IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

A Gatekeeper – Enhanced Deep Learning System of Lung Cancer Prognosis using Histopathology Images

M. S. Kale[1], A. S. Aware[2], D. B. Berad[3], P. A. Chavan[4], and S. K. Yemul[5]
Department of Computer Engineering[1-5]
Dr. Vithalrao Vikhe Patil College of Engineering, Ahilyanagar, India
[1] madhavi4186@gmail.com, [2] akshadaaware8055@gmail.com, [3] diptiberad@gmail.com
[4] preranac90@gmail.com, [5] sakshiyemul1524@gmail.com

Abstract: *Lung cancer is one of the biggest health problems for people all over the world. This shows how important it is to find it early, tell the difference between subtypes, and give a reliable prognosis. The current system shows a multi-level deep learning model that uses histopathology images to find cancer, classify it, and look at survival rates. The pipeline begins with a Gatekeeper module, which checks that the input is a valid histopathology slide. This makes the next steps more reliable. After being checked, the CancerD module uses a convolutional architecture that is meant to capture fine morphological details to find malignant tissue. The CancerSC module then uses optimized feature extraction to sort the tumors it finds into different types of lung cancer. Finally, the CancerSA module combines imagebased features with important patient information to figure out the risk of survival and help doctors make predictions about the future. The proposed system demonstrated a detection accuracy of 96.4% , an overall subtype classification accuracy of 92.1% , and a survival prediction performance with a concordance index of 0.71, based on experimental evaluation of publicly available lung cancer histopathology datasets. The proposed framework helps cut down on the amount of work that doctors have to do by hand, makes things more consistent, and gives lung cancer pathologists an effective AI-assisted decision support system.*

Keywords: Convolutional Neural Networks (CNNs), Deep Learning, Lung cancer, Histopathology images, Cancer-detection, Subtype classification, and Survival Analysis.

I. INTRODUCTION

Lung cancer remains one of the most difficult and fatal types of cancer worldwide, and its early diagnosis has proven to be one of the most impactful factors in improving patient outcomes [1]. The increasingly popular field of digital pathology enables the application of deep learning techniques to interpret histopathological images, offering automated and objective data analysis while enhancing the efficiency of pathologists [2]. In particular, Convolutional Neural Networks (CNNs) have demonstrated strong capability in learning complex visual patterns from tissue slides and are actively investigated for tasks such as tumor detection [3], subtype classification, and prognosis estimation [4].

Despite these advancements, several challenges hinder the development of clinically reliable AI-based pathology systems. Histopathological images often exhibit significant variations in color, staining, magnification, and region-of-interest (ROI) distribution, which may reduce model robustness if not adequately addressed [5]. Furthermore, lung cancer subtypes frequently share subtle morphological characteristics, requiring highly discriminative feature representations for accurate classification [9] [11] [12]. Patient survival prediction is also a complex task due to the nonlinear relationships among tissue architecture, tumor progression, and clinical outcomes, which traditional statistical approaches may fail to capture effectively [13].

DOI: 10.48175/IJARSCT-36736

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

To address these challenges, this paper proposes *CancerScan*, a multi-stage deep learning framework specifically designed for lung cancer analysis using histopathological images. The framework begins with a Gatekeeper Model that verifies whether the input image belongs to the histopathology domain, ensuring clean and domain-appropriate data flow. Subsequently, the *CancerD* module identifies malignant tissue regions using a set of CNN architectures. The *CancerSC* module performs subtype classification through multimodel feature fusion to distinguish morphologically similar lung cancer subtypes [14]. Finally, the *CancerSA* module integrates imagebased features with patient-related clinical data to estimate survival probability using deep survival learning techniques that model nonlinear risk patterns [15].

The proposed CancerScan framework aims to reduce diagnostic workload, enhance reliability, and provide a comprehensive AI-assisted decision support system for lung cancer pathology. Future work will focus on improving model explainability, conducting multicenter validation, and facilitating seamless integration into clinical environments. Recent AIdriven studies further confirm that automated medical image analysis can significantly enhance lung cancer detection efficiency [21], effectively handle data variability [22], ensure consistent predictive performance through proper model validation [23], and automate cancer detection and subtype differentiation using histopathology images [24].

