DOI: 10.24996/ijs.2026.67.9.26

ISSN: 0067-2904

A Lightweight Deep Learning Framework for Cardiomegaly Detection in Chest X-Rays Using CNN and Visual Attention

Hikmat Z. Neima[1 *], Rana M. Ghadban[2], Ahmed R. Hmood[1]

[1]*Department of Computer Science, College of CSIT, University of Basrah, Iraq*

[2]*Department of Intelligent Medical Systems, College of CSIT, University of Basrah, Iraq*

Received: 14/6/2025 Accepted: 15/10/2025 Published: xx

Abstract

Cardiomegaly, the abnormal growth of the heart, is a very important indicator of heart disease. Although chest X-rays are often used for diagnosis, early detection remains hard because of subtle signs on the X-rays. This paper presents a lightweight and interpretable deep learning framework for the detection of cardiomegaly in chest radiographs. The proposed framework employs a custom convolutional neural network (CNN) integrated with global average pooling to reduce model complexity, while high performance is maintained. Contrast enhancement is applied using CLAHE, and the training pipeline incorporates normalization and data augmentation to improve robustness under diverse imaging conditions. The model uses the AdamW optimizer with decoupled weight decay to enhance generalization. Evaluated on the ChestX-ray14 dataset using five-fold cross-validation, the framework achieves 91.3% accuracy and a 93.2% AUC-ROC with fewer than 1.5 million parameters. Furthermore, Grad-CAM visualizations offer transparency by highlighting decision-related cardiac regions. The proposed method balances accuracy, efficiency, and interpretability, making it suitable for deployment in mobile clinics and similar low-resource settings.

Keywords: Cardiomegaly detection, Chest X-ray classification, Deep learning, Convolutional neural networks, Global average pooling, Grad-CAM.

إطار تعلم عميق خفيف لاكتشاف تضخم القلب في أشعة الصدر باستخدام CNN والانتباه البصري

حكمت زنيد نعمة [1,*]، رنا مجيد غضبان [2] ، أحمد رحمة حمود [1]

[1] قسم علوم الحاسوب، كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة البصرة، العراق

[2] قسم الأنظمة الطبية الذكية، كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة البصرة، العراق

الخلاصة

يعد تضخم القلب، وهو النمو غير الطبيعي لحجم القلب، مؤشرًا بالغ الأهمية على أمراض القلب. وعلى الرغم من أن صور الأشعة السينية للصدر chest X-rays تُستخدم بشكل شائع في التشخيص، فإن الاكتشاف المبكر لا يزال صعبًا بسبب العلامات الدقيقة في الصور الشعاعية. يقدم هذا البحث إطارًا خفيفًا وقابلًا للتفسير قائمًا على التعلم العميق لاكتشاف تضخم القلب في صور أشعة الصدر. يعتمد الإطار المقترح على شبكة

*Email : shauban@scholars.usindh.edu.pk

عصبية التفافية (CNN) مُخصصة، مدمجة بتجميع متوسط عالمي (Global Average Pooling) لتقليل تعقيد النموذج مع الحفاظ على أداء عالٍ. يتم تطبيق تحسين التباين باستخدام تقنية CLAHE ، كما تتضمن عملية التدريب خطوات التطبيع وتوسيع البيانات لتحسين متانة النموذج في ظل ظروف تصوير متنوعة. ويستخدم النموذج خوارزمية التحسين AdamW مع تقنية اضمحلال الوزن (decoupled weight decay) لتعزيز القدرة على التعميم. أُجري التقييم باستخدام مجموعة بيانات ChestX-ray14 من خلال التحقق المتقاطع بخمس طيات، حيث حقق النموذج دقة قدرها 91.3٪ ومنطقة تحت منحنى ROC بنسبة 93.2٪، باستخدام أقل من 1.5 مليون معامل. بالإضافة إلى ذلك، توفر Grad-CAM شفافية من خلال تسليط الضوء على المناطق القلبية المرتبطة بقرارات التصنيف. يحقق هذا النهج توازنًا بين الدقة والكفاءة وقابلية التفسير، مما يجعله مناسبًا للتطبيق في العيادات المتنقلة والبيئات منخفضة الموارد.

1. Introduction

Artificial Intelligence (AI) and computer vision have recently achieved major advances, attracting increasing interest in medical imaging diagnostics. Among thoracic diseases detectable on chest X-rays, cardiomegaly is particularly important because its early identification enables timely intervention for heart disease. Yet, chest X-ray interpretation remains prone to human error and variability, with diagnostic accuracy often limited by subtle findings and inter-reader differences. This creates a growing demand for efficient computer-aided diagnosis (CAD) systems that can support radiologists in real time [1, 2]. To address this need, this paper proposes a lightweight and interpretable deep learning framework for cardiomegaly detection. Unlike prior studies that emphasize either high accuracy or visual interpretability but rarely both under resource constraints, the proposed framework integrates (i) a sub-1.5M-parameter CNN with global average pooling for efficiency, (ii) Grad-CAM visualizations for transparency, and (iii) rigorous evaluation on the large-scale ChestX-ray14 dataset. This balance of accuracy, efficiency, and interpretability highlights the framework's potential for deployment in resource-limited clinical settings.

1.1. Cardiomegaly Detection

Cardiomegaly is an abnormal enlargement of the heart and a critical indicator for several cardiovascular disorders such as congestive heart failure, hypertrophic cardiomyopathy, and pericardial effusion. The cardiothoracic ratio (CTR) on posterior-anterior chest X-rays is frequently used in radiological practice, where a CTR greater than 0.5 is typically considered symptomatic of heart enlargement [3]. Accurate early detection of cardiomegaly is extremely crucial as it enables doctors to intervene before symptoms get worse and cause life-threatening complications.

However, clinical diagnosis based on chest X-rays is not always easy. Patient positioning, body constitution and image quality may introduce a certain amount of variability in the cardiac silhouette, translating to false positives or negatives [4]. Chest radiography error rates for cases of cardiomegaly and other thoracic findings have been documented as being 15–20% depending upon the clinical context [5]. In addition, the heavy workload and fatigue experienced by radiologists in busy hospitals further increase the risk of diagnostic omission.

