International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

Review of AI Techniques based Lung Cancer Detection

[1]Rahul Kumar, [2]Mr. Dilip Kumar

[1]Research Scholar, Department of CSE, Vishveshwarya Group of Institutions Gautam Buddh Nagar, India
[2]Assistant Professor, Department of CSE, Vishveshwarya Group of Institutions Gautam Buddh Nagar, India

***Abstract*— This review paper explores the recent advancements in artificial intelligence (AI) techniques applied to lung cancer detection, emphasizing their impact on improving diagnostic accuracy and early disease identification. Various AI methods, including machine learning, deep learning, and hybrid models, are analyzed for their ability to process and interpret medical imaging data such as CT scans and X-rays. The review highlights key algorithms like convolutional neural networks (CNNs), support vector machines (SVMs), and ensemble methods that have demonstrated promising results in detecting lung nodules and classifying malignant tumors. Challenges such as data scarcity, variability in imaging quality, and the need for large annotated datasets are discussed. Furthermore, the review outlines future research directions aimed at enhancing model robustness, interpretability, and integration into clinical workflows to support timely and accurate lung cancer diagnosis.**

***Keywords*—*CT, CNN, SVMs, X-rays*.**

I. Introduction

Lung cancer remains one of the leading causes of cancer-related deaths worldwide, accounting for a significant portion of the global cancer burden. Early detection of lung cancer is critical, as it greatly improves the chances of successful treatment and patient survival. However, lung cancer is often diagnosed at advanced stages due to the subtle and non-specific nature of early symptoms, making timely identification a major challenge in clinical practice[1]. Traditional diagnostic methods rely heavily on imaging techniques such as chest X-rays, computed tomography (CT) scans, and biopsy procedures. While these methods provide essential information, they are often time-consuming, expensive, and dependent on the expertise of radiologists and pathologists, which may introduce subjectivity and variability in diagnosis[2].

In recent years, advancements in medical imaging technology have produced vast amounts of data, presenting both opportunities and challenges for healthcare professionals. The sheer volume and complexity of imaging data necessitate automated and efficient diagnostic tools that can assist clinicians in accurately detecting lung cancer at its earliest stages. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) techniques, has emerged as a powerful solution for analyzing medical images and extracting meaningful patterns that might be imperceptible to the human eye[3]. These AI-driven approaches enable the development of computer-aided detection (CAD) systems that can significantly enhance the sensitivity and specificity of lung cancer screening [4].

Machine learning algorithms such as support vector machines (SVM), random forests, and decision trees have been widely applied to classify lung nodules based on features extracted from imaging data[5]. More recently, deep learning models, especially convolutional neural networks (CNNs), have shown remarkable success in automatically learning hierarchical features from raw images, thereby eliminating the need for manual feature engineering. CNN-based systems can detect, segment, and classify lung nodules with high accuracy, enabling early intervention and personalized treatment planning. Moreover, hybrid models that combine multiple AI techniques are being explored to further improve diagnostic performance by leveraging complementary strengths[6].

Despite the promising advancements, several challenges persist in lung cancer detection using AI. These include the scarcity of large, annotated medical datasets, variability in imaging protocols across different institutions, and the black-box nature of many AI models that limits interpretability and clinical trust. Addressing these challenges requires multidisciplinary collaboration among clinicians, data scientists, and engineers to develop robust, generalizable, and explainable AI solutions[7].

This introduction sets the stage for an in-depth exploration of lung cancer detection methodologies, emphasizing the pivotal role of AI in revolutionizing early diagnosis and improving patient outcomes. The integration of AI techniques with traditional diagnostic workflows holds immense potential to transform lung cancer care, ultimately reducing mortality rates and enhancing quality of life for patients worldwide[8].

Building upon the foundational advancements in AI-driven lung cancer detection, ongoing research is increasingly focusing on integrating multimodal data sources to enhance diagnostic accuracy. Besides imaging data, clinical parameters such as patient history, genetic information, and biomarkers are being incorporated into predictive models [9]. This holistic approach allows for more comprehensive risk assessment and personalized screening strategies. For instance, combining CT scan analysis with molecular markers can improve differentiation between benign and malignant nodules, reducing unnecessary invasive procedures. Additionally, advancements in natural language processing (NLP) enable the extraction of valuable insights from unstructured clinical notes and radiology reports, further enriching the data landscape for AI models [10].

II. LITERATURE SURVEY

B. S., P. R., et al., [1] Lung disease is one of the most prevalent diseases that is impacted in the early stages, thereby enhancing the survival rate of patients.The most difficult aspect of the job for a radiologist is the cancer diagnosis.Radiologists greatly benefit from an intelligent computer-aided system. A variety of studies have been conducted to detect lung cancer using machine learning techniques. The multi-stage classification is the primary method employed to predict lung cancer.The classification system employed for data enhancement and segmentation has been implemented. The segmentation method employs a binary classifier for classification and a threshold and marker-controlled watershed for segmentation. The accuracy of lung cancer detection is significantly greater.

