Politecnico di Torino

Politecnico di Torino

Corso di Laurea Magistrale in
Ingegneria Biomedica

A.a. 2021/2022
Sessione di Laurea Ottobre 2022

Chest X-ray Image Processing: Clipped Anatomy Detection and Cardiothoracic Ratio Estimation

Relatore:
Filippo Molinari

Candidato:
Matilde Bodritti - 289487

Correlatori:
Jan Aelterman (UGent)
Adriyana Danudibroto (Agfa)

Permission of use on loan

"The author gives permission to make this master dissertation available for consultation and to copy parts of this master dissertation for personal use. In all cases of other use, the copyright terms have to be respected, in particular with regard to the obligation to state explicitly the source when quoting results from this master dissertation."
"This master's dissertation is part of an exam. Any comments formulated by the assessment committee during the oral presentation of the master's dissertation are not included in this text."

Matilde Bodritti, August 2022

Acknowledgement

First of all, I would like to thank my supervisor Prof. Filippo Molinari for allowing me to do my thesis in this interesting field, during my Erasmus in Ghent.

I would like to express my deepest appreciation to my supervisors from Belgium Prof. dr. ir. Jan Aelterman (from Ghent Univerisity) and Dr. Adriyana Danudibroto (from Agfa Healthcare department), for their guidance and patience throughout the duration of this thesis work. They gave me the opportunity to enrich my knowledge in this challenging project and helped me in every step. I would also like to express my gratitude for the opportunity I had to collaborate with Agfa Healthcare department and for the helpful advice received during the meetings with them.

Special thanks is reserved for my family and friends whose moral support during the writing of this thesis was essential.

Sommario

Il rapporto cardiotoracico (CTR) è una misurazione clinica ottenuta da immagini radiografiche del torace utilizzato per stimare le dimensioni del cuore e le possibili anomalie collegate, quale cardiomegalia. La valutazione visiva del CTR nella pratica clinica richiede molto tempo e può presentare variazioni tra gli interpreti. Ottenere una misurazione oggettiva del CTR in modo automatico mira a ridurre la soggettività della valutazione del radiologo e formire un maggiore supporto alla diagnosi durante gli esami di follow-up. L'obiettivo di questa tesi è studiare la misurazione automatica del CTR. Sono stati proposti due approcci per il calcolo del CTR basati sulla segmentazione anatomica a partire da immagini radiografiche del torace. Un primo metodo stima il CTR basandosi solo sulla segmentazione dei polmoni, mentre un secondo metodo aggiunge la segmentazione del cuore per la stima del CTR. Questi metodi sono stati sviluppati per essere robusti anche in caso di polmoni parzialmente fuori dal campo visivo dell'immagine: il detettamento di tali casi è uno dei controlli di qualità che dovrebbero essere effettuati prima di ogni valutazione di immagini radiografiche del torace. In questa tesi, il calcolo del CTR a partire dalla sola segmentazione polmonare ha ottenuto risultati migliori rispetto al calcolo del CTR a partire da entrambe le segmentazioni polmonare e cardiaca. L'errore medio e la deviazione standard dell'errore assoluto sono risultati pari a 0.038 ± 0.040 quando testati su immagini non clippate e 0.058 ± 0.057 quando testati su immagini con polmoni clippati. Sulla base di questo modello di segmentazione polmonare, è stato proposto un algoritmo per rilevare la presenza di anatomia tagliata, il quale ha mostrato risultati promettenti. Grazie alla possibilità di calcolare automaticamente e rapidamente il CTR, è stato condotto un breve studio di popolazione su 25369 immagini radiografiche del torace appartenenti a un dataset pubblicamente disponibile (CheXpert dataset). Ciò ha permesso di valutare come valori di CTR presumibilmente normali varino in base all'età e al sesso. È stato riportato un aumento del CTR medio da 0.448 a 0.562 in pazienti da 18 anni a 90 anni, con un aumento maggiore nelle donne rispetto agli uomini.

Parole chiave: rapporto cardiotoracico, segmentazione anatomica del torace, raggi X del torace, anatomia tagliata

Abstract

The cardiothoracic ratio (CTR) is a clinical criteria used to estimate heart size and possible linked abnormality, such as cardiomegaly, from chest X-ray (CXR) images. Visual evaluation of CTR in clinical practice is time-consuming and may introduce variation across interpreters. Obtaining the objective measurement of CTR in an automatic way will decrease the subjectivity of the radiologist's evaluation and will give more support to their diagnosis during follow-up examinations. The goal of this thesis was to investigate the automatic measurement of the CTR. Two segmentation-based approaches have been proposed for the calculation of the CTR from CXR images. A first method estimates the CTR based only on lung segmentation, and a second method adds the segmentation of the heart for the estimation of the CTR. These methods were also developed to be robust to the presence of clipped lung on the image, which is one of the quality checks that should be observed before each CXR analysis. The calculation of the CTR from only lung segmentation obtains better results in this thesis when compared to the CTR calculation from both lung and heart segmentations. It shows a mean and a standard deviation of the absolute error of 0.038 ± 0.040 when tested on non-clipped CXRs and 0.058 ± 0.057 when tested on CXRs with clipped lungs. Based on this lung segmentation model, an algorithm to detect the presence of clipped anatomy is proposed, showing promising results. Thanks to the possibility to quickly apply the automatic CTR calculation to a large number of CXRs, a short population study was carried out on 25369 CXRs from the CheXpert dataset. It allows to study how presumably normal CTR values change due to age and gender. An increase in mean CTR from 0.448 to 0.562 from 18-years-old to 90-years-old patients has been reported, with an higher increase for female than male.

Keywords: cardiothoracic ratio, chest anatomy segmentation, chest X-ray, clipped anatomy

Chest X-ray Image Processing: Clipped Anatomy Detection and Cardiothoracic Ratio Estimation

Matilde Bodritti

Supervisors: Filippo Molinari (PoliTo), Jan Aelterman (UGent), Adriyana Danudibroto (Agfa)

***Abstract*—Cardiothoracic ratio (CTR) plays an important role in early detection of cardiac enlargement related disease in chest X-ray (CXR) examinations. However, its measurement in clinical practice is highly subjective and time-consuming. This thesis proposes two segmentation-based approaches for automatic measurement of the CTR based respectively on the estimation of CTR from only the lung segmentation mask and from lung and heart segmentation masks. Both methods are based on modification of the U-Net architecture [1]: a convolutional neural network that was developed specifically for biomedical image segmentation. The proposed methods have been developed to be robust to the presence of clipped lung (one of the many quality control to check before CXR's examination) and to be able to detect such cases. The best proposed method have been applied to a large dataset showing how CTR usually increase with the age of the patient, particularly in women.**

***Index Terms*—cardiothoracic ratio, chest anatomy segmentation, chest X-ray, clipped anatomy**

I. Introduction

Medical images are widely used for disease diagnosis and response monitoring. The history of diagnostic images began with the first radiography, in 1895 when X-rays have been discovered. Although radiology is the oldest imaging technique, it is still widely used nowadays. Chest X-ray (CXR) is the most commonly performed diagnostic X-ray examination [2]. It is used in everyday clinical practice to analyze heart, lungs, blood vessels, airways, ribs and spine. Its huge popularity is due to the fact that it is dose-effective and fast compared to other imaging tools, non-invasive, relatively cheap and high accessible. Moreover, a wide range of pathologies can be identified from CXR evaluation, such as cardiomegaly: a medical condition that refers to an enlargement of the heart [3]. Although the huge progress from 1895, CXR has also been criticized for its low diagnostic sensitivity when compared to cross-sectional techniques [2], that needs to be counterbalanced by an accurate and time-consuming radiologist interpretation.

Since the notoriously difficult interpretation of CXRs, computer-aided technology has been investigated to help clinicians in their diagnoses since the birth of artificial intelligence. The interest in this topic increases recently due to the availability of large amount of data and the development of deep learning techniques [4]. Different levels of automation exist for a system application in the interpretation of medical images. One way to reduce the over-reliance of the clinician on the system, is to develop systems that produced autonomous information instead of autonomous decision, characterized by a separation between what the device and the clinician each contribute to the decision [5]. Instead of outputting directly the disease inferred from the CXR, it is possible to output some measurements (as objective as possible) that will help the clinician to formulate his/her diagnosis. This will also help to overcome interobserver and intraobserver variability in the subjective reading of the CXRs [6].

An example of objective measurement is the cardiothoracic ratio (CTR). CTR is a screening tool which allows to evaluate the size of the heart's silhouette and thus the presence of cardiomegaly from CXR [7]. CXR can be done in different possible projections but the gold-standard for CTR evaluation is the posteroanterior projection (PA), with the X-ray beam passing through the patient from the back to the front. This avoids possible enlargement of cardiac silhouette, since the heart is an anterior structure [8]. The theoretical definition of the CTR involves measuring the maximum horizontal thoracic diameter (Dthorax), measured at the inner edge of ribs, as well as the maximum horizontal heart diameter (Dheart). An example is shown in Figure 1. Consequently, the following formula can be applied:

$$CTR = \frac{D_{heart}}{D_{thorax}}. \tag{1}$$

In clinical practice, a visual analysis of the image is usually considered sufficient to determine the presence of cardiac problems related to the heart size. The precise value of CTR is usually not explicitly calculated primarily for timing reasons. An automatic measurement of CTR would support the clinician's diagnoses and decrease the subjectivity of the evaluation.

Before carrying out such evaluations, radiologists usually perform a qualitative assessment of the image. Specific criteria, listed in the "European guidelines on quality criteria for diagnostic radiographic images" [9], need to be fulfilled. One of those is that the image need to be a *"reproduction of the whole rib cage above the diaphragm"*. The whole lungs need to be in the field of view of the image, otherwise the radiograph is usually rejected.

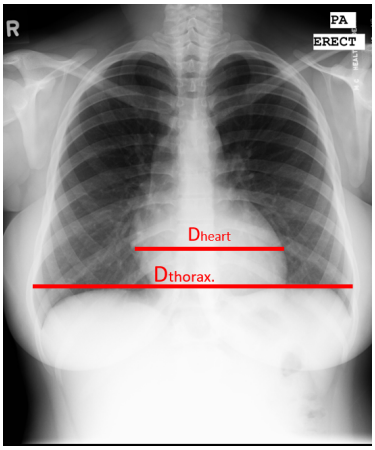

Fig. 1. Measurement of maximum horizontal cardiac diameter (Dheart) and maximum horizontal thoracic diameter(Dthorax) from a CXR.

This study aims to propose a segmentation-based method able to automatically estimate the CTR from CXRs. Most of the publicly available segmentation algorithms are developed to perform tasks with the underlying assumption that the images are taken from correct acquisition. Unlike them, the proposed algorithms, have been developed to be robust to clipped anatomy. In this way, it would be possible to extract information about the CTR even in suboptimal CXRs. Thanks to the possibility of quickly applying the automatic CTR calculation to a large number of CXRs, a population study was carried out, showing how the average CTR increases with age and particularly in women.

