The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 21(1) S.I (1) 445-467, 2026 www.thebioscan.com

Explainable Lung Cancer Classification Framework Using Attention-LiteVoxResNet and Grad-CAM Visualization

T. Evangeline Dhivya[1], Dr. R. Balasubramanian[2]

1. Research Scholar, 23114012302034, Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli

2. Professor Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli

*Corresponding Author: evangelinedhivya94@gmail.com

DOI: https://doi.org/10.63001/tbs.2026.v21.i01.S.I(1).pp445-467

KEYWORDS

VoxResNet, Lung Cancer, 3D CNN, CT image, Grad-CAM, Grad-CAM++

Received on: 27-12-2025

Accepted on: 12-02-2026

Published on:26-02-2026

Abstract

Lung cancer is one of the foremost reasons of cancer-related mortality worldwide, underscoring the essential for early and accurate diagnosis. This study introduces Attention-LiteVoxResNet, a lightweight 3D convolutional neural network designed for analyzing volumetric CT scans. The framework follows a two-stage process: first, it differentiates normal lung tissue from Cancerous cases; second, it classifies non-small cell lung Cancer (NSCLC) into adenocarcinoma, large cell carcinoma, or squamous cell carcinoma. By integrating residual learning with attention mechanisms, the model captures fine spatial details while keeping computational demands low. To provide greater transparency, explainable AI methods—Grad-CAM and Grad-CAM++ applied. Both highlight the regions most responsible for the model's predictions, with Grad-CAM++ offering more precise localization of subtle structural patterns. Experiments conducted on CT data from The Cancer Imaging Archive (TCIA) achieved 95% accuracy in Cancer detection and 91% in subtype classification.

INTRODUCTION

Across the world, lung cancer remains a major contributor to cancer-related deaths, with non-small cell lung cancer (NSCLC) accounting for approximately 85% of diagnosed cases [1–4]. Among NSCLC subtypes, adenocarcinoma, squamous cell carcinoma, and large cell carcinoma exhibit different biological characteristics. Therefore, accurate and early differentiation among these subtypes is vital for treatment and predictions [5-6].

Conventional diagnostic methods rely heavily on biopsy and histopathological investigation, which is time-consuming. Nowadays, chest computed tomography (CT) imaging has become a broadly used non-invasive tool for detecting lung abnormalities. Interpreting 3D CT scans manually can be time-consuming and prone to inter-observer variability, particularly when identifying slight structural variations across different Cancer subtypes. Moreover, variability in tumor shape, location, and size across patients presents significant challenges for automated classification systems.

Deep learning methods have proven effective in analyzing volumetric data. Among them, 3D convolutional neural networks (3D CNNs) particularly useful because they can capture complex spatial features. However, most existing models computationally intensive and may not focus effectively on the most informative regions within CT scans. This can lead to suboptimal performance, especially when training on limited medical datasets [8-11].

To overcome these limitations, we propose Attention-LiteVoxResNet, a lightweight yet powerful 3D CNN framework designed for the two-stage analysis of chest CT scans. In the first stage, the model distinguishes normal lung tissue from lung Cancer cases. If Cancer is detected, the second stage further classifies it into one of three non-small cell lung Cancer subtypes: adenocarcinoma, large cell carcinoma, or squamous cell carcinoma. The architecture integrates residual learning for better feature propagation and a channel-wise attention mechanism to highlight clinically significant regions, improving the model's ability to separate visually similar conditions. By maintaining a compact design, Attention-LiteVoxResNet achieves high accuracy with lower computational cost, making it well-suited for practical use in clinical workflows.

2. LITERATURE REVIEW

With advancement in artificial intelligence, computer-aided diagnostic tools have been developed. This helps to classify lung Cancer at an earlier stage. Many studies have investigated the use of 3D CNNs to process volumetric CT data for lung disease detection [12-15]. Shen et al. [16] introduced a multi-scale CNN framework to detect lung nodules, demonstrating the importance of hierarchical feature learning. Similarly, Setio et al. [17] evaluated various deep learning models on the LIDC-IDRI dataset,

highlighting the potential of 3D architectures to capture volumetric features.

The use of residual learning is initiated by He et al. [18] and has since been adopted in medical imaging models to mitigate vanishing gradient issues during deep network training. In the context of volumetric data, VoxResNet is proposed by Chen et al. [19] for 3D brain segmentation, using stacked residual modules tailored for voxel-wise operations. Their design efficiently propagated features across depth, height, and width while maintaining gradient flow.

However, standard deep networks often treat all spatial regions with equal importance, which may dilute diagnostically relevant features. To overcome this, attention mechanisms have gained attention in the field of medical imaging. Oktay et al. [20] introduced an attention U-Net for organ segmentation, demonstrating that spatial attention helps the network prioritize relevant regions. Similarly, Roy et al. [21] implemented squeeze-and-excitation (SE) blocks in their medical segmentation models to reweight channel-wise features, improving performance in limited data scenarios.

In lung Cancer classification, Deshpande et al. [22] proposed a deep feature fusion method combining CNN features with handcrafted radiomics, showing improved performance in subtype classification. However, such methods often require separate feature extraction and classification stages. More recently, end-to-end 3D CNN models such as those by Liang et al. [23] have been employed for NSCLC classification. Yet, these models typically large in size and computationally expensive, limiting their clinical applicability.

Given these gaps, there is a growing interest in lightweight architectures that balance efficiency and accuracy. The proposed Attention-LiteVoxResNet in this study builds upon the strengths of residual learning and 3D convolutions, while introducing an attention-augmented residual block to focus on informative regions in CT volumes. This design aims to offer both interpretability and improved classification accuracy for lung Cancer subtypes**.**

3. METHODOLOGY

The proposed Attention-LiteVoxResNet framework performs lung CT scan analysis in two stages:

1. Stage 1 – Binary classification: Determine whether a volumetric CT scan is *normal* or *Cancerous*.
2. Stage 2 – Subtype classification: If a scan is identified as Cancerous,

classify it into one of three subtypes—adenocarcinoma, large cell carcinoma, or squamous cell carcinoma.

Both stages the same Attention-LiteVoxResNet backbone architecture, which combines residual learning and attention mechanisms for efficient 3D feature extraction. The only difference lies in the final classification layer—binary in Stage 1 and three-class softmax in Stage 2.

(a) Overall Attention-LiteVoxResNet architecture

(b) Standard VoxRes module

(c) Proposed attention-enhanced VoxRes block

The overall architecture of the proposed Attention-LiteVoxResNet is illustrated in Figure 1.

