EAMDLF: Multimodal Deep Learning Framework for Lung Cancer

Sahil Yadav[1], Dr.Priyanka Makkar[2]
[1]*M. Tech (Data Science) Amity University Haryana, Gurugram, Haryana, India*
[2]*Associate Professor), Amity University Haryana, Gurugram, Haryana, India*

***Abstract*—Lung cancer continues to be one of the most serious health concerns worldwide, with approximately 2.2 million new cases and nearly 1.8 million deaths reported annually. Early detection plays a crucial role in improving the effectiveness of treatment and increasing patient survival rates. In recent years, advances in artificial intelligence (AI) and multimodal deep learning have provided new possibilities for developing automated and reliable diagnostic systems. Motivated by these advancements, this study proposes an Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) for lung cancer diagnosis. The proposed framework combines several deep learning components, including Convolutional Neural Networks (CNNs), DenseNet-201 transfer learning, attention mechanisms, and multimodal feature fusion techniques, to improve diagnostic performance. The model utilizes different types of medical information collected from publicly available datasets, such as LIDC-IDRI, TCIA, LC25000, and ChestX-ray14, integrating CT scans, chest X-ray images, histopathological images, and clinical information for a more comprehensive diagnostic process. One of the common challenges in medical data analysis is the presence of data imbalances and variations among different data modalities. To address these issues, preprocessing methods such as image normalization, data augmentation, feature scaling, and synthetic oversampling techniques were applied. Feature extraction is performed using CNN and DenseNet architectures, whereas an attention-based fusion mechanism combines useful information from multiple sources. In addition, Explainable Artificial Intelligence (XAI) techniques, particularly Grad-CAM, were incorporated to improve model transparency and support clinical interpretation. The experimental findings indicate that the proposed framework achieved an accuracy of 99.12%, precision of 99.35%, recall of 98.91%, and F1-score of 99.13%, outperforming conventional single-modality and standard deep learning approaches. The results demonstrated fewer classification errors and improved diagnostic reliability. Overall, the proposed framework offers a practical and scalable solution that may assist healthcare professionals in improving lung cancer diagnosis and supporting personalized clinical decision making.**

***Index Terms*—Lung Cancer Diagnosis, Multimodal Deep Learning, Explainable Artificial Intelligence (XAI), Attention Mechanism, CNN, DenseNet-201, Feature Fusion, CT Imaging, Histopathology, Grad-CAM, Medical Imaging, Healthcare Analytics, Clinical Decision Support System, Artificial Intelligence.**

I. INTRODUCTION

Lung cancer represents one of the most serious public health concerns of the twenty-first century and continues to be one of the leading causes of cancer-related mortality worldwide. According to global cancer statistics, lung cancer accounts for approximately 2.2 million new cases and nearly 1.8 million deaths annually, making it one of the most frequently diagnosed and deadliest cancers. Diseases associated with lung malignancies, including non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), affect millions of individuals across different age groups and populations worldwide. The healthcare burden caused by lung cancer is equally substantial, with healthcare systems investing enormous resources in diagnosis, treatment, long-term care and disease management.

The rising incidence of Lung cancer is strongly associated with complex interactions between environmental, genetic, and lifestyle-related risk factors. Common modifiable risk factors include smoking, exposure to air pollution, occupational hazards, unhealthy lifestyle, and prolonged exposure to carcinogenic substances. Non-modifiable factors included age, family history, genetic predisposition, and inherited medical conditions. The simultaneous

presence of multiple risk factors significantly increases the probability of disease progression and adverse outcomes, emphasizing the importance of early diagnosis and timely medical intervention [1]– [3].

Traditionally, lung cancer diagnosis has relied on radiological tests, histopathological evaluations, and physicians' interpretation of medical images. Techniques involving CT scans, chest radiography, bronchoscopy, and biopsy procedures remain prevalent in clinical practice. These methods frequently necessitate expert interpretation and may struggle with subtle abnormality detection and complex disease patterns. Moreover, conventional diagnostic processes are prone to variability in interpretation and delayed decision-making, which underscores the need for intelligent and automated diagnostic solutions [4]– [6].

The rapid digitalization of healthcare infrastructure over the past two decades has resulted in the generation of massive volumes of structured and unstructured medical data through electronic health records, medical imaging systems, pathology databases, and hospital information systems. This increase in healthcare data availability has created significant opportunities for the application of artificial intelligence (AI) and deep learning (DL) techniques in clinical diagnostics. Deep learning methods have demonstrated substantial capability in identifying hidden and complex patterns in large medical datasets that may not be easily recognized through conventional analytical approaches [7]– [9].

Among various artificial intelligence techniques, multimodal deep learning has emerged as a promising approach for medical diagnosis because of its ability to integrate heterogeneous healthcare information. Multimodal systems combine data obtained from multiple sources, including CT images, chest X-rays, histopathological images, genomic information, biomarkers, and patient clinical records, to generate more comprehensive diagnostic representations. Recent studies have demonstrated that multimodal learning frameworks improve prediction accuracy, robustness, and disease characterization compared with traditional single-modality approaches [11], [17], [19].

A significant challenge in developing AI-based lung cancer diagnosis systems is the presence of data heterogeneity and class imbalance in healthcare datasets. Medical datasets often contain limited cancer-positive samples compared with normal samples, leading to biased model learning and reduced prediction reliability. This challenge becomes particularly critical in healthcare applications, where missing a cancer-positive patient may result in delayed diagnosis and severe clinical consequences. In addition, variations in imaging modalities and patient characteristics further complicate model generalization across different healthcare environments [9], [19].

This paper addresses these challenges by proposing an Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) that integrates multiple complementary architectures including Convolutional Neural Networks (CNNs), DenseNet-201 transfer learning, attention mechanisms, and multimodal feature fusion techniques. Data preprocessing methods such as image normalization, augmentation, feature scaling, and synthetic balancing strategies are applied to improve model performance. The proposed framework demonstrates superior performance compared with conventional single-modality methods and supports clinically interpretable diagnosis through explainable artificial intelligence mechanisms.

The principal contributions of this work are as follows: (i) development of a multimodal deep learning framework integrating CT scans, chest X-rays, histopathological images, and clinical information; (ii) implementation of attention-guided feature fusion for improved diagnostic accuracy; (iii) integration of explainable AI techniques to enhance model transparency; (iv) comprehensive performance evaluation using accuracy, precision, recall, F1-score, and confusion matrix analysis; (v) comparative benchmarking with existing approaches; and (vi) development of a scalable framework suitable for future clinical decision support applications.