II. CTR ESTIMATION

Two different segmentation-based methods have been investigated, based respectively on the calculation of the CTR from lung segmentation mask or lung and heart segmentation masks. The models tried to recover the lung segmentation also from the possible clipped part of the image by assuming their ability to learn the general shape of the lungs.

A. CTR FROM LUNG SEGMENTATION

Starting from the CXR image, the lung segmentation mask is extracted. From the lung segmentation mask, both Dheart and Dthorax are extracted and the CTR is calculated as described in equation 1. The publicly available model from the official implementation of the work by Selvan et. al [10] was used as a starting point to perform the lung segmentation task. It employs a segmentation network similar to the U-Net [1] but adds a Variational Autoencoder [11]. The released model has been trained and validate on a total of 704 images from Shenzhen and Montgomery hospitals [12] and has in output a segmentation mask of the same size of the input. For this reason, a padded version of the image is needed in input to try to reconstruct part of the lungs that are clipped. This model has been trained 3 different times, which differ from the augmentation technique used:

- **Model 1**: with standard augmentation (rotation, flipping);
- **Model 2**: with block masks and diffuse noise masks in addition to standard augmentation;
- **Model 3**: with clipping and padding augmentation in addition to previous augmentation techniques.

The first two models replicate the work from Selvan et al. [10]. Subsequently, some modifications to the U-Net with VAE architecture have been done, to allow the output of the model to have a field of view larger of 128 pixels on each side with respect to the field of view of the input image. This allows the fourth model (trained on the same dataset) to input directly clipped images (not padded) that are more faithful to the realistic ones. The last model can be defined as:

- **Model 4**: U-Net with VAE with wider output with clipping augmentation in addition to augmentation techniques used in model 2.

The four methods have been tested on 247 images from JSRT [13] dataset. Corresponding manually generated lung (and also heart) field masks are provided by Ginneken et al. [14]. Models 1, 2 and 3 have also been tested on a clipped & padded version of the JSRT dataset (manually created), while model 4 has been tested on a clipped (and not padded) version of the JSRT dataset, to test their performances with clipped lungs. The Dice index was used to evaluate the segmentation performances. It measures the degree of overlap between the ground truth mask (G) and the predicted segmentation mask (P) and it is defined as:

$$Dice(G, P) = \frac{2|G \cap P|}{|G| + |P|}. \quad (2)$$

Dice index was calculated for left and right lung separately. The best performance on clipped CXRs have been obtained with model 4, which has also shown good results on non-clipped images. Results are shown in Figure 2.

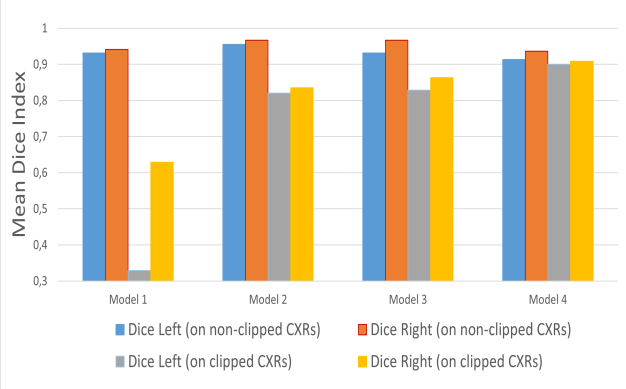

Fig. 2. Dice indexes of the four lung segmentation models developed, tested on clipped and non-clipped JSRT dataset.

From the lung segmentation masks obtained applying model 4, the CTR values of clipped and non-clipped testing images have been calculated, following equation 1. Dheart in this case has been defined as the maximum horizontal distance between the two lungs above the vertex of the cardiophrenic

angle (angle between the heart and the diaphragm) of the right lung. This approximation is done because, for its gold standard definition, the lung segmentation contour should follow the contour of the surrounding anatomical parts including the heart. Dthorax is defined as the maximum horizontal width of the lungs mask. An illustration of CTR calculation from lung segmentation mask is shown in Figure 3. Results are shown in Table 1. Most of the results achieved with the proposed method are in the same order of magnitude of the state-of-art method based on the estimation of CTR from lung segmentation prediction [15].

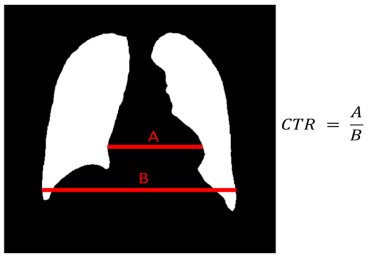

Fig. 3. Illustration of CTR calculation from lung segmentation mask.

B. CTR FROM LUNG AND HEART SEGMENTATION

Since the strict definition of the CTR involves also the heart boundary, improvement in CTR calculation is expected when the cardiac diameter is extracted from the segmentation mask of the heart mask itself, instead of from the lung mask. For this reason, the CTR estimation from lung and heart masks have been investigated. This approach is the same followed in the majority of works in literature related to this topic, such as [16].

Starting from CXR images, the lung and heart segmentation masks are extracted. Two methods have been investigated:

A Lung segmentation masks are obtained applying model 4 described in the previous section. A new model have been trained to output heart segmentation mask: model 2 described in the previous section have been retrain from scratch using CXRs and associated heart masks. With this approach two different U-Net with VAE architectures are used to output heart and lung masks.

B Model 4 described in the previous section was modified to allow multi-label outputs. The already trained lung segmentation model was fine-tuned with heart segmentation masks. With this approach one U-Net with wider output with VAE was used to obtain both lungs and heart segmentation masks.

To train the heart segmentation model from method A and to fine-tune model B, the Wingspan dataset [17] was used. It contained 259 CXRs with lung and heart masks annotation. Both the methods were tested on JSRT dataset. For the heart segmentation, model B was demonstrated to work better in terms of Dice index: it seems to be able to derive information on the relative position between the heart and the lung. Model B was then tested both on clipped and non-clipped JSRT images, showing a decrease in the heart segmentation performance in clipped lung images, supporting the hypothesis that, in this case, the segmentation of the lung also influences the segmentation of the heart.

Subsequently, the CTR was calculated from lung and heart segmentation masks obtained from model B. The performances are reported in Table 1 and compared with the results from the previous section. Absolute error and root mean square error get worse when also the heart segmentation is considered in the estimation of the CTR. This results are not in line with the initial hypothesis. However, it was possible to notice that heart masks of JSRT dataset are more ”circular shaped” when compared to the Wingspan’s heart masks that are more ”triangular shaped”. This could be due to the radiologists using different way of performing the annotation and could be the main reason of this decrease in performance.

To overcome this issue a different metric related to the CTR have been proposed: the Rotational Invariant CTR (**RI _CTR**). It is defined as the CTR in equation 1, but Dheart and Dthorax have a different meaning. Dheart is now defined as the diameter of the maximum circle inscribed in the heart mask, while Dthorax is now defined as the maximum horizontal width of the rotated lung. Lung have been rotated to remove possible wrong orientation of the patient. Performance of this metric are reported in Table 1. **RI _CTR** seems to be less dependent on different ways of annotating heart masks, since correlation coefficient with the RI_CTR calculated from ground truth segmentation masks is much higher. However, its clinical relevance is yet to be observed and could be the direction of future works. For this reason, the most promising results in this thesis for the CTR calculation have been shown by the calculation of the CTR only from lung segmentation masks. An illustration of both CTR and RI_CTR calculation from lung and heart segmentation masks is shown in Figure 4.

TABLE 1

	AE (mean $\pm$ std)	RMSE	Corr. Coeff.
CTR Lung	0.038 $\pm$ 0.040	0.055	0.677
CTR Lungs and Heart	0.062 $\pm$ 0.059	0.086	0.558
RI_CTR	0.024 $\pm$ 0.021	0.031	0.754

Performance of different methods used for CTR estimation on JSRT dataset in terms of Absolute Error (AE), Root Mean Square Error (RMSE) and Correlation Coefficient.

III. CLIPPING DETECTION

Assuming a clinical application of the automatic CTR measurement, informing clinicians if the predicted CTR comes from clipped or non-clipped CXRs would be useful information to give an idea about the reliability of the estimated CTR. For this reason, a possible method to automatically detect the presence of clipped lung in CXRs has been proposed by exploiting the ability of the U-Net with VAE with wider output.

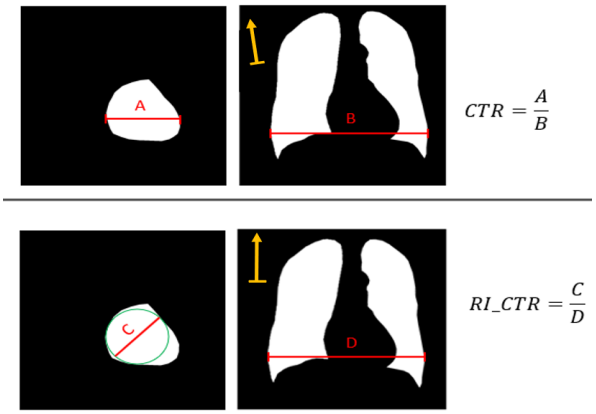

Fig. 4. Illustration of CTR and RI_CTR calculation from lung and heart segmentation masks. The yellow arrow represents the mayor axis of the lung mask.

From the binary segmentation of the lung obtained from model 4 described previously, the bounding boxes of both the lungs were obtained. The distances of the corresponding bounding boxes from the border of the initial field of view of the input image is calculated. The maximum distance found have been compared with a fixed threshold to classify if the lung is clipped or not. The best threshold has been found on a training dataset formed by Shenzhen and Montgomery datasets, 50% of which have been randomly clipped to simulate clipped lungs. When tested on a dataset formed by clipped and non-clipped JSRT images, this method obtained an accuracy of 0.97.

IV. HOW CTR CHANGES WITH AGE AND GENDER

It is generally accepted that the upper limit of normal heart size corresponds to the 50% of thoracic size. However, available literature demonstrates that the CTR threshold for cardiomegaly can vary when related to age and gender. These types of population studies are usually difficult to carry out on large scale, because of the need of clear and structured radiologists annotations for each CXR. The automatic calculation of CTR can make this process faster and easily accessible.

The best method discussed in this thesis to predict CTR from CXRs was model 4 described in subsection II.A. This method has been applied on a subset of the CheXpert dataset [18]: a large public dataset with CXRs labeled for the presence of 14 observations. In this subset (called in this research CheXpertCTR dataset):

- only PA images are extracted,
- images labeled as positive to cardiomegaly and enlarged cardiomediastinum are excluded.

It results in 25369 CXRs with 34% female and 66% male, from 18 to 90-year-old. The mean age is 56.5±17.1.

