Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable CNN for Automated Lung Opacity Detection Using Fine-Tuned VGG-19 with Grad-CAM

Dr Kazi Abdul Mannan

drkaziabdulmannan@gmail.com

Center for Academic & Professional Career Development and Research (CAPCDR)
https://orcid.org/0000-0002-7123-132X

Arifa Akter Shapla

Shanto-Mariam University of Creative Technology https://orcid.org/0009-0003-8555-4095

Research Article

Keywords: Pneumonia Detection, Chest X-ray, Deep Learning, Transfer Learning, VGG19, Grad-CAM

Posted Date: August 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10601281/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Pneumonia is a serious respiratory disease and one of the leading causes of death among children under five years of age. Manual interpretation of chest X-ray images by radiologists is time-consuming and may lead to diagnostic errors. This study proposes an automated lung opacity detection system based on transfer learning using a fine-tuned VGG19 architecture. The convolutional backbone of the pretrained VGG19 model is frozen, while a new binary classification layer is trained to distinguish between normal and opacity chest X-ray images. A publicly available Kaggle dataset containing two classes, Normal and Opacity, is used for training and evaluation. The proposed model achieves an overall accuracy of **93%** and demonstrates high recall for opacity cases, making it suitable for supporting clinical diagnosis. To improve transparency and trustworthiness, Grad-CAM is integrated to visualize the regions responsible for the model's predictions through heatmaps. The experimental results indicate that the model shows stable learning behavior without significant overfitting and can serve as a reliable computer-aided diagnostic tool for pneumonia detection.

1. Introduction

Pneumonia is a serious lung infection that remains one of the leading causes of death among children worldwide, particularly those under five years of age. The disease is primarily caused by viruses, bacteria, or fungi and requires early diagnosis to reduce mortality and improve treatment outcomes. Chest X-ray imaging is one of the most widely used diagnostic methods because it is inexpensive, readily available, and capable of identifying lung abnormalities. However, diagnosing pneumonia through manual interpretation of chest X-ray images is time-consuming and depends heavily on the expertise of experienced radiologists. In many healthcare settings, the shortage of skilled radiologists delays diagnosis and treatment, increasing the risk of severe complications. Recent advancements in artificial intelligence, particularly deep learning, have demonstrated promising results in medical image analysis. Convolutional Neural Networks (CNNs) and transfer learning models such as VGG19, ResNet50, and MobileNetV2 have significantly improved automated disease detection by learning image features directly from chest X-ray images. These developments have created new opportunities for building efficient computer-aided diagnosis systems that can assist clinicians in detecting pneumonia more quickly and accurately.

Although deep learning has achieved considerable success in medical image classification, several challenges remain in automated pneumonia detection. Manual diagnosis of chest X-ray images is still widely practiced and is susceptible to delays and human error, particularly in hospitals with limited access to experienced radiologists. Existing automated systems often focus only on improving classification accuracy without providing sufficient explanation for their predictions. As a result, healthcare professionals may hesitate to rely on these systems because they cannot easily understand how the model reaches its decision. Furthermore, different deep learning architectures produce varying levels of performance, making it necessary to identify an approach that balances accuracy and interpretability. Therefore, there is a need for an automated detection system that not only achieves

reliable classification performance but also provides visual explanations to support clinical decision-making and increase trust in artificial intelligence-based diagnosis.

The increasing demand for accurate and explainable artificial intelligence in healthcare provides the motivation for this study. VGG19 has demonstrated strong feature extraction capability in image classification tasks and is suitable for transfer learning applications using chest X-ray images. Integrating Grad-CAM with the model enables visualization of the lung regions that influence the prediction, making the diagnostic process more transparent and interpretable. This combination supports clinicians by providing both automated classification and visual evidence of the model's decision, thereby increasing confidence in AI-assisted diagnosis.

The main objective of this research is to develop an explainable deep learning model for automated lung opacity detection using chest X-ray images. The specific objectives are:

- To develop an automated pneumonia detection system using a fine-tuned VGG19 transfer learning model.
- To classify chest X-ray images into Normal and Opacity categories.
- To improve the reliability of pneumonia detection through transfer learning.
- To integrate Grad-CAM for visual explanation of the model's predictions.
- To evaluate the model using standard performance measures such as accuracy, precision, recall, and F1-score.

This research contributes to the development of an explainable computer-aided diagnosis system for pneumonia detection. By combining a fine-tuned VGG19 architecture with Grad-CAM, the proposed model provides both accurate classification and visual explanations of its predictions. Such explainability can improve clinicians' understanding of the model's decisions and increase trust in artificial intelligence-based healthcare applications. The system may also help reduce diagnostic time and support healthcare professionals in making faster and more consistent clinical decisions, particularly in environments where experienced radiologists are limited. In addition, the study demonstrates the practical application of transfer learning and explainable AI techniques in medical image analysis, providing a foundation for future research in automated disease detection.

