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

A Review on Diagnostic Framework for Lung Cancer Detection

[1]Shweta Srivastava, [2]Mr. Gunjan Kumar Mishra
[1]Research Scholar, Department of ECE, SHEAT College of Engineering, Varanasi, India
[2]Assistant Professor, Department of ECE, SHEAT College of Engineering, Varanasi, India

***Abstract*— A Review on Diagnostic Framework for Lung Cancer Detection presents a comprehensive analysis of existing computational and clinical approaches used for early and accurate identification of lung cancer. This review examines traditional diagnostic techniques such as imaging-based screening, histopathological analysis, and biomarker evaluation, along with advanced artificial intelligence (AI) and machine learning (ML) frameworks including convolutional neural networks (CNN), deep learning models, and hybrid algorithms applied to CT scan and X-ray image datasets. The study discusses data preprocessing methods, feature extraction techniques, model optimization strategies, and performance evaluation metrics such as accuracy, sensitivity, specificity, and AUC. Furthermore, it highlights the integration of IoT-enabled medical devices and cloud-based systems for real-time analysis and remote diagnosis. The review identifies current challenges including data imbalance, false positives, interpretability issues, and limited clinical validation, while outlining future research directions toward explainable AI, multimodal data fusion, and personalized diagnostic systems to enhance early-stage lung cancer detection and improve patient survival rates.**

Keywords—Lung Cancer, CNN, ML, AI .

I. INTRODUCTION

Diagnostic Framework for Lung Cancer Detection plays a vital role in improving early diagnosis, reducing mortality rates, and enhancing patient survival outcomes. Lung cancer is one of the leading causes of cancer-related deaths worldwide, primarily due to late-stage detection and rapid disease progression[1]. The disease often remains asymptomatic in its early stages, which makes timely identification extremely challenging. Traditional diagnostic approaches such as chest X-rays, computed tomography (CT) scans, biopsy procedures, and histopathological examinations are widely used in clinical practice. However, these methods are often time-consuming, expensive, and dependent on expert interpretation. Variability in radiological assessment and limited access to specialized healthcare facilities further complicate early detection, especially in developing regions[2].

A diagnostic framework for lung cancer detection refers to a structured and systematic approach that integrates medical imaging, computational intelligence, clinical data analysis, and decision-support systems to improve diagnostic accuracy[3]. The framework generally consists of multiple stages, including data acquisition, preprocessing, segmentation, feature extraction, classification, and performance evaluation. Modern diagnostic systems increasingly rely on artificial intelligence (AI) and machine learning (ML) algorithms to automate these stages. Deep learning models, particularly convolutional neural networks (CNNs), have demonstrated remarkable capability in detecting pulmonary nodules from CT scans with high precision. These models can learn complex patterns from large datasets and reduce human error in image interpretation[4].

The integration of advanced imaging techniques such as low-dose CT (LDCT) has significantly enhanced early screening programs. LDCT allows detection of small nodules that may not be visible in conventional radiographs. However, high false-positive rates remain a concern. To address this issue, hybrid diagnostic frameworks combine radiomic features, clinical parameters (such as age, smoking history, and genetic predisposition), and biomarker analysis to improve predictive performance[5]. Radiomics converts medical images into quantitative data, enabling deeper analysis of tumor shape, texture, and intensity characteristics. When combined with AI-based classifiers such as Support Vector Machines (SVM), Random Forest (RF), Gradient Boosting, or deep neural networks, diagnostic reliability improves substantially[6].

In recent years, multimodal data fusion has emerged as an important component of lung cancer diagnostic frameworks. This approach integrates imaging data, genomic profiles, electronic health records (EHRs), and laboratory results to provide a comprehensive understanding of disease progression. The inclusion of biomarkers such as circulating tumor DNA (ctDNA) and protein signatures further enhances early-stage detection[7]. Additionally, cloud-based platforms and Internet of Things (IoT)-enabled healthcare devices facilitate remote monitoring, real-time data sharing, and collaborative diagnosis among medical professionals[8].

Despite significant technological advancements, several challenges remain in developing an efficient diagnostic framework. These include data imbalance, limited availability of annotated medical datasets, variability in imaging protocols, and lack of interpretability in deep learning models. Clinical validation and regulatory approval are also critical barriers before deployment in real-world healthcare systems. Therefore, research is increasingly focused on explainable AI (XAI) techniques that enhance transparency and build trust among clinicians[9].

The diagnostic framework for lung cancer detection represents a multidisciplinary convergence of medical science, imaging technology, and computational intelligence. By integrating AI-driven analysis with clinical expertise, such frameworks aim to enable early detection, reduce false diagnoses, optimize treatment planning, and ultimately improve patient survival rates. Continuous research, data standardization, and collaboration between technologists and healthcare professionals are essential to realize the full potential of intelligent diagnostic systems in combating lung cancer[10].

