



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

흉부 측면 X 선 영상에서 심혈관 경계의  
딥러닝 기반 자동 분석 소프트웨어  
개발 및 검증

Development and Validation of Deep Learning-Based Software  
for Automatic Analysis of Cardiovascular Borders in Lateral  
Chest X-ray Images

울산대학교 대학원

의과학과

정규준

흉부 측면 X 선 영상에서 심혈관 경계의  
딥러닝 기반 자동 분석 소프트웨어  
개발 및 검증

지도 교수 강수진, 이준구

이 논문을 공학석사 학위 논문으로 제출함

2024년 02월

울산대학교 대학원

의 과학과

정규준

정규준의 공학석사학위 논문을 인준함

심사위원 이 준 구 (인)

심사위원 강 수 진 (인)

심사위원 양 동 현 (인)

울 산 대 학 교 대 학 원

2024 년 02 월

## Abstract

Cardiovascular border (CB) analysis is a fundamental method for detecting and assessing the severity of heart diseases using chest X-ray (CXR). This study aimed to develop and validate a deep learning-based CB automatic analysis software algorithm for quantitative analysis of lateral chest X-ray images. For detecting CB in Lateral CXR images, we utilized the Mask R-CNN. This model is an algorithm capable of accurate object detection and segmentation. We proposed more precise results using reasonable pre-processing and post-processing algorithms with this model, and suggested indicator for quantitative analysis of images. The CB detection performance of the developed model was evaluated with a precision of 99.92%, recall of 99.89%, F1 score of 99.91%, and a false positive rate per patient of 0.48. We used a developed validation dataset to verify the reliability between the CB automatic label (CB\_auto) and the CB manual label (CB\_hand) using the proposed CB measure index. At the 4cm point in the proposed CB measure, the highest Intraclass correlation coefficient (ICC) (0.95-0.99) was observed. Comparing the absolute difference in CB measure between the normal control group and abnormal groups at the 4cm point, significant differences were found in Tricuspid valve disease (TD) 12.7mm, Mitral valve disease (MD) 11.2mm, Pulmonary valve disease (PD) 8.5mm, and Aortic valve disease (AD) 6.8mm. The success rate of the CB measure was 91.6% in normal, 83.0% in AD, 84.5% in MD, 84.4% in PD, and 77.7% in TD. Subsequently, we conducted statistical analysis on the measured CB measure results by gender (male, female) in the normal control group. The Mean and Standard deviation in the normal control group were  $25.5 \pm 9.7$  for males and  $16.7 \pm 8.9$  for females, and the median (Interquartile ranges (IQRs)) were 25.5 (18.9-31.9) for males and 16.7 (10.7-22.6) for females. Additionally, we compared the median (IQRs) values of z-scores after fitting the model for the normal control group using the GAMLSS library. For males, normal was 0.01 (-0.69 to +0.68), AD was -0.58 (-1.51 to +0.29), MD was -0.90 (-1.62 to -0.10), PD was -0.17 (-1.03 to +0.21), and TD was -1.09 (-1.78 to +0.30). For females, normal was 0.01 (-0.67 to +0.68), AD was -0.33 (-1.12 to +0.42), MD was -0.91 (-1.57 to -0.22), PD was -0.80 (-1.57 to +0.13), and TD was -0.90 (-1.47 to 0.00). When comparing normal control with abnormal groups, the abnormal groups generally showed lower values.

## 차 례

Abstract .....	i
차례 .....	ii
그림목차 .....	iii
1. Introduction .....	1
2. Background .....	3
2.1. Convolutional Neural Networks (CNNs) .....	3
2.1.1. CNNs .....	3
2.1.2. Residual Network (ResNet) .....	4
2.2. Feature Convolutional Network (FCN) .....	4
2.3. Feature Pyramid Network (FPN) .....	4
2.4. Region with CNN (R-CNN) Series .....	5
2.4.1. R-CNN .....	5
2.4.2. Fast R-CNN .....	5
2.4.3. Faster R-CNN .....	6
2.4.4. Mask R-CNN .....	6
3. Related Works .....	7
3.1. Previous studies .....	7
3.2. Automated cardiovascular border analysis .....	7
4. Proposed Methods .....	8
4.1. Model development .....	8
4.2. Implementation details .....	9
4.3. Pre-Processing: Image processing and Data augmentation .....	9
4.4. Post-Processing: Rule-based algorithms .....	11
4.5. Statistical analysis .....	12
5. Internal Results .....	14
5.1. Datasets .....	14
5.2. Pre-Processing for CB detection of Lateral CXR images .....	14
5.3. Backbone network of Mask R-CNN .....	15
5.4. Post-Processing for Enhanced performance of CB detection .....	15
5.5. Reliability of CB_auto .....	16
6. External Results .....	19
6.1. Datasets .....	19
6.2. CB measurement among the normal control and abnormal groups ..	19
7. Discussion .....	23
8. Conclusion .....	24
Reference .....	25
국문요약 .....	27

## 그림목차

Figure 1 .....	3
Figure 2 .....	8
Figure 3 .....	10
Figure 4 .....	13
Figure 5 .....	13
Figure 6 .....	16
Figure 7 .....	17
Figure 8 .....	18
Figure 9 .....	19
Figure 10 .....	21
Figure 11 .....	21

## ***1. Introduction***

Valvular Heart Disease (VHD) occurs when one or more of the heart valves do not open or close properly. This encompasses various conditions, including Stenosis, where the valve opening narrows, impeding the normal blood flow. Another manifestation is Prolapse, characterized by a displacement of the valve or improper closure of valve flaps. Additionally, Regurgitation occurs when blood flows backward through a valve. These valve dysfunctions contribute to the complexity of VHD. In the context of VHD, Heart Failure emerges as a serious medical condition, signifying the heart's inability to pump sufficient blood to meet the body's oxygen requirements. This can lead to an enlarged heart or Heart Failure. Treatment options include medication, surgery, and other procedures aimed at repairing or replacing the affected valve.

Cardiac Chamber Enlargement is important in the prediction of morbidity and mortality for a multitude of cardiovascular processes. The changes in the cardiac border and the emergence of new or different interfaces with adjacent lungs can be readily identified in chest radiographs. These radiographic images provide a valuable means to observe alterations in the cardiac border, contributing to the non-invasive and cost-effective evaluation and diagnosis of Cardiac chamber enlargement. Thus, leveraging the advantages of chest radiographs in assessing cardiac morphology becomes pivotal in enhancing our understanding and management of cardiovascular conditions, particularly those associated with cardiac chamber enlargement.

Recent integrations of deep learning technologies have catalyzed significant advancements in the domain of computer vision. Convolutional Neural Networks (CNNs), a predominant deep learning algorithm, are primarily harnessed for tasks related to image recognition and classification. Through the automatic extraction of high-level abstract features from images, CNNs have demonstrated unparalleled efficacy in image processing, particularly in discerning local patterns and structures [4]. In the realm of semantic segmentation, Fully Convolutional Networks (FCN) have emerged as a transformative approach, enabling end-to-end pixel-wise classification and facilitating dense predictions across images [17]. Furthermore, the Feature Pyramid Network (FPN) has been introduced to address the inherent challenges associated with detecting objects of diverse sizes [5]. By generating feature maps across multiple scales, the FPN augments the efficacy of object detection and classification. This enhancement is realized by amalgamating feature maps from disparate layers, thereby harnessing both high-resolution and multi-scale information.

These methodologies have paved the way for monumental progress in object detection algorithms, which are predominantly bifurcated into single-stage and two-stage paradigms. The single-stage approach amalgamates the processes of object candidate region extraction, classification, and bounding box regression [6,7]. Prototypical examples of this approach include You Only Look Once (YOLO) and Single Shot Multibox Detector (SSD), which prioritize computational efficiency, rendering them apt for real-time object detection [6,7]. Conversely, the two-stage paradigm, epitomized by the Region with Convolutional Neural Networks (RCNN) series, segregates the extraction of object candidate regions from images and their subsequent classification via distinct CNN models [9,10]. While this method boasts superior accuracy, it is encumbered by computational overheads. To mitigate this, Fast RCNN amalgamates object candidate extraction and classification [9], while Faster RCNN further streamlines the process by incorporating the Region Proposal Network (RPN) for

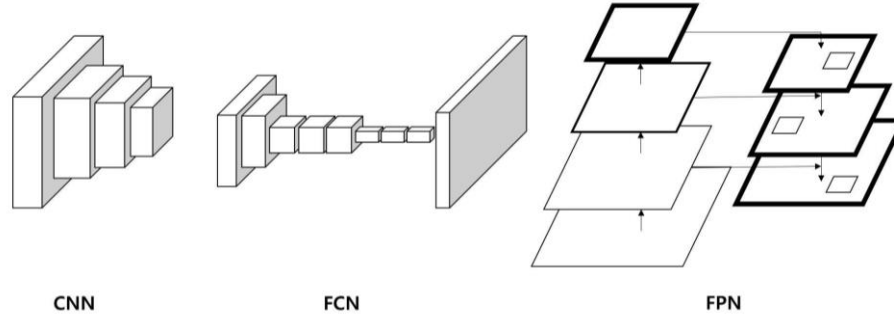


automated object candidate extraction [10]. Notably, Mask R-CNN, a salient member of the RCNN lineage, is distinguished by its capability to execute both object detection and pixel-level segmentation of individual entities [11]. This is achieved by integrating the FCN and FPN, which, during the processing phase of Mask R-CNN, furnishes rich spatial information and multi-scale context for each object candidate region [5,11,17].

In this study, our primary objective is to develop a deep learning-based software algorithm for detecting and segmenting borderlines in chest X-ray images using the Mask R-CNN framework. Building upon this, the study also aims to establish normal reference values for cardiovascular borders by applying the proposed cardiac border (CB) measurement method on Lateral chest X-ray images within a large dataset, and subsequently, to conduct a comparative analysis with an abnormal group.

## 2. Background

In the evolving landscape of computer vision, Convolutional Neural Networks (CNNs) have emerged as a foundational deep learning algorithm tailored for structured data like images [3]. Building upon CNNs, Fully Convolutional Networks (FCN) introduced a paradigm shift, targeting semantic segmentation with pixel-wise predictions [17]. To address the challenges of multi-scale object detection, the Feature Pyramid Network (FPN) was developed, utilizing a multi-scale pyramid of feature maps [4]. The Region with Convolutional Neural Networks (R-CNN) series [8, 9, 19, 11] achieved progress using superior techniques with CNN, FCN, and FPN for object detection, segmentation, and classification.



**Figure 1.** Overview of CNN, FCN and FPN.

### 2.1. Convolutional Neural Networks (CNNs)

**2.1.1. CNNs.** Convolutional Neural Networks (CNNs) have witnessed substantial evolution since their inception. The foundational model, LeNet-5, introduced by LeCun et al., was designed for handwritten digit recognition. It utilized convolution operations defined as

$$(I * F)(x, y) = \sum_{i=-\frac{k}{2}}^{\frac{k}{2}} \sum_{j=-\frac{k}{2}}^{\frac{k}{2}} I(x + i, y + j) \cdot F(i, j)$$

to extract features from images [2].

