INTERNATIONAL JOURNAL OF APPLIED METHODS IN ELECTRONICS AND COMPUTERS

www.ijamec.org

International Open Access

Volume 14
Issue 01
March, 2026

Research Article

Deep Learning-Based Lung Cancer Classification and Grad-Cam++ With Lime-Supported Explainability Analysis

Yasin DOGANSAH [a,*], ***Inayet Hakki CIZMECI*** [b], ***Murat KOKLU*** [c]

[a] *Department, of Computer Technologies, Afyon Vocational School, Afyon Kocatepe University, Afyonkarahisar, Türkiye*
[b] *Department of Computer Engineering, Afyon Kocatepe University, Afyonkarahisar, Türkiye.*
[c] *Department of Computer Engineering, Faculty of Technology, Selcuk University, Konya, Türkiye.*

ARTICLE INFO

Article history:
Received 10 February 2026
Accepted 25 March 2026

Keywords:
Explainable artificial intelligence,
Lung cancer,
Grad-CAM++,
LIME,
Medical image analysis.

ABSTRACT

This study aims to classify lung cancer using deep learning-based methods and to interpret the obtained model results using explainable artificial intelligence (XAI) approaches. The dataset used in this study consisted of 1933 chest computed tomography images, which were classified as normal, benign, and malignant. In the classification process, EfficientNet-B4, MobileNetV3-Large, and ResNet50 deep learning architectures were trained using a transfer learning approach, and their performance was evaluated using a 10-fold cross-validation method. The performance of the models was analyzed using accuracy, precision, recall, F1-score, and ROC-AUC metrics. According to the results obtained in the study, the MobileNetV3-Large model showed the highest overall performance with 97.05% accuracy and 99.70% ROC-AUC value. The EfficientNet-B4 and ResNet50 models also provided high and balanced performance values, achieving effective results in lung cancer classification. Grad-CAM++ and LIME methods were used to make model decisions more clinically reliable and interpretable. Grad-CAM++ analyses reveal that the models primarily focus on anatomically significant regions within the lung parenchyma during classification. LIME analyses, on the other hand, have enabled the superpixel-level explanation of local regions that contribute most to classroom decisions. Explainability maps obtained for normal, benign, and malignant classes showed that each class exhibited distinct spatial attention and contribution patterns. In conclusion, this study demonstrates that deep learning-based lung cancer classification offers a more reliable and transparent framework for clinical decision support systems, not only when evaluated with high-performance metrics but also when considered in conjunction with Grad-CAM++ and LIME-supported explainability analyses. The findings demonstrate that explainable artificial intelligence approaches play a significant role in improving model reliability in medical image analysis.

1. Introduction

Globally, lung cancer is one of the leading causes of death and morbidity among all cancer types. This type of cancer is increasing every year, with an average of 2.5 million new cases diagnosed annually and causing an average of 2 million deaths per year [1, 2]. When lung cancer is diagnosed early and cancerous areas are removed, it becomes more difficult to treat in advanced stages [3]. In lung cancer, early diagnosis is as vital as identifying the type of nodule detected early and removing it surgically [4]. Patients with advanced disease that cannot be surgically removed are treated with radiation therapy and chemotherapy. Despite this treatment, survival generally does not exceed 12 months. Due to the difficulty of treatment and rapid spread, the average 5-year survival rate for lung cancer is less than 20%. Therefore, early detection of lung nodules and complete removal of the mass before metastasis is critically important to increase survival rates [1, 3].

* Corresponding author. E-mail address: *ydogansah@aku.edu.tr*
DOI: 10.58190/ijamec.2026.164

Lung cancer has four different clinical stages. Stage I is the early diagnosis stage. Intervention at this stage increases the survival rate to 90%. Stage IV, the final stage, is the late diagnosis stage. At this stage, if treatment is delayed, the mortality rate reaches 90% [5]. Before initiating treatment, accurately determining the stage of the disease is critical for selecting the appropriate treatment method [6]. The first step in diagnosing lung cancer is a chest X-ray, which uses radiographs (chest X-rays) [7]. If a nodule is suspected after radiological evaluation of lung images, diagnosis is attempted using computed tomography (CT) or high resolution computed tomography (HRCT). In addition to these methods, other imaging techniques are also used for nodule characterization [8]. No matter how advanced these imaging methods are, lung anomalies can still be missed due to human error during the radiological evaluation process. Therefore, the use of computer assisted diagnostic systems aims to minimize human error. One of the primary goals of computer assisted diagnostic systems is to detect anomalies in images by using quantitative features extracted from those images. Additionally, by autonomously classifying the anomalies detected in the images as malignant or benign, it aims to support the clinical decision-making process in cases of differing opinions among experts [9-11]. In addition, computer aided systems such as deep learning approaches help analyze more data in a shorter time and make more accurate diagnoses compared to traditional anomaly detection methods [12]. Therefore, the success of deep learning in radiological image analysis and diagnoses is critical [13].

Explainable Artificial Intelligence (XAI), which offers transparency to users in AI-powered computerized diagnostic systems, refers to technical methods that enable users to understand the predictions produced by the models and the resulting outputs. [14]. Today, explainable artificial intelligence models are being used effectively in important fields such as finance, defense, and medicine. [15]. Artificial intelligence approaches, often described as "black boxes," can potentially lead to serious consequences, especially when used in vital fields such as medicine, due to the lack of transparency in decision-making processes and the inability to justify predictions. Therefore, the explainability of artificial intelligence models plays a significant role in medical applications. For healthcare professionals to trust artificial intelligence (AI) models during the diagnosis and treatment phase, the decision making capabilities of these AI models must be transparent, effective, and traceable [16].

This study aims to detect anomalies in the lungs using deep learning methods with computed tomography (CT) scans of the lungs. A deep learning model is used to first detect lung cancer and then determine whether the lung cancer is benign or malignant. It also aimed to ensure the explainability of the findings by interpreting the results. This will increase the reliability of diagnoses and assist healthcare professionals in their diagnosis [17].

2. Literature Review

Lung cancer is one of the leading causes of cancer-related deaths worldwide, and early diagnosis directly impacts patient survival. Radiological imaging techniques are used as primary diagnostic tools in the detection of nodules. In recent years, the use of deep learning models in the analysis of radiological images has enabled high success rates in this field. However, the black box nature of these models creates problems of reliability and transparency, especially in clinical applications. In this context, explainable artificial intelligence (XAI) approaches are gaining importance in making model decisions interpretable.

In their 2017 study, Wang et al. developed a poorly controlled multi-label image classification and disease localization framework on the “ChestX-ray8” dataset, which contains 108,948 chest X-ray images from 32,717 unique patients. Natural Language Processing (NLP) techniques were used to automatically extract eight different disease labels from radiology reports: Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, and Pneumothorax. In their study comparing different deep learning architectures, they achieved AUC values of 0.8141 for cardiomegaly, 0.7891 for pneumothorax, and 0.7362 for effusion using the ResNet-50 based model [18]. In 2017, Li and colleagues, using computed tomography (CT) images that provide detailed visualizations for more thorough examination and diagnosis of lung diseases, achieved an 83.03% success rate in predicting malignant lung cancer nodules with a hybrid model using 2620 CT images [19]. Mohandass and colleagues achieved 99.43% accuracy for cancer classification and 99.34% accuracy for non-cancerous classification with their AtCNN-DenseNet-201-TL-NBOA-CT hybrid model, developed using 888 CT scans and 1,186 lung nodules on the LUNA16 dataset [20].

In studies on lung cancer classification, Sivakumar et al. demonstrated high performance in 2024 by achieving 99.87% accuracy and 99.75% sensitivity with their ADBN + LightGBM (Adaptive Deep Belief Network) model applied to the LUNA16 and Kaggle DSB datasets [21]. Mohandass and colleagues achieved 99.43% accuracy in cancer classification and 99.34% accuracy in non-cancer classification with the AtCNN-DenseNet-201 TL-NBOA-CT optimized hybrid model developed using 888 CT scans and 1186 images from the LUNA16 dataset in 2024 [20]. In the field of segmentation and nodule detection, Budati and Karumuri achieved a 99.75% CPM in 2024 with their SbYSF (Sailfish + YOLO) based detection model developed using the LUNA16 dataset [22]. In terms of

segmentation, Chen et al. (2020) achieved a segmentation accuracy of over 99% with their proposed LDDNet model on the LIDC-IDRI dataset [23].

Studies on model efficiency and the use of different data types also hold a significant place in the literature. In their 2020 study, Masud and colleagues achieved 97.90% accuracy in terms of mobile device compatibility using a light CNN-based Light Deep Model on 1,279 images in the LIDC dataset [24, 25]. Focusing on multimodal imaging, Jose and her colleagues achieved 96.4% accuracy, 83.3% sensitivity, and 91.7% specificity with their DenseNet-based CNN model containing dilated convolutions, developed using 6,034 CT / X-ray images from the GitHub dataset [26]. In addition, Singh (2024), who investigated the effectiveness of hybrid architectures using attention mechanisms, obtained AUC values of 0.8556 for ChestX-ray14 and 0.904 for five pathologies in the CheXpert dataset with the SA-DenseNet121 model she developed on ChestX-ray14 and CheXpert datasets [27]. In their study conducted on a dataset containing 7560 chest X-ray (CXR) images, Veeramani et al. (2025) achieved 98% precision and 97% accuracy in diagnosing multi-class lung disease using an improved U-Net architecture for separating lung regions and an Inception-based XAI-TRANS transfer learning model for classification. [28]. Megat Ramli et al. (2025), supporting these high success rates obtained in individual studies at the general literature level, evaluated 34 studies on the detection of lung nodules in chest X-ray images and found that artificial intelligence algorithms showed significant potential in detecting small nodules, especially compared to radiologists (average AUC of 0.81), with sensitivity ranging from 56.4% to 95.7%, specificity ranging from 71.9% to 97.5%, and AUROC values ranging from 0.89 to 0.99 [29].

