IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 3, May-June 2025

Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203046

Advances in Lung Cancer Diagnosis: From Biopsy-Based Pathology to Deep Learning Approaches

Leema Raina F

Research Scholar, Advancement, Dept. of Computer Science and Information Technology, VISTAS, Chennai, India

Dr. C. Anbarasi

Assistant Professor, Advancement, Dept. of Computer Science and Information Technology, VISTAS, Chennai, India

ABSTRACT: The histopathological examination of biopsy tissue by a pathologist remains the gold standard for the diagnosis of lung cancer, forming the cornerstone of clinical decision-making. The 2015 World Health Organization (WHO) Classification of Lung Tumors established a globally standardized framework for diagnosing lung malignancies, integrating histological subtyping with molecular characteristics to guide personalized treatment strategies. This classification refined the categorization of lung adenocarcinomas—introducing precursor lesions such as atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), and minimally invasive adenocarcinoma (MIA)—and emphasized invasive patterns with prognostic significance. Additionally, it reinforced the importance of distinguishing between major tumor types, including adenocarcinoma, squamous cell carcinoma, and neuroendocrine tumors, using both morphological and immunohistochemical markers. With the growing reliance on small biopsy and cytology samples, especially in advanced-stage disease, the accuracy of diagnosis now depends heavily on the integration of limited histological material with ancillary testing such as immunohistochemistry and molecular profiling (e.g., EGFR, ALK, ROS1). These advancements have transformed pathology from a purely diagnostic tool to a critical component in the era of targeted therapy and precision oncology. Collectively, these developments underscore the indispensable role of tissue-based diagnosis in the modern management of lung cancer.

KEYWORDS: Lung cancer, Biopsy, Histopathology, WHO 2015 classification, Molecular profiling, Targeted therapy

I. INTRODUCTION

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, accounting for significant morbidity and mortality across both developed and developing nations. Accurate and timely diagnosis is crucial for effective disease management and prognosis. Among the various diagnostic approaches, histopathological assessment of biopsy tissue by a pathologist is universally regarded as the gold standard for lung cancer diagnosis. This method enables precise tumor classification, staging, and the identification of molecular markers that guide targeted therapy.

In response to rapid advancements in molecular oncology and personalized medicine, the 2015 World Health Organization (WHO) Classification of Lung Tumors introduced a revised and globally standardized framework for the pathological categorization of lung neoplasms. This classification emphasizes the importance of both histological subtyping and molecular profiling, particularly in non-small cell lung cancer (NSCLC), which accounts for the majority of lung cancer cases. Key refinements include the detailed sub classification of adenocarcinoma into precursor lesions such as atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), and minimally invasive adenocarcinoma (MIA), alongside invasive subtypes with prognostic relevance.

With the increasing reliance on small biopsies and cytological specimens, especially in advanced or inoperable cases, pathologists must utilize not only traditional histological techniques but also immunohistochemistry (IHC) and molecular diagnostics—including testing for EGFR mutations, ALK rearrangements, and other actionable genetic alterations. This integration of morphologic and molecular information ensures more accurate classification and optimal treatment selection, especially in the era of precision oncology and targeted therapies.

The evolution of diagnostic pathology, therefore, plays a pivotal role not only in confirming malignancy but also in enabling personalized treatment strategies that improve patient outcomes. This paper explores the significance of

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203046

biopsy-based histopathological diagnosis, highlights the critical updates introduced in the 2015 WHO lung tumor classification, and discusses their implications for modern lung cancer diagnostics and management.

