

Article

Advanced Clinical Analysis Based on Automated Cardiothoracic Ratio Estimation Using Attention-Gated Deep Learning

Hye-Won Jung ^{1,*} , Sang-Heon Lee ^{2,*}, Hyung-Jeong Yang ³  and Roger Davies ¹¹ AdelaideMRI, Adelaide, SA 5070, Australia; rdavies@adelaidemri.com² Mawson Lakes Campus, Adelaide University, Adelaide, SA 5095, Australia³ Smart Computing Laboratory, Chonnam National University, Gwangju 61186, Republic of Korea; hjyang@jnu.ac.kr

* Correspondence: hye-won.jung@adelaidemri.com (H.-W.J.); sang-heon.lee@adelaide.edu.au (S.-H.L.)

Abstract

Chest X-rays (CXR) are essential for diagnosing various chest conditions, including pneumonia, congestive heart failure, and cardiomegaly. Cardiomegaly is clinically defined by a Cardiothoracic Ratio (CTR) exceeding 0.5. Traditional CTR assessment relies on manual measurement, which increases radiologists' workload and susceptibility to inter-observer variability. This study proposes a segmentation framework integrating the Efficient and Lightweight U-Net (ELU-Net) with the Convolutional Block Attention Module (CBAM). Performance was benchmarked against standard U-Net and several advanced variants with attention mechanisms. To assess clinical translatability, the proposed model was validated on an external, independent dataset of 1240 images from a private clinic (AdelaideMRI), comparing automated CTR values with expert manual assessment. The proposed method achieved superior segmentation performance, recording a DSC of 95.04%, IOU of 92.96%, and F1-score of 96.39%, outperforming all comparative architectures. In the external clinical validation, the method demonstrated a strong correlation with expert measurements (Pearson coefficient = 0.853) and a low Mean Absolute Error (MAE) of 0.0070. The system correctly classified 97% of cardiomegaly cases (CTR > 0.5). The proposed method significantly enhances the segmentation accuracy of cardiothoracic structures, demonstrating high reliability and generalizability when applied to diverse, real-world clinical data.

Keywords: attention mechanism; cardiomegaly; cardiothoracic ratio (CTR); clinical evaluation; deep learning

1. Introduction

Chest X-rays (CXR) are a fundamental diagnostic tool extensively used to diagnose numerous conditions involving the lungs, heart, and great vessels [1–3]. While frequently identifying conditions like pneumonia and congestive heart failure, CXR often serves as a preliminary screening tool rather than a definitive diagnostic method. A critical application of CXR is assessing cardiomegaly, defined as a heart width exceeding 50% of the internal thoracic diameter, quantified by the Cardiothoracic Ratio (CTR). Cardiomegaly can be caused by various underlying conditions such as coronary artery disease, kidney disease, hypertension, inherited disorders, infections, and cardiomyopathy [2,4,5]. Studies show that CTR predicts cardiomegaly with high accuracy, making it a valuable diagnostic tool [6–8]. In clinical practice, manual CTR measurement is labour-intensive and susceptible to inter-observer variability.



Academic Editor: Mara Pistellato

Received: 28 May 2026

Revised: 29 June 2026

Accepted: 9 July 2026

Published: 13 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Recent advancements of deep learning in medical imaging have significantly improved disease detection and image segmentation [4,9–12]. Image segmentation involves isolating and describing specific areas of interest within a medical image, such as organs or tumors. This process is essential for accurately diagnosing various abnormalities or diseases. In the context of cardiomegaly, deep learning methods have been employed to automatically determine the CTR from segmented images of the lung and heart. These automated processes depend heavily on the performance of image segmentation algorithms. UNet has achieved remarkable results in medical image segmentation tasks [9,13–16]. UNet’s encoder–decoder architecture is especially effective in segmentation tasks, as it incorporates skip connections that concatenate the encoder features with corresponding decoder features, providing a more robust learning process [2–4,17–20]. However, UNet faces challenges such as lower accuracy in regions with complex or ambiguous boundaries, as well as issues with learning speed and overfitting [2,9,21]. To mitigate these issues, several expanded UNet architectures have been introduced to improve segmentation performance across different applications [10,22–27]. Their modifications enhance feature extraction and multi-scale representation, ultimately improving overall segmentation performance. On the other hand, attention mechanisms have been integrated to assign adaptive weights to feature maps, allowing models to prioritize salient regions while suppressing background artifacts [21,28–34]. While mechanisms such as Squeeze-and-Excitation (SE) [35] and Efficient Channel Attention (ECA) [36] focus effectively on channel-wise feature interaction, they often neglect spatial dependencies. Conversely, the Convolutional Block Attention Module (CBAM) [37] uniquely integrates both spatial and channel attention, enabling a more comprehensive feature refinement process that is theoretically superior for delineating anatomical boundaries in CXR. Other variations, such as Global Context (GC) [38] networks and Gated Channel Transformation (GCT) [39], further attempt to generalize these interactions, though their specific efficacy in cardiothoracic segmentation remains variable.

Traditional U-Net variants often suffer from semantic gaps between encoder and decoder features, leading to blurred boundary definitions in complex structures like the cardiac apex. While architectures like Attention U-Net utilize grid-based gating to mitigate this, they scale up computational overhead without fully capturing cross-channel interactions. Furthermore, channel-only attention mechanisms (e.g., SE and ECA blocks) frequently fail to preserve spatial topography, leading to suboptimal segmentation of the thin pleural and mediastinal interfaces critical for accurate CTR computation. Conversely, spatial-only alternatives lack the capacity to suppress channel-wise redundant noise inherent in high-frequency clinical X-ray artifacts. Therefore, addressing these localized boundary and noise limitations requires an integrated approach that simultaneously refines both spatial and channel-wise feature dependencies. Despite the emergence of sophisticated architectures and attention modules, their most effective synergistic combination for segmentation has not been fully established. Consequently, this study identifies the most effective segmentation framework for automated CTR measurement by systematically evaluating combinations of advanced U-Net variants and attention mechanisms. We propose an ELU-Net model augmented with CBAM, which demonstrated superior performance in our preliminary analysis. Unlike studies relying solely on standardized training sets, this research bridges the gap to clinical applicability by validating the proposed model on a large-scale, heterogeneous clinical population while detailing the practical inference principles required for low-latency integration into real-world hospital workflows.