The groundbreaking AlexNet by Krizhevsky et al. significantly outperformed other models in the ImageNet challenge in 2012. This model introduced the use of Rectified Linear Units (ReLUs) as activation functions, defined as

$$f(x) = \max(0, x)$$

, and Dropout layers to prevent overfitting [3].

After AlexNet, the VGGNet further deepened the network architecture, employing multiple 3x3 convolutional layers to capture intricate patterns [12]. GoogLeNet, also known as the Inception network, introduced the "Inception module", a novel structure that allowed for increased depth without a substantial rise in computational cost [13]. These developments set the stage for the introduction of ResNet, which tackled the vanishing gradient problem inherent in very deep networks through the use of Skip Connections.

**2.1.2. Residual Network (ResNet).** Residual Network (ResNet) introduced a novel architecture that revolutionized deep learning by allowing the training of much deeper networks than was previously feasible [14]. The core idea behind ResNet is the introduction of "skip" or "shortcut" connections that bypass one or more layers. The fundamental building block of ResNet is the Residual Block. Instead of trying to learn an underlying mapping  $H(x)$  directly, ResNets aim to approximate the residual function:

$$F(x) = H(x) - x$$

The original input  $x$  is then added back to this residual, resulting in the final output [14]:

$$H(x) = F(x) + x$$

This simple yet powerful concept of adding the input to the output helps in mitigating the vanishing gradient problem, allowing for the training of networks with depths of even 100 layers or more. The introduction of these residual connections led to significant improvements in training deep networks. Not only did ResNet achieve state-of-the-art performance on various benchmarks, but they also highlighted the importance of network architecture design in deep learning. The residual learning principle has since become a foundational concept, inspiring numerous subsequent architectures and variants [15].

**2.2. Full Convolutional Network (FCN).** Fully Convolutional Networks (FCN) have revolutionized the domain of semantic segmentation, where the goal is to assign a class label to each pixel in an image. Unlike traditional CNNs that produce a fixed-size output (typically a vector of class probabilities), FCNs are designed to produce outputs (segmentation maps) that are of the same spatial dimensions as the input image. This is achieved by replacing the Fully connected layers in CNNs with Convolutional layers. Given an input image  $I$ , an FCN produces a segmentation map  $S$  where each pixel  $S_{i,j}$  represents the probability of the pixel  $I_{i,j}$  belonging to a particular class. The transformation from  $I$  to  $S$  is governed by the convolutional operations and learned filters in the FCN. Additionally, FCNs employ Up sampling layers to upscale coarse feature maps to the original input size, ensuring dense pixel-wise predictions. The loss function for training an FCN typically involves comparing the predicted segmentation map  $S$  with the ground truth map  $G$  using a pixel-wise cross-entropy loss:

$$L = - \sum_{i,j} G_{i,j} \log(S_{i,j})$$

This loss ensures that the FCN learns to produce accurate pixel-wise class probabilities. Through this architecture and training mechanism, FCNs have set new benchmarks in semantic segmentation tasks [17].

**2.3. Feature Pyramid Network (FPN).** Feature Pyramid Network (FPN) is a novel neural network architecture designed to enhance the scale invariance in object detection tasks [4]. Traditional convolutional networks tend to produce feature maps at a single scale, which may not be optimal for detecting objects of varying sizes. FPN addresses this by constructing a multi-scale pyramid of feature maps.

The FPN architecture builds a Top-down pathway with Lateral connections to the Bottom-up computed feature hierarchy. For a given feature map level  $l$ , the Top-down feature map is up sampled by a factor of 2 and then merged with the corresponding Bottom-up feature map through an Element-wise addition:

$$M_l = U(M_{l+1}) + L_l$$

Where  $M_l$  is the Merged feature map at level  $l$ ,  $U$  is the Up-sampling Operation, and  $L_l$  is the Lateral connection from the Bottom-up pathway [4]. This process is repeated across all levels of the pyramid, ensuring high-resolution features are enriched with semantic information from deeper layers.

The introduction of FPNs has significantly improved the performance of object detection models, especially for detecting objects across a wide range of scales. By leveraging multi-scale feature representations, FPNs provide a more comprehensive and semantically rich feature set for object detection tasks [11].

**2.4. Region with CNN (R-CNN) Series.** The Regions with CNN Features (R-CNN) series of models have revolutionized object detection by seamlessly integrating Region proposal methods with deep convolutional networks. Starting with the foundational R-CNN, which introduced Region-based Object detection, the lineage evolved with Fast R-CNN and Faster R-CNN, streamlining the detection process and enhancing computational efficiency. Mask R-CNN further extended this lineage by adding Pixel-level Segmentation capabilities, showcasing the continual advancements in Object detection and Instance segmentation.

**2.4.1. R-CNN.** R-CNN marked a significant advancement in object detection by seamlessly integrating Region proposal methods with CNNs [8]. Traditional object detection techniques often employed a sliding window approach, which, while comprehensive, was computationally intensive and lacked precision. R-CNN innovatively addressed this limitation by introducing a two-stage method.

Initially, for each image, approximately 2000 Region proposals are generated using the Selective Search [16], a technique that clusters adjacent pixels based on texture, color, and intensity to identify potential object bounding boxes. Mathematically, for an image  $I$ , the set of region proposals  $R = \{R_1, R_2, \dots, R_n\}$  is obtained using:

$$R = \text{SelectiveSearch}(I)$$

Each proposed region  $R_i$  is resized and processed through a Pre-Trained CNN, typically AlexNet, to derive a fixed-length feature vector  $F_i$ :

$$F_i = \text{CNN}(R_i)$$

Subsequently, these feature vectors are classified using Class-specific linear SVMs. Additionally, a Bounding box regressor refines the object's location based on the extracted features.

R-CNN's approach revolutionized object detection, setting new performance benchmarks. Its methodology of combining region proposals with deep learning laid the foundation for future advancements in the domain.

**2.4.2. Fast R-CNN.** Fast R-CNN emerged as an optimized successor to R-CNN, addressing its predecessor's computational inefficiencies by introducing a unified model that uses a single network for both Object classification and Bounding box regression [9]. Unlike R-CNN, which processed regions independently, leading to redundant computations, Fast R-CNN streamlined the process. The entire image is first passed through a convolutional network to produce a feature map. Then, using a technique called Region of Interest (RoI) Pooling, proposed regions (typically obtained from an external method like Selective Search) are mapped onto this feature map. This RoI Pooling layer standardizes the varying sizes of the proposed regions into a fixed-size feature vector.

A pivotal innovation in Fast R-CNN is its multi-task loss function. For each RoI, the network predicts a set of class probabilities and bounding box offsets. The Multi-task loss  $L$  is defined as:

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{loc}(t^u, v)$$

Where  $p$  represents the predicted probability distribution over classes.  $u$  denotes the ground-truth class label.  $t^u$  is the predicted bounding box offset for class  $u$ .  $v$  stands for the ground-truth bounding box offset.  $L_{cls}$  and  $L_{loc}$  are the Classification and Bounding box regression losses, respectively.  $\lambda$  is a balancing parameter. The term  $[u \geq 1]$  ensures that the bounding box regression loss is only considered for RoIs containing an object [9]. This combined loss function allows for end-to-end training, optimizing both classification and bounding box regression simultaneously, leading to enhanced accuracy and efficiency.

**2.4.3. Faster R-CNN.** Faster R-CNN's pivotal innovation is the Region Proposal Network (RPN), a Fully convolutional network that predicts Object bounds and Objectness scores simultaneously [10]. The RPN operates on the feature map obtained from the primary convolutional layers of the network.

For each position in the feature map, the RPN predicts multiple Region proposals based on predefined Anchor boxes. These anchor boxes are of various scales and aspect ratios, providing a set of reference boxes over which the predictions are made. Given a feature map  $F$ , for each anchor  $A$  at position  $p$ , the RPN predicts: An objectness score  $o_{p,A}$  indicating the likelihood that the anchor contains an object. Four values  $t_{p,A} = (t_x, t_y, t_w, t_h)$  representing the predicted bounding box's offsets and scales relative to the anchor. the objectness score is given by:

$$o_{p,A} = \sigma(f_o(F_p))$$

Where  $\sigma$  is the sigmoid function and  $f_o$  is the part of the RPN predicting objectness. The bounding box offsets and scales are predicted as:

$$t_{p,A} = f_t(F_p)$$

Where  $f_t$  is the part of the RPN predicting box coordinates.

The proposals generated by the RPN, based on these scores and offsets, are then processed further for object detection. The combined multi-task loss of the RPN considers both the Objectness loss (binary logistic regression) and the Bounding box regression loss. Faster R-CNN's introduction of the RPN streamlined the object detection pipeline, making it both efficient and accurate by generating high-quality region proposals.

**2.4.4. Mask R-CNN.** A pioneering architecture in Instance segmentation ingeniously builds upon the principles of FCN and FPN to achieve unparalleled precision in delineating object boundaries [5,11,17]. FCN revolutionized semantic segmentation by enabling end-to-end pixel-wise classification, while FPN addressed the challenge of detecting objects across varying scales by generating multi-scale feature maps. Integrating these concepts, Mask R-CNN employs an FPN backbone for robust feature extraction and appends an FCN-inspired branch for mask prediction. For each Region of Interest (RoI), this Mask branch predicts a Binary mask, capturing the object's spatial contours at the pixel level. The overall loss function  $L$  in Mask R-CNN is a combination of the Classification loss  $L_{cls}$ , Bounding box regression loss  $L_{box}$ , and Mask binary classification Loss  $L_{mask}$ :

$$L = L_{cls} + L_{box} + L_{mask}$$

This multi-task loss ensures that the network simultaneously optimizes object classification, bounding box regression, and mask prediction. By leveraging the strengths of FCN and FPN and optimizing this composite loss, Mask R-CNN sets a benchmark in instance segmentation, showcasing its versatility and effectiveness.

### **3. Related Works**

#### **3.1. Previous studies**

In earlier studies, it's important to recognize that while initial models laid the foundation for the use of CNNs in classifying thoracic diseases, including cardiomegaly, they may have had limitations in the specificity and precision of cardiac border segmentation. For instance, Ferreira Junior et al. [22] achieved substantial results with their DenseNet model but possibly faced challenges in model generalization across diverse datasets, which is crucial for robust diagnostic tools. The study by Arunachalam and Bhavathankar [23] advanced the field by modifying DenseNet121 to improve classification performance, yet it may not have fully addressed the complexities of cardiac structure delineation, which is critical for accurate heart disease diagnosis. In contrast, the research by Pi-Yun Chen et al. [24] introduced an innovative high-dimensional multiple regression estimator, signaling a shift towards more sophisticated analytical tools that can rapidly screen and differentiate cardiomegaly from normal anatomical variations. This approach likely offers improvements in the automated and assistive aspects of chest X-ray analysis, addressing labor shortages and the need for more efficient diagnostic workflows. The study by C. Kim et al. [1] represents a significant advancement by not only focusing on cardiomegaly detection but also providing detailed cardiovascular border analysis. Their deep learning-based algorithm was specifically designed to diagnose and quantitatively evaluate valvular heart disease, presenting a highly reliable system with a strong correlation to echocardiographic measurements. This research likely overcomes many of the earlier limitations by delivering precise cardiac border segmentation and comprehensive diagnostic insights.