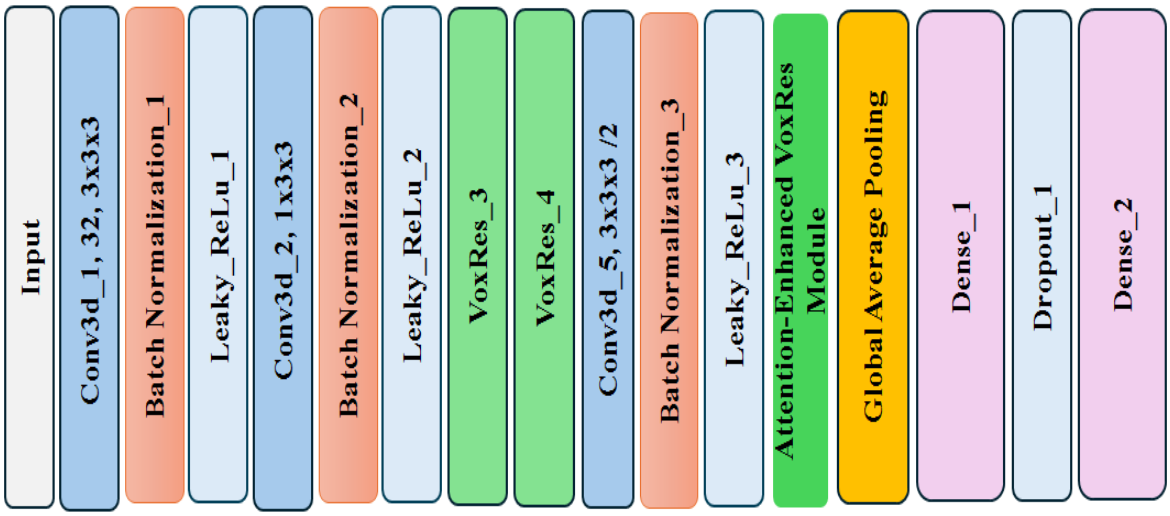

Figure 1 (a) Architecture of the Proposed Attention-LiteVoxResNet

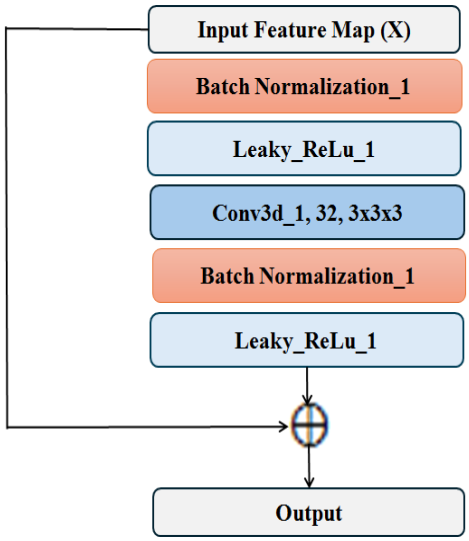

Figure 1 **(b) VoxRes Module**

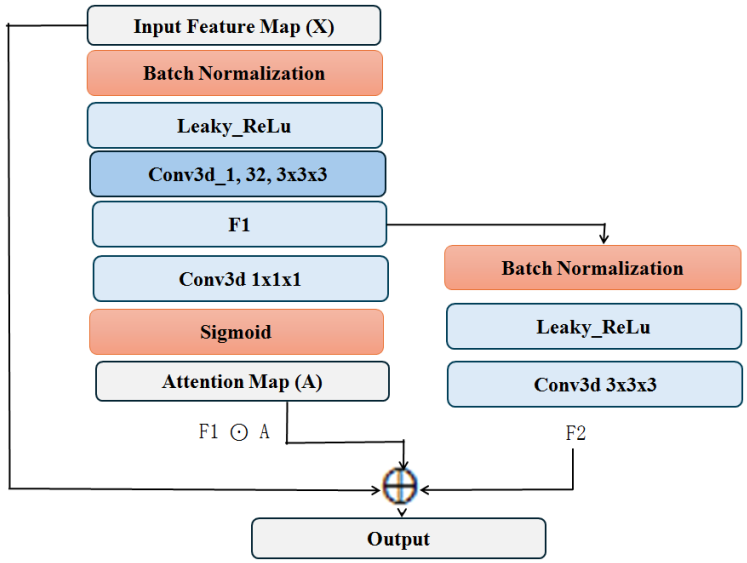

Figure 1 (c) Proposed Attention-Enhanced VoxRes Block

3.1 Data augmentation

To prep the input for the Attention-LiteVoxResNet model, volumetric chest CT scans are collected for three lung Cancer subtypes. A fixed number of 32 axial slices are selected from each scan to represent the lung region effectively. These slices are resized to a standard resolution of 256×256 pixels to maintain uniformity in spatial dimensions across all samples. Pixel intensities are normalized to a range between 0 and 1 to improve numerical stability during training. Finally, the slices are stacked in sequential order to form a 3D input tensor of shape (32, 256, 256, 1), which served as input to the proposed 3D CNN model.

To enhance the model's capability, data augmentation techniques applied. This is done by transforming existing CT images in various ways like rotating the images by random angles such as 90°, 180°, or 270° introduces orientation changes, while horizontal and vertical flipping helps the model learn from mirrored anatomical structures. Adjusting the contrast simulates variations in imaging conditions, making the model more adaptable to different scan qualities. Adding Gaussian noise helps the model become resilient to sensor-level distortions and minor artifacts. Additionally, blending two images through interpolation creates new training samples, providing smoother transitions and

improving robustness across classes. Some sample enhanced images shown in Figure 2.