In studies to ensure the explainability of the models used, Naz et al. (2023) achieved a classification accuracy of 93% in the COVID-CT dataset and 97% in the COVID-Net dataset with a ResNet50-based transfer learning model they trained using COVID-CT and COVID-Net datasets to interpret lung diseases such as edema, tuberculosis, nodules, and pneumonia. They used the LIME technique to visualize the regions the model focused on, explain the classification results, and build clinical confidence [30]. Using a similar explainable artificial intelligence approach, Alqhatani et al. (2025) classified lung cancer stages (benign, malignant, and normal) on 1190 computed tomography (CT) images from the IQ-OTH/NCCD dataset using the EfficientNet-B0 architecture, achieving a high accuracy and precision rate of 99%, while generating visual heat maps using the Grad-CAM technique to make the model's decision processes transparent [31].

This study aims to analyze CT images of benign, malignant, and normal lung cancer using deep learning-based methods and to interpret the resulting classification decisions using explainable artificial intelligence approaches. Although numerous classification studies on lung CT images exist in the literature, studies that explain the decision-making mechanism in a clinically transparent manner, as well as the model performance, are limited. In this context, the multi-class classification process, implemented using convolutional-based deep learning models, was supported by Grad-CAM++ and LIME methods, and the anatomical and pathological regions that the models focused on during decision-making were visually analyzed. The proposed approach contributes to the literature by not only improving classification success but also providing a explainable decision support framework that allows clinicians to evaluate the reliability of model outputs.

This article is structured as follows. The first section discusses the clinical significance of lung cancer, the necessity of early diagnosis, and the role of deep learning-based artificial intelligence approaches in this field. The second section summarizes current literature studies on lung cancer classification and explainable artificial intelligence. The third section details the dataset used, preprocessing steps, deep learning architectures, experimental environment, performance metrics, and explainability methods. In the fourth chapter, the experimental results obtained and the Grad-CAM++ and LIME-based explainability analyses are presented and discussed comparatively. The final section evaluates the overall results of the study and provides suggestions for future research.

3. Material and Methods

In this study, computed tomography (CT) images of lung cancer were obtained and analyzed using deep learning. In the first stage, the images were preprocessed to be compatible with deep learning models, and then multi-class classification was performed using EfficientNet-B4, MobileNetV3-Large, and ResNet50 architectures. The models' performance was analyzed using standard evaluation criteria such as accuracy, precision, sensitivity, F1-score, and ROC-AUC. The accuracy and generalizability of the proposed approach were evaluated using a 10-fold cross-validation method. Furthermore, to improve the interpretability of model predictions, Grad-CAM++ and LIME explanatory AI methods were applied, and regions contributing to classification decisions were visually analyzed.

3.1. Dataset Description

The dataset used in this study was created from chest computed tomography (CT) images collected from the National Chest Diseases Hospital (NIDCH) in Dhaka, Bangladesh, with the aim of supporting medical imaging research focused on the classification and early diagnosis of lung-related diseases through artificial intelligence and deep learning methods. This dataset, divided into three

diagnostic categories, contains a total of 1933 images: 1050 malignant, 467 benign, and 416 normal cases [32]. This dataset is important for training, validating, and comparing AI-based models developed for autonomous diagnosis to detect respiratory diseases and lung anomalies. In healthcare settings, the software contributes to the development of clinical decision support systems.

3.2. Dataset Preprocessing

In this study, a dataset consisting of lung computed tomography scans representing different pathological conditions was first obtained. The dataset, after being checked, was uploaded to Google Drive and transferred to the Google Colab platform. Next, the images were processed through standard data preprocessing steps to meet the input requirements of deep learning models, and the dataset was divided into training and validation subsets. Lung CT images were converted to RGB format, rescaled to fixed sizes depending on the architecture used (380×380 for EfficientNet-B4, 224×224 for MobileNetV3-Large and ResNet50), and converted to tensor format. To ensure compatibility with pre-trained ImageNet weights, the images were normalized using the mean and standard deviation values from the ImageNet dataset.

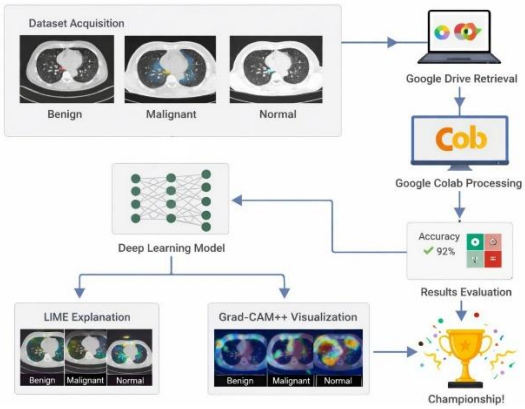

Figure 1. Data preprocessing

Figure 1 presents a deep learning and explainable artificial intelligence workflow (AI-generated) for autonomously distinguishing benign, malignant, and normal classes from lung CT images.

In the classification phase, three different deep learning-based convolutional neural network architectures—EfficientNet-B4, MobileNetV3-Large, and ResNet50—were used, and the models were trained using transfer learning. These architectures are optimized to autonomously learn distinctive structural and textural features in lung computed tomography images. Following the training process, the performance of the models was analyzed using evaluation metrics such as accuracy, precision, sensitivity, F1-score, and ROC-AUC. To improve the interpretability of model decisions, explainable artificial intelligence approaches have been utilized. Using the Grad-CAM++ method, a demonstrable artificial intelligence approach, critical regions that the models focused on during classification were visualized through heat maps. With the LIME method, images were analyzed on a superpixel basis to identify local regions that contributed most to classification decisions. These visualizations have allowed for a more transparent and meaningful clinical evaluation of the models' decision-making mechanisms. In conclusion, these proposed approaches ensure not only high classification performance but also that the model outputs are explainable and reliable.

Table 1. Model Structure and Training Parameters

Parameter	EfficientNet-B4	MobileNetV3-Large	ResNet50
Model Architecture	EfficientNet-B4	MobileNetV3-Large	ResNet50
Pretrained Weights	ImageNet (Imagenet1k_V1)	ImageNet (Imagenet1k_V1)	ImageNet (pretrained=True)
Input Image Resolution	380 × 380	224 × 224	224 × 224
Number of Classes	Dataset-derived (ImageFolder)	Dataset-derived (ImageFolder)	3 (Benign, Malignant, Normal)
Loss Function	CrossEntropyLoss	CrossEntropyLoss	CrossEntropyLoss
Optimizer	AdamW	AdamW	Adam
Learning Rate	1×10^{-4}	1×10^{-4}	1×10^{-4}
Batch Size	16	16	16
Number of Epochs	3	5	5
Cross-Validation Strategy	10-Fold	10-Fold	10-Fold Stratified
Evaluation Metrics	Accuracy, F1-score, Precision, ROC-AUC Recall,	Accuracy, F1-score, Precision, ROC-AUC Recall,	Accuracy, F1-score, Precision, ROC-AUC Recall,
ROC-AUC Calculation	OvR (multi-class)	OvR (multi-class)	OvR (multi-class)

Table 1 presents the model structure and training parameters of the EfficientNet-B4, MobileNetV3-Large, and ResNet-50 deep learning architectures used for multi-class classification of computed tomography images of lung cancer. Each model is configured taking into account the input image resolution specific to its architectural structure. Computed tomography images have been rescaled to match the input dimensions expected by the relevant networks. The models were trained on the ImageNet dataset using pre-trained weights, and their performance was evaluated using standard grading measures such as accuracy, precision, sensitivity, F1-score, and ROC-AUC.

3.3. Experimental Environment and Application Details for Lung Cancer CT Imaging

Analysis of lung computed tomography images using deep learning models was performed using data access provided via Google Drive and in the Google Colab

environment with A100 GPU infrastructure.

Table 2. Computing Environment and Software Structure

Environment Details	Specifications (According to the Code)
Platform	Google Colab
Storage	Google Drive (mounted dataset directory)
Backend Framework	PyTorch
Programming Language	Python 3.10
GPU	NVIDIA A100 (CUDA-enabled, Google Colab)
Device Selection	Automatic (CUDA if available, else CPU)
Operating System	Linux (Google Colab environment)
Deep Learning Libraries	Torch, TorchVision
Data Processing	TorchVision Transforms
Visualization Library	Matplotlib
Evaluation & Metrics	Scikit-learn
Statistical Analysis	SciPy (paired t-test)
Cross-Validation Method	10-Fold Cross-Validation

Table 2 shows that the analysis studies performed using CT images of lung cancer were carried out in the Google Colab environment using Python 3.10 and a PyTorch-based deep learning infrastructure. TorchVision was used for data preprocessing, Matplotlib for visualization, Scikit-learn for performance evaluation, and SciPy for statistical analysis. All experiments were conducted on a CUDA-enabled NVIDIA A100 GPU using a 10-fold K-Fold cross-validation approach.