2. Materials and Methods

CTR is calculated using Posteroanterior (PA) or Anteroposterior (AP) chest X-rays to visualize the cardiac silhouette; this study utilizes both to accurately delineate cardiac and thoracic outlines. For objective comparison, we utilized 704 publicly available chest X-ray images: 566 from the Shenzhen dataset and 138 from the Montgomery County dataset [40]. The training datasets include both normal and abnormal cases featuring manifestations of tuberculosis. As the public datasets provided incomplete masks, we manually generated comprehensive heart and lung masks using the VGG Image Annotator (VIA) tool by medical experts [41]. While increasing the amount of training data generally enhances performance, manual segmentation tasks are expensive and time-consuming. Achieving high accuracy with minimal training images is therefore crucial to reduce cost and effort. Consistent with Sogancioglu et al. who established robust segmentation with 500 images [4], this study utilizes 704 images, divided with 80% allocated for training and validation, and the remaining 20% reserved for testing the model. To account for natural variations in body size, habitus, and age across cohorts, our pipeline relies on uniform image resizing and global contrast normalization; specifically, all images were resized to 512×512 pixels, followed by the application of histogram equalization to normalize brightness and contrast. To ensure objective model comparison, transfer learning and augmentation (e.g., cropping, rotation) were deliberately excluded. For external validation, an independent clinical dataset was utilized; these images were strictly excluded from the training process.

This research systematically investigates a range of combinations involving several advanced UNet models and attention mechanisms to identify a highly optimized, computationally efficient framework for automated clinical screening. While modern vision transformers and generative diffusion models offer powerful segmentation capabilities, they often incur prohibitive computational overhead, high data requirements, and elevated hardware costs that limit their utility in real-time, high-throughput clinical pipelines. Therefore, to evaluate solutions that maximize boundary localization accuracy on modest medical datasets, our initial experiments focused on seven established, highly representative convolutional UNet variants (UNet [2–4,17], UNet3+ [10,22], Double UNet (D-UNet) [23] and UNet++ [24], Inf-Net [25], DTA-UNet [26] and ELU-Net [27]). This optimized architectural choice ensures that the framework can be seamlessly integrated into standard clinical workstations without requiring specialized, cost-prohibitive hardware upgrades. Afterwards, the secondary experiments were conducted by incorporating four different attention gates (ECA [36], CBAM [37], GCT [39], and GC [38]) to enhance performance.

Our experiments revealed that ELU-Net combined with the CBAM attention gate achieved the highest accuracy (Figure 1). ELU-Net enhances the efficiency and lightweight nature of the UNet by utilizing deep skip connections that fully leverage full-scale features from the encoder. These skip connections incorporate both same and large-scale feature maps, resulting in an effective architecture for segmentation tasks. The decoder integrates several resolution feature maps from both the encoder and decoder, seamlessly combining shallow, detailed information with deeper semantic data to improve segmentation accuracy. This integration allows the model to capture both fine-grained details and coarse-grained semantics across all scales, making it highly adaptable to various segmentation challenges. The model further incorporates an attention gate into ELU-Net prior to concatenation at the end of each layer (X^1, X^2, X^3, X^4), enabling the decoder to prioritize key information and suppress irrelevant details.

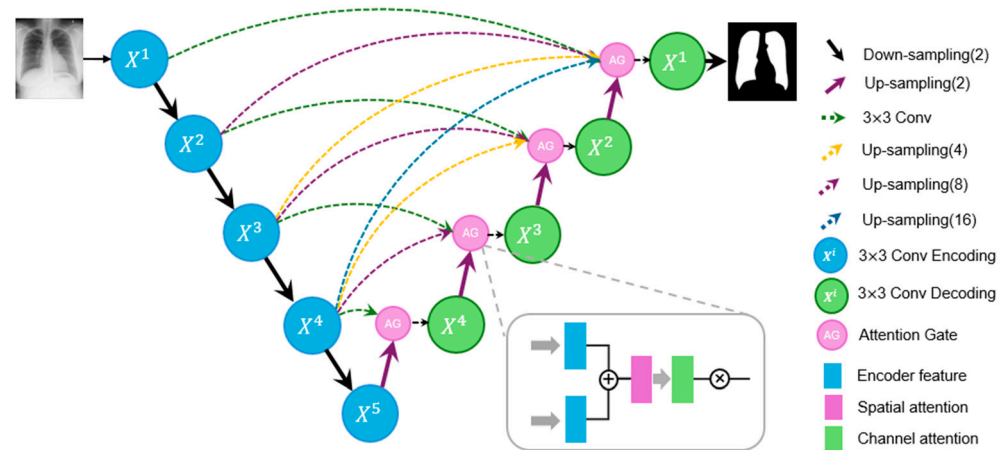


Figure 1. Illustration of the architecture of the proposed method with attention gate.

Initial segmentation results of the lung and heart were suboptimal due to the presence of several small-sized artifacts. To address this issue, morphological operation was applied to fill any holes in the resulting masks and remove small-sized noise. The two lung regions and one heart region were identified based on the size and location of segmentations, with all other smaller regions treated as artifacts. The postprocessing method ensures accurate calculation of the CTR by measuring the width of both the lungs and heart. To assess the accuracy of the segmentation models, this study employs three semantic segmentation metrics: Dice Similarity Coefficient (DSC), Intersection Over Union (IOU) and F1-score. The DSC is calculated as twice the area of overlap between the predicted and the ground truth regions divided by the total number of pixels in both regions, as shown in Equation (1). This score measures true positives while penalizing false positives, making it more similar to precision than accuracy.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}, IOU = \frac{|X \cap Y|}{|X \cup Y|}, F1 - score = \frac{precision \times recall}{precision + recall} \quad (1)$$

Here, $\cap$ and $\cup$ denote the intersection and the union operators, while X and Y represent the predicted segmented region and the ground truth, respectively. The IOU measures the overlap area between the predicted bounding box and the ground truth bounding box. The F1-score is a measure of predictive performance which is calculated from the precision and recall of the test.

