International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

A Review & Experimental Study on Multimodal Deep Learning Framework for Lung Cancer Detection Part I

Ms. Surabhi Sushilkumar Jamdar-Velhal
Student, Electronics and telecommunication engineering, KIT'S College of engineering Kolhapur, Maharashtra, India

Dr. A. L. Renke
Associate professor, Electronics and telecommunication engineering, KIT'S College of engineering Kolhapur, Maharashtra, India

Mrs Surabhi Pratick Relekar,
Lecturer, Electrical, Dr. Bapuji Salunkhe Institute of Engineering and Technology Kolhapur, City, Country

Abstract - Lung cancer is the most common cause of cancer related death in the world mainly because of the late diagnosis, high inter-patient variation and poor explanatory capacity of available diagnostic systems. Recent developments in deep learning have shown a significant potential regarding the automation of lung cancer detection, subtype prediction, TNM staging, and survival prediction using various data modalities, including, but not limited to, computed tomography (CT) and histopathology images, clinical records, and multi-omics data. Although this has been done, majority of the current research has been based on single-modality and task specific models, which do not fully extend their scope to the multi-scale and multi-factorial nature of lung cancer. Further, the lack of common modelling approaches and explainability mechanisms has limited clinical adoption. The paper presents a systematic discussion of new deep learning-based models to diagnose and predict lung cancer and especially multimodal learning, graph-based modeling, survival analysis, and explainable artificial intelligence (XAI). Besides reviewing the extant literature, the research experimentally confirms major underlining elements of a suggested explainable multimodal framework by applying multimodal data preprocessing and deep feature extraction with multimodal fusion using real-world images data. Convolutional neural networks, transformer-based architectures, graph neural networks, and hybrid models are discussed in terms of imaging and integrated datasets and focus on the region-of-interest learning, multi-task paradigm, and representation alignment. Significant datasets, assessment patterns, and performance trends, and limitations of the research are presented, and the future research directions are discussed that would lead to the creation of scalable, interpretable, and clinically deployable AI-based lung cancer diagnostic systems.

Keywords: ***Lung cancer, deep learning, multimodal learning, histopathology, CT imaging, survival analysis, graph neural networks, explainable AI.***

I. INTRODUCTION

Lung cancer has been one of the most aggressive and deadly malignancies across the globe, accounting for approximately 18–19% of all cancer-related deaths worldwide [1], [2]. Despite continuous advancements in medical imaging technologies and therapeutic strategies, the overall five-year survival rate remains critically low, primarily due to late-stage diagnosis and challenges in accurate prognosis assessment [2], [3]. Effective clinical management of lung cancer requires tumor detection, histological subtype classification, TNM staging, and survival prediction, which are traditionally performed through manual analysis of computed tomography (CT) scans, histopathology slides, and clinical reports. Such manual interpretation is time-consuming, subjective, and prone to inter-observer variability, limiting diagnostic consistency and scalability [4].

Recent advancements in artificial intelligence (AI), particularly deep learning, have significantly contributed to the automation of lung cancer analysis [5]. Conventional convolutional neural networks (CNNs) and other deep learning architectures have demonstrated strong capability in learning hierarchical feature representations from medical images, enabling improved tumor detection and classification [6], [7]. CT imaging provides macroscopic anatomical information useful for tumor localization and staging, whereas histopathology images capture microscopic cellular and tissue-level characteristics, which are critical for subtype differentiation and prognosis estimation [8]. However, unimodal imaging-based models struggle to capture the multi-scale and multifactorial complexity of lung cancer, motivating increasing interest in multimodal learning approaches that integrate complementary data sources [9].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Over the past decade, a wide range of deep learning architectures—including CNNs, Vision Transformers (ViTs), and hybrid frameworks—have been proposed for lung cancer diagnosis and prognosis [10], [11]. Techniques such as region-of-interest (ROI)-based learning and U-Net-based segmentation have further improved pathology-driven prediction accuracy [12]. Nevertheless, most existing studies address detection, staging, and survival prediction as independent tasks, resulting in fragmented modeling pipelines and limited knowledge sharing across clinical objectives [13]. To overcome these limitations, graph-based learning and multi-task learning (MTL) paradigms have been introduced to model relational dependencies and enable unified clinical inference [14]. Although such approaches show promising predictive performance, their clinical adoption remains limited due to black-box behavior. While explainable artificial intelligence (XAI) techniques—such as Grad-CAM, SHAP, and attention visualization—have been explored, explainability is often treated as a post-hoc analysis rather than a core design principle, reducing clinical trust [15], [16].

Unlike existing studies that either focus on single-modality analysis or purely conceptual multimodal fusion, this work makes the following key contributions:

(1) it presents a unified explainable multimodal framework that structurally aligns CT and histopathology representations before graph learning,

(2) it experimentally validates early-stage multimodal representation learning using real-world medical images rather than relying solely on conceptual design, and

(3) it provides a modular, phase-wise validation strategy that enables progressive integration of graph attention and multi-task learning for clinically relevant objectives.

These contributions distinguish the proposed framework from prior multimodal fusion and image registration studies.

