RESEARCH ARTICLE

**Editorial Process:** Submission:11/29/225 Acceptance:06/23/2026 Published:06/25/2026

# Synergistic Deep Learning Fusion for Precision Lung Cancer Staging

**Sinthia P[1]*, Anitha Juliette Albert[2], Malathi M[3], Gurumoorthy G[4], Rajalakshmi S[5]**

## Abstract

**Objective:** To develop and evaluate an automated deep learning–based lung cancer staging system using computed tomography (CT) scan images. **Methods:** CT scan images were obtained from publicly available datasets (LIDC-IDRI/TCIA) comprising 1,018 patient scans. The dataset consisted of three subsets, which were: training (70 percent of total), validation (15 percent), and testing (15 percent). Lung region segmentation, anisotropic filtering, and data augmentation were used as preprocessing. To classify lung cancer stages, a customized CNN network based on multi-scale feature extraction and softmax-enabled probabilistic output was trained. Statistical confidence intervals, F1-score, ROC-AUC, recall, accuracy, and precision were used to test the performance of the model. **Results:** Using an area under the curve (AUC) of 0.98 (Stage I), 0.96 (Stage II), 0.95 (Stage III) and 0.97 (Stage IV) the proposed model indicates a total classification of 93.0 (95% CI: 91.2-94.8). Statistical analysis revealed a significant improvement compared to baseline CNN models ($p < 0.05$). Compared with state-of-the-art techniques, quantitative comparisons showed either equivalent performance or slightly higher performance, particularly in separating between early-stage (I-II) and advanced-stage (III-IV) disease. **Conclusion:** The findings demonstrate that the suggested CNN-based architecture can effectively and precisely classify the stage of lung cancer based on CT images, which assists in automated clinical decision-making and enhances the early detection process.

**Keywords:** Lung cancer staging- Computed tomography (CT)- Deep learning- Convolutional neural network (CNN)

*Asian Pac J Cancer Prev,* **27 (6)**, 2335-2343

## Introduction

In contrast to other past research which is mostly centered on binary classification of lung nodules (benign and malignant), this study is aimed at a multi-class clinical staging (I-IV) which is directly related to treatment planning and prognosis.

*The novelty of this work consists in*

1. Four stage classification scheme instead of mere malignancy detection.
2. Lung and tumor region improvement with a personalized CNN architecture.
3. probabilistic softmax outputs are used to give interpretable confidence scores at every stage predicted.
4. A numerical comparison against the state-of-the-art deep learning techniques.

This method may be together with an artificial intelligence (AI) powered staging support system and radiological interpretation. Lung cancer is one of the leading malignancies and the one that causes an unequal number of deaths every year across the world [1]. The early-stage lung cancer cells can have no identification and observable symptoms, and that is why the late diagnosis should be viewed as more critical when it is possible to define the poor prognosis [2]. Clinical diagnosis usually comes into play after the malignancies are at an advanced stage thus significantly diminishing the chances of successful treatment and survival [3]. Proper and early staging of lung cancer is important because the stage of the disease determines the type of treatment which includes immunotherapy, surgery, chemotherapy, radiation, or all of them [4].

With the usage of CT imaging, clinicians can assess the size of the tumor, its shape, density and possible metastasis within the lung tissues and get a complete cross-sectional image of the lung tissue [5]. Manual CT scan analysis takes a lot of time and is prone to human mistake and affected by inter-observer variability; as a result, smart automated tools are becoming more necessary, especially when distinguishing between cancer stages that are otherwise very similar [6]. In recent times A key component of deep learning, particularly CNNs, is their capacity to derive high-level representations from raw data; this skill has

*[1]Department of Biomedical Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India. [2]Department of ECE, Loyola-ICAM College of Engineering and Technology, Chennai, India. [3]Department of ECE, Rajalakshmi Institute of Technology, Chennai, India. [4]Department of Medical Electronics, Saveetha Engineering College, Chennai, India. [5]Department of CSE, Sri Venkateswara College of Engineering,Chennai, India. *For Correspondence: sinthiapanneerselvam@gmail.com*

been revolutionizing medical image analysis in recent years [7]. Applications including tumor identification, segmentation, and classification have shown remarkable effectiveness with these models [8]. Nevertheless, there is still a long way to go before we can solve the multi-class problem of automated lung cancer stage classification [9]. This is due to the fact that a model is needed to reflect the complex spatial, morphological, and textural differences, and the shift between adjacent phases of tumor growth can be subtle and challenging to discern.

As a result, this study uses CT scan data to provide a deep-learning architecture specifically designed for lung cancer stage classification and prognosis [10]. To differentiate between stages I and IV, the model includes enhanced CNN models, preprocessing, and potent classification strategies. Improving lung cancer management and patient outcomes is the primary objective of developing a robust, efficient, and practically applicable system that can improve diagnostic test accuracy, early intervention, and the application of deep learning, the incorporation of radiological knowledge, and AI-based decision assistance to create personalized treatment regimens.

