



Multimodal AI framework for lung cancer diagnosis: Integrating CNN and ANN models for imaging and clinical data analysis



Emir Oncu ^{a,b,*} , Fatih Ciftci ^{b,c,d,**}

^a Department of Bioengineering, Faculty of Chemical and Metallurgical Engineering, Yildiz Technical University, Istanbul, 34210, Turkey

^b BioriginAI Research Group, Department of Biomedical Engineering, Fatih Sultan Mehmet Vakif University, Istanbul, 34015, Turkey

^c Faculty of Engineering, Department of Biomedical Engineering, Fatih Sultan Mehmet Vakif University, Istanbul, 34015, Turkey

^d Department of Technology Transfer Office, Fatih Sultan Mehmet Vakif University, Istanbul, 34015, Turkey

ARTICLE INFO

Keywords:

Lung cancer detection
Convolutional Neural Networks (CNN)
Artificial Neural Networks (ANN)
Medical image analysis
MRI imaging
Clinical data integration
Early diagnosis
Lung cancer subtypes
Image classification
Machine learning in healthcare
Tumor detection
Personalized treatment strategies

ABSTRACT

Lung cancer remains a leading cause of cancer-related mortality worldwide, emphasizing the critical need for accurate and early diagnostic solutions. This study introduces a novel multimodal artificial intelligence (AI) framework that integrates Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs) to improve lung cancer classification and severity assessment. The CNN model, trained on 1019 preprocessed CT images, classified lung tissue into four histological categories, adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal, with a weighted accuracy of 92 %. Interpretability is enhanced using Gradient-weighted Class Activation Mapping (Grad-CAM), which highlights the salient image regions influencing the model's predictions. In parallel, an ANN trained on clinical data from 999 patients—spanning 24 key features such as demographic, symptomatic, and genetic factors—achieves 99 % accuracy in predicting cancer severity (low, medium, high). SHapley Additive exPlanations (SHAP) are employed to provide both global and local interpretability of the ANN model, enabling transparent decision-making. Both models were rigorously validated using k-fold cross-validation to ensure robustness and reduce overfitting. This hybrid approach effectively combines spatial imaging data and structured clinical information, demonstrating strong predictive performance and offering an interpretable and comprehensive AI-based solution for lung cancer diagnosis and management.

1. Introduction

Lung cancer accounts for 21 % of all cancer-related fatalities globally and it is the second most frequent malignancy in all genders [1]. Lung cancer has a high death rate when compared to all other cancer types, and early diagnosis is still difficult despite modern screening techniques. Only about 20 % of cases are diagnosed at stage I. Statistics demonstrate that they have remained unchanged for years, highlighting ongoing challenges clinicians face in prognosis assessment and treatment planning [2]. Despite significant advancements in immunotherapy and targeted treatments for lung cancer, their efficacy remains unpredictable, with response rates varying among patients [3]. As a result, early detection of lung cancer requires the use of extremely sensitive and specific diagnostic instruments [4]. In 2017, multiple studies [5] highlighted the importance of analyzing the morphological characteristics of

pathological slides for accurate diagnosis and prognosis of lung cancer [6]. These findings underscored the crucial role that computer-aided image analysis plays in enhancing the accuracy and effectiveness of lung cancer prognosis.

Adenocarcinoma, the most common subtype of non-small cell lung cancer (NSCLC), accounts for approximately 40 % of all lung cancer cases [7]. It primarily originates in the peripheral regions of the lungs and is strongly associated with smoking, though it can also occur in non-smokers. This subtype is distinguished by glandular differentiation and mucin production, both of which are observable in histopathological evaluations. Adenocarcinoma frequently progresses asymptotically in its initial stages, leading to delayed diagnosis and subsequent treatment challenges [7].

Squamous cell carcinoma (SCC), a major subtype of NSCLC, accounts for 25–30 % of lung cancer cases [8]. Typically arising in the central

* Corresponding author. BioriginAI Research Group, Department of Bioengineering, Faculty of Chemical and Metallurgical Engineering, Yildiz Technical University, Istanbul, 34210, Turkey.

** Corresponding author. BioriginAI Research Group, Department of Biomedical Engineering, Fatih Sultan Mehmet Vakif University, Istanbul, 34015, Turkey.

E-mail addresses: biomedical.emr@gmail.com (E. Oncu), faciftcii@gmail.com, fciftci@fsm.edu.tr (F. Ciftci).

regions of the lungs, particularly the bronchi, squamous cell carcinoma is strongly associated with a history of smoking. Histologically, it is characterized by the presence of keratin pearls and intercellular bridges, indicating its origin from the squamous epithelial cells lining the respiratory tract [9].

Large cell carcinoma (LCC) is a less common subtype of NSCLC, accounting for approximately 10–15 % of lung cancer cases. Under microscopic examination, it is characterized by large, undifferentiated cells that lack the glandular features of adenocarcinoma and the squamous characteristics of squamous cell carcinoma [10]. LCC can occur in any part of the lung but is most frequently found in the peripheral regions. Clinically, large cell carcinoma is aggressive, exhibiting rapid growth and early metastasis, which significantly contributes to its poor prognosis [11]. Patients often exhibit nonspecific symptoms, including cough, chest pain, and weight loss, which can contribute to delayed diagnosis [12].

Given the challenges associated with diagnosing and treating aggressive cancers such as large cell carcinoma (LCC), there is increasing interest in integrating advanced technologies to improve clinical outcomes. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), has shown significant potential in revolutionizing cancer diagnostics and treatment [13]. AI is broadly classified into two main domains: virtual and physical. The virtual domain is further subdivided into ML and DL. [14]. ML refers to a system's ability to independently learn from data without explicit programming [15]. DL, a specialized subset of ML, employs multiple layers simultaneously to perform feature extraction and model optimization [16]. Machine learning has revolutionized multiple fields, including natural language processing, autonomous vehicles, healthcare, and image recognition. It excels at efficiently processing complex datasets, identifying hidden patterns beyond human perception, and delivering highly accurate predictions [17–19].

ML is centered around advanced neural network architectures, including Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs), which are designed to process and analyze complex data patterns [20]. CNNs are a specialized type of ANNs designed to process and analyze visual data, such as images. Unlike traditional ANNs, CNNs utilize convolutional layers that apply filters (kernels) to extract spatial and hierarchical features, such as edges, textures, and shapes [21]. These convolutional layers are followed by pooling layers,

which reduce the spatial dimensions of the data, enhancing computational efficiency and helping to prevent overfitting [22]. CNNs excel in tasks such as image recognition, object detection, and medical imaging due to their ability to capture spatial relationships and intricate patterns within data [23]. ANNs are inspired by the structure and function of the human brain, consisting of layers of interconnected nodes (neurons) that process and transmit information [24]. These nodes process input data by applying weights, biases, and activation functions, which allow the network to learn from data and make predictions.

Prior studies have demonstrated that ANNs can effectively analyze patient-reported symptoms and medical history to predict lung cancer presence with promising results [25]. Table 1 shows the previous studies about AI aided lung cancer detection. No prior work has effectively combined ANN models trained on clinical data with CNN models trained on MRI/CT scans to enhance the reliability of AI-driven lung cancer detection directly.

This paper introduces a novel multimodal AI framework that enhances lung cancer detection and classification by integrating CNN and ANNs. By combining imaging data from CT scans with structured clinical patient data, this approach captures both spatial tumor characteristics and patient-specific risk factors, addressing the challenges of early diagnosis and accurate prognosis of lung cancer subtypes, including adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. The CNN model specializes in extracting radiological features critical for tumor identification, while the ANN model analyzes demographic, symptomatic, and genetic variables to assess cancer severity. This integration not only improves classification precision but also enhances interpretability, allowing for a more transparent and clinically relevant diagnostic process. By leveraging both imaging and clinical data, the proposed framework offers a more comprehensive assessment of lung cancer, facilitating personalized treatment strategies and advancing AI-assisted decision-making in oncology.

2. Methods

The following sections detail the datasets, preprocessing techniques, model architecture, and evaluation metrics employed in this study. By integrating image-based and clinical data-driven AI models, this study presents a robust diagnostic framework that enhances traditional lung cancer detection methods. The CNN model excels in extracting spatial

Table 1
Summary of the previous studies.

