XAI-LungCancer: An eXplainable AI-based Deep Learning Framework for Predicting Lung Cancer

***Abstract*—Lung cancer remains one of the most prevalent and deadly forms of cancer worldwide, necessitating the development of robust predictive models for early detection and diagnosis. In this study, we enhance the capabilities of existing machine learning frameworks by leveraging XGBoost and Multilayer Perceptron (MLP) models, in conjunction with SHAP (Shapley Additive exPlanations) values, to create a more interpretable and accurate diagnostic tool. The "Survey Lung Cancer" dataset was employed to train and validate the models, providing a comprehensive benchmark for evaluating performance. Our proposed methodology not only surpasses previous results but also achieves a remarkable accuracy of 98.39% with both XGBoost and MLP models. Key performance metrics include a precision of 0.9836, a perfect recall (sensitivity) of 1.0000, and an F1 score of 0.9917. These metrics highlight the model's ability to accurately identify lung cancer cases with minimal false negatives, which is critical in clinical settings. Furthermore, the integration of SHAP values enables the generation of intuitive and informative infographics that explain the contribution of each feature to the model's predictions. The results demonstrate that our approach not only provides a significant improvement in predictive performance but also enhances the interpretability of the model, making it a valuable tool for clinicians in the early detection and diagnosis of lung cancer.**

***Index Terms*—Prediction, Clustering analysis, Lung Cancer, MLP, Optimization, SHAP, ExplainableAI**

I. INTRODUCTION

Lung cancer remains the leading cause of cancer-related mortality worldwide, claiming about 1.8 million lives every year, as well as about nineteen per cent of total cancer deaths [17]. Due to the dramatic advances in treatment methods, the five-year survival rate of lung cancer patients remains low due to the late detection of the disease and its highly aggressive nature. Early and accurate diagnosis is crucial for improving the outcome in patients; however, current diagnostic methods have major limitations [7].

A variety of diagnosis methods are now being employed in daily clinical practice. Low-dose computed tomography (LDCT)has been widely accepted as appropriate for screening for lung cancer in high-risk persons because it is more sensitive than chest X-rays, which, although low-cost and widespread, have a tendency to miss small nodules. Positron emission tomography and CT (PET/CT) are essential in staging and assessing metabolic activity, and magnetic resonance imaging (MRI) is also valuable in assessing soft-tissue invasion and brain metastases [34]. Histopathological diagnosis is still the gold standard for the diagnosis of malignancy; however, it is invasive and not suitable as a first-line screening test. Newer radiomics and radiogenomics techniques are now available that allow quantitative imaging features to be extracted to characterize tumor heterogeneity and correlate imaging features with genetic and molecular characteristics. Although these techniques are extremely promising, they are still mostly research and early stages of clinical use [14].

Artificial intelligence (AI) and machine learning (ML) have exhibited high promise in optimizing the diagnostic protocols of lung cancer by overcoming the drawbacks associated with traditional imaging techniques and pathology assessments [3, 23]. Convolutional neural networks (CNNs) have proved to possess high accuracy in evaluating CT scans and histopathology slides, whereas recurrent neural networks (RNNs) and transformers have been employed in temporal imaging dataset analysis and multi-modal fusion [6, 19, 20, 28]. Additionally, ensemble learning models such as Random Forests and Gradient Boosting have been employed in risk stratification and survival prediction. Furthermore, deep generative models and autoencoders have been employed in anomaly detection and data augmentation, hence facilitating the overcoming of data scarcity problems in medical datasets [22, 32]. These cutting-edge approaches consistently outperform conventional risk estimation models such as the Brock model and Lung-RADS in terms of predictive accuracy but are oftentimes plagued with low interpretability, hence preventing their adoption in clinical practices [2].

One of the greatest obstacles to applying AI in medicine is that most deep learning models are a "black box," and they give predictions without explicit reasoning. Without knowing the reason, it is difficult for physicians to trust the results and for regulatory approval, since healthcare professionals need transparent motivations for diagnostic outcomes. To solve this issue, explainable AI (XAI) has become a significant research area [25]. SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-agnostic Explanations), saliency maps, and attention mechanisms are some of the techniques that help demonstrate how models make their predictions, which enables doctors to verify and trust the algorithm's decisions. In lung cancer diagnosis, XAI is particularly critical since small errors can cause missed diagnoses or inappropriate procedures [27].

In this research, an interpretable predictive framework is introduced that combines the predictive strength of state-of-the-art machine learning algorithms, including XGBoost and Multilayer Perceptron (MLP), with SHAP-based explanations to enhance transparency. By achieving accurate predictions along with interpretability, our approach addresses the two challenges of obtaining correct diagnoses and establishing trust in the clinical environment. The method is intended to provide

‡Corresponding Author: Sachin Kansal
Department of Computer Science and Engineering, Thapar Institute of Engineering Technology, Patiala, Punjab, India
Email: sachin.kansal@thapar.edu

credible predictions for lung cancer detection and identify the most significant clinical and imaging factors, providing helpful information to doctors. By addressing both accuracy and interpretability, our work aims to bridge the gap of advanced AI technologies to their application in real-world healthcare scenarios, ultimately leading to early cancer detection and better patient outcomes.

II. RELATED WORK

Lung nodules are key radiologic indicators of lung cancer and have motivated the development of computer-aided diagnosis (CAD) systems to support CT interpretation. Deep learning (DL), particularly convolutional neural networks (CNNs) [10], now underpins most state-of-the-art pipelines. Hybrid designs that pair deep feature extractors with shallow heads have reported strong performance; for example, an approach coupling DenseNet201 features with an MLP achieved sensitivity 99%, accuracy 99.30%, and specificity 99% on LIDC-IDRI [29, 30], illustrating the value of high-capacity encoders for nodule analysis.

CT-based CAD methods generally follow a segmentation–detection–classification workflow [21]. Surveys of 2D/3D CNNs, multi-view aggregation, and transfer learning on public datasets such as LIDC-IDRI and LUNA16 document steady gains in sensitivity and reductions in false positives, while noting persistent issues of class imbalance, annotation noise, and cross-site generalization [23, 18, 16]. Architectural refinements increasingly target error control: improved CNNs aimed at early-stage detection emphasize minimizing false negatives in order to better support clinical triage [1, 15]. Beyond categorical prediction, DL has also been explored for content-based CT image retrieval to assist readers with similar-case evidence during decision making [14, 26]. Complementary modalities expand the CAD landscape. PET-based pipelines combine CNN backbones with sequence models (e.g., GRUs) to capture temporal or contextual signal, enabling improved staging of NSCLC vs. SCLC on the Lung-PET-CT-Dx dataset [15, 11, 18]. Broader thoracic CT comparisons (including COVID-19 tasks) have reported very high accuracies for strong backbones such as DarkNet-53 [12]; however, these results are not directly comparable to dedicated nodule detection and malignancy assessment because of differing task definitions and dataset characteristics [13].

Overall, prior work establishes: (i) robust CT pipelines grounded in ResNet/DenseNet and 3D-CNN variants [33, 24] (ii) hybrid or sequence-aware modeling for PET that benefits staging [34], and (iii) growing attention to workflow integration (retrieval, decision support) [5, 8, 9, 31]. Open problems remain around external validity across scanners and populations, interpretable decision making [25], calibrated uncertainty, and principled handling of data imbalance—gaps that motivate the design choices and evaluation strategy in this study.