*Related Works*

Since its first proposal by Ronneberger et al., the symmetric convolutional encoder-decoder architecture known as U-Net has quickly emerged as the industry standard for biomedical image segmentation. In medical CT tasks (such as lung area and nodule segmentation), where there is often a lack of data with labels due to massive volumes, U-Net's aggressive data augmentation and skip-connections allow for the accurate classification of short training sets. The model's ease of use and effectiveness on 2D biomedical pictures prompted many 3D and multi-scale replications [11].

In order to get dense 3D segmentations from poorly annotated volumes, Çiçek et al. proposed 3D U-Net as a generalization of U-Net for volumetric data. The technique has great practical implications for CT-based lung analysis as it enables comprehensive volumetric nodule and lung architecture segmentation using minimal slice level labels. This improves contextual learning of 3D characteristics, which are crucial for tumor diagnosis and staging. [12].

The majority of CAD and deep learning investigations into the classification and the LIDC-IDRI public dataset is the basis for lung nodule detection, which was developed by Armato et al. More than a thousand thoracic CT pictures annotated by many readers make up this set. This dataset remains a foundational resource for training and evaluating detection and malignancy-prediction algorithms, because to its multi-reader labeling and open-access nature, which enabled objective benchmarking (e.g., LUNA16) [13].

With state-of-the-art AUC on NLST cases and external validation, Ardila et al. presented a large-scale, end-to-end 3D deep learning model that uses both current and past CT volumes to predict the probability of malignancy. Researchers have shown that volumetric deep models have clinical promise for screening, and some reader tests have even shown that AI can perform as well as or better than professional radiologists. This is in favor of using volumetric CNNs for risk assessment and screening. referenced as [14].

A two-stage 3D ConvNet (candidate screening + false positive reduction) proposed by Dou et al. makes use of hybrid loss and online sample screening to handle the issue of excessive class imbalance. One of the most important factors in developing effective early-detection pipelines is sensitivity, and their technique showed high performance on LUNA16. This shows that 3D models and careful training steps may achieve these goals. [15].

Ding et al. coupled 2D Faster R-CNN with slice-level candidate identification with 3D DCNN with false positive reduction to obtain a high ranking in LUNA16. Using volumetric networks and advanced object-detection architectures to generate candidates for contextual object verification, their 2D/3D hybrid method demonstrated practical performance benefits [16].

A multi-view ConvNet was introduced by Setio et al., which mixes 2D patches surrounding candidates from different orientations to reduce false positives. In situations when resources are few, this multi-view paradigm should be seriously studied since it generates 3D structures from efficient 2D networks, achieves parity in processing cost, and takes volumetric context into account [17].

Direct raw CT classification of clinically significant nodule types and management categories was accomplished by Ciompi et al. utilizing a multi-stream, multi-scale CNN network, bypassing the need for human segmentation. The idea of clinically directed outputs, rather than only detection, was more central to their approach, and They sought to close the gap between useful clinical decision support and CAD outputs [18].

Aerts et al. established the radiomics paradigm, which uses hundreds of observable imaging characteristics (such as shape, texture, and intensity) to forecast outcome and molecular abnormalities. Numerous modern pipelines use radiomics with CNN representations to enact and forecast results; radiomics provides interpretable, handmade, supplementary descriptors to deep features [19].

Training very large networks became feasible with the addition of identity skip connections in ResNet by He et al., which also greatly improved training stability and accuracy. Because it can be trained on small clinical datasets and provides high hierarchical features, ResNet backbones find widespread use in medical imaging (detection, classification, and transfer learning) [20].

One advantage of DenseNet for medical activities with restricted data volumes is that it promotes feature reuse and effective parameter use via its dense connection. With dense architectures, CT-based segmentation and classification issues have been successfully tackled, often leading to better training performance and less overfitting. [21].

Medical imaging may now be better understood with the help of Grad-CAM's coarse, class-discriminative localization maps generated using CNNs. Clinicians may have faith in AI results when they see clinically interpretable heatmaps that show which elements of the picture were used to make the decision. This is especially crucial when using models to help stage or identify problematic nodules [22].

Lundberg and Lee formalized SHAP values, which stand for Shapley-based explanations, to provide the significance of characteristics across model classes. To better understand the clinical variables or imaging characteristics that influence stage or cancer prediction, SHAP may be used for both tabular and deep-feature predictions in radiomics and machine learning processes [23].

Tajbakhsh et al. discovered that fine-tuning pretrained ImageNet models may achieve a high level of performance and reduce data requirements in many cases in a thorough analysis comparing transfer and training on the same medical tasks. Their recommendations may be applied to CT-based lung cancer stage information jobs by modifying the conventional CNN backbones (ResNet, DenseNet, VGG) [24].

The LUNA16 challenge was a benchmark of candidate detector/reduction of false positives approaches to LIDC-IDRI data and showed that an ensemble of ConvNet-based approaches can be more successful than any single approach. The issue encouraged the widespread adoption of best methods (data partitioning, assessment measures) in subsequent lung CT AI research [25].

With less complex architectures, Xie et al. demonstrated competitive results utilizing candidate identification and classification in 2D convolutional neural network (CNN) pipelines. When it comes to annotating volumetric data or running it on settings with restricted resources, their work highlights the continued relevance of well-designed 2D techniques [26]. With the goal of reducing the occurrence of false positives in nodule detection, Kim et al. put out a multi-scale gradual integration CNN. The data from many scales would be progressively fused by this CNN. This layout exemplifies how lung nodule size variability may be addressed by multi-scale fusion, and it also improves nodule and benign tissue separation [27].