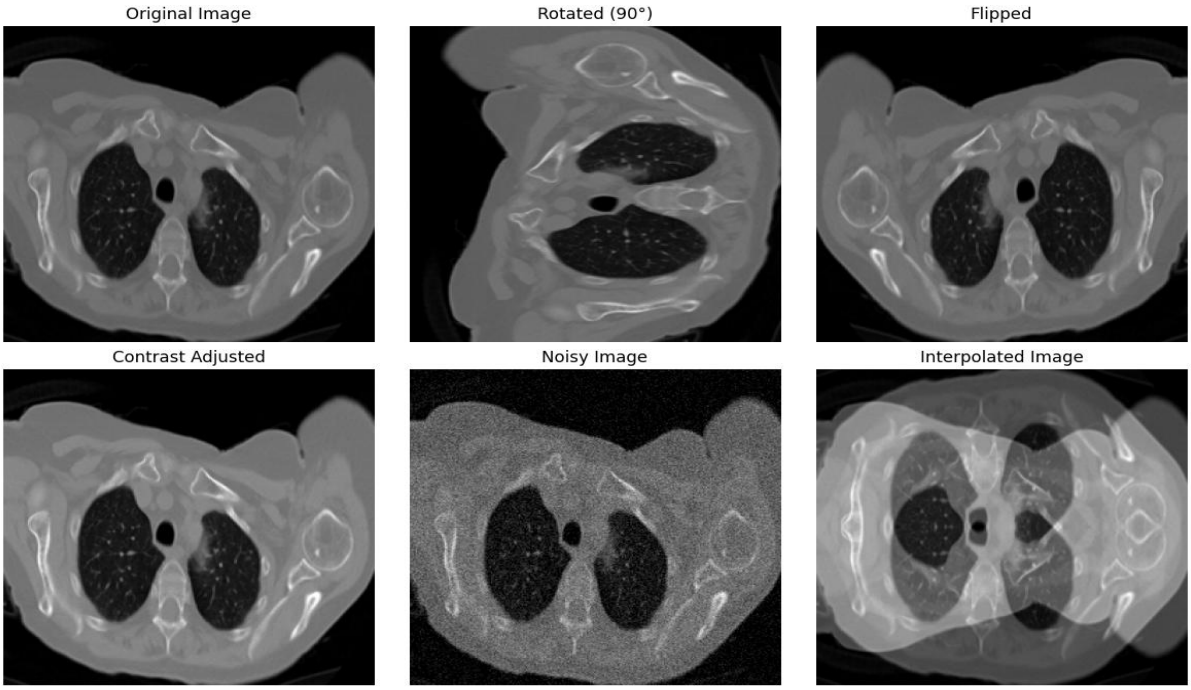

Figure 2 Illustration of preprocessing and augmentation techniques applied to chest CT images

3.2 Feature Extraction using LiteVoxNet

The input to the model is a volumetric CT tensor $X\epsilon\ \mathbb{R}^{32\times256\times256\times1}$, representing 32 axial slices per patient. A 3D convolutional layer with 32 filters of size 3×3×3 is first applied to extract local spatial and depth-wise features:

$$X_1 = Conv3D_{3\times3\times3}(32)(X) \quad (1)$$

This is followed by batch normalization and LeakyReLU activation. A second 3D convolution with 64 filters and a stride of 2 is then used to reduce spatial resolution and increase feature depth:

$$X_2 = Conv3D_{3\times3\times3}(64, stride = 2)(LeakyReLU\big(BatchNorm(X_1)\big)) \quad (2)$$

Two stacked VoxRes blocks applied to learn deeper feature representations through residual learning. Each block contains a pair of 3D convolutional layers connected through residual links.

$$Z_1 = Conv3D_{3\times3\times3}(LeakyReLU(BatchNorm(X_2))) \quad (3)$$

$$Z_2 = Conv3D_{3\times3\times3}(LeakyReLU(BatchNorm(Z_1))) \quad (4)$$

$$Y = X_2 + Z_2 \quad (5)$$

The residual connection permits the network to learn the mapping $F(X) = Z_2$, enabling deeper training without gradient vanishing, using the identity:

$$Y = X + F(X) \quad (6)$$

To further capture global context, another downsampling convolution is applied:

$$X_3 = Conv3D_{3\times3\times3}(128, stride = 2)(Y) \quad (7)$$

This increases the feature channels while reducing spatial dimensions, preparing the feature map for attention-based refinement.

3.3 Attention-Augmented VoxRes Block

A channel-wise attention mechanism is added to the VoxRes block to enhance the model's focus on important diagnostic as within 3D CT scans. This helps the model give more weight to features that critical for accurate interpretation. This mechanism enables the network to prioritize spatial features that most relevant.

The input feature map (from the previous layer) undergoes batch normalization followed by LeakyReLU activation. Then, a 3D convolution is applied. This helps extract richer spatial features from the input volume.

$$F_1 = Conv3D_{3\times3\times3}(LeakyReLU(BatchNorm(X_3))) \quad (8)$$

To compute the attention weights, a separate 1×1×1 convolution is applied on the output of the previous step, trailed by a sigmoid activation. This results in a channel-wise attention map with values between 0 and 1, which represents the significance of each voxel in the feature map.

$$A = \sigma(Conv3D_{3\times3\times3}(F_1)), A\epsilon[0,1]^{D\times H\times W\times C} \quad (9)$$

The generated attention map is then multiplied element-wise with the feature map. This operation emphasizes the

informative parts of the feature map while suppressing the less relevant ones.

$$F_1' = F_1 \odot A \quad (10)$$

The reweighted feature map is again passed through batch normalization, LeakyReLU, and a second 3D convolution. This step helps refine the spatial representation learned so far.

$$F_2 = Conv3D_{3\times3\times3}(LeakyReLU(BatchNorm(F_1'))) \quad (11)$$

Finally, the original input to this block is added back to the output (residual addition), preserving essential features while allowing the new layer to learn corrections. This helps with stable and deep learning.

$$Y_{att} = X_3 + F_2 \quad (12)$$

After the attention-augmented VoxRes block, the resulting 3D feature map is reduced to a 1D vector using Global Average Pooling. This step calculates the mean value of each channel over the entire spatial a, creating a concise representation of the full volume. This vector captures the most relevant global features and has a fixed dimension (e.g., 128).

The pooled feature vector is passed through a fully connected (dense) layer containing 128 neurons, followed by a ReLU activation. To reduce the risk of overfitting, a dropout layer with a rate of 0.4 is applied during training. The final classification is performed using another dense layer with 3 output neurons—one for each lung Cancer subtype. A softmax activation is used to convert the outputs into a probability distribution over the three classes. The summary of the proposed architecture is shown in Table 1.

Table 1 Proposed Attention-LiteVoxResNet Architecture

Step	Operation	Output Shape
1	Input CT Tensor	(32, 256, 256, 1)
2	Conv3D(32) → Conv3D(64, stride=2)	(16, 128, 128, 64)
3	VoxRes Block ×2 (Residual Add)	(16, 128, 128, 64)

4	Conv3D(128, stride=2)	(8, 64, 64, 128)
5	Attention-Enhanced VoxRes	(8, 64, 64, 128)
6	GlobalAvgPooling3D	(128,)
7	Stage 1:Dense(1, sigmoid)	(1,)
8	Stage 2: Dense(3, softmax)	(3,)

3.4 Stage1–Binary Classification Head

In the first stage, the global pooled features from the Attention-LiteVoxResNet backbone passed through a dense layer with 128 units and ReLU activation, followed by a dropout layer with a rate of 0.4 to reduce overfitting. The final output layer contains a single neuron with sigmoid activation, generating a probability score that determines whether the scan is normal or cancerous. The model is optimized using the binary cross-entropy loss function.

3.5 Stage 2 – Subtype Classification Head

In the second stage, scans predicted as Cancerous by Stage 1 re-evaluated using the same Attention-LiteVoxResNet backbone. The only modification is in the classification head: the final layer now contains three neurons with softmax activation, outputting probabilities for each NSCLC subtype—adenocarcinoma, large cell carcinoma, and squamous cell carcinoma. This stage is trained using the categorical cross-entropy loss function.