Study	Objective	Dataset	Techniques	Results
[26]	To accurately identify four types of lung conditions (adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal) using CT scan images.	Kaggle chest CT-scan images dataset	Modified DenseNet201 with additional layers; feature selection; ML classifiers; evaluated using confusion matrix, ROC, MCC, Kappa, p-value, and 5-fold cross-validation	Achieved 100 % maximum accuracy, 95 % average accuracy: p-value <0.001
[27]	Developing a computer-aided detection (CADe) system for early detection of lung nodules from low-dose CT images.	I-ELCAP (International Early Lung Cancer Action Project), 320 LDCT images from 50 subjects	Preprocessing to enhance contrast; feature extraction using AlexNet, VGG16, and VGG19; feature optimization with Genetic Algorithm; classification using multiple classifiers including SVM	Best accuracy of 96.25 %, sensitivity of 97.5 %, and specificity of 95 % using VGG19 + SVM
[28]	Developing an interpretable AI model for lung cancer detection that provides both accurate predictions and clear explanations.	Survey Lung Cancer (open-source dataset)	Hybrid model called "DeepExplainer" combining CNN for feature extraction, XGBoost for classification, and SHAP for prediction explainability	Achieved 97.43 % accuracy, 98.71 % sensitivity, and 98.08 F1-score; provided local and global explanations for predictions using SHAP
[29]	To develop and optimize a CNN-based model for accurate detection of abnormalities in chest X-ray (CXR) images.	4538 CXR images	CNN with hyperparameter optimization (Adam, RMSprop; tuning pooling, convolutional, dropout layers, target size, and epochs)	Achieved best accuracy of 97.94 % through testing 32 model variations; results validated using a confusion matrix
[30]	To develop a fast, non-invasive CNN-based model for accurate lung cancer stage classification	IQ-OTHNCCD Lung Cancer CT Scan Dataset	CNN with preprocessing (resizing, normalization, Gaussian blurring) and SMOTE for class balancing	Achieved 99.64 % accuracy: precision, recall, and F1-score all above 98 %
[31]	Enhancing CNN performance for medical image segmentation by applying image preprocessing techniques to improve anatomical structure detection.	Lung CT-Scan (267 images) and Chest X-ray (3616 images) with ground truth	CNN with image enhancement using Histogram Equalization (HE), CLAHE, and hybrid approaches (CLAHE-HE, HE-CLAHE)	CLAHE-HE approach improved segmentation accuracy by up to 3.22 % on Lung CT-Scan and 0.96 % on Chest X-rays; achieved DSC ≈ 0.92 and SSIM ≈ 0.97

patterns from CT images, while the ANN provides accurate predictions based on clinical attributes. Together, these models highlight the versatility of deep learning techniques in improving diagnostic accuracy and facilitating more precise lung cancer classification.

2.1. Data information

We developed two distinct AI models using a combination of image-based and clinical data to predict lung cancer status. In this study, we developed two distinct AI models to predict lung cancer types and assess their potential malignancy. The first model employed a CNN trained on 1019 CT images, and the second utilized an ANN trained on clinical data from 999 patients. The primary aim was to design an integrated diagnostic approach capable of accurately classifying lung cancer into four categories: adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal tissue.

The CT dataset consisted of 1019 (obtained from two different publicly available datasets on Kaggle) labeled images divided into these four classes. Each image underwent preprocessing to ensure uniformity and compatibility with the CNN model, including resizing to a standard resolution of 256x256 pixels and normalization to enhance computational efficiency while preserving key diagnostic features. CNN predicts the class of the uploaded image along with a probability score for each category, providing valuable insights into the likelihood of a specific diagnosis. Additionally, Grad-CAM was implemented to generate heatmap overlays on the input images, highlighting regions critical to the model's predictions. These heatmaps offer visual interpretability, making the system more transparent and clinically applicable.

The second phase of the study involved an ANN trained on clinical data collected from 999 patients (obtained from one publicly available dataset on Kaggle). The dataset consisted of 24 features, capturing various demographic, lifestyle, genetic, and symptomatic factors associated with lung cancer. These features included:

- **Age and Gender:** Basic demographic variables critical for understanding population-specific cancer risks.
- **Air Pollution and Smoking:** Indicators of environmental and behavioral exposure to carcinogens, with smoking being a primary risk factor for lung cancer.
- **Alcohol Use:** Reflects the patient's alcohol consumption, which may compound the risk of cancer in smokers.
- **Dust Allergy and Occupational Hazards:** Variables assessing exposure to dust particles or workplace environments with known lung irritants or carcinogens.
- **Genetic Risk and Chronic Lung Disease:** Capture hereditary predisposition and pre-existing respiratory conditions that increase cancer susceptibility.
- **Balanced Diet and Obesity:** Reflect the role of nutrition and body mass index in overall health and immune response.
- **Passive Smoker:** Accounts for indirect exposure to second-hand smoke, an established risk factor for lung cancer.
- **Chest Pain, Coughing of Blood, Fatigue, and Weight Loss:** Common symptoms associated with lung cancer progression.
- **Shortness of Breath, Wheezing, and Swallowing Difficulty:** Respiratory symptoms indicative of tumor obstruction or advanced disease.
- **Clubbing of Fingernails:** A physical manifestation often linked to chronic hypoxia in lung diseases, including cancer.
- **Frequent Cold and Dry Cough:** Represent the recurring respiratory issues that could signal early lung disease.
- **Snoring:** Included as a potential indirect factor linked to airway obstruction or respiratory abnormalities.
- **Level:** Represents the severity or progression stage of lung cancer, providing additional context for diagnosis.

These features were standardized and preprocessed to ensure

compatibility with the ANN model. By integrating demographic, symptomatic, and lifestyle data, the ANN was designed to classify patients into potential lung cancer categories. This dual-model approach combining CNN for imaging and ANN for clinical data analysis presents a robust and comprehensive diagnostic framework for lung cancer detection and classification. The goal was to create a multimodal approach that integrates different diagnostic techniques for more accurate and comprehensive predictions.

2.2. Machine learning models

A CNN was developed to classify lung cancer from CT images into four categories: Class 0 (squamous cell carcinoma), Class 1 (normal), Class 2 (adenocarcinoma), and Class 3 (large-cell carcinoma). The model was implemented using Python and the Keras deep learning library. To preserve critical spatial and contextual information vital for medical image interpretation, images were processed in their 2D format rather than being flattened into 1D arrays.

Preprocessing involved resizing all input images to 256×256 pixels and normalizing pixel intensity values to the range [0, 1], standardizing them for CNN input and ensuring compatibility across batches. To enhance generalization, the dataset was augmented using random horizontal and vertical flipping, as well as rotation, creating varied representations of each class and mitigating the risk of overfitting—particularly important given the limited size of annotated medical datasets.

The CNN architecture (illustrated in Fig. 1) consisted of five convolutional blocks. Each block included a Conv2D layer activated by ReLU, followed by max pooling for dimensionality reduction and batch normalization to stabilize learning and accelerate convergence. These layers allowed the model to progressively extract hierarchical features from the image data. The extracted features were then passed through fully connected layers, with dropout regularization applied to mitigate overfitting. The final layer used a SoftMax activation to output class probabilities for the four target categories.

Model training was conducted using the Adam optimizer, chosen for its adaptive learning rate capabilities and robustness across a wide range of deep learning tasks. The loss function employed was sparse categorical cross-entropy, suitable for multi-class classification where target labels are integer-encoded. Training was performed over 75 epochs per fold with a batch size of 32, balancing training stability and computational efficiency.

To ensure robust performance evaluation and mitigate data bias, a 5-fold cross-validation strategy was used. The entire dataset was shuffled with a fixed seed to ensure reproducibility and then split into five equal parts. In each iteration, 80 % of the data was used for training, while the remaining 20 % was split equally between validation and testing. This rotation allowed every sample to be used in both training and evaluation phases, enhancing generalizability assessments.

Performance was evaluated using accuracy, precision, recall, and F1-score, calculated per fold and averaged to obtain overall model metrics (Table 1). Confusion matrices were used to visualize the alignment between predicted and true labels. Additionally, training and validation accuracy/loss curves were plotted for each fold to monitor model learning dynamics, detect overfitting or underfitting patterns, and inform potential hyperparameter tuning.

TP (True Positives): Correctly predicted positive cases, TN (True Negatives): Correctly predicted negative cases, FP (False Positives): Incorrectly predicted positive cases, and FN (False Negatives): Incorrectly predicted negative cases.

The ANN model in this study was developed as part of a hybrid diagnostic framework aimed at predicting lung cancer severity from structured clinical data. The dataset consisted of 24 input features, including demographic details, behavioral risk factors (e.g., smoking history), genetic predispositions, and symptom profiles. All input features were standardized using z-score normalization to facilitate stable

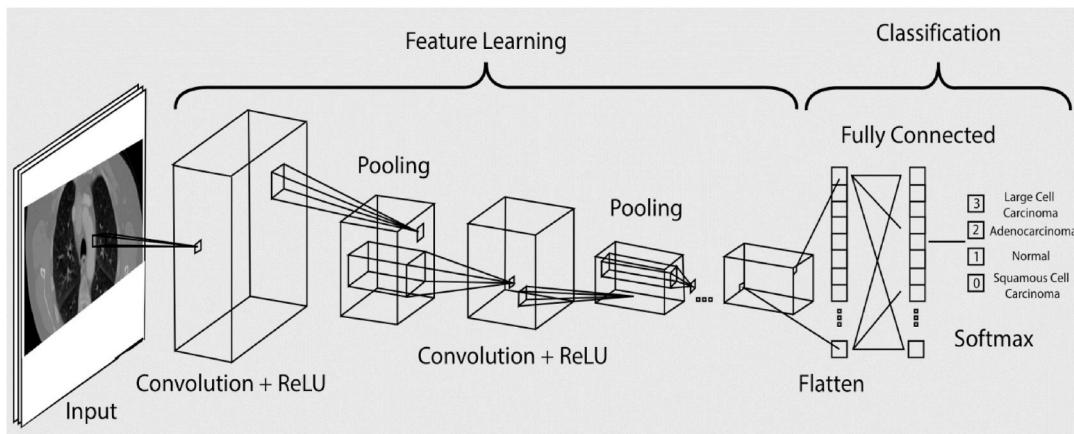


Fig. 1. Graphical representation of the CNN architecture.

and efficient training.