II. LITERATURE SURVEY

Deep learning has made significant contributions to automated lung cancer analysis. A multimodel framework combining subtype classification and survival estimation was proposed in [1]. Subsequently, improved survival prediction using CNN-based models, outperforming traditional clinical staging systems, was presented in [2]. The integration of deep and handcrafted features achieved higher accuracy in CT-based prognostics [3], while CNN classifiers demonstrated reliable performance in lung cancer subtype classification [4].

The incorporation of multiomics data further enhanced survival subtype classification capabilities [5], and extended evaluations highlighted the effectiveness of deep learning across diverse cancer datasets [6]. Multimodal deep learning models improved long-term survival prediction [7], whereas unsupervised feature extraction methods [8] and CNN-based survival analysis approaches [9] contributed to more robust risk modeling. CT-based survival prediction systems were validated in [10].

Survival models derived from histopathological images demonstrated strong generalizability [12], and variational autoencoder (VAE)-based survival models showed improved performance on limited clinical datasets [13]. Personalized recurrence and survival prediction frameworks based on machine learning were emphasized in [14], while transfer learning approaches using gene expression features further enhanced survival estimation accuracy [15].

Recent studies continue to reinforce the importance of deep learning in lung cancer analysis. CNN-based approaches for improved subtype classification were proposed in [18]. Additionally, handcrafted GLCM features combined with SVM classifiers demonstrated competitive performance in CT-based lung cancer detection [19]. Transfer learning and hybrid CNN architectures were shown to be more robust across multiple lung cancer types in a comparative evaluation presented in [20]. Overall, these studies emphasize the growing importance of multimodal integration and deep learning techniques for effective lung cancer diagnosis and prognosis.

In parallel, early lung cancer detection has gained increasing attention in machine learning research. Data-driven biomarker discovery methods have demonstrated strong potential for enhancing prognostic evaluation and supporting early diagnosis through advanced learning techniques [25]. Collectively, these findings further validate the expanding role of deep learning and multimodal analysis in lung cancer research.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

III. METHODOLOGY

The proposed work is a multi stage, deep learning-based, comprehensive system to autonomously identify lung cancer, subtypes, and survival outcomes based on histopathology images. The design of the methodology is in such a way that it guarantees the validity of input , capability to predict cancer accurately using multi classes, and provide clinically meaningful levels of survival estimation.The overall workflow consists of preprocessing, Gatekeeper validation, model pipelines (CancerD, CancerSC, CancerSA), and final risk prediction as described in Fig.1.

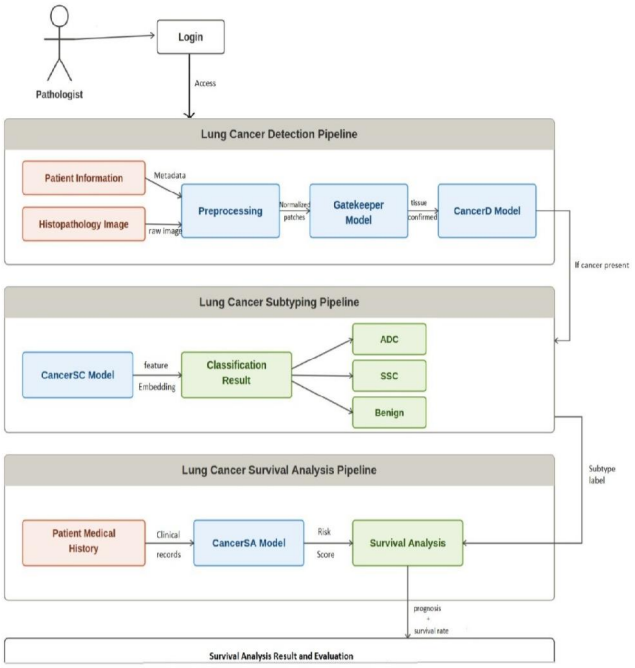

Fig. 1: System Architecture

A. Dataset and Preprocessing

We are using publicly available histopathology datasets, including the Lung Cancer Histopathology Image Dataset and TCGA dataset [16], [17]. These data sets have a variety of lung cancer subtypes images (Benign, ADC, SCC, SCLC) as illustrated in Fig.2 and Fig.3 and, all the images were resized to 224 × 224 to fit the input layer of the deep learning models. The pixels are scaled to the range of [0, 1] to allow the models to be trained much faster and enhance numerical stability.