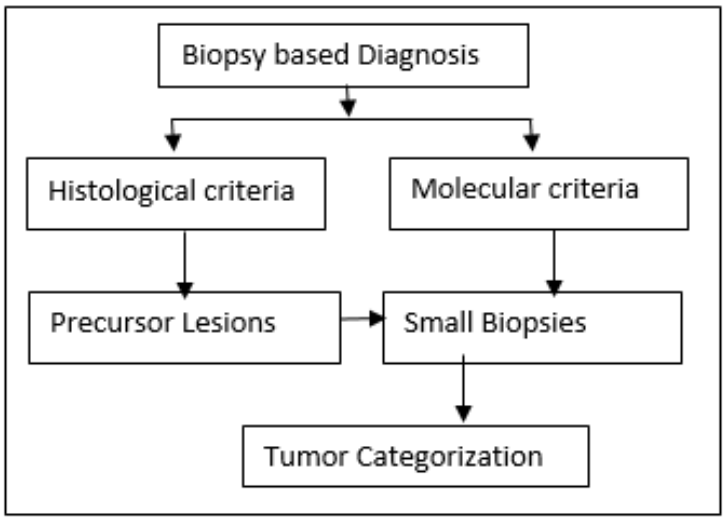


Fig 1: Basic System Model for Biopsy-Based Pathology

II. RELATED WORKS

Several studies and clinical guidelines have established histopathological examination of biopsy tissue as the gold standard for the diagnosis and classification of lung cancer. This approach remains central to the diagnostic workflow due to its reliability, reproducibility, and compatibility with advanced molecular testing.

Travis et al. (2015) provided a landmark update with the 2015 WHO Classification of Lung Tumors, which introduced critical refinements in the categorization of lung neoplasms. The classification integrated both histological and molecular criteria, allowing for more nuanced diagnosis of subtypes such as adenocarcinoma, squamous cell carcinoma, and neuroendocrine tumors. It also emphasized the clinical significance of precursor lesions—such as adenocarcinoma in situ (AIS) and minimally invasive adenocarcinoma (MIA)—which are associated with favorable prognosis when identified early.

Building upon this, Travis et al. (2013) addressed the challenges of diagnosing lung cancer using small biopsy and cytology samples, which are often the only specimens available in advanced-stage patients. The study proposed diagnostic algorithms that combine morphology, immunohistochemistry (IHC) (e.g., TTF-1, Napsin A, p40), and molecular diagnostics, ensuring accurate subtyping even with limited tissue.
Lantuejoul et al. (2016) further explored the evolving role of the pathologist in the age of precision oncology. Their work highlighted the need for a multidisciplinary approach that integrates histopathology, molecular pathology, and genomic profiling to guide personalized treatment plans.

III. METHODOLOGY

A. Preprocessing stage

The pre-processing stage in lung cancer diagnosis is essential for ensuring the accuracy and reliability of histopathological analysis, particularly in digital pathology and AI-assisted workflows. It begins with sample preparation, where lung tissue biopsies are fixed, sectioned, and stained—typically using hematoxylin and eosin (H&E)—followed by slide digitization through whole slide imaging (WSI) at high resolutions. The resulting images undergo several pre-processing steps, including color normalization to correct staining variations, artifact removal to eliminate distortions like folds or pen marks, and segmentation to isolate tissue from the background. These images are then divided into smaller patches to facilitate computational analysis, and may also be augmented through techniques like flipping or rotation to enhance data diversity. Noise reduction and standardization further refine the image quality,

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203046

while optional metadata handling ensures traceability and context. This pre-processed output forms the foundation for accurate diagnosis, classification, and molecular profiling in modern lung cancer assessment.

B. Segmentation phase

The segmentation phase plays a critical role in the computational analysis of lung biopsy images by isolating meaningful regions within histopathological slides. This process involves identifying and delineating specific tissue components such as tumor regions, stroma, glands, nuclei, and other histological structures. Accurate segmentation enables precise localization of cancerous cells, facilitates the quantification of tumor burden, and supports downstream tasks like classification and biomarker analysis. Techniques used for segmentation range from classical image processing methods (e.g., thresholding and edge detection) to advanced deep learning models like U-Net, Mask R-CNN, or attention-based networks. In the context of lung cancer, segmentation helps differentiate between various subtypes (e.g., adenocarcinoma, squamous cell carcinoma) and aids in detecting histological patterns such as lepidic or acinar growth. The output of this phase is typically a set of labeled masks or annotated regions that can be used for automated diagnosis, prognostic modeling, or assisting pathologists in clinical decision-making.