From the CheXpertCTR dataset an average CTR of 0.498 ± 0.089 was reported. A general increase in mean CTR is reported as the age of the patients increases: from 0.448 in 18-year-old to 0.562 in 90-year-old patients, showing an increase of the 25%. A mean value and a standard deviation of 0.507 ± 0.094 and 0.492 ± 0.085 has been reported for female and male respectively, showing a slightly higher mean CTR for females. This difference has been demonstrated to be significant (p-value < 0.5). Since a significant difference was reported, CXRs of males and females have been considered separately. Figure 5 shows how the predicted CTR changes according to age and gender. Both the trends appear to be fairly linear, showing similar mean values for younger patients, while females reach higher mean CTR values with increasing age.

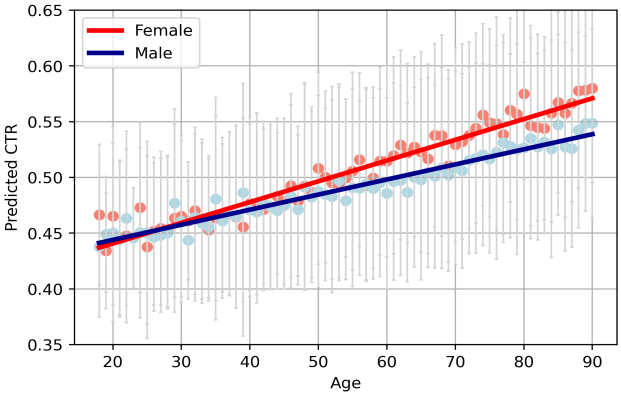

Fig. 5. Predicted CTR as a function of patient age on female and male CXRs from CheXpertCTR dataset.

The results obtained reflects the observations of preliminary studies on age and gender dependency of CTR measurements. In their work, Brakohiapa et al. [19], reported a significant difference in the overall CTR between male and female too, with a slightly higher mean CTR for female. A comparable result has been reported by [6]. In addition, Brakohiapa et al. [19] reported a higher increase in mean CTR values for female as age increases, which also reflects the obtained results.

V. CONCLUSION

This thesis proposed a method to automatically calculate CTR from CXR images. The best method found in this research is based only on lung segmentation. The obtained results have errors in the same order of magnitude of the errors of the state-of-art methods based on the estimation of CTR from lung segmentation prediction, with the advantage that the proposed method also deals with clipped anatomy. A promising method to detect such clipped anatomy cases has been proposed. Moreover, the automatic measurement of the CTR has allowed to easily perform a population study on a large dataset. This study reflects that the proposed method could be suitable for such population studies and shows how choosing a fixed and unique threshold to detect cardiomegaly would not be consistent with the normal variation that the CTR shows to have between age and genders.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[2] C. Schaefer-Prokop, U. Neitzel, H. Venema, M. Uffmann, and M. Prokop, "Digital chest radiography: An update on modern technology, dose containment and control of image quality," *European radiology*, vol. 18, pp. 1818–30, 05 2008.

[3] M. A. Jones, J., "Chest radiograph," *Reference article, Radiopaedia.org.*

[4] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. http://www.deeplearningbook.org.

[5] M. Sujan, C. Baber, P. Salomon, R. Pool, N. Chozos, C. Aceves-González, M. Cooke, C. Escobar-Galino, C. Flashman, G. Frau, R. Hawkins, H. Hughes, G. Mejia, G. Kaya, B. Kirby, I. C. Landa-Avila, K. Laudanski, P. Lewis, F. Magrabi, and S. White, "Human factors and ergonomics in healthcare ai," 09 2021.

[6] E. Obikili and I. Okoye, "Transverse cardiac diameter in frontal chest radiographs of a normal adult nigerian population," *Nigerian Journal of Medicine*, vol. 14, no. 3, pp. 295–298, 2005.

[7] K. Truszkiewicz, R. Poreba, and P. Gać, "Radiological cardiothoracic ratio in evidence-based medicine," *Journal of Clinical Medicine*, vol. 10, no. 9, p. 2016, 2021.

[8] D. G. Lloyd-Jones, "Chest x-ray quality projection." Salisbury NHS Foundation Trust UK - www.radiologymasterclass.co.uk.

[9] J. H. E. Carmichael, "European guidelines on quality criteriafor diagnostic radiographic images," Officer for Official Publication of the European Communities, 1996.

[10] R. Selvan, E. B. Dam, N. S. Detlefsen, S. Rischel, K. Sheng, M. Nielsen, and A. Pai, "Lung segmentation from chest x-rays using variational data imputation," 2020.

[11] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013.

[12] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, 2014.

[13] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000. PMID: 10628457.

[14] B. van Ginneken, M. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006.

[15] A. H. Dallal, C. Agarwal, M. R. Arbabshirani, A. Patel, and G. Moore, "Automatic estimation of heart boundaries and cardiothoracic ratio from chest x-ray images," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, pp. 134–143, SPIE, 2017.

[16] E. Sogancioglu, K. Murphy, E. Calli, E. Scholten, S. Schalekamp, and B. Ginneken, "Cardiomegaly detection on chest radiographs: Segmentation versus classification," *IEEE Access*, vol. PP, pp. 1–1, 05 2020.

[17] N. Dong, M. C. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. P. Xing, "Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio," in *MICCAI*, 2018.

[18] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 590–597, 2019.

[19] E. K. Brakohiapa, B. O. Botwe, and B. D. Sarkodie, "Gender and age differences in cardiac size parameters of ghanaian adults: Can one parameter fit all? part two," *Ethiopian Journal of Health Sciences*, vol. 31, no. 3, 2021.

Contents

List of Figures

List of Tables

List of Abbreviations

AE Absolute Error

AI Artificial Intelligence

AP Anteroposterior

AUC Area Under the ROC Curve

CAD Computer-aided Design

CTR Cardiothoracic Ratio

CXR Chest X-ray

Dheart Diameter of the heart

Dthorax Diameter of the thorax

FN False Negative

FP False Positive

FPR False Positive Rate

JSRT Japanese Society of Radiological Technology

MAE Mean Absolute Error

PA Posteroanterior

PE Percentage Error

RICTR Rotational Invariant Cardiothoracic Ratio

RMSE Root Man Square Error

ROC Receiver Operating Characteristic

ROI Region Of Interest

SCR Segmentation in Chest Radiographs

TPR True Positive Rate

VAE Variational Autoencoder

Chapter 1

Introduction

1.1 Motivation

Chest radiography is the most commonly used modality in clinical practice to detect lung and heart pathologies. However, it is one of the most difficult to interpret exams. For this reason, many studies are focused on automatic interpretation of chest X-rays (CXRs) to help clinician's evaluation by showing directly the detected disease. However, such systems can cause an over-reliance on the technology and the tendency to blindly accept the algorithm's output. A possible way to cope with this problem is to provide clinicians with algorithms that output objective measurements from CXRs, leaving the interpretation to them and increasing the interaction between their opinion and the algorithm's findings. One of the most important measurements from CXRs is the calculation of the cardiothoracic ratio (CTR). Increased CTR often indicates abnormalities and cardiac enlargement related diseases. The automatic estimation of this objective measurement from CXR would be beneficial for a number of aspects:

- helping the clinician in the diagnosis of the patient;
- helping the clinician during the follow-up of the patient by allowing an easily comparison;
- decreasing the subjectivity of the evaluation;
- allowing the calculation of such measurements in a short time, enabling population studies on large datasets.

Moreover, a quality check of the image is always done by first the radiographers and then by the radiologists prior to any clinical interpretation and specifically in this case to assure that the whole lungs are in the field of view of the image, otherwise the radiograph is usually rejected and repeated. A CTR estimation algorithm robust to clipped anatomy would allow the extraction of useful information also from such suboptimal CXRs.

1.2 Objective

This thesis focuses on the automatic calculation of the CTR from CXRs. The aim is to propose a method that minimises the CTR error and that is able to estimate the CTR when clipped anatomy occurs in addition to detecting such cases. The last goal is to apply the proposed method on a large dataset to perform a short population study evaluating how the CTR changes with age and gender.

1.3 Description of the remaining chapters

The thesis is organized as follows.
Chapter 2 describes the fundamental concepts underlying radiographic images focusing then on CXRs. The problem of objective measurements obtained from CXRs is addressed and the definition of CTR measurements is given. An overview of works related with the topic of this thesis is presented, together with a list of publicly available datasets related to the problem of anatomical segmentation for CTR estimation. **Chapter 3** describes four different methods to obtain lung segmentation mask and how to calculate CTR using only the lung segmentation. The best method that addresses also the problem of clipped anatomy is determined. In **Chapter 4** the segmentation of the heart is introduced since it was expected to improve the performance of CTR estimation. An alternative metric correlated to the CTR is also proposed. Subsequently in **Chapter 5** a method to detect the presence of clipped anatomy is discussed, based on lung segmentation model described in Chapter 3. **Chapter 6** describes the application of the automatic CTR measurement on a large dataset, showing how normal CTR values change with the age and gender of the patient. Lastly, conclusions can be found in **Chapter 7**, while some supplementary information about ethical aspects that apply to this research is described in **Appendix A**.

Chapter 2

Background

As mentioned in Chapter 1, the work presented here is focused on performing a measurement in CXR images. Before going into the technical details, some background information related to CXR imaging is presented in this chapter. First, general information about medical imaging is presented, followed by CXR in particular. Then, the definition of the measurement is laid out. Lastly, a list of related research is reviewed.

2.1 Chest X-ray Imaging

2.1.1 Introduction to medical imaging

A medical image is defined as a figure related to the anatomy or physiology of internal body parts. Various technological and physical approaches are now employed to obtain images of living creatures' anatomical structures for diagnostic purposes. Diagnostic imaging is the medical discipline concerned with the application of these techniques. Diagnostic images are utilised for two distinct purposes: disease diagnosis and response monitoring [11]. In specific situations, diagnostic images are used for screening applications, namely for identifying diseases prior to clinical manifestations, i.e. before the development of symptoms and signs.

The history of diagnostic imaging began in 1895, when German scientist Wilhelm Conrad Röntgen discovered X-rays (1845-1923) [12]. Radiography, the diagnostic imaging technology based on X-rays, was the beginning of a series of discoveries and inventions that culminated in the 1970s with the development of the first diagnostic image reconstruction system based on computers. Godfrey N. Hounsfield developed this technique, which was known as computed axial tomography (usually abbreviated as CT, computed tomography) [12]. Numerous medical imaging techniques are now available apart from conventional radiography and CT, such as magnetic resonance imaging (MRI), ultrasound, nuclear medicine techniques, positron emission tomography (PET) and single photon emission computed tomography (SPECT). All imaging tools have one common element: they appropriately exploit the interaction of some kind of energy with the human body. This

interaction can be of different types: absorption, diffusion or reflection and the medical image will represent only a partial view of reality, resulting from this specific interaction between a form of energy and a tissue property.