CTR is defined as the ratio between the transverse diameter of the cardiac silhouette and the distance of the internal rib margins or the maximal diameter of the lung field. Alternatively, CTR is calculated from CXR images using two measurements: HD (the Heart Diameter) and ID (the Internal Diameter of the chest) as shown in Figure 2. HD was calculated as the maximum horizontal width of the heart mask and ID was determined by identifying the maximum horizontal width of the combined lung masks. The equation for calculating CTR is presented as Equation (2).

$$CTR = HD / ID \quad (2)$$

In typical diagnostic practice, a normal CTR measurement should be less than 0.5, with a CTR between 0.5 and 0.55 indicating mild cardiomegaly, and a CTR greater than 0.55 suggesting cardiomegaly [3]. Since mild cardiomegaly is also noted in radiologist reports and is regarded as a form of cardiomegaly, this study adopts a ratio of 0.5 as the cutoff value for evaluating cardiomegaly. To ensure accurate ground truth for CTR, an experienced medical expert independently measured both HD and ID.

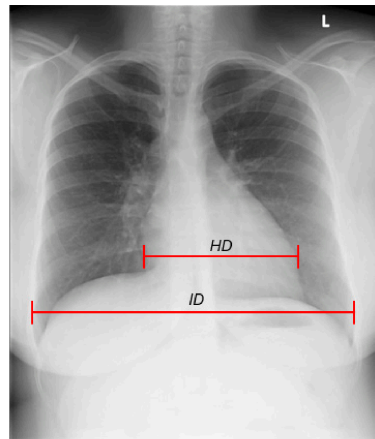


Figure 2. Measurement of Cardiothoracic ratio (CTR) in the chest radiograph. CTR is calculated using two measurements: HD (the Heart Diameter) and ID (the Internal Diameter of the chest).

This observed CTR outcome was used as the reference standard for comparison. To evaluate CTRs between predictions and the ground truth measurements, the Mean Absolute Error (MAE) was used to evaluate the performance:

$$MAE = \frac{1}{N} \sum_{i=1}^N |e_i|, \quad (3)$$

where N denotes the number of images, and e_i represents the difference between the predicted CTR and the observed CTR measurements. In addition, the Pearson correlation coefficient was calculated to quantify the linear relationship between measurements.

3. Results

3.1. Experimental Outcomes

The overall performance of different combinations of models was determined using the cross-validation method. Three-fold cross-validation ensured fair comparison across the training dataset. The experiments were conducted using the PyTorch(2.12.0) framework, with the model applying the Adam optimizer at a learning rate of 0.001, a batch size of 2 and the soft dice loss function. Each fold was trained for 100 epochs because this duration allowed the model to learn effectively without overfitting. This study compared the standard UNet against the six expanded UNet variants. As described in Table 1, the results were obtained based on utilizing the test dataset. Results show improved accuracy for most expanded UNet models (UNet++, D-UNet, Inf-Net, ELU-Net) compared to standard UNet. Among the tested models, the ELU-Net model achieved the highest average accuracy across all metrics, with a DSC of 94.82%, IOU of 90.47%, and F1-score of 96.21%. Based on these findings, ELU-Net was selected as the model for further research, particularly for incorporating attention gates to further enhance segmentation performance.

To identify the synergistic combination of ELU-Net with an attention gate, four different attention gates (ECA, CBAM, GC, and GCT) were applied. The experimental results obtained using the test dataset are presented in Table 2. These results revealed that incorporating ELU-Net with the CBAM achieved the highest average performance metrics, with DSC of 95.04%, IOU of 90.73% and F1-score of 96.39%.

Figure 3 presents an example of heart and lung segmentation results obtained from expanded UNet models. In images with clear boundaries and high-contrast images, segmentation results achieved high accuracy, demonstrating the robustness of the models under favorable conditions. However, for images with abnormal lung area or unclear

boundaries, some models struggled to produce accurate segmentations. Figure 3 includes challenging cases, where the inferior border of the heart shows significant segmentation inaccuracies, particularly in the fourth and fifth rows. The proposed method yielded the most consistent and smooth segmentation results relative to the ground truth. This highlights better capability in handling complex cases with unclear boundaries.

Table 1. Experiment results of expanded UNet models.

	Lung			Heart			AVG		
	DSC	IOU	F1 Score	DSC	IOU	F1 Score	DSC	IOU	F1 Score
UNet	96.12%	92.65%	97.40%	93.24%	87.83%	94.77%	94.68%	90.24%	96.09%
UNet3+	96.18%	92.73%	97.46%	92.99%	87.45%	94.55%	94.58%	90.09%	96.00%
UNet++	96.23%	92.83%	97.50%	93.37%	87.99%	94.87%	94.80%	90.41%	96.18%
D-UNet	96.15%	92.70%	97.42%	93.24%	87.90%	94.75%	94.69%	90.30%	96.09%
Inf-Net	96.24%	92.85%	97.51%	93.22%	87.86%	94.79%	94.73%	90.36%	96.15%
DTA-UNet	95.30%	91.15%	96.65%	90.61%	83.39%	91.97%	92.96%	87.27%	94.31%
ELU-Net	96.16%	92.68%	97.44%	93.49%	88.25%	94.99%	94.82%	90.47%	96.21%

Table 2. Experiment results of ELU-Net with attention gates.

	Lung			Heart			AVG		
	DSC	IOU	F1 Score	DSC	IOU	F1 Score	DSC	IOU	F1 Score
ELU-Net with ECA	96.14%	92.66%	97.40%	93.78%	88.48%	95.20%	94.96%	90.57%	96.30%
ELU-Net with CBAM	96.22%	92.81%	97.47%	93.88%	88.64%	95.30%	95.05%	90.73%	96.39%
ELU-Net with GC	96.06%	92.52%	97.34%	93.51%	88.05%	94.95%	94.79%	90.29%	96.14%
ELU-Net with GCT	96.18%	92.74%	97.46%	93.17%	87.53%	94.66%	94.67%	90.14%	96.06%