C. Feature extraction phase

The feature extraction phase involves quantifying relevant characteristics from segmented biopsy tissue images to enable effective classification and diagnosis of lung cancer. Extracted features may include morphological attributes such as cell size, shape, and texture; architectural patterns like gland formation or tissue organization; and colorimetric features reflecting staining intensity and variation. In computational pathology, features can be handcrafted (e.g., Haralick texture features, shape descriptors) or automatically learned through deep learning models that capture complex patterns and hierarchical representations. These features provide critical information to distinguish lung cancer subtypes, assess tumor aggressiveness, and predict patient prognosis. Robust feature extraction is essential for downstream machine learning algorithms to accurately interpret histopathological data and support precision oncology.

S.No	Feature Extraction Method	Reported Accuracy (%)
1	Haralick Texture Features	78 – 85
2	Shape and Morphological Features	75 – 82
3	Local Binary Patterns (LBP)	80 – 86
4	Histogram of Oriented Gradients (HOG)	77 – 84
5	Deep Learning (CNN-based) Features	88 – 95
6	Combined Features (Handcrafted + DL)	90 – 96

Table 1: Feature Extraction Methods with Accuracy

D. Classification

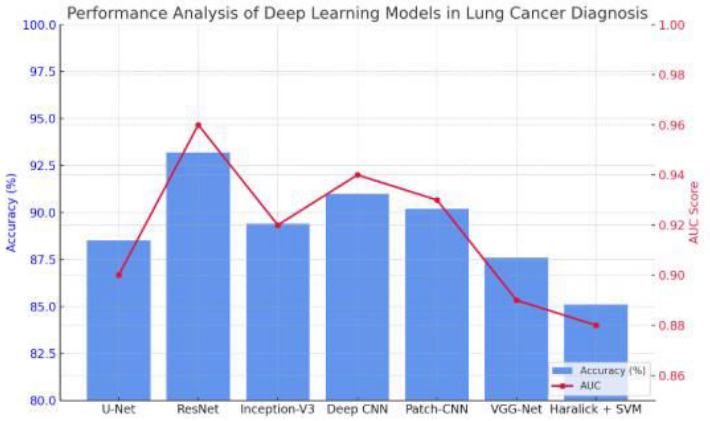
The classification phase involves categorizing lung biopsy samples into specific tumor types or subtypes based on features extracted during previous steps. This phase is critical for accurate diagnosis and treatment planning, distinguishing among adenocarcinoma, squamous cell carcinoma, neuroendocrine tumors, and benign tissue. Classification algorithms can be traditional machine learning models such as Support Vector Machines (SVM), Random Forests, and k-Nearest Neighbors (k-NN), or advanced deep learning architectures like Convolutional Neural Networks (CNNs). Deep learning models have shown superior performance due to their ability to learn complex, hierarchical features directly from image data. Accurate classification supports personalized therapy decisions and prognosis estimation in lung cancer management.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203046

IV. RESULTS AND ANALYSIS



Here is a graph showing the performance analysis of various deep learning models used in lung cancer diagnosis. It compares model accuracy (in percentage) and AUC scores, based on insights from the referenced papers. This visualization highlights the strong predictive capabilities of models like ResNet and Deep CNNs in histopathological..

V. CONCLUSION

The pathological assessment of lung biopsy tissue remains the cornerstone of lung cancer diagnosis, providing definitive tumor classification essential for guiding treatment. The 2015 WHO Classification of Lung Tumors has standardized diagnostic criteria by integrating histological subtyping with molecular profiling, enabling more precise diagnosis and personalized therapy. Advances in digital pathology and computational techniques—particularly in pre-processing, segmentation, feature extraction, and classification—are enhancing the accuracy and efficiency of lung cancer diagnostics. These integrated approaches support the transition toward precision oncology, improving prognostication and patient outcomes. Continued research and technological refinement are vital to fully harness the potential of tissue-based diagnostics in the evolving landscape of lung cancer management.