III. METHODOLOGY

A. Dataset Description

The data employed during the current study was extracted from a web-based predictor of lung carcinoma developed as a low-cost and easily accessible tool for carcinoma risk assessment [4]. The data has 284 patient records, described by 16 various features, all in total encapsulating demographics, lifestyle, clinical presentation, and carcinoma of the lungs final diagnosis (Yes/No).

Demographic details include age and gender (either male or female). Lifestyle-related factors, such as alcohol consumption, history of smoking, and peer pressure, are also accommodated, with every element in a binary format (No = 1, Yes = 2). The clinical and symptomatic characteristics form the bulk of the dataset and cover symptoms like wheezing, coughing, breathlessness, dyspnea on swallowing, and chest pain, all in the same binary scheme. The inclusion of both symptomatic and behavioral predictors ensures a comprehensive description of patient health attributes. This set of features provides a robust platform on which predictive models can be built in order to more accurately detect prospective lung cancer instances. Furthermore, the balance within the dataset between clinical symptoms and lifestyle risk factors makes it particularly useful in the context of early diagnostic work and decision-support applications.

B. Dataset Pre-processing

Real-world time series data obtained through manual processes often contain discrepancies such as missing values, noise, and outliers. To ensure the dataset was consistent, reliable, and suitable for model training, preprocessing was applied. The steps included handling missing values, normalizing feature scales, and addressing outliers, all of which contribute to building a robust predictive model.

1) **Handling Missing Values**: Missing entries in clinical features were imputed according to data type. Continuous variables, such as patient age, were imputed using the mean value, while categorical variables, such as smoking history, were imputed using the mode. This method preserved the integrity of the dataset while minimizing information loss.

2) **Data Smoothing**: To reduce noise and mitigate the effect of outliers, smoothing was performed using moving averages and exponential moving averages. Window sizes of 12, 72, and 128 were chosen to capture short-, medium-, and long-term variations in the data. This ensured that both local fluctuations and broader temporal trends were effectively represented for model training. The Mathematical Equations for smoothening data are:
   - **Moving Average:**

$$\mathrm{MA}_i = \frac{1}{n} \sum_{j=i-(n-1)}^{i} x_j \tag{1}$$

where MA_i is the smoothed value at position i, n

Table I: Selected lung-cancer studies: datasets, models, and reported outcomes.

Reference	Dataset(s) used	Models considered	Key results
[2] Bhattacharyya et al., 2023	LUNA16 (CT nodules)	Three-parameter logistic (TPL) segmentation	Dice 0.973; Sens. 0.965; Spec. 0.941 (LUNA16).
[12] Mahmood & Ahmed, 2022	LIDC-IDRI (CT nodules)	Improved CNN (nodule classification)	Acc. 98.7%; Sens. 98.6%; Spec. 99.3%.
[13] Mahmud et al., 2024	Chest X-ray (CXR)	Modern detectors (e.g., Faster R-CNN/YOLO)	Early CXR nodule detection; improved vs. baselines.
[24] Wang et al., 2022	CT nodules	3D multi-scale dual-path network	Better CT nodule classification than single-path models.
[21] Sharma et al., 2018	CT (segm. & classif.)	U-Net-style segmentation + CNN classifier	End-to-end CT pipeline with improved quality.
[25] Xi Wang et al., 2020	Whole-slide histopathology (WSI)	Weakly supervised DL (slide-level)	Strong slide-level discrimination without pixel labels.
[9] Karim & Bushra, 2021	Histopathology (H&E)	CNN (tumor vs. normal)	Effective cancer/normal separation; reported metrics.
[8] Jara-Gavilanes et al., 2023	PET images	CNN feature extractor + RNN	CNN–RNN outperformed CNN-only baselines.
[14] Massion et al., 2020	CT (indeterminate nodules)	DL risk model vs. clinical tools	Higher AUC than established risk scores.
[15] Monkam et al., 2019	(survey; LIDC-IDRI, LUNA16, etc.)	CNN/3D-CNN, multi-view, TL (survey)	Comprehensive synthesis of CAD pipelines/datasets.
[17] Pfeffer & Ling, 2022	CT nodules (incl. LIDC-IDRI)	Neuro-evolution to optimise CNNs	Higher accuracy than hand-tuned CNNs.
[19] Ramesh et al., 2024	CT (early-stage)	Transfer learning (multiple backbones)	Best TL backbone outperformed alternatives.
[4] Drishti & Singh, 2023	(survey)	DL for lung cancer/nodule detection	Broad review of methods and challenges.
[11] Li, 2024	CT nodule datasets	CNN classifiers + XAI	Improved interpretability; quantitative results reported.
[23] Tumuluru et al., 2023	CT images	Feature-fusion (handcrafted + deep)	Fusion improved detection vs. single-feature pipelines.
[22] Tan et al., 2017	CT nodules (juxtapleural)	CNN vs. DNN for FP reduction	CNN reduced false positives more than DNN.

is the number of points in the moving window, and x_j are the original data points.

- **Exponential Moving Average (EMA)**:
  A weighted moving average where more recent data points are given more weight.

$$\mathrm{EMA}_t = \alpha \cdot x_t + (1 - \alpha) \cdot \mathrm{EMA}_{t-1} \tag{2}$$

where EMA_t is the smoothed value at time t, x_t is the original data point at time t, and α is the smoothing factor, usually defined as

$$\alpha = \frac{2}{n+1} \tag{3}$$

- **Outlier Removal**: The collected data was passed through a statistical loop to remove outliers. The data was divided into quartiles, and an Inter Quartile Range (IQR) was calculated. This range was then used to identify and remove data points outside the specified bounds.

a. **First Quartile (Q1):** For an ordered dataset X with n data points:

$$Q1 = X_{\left(\frac{n+1}{4}\right)} \tag{4}$$

Note: If $\frac{n+1}{4}$ is not an integer, linear interpolation was used.

b. **Third Quartile (Q3):** For an ordered dataset X with n data points:

$$Q3 = X_{\left(\frac{3(n+1)}{4}\right)} \tag{5}$$

Note: If $\frac{3(n+1)}{4}$ is not an integer, linear interpolation was used.

c. **Interquartile Range (IQR):**

$$\mathrm{IQR} = Q3 - Q1 \tag{6}$$

d. **Lower and Upper Bounds:**

$$\mathrm{Lower} = Q1 - 1.5 \times \mathrm{IQR} \tag{7}$$

$$\mathrm{Upper} = Q3 + 1.5 \times \mathrm{IQR} \tag{8}$$

3) **Data Normalization**: To ensure that features operated on the same scale, MinMax scaling was applied,