This preprocessing pipeline will be used to make sure that images uploaded to deployment are of the same standardized format as the training data and will give consistent and reliable model results.

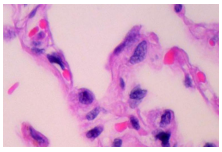

Fig. 2: Non-Cancerous Histopathology Image

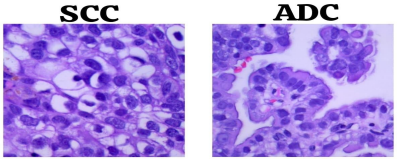

Fig. 3: Cancerous Histopathology Images

DOI: 10.48175/IJARSCT-36736

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

B. Gatekeeper Validation Module

The Gatekeeper module is trained as a lightweight CNNbased binary classifier which differentiates between valid histopathology images and non histopathology or low quality resulting images. The model takes normal images and imput , resized to 224x224 and normalised image in the 0-1 range. It is made up of a stack of few convolutional layers with ReLU activation and max pooling followed by a fully connected layer and a sigmoid output in case of binary classification. In case the probability that is predicted fails to satisfy the histopathology condition, the image is discarded and is not sent to the CancerD model. This guarantees that valid and relevant tissue images will only make it into the main pipeline and avoid unnecessary errors within the downstream modules.

C. Cancer Detection and Subtype Classification

Once the validation process is completed, the images are sent to CancerD, a convolutional neural network developed to identify malignancy in the images. If the images are determined to be malign, they are routed to CancerSC, a multiclass cancer subtype classifier based on LCSCNet. There are several layers of enhanced CNN. The architecture consists of several stacked convolutional blocks, each of which has the following:

a ReLU activation function, nonlinear activation max pooling, spatial down sampling a convolutional layer, which extracts features locally dropout, which relaxes overfitting to avoid (0.3). The flattened feature maps are sent to a fully connected layer. The classification layer is a softmax layer which generates the probabilities of each subtype. This design enables the sharp detection of cancer to be followed up with the reliable classification of the subtype of the cancer, which is quite beneficial.

D. Survival Prediction (CancerSA)

In the case of malignant cases, the CancerSA module does the survial prediction on an improved deep learning model inspired by LCSANet [1]. High level feature embeddings from the subtype classification network are aggregated to obtain a slide level representation. When available, patient information is integrated with these embeddings.

The model outputs a risk score representing the likelihood of reduced survival. Performance is evaluated using the concordance index (c-index), which measures how well predicted risk rankings match actual patient outcomes.

E. Deployment and Output

The following Fig. 4 shows the flow of system.

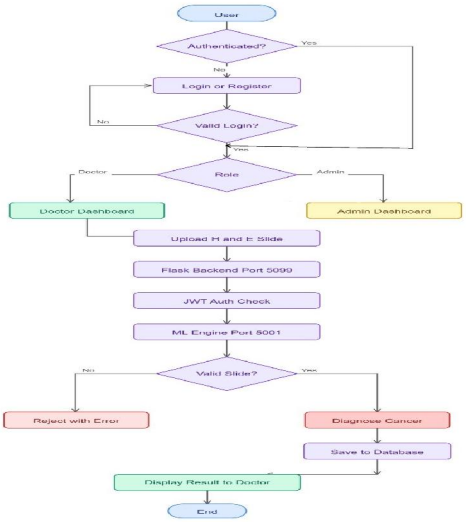

Fig. 4: Cancerscan System Flowchart

DOI: 10.48175/IJARSCT-36736

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

The complete system is deployed as a desktop application designed for pathologists. Pathologist can upload histopathology images along with clinical information such as age,smoking history and gender. The Fig. 5 shows the Pathologist Dashboard from Cancerscan system. The application applies the same preprocessing pipeline used during training, ensuring consistency and reliable inference across real world cases. Fig. 6 and Fig. 7 shows the final report of a patient having lung cancer along with Lung cancer subtype, survival rate and clinical findings.