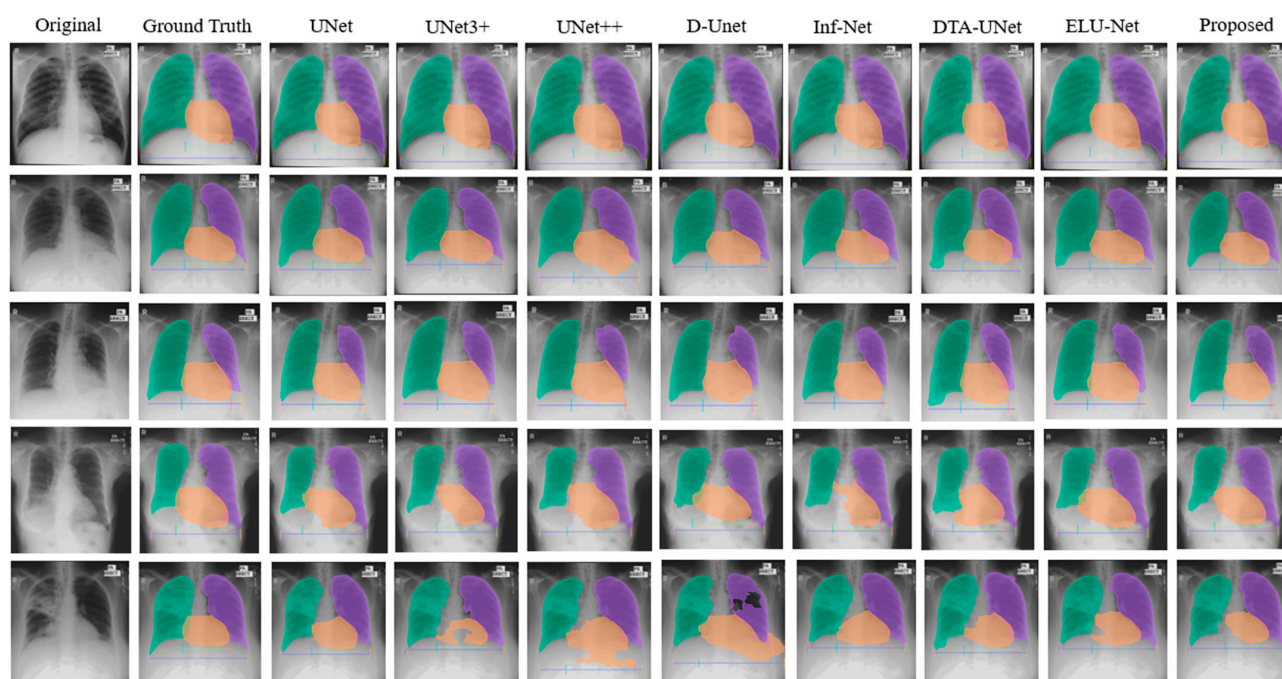


Figure 3. Qualitative comparison of segmentation results across U-Net variants. The heart is segmented in orange, the right lung in green, and the left lung in purple. The proposed model demonstrates superior boundary delineation, particularly in challenging cases with unclear inferior heart borders of the fourth and fifth rows.

3.2. CTR Evaluation Study

To validate the proposed method's suitability for clinical applications, it compares the predicted CTR measurements with the observed references obtained from manual measurements. For this CTR analysis, 1230 X-ray images were collected from October 2022 to July 2023 at an independent medical clinic, AdelaideMRI. This data collection period was randomly selected, and all patient information was depersonalized to ensure privacy. Table 3 presents the demographic distribution of the dataset, revealing a normal age distribution with the mean age 53.1 ± 20.3 . Additionally, the dataset contains a slightly larger proportion of female samples compared to male samples.

Table 3. Patient demographic data of CTR evaluation.

Age Group	0–4	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99	Total
N	4	7	16	44	37	68	79	83	97	93	98	98	112	112	81	67	56	54	18	5	1230 M: 575, F: 654
% Total	0.3	0.6	1.3	3.6	3	5.5	6.4	6.8	7.9	7.6	8	8	9.1	9.1	6.6	5.5	4.6	4.4	1.5	0.4	100 M: 46.8, F: 53.2

The predicted CTR values were automatically calculated based on segmentations from the proposed method. Figure 4 illustrates the comparative analysis, with Figure 4a showing scatter plots of CTR values between the predicted segmentations and the observed measurements.

