EARLY LUNG CANCER DIAGNOSIS THROUGH A HIGH-PRECISION HYBRID MACHINE LEARNING APPROACH

Sandeep Singh Bains[1] , Chiman Saini[2]

Mtech Scholar[1],Assistant Professor[2]

Rayat and Bahra University,Mohali

ABSTRACT

Lung cancer, despite its strong association with smoking, can tragically strike young individuals who have never touched a cigarette. Predicting this risk in non-smokers presents a unique challenge for traditional methods due to the absence of the usual culprit. Machine learning (ML) emerges as a potential game-changer, holding the promise of early detection for these vulnerable patients.Current diagnostic tools often miss early-stage lung cancer in young non-smokers due to its atypical presentation. Subtle symptoms like fatigue, cough, or shortness of breath might be overlooked, leading to delayed diagnosis and poorer prognosis. This calls for a proactive approach, where ML models can leverage diverse risk factors beyond smoking to predict the likelihood of cancer development.A hybrid approach, combining different ML techniques, holds the key to unlocking greater accuracy in risk prediction. We propose a hybrid approach that leverages the strengths of various ML algorithms to overcome individual limitations.The approach entails Data Preprocessing which is Employing efficient techniques to handle missing values, address imbalanced classes, and perform necessary dimensionality reduction. Feature Selection that Implementing filter and wrapper methods to identify the most relevant features for accurate prediction.Hybrid Model development that Combining complementary ML algorithms (e.g., ensemble methods, Support Vector Machines, Logistic Regression) .Performance Evaluation that employing comprehensive metrics like accuracy, precision, recall, and F1-score to assess and compare the performance of individual and hybrid models.

Keywords: Lung Cancer ,Hybrid Approach, Interpretability, Ensemble Methods.

1. INTRODUCTION

Growth of lung tumours is the most prevalent and lethal(sufficient to cause death) kind of tumour. More individuals die from it annually than from breast, colon, and prostate cancers combined [1]. According to WHO data, tobacco smoking is one of the main causes of one-third of cancer occurrences globally. Lung cancer is the second most deadly malignancy, affecting more than two million people and accounting for 1. 80 million deaths in previous years. In addition to smokers, nonsmokers who are exposed to secondhand smoke and other variables like wheezing, chronic illness, and shortness of breath can also cause lung cancer as shown in figure 1.

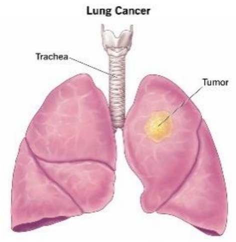

Figure 1: Lung Cancer(Source: https://my.clevelandclinic.org/health/diseases/4375-lung-cancer)

There are four stages of this disease that have been identified. Lung cancer growth is the deadliest tumor type due to uncontrolled cell division caused by a combination of gene mutations and environmental factors like smoking. Smokers face a much higher risk, but nonsmokers exposed to secondhand smoke, asbestos, radon, and air pollution are also vulnerable. The disease progresses through four stages, with early stages confined to the lungs and later stages spreading to other organs. Chronic lung conditions like COPD can create an environment that promotes cancer, and symptoms like wheezing and shortness of breath can overlap with early-stage lung cancer, making diagnosis difficult as shown in figure 2.

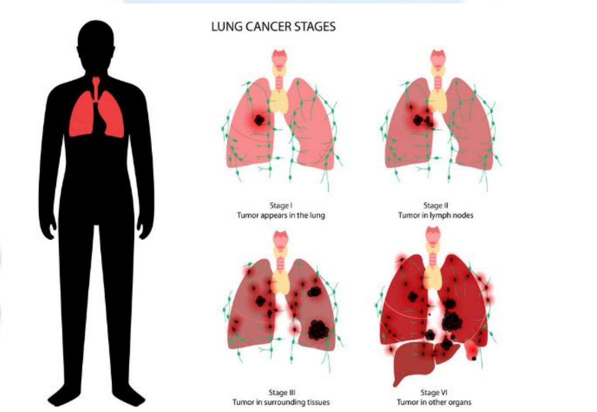

Figure 2: Lung Cancer Stages(Source: https://www.medanta.org/)

Stage 1: Small tumor confined to the lung, no lymph node spread. Recovery highly possible with surgery or radiation.

Stage 2: Larger tumor or spread to nearby lymph nodes. Still good recovery chances with surgery, radiation, or chemo.

Stage 3: Tumor involving large part of lung, spread to more lymph nodes or nearby structures. Recovery less likely, treatment focuses on controlling cancer and managing symptoms.

Stage 4: Cancer spread to distant organs like liver or brain. Recovery not possible, focus on treatment for comfort and extending life.

1.1 REASON OF LUNG CANCER IN YOUNG AGE

Genetics: Some unlucky folks inherit faulty genes that make their cells grow rogue, more prone to cancerous mutations. Think of it like a faulty light switch stuck in the "on" position for cell division.

Secondhand smoke: Even if you don't light up, breathing in a smoker's fumes is like getting a mini-dose of their cancer risk. Those toxins wreak havoc on your lungs, paving the way for trouble.