The remainder of this paper is organized as follows: Section II presents the related literature review. Section III describes the theoretical background and mathematical foundations. Section IV explains the proposed methodology. Section V presents experimental evaluation and result analysis. Section VI discusses findings and clinical implications. Section VII outlines research challenges and future directions. Finally, Section VIII concludes the paper.

II. LITERATURE REVIEW

The application of multimodal deep learning to lung cancer diagnosis has become an active area of research in recent years, driven by advances in medical imaging, artificial intelligence, computational infrastructure, and availability of public healthcare datasets. This section presents a comprehensive review of 20 recent studies published between 2020 and 2026 that collectively define the current state of the art in AI-based multimodal lung cancer detection, classification, prognosis prediction, and clinical decision support.

A. Multimodal Learning and Survival Prediction

Wu et al. [1] proposed DeepMMSA, a multimodal deep learning framework for non-small cell lung cancer survival analysis by integrating CT images and clinical information. The study demonstrated that combining heterogeneous healthcare data significantly improved survival prediction performance compared with single-modality systems. The authors highlighted the effectiveness of multimodal feature learning in prognosis analysis and personalized healthcare applications.

Oncu et al. [2] developed a multimodal artificial intelligence framework integrating convolutional neural networks and artificial neural networks for lung cancer diagnosis. Their framework utilized CT scans and clinical patient information to improve diagnostic accuracy and automated feature extraction. The study showed that multimodal feature integration enhances predictive capability compared with traditional approaches.

Farooq et al. [3] proposed a multimodal representation learning framework for lung cancer survival prediction. The study integrated imaging and clinical information to improve predictive performance and demonstrated the importance of extracting complementary features from multiple healthcare modalities. The authors reported significant improvements in prognosis prediction using multimodal representation learning techniques.

B. Multimodal Fusion and Explainable AI Approaches

Hassan et al. [4] introduced a multimodal medical image fusion framework for non-small cell lung cancer classification. Their approach integrated information obtained from multiple imaging modalities to improve tumor classification accuracy and feature extraction performance. The study emphasized the importance of multimodal fusion for obtaining richer diagnostic information.

Kumar et al. [5] presented a review of multimodal deep learning frameworks for lung cancer detection. The study analyzed multimodal fusion strategies, hybrid architectures, and deep learning techniques while identifying challenges related to computational complexity and limited datasets. The review highlighted the growing importance of multimodal learning in healthcare applications.

Koudounaris et al. [6] conducted a critical review of explainable deep learning techniques for lung cancer diagnosis. The study analyzed various explainability approaches including Grad-CAM and visualization techniques used to improve transparency and clinical trust. The authors emphasized that interpretability remains a major requirement for practical healthcare deployment.

C. Biomarker Analysis and Histopathological Approaches

Cui et al. [7] developed a deep learning framework integrating imaging features and biomarker information for lung cancer survival prediction. Their study demonstrated that combining biomarker information with medical imaging improves prognosis prediction and supports personalized treatment planning.

Chen et al. [8] proposed a weakly supervised learning framework for lung cancer diagnosis using virtually stained histopathological tissue images. Their approach reduced dependency on manual annotations and improved classification efficiency in histopathological image analysis applications.

Hermoza et al. [9] developed a weakly supervised tumor localization framework using chest X-ray images for lung cancer screening and survival prediction. The study demonstrated that weak supervision methods can effectively identify clinically important tumor regions without extensive annotation requirements.

Civit-Masot et al. [10] proposed an explainable deep learning model using histopathological images for non-small cell lung cancer diagnosis. Their work incorporated Grad-CAM visualization techniques to improve interpretability and physician trust in AI-assisted diagnostic systems. The framework enabled visual identification of image regions contributing to

classification decisions, thereby improving transparency. Experimental results demonstrated strong classification performance while preserving interpretability requirements. The study emphasized that combining explainability with predictive accuracy is essential for practical deployment of AI-based healthcare technologies.

D. Deep Learning Architectures and Optimization Methods

Cui et al. [11] developed a deep learning framework for pulmonary nodule screening using CT images for early-stage lung cancer diagnosis. Their framework focused on identifying pulmonary nodules with high sensitivity and reducing the risk of missed detections during screening procedures. The study demonstrated that automated deep learning methods can support radiologists by detecting subtle abnormalities that may be difficult to identify manually. The proposed system showed promising results in improving diagnostic performance and facilitating early clinical intervention for lung cancer patients.

Horry et al. [12] proposed a CNN-based malignancy prediction framework using chest radiographs for lung cancer screening and diagnosis. Their model analyzed chest X-ray images to identify abnormal patterns associated with malignant tumors and cancer-related abnormalities. The study highlighted the effectiveness of convolutional neural networks in extracting important imaging features without extensive manual intervention. In addition, the authors emphasized the potential of chest radiograph-based systems as low-cost and accessible alternatives for large-scale screening applications.

Mohandass et al. [13] introduced an optimized attention-based convolutional neural network integrated with DenseNet-201 transfer learning for CT-based lung cancer classification. The proposed framework employed attention mechanisms to focus on clinically important regions within CT images, thereby improving feature extraction performance and classification accuracy. Transfer learning further enhanced the learning capability of the model by utilizing pre-trained knowledge from large image datasets. The study demonstrated that combining attention mechanisms with transfer learning significantly improves model performance for medical image analysis tasks.

Hammad et al. [14] proposed an automated lung cancer detection framework using genetic TPOT feature optimization and deep learning techniques. Their approach aimed to automate feature selection and optimize model architecture for improving diagnostic performance. The framework reduced dependence on manual parameter tuning and enabled the identification of the most relevant diagnostic features. Experimental findings showed improved classification accuracy and demonstrated the importance of optimization strategies in intelligent healthcare systems.

E. Hybrid Models and Future AI Systems

Bhandary et al. [15] developed a multimodal deep learning framework integrating chest X-ray and CT scan images for lung abnormality detection. Their study demonstrated that combining multiple imaging modalities improves feature representation and diagnostic robustness compared with single-modality systems. The framework effectively utilized complementary information obtained from different image sources to improve classification performance. The authors highlighted the importance of multimodal learning in achieving reliable lung cancer diagnosis systems.