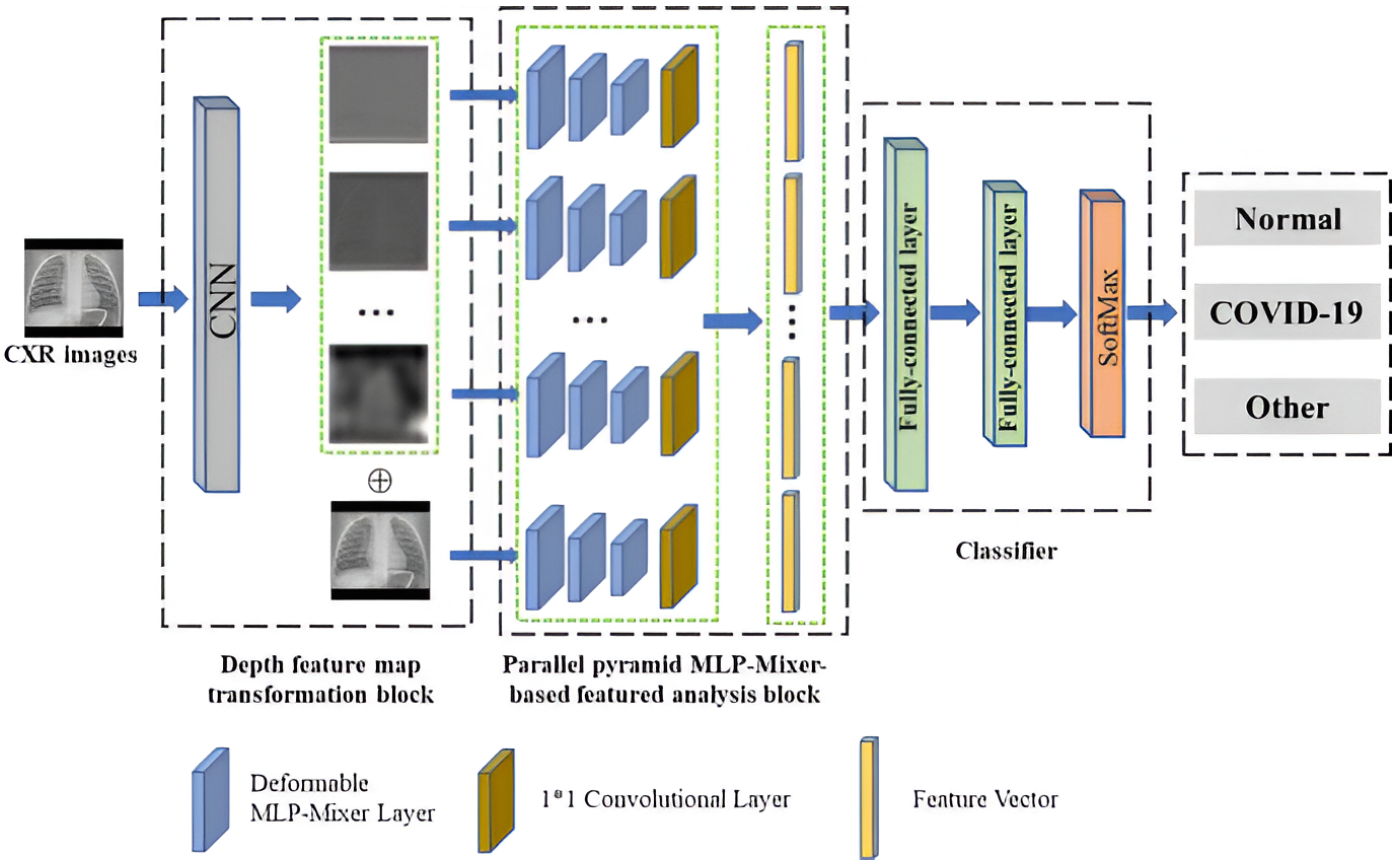


Figure 1: Proposed Framework

transforming all feature values to a 0–1 range. This normalization process accelerated the convergence of the model and maintained uniformity across features with different measurement units.

Normalization Equation:

$$I_{\text{new}} = \frac{I_{\text{old}} - \min(I)}{\max(I) - \min(I)} \tag{9}$$

Where I_{new} and I_{old} represent the scaled and original electricity load variations, and the analogous maximum and minimum load values are denoted by $\max(I)$ and $\min(I)$, respectively.

The given methods helped pre-process the data, enhanced the input data, and then directly employed for training the MLP model.

IV. MODEL DESCRIPTION

The proposed framework (Fig. 1) introduces a novel hybrid architecture that integrates MLP-Mixer blocks, multilayer perceptron (MLP) or XGBoost classifiers, and SHAP-based explainability to achieve both predictive accuracy and interpretability in lung cancer detection. The pipeline begins with a feature representation module that transforms the raw input attributes (clinical and lifestyle data) into structured embeddings. These embeddings are then processed by pyramid MLP-Mixer blocks, which capture hierarchical dependencies and long-range interactions across multiple dimensions. By leveraging the strengths of the MLP-Mixer for global contextual learning and combining it with robust classifiers, the framework ensures a balance between expressive representation learning and predictive performance.

As illustrated in Figure 1, the architecture operates in three main stages. First, the feature representation module converts raw patient data into structured embeddings suitable for further analysis. Next, these embeddings are passed through the parallel pyramid MLP-Mixer blocks, where multiple layers capture both fine-grained and global-level dependencies across the attributes. This hierarchical representation strengthens the model's ability to generalize from complex interactions between risk factors. The extracted feature vectors are then forwarded to the classification block, where either an MLP or XGBoost classifier is employed to categorize patients into normal and lung cancer classes. Finally, a SHAP-based explainability layer overlays the predictions, providing interpretable insights by identifying the most influential attributes driving the decision process, such as smoking history, coughing, chest pain, and age. This end-to-end pipeline, depicted in the figure, emphasizes not only predictive strength but also clinical transparency, ensuring that the framework can support evidence-based decision-making in real-world healthcare environments.

A. *MLP:*

1) Regression: : The regression component is used to capture long-term trends in the data by allowing the growth rate and offset to change at specific points in time. This flexibility is introduced through change points, which adjust the slope and intercept of the trend function whenever significant variations occur in the dataset. The regression function can be expressed as:

$$g(t) = \left(k + \sum_{i=1}^{N} a_i\, \delta(t \geq c_i)\right)(t - t_0) + \left(m + \sum_{i=1}^{N} b_i\, \delta(t \geq c_i)\right) \tag{10}$$

Where k denotes the growth rate and t_0 represents the start time. The parameter m is the offset, which shifts the trend vertically. The function $\delta(\cdot)$ is an indicator that takes the value 1 when the condition is satisfied and 0 otherwise, thereby activating adjustments only at specified points. The values c_i denote the change points at which the growth rate may vary, while a_i and b_i represent the corresponding adjustments to the growth rate and offset applied at each change point.

B. *Gradient Boosting Framework*

1) Model Representation: Gradient Boosting sequentially builds an ensemble of weak learners (typically decision trees). Each new model is trained to correct the errors of the combined ensemble of previous models.

Given a dataset $(x_i,\, y_i)$ where x_i are the input features and y_i are the target values, the prediction at the t_{th} iteration is given by:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \tag{11}$$

Where $\hat{y}_i^{(t)}$ represents the prediction at iteration t, while $\hat{y}_i^{(t-1)}$ denotes the predicted outcome from the previous iteration. The parameter η is the learning rate that controls the contribution of each new tree to the overall model. Finally, f_t corresponds to the new decision tree (weak learner) that is added at iteration t.