Computers have played such an important part in the field of medical imaging that it is now almost completely dependent on computer technology for data collecting and processing, as well as patient data management, storage, retrieval, and transfer. Computers can also help with the interpretation of diagnostic images, which is still primarily done by expert examiners. In practice, there is no field of clinical care where diagnostic imaging does not provide useful information.

2.1.2 The physics of radiology

Although radiology is the oldest imaging technique, it is still the most widely used today. X-rays are a type of ionising radiation. Radiation is defined as the emission or transmission of energy in the form of waves or particles through space. A radiation is said to be ionising if it possesses the ability to break atomic and molecular bonds of the target cell and release energy. The energy released by ionising radiation inside the body is expressed by the dose: this is measured in *Gray* ($1\ Gy = 1\ J/kg$) [13]. However, since not all types of radiation produce the same biological effect, the effective dose is often used instead of the absorbed dose, and it is measured in millisievert (mSv) [13]. It takes into account the different sensitivity of tissues to radiation. The effective doses in the most commonly used X-ray examinations (bone, chest, mammography, digestive system, etc.) are between 0.05 and 3 mS, while in highly demanding examinations such as CT scans of large districts (chest, abdomen) or arteriography effective doses are several times higher [14].

X-rays are ionising radiations since they are high-energy electromagnetic waves. Wavelength of X-rays is much shorter than for example that of radio or visible light waves and belongs to the band between 0.001 and 10 nm (as shown on Figure 2.1). Such short wavelength is related to an high energy thanks to Planck's hypothesis [15]. For Plank's hypothesis, all the electromagnetic radiation, such as the X-rays, occur in finite "bundles" of energy called photons. Each photon has a precise energy (E) given by the product between the Plank's constant h= 6.96261 x $10^{-}34$ Js and the frequency of the waves υ. This relationship can be summarized in the following equation:

$$E = \frac{hc}{\lambda} = h\upsilon [16], \tag{2.1}$$

where $c = 3x10^8\ m/s$ is the speed of light and λ is the wavelength.

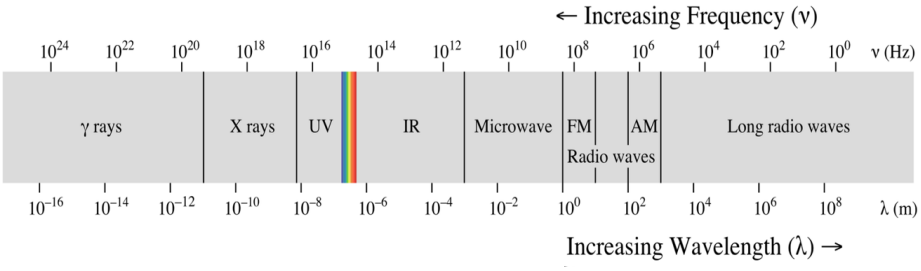

Figure 2.1: Electromagnetic spectrum [1].

When an X-ray passes through a material with thickness s there is a loss of this incoming energy intensity I_0 by absorption:

$$I = I_0 exp(-\mu s). \tag{2.2}$$

Equation 2.2 is known as the Lambert-Beer law [17] and it is at the base of imaging with X-rays. μ represents the linear attenuation coefficient. It is mostly not uniform: it is the product of the density and the mass-absorption coefficient, which depend on the local element composition. This loss of energy from the X-ray beam is what happened when it pass trough the body during an X-ray examination. The usual X-ray set up is shown in Figure 2.2. The X-ray beam, created from an X-ray source is directed trough the patient, which is positioned in front of an X-ray panel. The X-ray panel will collect the remaining intensity as the beam passes through the patient [18]. From the collected remaining intensity, it is possible to reconstruct the mapping of the attenuation coefficients, creating the X-ray image. The intensity of the beam, in fact, will gradually decrease when passing through material due to the attenuation mainly caused by photoelectric absorption and Compton scattering.

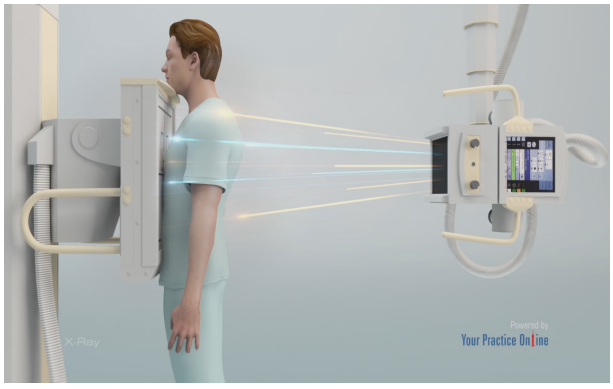

Figure 2.2: Illustration of the set up of an X-ray exam [2].

Low energy photons will be almost completely absorbed by the patient and will not contribute to the projection image. They will contribute only to the dose of the patient. Therefore these low energy photons are preferentially absorbed before hitting the patient in order to reduce the dose. This is achieved by placing a plate of dense material (typically

aluminium) between the X-ray and the patient, called filter. A second component also reduces the dose to the patient: the collimator. It limits the area of the patient that is irradiated.

The effect of X-rays on matter is ionisation of atoms with formation of free radicals. In a resting condition, most body tissues are neutral. When irradiated with X-rays, the presence of induced charge can be seen. The presence of charged ions in the circulation, if in large amounts, can cause:

- long-term effects: examples of long-term effects can be solid tumors (somatic cell mutations), non-solid tumors (lymphomas, leukemias), germline mutations.
- acute effects: immediately occurring effects from high radiation include cell necrosis (especially on the skin), premature aging, and death.

Every individual is capable of non-hazardously absorbing a certain dose, which is internationally normed. X-rays used with diagnostic purpose caused a limited dose, that is not dangerous for the patient. However, radiologists should always follow the ALARA principle: patients should always receive radiation doses "As Low As Reasonably Achievable" [19].

2.1.3 Chest radiography

The chest radiography is used in everyday clinical practice to analyze heart, lungs, blood vessels, airways, ribs and spine. It is the most commonly performed diagnostic X-ray examination: it represent around 30-40% of all X-ray examinations performed [20].
The main advantages of using CXRs are:

- the speed at which they can be acquired;
- the wide range of pathologies that can be identified [21];
- the low cost;
- being non-invasive;
- being dose-effective compared to other imaging tools [22]: a single chest X-ray exposes the patient to about 0.1 mSv, which is about the same amount of radiation people are exposed to naturally in about 10 days [23];
- the high accessibility (e.g. under-resourced regions of the world that also have to face a heavy burden of infectious diseases, such as tuberculosis (TB), commonly use CXR as frontline diagnostic imaging due to lower infrastructure setup, operational costs, and portability [24]).

From 1895, CXR has faced a huge progress but has also been criticized for its low diagnostic sensitivity, when compared to cross-sectional techniques [20] which have diagnostic superiority and increasing availability. The limitation that comes from the use of such a simple technology, needs to be counterbalanced by an accurate, detailed, and time-consuming radiologist interpretation. CXRs are, in fact, one of the most complex imaging modalities to interpret and the interpretation is linked to the level of training and experience of the physician: in one study it was demonstrated how discrepancy rate was higher in less experienced physicians [25]. Since the clinical outcome is based on the complete understanding of the CXR, emerging computerized tools try to help in improving diagnosis and simplify clinicians work. These computer-aided diagnosis (CADx) systems try to automatically interpret CXRs and assist medical practitioners in decision-making.

2.1.4 CXR views

CXR is usually done in different possible views or projections depending on the relative position between the patient and the X-ray beam. Most common positions are:

- anteroposterior projection (AP): it can be performed with the patient in an erect position but also on the bed. X-rays pass from the anterior (the front) to the posterior (the back) of the patient. Since the heart is an anterior structure within the chest, it will be magnified by an AP view [26]. The same occurs with the mediastinum. For bedside imaging, this view has the advantage to be performed also outside the radiology department thanks to mobile X-ray units.

- posteroanterior projection (PA): it is the standard for CXRs. It is performed with the patient in an erect position with the X-ray beam passing through the patient from posterior (the back) to anterior (the front). Patient is in full inspiration, hugging the detector to keep them from overlapping with lung field. It allows a technically excellent visualisation of the mediastinum and lungs. In this case, magnification will not occur resulting in an accurate assessment of heart size [26].

- lateral projection: performed with the patient standing upright with the left side of the thorax adjacent to the image receptor. The patient is asked to raise hands over their head. It may be performed as an adjunct in cases where there is diagnostic uncertainty. The lateral chest view can be particularly useful in assessing the retrosternal and retrocardiac airspaces [21].

An example of the different views is shown in Figure 2.3.

Figure 2.3: From left to right: AP view chest radiograph, PA view chest radiograph, lateral view chest radiograph.

2.1.5 Chest Anatomy in CXRs

When approaching a CXR, either frontal (AP/PA) or lateral projection, a systematic approach need to be used to evaluate the visual appearance of all the anatomical structures. A popular method to do this is called the 'ABCs' rules. The following anatomical compartment can be analyze from a CXR [27, 28, 29]:

A. Airway: airway can be inspected analyzing the trachea and mainstream bronchi. Airways have a lower density if compared with the surrounding soft tissues because contain air. Thus, they appear darker on a CXR.
B. Bones: the bones visible on a CXR include clavicles, ribs, part of the spine, scapula and the proximal humeri [3].
C. Cardiomediastinal region: defined as the area between the lungs, formed by blood vessels, trachea, muscular esophagus, thymus gland and the heart. Most of the cadiomediastinal region are not clearly visible on CXR, except the heart. Cardiomediastinal profile is important to diagnose various diseases and to assess the size and the contour of the cardiac silhouette.
D. Diaphragm: the diaphragm's shape can also reveal a significant information about a patient's present health. The right hemidiaphragm is higher than the left because it rests on top of the liver. The diaphragm is typically curved. The patient may have persistent asthma or chronic obstructive lung disease if the diaphragm appears flattened [3].
E. Edges: it should be easy to see the cardiophrenic and costophrenic angles. The costophrenic angles are formed by contours of the chest wall and diaphragm. On the frontal CXR the costophrenic angles should form acute angles. The area in the lower edges of the lungs which contact the diaphragm is called costophrenic recesses. The cardiophrenic angle is the angle between the heart and the diaphragm.
F. Fields, Foreign bodies: the presence of opaque masses, consolidation, or fluid can be checked on the lung field.
G. Great vessels, gastric bubbles: aorta and pulmonary vessels can be assess. The aortic knob should be visible. Under the left diaphragm, a typical gastric bubble is usually

visible.

H. Hilium: it is what connects the lungs to their supporting structures and where pulmonary vessels enter and exit the lungs. The left hilum is normally higher than the right one.

The visualisation of different structures on CXRs is shown on Figure 2.4.