3.6 Comparative Analysis of Grad-CAM and Grad-CAM++ for 3D CT Volume Analysis

This section presents a comparative study of two explainable AI techniques—Gradient-weighted Class Activation Mapping (Grad-CAM) and its enhanced variant (Grad-CAM++)—applied to lung Cancer classification from 3D CT scans. Both methods produce heatmaps that highlight the most important regions contributing to the model's decision, but Grad-CAM++ introduces refined weighting mechanisms that improve localization and capture finer details, particularly in medical imaging contexts.

3.6.1 Methodology for Comparison

To ensure a fair and consistent evaluation, both Grad-CAM and Grad-CAM++ are applied to the same trained Attention-LiteVoxResNet model using identical 3D chest CT inputs. The final convolutional layer is chosen as the target layer for feature map extraction, as it retains rich semantic information while preserving spatial structure. This allowed both methods to

generate heatmaps based on comparable high-level representations of lung features.

3.6.1.1 Grad-CAM Layer-Wise Workflow for 3D CT Volume Analysis

Grad-CAM is a technique used to visualize which regions of an input image have the greatest influence on a CNN's prediction for a specific class. It produces a heatmap overlay that emphasis the key regions influencing the model's prediction. The explainable lung Cancer classification workflow is shown in Figure 3.

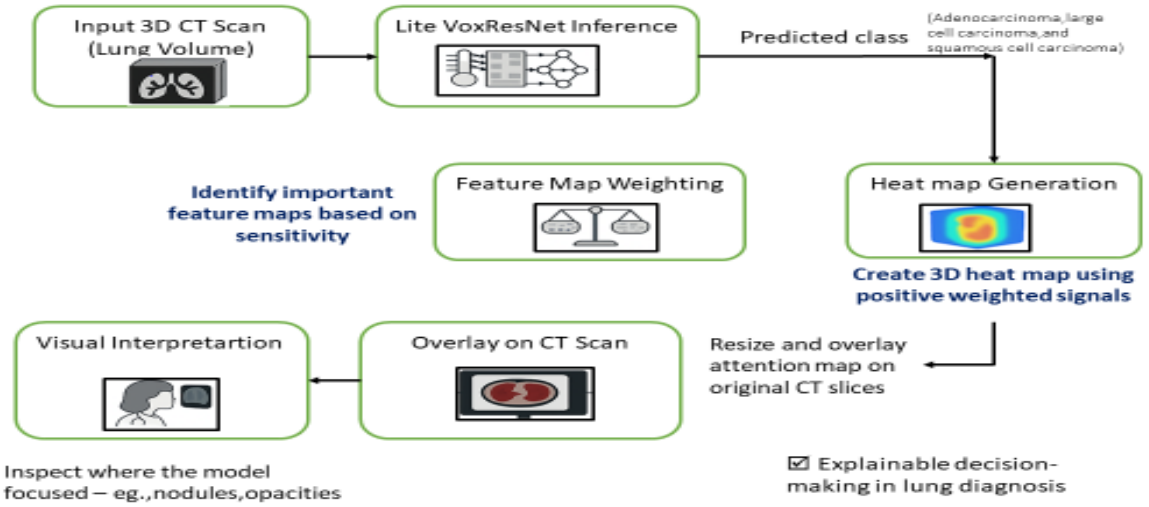

Figure 3 Explainable lung Cancer classification using Attention-LiteVoxResNet with Grad-CAM visualization

In the proposed approach, a 3D computed tomography (CT) scan is first processed by the trained Attention-LiteVoxResNet model, which captures spatial lung features to generate a classification output. To identify the regions that contributed to the prediction, the gradient of the class score is calculated with respect to the feature maps of the final convolutional layer. These gradients then spatially averaged to assign significance weights to each feature map, indicating their relative contribution to the decision. Using these weights, a three-dimensional attention map is constructed by executing a weighted sum of the feature maps, followed by the application of a ReLU activation to emphasize the most informative regions. This heatmap is then scaled to match the original CT slice dimensions and overlaid onto the scan, enabling clear visualization of the as where the model concentrated its attention during classification.

Algorithm 1: Grad-CAM-Based Lung Cancer Classification

- **Input:** 3D CT scan dataset D_s

- **Output:** Grad-CAM heatmap H for target class ccc

Steps:

1. **Data Preprocessing:** Normalize, denoise, and resize CT volumes to match the Attention-LiteVoxResNet input size.
2. **Data Splitting:** Partition D_s into training (80%), validation (10%), and testing (10%) sets.
3. **Model Training:** Train Attention- LiteVoxResNet on the training set and save the trained model N_k.
4. **Forward Pass:** Pass a CT volume through N_k to obtain the predicted score y^c for class c.
5. **Gradient Computation:** Calculate the gradient of y^c with respect to the activations A^kin the last convolutional layer:

$$\frac{\partial y^c}{\partial A^k} \tag{13}$$

6. **Weight Calculation:** Compute channel weights by global average pooling over the gradients:

$$\partial_k^c = \frac{1}{Z}\sum_i \sum_j \sum_l \frac{\partial y^c}{\partial A_{i,j,l}^k} \tag{14}$$

7. **Heatmap Generation:** Create the Grad-CAM heatmap:

$$L_{Grad-CAM}^c = ReLu(\sum_k \alpha_k^c \, A^k) \tag{15}$$

The heatmap generation process continues until all feature maps A^kcontributing to the target class c

8. **Rescaling:** Resize the heatmap to the original CT volume dimensions.
9. **Overlay:** Overlay the heatmap onto the CT slices for visual explanation.
10. **Prediction Output:** Provide both the class prediction and the visualization for interpretability.

3.6.1.2 Grad-CAM++ Layer-Wise Workflow for 3D CT Volume Analysis

Grad-CAM++ is an enhanced visualization technique designed to overcome certain limitations of Grad-CAM, particularly in scenarios where multiple distinct regions

independently influence the model's prediction. By introducing pixel-wise weighting of gradients, Grad-CAM++ generates more precise and finely localized heatmaps. The workflow for applying Grad-CAM++ to the Attention-LiteVoxResNet model is illustrated in Figure 4.

In this approach, a 3D computed tomography (CT) scan is first processed through the trained Attention-LiteVoxResNet, which encodes spatial and semantic features of the lung. To determine which regions most influenced the prediction, the partial derivatives of the class score are computed. These derivatives are calculated with respect to the feature maps of the final convolutional layer. Unlike Grad-CAM, Grad-CAM++ calculates higher-order derivatives and applies pixel-wise weighting to these gradients. This refinement allows the method to capture the contribution of each spatial location more accurately, especially when the target object appears in multiple small regions.

The weighted contributions then combined to form a three-dimensional attention map, trailed by the application of a ReLU activation to highlight only the most relevant as. This heatmap is resized to match the original CT slice dimensions and overlaid on the scan, producing a cler and more precise visualization of the as where the model focused its attention during classification. Figure 4 shows representative examples of Grad-CAM++ visualizations for different lung Cancer subtypes.