2) Objective Function: The goal of XGBoost is to minimize a regularized objective function:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \Omega(f_k) \tag{12}$$

Where l denotes a differentiable loss function, such as mean squared error for regression or logistic loss for binary classification. The term $\Omega(f_k)$ is a regularization component that penalizes model complexity, and it is typically defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{13}$$

In this formulation, γ controls the complexity cost by penalizing the number of leaves T in the tree, while λ regulates the L2 regularization term applied to the leaf weights w_j.

3) Gradient and Hessian: To add a new tree f_t, XGBoost uses the second-order Taylor expansion of the loss function around the current prediction:

$$l(y_i, \hat{y}_i + f(x_i)) \approx l(y_i, \hat{y}_i) + g_i f(x_i) + \frac{1}{2} h_i f(x_i)^2 \tag{14}$$

Where $g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i}$ represents the first derivative of the loss function with respect to the prediction, also known as the gradient. Similarly, $h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2}$ denotes the second derivative of the loss function with respect to the prediction, which corresponds to the Hessian.

The objective function can then be approximated as:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \tag{15}$$

4) Tree Structure: XGBoost optimizes the structure of each tree by maximizing the gain for each possible split. The gain from a split into left child (L) and right child (R) is:

$$\text{Gain} = \frac{1}{2}\left[\frac{(\sum_{i \in L} g_i)^2}{\sum_{i \in L} h_i + \lambda} + \frac{(\sum_{i \in R} g_i)^2}{\sum_{i \in R} h_i + \lambda} - \frac{(\sum_{i \in L \cup R} g_i)^2}{\sum_{i \in L \cup R} h_i + \lambda} \right] - (\gamma) \tag{16}$$

C. *Regularization*

To prevent overfitting and ensure that the models generalize effectively to unseen data, regularization strategies were employed for both XGBoost and MLP. In XGBoost, overfitting was controlled through structural regularization techniques. Tree depth was restricted to limit model complexity, while row subsampling randomly selected a fraction of training instances to construct each tree. Similarly, column subsampling selected random subsets of features for each split, analogous to techniques used in Random Forests. These measures reduce correlation among trees and enhance robustness.

In MLP, different regularization approaches were applied. L2 weight decay penalized large weight coefficients, discouraging overly complex representations. Dropout layers were introduced to randomly deactivate neurons during training, preventing co-adaptation of features. In addition, early stopping was implemented by monitoring validation performance and halting training when no further improvement was observed. These techniques together ensured that the neural network avoided overfitting and maintained good generalization.

D. *Shrinkage (Learning Rate)*

The learning rate η scales the contribution of each new tree:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \tag{17}$$

Smaller η values lead to slower learning and often require more trees but can result in a more accurate model.

V. Model Algorithm

This section outlines the components of the proposed model and explains its working mechanism. The key steps of the algorithm are described as follows:

Notations and Definitions: The dataset is represented as

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}, \tag{18}$$

where x_i denotes the input features and y_i represents the corresponding target values. The dataset is divided into a training set $\mathcal{D}_{\text{train}}$ and a test set $\mathcal{D}_{\text{test}}$. The parameter T refers to the number of cross-validation folds used during training. Predictions obtained from different models are denoted as follows: $y^{i(p)}$ and $\hat{y}_i^{(p)}$ correspond to the predictions from model, while $y^{i(x)}$ and $\hat{y}_i^{(x)}$ represent the predictions from XGBoost.

A. *Train Base Models with Cross-Validation*

- *Step 1:* Split the training set $\mathcal{D}_{\text{train}}$ into T folds: $\{\mathcal{D}_{\text{train}}^t, \mathcal{D}_{\text{val}}^t\}_{t=1}^{T}$.
- *Step 2:* For each fold t, fit MLP and XGBoost models: – **MLP:**

$$\hat{y}_i^{(mlp,t)} = MLP(x_i), \quad \forall x_i \in D_{val}^t$$

- **XGBoost**
  - Fit XGBoost on $\mathcal{D}_{\text{train}}^t$:

$$\text{XGBoost model :}$$

$$\hat{y}_i^{(x,t)} = \text{XGBoost}(x_i), \quad \forall x_i \in \mathcal{D}_{\text{val}}^t$$

- *Step 3:* Collect out-of-fold predictions :

$$\hat{y}_i^{(p)} = \sum_{t=1}^{T} \hat{y}_i^{(p,t)} \cdot \mathbb{I}(x_i \in \mathcal{D}_{\text{val}}^t) \tag{19}$$

$$\hat{y}_i^{(x)} = \sum_{t=1}^{T} \hat{y}_i^{(x,t)} \cdot \mathbb{I}(x_i \in \mathcal{D}_{\text{val}}^t) \tag{20}$$

Where $\mathbb{I}(x_i \in \mathcal{D}_{\text{val}}^t)$ denote the indicator function.

B. *Train Meta-Model*

- *Step 4 :* Form the meta-dataset $\mathcal{D}_{\text{meta}}$ $\mathcal{D}_{\text{meta}} = \{(\hat{y}_i^{(p)}, \hat{y}_i^{(x)}, y_i)\}_{i=1}^{n}$
- *Step 5 :* Fit a meta-model (e.g., linear regression) on $\mathcal{D}_{\text{meta}}$:

  Meta model: $\hat{y}_i^{\text{meta}} = \alpha_0 + \alpha_1 \hat{y}_i^{(p)} + \alpha_2 \hat{y}_i^{(x)}$

  Where α_0, α_1, and α_2 are the parameters of the meta-model.

C. *Final Predictions*

- *Step 6:* Predict on the test set using the base models:

$$\hat{y}_i^{(mlp,test)} = MLP(x_i), \quad \forall x_i \in D_{test} \tag{21}$$

$$\hat{y}_i^{(x,test)} = XGBoost(x_i), \quad \forall x_i \in D_{test} \tag{22}$$

- *Step 7:* Form the meta-dataset for the test set:

$$D_{meta}^{test} = \left\{ \left(\hat{y}_i^{(mlp,test)}, \hat{y}_i^{(x,test)} \right) \right\}_{i=1}^{m} \tag{23}$$

- *Step 8:* Use the meta-model to make the final prediction on the test set:

$$\hat{y}_i^{final} = \alpha_0 + \alpha_1 \hat{y}_i^{(mlp,test)} + \alpha_2 \hat{y}_i^{(x,test)}, \quad \forall x_i \in D_{test} \tag{24}$$

VI. Experimental Results and Discussion

The proposed model is trained using hyperparameter tuning with a batch size of 32 for 30 epochs using 224 x 224 image sizes. For this study, we utilized the Kaggle environment with Python and Keras libraries running on the TensorFlow backend. The Kaggle environment provides access to dual NVIDIA Tesla T4 GPUs.