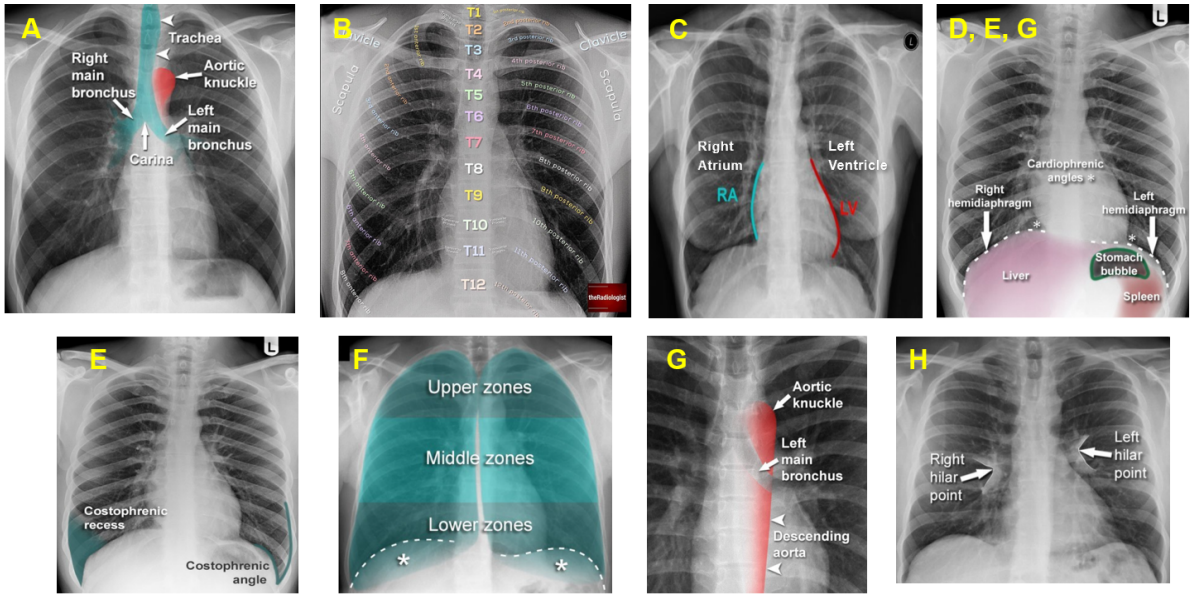

Figure 2.4: Visualisation of different structures on CXRs: Airways (A), Bones (B), Cardiac silhouette (C), Diaphragm / Edges: cardiophrenic angles / Gastric bubble (D,E,G), Edges: costophrenic angles (E), Field of the lung (F), Grand vessels and aortic knocke (G), Hilium (H) [3].

2.1.6 Lung and heart shapes on CXR

Since the lung and the heart are the organs of interest for this thesis, a more in depth analysis of their shape on CXR is presented. Right and left lungs are located on either side of the heart near the backbone. The heart is usually tilted to the left: for this reason, the left lung has an indentation called the cardiac impression to accommodate the heart. This will lead to the left lung to be smaller, narrower and longer compared to the right lung which is wider and shorter [30]. Another difference between the two lungs is their base: the right lung's base is more concave than that of the left lung [30]. Furthermore, the human heart can appear as a variety of shapes as for example elliptical, round, conical or trapezoidal. The shape of the cardiac silhouette can also be used as clues to the underlying disease: a "water bottle" configuration can be linked with pericardial effusion or generalized cardiomyopathy; left ventricular or "Shmoo" configuration describes

lengthening and rounding of the left heart border with a downward extension of the apex causing a left ventricular enlargement; "straightening" of the left heart border is linked with rheumatic heart disease and mitral stenosis [31]. Examples of these possible cardiac shapes are shown in Figure 2.5

Figure 2.5: Three types of abnormal cardiac silhouette. From left to right: water bottle silhouette [4], "Shmoo" silhouette [5], "straightening" of the left heart border [6]

It is important to notice how a large variability in the heart silhouette reflects a large variability in the lung silhouette, with more impact on the left lung, for the reasons explained before.

Despite what normally occurs, in some patients the heart can be positioned on the right side because of abnormalities such as dextrocardia or dextroposition. Dextrocardia is an intrinsic cardiac positional anomaly in which the heart is located in the right hemithorax with its base-to-apex axis directed to the right. Dextrocardia is often misinterpreted as cardiac dextroposition, which also refers to a displacement of the heart to the right, but in this case it is due to extracardiac abnormalities such as right lung hypoplasia, right pneumonectomy, or diaphragmatic hernia [32].

2.1.7 Cardiomegaly

A large variety of cardiothoracic abnormality can be observed from a CXR which are mainly heart and lung pathologies, e.g., atelectasis, consolidation, pneumothorax, pleural and pericardial effusion, cardiac hypertrophy and hyperinflation [22]. Many of these pathologies are clearly visible due to the deformation of heart and lung region: structural cardiac abnormalities and cardiac enlargement has been shown to be associated with functional status and adverse clinical outcomes [33]. "Cardiomegaly" is the term that refers to an enlarged heart. Cardiac enlargement can either refer to the dilation of a heart chamber or the hypertrophy of the heart muscle. If the heart chamber dilates, the heart muscle is stretched, causing the chamber to grow larger. In cardiac hypertrophy, the heart's muscular fibers actually increase in size, which causes the chamber to enlarge. The overall number of heart muscle fibers does not grow during cardiac hypertrophy; rather, each fiber

gets bigger [34]. The cardiac enlargement itself is not a disease but rather a sign of possible conditions such as: congenital heart defect, damage from a heart attack, cardiomyopathy, pericardial effusion, heart valve disease, hypertension, pulmonary hypertension, anemia, thyroid disorders, hemochromatosis or cardiac amyloidosis [35]. An enlarged heart can lead to high risk of complication such as heart failure, blood clots, leaky heart valve or even cardiac arrest [35]. In Figure 2.6 an example of CXR of a patient with enlarged heart is shown with the follow-up for the same patient after medication that reveal a restoration of a normal heart size.

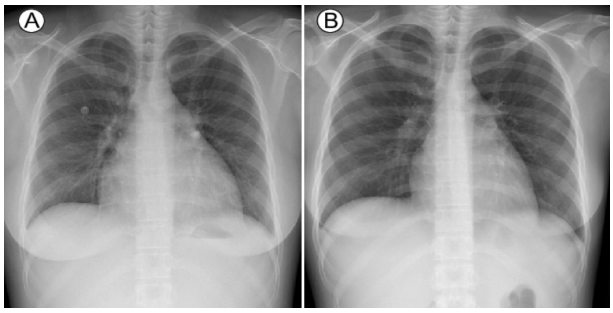

Figure 2.6: (A) A chest radiograph taken on admission revealed mild cardiomegaly and pulmonary congestion. (B) After methylprednisolone pulse therapy, chest radiography showed that the heart was of normal size and the lung parenchyma of normal appearance [7].

Usually to detect cardiomegaly the CTR is evaluated: it refers to the transverse diameter of the heart, relative to that of the rib cage. Although there is still no general consensus, in the literature an indicative threshold of 0.5 is often used to detect the presence of cardiomegaly, regardless of the patient's age, gender and race [36].

2.2 Measurements on Chest X-rays

Since the notoriously difficult interpretation of CXRs, computer-aided technology has been investigated to help clinicians in the diagnoses since the birth of artificial intelligence (AI). The first articles about computer analysis of radiographic images appeared in the 1960s while techniques specifically designed for computerized detection of abnormalities in chest radiographs began to appear in the 1970s [37]. Nowadays, due to the availability of large amount of data and the development of deep learning techniques [38], interest has been stimulated in applying deep learning to medical imaging tasks (including cardiomegaly detection).

Studies have shown that the detection accuracy for the chest disease is improved when using X-ray CAD system as an assistant for the radiologist [39]. However, there are many side effect to radiologists relying too heavily on new AI technologies. A study has doc-

umented possible "automation bias" effects in CAD that degrade radiologists' decision making [40]. Automation bias is referred to the over-reliance on the technology and the tendency to blindly accept AI output. The danger is when the system does not generalize well to unusual cases. This reliance on technology may reduce attention and perceptive skills of clinicians [40]. It is indeed important that technology's design supports an high level of interaction between them and the system. The AI should supplement and not replace their work. Since different levels of automation exists for a system application in clinical field, one way to reduce the automation bias is to use systems that produce autonomous information instead of autonomous decision, characterized by a separation between what the device and the clinician each contribute to the decision [40]. For example, instead of outputting directly the disease inferred from the CXR, it is possible to output some measurements (as objective as possible) that will help the clinician to formulate his/her diagnosis. This will also help the clinicians to overcome interobserver and intraobserver variability in the subjective reading of the CXRs [41]. Some examples of objective measurements are the cardiac transverse diameter, the cardiac volume using both the frontal and lateral view, or the CTR. The CTR, in particular, as mentioned in the previous subsection, refers to the size of the heart compared to the size of the thoracic cavity. Reporting only the CTR is an example of autonomous information, while reporting the clinical diagnosis linked with this information (that in this case would be the presence of cardiomegaly) is an example of autonomous decision.

Later in this chapter, the prerequisites to extract valuable measurements from a CXR are presented, followed by a clinical and a more theoretical approaches for the calculation of the CTR.

2.2.1 Quality checks on CXRs: clipped lung detection

The evaluation of the image quality is critical before interpreting a CXR. A deviation from quality standards may lead to misdiagnosis and hold legal risk [42]. Specific criteria are listed in the "European guidelines on quality criteria for diagnostic radiographic images" [43]. It reports that the image has to fulfill the following criteria:

- *"performed at full inspiration (as assessed by the position of the ribs above the diaphragm — either 6 anteriorly or 10 posteriorly) and with suspended respiration"*,
- *"symmetrical reproduction of the thorax as shown by central position of the spinous process between the medial ends of the clavicles"*,
- *"medial border of the scapulae to be outside the lung fields"*,
- *"reproduction of the whole rib cage above the diaphragm"*,
- *"visually sharp reproduction of the vascular pattern in the whole lung, particularly the peripheral vessels"*,

- *"visually sharp reproduction of the trachea and proximal bronchi, the borders of the heart and aorta, the diaphragm and lateral costo-phrenic angles"*,
- *"visualisation of the retrocardiac lung and the mediastinum"*,
- *"visualisation of the spine through the heart shadow"*.

Usually, radiographers execute the quality assessment procedure visually, which is inherently subjective. Since the visual quality judgement process is subjective, radiographic technologists may disagree on whether to reject and retake an image [44]. A method to automatically quantify the quality of the CXRs has been proposed by Von Berg et al. [42], where the aspect of collimation, patient rotation and inhalation state of PA CXRs were taken into account as defined as the three most relevant quality aspects. These factors are particularly important for the evaluation of the CTR. For this measurement, a PA projection is the standard projection used, for the reasons explained in Section 2.1.4. It is possible to evaluate the CTR also from AP projection, but it is needed to take into account that the heart will appear enlarged due to magnification effect. Moreover, wrong positioning of the patient may cause the repetition of the acquisition when part of the Region Of Interest (ROI) is outside the field of view of the image resulting in clipped anatomy. For CXRs, the ROI is the whole lung: as stated before in this subsection, the CXR should be a reproduction of the whole rib cage above the diaphragm. In clinical practice, no information is extracted from a clipped CXR and the acquisition is always retaken [44, 45].