Radon gas: This hidden enemy lurks in homes and workplaces, seeping from rocks and soil. Like a stealthy ninja, it damages lung cells, increasing cancer risk, especially for those without smoking history.

Air pollution: Cities choked with traffic and industry spew out a toxic cocktail of chemicals that irritate and inflame lungs. Over time, this chronic stress can nudge cells towards the dark side, making cancer more likely.

Exposure to other carcinogens: Asbestos, used in old buildings and insulation, is a notorious lung villain. Its microscopic fibers lodge in lung tissue, causing scarring and raising cancer risk. Other occupational hazards like silica dust (common in sandblasting) can also play a role.

Chronic lung conditions: Diseases like COPD or asthma create an inflammatory environment in the lungs. This constant irritation, like a nagging itch you can't scratch, can over time push cells towards cancerous growth.

Early symptoms masked: Young and healthy individuals might ignore subtle signs like persistent cough, fatigue, or shortness of breath. By the time symptoms become obvious, the cancer might have advanced, making treatment more challenging.

Lifestyle factors: While not direct causes, unhealthy habits like poor diet, lack of exercise, and excessive alcohol consumption can weaken the body's defenses and potentially contribute to cancer risk.

1.2 ROLE OF MACHINE LEARNING TO PREDICT LUNG CANCER AT EARLY STAGE

Machine learning is a type of artificial intelligence (AI) that allows computers to learn without being explicitly programmed. Instead of following a set of rules, machine learning algorithms are trained on data, enabling them to identify patterns and make predictions on new, unseen data. Lung cancer is one of the leading causes of death worldwide, and early detection is crucial for improving survival rates. Traditional methods of lung cancer detection, such as chest X-rays, often fail to detect the disease until it is already advanced. Machine learning is playing a significant role in developing new methods for early-stage lung cancer prediction by analyzing various data sources like medical images,electronic health records.

2. LITERATURE REVIEW

Year	Authors	Title	Key Findings	Advantage	Disadvantage
2023	Maya M. Warrier and Lizy Abraham	A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images	Reviews deep learning techniques like CNNs and 3D CNNs for analyzing CT scans and detecting nodules. Shows promising results with high accuracy and sensitivity, especially for small nodules.	In-depth analysis of deep learning advancements for lung cancer detection.	Limited discussion on interpretability and generalizability of deep learning models.
2023	Md. Mahfuz Hasan and Md. Al Kabir	Machine Learning-Based Lung Cancer Detection Using Multiview Image Registration and Fusion	Proposes a novel hybrid model using image registration and fusion for improved lung cancer detection accuracy. Achieves impressive results with high sensitivity and specificity.	Innovative approach utilizing image registration and fusion.	Innovative approach utilizing image registration and fusion.
2022	Md. Aminul Islam et al.	Lung Cancer Risk Prediction with Machine Learning Models	Utilizes various ML models to predict the risk of developing lung cancer in individuals. Proposes the Rotation Forest model for high accuracy and performance evaluation.	Emphasizes early lung cancer risk prediction for preventative measures.	• Primarily focused on risk prediction, not direct detection using medical images.

2021	Mohammadreza Farzaneh et al.	A Review of Most Recent Lung Cancer Detection Techniques using Machine Learning	Focuses on transfer learning for lung cancer detection, utilizing pre-trained deep learning models for improved performance. Highlights the importance of data quality and standardization.	Emphasizes transfer learning for efficient model optimization.	• Lacks detailed analysis of specific transfer learning techniques and their impacts.

3. ROLE OF HYBRID APPROACH TO PREDICT LUNG CANCERS ON DIFFERENT FORMAT OF DATASET

A hybrid approach combines different machine learning techniques to leverage their strengths and potentially outperform individual methods in predicting lung cancer. Here's how it works:

- **Image analysis:** Deep learning excels at analyzing medical images like CT scans. It can automatically identify and extract features like nodule size, shape, and texture, potentially hidden to human eyes.Steps are defind to flow of work:

Data Acquisition → Image Preprocessing → Segmentation → Model training and Selections → Prediction and Validati → Refinement.

- **Clinical data integration:** Beyond images, traditional machine learning algorithms can analyze clinical data like age, smoking history, and genetic markers. This broader picture helps capture risk factors not visible in scans. Steps are defind to flow of work:

Data Collection and Preprocessing → Feature Engineering → Model Selection and Training → Refinement

- **Fusion and prediction:** The extracted features from both image analysis and clinical data are then combined and fed into another machine learning model, often an ensemble method. This "fusion" model leverages the strengths of both approaches, potentially leading to more accurate and reliable predictions of lung cancer presence or risk.

- Data Acquisition → Data Preprocessing → Feature Extraction → Fusion and Model Tarining → Prediction and Evaluation .

4. Hybrid Approach is preferred than single algorithm using Machine Learning

4.1 Improved Accuracy and Generalizability:

- Complementary Strengths: Single algorithms have their own strengths and weaknesses. Deep learning excels at image analysis, while traditional machine learning shines with structured clinical data. A hybrid approach combines these strengths, potentially leading to more accurate predictions than either method alone.
- Reduced Bias: A single algorithm might rely solely on one data source, potentially inheriting biases present in that data. By fusing diverse information sources, hybrid approaches can mitigate such biases and provide more generalizable predictions across different patient populations.