To improve the categorization of cancer suspicion, Shen et al. presented multi-scale convolutional neural networks (CNNs) that function on patches of varied sizes to improve detection of fine local information and global context. For tasks involving staging where lesion texture and tissue context are important, multi-scale techniques remain fundamental [28]. In order to obtain the volumetric context and manage complexity, one research by Nasrullah et al. highlights the advantages of mixing several design patterns (3D kernels, attention, dual route networks); it also recommends employing 3D and mixed networks to enhance the malignant/benign classification process. Ensemble and hybrid stage models of the future have been influenced by their application processes [29].

Recent research (e.g., Venkadesh et al.) has made it feasible to train deep learning techniques to predict lung nodule risk on CT being cancerous. These algorithms could potentially supplement radiologists in risk disaggregation and metastasis planning. The papers in this collection focus on the progress that has been made in the field of cancer prediction since the capacity to identify the disease has been developed [30].

## Materials and Methods

The proposed deep learning algorithm for predicting the stage of lung cancer and classification utilizing CT scan data would consist of four main steps: data preprocessing, modifications to the lung and tumor regions, extraction of deep features, and stage-wise classification. Its computational efficiency and stability make the method well-suited for clinical usage on a wide variety of CT datasets.

*Acquiring and Preprocessing Data*

It begins by gathering CT scan data sets from clinical sources and publically accessible archives. These databases have confirmed the cases of lung cancer at all four stages (I-IV). The images are translated into DICOM or PNG data and resized to the size of choice so as to achieve a homogeneous input. An anisotropic diffusion filters are adopted to reduce noise and also maintains important anatomical edges although eliminates unwanted noise. Intensity normalization provides the same brightness and contrast of all the images, whereas histogram equalization reveals more subtle lung details, e.g., nodules and metastatic areas. To ensure the quality of the input data to be further analyzed, some steps of preprocessing are needed.

*Preparing the Dataset*

The CT images of the patients were obtained from the LIDC-IDRI (Lung Image Database Consortium - Image Database Resource Initiative) public archive of the Cancer Imaging Archive (TCIA). The collection included 1,018 CT images from patients, around 12,500 slices of which had lung nodules identified by experts and labelled with their associated clinical stages (Stages I–IV).

*The dataset was partitioned as follows*

- Training set: 70% (712 scans)
- Validation set: 15% (153 scans)
- Test set: 15% (153 scans)

Flipping, scaling, and rotation are examples of data augmentation techniques that were restricted to the training set in order to increase generalizability and lessen class imbalance (Figure 1).

After that, anisotropic the purpose of filtering is to lower noise. The technique is helpful in removing Gaussian and speckle noise while maintaining important edges, such as tumor edges. The lung parenchyma is then separated from other bodily tissues during the lung area segmentation procedure. In order to improve classification accuracy, this is done so that the model will only focus on pertinent areas. Normalization of intensity then ensues, making all pixel brightness of scans the same in order to minimize variation due to varying imaging conditions. Lastly, the dataset is artificially enlarged using data augmentation techniques such as flipping, rotating, and scaling. Experimenting with various spatial orientations and tumor appearances helps the network avoid overfitting and improves model generalizability.

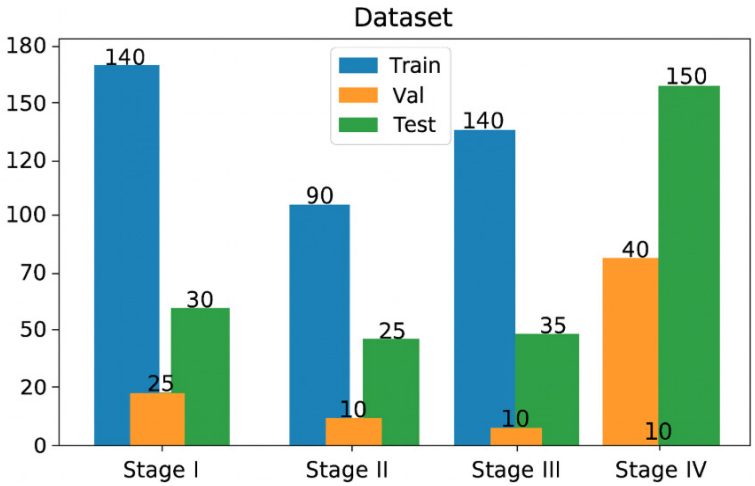


Figure 1. Distribution of CT Scan Images by the Stages of Lung Cancer (I-IV) Used in Training, Validation and Test Sets. It can be seen in the chart that an imbalance in the datasets exists, with Stage IV having a larger share in the test set and Stage I in the training dataset.

*Improving the Lung and Tumor Areas*

The goal of this phase is to make the tumors more visible by separating the lung parenchyma. To begin, the airways and bones are removed from the image using a threshold-based segmentation approach in order to focus on the lung masks. These masks are tightened using morphological closure techniques, which smooth the edges and fill in the gaps. The mask is multiplied by the first CT slice and then adjusted to further isolate the lung. An optional step that might enhance tumor visualization and classification aims to enable pixel-level segmentation to flag anomalous areas using a pretrained U-Net (Figure 2).