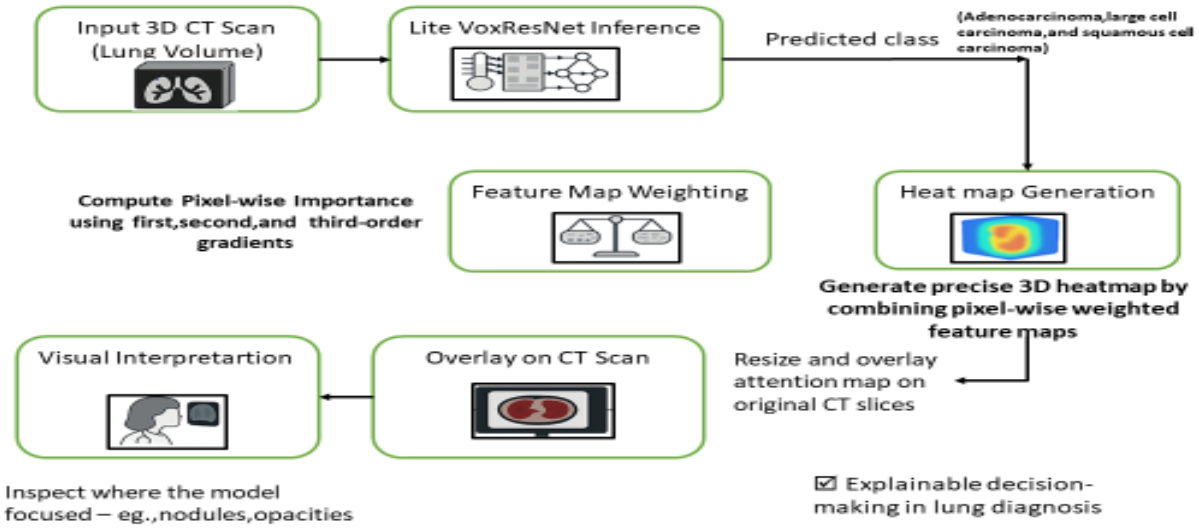

Figure 4 Explainable lung Cancer classification using Attention-LiteVoxResNet with Grad-CAM++ visualization

Figure 5 presents a comparative visualization of Grad-CAM and Grad-CAM++ outputs for the proposed Attention-LiteVoxResNet model representing three different classes. Across all classes, Grad-CAM generates broad activation regions, capturing the general areas as contributing to the model's decision. In contrast, Grad-CAM++ produces more focused and sharper attention maps, effectively highlighting smaller and clinically relevant structures, such as nodules or localized opacities. For instance, in Class 1 and Class 2 cases, Grad-CAM++ shows concentrated activations around lesion boundaries, which may aid radiologists in better understanding the decision rationale. This comparative analysis demonstrates that Grad-CAM++, due to its improved sensitivity to multiple object instances, can provide more interpretable and diagnostically relevant explanations for model predictions.

Algorithm 2: Grad-CAM++-Based Lung Cancer Classification

- **Input:** 3D CT scan dataset D_s
- **Output:** Grad-CAM heatmap H for target class c

Steps:

1. **Data Preprocessing:** Normalize, denoise, and resize CT volumes to match the Attention-LiteVoxResNet input size.
2. **Data Splitting:** Partition D_s into training (80%), validation (10%), and testing (10%) sets.
3. **Model Training:** Train Attention-LiteVoxResNet on the training set and save the trained model N_k.
4. **Forward Pass:** Pass a CT volume through N_k to obtain the predicted score y^c
5. **First-Order Gradient Computation**: Calculate

$$\frac{\partial y^c}{\partial A^k_{i,j,l}} \tag{16}$$

for each spatial location in the last convolutional layer.

6. **Higher-Order Gradient Computation:** Calculate

$$\frac{\partial^2 y^c}{(\partial A^k_{i,j,l})^2} \; and \; \frac{\partial^3 y^c}{(\partial A^k_{i,j,l})^3} \tag{17}$$

to capture pixel-level contributions.

7. **Pixel-wise Weighting**: Compute spatial weights:

$$w_{i,j,l}^{k} = \frac{\frac{\partial^2 y^c}{\left(\partial A_{i,j,l}^{k}\right)^2}}{\left(2\frac{\partial^2 y^c}{\left(\partial A_{i,j,l}^{k}\right)^2} + \sum_{a,b,d} A_{a,b,d}^{k} \frac{\partial^3 y^c}{\left(\partial A_{i,j,l}^{k}\right)^3}\right)} \quad (18)$$

8. **Final Weight Calculation:**

$$\alpha_k^c = \sum_{i,j,l} w_{i,j,l}^{k} \,.\, ReLu\left(\frac{\partial y^c}{\partial A_{i,j,l}^{k}}\right) \quad (19)$$

9. **Heatmap Generation:**

$$L_{Grad-CAM++}^{c} = ReLu\left(\sum_k \alpha_k^c \; A^k\right) \quad (20)$$

The heatmap generation process terminates when all activation maps A^k have been combined with their respective weights α_k^c and the ReLU activation has been applied.

10. **Rescaling:** Resize the heatmap to match the original CT volume dimensions.
11. **Overlay:** Overlay the heatmap on the CT slices for fine-grained visual explanation.
12. **Prediction Output:** Present the classification result alongside the Grad-CAM++ visualization.