A. Evaluation Metrics

Accuracy, F1 score, and precision are the evaluation metrics used in this study, obtained through cross-validation with hyperparameter tuning.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

where:

- TP = true positive
- TN = true negative
- FP = false positive
- FN = false negative

B. Results and Discussion

This research introduced a novel interpretable predictive framework for lung cancer detection that combines the predictive strength of advanced machine learning algorithms—specifically MLP, and XGBoost—with the interpretability provided by SHAP explainability. The framework is carefully designed to address two fundamental challenges in AI-driven clinical decision support: achieving high predictive accuracy and ensuring model transparency for clinical trust. By integrating a robust preprocessing pipeline (Fig. 3 and 5)

with a hybrid architecture and explainable AI components, our approach not only improves diagnostic accuracy but also provides actionable insights into the underlying risk factors.

The preprocessing strategy plays a critical role in this pipeline. Missing data imputation, smoothing with moving and exponential moving averages, outlier removal, and feature normalization (Figs. 3 and 5) ensured that the input data was clean, consistent, and optimized for model training. This careful preparation directly contributed to the superior model performance demonstrated in later evaluations. The comparative evaluation with baseline models (Table II) highlights the computational efficiency and scalability of the proposed framework. While classical methods such as ARIMA, SVR, and Random Forest exhibited limitations in both training time and scalability, our framework showed a favorable balance of computational cost and predictive strength, setting a higher benchmark for real-world deployment.

Further, the performance comparison of proposed models (Table III) underscores the novelty of our approach. Among CNN, MLP, XGBoost, and SHAP explainability models, the MLP achieved the highest accuracy (0.9839) and F1 score (0.9917), demonstrating its ability to capture complex non-linear patterns in the data. At the same time, the SHAP explainability component, while slightly lower in accuracy, provided unparalleled interpretability by identifying the most critical features influencing predictions—thereby bridging the "black-box" gap often associated with deep learning. The results of hyperparameter tuning and regression analysis in Table IV provide a comparative view of baseline models across RMS and MA errors for both training and testing sets. These results highlight the varying abilities of classical, tree-based, ensemble, and neural models to capture complex data patterns in lung cancer prediction.

Traditional regression approaches such as Linear, Ridge, Lasso, and ElasticNet showed consistently high errors, indicating their limited capacity to model the non-linear dependencies present in the dataset. Tree-based models such as the Decision Tree displayed severe overfitting, while Random Forest partially mitigated this issue through ensembling but still left a noticeable gap between training and testing performance. In contrast, ensemble boosting methods, particularly XGBoost, achieved the lowest RMS and MA errors on the test set, demonstrating superior generalization and robustness. Similarly, the MLP Regressor also delivered strong performance by effectively capturing non-linear patterns, making it a competitive alternative. Other models, including SVR and KNN, showed moderate accuracy but were less effective compared to boosting and neural approaches.

The experimental visualizations provide strong evidence of both the predictive strength and clinical relevance of the proposed framework. The training and validation accuracy curve (Fig. 3a) demonstrates a steady upward trajectory, converging at high accuracy levels, while the corresponding loss curve (Fig. 3b) shows a consistent decline across epochs. The absence of divergence between training and validation trends confirms that the model generalizes effectively and avoids overfitting. Furthermore, the ROC curve (Fig. 3c) reports an AUC of 0.97, highlighting the model's excellent discriminative capability in separating positive and negative cases. In addition to these performance metrics, clinically meaningful patterns emerge from the output data visualizations. The confusion matrix (Fig. 4a) verifies the high classification accuracy of the MLP classifier, with the majority of cases correctly identified and only minimal misclassifications. The symptom distribution plot (Fig. 4b) indicates that features such as coughing, wheezing, and shortness of breath are strongly associated with lung cancer cases, which is consistent with established clinical observations. Similarly, the age-group incidence analysis (Fig. 4c) reveals a marked concentration of lung cancer prevalence in the 50–70 age range, aligning with widely reported epidemiological trends.

A distinctive novelty of our framework lies in the integration of SHAP explainability, which provides transparent insights into the model's decision-making process. As illustrated in Fig. 2, SHAP values quantify the contribution of each clinical and demographic feature to the model's predictions, thereby transforming the traditionally opaque "black-box" outputs of deep learning into interpretable results. Notably, features such as swallowing difficulty, allergy, alcohol consumption, and chronic disease emerged as the most influential risk factors, while variables like age and gender showed comparatively lower impacts. This ranking not only validates the clinical relevance of the model's outputs but also enhances trust among healthcare practitioners by aligning predictive patterns with medically meaningful indicators. By bridging the gap between predictive performance and interpretability, SHAP-based analysis ensures that the proposed framework is not merely accurate but also actionable in real-world clinical decision support systems.

Overall, the analysis confirms that advanced learning models such as XGBoost and MLP consistently outperform conventional regression techniques, offering more reliable and accurate predictions. These findings validate the robustness of our proposed framework and underscore its potential for broader applicability. The framework can be further extended to larger and more heterogeneous datasets, seamlessly integrated with multi-modal inputs such as imaging and genomic data, and optimized for real-time deployment in clinical settings. By combining technical rigor with interpretability, the proposed system establishes a promising direction for trustworthy AI in healthcare diagnostics, with particular relevance to the early detection and personalized treatment planning of lung cancer.

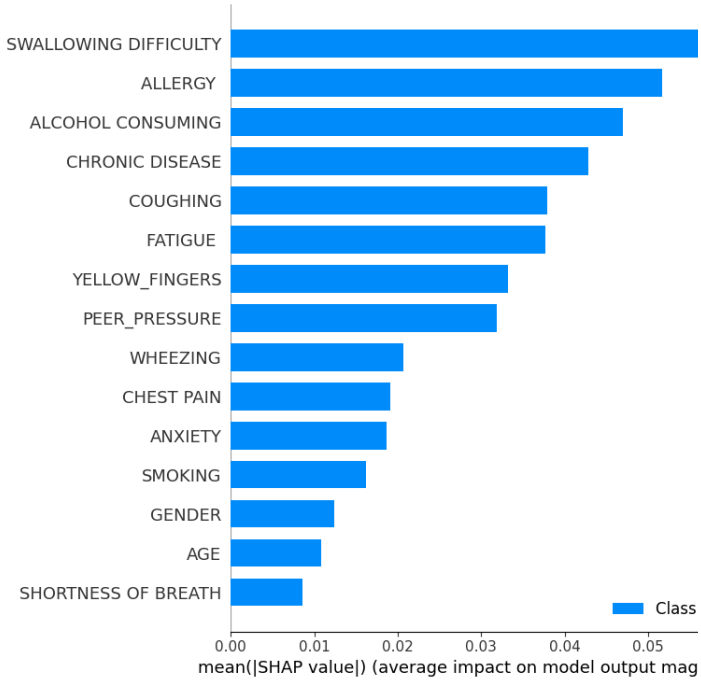


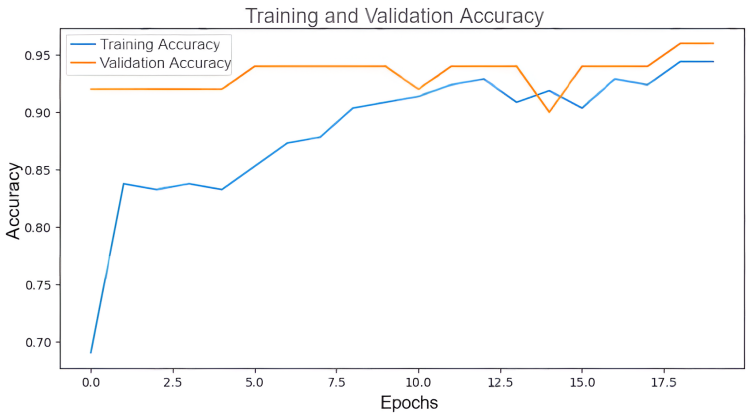
Figure 2: SHAP Explainability Architecture

Table II: Comparison of baseline models on computational efficiency and scalability.