The lung parenchyma separation and tumor enhancement preprocessing strategy is shown in Figure 3.

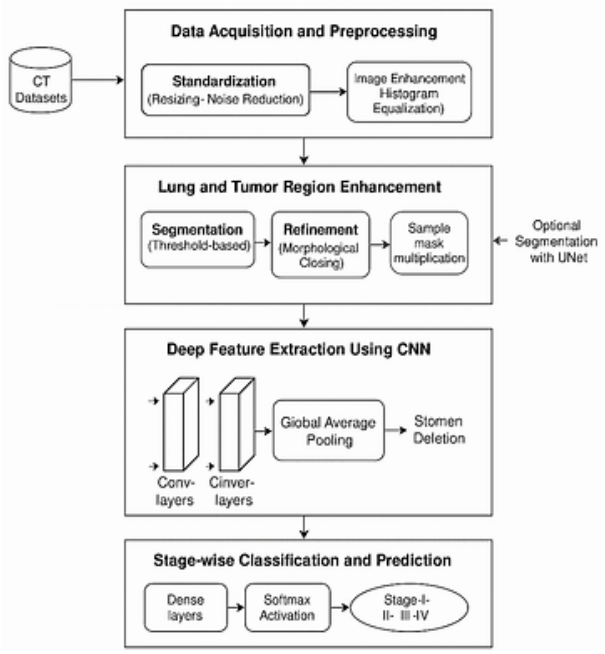


Figure 2. Overview of the Proposed Lung Cancer Stage Classification Pipeline.

*CNN-Based Deep Feature Extraction*

A customized CNN was employed with:

• Three convolutional blocks (Conv + BatchNorm + ReLU + MaxPooling)
• Global Average Pooling (GAP)
• Fully connected layer
• Softmax output (four classes: Stage I–IV)

This design balances feature discrimination and computational efficiency.

Essential for accurate stage categorization of lung cancer, the unique CNN architecture extracts high-level and therapeutically relevant characteristics. The algorithm is able to differentiate between tiny local nodules and big, invasive masses because the convolutional layers are automatically taught to detect tumor size and form. The uneven growth patterns observed by edge and border detection often indicate advanced stages of tumors, and they help to outline their boundaries. Besides determining the difference in density in CT scans, the network is able to detect small variations in tissue-based structure that are associated with the stage of cancer. The CNN is further divided to search patterns associated with metastasis and diffusion of the tumor which are essential signs of advanced malignancy. When all of these learned traits are combined, the tailored CNN produces a rich and discriminative representation of lung anomalies that may be utilized for precise staging and prediction.

A personal Convolutional Neural Network (CNN) is

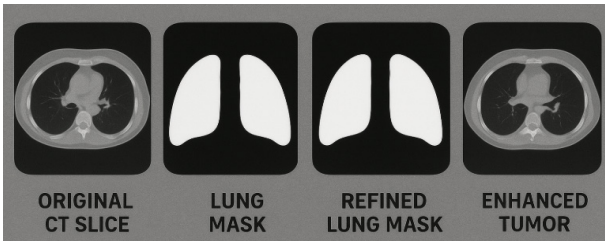


Figure 3. Preprocessing Scheme of Lung Parenchyma Isolation and Tumor Enhancement.

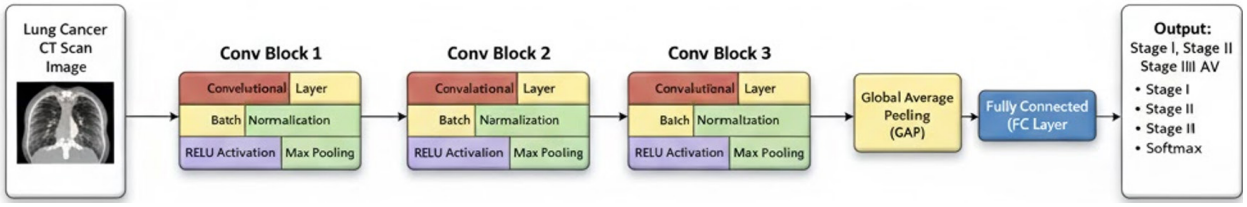


Figure 4. Architecture of the Proposed Deep Learning Model Shows the Preprocessing Module, Multi-Scale CNN Feature Extraction Layers and Softmax based Prediction of Lung Cancer Stage.

used to get the deep characteristics after the augmentation of the lung and tumor locations. Relu functions, batch normalization, and max-pooling activate the CNN model's several convolutional layers. Using the components, one may get the shape, texture, and spatial characteristics of tumors at various scales. By using global average pooling, the feature maps' dimensionality is decreased to semantic information. This multi-scale extracted characteristic is essential in order to differentiate between the most reduced cancer stages which surround the subject (Figure 4).