3.4. Performance Metrics

Performance metrics are criteria used to measure the accuracy and effectiveness of the artificial intelligence model being used, and they vary depending on different machine learning tasks such as classification, regression, and clustering [33]. The success of artificial intelligence models used in lung cancer diagnosis and medical imaging analysis is typically evaluated using statistical metrics calculated from an error matrix [31]. This matrix consists of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values [21].

3.4.1. Accuracy

Accuracy is a key evaluation metric that represents the percentage of correct predictions among all predictions made by the model and summarizes its overall performance (Veeramani et al., 2025). In medical diagnoses, it is calculated by dividing the number of correctly classified cases by the total number of cases [31]. The accuracy formula is presented in Equation (1)[34].

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (1)$$

3.4.2. Precision

Precision identifies the percentage of true positives among all cases that an AI model classifies as positive, reflecting the model's ability to make accurate predictions [22]. This metric is used to evaluate the quality and reliability of the model, especially in situations where keeping the false positive rate low and preventing misdiagnosis is important [21, 22]. The equation of certainty is given in Equation (2) [35].

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (2)$$

3.4.3. Recall

Recall demonstrates how well a model can capture truly positive examples (Tian et al., 2024). Since it is vital not to miss a disease in medical diagnosis, the recall value is one of the most important parameters determining the sensitivity of the model, and the sensitivity equation is given in equation (3) [28, 36, 37].

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (3)$$

3.4.4. F1-Score

The F1-Score is the harmonic mean of the precision and recall metrics, and the F1-Score equation is given in equation (4) [38]. It provides a more reliable performance measurement than accuracy rate alone, especially when the classes in the dataset are unevenly distributed [13, 39].

$$F1\text{-}Score = 2 \times \frac{\mathrm{Precision \times Recall}}{\mathrm{Precision+Recall}} \quad (4)$$

3.4.5. Specificity

Specificity is the model's ability to accurately identify true negative cases (Sivakumar et al., 2024). It is used to exclude the possibility of a healthy individual being mistakenly classified as ill [26]. The specificity equation is given in equation (5) [40].

$$Specificity = \frac{\mathrm{TN}}{\mathrm{TN+FP}} \quad (5)$$

3.5. Cross Validation

Cross-validation improves the generalization capabilities of machine learning models by evaluating their performance. It is a resampling method that allows for the acquisition of statistically reliable results [41]. This approach works on the principle of dividing the dataset into K equal-sized subsets called folds [42]. Throughout the process, in each iteration, one of these subsets is reserved as the validation set, while the remaining K-1 subsets are used for training the model [43]. This cyclical structure increases data utilization efficiency by ensuring that each instance in the dataset is included in both the training and testing phases throughout the process [9]. Furthermore, cross-validation allows the average of results obtained from multiple evaluations to minimize biases that may arise from irregularities in data distribution, thus ensuring that the model exhibits more stable and reliable performance [13].

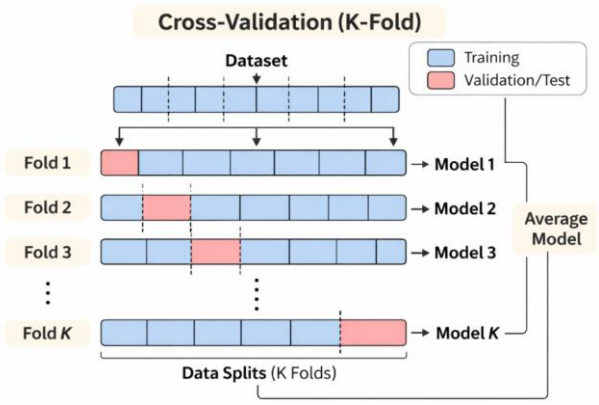

Figure 2. Schematic representation of the K-fold cross-validation process on lung CT images.

Figure 2 shows the k-fold cross-validation process. The dataset is divided into multiple folds. At each step, one fold is used for validation or testing, while the remaining folds are used for training. At each step, one fold is used for validation or testing, while the remaining folds are used for training. The images selected for validation or testing are changed in each iteration, ensuring the model is evaluated fairly and reliably across the entire dataset [44].

3.6. Explainable Artificial Intelligence (XAI)

Explainable AI is a method that makes the closed, often described as black box, nature of complex deep learning models transparent and understandable to humans [28]. In life-threatening medical diagnostic processes, such as lung cancer, a model that only produces predictions is not sufficient. Allowing physicians to see the steps involved in reaching this result increases both trust in the system and clinical accuracy [30]. XAI techniques aim to transform artificial intelligence from a decision-maker into a system that can present its reasoning and support it with visual evidence [28]. This approach helps artificial intelligence gain wider acceptance in healthcare by solving critical issues such as reliability, transparency, and algorithmic bias detection in diagnostic processes [28, 30].

3.6.1. LIME

Local Interpretable Model-Independent Explanations (LIME) is a model-independent method that explains the prediction of any AI model on a given sample at the local level. [31, 45]. This method divides a given image into smaller pieces that share similar visual characteristics and are called superpixels. Then, by randomly swapping these components, they analyze the model's responses [30, 31]. The LIME method creates a simple surrogate model that represents a complex pattern at the local level, highlighting the regions that contribute most to the prediction result through colored patches [30, 45]. In conclusion, LIME highlights key features of the model's decision-making process by directing the physician's focus to areas of pathological significance [30].

The mathematical optimization formula that forms the basis of the LIME architecture is given in equation (6) [30, 45]

$$\xi(x) = arg \min_{g \varepsilon G} L\ (f, g, \pi_x)\ + \Omega(g) \quad (6)$$

Definitions of the components included in the formula;

f : The actual black box that needs to be explained represents the classifier or model.

g∈ G : It creates a simple surrogate model, such as linear regression, selected from the class of interpretable models (G), which represents the original model at the local level.

L (f, g, π x) : The loss function measures the difference between the original model f and the surrogate model g, indicating how faithful the surrogate model is to the original model.

πx : It is a weighting mechanism that determines the neighborhood and proximity of a given sample.

Ω (g) : This term measures the complexity of a proxy model (e.g., number of rules or density of weights) so that the explanation remains understandable to humans [30, 45].

3.6.2. Gradient Weighted Class Activation Mapping (Grad-CAM)

Grad-CAM is an explainable artificial intelligence technique used to visualize complex decision-making processes in convolutional neural networks (CNNs) and reveal which regions directly influence model predictions [41]. In medical imaging, this method allows the model to move from being a black box to a transparent model by identifying pathological areas, such as nodules or lesions in the lung, that give rise to its prediction [12, 31].

The Grad-CAM method combines feature maps with gradient signals by retrospectively tracing the model's decision-making process. This approach contributes to evaluating the reliability of the diagnosis by visually demonstrating that the network focuses on medically critical regions. The Grad-CAM method preserves the spatial features in the final convolutional layers of the model, revealing which pixels contribute to the prediction score via a transparent heat map [27, 30]. The integration of these visualization techniques hierarchically explains the black-box nature of AI, thereby increasing healthcare professionals' clinical confidence in the model [12]. Heat maps generated by Grad-CAM help to verify the reliability of diagnosis by visually proving whether artificial intelligence is focusing on medically significant tumor or lesion areas [28, 31].

The Grad-CAM process mathematically takes place in two fundamental stages:

Calculation of Neuron Significance Weights (α_k^c): First, the gradient of y^c, which is the score of a class such as the target, is taken with respect to the activations of feature map number k (A^k) in the last convolutional layer of the model [28]. These gradients are subjected to a global average pooling process to obtain a single significance value for each feature map (Alqhatani et al., 2025).

The weight formula is given in equation (7).

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (7)$$

In equation (7), A_{ij}^k k represents the activation map of the k convolution channel, y^c represents the model output of the c class, Z represents the total number of pixels in the activation map, and α_k^c represents the class specific weight of the relevant channel [46].

Creating a Heat Map: Secondly, the calculated weights are multiplied by the corresponding feature maps and then summed. Based on the obtained result, the ReLU (Linear Smoothing Unit) activation function is applied (Veeramani et al., 2025). The ReLU function ensures that spatial regions that positively contribute to the prediction of the target class are preserved, preventing the inclusion of negative activations that reduce the class score in the visualization (Alqhatani et al., 2025).

The heat map formula is given in equation (8) [47].

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c\ A^k\right) \quad (8)$$

The heat maps generated through Equation (8) are overlaid on the original input image to visualize the critical regions that the model gives weight to in the decision mechanism [12]. In clinical practice, this visual evidence allows physicians to verify that artificial intelligence is focusing on medically relevant areas and to identify errors [9].

3.6.3. Grad-CAM++

Grad-CAM++ is an enhanced extension of the standard Grad-CAM method, specifically designed to more precisely determine the position of objects within an image and to identify situations where multiple disease markers are present [9, 30]. This enhanced version uses higher order combinations of gradients to more clearly and comprehensively visualize complex pathological features in medical images, such as ground glass opacities. Compared to the standard method, it offers physicians a more detailed and explanatory visual analysis by marking the regions contributing to the model's prediction with more accurate coverage [30]. Especially in cases where multiple lesions are present in a single image, it provides superior interpretability in clinical decision support processes by better differentiating the contribution of each object [9].