The use of chest X-ray (CXR) imaging itself presents several advantages for cardiac border analysis. It is a non-invasive, cost-effective, and widely accessible method that allows for the rapid screening of cardiopulmonary diseases. Compared to more complex imaging modalities like cardiac CT or MRI, CXRs can be performed quickly and do not require extensive preparation or high operational costs, making them suitable for routine and follow-up health examinations. Additionally, CXRs can reveal asymptomatic stages of diseases like cardiomegaly, which might not be detectable with other methods such as electrocardiography.

In summary, while earlier studies laid the groundwork for the use of CNNs in thoracic disease classification, recent advancements have refined these approaches to offer more detailed cardiac analyses, addressing previous limitations and enhancing the overall utility of CXR imaging in cardiac diagnostics.

#### **3.2. Automated cardiovascular border analysis**

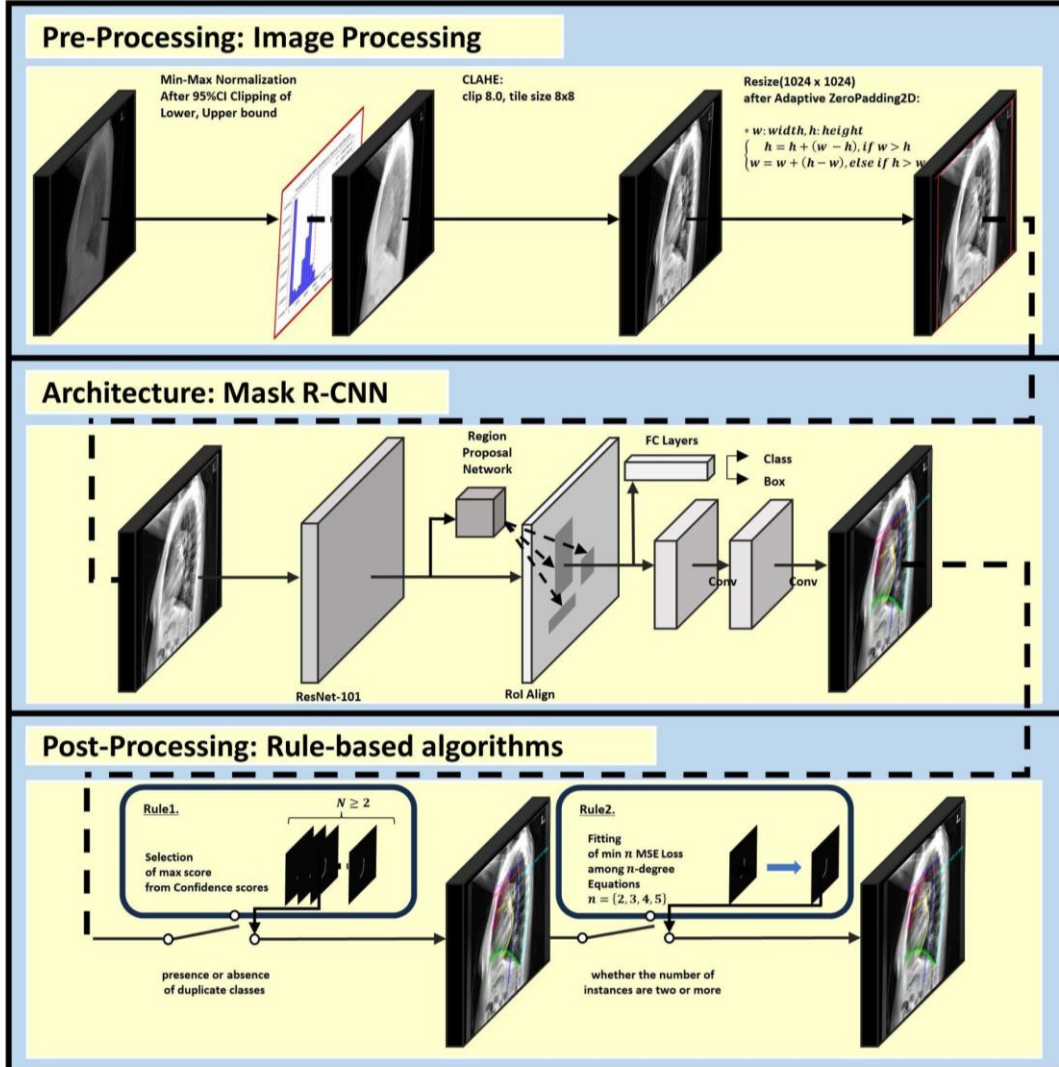
In the research presented herein, the application of deep learning techniques to medical image analysis explores new potentials for significantly enhancing diagnostic accuracy. A custom deep-learning methodology for the automatic analysis of cardiovascular borders in chest radiographs is introduced [1]. This methodology has been developed to improve the effectiveness of diagnosing valvular heart disease.

Our team has rigorously trained and validated the model across diverse patient cohorts to ensure its robustness and generalizability. The findings demonstrate the potential of the proposed methodology to surpass traditional radiographic analysis in specific scenarios, thereby increasing diagnostic accuracy for heart disease. This underscores the innovative impact of deep learning in medical imaging, especially in tasks requiring precision and consistency. Moreover, the study emphasizes the need for integrating advanced computational techniques into clinical practice to enable more efficient diagnostic processes. Our current aim is to develop a CB\_auto system specifically for Lateral CXRs.

Consequently, following our preliminary research on frontal images, we explored the potential to deepen our understanding of heart disease from different perspectives by detecting and analyzing cardiac borders in lateral images.

#### 4. Proposed Methods

We propose methodologies aimed at enhancing and refining the detection and segmentation capabilities. This paper presents Figure 2 illustrating the architecture of the CB\_auto system developed in our study.



**Figure 2.** Overall procedure on CB detection in Lateral chest X-ray.

##### 4.1. Model development

The input image is resized to a dimension of 1024 x 1024 with 3 channels. serving as the model's input image. The operation of this model can be divided into three key stages. Firstly, for object detection and segmentation, the ResNet101 architecture is chosen as the backbone network and is initialized with pre-trained weights from the MSCOCO dataset. The ResNet model offers two variants in terms of size, namely ResNet50 and ResNet101. The chosen model facilitates the capture of even richer features. ResNet, a prevalent convolutional neural network structure in deep learning, mitigates the problem of vanishing gradients in deep networks through the integration of residual connections, thereby enabling efficient training. In the second stage, the extracted feature maps are passed through a RPN to predict object positions and classes. Leveraging this information, Region of Interest Align (RoIAlign) is employed to accurately extract regions of detected objects. Interposed within these stages, the

FPN is strategically employed to effectively recognize objects of varying sizes and generate robust feature representations. FPN adopts a pyramid structure to generate feature maps of diverse resolutions, enabling the model to adapt to objects of varying scales. Consequently, the Mask R-CNN framework excels in accurately detecting objects and performing pixel-level segmentation within input images, underscored by its formidable capabilities.

#### ***4.2. Implementation details***

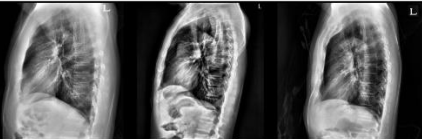

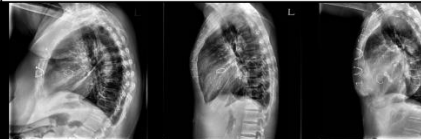


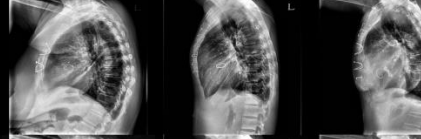









The parameters applied to proceed learning are as follows. We used a mini-batch size of 8 images and a stochastic gradient descent (SGD) optimizer for 100 epochs was applied with an initial learning rate of 0.001 with weight decay, where the learning rate drops by a factor of 0.1 every 25 epochs. The learning momentum is used 0.9. Anchors per image of RPN network used 256. The number of ROIs per image to feed to Classifier and Mask heads used 256. Due to the nature of multiple classes, Cross entropy was used as a loss function. Model learning required approximately 1 to 2 hours per epoch and approximately 5 to 6 days of total time spent. The inference time of CB detection required approximately 2 to 3 seconds per image. The algorithm was assessed by using Computer specification; Intel Xeon® Silver 4214R CPU @ 2.40GHz 49GB RAM, NVIDIA® Quadro RTX 8000 with 49GB, Ubuntu 22.04.2 LTS OS.

#### ***4.3. Pre-Processing: Image processing and Data augmentation***

To enhance the image quality conveyed to the model, a preprocessing procedure is applied to the input images. The process involves calculating the 95% confidence interval from the histogram distribution of the image and then clipping pixel values to predefined lower and upper bounds. This effectively mitigates the impact of outliers or extreme values, resulting in clipped min-max normalization of the image data. Next, Contrast limited adaptive histogram equalization (CLAHE) is employed to enhance contrast and finer details within the image. Parameters such as an 8 x 8 tile size and a clipping limit of 8.0 are experimentally selected from various candidates. As a result, the resultant images showcase improved features and a balanced brightness distribution. This aids in visualizing previously concealed subtle patterns within the original image, now refined. Building on this foundation, a final step involves geometric adjustments to maintain the aspect ratio of the image. The absolute difference between the width and height values is computed, and zero padding is added to the smaller dimension based on this calculated absolute difference. This ensures the preservation of the image's aspect ratio as it is resized to 1024 x 1024, conforming to the model's input specifications. This integration not only enhances the structural accuracy of the image but also prepares it in an optimally suitable form for the model's interpretative capabilities.

In the data augmentation process, a series of techniques were strategically employed to enhance the diversity and robustness of the training dataset. Specifically, Gaussian blur was applied with a sigma parameter set to the range of (0.0, 1.5) to introduce subtle blurring effects. To incorporate controlled noise variations, Adaptive gaussian noise was utilized with a scale parameter of (0, 0.1 \* 255) and was set to be consistent across all channels (per\_channel=False). To further augment the dataset's spatial diversity, Horizontal flip was incorporated for mirroring, while Affine transformations were judiciously applied, encompassing rotations between (-15, 15) degrees, scaling adjustments in both x and y axes ranging from 0.8 to 1.2, and translational shifts spanning from -1.5 to 1.5 in both dimensions. Notably, to ensure a balanced and varied application of these techniques, two random augmentations were selected and applied to each mini-batch during the training phase. The applied augmentation is visualized in Figure 3.