2. Literature Review

Deep learning has become one of the most effective approaches for pneumonia detection from chest X-ray images because it improves diagnostic accuracy while reducing the time required for manual interpretation. Many recent studies have focused on developing automated systems that assist radiologists in identifying lung diseases using artificial intelligence. Most of these studies have been trained and evaluated using the same publicly available pediatric chest X-ray dataset from Guangzhou, which is distributed through Kaggle. The availability of a common benchmark dataset allows researchers to compare the performance of different deep learning models under similar experimental conditions.

However, most existing studies have not validated their models on datasets collected from different hospitals, age groups, or imaging protocols, limiting their generalizability. Various deep learning architectures, including conventional CNNs, transfer learning models, Vision Transformers (ViTs), and explainable AI techniques, have been investigated to improve pneumonia detection. While these approaches have achieved promising classification performance, differences remain in computational efficiency, interpretability, and clinical applicability. Therefore, reviewing these studies is essential for understanding the strengths, limitations, and future direction of automated pneumonia detection systems.

2.1 CNN-Based Approaches

CNN-based methods represent the earliest and most widely adopted deep learning techniques for pneumonia detection. Szepesi and Szilágyi developed a custom CNN architecture that achieved 97.21% accuracy, 97.34% recall, and an AUC of 0.982, while maintaining faster inference compared with heavier transfer learning models. Similarly, Sharma and Guleria used VGG16 as a feature extractor combined with a neural network classifier, obtaining 92.15% accuracy for binary pneumonia classification and 95.4% accuracy for a multiclass dataset. Gill et al. demonstrated that fine-tuning MobileNetV2 improved classification accuracy to approximately 93%. Beri et al. also reported class-wise accuracy between 92% and 93%, with better recall for pneumonia cases than for normal images. These studies indicate that CNN-based architectures remain practical for medical image analysis because they provide a good balance between computational efficiency and diagnostic performance.

2.2 Vision Transformer-Based Approaches

Vision Transformer (ViT) models represent a more recent development in pneumonia detection. Unlike CNNs, ViTs capture global spatial relationships through self-attention mechanisms instead of convolution operations. Singh et al. reported a ViT-based framework that achieved 97.61% test accuracy and high specificity, outperforming several CNN baselines on the same pediatric chest X-ray dataset. Aslan also proposed a Vision Transformer model that reached 95.67% accuracy and 99.52% recall, successfully identifying almost all pneumonia cases despite producing a higher number of false-positive predictions. However, Vision Transformer models generally require greater computational resources. The reviewed study also noted that the larger ViT-H model could not be trained because of GPU memory limitations.

2.3 Hybrid and Multi-Class Models

Several researchers have explored hybrid deep learning frameworks that combine different architectures to improve classification performance. Fu et al. introduced DAViT, a domain-adapted Vision Transformer integrated with a shallow CNN. Their model achieved approximately 97% F1-score and 96% AUC for binary pneumonia detection while also supporting bacterial-versus-viral classification. Rafi et al. extended the classification problem to three categories: Normal, Pneumonia, and Tuberculosis. Their results showed that the Vision Transformer slightly outperformed CNN in test accuracy, whereas CNN

demonstrated more stable performance during cross-validation. These studies suggest that more complex clinical tasks reduce the performance gap between different architectures and highlight the importance of evaluating both classification accuracy and model stability.

2.4 Explainable Artificial Intelligence

Only a limited number of reviewed studies incorporated explainability into pneumonia detection models. DAViT generated Grad-CAM heatmaps that successfully highlighted clinically meaningful lung regions. Rafi et al. employed saliency mapping and reported strong agreement between model explanations and expert-annotated regions. Another comparative CNN study demonstrated that some architectures produced clearer localization of diseased lung regions even when their classification accuracy was not the highest. However, many studies concentrated primarily on classification metrics without providing interpretable visual explanations. The literature therefore indicates that explainability remains an important requirement for increasing clinician trust and supporting practical deployment in healthcare environments.

The reviewed literature demonstrates that deep learning has significantly improved automated pneumonia detection from chest X-ray images. Nevertheless, several limitations remain. Most previous studies rely on the same public pediatric chest X-ray dataset, with little external validation using images collected from different hospitals or patient populations. In addition, although CNNs and Vision Transformers have achieved high classification accuracy, only a small number of studies integrate explainable AI techniques such as Grad-CAM to improve model transparency. Many studies emphasize classification performance while providing limited visual interpretation of model predictions. These limitations highlight the need for an explainable deep learning framework that combines reliable classification with interpretable visual evidence to support clinical decision-making.

The reviewed studies demonstrate that deep learning has become a powerful tool for automated pneumonia detection from chest X-ray images. Conventional CNN models provide efficient feature extraction and competitive diagnostic performance, while Vision Transformer-based approaches improve the ability to capture global image relationships. Hybrid architectures further enhance classification accuracy by combining complementary feature representations. However, most existing studies concentrate primarily on improving predictive performance and often overlook model explainability and external validation. The limited use of Grad-CAM and other explainable AI techniques indicates that transparency remains an important challenge in medical image analysis. Based on these observations, the proposed study adopts a fine-tuned VGG19 transfer learning model integrated with Grad-CAM to provide both accurate classification and interpretable visual explanations. This approach is intended to support clinicians by improving confidence in automated pneumonia detection while maintaining reliable diagnostic performance.