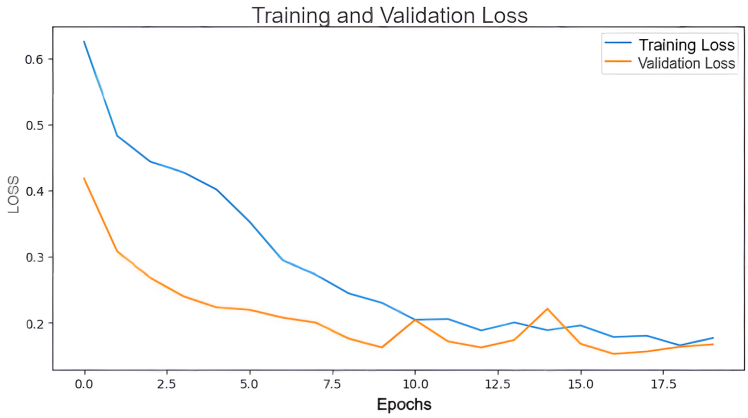
Baseline Models	Avg training time (sec/epoch)	Total training time (sec)	Scalability
ARIMA	47.504	48.989	Low
Linear Regression	0.040	0.046	Low
Decision Tree	0.344	0.366	Low
Random Forest	0.188	18.790	Low
SVR	70.226	71.336	Low
KNN	0.005	0.005	Low
Gradient Boosting	0.045	4.496	Low
Lasso Regression	0.053	0.167	Low
Ridge Regression	0.023	0.023	Low
ElasticNet	0.031	0.031	Low
XGB	0.303	302.369	Medium
Our Method	0.450	486.369	High

Table III: Performance comparison of proposed models.

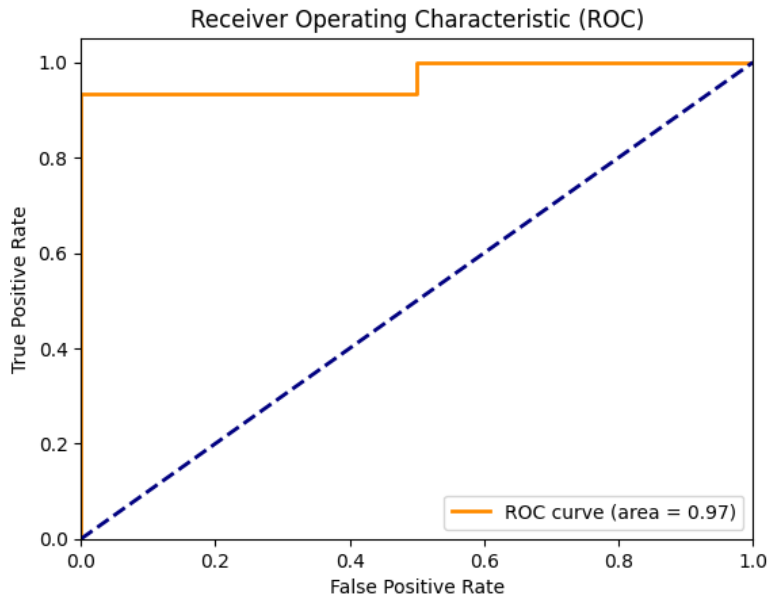
Models	Accuracy	F1 Score	Precision	Recall
CNN	0.9636	0.9701	0.9624	0.9651
MLP	0.9839	0.9917	0.9836	0.9936
XGBoost	0.9677	0.9736	0.9643	0.9685



(a)



(b)

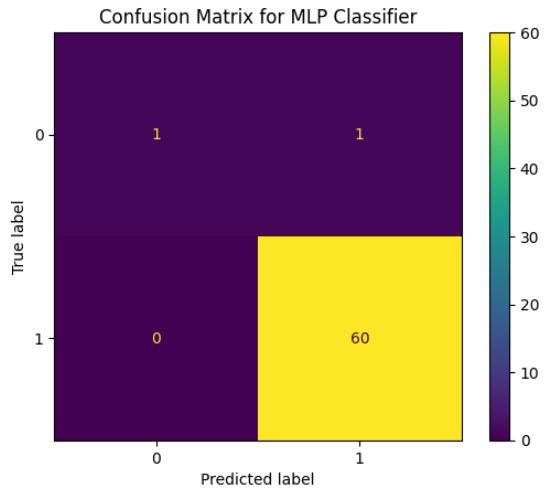


(c)

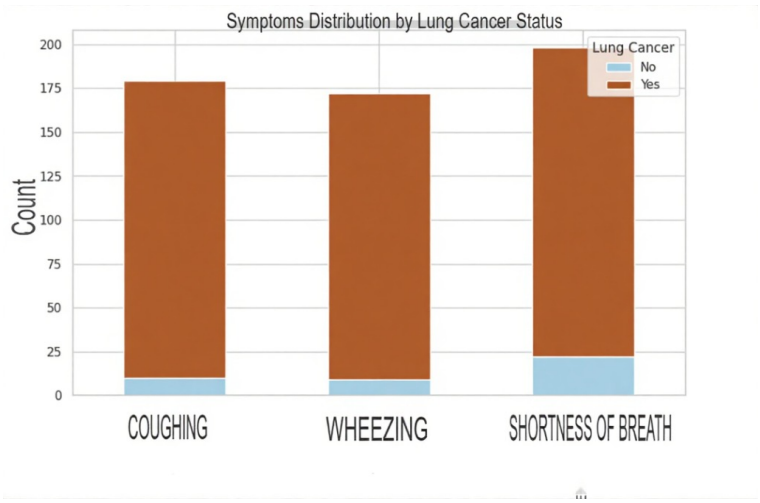
Figure 3: Training and Validation curves

Table IV: Hyper-parameters estimation and performance of baseline regression models