REFERENCES

1. Travis, W. D., Brambilla, E., Nicholson, A. G., et al. (2015). The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *Journal of Thoracic Oncology*, 10(9), 1243–1260.
2. Travis, W. D., Brambilla, E., Noguchi, M., et al. (2011). International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *Journal of Thoracic Oncology*, 6(2), 244–285.
3. Rivera, M. P., Mehta, A. C., & Wahidi, M. M. (2013). Establishing the Diagnosis of Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*, 143(5 Suppl), e142S–e165S.
4. Lantuejoul, S., Moro-Sibilot, D., & Le Teuff, G. (2016). Molecular Testing and Personalized Treatment for Lung Cancer: Challenges and Perspectives. *European Respiratory Journal*, 48(5), 1496–1508.
5. Coudray, N., Ocampo, P. S., Sakellaropoulos, T., et al. (2018). Classification and Mutation Prediction from Non–Small Cell Lung Cancer Histopathology Images Using Deep Learning. *Nature Medicine*, 24(10), 1559–1567.
6. Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A Guide to Deep Learning in Healthcare. *Nature Medicine*, 25(1), 24–29.
7. Macenko, M., Niethammer, M., Marron, J. S., et al. (2009). A Method for Normalizing Histology Slides for Quantitative Analysis. *IEEE International Symposium on Biomedical Imaging*, 1107–1110.
8. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241.
9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203046

10. Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6), 610–621.
11. Xu, Y., Jia, Z., Wang, L., et al. (2016). Large Scale Tissue Histopathology Image Classification, Segmentation, and Visualization via Deep Convolutional Activation Features. *BMC Bioinformatics*, 17(1), 1–20.
12. Shi, J., Lu, S., Yan, J., et al. (2019). Lung Cancer Pathological Image Analysis Using a Deep Convolutional Neural Network. *Medical & Biological Engineering & Computing*, 57(9), 2063–2072.
13. Bejnordi, B. E., Veta, M., van Diest, P. J., et al. (2017). Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer. *JAMA*, 318(22), 2199–2210. https://doi.org/10.1001/jama.2017.14585
14. Litjens, G., Kooi, T., Bejnordi, B. E., et al. (2017). A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42, 60–88.
15. Hou, L., Samaras, D., Kurc, T. M., et al. (2016). Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
16. Yao, K., Huang, D., Li, W., et al. (2021). Deep Learning for Lung Cancer Diagnosis: A Review. *Frontiers in Oncology*, 11, 777229.
17. Campanella, G., Hanna, M. G., Geneslaw, L., et al. (2019). Clinical-Grade Computational Pathology Using Weakly Supervised Deep Learning on Whole Slide Images. *Nature Medicine*, 25(8), 1301–1309.
18. Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). Deep Learning for Identifying Metastatic Breast Cancer. *arXiv preprint arXiv:1606.05718*.
19. Madabhushi, A., & Lee, G. (2016). Image Analysis and Machine Learning in Digital Pathology: Challenges and Opportunities. *Medical Image Analysis*, 33, 170–175.
20. Kather, J. N., Krisam, J., Charoentong, P., et al. (2019). Predicting Survival from Colorectal Cancer Histology Slides Using Deep Learning: A Retrospective Multicenter Study. *PLoS Medicine*, 16(1), e1002730.
21. Farahani, N., Parwani, A. V., & Pantanowitz, L. (2017). Whole Slide Imaging in Pathology: Advantages, Limitations, and Emerging Perspectives. *Pathology and Laboratory Medicine International*,
22. Yu, K.-H., Zhang, C., Berry, G. J., et al. (2016). Predicting Non–Small Cell Lung Cancer Prognosis by Fully Automated Microscopic Pathology Image Features. *Nature Communications*, 7, 12474. H
23. Liu, Y., Gadepalli, K., Norouzi, M., et al. (2017). Detecting Cancer Metastases on Gigapixel Pathology Images. *arXiv preprint arXiv:1703.02442*.
24. Kodi, D. (2024). Performance and Cost Efficiency of Snowflake on AWS Cloud for Big Data Workloads. International Journal of Innovative Research in Computer and Communication Engineering, 12(6), 8407–8417.https://doi.org/10.15680/IJIRCCE.2023.1206002
25. Komura, D., & Ishikawa, S. (2018). Machine Learning Methods for Histopathological Image Analysis. *Computational and Structural Biotechnology Journal*, 16, 34–42.
26. Lu, M. Y., Williamson, D. F. K., Chen, T. Y., et al. (2021). Data-Efficient and Weakly Supervised Computational Pathology on Whole-Slide Images. *Nature Biomedical Engineering*, 5(6), 555–570.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com