3. Theoretical and Conceptual Framework

The proposed study is based on three fundamental concepts: Deep Learning, Transfer Learning, and Explainable Artificial Intelligence (XAI). These theories provide the foundation for developing an automated and explainable pneumonia detection system using chest X-ray images.

Deep Learning enables computers to learn hierarchical image features automatically through multiple neural network layers. Among various deep learning architectures, Convolutional Neural Networks (CNNs) are widely used in medical image analysis because they can identify complex visual patterns without manual feature extraction. CNNs have demonstrated excellent performance in image classification tasks, particularly for detecting abnormalities in chest X-ray images.

Transfer Learning further enhances model performance by utilizing knowledge learned from large-scale image datasets. In this research, the pretrained VGG19 model is employed as the backbone network. The convolutional layers of VGG19 remain frozen to preserve the feature representations learned from ImageNet, while a new binary classification layer is trained specifically for lung opacity detection. This approach reduces training time, minimizes overfitting, and improves classification performance when the available medical dataset is limited.

The third theoretical foundation is Explainable Artificial Intelligence (XAI). Although deep learning models often achieve high prediction accuracy, their decision-making process is generally difficult to interpret. To improve transparency, this study integrates Grad-CAM, which generates heatmaps highlighting the lung regions responsible for each prediction. These visual explanations help clinicians understand why the model classified an image as normal or opacity, thereby increasing confidence in AI-assisted diagnosis. By combining transfer learning with explainable AI, the proposed framework aims to achieve both reliable performance and improved interpretability.

The conceptual framework illustrates the overall workflow of the proposed automated lung opacity detection system. The process begins with collecting chest X-ray images from a publicly available Kaggle dataset. The dataset contains two image categories: Normal **and** Opacity. Before model training, each image is converted into RGB format, resized to 224 × 224 pixels, and divided into training, validation, and testing sets.

The preprocessed images are then passed through the pretrained VGG19 convolutional backbone, where feature extraction is performed. The pretrained convolutional layers remain frozen, while a custom binary classifier is trained to distinguish between the two classes. The classifier produces the final prediction indicating whether the input image belongs to the normal or opacity category.

To improve interpretability, Grad-CAM is applied to the last convolutional layer of the network. The generated heatmap is superimposed on the original chest X-ray image to highlight the regions that contribute most significantly to the prediction. Finally, the performance of the proposed framework is evaluated using Accuracy, Precision, Recall, F1-score, Confusion Matrix, and Training–Validation performance analysis. This framework combines automated classification with explainable visualization, making it suitable for supporting clinical decision-making.

4. Methodology

This study follows a quantitative research methodology using a publicly available secondary dataset of chest X-ray images for automated lung opacity detection. The proposed framework employs transfer learning with a pretrained VGG19 model and integrates Grad-CAM to provide visual explanations of the classification results. The overall research process consists of dataset preparation, image preprocessing, feature extraction, model training, prediction, explainability, and performance evaluation.

The dataset used in this study is a publicly available Kaggle chest X-ray dataset containing 3,943 labeled images, including 3,319 images for training and 624 images for testing. The dataset consists of two classes: Normal and Opacity, which are used to train the binary classification model.

Before model training, all chest X-ray images are converted into RGB format and resized to 224 × 224 pixels to ensure compatibility with the pretrained VGG19 architecture. Additionally, 20% of the training dataset is separated as a validation set to monitor model performance during training. This preprocessing step ensures consistency in image dimensions and improves training efficiency.

The proposed framework utilizes the pretrained VGG19 model as the feature extraction backbone. All convolutional layers of VGG19 are frozen to preserve the pretrained ImageNet feature representations, while a custom binary classifier replaces the original fully connected classifier. Since the dataset contains only two classes, the classifier is modified to perform binary classification between Normal and Opacity images. Only the newly added classification layer remains trainable during model optimization.

Model training is performed using a batch size of 32 for 30 epochs. The Adam optimizer is applied with a learning rate of 0.001, and Early Stopping is used to prevent unnecessary training after performance stabilization. Model checkpoints are saved whenever the validation loss improves, ensuring that the best-performing model is retained.

To improve interpretability, Grad-CAM is applied to the final convolutional layer of VGG19. The resulting heatmap highlights the image regions that contribute most strongly to the model's prediction, enabling visual interpretation of lung opacity detection and supporting clinical decision-making.

4.1 Variable Selection

The primary input variable in this study is the chest X-ray image obtained from the Kaggle dataset. Each image belongs to one of two classes: Normal or Opacity. Before classification, every image undergoes preprocessing through RGB conversion and resizing to 224 × 224 pixels.

The main model variable is the pretrained VGG19 architecture, which serves as the feature extractor. A custom binary classification layer is added to replace the original classifier of VGG19. During training, the convolutional layers remain frozen, while only the custom classifier is updated.

The output variable is the predicted class generated by the model, indicating whether the chest X-ray image belongs to the Normal or Opacity category. The model performance is measured using evaluation metrics such as Accuracy, Precision, Recall, F1-score, and Confusion Matrix analysis.