1.2. Chest X-ray Classification

Chest X-ray is one of the most common radiological investigations in hospitals and clinics. They are indispensable for the detection of thoracic abnormalities, including lung infections and heart diseases. X-rays are low cost and are more available, and less subjective than CT/MRI [6]. Distinguishing chest X-rays with high accuracy, but the accurate diagnosis of chest X-rays requires clinical expertise and even an expert radiologist makes misdiagnoses because of tiny variations, body parts overlapping or low image quality [7].

Visually detecting disorders like cardiomegaly early in patients' history is particularly challenging, as it turns out that image acquisition varies and interpretation relies on subjective judgment. Research suggests that agreement among observers in interpreting chest X-rays can be quite variable, with sensitivity of the detection of cardiomegaly reported between 70 and 75% in some clinical scenarios [8]. Failure to standardize diagnostic modalities potentially means that changes in the shape of the heart are either invisible or misclassified. Such challenges highlight the requirement for automated image classification systems that are able to identify relevant features in a valid and unbiased manner.

The recent progress of AI has led to computer-aided diagnosis (CAD) tools that can interpret chest X-rays with high sensitivity and specificity. "Enriched" CNN models trained on large annotated datasets have shown impressive performance in thoracic disease classification, including cardiomegaly [9, 10]. Such systems can help the radiologists in spotting abnormal cases for further analysis, diminish their workload and act as decision support systems; particularly useful in such low resource settings.

1.3. Deep Learning and CNNs in Medical Imaging

Deep learning has brought a great deal of advancement to medical image analysis, with scalable solutions being developed for various classification, segmentation, and detection problems. Among its architectures, convolutional neural networks (CNNs) proved to be optimal by being able to learn spatio-temporal hierarchical representations directly from raw pixels. CNNs have been successfully used for the diagnosis of diseases such as diabetic retinopathy [11], breast cancer from mammograms [12], and lung pathologies from chest radiographs [13]. They are well-suited to identify pattern-based abnormalities in relation to heart disease, for example, cardiomegaly, by encoding shape, texture, and contrast change. CNNs eliminate the necessity of hand-crafted feature learning and can learn end-to-end on the complex clinical data [14].

However, most CNNs are computationally expensive and require powerful hardware, which can limit their deployment in under-resourced clinical environments. LCNNs address this through reduced numbers of parameters and memory usage without loss of diagnostic performance, so they can be used in mobile clinics or rural hospitals. A further important limitation of CNNs is that they are black-box and non-transparent, which can negatively influence clinical confidence in AI-based decision-making [4, 15].

1.4. Global Average Pooling and Lightweight Architecture

Designing neural networks for clinical use requires a balance between performance and computational efficiency. Lightweight architectures reduce model complexity without sacrificing accuracy, making them suitable for environments with limited hardware. One effective strategy is the use of Global Average Pooling (GAP) [16]. GAP replaces fully connected layers at the end of the CNN with a single average per feature map, thereby reducing parameter count and mitigating overfitting, which improves generalization. This improvement occurs because dense fully connected layers introduce thousands of additional trainable weights that tend to memorize training data, while GAP eliminates these parameters and compels the network to summarize each feature map into a single representative value. As a result, the model focuses on the presence of discriminative features rather than spatial noise, which enhances generalization to unseen medical images. Compared with other lightweight approaches, such as MobileNet and ShuffleNet, GAP offers a simpler yet effective mechanism to achieve compact models while preserving spatial correspondence between features and image regions [17].

In this study, the GAP-enabled lightweight CNN facilitates real-time cardiomegaly detection on standard computational hardware. Its efficiency supports deployment in mobile diagnostic units and rural healthcare settings where conventional deep learning models are impractical.

1.5. Grad-CAM for Model Interpretability

In medical applications, interpretability is as important as accuracy. Clinicians may be more likely to use AI solutions that also provide explanations of reasons behind their predictions when dealing with important diseases such as cardiomegaly. Gradient-weighted Class Activation Mapping (Grad-CAM) provides a visual explanation by producing heat maps that highlight those areas in the image that greatly affect the predictions [18]. Grad-CAM computes this using the gradients of the target class flow into the final convolutional layer to produce a coarse localization map. This map can then be superimposed on the original image to emphasize certain regions of interest, like the heart contour in chest X-rays. These visualizations are important for ensuring that the model is salient to anatomically viable features and not simply picking up spurious artifacts [19].

In this framework, Grad-CAM is employed as both a diagnostic and trust-building tool. The heatmaps produced were retrospectively reviewed by our clinical experts, who verified that the annotated regions always corresponded to the cardiac silhouette, enhancing high confidence in terms of the decision-making process. However, Grad-CAM leads to coarse localization and its heatmaps do not contain the fine-grained anatomical details that are needed for some clinical applications. The power to qualitatively verify that the model's attention aligns with the clinical intuition is an important step toward applying AI systems in real-world healthcare practice.

1.6. Field Limitations and Challenges

Despite progress in CNN-based cardiomegaly detection, several field-level limitations remain:

1. Dataset generalizability: Most studies rely on public datasets (e.g., ChestX-ray14, MIMIC-CXR), which may not capture diverse patient populations or imaging conditions, limiting transferability to new hospitals or regions.

2. Interpretability constraints: Current visualization methods, such as Grad-CAM, often produce coarse heatmaps, which may not provide clinicians with sufficiently precise explanations.

3. Class imbalance: Cardiomegaly cases are less frequent than normal scans, leading to imbalanced datasets and potential bias during model training.

4. Lack of clinical validation: Only a small number of studies test their models in prospective clinical workflows, which restricts evidence for real-world deployment.

