Available online www.worldscientificnews.com

World Scientific News

An International Scientific Journal

WSN 213 (2026) 149-169 EISSN 2392-2192

Attention-Guided Diagnostic Intelligence: Towards Trustworthy and Explainable Lung Disease Classification

Simon Onuwa Agbonifo[1], Gabriel Chukuemeke Agbonifo[2], Isaac Nosakhare Agbonifo[3], Abraham Osemeke Agbonifo[4], Prisca Chimezie Opara[5], Daniel Agbonifo[6], Sunday Chukwuebuka Uduogu[7], Happy Nkanta Monday[8*]

[1]Department of Chemistry, University of Benin, Nigeria

[2]Department of Mechanical Engineering Technology, Auchi polytechnic, Auchi, Nigeria

[3]School of Business Studies, Auchi Polytechnic, Auchi, Nigeria

[4]Department of Guidance and Counselling, Delta State University, Abraka, Nigeria

[5]Faculty of Law, University of Nigeria, Nsukka, Nigeria

[6]College of Geophysics, Chengdu University of Technology, China

[7]Department of Medical Biochemistry, Cross River State University, Nigeria

[8]School of International Education, Chengdu University of Technology, China

*Author for Correspondence: happy.monday@zy.cdut.edu.cn

https://doi.org/10.65770/MSPI4814

ABSTRACT

The rapid advancement of medical imaging technology has revolutionized diagnostics; however, manual interpretation remains subjective and time-intensive. This study presents an attention-based deep learning framework, the Mask-Guided Convolutional Neural Network (MG-CNN), designed to enhance both the classification accuracy and explainability of chest X-ray analysis for COVID-19 and pneumonia.

(Received 12 January 2026; Accepted 21 February 2026; Date of Publication 15 March 2026)

Leveraging the COVID-19 Radiology Dataset, the model integrates a segmentation-based attention mechanism that dynamically prioritizes relevant pulmonary regions while suppressing background noise. Experimental results demonstrate a test accuracy of 89.97%. Crucially, the integration of Gradient-weighted Class Activation Mapping (Grad-CAM) validates that the model's decision-making aligns with clinical pathology, focusing on lung parenchyma rather than irrelevant artifacts. This work addresses the "black box" limitation of traditional deep learning, offering a transparent, trustworthy Clinical Decision Support System (CDSS) for pulmonary medicine.

Keywords: COVID-19 Detection, Explainable AI (XAI), Mask-Guided Attention, Deep Learning, Chest X-ray Analysis, Grad-CAM, Medical Image Classification.

1. INTRODUCTION

The field of medical imaging has long been at the forefront of technological advancement in healthcare, with the primary objective of enhancing diagnostic accuracy and patient outcomes. Modalities such as X-rays, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) provide critical insights into internal anatomical structures, enabling physicians to detect, diagnose, and monitor a spectrum of pathologies. Thoracic imaging is particularly vital given the global prevalence of respiratory conditions, including lung cancer, which remains the leading cause of cancer-related mortality worldwide (Siegel et al., 2023). Traditionally, medical image analysis relies on manual interpretation by radiologists. This process, while standard, is time-consuming and susceptible to inter-observer and intra-observer variability. Furthermore, the exponential growth in medical imaging data volume has outpaced the capacity of human experts for timely analysis, necessitating the development of automated diagnostic aids (Esteva et al., 2017).

Deep Learning (DL), a subset of machine learning, has emerged as a powerful paradigm for medical image analysis. Its capacity to learn complex, non-linear patterns directly from data has significantly improved performance in classification tasks (Litjens et al., 2017). Convolutional Neural Networks (CNNs), in particular, are widely adopted for image classification due to their proficiency in capturing spatial hierarchies (LeCun et al., 2015). However, a critical limitation of standard CNNs is their lack of explainability—often termed the "black box" problem. The opacity of decision-making in deep learning models presents a significant barrier to clinical adoption, where understanding the rationale behind a prediction is essential for trust and regulatory approval (Rajkomar et al., 2018). To mitigate this, attention mechanisms have been introduced, allowing models to dynamically focus on image regions most relevant to the classification task. This approach not only enhances predictive performance but also provides interpretability by highlighting features driving the decision (Miao, 2019). While integrating attention mechanisms into medical imaging is a nascent field, it holds immense promise for bridging the gap between AI accuracy and clinical trust (Chilamkurthy et al., 2018).

Despite the success of deep learning in detecting lung diseases, standard models often learn from irrelevant background noise rather than pathological features. Consequently, there is an urgent need for diagnostic tools that are both accurate and transparent. The primary motivation for this work is to develop a model that accurately classifies lung pathologies (specifically distinguishing COVID-19 from normal cases) while simultaneously elucidating its decision-making process. By integrating an external mask attention mechanism, this study aims to guide the model's focus strictly to the anatomical Regions of Interest (ROI), thereby reducing background interference and improving diagnostic precision.

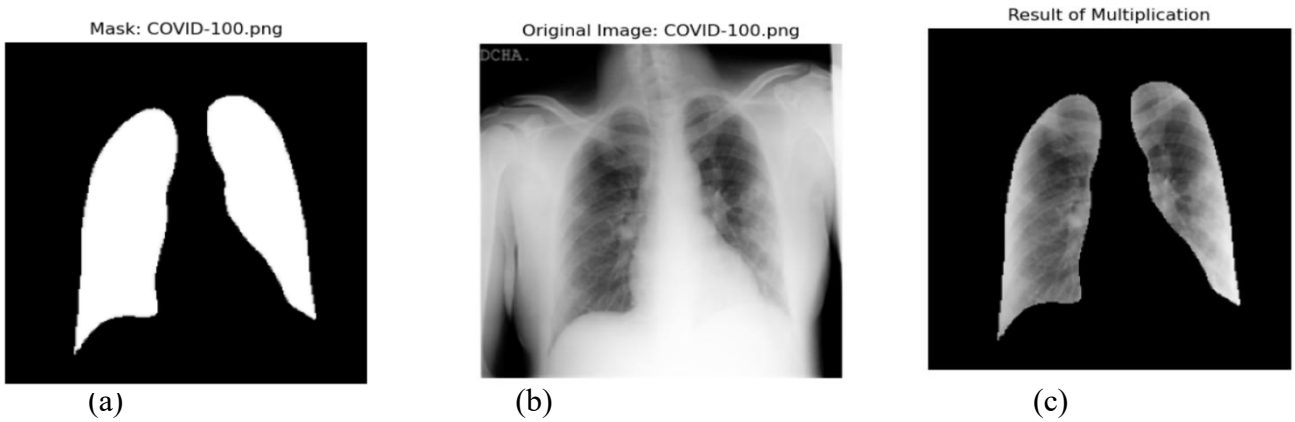

Figure 1. Showing of sample COVID-19 Images. (a)Unprocessed Image, (b)Lung Segmentation Mask, and (c)final Processed Image.

To achieve this, the study develops a Mask-Guided Convolutional Neural Network (MG-CNN) for the binary classification of chest X-rays. The research methodology involves implementing a robust preprocessing pipeline that standardizes Kaggle lung datasets, including grayscale conversion and the application of lung segmentation masks. A custom CNN architecture is designed to prioritize lung regions during feature extraction, trained using a custom data generator and optimized hyperparameters such as the Adam optimizer and binary cross-entropy loss. To ensure rigor, the model is assessed using comprehensive metrics including Precision, Recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Furthermore, the study validates the model's explainability by utilizing Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize decision heatmaps, verifying that the model's focus aligns with clinically relevant lung areas rather than artifacts.