4.2 Cross Validation

The original thesis paper does not mention the use of k-fold cross-validation or any other cross-validation technique. Instead, the study separates 20% of the training dataset as a validation set to monitor model performance during training. This validation set is used to evaluate the model after each training epoch and to determine whether model performance improves.

In addition, Early Stopping is employed to prevent overfitting by stopping the training process when validation performance no longer improves. Model checkpoints are also saved whenever the validation loss decreases, ensuring that the best-performing model is retained for final evaluation. This validation strategy provides continuous monitoring of training performance and supports reliable model optimization.

4.3 Model Evaluation

The performance of the proposed model is evaluated using several standard classification metrics. The Confusion Matrix is used to analyze the number of correct and incorrect predictions made by the model for both Normal and Opacity classes.

The Classification Report provides Precision, Recall, F1-score, and Support for each class, allowing detailed evaluation of model performance. Training and Validation Loss curves are analyzed to determine whether the model learns effectively without overfitting. Similarly, the Training and Validation Accuracy curves are used to monitor the learning progress throughout the training process.

The overall classification performance is summarized using an accuracy of **93%**, together with detailed analysis of precision, recall, and F1-score for both classes.

4.4 Software

The original thesis paper does not explicitly specify the software environment used for model development. However, it clearly states that the pretrained VGG19 model is loaded from PyTorch and TorchVision and trained using the Adam optimizer with Early Stopping and Model Checkpoint techniques.

4.5 Ethical Issues

The study utilizes a publicly available Kaggle chest X-ray dataset for research purposes. Since the research is based on secondary data and does not involve direct interaction with patients or the collection of personal medical information, the ethical risk is minimal. The proposed explainable AI framework is intended to support healthcare professionals by providing transparent visual explanations through Grad-CAM rather than replacing clinical judgment.

5. Results

The proposed VGG19-based transfer learning model was evaluated using a publicly available chest X-ray dataset containing Normal and Opacity images. The experimental results demonstrate that the model achieved satisfactory classification performance and effectively identified pneumonia-related lung opacity. Performance evaluation was conducted using a Confusion Matrix, Classification Report, and Training–Validation performance analysis. The overall classification accuracy of the proposed model was 93%, indicating that the fine-tuned VGG19 architecture can effectively distinguish between normal and opacity chest X-ray images.

5.1 Confusion Matrix Analysis

The confusion matrix provides a detailed summary of the model's classification performance by comparing the predicted class with the actual class. The rows represent the actual labels, while the columns represent the predicted labels.

The proposed model correctly classified 169 normal chest X-ray images as normal. However, 65 normal images were incorrectly classified as opacity, representing false-positive predictions. Similarly, the model correctly identified 365 opacity images, while 25 opacity images were incorrectly classified as normal. These misclassified opacity cases represent false-negative predictions, which are considered more critical in clinical diagnosis because they may delay appropriate treatment for pneumonia patients.

The confusion matrix also shows that the diagonal values are significantly higher than the off-diagonal values, indicating that the majority of the chest X-ray images were classified correctly. Although a limited number of misclassifications occurred, the overall confusion matrix demonstrates that the proposed VGG19 model provides reliable and consistent classification performance for automated lung opacity detection.

5.2 Classification Report Analysis

The classification report evaluates the proposed model using Precision, Recall, F1-score, and Support for both classes. For the Normal class, the model achieved a precision of 0.87 and a recall of 0.72, indicating that most normal images were correctly identified, although some were incorrectly classified as opacity. The lower recall suggests that a number of healthy chest X-ray images were predicted as pneumonia cases.

For the Opacity class, the model achieved a precision of 0.85, recall of 0.94, and F1-score of 0.89. The high recall demonstrates that the proposed model successfully detected the majority of pneumonia cases while missing only a small number of patients. This characteristic is particularly important in clinical applications because failing to detect pneumonia may lead to delayed treatment and increased health risks.

The testing dataset consisted of 234 normal and 390 opacity chest X-ray images. Since the opacity class contained more samples than the normal class, the weighted average performance was influenced more by the opacity category. Overall, the classification report confirms that the proposed model performs effectively for automated lung opacity detection.

5.3 Training and Validation Performance

Training and validation performance were analyzed using loss and accuracy curves generated during model training.

5.3.1 Training and Validation Loss

The loss curve shows that the training loss decreased from approximately 0.075 during the initial epoch to around 0.020 in the later epochs. This gradual reduction indicates that the model effectively learned meaningful image features throughout the training process. Furthermore, the training loss and validation loss remained very close to each other, suggesting that the model generalized well to unseen validation data without significant overfitting. The loss curve also shows that performance improvements became minimal after approximately the tenth to twelfth training epoch, indicating that additional training would provide only limited improvement.

5.3.2 Training and Validation Accuracy

The accuracy curve illustrates continuous improvement in model performance throughout the training process. Training accuracy increased from approximately 80% during the first epoch to around 94% by the final epoch. Validation accuracy fluctuated slightly during the early training stages, which is expected while the model learns representative image features. As training progressed, the validation accuracy stabilized between 92% and 95%, indicating consistent performance on unseen validation data. The close relationship between training accuracy and validation accuracy further suggests that the proposed model does not suffer from noticeable overfitting and is capable of producing reliable predictions for chest X-ray classification.