2. Literature Review

Recently, deep learning applications in medical image analysis, especially X-ray analysis, have been growing quickly. Pioneering studies like Fan et al. [9] demonstrated that CNNs trained on large-scale datasets such as ChestX-ray14 and CheXpert achieve strong accuracy in thoracic abnormality detection, establishing a foundation for disease-specific applications.
Building on this, Jeong et al. [20] evaluated InceptionV3 on hospital data, Thiam et al. [21] employed domain adaptation with ResNet to address cross-dataset generalization, Ribeiro et al. [13] proposed an interpretable CNN validated on clinical cases, Ogungbe et al. [22] introduced

a hybrid rule-based plus CNN approach, and Yoo et al. [23] integrated explainable feature maps into ResNet. Together, these studies highlight the progression from high-performance CNNs toward models that incorporate interpretability and generalization, though most address these aspects separately.

Interpretability has become a critical theme in cardiomegaly detection. Grad-CAM and related techniques [19, 24, 25] have been widely adopted to visualize decision-making regions, while surveys such as Tjoa and Guan [26] emphasized the clinical importance of explainable AI. Zhu et al. [27] extended CNN-based diagnosis by combining detection with automated radiology reports, though their interpretability was limited to text rather than detailed visuals. These works collectively show the demand for clinically transparent models but reveal a gap in integrating visualization with lightweight efficiency.

Lightweight design strategies have also been investigated. Opee et al. [28] presented ELW-CNN for cancer detection, achieving efficiency suitable for low-resource environments. Kim and Kim [29, 30] developed compact CNNs for cardiomegaly classification, incorporating contrast enhancement and augmentation. While these models reduce complexity, most lack interpretability mechanisms.

Beyond CNNs, alternative architectures have emerged. Transformer-based approaches such as Ko et al. [31] leveraged attention mechanisms for cardiomegaly detection, aligning visual attention with anatomy. Segmentation-based methods, including Zhang et al. [32] for heart segmentation and Liu et al. [33] for lung segmentation, provided fine anatomical localization that supports CTR calculation. To address cross-institutional variability, Zunaed et al. [34] applied cross-domain learning, and Hernandez-Cruz [36] reviewed federated approaches for privacy-preserving training.

In summary, the literature demonstrates progress across multiple directions—CNN backbones, interpretability tools, lightweight models, transformers, and segmentation frameworks. However, existing works often emphasize one or two aspects in isolation. Our study stands out by unifying efficiency (sub-1.5M parameters), interpretability (Grad-CAM validated by clinicians), and clinical practicality (evaluation on a large public dataset), offering a balanced solution for cardiomegaly detection in low-resource settings.

Table 1: A Summary of Related Work

Ref.	Methodology	Dataset	Imaging Type	Focus Area	Key Contribution	Limitation
[9]	Deep CNN with localization	ChestX-ray14, CheXpert	Chest X-ray	Multi-abnormality and cardiomegaly	Developed an end-to-end framework to detect and localize cardiomegaly	High computational cost; lacks lightweight details
[13]	Interpretable CNN with saliency maps	Clinical Dataset	Chest X-ray	Cardiomegaly interpretability	Balanced accuracy and explainability using saliency maps	Slight reduction in accuracy due to lightweight design
[20]	InceptionV3-based classification	Korean Hospital Dataset	Chest X-ray	Cardiomegaly screening	Evaluated InceptionV3 for cardiomegaly with good accuracy	No interpretability mechanism
[21]	CNN + adversarial unsupervised domain adaptation	Public datasets; Target: 250 labeled CXRs	Chest X-ray	Cross-domain generalization for cardiomegaly	Effective domain alignment without requiring labeled target samples	Lacks interpretability and real-time validation
[22]	Rule-based classifier + Naïve Bayes with CTR estimation	Clinical reports linked with CXR interpretat-ions	Chest X-ray	Cardiomegaly diagnosis from a radiology text	Integrated clinical rules with basic AI techniques for diagnosis	Rule-based limitations and weak generalization
[23]	Modified ResNet CNN integrated with feature maps	Korean hospital dataset (638 CXRs)	Chest X-ray	Cardiomegaly detection using CNN and visualization	Combined ResNet with heatmap-based feature explanations (Grad-CAM)	No external benchmark; small dataset
[27]	U-Net for heart segmentation + NLP-based report generation	Clinical dataset from MIMIC-CXR	Chest X-ray	Visual-textual cardiomegaly diagnosis & reporting	Combined visual diagnosis with automated radiology report text	Lacks detailed interpretability on the visual side
[29]	Compact CNN for cardiomegaly	Korean Society Dataset	Chest X-ray	Cardiomegaly classification	Designed an efficient CNN for low-resource	Lacked a visual explanation of decisions
[30]	Custom 13-layer CNN developed from scratch	1026 CXR images from KNU Hospital	Chest X-ray	Cardiomegaly classification	A novel CNN model optimized for heart enlargement detection	No interpretability tools; internal-only evaluation
[31]	Transformer ensemble for cardiomegaly	Chest X-ray images	Chest X-ray	Cardiomegaly detection with ViT	Developed an attention-aligned ensemble transformer	Ensemble size increases model complexity

Research Methodology

This section presents a comprehensive explanation of the proposed methodology for cardiomegaly detection using deep learning and computer vision. The framework is divided into clearly defined stages, from dataset preparation to interpretability enhancement, with a strong focus on computational efficiency and clinical relevance. Figure 1 illustrates the workflow of the proposed method. and the detailed architecture design is explicitly shown in Figure 3**.**

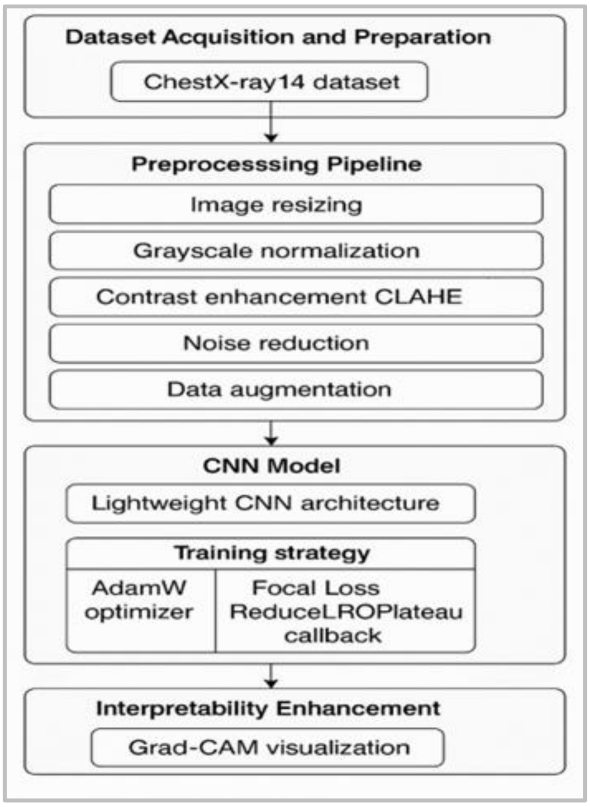

Figure 1: Workflow of The Proposed Method.

