International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 15, Issue 2, February 2026)

Review of Lung Cancer Classification and Prediction With Deep Learning

Asheesh Mishra[1], Dr. Devendra Singh Rathore[2], Dr. Vivek Richhariya[3]
[1]Research Scholar, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India
[2]Associate Professor, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India
[3]Professor, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India

***Abstract*— Lung cancer remains one of the leading causes of cancer-related deaths globally, primarily due to late-stage diagnosis and limited early detection methods. Recent advancements in deep learning have shown immense potential in improving the classification and prediction of lung cancer through automated and accurate analysis of medical imaging data, such as CT scans, X-rays, and histopathological images. This review explores the latest developments in deep learning techniques applied to lung cancer detection, discussing various architectures like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models. It highlights the advantages, challenges, and future directions in adopting deep learning for clinical practice, emphasizing the critical role of robust datasets, model interpretability, and real-world validation. The review aims to provide a comprehensive understanding of how deep learning is reshaping lung cancer diagnostics and predictive modeling, ultimately contributing to earlier detection, personalized treatment, and improved patient outcomes.**

Keywords—Lung, Cancer, X-rays, CNN, RNN, AI, Deep learning, ML.

I. INTRODUCTION

Lung cancer continues to be a major global health challenge, accounting for a significant number of cancer deaths each year. Despite advancements in medical science, the survival rate of lung cancer patients remains low, largely due to the difficulties in early detection and the aggressive nature of the disease [1]. Traditional diagnostic methods, such as chest X-rays, CT scans, biopsies, and sputum cytology, often suffer from limitations like human error, variability in interpretation, and sensitivity to tumor size and location. These challenges have created a strong demand for automated, accurate, and early diagnostic tools that can assist clinicians in improving lung cancer prognosis [2].

In this context, deep learning—a subset of artificial intelligence that mimics the human brain's neural networks—has emerged as a transformative technology. Deep learning models, particularly convolutional neural networks (CNNs), have demonstrated outstanding capabilities in image classification, segmentation, and feature extraction, making them highly suitable for medical imaging analysis [3]. Their ability to learn hierarchical feature representations directly from raw data reduces the dependency on manual feature engineering and enhances the detection of subtle patterns that may be overlooked by human observers [4].

Several studies have successfully applied deep learning algorithms to various aspects of lung cancer detection and classification, including nodule detection, malignancy prediction, histopathological image analysis, and even genetic marker identification [5]. CNNs have been extensively used to classify lung nodules as benign or malignant, while recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) have been explored for time-series analysis of sequential imaging data. Furthermore, hybrid models that combine deep learning with traditional machine learning techniques, such as support vector machines (SVMs) and random forests, have also been proposed to boost performance [6].

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 15, Issue 2, February 2026)

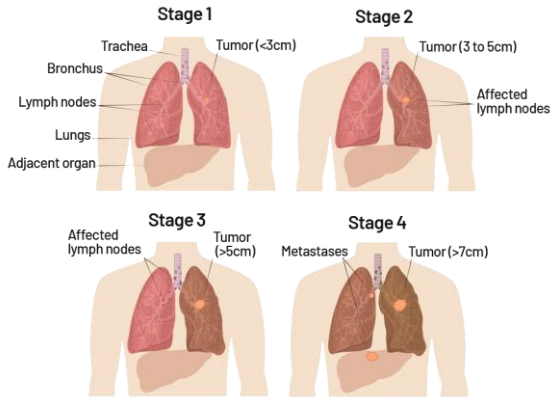

Figure 1: Lung cancer

However, despite promising results, there are several challenges that need to be addressed before deep learning can be fully integrated into routine clinical practice. These include the need for large and diverse annotated datasets, ensuring model interpretability for clinical acceptance, managing data privacy concerns, and mitigating biases that could affect diagnostic equity [7]. Additionally, real-world validation through clinical trials and cross-institutional collaborations is crucial for assessing the robustness and generalizability of deep learning models across different patient populations and imaging modalities [8].

The objective of this review is to systematically analyze and summarize the recent research contributions in lung cancer classification and prediction using deep learning techniques. It focuses on the different deep learning architectures employed, the types of datasets utilized, preprocessing techniques, performance metrics, and real-world applicability. Furthermore, the review discusses the critical challenges faced by researchers and proposes potential future directions for developing more reliable and interpretable AI-driven solutions in lung cancer care [9].