The ANN model was designed as a lightweight, interpretable classifier for three severity levels: Low (0), Medium (1), and High (2). Its architecture followed a fully connected feedforward design, consisting of an input layer for 24 features, a single hidden layer with 24 neurons and ReLU activation, and an output layer with three units using SoftMax activation to support multi-class classification. The model was optimized using the Adam optimizer and sparse categorical cross-entropy loss, which provided effective weight updates for multi-class predictions.

While the ANN architecture is intentionally kept simple for clinical interpretability, this study advances prior work by integrating rigorous model explainability techniques. Specifically, SHAP (SHapley Additive exPlanations) was employed to interpret the model's outputs both globally and locally. SHAP enabled clear identification of the most influential clinical features for each severity class, thus supporting real-world clinical decision-making by highlighting patient-specific risk drivers. The ANN was evaluated using standard metrics including accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC). A confusion matrix and multi-class ROC-AUC curves were also analyzed to validate the robustness and discrimination of the model across classes.

Although the model does not introduce a new ensemble architecture or optimization algorithm, its novelty lies in its modular integration with a CNN-based image classifier and its emphasis on interpretable clinical AI. As illustrated in Fig. 2, the overall system first uses a CNN to classify lung CT images. If cancer is detected, the ANN is triggered to assess the severity level using the patient's clinical data. This two-stage pipeline reflects a real-world diagnostic workflow and enables both image-based detection and data-driven risk stratification.

This approach demonstrates how a combination of deep learning models—despite being standard individually—can be orchestrated to deliver personalized, explainable insights into both lung cancer detection and severity classification. Future work may explore ensemble strategies, feature-level fusion, or hybrid attention mechanisms to further enhance the performance and adaptability of this framework.

3. Results

The results of this study provide a comprehensive evaluation of the developed models, demonstrating their performance across a range of critical metrics and visualizations. These analyses encompass key aspects such as classification accuracy, interpretability, and reliability, highlighting the models' effectiveness in handling complex datasets. Through a detailed exploration of training dynamics, precision-recall characteristics, and class-wise discrimination capabilities, the findings underscore the robustness and clinical relevance of the proposed

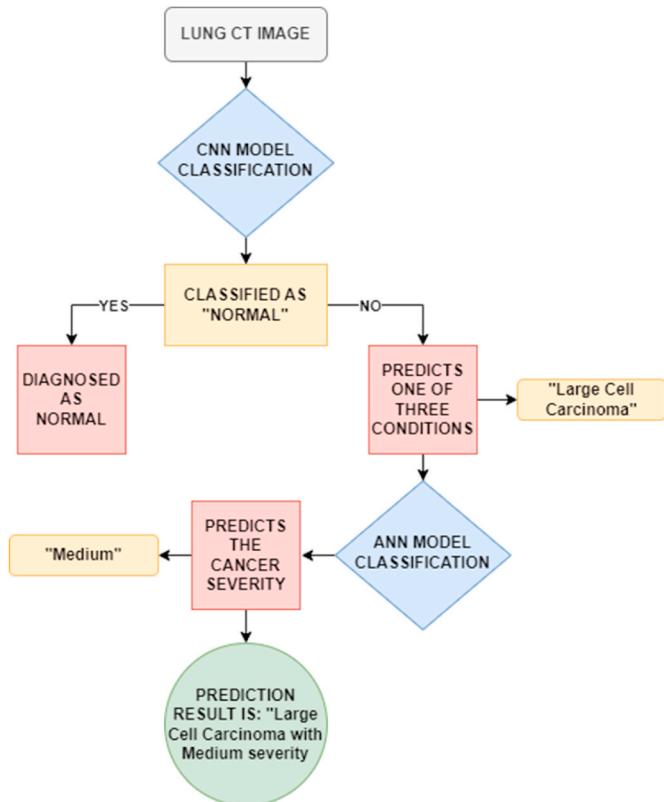


Fig. 2. Hybrid deep learning framework for lung cancer classification and severity prediction.

systems. The figures and tables presented in this section serve to substantiate these outcomes, offering valuable insights into the models' strengths and areas for potential improvement.

Fig. 3 illustrates the training and validation accuracy, as well as the corresponding loss, over 75 epochs. The accuracy graph demonstrates a steady increase in training accuracy, ultimately reaching above 90 %, while validation accuracy exhibits fluctuations but follows an overall upward trend, indicating generalization capability. The loss graph shows a consistent decline in both training and validation loss, suggesting effective learning. However, the variations in validation accuracy and loss indicate some degree of instability, potentially due to dataset variability or model sensitivity. These results confirm that the model is effectively learning patterns in the data while maintaining generalization, supporting its robustness for lung cancer classification.

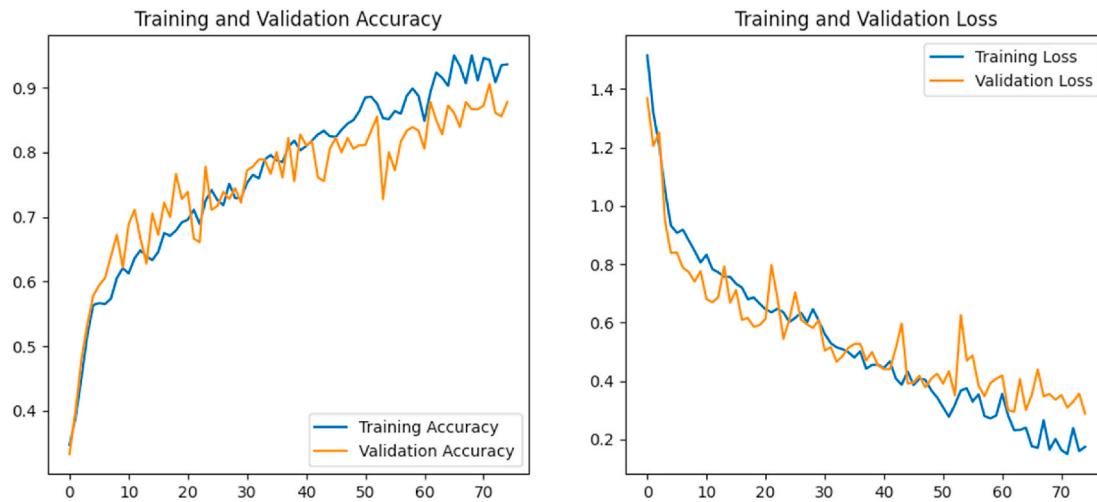


Fig. 3. Training and validation accuracy - loss curves over 75 epochs with K-fold = 5.

The classification performance of the model is further validated using Receiver Operating Characteristic (ROC) curves, as depicted in Fig. 4. The ROC curves for each class are presented, with the areas under the curve (AUC) values of 0.93 for adenocarcinoma, 0.98 for large cell carcinoma, 1.00 for normal, and 0.94 for squamous cell carcinoma. These results demonstrate the model's excellent discrimination capabilities across all categories. The near-perfect AUC values for multiple classes underscore its efficacy in differentiating between malignant and normal cases, supporting its application in real-world clinical scenarios. The performance metrics for the CNN are detailed in Table 2. The model achieved a weighted average accuracy of 87.78 %, with a Matthews Correlation Coefficient (MCC) of 0.84.27. These metrics suggest that the model performs well overall, though its ability to differentiate between certain classes remains limited, as indicated by the ROC AUC scores.

To evaluate the robustness and generalizability of the proposed lung

Table 2

Summary of performance metrics used in the study and their calculation methods.