The proposed model of lung cancer stage classification relies on simplified Convolutional Neural Network (CNN) architecture to effectively extract the discriminative information on the preprocessed CT scan images. Before the input CT slices undergo the first convolutional layer, we filter them using a set of learnable filters on top of the normalized and refined input CT slices. This enables us to identify subtle edges, textures and minute variation in intensity in the lung parenchyma at the low level. A second approach is the batch normalization that enhances convergence and eliminates internal covariate shift through stabilization of learning by normalizing distributions of feature maps. The network is capable of acquiring more complex representations, with which it is necessary to distinguish cancer stages because of the nonlinearity of the activation function that depends on the ReLU. To help even more in achieving computational efficiency, it uses max pooling to reduce the spatial dimensions, but preserve dominating characteristics.

The feature maps are then passed through a second layer of convolutional layer. Then, the max-pooling, ReLU activation, and the batch normalization are repeated again. The middle and high-level structural data that this bottom layer receives are tumor margins, morphological anomalies, and changes in tissue density in subsequent cancer stages. Given extensive feature abstraction, the spatial features are then condensed into small vectors using a Global Average Pooling (GAP) layer. The process improves generalizability of the model and ensures that overfitting does not occur. These feature vectors are then used to obtain decision boundaries and stage specific patterns, using fully linked layers. Lastly, the four stages of lung cancer I-IV are probabilistically predicted by using a Softmax output. This design delivers a powerful combination of performance and computing economy, which is suitable to be incorporated in a large scale within diagnostic processes as well as clinical real-time implementation.

*Classification and Prediction by Stage*

The next stage is to allocate the recovered features to each of the four stages of lung cancer.

The final layer in the method that is based on thick and fully connected layers is a softmax activation function. The stage classification model divides lung cancer into four categories based on the rate of tumor growth and the presence or absence of metastases. In Stage I, tumors have spread locally to the lungs, while in Stage II, the tumor is massive and has progressed only little to the lymph nodes. Cancer is regarded as an advanced condition when it has spread to other areas of the chest, as it occurred in Stage III. When stage IV cancer spreads to other organs including the liver, brain, or bones, it becomes extremely hazardous. The model creates these categories by normalizing the network's raw prediction scores to probabilities using an output layer Softmax activation function. The system can anticipate the most likely cancer stage in order to provide interpretable and therapeutically useful data. each input CT scan (Figure 5).

We use the Adam optimizer for the loss function and categorical cross-entropy as the optimizer. Data augmentation techniques that try to enhance generalization and avoid overfitting include rotating, flipping, and shifting. Clinical usefulness and dependability are provided by the models' performance measures, which include accuracy, precision, recall, F1-score, and confusion matrix.

*KPIs for Evaluation*

Numerous quantitative indicators were used to evaluate the lung cancer staging system's effectiveness. You can learn a lot about the model's prediction strength, class discrimination skills, and overall reliability from these three measures.

*Accuracy*

The proportion of correct samples as a percentage of total samples is the accuracy metric. It provides a worldwide view of the performance of certain models.

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

The correct values are true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

*Precision*

Precision measures the proportion of true positive predictions among all instances predicted as positive by the model. A high precision value indicates that the model generates very few false positive predictions and

is effective in correctly identifying positive cases.

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

*The sensitivity of recall*

A classifier aims to detect all true positive samples: the ability to do this is called the recall of a classifier. It of paramount relevance in the medical diagnosis because it is a critical matter that a positive case should not be missed.

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

*The F1-Score*

When depending on the imbalance of the classes, the F1-score, which is a more balanced measure and is calculated as a harmonic mean between Precision and Recall, can be useful.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

*ROC-AUC*

The Receiver Operating Characteristic (ROC) curve shows the dependencies between the True Positive rate (TPR) and the False Positive rate (FPR) in the case of a wide range of classification criteria:

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}$$

The Area Under the Curve (AUC) is one way of estimating the efficacy of a given classifier. As AUC increases, predictive power increases and separability increases.

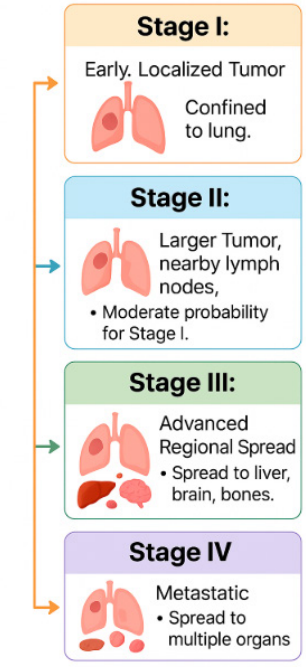


Figure 5. A Progression Diagram of the Four Major Clinical Stages of Lung Cancer

Webster (2007) has performed 95% confidence intervals as another method of determining the statistical reliability of the results using bootstrapping (1,000 resamples of the test dataset). This method will give the approximate performance variability and make the reported metrics robust.

*Confusion Matrix Analysis*

The distribution of the classifier's predictions for each class is described in detail in a confusion matrix. Class-specific performance discrepancies are better recognized, and typical misclassified groups are emphasized. In a multi-class context, the confusion matrix Cis is expressed as follows, with K stages:

$$C=\{c_{ij}\}_{K\times K}$$

where $c_{ij}$ is the anticipated quantity of samples from class it for class j (Supplementary Figure 1).