2.2.2 CTR measurements

CTR is a screening tool to evaluate the size of the heart's silhouette. The most straightforward definition of the CTR involves measuring the maximum horizontal thoracic diameter (D_{thorax}), measured at the inner edge of ribs, as well as the maximum horizontal heart diameter (D_{heart}) [8, 46, 45]. An example is shown in Figure 2.7. After recognizing the contours of these two anatomical structures, the following formula can be applied:

$$CTR = \frac{D_{heart}}{D_{thorax}}. \tag{2.3}$$

The range of the CTR values that indicates a normal condition are usually between 0.42 and 0.50 [46]. Values greater than 0.50 are generally linked with pathological condition such as cardiomegaly, even if there is no consensus for optimal threshold [22], as mentioned in Section 2.1.7. In fact, numerous criticisms have been made to the use of CTR as cardiomegaly indicator itself. It has been demonstrated by Brakohiapa et al. [36] that the CTR threshold for cardiomegaly can vary when related to age and gender. In their study, the mean CTR increases gradually with age and males usually have a slightly lower mean CTR than females. In addition, it is important to remember that CXR are only a

2D section of 3D structures: other advanced methods exist and can provide more useful detailed information about heart size (such as CT). Despite the criticisms, CXRs are still used as the easiest way to detect cardiomegaly thanks to the advantages described in Section 2.1.3.

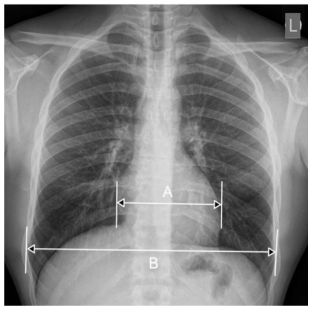

Figure 2.7: Horizontal heart diameter (A) and horizontal thoracic diameter (B) measured on a PA CXR for CTR calculation [8].

2.2.3 CTR in clinical practice

In everyday clinical practice, the precise value of CTR is usually not calculated. A visual analysis of the image and a great deal of knowledge in CXR evaluation turns out to be sufficient to determine the presence or absence of cardiac problems related to the heart size. Radiologists usually perform a qualitative evaluation of the image, taking into account the history of the patient. The follow-up of the patient and the evaluation of how the size of the heart has changed over time has more clinical relevance than the CTR value. Even if this specific CTR value is not calculated in clinical practice mainly due to timing reasons, it is still clinically interesting to calculate this measurement in an automatic and fast way. It can be useful in clinical practice and may lead to certain advantages. Besides being time efficient for the radiologists, showing the actual value of the CTR helps radiologists to make more objective comparison in patient follow-up. A comparison between the CTR value detected in the follow-up of the patient and the CTR value from the previous examination would result in a more objective evaluation of how the size of the heart has changed through time. This could lead to a decreased subjectivity of the evaluation, when conducted by different radiologists.

2.3 Related Research

Earlier methods for automated CTR calculation used traditional image manipulation techniques.

Becker et al. [47] in 1964 proposed to estimate the maximum diameters of heart and rib cage by vertical intensity histogram analysis. Later, almost all the works regarding CTR estimation were segmentation-based solution, where heart and lung boundary were de-

tected [46, 33, 45]. It was demonstrated that the correctness of the CTR value is directly proportional to the correctness of the segmentation [48]. Some works as [49, 50], based the calculation only on the segmentation of the lung region, with the assumption that it is possible to extract heart's shape information from the lung contour. The gold standard for lung boundary detection, in fact, follows the contour of the surrounding anatomical parts including the heart (as shown in Figure 2.8). From the segmented lung field, it is possible to detect the vertex of the cardiophrenic angle (see Section 2.1.4), usually better detectable on the right lung. The maximum diameter of the heart, D_{heart} can be searched above this point (point C on the right lung in Figure 2.8), as the maximum distance between the two lungs. The maximum diameter of the thorax, D_{thorax} can be calculated from maximum horizontal width of the lungs segmentation mask.

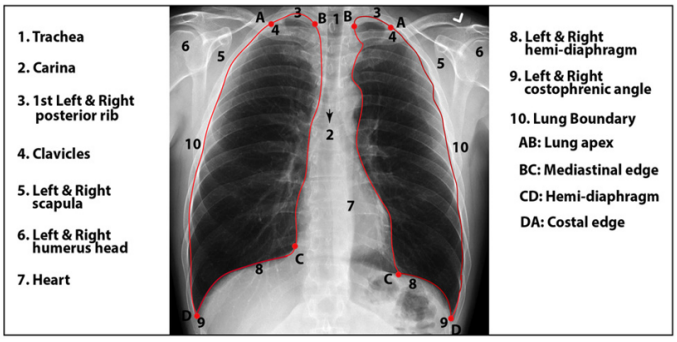

Figure 2.8: Example of lung boundary definition [9].

Nowadays, deep learning (as a type of machine learning) is leading medical imaging analysis. Convolutional neural networks (CNNs), as a popular deep learning technique, has demonstrated to be a powerful method for image processing [33]. It automatically learn mid- and high-level abstractions derived from unprocessed data. Many works are based on the application of deep learning to segment anatomical structures used for CTR estimation. Between deep learning methods, U-Net[10] models show excellent results to extract anatomical boundaries: they use expanding and contracting paths, where an encoder part performs feature extraction from the input, while a decoder part reconstruct the output mask processing the features, as shown in Figure 2.9. These pathways are also linked with each other with skipped connections: these are perhaps the most innovative component of U-Net, that enable the network to restore spatial information that was lost during downsampling operations. With the promising performance of medical image segmentation with U-Net, a series of CTR estimation approaches using U-Net to segment lung and heart fields has followed. The first was done by Que et al. where they used the classical U-Net on a limited dataset, showing promising results. Inspired by that approach, Li et al. [33] developed U-Net inspired model in 2019. A similar technique was developed by Chamveha et. al [51]. In their work a U-Net [10] architecture with VGG-16 Encoder [52] was used.

Many works related to the heart size assessment from CXRs subject used deep learning to directly detect the presence of cardiomegaly [53, 22, 54]. The factors that influence the linking of the size of the heart to the presence of cardiomegaly (such as age or gender as discussed in Section 2.2.2.) makes the explainability of this detection more questionable. Since the diagnosis of cardiomegaly involved complex interaction between various factors related to sensitive patient information, this topic is excluded from the scope of this thesis. In relation to CTR measurement from CXRs with clipped anatomy, at the time of the thesis writing no works have been found in the literature.

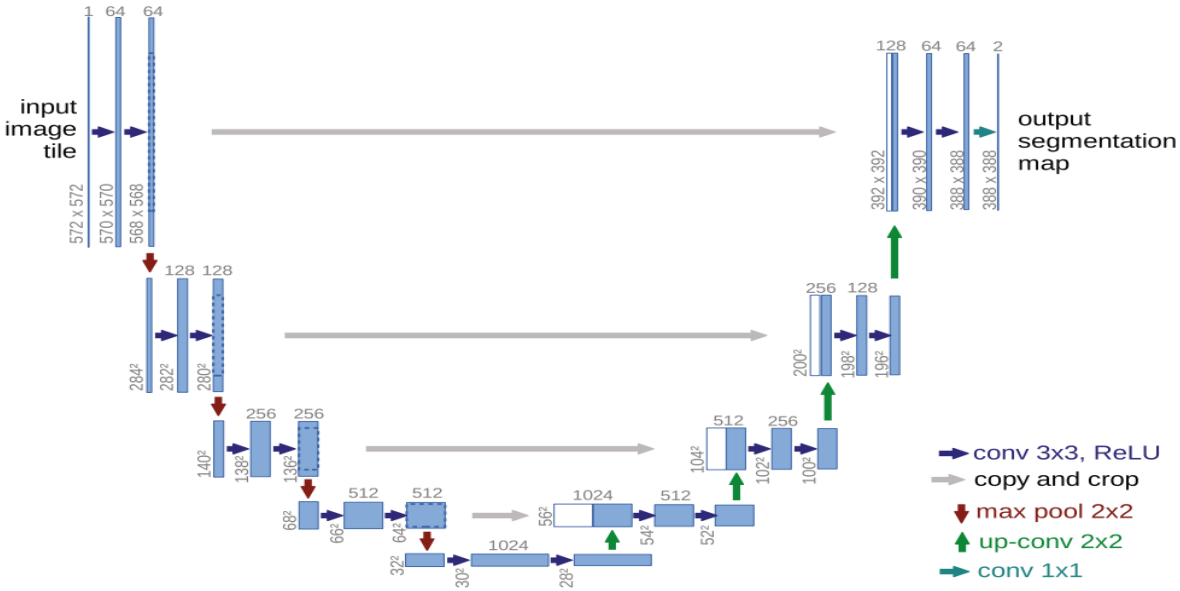

Figure 2.9: The structure of U-Net architecture from Ronneberger et al. [10]

2.3.1 Public datasets

Multiple CXR datasets are publicly available, but only few of them has lung or heart masks annotations. The datasets with lung or heart masks annotation used in this work are described below and summarized in Table 2.1

Two X-ray datasets of PA chest radiographs are made available by the U.S.National Library of Medicine [55] and curated for tuberculosis (TB) detection [56, 57]. The radiographs were acquired from the Department of Health and Human Services, Montgomery County, Maryland, USA and Shenzhen No. 3 People's Hospital in China. Both datasets contain normal and abnormal CXRs with manifestations of TB and include associated radiologist readings. Details are as follows:

- **Montgomery County X-ray set** : developed in 2014, it includes 138 frontal CXRs from the Montgomery County TB screening program, of which 80 are normal

and 58 show signs of TB. Images come in PNG format and have a 12-bit grey level. The size of the X-rays is either 4,020×4,892 or 4,892×4,020 pixels. Binary lung masks are available separately for the left and the right lungs. Lung were segmented under the supervision of a radiologist, following anatomical landmarks such as the boundary of the heart, aortic arc/line, pericardium line and a sharp costophrenic angle that follow the diaphragm boundary [55], as shown in Figure 2.8. Both posterior and anterior ribs are readily visible in the CXRs. The area behind the heart and the diaphragm was excluded as the "gold standard" segmentation is defined. It is possible to download the dataset from the National Institute of Health's web page [1].

- **Shenzhen chest X-ray set**: captured within a one month period, mostly in September 2012, as part of the daily routine at Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China. It contains 662 frontal CXRs, of which 326 are normal cases and 336 are cases with manifestations of TB. The X-rays are provided in PNG format. Their size can vary but on average it is approximately 3K × 3K pixels. Manually segmented masks are available thanks to Stirenko et al. [58].