By bridging the gap between cutting-edge technology and clinical needs, deep learning has the potential to revolutionize the landscape of lung cancer diagnostics and prognosis. This review aims to provide researchers, clinicians, and healthcare stakeholders with an updated, comprehensive understanding of how deep learning can contribute to the early detection, accurate classification, and personalized management of lung cancer, ultimately leading to better survival rates and quality of life for patients [10].

II. LITERATURE SURVEY

P. S et al., [1] Due to unchecked cell development in the lungs, lung cancer typically affects both men and women. This seriously impairs one's ability to breathe in and out of the chest. According to the World Health Organisation, the main causes of lung cancer are cigarettes and passive smoking. Compared to other cancers, the death rate from lung cancer is rising daily among both young and old people. Despite the availability of advanced medical facilities for accurate diagnosis and effective medical care, the mortality rate is still not effectively under control.

T. I. A. Mohamed et al., [2] presented an innovative deep-learning model for lung cancer detection by integrating markers from mRNA, miRNA, and DNA methylation. Subsequently, integration of all prepared omics data types was achieved by selecting common samples, resulting in a consolidated dataset comprising 448 samples and 8228 features (genes). To streamline features, principal components analysis (PCA) was implemented, and the synthetic minority over-sampling technique (SMOTE) algorithm was applied to ensure class balance.

Al-Tamimi et al. [3] employed 3D convolutional neural networks (3D-CNNs) for volumetric CT data analysis. Unlike 2D methods, the 3D-CNN captured spatial context more effectively, leading to improved sensitivity in detecting small malignant nodules. The study also introduced augmentation techniques specific to volumetric data, which further enhanced model generalization on external test sets.

Kim et al. [4] presented a dual-path network architecture combining CNN-based feature extraction with graph convolutional networks (GCNs) to model spatial relationships between detected lung nodules. This combination allowed the system to consider not only the individual characteristics of nodules but also their spatial distributions, resulting in superior

malignancy classification compared to conventional single-path models.

Liu developed [5] a multi-task deep learning model capable of performing both lung nodule segmentation and classification simultaneously. By sharing the feature extraction layers between tasks, their model achieved better performance with fewer parameters. This joint learning approach demonstrated that multitasking could boost overall system robustness in clinical imaging workflows.

Ypsilantis and Montana [6] explored the application of deep reinforcement learning (DRL) for optimizing lung cancer screening strategies. Rather than only focusing on image classification, their DRL-based system learned to recommend personalized screening intervals based on patient risk profiles and previous imaging results, introducing an intelligent decision-making layer into lung cancer management.

Shen et al., [7] performed one of the early comprehensive studies applying deep learning to lung nodule detection and classification. They proposed a multi-crop CNN method that analyzed nodules at different scales to capture various levels of detail. Their approach significantly reduced the false positive rate, a major issue in automated lung cancer diagnosis.

Setio et al., [8] introduced the LUNA16 challenge dataset and benchmarked several deep learning models for lung nodule detection. Their work laid a foundation for standardized evaluation in lung cancer imaging research. They demonstrated that ensemble approaches combining multiple CNN models could significantly outperform individual models, pushing detection sensitivity beyond 90%.

Hua et al., [9] investigated the application of deep belief networks (DBNs) for lung cancer classification. Although CNNs later became dominant, their early work showed that DBNs could effectively extract hierarchical features from CT scans and deliver promising classification results, emphasizing the early potential of deep learning in medical imaging.

Hussein et al., [10] pioneered one of the first uses of 3D patches from CT images in conjunction with CNNs to classify lung nodules. Their patch-based approach addressed the imbalance between malignant and benign samples and laid important groundwork for future 3D deep learning applications in lung cancer detection.

U. G. et al., [11] emphasized the urgency of addressing non-communicable diseases, including cancer, through international collaboration, policy implementation, and research investment. This global initiative has accelerated the development of computational tools, including artificial intelligence (AI) and deep learning, for cancer diagnosis and management. As a result, research attention has shifted toward algorithm-driven cancer classification systems capable of early detection.

Fitipaldi and Franks et al., [12] examined genome-wide association studies (GWAS) between 2005 and 2022, uncovering critical biases in the data related to ethnicity, gender, and other socio-demographic factors. These biases compromise the fairness and generalizability of machine learning and deep learning models used in cancer genomics. Addressing such disparities is essential for building robust DL classifiers that can perform equitably across patient populations.