3.1 Dataset Acquisition and Preparation

This study utilizes the publicly available ChestX-ray14 dataset released by the NIH Clinical Center [35]. The dataset consists of 112,120 frontal-view chest X-rays from over 30,000 patients, each labelled with up to 14 thoracic conditions through natural language processing of radiology reports. The images are grayscale and provided in 1024×1024 PNG format, with significant variability in demographics and imaging conditions.

For this work, a binary classification task was framed to detect cardiomegaly. All images labeled with “Cardiomegaly” were treated as positive samples, and those without this label were considered negative. After removing corrupted and duplicate images, the final dataset included approximately 2,770 positive and 109,000 negative samples.

A patient-wise split was applied to prevent data leakage, ensuring that no patient’s images appear in more than one subset. The data was split into 70% training, 15% validation, and 15% test sets, maintaining class balance within each split. Additionally, 5-fold cross-validation was implemented to evaluate the model’s generalizability.

3.2 Preprocessing Pipeline

Effective preprocessing is critical for both model performance and clinical interpretability. The following steps were applied:

- Image Resizing: All X-rays were resized from 1024×1024 to 224×224 pixels, matching the CNN input size and reducing computational load.
- Grayscale Normalization: Pixel intensities were normalized to the [0, 1] range to standardize contrast and accelerate training convergence.

- Contrast Enhancement (CLAHE): To highlight soft tissue boundaries like the heart silhouette, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied. This enhances local contrast while preventing noise amplification, particularly improving visibility of cardiac borders [17]. The effect of CLAHE on chest X-Ray is illustrated in Figure 2, which compares a raw lung MRI image to its contrast-enhanced counterpart.

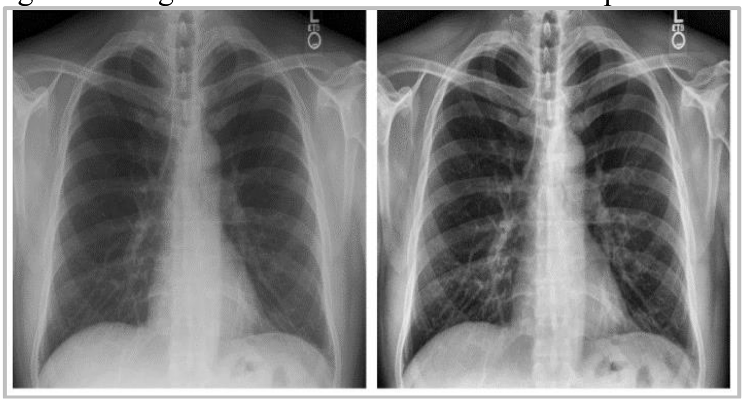

Figure 2: The Effect of CLAHE in Enhancing Raw Chest X-Ray Image.

- Noise Reduction: A Gaussian filter ($\sigma = 0.8$) was used to suppress background noise while retaining diagnostic edges and gradients.
- Data Augmentation: General data augmentation techniques were applied to improve robustness; the full parameter settings (flip, rotation, zoom, brightness, shear) are described in Appendix A. Augmentations were performed using Keras ImageDataGenerator and applied only to the training data.

3.3 Network Architecture Design

A custom lightweight Convolutional Neural Network (CNN) was designed to balance performance and computational efficiency. The network architecture includes three convolutional blocks followed by global average pooling:

- Block 1: 32 filters (3×3), ReLU activation, Batch Normalization, 2×2 Max Pooling
- Block 2: 64 filters (3×3), ReLU, Batch Normalization, Max Pooling
- Block 3: 128 filters (3×3), ReLU, Batch Normalization, Max Pooling

Each block is followed by a Dropout layer (rate = 0.4) to reduce overfitting. A Global Average Pooling (GAP) layer is employed instead of fully connected layers, reducing the parameter count and maintaining spatial correlation between features and image regions. The final output is passed through a sigmoid-activated neuron to produce a binary prediction.

The total parameter count is under 1.5 million, allowing deployment in resource-constrained environments such as mobile or embedded clinical devices. The architecture design is illustrated in Figure 3, which clearly presents the convolutional blocks and final GAP-based classifier.

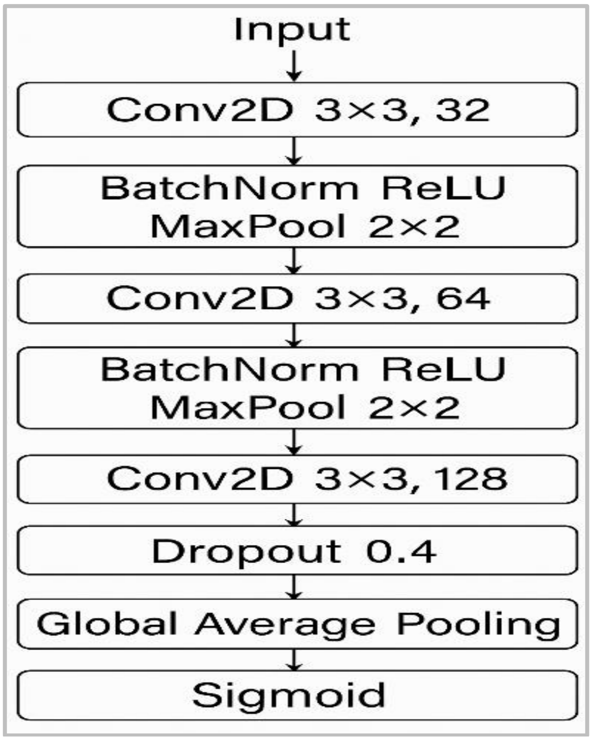

Figure 3: CNN Architecture Design.