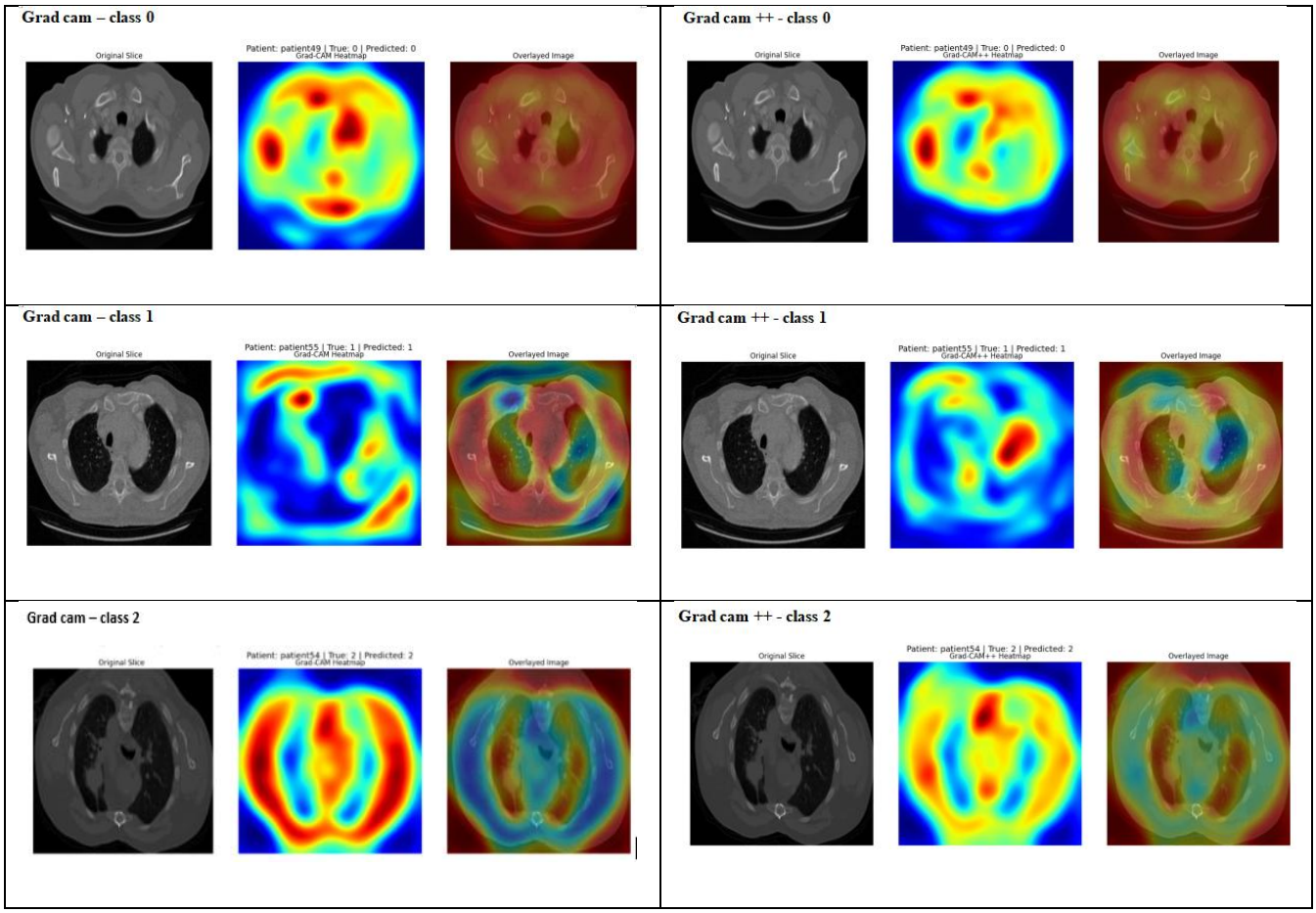

Figure 5. Comparison of Grad-CAM and Grad-CAM++ visualizations for Attention-LiteVoxResNet predictions on representative 3D CT slices across three classes (Class 0, Class 1, and Class 2).

For each class, the first column shows the original CT slice, the second column displays the generated heatmap, and the third column presents the heatmap overlaid on the original image. Grad-CAM++ produces more localized and fine-grained attention maps compd to Grad-CAM, highlighting subtle regions that contribute to the classification decision.

4. EXPERIMENTAL RESULTS

4.1 Experimental Setup

The performance of the proposed Attention-LiteVoxResNet model for lung Cancer classification is assessed using the kaggle platform with GPU acceleration. All experiments are executed in a Kaggle notebook environment equipped with NVIDIA Tesla P100/T4 GPUs and 16–32 GB of RAM. Training is performed using a batch size of 8 over 100 epochs. All CT volumes are pre-processed to include 32 axial slices per scan, resized to 256×256 pixels.

4.2 Dataset

This study utilized a combined CT scan dataset compiled from multiple publicly available sources. Normal lung CT images are obtained from the COVID-Net Open Source Initiative on Kaggle, while Cancer cases are collected from the NSCLC Radiomics dataset available through The Cancer Imaging Archive (TCIA) [https://wiki.Cancerimagingarchive.net/display/Public/NSCLC-Radiomics].

A total of 19,450 CT images are used, corresponding to Normal and Lung Cancer cases for Stage 1 classification, and three NSCLC subtypes for Stage 2 classification. The dataset is split into 80% for training, 10% for validation and 10% for testing. Table 2 shows the detailed distribution.

Table 2 Dataset Distribution

Class		Total Images	Training Set Images	Testing Set Images
Normal		2230	1784	446
Lung Cancer	Adenocarcinoma	5500	4400	1100
	Large Cell Carcinoma	6200	4960	1240
	Squamous Cell Carcinoma	5520	4416	1104

4.3 Performance Metrices

4.3.1 Accuracy

Accuracy is the simplest but significant performance metric which is a ratio of accurately predicted observations to total observations

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

4.3.2. Precision

Precision measures the ratio of true positive predictions to all instances that the model classified as positive. It indicates how effectively the model minimizes false positive predictions.

$$Precision = \frac{TP}{TP+FP} \quad (22)$$

4.3.3. Recall (Sensitivity)

Recall, quantifies the proportion of actual positive cases that the model correctly identifies. It indicates the model's effectiveness in reducing false negatives.

$$Recall = \frac{TP}{TP+FN} \quad (23)$$

4.3.4. F1-Score

The F1-score is the harmonic mean of precision and recall, offering a balanced metric when both are equally important.

$$F1 = \frac{2 \times precision \times Recall}{precision + Recall} \quad (24)$$

4.3.5. Cohen's Kappa (κ)

Cohen's Kappa quantifies the agreement between predicted and true classes, adjusting for the agreement that could happen by chance. It is more robust than accuracy in imbalanced datasets.

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (25)$$

p_o represents the observed agreement (accuracy), and p_e denotes the agreement expected by chance. The values range from –1, indicating complete disagreement, to +1, representing perfect agreement, with 0 corresponding to chance-level prediction.

4.3.6. Matthews Correlation Coefficient (MCC)

MCC incorporates true positives, true negatives, false positives, and false negatives. It is particularly valuable for evaluating performance on imbalanced datasets.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (26)$$

The coefficient ranges from –1 to +1, where +1 signifies perfect classification, 0 indicates random prediction, and –1 represents complete disagreement.

4.3.7. F2-Score

The F2-score is a variant of the F-measure that emphasizes recall more than precision, making it suitable for medical applications where minimizing false negatives is essential.

$$F_\beta = \frac{(1+\beta^2).Precision.Recall}{(\beta^2.Precision)+Recall}, \beta = 2 \quad (27)$$

By setting β=2, recall is given four times more weight than precision. This ensures that the model prioritizes detecting as many true positive cases as possible.

4.4 Performance Analysis of the Proposed Attention-LiteVoxResNet

In Table 3 the first stage, is trained to regulate whether a CT scan showed normal lung tissue or contained evidence of lung Cancer. The model achieved an overall accuracy of 95%, with a precision of 0.96, recall of 0.94, and an F1-score of 0.95.In the second stage, it trained to determine the Subtype Classification of Lung Cancer.The model achieved an overall accuracy of 0.91%.This shows a strong capability of the model to distinguish this subtype from others.