Yuan et al., [13] explored gene expression profiles of lung cancer subtypes using machine learning algorithms, demonstrating that classification accuracy could be significantly improved through supervised learning techniques. This study laid foundational work for integrating such models with deep learning architectures like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) for more scalable and automated subtype prediction.

Huang et al., [14] revealed that alternative polyadenylation, a post-transcriptional gene regulation mechanism, drives oncogenic gene expression in non-small cell lung cancer (NSCLC). Deep learning models that process transcriptomic data have the

potential to detect these nuanced gene expression shifts, enabling early intervention strategies and patient stratification.

J. N. Bodor et al., [15] focused on biomarkers predictive of immune checkpoint inhibition effectiveness in NSCLC. These biomarkers, such as PD-L1 expression and tumor mutational burden, have been used as input features in DL models to predict patient response to immunotherapy, advancing personalized medicine in oncology.

Ginn et al., [16] delved into the functional relevance of long non-coding RNAs (lncRNAs) in NSCLC. The complex expression patterns of lncRNAs are ideally suited for deep learning methods, particularly autoencoders and hybrid architectures, which can uncover hidden representations in high-dimensional data and enhance classification outcomes.

Cai et al., [17] developed the Lung Cancer Explorer (LCE), a comprehensive portal integrating gene expression and clinical data. Such platforms enable seamless data access and foster the development of explainable DL models that integrate both genomic and phenotypic information, driving innovations in both classification and prediction tasks.

Osama et al., [18] provided a detailed review of gene selection and reduction techniques applied in conjunction with machine learning models. Their findings highlight the importance of dimensionality reduction before feeding input into deep neural networks (DNNs), as irrelevant features can degrade performance in cancer prediction models.

Alharbi and Vakanski [19] conducted a survey on machine learning methods applied to cancer classification using gene expression datasets. Their analysis revealed that DL architectures, especially deep belief networks and ensemble deep learners, outperform traditional methods when trained on sufficient and well-curated genomic data.

Xue et al., [20] employed single-cell RNA sequencing to uncover immunotherapeutic targets in lung cancer. The granularity of single-cell data makes it particularly compatible with attention-based DL models, which can detect subtle variations in cell populations and predict treatment outcomes with higher accuracy.

III. CHALLENGES

Limited Annotated Datasets: One of the fundamental challenges in lung cancer classification is the scarcity of labeled medical imaging data. Deep learning models require vast amounts of annotated data to effectively train and generalize. However, annotating medical images, particularly CT scans, is a labor-intensive process that demands highly skilled radiologists, making it costly and time-consuming. Additionally, the quality of annotations can vary between experts, leading to inconsistencies in the dataset. This limited availability of annotated datasets significantly hinders the development of robust deep learning models, especially when dealing with rare conditions like early-stage lung cancer.

Class Imbalance: Class imbalance is another significant challenge faced in lung cancer prediction. In typical datasets, benign nodules outnumber malignant ones by a large margin, which skews the performance of machine learning models. When trained on such imbalanced data, models tend to classify most nodules as benign, resulting in lower sensitivity and accuracy in detecting malignant nodules, particularly in early stages of cancer. This imbalance can lead to critical misclassifications, as detecting lung cancer early is essential for improving survival rates.

Variability in Imaging Protocols: Inconsistent imaging protocols across different healthcare facilities present a major challenge for developing universal lung cancer classification models. Variations in CT scanner models, imaging resolutions, slice thicknesses,

and contrast agents used during imaging all contribute to differences in the quality and characteristics of the images. These variations can lead to domain shifts, where models trained on data from one scanner type or hospital might not perform well on data from a different source. This issue of poor generalization across diverse clinical environments impedes the widespread adoption of deep learning models in real-world healthcare settings.

Lack of Interpretability: Deep learning models, particularly Convolutional Neural Networks (CNNs) and newer architectures like Vision Transformers, are often considered black boxes due to their lack of interpretability. In medical applications, it is crucial for clinicians to understand the reasoning behind an AI-driven diagnosis or prediction. Without clear explanations for why a model classifies a nodule as malignant or benign, the model's trustworthiness and adoption by healthcare professionals are significantly compromised. This lack of transparency in model decision-making hinders its integration into clinical workflows and patient management systems.