Metrics	Calculation
Accuracy [32]	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision [33]	$\frac{TP}{TP + FP}$
Recall [34]	$\frac{TP}{TP + FN}$
F1-Score [34]	$\frac{Precision \times Recall}{Precision + Recall}$
ROC AUC [35]	$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + FN}$
Matthews Correlation Coefficient [35]	$\frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(FP + FN)(TN + FN)}}$

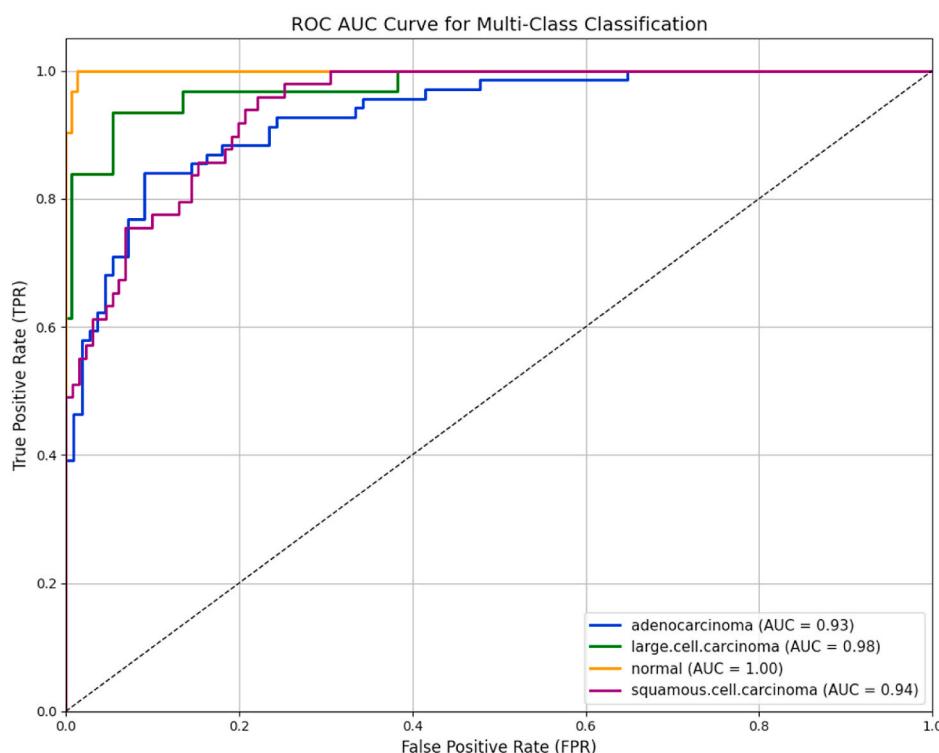


Fig. 4. The ROC curves for each CNN Model's class.

cancer detection model, a 5-fold cross-validation strategy was employed, with results summarized in Fig. 5. This technique partitions the dataset into five equally sized subsets, iteratively training the model on four folds while validating on the remaining one, thereby ensuring comprehensive utilization of the data and mitigating potential bias due to random sampling. The model demonstrated consistent performance across all folds, achieving validation accuracies ranging from approximately 75 %–89 %, with an overall trend indicating stable predictive capability. The highest accuracy was observed in Fold 5, reflecting optimal convergence and model generalization in that subset, while the minor variability across folds suggests sensitivity to data distribution but no signs of overfitting. These findings validate the reliability of the model in detecting diverse histological patterns of lung cancer under varying training conditions, reinforcing its potential for deployment in real-world clinical settings where dataset heterogeneity is inevitable.

Table 2 presents the classification performance metrics, including precision, recall, F1-score, and support for each class. The model achieved an overall accuracy of 92 %, demonstrating high reliability in lung cancer classification. Among individual classes, Adenocarcinoma attained a precision of 0.89 but had a slightly lower recall of 0.88, indicating some misclassification. Large Cell Carcinoma exhibited balanced precision and recall at 0.86, suggesting stable detection. Normal cases were classified with perfect recall (1.00) and a high F1-score of 0.98, showing the model's ability to correctly identify non-cancerous cases. Squamous Cell Carcinoma also achieved perfect recall, reinforcing its distinguishability. The macro and weighted averages of precision, recall, and F1-score remain consistently high, confirming the model's robustness across different lung cancer subtypes.

Fig. 6 illustrates the confusion matrix reflecting the classification performance of the proposed model across four histopathological classes: Adenocarcinoma (AC), Large Cell Carcinoma (LCC), Normal, and Squamous Cell Carcinoma (SCC). The model demonstrated high discriminative capability, particularly for the Normal and SCC categories, both of which exhibited near-perfect classification, with 100 % accuracy for Normal (31/31) and 50 out of 56 correct predictions for SCC. AC was correctly identified in 53 out of 60 instances, with a degree of misclassification primarily into the SCC category, suggesting morphological overlap between these tumor types that may challenge boundary distinctions. Similarly, LCC cases were accurately predicted in 24 out of 32 instances, with notable misclassification into AC, indicating

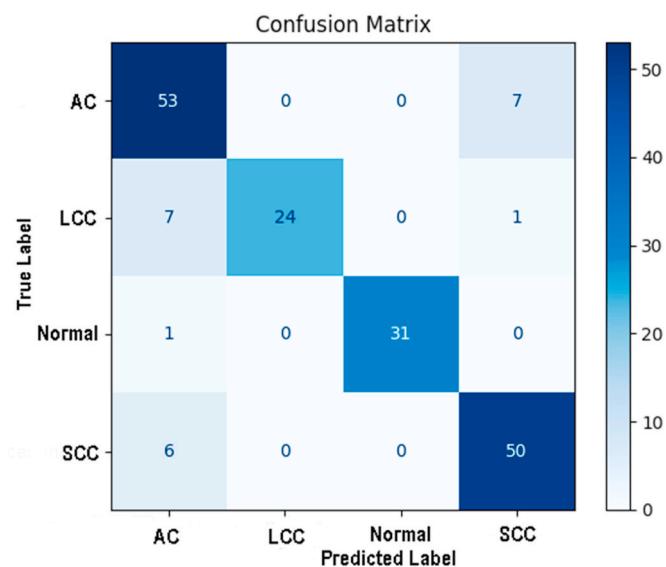


Fig. 6. Confusion matrix of CNN Model's classes.

potential feature space intersection.

Fig. 7 presents representative CT slices from correctly classified samples of Large Cell Carcinoma (LCC) across different test instances within Fold 5 of cross-validation, visually reinforcing the model's predictive consistency for this class. Despite the inherent heterogeneity and relatively lower prevalence of LCC compared to other lung cancer subtypes, the model successfully identified key imaging signatures indicative of LCC, as evidenced by uniform classification outcomes across all six samples shown. The consistent agreement between ground truth and predicted labels, visually annotated in each subpanel, underscores the model's capacity to generalize across inter-patient anatomical variations and varying tumor presentations. This performance suggests that the convolutional layers effectively captured radiographic features unique to LCC, such as peripheral localization, irregular margins, and associated parenchymal distortion, which are often subtle and challenging even for expert radiologists.

Fig. 8 highlights the Grad-CAM (Gradient-weighted Class Activation

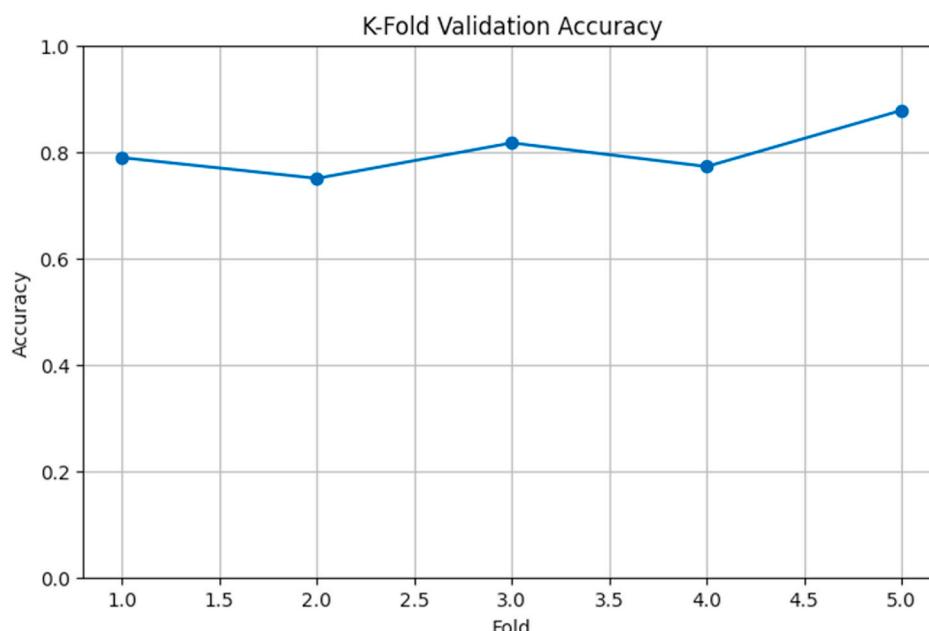


Fig. 5. Five-fold cross-validation accuracy for lung cancer classification model.