	Normal	Valve Prosthetic					
		PreOP			PostOP		
Original Image							
Gaussian Blur							
Adaptive Gaussian Noise							
Horizontal Flip							
Affine transformation (rotation, scale, shift)							

**Figure 3.** Augmentation techniques applied to the model training of the CB\_auto system.

#### 4.4. Post-Processing: Rule-based algorithms

As part of our post-processing strategy, we applied two rule-based algorithms. The first algorithm focuses on cases where there can only be one instance of the CB class within an image. Following this rule, if duplicate detections occur within the same class, we retain the result with the highest Confidence score and discard the other duplicates. Applying this post-processing step reduces false positives and yields more accurate results from Bounding box regression information. The pseudo-code is articulated in Algorithm 1.

---

**Algorithm1: Post-Processing for Duplicate detections**

---

**Input** List of detected objects D

**Output** Updated list of detected objects D'

Initialize an empty list D'

**for** each class c **do**

    Initialize a temporary list T

**for** each object o **in** D **do**

**if** o's class is c **then**

            Add o to T

**end for**

**if** length of T > 1 **then**

        Find object o\_max with the highest confidence score in T

        Add o\_max to D'

**else**

        Add all objects in T to D'

**end if**

**end for**

**return** D'

---

The second algorithm addresses instances where extracted CB masks lack consecutive information. This situation can arise due to insufficient contrast or brightness quality along the heart's borders in the image. To address this, we initially utilize Connected-component analysis (CCA) to determine whether the mask contains consecutive information. In this assessment, pixel connectivity is evaluated in a 2D mask using the 8-adjacency criteria. If the mask is found to be non-consecutive, we apply a fitting algorithm for n-degree equations to enhance segmentation performance. The fitting of the Nth-degree polynomial function fitting can be achieved using the "polyfit" and "polyval" functions from the Numpy library. Before applying the Fitting algorithm, we obtain the coordinates of the centerline of the mask. The definition of the centerline coordinates is either the average value of y for the same x value or the average value of x for the same y value among the coordinates where the mask has pixels. The better representation is predefined for each class. Afterward, we apply the fitting algorithm. The operational sequence can be explained in three main steps. First, the equation for the X, Y coordinates of the Centerline values is obtained using the "polyfit" function. The information of the axis that can better represent consecutive data is used as input, while the values from the other axis are used as output. Equations are generated from quadratic to fifth-degree equations, and among the generated equations, the one with the lowest Mean squared error (MSE) loss is selected. Next, the "polyval" function is used to calculate the values of the other axis based on the input axis values. It's important that the values of the input axis used in this step should be continuous from the beginning to the end. This means that were

disconnected should be included as input. Finally, the “polylines” function provided by the OpenCV library is used to fit the CB mask, allowing it to follow the calculated Polynomial centerline equation. The pseudo-code is articulated in Algorithm 2.

---

**Algorithm2: Post-Processing for Enhance segmentation with Polynomial fitting**

---

**Input** CB mask M

**Output** Enhanced CB mask M'

Initialize an empty mask M'

Perform CCA on M to check for consecutive information

**if** mask M is found to be non-consecutive **then**

    Obtain the coordinates of the centerline of mask M

    Predefine the better representation axis for each class

**for** degree n from 2 **to** 5 **do**

        Fit a polynomial of degree n to the centerline coordinates using `numpy.polyfit`

        Calculate MSE loss for the fitted polynomial

**end for**

Select the polynomial with the lowest MSE loss

Use `numpy.polyval` to calculate the values of the other axis based on the selected polynomial

Use `Opencv.polylines` to fit the CB mask M to follow the calculated polynomial centerline equation, update M'

**end if**

**return** enhanced CB mask M'

---

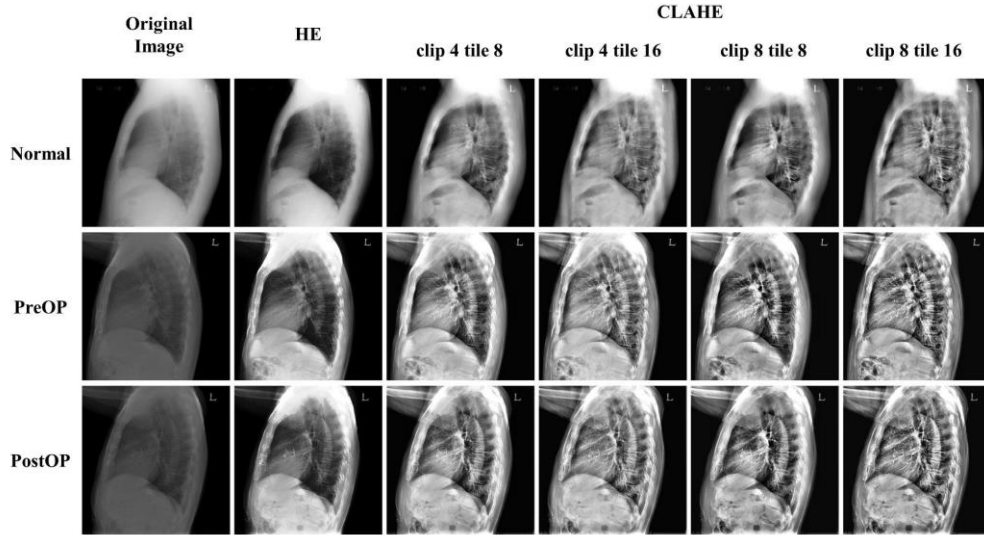
#### 4.5. Statistical analysis

For model performance evaluation, the detection or segmentation performance for all Total CBs was evaluated. The detection performance was evaluated using True Positive (TP), False Positive (FP), False Negative (FN), Precision, Recall, F1 score, and False Positive Ratio (FPR). The segmentation performance was assessed using the Jaccard Coefficient and Dice Similarity Coefficient (DSC).

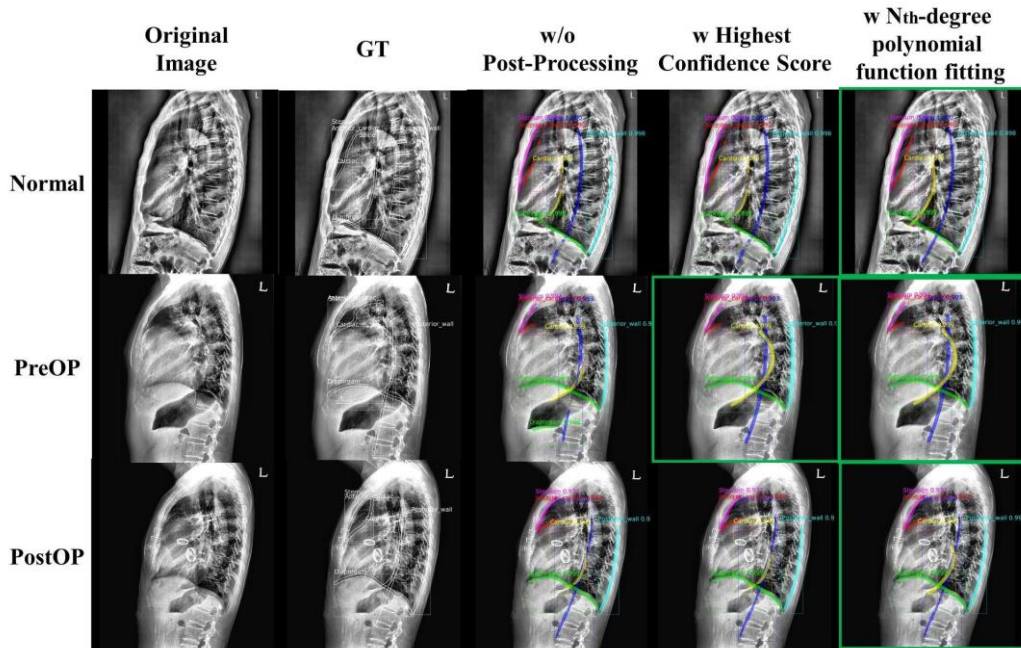
The CB parameters, using CB\_auto and CB\_hand, were calculated at 1, 2, 3, 4, and 5 cm above the intersection point of the diaphragm and cardiac borderline. This CB measure is visualized in Figure 5. The distances from the cardiac to the spine borderline were measured at these points. The reliability of the CB parameters determined by CB\_auto was evaluated using the Intraclass correlation coefficient (ICC) and 95% limits of agreement by Bland-Altman analysis. To evaluate the difference between CB\_auto and CB\_hand, the Absolute error was calculated using the following formula:  $|(CB\_auto)-(CB\_hand)|$ . The differences in CB parameters determined by CB\_auto or CB\_hand between normal controls and abnormal groups were evaluated using paired t-tests or ANOVA. A p-value of less than 0.05 was considered statistically significant.

Characteristics of the study population were described using means, standard deviations, medians, and Interquartile ranges (IQRs) for continuous variables and absolute (n) and relative (%) frequencies for categorical ones. additionally, modeling strategies were considered for the construction of the charts called Generalized Additive Models for Location, Scale, and Shape (GAMLSS), which is available in the GAMLSS package in R software, was used. GAMLSS are distributional regression models, where all the parameters of the assumed distribution for the response can be modeled as functions of the explanatory variables. The models are appropriate when the focus is not only on the mean (or location) of the distribution but possibly other parts of the distribution like variance, quantiles and skewness and kurtosis or tails. We

aimed to gain a comprehensive understanding of the characteristics and patterns of CB measure by fitting the Box-Cox  $t$  distribution (BCTo) model. To achieve this, we utilized Quantile-Quantile (QQ) plots for each percentile (1<sup>st</sup>, 2.5<sup>th</sup>, 5<sup>th</sup>, 10<sup>th</sup>, 20<sup>th</sup>, 30<sup>th</sup>, 50<sup>th</sup>, 70<sup>th</sup>, 80<sup>th</sup>, 90<sup>th</sup>, 95<sup>th</sup>, 97.5<sup>th</sup>, 99<sup>th</sup>) providing an intuitive visualization of the distribution and variations in CB measure. Finally, to rigorously validate our findings and assess their significance, z-scores were computed for comparative analysis across diverse cohorts.



**Figure 4.** Visualization of Pre-Processed images in Lateral CXR images.



**Figure 5.** Visualization of Post-Processing according to Rule-based algorithms. The figure delineates the before and after scenarios using PreOP and PostOP cases from the same patient.

## 5. Internal Results

We conducted four experiments for our study. Firstly, in the Pre-Processing stage before model input, a comparative experiment was performed using the histogram of images to improve the contrast of the images. Secondly, a comparative experiment of well-known CNN models was conducted for the Encoder backbone network of the model. Thirdly, we propose to enhance detection and segmentation performance by applying Rule-based algorithms for Post-Processing, utilizing the result values obtained from the model output. Lastly, image analysis is performed using the Proposed metric values. Our experiments were conducted in phases: Pre-Processing, Model backbone, Post-Processing, and Quantitative image analysis.