S. S. Araújo et al.,[2] The features extraction on thorax computerized tomography images is performed using the radiological densities of human tissues in Hounsfield Units. In order to evaluate the efficacy of the proposed method in conjunction with four machine learning classifiers, we conducted a comparison with the Gray Level Co-occurrence Matrix and Statistical Moments. In general, the findings indicated that the proposal obtained the highest accuracy ratios and the shortest duration among all experiments conducted. Consequently, we regard our proposal as a viable alternative that can be implemented in real-time applications.

The objective of T. K. K. Ho et al.,[3] is to either convert knowledge from ponderous teacher models into lightweight student models or to self-train these student models in order to produce weakly supervised multi-label lung disease classifications. The visualizations employed in saliency maps of the pathological regions where an abnormality was located were supported by multi-task deep learning architectures, in addition to multi-class classification, which was the foundation of our approach. A self-training KD framework, in which the model learnt from itself, was demonstrated to outperform both the well-established baseline training procedure and the conventional KD, attaining AUC improvements of up to 6.39% and 3.89%, respectively. In comparison to the current deep learning baselines, our approach effectively surmounted the interdependency of 14 inadequately annotated thorax maladies and facilitated the state-of-the-art classification through application to the publicly available ChestX-ray14 dataset.

AFS-DF on the lungs diseases -19 dataset is presented by L. Sun et al.[4]. The dataset includes 1495 patients with lungs diseases -19 and 1027 patients with community-acquired pneumonia (CAP). Our method has demonstrated accuracy (ACC), sensitivity (SEN), specificity (SPE), AUC, precision, and F1-score of 91.79%, 93.05%, 89.95%, 96.35%, 93.10%, and 93.07%, respectively. The experimental results on the lungs diseases dataset indicate that the proposed AFS-DF outperforms four widely used machine learning methods in the classification of lungs diseases vs. CAP.

C. Baloescu et al.,[5] In order to develop and evaluate the deep learning algorithm based on deep convolutional neural networks, we employed ultrasound recordings (n = 400) from an existing database of ED patients to serve as training and test sets. Expert human interpretations of binary and severity classifications (a scale of 0-4) were contrasted with the algorithmic interpretations of the images. In comparison to an expert read, our model demonstrated a sensitivity of 93% (95% confidence interval (CI) 81%-98%) and a specificity of 96% (95% CI 84%-99%) for the presence or absence of B-lines. The kappa ratio was 0.88 (95% CI 0.79-0.97). A weighted kappa of 0.65 (95% CI 0.56- 074) was obtained for the severity classification model to expert agreement. In general, the DL algorithm demonstrated satisfactory performance and could be incorporated into an ultrasound system to assist in the diagnosis and monitoring of B-line severity. The algorithm is more effective at distinguishing the presence from the absence of B-lines, but it can also be effectively employed to differentiate between the severity of B-lines. These methods have the potential to reduce variability and establish a standardized approach to enhance diagnosis and outcome.

J.-X. Wu et al.,[6] Subjects with typical lung diseases are screened using a multilayer machine vision classifier that incorporates a radial Bayesian network and gray relational analysis. The NIH chest X-ray database (NIH Clinical Center) is utilized to enroll anterior-posterior chest X-ray images. The proposed multilayer machine vision classifier is employed to assist in the diagnosis of common lung diseases on specific bounding ROIs using digital chest X-ray images. The performance of the proposed multilayer classifier for the rapid screening of lung lesions on digital chest X-ray images is evaluated using mean recall (%), mean precision (%), mean accuracy (%), and mean F1 score (0.8981), respectively, with K-fold cross-validation.

S. Roy et al.[7] propose a novel approach to the video-level aggregation of frame scores that is based on uninorms. Lastly, we evaluate the performance of cutting-edge deep models in the estimation of pixel-level segmentations of lung disease imaging biomarkers. The proposed dataset has yielded satisfactory results in all of the tasks that were examined, which will facilitate future research on deep learning for the assisted diagnosis of pulmonary diseases from LUS data.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

S. Pang et al.[8] suggest a deep learning model for the identification of lung cancer type from CT images in patients at Shandong Provincial Hospital. It faces a dual challenge: the limited number of patient data acquired and the inadequacy of artificial intelligent models trained on public datasets in meeting these practical requirements. The two-fold problem is resolved by employing image rotation, translation, and transformation methods to expand and balance our training data. Subsequently, densely connected convolutional networks (DenseNet) are employed to classify malignant tumors from images collected. Finally, the adaptive boosting (adaboost) algorithm is employed to aggregate multiple classification results in order to enhance classification performance. The experimental results indicate that our method is capable of achieving an identifying accuracy of 89.85%, surpassing DenseNet without adaboost, ResNet, VGG16, and AlexNet. This offers a non-invasive, efficient detection instrument for the pathological diagnosis of lung cancer.