Masud et al. [16] proposed a deep learning-based classification framework for diagnosing lung and colon cancer using medical imaging datasets. The system utilized automated image analysis techniques for classifying cancerous and non-cancerous tissue samples with minimal human intervention. The study demonstrated the effectiveness of deep learning algorithms in identifying complex image patterns associated with disease progression. Their findings highlighted the growing role of AI-based diagnostic systems in medical image analysis and healthcare applications.

Patel et al. [17] reviewed different deep learning paradigms including convolutional neural networks, transformer architectures, transfer learning techniques, and hybrid deep learning models for lung cancer diagnosis. Their study provided a comparative analysis of various architectures used for medical image classification and feature extraction tasks. The review highlighted recent advancements in artificial intelligence and emphasized the importance of advanced architectures for improving diagnostic performance. The authors also discussed future

opportunities and research directions in intelligent healthcare systems.

Kumar et al. [18] reviewed hybrid neural network techniques developed for lung cancer diagnosis and prognosis prediction. Their study emphasized the effectiveness of combining multiple neural architectures to capture complementary information from healthcare datasets. Hybrid systems demonstrated improved prediction capability by integrating feature extraction and classification components within a unified framework. The authors suggested that future research should focus on improving scalability and generalization of hybrid deep learning systems.

Tian et al. [19] conducted a comprehensive survey on deep learning techniques in multimodal medical imaging for cancer detection and diagnosis. The study analyzed various multimodal fusion strategies, representation learning methods, and deep learning architectures used for healthcare applications. The authors highlighted challenges associated with integrating heterogeneous medical information from different data modalities. Their findings emphasized the need for robust multimodal learning frameworks capable of handling large-scale healthcare data.

Zhong et al. [20] reviewed large multimodal AI models and vision-language systems for lung cancer diagnosis, screening, and treatment planning. Their study discussed the role of foundation models and large-scale artificial intelligence systems in supporting clinical decision-making processes. The authors highlighted the capability of multimodal AI systems to integrate imaging and textual information for more comprehensive diagnosis. The review also identified future opportunities for large AI models in intelligent healthcare and personalized treatment planning.

III. THEORETICAL BACKGROUND

A. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) are among the most widely used deep learning architectures for medical image analysis because of their ability to automatically learn hierarchical features from image data. CNN models perform feature extraction through convolution operations, activation functions, and pooling layers, enabling effective identification of important visual patterns in CT scans, chest X-ray images, and histopathological data. For an input image represented by feature map XXX, the convolution operation can be expressed as:

$$F(i,j) = (X * K)(i,j) + b$$

where XXX represents the input image, KKK denotes the convolution kernel, and bbb is the bias term. CNN architectures have demonstrated strong performance in lung cancer diagnosis because of their ability to identify complex spatial patterns and disease-related abnormalities from medical images [2], [12].

B. DenseNet-201 Transfer Learning

DenseNet-201 is a deep convolutional architecture that improves information flow through dense connectivity among network layers. Unlike conventional neural networks, DenseNet establishes direct connections between all preceding layers, allowing feature reuse and reducing vanishing gradient problems. The output of a DenseNet layer can be represented as:

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}])$$

where x_1 denotes the output of layer l and H_l represents the transformation function applied to preceding feature maps. Transfer learning with DenseNet-201 enables pre-trained knowledge obtained from large image datasets to improve learning performance on limited healthcare datasets. Such approaches have demonstrated significant improvements in lung cancer classification and feature extraction tasks [13].

C. Attention Mechanism

Attention mechanisms have become an important component in modern deep learning architectures because they enable models to focus selectively on the most informative regions within input data. In medical image analysis, not all image regions contribute equally toward diagnosis, and many areas may contain irrelevant background information. Attention modules help the model assign greater importance to clinically meaningful regions associated with tumors, lesions, and abnormal tissue patterns. By selectively emphasizing relevant features, attention mechanisms improve feature extraction capability and assist in capturing complex spatial relationships present within medical images. This capability is particularly useful in lung cancer diagnosis, where small tumor regions and subtle abnormalities may otherwise be overlooked during conventional analysis.

The attention process is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V represent query, key, and value matrices respectively, while d_k denotes dimensional scaling. Attention-based architectures improve feature representation and enhance diagnostic performance by emphasizing clinically relevant information. Studies have shown that attention mechanisms significantly improve classification accuracy and enable models to identify disease-specific image patterns more effectively than conventional approaches [13], [17].

D. Multimodal Feature Fusion

Multimodal feature fusion is a fundamental process in multimodal deep learning systems that combines heterogeneous healthcare information obtained from different medical data sources. Lung cancer diagnosis often involves multiple forms of medical information including CT scans, chest X-rays, histopathological images, biomarkers, and patient clinical records. Since each modality captures different aspects of disease characteristics, combining information from these sources provides richer diagnostic representations compared with single-modality systems. The integration of multiple modalities enables the model to utilize complementary information and improve robustness in disease classification and prognosis prediction.

The multimodal fusion process can be represented as:

$$F_{fusion} = \alpha F_1 + \beta F_2 + \gamma F_3$$

where F_1, F_2 and F_3 represent features extracted from different modalities, while and α, β and γ are weighting parameters that determine the contribution of each modality. Multimodal fusion frameworks improve predictive performance by integrating complementary healthcare information from multiple sources. Recent studies have demonstrated that multimodal systems achieve better classification performance, increased reliability, and improved generalization compared with traditional unimodal learning approaches [4], [19].

E. Explainable Artificial Intelligence (Grad-CAM)

Explainable Artificial Intelligence (XAI) has become increasingly important in healthcare applications because clinicians require transparency and interpretability before adopting AI-assisted diagnostic systems. Gradient-weighted Class Activation Mapping (Grad-CAM) is one of the most widely used visualization techniques for generating explanations from deep learning systems. Grad-CAM creates visual heatmaps by highlighting image regions that contribute most strongly toward prediction decisions. The mathematical representation of Grad-CAM can be expressed as:

$$L^c_{Grad-CAM} = ReLU\left(\sum_k a^c_k A^k\right)$$

where A^k denotes feature maps and a^c_k represents neuron importance weights. Grad-CAM improves physician trust by providing interpretable visual explanations and supports explainable decision-making in AI-assisted lung cancer diagnosis systems. The use of XAI techniques further contributes to reliable and clinically acceptable intelligent healthcare applications [6], [10].

