Clustering” Hindawi publishing corporation, Journal of Biomedicine and Biotechnology Volume 2009, Article Id 632786.

[3] Mallika Rangasamy And Saravanan Venketraman,"An Efficient Statistical Model Based Classification Algorithm For Classifying Cancer Gene Expression Data With Minimal Gene Subsets‖, International Journal Of Cyber Society And Education, Vol. 2, No. 2, Pp.51-66, 2009.

[4] Wang, X., And Gotoh, O., Microarray-Based Cancer Prediction Using Soft Computing Approach, Cancer Informatics, 7:123–139, 2009.

[5] Zainuddin. Z and Pauline. O, "Improved wavelet neural network for early diagnosis of cancer patients using microarray gene expression data", International Joint Conference on Neural Networks, 2009. IJCNN 2009.

[6] Osareh. A and Shadgar. B, "Microarray data analysis for cancer classification", 5th International Symposium on Health Informatics and Bioinformatics (HIBIT), 2010.

[7] Kai-Lin Tang, Wei-Jia Yao, Tong-Hua Li, Yi-Xue Li And Zhi-Wei Cao, ―Cancer Classification From The Gene Expression Profiles By Discriminant Kernel-Pls‖,Journal Of Bioinformatics And Computational Biology,Vol.8,Suppl.1(2010) 147-160.

[8] A. Bharathi And Dr.A.M.Natarajan,"Cancer Classification Of Bioinformatics Data Using ANOVA ", International Journal Of Computer Theory And Engineering, Vol. 2, No. 3, 1793-8201,2010.

[9] R. Mallika, And V. Saravanan, "An SVM Based Classification Method For Cancer Data Using Minimum Microarray Gene Expressions", World Academy Of Science, Engineering And Technology 62 2010.

[10] N. Revathy And R. Amalraj, "Accurate Cancer Classification Using Expressions Of Very Few Genes", International Journal Of Computer Applications, Vol.14, No.4, 2010.

[11] Ritu Chauhan “Data clustering method for Discovering clusters in spatial cancer databases” International Journal of Computer Applications (0975-8887) Volume 10-No.6, November 2010.

[12] Rajashree Dash “A hybridized K-means clustering approach for high dimensional dataset” International Journal of Engineering, Science and Technology Vol. 2, No. 2, 2010, pp. 59-66.

[13] G. Rajkumar “ Intelligent Pattern Mining and Data Clustering for Pattern Cluster Analysis using Cancer Data” International journal of Engineering Science and Technology Vol. 2(12), 2010, ISSN: 7459-7469

[14] Santanu Ghorai, Anirban Mukherjee, Sanghamitra Sengupta, And Pranab K. Dutta, " Cancer Classification From Gene Expression Data By NPPC Ensemble", IEEE/Acm Transactions On Computational Biology And Bioinformatics, Vol. 8, No. 3, May/June 2011.

[15] K.Kalaivani “Childhood Cancer-a Hospital based study using Decision Tree Techniques” Journal of Computer Science 7(12): 1819-1823, 2011 ISSN: 1549-3636

[16] Shyi-Ching Liang, Yen-Chun Lee and Pei-Chiang Lee , “The Application of Ant Colony Optimization to the Classification Rule Problem”, 2011 IEEE International Conference on Granular Computing.

[17] Arezoo Modiri and Kamran Kiasaleh," Permittivity Estimation for Breast Cancer Detection Using Particle Swarm Optimization Algorithm", 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA, August 30 - September 3, 2011

[18] Yao Liu and Yuk Ying Chung, "Mining Cancer data with Discrete Particle Swarm Optimization and Rule Pruning", IEEE 2011.

[19] Hnin Wint Khaing," Data Mining based Fragmentation and Prediction of Medical Data", IEEE 2011.

[20] M. H. Mehta," Hybrid Genetic Algorithm with PSO Effect for Combinatorial Optimisation Problems", International Journal of Advanced Computer Research (IJACR), Volume-2 Number-4 Issue-6 December- 2012.

[21] Parag Deoskar, Dr. Divakar Singh, Dr. Anju Singh, "Mining Lung Cancer Data And Other Diseases Data using Data Mining Techniques: A Survey" Volume 4, Issue 2, March – April (2013).

[22] T.Karthikeyan, P.Thangaraju, "Analysis of Classification Algorithms Applied to Hepatitis Patients", International Journal of Computer Applications, volume-62-No.5, January 2013.

[23] Chinnappan Ravinder Singh, Kandasamy Kathiresan, "Molecular understanding of lung cancers-A review" Centre of Advanced Study in Marine Biology, Faculty of Marine Sciences, Annamalai University, Parangipettai-608 502, Tamil Nadu, India, 2014.

[24] Lawrence A. Loeb, Virginia L. Ernster, Kenneth E. Warner, John Abbotts, and John Laszlo "Smoking and Lung Cancer", on July 17, 2014.