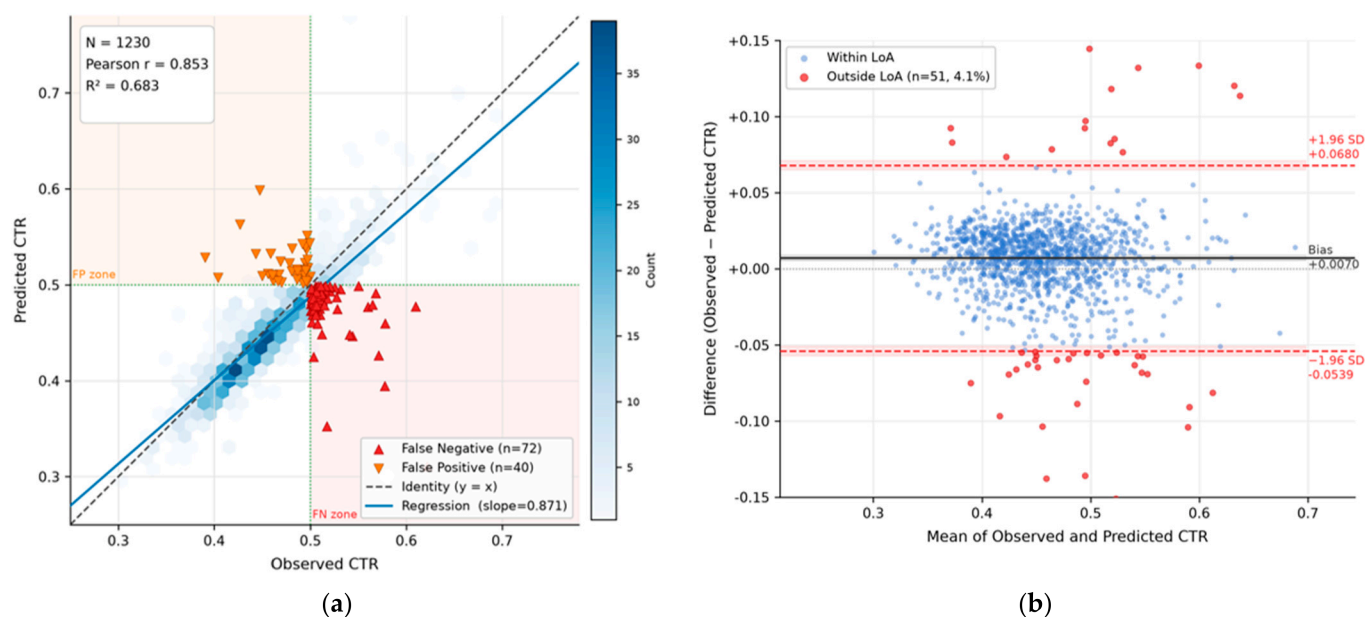


Figure 4. Clinical CTR validation between predictions and observed references. (a) Scatter plot of predicted and observed CTR. True Positive ($n = 180$) and True Negative ($n = 928$) was classified as concordant cardiomegaly. In contrast, False Negative ($n = 72$) and False Positive ($n = 40$) indicate discordant cardiomegaly classifications. (b) Bland–Altman plot of prediction errors. A total of 4.1% of data points ($n = 51$) fall outside the limits of agreement (LoA).

Table 4 presents the performance evaluation metrics for the CTR model evaluated on the AdelaideMRI dataset, demonstrating strong overall predictive accuracy and diagnostic reliability. The model achieved a high Pearson correlation coefficient ($r = 0.8527$, $p < 0.001$), indicating a statistically significant, robust linear relationship between the predicted and actual values, which is further supported by an R-square of 0.6831 (explaining approximately 68.3% of the variance). Furthermore, classification performance remains highly robust, yielding an overall accuracy of 90.2%. While the model demonstrates a solid sensitivity

of 71.7% in correctly identifying positive instances, it exhibits exceptional specificity at 95.9%, highlighting its strong capacity for minimizing false positives and ensuring highly dependable negative predictions. Some outliers are also present, potentially caused by unexpected artifacts in the images, suggesting that further refinement could improve this value—a potential target for future research.

Table 4. CTR evaluation metrics.

Metric	Value
Pearson r	0.8527 ($p < 0.001$)
R-square	0.6831
Sensitivity	71.7%
Specificity	95.9%
Accuracy	90.2%

Figure 4b illustrates a Bland–Altman plot comparing observed CTR values against those predicted, where each point represents the difference plotted against the pair's mean. The solid line indicated the mean bias, the dashed red lines mark the 95% limits of agreement (± 1.96 SD), and points outside these limits are highlighted in red to visualize systematic error and spread.

Figure 5 presents boxplots comparing the CTR values predicted by the proposed method to the observed references across 5-year age groups. Figure 5a displays the distribution of predicted CTR values per age group as box plots overlaid with a mean trend line (95% CI) and a dashed cardiomegaly threshold at CTR = 0.50, revealing how predicted CTR varies with patient age. Figure 5b shows the corresponding prediction error per age group against a zero-error reference line, evaluating whether the model exhibits any systematic age-dependent bias in its CTR estimates. The two graphs demonstrate the model's consistency and accuracy across the full age spectrum, with overall performance summarized by the CTR evaluation metrics.

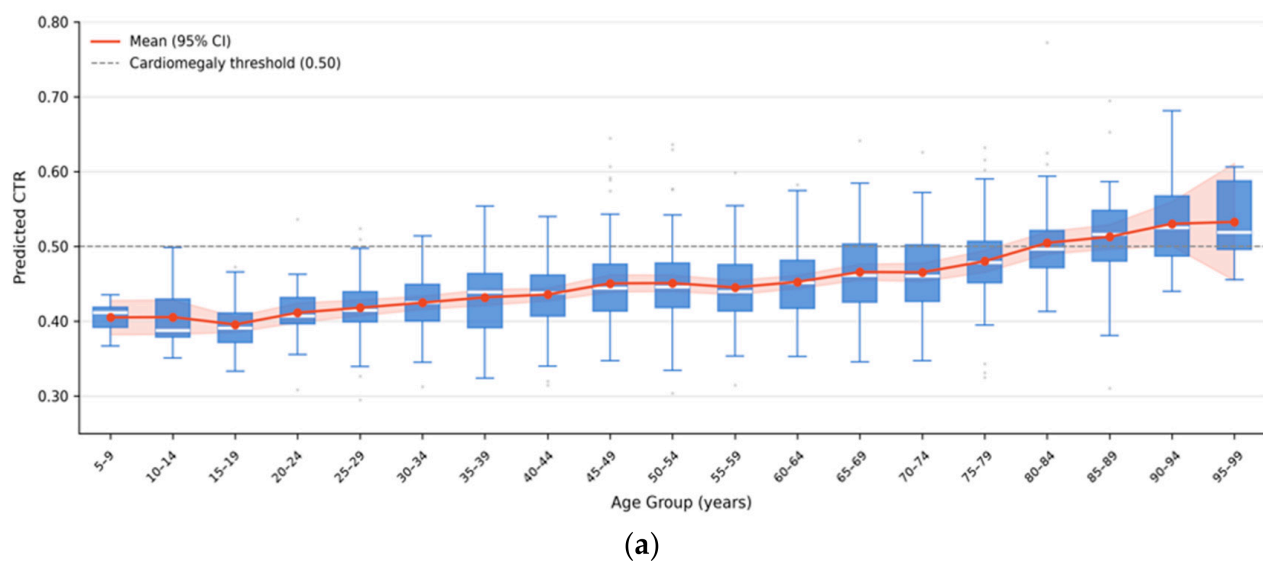


Figure 5. Cont.