This research addresses a critical gap in pulmonary medicine by proposing a solution that is both rapid and interpretable. The ability to classify lung images with high precision can significantly reduce diagnostic turnaround times, a capability that is crucial during health crises. The implications of this work extend to multiple stakeholders. Clinical practitioners, such as radiologists, benefit from a "second opinion" system that improves diagnostic efficiency and reduces fatigue-related errors. Healthcare institutions can streamline diagnostic workflows to potentially improve patient throughput, while public health organizations can leverage such automated tools for rapid disease surveillance. Finally, by prioritizing explainability, this work contributes to the development of "Trustworthy AI," aiding regulatory bodies in establishing standards for safe medical AI deployment.

2. BACKGROUND REVIEW

Medical image classification is a cornerstone of automated disease diagnosis, enabling timely and accurate clinical interventions (Wang et al., 2021). The integration of Deep Learning (DL), specifically Convolutional Neural Networks (CNNs), has revolutionized this domain by significantly enhancing the accuracy and efficiency of radiological analysis (Bhattacharjee et al., 2022).

These advancements are particularly critical for identifying subtle abnormalities in lung pathologies—such as early-stage tumors or viral pneumonia—that may elude visual detection by human observers. Consequently, embedding automated classification tools within healthcare ecosystems holds the potential to streamline diagnostic workflows and markedly improve patient prognoses (Islam et al., 2024; Meeradevi et al., 2024; Alazzam et al., 2024).

Deep learning has precipitated a paradigm shift in medical imaging. CNNs are uniquely adept at learning hierarchical image representations, making them highly effective for tasks ranging from segmentation to disease detection (Monday et al., 2022a; Nneji et al., 2022a). In the context of pulmonary medicine, CNNs have been successfully deployed to automatically detect conditions such as lung cancer and pneumonia by extracting complex feature sets from X-rays and CT scans (Nneji et al., 2022c; Nneji et al., 2022d). Furthermore, recent innovations in transfer learning and the utilization of pre-trained models have expanded the applicability of these systems, offering robust solutions to real-world diagnostic challenges (Li et al., 2025; Monday et al., 2025).

While standard CNNs excel at feature extraction, they often treat all image regions with equal importance, which can lead to learning from irrelevant background noise. To mitigate this, attention mechanisms have emerged as a critical enhancement, allowing models to dynamically focus on diagnostically relevant regions (Nneji et al., 2025). In lung disease classification, this enables the network to prioritize the pulmonary parenchyma while suppressing artifacts, thereby improving both accuracy and noise immunity (Monday et al., 2022b; Monday et al., 2022c). Concurrently, the "black box" nature of deep learning remains a significant barrier to clinical adoption. Explainability is essential for fostering trust among clinicians and patients (Rajkomar et al., 2018). This research addresses this challenge by integrating an external mask attention mechanism and Gradient-weighted Class Activation Mapping (Grad-CAM). By visualizing the specific regions driving a model's prediction, this approach provides inherent interpretability, ensuring that diagnostic decisions are based on clinically valid anatomical features rather than spurious correlations.

The development of robust diagnostic models requires high-quality, diverse datasets. The COVID-19 Radiology Database serves as a vital resource, containing a comprehensive repository of chest X-rays spanning COVID-19 positive cases, viral pneumonia, and normal controls. This diversity is essential for training models to distinguish between visually similar respiratory pathologies. Leveraging this dataset, this study aims to develop a mask-guided classification framework capable of delivering high-sensitivity detection in pandemic scenarios.

Despite significant progress, medical image classification faces persistent challenges. Data scarcity and class imbalance often result in biased models that generalize poorly to underrepresented categories. In addition, variations in image quality and acquisition protocols can degrade model robustness. The proposed mask-guided approach mitigates these issues by restricting the model's learning scope to the Region of Interest (ROI), effectively reducing the influence of background variability and enhancing generalization performance across diverse clinical samples.

Table 1. Summary of Relevant Literature.

Author(s) & Year	Methodology	Key Findings
Wang et al. (2021)	Review of Deep Learning in Radiology	Analyzed the role of DL in lung cancer diagnosis, emphasizing the necessity of accurate early detection for effective treatment planning.
Bhattacharjee et al. (2022)	Deep Neural Network for CT Analysis	Proposed an automated model for lung nodule classification achieving 99% accuracy, demonstrating the efficacy of DL in nodule evaluation.
Islam et al. (2024)	Explainable DL for Lung Cancer	Investigated explainable methods (XAI) for lung cancer detection in CT scans, focusing on feature analysis to support early treatment decisions.
Alazzam et al. (2024)	CNN Model Comparison	Evaluated CNN architectures (ResNet, MobileNetV2, VGG-16) for lung cancer detection, highlighting challenges regarding data scarcity and bias.
Meeradevi et al. (2024)	Performance Analysis of Deep Models	Compared the computational efficiency and accuracy of various DL models (MobileNetV2, ResNet50) for early lung cancer prediction.

3. METHODOLOGY

3.1. Research Approach

This study employs a quantitative experimental design to develop a Mask-Guided Convolutional Neural Network (MG-CNN) for the binary classification of COVID-19 and normal chest X-rays. The methodological framework is divided into four distinct phases: (1) Data Curation and Preprocessing, specifically focusing on ROI isolation; (2) Model Architecture Design, integrating an external attention mechanism; (3) Training and Optimization; and (4) Comprehensive Performance Evaluation. Figure 2 shows the roadmap, which outlines the methods and processes applied throughout the study.

