IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Development Of A Lung Cancer Prediction System Using Machine Learning Techniques

R.Jyosna Devi[1] ,A.V BhanuRavali[2]

Assistant Professor [12], Department of MCA[1], Dept.of CSE[2]

Sree Venkateswara College of Engineering, Nellore, AP[12].

Abstract:

Cancer remains the leading cause of mortality among both genders. The timely identification of cancer significantly enhances the chances of complete recovery. Consequently, there is a growing demand for methods that can detect cancer nodules at an early stage. Lung cancer, in particular, is frequently misdiagnosed. Early detection of lung cancer is crucial, as it can save countless lives; failure to do so may result in severe complications and potentially fatal outcomes. The prognosis and treatment success largely hinge on the disease's early detection and diagnosis. Diagnostic errors rank among the most prevalent forms of medical malpractice worldwide. The fields of knowledge discovery and data mining have numerous applications in both business and scientific research. The implementation of data mining techniques in the healthcare sector can yield valuable insights. This study provides a brief overview of the potential application of classification-based data mining techniques, including Rule-based systems, Decision Trees, Naïve Bayes, and Artificial Neural Networks, to analyze extensive healthcare datasets.

Keywords: Lung cancer, Naive Bayes, Decision Tree, ANN, Data Mining, Classification.

I. INTRODUCTION

Lung cancer ranks among the foremost causes of cancer-related fatalities in both men and women. In many instances, the early symptoms of lung cancer become apparent in the patient's body. The approach to treatment and the prognosis are influenced by the histological type of cancer, the stage of progression, and the patient's overall health status. Treatment options may encompass surgery, chemotherapy, and radiotherapy. Survival rates are contingent upon the stage of the disease, the patient's general health, and various other factors; however, the overall five-year survival rate for individuals diagnosed with lung cancer is only 14%.

Symptoms that may suggest lung cancer include:

- dyspnea (shortness of breath with activity),
- hemoptysis (coughing up blood),
- chronic coughing or change in regular coughing pattern,
- wheezing,
- chest pain or pain in the abdomen,
- cachexia (weight loss, fatigue, and loss of appetite),
- dysphonia (hoarse voice),
- clubbing of the fingernails(uncommon),
- dysphasia(difficulty swallowing),
- Pain in shoulder ,chest , arm,
- Bronchitis or pneumonia,
- Decline in Health and unexplained weight loss.

The incidence of mortality and morbidity associated with tobacco consumption is significantly elevated. Typically, lung cancer originates within the walls or epithelium of the bronchial tree; however, it can initiate in any region of the lungs and impact various areas of the respiratory system. This form of cancer predominantly affects individuals aged 55 to 65 and frequently requires many years for its development.

II. MAJOR TYPES OF LUNG CANCER.

Lung cancer is classified into two primary types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), also known as oat cell cancer. Each type exhibits distinct growth patterns and methods of metastasis, necessitating different treatment approaches. In cases where the cancer displays characteristics of both types, it is referred to as mixed small cell/large cell cancer. NSCLC is more prevalent than SCLC and typically progresses at a slower rate. In contrast, SCLC is closely associated with smoking, tends to grow more rapidly, and often results in large tumors that can disseminate extensively throughout the body. These tumors frequently originate in the bronchi, located near the center of the chest. The mortality rate associated with lung cancer is significantly influenced by exposure to various environmental carcinogens.

III. PRE-DIAGNOSIS TECHNIQUES

Pre-diagnosis plays a crucial role in identifying or refining the potential for lung cancer screening. Factors such as symptoms and risk elements—including smoking, alcohol use, obesity, and insulin resistance demonstrated a statistically significant impact during the pre-diagnosis phase. The challenges associated with diagnosing and prognosticating lung cancer primarily revolve around the widely debated classification issues. These challenges have garnered considerable attention from researchers in the fields of computational intelligence, data mining, and statistics. While cancer research is predominantly clinical and/or biological, data-driven statistical research has increasingly become a valuable complement. One of the most intriguing and

complex endeavors in this domain is predicting disease outcomes, which presents opportunities for the development of data mining applications. With the advent of computers equipped with automated tools, vast amounts of medical data are being collected and made accessible to medical research teams.