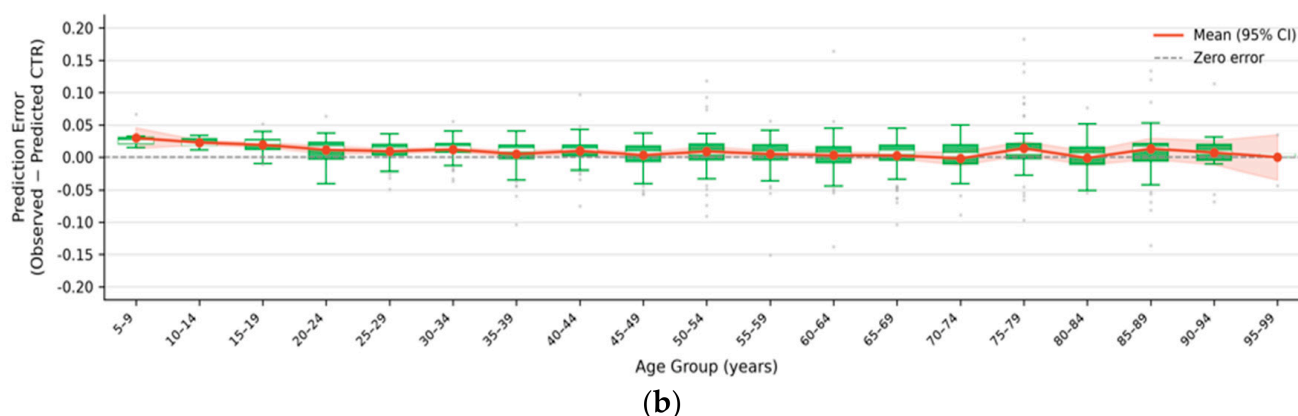


Figure 5. CTR prediction performance across age group. Box plots represent the interquartile range. The pink band represents the 95% confidence interval (CI) for the mean of each age group; a narrower band indicates higher precision, whereas a wider band reflects greater variability. (a) Predicted CTR by age group. (b) Prediction error (observed–prediction CTR) by age group.

The plots reveal a gradual increase in CTR values with advancing age. For individuals up to middle age (44 years), the average CTR remains within normal ranges. However, a noticeable rise in mean CTR values occurs between ages 45 to 79. A notable prevalence of cardiomegaly emerges in the 65-year age group, where the interquartile range exceeds the 0.5 CTR threshold. Cardiomegaly becomes evident in individuals over 80 years, where the mean CTR value surpasses 0.5.

In Table 5, the CTR values obtained from experimented models were compared with the observed CTR references for further validation based on the mean CTR, Mean Absolute Error (MAE), Pearson correlation. *p*-values were calculated using a paired *t*-test for each model. The proposed model achieved the highest overall agreement with reference measurements, demonstrating a superior Pearson correlation coefficient of 0.8527, which indicates greater linear consistency across diverse patient anatomies than baseline networks like UNet ($r = 0.8211$) or D-UNet ($r = 0.6798$). Although the paired *t*-test indicates a statistically significant difference ($p = 0.0024$), the exceptionally low MAE of 0.0070 ± 0.0311 confirms that this deviation is minimal. Crucially, while UNet++ yields a marginally lower absolute MAE (0.0060), the proposed method achieves a tighter standard deviation (± 0.0311 vs. ± 0.0394), reflecting superior error stability. Given that a CTR variance of just 0.01 to 0.02 can reclassify borderline cases near the critical cardiomegaly diagnostic threshold (≤ 0.50), this combination of high correlation and minimized error variance confirms that the model's minor statistical deviation is clinically negligible and highly dependable for automated screening workflows.

Table 5. Paired evaluation between the predictions and references.

Model	Mean CTR	MAE	Pearson Correlation (<i>r</i>)	<i>p</i> -Value
UNet	0.4443 ± 0.0577	0.0140 ± 0.0342	0.8211	<0.0001
UNet3+	0.4476 ± 0.0574	0.0113 ± 0.0386	0.7718	<0.0001
UNet++	0.4530 ± 0.0644	0.0060 ± 0.0394	0.7963	0.0147
D-UNet	0.4489 ± 0.0696	0.0092 ± 0.0518	0.6798	0.0004
Inf-Net	0.4515 ± 0.0644	0.0069 ± 0.0437	0.7462	0.0047
DTA-UNet	0.4460 ± 0.0519	0.0073 ± 0.0310	0.8045	0.1368
ELU-Net	0.4503 ± 0.0638	0.0085 ± 0.0368	0.8211	0.0006
The proposed method	0.4512 ± 0.0579	0.0070 ± 0.0311	0.8527	0.0024

4. Discussion

The experimental results indicate that most expanded UNet models, including UNet++, D-UNet, Inf-Net, and ELU-Net, outperformed the UNet in terms of segmentation accuracy. This performance gain is attributed to architectures designed for enhanced feature extraction. Among these, the ELU-Net model achieved the highest overall accuracy.

Notably, lung detection accuracy (DSC) was ~3% higher than heart detection, likely due to ambiguous cardiac boundaries and artifacts. For instance, Inf-Net achieved the highest lung segmentation accuracy but performed poorly in heart segmentation. This discrepancy is likely due to the inherent challenges in heart segmentation. These findings underscore the importance of tailoring model architectures to specific anatomical structures and their unique characteristics.

The integration of attention mechanisms with the CBAM into the ELU-Net model further improved segmentation accuracy. This is particularly beneficial in medical imaging, where subtle details and boundaries are critical for accurate segmentation. Nevertheless, not all attention mechanisms yielded improvements. While ECA and CBAM enhanced accuracy, the GC and GCT attention gates resulted in slightly lower performance compared to the standalone ELU-Net model. This illustrates that the effectiveness of attention mechanisms depends on their design and compatibility with the base model. Additionally, the limited representational capacity of the attention blocks may have constrained further performance gains, while the already high accuracy of the ELU-Net model leaves limited scope for additional enhancement.

Diagnosis of cardiomegaly remains subjective and variable among radiologists, limiting classification consistency due to a lack of interpretability and objectivity. However, the proposed method demonstrated strong performance in CTR measurement, a critical metric for the diagnosis of cardiomegaly. The high Pearson correlation coefficient and low MAE demonstrate the model's reliability, supporting its suitability for practical clinical applications. These findings suggest that the proposed method can serve as a valuable adjunct for radiologists by providing consistent and objective CTR measurements that complement conventional diagnostic approaches.