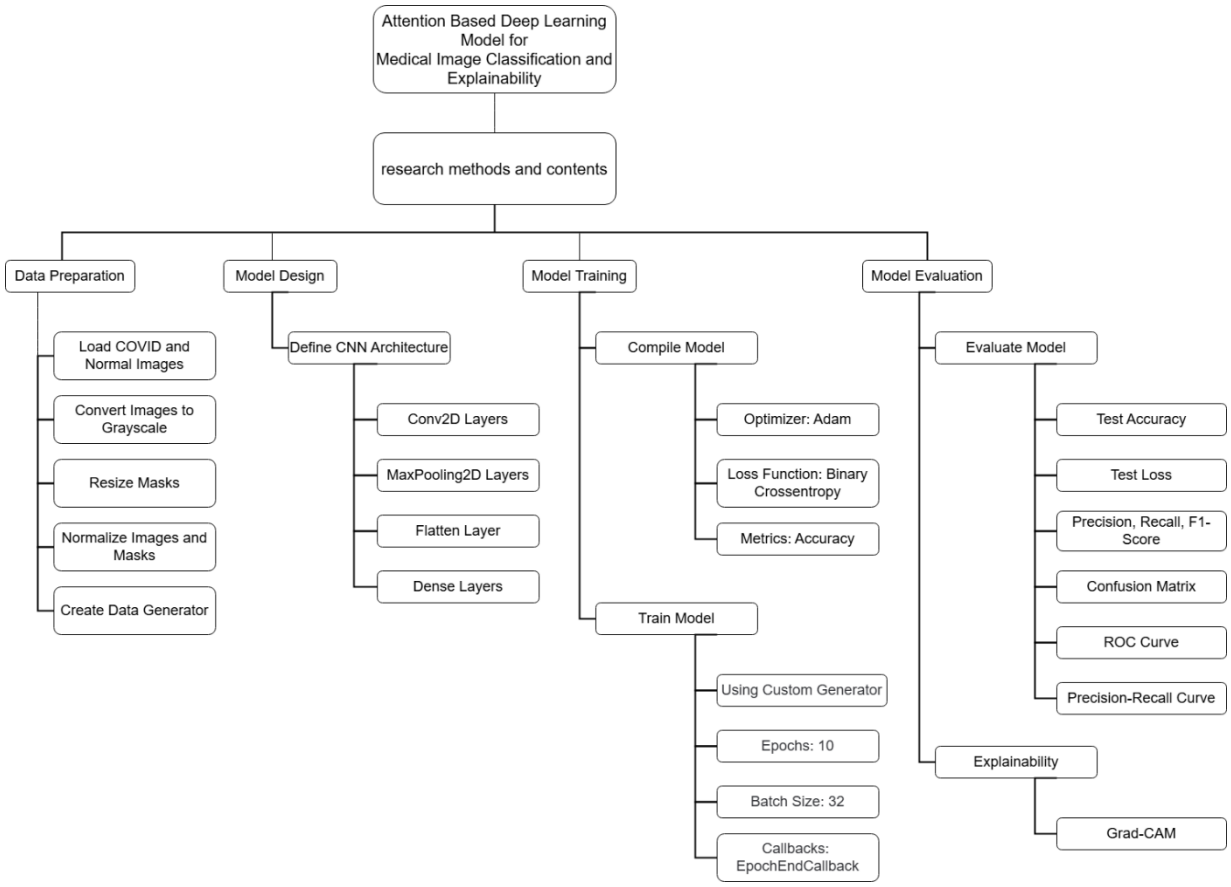

Figure 2. Technology roadmap outlining the experimental workflow.

3.2. Dataset Description

The Dataset used in this study is named COVID-19 Radiology Dataset, which consists of two major parts: the chest X-ray dataset of COVID-19 patients and the normal chest X-ray dataset. The dataset is from Kaggle and is used to classify chest X-ray images based on the presence or absence of COVID-19.

The COVID-19 radiology dataset is a collection of chest X-ray images of patients diagnosed with COVID-19. This dataset consists of images of different severity levels and is designed to assist in training deep-learning models for the automatic diagnosis of COVID-19. This dataset includes 3,616 images of COVID-19-positive patients. The resolution of each image is 299 × 299 pixels, and the size of each image is approximately 40KB. The images are preprocessed to ensure uniformity. The dataset includes X-ray images and the corresponding segmentation masks for region of interest (ROI) identification.

The normal chest X-ray dataset contains 10,192 images of healthy individuals without any respiratory diseases, providing a negative class for training and validation. These images have been balanced with the COVID-19 dataset and provide various X-ray scans. Each image is 299× 299 pixels in size and has undergone preprocessing to maintain consistency and ensure compatibility with deep learning models. Like the COVID-19 dataset, these images are accompanied by corresponding segmentation masks, highlighting the key areas within the lung regions.

Table 2. Dataset Specifications.

Dataset subfile Name	Number of Images	Image Size	Image Resolution	Segmentation Masks
COVID	3,616	~40KB	299 × 299 pixels	Yes
Normal	10,192	~40KB	299 × 299 pixels	Yes

3.3. Data pre-processing

3.3.1. Image size adjustment and Normalization

Raw X-ray images and their corresponding masks were resized to a uniform dimension of 299×299 pixels to align with the CNN input requirements. Subsequently, pixel intensity values were normalized to the range $[0, 1]$ by dividing by 255.0. This step is essential for stabilizing gradient descent and accelerating model convergence.

3.3.2. Mask-Guided ROI Isolation

To enhance the model's focus on the relevant anatomical structures, each chest X-ray image was multiplied by the corresponding lung segmentation mask element by element. The mask protrudes from the lung area (the area of interest, ROI) and suppresses irrelevant background areas such as bones, the heart, and surrounding tissues. This external attention-blocking mechanism enables the model to specifically focus its learning efforts on areas where anomalies caused by COVID-19 may occur, thereby improving classification accuracy and interpretability.

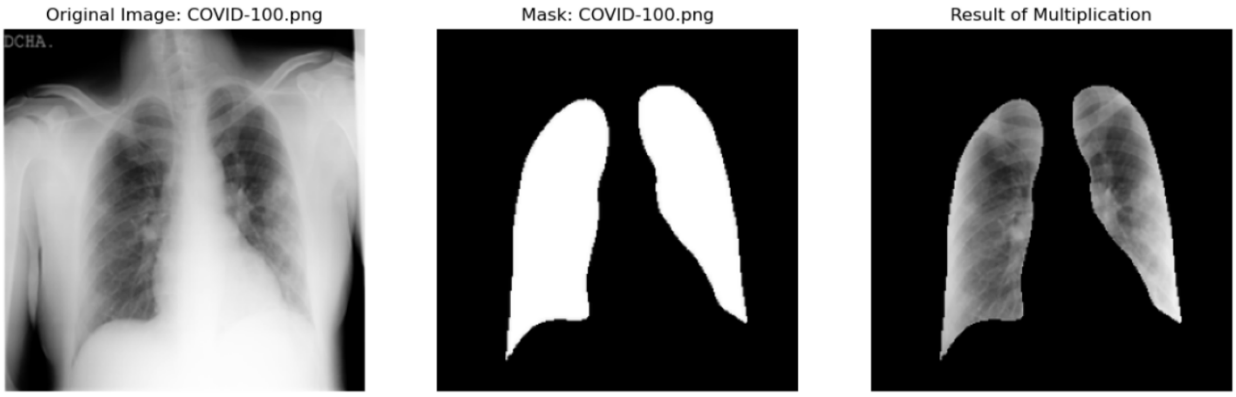

Figure 3. Processed Image with ROI.

3.3.3. Stratified Splitting and Augmentation

The dataset was partitioned into Training (72%), Validation (8%), and Testing (20%) sets. To prevent bias, stratified sampling was employed to maintain the ratio of COVID-19 to Normal cases across all subsets. Given the inherent data scarcity in medical imaging, real-time data augmentation was applied during training. Techniques included horizontal/vertical flips, random rotations ($\pm$ 10°), translations, and zooming to enhance model robustness and prevent overfitting.

3.4. Model Architecture

This model is named Mask-Guided Convolutional Neural Network (MG-CNN). Through the built-in attention mechanism, it performs binary classification of lung X-ray images (COVID vs. Normal) using lung segmentation mask. This architecture begins with a dual-input preprocessing step, where each input image is paired with the corresponding lung mask. These masks are used to perform element multiplication with the original grayscale X-ray image, effectively suppressing background structures (such as ribs and spine) and enhancing the focus on the lung region. This external attention mechanism ensures that the model mainly learns from clinical-related fields, thereby enhancing interpretability and reducing noise.