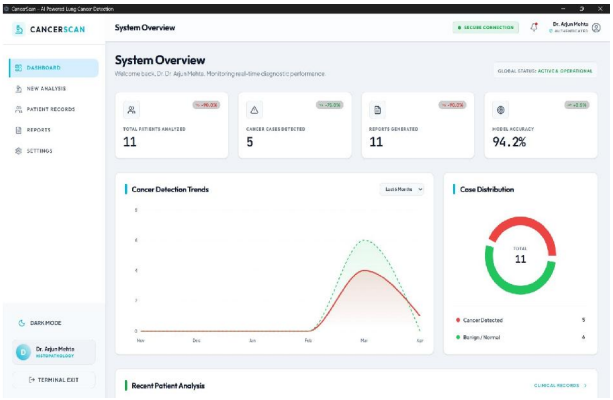

Fig. 5: Cancerscan System Pathologist Dashboard

The Pathologist Dashboard provides a comprehensive overview of diagnostic activities, patient records, and cancer detection statistics. It acts as the primary workspace where pathologists can manage analyses and review results efficiently.pathologists can upload histopathology images and enter rel evant patient information for analysis. The CancerScan framework processes the provided data and performs automated cancer detection, subtype classification, and survival predic tion.

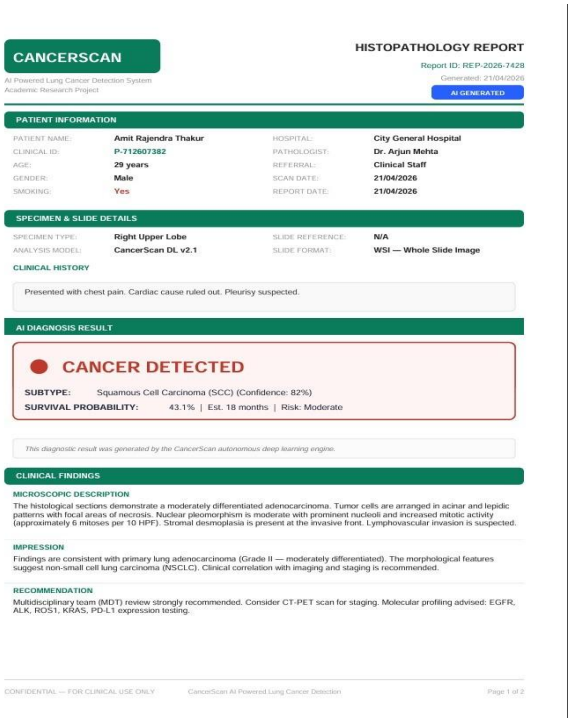

Fig. 6: Cancerscan System Report Page-1

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

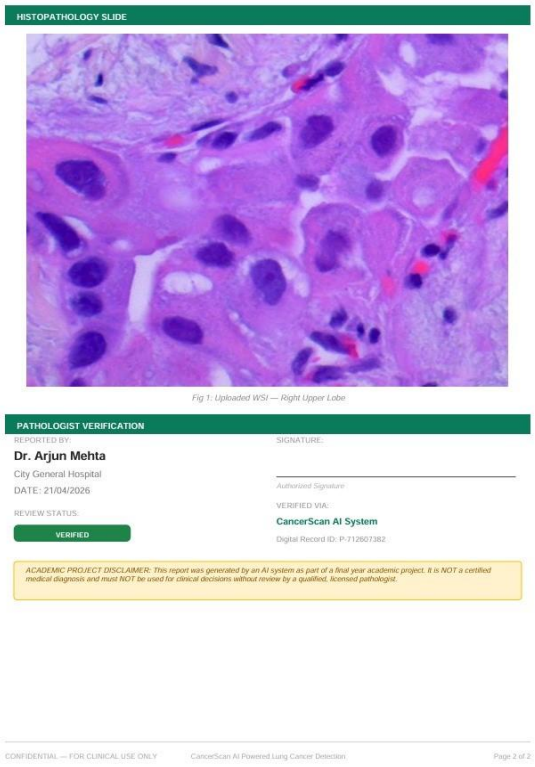

Fig. 7: Cancerscan System Report Page-2

IV. MATHEMATICAL MODEL

The Fig. 8 mathematically shows the multi stage framework for lung cancer detection, subtype classification, and survival analysis in this section.