High Computational Requirements: Deep learning models, especially those using 3D convolution or hybrid architectures, require significant computational resources. Training these models involves processing large volumes of high-resolution imaging data, which demands high-performance hardware, such as powerful GPUs. This increases the cost of developing and deploying AI models, making them less accessible for smaller clinics or hospitals with limited resources. Moreover, the need for high-end infrastructure also delays the real-time application of these models, limiting their use for immediate clinical decision-making.

False Positives: False positives are a persistent issue in lung cancer classification. Many deep learning models tend to misclassify benign nodules as malignant, which results in unnecessary procedures such as biopsies, surgeries, and follow-up tests. These procedures can cause psychological stress for patients, as well as financial and medical burden. Reducing false positives is essential to make AI models clinically viable. However, this challenge is difficult to overcome, as high sensitivity (to detect all potential malignancies) often comes at the cost of a higher false positive rate.

Poor Generalization Across Populations: Most deep learning models for lung cancer detection are trained on datasets derived from specific populations, often focusing on a single demographic group or geographical area. This can cause the models to be biased toward that group, leading to reduced performance when applied to other populations with different genetic, environmental, and lifestyle factors. Such population bias poses a major limitation for the broader applicability of AI models in global health contexts, where lung cancer manifests differently depending on a variety of external factors.

Integration into Clinical Workflows: Many proposed deep learning models are not designed with clinical integration in mind. Radiology departments often rely on specific hospital systems, such as PACS (Picture Archiving and Communication Systems) or RIS (Radiology Information Systems), to manage medical imaging data. For AI models to be useful in practice, they must seamlessly integrate with these systems, allowing doctors to use them without disrupting their workflow.

IV. PROPOSED PLAN

To address the challenges identified in the review of lung cancer classification using deep learning, several strategic approaches can be pursued. First, efforts should focus on the augmentation of datasets to mitigate the issue of limited labeled data. This can be achieved through synthetic data generation, leveraging data augmentation techniques, or utilizing transfer

learning to adapt models trained on large, publicly available datasets to smaller, specialized datasets. Additionally, techniques like semi-supervised learning can be explored to leverage unlabeled data effectively.

To handle the class imbalance problem, advanced sampling techniques like SMOTE (Synthetic Minority Over-sampling Technique) or cost-sensitive learning can be employed. These techniques adjust the model's loss function or the dataset itself to ensure that minority classes (malignant nodules) are given sufficient importance during training.

Furthermore, the development of interpretable deep learning models using methods like attention mechanisms or Grad-CAM (Gradient-weighted Class Activation Mapping) can help address the transparency issue, enabling clinicians to trust and understand the model's predictions. These methods can provide heatmaps that highlight areas of the image that the model deems most important for its classification decision.

Another avenue of improvement is the integration of multi-modal data. Future models should be capable of fusing not only imaging data but also genomic, clinical, and patient history data to improve prediction accuracy. This requires designing hybrid architectures that can effectively learn from diverse data types.

Finally, the deployment of deep learning models in clinical settings requires a focus on user-friendly interfaces and integration with existing hospital information systems. By ensuring that AI models seamlessly fit into clinical workflows, their adoption by healthcare professionals will be significantly enhanced. Moreover, ethical and regulatory concerns should be addressed from the outset, ensuring that patient data privacy is maintained and that the model complies with healthcare regulations such as HIPAA.

V. CONCLUSION

The deep learning techniques offer great promise for improving lung cancer classification and prediction, potentially leading to earlier detection and better patient outcomes. However, significant challenges remain, including data limitations, model interpretability, and integration into clinical workflows. Addressing these challenges through innovative solutions, such as data augmentation, transfer learning, and the fusion of multi-modal data, will be key to making these models viable for real-world clinical applications. Furthermore, ensuring transparency, fairness, and compliance with ethical standards will be crucial for building trust among healthcare professionals and patients. With continued research and development, deep learning-based systems could revolutionize the way lung cancer is diagnosed and treated, ultimately contributing to better healthcare delivery worldwide.