Masked images are passed through a custom data generator, which dynamically prepares batches during the training process, including preprocessing and label assignment. Then the processed input is input into a deep convolutional neural network (CNN) composed of five convolutional blocks. Each block consists of a Conv2D layer with ReLU activation, followed by a MaxPooling2D layer, which gradually captures spatial features while reducing the dimension. The convolutional backbone is composed of a Flatten layer, a 512-unit dense layer and a Dropout layer to prevent overfitting. The final classification is performed using a Dense layer with binary output activated by sigmoid.

To improve transparency, Grad-CAM is applied to the final convolutional layer (conv2d_14) after training to visualize the spatial focus predicted by the model. The evaluation metrics include accuracy, precision, recall rate, F1-score, and confusion matrix, which are used to quantify the performance of the model. The entire architecture combines the benefits of spatial attention through lung masking and deep convolutional learning, generating a model that is not only accurate but also interpretable.

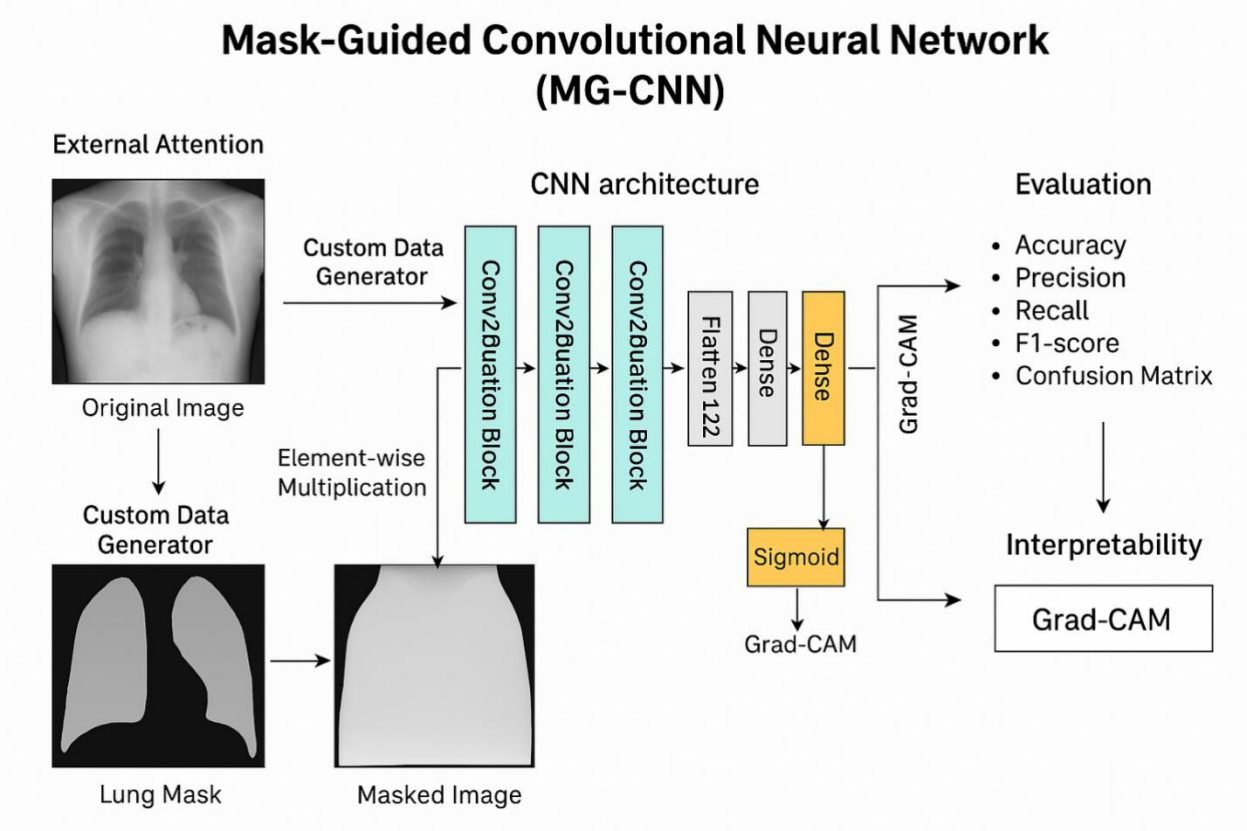

Figure 4. Model Architecture.

Extract hierarchical features from the input chest X-ray images using multiple convolutional layers. These layers apply two-dimensional convolution to capture local spatial patterns, such as edges, textures, and shapes within the lung region.

$$H = \partial W * X + b \quad (1)$$

The pooling operation (especially Max pooling) is introduced to reduce the spatial dimension of the feature mapping, thereby reducing the computational complexity and enhancing the robustness of the features to small spatial variations.

$$P = pooling\ H \quad (2)$$

After each convolutional layer, the rectifier Linear unit (ReLU) activation function is applied to introduce nonlinearity into the model, enabling it to learn complex patterns. An external mask-based attention mechanism was incorporated into the model. This model does not purely learn the attention weights from the internal features but uses the pre-provided lung segmentation mask to guide its focus to the region of interest (ROI). The external mask is dynamically applied during the training and inference process to ensure that the model prioritizes the lung regions and reduces the influence of irrelevant background information. Mathematically, note that the enhanced feature map is calculated as:

$$X' = X \times (1 + \alpha \cdot M) \quad (3)$$

Where X is the original feature map, M is the external lung mask, and α is a learnable scaling parameter.

$$CA(X) = X * Sigmoid\,(favg(X) + fmax(X)) \quad (4)$$

Indicates the channel attention mechanism. CA is the output of the channel attention.

$$SA(X) = X * Sigmoid(fconcat(concat(favg(X), fmax(X)))) \quad (5)$$

Spatial attention mechanism, SA is the output of spatial attention. After convolution and pooling operations, the multi-dimensional feature map is flattened into a one-dimensional vector and serves as the input of the dense layer. Flat feature vectors are fed into one or more dense (fully connected) layers. These layers perform advanced reasoning and integrate the information throughout the entire image to support classification. The final output layer consists of a single neuron with an S-shaped activation function and applies to binary classification (COVID-19 positive and healthy). The final classification probability is calculated as:

$$Output = Softmax(FC(X)) \quad (6)$$

3.5. Custom Data Generator

To effectively handle large datasets and integrate external lung masks, a custom data generator has been implemented. Batch load and process images, masks, and labels to optimize memory usage. Shuffle the data at the beginning of each epoch to prevent the model from learning the order of the data and improve generalization. During the batch generation process, the lung mask is dynamically loaded and applied to the corresponding images to ensure that the model focuses on the lung region during both the training and testing phases.

The small-batch gradient descent method is adopted for training. Batch load images and their corresponding lung masks through a custom generator. Each batch contains a balanced mixture of COVID-19 and normal samples, which promotes stable learning. Before each epoch, the training data is shuffled to prevent the model from overfitting the order of the samples and during the training process, the lung mask is dynamically applied to the input images. This External Mask Attention mechanism helps the model focus on the lung regions, suppress irrelevant background regions, and improve the quality of feature extraction.

3.6. Hyperparameters and Settings

This model adopts the Adam optimizer. Due to its adaptive learning rate and strong convergence, especially on noisy medical imaging datasets. Initial learning rate: 0.001, Beta1: 0.9, Beta2: 0.999. The binary cross-entropy loss was used to guide the optimization process, fitting naturally with the binary classification task. To balance the training speed and GPU memory usage, 32 batch processing sizes were selected. The model was trained for 10 epochs. Based on the previous experiments, the number of epochs is selected to ensure adequate learning without obvious overfitting. All input images and masks have been adjusted to 299 × 299 pixels to maintain consistency with the model architecture.