The observed R-square value of 0.6831 reflects the inherent complexity of CTR estimation within a diverse clinical environment. This variance is attributable to several intersecting factors that challenge the precision of automated boundary regression. First, significant anatomical and structural variability exists; biological differences in chest geometry and cardiac positioning across varied demographic and age cohorts introduce structural variance that a singular baseline model may not fully encapsulate. Second, boundary ambiguity frequently arises in complex pathological cases, where the mathematical precision of cardiac border detection is constrained by the partial obscuration of anatomy due to pleural effusions, consolidation, or variations in diaphragmatic position. Finally, differences in dataset characteristics—specifically the distinct image acquisition parameters and cohort distributions of the AdelaideMRI dataset compared to the training data—introduce clinical variance. Collectively, these factors suggest that while the model maintains robust diagnostic utility, its predictive limits are shaped by both biological heterogeneity and real-world imaging variability. To address this issue, the training dataset needs to incorporate data from multiple sources, along with image augmentations, to improve the model's generalization and performance across diverse imaging conditions.

5. Conclusions

Chest X-rays remain an invaluable tool in diagnosing thoracic conditions, and automated CTR calculation using deep learning holds significant potential for maintaining high diagnostic accuracy. Through systematic benchmarking, this study establishes that

the ELU-Net model augmented with a CBAM attention gate achieves the highest performance, yielding an average Dice Similarity Coefficient (DSC) of 95.05% and a mean Intersection over Union (IOU) of 90.73%. External validation on a heterogeneous clinical population demonstrated strong alignment with expert manual measurements, evidenced by a superior Pearson correlation coefficient ($r = 0.8527$) and a low Mean Absolute Error (MAE) of 0.0070 ± 0.0311 . To ensure real-world clinical utility, the proposed framework is designed to interface directly with existing clinical systems, operating as an automated, high-throughput triage tool that flags abnormal cardiothoracic ratios prior to radiologist review. Ultimately, this approach improves spatial interpretability and provides a stable, scalable tool to optimize diagnostic workflows and reduce radiologist workload.

Author Contributions: H.-W.J. served as the principal investigators, contributing to model design, result analysis, and drafting and revising the manuscript. H.-W.J. and S.-H.L. conducted cardiothoracic ratio experiments and analyzed the results. H.-J.Y. offered technical and methodological support for the deep learning model. R.D. supervised the experiment. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study complied with the Declaration of Helsinki and was approved by the University of South Australia's Human Research Ethics Committee (Application ID: 206982, approval date: 12 June 2025) registered with the National Health and Medical Research Council (NHMRC).

Informed Consent Statement: Informed consent to participate was waived due to the retrospective nature of the study.

Data Availability Statement: The independent clinical datasets analyzed in this study are not publicly available due to the privacy issues of patients, but they are available from the corresponding authors on reasonable request.

Acknowledgments: The authors wish to thank AdelaideMRI for providing the clinical datasets and computational resources necessary to conduct this research. During the preparation of this work, H.-W.J. used Gemini 3 for manuscript revision and proofreading. After using this tool, H.-W.J. reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Conflicts of Interest: Authors Hye-Won Jung and Roger Davies were employed by the company AdelaideMRI. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Alghamdi, S.S.; Abdelaziz, I.; Albadri, M.; Alyanbaawi, S.; Aljondi, R.; Tajaldeen, A. Study of cardiomegaly using chest x-ray. *J. Radiat. Res. Appl. Sci.* **2020**, *13*, 460–467. [\[CrossRef\]](#)
2. Chaisangmongkon, W.; Chamveha, I.; Promwiset, T.; Saiviroonporn, P.; Tongdee, T. External Validation of Deep Learning Algorithms for Cardiothoracic Ratio Measurement. *IEEE Access* **2021**, *9*, 110287–110298. [\[CrossRef\]](#)
3. Kim, D.; Lee, J.H.; Jang, M.-j.; Park, J.; Hong, W.; Lee, C.S.; Yang, S.Y.; Park, C.M. The Performance of a Deep Learning-Based Automatic Measurement Model for Measuring the Cardiothoracic Ratio on Chest Radiographs. *Bioengineering* **2023**, *10*, 1077. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Sogancioglu, E.; Murphy, K.; Calli, E.; Scholten, E.T.; Schalekamp, S.; Van Ginneken, B. Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification. *IEEE Access* **2020**, *8*, 94631–94642. [\[CrossRef\]](#)
5. Kiselev, S.; Maksudov, B.; Mustafaev, T.; Kuleev, R.; Ibragimov, B. Automating cardiothoracic ratio measurements in chest X-rays. In *2021 International Conference "Nonlinearity, Information and Robotics" (NIR)*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–4.
6. Gupte, T.; Niljekar, M.; Gawali, M.; Kulkarni, V.; Kharat, A.; Pant, A. Deep learning models for calculation of cardiothoracic ratio from chest radiographs for assisted diagnosis of cardiomegaly. In *2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
7. Calli, E.; Sogancioglu, E.; van Ginneken, B.; van Leeuwen, K.G.; Murphy, K. Deep learning for chest X-ray analysis: A survey. *Med. Image Anal.* **2021**, *72*, 102125. [\[CrossRef\]](#) [\[PubMed\]](#)