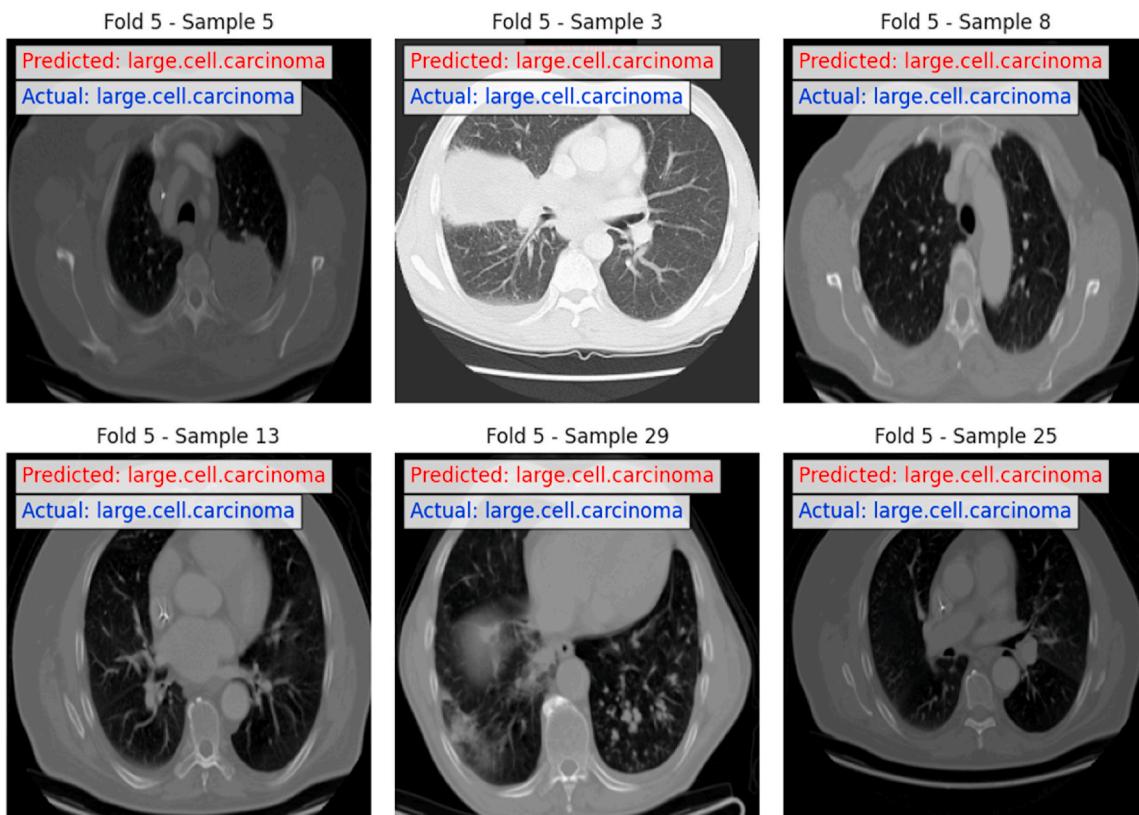


Fig. 7. Example of predicted images of large cell carcinoma class.

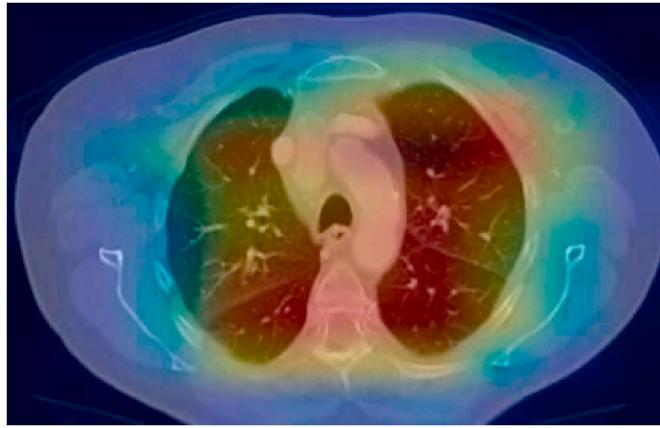


Fig. 8. Heatmap mask of predicted image.

Mapping) visualization, which identifies the regions of the CT image that contribute most significantly to the model's predictions. This visualization serves as an essential interpretability tool, offering insights into the CNN's decision-making process. By localizing critical features in the image, such as tumor regions, Grad-CAM masking enhances the model's transparency and trustworthiness, an essential aspect in clinical AI applications.

Fig. 9 illustrates the training dynamics of the proposed ANN model over 25 epochs, providing a comprehensive view of its learning progression in terms of both accuracy and loss. As observed in the left panel, the model exhibits a steady and substantial increase in both training and validation accuracy, with the validation curve surpassing 95 % accuracy by epoch 10 and maintaining a near-saturation plateau, thereafter, indicating strong generalization performance. Simultaneously, the training accuracy demonstrates a progressive ascent, reaching

convergence close to the validation trend without abrupt fluctuations, thereby suggesting stability in learning without overfitting. The corresponding loss curves (right panel) show a monotonic decrease for both training and validation sets, further confirming that the model effectively minimizes the prediction error without signs of overfitting. Notably, the validation loss remains consistently lower than the training loss across epochs, a phenomenon that may be attributed to regularization effects or batch normalization contributing to smoother gradient flow.

Fig. 10 presents the multi-class Receiver Operating Characteristic (ROC) curves for the proposed ANN model, showcasing its discriminatory performance across three distinct classes. Each class-specific ROC curve closely adheres to the top-left corner of the plot, indicating near-perfect classification. The Area Under the Curve (AUC) values for all three classes are reported as 1.00, reflecting exceptional model sensitivity and specificity with zero trade-off between true positive and false positive rates. This level of performance suggests that the ANN model is not only highly accurate but also capable of maintaining class-wise balance without bias. The pronounced separation from the diagonal line (representing random guessing) underscores the model's ability to learn highly discriminative feature representations, even in the presence of class imbalance or inter-class overlap.

Fig. 11 illustrates the confusion matrix for the ANN model, capturing its classification performance across three lung cancer-related classes. The model demonstrates exceptional predictive accuracy, correctly identifying 55 out of 60 samples in Class 0, achieving perfect classification for Class 1 with 69 out of 69 correct predictions, and similarly for Class 2 with 71 out of 71 instances classified without error. The only misclassifications observed involve five instances of Class 0 erroneously labeled as Class 1, suggesting minimal overlap between these two classes in feature space. Notably, the complete absence of false positives and false negatives in Classes 1 and 2 underscores the model's ability to learn highly discriminative representations for these categories.

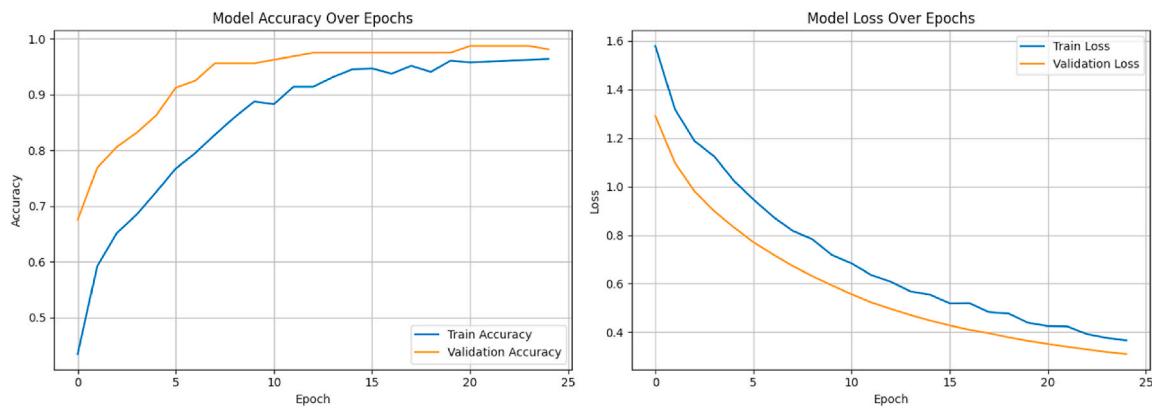


Fig. 9. Training and validation accuracy of ANN model.

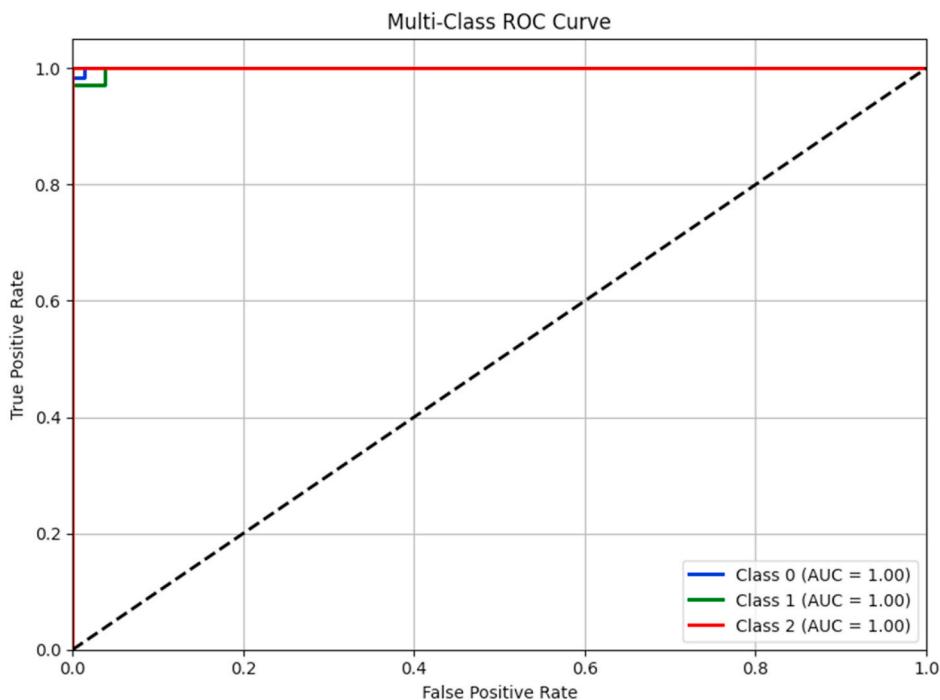


Fig. 10. ROC curves of ANN model for 3 classes.