Knowledge Discovery in Databases (KDD), which encompasses data mining methodologies, has emerged as a widely utilized research instrument for medical scientists aiming to uncover and leverage patterns and correlations among a vast array of variables. This approach enables researchers to forecast disease outcomes based on historical cases archived in datasets. The purpose of this study is to provide a comprehensive summary of various review and technical articles focused on lung cancer diagnosis. It presents an overview of ongoing research involving different lung cancer datasets, utilizing data mining techniques to improve diagnostic accuracy for lung cancer. Key preventive measures include smoking cessation, dietary changes, and chemoprevention.

Screening serves as a strategy for secondary prevention. Our approach to identifying potential lung cancer patients relies on a thorough examination of symptoms and risk factors. Non-clinical symptoms and risk factors are recognized as general indicators of cancer-related diseases. Environmental influences significantly contribute to the development of cancer in humans. Numerous carcinogens exist in the air we inhale, the food we consume, and the water we utilize. The persistent and often inescapable exposure to these environmental carcinogens complicates the analysis of cancer causation in humans.

IV.LATEST TRENDS IN LUNGS CANCER DETECTION

The main objective of the research presented is to illustrate the methodologies and strategies utilized in the deep learning classification of lung cancer. The deep learning algorithm represents the latest advancement that aids healthcare professionals in diagnosing diseases and assists radiologists in identifying challenging conditions such as lung cancer. This study focuses on the prediction and classification of cancer. Various techniques associated with deep learning algorithms and concepts are discussed in this paper (see Fig. 5). The Convolutional Neural Network (CNN) is composed of several layers, including a convolutional layer that extracts image features, a pooling layer that selects these features, and a fully connected (FC) layer that integrates the extracted features. The Recurrent Neural Network (RNN) is designed for sequential data and is primarily applied in audio, video, and text processing. The Deep Belief Network (DBN) is made up of multiple Restricted Boltzmann Machines (RBMs), which are probabilistic generative models.

DBN encompasses a variety of models. The Support Vector Machine (SVM) operates based on statistical theory. The Artificial Neural Network (ANN) mimics the structure of human brains, utilizing neurons, which is why it is referred to as a biologically inspired network. The Deep Neural Network (DNN) represents a cutting-edge advancement in artificial intelligence, capable of addressing complex nonlinear relationships. DNA-binding proteins are closely associated with various human disorders, such as AIDS, cancer, and asthma (Ali et al. 2022b). The DBP-DeepCNN model could play a significant role in creating more

effective therapeutic strategies for managing chronic diseases (Ali et al. 2022a), particularly as individuals suffering from chronic depressive disorders often face challenges in their social interactions (Gheisari 2016).

The integration of cloud computing technology with wireless body area networks (WBANs) to establish a sensor-cloud infrastructure (S-CI) is significantly benefiting the healthcare sector by facilitating the early detection of diseases and enabling real-time patient monitoring (Masood et al. 2023, 2018a). It is essential to ensure that patient privacy is maintained throughout this process (Masood et al. 2018b). A well-developed deep learning model has the potential to mitigate the risks of misdiagnosis and reduce time inefficiencies (Javed et al. 2023). Deep machine learning techniques can be utilized for the initial processing of medical images, the segmentation of these images to highlight the diagnostic elements of interest, and the classification of these elements to determine their benign or malignant characteristics (Jamshaid Iqbal Janjua et al. 2022). Predicting human diseases, particularly cancer, poses significant challenges, yet it is crucial for providing more effective and timely medical care. Cancer is a serious illness that can impact various organs and systems within the human body (Abbas et al. 2023b).

Table 1 Comprehensive overview and comparative results

Author	Method	Accuracy	Sensitivity	Specificity	Dataset
(Lang et al. 2019)	CLSTM	0.75–0.84	0.75	0.968	MRI
(Shah et al. 2023)	CNN	0.95	-	-	CT
(Pandit et al. 2022)	DT	0.995	-	-	CT
(Heuvelmans et al. 2021)	ANN	0.99	0.221	-	CT
(Jena et al. 2021)	RF	0.8779	-		CT

V.CONCLUSION

In certain instances, even patients with advanced lung cancer may not exhibit the typical symptoms associated with the disease. The prevalence of lung cancer is notably high in India, particularly in rural areas, where early detection is often hindered by a lack of awareness. Additionally, it is challenging for voluntary organizations to conduct screenings for the entire population. This study aims to identify specific groups of individuals who require further screening for lung cancer, with the goal of reducing both prevalence and mortality rates. The lung cancer prediction system has the potential for further enhancement and expansion, including the integration of various data mining techniques such as time series analysis, clustering, and association rules. Furthermore, continuous data could be utilized alongside categorical data. Another promising avenue is the application of text mining to analyze the extensive unstructured data present in healthcare databases. A significant challenge remains in effectively integrating data mining and text mining methodologies.