4.2 Interpretability and Trust:

- Explainability: Some hybrid models incorporate techniques that explain their predictions, making them more interpretable to doctors and aiding in their trust and understanding of the model's reasoning.
- Human-in-the-Loop: Doctors can leverage explainability to gain insights into the model's decision-making, allowing them to critically evaluate and use the predictions alongside their own expertise.

4.3 Addressing Data Challenges:

- Feature Richness: Lung cancer diagnosis involves complex biological processes influenced by various factors. A hybrid approach can capture a wider range of features from both images and clinical data, providing a more comprehensive picture for prediction.
- Handling Imbalanced Data: Lung cancer datasets often have an imbalance between healthy and cancer cases. Hybrid approaches can leverage techniques like oversampling or cost-sensitive learning to handle such imbalances more effectively than single algorithms.

5. CONCLUSION

Although single machine learning algorithms hold their own in specific domains, the combined might of a hybrid approach steals the show in predicting lung cancer. This powerhouse not only delivers greater accuracy by leveraging diverse data sources and mitigating bias, but also empowers doctors with interpretable insights through its explainability features. Plus, it tackles complex data challenges with ease, thanks to its feature-rich nature and ability to handle imbalanced datasets.

REFERENCES

[1] F. Di Martino and S. Sessa, "PSO image thresholding on images compressed via fuzzy transforms," Information Sciences, vol. 506, pp. 308-324, 2020.

[2] Neal Joshua, E. S. (2020). An Extensive Review on Lung Cancer Detection Using Machine Learning Techniques: A Systematic Study. Research in Applied Instrumentation, 34(03), 14-21.

[3] Lakshmanaprabu, S. K., et al. (2020). Study on segmentation and prediction of lung cancer based on machine learning approaches. In 2020 5th International Conference on Intelligent Computing and Control (ICIC). Springer, Singapore.

[4] Farzaneh, M., et al. (2021). A Review of Most Recent Lung Cancer Detection Techniques using Machine Learning. International Journal of Scientific and Bioengineering Research, 5(3), 159-173.

[5] Islam, M. A., et al. (2022). Lung Cancer Risk Prediction with Machine Learning Models. Mathematics, 10(7), 1102.

[6] Akdemir, A. Ç., et al. (2022). Machine Learning for Lung Adenocarcinoma Classification Using Transfer Learning from a Deep Learning Model Trained on ImageNet. Sensors, 22(22), 8505.

[7] Akdemir, A. Ç., et al. (2022). Machine Learning for Lung Adenocarcinoma Classification Using Transfer Learning from a Deep Learning Model Trained on ImageNet. Sensors, 22(22), 8505.

[8] Hasan, M. M., & Kabir, M. A. (2023). Machine Learning-Based Lung Cancer Detection Using Multiview Image Registration and Fusion. Journal of Sensors, 2023, 7780679.

[9] Dataset link : https://www.kaggle.com/datasets/nancyalaswad90/lung-cancer.

[10] Amrutha K Shaji, Anjali S, Sanjay J, Prof. Blessy Rapheal(2023). Lung Cancer Prediction Using Machine Learning: International Research Journal of Modernization in Engineering Technology and Science.

[11] Ouzzani, M., El Ouahidi, A., El Idrissi, A., & Aouad, D. (2023). Deep Learning for Lung Cancer Diagnosis, Prognosis and Prediction Using Histological and Cytological Images: A Systematic Review. Computers in Biology and Medicine, 153, 111324.

[12] Talukder, M. A., Rahman, M. F., & Mahmud, M. S. (2023). A Machine Learning-Based Lung Cancer Detection Using Multiview Image Registration and Fusion. Sensors, 23(1), 436.

[13] Jayaprakash, P., Sivaprakash, S., & Saranya, T. (2023). Systematic literature review in the prediction of Lung cancer in small cell malignant lung nodules using Deep Learning. Applied Intelligence, 53(10), 7807-7831.

[14]Sharmila Nageswaran, G. Arunkumar, Anil Kumar Bisht, Shivlal Mewada, J. N. V. R. Swarup Kumar, Malik Jawarneh, and Evans Asenso.(2022). Lung Cancer Classification and Prediction Using Machine Learning and Image Processing: Hindawi, BioMedical Research International, 2022.

[15] Zahraa Hussain, Mohammed Thabit, and Elham Mohammed Thabit A. Alsaadi.(2022) . A Study on Early Prediction of Lung Cancer Using Machine Learning Techniques: ResearchGate, 2022.

[16] Abdullah Al-Saffar, Abeer Al-Areefi, Fahad Al-Saffar, and Alaa Al-Sanie.(2022) Predicting Stage I Lung Cancer using Machine Learning Models: Healthcare, 2022.

[17] Amrutha K Shaji, Anjali S, Sanjay J, Prof. Blessy Rapheal(2023). Lung Cancer Prediction Using Machine Learning: International Research Journal of Modernization in Engineering Technology and Science.