3.7. Deep Learning Algorithms

Deep learning algorithms with different architectural features have been used in the multi-class classification of CT images of lung cancer. The selected models were chosen for their ability to automatically learn distinctive features from images, effectively represent structural and textural information at different scales, and allow for comparative evaluation of their classification performance. Each architecture was structured according to its own design principles and trained on the same dataset, and its performance was analyzed using standard evaluation criteria. In this context, EfficientNet, which stands out with its combined scaling approach, ResNet50, which aims to reduce the gradient loss problem thanks to its deep network structures, and MobileNetV3 Large, which provides computational efficiency with its lightweight structure, have been examined.

3.7.1. EfficientNet

The EfficientNet architecture is built on a unique composite scaling strategy that scales the network's depth, width, and resolution simultaneously and in a balanced manner [31]. EfficientNet-B4, as a member of this model family ranging from B0 to B7, is a mid-to-high-end model that aims to deliver higher performance while optimizing the increase in computing cost [25, 31].

3.7.2. Residual Network 50 (ResNet50)

ResNet50 is an advanced convolutional neural network architecture with a depth of 50 layers, developed in 2015 by Kaiming He and colleagues at Microsoft Research Asia. This network structure consists of 48 convolution layers, one maximization pooling layer, and one averaging pooling layer [28]. The most distinctive feature of the ResNet50 architecture, hopping connections, allows input to be passed directly to subsequent layers by skipping some layers. This solves the gradient disappearance problem encountered in deep networks [28, 48]. ResNet50 has over 23 million trainable parameters and is commonly used for transfer learning with pre-trained weights on massive datasets like ImageNet [12, 28]. Thanks to its identity matching capability, the model reduces the risk of overfitting the dataset by bypassing layers it deems unnecessary during training [48]. In medical imaging studies, it is frequently used as a teaching model due to its high accuracy and rapid convergence capabilities, particularly in complex tasks such as the classification of lung cancer subtypes and pathological image analysis [12, 13].

3.7.3. MobileNetV3-Large

The MobileNetV3-Large architecture is built upon depth-based decomposable convolutions and linear bottleneck inverted residual structures, which formed the basis of its predecessors, the V1 and V2 models [42]. This architecture is structured to include more layers and parameters compared to the lighter, smaller version of the family, specifically to achieve high accuracy in complex visual recognition tasks. The NetAdapt algorithms used in the design process allow each layer of the network to be optimized to operate most efficiently on the hardware. This effectively reduces calculation costs and latency. Thanks to its lightweight and modular design, MobileNetV3-Large can offer high performance with low resource consumption in a wide range of applications, from medical diagnostics to industrial quality control

processes [41]. The architectural advantages offered by MobileNetV3-Large provide an effective solution, especially in environments with limited computing resources such as mobile devices, thanks to the balanced structure it establishes between accuracy and latency [13]. The depth-based decomposable convolution structure reduces mesh parameters, increasing speed in computationally intensive fields such as medical image analysis and enabling more efficient processing of complex data [13, 20]. In critical clinical tasks such as lung cancer diagnosis, the use of such optimized models accelerates diagnostic processes, provides reliable opinions for radiologists, and strengthens decision support mechanisms [13].

4. Result and Discussion

Computed tomography images of lung cancer were classified into benign, malignant, and normal categories, and EfficientNet-B4, ResNet50, and MobileNetV3-Large models were applied to these images. The experimental results obtained were compared using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics in the multi-class classification problem and are presented in Table 3.

Table 3 Average Performance Comparison of Models

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
MobileNetV3-Large	0.9705 ± 0.0179	0.9688 ± 0.0206	0.9744 ± 0.0138	0.9706 ± 0.0165	0.9970 ± 0.0027
EfficientNet-B4	0.9617 ± 0.0161	0.9592 ± 0.0154	0.9624 ± 0.0266	0.9599 ± 0.0209	0.9704 ± 0.0193
ResNet50	0.9555 ± 0.0174	0.9577 ± 0.0162	0.9555 ± 0.0174	0.9555 ± 0.0172	0.9924 ± 0.0045

The MobileNetV3-Large model achieved the highest overall classification success among all models examined, with an accuracy rate of 97.05 ± 1.79%. The fact that the model's Precision (96.88% ± 2.06%), Recall (97.44% ± 1.38%), and F1-Score (97.06% ± 1.65%) values are quite close to each other indicates a balanced distribution between false positive and false negative predictions and that the model has a consistent decision-making mechanism. Furthermore, the 99.70% ± 0.27 ROC-AUC value reveals that the MobileNetV3-Large model has a very high inter-class discrimination capability.

The EfficientNet-B4 model achieved a performance close to MobileNetV3-Large with an accuracy rate of 96.17% ± 1.61. The fact that the Precision (95.92% ± 1.54%), Recall (96.24% ± 2.66%), and F1-Score (95.99% ± 2.09%) metrics are close to each other indicates that the model exhibits balanced and stable classification behavior. However, the ROC-AUC value of 97.04% ± 1.93 indicates that the EfficientNet-B4 model successfully distinguishes between classes.

The ResNet50 model, with an accuracy rate of 95.55 ± 1.74%, offered a relatively lower overall accuracy compared to the other two models. The fact that the Precision (95.77% ± 1.62%), Recall (95.55% ± 1.74%), and F1-Score (95.55% ± 1.72%) values are very close to each other indicates that the model exhibits stable classification performance. However, the ROC-AUC value (99.24% ± 0.45%) indicates that the ResNet50 model demonstrates strong performance in terms of inter-class discrimination.

Overall, the MobileNetV3-Large model stands out in terms of overall performance, achieving both the highest accuracy and the highest ROC-AUC values. While the EfficientNet-B4 model offers balanced and high performance, the ResNet50 model provides a strong distinction, especially in the ROC-AUC metric. These results demonstrate that lightweight and efficient convolutional architectures (MobileNet and EfficientNet) offer effective solutions in terms of both accuracy and stability on the relevant dataset.

4.1. Analysis of Lung CT Images Using Grad-CAM++ Method

Explainable artificial intelligence approaches have been applied to make the predictions obtained from the deep learning-based classification model more reliable and interpretable from a clinical perspective. The Grad-CAM++ method, which generates class-separated spatial attention maps to reveal which regions of the image the model focuses on when making decisions, was used. In addition, LIME analysis was performed by dividing the image into superpixel regions, locally explaining the contribution of each region to the classification result. Thus, for all three classes, Grad-CAM++ shows both where the model concentrates and how each region influences the class decision. With LIME, the model's decision-making mechanism is supported by visual evidence through comparative evaluation.

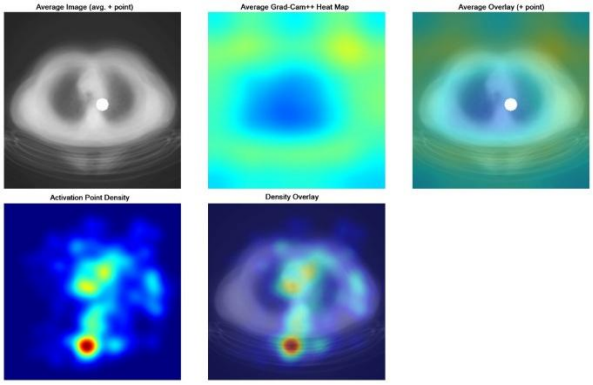

Figure 3. Average Grad-CAM++ activation and density analyses belonging to the normal class.

Figure 3 shows the average Grad-CAM++ activation maps generated from all lung CT images belonging to the Normal class (images without disease symptoms). The images obtained revealed low-intensity and diffuse

activation patterns throughout the lung parenchyma. The model does not appear to focus strongly on a specific anatomical region; the classification decision is based on the absence of pathological structural cues. This indicates that the model does not produce abnormal focuses or random attention areas when making decisions in normal cases.

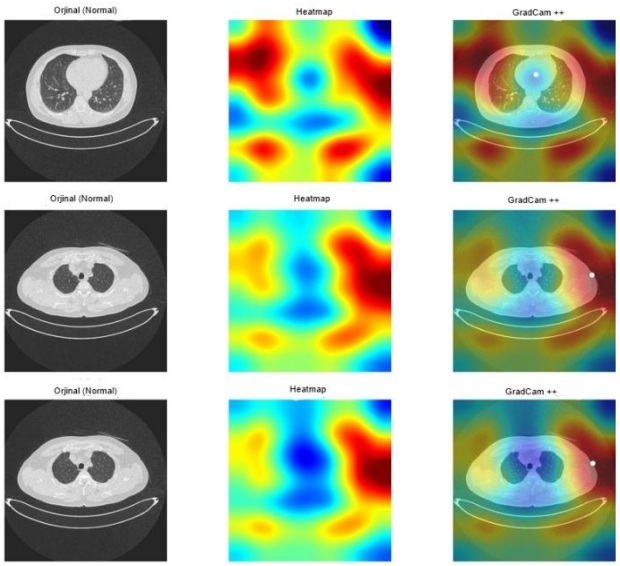

Figure 4. Example-based Grad-CAM++ visualizations belonging to the standard class.