The experimental findings indicate that the proposed transfer learning framework based on VGG19 achieved strong performance for automated lung opacity detection. The model attained an overall accuracy of 93%, demonstrated high recall for opacity cases, and maintained stable training and validation performance without significant overfitting. In addition, the confusion matrix and classification report confirmed that the model effectively classified most chest X-ray images while minimizing clinically critical false-negative predictions. These findings suggest that the proposed explainable VGG19 framework is suitable for supporting computer-aided pneumonia diagnosis in clinical environments.

6. Discussion

The experimental results demonstrate that the proposed transfer learning framework based on the pretrained VGG19 architecture is capable of accurately detecting lung opacity from chest X-ray images.

The model achieved an overall classification accuracy of 93%, indicating that transfer learning can effectively classify normal and opacity chest X-ray images using a relatively simple binary classification framework. The findings suggest that preserving the pretrained convolutional layers while training only the custom classifier successfully extracts meaningful image features without requiring complete retraining of the deep neural network.

The confusion matrix analysis shows that the majority of chest X-ray images were correctly classified into their respective categories. Although some normal images were incorrectly classified as opacity and a small number of opacity images were predicted as normal, the overall prediction performance remained satisfactory. From a clinical perspective, minimizing false-negative predictions is particularly important because an undetected pneumonia patient may not receive timely treatment. The proposed model achieved a high recall for the opacity class, indicating that it successfully identified most pneumonia cases. This characteristic increases the practical value of the proposed framework in supporting clinical diagnosis.

The classification report further confirms the effectiveness of the proposed model. The Normal class achieved higher precision, while the Opacity class demonstrated stronger recall. This result indicates that the model is more sensitive to detecting diseased patients than healthy individuals. In medical diagnosis, such behaviour is generally preferable because identifying a healthy patient as potentially infected is considered less harmful than failing to detect an actual pneumonia case. Therefore, the high recall achieved for the opacity class enhances the reliability of the proposed diagnostic system for real-world healthcare applications.

The training and validation curves demonstrate stable learning throughout the training process. Both loss curves decreased gradually and remained close to each other, indicating that the model learned representative image features without experiencing significant overfitting. Similarly, training and validation accuracy improved consistently during successive epochs before reaching stable performance. These observations suggest that the selected training strategy, including frozen convolutional layers, Adam optimizer, Early Stopping, and model checkpointing, effectively optimized the proposed model while maintaining good generalization performance.

An important contribution of this study is the integration of Grad-CAM into the pneumonia detection framework. Unlike conventional deep learning models that provide only classification outputs, Grad-CAM generates visual heatmaps highlighting the lung regions responsible for each prediction. These visual explanations improve the interpretability of the proposed model and allow clinicians to verify whether the prediction is based on clinically relevant image regions. Such explainability increases trust in AI-assisted diagnosis and supports more transparent clinical decision-making.

Despite the encouraging results, several limitations remain. The proposed framework was evaluated using a publicly available Kaggle dataset containing only two image categories. Consequently, the model was not validated using images collected from multiple hospitals or different imaging environments. In addition, the study focused exclusively on binary classification and did not investigate more complex

diagnostic tasks involving multiple pulmonary diseases. These limitations indicate opportunities for future research to improve the generalizability and practical applicability of the proposed framework.

Overall, the discussion indicates that the proposed fine-tuned VGG19 architecture provides a reliable balance between classification accuracy and explainability. The integration of transfer learning with Grad-CAM enables both effective automated detection and transparent visualization of model decisions. Therefore, the proposed framework has strong potential to serve as a supportive computer-aided diagnosis system for pneumonia detection while assisting healthcare professionals in making faster and more informed clinical decisions.

7. Conclusion

This study proposed an explainable transfer learning framework based on a fine-tuned VGG19 architecture for automated lung opacity detection using chest X-ray images. The integration of Grad-CAM improved the interpretability of the model by providing visual explanations of the classification results.

The primary objective of this research was to develop an automated and explainable pneumonia detection system using a pretrained VGG19 model. The study also aimed to improve diagnostic reliability through transfer learning and Grad-CAM-based visual interpretation of chest X-ray images.

The experimental results demonstrate that the proposed transfer learning model successfully classified chest X-ray images into Normal and Opacity categories with an overall accuracy of 93%. The classification report showed strong performance, particularly for the opacity class, which achieved a high recall rate and effectively identified most pneumonia cases. Analysis of the confusion matrix indicated that the majority of chest X-ray images were correctly classified, while the training and validation curves confirmed stable learning without significant overfitting. Furthermore, the integration of Grad-CAM enabled visualization of the lung regions responsible for each prediction, improving the transparency and interpretability of the proposed system. These findings indicate that the framework can effectively support automated pneumonia detection while providing explainable predictions for clinical use.