V. REFERENCES

[1]. Abbas S, Issa GF, Fatima A et al (2023) Fused Weighted Federated Deep Extreme Machine Learning Based on Intelligent Lung Cancer Disease Prediction Model for Healthcare 5.0. Int J Intell Syst 2023:1−14.https:// doi. org/ 10. 1155/ 2023/ 25991 61

[2]. Abbas T, Fatima A, Shahzad T et al (2023) Secure IoMT for Disease Prediction Empowered With Transfer Learning in Healthcare 5.0, the Concept and Case Study. IEEE Access 11:39418−39430. https:// doi.org/ 10. 1109/ access. 2023. 32661 56

[3]. Abbas Q (2017) Lung-Deep: a Computerized Tool for detection of lung nodule patterns using deep learning algorithms detection of lung nodules patterns. Int J Adv Comput Sci Appl 8. https:// doi. org/ 10. 14569/ijacsa. 2017. 081015

[4]. Affonso C, de Nedjah NL (2020) Detection and classification of pulmonary nodules using deep learning and swarm intelligence. Multimedia Tools Appl 79:15437−15465. https:// doi. org/ 10. 1007/s11042- 019-7473-z

[5]. Alharbey R, Kim JI, Daud A et al (2022) Indexing important drugs from medical literature. Scientometrics 127:2661−2681. https:// doi. org/ 10. 1007/ s11192- 022- 04340-7

[6]. Ali F, Kumar H, Patil S et al (2022) DBP-DeepCNN: Prediction of DNA-binding proteins using waveletbased denoising and deep learning. Chemom Intell Lab Syst 229:104639−104639. https:// doi. org/ 10.1016/j. chemo lab. 2022. 104639.

[7]. Harleen Kaur and Siri Krishan Wasan, Empirical Study on Applications of Data Mining Techniques in Healthcare, Journal of Computer Science 2 (2): 194-200, 2006ISSN 1549-3636.

[8] J.R. Quinlan. Induction of decision trees. Machine learning, 1(1):81–106, 1986.

[9] J.R. Quinlan. C4. 5: Programming for machine learning. Morgan Kauffmann, 1993.

[10] L. Breiman. Random forests. Machine learning, 45(1):5–32, 2001.

[11] R. D´ıaz-Uriarte and A. de André's. Gene selection and classification of microarray data using random forest. BMC bioinformatics, 7(1):3, 2006.

[12] R.S. Michal ski and K. Kaufman. Learning patterns in noisy data: TheAQ approach. Machine Learning and its Applications, Springer- Verlag, pages 22–38, 2001.

[13] R. Linder, T. Richards, and M. Wagner. Microarray data classified by artificial neural networks. METHODS IN MOLECULAR BIOLOGYCLIFTON THEN TOTOWA-, 382:345, 2007.

[14] Murat Karabhatak, M.Cevdet Ince 2008. Expert system for detection of breast cancer based on association rules and neural network. Journal: Expert systems with Applications.

[15] Han, J. and M. Kamber, 2001. Data Mining: Concepts and Techniques. San Francisco, Morgan Kauffmann Publishers.

[16] [by Two Crows Corporation Introduction to Data Mining and Knowledge Discovery .Third Edition,2005. ISBN: 1-892095-02-5, Pages 10, 11.

[17] Maria-Luiza Antonie, Osmar R. Za¨ıane, Alexandru Coman Application of Data Mining Techniques for Medical Image Classification.Page 97

[18] Heckerman, D., A Tutorial on Learning with Bayesian Networks.1995, Microsoft Research.

[19] Neapolitan, R., Learning Bayesian Networks. 2004, London: Pearson Printice Hall.

[20] Neapolitan, R., Learning Bayesian Networks. 2004, London: Pearson Prentice Hall.

[21] Krishnapuram, B., et al., A Bayesian approach to joint feature selection and classifier design. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2004.