Table 3 Performance analysis of the Proposed Attention-LiteVoxResNet Model

Classification	Accuracy	Precision	Recall	F1-score	Kappa	MCC	F2-score
Stage1 Binary Classification (Normal, Lung Cancer)	0.95	0.96	0.94	0.95	0.93	0.94	0.95
Stage2 Triclass Classification	0.91	0.93	0.92	0.92	0.90	0.89	0.92

(Adenocarcinoma, LargeCell Carcinoma, SquamousCell Carcinoma)							

Table 4 provides a comparative analysis of several 3DCNN architectures for classifying lung Cancer. The 3D-VGG16 model achieved an accuracy of 82% reflecting consistent performance across all measures. The 3D-AlexNet showed a slight improvement, reaching 84% accuracy. In comparison, 3D-GoogleNet attained 87% accuracy, with higher recall (0.74). The 3D-VoxResNet model, which incorporates residual learning, performed better overall with 90% accuracy. Among all the models, the proposed Attention-LiteVoxResNet showed the best performance, achieving 91% accuracy, and maintaining 0.93 in precision.

Table 4 Performance comparison of various 3D CNN models

3DCNN MODELS	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**Kappa**	**MCC**	**F2-score**
3D-VGG16	82%	0.84	0.78	0.81	0.75	0.72	0.79
3D-Alexnet	84%	0.86	0.81	0.83	0.78	0.76	0.82
3D-GoogleNet	87%	0.89	0.84	0.86	0.82	0.80	0.85
3D-VoxResNet	90%	0.92	0.88	0.90	0.86	0.84	0.89
Proposed Attention-LiteVoxResNet	91%	0.93	0.92	0.92	0.90	0.89	0.92

Figure 6 illustrates the confusion matrix of the model's effectiveness. The highest number of correctly identified cases belongs to Squamous Cell Carcinoma, with 179 instances accurately predicted. Large Cell Carcinoma follows with 176 correct classifications, although some overlap with Adenocarcinoma is evident. Adenocarcinoma had the most misclassifications, often being confused with Large Cell Carcinoma.

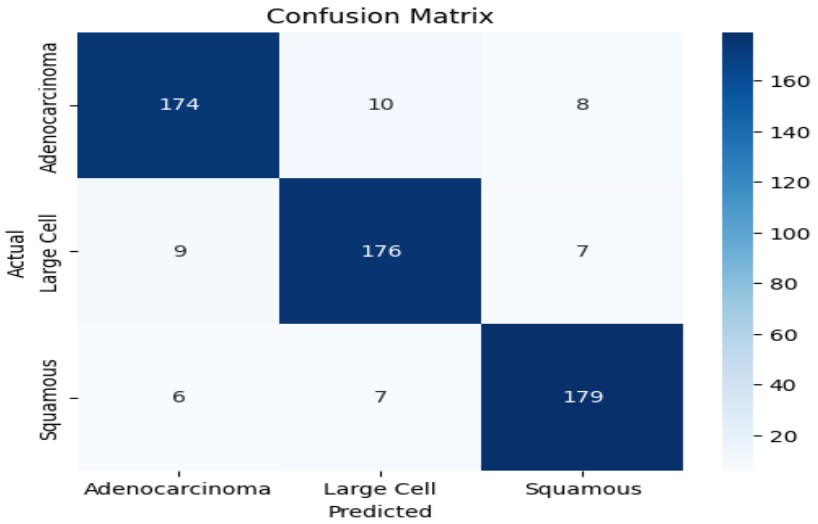

Figure 6 Confusion Matrix

The multiclass ROC curve is illustrated in Figure 7. Class 0 shows an excellent AUC of 1.00, indicating perfect sensitivity and specificity without any misclassification. Class 1 achieves an AUC of 0.90, reflecting strong classification accuracy with minimal false positives. Class 2 follows closely with an AUC of 0.88, also demonstrating effective separation from other classes.

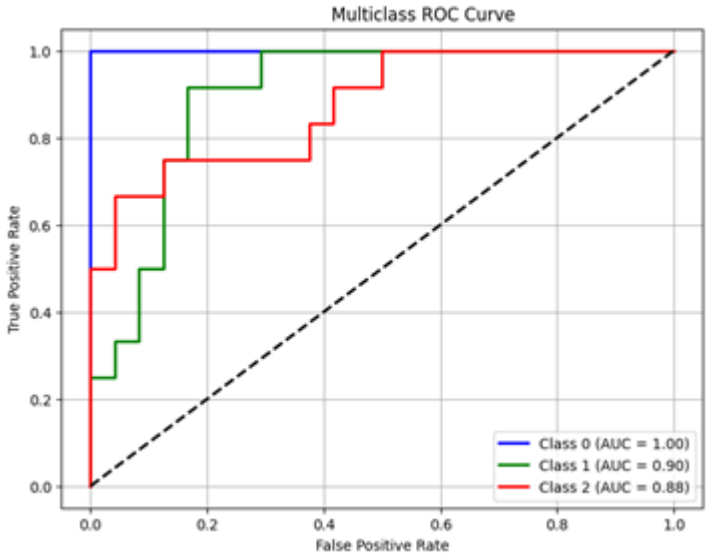

Figure 7 Multiclass ROC curve

Figure 8 illustrates the training and validation performance of the classification model over 20 epochs. The Accuracy over Epochs graph shows a consistent rise in both training and validation accuracy throughout the course of training. The training accuracy rises consistently,

reaching approximately 80% by the final epoch. The validation accuracy also improves, though it fluctuates, peaking at around 87% near the 10th epoch before stabilizing.

Training and validation loss steadily decrease over the course of the epochs. In contrast, the validation loss displays noticeable fluctuations, particularly between epochs 3 and 6. From epoch 7 onward, it decreases sharply and aligns closely with the training loss, reflecting improved model generalization.