Other datasets used are:

- **JSRT dataset** [59]: this dataset of CXRs with and without lung nodule was developed in 1998 by the Academic Committee of the Japanese Society of Radiological Technology (JSRT). It contains 247 images from scanned films: 154 X-rays with lung nodules and 93 without a nodule. All CXR images have a size of 2048 × 2048 pixels and a gray-scale color depth of 12 bits. Manually generated lung and heart field masks are provided by van Ginneken et al. [60] in the **Segmentation in Chest Radiographs (SCR) dataset** [60]. Even if the annotated ground truth of the CTR is not available for this dataset, since both heart and lung masks are present, it is possible to calculate it from the segmentations. The assumption that all the CXRs are well oriented is made after visual inspection since no CXR shows a significant rotation of the patient orientation. The ratio between the maximum horizontal width of the heart and the maximum horizontal width of the lung was calculated for each sample and the obtained CTRs were used as the ground truth.

- **Wingspan dataset**[49]: it is provided by a private research institute, Wingspan Technology. The dataset contains 221 grayscale CXRs for adult patients with annotated CTR. The extended version of Wingspan dataset includes 38 additional annotated images, for a total of 259 images. Each image was annotated by two licensed radiologists independently, and the annotations were accepted by both an-

[1] https://data.lhncbc.nlm.nih.gov/public/Tuberculosis-Chest-X-ray-Datasets/Montgomery-County-CXR-Set/MontgomerySet/index.html

notators and an independent reviewer. The de-identified data were collected from 6 hospitals, which have different imaging protocols. The image sizes, pixel spacing and clinical setup vary for each CXR.

Dataset	# CXRs	Lung masks	Heart masks
Montgomery	138	✓	
Shenzen	662	✓	
JSRT	247	✓	✓
Wingspan	259	✓	✓

Table 2.1: Summary of available datasets used in this work.

Chapter 3

Estimate CTR from Lung Segmentation

In the literature, many approaches to automatically extract the CTR from CXRs exist and almost all of them are based on an anatomical structures' segmentation. In this chapter, an approach of CTR estimation from lung segmentation is presented. To obtain the segmentation of anatomical structures from CXRs, a wide range of possibilities is open. Cadmir et al.[24] classified the segmentation algorithms into 5 groups: (1) rule-based methods, (2) pixel classification-based methods, (3) model-based methods, (4) hybrid methods (which are a combination of the previous ones), and (5) deep-learning methods. After a large study of the rich literature, they concluded that hybrid methods and deep learning methods surpass the algorithms in other categories and have segmentation performances as good as inter-observer segmentation performance. This means that recent works using deep learning approaches can provide a more abstract learning, resulting in better performances and higher accuracy compared to traditional image processing methods [61].

Several segmentation algorithms are publicly available. However, they all have the specific task of obtaining the segmentation of specific anatomical structures, e.g. the lungs. Most of the algorithms are developed to perform the segmentation task with underlying assumption that the images are taken from correct acquisitions. The computation of CTR in case of clipped anatomy is a mostly unknown subject. The presence of the area of interest in the image is normally not in dispute. Wrong positioning of the patient is one possible cause of the presence of clipped lungs in the image: it will cause the current approaches to fail.

In this chapter, four different models that produce lung segmentation masks and a method to extract CTR from them are presented. With the aim of obtaining a CTR measurement in the most robust and accurate way, the models tried also to recover the lung segmentation from the clipped part of the image by assuming that they can learn the

general shape of the lungs. CTR can be then calculated from the lung contours.

The objective of this chapter is described in Section 3.1. A description of the four models and CTR calculation are described in Section 3.2, followed by the implementation details in Section 3.3. The different models are tested on clipped and non-clipped datasets: details about the experiments and discussions are shown in Section 3.4. On Section 3.5 the conclusion can be found.

3.1 Objective

This research aims to develop a method to extract the variables to compute CTR from CXRs starting from lungs segmentation. During the computation of CTR, clipped lung will be an issue since part of the lung is out of the field of view. In case X-ray image retake is not possible, the goal for the algorithm should be to reconstruct the part of the lung that is clipped.

Four models have been developed and applied on both normal and clipped datasets. The goal was to test their robustness to the presence of clipping and to see if the accuracy obtained on clipped images improved between different models while remaining good on unclipped images.

3.2 Methodology

To extract the CTR from CXR images, the approach has been split into two parts:

1. Getting the binary segmentation of the lung.
2. Computing the CTR from the binary lung segmentation. Based on the medical definition of CTR as mentioned in Section 2.2, two variables need to be known: maximum diameter of the heart and maximum diameter of the chest.

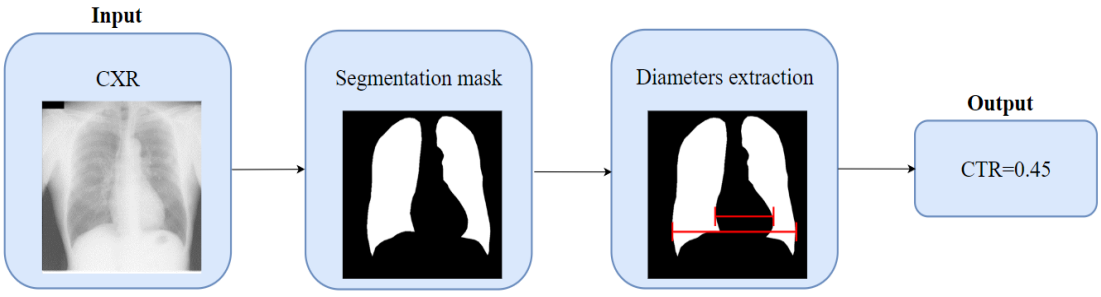

Figure 3.1: Overview of the methodology (estimate CTR from lung segmentation).

The publicly available model from the official implementation of "Lung Segmentation from Chest X-rays using Variational Data Imputation", from Selvan et al. [62] was used as a starting point. The aim of their work was to segment lungs with pulmonary opacification that render regions of lungs imperceptible. Although opacification is not within the scope of this research, their work was selected as a starting point because clipped lung involves the problem of missing information, which in a sense is of similar nature to opacification. It employs a segmentation network similar to U-Net[10] but adds a Variational Autoencoder (VAE)[63]. Autoencoders are used as an efficient way to code unlabelled data in unsupervised learning. Input data are mapped to a latent representation with very low dimension. An encoder network will output a single value for each encoding dimension, while a decoder will reconstruct the latent representation back to input space, trying to minimise the loss. To do so, the most important features variations need to be learnt by the network. Figure 3.2 shown an example of autoencoder architecture.

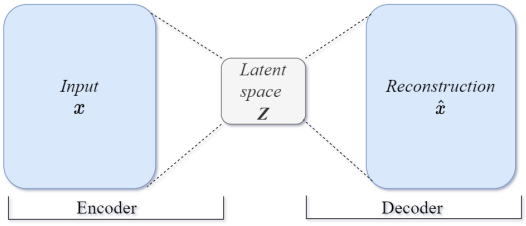

Figure 3.2: Autoencoder illustration.

VAEs are a specific type of autoencoders that rather than outputs a single value to describe each latent state attribute, they will describe a probability distribution for each latent attribute. Assuming that the input and the latent space are random variables, the latent variables are sampled from the distribution. Figure 3.3 shown an example of VAE architecture.

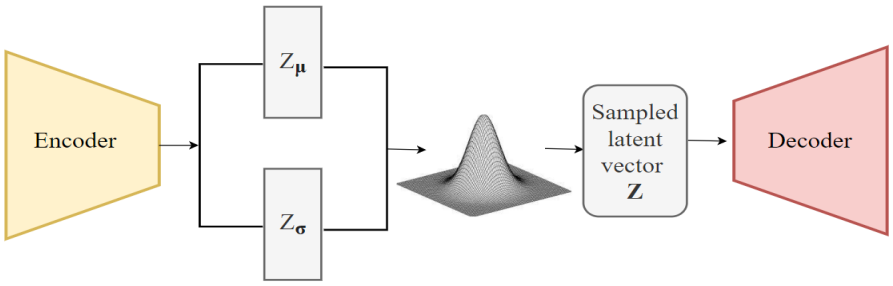

Figure 3.3: VAE illustration.

In their work, a latent representation of the data was used to impute high opacification regions on CXRs.

The released model of Selvan et al. [62] had been trained using CXR datasets from

Shenzhen and Montgomery hospitals (see Section 2.3). From 704 images, the 75% was used for training and 25% for validation purposes, but the indexes for the splitting were not available. For this reason, a random splitting with the same percentage of training and validation was done and the model was retrained from scratch on the same datasets. Starting from their work, two models have been replicated (Model 1 and Model 2), which differ in augmentation techniques used. To improve the segmentation of the clipped region of the lung, a third model was trained using a new augmentation technique, namely Clipping augmentation (Model 3). In these 3 models, the field of view of the segmented mask is always the same field of view of the input CXR. For this reason, a padded version of the image is necessary in input to predict the clipped part of the lung, as it will be better described later in this chapter. Subsequently, the idea was to have an algorithm capable of producing an output with a greater field of view than the initial image in order to input clipped images that are more faithful to the realistic ones (without padding in the input image). This leads to the last proposed model (Model 4). The architecture and the augmentation techniques used for this four models are summarised in Table 3.1 and will be better described in Section 3.2.2 and Section 3.2.3.

Model name	Architecture	Standard Aug.	Block and Diffuse Aug	Clipping Aug.	Realistic Aug.	Padded input image
Model 1	U-Net with VAE	✓				✓
Model 2	U-Net with VAE	✓	✓			✓
Model 3	U-Net with VAE	✓	✓	✓		✓
Model 4	U-Net with VAE with wider output	✓	✓		✓	

Table 3.1: Summary of lung segmentation models.

All the four methods were tested on both normal and clipped datasets. As a last step, the CTR was calculated from cardiac and thoracic diameter extracted from the binary output of the models.

3.2.1 Dataset

Different datasets were used for training, validation and test purposes.

The dataset used for the training and validation was formed by combining Montgomery County X-ray set and Shenzhen chest X-ray set, described in Section 2.3.1. The splitting of the dataset follows the same structure as the one used by Selvan et al. [62]. From the combination of the two datasets, 704 images are selected: 528 CXRs are used for training while 176 are used for validation purpose. To test the performances for both the lung segmentation and the CTR calculation, the JSRT dataset (also described in Section 2.4.1)

was used.

To evaluate the performances of different models on clipped lungs, it was necessary to build appropriate test sets. As discussed in Section 3.2, the output of models 1, 2 and 3 have the same field of view as the input images. To detect the segmentation of the clipped part of the lung, it is necessary to test them on padded images. For this reason, the clipped & padded JSRT dataset was created. The padding of input is not necessary for the U-Net with wider output model (model 4) to predict the clipped part of the lung. For this reason, to test this model the Clipped JSRT dataset was created.