The CNN model is initialized and compiled by using the Adam optimizer and the binary cross-entropy loss. The custom data generator dynamically loads images, masks and labels for the training set and validation set. For each epoch, the model updates its weights according to the training batch. After each epoch, the model evaluates the performance on the validation set to monitor overfitting and convergence trends. Before being input into the model, each X-ray image is shielded with a corresponding lung mask to highlight the area of interest and reduce the influence of irrelevant areas.

Table 3. Hyperparameter Summary.

Parameter	Value	Description
Input Shape	299×299×1	Grayscale ROI-masked image
Batch Size	32	Optimized for GPU memory stability
Optimizer	Adam	"β1=0.9 β2=0.999"
Loss Function	Binary Cross-Entropy	Standard for two-class problems
Dropout	0.5	Applied to the dense layer to reduce overfitting

3.7. Performance Evaluation Metrics

The performance of the trained model is evaluated using a test set that is completely separate from the training and validation datasets. This ensures that the evaluation results accurately reflect the model's generalization ability.

Accuracy measures the proportion of correctly classified images among all test images. It gives an overall assessment of how often the model's predictions are correct.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP+ FP} + \text{TN} + \text{FN}} \quad (7)$$

Where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative.

Precision evaluates the correctness of positive predictions. It answers: out of all cases predicted as COVID-19, how many were COVID-19 positive?

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (8)$$

Recall measures the model's ability to detect positive cases. It answers: out of all actual COVID-19 positive cases, how many did the model correctly identify?

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (9)$$

The F1-score provides a harmonic mean between precision and recall, giving a balanced measure that is especially important in medical image classification where both false positives and false negatives can have serious consequences.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (10)$$

Confusion Matrix: In this work, a confusion matrix is used to visualize classification performance. The confusion matrix is a 2×2 table since the classification task involves two classes: COVID-19 positive and Normal.

- True Positive (TP): Correctly predicted COVID-19 positive cases.
- False Positive (FP): Normal cases incorrectly predicted as COVID-19 positive.
- False Negative (FN): COVID-19 positive cases incorrectly predicted as normal.
- True Negative (TN): Correctly predicted normal cases.

The confusion matrix provides intuitive insights into the types of errors the model makes and highlights any bias towards a particular class. To further analyze the model behaviour, a set of 100 chest X-ray images was visualized from the test set. For each image, it is displayed as follows:

- Grayscale X-ray image.
- Prediction labels and confidence scores.
- A basic truth label for direct comparison.

This qualitative assessment allows for the detection of system errors or failure cases that may not be fully captured through digital metrics.

4. IMPLEMENTATION AND RESULTS

4.1. Experimental Overview

This chapter presents the experimental evaluation of the Mask-Guided Convolutional Neural Network (MG-CNN) developed for the binary classification of COVID-19 and normal chest X-rays. The model was assessed using a comprehensive suite of metrics, including accuracy, loss, precision, recall, and the F1-score to validate its efficacy in a real-world diagnostic context. Furthermore, the study utilizes Gradient-weighted Class Activation Mapping (Grad-CAM) to move beyond quantitative metrics and qualitatively assess the model's interpretability.

4.2. Results

The attention-based convolutional neural network model proposed in this paper has been trained for 10 epochs. The final training accuracy rate is 94.75% and the verification accuracy rate is 91.82%. Evaluated on the independent test set, the test accuracy rate of this model was 89.97% and the test loss was 0.2807. However, a more detailed analysis using accuracy, recall rate, F1-Score and confusion matrix indicates that although the model can effectively classify normal cases, it faces challenges in accurately detecting positive cases of the novel coronavirus. The specific evaluation indicators are shown in Table 4.

Table 4. Evaluation Metrics on Test Set.

Metric	Value
Test Accuracy (Global)	0.8997
Manual Accuracy (Confusion Matrix)	0.6231
Precision (COVID-19)	0.2554
Recall (COVID-19)	0.2296
F1-Score (COVID-19)	0.2418
Precision (Normal)	0.7400
Recall (Normal)	0.7600
F1-Score (Normal)	0.7500

The model accuracy graph (figure 5—left), shows the variation of the model's accuracy on the training set and validation set with the number of training rounds (epochs). It can be seen that as the training proceeds, the accuracy of the model gradually increases, indicating that the performance of the model is constantly improving during the learning process. The model loss graph (figure 5—right) shows the variation of the loss values of the model on the training set and the validation set with the training rounds. The decrease in the loss value indicates that the prediction error of the model is reducing, and the model is learning effectively.

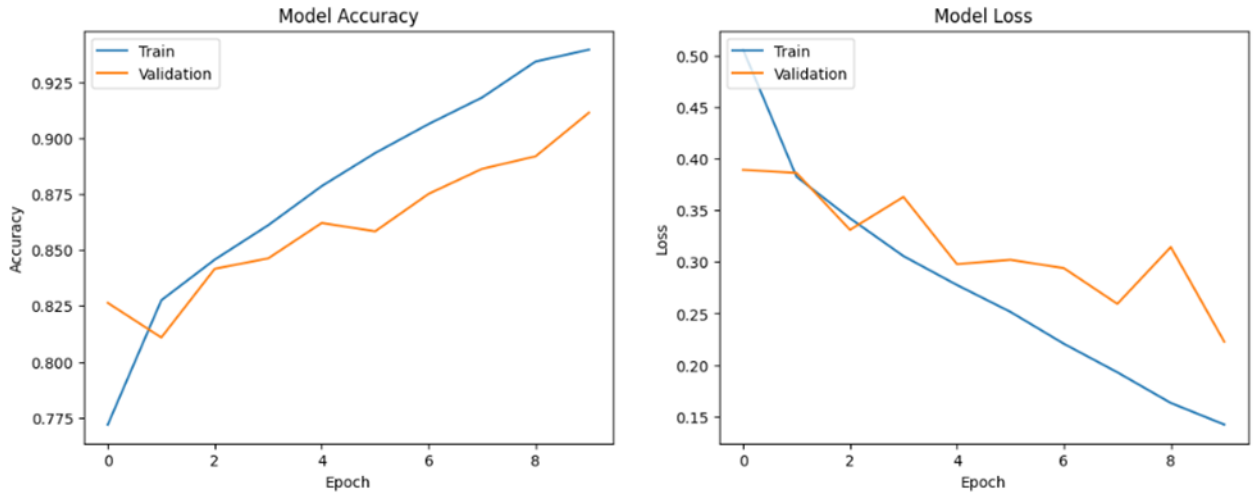

Figure 5. (left)Accuracy and (right)Loss.

The confusion matrix diagram (figure 6) illustrates the classification performance of the model for the two categories of COVID-19 and normal (non-COVID) on the test set: True Positives (TP): 166 COVID-19 cases were correctly identified as COVID-19 by the model. True Negatives (TN): 1,555 normal (non-COVID-19) cases were correctly identified as normal by the model. False Positives (FP): 484 normal cases were wrongly identified as COVID-19 by the model, that is, false positives occurred. False Negatives (FN): 557 COVID-19 cases were wrongly identified as normal by the model, that is, underreported.