For all the experiments, pre-trained weights from the MSCOCO dataset were employed. All the results were obtained for model evaluation parameters with a confidence score greater than 0, and the evaluation was carried out with Bounding Box (BBox) detection Intersection over Union (IoU) values greater than 0.05 between Ground truth and Prediction. In the Pre-Processing experiment, ResNet101 was utilized as the Backbone network for Mask R-CNN.

The two terms carry the following meanings: CB\_auto represents results obtained using a software algorithm based on deep learning. On the other hand, CB\_hand corresponds to labeling results drawn by expert radiologists.

**5.1. Datasets.** For the development of this study, we utilized a total of 2,998 Lateral chest X-ray (CXR) images, spanning from 2002 to 2020. These images were sourced from Asan Medical Center. The dataset comprises a mix of normal CXRs and those involving valve prosthetics, both pre- and post-operation. In terms of dataset distribution, for training purposes, we included 793 Normal, 630 Pre-Operation (PreOP), and 952 Post-Operation (PostOP) images. For validation and testing, the dataset was composed of 199 normal, 212 PreOP, and 212 PostOP images. Notably, the valve prosthetic images in the validation and test sets consisted of paired patient images, offering a comprehensive view of pre- and post-operation states. This dataset enabled us to train our model to recognize and understand the diversity in heart size and shape. Furthermore, our objective was to extract specific measurements related to the Heart from CXR images. To achieve this, we focused on detecting various anatomical lines, including the cardiac, spine, posterior wall, diaphragm, anterior cardiac, and sternum. The image labeling process was carried out with the assistance of a senior radiologist, utilizing the ImageJ labeling tool.

**5.2. Pre-Processing for CB detection of Lateral CXR images.** Histogram Equalization (HE) and CLAHE share a commonality in utilizing the histogram of images to improve contrast. Initially, HE adjusts the distribution of pixel intensities to enhance the contrast between darker and lighter regions. Subsequently, CLAHE divides the image into small tiles and applies HE independently to each tile. While HE applies uniform adjustments across the entire image, CLAHE operates independently on each tile. Examples of image processing using each technique are illustrated in Figure 3, and the performance outcomes are presented in Table 1. Consequently, prioritizing the False Negative and Recall performance, we were able to select the result values of CLAHE (parameter settings: clip 8, tile 8x8) and applied them uniformly in subsequent experiments.

	Total CBs	TP	FP	FN	Precision	Recall	F1 score	FPR†
<b>w/o Pre-Processing</b>	3738	3732	9	6	99.76	99.84	<b>99.80</b>	1.28
<b>w HE</b>	3738	3731	11	7	99.71	99.81	99.76	1.77
<b>clip 4 tile 8</b>	3738	3729	9	9	99.76	99.76	99.76	1.44
<b>clip 4 tile 16</b>	3738	3731	7	7	<b>99.81</b>	99.81	<b>99.80</b>	<b>1.12</b>
<b>w CLAHE</b> <b>clip 8 tile 8</b>	3738	<b>3734</b>	11	<b>4</b>	99.71	<b>99.89</b>	<b>99.80</b>	1.77
<b>clip 8 tile 16</b>	3738	3728	7	10	<b>99.81</b>	99.73	99.77	<b>1.12</b>

**Table 1.** Detection performance based on each Pre-Processing technique. Herein, TP denotes True Positive, FP signifies False Positive, FN stands for False Negative, and FPR† represents the False Positive Ratio on a Per-Patient basis. The metrics Precision, Recall, F1 Score, and FPR have been expressed by multiplying by 100 for better representation and clarity. This configuration was maintained consistently in subsequent experiments.

**5.3. Backbone network of Mask R-CNN.** The backbone network serves as the Encoder segment of the model, tasked with extracting feature values from the input images. Accordingly, we selected commonly utilized backbone networks [14, 18, 19] for a comparative examination. The performance outcomes are tabulated in Table 2. Based on the results, we were able to select the ResNet101, which was then uniformly applied in the ensuing experiments.

Backbone Network	Total CBs	TP	FP	FN	Precision	Recall	F1 score	FPR†
<b>ResNet50</b>	3738	3721	5	17	99.87	99.55	99.71	0.80
<b>ResNet101</b>	3738	<b>3734</b>	11	<b>4</b>	99.71	<b>99.89</b>	<b>99.80</b>	1.77
<b>DenseNet121*</b>	3738	3709	<b>4</b>	29	<b>99.89</b>	99.22	99.56	<b>0.64</b>
<b>EfficientNet-B3</b>	3738	3642	16	96	99.56	97.43	98.49	2.57

**Table 2.** Detection performance according to the Backbone network of the Mask R-CNN framework. The growth rate of the final dense layer's block in DenseNet121\* has been adjusted from 16 to 48 to achieve channel size matching.

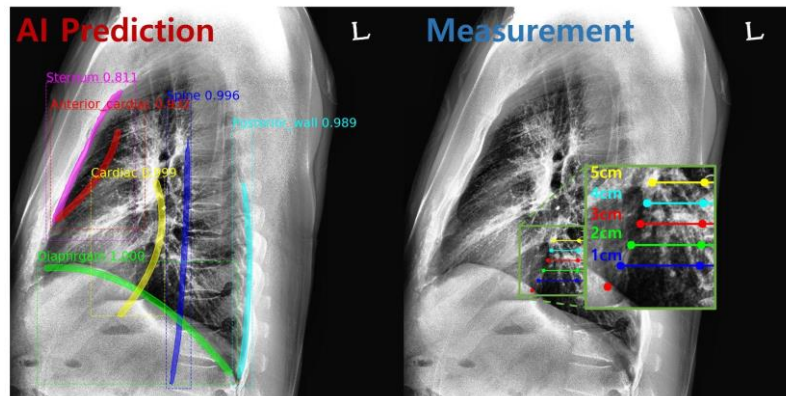
**5.4. Post-Processing for Enhanced performance of CB detection.** The configuration incorporating CLAHE (clip 8, tile 8x8) along with the ResNet101 backbone network was employed. In this section, two Rule-based algorithms are elucidated. Firstly, a method is delineated to bolster detection performance by retaining the detection with the highest Confidence score within a specific class while discarding the others, in scenarios where

multiple detections transpire. Secondly, in instances where the mask values in the imagery are discontinuous, a method is proposed to augment segmentation performance using Nth-degree polynomial function fitting. Detailed discussions on these methodologies are furnished in Section 3.5. The result is visualized in Figure 4 and Table 3.

	Total CBs	Detection						Segmentation		
		TP	FP	FN	Precision	Recall	F1 score	FPR†	Jaccard	DSC
<b>w/o Post-Processing</b>	3738	3734	11	4	99.71	99.89	99.80	1.77	32.73	46.16
<b>w Highest Confidence score</b>	3738	3734	<b>3</b>	4	<b>99.92</b>	99.89	<b>99.91</b>	<b>0.48</b>	33.06	46.55
<b>w Highest Confidence score + Nth-degree polynomial function fitting</b>	3738	3734	3	4	99.92	99.89	99.91	0.48	<b>35.32</b>	<b>49.35</b>

**Table 3.** Performance Evaluation of Detection and Segmentation according to Post-Processing as Per Rule-based algorithms.

**5.5. Reliability of CB\_auto.** In terms of accordance with CB\_hand measurements, the CB measure of 4cm’s cardiac to spine borderline point above diaphragm and cardiac borderline delineated by CB\_auto yielded excellent reliability (ICC>0.95) in the development validation group(Normal: 1cm 0.89, 2cm 0.95, 3cm 0.70, 4cm 0.97, 5cm 0.88; PreOP: 1cm 0.97, 2cm 0.43, 3cm 0.94, 4cm 0.99, 5cm 0.97; PostOP: 1cm 0.95, 2cm 0.76, 3cm 0.91, 4cm 0.95, 5cm 0.91)(all  $p < 0.05$ ).

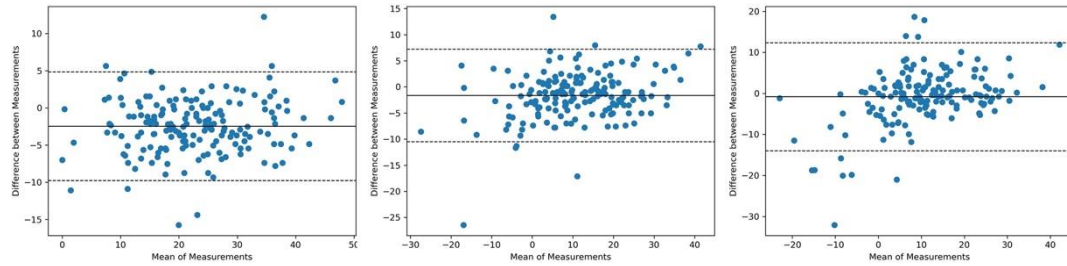


**Figure 6.** Visualization of Example CB indicator measurement in Lateral CXR images.

	1cm	2cm	3cm	4cm	5cm
<b>Normal</b>	4.9±3.5	4.7±3.3	4.0±3.1	3.5±2.8	3.2±2.7
<b>ValveProsthetic(PreOP)</b>	4.5±4.1	4.3±3.8	3.8±3.5	3.5±3.3	3.5±3.4
<b>ValveProsthetic(PostOP)</b>	4.6±5.0	4.5±5.0	4.4±5.1	4.5±5.0	4.7±4.8

**Table 4.** Mean and Standard deviation of Absolute error measurement of Cardiovascular border parameter between CB\_auto and CB\_hand.

On Bland-Altman Plot analysis, the 95% limits of agreement for the CB measure of 4cm's cardiac to spine borderline was Normal  $-2.46 \pm 3.73$ , PreOP  $-1.64 \pm 4.53$  and, PostOP  $-0.81 \pm 6.72$ . The Bland-Altman plots are visualized in Figure 6. The difference between CB\_auto and CB\_hand as defined by the Absolute error was less than 4.9. The AE of CB measure is represented by the Mean and Standard deviation in Table 4. In the validation dataset of VHD and the Public dataset, the difference between CB\_auto and CB\_hand was not calculated due to the absence of CB\_hand.