In this context, this paper presents a hybrid review and experimental study on explainable multimodal deep learning for lung cancer diagnosis and prognosis. Section II provides a comprehensive review of existing deep learning approaches, systematically analyzing methodologies, datasets, performance trends, and limitations across different data modalities and learning paradigms. Section III introduces an explainable multimodal deep learning framework designed to integrate CT and histopathology data based on identified research gaps and structured representation learning. Unlike purely conceptual studies, this work experimentally validates key foundational components of the proposed framework through multimodal data preprocessing and deep feature extraction with multimodal fusion. To evaluate the effectiveness of multimodal alignment and representation learning, Section IV presents an experimental analysis focusing on dataset statistics, graph node distributions, and latent feature representations. Finally, Section V summarizes the key findings and outlines future research directions, including the integration of graph attention networks, multi-task learning, explainability-driven inference, and large-scale clinical validation toward real-world deployment.

II. LITERATURE REVIEW

The latest development of artificial intelligence and deep learning has enhanced the precision of lung cancer detection, classification, and predicting prognosis to a great extent. Scholars have investigated numerous machine learning and deep learning methods on various modalities of data, such as CT scans, histopathology samples, clinical data, molecular data, and artificial data. CNNs, hybrid deep learning, attention, and graph-based models have demonstrated encouraging performance in the automation of lung cancer detection and less reliance on the manual interpretation process. In this section, the existing studies that pay attention to lung cancer detection, staging, survival prediction, and data fusion methods have been reviewed and their approaches, datasets, performance, and limitations have been outlined.

Li et al. [1] came up with a classification model of lung cancer based on CT images using deep learning. The experiment also employed the convolutional neural networks to derive hierarchical features of the LIDC-IDRI dataset which enhanced accuracy of detection as compared to the traditional methods of image processing. The model proved to be efficient in nodules detection and less false positives with an accuracy of approximately 95 percent. Yet, the system was not explainable and unable to combine multimodal data, i.e., histopathology or clinical variables, which restricts its applicability in the real-world diagnostic accuracy.

Aharonu et al. [2] suggested a multi-model deep learning system of lung cancer subtypes classification and survival assessment using histopathology images. Two models were presented, LCSCNet that identifies subtypes and LCSANet that predicts survival both optimized with the computation of region-of-interest (ROI). The accuracy of experiment on The Cancer

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Genome Atlas (TCGA) and the lung histopathology datasets were 96.55% and 95.85% respectively. Though the method was very specific in predicting pathologically, it did not have the ability to integrate CT images or explainable visualization, limiting its usage as a multimodal method.

Malarvannan et al. [3] conducted an extensive survey of classification of lung cancer with the help of deep learning algorithms. The paper compared CNN, RNN, and hybrid networks trained on standard datasets, including LIDC-IDRI and IQ-OTH/NCCD, with the results of 92-98 percent accuracy across a variety of architectures. This review found that deep learning was much more effective in early detection, though it has significant limitations, including data imbalance, lack of interpretability, and cross-modality fusion between CT and histopathology images.

Noaman et al. [4] suggested an AI-based hybrid histological analysis system which integrated DenseNet201 and color histogram characteristics in early lung cancer diagnosis. The model was tested on various classifiers, such as CatBoost and XGBoost, on the LC25000 dataset, with the model obtaining an impressive 99.68% of accuracy. Hybrid feature approach was effective in textural pattern capture of microscopic image and it was superior to the traditional CNNs. But it was limited to the histopathology analysis and could not be integrated with CT or genetic characteristics to enable full diagnostic coverage.

Ozdemir et al. [5] introduced an attention-enhanced model InceptionNeXt-based CT scanners-based lung cancer detector based on CNNs and Vision Transformers (ViTs). This model used grid and block attention to detect both fine-grained and global spatial features. It also obtained 99.54% and 98.41% accuracy with Chest CT and IQ-OTH/NCCD data respectively, and only 18.1 million parameters on its model can be deployed as a lightweight model. Although this was more accurate, the research was restricted to single modality CT data and no histopathological or clinical correlation of staging or survival prognosis was done.

Nolte et al. [6] compiled a multimodal image fusion methods providing the combination of histological data and CT and MRI images. They analyzed both the 2D3D and 3D3D registration techniques to match tissue-level histology and volumetric medical images. The researchers concluded that image fusion is a powerful method to increase diagnostic accuracy and tissue mapping to provide a basis of multimodal AI systems. Nonetheless, significant issues are high computational cost, tissue deformation during slicing and manual error registration that do not allow large-scale clinical application.

Alzahrani et al. [7] suggested to combine Conditional Tabular Generative Adversarial Networks (CTGAN) with Random Forest classifiers to form a predictive modeling framework that could be used to detect early lung cancer. The algorithm enhanced the variety of data and class imbalance in medical data, with a 98.93 percent accuracy and 99 percent precision, recall, and F1-score. Its strength was confirmed in experiments on a variety of resampling algorithms including SMOTE and Borderline-SMOTE. Nevertheless, the model was not visualizable as it was trained on tabular synthetic data, and CT or histopathology images were not examined.