REFERENCES

1. P. S, V. B, L. Krishnasamy, T. P, P. R. M and S. S, "Lung Cancer Prediction using Machine Learning," *2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI)*, Greater Noida, India, 2025, pp. 1604-1608, doi: 10.1109/ICCSAI64074.2025.11063814.
2. T. I. A. Mohamed and A. E. -S. Ezugwu, "Enhancing Lung Cancer Classification and Prediction With Deep Learning and Multi-Omics Data," in *IEEE Access*, vol. 12, pp. 59880-59892, 2024, doi: 10.1109/ACCESS.2024.3394030.
3. Al-Tamimi, M., Noor, A., & Zahir, M. (2022). 3D Convolutional Neural Networks for Early Lung Cancer Detection Using Volumetric Data. IEEE Access, 10, 112234–112245.
4. Kim, D., Lee, S., & Kang, M. (2021). Dual-Path Deep Learning Framework Using GCN for Lung Cancer Prediction. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 5021–5032.
5. Liu, Y., Xie, Y., & Zhang, K. (2020). Multi-Task Deep Learning Model for Joint Lung Nodule Segmentation and Classification. IEEE Transactions on Medical Imaging, 39(12), 4034–4045.
6. Ypsilantis, P.P., & Montana, G. (2019). Deep Reinforcement Learning for Optimizing Lung Cancer Screening. IEEE Transactions on Medical Imaging, 38(4), 1075–1085.
7. Shen, W., Zhou, M., Yang, F., & Tian, J. (2018). Multi-Crop Convolutional Neural Networks for Lung Nodule Malignancy Classification. IEEE Transactions on Biomedical Engineering, 65(5), 1040–1050.

8. Setio, A.A.A., Traverso, A., De Bel, T., et al. (2017). Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images: The LUNA16 Challenge. IEEE Transactions on Medical Imaging, 36(10), 2050–2061.
9. Hua, K.L., Hsu, C.H., Hidayati, S.C., Cheng, W.H., & Chen, Y.J. (2016). Computer-Aided Classification of Lung Nodules on Computed Tomography Images via Deep Learning Technique. OncoTargets and Therapy, 9, 3711–3720.
10. Hussein, S., Cao, K., Song, Q., & Bagci, U. (2015). Risk Stratification of Lung Nodules Using 3D CNN-Based Multi-Task Learning. IEEE International Symposium on Biomedical Imaging (ISBI), 279–283.
11. U. G. Assembly, "Political declaration of the third high-level meeting of the General Assembly on the prevention and control of non-communicable diseases" in Resolution Adopted by the General Assembly, New York, NY, USA:United Nations Digital Library, Oct. 2018.
12. H. Fitipaldi and P. W. Franks, "Ethnic gender and other sociodemographic biases in genome-wide association studies for the most burdensome noncommunicable diseases: 2005–2022", Human Mol. Genet., vol. 32, no. 3, pp. 520-532, Jan. 2023.
13. F. Yuan, L. Lu and Q. Zou, "Analysis of gene expression profiles of lung cancer subtypes with machine learning algorithms", Biochimica et Biophys. Acta (BBA) Mol. Basis Disease, vol. 1866, no. 8, Aug. 2020.
14. K. Huang, Y. Zhang, X. Shi, Z. Yin, W. Zhao, L. Huang, et al., "Cell-type-specific alternative polyadenylation promotes oncogenic gene expression in non-small cell lung cancer progression", Mol. Therapy Nucleic Acids, vol. 33, pp. 816-831, Sep. 2023.
15. J. N. Bodor, Y. Boumber and H. Borghaei, "Biomarkers for immune checkpoint inhibition in non–small cell lung cancer (NSCLC)", Cancer, vol. 126, no. 2, pp. 260-270, Jan. 2020.
16. L. Ginn, L. Shi, M. La Montagna and M. Garofalo, "LncRNAs in non-small-cell lung cancer", Non-Coding RNA, vol. 6, no. 3, pp. 25, Jun. 2020.
17. L. Cai, S. Lin, L. Girard, Y. Zhou, L. Yang, B. Ci, et al., "LCE: An open web portal to explore gene expression and clinical associations in lung cancer", Oncogene, vol. 38, no. 14, pp. 2551-2564, Apr. 2019.
18. S. Osama, H. Shaban and A. A. Ali, "Gene reduction and machine learning algorithms for cancer classification based on microarray gene expression data: A comprehensive review", Expert Syst. Appl., vol. 213, Mar. 2023.
19. F. Alharbi and A. Vakanski, "Machine learning methods for cancer classification using gene expression data: A review", Bioengineering, vol. 10, no. 2, pp. 173, Jan. 2023.
20. Q. Xue, W. Peng, S. Zhang, X. Wei, L. Ye, Z. Wang, et al., "Promising immunotherapeutic targets in lung cancer based on single-cell RNA sequencing", Frontiers Immunol., vol. 14, Apr. 2023.