II. LITERATURE SURVEY

B. S. et al.,[1] presented a machine learning-based framework for lung cancer detection using CT scan images obtained from a benchmark dataset. The study applied preprocessing techniques such as noise filtering, normalization, and image enhancement before feature extraction. Texture, shape, and statistical features were extracted and evaluated using classifiers including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest (RF). Experimental findings showed that SVM achieved an accuracy of 96.8%, while Random Forest obtained 95.4% accuracy. The reported sensitivity and specificity were 97.1% and 94.9%, respectively. The results demonstrated that combining multiple features improved diagnostic reliability and reduced misclassification rates.

Araújo Alves et al., [2] developed a lung disease classification approach based on lung tissue density analysis from CT images. The method extracted quantitative density parameters and applied statistical learning techniques for categorization. The framework effectively distinguished between normal and abnormal lung tissues. Experimental evaluation reported an overall accuracy of 93.2% with a precision of 92.5%. False positive rates were reduced by approximately 8% compared to traditional threshold-based segmentation. The findings highlighted that density-based metrics can significantly enhance early-stage detection accuracy.

Ho et al., [3] utilized knowledge distillation in deep learning for classification of chest X-ray abnormalities. A large teacher model transferred learned representations to a compact student network, reducing computational complexity. The distilled model preserved high classification performance while lowering memory requirements. The system achieved 94.6% accuracy with a sensitivity of 95.1% for abnormality detection. Inference speed improved by nearly 30% compared to the larger model. The study emphasized efficient deep learning deployment in resource-limited healthcare settings.

Sun et al., [4] introduced an adaptive feature selection-guided deep forest method for lung disease classification using chest CT scans. The approach dynamically selected relevant features during training to minimize redundancy. Ensemble learning techniques enhanced classification robustness and reduced overfitting. Experimental outcomes showed 97.3% accuracy with an AUC value of 0.98. Performance stability improved by around 5% when compared with conventional deep learning architectures. The study demonstrated that adaptive feature optimization strengthens diagnostic frameworks.

Baloescu et al., [5] focused on automated lung ultrasound B-line assessment through deep learning techniques. A convolutional neural network analyzed ultrasound images to detect B-line artifacts associated with pulmonary abnormalities. The system achieved 91.7% detection accuracy with a recall rate of 92.4%. Diagnostic time was reduced by approximately 40% when compared with manual evaluation. Agreement with expert radiologists remained consistently high. The findings supported the use of AI-driven ultrasound systems for rapid lung screening.

Wu et al., [6] developed a multilayer fractional-order machine vision classifier for rapid lung disease screening using digital chest X-ray images. Fractional-order texture features were extracted to improve representation of subtle image variations. A multilayer classification structure enhanced robustness against noise and imaging inconsistencies. The framework achieved 95.9% overall accuracy with 96.3% sensitivity. Comparative analysis indicated superior performance over traditional CNN-based

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

systems. The results suggested suitability for large-scale diagnostic applications.

Roy et al., [7] designed a deep learning system for classification and localization of lung disease markers in point-of-care ultrasound images. The architecture integrated segmentation and classification processes in a unified pipeline. Marker classification accuracy reached 96.1%, while localization precision was 94.8%. The model reduced misclassification rates by nearly 6% compared to baseline methods. Diagnostic consistency among clinicians improved significantly with AI assistance. The framework demonstrated the feasibility of portable and intelligent lung disease screening systems.

Pang et al., [8] combined densely connected convolutional networks (DenseNet) with adaptive boosting for lung cancer type identification. Dense connections enhanced feature reuse, while boosting improved classification stability. The hybrid architecture achieved 98.2% accuracy with an F1-score of 97.9%. Performance increased by approximately 4% compared to standalone DenseNet. The method effectively differentiated lung cancer subtypes with reduced overfitting. The study highlighted the advantage of ensemble-based deep learning frameworks in medical diagnostics.

Yazdani et al., [9] applied a bounded fuzzy possibilistic clustering method to analyze metabolomics data for lung cancer detection. The technique handled uncertainty and overlapping feature distributions effectively. Biomarker-based clustering improved separation between cancerous and non-cancerous samples. The framework achieved 92.6% classification accuracy and enhanced clustering validity index by 10%. Sensitivity for early-stage cancer detection improved significantly. The findings emphasized the role of metabolomic analysis in complementary diagnostic frameworks.