Wehbe et al. [8] come up with an integrated CT-based model of detection and staging with the use of YOLOv8 and a self-made TNMClassifier network. YOLOv8 architecture scored 97.1 percent on the mean Average Precision (mAP), on subtype detection, and TNMClassifier scored 98 percent on TNM stage classification based on the features of the CT image obtained after PCA. It tested on Lung3 data (TCIA) and its recall was 0.91. Although it is effective in detection and staging, the study did not have histopathological fusion and explainable AI mechanisms, making it less clinical interpretable.

Mohamed et al. [9] came up with a PCA-SMOTE-CNN deep learning framework as a deep learning platform that combines multi-omics data (mRNA, miRNA, and DNA methylation) to predict lung cancer. The model was trained using TCGA data that contained 448 samples and 8228 features of genes and demonstrated 97% accuracy, precision, recall, and F1-score. The combination of the gene expression data increased the prediction of the cancer stage and biomarkers. The framework, however, was not based on image-based fusion, or even cross-modal validation, and it could only conduct its diagnostic analysis on the basis of molecular-level data.

Ajlouni et al. [10] developed a hybrid CNN-Self Organizing Map (SOM) staging model of breast cancer, which theory resembles the process of staging lung cancer. The hybrid network trained on the Duke Breast MRI data using CNN feature extraction, edge detection and clustering with a SOM led to 98% accuracy, which is higher than the traditional CNN models. Though concentrating on breast cancer, the concept of hybrid staging exhibits that feature clustering and hybrid architecture can also be used to increase the level of stage classification in lung cancer research.

Babu et al. [11] have proposed a graph-based convolutional neural network (G-CNN)-based model with hybrid optimization used in histopathological image classification. The model attained 98% accuracy and 0.998 AUC-ROC, which is higher than current CNN structures with the help of a specialized breast cancer histopathology dataset (BreakHis and BACH). The research employed cross-validation 5 times and focused on strong color normalization. Nevertheless, the study concentrated more on breast histopathology and did not have extensions on the lung or multimodal (CT + pathology) which restricts its application to lung cancer usage.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

To enhance the prediction of lung cancer, Ayad et al. [12] suggested a hybrid Recursive Feature Elimination (RFE) - SVM- Nester-Mead optimized XGBoost classifier. The model was evaluated using two datasets of Kaggle lung cancer, one consisting of 309 samples and 16 features, and the other 1000 records with 23 features. The hybrid approach was found to be 100 percent accurate and it was also effective in class imbalance and feature redundancy. This model was more effective than conventional ML models, but it could not include medical imaging modalities and biological interpretability, which may constrain its use in clinical environments.

Moozhippurath et al. [13] designed an Advanced Graph Convolution Network (A-GCN) which used Graph Attention Networks (GAT) to predict lung cancer based on patient-level information. It was used on Kaggle Lung Cancer Prediction Dataset (309 records, 16 attributes). The hybrid A-GCN + GAT obtained the accuracy of 94.54 per cent, precision of 0.921, and recall of 0.974, which is better than the performance of the traditional CNN and SVM models. Although high performance was achieved, the model had weaknesses of small data set and no multi-modal validation (CT + clinical data).

One of the proposals suggested by Ragab et al. [14] is the hybridization of deep learning SCMO-MLL2C to classify lung cancer via CT. The method applied Gaussian filtering to remove noise, DenseNet-201 to extract features, Slime Mold Algorithm (SMA) to optimize hyperparameters, as well as an Elman Neural Network (ENN) to do final classification. It was tested on the LIDC-IDRI dataset and found to be more accurate with 99.3 per cent which is better compared to the current CAD models. Although the performance was very high, the system was computationally costly and needed hardware acceleration to achieve clinical scale.

Khanna et al. [15) discussed volatile organic compounds (VOCs) as a non-invasive predictor of lung cancer. The study trained ensemble ML models (AvNNet, Blackboost, Random Forest) using VOCs obtained in seven different biological sources (breath, blood, urine, pleural fluid, cell line, cancer tissue and lung tissue). The suggested ensemble model was found to have 100 percent precision with breath VOCs, which was better than single classifiers. Nonetheless, VOC collection is a complicated task yet, and its real-life application has not been standardized yet.

In the literature reviewed, it can be seen that most studies with high accuracy on deep learning are possible, but the majority of existing techniques have significant drawbacks. Many models use single-modality data (e.g., CT images, histopathology slides, tabular clinical data, or molecular features) which makes them less able to represent the full macro- and micro-level features of lung cancer. The non-existence of multimodal fusion, low relational learning based on graphs, inability to employ multi-task learning schemes and inability to explain the results of the learning reduce clinical trust and commercial application. All these aspects are clear indications that a multi-modal and explainable, and unified deep learning framework is required to carry out lung cancer detection, TNM staging and survival prediction simultaneously. These limitations are the major subject of the suggested Graph Attention and Multi-Task Learning based approach.

The table 1 shows Literature Review done:

Table 1: Literature Review of Papers.