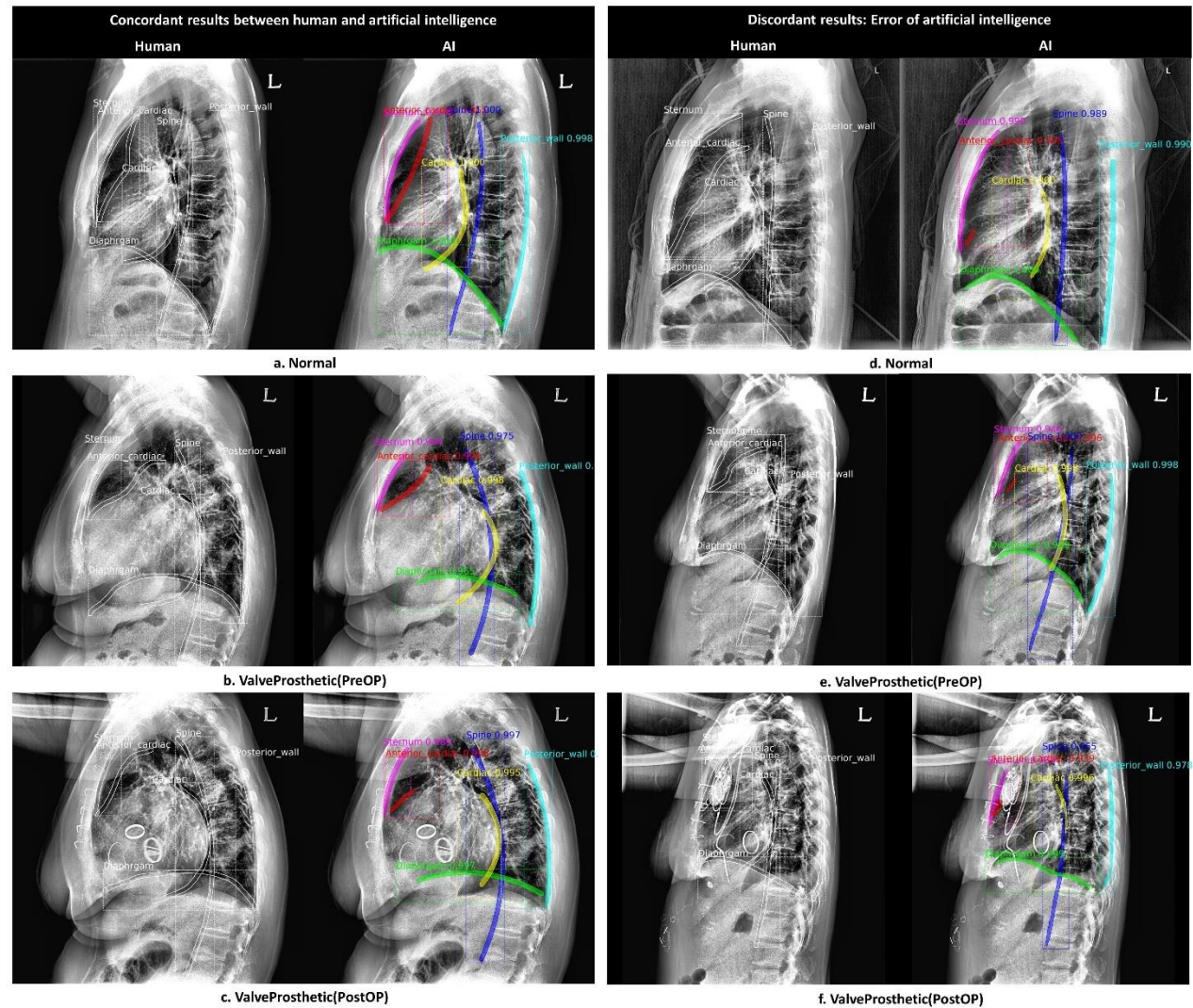
**Figure 7.** Bland-Altman of 4cm point indicator value above the diaphragm and cardiac borderline intersection among Normal, PreOP and, PostOP.

When categorizing CBs into Concordant and Discordant situations, it was possible to visualize them as shown in Figure 7. Typically, concordant results between CB\_hand and CB\_auto are demonstrated. However, in cases where the contrast of certain images is low or when there is overlap with other borderlines, discordant results were observed, especially in the presence of artifacts such as device. In discordant situations, it is challenging to locate the intersection point of the diaphragm and cardiac borderline. Consequently, the measurement of CB becomes unattainable, necessitating the proposal of diverse and robust CB measures for such cases.

In the success rate of the development dataset, all points of CB measure were successfully drawn: Normal 173(86.9%), PreOP 175(82.5%), PostOP 160(75.5%).

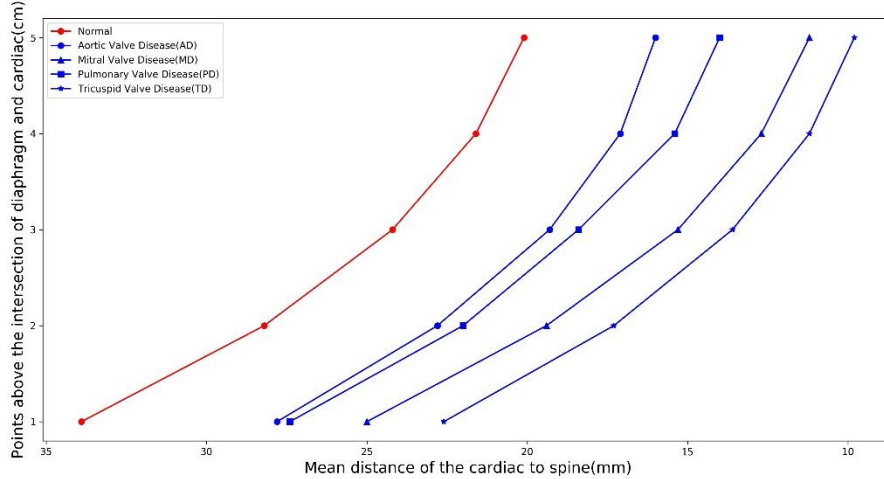


**Figure 8.** Examples of concordant and discordant results between CB\_hand and CB\_auto. **a b c** Concordant case between human measurement and the deep learning algorithm. The cardiovascular borders (CBs) drawn by the deep learning algorithm are almost same as those drawn manually. **b c** the two images belong to the same patient. **c e** Case showing a little discordant result for the anterior cardiac. **d f** Discordant case between human measurement and the deep learning algorithm, especially in the drawing of the anterior cardiac, cardiac. This case is quite difficult even for an observer because the identification of each CB is not easy due to the decreased overall density of the cardiac border.



## 6. External Results

The comparisons among the Validation datasets, Mean and Standard deviation of CB measure calculated by CB\_auto among Normal, Aortic valve disease (AD), Mitral valve disease (MD), Pulmonary valve disease (PD), and Tricuspid valve disease (TD), which are presented in Table 5. Figure 8 illustrates the plotted average CB measure values, while the y-axis indicates points 1, 2, 3, 4, and 5cm above the intersection of the diaphragm and cardiac borderline. This plot allows the visualization of the average cardiac borderline for each validation dataset, facilitating the observation of differences in CB measure average values.



**Figure 9.** The CB measure plot of CB measures (1, 2, 3, 4, 5 cm points) using CB\_auto.

**6.1. Datasets.** The external large dataset validated a total of 32,875 Normal CXR images and a total of 6,957 VHD CXR images, with 2,564 images related to Aortic valve disease (AD), 3,644 images related to Mitral valve disease (MD), 64 images related to Pulmonary valve disease (PD), and 685 images related to Tricuspid valve disease (TD).

The external public dataset, CheXpert [20], provided a collection of images, 402 from a normal cardiac condition group and 284 from an abnormal cardiac condition group, for analysis. The Normal group comprised subjects with no Lung Lesion and either no Enlarged Cardiom and Cardiomegaly or no mention of these conditions in the radiology reports. Conversely, the abnormal group included subjects with either no Lung Lesion or no mention of it in the radiology reports, but with the presence of Enlarged Cardiom and Cardiomegaly.

### 6.2. CB measurement among the normal control and abnormal groups.

The mean CB measure's Absolute error is calculated between normal control and abnormal groups in all cases (vs. AD: 1cm 6.1mm, 2cm 5.4mm, 3cm 4.9mm, 4cm 4.5mm, 5cm 4.1mm; vs. MD: 1cm 8.9mm, 2cm 8.8mm, 3cm 8.9mm, 4cm 8.9mm, 5cm 8.9mm; vs. PD: 1cm 6.5mm, 2cm 6.2mm, 3cm 5.8mm, 4cm 6.2mm, 5cm 6.1mm; vs. TD: 1cm 11.3mm, 2cm 10.9mm, 3cm 10.6mm, 4cm 10.4mm, 5cm 10.3mm). normal control measured larger than the abnormal groups (all  $p < 0.05$ ). At the 4cm point, where the ICC was high, significant differences were observed in the following order when distinguishing between normal control and abnormalities: TD, MD, PD, and AD. The corresponding Absolute error measurement was as follows: TD 12.7mm, MD 11.2mm, PD 8.5mm and, AD 6.8mm. At normal and abnormal condition of the public dataset, normal control had larger CB measurements than abnormal group ( $p < 0.05$ ). Due to the absence of unit vector values in the public dataset, a direct comparison with other validation datasets could not be conducted.

In the success rate of validation and public dataset, all points of CB measure were successfully drawn Normal 30131 (91.6%), AD 2128 (83.0%), MD 3080 (84.5%), PD 54 (84.4%), TD 532 (77.7%), Normal and Abnormal of the public dataset are 317 (78.9%) and

192 (67.6%).

	1cm	2cm	3cm	4cm	5cm
<b>Normal</b>	33.9 ± 12.8	28.2 ± 11.7	24.2 ± 10.9	21.6 ± 10.3	20.1 ± 9.8
<b>Aortic Valve Disease (AD)</b>	27.8 ± 14.2	22.8 ± 13.1	19.3 ± 12.3	17.1 ± 11.7	16.0 ± 11.3
<b>Mitral Valve Disease (MD)</b>	25.0 ± 12.9	19.4 ± 11.8	15.3 ± 11.0	12.7 ± 10.5	11.2 ± 10.2
<b>Pulmonary Valve Disease (PD)</b>	27.4 ± 12.9	22.0 ± 11.6	18.4 ± 10.4	15.4 ± 10.0	14.0 ± 9.7
<b>Tricuspid Valve Disease (TD)</b>	22.6 ± 13.0	17.3 ± 12.0	13.6 ± 11.4	11.2 ± 10.8	9.8 ± 10.6