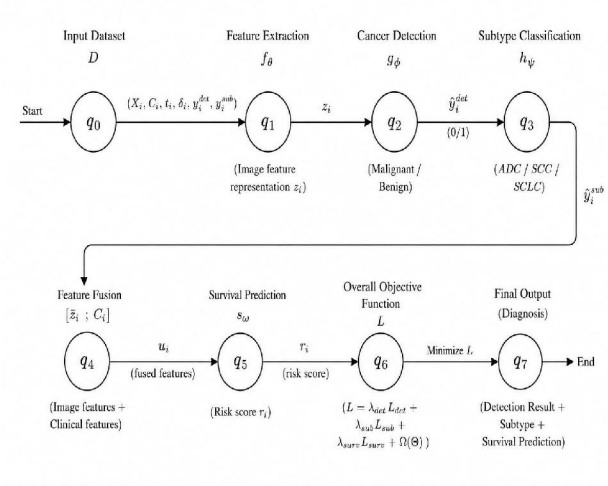

Fig. 8: Mathematical Model

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

A. Notation **Consider a dataset:**

$D = \{(X_i, C_i, t_i, \delta_i, y_i^{det}, y_i^{sub})\}_{i=1}^{N}$

where:

- N: total number of patients.
- X_i: whole slide histopathology image (WSI) of patient i.
- $C_i \in R^p$: optional clinical covariates (age, tumor stage, etc.).
- t_i: observed survival/censoring time.
- $\delta_i \in \{0,1\}$: event indicator (1 = event occurred, 0 = censored).
- $y_i^{det} \in \{0,1\}$: detection label (1 = malignant, 0 = benign).
- $y_i^{sub} \in \{1,...,K\}$: subtype class label ($K$ = total subtypes).

B. Patch Extraction and Feature Encoding **Each**

WSI is divided into patches:

$$P(X_i) = \{p_{i1}, p_{i2}, ..., p_{iKi}\}$$

where K_i is the number of extracted patches.

A CNN encoder produces a feature embedding:

$$zik = f\theta(pik) \in Rd$$

where d is the embedding dimension and θ denotes learnable parameters.

C. Patch Aggregation

Patch embeddings are aggregated to obtain a patient level representation:

$$\bar{z}_i = A(\{z_{ik}\})$$

where $A(\cdot)$ is an aggregation operator (mean pooling or attention based pooling).

D. Prediction Heads

1) Detection Head:

$$\hat{y}idet = g\phi(\bar{z}i)$$

where g_ϕ is a binary classifier with parameters ϕ.

2) Subtype Classification Head:

$$\hat{y}isub = h\psi(\bar{z}i) \in \Delta K-1$$

where h_ψ is a softmax classifier with parameters ψ.

3) Survival Risk Head:

$$r_i = s_\omega([\bar{z}_i; C_i])$$

where s_ω is the risk scoring function and $[\cdot;\cdot]$ denotes feature concatenation.

E. Overall Objective

$$L = \lambda detLdet + \lambda subLsub + \lambda survLsurv + \Omega(\Theta)$$

where:

- $\lambda_{det}, \lambda_{sub}, \lambda_{surv}$: task specific weights.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

- $\Omega(\Theta)$: L_2 regularization term.
- $\Theta = \{\theta, \phi, \psi, \omega\}$: all learnable parameters.

F. Training Procedure

1) Train the encoder f_θ and detection head g_ϕ using L_{det}.
2) Fine-tune encoder f_θ and train subtype head h_ψ using *Lsub*.
3) Freeze or lightly tune f_θ, compute $\bar{z}_i$, and train survival head s_ω using L_{surv}.