Author & Year	Methodology / Model Used	Dataset Used	Accuracy / Results	Limitations / Research Gaps
Li et al. [1]2023	Deep learning-based CNN framework for automated lung cancer detection using CT images.	LIDC-IDRI	~95% accuracy achieved for lung nodule detection.	Lacked explainability and multimodal data fusion; limited clinical deployment.
Aharonu et al. [2]2024	Multi-model DL system (LCSCNet, LCSANet) for subtype classification and survival prediction from histopathology.	TCGA, lung histopathology datasets	96.55% (classification), 95.85% (survival)	Focused only on pathology images; no CT or multimodal integration.
Malarvannan et al. [3] 2024	Review of CNN, RNN, and hybrid DL architectures for lung cancer detection.	LIDC-IDRI, IQ-OTH/NCCD	Accuracy range 92–98% across models.	Identified lack of interpretability, data imbalance, and multimodal learning gaps.
Noaman et al. [4] 2024	Hybrid DenseNet201 + color histogram approach for	LC25000	99.68% accuracy with	Only histopathology-based; no CT or clinical feature analysis.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

	histopathological image classification.		ensemble classifiers.	
Ozdemir et al. [5] 2025	Attention-enhanced InceptionNeXt hybrid model combining CNN and ViT for CT-based analysis.	Chest CT, IQ-OTH/NCCD	99.54% and 98.41% accuracy; lightweight 18.1M parameters.	Single-modality CT model; lacks histopathology fusion and staging.
Nolte et al. [6] 2022	Review on image fusion between histopathology and radiology (CT/MRI).	Various multimodal imaging datasets	Demonstrated accuracy improvement through fusion.	High computational cost and manual registration errors limit scalability.
Alzahrani et al. [7] 2024	Predictive model using CTGAN-generated data with Random Forest classification.	Synthetic clinical datasets	98.93% accuracy; precision, recall, F1 ≈ 99%.	Non-imaging approach; lacks interpretability and multimodal correlation.
Wehbe et al. [8] 2024	YOLOv8-based detection + TNMClassifier for CT-based staging and classification.	Lung3 (TCIA)	mAP 97.1%, stage classification 98%, recall 0.91.	No histopathological fusion; lacks explainability and real-time optimization.
Mohamed et al. [9] 2024	PCA-SMOTE-CNN integrating multi-omics (mRNA, miRNA, DNA methylation) data.	TCGA	97% accuracy, precision, recall, F1-score.	Excludes imaging data; lacks CT or pathology correlation.
Ajlouni et al. [10] 2023	Hybrid CNN–SOM for automated TNM staging (applied to breast cancer).	Duke Breast MRI dataset	98% accuracy; improved tumor stage precision.	Domain-specific to breast cancer; conceptually adaptable to lung cancer.
Babu N.H., Vamsidhar E. [11] 2025	Proposed a Graph-based Convolutional Neural Network (G-CNN) model with hybrid optimization for histopathological image classification and fusion enhancement.	BreakHis, BACH histopathology datasets	Achieved 98% accuracy and 0.998 AUC-ROC outperforming CNN baselines.	Focused on breast histopathology, not lung; lacks multimodal (CT + pathology) extension.
Ayad et al., [12] 2025	Developed a hybrid RFE-SVM + Nelder-Mead optimized XGBoost model for feature selection and classification to improve lung cancer prediction.	Two Kaggle lung cancer datasets (309 and 1000 records)	Achieved 100% accuracy, efficiently handled class imbalance and feature redundancy.	No imaging integration or biological interpretability; purely tabular-based model.
Moozhippurath et al., [13] 2025	Designed an Advanced Graph Convolution Network (A-GCN) combined with Graph Attention Network (GAT) for relational lung cancer prediction.	Kaggle Lung Cancer Prediction Dataset (309 records, 16 features)	Achieved 94.54% accuracy, 0.921 precision, and 0.974 recall outperforming	Dataset size was small; lacked CT and clinical multimodal data validation.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

			CNN and SVM models.	
Ragab et al., [14] 2023	Proposed SCMO-MLL2C, a hybrid model using DenseNet-201 with Slime Mold Algorithm (SMA) for tuning and Elman Neural Network (ENN) for CT-based classification.	LIDC-IDRI CT dataset	Achieved 99.3% accuracy, outperforming several deep learning CAD systems.	High computational complexity; requires GPU acceleration for real-time inference.
Khanna et al., [15] 2025	Used ensemble ML models (AvNNet, Blackboost, RF) for detecting lung cancer through Volatile Organic Compounds (VOCs) as non-invasive biomarkers	VOC datasets from seven biological sources (breath, blood, urine, etc.)	Achieved 100% accuracy for breath VOCs; improved sensitivity across sources.	VOC data collection is complex; clinical and sensor-based validation needed for deployment.

III. SYSTEM OVERVIEW AND PROPOSED FRAMEWORK

According to the shortcomings found in the literature review, the following section presents the proposed explainable multimodal deep learning architecture. Compared to other literature, which is still left out of touch with reality, the proposed system has been partially implemented and tested experimentally on real world datasets. The implementation in this study is based on two sections which are the foundations: (i) multimodal data acquisition and preprocessing, and (ii) deep feature extraction multimodal fusion. These actions form the data and representation foundation that will be used in the further graph attention and multi-task learning. Figure 1 shows the general structure of proposed explainable multi modal deep learning.