**Table 5.** Mean and Standard deviation of CB measures (1, 2, 3, 4, 5 cm points) using CB\_auto.

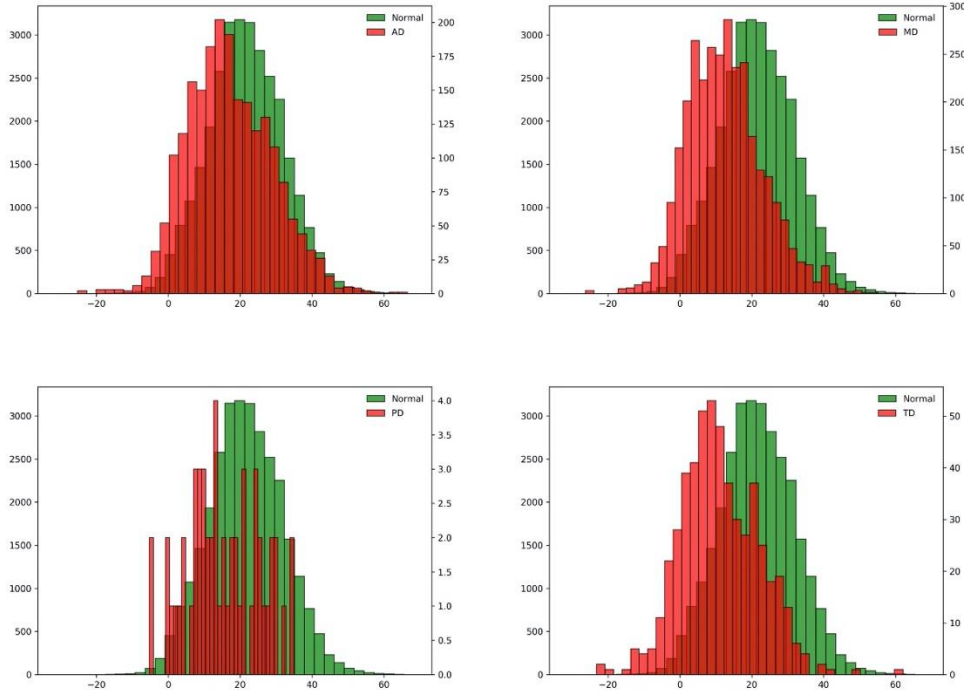
For the CB measure values measured at the 4cm point under normal controls, the Mean and Standard deviation by gender are as follows: for males, it is 25.5±9.7, and for females, it is 16.7±8.9. The median (IQR) is 25.5 (18.9-31.9) for males and 16.7 (10.7-22.6) for females. These results suggest that the reference values for CB measure in females are generally lower than in males. CB measure values were analyzed by Gender (male, female) and Age groups (under 39, 40-59, 60-79, 80 and above). For males, the Mean ± Standard deviation of CB measure values was as follows: under 39 years, 24.6±9.7; 40-59 years, 25.6±9.4; 60-79 years, 25.7±10.3; and 80 years and above, 24.5±12.9. For females, these values were: under 39 years, 17.2±8.5; 40-59 years, 16.8±8.7; 60-79 years, 16.4±9.4; and 80 years and above, 18.2±10.8. The median (IQR) values for males were: under 39 years, 24.9 (18.5-30.9); 40-59 years, 25.7 (19.1-31.9); 60-79 years, 25.5 (18.5-32.5); and 80 years and above, 23.8 (14.8-33.7). For females, these were: under 39 years, 17.7 (12.3-23.0); 40-59 years, 16.9 (10.9-22.8); 60-79 years, 16.1 (9.9-22.4); and 80 years and above, 17.7 (8.6-26.3). These findings indicate that CB measures reference values are generally lower in females across all age groups when compared to males. The reference values for CB measure by Gender and Age under normal controls are detailed in Table 6.

	male				female			
	-39 year	40-59 year	60-79 year	80- year	-39 year	40-59 year	60-79 year	80- year
<b>Baseline Characteristics: mean ± standard deviation</b>								
Age, years	52.7 ± 10.7				53.3 ± 10.3			
gender, n(%)	1825 (10.7)	11053 (65.1)	4008 (23.6)	101 (0.6)	1176 (8.7)	8979 (66.3)	3306 (24.4)	80 (0.6)
<b>CB measures: mean ± standard deviation</b>								
1cm	38.8±12.4	39.2±11.9	38.6±12.7	37.0±14.5	29.2±10.1	27.6±10.1	26.4±10.9	27.8±13.1
2cm	32.5±11.3	33.1±10.8	32.8±11.7	31.4±13.8	23.7±9.4	22.5±9.5	21.5±10.2	22.9±12.1
3cm	27.8±10.4	28.6±10.0	28.5±10.9	27.3±13.2	19.7±8.9	18.9±9.1	18.3±9.8	19.7±11.4
4cm	24.6±9.7	25.6±9.4	25.7±10.3	24.5±12.9	17.2±8.5	16.8±8.7	16.4±9.4	18.2±10.8
5cm	22.7±9.1	23.7±9.0	23.9±9.9	22.8±12.6	15.9±8.1	15.7±8.5	15.5±9.2	17.2±9.9
<b>CB measures: median(IQR ranges)</b>								
1cm	38.1 (30.4 - 46.7)	38.9 (31.0 - 46.9)	38.3 (29.8 - 46.5)	36.8 (27.0 - 46.2)	29.2 (22.6 - 35.5)	27.4 (21.0 - 34.2)	26.1 (18.9 - 33.7)	28.0 (17.6 - 38.1)
2cm	31.9 (24.7 - 39.9)	32.9 (25.7 - 40.3)	32.5 (24.7 - 40.3)	31.6 (20.0 - 41.0)	23.7 (17.9 - 29.8)	22.4 (16.1 - 28.8)	21.2 (14.4 - 28.2)	23.2 (13.4 - 33.2)
3cm	27.6 (21.0 - 34.8)	28.6 (21.6 - 35.2)	28.4 (20.8 - 35.8)	26.7 (16.8 - 37.0)	20.0 (14.2 - 25.7)	18.9 (12.8 - 25.1)	18.1 (11.5 - 24.7)	20.2 (9.9 - 28.3)
4cm	24.9 (18.5 - 30.9)	25.7 (19.1 - 31.9)	25.5 (18.5 - 32.5)	23.8 (14.8 - 33.7)	17.7 (12.3 - 23.0)	16.9 (10.9 - 22.8)	16.1 (9.9 - 22.4)	17.7 (8.6 - 26.3)
5cm	23.0 (17.1 - 28.9)	23.9 (17.9 - 29.8)	23.9 (17.1 - 30.4)	21.4 (13.0 - 31.1)	16.5 (11.0 - 21.4)	16.0 (10.1 - 21.4)	15.4 (9.1 - 21.4)	16.6 (8.7 - 24.5)

**Table 6.** Normal reference value according to gender and age of cardiovascular borders in chest X-ray.

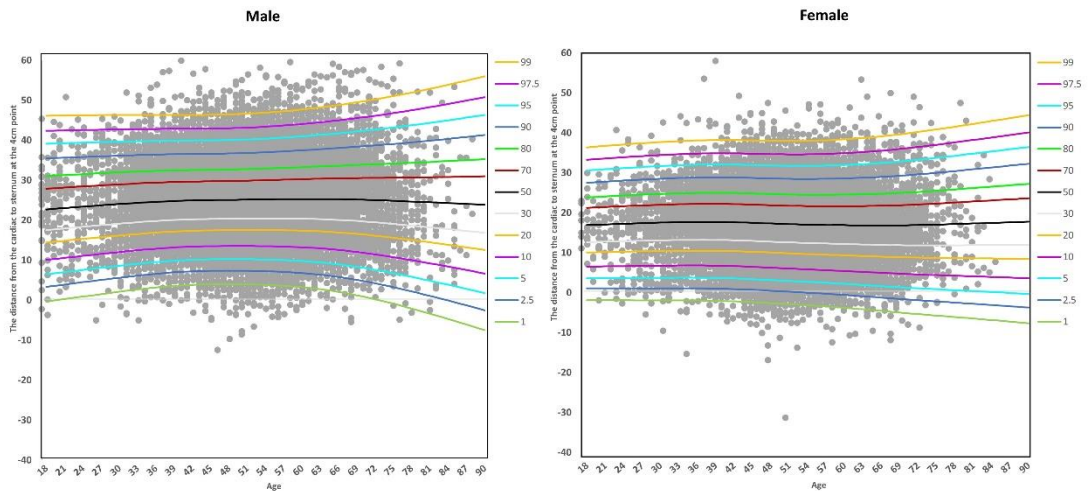
In the CB measures, the value at the 4cm point, which showed the highest reliability, was used as a basis for comparing normal controls and abnormal groups. Firstly, distributions that can

be visually compared are visualized in Figure 10. Additionally, Table 5 presents the Mean and Standard deviation of the distributions for each category in the validation dataset.



**Figure 10.** The Comparison distribution of normal control vs abnormal groups.

Then, for the analysis using Generalized Additive Models for Location, Scale, and Shape (GAMLSS), the CB measure values were subjected to a Shift transformation of +40mm. After fitting the model with these adjusted values, statistical analysis was conducted on the restored values. For comparative analysis, the CB measure values from the normal control group in the validation dataset were divided by gender (male, female) and fitted to the model. The Quantile-Quantile (QQ) plots for each percentile (1st, 2.5th, 5th, 10th, 20th, 30th, 50th, 70th, 80th, 90th, 95th, 97.5th, 99th) of this model are visualized in Figure 11. According to the percentile ranges, the distribution among all males was 4,561 (27.2%) for less than -30%, 7,130 (42.4%) for 30-70%, and 5,115 (30.4%) for more than 70%. For all females, the distribution was 3,906 (29.3%) for less than -30%, 5,561 (41.7%) for 30-70%, and 3,858 (29.0%) for more than 70%.



**Figure 11.** Normal Quantile-Quantile (QQ) plots of the fitted model using GAMLSS library.

Using the fitted Normal distribution model, the z-scores for percentile ranges were calculated. For the entire male population, the median (IQRs) was 0.01 (-0.69 to +0.68), for the average below -30% range it was -1.09 (-1.49 to -0.84), for the 30-70% range it was -0.04 (-0.31 to +0.24), and for above 70% it was 1.02 (+0.75 to +1.40). For the entire female population, the median (IQRs) was 0.01 (-0.67 to +0.68), for the average below -30% range it was -1.08 (-1.49 to -0.77), for the 30-70% range it was 0.01 (-0.25 to +0.29), and for above 70% it was 1.05 (+0.78 to +1.44).

The z-scores for the defined percentile ranges of abnormal groups are as follows, presented as median (IQRs). For AD, the overall male population showed -0.58 (-1.51 to +0.29), average below -30% was -1.34 (-1.81 to -0.94), 30-70% was -0.07 (-0.26 to +0.23), and above 70% was 0.90 (+0.68 to +1.38). The overall female population showed -0.33 (-1.12 to +0.42), average below -30% was -1.20 (-1.66 to -0.88), 30-70% was -0.05 (-0.30 to +0.19), and above 70% was 1.05 (+0.75 to +1.52). For MD, the overall male population was -0.90 (-1.62 to -0.10), average below -30% was -1.40 (-1.99 to -1.00), 30-70% was -0.05 (-0.32 to -0.20), and above 70% was 1.05 (+0.77 to +1.47). The overall female population was -0.91 (-1.57 to -0.22), average below -30% was -1.39 (-1.82 to -0.95), 30-70% was -0.12 (-0.34 to +0.14), and above 70% was 0.99 (+0.74 to +1.32). For PD, the overall male population was -0.17 (-1.03 to +0.21), average below -30% was -1.40 (-1.53 to +1.17), 30-70% was -0.04 (-0.21 to +0.21), and above 70% was 0.89 (+0.87 to +0.91). The overall female population was -0.80 (-1.57 to +0.13), average below -30% was -1.57 (-1.80 to -0.94), 30-70% was -0.05 (-0.30 to +0.17), and above 70% was 1.43 (+1.30 to +1.61). For TD, the overall male population was -1.09 (-1.78 to +0.30), average below -30% was -1.57 (-2.06 to -1.16), 30-70% was -0.19 (-0.37 to +0.13), and above 70% was 0.84 (+0.63 to +1.16). The overall female population was -0.90 (-1.47 to -0.00), average below -30% was -1.36 (-1.82 to -0.99), 30-70% was -0.89 (-1.43 to -0.01), and above 70% was 0.92 (+0.73 to +1.10). It was observed that the z-score values in the abnormal groups were lower compared to the normal control.