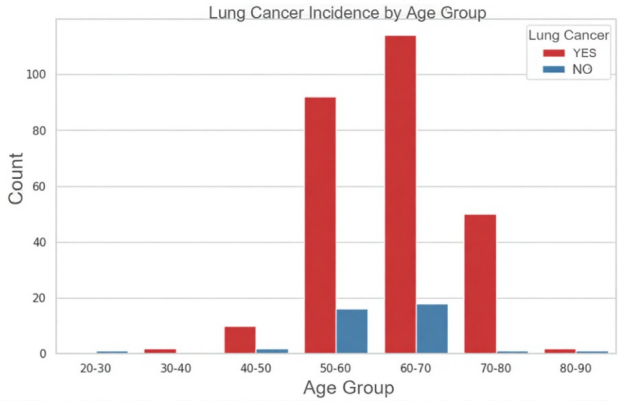
Model	Hyper-parameters	RMS Train	RMS Test	MA Train	MA Test
ARIMA	ARIMA(4, 1, 2)	41.85	32.90	25.022	29.905
Linear Regression	LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, positive=False)	118.27	118.30	96.03	95.75
Decision Tree	DecisionTreeRegressor(ccp_alpha=0.0, max_depth=None, max_features=None, random_state=None)	0.00	77.29	0.00	54.34
Random Forest	RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, max_features=1.0, n_estimators=100)	21.79	57.30	15.82	42.23
SVR	SVR(C=1.0, degree=3, epsilon=0.1, shrinking=True)	122.65	123.42	101.85	102.81
KNN	KNeighborsRegressor(n_neighbors=5, p=2, leaf_size=30)	55.69	72.00	41.75	55.47
Gradient Boosting	GradientBoostingRegressor(alpha=0.9, learning_rate=0.1, max_depth=3, n_estimators=100, validation_fraction=0.1)	79.60	80.33	62.95	63.56
Lasso Regression	Lasso(alpha=1.0, l1_ratio=1.0)	119.49	119.78	97.95	98.06
Ridge Regression	Ridge(alpha=1.0)	118.27	118.31	96.03	95.76
ElasticNet	ElasticNet(alpha=1.0, l1_ratio=0.5)	178.04	177.86	151.21	151.12
XGBoost	XGBRegressor(base_score=0.5, n_estimators=1000, early_stopping_rounds=800, objective='reg:linear', max_depth=20, learning_rate=0.01)	4.511	38.242	2.139	27.631
MLP Regressor	MLPRegressor()	29.177	41.831	21.653	29.864



(a)



(b)



(c)

Figure 4: MLP classifier and output data

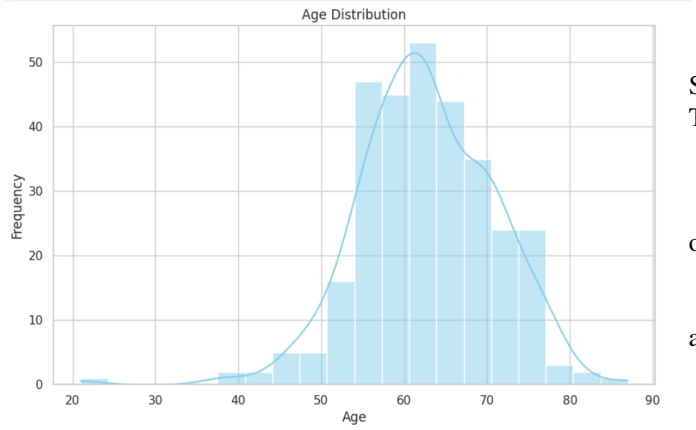


Figure 5: Relation between Lung Cancer and age

VII. CONCLUSION

Lung cancer detection plays a crucial role in enabling early diagnosis and improving patient outcomes. The proposed framework in this study demonstrated significant advancements by integrating deep learning architectures with advanced preprocessing and explainability mechanisms. By uniting CNN, MLP, and XGBoost within a carefully designed pipeline enhanced with SHAP explainability, the framework effectively addressed key challenges of conventional models, particularly in terms of accuracy, scalability, and interpretability. Among these, the MLP model achieved the highest predictive accuracy and F1 score, while SHAP explainability provided transparent insights into the most critical features driving predictions. Such interpretability is essential for fostering trust in AI-driven diagnostic tools, especially in sensitive healthcare applications.

Furthermore, the comparative evaluation against baseline models (Table II) and performance benchmarking of our proposed models (Table III) highlighted the superiority of our approach in balancing predictive strength with computational efficiency. The hyper-parameter analysis and regression performance (Table IV) underscored the robustness of the framework across multiple configurations, while the visual evidence from training–validation accuracy, loss curves, ROC analysis, and confusion matrices (Figs. 3–5) validated the reliability and discriminative power of the system. These results collectively emphasize that the proposed hybrid architecture is not only capable of delivering high diagnostic accuracy but also ensures clinical transparency and usability.

By prioritizing explainability and generalization alongside predictive performance, this work contributes to the development of more trustworthy and clinically applicable AI systems for lung cancer detection. Looking forward, the framework can be extended to larger and more diverse datasets, incorporating multi-modal imaging and genomic information, and further optimized for real-time deployment in clinical environments. Ultimately, the proposed system represents a promising step toward building reliable AI-based diagnostic support that can improve early detection, enhance treatment planning, and foster greater clinical adoption of AI in healthcare.

VIII. ACKNOWLEDGEMENT

We appreciate the assistance of the Department of Computer Science and Engineering at the Thapar Institute of Engineering Technology in Patiala, Punjab, in making the facility available.

Compliance with ethical standards

Conflict of interest: The authors do not have any conflict of interest.

The datasets used for the current study's analysis are accessible upon request.

REFERENCES

[1] Yash Aggarwal and Neetu Mittal. "Enhancing Lung Cancer Diagnosis Through Convolutional Neural Networks: A Comprehensive Study on Image-Based Detection and Classification". In: ed. by Shukla B. et al. 2024. ISBN: 979-835035035-7. DOI: 10.1109/ICRITO61523.2024.10522127.

[2] Najah M. Alsubaie, David Snead, and Nasir M. Rajpoot. "Tumour Nuclear Morphometrics Predict Survival in Lung Adenocarcinoma". In: *IEEE Access* 9 (2021), pp. 12322–12331. ISSN: 21693536. DOI: 10.1109/ACCESS.2021.3049582.

[3] Qanita Bani Baker et al. "Lung Cancer Survival Time Prediction Using Machine Learning and Deep Learning Techniques". In: ed. by Alsmirat M. et al. 2023, pp. 35–42. ISBN: 979-835033925-3. DOI: 10.1109/IDSTA58916.2023.10317871.

[4] Sarahmad Bhat. *Lung cancer*. https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer. Accessed: 2025-09-08. 2021.

[5] Gresha Bhatia et al. "Detection of Lung Carcinoma using Volumetric Convolution (V-Net)". In: 2022. ISBN: 978-166549499-1. DOI: 10.1109/INCET54531.2022.9825358.

[6] Debnath Bhattacharyya et al. "Lung Cancer Segmentation with Three-Parameter Logistic Type Distribution Model". In: *Computers, Materials and Continua* 75.1 (2023), pp. 1447–1465. DOI: 10.32604/cmc.2023.031878.

[7] Debabrata Dansana et al. "Early diagnosis of COVID-19-affected patients based on X-ray and computed tomography images using deep learning algorithm". In: *Soft Computing* 27.5 (2023), pp. 2635–2643. ISSN: 14327643. DOI: 10.1007/s00500-020-05275-y.

[8] Drishti and Jaspreet Singh. "Lung Cancer and Pulmonary Node detection using Deep Learning: A survey". In: 2023. ISBN: 978-166549294-2. DOI: 10.1109/OTCON56053.2023.10113936.

[9] L. Duranti et al. "New Perspectives on Lung Cancer Screening and Artificial Intelligence". In: *Life* 15.3 (2025), p. 498. DOI: 10.3390/life15030498.