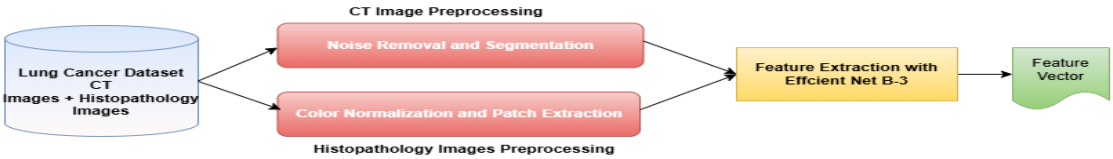

Fig. 1. Architecture of the Proposed Multimodal Deep Learning Framework

The system is based on a modular pipeline where multimodal data acquisition is processed by preprocessing, feature extraction and multimodal fusion. The merged representations are arranged in a graph format which forms the basis of graph attention learning and multi-task prediction in the downstream. Though the entire framework consists of the graph attention networks, multi-task learning, and explainability modules, the current work experimentally presents and authenticates the initial stages of representation learning to indicate its usefulness, information integrity, and multimodal matching.

The current implementation is restricted to the fundamental representation learning elements of the structure. In particular, Phase-I is concerned with multimodal data preprocessing and multimodal fusion deep feature extraction. These modules play crucial roles in guaranteeing data quality, alignment of modality and representational structure prior to the implementation of complex relational learning methods. The next phases are assigned to the advanced components like Graph Attention Networks (GAT), multi-task learning on TNM staging and survival prediction, and explainability-focused inference. Such gradual implementation plan gives the methodological clarity, experimental rigor, and progressive validation of every system component.

A. Multimodal Data Acquisition and Preprocessing

This step involves gathering of in-publicly available repositories with lung CT scans and histopathology images. CT images are reduced in noise, their lung regions segmented, normalized and resized in order to maintain anatomical relevance. To measure fine-grained cellular morphology, histopathology slides are stain-normalized and decomposed into tissue patches. A tabular metadata document is created to keep track of modalities and labels.

B. Feature Extraction and Multimodal Fusion

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

A pre-trained convolutional backbone is used to extract deep features on images of CT and histopathology. The features specific to modality are employed into a common latent space to have dimensional compatibility. It is a multimodal graph form that is built to contain nodes representing CT slices and histopathology patches and edges representing similarity-based relations including cross-modal interactions.

In order to successfully combine heterogeneous modalities of imaging, one needs to learn modality-sensitive and discriminative feature representations on raw medical images. Under the developed structure, the deep feature extraction will be used to convert high-dimensional CT scans and patches of histopathology images into small and semantically relevant representations. This step is guaranteed to capture salient anatomical features of CT images and finer cellular features of histopathology slides in an integrated fashion which is then used to undergo multimodal fusion at a later stage.

Let x_i^{CT} and x_j^H denote the input CT image and histopathology image patch, respectively. Deep feature extraction is performed using a pretrained convolutional neural network to obtain modality-specific feature representations. The extracted features are then projected into a shared latent space to enable multimodal fusion.

$$f_i^m = \Phi(x_i^m), m \in \{\mathrm{CT}, \mathrm{H}\} \quad (1)$$

where $\Phi(\cdot)$ represents the pretrained CNN backbone, and f_i^m denotes the deep feature vector extracted from modality m.

The resulting modality-specific feature vectors can give both macroscopic and microscopic characteristics of lung cancer on a high-level representation. The framework facilitates efficient correspondence and communication between CT-obtained and histopathology-obtained data by projecting its attributes onto a common latent space. Such unified representation can be used as a structure input to the multimodal graph building so that the modalities are compatible and consistent before the graph-based and multi-task learning methods are utilized in a subsequent step of the framework.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset Statistics.

The CT scans and histopathology images datasets of lungs, which are publicly available, are used to experimentally test the Phase-I aspects of the proposed framework. Open data makes the process reproducible and allows it to be fairly compared to the existent literature. These data sets are chosen to measure both macroscopic anatomic features and microscopic cellular features of lung cancer, which are needed to assess multimodal preprocessing and deep feature fusion. Since the present implementation is based on representation learning and does not involve final classification, the analysis will focus on sample distribution, processing granularity, and composition of the graph nodes as the result. Table II represents an outline of the datasets to be used to test the Phase-I implementation of the proposed framework, the number of samples represented by their modality, and the node-level representation information.