This study contributes to the field of computer-aided medical diagnosis by proposing an explainable transfer learning framework based on the VGG19 architecture. Unlike conventional image classification systems that provide only prediction results, the proposed framework incorporates Grad-CAM to generate visual explanations highlighting the lung regions influencing the model's decision. The research also demonstrates the effectiveness of replacing the original VGG19 classifier with a custom binary classification layer while keeping the pretrained convolutional backbone frozen. This approach improves computational efficiency and provides reliable performance for automated lung opacity detection. Overall, the study illustrates how transfer learning and explainable AI can be integrated to support transparent and accurate pneumonia diagnosis.

The findings of this study suggest that explainable artificial intelligence can play an important role in supporting healthcare services, particularly in environments where experienced radiologists are limited. An automated lung opacity detection system can assist healthcare professionals by providing rapid preliminary screening while maintaining transparency through visual explanations. Such systems may reduce diagnostic workload, improve consistency in image interpretation, and support timely clinical decision-making. However, the proposed framework should be considered as a decision-support tool rather than a replacement for professional medical judgment. Wider adoption of explainable AI technologies may contribute to improving healthcare accessibility and increasing confidence in AI-assisted diagnosis. This paragraph is an implication derived from the study's findings rather than a direct statement from the paper. It is consistent with the paper's overall conclusions but is not explicitly written in the source.

Future research may focus on improving the proposed framework through advanced image preprocessing techniques, including contrast enhancement and additional augmentation methods to improve feature learning. The authors also propose investigating reinforcement learning approaches for pneumonia detection to further enhance model performance. In addition, deploying the complete system as a web-based application would increase accessibility and allow healthcare professionals to use the model more conveniently in practical clinical environments. Future studies may also evaluate the framework using larger and more diverse datasets to improve its generalizability and reliability across different healthcare settings.

Declarations

The data used in our study are publicly available from the Kaggle repository. The dataset can be accessed at the following link: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

References

1. World Health Organization (WHO) Pneumonia Fact Sheet. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia
2. Rayasa T, Shetty DK, Shukla SRPVVK, Pullela PK (2025) AI-Driven Pneumonia Detection: Evaluating CNN Models on Pediatric Chest X-Rays, in Proc. IEEE ICCIKE, pp. 387–392
3. Beri M, Gill KS, Upadhyay D, Devliyal S (2024) AI-powered Pneumonia Detection: Enhanced Chest X-ray Interpretation with CNNs, in Proc. IEEE ICECSP, pp. 1–5
4. Singh S, Kumar M, Kumar A, Verma BK, Abhishek K, Selvarajan S (2024) Efficient Pneumonia Detection Using Vision Transformers on Chest X-rays. Sci Rep, 14, 1
5. Fu M, Tantithamthavorn C, Le T (2025) DAViT: A Domain-Adapted Vision Transformer for Automated Pneumonia Detection and Explanation Using Chest X-ray Images. IEEE Access

6. Aslan E (2024) Diagnosis of Pneumonia from Chest X-ray Images with Vision Transformer Approach. Gazi Univ J Sci Part A 11(2):324–334
7. Rafi MSM, Ali MS, Hossen MM, Sneha SZR, Arefin MS (2025) Chest X-ray Analysis for Pneumonia and Tuberculosis Detection using CNNs and Vision Transformers with Explainable AI, in Proc. IEEE COMPAS
8. Mathew L, Murugeswari R (2025) AI-Driven Pneumonia Detection: A Comparative Study of CNN-Based Architectures on Chest X-ray Images, in Proc. IEEE ICCPCT
9. Szepesi P, Szilágyi L (2022) Detection of Pneumonia Using Convolutional Neural Networks and Deep Learning. Biocybernetics Biomedical Eng 42(3):1012–1022
10. Sharma S, Guleria K (2023) A Deep Learning-Based Model for the Detection of Pneumonia from Chest X-ray Images Using VGG16 and Neural Networks. Procedia Comput Sci 218:357–366
11. Gill KS, Anand V, Chauhan R, D., GuptaUsing Deep Learning and MobileNetV2 CNN Model to Classify Chest X-ray Images for Pneumonia Disease Detectionin Proc. IEEE INCOFT, (2023)