In Figure 4, it is observed that the average activation point (Average activation point: 124.16, 117.36) is concentrated in the coordinates of the model input scale. The fact that this point corresponds to the central regions of the lung area indicates that the model has developed an anatomically consistent and non-random attention mechanism for the normal class. In the images, the white dot represents the average activation center calculated from the Grad-CAM++ heatmap. The model shows the centroid of the region that contributes most spatially to the classification decision.

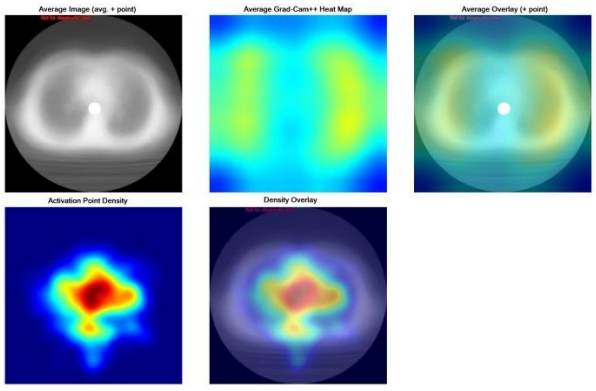

Figure 5. Average Grad-CAM++ activation and density analyses belonging to the malignant class.

Figure 5 presents the average images, Grad-CAM++ heat maps, and activation intensity analyses obtained from the analysis of malignant class lung CT images using a deep learning model. The average CT images and corresponding Grad-CAM heat maps presented in the upper panel demonstrate that the model focuses its decision-making attention on distinct, asymmetric, and localized pathological regions within the lung parenchyma. The activation point density and density overlay visualizations in the lower panel clearly show that the regions contributing most to the classification decision emerge as spatially consistent hotspots. This indicates that, in contrast to the homogeneous and symmetrical attention distribution observed in the normal class, the structural distortions and lesion-like morphological abnormalities specific to the malignant class are effectively distinguished by the model. The analysis reveals that by highlighting the critical regions focused on by the deep learning-based diagnostic model, both the model's transparency is increased and it is possible to interpret the locations where pathological findings are most frequently concentrated in a clinically meaningful way.

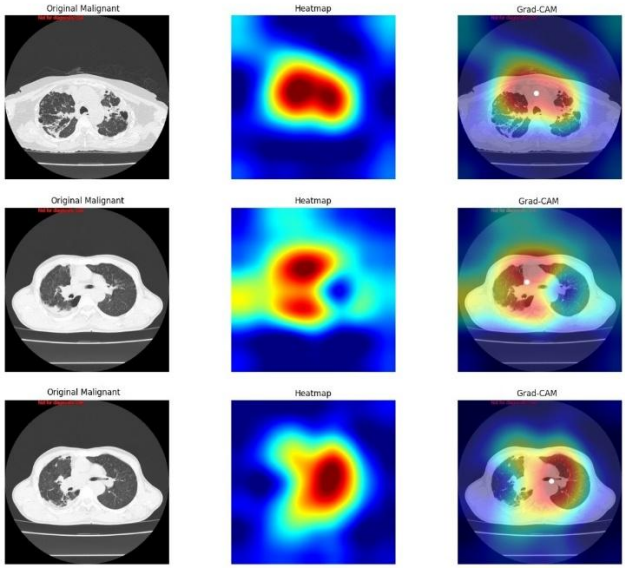

Figure 6. Example-based Grad-CAM++ visualizations of the Malignant class.

Figure 6 shows three different axial CT images of malignant lung cancer. The Grad-CAM results obtained from these images significantly reflect the model's decision-making mechanism. In the first example, the model's production of high red activation in the middle and lower regions of the lung parenchyma suggests that it is focusing on potentially malignant areas. The blue tones observed in extrapulmonary structures indicate that these regions have low importance in the decision-making process. In the second example, while significant activations were observed in pathological regions located in both lungs, the presence of low-level activations in peripheral and non-tumor-related areas indicates that the model exhibits careful and selective focusing. In the third example, the activations are predominantly concentrated within the lung parenchyma, with weaker activations on the opposite side. These findings demonstrate that the

model is able to accurately distinguish anatomical regions associated with malignancy and exhibits generally consistent classification behavior.

Grad-CAM++ heat maps show that the model primarily focuses on regions corresponding to abnormal tissue patterns and potential lung nodules within the lung parenchyma. In malignant cases, highlighted areas generally overlap with regions where nodular structures are visually prominent, i.e., regions with varying densities. These findings are consistent with pathological patterns reported in the healthcare field for lung anomaly detection. This observation demonstrates that the model's decision-making process relies on clinically significant image features (density differences) rather than irrelevant regions, thus supporting the reliability of the proposed explainable AI framework.

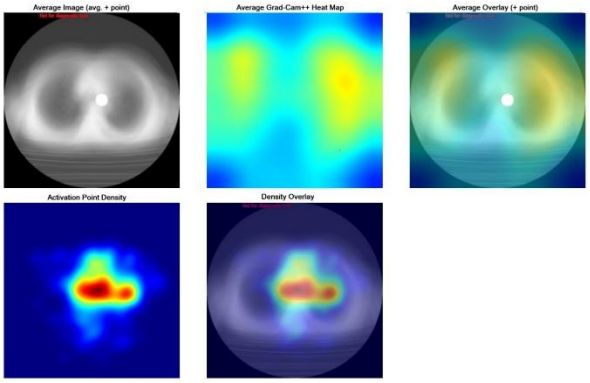

Figure 7. Average Grad-CAM activation and intensity analyses belonging to the benign class.

Figure 7 shows the average image obtained from benign class lung CT sections, the corresponding average Grad-CAM++ heat map, overlay, and activation point density. The average Grad-CAM++ map reveals that, unlike the sharp, singular hotspot structure seen in the malignant grade, the model's attention is focused on a broader region extending into the central parenchymal areas of the lungs. Particularly in the overlay panel, it is observed that the activation largely corresponds to intrapulmonary structures, while there is no significant concentration in extrapulmonary regions such as the chest wall and background. The fact that the hot regions in the activation point density map exhibit a smoother, more balanced profile suggests that the model's decision in the benign class is supported by the overall textural arrangement of the parenchyma and widespread structural cues, rather than by a single mass-like focus. The calculated average activation point (mean activation point: 123.58, 110.73) quantitatively confirms that this focus corresponds to the middle sections of the lungs. Thus, it appears that the benign class is represented by a more stable and diffused attention distribution compared to the malignant class.

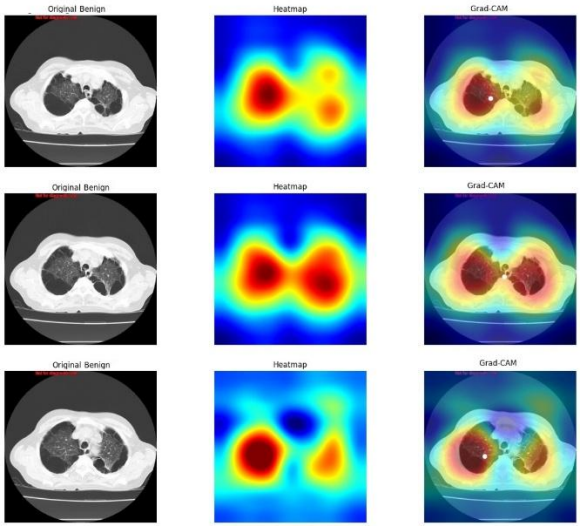

Figure 8. Example-based Grad-CAM visualizations belonging to the benign class.

Grad-CAM++ based attention maps of three benign (non-cancerous) lung CT samples presented in Figure 8 show that the model's classification decision is predominantly based on structures within the lung parenchyma. In the first example, although the dominant activation is more pronounced on one side, a secondary contribution is also observed in the opposite lung, suggesting that the decision is formed by a combination of features distributed throughout the parenchyma rather than a single, sharp focus. In the second example, the appearance of activation areas of similar size and intensity in both lungs suggests the adoption of a more balanced spatial attention pattern characteristic of the benign class. In the third example, although activation is more dominant in one lobe, it does not completely subside in the opposite lobe but persists at a moderate level, supporting the idea that the model makes its decision by evaluating bilateral tissue patterns together, rather than locking onto a local aggressive focus. Overall, the concentration of the highest contribution in the parenchyma rather than in extrapulmonary structures indicates that the obtained Grad-CAM++ samples are anatomically consistent and strengthens the interpretability of the benign class.

4.2. Analysis of Lung CT Images Using the LIME Method

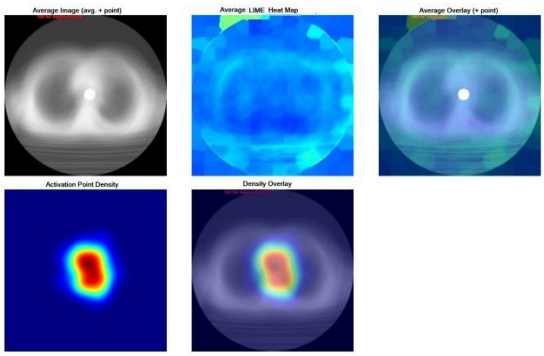

Figure 9. Average LIME activation and intensity analyses belonging to the Bening class.