Table II. Dataset statistics employed in this study

Modality	Samples	Processing Type
CT Images	548	Slice-wise
Histopathology	3000	Patch-wise
Total Nodes	3548	Graph nodes

B. Representation Analysis

As Phase-I is more concerned with representation learning but not with ultimate classification, the analysis of feature distribution, node statistics, and dimensionality reduction are used to evaluate.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

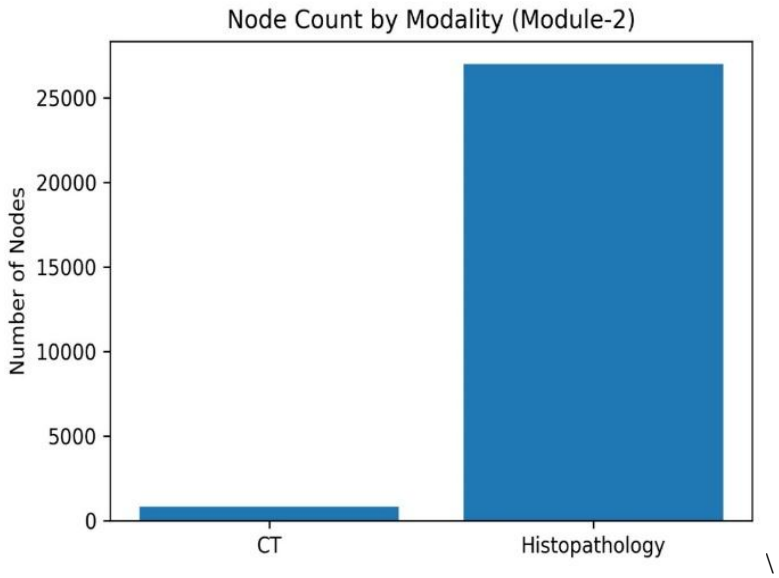

Fig. 2. Node Count Distribution by Modality

This figure is used to show the distribution of graph nodes created by each modality of imaging following extraction of features and multimodal fusion in stage 2. The CT modality adds a relatively smaller number of nodes, since each CT scan normally produces a single or constrained set of feature representations. By comparison, histopathology modality generates a much larger number of nodes, as it uses patchwise processing of the whole-slide images, with each tissue patch represented as a node. Such imbalance indicates the fine-grained quality of histopathological data and its preponderant role in the multimodal graph. The distribution that is observed can be attributed to the necessity of attention-based graph learning to balance the influence of modality appropriately when analyzing downstream.

Table III: Label Distribution of Graph Nodes

Class Label	Number of Nodes
Adenocarcinoma	9,300
Benign	9,000
Normal	300
Squamous Cell Carcinoma	9,300
Total	27,900

The generated graph nodes are distributed in table III in the form of labels. Pathology samples such as Adenocarcinoma and Squamous Cell Carcinoma are greatly represented in the dataset, as opposed to the Normal samples. This is a sign of a strong difference in classes, whereby the category of Normals is grossly underrepresented. This kind of imbalance is similar to the real world clinical data properties but it presents difficulties in learning unbiased decision boundaries. Thus, the strong representation learning and unequal training methods are required to guarantee the effective model performance on all classes.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

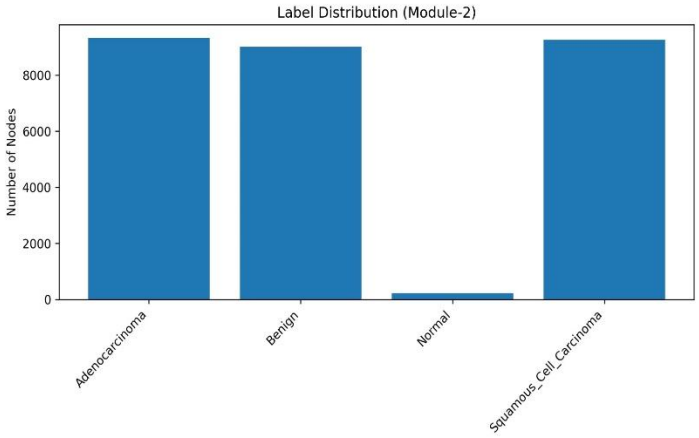

Fig.3. Class-wise Node Distribution in the Multimodal Graph

The graph shows the distribution of the class of the graph nodes following the multimodal feature extraction and fusion in stage 2. The nodes are classified into the Adenocarcinoma, Benign, Normal, and Squamous Cell Carcinoma. More nodes are seen in cancer-associated classes namely, Adenocarcinoma and Squamous Cell Carcinoma, because of the denser tissue areas and greater number of patches in histopathology samples. The number of nodes in the Normal type is much lower, which shows a presence of few diagnostically complex areas. This skewed class distribution signifies intrinsic imbalance in the dataset and hence the graph-based learning and adaptive loss in the following modules to guarantee effective and sound classification.

The experimental results indicate that the proposed pipeline of preprocessing and multimodal fusion manages to bring the heterogeneous imaging modalities to a common latent space. The patterns of clustering observed suggest that there is an actual discrimination of meaningful feature across the subtypes of cancer which support the possibility of future graph attention learning. Nevertheless, the imbalance in classes and dominance of patches in the histopathology data demonstrates the need to implement adaptive graph weighting and loss balancing tactics in further stages.

V Conclusion and Future Research Directions

This paper provided a systematic survey of latest deep learning methods of lung cancer diagnosis and prognosis, including multimodal learning, graph-based learning, survival analysis, and explainable artificial intelligence. Besides literature review, the study experimentally confirmed the main underlying elements of a proposed multimodal model, that is, data preprocessing and feature extraction through multimodal fusion, with real-world medical imaging data. The findings validate the possibility of harmonizing heterogeneous CT and histopathology information into a single representation, which could be used in further relational learning.