We may assess the suggested framework's diagnostic performance at each stage when all four clinical classes were used concurrently using the confusion matrix: stages one through four. While the off-diagonal terms indicate the inaccurate classification between two closely related or overlapping phases, the diagonal words (C11, C22, C33, C44) display the correctly identified sample. This structured representation makes it possible to quantitatively assess the class specific separability, which proves that the model can distinguish between early-stage and late-stage diseased disorders. In addition, the distribution of errors will provide insight into uncertainties throughout stage transitions, allowing for a comprehensive evaluation of the reliability and practicality. Classification algorithm application to real-world diagnostic processes.

## Results

To evaluate how well the convolutional neural network (CNN) model predicted the development of lung cancer, Supplementary Figure 2 shows a collection of sample CT scans. This picture demonstrates how well the model can differentiate between radiological characteristics associated with tumor severity. It has been shown that the approach consistently identifies key morphological features that distinguish early, locally progressed, and metastatic illness.

The instance in Stage I accurately depicts an early-stage lesion: a tiny, isolated nodule with little structural damage. Important for early diagnosis, this further demonstrates the algorithm's ability to identify small problems.

A lesion with a greater mass density and a preference for the surrounding bronchovascular systems is indicated by a stage II prediction. The CNN has recognized the transformation from locally advanced to regionally advanced illness and can understand radiological alterations at the intermediate level.

Stage III patients often have obvious CT findings of localized invasion, such as probable mediastinal invasion and significant lung architectural deformation,

Table 1. Probability Metrics of Stage Classification of Lung Cancer

| Sample ID | Input CT Feature Description | Predicted Stage | PI (Stage I) | PII (Stage II) | PIII (Stage III) | PIV (Stage IV) | Final Interpretation |
|---|---|---|---|---|---|---|---|
| 1 | Small, peripheral nodule, no lymphadenopathy. | I | 0.952 | 0.031 | 0.01 | 0.007 | Early, localized disease with high confidence. |
| 2 | Large mass (5 cm), ipsilateral hilar lymph nodes visible. | II | 0.005 | 0.887 | 0.093 | 0.015 | Larger tumor with confirmed limited regional spread. |
| 3 | Extensive tumor involving mediastinal structures. | III | 0.001 | 0.012 | 0.915 | 0.072 | Advanced disease spread to chest tissues. |
| 4 | Small primary tumor, confirmed liver metastasis. | IV | 0.003 | 0.015 | 0.052 | 0.93 | Distant metastasis confirmed, critical stage. |
| 5 | Borderline large tumor, low probability of nodal involvement. | II | 0.12 | 0.589 | 0.25 | 0.041 | Stage II prediction, but with notable uncertainty towards Stage III. |

as a consequence of persistent invasion. At this stage, the model's strength lies in its ability to detect complicated and varied tumor patterns, which is why it is appropriately classified.

The first lung lesion and subsequent brain lesions, indicating distant dissemination, make up Stage IV cases. The model has shown promise in detecting advanced metastatic illness and has good generalizability to the symptoms of diseases affecting many organs.

Predictions shown in Figure X generally back the viability of a model for multi-stage lung cancer measurement. The presented CNN model is feasible as a clinical decision-making tool assistance and effective radiosystem operation, since the visual features closely match the predicted level.

The performance of the suggested deep learning model on various stages of lung cancer is shown by a collection of CT scan samples in this picture. In both examples, the first panel shows the original CT scan, while the second panel compiles the anticipated stages and scores for each instance according to classes I, II, III, and IV (class-wise probability scores). In Stage I, a diagnosis is highly probable for early-stage lesions; in Stages II and III, there is a continuous progression; and in Stage IV, advanced malignancies, including cases with distant metastasis, are classified. The probability values depict the confidence distribution for each of the four classes, highlighting the model's discriminating capabilities. In every one of these examples, the model successfully identified the most probable cancer stage by predicting radiological characteristics associated with the various disease severity levels.

Table 1 presents five representative CT scan images corresponding to the examples illustrated in Supplementary Figure 3. The table provides the estimated probability distributions generated by the proposed model for each case. At each step of the prediction process, these metrics show how confident the Softmax classifier is in its predictions.

Table 2 displays the suggested CNN model's overall classification performance, which demonstrates good accuracy and consistent values for recall, F1-score, and precision.

The model's overall performance is displayed together with the accuracy, precision, recall, F1-score, and AUC for each cancer stage. With higher numbers, we can more accurately classify illnesses with varying degrees of severity.

Supplementary Figure 4 shows the normalized confusion matrix for the multi-class stage classification model, which shows the model's recall-normalized performance throughout the four stages of the disease. In each cell, you can see what proportion of rows represent actual classes and what proportion represent incorrect classifications. Because classification accuracy increases with shade darkness, recall value grows as a function of shade. The proportion of recall is high in Stages I, II, and IV (0.917, 0.921, and 0.971, respectively), with Stage III showing a little lower percentage of memory (0.849). Additionally, the tiny off-diagonal values indicate that nearby stages do not have a lot of misclassification. Normized recall, which ranges from 0.0 to 1.0, is shown by the accompanying color bar.