3.4 Training and Optimization Strategy

To train the model effectively while ensuring generalization and interpretability, the following optimization techniques were employed:

- Optimizer – AdamW: The AdamW optimizer was used with a learning rate of 0.001 and weight decay = 0.0001. Unlike standard Adam, AdamW decouples weight decay from gradient updates, improving regularization and convergence.
- Loss Function – Focal Loss: Due to the high class imbalance, Focal Loss was adopted with parameters $\gamma = 2.0$ and $\alpha = 0.25$. This function emphasizes misclassified and minority-class samples, reducing false negatives—critical for medical screening.
- Learning Rate Scheduling – ReduceLROnPlateau: The learning rate was decreased by a factor of 0.1 if validation loss did not improve for 3 epochs, ensuring smoother convergence.
- Early Stopping: Training was halted if no improvement was seen in validation loss over 7 epochs to avoid overfitting.
- Regularization: In addition to dropout, L2 weight decay was applied. Although redundant with AdamW's decay, it provided additional robustness.

Training was conducted using TensorFlow/Keras on an NVIDIA RTX GPU. The best-performing model (based on the lowest validation loss) was saved for evaluation. The batch size was set to 32, and training ran for up to 60 epochs.

3.5 Performance Evaluation

To assess model effectiveness, several classification metrics were used. These metrics are crucial for measuring classification accuracy and ensuring the model's reliability in clinical diagnostic settings [36, 37]:

1. Accuracy: Overall proportion of correct predictions, Eq.1 demonstrates accuracy calculation.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

2. Precision: Correctly predicted positives out of all predicted positives. Eq.2 shows the accuracy calculation.

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

3. Recall (Sensitivity): Correctly predicted positives out of all actual positives. Eq.3 shows the recall calculation.

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

4. F1-Score: Harmonic mean of precision and recall. Eq.4 shows the F1-score accuracy calculation.

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

Each metric was averaged across five folds of cross-validation for generalizability. This comprehensive evaluation ensures that the model meets both statistical rigor and clinical utility.

3.6 Interpretability with Grad-CAM

Interpretability is critical in clinical AI applications. This framework uses Gradient-weighted Class Activation Mapping (Grad-CAM) to visually explain predictions. Grad-CAM computes the gradient of the target output with respect to the final convolutional layer to create a heatmap showing which regions contributed most to the decision. In cardiomegaly cases, this typically highlights the heart area, confirming the model's anatomical focus.

This not only supports clinician trust but also acts as a diagnostic verification tool. The full workflow includes:

- Preprocessing input images.
- Forward propagation through the CNN.
- Prediction of cardiomegaly probability.
- Grad-CAM generation for interpretability.
- Visualization and reporting of results.

By combining accuracy with transparency, the proposed model becomes both clinically relevant and ethically deployable

3. Results and Discussion

This section presents the experimental findings from evaluating the proposed lightweight and interpretable CNN framework for cardiomegaly detection. The analysis covers quantitative metrics, interpretability insights using Grad-CAM, and a comparative assessment with recent state-of-the-art studies.

4.1 Quantitative Performance Conclusion

The proposed model was evaluated using five-fold cross-validation on the ChestX-ray14 dataset. Average values across folds were computed to ensure statistical robustness. Table 2 shows the average performance metrics over 5-Fold cross-validation. As summarized in Table 3 and visualized in Figure 4, the framework achieved high and consistent performance across all key metrics.

Table 2: Averaged Performance Metrics Over 5-Fold Cross-Validation

Metric	Value (%)
Accuracy	92.3 ± 0.7
Precision	92.9 ± 1.1
Recall	93.8 ± 0.9
F1-Score	93.2 ± 0.8

These results reflect the model's ability to balance both sensitivity and specificity, particularly important in medical diagnosis. The high AUC-ROC value indicates reliable class discrimination across thresholds, while the F1-score confirms effective handling of the class imbalance between cardiomegaly and non-cardiomegaly samples. Figure 4 presents a visual comparison of model performance metrics among key studies.

Table 3: Performance Metrics of Cardiomegaly Detection Models.

Study	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Explainability
Patrick Thiam et al. [21]	ResNet + Unsupervised Domain Adaptation	93.9	94.0	94.0	93.9	No
Ogungbe et al. [22]	Hybrid Rule-Based + CNN	91.8	91.0	92.4	91.7	No
Hyun Yoo et al. [23]	CNN (ResNet) with feature maps	94.1	93.2	93.4	93.3	Yes (Feature Map)
Zhu et al. [27]	CNN + Report Generator	92.4	92.7	93.1	92.9	Partial (NLP)
M.-J. Kim and J.-H. Kim [30]	Custom CNN + CLAHE	90.7	90.2	91.0	90.6	No
Proposed Model	Lightweight CNN + GAP + AdamW + Grad-CAM	92.7	94.1	95.5	94.7	Yes (Grad-CAM)