V. EXPERIMENTAL RESULTS AND ANALYSIS

An 80:20 training testing split was employed to evaluate the proposed *CancerScan* framework on a publicly available Lung Histopathology Image Dataset. All experiments were conducted using identical stain normalization and preprocessing techniques to ensure fair comparison.

A. Cancer Detection and Subtype Classification

The *CancerD* module performed binary classification to distinguish between benign and malignant tissue samples. Model performance was evaluated using accuracy, precision, recall, and F1-score. The results are summarized in Table I.

TABLE I: Cancer Detection Performance

Metric	Value (%)
Accuracy	96.4
Precision	95.8
Recall	97.1
F1-score	96.4

For malignant samples, the *CancerSC* module further classified lung cancer into Adenocarcinoma (ADC), Squamous Cell Carcinoma (SCC), and Small Cell Lung Cancer (SCLC). Subtype-wise and overall classification accuracy are reported in Table II.

TABLE II: Subtype Classification Accuracy

Subtype	Accuracy (%)
ADC	93.2
SCC	91.8
SCLC	90.4
Overall	92.1

B. Survival Prediction and Baseline Comparison

The *CancerSA* module predicted patient survival risk using clinical attributes such as age and gender. The survival prediction performance achieved a concordance index (Cindex) of 0.71, indicating effective patient risk stratification. A comparison between the proposed framework and baseline models is presented in Table III.

TABLE III: Baseline Model Comparison

Model	Accuracy (%)
Basic CNN	88.6
VGG16 / ResNet	92.3
Proposed	96.4

DOI: 10.48175/IJARSCT-36736

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

CancerScan	

VI. CONCLUSION

A multiphase deep learning framework for automated lung cancer analysis using histopathology images was successfully developed and assessed in this work. In addition to a Gatekeeper module that guarantees input reliability prior to analysis, the proposed system incorporates three core modules: CancerD for cancer detection, CancerSC for subtype classification, and CancerSA for survival prediction. The framework provides a reliable and efficient solution for thorough lung cancer assessment by combining these elements into a single end to end pipeline. The experimental findings show that the suggested method successfully improves diagnostic precision while lowering manual labor, enabling quicker and better clinical decision-making. In addition to increasing reliability, the system's modular design permits future integration of more diagnostic features and scalability. Overall, this study shows how deep learning-driven histopathological analysis can help pathologists and clinicians make better decisions by improving lung cancer diagnosis, prognosis, and treatment planning.

VII. FUTURE WORK

Further research will be undertaken to enhance stain normalization, ROI extraction, and use of other augmentation methods to strengthen the model. Heatmaps and attention-based visualization as explainability mechanisms will be included in order to enhance clinical transparency. Subsequent validation on multicenter data and co-operation with the medical practitioners will aid to test the clinical viability of the system. Finally, it aims at creating a completely trustworthy AI-based diagnostic system applicable in real-world pathology facilities.

REFERENCES

1. M. Aharonu and L. Ramasamy, "Deep learning–driven multi-model approach for predicting survival rates across lung cancer subtypes," *IEEE Access*, 2024.
2. Y. She *et al.*, "A neural-network-based method for estimating survival outcomes in non-small cell lung cancer," *JAMA Network Open*, vol. 3, no. 6, 2020, Art. no. e205842.
3. R. Paul *et al.*, "Integrating CNN-derived and handcrafted image features to improve lung cancer survival forecasting using CT scans," in *Proc. IEEE SMC*, 2016, pp. 2570–2575.
4. M. Kriegsmann *et al.*, "Automated identification of small- and nonsmall-cell lung cancers using deep learning classification," *Cancers*, vol. 12, no. 6, p. 1604, 2020.
5. S. Takahashi *et al.*, "Multi-omics fusion using machine learning for stratifying lung cancer survival groups," *Biomolecules*, vol. 10, no. 10, p. 1460, 2020.
6. S. Gupta *et al.*, "Assessment of deep learning models for colon cancer survival prediction using SEER data," *BioMed Research International*, vol. 2022, pp. 1–12, 2022.
7. L. A. Vale-Silva and K. Rohr, "Predicting long-term cancer outcomes with multimodal deep architectures," *Scientific Reports*, vol. 11, 2021.
8. S. Wang *et al.*, "Feature extraction using unsupervised deep learning for overall survival analysis in lung cancer," in *Proc. EMBC*, 2018, pp. 2583–2586.
9. C. Haarburger *et al.*, "CNN-based survival modeling for patients with lung cancer," in *Proc. IEEE ISBI*, 2019, pp. 1197–1201.
10. N. Cherukuri *et al.*, "Deep learning framework for diagnosing lung cancer using CT scans from NSCLC patients," in *Proc. ICAIS*, 2021, pp. 325–330.
11. H. Wang *et al.*, "Discovering new imaging biomarkers for NSCLC classification and survival determination," *BMC Bioinformatics*, vol. 15, 2014.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