## 7. Discussion

We developed a novel deep learning-driven CB\_auto system and validated it using Datasets from the Internal datasets using a manually generated reference standard (CB\_hand) and, studied external datasets. The major findings of this study are as follows: (1) The development of software for CB detection involved three fundamental stages. Firstly, Pre-Processing was conducted to enhance image contrast. Secondly, an efficient model backbone network was selected and applied for feature extraction. Lastly, Rule-based algorithms were employed as a post-processing step to enhance overall performance. Ultimately, this comprehensive approach resulted in the successful development of software capable of superior CB detection in Lateral CXR images; (2) CB\_auto exhibited high reliability for determining CB measurement at 4cm point above between diaphragm and cardiac (ICC range, 0.95-0.99) and yielded narrow 95% limits of agreement ( $-2.46 \pm 3.73$ mm for the Normal,  $-1.64 \pm 4.53$  for the PreOP,  $-0.81 \pm 6.72$  for the PostOP); (3) the CB measure, normal control measures larger than abnormal groups (all  $p < 0.05$ ).

We consider the contributions of this study to be as follows: First, if deep learning can perform automatic and comprehensive CB analysis, it would be possible to establish sex- and age-specific CB reference values and to more objectively describe CB abnormalities according to the pattern of cardiovascular disease. Second, for experts, understanding the cardiac border in CXR images is facilitated not only by the frontal view but also by integrating lateral images to provide additional visual information. Third, understanding the cardiac border in CXR images is facilitated not only by the frontal view but also by integrating lateral images to provide additional visual information.

In our study, although the difference between CB\_auto and CB\_hand as defined by the Absolute error was less than 4.9mm for all points. The accurate representation of changes in cardiac enlargement through the variation between PreOP and PostOP of the valve prosthetic images was not achieved. Therefore, to apply this deep learning-based algorithm to actual clinical practice, further optimization is still necessary to reduce this difference.

This study had several limitations. First, Accurate and diverse CB measures that can precisely represent the borders of the heart need to be proposed and validated. Therefore, this CB measurement should be secured in future studies in more detail. Second, there is currently insufficient correlation information between CB parameters and Echocardiography (ECHO) parameters. We plan to conduct further research on this matter. Third, Sufficient analysis has not been conducted on external datasets and a large, diverse cohort of various diseases. Some Lateral CXR images present challenges in accurately distinguishing the Cardiac border due to either the presence of diseases or lower image quality. Removing such types of evaluations can lead to improved overall performance. Finally, the possibility of human error cannot be excluded in this study setting. Since some parameters of the CBs, overlap with several anatomical structures, line delineation is difficult in some cases although an expert draws the line. Moreover, since the CBs are drawn with free curves rather than straight lines, different experts may draw the CBs differently even if they intend to delineate the same structure. In these cases, the line depicted by the deep learning algorithm cannot be determined to be wrong. If different experts draw the same structure, there is a possibility that it might have been drawn differently. Therefore, if the deep learning algorithm provides consistent results for parts that are inconsistent among humans, it is expected to reduce human error from the point of view of use.

Our proposed algorithm and image analysis results contribute significantly to the assessment and analysis of the severity of cardiovascular diseases by measuring CB in CXR images, aiding in image interpretation and analysis.

## ***8. Conclusion***

Our goal was to develop the CB\_auto system, and towards this end, we have successfully created an algorithm that automatically detects borders in Lateral Images. Furthermore, our deep learning-driven CB\_auto system provided high reliability in terms of CB Measurement. We conducted a quantitative analysis of images using the proposed CB measurement method, which can demonstrate its potential for effective application in clinical settings. Finally, this study not only introduces a novel CB measurement method but also seeks to establish normal reference values for cardiovascular borders on Lateral CXR images by applying this method to a large dataset. Furthermore, the study endeavors to provide a comprehensive comparative analysis by contrasting these normal reference value with abnormal groups.

## Reference

- [1] C. Kim, G. Lee, H. Oh, G. Jeong, S. W. Kim, E. J. Chun, Y. H. Kim, J. G. Lee, D. H. Yang, A deep learning–based automatic analysis of cardiovascular borders on chest radiographs for diagnosing valvular heart disease. vol. 32, no. 3, pp. 1423-1431, 2021. DOI: 10.1007/s00330-021-08296-9, *European Radiology*.
- [2] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. 86(11), 2278-2324, *Proceedings of the IEEE*.
- [3] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning.521(7553), 436-444, *Nature*.
- [4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks.25, 1097-1105, *Advances in neural information processing systems*.
- [5] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection.2117-2125, *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [6] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. 779-788, *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. 21-37, *European conference on computer vision*.
- [8] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. 580-587, *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [9] Girshick, R. (2015). Fast R-CNN. 1440-1448, *Proceedings of the IEEE international conference on computer vision*.
- [10] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. 28, 91-99, *Advances in neural information processing systems*.
- [11] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. 2961-2969, *Proceedings of the IEEE international conference on computer vision*.
- [12] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556, arXiv preprint*.
- [13] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. 1-9, *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [14] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. 770-778, *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [15] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. 630-645, *European conference on computer vision*.
- [16] Uijlings, J. R., van de Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. 104(2), 154-171, *International journal of computer vision*.
- [17] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. 3431-3440, *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [18] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger. (2017). Densely Connected Convolutional Networks. pp. 2261-2269, 2017. DOI: 10.1109/CVPR.2017.243, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.



- [19] M. Tan, Q. V. Le. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. vol. 97, pp. 6105-6114, 2019, in Proceedings of the 36th International Conference on Machine Learning.
- [20] J. Irvin et al. (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. arXiv:1901.07031, arXiv preprint.
- [21] Spear, M. E. (1952). Charting Statistics. New York: McGraw-Hill. This reference is for Mary Eleanor Spear's original work where she introduced an early form of the box plot, known as the "Range Bar" in her book.
- [22] J. F. Ferreira Junior et al. (2021). A general fully automated deep-learning method to detect cardiomegaly in chest x-rays. vol. 11597, pp. 2581980, in Proceedings of SPIE Medical Imaging 2021: Computer-Aided Diagnosis.
- [23] Siddhika Arunachalam, P. Bhavathankar. (2021). Deep Learning Approach for Diagnosis of Thoracic Diseases from Chest X-Ray Images. pp. 9510133, in Proceedings of the 2021 International Conference on Computational Intelligence and Communication Technologies.
- [24] Pi-Yun Chen et al. (2023). Deep learning-based high-dimensional multiple regression estimator for chest x-ray image classification in rapid cardiomegaly screening. vol. 13, no. 9, pp. 1338, in Translational Biomedicine.
- [25] Rigby, R. A., & Stasinopoulos, D. M. (2006). Using Box-Cox t distribution in GAMLSS to model skewness and kurtosis. vol. 6, 2006, pp. 209-229, Statistical Modeling.

## 국문 요약

Cardiovascular border(CB) 분석은 흉부 엑스레이를 이용하여 심장 질환의 감지 및 심각도를 평가하는 기본 방법이다. 이 연구의 목적은 옆면 흉부 엑스레이의 정량적인 영상 분석을 위한 딥러닝 기반 CB 자동 분석 소프트웨어 알고리즘을 개발하고 검증하는 것이다. Lateral CXR 영상에서 CB 검출을 위해, 우리는 Mask R-CNN 을 이용했다. 이 모델은 정확한 객체 탐지 및 분할을 수행할 수 있는 알고리즘이다. 이 모델을 이용한 합리적인 전처리, 후처리 알고리즘을 이용하여 더 정확한 결과를 얻기 위한 결과를 제안하며, 영상의 정량적인 분석을 위한 지표를 제안한다. 개발된 모델의 CB 검출 성능은 정밀도 99.92, 재현율 99.89, F1 점수 99.91, 환자 당 거짓 양성률 0.48 결과로 평가되었다. 제안된 CB measure 지표를 이용하여 CB\_auto 와 CB\_hand 간의 신뢰성을 확인하기 위해 개발 검증 데이터셋을 사용했습니다. 제안된 심장 보더 척도에서 4cm 지점은 가장 높은 급내상관계수 (0.95-0.99) 를 보였습니다. 심장 보더 척도 중에서 4cm 지점에서 정상 대조군과 비정상 집단 간의 CB measure 의 절댓값 차이를 비교한 결과 삼첨판막질환 12.7mm, 승모판막질환 11.2mm, 폐동맥판막질환 8.5mm 및 대동맥판막질환 6.8mm 로 상당한 차이가 있었습니다. CB measure 성공적으로 측정된 것은 정상 대조군 91.6%, 대동맥판막질환 83.0%, 승모판막질환 84.5%, 폐동맥판막질환 84.4%, 삼첨판막질환 77.7% 이었습니다. 그 다음으로, 측정된 CB measure 결과값을 이용하여 정상 대조군의 성별 (남성, 여성) 을 구분하여 통계 분석을 진행했습니다. 먼저, 정상 대조군의 평균과 표준편차는 남성  $25.5 \pm 9.7$  이었으며, 여성은  $16.7 \pm 8.9$  이었습니다. 그리고 중간값(IQRs)는 남성 25.5(18.9-31.9) 이었으며, 여성은 16.7(10.7-22.6) 이었습니다. 그 다음으로, GAMLSS 라이브러리를 이용하여 정상 대조군의 모형 적합을 진행하여 z-score 의 중간값(사분위수 범위) 값을 비교했습니다. 남성의 경우 정상 대조군 0.01(-0.69 to +0.68), 대동맥판막질환 -0.58(-1.51 to +0.29), 승모판막질환 -0.90(-1.62 to -0.10), 폐동맥판막질환 -0.17(-1.03 to +0.21), 삼첨판막질환 -1.09(-1.78 to +0.30) 이었습니다. 여성의 경우 정상 대조군 0.01(-0.67 to +0.68), 대동맥판막질환 -0.33(-1.12 to +0.42), 승모판막질환 -0.91(-1.57 to -0.22) 폐동맥판막질환 -0.80(-1.57 to +0.13), 삼첨판막질환 -0.90(-1.47 to -0.00) 이었습니다. 정상 대조군 과 비정상 그룹을 비교할 때 비정상 그룹이 전반적으로 낮은 값을 가지고 있었습니다.