Figures

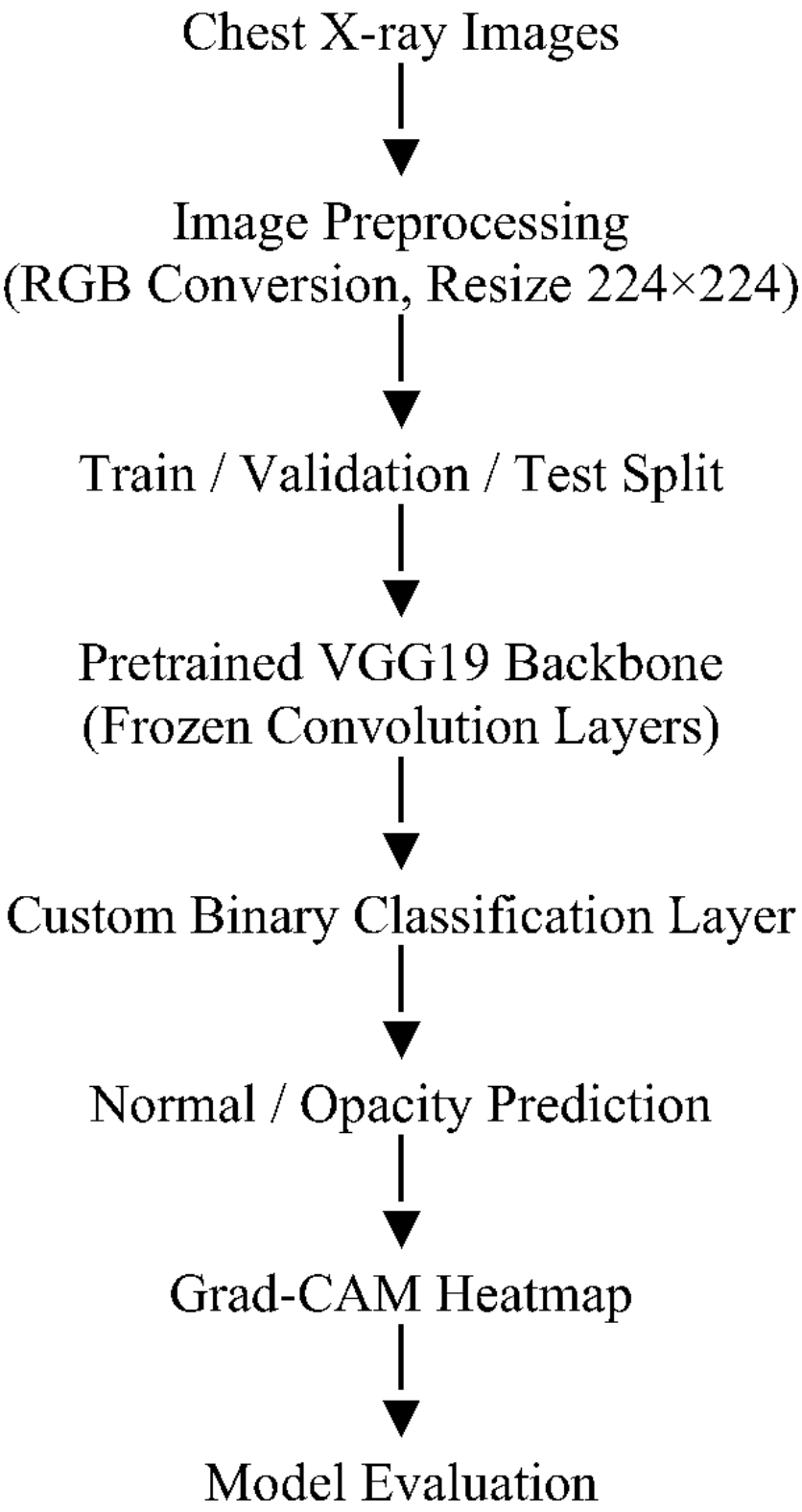


Figure 1

Figure 3.1: Conceptual framework of the proposed explainable VGG19-based lung opacity detection system