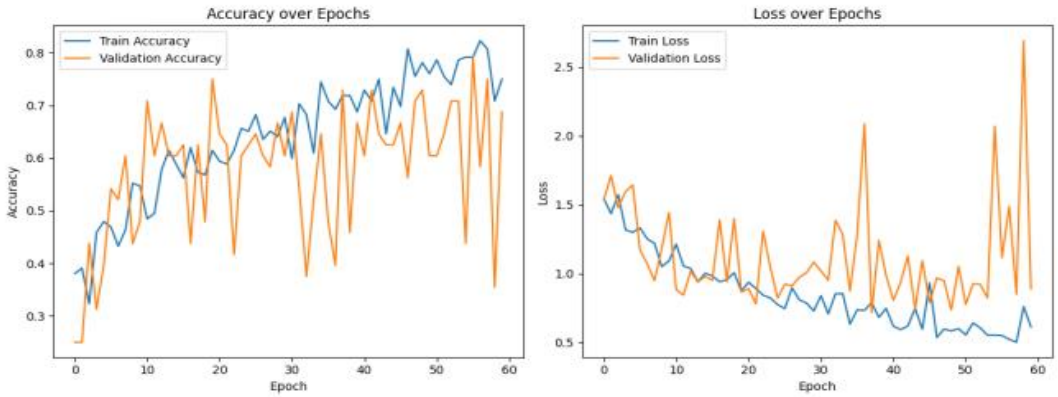

Figure 8 Accuracy/Loss curve

4.5 Ablation Study

Table 5 presents the results of the ablation study, conducted to evaluate the impact of the attention mechanism on the proposed Attention-LiteVoxResNet model. The Attention-LiteVoxResNet without the attention block attained an accuracy of 87%. In contrast, the model incorporating the attention block attained an increased accuracy of 91%. This comparison validates the effectiveness of the model with the integration of attention mechanism.

Table 5 Ablation Study on Attention Integration

Model Variant	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**Kappa**	**MCC**	**F2-score**
VoxResNet Without Attention Block	87%	0.89	0.87	0.88	0.86	0.84	0.87
Attention-LiteVoxResNet	91%	0.92	0.89	0.90	0.90	0.89	0.90

5. Conclusion

This study proposed an efficient 3D convolutional architecture, Attention-LiteVoxResNet, for classifying lung Cancer using volumetric CT data. The model is designed with integrating residual connections and an attention mechanism to highlight critical spatial features. The experimental results demonstrated that the proposed model performs well in distinguishing the sub types. Notably, the inclusion of the attention block led to a marked improvement in classification metrics. Overall, Attention-LiteVoxResNet proves to be a reliable and lightweight solution for lung Cancer subtype classification. Additionally, visualization techniques such as Grad-CAM and Grad-CAM++ were applied to enhance the interpretability of the model. Grad-CAM++ produced more focused and sharper attention maps. Overall, Attention-LiteVoxResNet proves to be a reliable, lightweight, and interpretable solution for lung Cancer subtype classification. Grad-CAM++ adds to its potential as a powerful tool for medical diagnosis and treatment planning.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 Cancers in 185 countries. CA Cancer J Clin. 2021;71:209- 249.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72:7-33.
3. Ettinger DS. Ten years of progress in non-small cell lung Cancer. J Natl Compr Cancer Netw. 2012;10:292-29
4. Gettinger S, Horn L, Jackman D, et al. Five-year follow-up of nivolumab in previously treated advanced non–small-cell lung Cancer: results from the CA209-003 study. J Clin Oncol. 2018;36:1675-1684.
5. Lococo, F., Ghaly, G., Chiappetta, M., Flamini, S., Evangelista, J., Bria, E., ... & Mohammed, A. (2024). Implementation of artificial intelligence in personalized prognostic assessment of lung Cancer: a narrative review. *Cancers*, *16*(10), 1832.
6. Siddique, F., Shehata, M., Ghazal, M., Contractor, S., & El-Baz, A. (2024). Lung Cancer Subtyping: A Short Review. *Cancers*, *16*(15), 2643.
7. Dunn, B., Pierobon, M., & Wei, Q. (2023). Automated classification of lung Cancer subtypes using deep learning and CT-scan based radiomic analysis. *Bioengineering*, *10*(6), 690.
8. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, Tse D, Etemadi M, Ye W, Corrado G, Naidich DP, Shetty S (2019) End-

to-end lung Cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 25(6):954–961.

9. Huang, X., Sun, Y., Tan, M., Ma, W., Gao, P., Qi, L., ... & Li, M. (2022). Three-dimensional convolutional neural network-based prediction of epidermal growth factor receptor expression status in patients with non-small cell lung Cancer. *Frontiers in oncology*, *12*, 772770.
10. Patharia, P., Sethy, P. K., & Nanthaamornphong, A. (2024). Advancements and Challenges in the Image-Based Diagnosis of Lung and Colon Cancer: A Comprehensive Review. *Cancer Informatics*, *23*, 11769351241290608.
11. Pan, L., Feng, Z., & Peng, S. (2022). A review of machine learning approaches, challenges and prospects for computational tumor pathology. *arXiv preprint arXiv:2206.01728*.
12. Yu, J., Yang, B., Wang, J., Leader, J., Wilson, D., & Pu, J. (2020). 2D CNN versus 3D CNN for false-positive reduction in lung Cancer screening. *Journal of Medical Imaging*, *7*(5), 051202-051202.
13. Guo, Y., Song, Q., Jiang, M., Guo, Y., Xu, P., Zhang, Y., ... & Yao, X. (2021). Histological subtypes classification of lung Cancers on CT images using 3D deep learning and radiomics. *Academic radiology*, *28*(9), e258-e266.
14. Muñoz-Aseguinolaza, U., Fernandez-Iriondo, I., Rodríguez-Moreno, I., Aginako, N., & Sierra, B. (2023). Convolutional neural network-based classification and monitoring models for lung Cancer detection: 3D perspective approach. *Heliyon*, *9*(11).
15. Yanagawa, M., Niioka, H., Kusumoto, M., Awai, K., Tsubamoto, M., Satoh, Y., ... & Tomiyama, N. (2021). Diagnostic performance for pulmonary adenocarcinoma on CT: comparison of radiologists with and without three-dimensional convolutional neural network. *European Radiology*, *31*, 1978-1986
16. Shen, W., Zhou, M., Yang, F., et al. (2015). Multi-scale Convolutional Neural Networks for Lung Nodule Classification. *Information Processing in Medical Imaging*.
17. Setio, A. A. A., Traverso, A., de Bel, T., et al. (2017). Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in CT Images. *Medical Image Analysis*, 42, 287–300.
18. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *CVPR*.
19. Chen, H., Dou, Q., Yu, L., & Heng, P.-A. (2018). VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, 170, 446–455.
20. Oktay, O., Schlemper, J., Folgoc, L. L., et al. (2018). Attention U-Net:

Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

21. Roy, A. G., Navab, N., & Wachinger, C. (2018). Concurrent Spatial and Channel Squeeze & Excitation in Fully Convolutional Networks. *MICCAI*.

22. Deshpande, P., Bhatt, M. W., Shinde, S. K., Labhade-Kumar, N., Ashokkumar, N., Venkatesan, K. G. S., & Shadrach, F. D. (2024). Combining handcrafted features and deep learning for automatic classification of lung Cancer on CT scans. *Journal of Artificial Intelligence and Technology*, *4*(2), 102-113.

23. Liang, B., Tong, C., Nong, J., & Zhang, Y. (2024). Histological Subtype Classification of Non-Small Cell Lung Cancer with Radiomics and 3D Convolutional Neural Networks. *Journal of Imaging Informatics in Medicine*, 1-15.