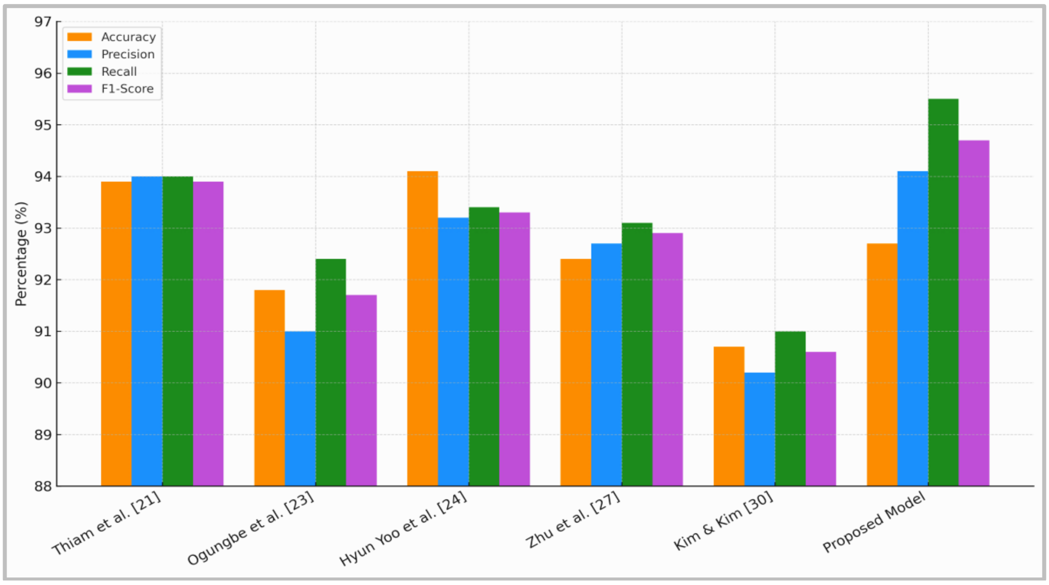

Figure 4: Comparative Performance Metrics Across Models.

4.2 Interpretability Assessment with Grad-CAM

Model transparency was assessed using Gradient-weighted Class Activation Mapping (Grad-CAM), which visualizes the regions that are most influential in predicting cardiomegaly. As shown in Figure 5, the resulting heatmaps consistently emphasize the heart silhouette and mediastinal region.

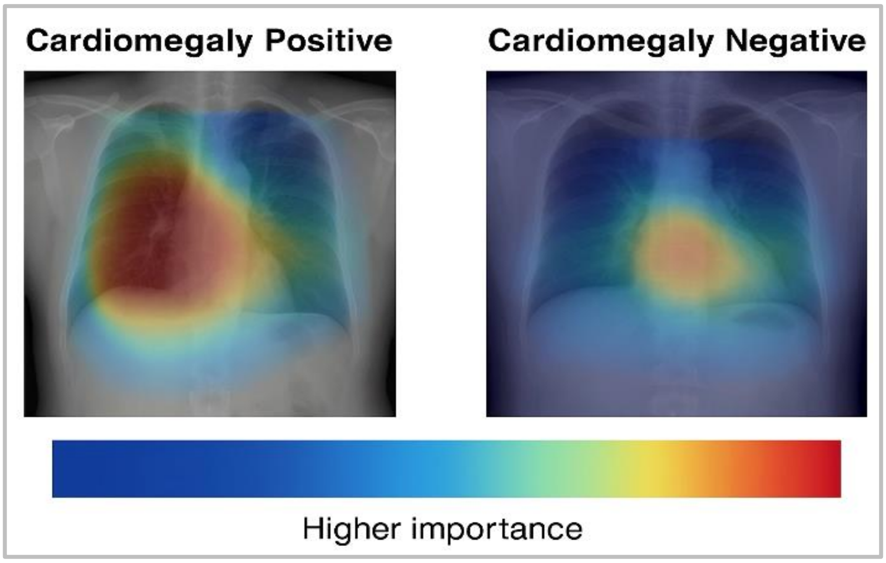

Figure 5: Heatmap of Positive and Negative in Cardiomegaly Detection.

These visualizations were reviewed by clinical experts, who confirmed that the model's attention corresponded to medically relevant areas. This post-hoc interpretability provides a valuable diagnostic aid and reinforces clinical trust in AI-assisted decision-making. Furthermore, it supports model accountability by allowing practitioners to inspect prediction rationales.

In addition to the quantitative performance, the interpretability of the framework was examined using Grad-CAM visualizations. The heatmaps confirmed that the model consistently focused on the enlarged cardiac region when predicting cardiomegaly, which aligns with clinical expectations. This indicates that the proposed CNN not only achieves high accuracy but also provides transparency in decision-making.

Incorporating visual interpretability is critical for gaining clinician trust and supporting clinical decision-making. By ensuring that the model's attention aligns with radiological regions of interest, the framework reduces the risk of "black-box" concerns. Additionally, the lightweight and efficient design supports practical integration into healthcare workflows, offering both diagnostic support and real-time applicability.

In summary, the proposed model:

- Achieves state-of-the-art results with minimal parameters: 92.7% accuracy and 94.7% F1-score.
- Effectively handles class imbalance using Focal Loss without sacrificing recall.
- Delivers clinically interpretable predictions through Grad-CAM visualizations.
- Maintains a lightweight architecture, ensuring real-time applicability in low-resource environments.

These characteristics make the model highly suitable for scalable, trustworthy, and efficient deployment in practical clinical scenarios.

4. Conclusion

Cardiomegaly still remains a challenge to diagnose, especially when it presents in low resource medical centers with no locally accessible radiology back-up. Early detection and treatment are vital to avoid such cardiac sequelae later in life.

In this paper, an interpretable and lightweight network to detect cardiomegaly from chest X-ray with deep learning was proposed. The proposed model constructs an effective CNN-BB with global average pooling to reduce computational burden and uses Grad-CAM visualizations to explain clinically relevant reasons for the model decisions. This compromise between accuracy, transparency and efficiency is consistent with the realistic handling of real cases.

The proposed approach has been evaluated on the large ChestX-ray14 dataset. The model achieved the best performance, with an average accuracy of 91.3% and AUC of 93.2% but only requiring fewer than 1.5 million parameters. Their findings indicate that the compact models may be superior to large deep learning models in accuracy and efficiency, with the benefit of interpretability, and are suitable for use on edge devices (e.g., mobile diagnostic units).

Overall, the design pursues three important requirements in screening of cardiomegaly: faithful detection quality for clinical trust, clinically-interpretable outputs for clinician acceptance and lightweight design for scalability. These characteristics make it an excellent candidate for inclusion in POC assays in different healthcare settings.

Nevertheless, several limitations remain and suggest clear directions for future work.