Table 3 presents the performance metrics of the ANN model, showcasing its high classification capability across three classes. The model achieved an overall accuracy of 97.5 %, accompanied by a macro-averaged precision of 0.977, recall of 0.972, and F1-score of 0.974, indicating strong consistency across all classes regardless of sample imbalance. Class-wise, the ANN demonstrated perfect precision (1.00) for all categories, with recall values of 0.92, 1.00, and 1.00 for Classes 0, 1, and 2, respectively. The slightly lower recall in Class 0 suggests a small number of false negatives, which aligns with the confusion matrix analysis. Importantly, the F1-scores—balancing precision and recall—remained high for all classes (0.96–1.00), further emphasizing the robustness of the model. Both macro and weighted averages confirm that the ANN performs equitably across class distributions. These results underscore the ANN's effectiveness as a reliable tool for multi-class lung cancer classification, demonstrating both sensitivity and specificity critical for real-world clinical deployment (see **Table 4**).

Fig. 12 illustrates the SHAP analysis of the ANN model, providing a transparent interpretation of feature contributions toward the multi-class lung cancer classification. The plot ranks input features by their average impact on the model's output, separately for each class (Class 0,

Class 1, and Class 2). Notably, Coughing of Blood, Wheezing, and Passive Smoker emerged as the top three most influential predictors across all classes, with high SHAP values indicating strong discriminative power in driving model predictions. For Class 0 (olive), these features exerted the greatest influence, underscoring their relevance in distinguishing that subtype. Fatigue, Snoring, and Smoking also showed significant impact, particularly for Class 1 (blue) and Class 2 (pink), suggesting these symptoms are vital indicators for their respective cancer classes. Interestingly, lifestyle and environmental factors such as Air Pollution, Balanced Diet, and Obesity also demonstrated moderate contributions, reinforcing the ANN model's ability to integrate both clinical symptoms and broader risk factors. The comprehensive SHAP analysis validates the ANN's interpretability and emphasizes its clinical relevance by highlighting feature importance in alignment with domain knowledge, thereby bolstering trust in the model's decision-making process for lung cancer subtype prediction.

4. Discussion

The comparative analysis between the CNN and ANN models reveals

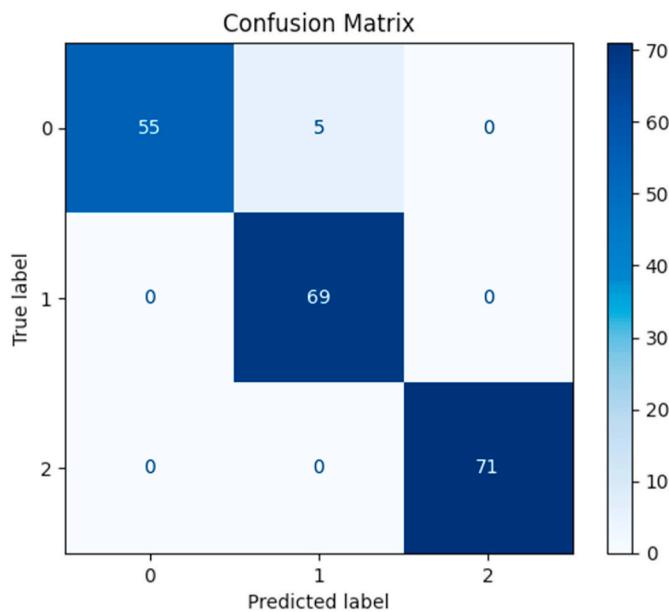


Fig. 11. The confusion matrix of ANN model.

Table 3
CNN Model's classification report with precision, recall, and F1-Scores.

	Precision	Recall	F1-Score	Support
Adenocarcinoma	0.75	0.88	0.83	60
Large Cell Carcinoma	1.00	0.86	0.86	32
Normal	1.00	0.97	0.98	32
Squamous Cell Carcinoma	0.86	0.89	0.88	56
Accuracy	0.92	0.89	0.89	180
Macro Avg.	0.92	0.98	0.89	180
Weighted Avg.	0.89	0.89	0.89	180

Table 4
Performance metrics of ANN model.

	Precision	Recall	F1-Score	Support
Low Level	1.00	0.92	0.96	60
Medium Level	0.93	1.00	0.97	69
High Level	1.00	1.00	1.00	71
Accuracy	0.98	0.97	0.97	200
Macro Avg.	0.98	0.97	0.97	200
Weighted Avg.	0.98	0.97	0.97	200

distinct strengths. The CNN model achieved a weighted average accuracy of 87.78 % and demonstrated high AUC scores across all lung cancer subtypes, including perfect discrimination for normal cases. The ANN model, on the other hand, outperformed the CNN in overall classification, attaining a remarkable 97.5 % accuracy with balanced precision and recall across all classes. Notably, both models exhibited strong generalization, as evidenced by high validation accuracy and stable training dynamics, with no significant signs of overfitting.

However, despite these strengths, both models showed susceptibility to certain classification errors—specifically, false positives and false negatives—which are critical to address in clinical AI. In the CNN model, adenocarcinoma and large cell carcinoma were occasionally misclassified as squamous cell carcinoma and vice versa. These errors may be attributed to overlapping radiographic features among these subtypes, as seen in the confusion matrix where 7 adenocarcinoma samples were misclassified, primarily into SCC. Similarly, for LCC, 8 instances were misclassified, often into AC, highlighting the need for improved feature disentanglement in these regions of the feature space. False positives, especially in classifying normal tissue as malignant, can lead to unnecessary anxiety and invasive diagnostic procedures, while false

negatives pose a greater risk by potentially delaying treatment for actual cancer cases. Therefore, minimizing such misclassifications is vital to enhance the clinical reliability of AI-assisted diagnostics.

In contrast, the ANN model showed near-perfect performance, particularly with no false positives or negatives in two of the three classes. The minor misclassification observed in Class 0—where 5 samples were incorrectly labeled as Class 1—suggests a need for refinement in distinguishing early or subtle pathological changes. Importantly, interpretability tools such as Grad-CAM and SHAP provided valuable insights into model decision-making, with Grad-CAM localizing tumor regions effectively and SHAP highlighting clinically relevant features like coughing blood, wheezing, and passive smoking. These tools not only improve transparency but also support clinicians in understanding model predictions, thereby enhancing trust in AI systems. To enhance the interpretability and transparency of the deep learning models, two complementary visualization techniques were employed: Grad-CAM for the CNN model and SHAP for the ANN model. Fig. 8, through Grad-CAM, highlights the spatial regions within CT scans that contribute most to the CNN's predictions, effectively localizing tumor areas and reinforcing the clinical relevance of the model's learned features. In contrast, Fig. 12 presents a SHAP (SHapley Additive exPlanations) summary plot for the ANN model, showcasing the average impact of each input feature on the model's output across three lung cancer classes. The SHAP analysis identifies Coughing of Blood, Wheezing, and Passive Smoker as the most influential predictors across all classes. Notably, Class 0 (olive) is heavily impacted by Coughing of Blood and Wheezing, while Class 1 (blue) and Class 2 (pink) are more influenced by features such as Fatigue, Snoring, and Smoking. This feature-level transparency not only aligns with known clinical risk factors but also underscores the ANN model's capacity to integrate both symptomatic and environmental variables into its decision-making process. By quantifying the contribution of each feature per class, SHAP supports model accountability and builds trust, which is essential for clinical deployment.

Several studies have reported promising results using CT scan and chest X-ray imaging combined with deep learning models, each addressing unique challenges related to accuracy, interpretability, data limitations, or early detection capabilities. For instance, the study in Ref. [26] proposed a modified DenseNet201 architecture for classifying four types of lung conditions and achieved a remarkable maximum accuracy of 100 % using multiple statistical evaluation metrics. However, the lack of detailed explanation methods for the predictions limits its clinical interpretability, a critical component for real-world deployment. Similarly [27], focused on developing a computer-aided detection (CADe) system using low-dose CT images and achieved 96.25 % accuracy by combining deep features from VGG networks with genetic algorithm-based optimization. While this method demonstrated excellent diagnostic performance, it also did not incorporate interpretability frameworks, which limits its utility in a transparent healthcare setting.