Figure 9 shows the results of the average activation analysis obtained using the LIME method for the benign lung class. The average CT image in the upper left represents the spatial common structure of benign samples in the dataset, while the marked center point reveals the location most frequently referenced by the model in the decision-making process. The average LIME shows that the additive regions in the heat map exhibit a superpixel-based distribution with smooth transitions, without forming sharp boundaries. This suggests that the model, when defining the benign class, relies on structural patterns formed by neighboring tissues rather than a single dominant lesion. In the overlay visualization, it is noteworthy that these contribution areas remain largely within the lung parenchyma and extend only minimally into surrounding anatomical structures. The activation point density map reveals that the high contribution is concentrated in a narrow and compact region, but exhibits a profile different from the aggressive and widely distributed foci observed in malignant classes. Finally, the density overlay panel shows that the model uses a central and repetitive spatial attention mechanism when making decisions for the benign class, but that this attention is widely and evenly distributed. These results indicate that LIME-based average analyses offer a more stable model and structurally-driven explainability in benign samples.

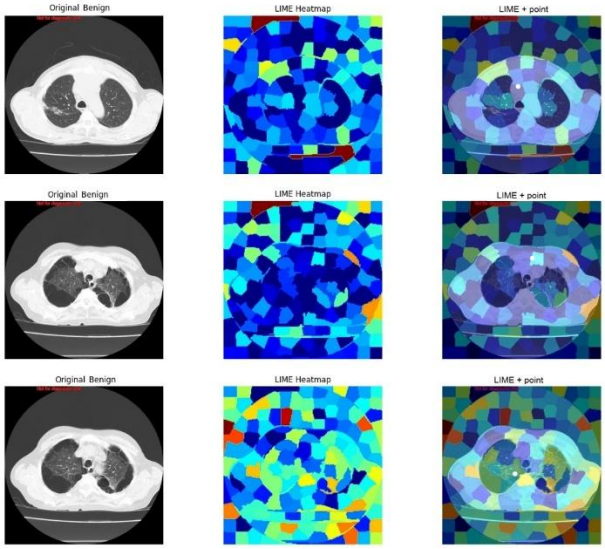

Figure 10. Bening class-specific example-based LIME visualizations

Figure 10 shows an example of a benign lung computed tomography image, illustrating how the LIME-based explainability approach shapes model decisions. In the first example, the general structure of the lung parenchyma is preserved in the original CT scan, and no significant aggressive lesion pattern is observed. In contrast, the LIME heatmap shows the contributions spread across numerous small superpixels, with low to medium dispersion. The LIME visualization shows that the model's decision center is located in the central region of the lungs, but the decision is not based on a single dominant region. In the second example, while increased heterogeneity in the parenchyma is noticeable in the original image, it is striking that the contributions in the LIME heat map are mostly limited to intrapulmonary tissues, and peripheral areas are represented with low significance. This suggests that the model bases its definition of the benign class on structural tissue arrangement. The LIME output shows that the insertion center is located within the parenchyma, supporting anatomical consistency. In the third example, although there are more pronounced density differences in the original section, the LIME heat map shows that the contributions are spread over a wide area and do not form a sharp focus. In the images in the third column, the location of the maximum contribution, marked with a dot, reveals that the model based its decision on the combined effect of environmental superpixel groups rather than a specific local structure. Overall, these three examples demonstrate that the LIME method explains the model's decisions for the benign class in a piecemeal, distributed, and structurally-based manner, deviating significantly from the dense and localized contribution patterns expected in the malignant classes.

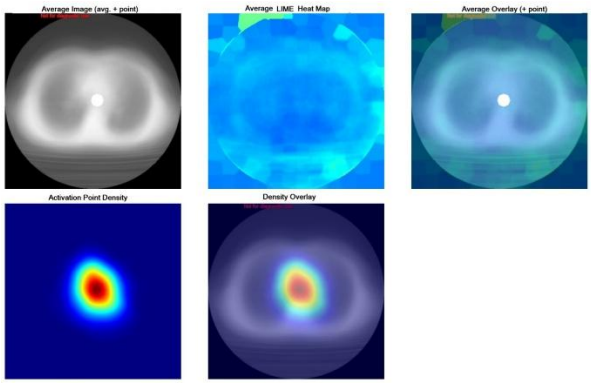

Figure 11. Average LIME activation and intensity analyses belonging to the Maling class.

Figure 11 shows the average activation analysis obtained using the LIME method for the maling class. The average CT image in the upper left panel represents the common spatial appearance of malignant specimens, while the marked point indicates the most frequently observed decision center within each class. In the average LIME heat map (upper center), the contribution regions are distributed with a superpixel-based and fragmented structure, but the most prominent contribution is concentrated in regions near the center of the image. This suggests that the model's decision is based on the combined effect of multiple neighboring structural components, rather than a single small region. The overlay panel in the upper right visually demonstrates how the model's attention is drawn to the central structures within the image by presenting these contribution areas superimposed on the anatomical structure. The activation

point density map in the lower left shows that activations are clustered distinctly in a single core region, and the calculated average activation point (average activation points: 116.52, 111.05) quantitatively confirms the center of this clustering. Finally, the density overlay panel (lower middle) clarifies where the high contribution is repeated on the image, demonstrating that the model has developed a more stable spatial focus in the lung malignant nodule class.

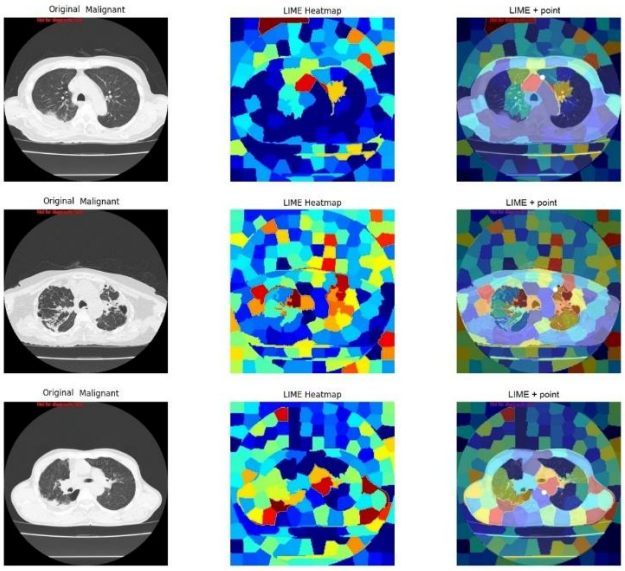

Figure 12. Example-based LIME visualizations of the malignant class.

Figure 12 illustrates how the model justifies its decision in the malignant classification of lung cancer using LIME, through three different examples. The first original lung CT image is shown, followed by a LIME superpixel heat map, and finally, an overlay of the additives, showing the point of highest additiveness. In the first example, most of the hot superpixels are clustered in a narrow area near the central/hilar region of the lungs, and the spot is located very close to this cluster; this suggests that the model supports the malignant diagnosis more through a local focus. In the second example, the contributions are distinctly spread across both hemithoraxes. The widespread distribution of hot patches throughout the parenchyma on the LIME map and the concentration of these patches in intrapulmonary structures in the overlay image indicate that the model's decision is based on widespread tissue patterns/heterogeneity rather than a single lesion. In the third example, hot superpixels are seen to be more dominant on one side and sparser on the other. The fact that the point falls within the dominant contribution region indicates that the model gives more weight to parenchymal changes showing asymmetry for the malignant class in this section. Overall, in all three examples, the strongest contributing areas remained within the lung parenchyma, and the marked point overlapped with these areas, suggesting that the LIME description indicates that the model's malignant classification is based not on randomness, but on the combined effect of specific parenchymal regions within the image.

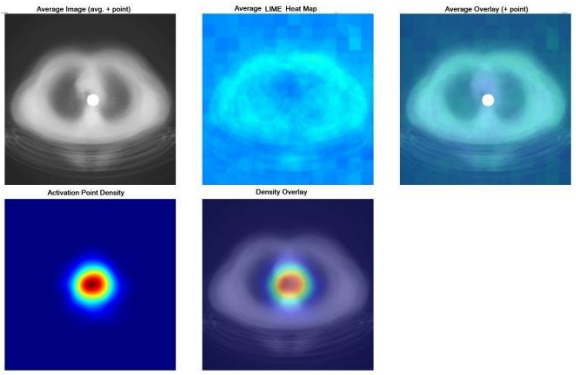

Figure 13. Average LIME activation and intensity analyses belonging to the normal class.

Figure 13 shows the LIME description generated from the average of the samples belonging to the normal class. Since no obvious findings indicating pathology are expected in the original average image, the foci seen on the LIME map do not represent the location of the disease. The model highlights the anatomical cues it references here to define the normal class. The clustering of the heat map and density overlay panel, particularly in a region close to the center of the lungs, suggests that the model frequently checks structural features such as airway/mediastinum periphery, hilus line, and overall symmetry contrast scheme when making decisions. The clustering of activation point densities in a single core region indicates that the model repeatedly examines a similar control region across different normal samples, supporting the normal class decision through a consistent spatial pattern. Consequently, these focal points should be interpreted not as pathological lesions, but as stable anatomical reference regions that the model uses to confirm its normality.