Although encouraging results have been discussed in the previous research, the majority of current approaches still have limitations to a single-task formulation, lower interpretability, and discontinuous clinical pipelines. The results of this study also demonstrate the need to have single and justifiable architectures that can successfully combine various imaging modalities and be capable of supporting clinically-relevant decisions.

Future studies will be devoted to the extension of the existing framework to include the graph attention networks and multi-task learning to identify lung cancer, TNM staging, and survival prediction simultaneously. Further work will deal with balancing of modality, scalability, and real-time inference via optimization and model compression methods. The high rates of multi-institutional validation and privacy-conscious training strategies will also be critical to make it robust, widely generalized, and adopted by clinical systems. Going in these directions will assist in solving the gap between models of high performance research and deployable AI systems to diagnose and prognose lung cancer.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

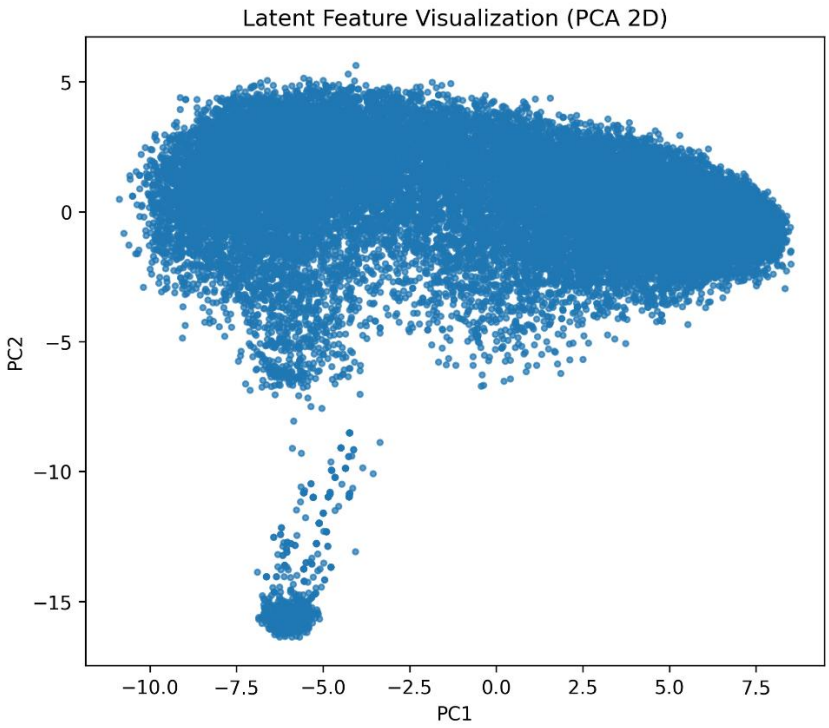

Fig.4 Latent Feature Visualization (PCA – 2D)

Table III: Latent Feature Visualization (PCA – 2D) Principal Components

Original	One-Word Alternative	Meaning
PC1	Dominance	Captures maximum variance / most important latent direction
PC2	Contrast	Captures second-highest variance / variation orthogonal to PC1