[10] Susmitha Valli Gogula et al. "Applying CNN on Lung Images for Screening Initial Cancer Stages". In: 2022, pp. 285–289. ISBN: 978-166549710-7. DOI: 10.1109/ICAAIC53929.2022.9793298.

[11] Ross Gruetzemacher and Ashish Gupta. "Using deep learning for pulmonary nodule detection diagnosis". In: 2016.

[12] Dua Hisam and Enes Hisam. "Deep learning models for classifying cancer and COVID-19 lung diseases". In: 2021. ISBN: 978-166543405-8. DOI: 10.1109/ASYU52992.2021.9598993.

[13] Hexiuli Huang et al. "Diagnosis of Lung Cancer Based on CT Scans Using Convolutional Neural Networks". In: 2022, pp. 338–341. ISBN: 978-166545470-4. DOI: 10.1109/ICDACAI57211.2022.00073.

[14] Daniel Perez Ibanez et al. "Deep learning for pulmonary nodule CT image retrieval - An online assistance system for novice radiologists". In: ed. by Gottumukkala R. et al. Vol. 2017-November. 2017, pp. 1112–1121. ISBN: 978-153861480-8. DOI: 10.1109/ICDMW.2017.158.

[15] Adolfo Jara-Gavilanes, Remigio Hurtado-Ortiz, and Stefania Guzman-Ortiz. "Lung Cancer Detection Using Positron Emission Tomography Images Through Convolutional and Recurrent Neural Networks". In: 2023. ISBN: 979-835031887-6. DOI: 10.1109/CLEI60451.2023.10346202.

[16] Dewan Ziaul Karim and Tasfia Anika Bushra. "Detecting Lung Cancer from Histopathological Images using Convolution Neural Network". In: vol. 2021-December. 2021, pp. 626–631. ISBN: 978-166549532-5. DOI: 10.1109/TENCON54134.2021.9707242.

[17] Hassan et al Lemjabbar-Alaoui. *Lung cancer: Biology and treatment options*. https://pmc.ncbi.nlm.nih.gov/articles/PMC4663145//. [] 2015.

[18] Juliana Li. "Deep Learning Modeling and Increasing Interpretability of Lung Nodule Classification". In: 2024. ISBN: 979-835037115-4. DOI: 10.1109/ECAI61503.2024.10607434.

[19] Sozan Abdullah Mahmood and Hunar Abubakir Ahmed. "An improved CNN-based architecture for automatic lung nodule classification". In: *Medical and Biological Engineering and Computing* 60.7 (2022), pp. 1977–1986. ISSN: 01400118. DOI: 10.1007/s11517-022-02578-0.

[20] Md. Tareq Mahmud et al. "Leveraging Deep Object Detection Models for Early Detection of Cancerous Lung Nodules in Chest X-Rays". In: *Lecture Notes in Networks and Systems* 935 LNNS (2024). Ed. by Silhavy R. and Silhavy P., pp. 79–98. ISSN: 23673370. DOI: 10.1007/978-3-031-54820-8_9.

[21] Sabilla Halimatus Mahmud, Indah Soesanti, and Rudy Hartanto. "Deep Learning Techniques for Lung Cancer Detection: A Systematic Literature Review". In: ed. by Dahlan A., Pristyanto Y., and Aziza P.F.A. 2023, pp. 200–205. ISBN: 979-835031563-9. DOI: 10.1109/ICOIACT59844.2023.10455848.

[22] Pierre P. Massion et al. "Assessing the accuracy of a deep learning method to risk stratify indeterminate pulmonary nodules". In: *American Journal of Respiratory and Critical Care Medicine* 202.2 (2020), pp. 241–249. ISSN: 1073449X. DOI: 10.1164/rccm.201903-0505OC.

[23] Patrice Monkam et al. "Detection and Classification of Pulmonary Nodules Using Convolutional Neural Networks: A Survey". In: *IEEE Access* 7 (2019), pp. 78075–78091. ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2920980.

[24] Tushar Nayak, Niranjana Sampathila, and Nitila Gokulkrishnan. "Processing and Detection of Lung and Colon Cancer from Histopathological Images using Deep Residual Networks". In: 2023. ISBN: 979-835033439-5. DOI: 10.1109/CONECCT57959.2023.10234757.

[25] Jatin Bedi Niyaz Ahmad Wani Ravinder Kumar. *DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence, Computer Methods and Programs in Biomedicine,* https://doi.org/10.1016/j.cmpb.2023.107879/. 2024.

[26] Janice Oentoro et al. "Machine Learning Implementation in Lung Cancer Prediction - A Systematic Literature Review". In: 2023, pp. 435–439. ISBN: 978-166545645-6. DOI: 10.1109/ICAIIC57133.2023.10067128. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85152092982&doi=10.1109%2fICAIIC57133.2023.10067128&partnerID=40&md5=80b694e4ad70249d80a94eaebba66a4c.

[27] Amy Rafferty, Rishi Ramaesh, and Ajitha Rajan. *Explainability Through Human-Centric Design for XAI in Lung Cancer Detection*. https://arxiv.org/abs/2505.09755/. 2025.

[28] Janjhyam Venkata Naga Ramesh et al. "Application of Several Transfer Learning Approach for Early Classification of Lung Cancer". In: *EAI Endorsed Transactions on Pervasive Health and Technology* 10 (2024). ISSN: 24117145. DOI: 10.4108/eetpht.10.5434.

[29] Abeer Mohammed Shanshool, Mariam Bouchakwa, and Ikram Amous-Ben Amor. "Improved accuracy of pulmonary nodule classification on LIDC-IDRI dataset using deep learning". In: vol. 225. 2023, pp. 394–403. DOI: 10.1016/j.procs.2023.10.024.

[30] Manu Sharma, Jignesh S. Bhatt, and Manjunath V. Joshi. "Early detection of lung cancer from CT images: Nodule segmentation and classification using deep learning". In: ed. by Zhou J. et al. Vol. 10696. 2018. ISBN: 978-151061941-8. DOI: 10.1117/12.2309530.

[31] Praveen Tumuluru et al. "Combining Multi-Features for Lung Cancer Detection in Computed Tomography Images: A Feature Fusion Model". In: 2023, pp. 627–632. ISBN: 978-166549199-0. DOI: 10.1109/ICSCDS56580.2023.10104741.

[32] Shengsheng Wang et al. "Deep 3D multi-scale dual path network for automatic lung nodule classification". In: *International Journal of Biomedical Engineering and Technology* 39.2 (2022), pp. 149–169. ISSN: 17526418. DOI: 10.1504/IJBET.2022.124016.

[33] Xi Wang et al. "Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis". In: *IEEE Transactions on Cybernetics* 50.9 (2020), pp. 3950–

3962. ISSN: 21682267. DOI: 10.1109/TCYB.2019.2935141.

[34] Katherine A. Zukotynski et al. “Clinical Applications of Artificial Intelligence in Positron Emission Tomography of Lung Cancer”. In: *PET Clinics* 17.1 (2022), pp. 77–84. ISSN: 15568598. DOI: 10.1016/j.cpet.2021.09.001.