In contrast, interpretability was a central objective in Ref. [28], where the authors introduced a hybrid “DeepExplainer” model integrating CNNs with XGBoost and SHAP for explainable lung cancer detection. The study reported high accuracy (97.43 %) along with strong sensitivity and F1-score and emphasized local and global interpretability. This aligns closely with the direction of our study, which leverages SHAP to provide fine-grained insights into the ANN model's decision process across multiple lung cancer classes. Our approach, however, distinguishes itself by integrating both SHAP and Grad-CAM, offering interpretability at both feature and spatial levels—a dual-modality that enhances transparency in both image and tabular domains. Further [29], optimized CNN architectures for chest X-ray classification, achieving 97.94 % accuracy through rigorous hyperparameter tuning. While comprehensive in model optimization, this study is limited to binary classification and does not address multiclass cancer staging. Our study expands on this by targeting multiclass lung cancer prediction, a more clinically nuanced task, supported by both

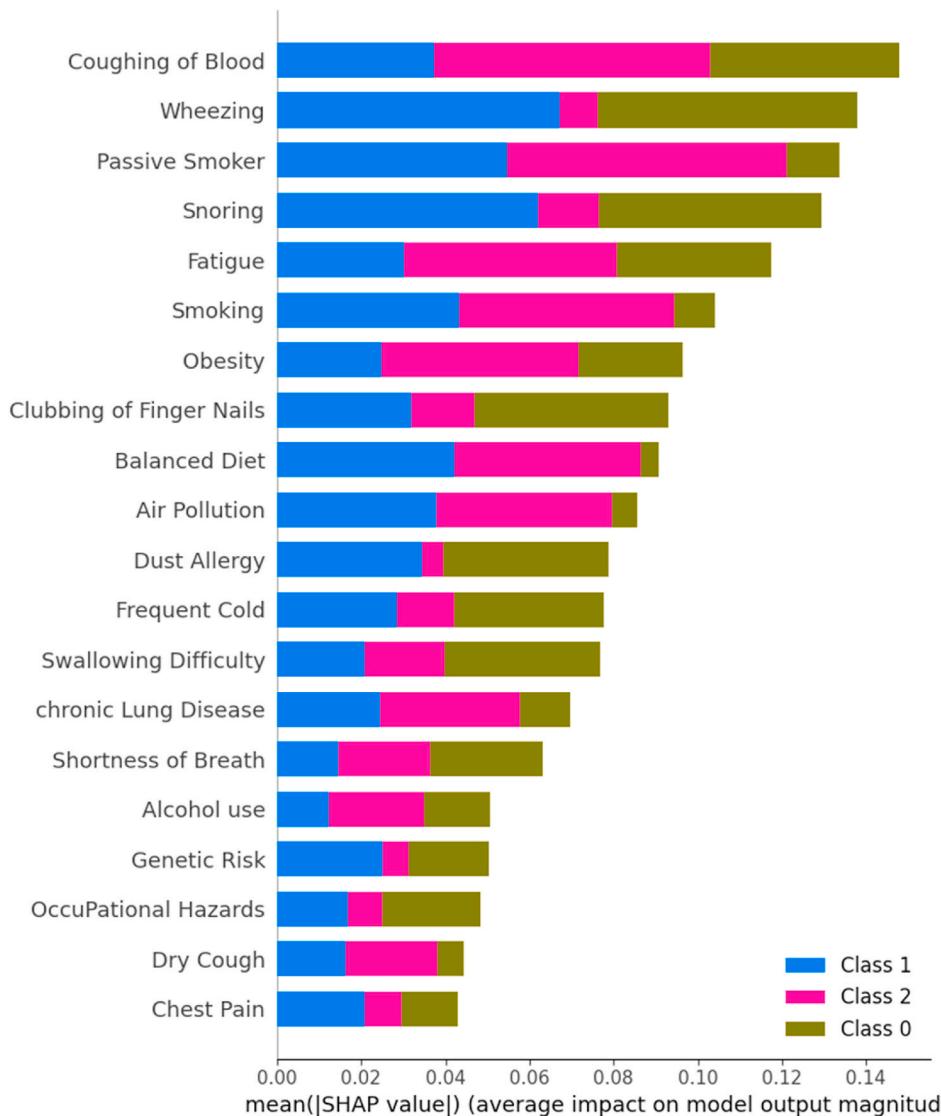


Fig. 12. Feature importance analysis using SHAP for multi-class lung cancer classification with ANN model.

ANN and CNN models enriched with interpretability techniques. The study in Ref. [30] focused on fast and non-invasive staging of lung cancer using CNNs and achieved 99.64 % accuracy. Despite the impressive results, the absence of model explainability and reliance on a specific dataset (IQ-OTHNCCD) limits generalizability. Our use of SHAP and Grad-CAM adds a crucial interpretive layer that fosters model trustworthiness, especially in high-stakes diagnostic decisions. Finally [31], concentrated on enhancing image segmentation accuracy via preprocessing techniques. Their work demonstrated measurable improvements in structural recognition in CT and X-ray images, achieving high Dice Similarity Coefficient (DSC) and Structural Similarity Index (SSIM). While complementary to our image preprocessing step, their study focused primarily on segmentation rather than classification and lacked integration with structured clinical features, which are integral to our ANN-based pipeline.

The integration of our explainable AI-based lung cancer diagnostic system into hospital workflows has the potential to enhance early detection, streamline radiological assessments, and support oncologists in making evidence-based decisions. The system could be deployed as a decision support tool embedded within existing Picture Archiving and Communication Systems (PACS) or Electronic Health Records (EHRs), where it can automatically analyze incoming CT or X-ray scans and flag high-risk cases based on learned patterns. The inclusion of

interpretability techniques such as SHAP and Grad-CAM is especially valuable in clinical settings, as it allows radiologists and physicians to understand and verify the rationale behind model predictions—thus facilitating trust, reducing diagnostic ambiguity, and aiding multidisciplinary team discussions. For successful deployment, collaboration with IT departments, radiology units, and clinical stakeholders would be essential to ensure seamless interoperability, data privacy compliance (e.g., HIPAA, GDPR), and user-friendly integration. However, clinical deployment also hinges on adherence to strict regulatory standards, such as FDA (U.S.) or CE (EU) approval pathways for Software as a Medical Device (SaMD). These require rigorous validation on diverse, real-world datasets, risk management documentation, usability studies, and post-market surveillance plans. Our current model, while demonstrating high accuracy and interpretability, would need to undergo prospective clinical trials to confirm generalizability across patient populations, scanner types, and healthcare environments. Future work will focus on expanding the dataset size and diversity, incorporating real-time model feedback loops, and potentially enabling a clinician-in-the-loop system that blends AI recommendations with human expertise. Additionally, efforts will be made to develop a user-friendly interface with adjustable confidence thresholds and alert systems, ensuring the AI tool serves as a supportive aid rather than a replacement, reinforcing ethical AI use in healthcare.

5. Conclusion

This study presents a novel AI-driven framework that integrates CNN and ANN to enhance lung cancer detection by combining radiological imaging with patient-specific clinical data. The results demonstrate that CNNs excel in capturing intricate spatial patterns in CT scans, while ANNs effectively interpret structured clinical parameters, allowing for a more holistic diagnostic approach. By integrating these models, the proposed system achieves high classification accuracy and predictive reliability, addressing key limitations observed in standalone deep learning models. The hybrid approach not only improves diagnostic performance but also enhances interpretability by incorporating multiple data modalities, offering a more comprehensive assessment of lung cancer progression and severity. The proposed system holds significant clinical relevance by aligning with the increasing demand for AI-assisted decision support tools in oncology, offering an automated yet interpretable diagnostic pathway for lung cancer classification.

Despite its promising results, certain challenges remain. Model performance is influenced by dataset quality, variations in imaging protocols, and class imbalances in patient data, which can impact generalizability across diverse populations. Additionally, computational demands pose challenges for real-time deployment in resource-constrained healthcare settings, necessitating further optimization for clinical implementation. Future research will focus on further refining the integration of CNN and ANN models, incorporating real-world clinical validation, and expanding the dataset to improve generalization.

CRediT authorship contribution statement

Emir Oncu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Investigation, Formal analysis, Data curation, Conceptualization. **Fatih Ciftci:** Writing – review & editing, Writing – original draft, Supervision, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Data availability

The datasets used in this study are publicly available on Kaggle. The clinical data utilized for the Artificial Neural Network (ANN) model can be accessed at "<https://www.kaggle.com/datasets/the-devastator/cancer-patients-and-air-pollution-a-new-link/data>". The imaging data employed for the Convolutional Neural Network (CNN) model is available at "<https://www.kaggle.com/datasets/mohamedhanyyy/chest-ct-scan-images>" and "<https://www.kaggle.com/datasets/borhanittrash/lung-cancer-ct-scan-dataset>". These datasets were used under their respective open-access licenses for research purposes. The code developed for this study is available upon reasonable request.

Ethical Statement for solid state ionics

- 1) This material is the author's own original work, which has not been previously published elsewhere.
- 2) The paper is not currently being considered for publication elsewhere.
- 3) The paper reflects the authors' own research and analysis in a truthful and complete manner.
- 4) The paper properly credits the meaningful contributions of co-authors and co-researchers.
- 5) The results are appropriately placed in the context of prior and existing research.
- 6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.