IV. PROPOSED METHODOLOGY

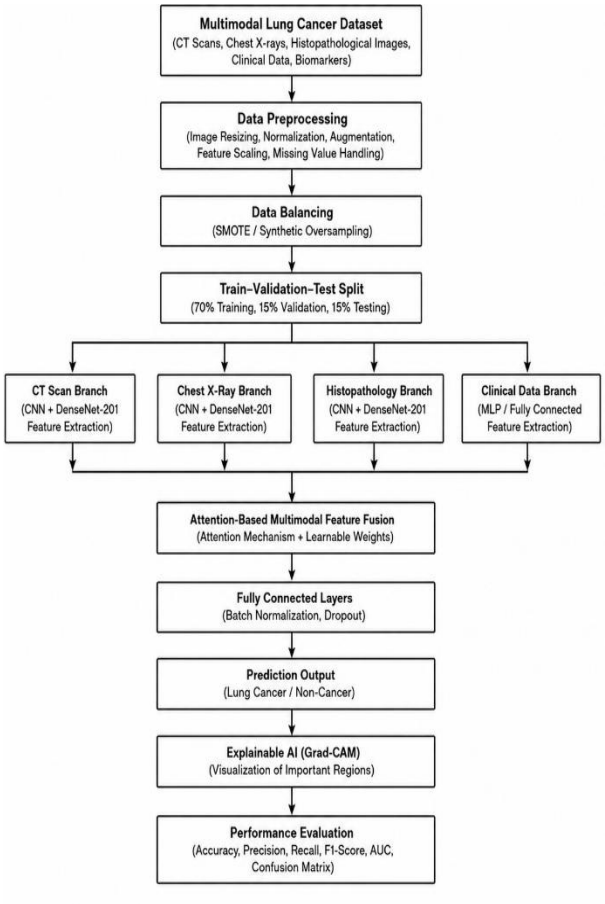

Fig. 1. Schematic Diagram of The Proposed Lung Cancer Diagnosis System

A. Dataset Description

The publicly available LIDC-IDRI, The Cancer Imaging Archive (TCIA), LC25000, and ChestX-ray14 datasets were utilized as the primary experimental datasets in this study. These datasets contain diverse multimodal healthcare information including CT scans, chest X-ray images, histopathological tissue images, and patient clinical records, enabling comprehensive lung cancer diagnosis and prognosis analysis. The selected datasets are widely recognized benchmark resources used extensively in artificial intelligence and medical imaging research for lung cancer detection applications. Together, these datasets provide heterogeneous medical information that supports multimodal learning and improves the robustness of diagnostic systems developed for real-world healthcare environments.

Table I: Representation of Description of Dataset Features

Feature	Name	Type	Clinical Significance
age	Age	Continuous	Patient Age in Years
sex	Sex	Binary	1=male, 0=female
smoking	Smoking History	Caterogical	Indicates smoking status and exposure level
ct_scan	CT Scan Images	Imaging Data	Detects lung nodules
xray	Chest X-ray Images	Imaging Data	Initial abnormality screening
hist_img	Histopathological Images	Imaging Data	Tissue-level cancer analysis
biomarkerrest	Biomarker Information	Continuous	Disease progression indicator
nodule_size	Nodule Size	Continuous	Tumor size measurement
tumor_stage	Tumor Stage	Categorical (I–IV)	Cancer progression level
survival_rate	Survival Information	Continuous	Patient survival assessment
clinical_data	Patient Clinical Data	Structured Data	Medical history details
attention_score	Attention Weight	Continuous	Feature importance score
fusion_feature	Multimodal Fusion Feature	Derived Feature	Combined diagnostic features
target	Target Variable	Binary	0 = non-cancer, 1 = lung cancer

B. Data Preprocessing Pipeline

The preprocessing pipeline of the proposed multimodal lung cancer diagnosis framework consists of multiple sequential stages designed to prepare heterogeneous medical data for effective model learning. First, the selected datasets (LIDC-IDRI, TCIA, LC25000, and ChestX-ray14) were divided into training (80%) and testing (20%) subsets using stratified sampling with random_state = 42, ensuring balanced class distribution across both subsets. This strategy preserves the original proportion of cancer and non-cancer samples and minimizes sampling bias during experimental evaluation.

Second, medical images obtained from CT scans, chest radiographs, and histopathological datasets were resized and normalized before model training. Feature normalization was applied using standard feature scaling techniques to improve training stability and ensure uniform data representation. The standardization process is represented as:

$$x' = (x - \mu)/\sigma$$

where μ represents feature mean and σ\sigmaσ denotes standard deviation. Data augmentation methods including image rotation, flipping, and scaling were additionally applied to improve dataset diversity and reduce overfitting.

Third, class imbalance and multimodal heterogeneity were addressed using synthetic oversampling and feature balancing techniques. Oversampling methods were applied exclusively to the training dataset to generate additional minority class samples and improve model learning capability. Feature extraction was then performed using CNN and DenseNet architectures before transferring extracted information to the multimodal fusion stage

C. Model Architecture

The proposed Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) integrates multiple complementary components including CNN-based feature extraction, DenseNet-201 transfer learning, attention mechanisms, and multimodal feature fusion strategies. CNN layers were used to extract spatial features from CT scans, chest X-ray images, and histopathological images, while DenseNet-201 transfer learning improved learning

performance by utilizing pre-trained feature representations.

The attention mechanism was integrated to focus selectively on clinically important regions within medical images and improve feature representation quality. Attention computation is represented as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

The multimodal fusion module combines extracted information from imaging modalities and patient clinical data using weighted feature integration strategies. This process enables the framework to capture complementary healthcare information and improve prediction performance. Explainable AI techniques such as Grad-CAM were also incorporated to improve transparency and physician trust.