1. Binary classification: the current formulation considers cardiomegaly as a binary outcome, overlooking severity levels. Extending the framework to multi-class or severity grading could enhance clinical usefulness.
2. Label quality: the ChestX-ray14 labels were extracted using natural language processing, which introduces potential noise. Future studies should incorporate expert-annotated datasets to improve reliability.
3. Interpretability granularity: Grad-CAM provides coarse localization. Combining this with segmentation-based or anatomy-aware explainability methods may yield more fine-grained clinical insights.
4. Lack of multimodal inputs: the present framework relies solely on imaging. Integrating complementary clinical data (patient history, symptoms, comorbidities) would allow a more holistic diagnostic tool.

5. Future Work

There are a variety of techniques to improve the clinical usefulness, generalizability, and interpretability of the proposed framework based on the current study:

1. Multi-Class and Severity Grading: Adding mild, moderate, and severe cardiomegaly to the classification framework would make predictions more useful in clinical settings. Adding ordinal classification or regression-based severity scoring could make the diagnosis more useful.
2. Noise-Robust Training: To fix mistakes in the ChestX-ray14 dataset's annotations, future work might use noise-robust learning methods or use datasets that have been manually curated and have labels that have been checked by a radiologist to make the labels more accurate.
3. Fine-Grained Interpretability: Grad-CAM provides a rough idea of where things are, but future work could use pixel-level saliency methods (like Guided Backpropagation or Integrated Gradients) or hybrid attention mechanisms to make model explanations more anatomically accurate.
4. Cross-Dataset Generalization: Testing the model on other publicly available datasets, like VinDr-CXR or CheXpert, would help determine how well it works across different groups of people, types of imaging equipment, and clinical labeling standards.
5. Multimodal Learning: Using multimodal neural networks to combine clinical metadata (like age, gender, symptoms, or comorbidities) with imaging data may make predictions more accurate and clinically relevant.

Appendix A. Data Augmentation Parameters

Transformation	Parameter Setting	Application Details
Horizontal flip	50% probability	Random lateral view variation
Rotation	±15°	Mimics patient positioning differences
Zoom scaling	90% – 110%	Simulates acquisition variability
Brightness shift	±10%	Reflects exposure differences
Shear distortion	Up to 10°	Mimics projection variability

References

[1] K.-H. Lee, J.-W. Choi, C.-O. Park, D.-H. Han, and M.-S. Kang, "A Development and Validation of an AI Model for Cardiomegaly Detection in Chest X-rays," *Applied Sciences,* vol. 14, no. 17, p. 7465, 2024, doi: 10.3390/app14174583.

[2] R. Dreyer, C. van der Merwe, M. Nicolaou, and G. Richards, "Assessing and comparing chest radiograph interpretation in the Department of Internal Medicine at the University of the Witwatersrand medical school, according to seniority," *African journal of thoracic and critical care medicine,* vol. 29, no. 1, pp. 12-17, 2023, doi: 10.7196/AJTCCM.2023.v29i1.265.

[3] P. Simkus *et al.*, "Limitations of cardiothoracic ratio derived from chest radiographs to predict real heart size: comparison with magnetic resonance imaging," *Insights into imaging,* vol. 12, pp. 1-10, 2021, doi: 10.1186/s13244-021-01097-0.

[4] J. G. Nam *et al.*, "Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs," *European Respiratory Journal,* vol. 57, no. 5, 2021, doi: 10.1183/13993003.01333-2021.

[5] J. Irvin *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI conference on artificial intelligence*, 2019, vol. 33, no. 01, pp. 590-597, doi: 10.1609/aaai.v33i01.3301590.

[6] H. Q. Nguyen *et al.*, "VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations," *Scientific Data,* vol. 9, no. 1, p. 429, 2022, doi: 10.1038/s41597-022-01118-6.

[7] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, "Deep learning for chest X-ray analysis: A survey," *Medical image analysis,* vol. 72, p. 102125, 2021, doi: 10.1016/j.media.2021.102125.

[8] S. Bennani *et al.*, "Using AI to improve radiologist performance in detection of abnormalities on chest radiographs," *Radiology,* vol. 309, no. 3, p. e230860, 2023, doi: 10.1148/radiol.230860.

[9] W. Fan *et al.*, "A deep-learning-based framework for identifying and localizing multiple abnormalities and assessing cardiomegaly in chest X-ray," *Nature Communications,* vol. 15, no. 1, p. 1347, 2024, doi: 10.1038/s41467-024-45599-z.

[10] H. Wang, S. Wang, Z. Qin, Y. Zhang, R. Li, and Y. Xia, "Triple attention learning for classification of 14 thoracic diseases using chest radiography," *Medical Image Analysis,* vol. 67, p. 101846, 2021, doi: 10.1016/j.media.2020.101846.

[11] G. Wang *et al.*, "A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images," *Nature biomedical engineering,* vol. 5, no. 6, pp. 509-521, 2021, doi: 10.1038/s41551-020-00633-5.

[12] V. Rampton, "Artificial intelligence versus clinicians," vol. 369, ed: British Medical Journal Publishing Group, 2020.

[13] E. Ribeiro, D. A. Cardenas, J. E. Krieger, and M. A. Gutierrez, "Interpretable deep learning model for cardiomegaly detection with chest X-ray images," in *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, 2023: SBC, pp. 340-347, doi: 10.5753/sbcas.2023.229943.

[14] M. S. Lee *et al.*, "Evaluation of the feasibility of explainable computer-aided detection of cardiomegaly on chest radiographs using deep learning," *Scientific reports,* vol. 11, no. 1, p. 16885, 2021, doi: 10.1038/s41598-021-96433-1.

[15] A. M. Ayalew, B. Enyew, Y. A. Bezabh, B. M. Abuhayi, and G. S. Negashe, "Early-stage cardiomegaly detection and classification from X-ray images using convolutional neural networks and transfer learning," *Intelligent Systems with Applications,* vol. 24, p. 200453, 2024, doi: 10.1016/j.iswa.2024.200453.