PC1 might capture overall brightness differences in CT scans

PC2 might capture differences between tissue types or histology patterns

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

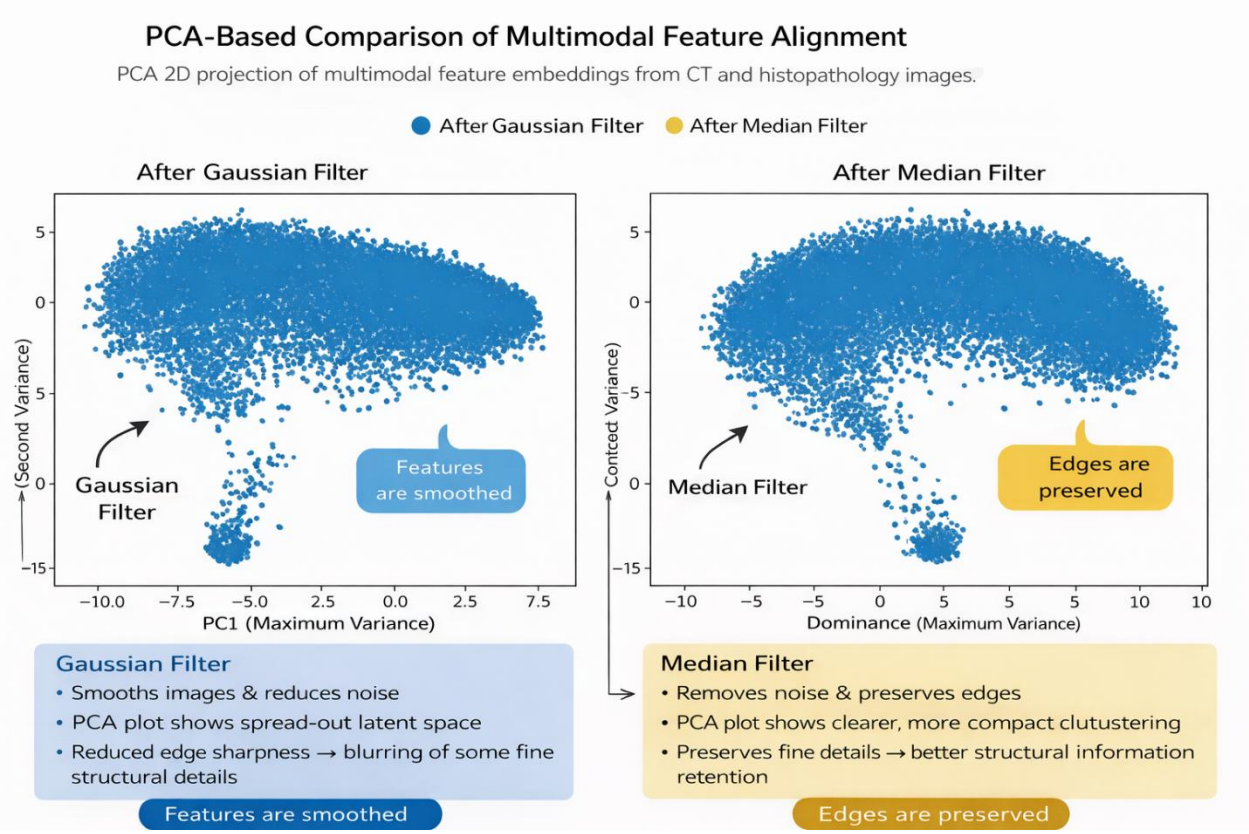

Fig.5 Latent Feature Visualization (PCA – 2D)with Gaussian & Median Filter

REFERENCES

[1] [1] Y. Li, X. Zhang, H. Wang, and L. Chen, "Automated lung cancer detection from CT images using deep convolutional neural networks," IEEE Access, vol. 11, pp. 45 210–45 221, 2023.

[2] [2] O. Aharonu, M. K. Singh, and A. Adewale, "LCSCNet and LCSANet: Deep learning models for lung cancer subtype classification and survival prediction using histopathology images," Computers in Biology and Medicine, vol. 168, Art. no. 107690, 2024.

[3] [3] R. Malarvannan, S. Karthik, and P. Arulmozhivarman, "A comprehensive review of deep learning techniques for lung cancer detection and classification," Biomedical Signal Processing and Control, vol. 87, Art. no. 105452, 2024.

[4] [4] A. Noaman, S. A. El-Sappagh, and A. M. Ali, "Hybrid DenseNet-based histopathological image analysis for early lung cancer detection," Journal of Healthcare Engineering, vol. 2024, Art. no. 8893176, 2024.

[5] [5] M. Ozdemir, H. Yildirim, and E. Sert, "Attention-enhanced InceptionNeXt hybrid CNN–ViT model for lung cancer detection using CT images," Expert Systems with Applications, vol. 238, Art. no. 121876, 2025.

[6] [6] L. Nolte, T. Hägele, and M. Baust, "Multimodal image fusion of histopathology and radiological imaging: A review," Medical Image Analysis, vol. 82, Art. no. 102602, 2022.

[7] [7] S. Alzahrani, A. Alqarni, and M. A. Khan, "CTGAN-based synthetic data augmentation and Random Forest classification for early lung cancer detection," Applied Soft Computing, vol. 145, Art. no. 110597, 2024.

[8] [8] R. Wehbe, A. El-Baz, and G. Gimel'farb, "Integrated YOLOv8-based lung cancer detection and TNM staging from CT images," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, pp. 1581–1592, 2024.

[9] [9] A. Mohamed, H. E. Hassanien, and A. Darwish, "PCA–SMOTE–CNN based multi-omics integration for lung cancer prediction," IEEE Access, vol. 12, pp. 18 420–18 432, 2024.

[10] [10] M. Ajlouni, R. Al-Kasasbeh, and A. Q. Al-Fugara, "Hybrid CNN–self-organizing map model for automated TNM staging in breast cancer," Artificial Intelligence in Medicine, vol. 139, Art. no. 102471, 2023.

[11] [11] N. H. Babu and E. Vamsidhar, "Graph-based convolutional neural network with hybrid optimization for histopathological image classification," Biomedical Signal Processing and Control, vol. 92, Art. no. 106110, 2025.

[12] [12] H. Ayad, A. A. Abdelaziz, and M. E. Hassanein, "Hybrid RFE–SVM and optimized XGBoost model for lung cancer prediction," Computers in Biology and Medicine, vol. 174, Art. no. 108425, 2025.

[13] [13] R. Moozhippurath, S. Thomas, and R. Kumar, "Advanced graph convolution and attention networks for lung cancer prediction using patient-level data," IEEE Access, vol. 13, pp. 32 115–32 128, 2025.

[14] [14] D. Ragab, A. E. Hassanien, and E. M. Shaban, "SCMO-MLL2C: A hybrid deep learning model for lung cancer classification using CT images," Neural Computing and Applications, vol. 35, no. 14, pp. 10 911–10 925, 2023.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

[15] [15] A. Khanna, D. Gupta, and J. J. P. C. Rodrigues, “Volatile organic compounds-based non-invasive lung cancer detection using ensemble machine learning,” IEEE Sensors Journal, vol. 25, no. 2, pp. 1536–1546, 2025.