Yan et al., [10] integrated Long Short-Term Memory (LSTM) networks with DenseNet for automatic annotation and classification of chest X-ray images. The architecture captured both spatial features and sequential dependencies in image representations. The system achieved 95.4% overall accuracy with an AUC of 0.97. Performance improved by around 3% compared to conventional CNN models. Cross-dataset validation demonstrated strong generalization capability. The study indicated that combining temporal learning with deep convolutional networks enhances automated lung disease diagnosis.

III. CHALLENGES

Despite Challenges in Diagnostic Framework for Lung Cancer Detection arise due to the complex nature of medical imaging, variability in patient data, and the limitations of computational models in real-world clinical environments. Although artificial intelligence and machine learning techniques have significantly improved early detection accuracy, practical deployment still faces multiple technical, clinical, and ethical barriers. Issues such as data scarcity, model interpretability, integration with hospital systems, and regulatory approval restrict large-scale adoption. Furthermore, lung cancer diagnosis requires high sensitivity and specificity because even minor errors may lead to delayed treatment or unnecessary invasive procedures. Therefore, addressing these challenges is essential to build reliable, scalable, and clinically acceptable diagnostic frameworks.

1. Limited Availability of High-Quality Annotated Data

Medical image datasets require expert annotation by radiologists and oncologists, which is time-consuming and expensive. Many publicly available datasets are small and lack diversity in terms of age groups, ethnicity, and disease stages. Insufficient data leads to overfitting and poor generalization in deep learning models. Moreover, privacy regulations restrict data sharing across hospitals, limiting the development of large-scale robust systems.

2. Class Imbalance Problem

In lung cancer datasets, the number of normal cases often exceeds malignant cases. This imbalance causes machine learning models to become biased toward the majority class. As a result, sensitivity for detecting cancerous nodules decreases. Handling imbalance requires advanced techniques such as oversampling, undersampling, synthetic data generation, or cost-sensitive learning, which add complexity to model development.

3. High False Positive Rates

Screening techniques such as low-dose CT scans often detect small pulmonary nodules that are not cancerous. AI models may incorrectly classify benign nodules as malignant, increasing false positives. This leads to unnecessary biopsies, patient anxiety, and higher healthcare costs. Reducing false alarms while maintaining high sensitivity remains a critical challenge.

4. Variability in Imaging Protocols

Different hospitals use different imaging devices, scanning resolutions, and acquisition protocols. Such variability affects image quality and feature representation. Models trained on one dataset may not perform well on images from another medical center. Standardization of imaging protocols and domain adaptation techniques are required to improve cross-institutional reliability.

5. Lack of Interpretability in AI Models

Deep learning models, particularly convolutional neural networks, operate as “black boxes.” Clinicians often hesitate to rely on predictions without understanding how decisions are made. Lack of transparency reduces trust and slows clinical adoption. Explainable AI (XAI) techniques are necessary to visualize important regions and justify diagnostic decisions.

6. Computational Complexity and Infrastructure Requirements

Advanced deep learning models require high computational power, GPUs, and cloud infrastructure. Many healthcare centers, especially in rural or developing regions, lack such resources. Large models also increase processing time and energy consumption. Lightweight and optimized architectures are needed for real-time clinical deployment.

7. Integration with Clinical Workflow

Diagnostic frameworks must seamlessly integrate with hospital information systems, electronic health records (EHRs), and radiology workflows. Poor integration can disrupt routine medical processes and reduce usability. Ensuring compatibility, interoperability, and user-friendly interfaces is a major implementation challenge.

8. Regulatory, Ethical, and Legal Issues

AI-based diagnostic systems must comply with medical regulations and safety standards before deployment. Issues related to patient data privacy, algorithm bias, accountability in case of misdiagnosis, and ethical concerns complicate approval processes. Clear regulatory guidelines and validation through clinical trials are essential for safe implementation.

IV. CONCLUSION

The diagnostic framework for lung cancer detection represents a significant advancement in modern healthcare by integrating medical imaging, machine learning, deep learning, and clinical data analysis into a unified decision-support system. The review of existing literature demonstrates that intelligent algorithms such as CNN, DenseNet, LSTM, ensemble models, and hybrid classifiers have achieved high accuracy, sensitivity, and specificity in detecting lung abnormalities and cancer subtypes. These frameworks enhance early-stage identification, reduce diagnostic time, and support clinicians in making more accurate treatment decisions. However, challenges such as data imbalance, high false positive rates, interpretability issues, and integration with real-world clinical workflows still limit widespread implementation. Continuous improvement through explainable AI, multimodal data fusion, standardized imaging protocols, and large-scale clinical validation is essential to ensure reliability and scalability. Overall, an optimized and clinically validated diagnostic framework has the potential to significantly reduce lung cancer mortality rates and improve patient survival outcomes through timely and precise detection.

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.