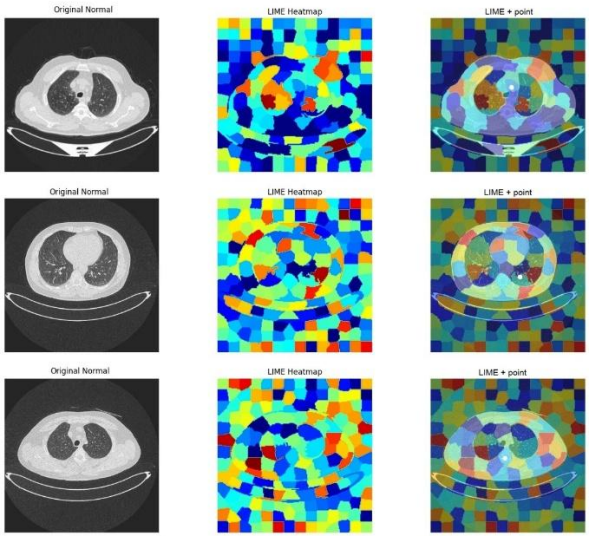

Figure 14. Example-based LIME visualizations belonging to the normal class.

Figure 14 illustrates the rationale behind the model's decision using the LIME method in three different axial thoracic CT sections belonging to the normal class. Each row contains the original image, a superpixel significance map, and an overlay of this map onto the image, along with a notation marking the point of highest contribution. Since no significant lesions are expected in the normal class, the highlighted areas reflect anatomical cues that the model confirms when making the normal decision, rather than a pathological focus. In the first example, the concentration of contributions around the upper central parenchyma and airway/bronchovascular structure suggests that the model references symmetry and typical tissue arrangement. In the second example, the peripheral contributions along the pleural line and parenchymal boundaries demonstrate that contour continuity and density transitions are distinctive for the normal class. Although the distribution is more diffuse in the third example, the fact that the highest contribution point remains in the intrapulmonary region supports the idea that the decision is based on anatomically significant areas. Overall, the model focuses not on a single focal point within the normal class, but on holistic normal anatomical features such as symmetry, borderline regularity, and typical parenchymal tissue.

5. Conclusion

This study aimed to automatically differentiate normal, benign, and malignant classes from lung CT scans. To improve the explanatory power of predictions obtained using a deep learning-based classification approach on a dataset containing a total of 1933 images, Grad-CAM++ and LIME methods were used together. The goal is not only to achieve highly accurate classification, but also to strengthen confidence in AI-based diagnostic support systems by making the model's decisions clinically interpretable.

The dataset used in this study was obtained from an existing medical image source. Therefore, the class distribution was left as is, without any changes, preserving the original structure of the dataset. The dataset contains 416 normal, 467 benign, and 1050 malignant CT images, resulting in a certain class imbalance. While such imbalances can affect model learning, different performance metrics such as precision, recall, and F1-score were used in addition to the accuracy metric to more reliably evaluate model performance.

When Grad-CAM++ outputs were evaluated on a class basis, it was observed that activations related to the normal class exhibited a low-intensity and diffuse distribution throughout the lung parenchyma, whereas the model did not tend towards a distinct pathological region. This suggests that, in normal cases, the decision is based not on the presence of a lesion, but on overall anatomical consistency supporting the absence of pathological findings. In contrast, the concentration of activations in the malignant tumor class as more localized, asymmetrical, and distinct hotspots indicated that the model focused more selectively on parenchymal changes that might be associated with malignancy. In the benign tumor class, a more widespread and balanced distribution of attention was observed compared to the malignant class, suggesting that the model considered broader structural cues together, such as the overall tissue arrangement of the parenchyma, rather than a single aggressive focus. This separation reveals that the three classes are represented by different spatial characteristics at different levels of explainability.

Image analysis using the LIME method provided local annotations that complemented the Grad-CAM++ results. In the normal (disease-free) category, it was observed that the regions highlighted by LIME did not represent a pathological focus; the model checked normal anatomical cues such as symmetry, contour continuity, pleural border regularity, and expected density transitions when making its normal decision. In benign lung imaging samples, it was observed that the contributions remained fragmented but distributed within the parenchyma at the superpixel level, and the resolution was not reduced to a single region. In malignant lung imaging samples, it was observed that the contributions were more focused in some sections and more widespread in others, but the fact that the strongest contribution areas mostly remained within the lung parenchyma supported the anatomical consistency of the explanations. Thus, while LIME shows which regions the model's decision is locally fed from, Grad-CAM++ simultaneously reveals the class-separating distractibility of the decision-making process, demonstrating that using the two methods together significantly improves interpretability.

In conclusion, the study findings demonstrate that deep learning-based lung CT classification becomes clinically more reliable and traceable when supported not only by performance outputs but also by explainability maps. The obtained Grad-CAM++ and LIME visualizations revealed that the model primarily referenced anatomically significant intrapulmonary regions when making decisions, and the observation of different attention/contribution patterns in normal, benign, and malignant classes increased the transparency of the model's decision-making mechanism.

However, to improve the clinical generalizability of these findings, it is recommended that future external validation be performed with larger datasets from different centers, integration of 3D volumetric assessments (3D CT), analysis of class imbalance and sensitivity to imaging protocols, and quantitative comparison of XAI outputs with radiologist assessments. Accordingly, it was concluded that explainable AI-assisted classification approaches have significant potential in the early detection of lung nodules and in supporting clinical decision-making

processes.

Disclosure Statement

During the writing process of this study, generative artificial intelligence-powered tools were used to improve the text linguistically and enhance its compliance with academic writing standards. This usage does not include content creation or scientific decision-making processes.

Declaration of Ethical Standards

The authors declare that this study was conducted in accordance with accepted ethical standards in research and publication. All authors have contributed honestly to the work, appropriate citations have been provided, and the manuscript represents original research that has not been published or submitted elsewhere.

Credit Authorship Contribution Statement

All authors contributed to the conception and design of the study. Data analysis and model development were performed collaboratively. All authors participated in the interpretation of results, drafting, and critical revision of the manuscript, and approved the final version for submission.

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.

Funding / Acknowledgements

No external funding was secured for this research.

Data Availability

In this study, a dataset obtained from the Mendeley platform was used. [32]. The relevant dataset can be accessed via the following link: https://data.mendeley.com/datasets/zr4fddh833/1

References

[1] Q. Xiao, M. Tan, G. Yan, and L. Peng, "Revolutionizing lung cancer treatment: harnessing exosomes as early diagnostic biomarkers, therapeutics and nano-delivery platforms," *J Nanobiotechnology*, vol. 23, no. 1, p. 232, Mar 21 2025, doi: 10.1186/s12951-025-03306-0.

[2] A. Tabur, "Akciğer Kanserlerinin Epidemiyolojik ve Klinik Özellikleri: Hastane Başvurularının Branş Dağılımlarına Göre İncelenmesi," *Journal of İzmir Chest Hospital*, pp. 26-35, 2025, doi: 10.14744/igh.2025.79188.

[3] K. L. Hua, C. H. Hsu, S. C. Hidayati, W. H. Cheng, and Y. J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *Onco Targets Ther*, vol. 8, pp. 2015-22, 2015, doi: 10.2147/OTT.S80733.

[4] K. Yasufuku, "Early diagnosis of lung cancer," *Clin Chest Med*, vol. 31, no. 1, pp. 39-47, Table of Contents, Mar 2010, doi: 10.1016/j.ccm.2009.08.004.

[5] J. Ning et al., "Early diagnosis of lung cancer: which is the optimal choice?," *Aging (Albany NY)*, vol. 13, no. 4, p. 6214, 2021. [Online]. Available: https://www.aging-us.com/article/202504/pdf.

[6] N. Gómez León, S. Escalona, B. Bandrés, C. Belda, D. Callejo, and J. A. Blasco, "18F-Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography Accuracy in the Staging of Non-Small Cell Lung Cancer: Review and Cost-Effectiveness," *Radiology Research and Practice*, vol. 2014, no. 1, p. 135934, 2014, doi: 10.1155/2014/135934.

[7] Y. Doğanşah and M. Köklü, "Use Of The Electromagnetic Spectrum In The Medical Field," *Gece Publishing*, 2022, ch. 9.

[8] T. Kobayashi et al., "HRCT findings of small cell lung cancer measuring 30 mm or less located in the peripheral lung," *Jpn J Radiol*, vol. 33, no. 2, pp. 67-75, Feb 2015, doi: 10.1007/s11604-014-0381-2.

[9] J. Chen et al., "Lung cancer diagnosis using deep attention-based multiple instance learning and radiomics," *Med Phys*, vol. 49, no. 5, pp. 3134-3143, May 2022, doi: 10.1002/mp.15539.

[10] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Comput Med Imaging Graph*, vol. 31, no. 4-5, pp. 198-211, Jun-Jul 2007, doi: 10.1016/j.compmedimag.2007.02.002.

[11] B. Van Ginneken, C. M. Schaefer-Prokop, and M. Prokop, "Computer-aided diagnosis: how to move from the laboratory to the clinic," *Radiology*, vol. 261, no. 3, pp. 719-732, 2011.

[12] X. Zhang et al., "Deep learning enhances precision iagnosis and treatment of non-small cell lung cancer: future prospects," *Transl Lung Cancer Res*, vol. 14, no. 8, pp. 3196-3215, Aug 31 2025, doi: 10.21037/tlcr-2025-187.