H. Yazdani et al.[9] introduce a method that assesses the migration of samples from one cluster to another. This method enables us to identify critical samples in advance that have the potential to be a part of other clusters in the near future. In a lung cancer case-control investigation, BFPM was implemented to analyze the metabolomics of individuals. Metabolomics may function as robust biomarkers of the current disease process by providing proximate molecular signals to the actual disease processes. The objective is to determine whether it is possible to distinguish between the serum metabolites of a healthy individual and those of a person with lung cancer. BFPM was employed to identify critical samples, evaluate pathology data, and observe certain discrepancies.

F. Yan et al.,[10] The chest X-ray is a straightforward and cost-effective medical tool for auxiliary diagnosis, and as a result, it has become a standard component of physical examinations for physicians. By utilizing deep learning techniques, we investigate the abnormality classification problem of chest X-rays using 40167 images of chest radiographs and corresponding reports. Initially, we suggest an annotation method that is based on the anomalous portion of the images, as radiology reports are typically templated by the aberrant physical regions. Secondly, we utilize the long short-term memory (LSTM) model to automatically annotate the remaining unlabeled data, building on a limited number of reports that are manually annotated by professional radiologists. The precision value, recall value, and F1-score all exceed 0.88 in the accurate annotation of images. Ultimately, we train convolutional neural networks to classify the abnormality in the chest X-rays, and the results indicate that the average AUC value is 0.835.

A. Rao et al.,[11] The accumulation of excessive air and water in the lungs results in the impairment of respiratory function and is a prevalent cause of patient hospitalization. Physicians can evaluate patients' respiratory conditions by employing non-invasive and compact methods to detect changes in lung fluid accumulation. In this study, a digital stethoscope instrument and an acoustic transducer are suggested as a targeted solution to address this clinical requirement. Measurable changes in the structure of the lungs can be employed to evaluate lung pathology. We standardize this procedure by transmitting a controlled signal through the airways of six healthy subjects and six patients with lung disease. Mel-frequency cepstral coefficients and spectroid audio features, which are frequently employed in classification for music retrieval, are extracted to differentiate between healthy and diseased subjects. We exhibit a 91.7% accuracy in the differentiation between healthy subjects and patients with lung pathology by employing the K-nearest neighbors algorithm.

O. P. Singh et al.,[12] At present, the clinic employs capnography to measure carbon dioxide (CO_2) waveforms in order to estimate respiratory rate and end-tidal CO2 ($EtCO_2$). Nevertheless, the asthmatic condition is significantly influenced by the morphology of the CO_2 signal. Previous research has demonstrated a robust correlation between a variety of features that quantitatively characterize the shape of the CO_2 signal and are employed to differentiate asthma from non-asthma using pulmonary function tests. However, no reliable progress has been made, and no translation into clinical practice has been achieved. Consequently, this study presents a signal processing algorithm that is relatively straightforward and can be used to automatically differentiate between asthma and non-asthma. The CO_2 signals were recorded from 30 non-asthmatic and 43 asthmatic patients. Features were computationally derived after each respiration cycle was decomposed into subcycles. Afterward, the area under the receiver operating characteristic curve analysis was employed to select the features.

III. CHALLENGES

Despite the remarkable progress, several challenges hinder the widespread adoption of machine learning in clinical settings. Key challenges include:

1. **Data Heterogeneity**: Variations in imaging protocols, patient demographics, and disease subtypes result in highly heterogeneous datasets, making it difficult to develop generalized models.
2. **Limited Annotated Datasets**: The scarcity of high-quality, annotated datasets for training and validation poses a significant obstacle to developing robust machine learning algorithms.

3. **Model Interpretability**: The "black-box" nature of many machine learning models, particularly deep learning approaches, raises concerns about their interpretability and trustworthiness in clinical decision-making.
4. **Computational Costs**: Training complex machine learning models requires significant computational resources, which may not be feasible for all healthcare institutions.
5. **Ethical Concerns**: Issues such as data privacy, algorithmic bias, and the potential misuse of patient data need to be addressed to ensure ethical deployment of ML technologies.
6. **Regulatory Challenges**: The lack of standardized evaluation frameworks and regulatory guidelines makes it challenging to validate and approve ML-based diagnostic tools for clinical use.
7. **Integration with Clinical Workflows**: Adapting machine learning tools to seamlessly integrate into existing clinical workflows without disrupting routine practices remains a critical challenge.
8. **Dynamic Nature of Data**: Lung cancer data, including imaging and molecular profiles, evolve over time, necessitating continuous updates and retraining of machine learning models to maintain their accuracy.