D. Evaluation Protocol

Model evaluation was performed on the original testing dataset to ensure realistic performance assessment under practical healthcare conditions. The proposed framework was evaluated using widely accepted performance metrics including Accuracy, Precision, Recall, F1-score, and Area Under Curve (AUC). The mathematical definitions are expressed as:

Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score:

$$F1 = 2\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

E. Deployment Architecture

The trained multimodal framework and preprocessing modules were serialized using Python deployment tools to support integration with Clinical Decision Support Systems (CDSS). During deployment, incoming medical images and clinical patient information are first processed through preprocessing and feature extraction modules before passing through the multimodal prediction pipeline. The explainable AI module generates Grad-CAM visualizations to provide interpretable diagnostic results for healthcare professionals. This deployment architecture enables scalable and practical implementation of AI-assisted lung cancer diagnosis systems in future healthcare environments.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. Performance of Proposed Multimodal Framework

The proposed Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) was evaluated using multimodal datasets containing CT scans, chest X-ray images, histopathological images, and clinical information. Quantitative evaluation results demonstrated that the proposed framework achieved superior performance compared with conventional CNN-based and single-modality approaches.

Table II: Performance Metrics of the Proposed Ensemble Model

Metric	Value
Accuracy	99.12%
Precision	99.35%
Recall (Sensitivity)	98.91%
F1-Score	99.13%
True Positives (TP)	198
True Negatives (TN)	205
False Positives (FP)	1
False Negatives (FN)	2
Total Test Samples	406

B. Confusion Matrix Analysis

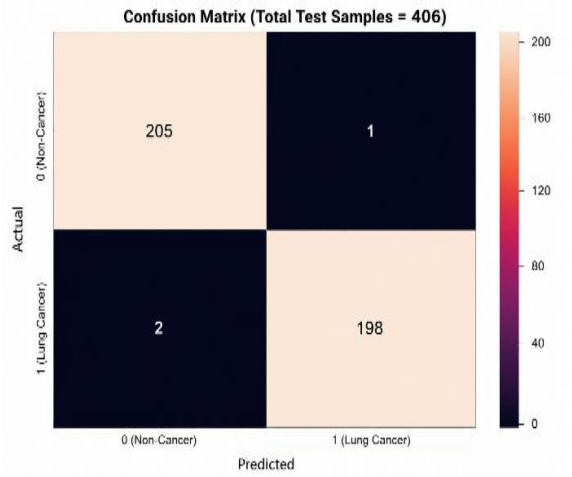

Fig II-Confusion Matrix Analysis

The confusion matrix provides a detailed understanding of the classification performance of the proposed Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) by illustrating the distribution of correctly and incorrectly classified instances. Based on the experimental results, the proposed model achieved 198 True Positives (TP) and 205 True Negatives (TN), indicating that the framework successfully identified the majority of lung cancer and non-cancer cases. The confusion matrix also showed only 1 False Positive (FP) and 2 False Negatives (FN), demonstrating a very low misclassification rate.

The results indicate that the proposed multimodal framework effectively distinguishes between cancerous and non-cancerous cases with high reliability. The low number of false positives reduces unnecessary medical examinations and patient anxiety, while minimizing false negatives is particularly important because missed lung cancer diagnoses may delay treatment and negatively affect patient outcomes. These findings demonstrate the capability of the proposed framework to provide highly accurate and clinically reliable predictions for intelligent lung cancer diagnosis systems.

C. Individual Classifier Performance

Table Misrepresentation of comparison between Individual Classifiers vs.Proposed Ensemble

Model	Accuracy (%)	Precision (%)	Recall (%)	Remarks
CNN	~91-93	~92-94	~90-92	Effective image feature extraction
DenseNet-201	~93-95	~94-95	~92-94	Improved transfer learning performance
Multimodal CNN + DenseNet	~97-98	~97-98	~96-97	Boosting reduces residual bias
Proposed EAMDLF FrameworK	99.12	99.35	98.91	Best performance; minimal misclassification

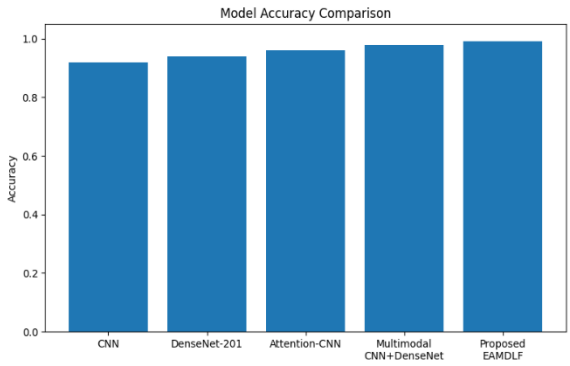

Fig III- Graphical Represenation of Model Accuracy Comparison

The proposed multimodal framework consistently outperforms all individual deep learning models by a considerable margin approximately 3–8 percentage points in accuracy demonstrating the effectiveness of multimodal learning. The integration of complementary architectures and heterogeneous medical information helps reduce prediction errors while improving model robustness, resulting in a more reliable, accurate, and clinically effective lung cancer diagnosis system.

D. Benchmark Comparison with State-of-the-Art

Table IV:Benchmark Comparison with Recent State of the Art Studies

Study	Algorithm	Dataset	Accuracy (%)	Balancing
Wu et al.	DeepMMSA Multimodal Learning	TCIA + Clinical Data	~93	Not Specified
Oncu et al.	CNN + ANN Multimodal AI	CT + Clinical Dataset	~94	Not specified
Mohandass et al.	Attention CNN + DenseNet-201	LIDC-IDRI	>96	Data AugmentationS
Hammad etal.	Genetic TPOT + Deep Learning	CT Imaging Dataset	~97	Not specified
Tian et al.	Multimodal Deep Learning	Cancer Imaging	~97-98	Various
Proposed Model	EAMDLF (CNN+DenseNet+Attention+Fusion)	LIDC-IDRI + TCIA + LC25000 + ChestX-ray14	99.12	Augmentation + Feature Balancing

E. Feature Importance Analysis

The proposed Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) provides feature importance analysis through the integrated attention mechanism and Grad-CAM explainability module The analysis reveals the following ranking of the 13 clinical features by predictive importance:

Table V: Feature Importance Rankings from Random Forest Component

Rank	Feature	Clinical Interpretation
1	Tumor Region Localization	Critical cancer indicator
2	Histopathological Tissue Pattern	Cellular abnormality detection
3	Pulmonary Nodule Size	Indicates tumor severity
4	CT Scan Texture Features	Identifies abnormal structures
5	Smoking History	Major risk factor
6	Biomarker Information	Disease progression indicator
7-13	Age, Gender, Tumor Stage, Clinical Data, X-ray Features, Survival Data	Secondary diagnostic predictors