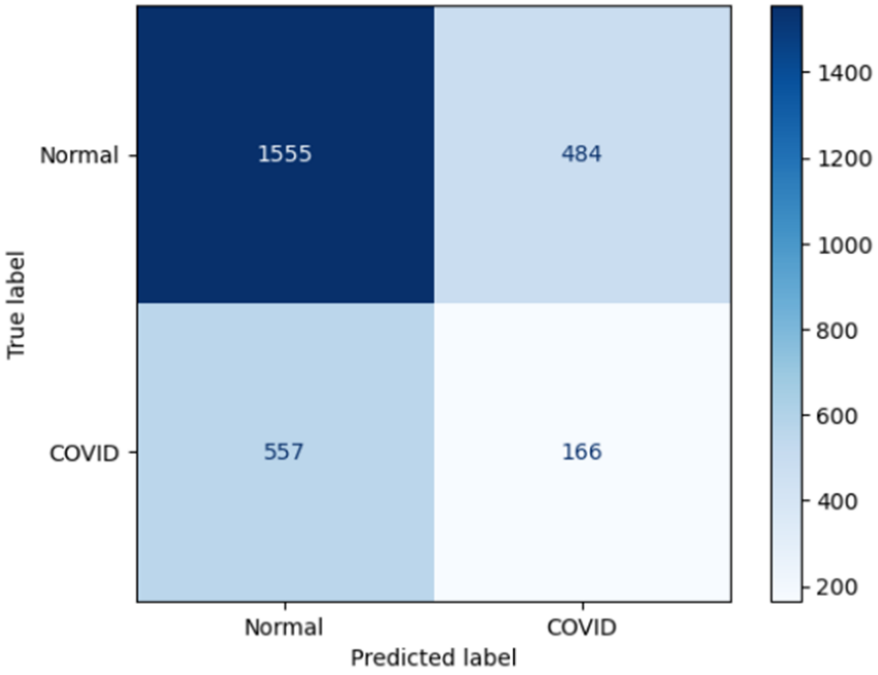

Figure 6. Confusion Matrix.

The Receiver Operating Characteristic (ROC) curve (figure 7) shows the performance of the model at different thresholds and is evaluated by the true positive rate (TPR) and the false positive rate (FPR). The area under the curve (AUC) is 0.5053, which indicates that the model has a certain discrimination ability, but the performance is not particularly ideal.

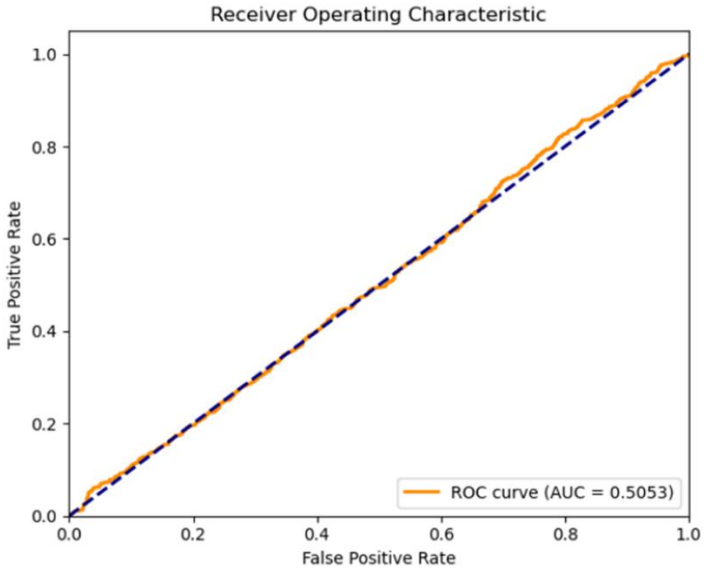

Figure 7. ROC Curve.

The Precision-Recall Curve (figure 8) shows the trade-off between the precision rate and the recall rate of the model. When the recall rate is low, the precision is very high, approaching 1. This indicates that in the case where the model is very conservative (that is, only predicting a small number of samples as positive classes), it can identify positive classes very accurately. When the recall rate increases, the precision rate decreases: As the recall rate increases, the precision rate decreases. This is usually because when improving the recall rate (that is, attempting to identify more positive class samples), the model tends to generate more False Positives. The accuracy rate of most areas of the curve remains between 0.2 and 0.4. This indicates that the model can identify positive class samples with moderate accuracy in most cases. The volatility of the curve indicates that the performance of the model is unstable at different thresholds, and the predictions for some samples may not be reliable enough.

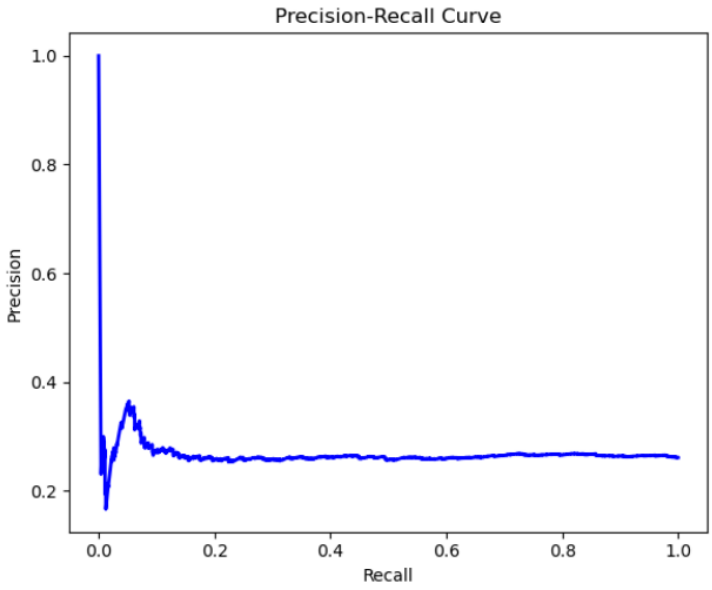

Figure 8. Precision-Recall Curve.

4.3. Effects of the Attention Mechanism

In this study, through the application of lung-splitting masks, an external attention mechanism was introduced to guide the focus of the model to clinically meaningful areas. Specifically, before being input into the convolutional neural network (CNN), each grayscale chest X-ray is multiplied by the corresponding binary lung mask. This effectively removes irrelevant areas outside the lungs (such as ribs, spine, and background), and highlights lung structures that are more likely to contain pathological signs of COVID-19.

The purpose of the attention mechanism:

- Point the model's attention only to the lung field
- Suppress visual noise and artefacts outside the region of interest
- Improve the quality of feature learning by reducing the interference from non-diagnostic areas

After 10 iterations, the training accuracy rate of this model is 94.75% and the verification accuracy rate is 91.82%. This indicates that the model has high training and validation accuracy and is effectively learning from the focused (masked) image regions, potentially benefiting from reduced overfitting due to the deletion of irrelevant information. The Grad-CAM heatmap (see Section 4.4) shows that in the correctly predicted COVID-19 cases, this model mainly focuses on the lung regions with abnormal textures, indicating that the attention mechanism is consistent with clinical reasoning and has high interpretability. The accuracy rate (0.74) and recall rate (0.76) of the normal category were relatively high, indicating that the model performed well in detecting non-COVID-19 cases and achieved better class separation under normal circumstances, especially when the lung regions were clear and the pathological patterns were not obstructed. COVID category testing remains challenging. Although the attention mechanism improves the spatial focus, the accuracy rate (0.255) and recall rate (0.229) of the COVID class are still very low. This indicates that although the model benefits from the attention mask, it is still challenged by class imbalance, subtle COVID features, and possible noise in the dataset labels.