8. Agrawal, T.; Choudhary, P. ReSE-Net: Enhanced UNet architecture for lung segmentation in chest radiography images. *Comput. Intell.* **2023**, *39*, 456–477. [\[CrossRef\]](#)
9. Saiviroonporn, P.; Wonglaksanapimon, S.; Chaisangmongkon, W.; Chamveha, I.; Yodprom, P.; Butnian, K.; Siriapisith, T.; Tongdee, T. A clinical evaluation study of cardiothoracic ratio measurement using artificial intelligence. *BMC Med. Imaging* **2022**, *22*, 46. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: Piscataway, NJ, USA, 2020; pp. 1055–1059.
11. Ehab, W.; Huang, L.; Li, Y. UNet and variants for medical image segmentation. *Int. J. Netw. Dyn. Intell.* **2024**, *3*, 100009. [\[CrossRef\]](#)
12. Rajamani, K.T.; Rani, P.; Siebert, H.; ElagiriRamalingam, R.; Heinrich, M.P. Attention-augmented U-Net (AA-U-Net) for semantic segmentation. *Signal Image Video Process.* **2023**, *17*, 981–989. [\[PubMed\]](#)
13. Wu, J.-X.; Pai, C.-C.; Kan, C.-D.; Chen, P.-Y.; Chen, W.-L.; Lin, C.-H. Chest X-ray image analysis with combining 2D and 1D convolutional neural network based classifier for rapid cardiomegaly screening. *IEEE Access* **2022**, *10*, 47824–47836. [\[CrossRef\]](#)
14. Kusakunniran, W.; Saiviroonporn, P.; Siriapisith, T.; Tongdee, T.; Uraiverotchanakorn, A.; Leesakul, S.; Thongnarintr, P.; Kuama, A.; Yodprom, P. Automatic measurement of cardiothoracic ratio in chest x-ray images with ProGAN-generated dataset. *Appl. Comput. Inform.* **2026**, *22*, 239–248.
15. Chamveha, I.; Promwiset, T.; Tongdee, T.; Saiviroonporn, P.; Chaisangmongkon, W. Automated cardiothoracic ratio calculation and cardiomegaly detection using deep learning approach. *arXiv* **2020**, arXiv:2002.07468.
16. Din, S.; Shoaib, M.; Serpedin, E. CXR-seg: A novel deep learning network for lung segmentation from chest X-ray images. *Bioengineering* **2025**, *12*, 167. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Li, Z.; Hou, Z.; Chen, C.; Hao, Z.; An, Y.; Liang, S.; Lu, B. Automatic cardiothoracic ratio calculation with deep learning. *IEEE Access* **2019**, *7*, 37749–37756. [\[CrossRef\]](#)
18. Wu, W.; Yan, J.; Zhao, Y.; Sun, Q.; Zhang, H.; Cheng, J.; Liang, D.; Chen, Y.; Zhang, Z.; Li, Z.-C. Multi-task learning for concurrent survival prediction and semi-supervised segmentation of gliomas in brain MRI. *Displays* **2023**, *78*, 102402.
19. Yu, X.; Liang, M.; Zeng, Q.; Ren, Z.; Kang, C.; Jiang, C.; Chen, G.; Liang, B.; Pu, B.; Zhu, N. ESE-UNet: An Extremely Simple yet Efficient UNet for medical image segmentation. *Biomed. Signal Process. Control* **2026**, *112*, 108634. [\[CrossRef\]](#)
20. Ding, T.; Zhang, Z.; Yan, J.; Sun, Q.; Zhao, Y.; Li, Z.-C. GaLNet: Weakly-Supervised Learning for Evidence-Based Tumor Grading and Localization in MR Imaging. In *International Forum on Digital TV and Wireless Multimedia Communications*; Springer: Singapore, 2020; pp. 249–258.
21. Khanh, T.L.B.; Dao, D.-P.; Ho, N.-H.; Yang, H.-J.; Baek, E.-T.; Lee, G.; Kim, S.-H.; Yoo, S.B. Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging. *Appl. Sci.* **2020**, *10*, 5729. [\[CrossRef\]](#)
22. Xu, Y.; Hou, S.; Wang, X.; Li, D.; Lu, L. A medical image segmentation method based on improved UNet 3+ network. *Diagnostics* **2023**, *13*, 576. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Jha, D.; Riegler, M.A.; Johansen, D.; Halvorsen, P.; Johansen, H.D. Doubleu-net: A deep convolutional neural network for medical image segmentation. In *IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*; IEEE: Piscataway, NJ, USA, 2020; pp. 558–564.
24. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA)*; Springer: Cham, Switzerland, 2018; pp. 3–11.
25. Fan, D.-P.; Zhou, T.; Ji, G.-P.; Zhou, Y.; Chen, G.; Fu, H.; Shen, J.; Shao, L. Inf-net: Automatic covid-19 lung infection segmentation from ct images. *IEEE Trans. Med. Imaging* **2020**, *39*, 2626–2637. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Li, Y.; Yan, B.; Hou, J.; Bai, B.; Huang, X.; Xu, C.; Fang, L. UNet based on dynamic convolution decomposition and triplet attention. *Sci. Rep.* **2024**, *14*, 271. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Deng, Y.; Hou, Y.; Yan, J.; Zeng, D. ELU-net: An efficient and lightweight U-net for medical image segmentation. *IEEE Access* **2022**, *10*, 35932–35941. [\[CrossRef\]](#)
28. Kim, M.; Lee, B.-D. Automatic lung segmentation on chest X-rays using self-attention deep neural network. *Sensors* **2021**, *21*, 369. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Liu, Y.; Zhang, Z.; Liu, X.; Lei, W.; Xia, X. Deep learning based mineral image classification combined with visual attention mechanism. *IEEE Access* **2021**, *9*, 98091–98109. [\[CrossRef\]](#)
30. He, Z.; He, Y.; Cao, W. Deformable image registration with attention-guided fusion of multi-scale deformation fields. *Appl. Intell.* **2023**, *53*, 2936–2950.
31. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Oktay, O. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.

33. Cai, Y.; Wang, Y. Ma-unet: An improved version of unet based on multi-scale and attention mechanism for medical image segmentation. In *Third International Conference on Electronics and Communication; Network and Computer Technology (ECNCT 2021)*; SPIE: Bellingham, WA, USA, 2022; pp. 205–211.
34. Mu, N.; Lyu, Z.; Rezaeitalashmahalleh, M.; Tang, J.; Jiang, J. An attention residual u-net with differential preprocessing and geometric postprocessing: Learning how to segment vasculature including intracranial aneurysms. *Med. Image Anal.* **2023**, *84*, 102697. [[CrossRef](#)] [[PubMed](#)]
35. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2018; pp. 7132–7141.
36. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2020; pp. 11534–11542.
37. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*; CVF: New York, NY, USA, 2018; pp. 3–19.
38. Ni, J.; Wu, J.; Tong, J.; Chen, Z.; Zhao, J. GC-Net: Global context network for medical image segmentation. *Comput. Methods Programs Biomed.* **2020**, *190*, 105121. [[CrossRef](#)] [[PubMed](#)]
39. Yang, Z.; Zhu, L.; Wu, Y.; Yang, Y. Gated channel transformation for visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2020; pp. 11794–11803.
40. Jaeger, S.; Candemir, S.; Antani, S.; Wang, Y.-X.J.; Lu, P.-X.; Thoma, G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant. Imaging Med. Surg.* **2014**, *4*, 475. [[PubMed](#)]
41. Dutta, A.; Zisserman, A. The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*; ACM Digital Library: New York, NY, USA, 2019; pp. 2276–2279.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.