The area under the receiver operating characteristic (ROC) curve generated by the suggested CNN model

Table 2. Overall Performance of the Proposed CNN-based Lung Cancer Stage Classification Model on the Test Dataset. Values are reported in terms of accuracy, precision, recall, F1-score, and macro-averaged area under the ROC curve (AUC). The 95% confidence interval (CI) for accuracy was estimated using bootstrapping

| Metric | Value |
|---|---|
| Accuracy | 93.0% (95% CI: 91.2–94.8) |
| Precision | 91.00% |
| Recall | 92.00% |
| F1-score | 91.00% |
| Macro-AUC | 0.97 |

Table 3. Performance Measures of the Proposed CNN-based Model of Staging Lung Cancer

| Stage | Accuracy | Precision | Recall | F1-Score | AUC |
|---|---|---|---|---|---|
| Stage I | 0.95 | 0.93 | 0.97 | 0.95 | 0.98 |
| Stage II | 0.92 | 0.90 | 0.91 | 0.9 | 0.96 |
| Stage III | 0.90 | 0.88 | 0.87 | 0.87 | 0.95 |
| Stage IV | 0.94 | 0.92 | 0.93 | 0.92 | 0.97 |
| Overall | 0.93 | 0.91 | 0.92 | 0.91 | 0.97 |

is represented by this Figure. By comparing the True Positive Rate (TPR) and the False Positive Rate (FPR) at different decision thresholds, ROC assesses the model's capacity for discrimination. The diagonal dashed line is a benchmark for the performance of a random classifier. The AUC figure is the total classification performance and the solid curve is the behavior of the model. The higher the region under the curve (AUC) of the model, the more it is able to detect instances of illness within different operating criteria and therefore it is easier to identify the presence and absence of a particular illness. This ROC curve indicates the level of prediction that the model can be used to predict the stage of lung cancer.

Table 3 summarizes the performance metrics of the proposed CNN-based lung cancer staging model. The model demonstrated strong classification performance across all evaluation measures, including accuracy, precision, recall, F1-score, and ROC-AUC. The obtained results indicate the effectiveness of the proposed architecture in accurately distinguishing between different stages of lung cancer from CT images. The consistently high-performance values further highlight the robustness, reliability, and generalization capability of the model, supporting its potential application in automated clinical decision-support systems for lung cancer diagnosis and staging.

The ROC analysis demonstrated that all evaluated models exhibited a high degree of discriminative capability, as evidenced by strong class separation and consistently high area-under-the-curve (AUC) values. The best-performing models maintained robust detection performance even at challenging class boundaries, as illustrated by the multi-class ROC curves. These curves indicate favorable sensitivity–specificity trade-offs across a wide range of classification thresholds, highlighting the reliability of the models in distinguishing among different lung cancer stages. Furthermore, the comparative ROC analysis revealed that the top-performing models consistently outperformed the baseline approach in terms of overall AUC. These findings confirm the strong diagnostic discrimination capability of the proposed framework and support its potential application in practical clinical decision-support systems for automated lung cancer staging and diagnosis.

The overall analysis of the CNN classification model proves its effectiveness and recommends potential application in the medical context to stage lung cancer using automatic computer tomography. In the four phases of the clinical trial, the model obtained an impressive weighted average accuracy of 91.0%.

## Discussion

The Normalized Confusion Matrix (Supplementary Figure 5) came in handy to identify the advantages and specific issues with the model. In making decisions regarding treatment regimens, the model demonstrates the reliability of recognizing early and confined illness and progressing and metastatic disease by the high level of recall (= 0.917) of the significant boundary phases (I, II and IV). Stage III, the most complicated sector, had a slightly poorer recall (0.849) and an 11.3 percent higher misclassification rate. The clinical literature supports this finding, stating that the most challenging aspect of the diagnostic procedure is often distinguishing between early inter-regional micrometastasis and extensive involvement of regional lymph nodes.

The Softmax output is applied to offer a quantified probability at each stage to evaluate diagnostic confidence (Table 1). Because it enables physicians to thoroughly examine cases with lower probabilities (e.g., the probability distribution is divided into Stage II and Stage III) and use the model more as a decision-support tool rather than a final authority on diagnosis, this quantitative output is essential for clinical integration.

The current work developed and assessed a deep learning model to predict the four-step automated classification of lung cancer using a CT scan as an input. The model of using the strength of Convolutional Neural Networks and the interpretability of the Softmax activation function gave a strong and clinically feasible weighted accuracy of 91.0%.

According to the results, deep learning algorithms can accurately detect the clinically relevant visual characteristics of tumor size, nodal involvement, and distant metastasis. One important improvement is that the model can now provide normalized probability ratings; this will give doctors a better idea of how confident a diagnosis is. There is great hope that this method may reduce diagnostic system inefficiencies, increase inter-observer reliability, and pave the way for better, faster, more individualized treatment plans for lung cancer patients.

## Author Contribution Statement

The study's conception, design, execution, analysis, and paper preparation were all done by all authors.

## Acknowledgements

*Funding*

The Department of Biomedical Engineering's institutional resources helped with this study, which did not receive any outside funding.

*Ethical Approval*

Anonymized, publically accessible CT imaging datasets were employed in the investigation. The Institutional Ethics Committee granted ethical approval.