[13] L. Tian et al., "Precise and automated lung cancer cell classification using deep neural network with multiscale features and model distillation," *Sci Rep*, vol. 14, no. 1, p. 10471, May 7 2024, doi: 10.1038/s41598-024-61101-7.

[14] Y. Zhang, Y. Weng, and J. Lund, "Applications of Explainable Artificial Intelligence in Diagnosis and Surgery," *Diagnostics (Basel)*, vol. 12, no. 2, p. 237, Jan 19 2022, doi: 10.3390/diagnostics12020237.

[15] D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen, "Explainable artificial intelligence: a comprehensive review," *Artificial Intelligence Review*, vol. 55, no. 5, pp. 3503-3568, 2022, doi: 10.1007/s10462-021-10088-y.

[16] Z. Sadeghi et al., "A brief review of explainable artificial intelligence in healthcare," *arXiv preprint arXiv:2304.01543*, 2023, doi: 10.48550/arXiv.2304.01543.

[17] N. Adil, P. Singh, and N. K. Nagwani, "Interpretable Lightweight CNN for Colon and Lung Cancer Classification with LIME Based Explainability," in *2024 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT)*, 2024: IEEE, pp. 122-127, doi: 10.1109/ICISSGT58904.2024.00034.

[18] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097-2106, doi: 10.1109/CVPR.2017.369.

[19] X. Li, Y. Kao, W. Shen, X. Li, and G. Xie, Lung nodule malignancy prediction using multi-task convolutional neural network (SPIE Medical Imaging). SPIE, 2017, doi: 10.1117/12.2253836.

[20] G. Mohandass, G. Hari Krishnan, D. Selvaraj, and C. Sridhathan, "Lung Cancer Classification using Optimized Attention-based Convolutional Neural Network with DenseNet-201 Transfer Learning Model on CT image," *Biomedical Signal Processing and Control*, vol. 95, p. 106330, Sep 2024, doi: 10.1016/j.bspc.2024.106330.

[21] M. Sivakumar, S. Chinnasamy, and T. Ms, "An efficient combined intelligent system for segmentation and classification of lung cancer computed tomography images," *PeerJ Comput Sci*, vol. 10, p. e1802, 2024, doi: 10.7717/peerj-cs.1802.

[22] M. Budati and R. Karumuri, "An intelligent lung nodule segmentation framework for early detection of lung cancer using an optimized deep neural system," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 34153-34174, 2024, doi: 10.1007/s11042-023-17791-8.

[23] Y. Chen, Y. Wang, F. Hu, and D. Wang, "A Lung Dense Deep Convolution Neural Network for Robust Lung Parenchyma Segmentation," *IEEE Access*, vol. 8, pp. 93527-93547, 2020,

doi: 10.1109/access.2020.2993953.

[24] M. Masud et al., "Light Deep Model for Pulmonary Nodule Detection from CT Scan Images for Mobile Devices," *Wireless Communications and Mobile Computing*, vol. 2020, no. 1, pp. 1-8, Jul 3 2020, doi: 10.1155/2020/8893494.

[25] G. Cai et al., "Medical artificial intelligence for early detection of lung cancer: A survey," *Engineering Applications of Artificial Intelligence*, vol. 159, p. 111577, Nov 8 2025, doi: 10.1016/j.engappai.2025.111577.

[26] P. S. H. Jose, J. E. Sagar, S. Abisheik, R. Nelson, M. Venkatesh, and B. Keerthana, "Leveraging Convolutional Neural Networks and Multimodal Imaging Data for Accurate and Early Lung Cancer Screening," in *2024 International Conference on Inventive Computation Technologies (ICICT)*, 2024: IEEE, pp. 1258-1264, doi: 10.1109/ICICT60155.2024.10544863.

[27] S. Singh, "Computer-aided diagnosis of thoracic diseases in chest X-rays using hybrid cnn-transformer architecture," *arXiv preprint arXiv:2404.11843*, 2024, doi: 10.48550/arXiv.2404.11843.

[28] N. Veeramani, A. R. S, S. P. S, S. S, and P. Jayaraman, "NextGen lung disease diagnosis with explainable artificial intelligence," *Sci Rep*, vol. 15, no. 1, p. 33052, Sep 26 2025, doi: 10.1038/s41598-025-07603-4.

[29] P. N. Megat Ramli, A. N. Aizuddin, N. Ahmad, Z. Abdul Hamid, and K. I. Ismail, "A Systematic Review: The Role of Artificial Intelligence in Lung Cancer Screening in Detecting Lung Nodules on Chest X-Rays," *Diagnostics (Basel)*, vol. 15, no. 3, Jan 22 2025, doi: 10.3390/diagnostics15030246.

[30] Z. Naz et al., "An Explainable AI-Enabled Framework for Interpreting Pulmonary Diseases from Chest Radiographs," *Cancers (Basel)*, vol. 15, no. 1, Jan 3 2023, doi: 10.3390/cancers15010314.

[31] A. Alqhatni, T. Babu, T. R. Mahesh, S. B. Khan, O. Saidani, and M. T. Quasim, "Automated classification and explainable AI analysis of lung cancer stages using EfficientNet and gradient-weighted class activation mapping," *Front Med (Lausanne)*, vol. 12, p. 1625183, 2025, doi: 10.3389/fmed.2025.1625183.

[32] T. H. Talha et al., "An Image Dataset of Advanced Imaging Techniques for Lung Cancer Diagnosis," *Mendeley Data*, 2025, doi: 10.17632/xxxxx.1.

[33] Y. Unal, E. T. Yasin, T. A. Cengel, and M. Koklu, "Classification of Turkish hazelnut (Corylus colurna L.) varieties: a comparative study of YOLOv8 and fine-tuned vision transformer," *Journal of Food Measurement and Characterization*, pp. 1-15, 2026, doi: 10.1007/s11694-025-03936-w.

[34] E. T. Yasin et al., "Detecting Driver Fatigue Using Artificial Intelligence on a Realistic Driving Images," *Journal of Future Artificial Intelligence and Technologies*, vol. 2, no. 4, pp. 648-660, 2026, doi: 10.62411/faith.3048-3719-299.

[35] I. A. Ozkan and M. Koklu, "Skin lesion classification using machine learning algorithms," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 5, no. 4, pp. 285-289, 2017, doi: 10.18201/ijisae.2017534420.

[36] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, "Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging," *Cancers (Basel)*, vol. 15, no. 16, p. 4172, Aug 18 2023, doi: 10.3390/cancers15164172.

[37] M. M. Saritas, O. Kilci, and M. Koklu, "Evaluation of CNN Models for Multi-Class Gear Fault Detection Using Waveform Images," in *Proceedings of International Conference on Intelligent Systems and New Applications*, 2025, vol. 3, pp. 31-40, doi: 10.58190/icisna.2025.137.

[38] I. Cinar and M. Koklu, "Identification of rice varieties using machine learning algorithms," *Journal of Agricultural Sciences*, pp. 9-9, 2022, doi: 10.15832/ankutbd.862482.

[39] F. Haque et al., "An End-to-End Concatenated CNN Attention Model for the Classification of Lung Cancer With XAI Techniques," *IEEE Access*, vol. 13, pp. 96317-96336, 2025, doi: 10.1109/access.2025.3572423.

[40] M. Koklu, R. Kursun, Y. S. Taspinar, and I. Cinar, "Classification of date fruits into genetic varieties using image analysis," *Mathematical Problems in Engineering*, vol. 2021, no. 1, p. 4793293, 2021, doi: 10.1155/2021/4793293.

[41] H. H. Aras, Y. Eryeşil, and M. Köklü, "An Explainable Deep Learning Framework for Agtron-Based Coffee Roast Classification Using Grad-CAM," in *Proceedings of International Conference on Intelligent Systems and New Applications*, 2025, vol. 3, pp. 51-57, doi: 10.58190/icisna.2025.139.

[42] B. Isgor and M. Koklu, "Lightweight Hybrid Model for Bone Fracture Detection Using MobileNetV2 Feature Extraction and Ensemble Learning," *Journal of Future Artificial Intelligence and Technologies*, vol. 2, no. 3, pp. 521-533, 2025, doi: 10.62411/faith.3048-3719-284.

[43] H. Incekara, I. H. Cizmeci, M. M. Saritas, and M. Koklu, "Classification of almond kernels with optuna hyper-parameter optimization using machine learning methods," *Journal of Food Science and Technology*, pp. 1-17, Nov 25 2025, doi: 10.1007/s13197-025-06494-7.

[44] İ. H. Çizmeci and H. İncekara, "Stacking Ensemble Based Hybrid Machine Learning Approach for Predicting Obesity Levels," in *2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA)*, 2025: IEEE, pp. 1-8, doi: 10.1109/ICHORA65333.2025.11017164.

[45] D. Garreau and U. Luxburg, "Explaining the explainer: A first theoretical analysis of LIME," in *International Conference on Artificial Intelligence and Statistics*, 2020: PMLR, pp. 1287-1296.

[46] M. D. İli and F. Özyurt, "Açiklanabilir Yapay Zeka Yöntemleriyle Mr Görüntülerinden Beyin Tümörü Tespiti," *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 28, no. 2, pp. 1092-1109, 2025.

[47] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.

[48] D. Theckedath and R. Sedamkar, "Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks," *SN Computer Science*, vol. 1, no. 2, p. 79, 2020, doi: 10.1007/s42979-020-0114-9.