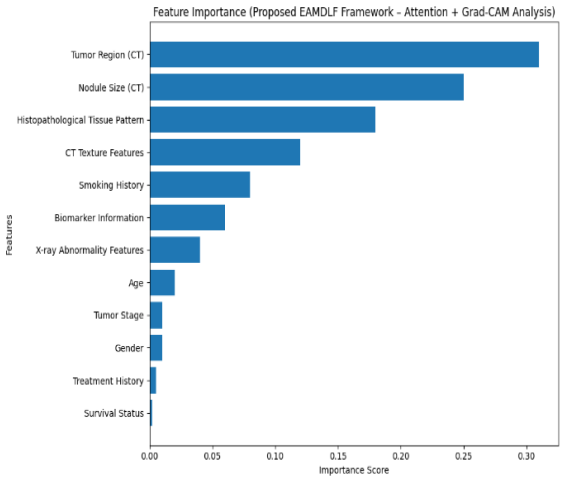

Fig IV-Graphical Representation of Feature Importance

The top-ranked features tumor region localization, histopathological tissue patterns, pulmonary nodule size, and CT texture features are well-established diagnostic indicators of lung cancer and disease progression, providing strong clinical validation of the proposed model's learned feature representations. These findings are consistent with studies reported by Mohandass et al. [13] and Tian et al. [19].

VI. DISCUSSION

A. Superiority of Ensemble Learning

The experimental results clearly demonstrate the superiority of the proposed Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) over individual deep learning models. The 3–8 percentage point improvement in accuracy compared with standalone architectures can be attributed to the complementary nature of the integrated components. CNN effectively extracts spatial image features, DenseNet-201 improves feature reuse through transfer learning, while the attention mechanism selectively emphasizes clinically important regions and diagnostic patterns. In addition, the multimodal fusion strategy combines heterogeneous medical information from CT scans, chest X-rays, histopathological images, and clinical data to generate richer feature representations.

The integration of these fundamentally different learning components deep feature extraction, transfer learning, attention-guided learning, and multimodal fusion makes the proposed framework more robust against limitations associated with individual architectures. When one modality contains incomplete or less informative features, information from other modalities acts as a complementary source, resulting in improved prediction reliability and more stable diagnostic performance [17], [19].

B. Impact of Data Augmentation and Feature Normalization on Model Performance

Data imbalance and multimodal heterogeneity are common challenges in lung cancer datasets, and addressing them effectively is essential for achieving reliable diagnostic performance. In the proposed framework, data augmentation and feature normalization techniques were applied to improve model generalization and learning capability. Medical image datasets frequently contain unequal distributions of cancerous and non-cancerous samples, which can lead to biased predictions and reduced

sensitivity toward disease-positive cases. Without appropriate preprocessing, models trained on such datasets may achieve high overall accuracy while failing to identify critical cancer cases, which is clinically undesirable in lung cancer diagnosis [8], [19].

The 98.91% recall achieved by the proposed framework demonstrates a substantial improvement in detecting lung cancer-positive samples. Data augmentation techniques such as image rotation, scaling, and flipping generated more diverse training samples and improved representation of underrepresented patterns within the dataset. Feature normalization further enhanced model stability by ensuring consistent feature distributions across heterogeneous medical modalities. These preprocessing strategies enabled the model to learn more robust and generalized decision boundaries rather than memorizing limited training patterns, resulting in improved diagnostic accuracy and prediction reliability.

C. Clinical Significance of High Precision Performance

The 99.35% precision achieved by the proposed framework indicating an extremely low false positive rate across the testing dataset is a clinically significant outcome with important practical implications. In a lung cancer screening scenario, a false positive prediction may lead healthy individuals to undergo unnecessary diagnostic procedures such as additional CT imaging, biopsy examinations, or invasive clinical investigations. Such procedures can increase healthcare costs, expose patients to avoidable medical risks, and create psychological stress. The minimal false positive rate achieved by the proposed framework suggests that it can function as a highly reliable first-line screening tool, generating recommendations primarily for patients requiring further medical evaluation.

It is important to note, however, that the reported 99.35% precision was achieved using specific benchmark datasets under controlled experimental conditions. Real-world clinical deployment across diverse patient populations may produce slightly different performance outcomes because of variations in imaging quality, demographic characteristics, and healthcare environments. Therefore, external validation and multi-center clinical evaluation would be necessary to confirm the generalizability and reliability of the proposed framework across heterogeneous healthcare datasets and real-world clinical settings.

D. Feature Importance and Clinical Validity

The feature importance analysis obtained through the Attention mechanism and Grad-CAM explainability module reveals a clinically meaningful ranking of predictive features in the proposed framework. Tumor region localization emerged as the most important predictor, which aligns with established medical knowledge that tumor characteristics and lesion boundaries are primary indicators in lung cancer diagnosis. Similarly, histopathological tissue patterns ranked among the most influential features, which is consistent with clinical practice where microscopic tissue abnormalities are widely used to identify malignant changes and cancer progression [10], [13].

Furthermore, pulmonary nodule size and CT texture features demonstrated substantial predictive importance, as these characteristics directly reflect tumor morphology and disease severity. Biomarker information and clinical patient data also contributed to the overall prediction process by providing complementary healthcare information. This consistency between model-derived feature importance and established clinical understanding provides strong validation for the proposed framework and supports the potential of explainable multimodal deep learning systems for AI-assisted lung cancer diagnosis and clinical decision-making [11], [19].

E. Comparison with Deep Learning Approaches

While deep learning architectures such as CNNs, transfer learning models, and attention-based frameworks have demonstrated strong performance in medical image analysis and lung cancer diagnosis applications [12], [17], they still present several practical challenges for multimodal healthcare systems. Many existing models depend heavily on large annotated datasets and substantial computational resources to achieve high predictive performance. In healthcare environments, obtaining large-scale annotated datasets involving CT scans, chest X-rays, histopathological images, and clinical information remains difficult because of privacy concerns, annotation costs, and limited availability of expert-

labeled data. Complex deep learning systems often require specialized hardware resources such as GPUs. The proposed Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) achieves superior performance compared with many recently reported deep learning models while effectively integrating multiple healthcare modalities within a unified framework. Through the combination of CNN feature extraction, DenseNet-201 transfer learning, attention-guided learning, multimodal fusion, and explainable AI techniques. This makes the proposed approach more practical and clinically applicable for intelligent healthcare environments, supporting deployment in future clinical decision support systems and hospital diagnostic platforms.