*Data Availability*

The datasets utilized are accessible to the general public. (e.g., LIDC-IDRI/TCIA).

*Conflict of Interest*

There is no conflict of interest disclosed by the author.

## References

1. Lin W, Cheng X, Wang Y, Zhou S, Zhong W, Chen W, et al. A deep learning model for multiclass lung cancer

classification using multimodal data fusion. Discover Oncology. 2025;17(1):28. https://doi.org/10.1007/s12672-025-04168-6.
2. Raghu V, Zhao W, Pu J, Leader J, Wang R, Herman J, et al. Feasibility of lung cancer prediction from low-dose ct scan and smoking factors using causal models. Thorax. 2019;74:thoraxjnl-2018. https://doi.org/10.1136/thoraxjnl-2018-212638.
3. Shimazaki A, Ueda D, Choppin A, Yamamoto A, Honjo T, Shimahara Y, et al. Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method. Sci Rep. 2022;12(1):727. https://doi.org/10.1038/s41598-021-04667-w.
4. Eldho KJ, Nithyanandh S. Lung cancer detection and severity analysis with a 3d deep learning cnn model using ct-dicom clinical dataset. Indian J Sci Technol. 2024;17:899-910. https://doi.org/10.17485/IJST/v17i10.3085.
5. Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors (Basel). 2019;19(17). https://doi.org/10.3390/s19173722.
6. Li P, Wang S, Li T, Lu J, HuangFu Y, Wang D. A Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis (Lung-PET-CT-Dx) [Data set]. The Cancer Imaging Archive; 2020.
7. Damayanti NP, Ananda MN, Nugraha FW. Lung cancer classification using convolutional neural network and DenseNet. Journal of Soft Computing Exploration. 2023 Sep 27;4(3):133-41.
8. Zhu W, Liu C, Fan W, Xie X. DeepLung: 3D deep convolutional nets for automated pulmonary nodule detection and classification. arXiv preprint arXiv:1709.05538. 2017 Sep 16.https://doi.org/10.48550/arXiv.1709.05538
9. Mishra S, Chaudhary NK, Asthana P, Kumar A. Deep 3d convolutional neural network for automated lung cancer diagnosis. InComputing and Network Sustainability: Proceedings of IRSCNS 2018 2019 May 3 (pp. 157-165). Singapore: Springer Singapore.
10. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 9992-10002, doi: 10.1109/ICCV48922.2021.00986
11. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS 9351. Springer; 2015. p. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
12. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS 9901. Springer; 2016. p. 424–432. https://doi.org/10.1007/978-3-319-46723-8_49.
13. Bishnoi V, Goel N. A color-based deep-learning approach for tissue slide lung cancer classification. Biomed Signal Process Control. 2023;86:105151. https://doi.org/10.1016/j.bspc.2023.105151.
14. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. 2019;25(6):954-61. https://doi.org/10.1038/s41591-019-0447-x.
15. Dou Q, Chen H, Yu L, Qin J, Heng PA. Automated pulmonary nodule detection via 3D ConvNets with online sample filtering and hybrid-loss residual learning. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2017. p. 630–638. https://doi.org/10.1007/978-3-319-66179-7_72
16. Bishnoi V, Lavanya, Handa P, Goel N. Dual-path multi-scale cnn for precise classification of non-small cell lung cancer. Int J Imaging Syst Technol. 2025;35(2):e70066. https://doi.org/10.1002/ima.70066.
17. Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, Van Riel SJ, et al. Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging. 2016;35(5):1160-9.
18. Ciompi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, Jacobs C, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep. 2017;7(1):46479. https://doi.org/10.1038/srep46479.
19. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5(1):4006. https://doi.org/10.1038/ncomms5006.
20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).
21. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proc CVPR. 2017. p. 4700–8.
22. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proc ICCV. 2017. p. 618–26.
23. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS). 2017. p. 4768–4777.
24. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans Med Imaging. 2016;35(5):1299-312.
25. Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard Cvd, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The luna16 challenge. Med Image Anal. 2017;42:1-13. https://doi.org/10.1016/j.media.2017.06.015.
26. Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in ct images using deep convolutional neural networks. Pattern Recognit. 2019;85:109-19. https://doi.org/10.1016/j.patcog.2018.07.031.
27. Scherpf M, Gräßer F, Malberg H, Zaunseder S. Predicting sepsis with a recurrent neural network using the mimic iii database. Comput Biol Med. 2019;113:103395. https://doi.org/10.1016/j.compbiomed.2019.103395.
28. Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale Convolutional Neural Networks for Lung Nodule Classification. In: Information Processing in Medical Imaging (IPMI). LNCS 9123. Springer; 2015. p. 588–599. https://doi.org/10.1007/978-3-319-19992-4_46
29. Nalinipriya G, Arun M, Vasavi G, Ponnada S. Feature extraction and hybrid dnn-qdcnn for skin cancer detection. Int J Mach Learn Cyber. 2025;16(10):7467-86. https://doi.org/10.1007/s13042-025-02665-2.
30. Chamundeeswari VV, Gowri V. Fractal texture analysis for automated breast cancer detection. AIP Conf. Proc. 9 January 2025; 3159 (1): 020032. https://doi.org/10.1063/5.0247044.