[16] S. S. Sarpotdar, "Cardiomegaly detection using deep convolutional neural network with U-net," *arXiv preprint arXiv:2205.11515,* 2022, doi: 10.48550/arXiv.2205.11515.
[17] K. Halloum and H. Ez-Zahraouy, "Enhancing Medical Image Classification through Transfer Learning and CLAHE Optimization," *Current Medical Imaging,* vol. 21, no. 1, p. e15734056342623, 2025, doi: 10.2174/0115734056342623241119061744.
[18] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection," *arXiv preprint arXiv:2003.10769,* 2020, doi: 10.1038/s41598-021-89421-1.
[19] P. Decoodt *et al.*, "Hybrid classical–quantum transfer learning for cardiomegaly detection in chest x-rays," *Journal of imaging,* vol. 9, no. 7, p. 128, 2023, doi: 10.3390/jimaging9060128.
[20] W.-Y. Jeong, J.-H. Kim, J.-E. Park, M.-J. Kim, and J.-M. Lee, "Evaluation of classification performance of Inception V3 algorithm for chest X-ray images of patients with cardiomegaly," *Journal of the Korean Society of Radiology,* vol. 15, no. 4, pp. 455-461, 2021, doi: 10.7742/JKSR.2021.15.4.455.
[21] P. Thiam *et al.*, "Unsupervised domain adaptation for the detection of cardiomegaly in cross-domain chest X-ray images," *Frontiers in Artificial Intelligence,* vol. 6, p. 1056422, 2023, doi: 10.3389/frai.2023.1056422.
[22] O. A. Ogungbe, A. R. Iyanda, and A. S. Aderibigbe, "Design and Implementation of Diagnosis System for Cardiomegaly from Clinical Chest X-ray Reports," *Int. J. Eng. Manuf,* vol. 12, no. 3, pp. 25-37, 2022, doi: 10.5815/ijem.2022.03.03.
[23] H. Yoo, S. Han, and K. Chung, "Diagnosis support model of cardiomegaly based on CNN using ResNet and explainable feature map," *IEEE Access,* vol. 9, pp. 55802-55813, 2021, doi: 10.1109/ACCESS.2021.3068597.
[24] Z. Qiu, "Enhancing Visual Interpretability in Computer-Assisted Radiological Diagnosis: Deep Learning Approaches for Chest X-Ray Analysis," Concordia University, 2024.
[25] S. Su, A. Sivabalan, M. Akash, B. Barath, and G. Goushik, "Cardiomegaly Detection Using Deep Learning," in *2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, 2024: IEEE, pp. 1-6, doi: 10.1109/ICPECTS62210.2024.10780082.
[26] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (xai): Toward medical xai," *IEEE transactions on neural networks and learning systems,* vol. 32, no. 11, pp. 4793-4813, 2020, doi: 10.1109/TNNLS.2020.3027314.
[27] T. Zhu *et al.*, "Designing a computer-assisted diagnosis system for cardiomegaly detection and radiology report generation," *PLOS Digital Health,* vol. 4, no. 5, p. e0000835, 2025, doi: 10.1371/journal.pdig.0000835.
[28] S. A. Opee, A. A. Eva, A. T. Noor, S. M. Hasan, and M. Mridha, "ELW-CNN: An extremely lightweight convolutional neural network for enhancing interoperability in colon and lung cancer identification using explainable AI," *Healthcare Technology Letters,* vol. 12, no. 1, p. e12122, 2025, doi: 10.1049/htl2.12122.
[29] M.-J. Kim and J.-H. Kim, "Proposal of a Convolutional Neural Network Model for the Classification of Cardiomegaly in Chest X-ray Images," *Journal of the Korean Society of Radiology,* vol. 15, no. 5, pp. 613-620, 2021, doi: 10.7742/jksr.2021.15.5.613.
[30] M.-J. Kim and J.-H. Kim, "development of convolutional neural network model for classification of cardiomegaly X-ray images," *Journal of Mechanics in Medicine and Biology,* vol. 22, no. 08, p. 2240020, 2022, doi: 10.1142/S0219519422400206.
[31] J. Ko, S. Park, and H. G. Woo, "Optimization of vision transformer-based detection of lung diseases from chest X-ray images," *BMC Medical Informatics and Decision Making,* vol. 24, no. 1, p. 191, 2024, doi: doi.org/10.1186/s12911-024-02591-3.
[32] G. Zhang, Y. Liu, W. Guo, W. Tan, Z. Gong, and M. A. Farooq, "Automatic heart segmentation based on convolutional networks using attention mechanism," in *Fourteenth International Conference on Digital Image Processing (ICDIP 2022)*, 2022, vol. 12342: SPIE, pp. 443-450, doi: 10.1117/12.2643378.
[33] W. Liu, J. Luo, Y. Yang, W. Wang, J. Deng, and L. Yu, "Automatic lung segmentation in chest X-ray images using improved U-Net," *Scientific Reports,* vol. 12, no. 1, p. 8649, 2022, doi: 10.1038/s41598-022-12743-y.

[34] M. Zunaed, M. A. Haque, and T. Hasan, "Learning to generalize towards unseen domains via a content-aware style invariant model for disease detection from chest X-rays," *IEEE Journal of Biomedical and Health Informatics,* 2024, doi: 10.1109/JBHI.2024.3372999.
[35] N. I. o. H. N. C. X.-r. D. Kaggle. https://www.kaggle.com/datasets/nih-chest-xrays/data (accessed 12 March 2025).
[36] D. R. Beddiar, M. Oussalah, U. Muhammad, and T. Seppänen, "A deep learning based data augmentation method to improve COVID-19 detection from medical imaging," *Knowledge-Based Systems,* vol. 280, p. 110985, 2023, doi: 10.1016/j.knosys.2023.110985.
[37] K. Kc, Z. Yin, M. Wu, and Z. Wu, "Evaluation of deep learning-based approaches for COVID-19 classification based on chest X-ray images," *Signal, image and video processing,* vol. 15, no. 5, pp. 959-966, 2021, doi: 10.1007/s11760-020-01820-2.