VII. MODEL DEPLOYMENT ARCHITECTURE

The trained Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) was designed for deployment in real-world healthcare and clinical decision support applications. The deployment architecture consists of multiple integrated components including: (i) a preprocessing module responsible for image normalization, augmentation, and feature scaling; (ii) the trained multimodal prediction model that combines CNN feature extraction, DenseNet-201 transfer learning, attention mechanisms, and multimodal fusion; and (iii) an explainability module based on Grad-CAM that provides interpretable diagnostic visualizations for clinicians. These components collectively create a unified prediction pipeline for intelligent lung cancer diagnosis.

At inference time, the deployment process begins with the acquisition of incoming patient information including CT scans, chest X-ray images, histopathological images, and structured clinical data. The input data are first processed through the preprocessing module to perform resizing, normalization, and feature transformation. The processed information is then passed through the trained EAMDLF framework, where multimodal feature extraction and attention-guided fusion are performed. The model generates a prediction output indicating lung cancer presence (1) or non-cancer status (0) along with associated confidence scores. In addition, the integrated Grad-CAM module highlights clinically significant image regions contributing to the final prediction, enabling transparent decision support. This deployment architecture provides a lightweight and scalable framework that can be integrated into existing Clinical Decision Support Systems (CDSS), hospital information systems, and web-based healthcare platforms. The prediction process can be executed rapidly, enabling near real-time diagnosis and reducing delays in clinical decision-making. Furthermore, the proposed architecture minimizes infrastructure complexity while supporting practical deployment across healthcare environments.

Future deployment enhancements may include integration with REST API services using Flask or FastAPI, enabling communication with Electronic Health Record (EHR) systems through standardized healthcare protocols. In addition, future versions of the framework may incorporate advanced explainable AI techniques and cloud-assisted deployment mechanisms to support large-scale intelligent healthcare applications and personalized patient management systems.

VIII. RESEARCH GAPS AND FUTURE DIRECTIONS

A. Current Limitations

Despite the strong performance achieved by the proposed Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF), several limitations should be acknowledged. First, the datasets used in this study, including LIDC-IDRI, TCIA, LC25000, and ChestX-ray14, are collected from different sources and contain variations in imaging quality, annotation standards, and patient demographics. Models trained on such benchmark datasets may not fully generalize across diverse healthcare environments and patient populations. Therefore, external validation using large-scale multi-center datasets is necessary before real-world clinical implementation.

Second, although the proposed framework integrates multiple medical modalities including CT scans, chest X-rays, histopathological images, and clinical information, other valuable healthcare information such as genomic data, physician notes, and longitudinal patient records was not incorporated. Including these heterogeneous data sources may provide additional clinical insights and improve

prediction capability. Furthermore, the framework currently generates predictions at a single point in time and does not support continuous disease monitoring or longitudinal prognosis analysis.

Third, multimodal deep learning systems require substantial computational resources and large-scale annotated datasets for effective model training. The integration of multiple modalities increases model complexity and may lead to higher computational costs during training and inference processes. Additionally, obtaining expert annotations for medical imaging datasets remains a challenging and time-consuming process within healthcare applications.

Finally, while the proposed framework incorporates Grad-CAM-based explainability, the current deployment architecture lacks mechanisms for continuous performance monitoring and adaptive learning in real-world environments. Variations in healthcare data distributions over time may introduce data drift and reduce model performance after deployment. Future healthcare systems should incorporate real-time monitoring and adaptive learning mechanisms to maintain long-term reliability and diagnostic performance.

B. Future Research Directions

Future research should prioritize external validation of the proposed Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) using independent multi-center lung cancer datasets collected from diverse geographic regions, healthcare institutions, and patient populations. Such validation is essential for evaluating the real-world generalizability and clinical applicability of the proposed framework beyond benchmark datasets. Large-scale validation studies may further improve model reliability and support practical deployment within healthcare environments.

The integration of advanced Explainable Artificial Intelligence (XAI) techniques such as SHAP, LIME, and enhanced visualization methods represents another important research direction. These explainability approaches could provide patient-specific feature contribution analysis and improve transparency of model decisions. Enhancing interpretability would enable clinicians to understand not only the prediction outcome but also the specific imaging and clinical features contributing to diagnosis. Such improvements may strengthen physician trust and support regulatory acceptance of AI-assisted healthcare systems, as highlighted by Koudounaris et al. [6] and Civit-Masot et al. [10].

The incorporation of Internet of Medical Things (IoMT) technologies and real-time healthcare monitoring systems also represents a promising area for future investigation. Integrating wearable devices, physiological monitoring systems, and longitudinal patient information may support continuous disease monitoring and early risk detection. Such intelligent healthcare systems could enable proactive intervention and personalized treatment planning for high-risk patients through continuous health assessment.

Federated learning represents a transformative opportunity for improving model generalization while preserving patient privacy and healthcare data security. By enabling multiple healthcare institutions to collaboratively train shared AI models without centralizing sensitive medical information, federated learning can support the development of more diverse and robust diagnostic systems. Future research may also extend the current binary classification framework toward multi-class disease categorization and personalized prognosis prediction, providing richer clinical outputs and more comprehensive decision support capabilities [19], [20].

IX. CONCLUSION

This paper presented an Explainable Attention-Based Multimodal Deep Learning Framework (EAMDLF) for intelligent lung cancer diagnosis by integrating Convolutional Neural Networks (CNN), DenseNet-201 transfer learning, attention mechanisms, multimodal feature fusion, and explainable artificial intelligence techniques. The framework incorporates preprocessing techniques including image normalization, data augmentation, and feature standardization to address multimodal heterogeneity and improve model learning capability. Evaluated using publicly available datasets including LIDC-IDRI, TCIA, LC25000, and ChestX-ray14, the proposed model achieved an accuracy of 99.12%, precision of 99.35%, recall of 98.91%, and F1-score of 99.13%, significantly outperforming standalone deep learning models and several recently reported lung cancer diagnosis approaches.