12. E. Wulczyn *et al.*, "Histopathology-based survival prediction across multiple cancer types using deep learning," *PLoS ONE*, vol. 15, no. 6, 2020.
13. T.-H. Vo *et al.*, "Survival estimation in lung cancer using a multitask variational autoencoder with limited clinical records," *Electronics*, vol. 10, no. 12, p. 1396, 2021.
14. Y. Yang *et al.*, "Machine-learning-enabled prediction of recurrence and survival in lung cancer patients," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 1811–1820, 2022.
15. G. Lopez-Garc´ıa *et al.*, "Enhancing cancer survival prediction using CNN-based transfer learning on gene expression data," *PLoS ONE*, vol. 15, no. 3, 2020.
16. Lung and Colon Cancer Histopathological Image Dataset. Accessed: Oct. 2023. [Online]. Available: academictorrents.com
17. The Cancer Genome Atlas (TCGA) Program. Accessed: Oct. 2023. [Online]. Available: cancer.gov
18. S. Z. Reem *et al.*, "Improved identification of lung cancer subtypes using CNN-based deep learning," in *Proc. ICECE*, 2024.
19. Q. Firdaus *et al.*, "CT-driven lung tumor detection using GLCM texture features and SVM classification," in *Proc. IES*, 2020.
20. A. Sultana, T. T. Khan, and T. H. Hossain, "Evaluation of transfer learning and hybrid CNN architectures for multi-type lung cancer classification," in *Proc. EICT*, 2021.
21. M. Pavanalaxmi, M. Praveen Kumar, R. Nayak, N. S. Prameela, and C. Singh, "AI-driven lung cancer detection for rapid analysis of medical imaging data," in Recent Advances in Signals and Systems (VSPICE 2023), Lecture Notes in Electrical Engineering, vol. 1227, Singapore: Springer, 2024, pp. 225–235.
22. K. S. Geethanjali, N. Umashankar, I. S. Rajesh, K. Jagannathan, M. S. Krishnamurthy, and C. Maithri, "A comprehensive exploration of AIbased approaches and various machine learning techniques for detecting lung cancer," in Smart Trends in Computing and Communications (SmartCom 2025), Lecture Notes in Networks and Systems, vol. 1464, Singapore: Springer, 2025, pp. 45–57.
23. S. Sharma, D. Joshi, D. Bhatia, U. Bharadwaj, and Rani, "Investigating deep learning models for lung cancer prediction," in Innovative Computing and Communications (ICICC 2025), Lecture Notes in Networks and Systems, vol. 1434, Singapore: Springer, 2025, pp. 101–111.
24. Md. M. Islam, R. I. Sony, A. S. Joy, M. E. Humayun, Md. A. Yousuf, and Md. S. Azam, "Automated lung cancer detection from histopathological image using deep neural networks," in Emerging Technologies in Computing (iCETiC 2024), Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 623, Cham, Switzerland: Springer, 2025, pp. 32–44.
25. Y. Xie et al., "Early lung cancer diagnostic biomarker discovery by machine learning methods," Translational Oncology, vol. 14, no. 1, Art. no. 100907, 2021.