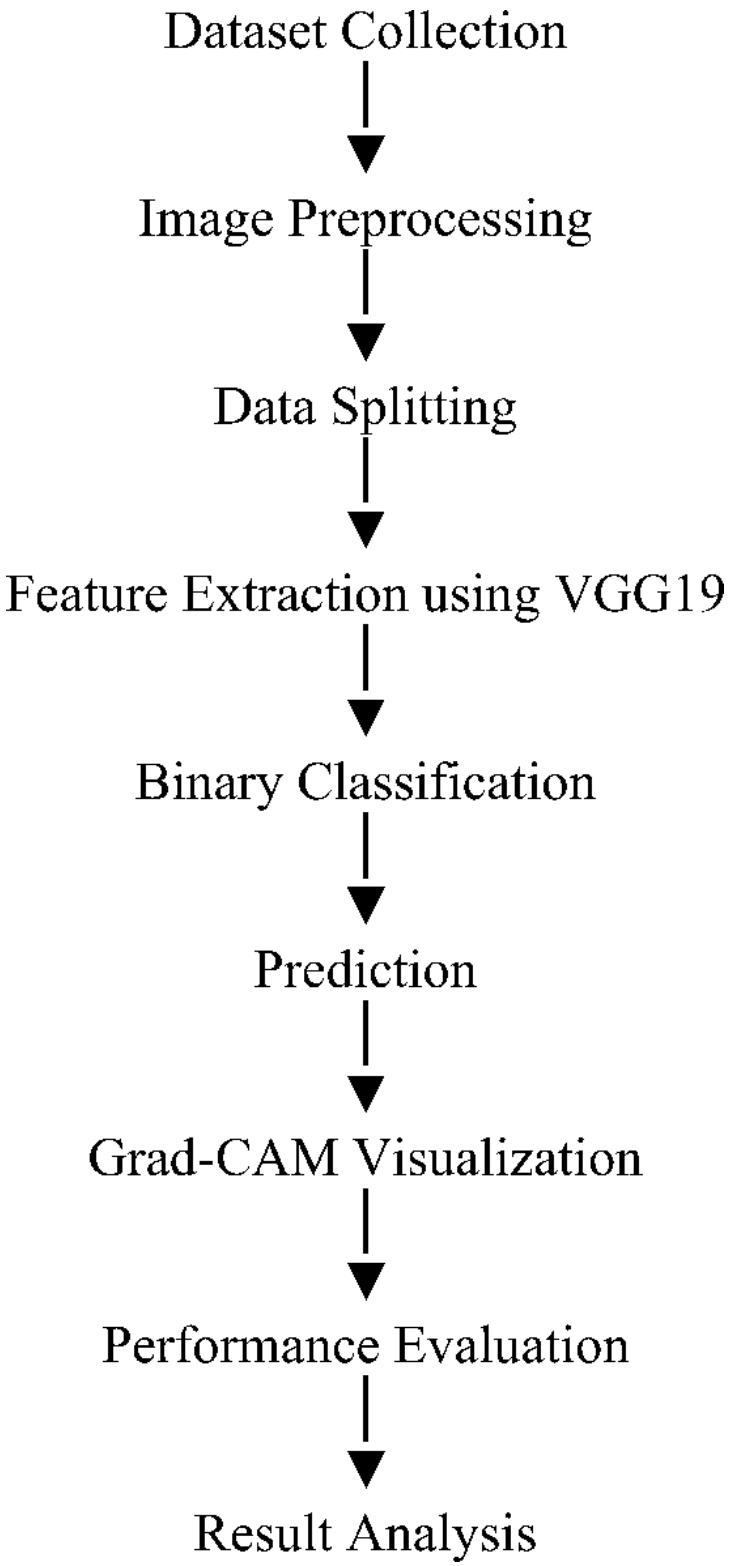


Figure 2

Figure 3.2: Integrated Research Framework

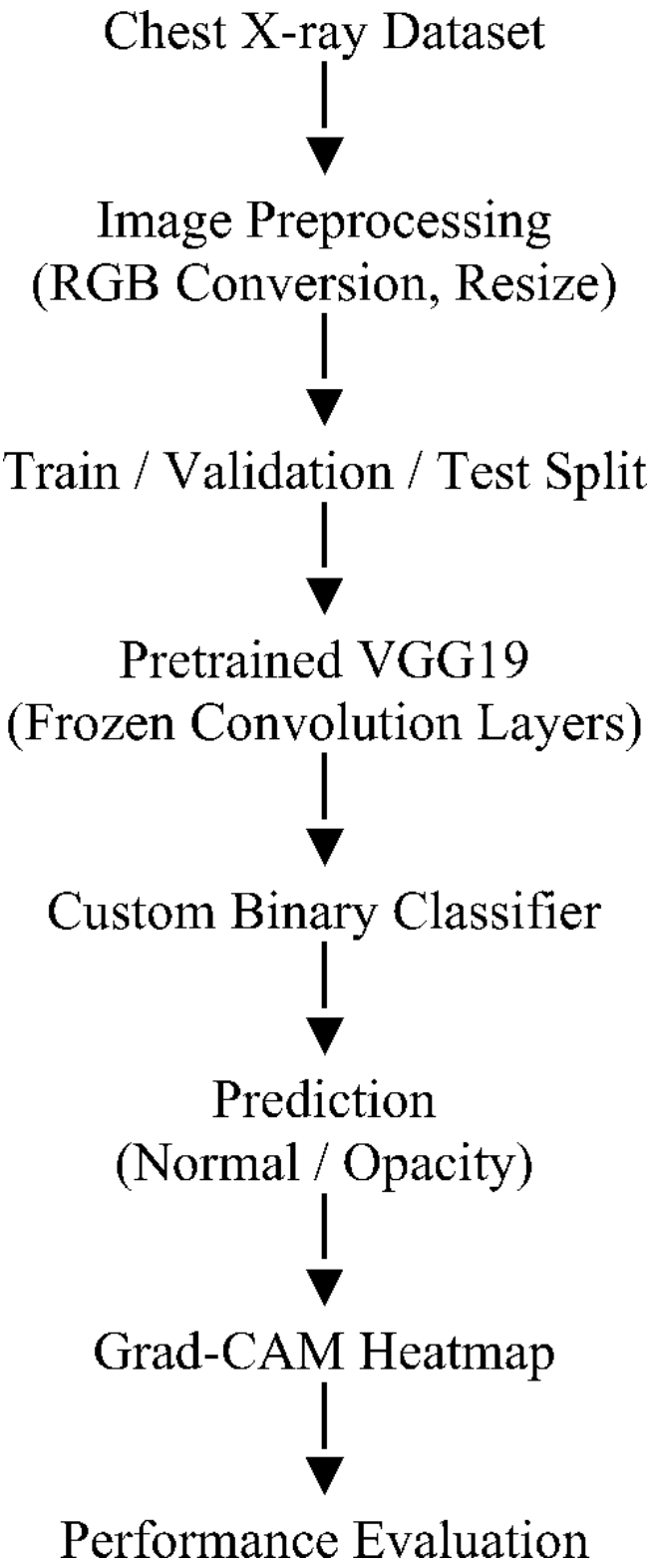


Figure 3

Figure 4.1: Research framework of the proposed VGG19-based explainable lung opacity detection system.