- 7) All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

The violation of the Ethical Statement rules may result in severe consequences.

To verify originality, your article may be checked by the originality detection software iThenticate. See also <http://www.elsevier.com/editors/plagdetect>.

I agree with the above statements and declare that this submission follows the policies of Solid-State Ionics as outlined in the Guide for Authors and in the Ethical Statement.

Declaration of generative AI and AI assisted technologies in the writing process

The authors used Grammarly and QuillBot solely for grammar correction and stylistic refinement. ChatGPT was employed strictly for language clarity suggestions and for improving the readability of technical content, without generating original scientific content or interpretations. All intellectual and scientific contributions—including study design, data analysis, interpretation, and manuscript content—were developed entirely by the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The endeavor was exclusively conducted using the organization's current staff and infrastructure, and all resources and assistance came from inside sources. Ethical approval is not applicable. The data supporting the study's conclusions are accessible inside the journal, according to the author. Upon a reasonable request, the corresponding author will provide the raw data supporting the study's findings.

References

- [1] R. Sharma, Mapping of global, regional and national incidence, mortality and mortality-to-incidence ratio of lung cancer in 2020 and 2050, *Int. J. Clin. Oncol.* 27 (4) (2022) 665–675.
- [2] M. Cellina, et al., Artificial Intelligence in Lung Cancer Imaging: Unfolding the Future, *Multidisciplinary Digital Publishing Institute (MDPI)*, Nov. 01, 2022, <https://doi.org/10.3390/diagnostics12112644>.
- [3] X. Yin, et al., Artificial intelligence-based prediction of clinical outcome in immunotherapy and targeted therapy of lung cancer, *Semin. Cancer Biol.* 86 (Nov. 2022) 146–159, <https://doi.org/10.1016/J.SEMCANCER.2022.08.002>.
- [4] K. Xu, et al., Progress of exosomes in the diagnosis and treatment of lung cancer, *Biomed. Pharmacother.* 134 (Feb. 2021) 111111, <https://doi.org/10.1016/J.BIOPHA.2020.111111>.
- [5] K.H. Yu, et al., Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features, *Nat. Commun.* 7 (Aug) (2016), <https://doi.org/10.1038/ncomms12474>.
- [6] Z. Gandhi, et al., Artificial Intelligence and Lung Cancer: Impact on Improving Patient Outcomes, *Multidisciplinary Digital Publishing Institute (MDPI)*, Nov. 01, 2023, <https://doi.org/10.3390/cancers15215236>.
- [7] W.D. Travis, et al., International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma, *J. Thorac. Oncol.* 6 (2) (Feb. 2011) 244–285, <https://doi.org/10.1097/JTO.0B013E318206A221>.
- [8] R.S. Heist, L.V. Sequist, J.A. Engelman, Genetic changes in squamous cell lung cancer: a review, *J. Thorac. Oncol.* 7 (5) (May 2012) 924–933, <https://doi.org/10.1097/JTO.0B013E31824CC334>.
- [9] D.R. Gandara, P.S. Hammerman, M.L. Sos, P.N. Lara, F.R. Hirsch, Squamous cell lung cancer: from tumor genomics to cancer therapeutics, *Clin. Cancer Res.* 21 (10) (May 2015) 2236–2243, <https://doi.org/10.1158/1078-0432.CCR-14-3039>.
- [10] J.M. Varlotto, et al., Should large cell neuroendocrine lung carcinoma be classified and treated as a small cell lung cancer or with other large cell carcinomas? *J. Thorac. Oncol.* 6 (6) (Jun. 2011) 1050–1058, <https://doi.org/10.1097/JTO.0B013E318217B6F8>.

- [11] R.J. Battafarano, et al., Large cell neuroendocrine carcinoma: an aggressive form of non-small cell lung cancer, *J. Thorac. Cardiovasc. Surg.* 130 (1) (Jul. 2005) 166–172, <https://doi.org/10.1016/J.JTCVS.2005.02.064>.
- [12] H. Takei, et al., Large cell neuroendocrine carcinoma of the lung: a clinicopathologic study of eighty-seven cases, *J. Thorac. Cardiovasc. Surg.* 124 (2) (Aug. 2002) 285–292, <https://doi.org/10.1067/MTC.2002.122523>.
- [13] P. Malik Amisha, M. Pathania, V. Rathaur, Overview of artificial intelligence in medicine, *J. Fam. Med. Prim. Care* 8 (7) (2019) 2328, https://doi.org/10.4103/jfmpc.jfmpc_440_19.
- [14] P. Hamet, J. Tremblay, Artificial intelligence in medicine, *Metabolism* 69 (Apr. 2017) S36–S40, <https://doi.org/10.1016/J.METABOL.2017.01.011>.
- [15] H. Goyal, et al., Application of Artificial Intelligence in Pancreaticobiliary Diseases, SAGE Publications Ltd, 2021, <https://doi.org/10.1177/2631774521993059>.
- [16] C.E. Lawson, et al., Machine learning for metabolic engineering: a review, *Metab. Eng.* 63 (Jan. 2021) 34–60, <https://doi.org/10.1016/J.YMBEN.2020.10.005>.
- [17] M.M. Taye, Understanding of machine learning with deep learning: architectures, workflow, applications and future directions, *Computers (Basel)* 12 (5) (2023) 91.
- [18] K. Seetharam, S. Shrestha, P.P. Sengupta, Cardiovascular imaging and intervention through the lens of artificial intelligence, *Intervent Cardiol. Rev. Res. Resour.* 16 (2021), <https://doi.org/10.15420/icr.2020.04>.
- [19] E.J. Topol, High-performance Medicine: the Convergence of Human and Artificial Intelligence, Nature Publishing Group, Jan. 01, 2019, <https://doi.org/10.1038/s41591-018-0300-7>.
- [20] L. Alzubaidi, et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, *J. Big Data* 8 (2021) 1–74.
- [21] R. Khanam, M. Hussain, R. Hill, P. Allen, A comprehensive review of convolutional neural networks for defect detection in industrial applications, *IEEE Access* 12 (2024) 94250–94295.
- [22] A. Zafar, et al., A comparison of pooling methods for convolutional neural networks, *Appl. Sci.* 12 (17) (2022) 8643.
- [23] R. Archana, P.S.E. Jeevaraj, Deep learning models for digital image processing: a review, *Artif. Intell. Rev.* 57 (1) (2024) 11.
- [24] M. Thorat, S. Pandit, S. Balote, Artificial neural network: a brief study, *Asian J. Conver. Tech.* 8 (3) (2022) 12–16.
- [25] N. Zeinali, N. Youn, A. Albashayreh, W. Fan, S. Gilbertson White, Machine learning approaches to predict symptoms in people with cancer: systematic review, *JMIR Cancer* 10 (2024) e52322.
- [26] M.G. Lanjewar, K.G. Panchbhai, P. Charanarur, Lung cancer detection from CT scans using modified DenseNet with feature selection methods and ML classifiers, *Expert Syst. Appl.* 224 (2023) 119961.
- [27] A. Elnakib, H.M. Amer, F.E.Z. Abou-Chadi, ‘Early Lung Cancer Detection Using Deep Learning Optimization’, 2020.
- [28] N.A. Wani, R. Kumar, J. Bedi, DeepExplainer: an interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence, *Comput. Methods Progr. Biomed.* 243 (2024) 107879.
- [29] S. Saifullah, B. Yuwono, H.C. Rustamaji, B. Saputra, F.A. Dwiyanto, R. Dreżewski, Detection of chest X-ray abnormalities using CNN based on hyperparameter optimization, *Eng. Proceed.* 56 (1) (2023), <https://doi.org/10.3390/ASEC2023-16260>.
- [30] M.M. Musthafa, I. Manimozhi, T.R. Mahesh, S. Guluwadi, Optimizing double-layered convolutional neural networks for efficient lung cancer classification through hyperparameter optimization and advanced image pre-processing techniques, *BMC Med. Inf. Decis. Making* 24 (1) (2024) 142, <https://doi.org/10.1186/s12911-024-02553-9>.
- [31] S. Saifullah, R. Dreżewski, Modified histogram equalization for improved CNN medical image segmentation, *Procedia Comput. Sci.* 225 (2023) 3021–3030, <https://doi.org/10.1016/j.procs.2023.10.295>.
- [32] V.M. Patro, M.R. Patra, Augmenting weighted average with confusion matrix to enhance classification accuracy, *Trans. Machine Learning Artificial Intell.* 2 (4) (2014) 77–91.
- [33] H. Dalianis, H. Dalianis, Evaluation Metrics and Evaluation’, *Clinical Text Mining: Secondary Use of Electronic Patient Records*, 2018, pp. 45–53.
- [34] C. Goutte, E. Gaussier, A probabilistic interpretation of precision, recall and F-score, with implication for evaluation, in: European Conference on Information Retrieval, Springer, 2005, pp. 345–359.
- [35] D. Chicco, G. Jurman, The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification, *BioData Min.* 16 (1) (2023) 4.