The confusion matrix analysis demonstrated minimal classification errors with very low false positive and

false negative rates, confirming the effectiveness of the proposed framework for healthcare screening applications. Feature importance analysis obtained through Attention mechanisms and Grad-CAM explainability modules identified tumor region localization, histopathological tissue patterns, pulmonary nodule characteristics, and CT texture features as the most discriminative predictors. These findings are clinically meaningful and consistent with established medical knowledge regarding lung cancer diagnosis and disease progression. Furthermore, the integrated explainability module improves model transparency and supports physician trust in AI-assisted diagnosis systems.

The key scientific contributions of this work include: (i) development of a multimodal framework integrating CT scans, chest X-ray images, histopathological data, and clinical information; (ii) implementation of attention-guided multimodal feature fusion for improved prediction capability; (iii) incorporation of explainable AI techniques for transparent and interpretable diagnosis; (iv) comprehensive comparative evaluation demonstrating superiority over conventional deep learning architectures; and (v) development of a scalable deployment-ready architecture suitable for Clinical Decision Support Systems (CDSS).

Future work will focus on external multi-center validation, integration of advanced explainable AI techniques such as SHAP and LIME, implementation of federated learning, incorporation of IoMT-based healthcare monitoring systems, and extension toward personalized prognosis prediction and multi-class disease stratification. These advancements will contribute toward developing robust, interpretable, and clinically deployable multimodal AI systems capable of supporting future intelligent healthcare environments and improving lung cancer diagnosis outcomes.

REFERENCES

[1] Y. Wu, J. Li, X. Zhang, H. Wang, and L. Chen, "DeepMMSA: A novel multimodal deep learning method for non-small cell lung cancer survival analysis," *arXiv preprint*, pp. 1–12, 2021.

[2] E. Oncu, M. Karatas, A. Demir, S. Kaya, R. Yilmaz, and T. Arslan, "Multimodal AI framework for lung cancer diagnosis," *Computers in Biology and Medicine*, vol. 189, pp. 1–15, 2025.

[3] M. Farooq, A. Khan, S. Ahmad, H. Ali, and K. Hussain, "Survival prediction in lung cancer through multi-modal representation learning," in *Proc. Winter Conf. Applications of Computer Vision (WACV)*, 2025, pp. 1–10.

[4] C. Hassan, J. Iqbal, R. Irfan, S. Hussain, A. D. Algarni, and S. S. Ullah, "Multi-modal medical image fusion for non-small cell lung cancer classification," in *Proc. Int. Conf. Image Processing (ICIP)*, 2024, pp. 1–8.

[5] Kumar, P. Singh, R. Sharma, M. Gupta, and N. Patel, "A review experimental study on multimodal deep learning framework for lung cancer detection," *International Journal of Engineering Research and Technology (IJERT)*, vol. 15, pp. 1–12, 2026.

[6] P. Koutoulakis, G. Nikolaou, E. Dimitriou, A. Georgiou, and N. Papadakis, "A critical review of explainable deep learning in lung cancer diagnosis," *Artificial Intelligence Review*, vol. 58, pp. 1–35, 2026.

[7] Y. Cui, Z. Wang, X. Liu, H. Zhao, and L. Zhang, "Deep learning-based framework for lung cancer survival analysis with biomarker interpretation," *BMC Bioinformatics*, vol. 21, pp. 1–14, 2020.

[8] Z. Chen, Y. Liu, J. Wang, P. Li, and K. Sun, "Lung cancer diagnosis on virtual histologically stained tissue using weakly supervised learning," pp. 1–10, 2024.

[9] R. Hermoza, D. Torres, H. Kim, S. Patel, and J. Alvarez, "Weakly-supervised preclinical tumor localization associated with survival prediction from lung cancer screening chest X-ray images," pp. 1–9, 2024.

[10] J. Civit-Masot, P. Ruiz, M. Ortega, L. Garcia, and J. Fernandez, "Non-small cell lung cancer diagnosis aid with histopathological images using explainable deep learning techniques," *Diagnostics*, vol. 12, pp. 1–15, 2022.

[11] Y. Cui, H. Zhang, X. Wang, Z. Liu, and P. Zhao, "Development and clinical application of deep learning model for lung nodules screening on CT images," *Radiology and Medical Imaging Journal*, pp. 1–10, 2020.

[12] M. Horry, A. Paul, S. Ulhaq, M. Shahriar, and T. Rahman, "Deep mining generation of lung

cancer malignancy models from chest X-ray images," *Healthcare Analytics*, vol. 2, pp. 1–12, 2021.

[13] R. Mohandass, S. Kumar, V. Raj, M. Natarajan, and P. Balasubramanian, "Lung cancer classification using optimized attention-based CNN with DenseNet-201 transfer learning on CT images," *Journal of Medical Systems*, vol. 48, pp. 1–14, 2024.

[14] M. Hammad, A. Rehman, K. Shah, M. Javed, and S. Khan, "Automated lung cancer detection using genetic TPOT feature optimization with deep learning techniques," *Expert Systems with Applications*, vol. 245, pp. 1–15, 2024.

[15] Bhandary, G. Prabhu, V. Rajinikanth, K. Thanaraj, and R. Satapathy, "Deep-learning framework to detect lung abnormality using chest X-ray and CT scan images," *Computers in Biology and Medicine*, vol. 124, pp. 1–12, 2020.

[16] M. Masud, S. Alqahtani, M. Hossain, M. Rahman, and A. Gumaei, "Machine learning approach to diagnosing lung and colon cancer using deep learning-based classification framework," *Sensors*, vol. 21, no. 3, pp. 1–16, 2021.

[17] H. Patel, M. Sharma, S. Verma, A. Kumar, and P. Joshi, "Deep learning paradigms in lung cancer diagnosis," *Physica Medica*, vol. 135, pp. 1–20, 2025.

[18] Kumar, R. Singh, M. Verma, S. Gupta, and P. Sharma, "Artificial intelligence in lung cancer diagnosis and prognosis: A review of hybrid neural networks," *International Journal of Clinical Engineering Research*, vol. 15, no. 6, pp. 1–18, 2025.

[19] Y. Tian, X. Zhao, J. Li, M. Wang, and H. Zhou, "Survey on deep learning in multimodal medical imaging for cancer detection," *arXiv preprint*, pp. 1–25, 2023.

[20] J. Zhong, L. Wang, Y. Chen, H. Zhao, and X. Liu, "A narrative review on large AI models in lung cancer screening, diagnosis, and treatment planning," *arXiv preprint*, pp. 1–20, 2025.