The integration of machine learning into lung cancer detection holds immense promise for revolutionizing diagnostic practices. By enabling early and accurate detection, these technologies can potentially reduce the global burden of lung cancer and improve patient survival rates. This review not only highlights the current advancements but also identifies areas requiring further research and development. By fostering collaboration between data scientists, clinicians, and policymakers, the potential of machine learning in lung cancer detection can be fully realized, bringing us closer to achieving precision medicine and personalized healthcare.

IV. CONCLUSION

Machine learning has demonstrated immense potential in transforming lung cancer detection by enabling early, accurate, and efficient diagnosis. By leveraging advanced algorithms and integrating diverse data sources such as medical imaging, histopathology, and biomarkers, machine learning tools can significantly enhance diagnostic precision and aid in personalized treatment planning. However, addressing challenges such as data heterogeneity, model interpretability, ethical concerns, and integration into clinical workflows is essential for realizing the full potential of these technologies. Future research should focus on developing standardized frameworks, improving the availability of annotated datasets, and fostering interdisciplinary collaboration between technologists, clinicians, and policymakers. With continued advancements, machine learning can play a pivotal role in reducing the global burden of lung cancer and improving patient outcomes.

REFERENCES

1. B. S, P. R and A. B, "Lung Cancer Detection using Machine Learning," *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2022, pp. 539-543, doi: 10.1109/ICAAIC53929.2022.9793061.
2. S. S. Araújo Alves, E. de Souza Rebouças, S. A. Freitas de Oliveira, A. Magalhães Braga and P. P. Rebouças Filho, "Lung Diseases Classification by Analysis of Lung Tissue Densities," in IEEE Latin America Transactions, vol. 18, no. 09, pp. 1329-1336, September 2020, doi: 10.1109/TLA.2020.9381790.
3. T. K. K. Ho and J. Gwak, "Utilizing Knowledge Distillation in Deep Learning for Classification of Chest X-Ray Abnormalities," in IEEE Access, vol. 8, pp. 160749-160761, 2020, doi: 10.1109/ACCESS.2020.3020802.
4. L. Sun et al., "Adaptive Feature Selection Guided Deep Forest for lungs diseases Classification With Chest CT," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 10, pp. 2798-2805, Oct. 2020, doi: 10.1109/JBHI.2020.3019505.
5. C. Baloescu et al., "Automated Lung Ultrasound B-Line Assessment Using a Deep Learning Algorithm," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 67, no. 11, pp. 2312-2320, Nov. 2020, doi: 10.1109/TUFFC.2020.3002249.
6. J. -X. Wu, P. -Y. Chen, C. -M. Li, Y. -C. Kuo, N. -S. Pai and C. -H. Lin, "Multilayer Fractional-Order Machine Vision Classifier for Rapid Typical Lung Diseases Screening on Digital Chest X-Ray Images," in IEEE Access, vol. 8, pp. 105886-105902, 2020, doi: 10.1109/ACCESS.2020.3000186.

7. S. Roy et al., "Deep Learning for Classification and Localization of lungs diseases -19 Markers in Point-of-Care Lung Ultrasound," in IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2676-2687, Aug. 2020, doi: 10.1109/TMI.2020.2994459.
8. S. Pang, Y. Zhang, M. Ding, X. Wang and X. Xie, "A Deep Model for Lung Cancer Type Identification by Densely Connected Convolutional Networks and Adaptive Boosting," in IEEE Access, vol. 8, pp. 4799-4805, 2020, doi: 10.1109/ACCESS.2019.2962862.
9. H. Yazdani, L. L. Cheng, D. C. Christiani and A. Yazdani, "Bounded Fuzzy Possibilistic Method Reveals Information about Lung Cancer through Analysis of Metabolomics," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 17, no. 2, pp. 526-535, 1 March-April 2020, doi: 10.1109/TCBB.2018.2869757.
10. F. Yan, X. Huang, Y. Yao, M. Lu and M. Li, "Combining LSTM and DenseNet for Automatic Annotation and Classification of Chest X-Ray Images," in IEEE Access, vol. 7, pp. 74181-74189, 2019, doi: 10.1109/ACCESS.2019.2920397.
11. A. Rao, S. Chu, N. Batlivala, S. Zetumer and S. Roy, "Improved Detection of Lung Fluid With Standardized Acoustic Stimulation of the Chest," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1-7, 2018, Art no. 3200107, doi: 10.1109/JTEHM.2018.2863366.
12. O. P. Singh, R. Palaniappan and M. Malarvili, "Automatic Quantitative Analysis of Human Respired Carbon Dioxide Waveform for Asthma and Non-Asthma Classification Using Support Vector Machine," in IEEE Access, vol. 6, pp. 55245-55256, 2018, doi: 10.1109/ACCESS.2018.2871091.