4.4. Explainability via Grad-CAM Visualization

To improve the interpretability of model predictions, in this study, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize the spatial regions in chest X-ray images that contribute the most to the model's classification decisions. Grad-CAM is applied to the final convolutional layer (conv2d_14) of the training model. This visualization technique enables us to understand whether the model's attention is consistent with known COVID-19 radiological patterns, such as bilateral ground-glass opacity or lung consolidation, and its performance when analyzing normal (healthy) lungs. Among the correctly predicted COVID-19 cases, the Grad-CAM heatmap (figure 9) shows strong activation in the lung areas, especially in the areas where radiologists expected abnormalities. These include peripheral or lower lobe regions, where opacity and infiltration are common in COVID-19-associated pneumonia. This indicates that the model is not making predictions arbitrarily but is learning to focus on features related to medicine.

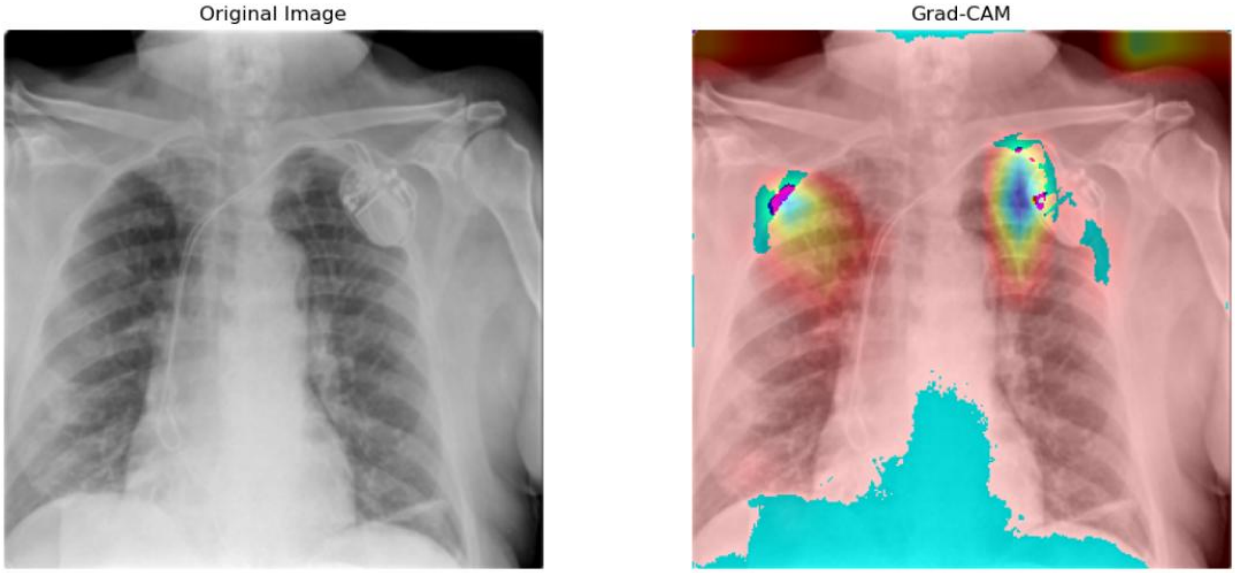

Figure 9. Grad-CAM Visualization-COVID-19.

In cases classified as normal, the Grad-CAM heatmap (figure 10) shows very low or diffuse activation, usually uniformly distributed or located outside the lung parenchyma. This indicates that the model did not detect any pathologically suspicious features in the image and confidently predicted that it was healthy.

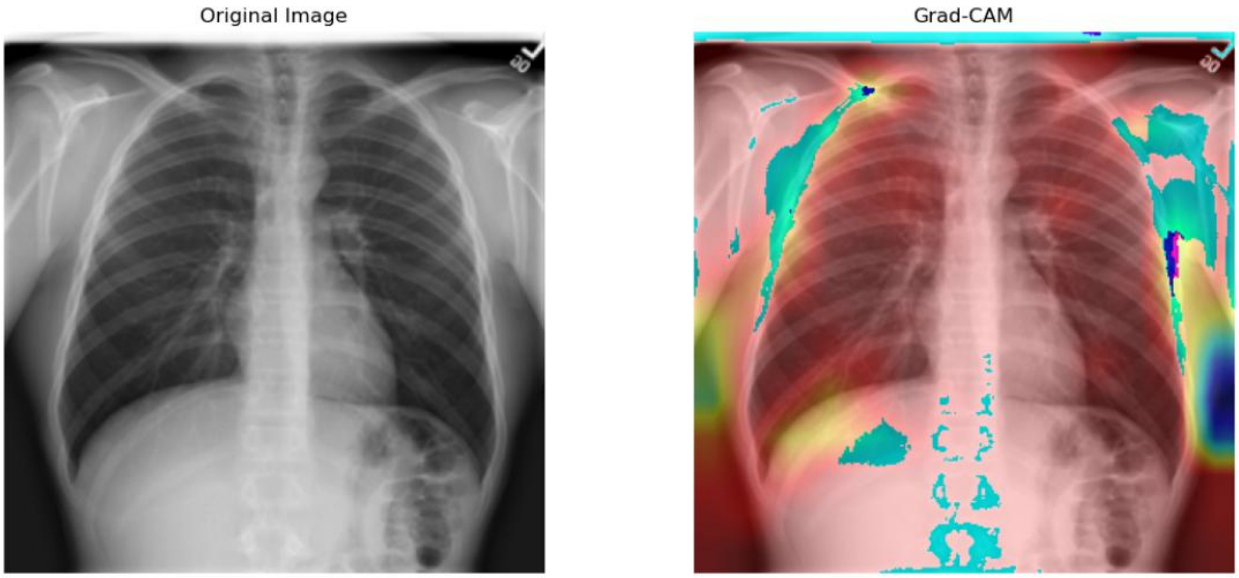

Figure 10. Grad-CAM Visualization-Normal.

Grad-CAM provides crucial insights into the decision-making process of deep learning models. It is confirmed that the model focuses on the lung field rather than the background or skeletal structure. It is consistent with the attention mechanism of pretreatment through a lung mask. Enhance trust in the predictions of clinical decision support models.

4.5. Chapter Summary

This chapter conducts a comprehensive evaluation of the attention-based deep learning model, which is used to classify COVID-19 and normal cases using chest X-ray images. This model is trained on masked lung images to suppress background noise and enhance the focus on diagnosis-related areas. The evaluation results show that this model has a relatively high training and validation accuracy rate (94.75% and 91.82%, respectively), and the test accuracy rate is 89.97%. However, a more detailed analysis shows that in terms of detecting COVID-19 cases, especially in terms of recall rate and F1-score, the performance is relatively low, highlighting the challenge of class imbalance in the dataset. The attention mechanism achieved through lung mask preprocessing improves the model's ability to learn spatially focused features and reduce irrelevant activations. The Grad-CAM visualization results support this point, confirming that the model mainly focuses on the lung field during normal and COVID-19 classification periods. These interpretable tools provide valuable insights into the decision-making process of the model and enhance trust in its predictions.

5. CONCLUSION

In conclusion, the study has successfully achieved its main goal and developed a powerful and accurate system to classify lung images into COVID-19, normal pneumonia and viral pneumonia. The integration of the attention mechanism achieved through the lung segmentation mask plays a crucial role in enhancing the predictive performance of the model and improving its interpretability, both of which are indispensable for clinical applicability.

The overall accuracy of this model is approximately 90%, indicating reliable generalization for unseen data. Its stable performance highlights its potential for deployment in actual diagnostic scenarios. More importantly, including interpretable tools such as Grad-CAM helps visualize the key areas of the model, ensuring that its decisions are consistent with clinically relevant characteristics, thereby building trust among medical professionals.

Despite these achievements, this model is not without limitations. Its performance essentially depends on the quality, diversity and balance of the training dataset. In particular, the classification of a few categories (such as COVID-19) remains a challenge, partly due to limited and unbalanced data. Solving this problem is crucial for improving sensitivity and reducing false negatives.

Future work should focus on expanding the dataset to incorporate broader patient demographics, imaging changes and clinical labels. Furthermore, exploring more advanced neural network architectures, such as transformer-based models or hybrid concern networks, may further enhance classification accuracy and model efficiency.

Overall, this study represents a meaningful step forward in the field of medical image analysis. The combination of high-precision and interpretable output provides practical value for healthcare practitioners by supporting faster, more accurate and interpretable diagnostic workflows. With the further improvement and adjustment of this model, its contribution to early disease detection and medical decision support is likely to be substantive and transformative.

Conflict of Interest

All authors declare no competing interest.

References

[1] Alazzam, M., Alazzam, H., & Mustafa, A. (2024). Lung cancer detection using deep learning and explainable methods. *Proceedings of the IEEE 25th International Conference on Circuits, Control, Communication, and Computing (14C),* 1–5.

[2] Bhattacharjee, R., Devi, S., & Sekar, K. (2022). Deep neural network based automatic detection and classification of lung nodules from CT images. *Proceedings of the International Conference on Smart Computing (SMARTCON),* 1–5.

[3] Chilamkurthy, S., Ghosh, R., & Dhara, S. (2018). Deep learning based lesion detection and segmentation for automated liver MRI analysis. *Medical Image Analysis, 53*, 169–181. https://doi.org/10.1016/j.media.2017.08.001

[4] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542(7639),* 115–118. https://doi.org/10.1038/nature21051

[5] Islam, R., Sumon, M., Alazzam, H., Mustafa, A., & Kim, H. (2024). Lung cancer detection using deep learning and explainable methods. *Proceedings of the 25th International Conference on Circuits, Control, Communication, and Computing (14C),* 1–5.

[6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521(7553),* 436–444. https://doi.org/10.1038/nature14539

[7] Li, X., Wang, Y., Monday, H. N., & Nneji, G. U. (2025). A novel residual learning of multi-scale feature extraction model for the classification of rice grain varieties. *Computers and Electronics in Agriculture, 237,* 110491.

[8] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2016.08.008

[9] Meeradevi, M., Mundada, M., Singh, S. P., Bhatn, S., & Srivastava, C. (2024). Performance analysis of various deep models in lung cancer detection and classification using medical images. *Proceedings of the IEEE 20th International Conference on Circuits, Control, Communication, and Computing (14C),* 1–5.

[10] Miao, S. (2019). A survey on convolutional neural networks for image classification. *Information, 10(2),* 47. https://doi.org/10.3390/information100201-C008

[11] Monday, H. N., Li, J., Nneji, G. U., Hossin, M. A., Nahar, S., Jackson, J., & Chikwendu, I. A. (2022). WMR-DepthwiseNet: A wavelet multi-resolution depthwise separable convolutional neural network for COVID-19 diagnosis. *Diagnostics, 12(3)*, 765. https://doi.org/10.3390/diagnostics12030765

[12] Monday, H. N., Li, J., Nneji, G. U., Hossin, M. A., Nahar, S., Jackson, J., & Ejiyi, C. J. (2022). COVID-19 diagnosis from chest X-ray images using a robust multi-resolution analysis Siamese neural network with super-resolution convolutional neural network. *Diagnostics, 12(3),* 741. https://doi.org/10.3390/DIAGNOSTICS12030741

[13] Monday, H. N., Li, J., Nneji, G. U., Ukwuoma, C. C., Cai, J., Chikwendu, I., & Oluwasanmi, A. (2022). A wavelet convolutional capsule network with modified super resolution generative adversarial network for fault diagnosis and classification. *Complex & Intelligent Systems, 8(1),* 1–15. https://doi.org/10.1007/s40747-022-00733-6

[14] Monday, H. N., Nneji, G. U., Hossin, M. A., Mark, K. D., Umana, E. S., Mgbejime, G. T., & Li, J. (2025). Enhancing ECG classification in cardiac diagnostics: A novel approach using adaptive focal cross-entropy loss function. IEEE Journal of Biomedical and Health Informatics.

[15] Nneji, G. U., Cai, J., Deng, J., Hossin, M. A., Nahar, S., & Jackson, J. (2022). Identification of diabetic retinopathy using weighted fusion deep learning based on dual-channel fundus scans. *Diagnostics, 12(2),* 540. https://doi.org/10.3390/diagnostics12020540

[16] Nneji, G. U., Cai, J., Deng, J., Monday, H. N., James, E. C., & Ukwuoma, C. C. (2022). Multi-channel based image processing scheme for pneumonia identification. *Diagnostics, 12(2),* 325. https://doi.org/10.3390/diagnostics12020325

[17] Nneji, G. U., Cai, J., Monday, H. N., Hossin, M. A., Nahar, S., Jackson, J., & Deng, J. (2022). Fine-tuned Siamese network with modified enhanced super-resolution GAN plus based on low quality chest X-ray images for COVID-19 identification. *Diagnostics, 12(3)*, 717. https://doi.org/10.3390/diagnostics12030717

[18] Nneji, G. U., Deng, J., Monday, H. N., Cai, J., Hossin, M. A., Nahar, S., & Jackson, J. (2022). COVID-19 identification from low-quality computed tomography using a modified enhanced super-resolution generative adversarial network plus and Siamese capsule network. *Healthcare, 10(2),* 403. https://doi.org/10.3390/healthcare10020403

[19] Nneji, G. U., Monday, H. N., Pathapati, V. S. R., Nahar, S., Mgbejime, G. T., Umana, E. S., & Hossin, M. A. (2025). FFS-IML: Fusion-based statistical feature selection for machine learning-driven interpretability of chronic kidney disease. *International Journal of Machine Learning and Cybernetics,* 1–34.

[20] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning–driven health care: The case for algorithmic audit trails. *Annals of Internal Medicine, 169(3),* 205–207. https://doi.org/10.7326/annals.000-2018-0003

[21] Siegel, R. L., Miller, K. D., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians, 73(1),* 5–32. https://doi.org/10.3322/ca.23.1

[22] Wang, H., & Xing, L. (2021). Deep learning's application on radiology and pathological image of lung cancer: A review. *Proceedings of the International Conference on Information Technology and Biomedical Engineering (ICiTBE),* 1–5.