- **Clipped&padded JSRT dataset**: Starting from the JSRT dataset (with corresponding masks from SCR dataset), the idea is to crop one side of the image in order to obtain a clipped lung. From each label, the bounding box of the lungs was identified; a random side of the image was chosen and, on this side, starting from the border of the bounding box, a percentage between 5 and 10% of the width or height of the lung was cropped. The image is then padded on all the sides, with 102 pixels (20% of the dimensions of the width and height of image, which were 512x512). The label is also padded with the correct number of pixels on each side, in order for the lung masks to be in the correct position with respect to the augmented image. A visual example of the clipped&padded JSRT images is given in Figure 3.4.

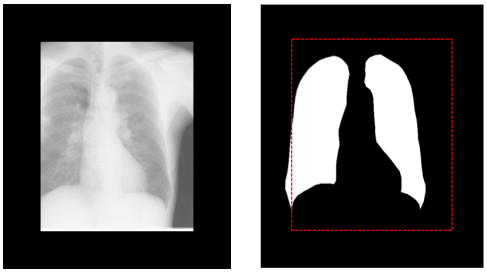

Figure 3.4: Example of image and label from JSRT clipped and padded dataset. The red dotted rectangle represents the initial field of view of the image

- **Clipped JSRT dataset**: Starting from JSRT dataset (with SCR masks), cropped CXR images with clipped lung are obtain as described in Clipped&padded JSRT. The mask of each image is positioned in the center of a black wider background image (with the same dimension of the label + 128 pixels on each side). The central masks is then translated on the background with an offset equal to the number of pixels that have been cropped on the corresponding image, on the same side. In this way, the field of view of the cropped CXR will always remain in the center of the label. An illustration of these steps is shown in Figure 3.5.

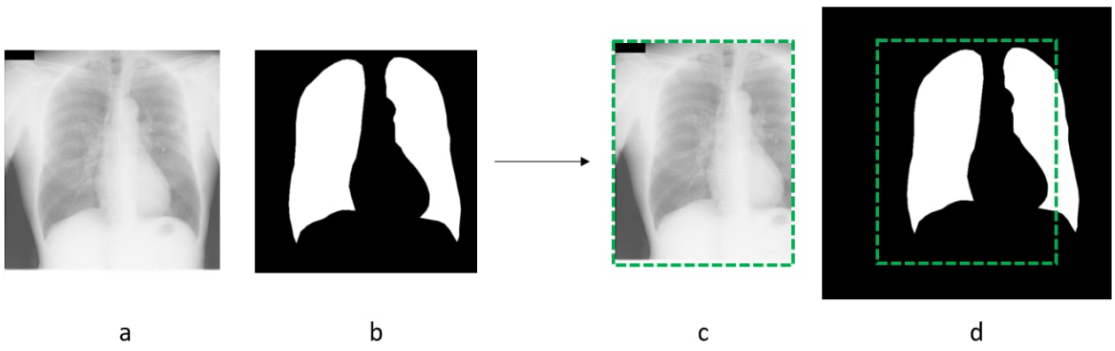

Figure 3.5: Example of JSRT image (a), its corresponding mask (b); cropped version of the image (c) and its corresponding wider mask (c). The green dotted rectangle outline the field of view of the cropped image.

3.2.2 Data Pre&Post-processing and Augmentation Techniques

To obtain the lung segmentations, all images are pre-processed by following these steps:

1. input images are rescaled to 640x512 pixels,
2. the histogram of the images are equalized in order to improve the contrast.

Post-processing steps on lung masks are as follow:

1. connected component analysis is used to exclude small erroneous regions,
2. binary morphological closing is done to fill holes up to 10 pixels in the segmentation.
3. since the output of every model is the segmentation of both the lungs, left and right lungs are identified by searching for the first and second-largest connected component. In some cases, when the algorithm does not work as it would, it is possible to obtain only one connected component in output, where a bridge is present, usually between the top of the two lungs. The segmentation is then cut in two parts: a vertical line is searched in the central part of the image (between the 2/5 and 3/5 of the width of the image). The vertical pixels of this region are summed in a vector, where the lowest value is searched to find the best position for the vertical line to obtain left and right lungs.

The pre-processing steps and steps 1 and 2 of the post-processing were already implemented in the work by Selvan et al. [62].

Data augmentation is essential to overcome the constraints of image data collections such as the limited number of collected data. Any proposed machine learning models may be less accurate if there is a shortage of data or erroneous labelling. The main advantages

of data augmentation are that it's cheaper than regular data collection and labelling and it helps in having less overfitting problems [64].

As mentioned in Section 3.2, different data augmentation techniques has been used to train different methods. To train the U-Net with VAE model [62] the following techniques were used:

- *Standard augmentation*: random rotation, vertical flipping and horizontal flipping.

- *Block and diffuse augmentation*: whitish masks are applied to vertically or horizontally cover one-half of the image, while to simulate high opacification regions, random sets of disks of varying radii smoothed with a Gaussian kernel are applied (200 precomputed masks were already available to be randomly applied on input images).

- *Clipping augmentation*: one side of the image is cropped to obtain a lung with a percentage of clipping between 5 and 10%. The result is then padded on all the sides with a number of pixels equal to the 20% of the width or the height of the image.

To train the wider-output U-Net model, the following technique has been used:

- *Realistic augmentation*: one side of the image is cropped to obtain a lung with a percentage of clipping between 5 and 10%. The central masks were then translated on a black wider background (with the same dimension of the label + 128 pixels on each side) with an offset equal to the number of pixels that have been cropped on the corresponding image, on the same side. The field of view of the cropped CXR will always remain in the centre of the label.

The different types of augmentation are summarised on Figure 3.6.

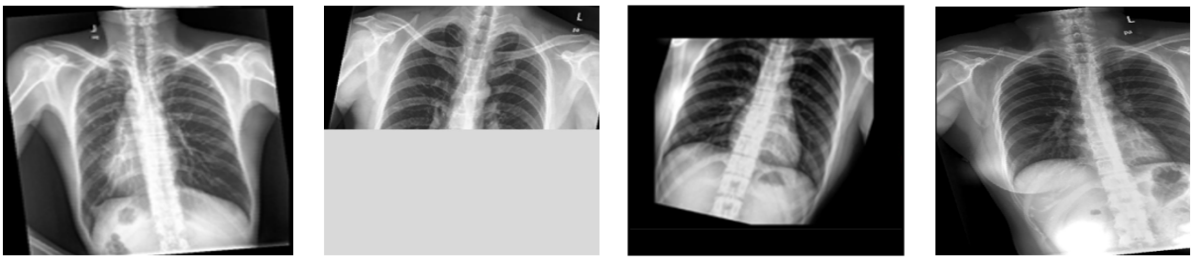

Figure 3.6: From left to right, examples of: standard augmentation, block and diffuse augmentation, clipping augmentation, realistic augmentation

3.2.3 U-Net Architecture and Variational Autoencoder

The architecture of models 1, 2 and 3 is an adaptation of the method by [62], which is the U-Net with VAE. Input of size 640 x 512 with 1 channel is used. The structure is

similar to the original U-Net[10] (Ronneberg et al., 2015) with some modifications. An illustration of the proposed architecture is presented in Figure 3.7. It operates at four resolutions, consisting of the repeated application of two 3x3 convolutions, each followed by the ReLu [65], and an average pooling operation. In this model, the first two resolution are obtained with a scaling factor of 4 and the other two by a factor of 2. The additional autoencoder used for data imputation has a similar structure with the encoder: it also operates at four resolutions, obtained with a scaling factor of 2. The 2D feature maps are then passed to a series of 4 1D convolutional layers to predict the variational density N (μ, σ^2), where N is a normal distribution with mean μ, and variance σ^2. The latent vector is then sampled from the distribution, with a latent dimension of 8. Results from the latent representation of the VAE are concatenated with the output of the encoder and both shared the same decoder. Skipping connections between the encoder and the decoder allow the U-Net constructs an image in the decoder part using fine-grained details learned in the encoder part. In this model, the output will have the same size as the input, resulting in a 640 x 512 mask with 1 channel.

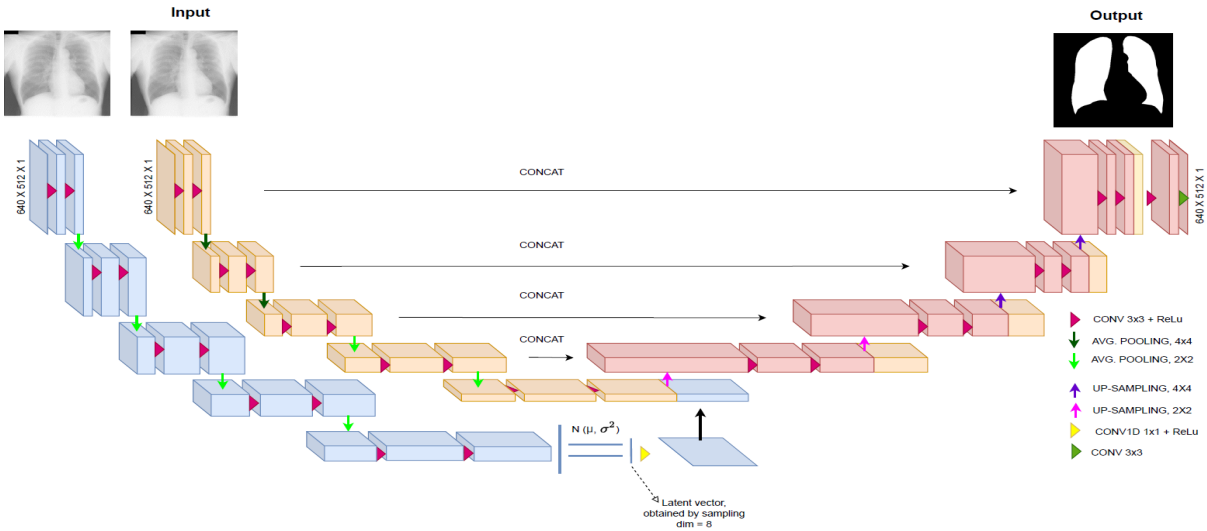

Figure 3.7: Architecture overview of U-Net with VAE model: the yellow part represents the encoder, the blue part represents the VAE and the red part represents the shared decoder.

Some modifications to the decoder path and to the skipped connection from the previous model have been done to build the U-Net with wider-output model (model 4). With an input of size 640 x 512 x 1, the segmentation will have a larger field of view of 128 pixels on each side of the image, for a total output size of 896 x 768 x 1. Skipped connections are made possible by padding with the correct number of pixels the output of each encoder layer (respectively for each layer: padding with 128, 32, 8, 4, 2 pixels). An illustration of the proposed architecture is shown in Figure 3.8.

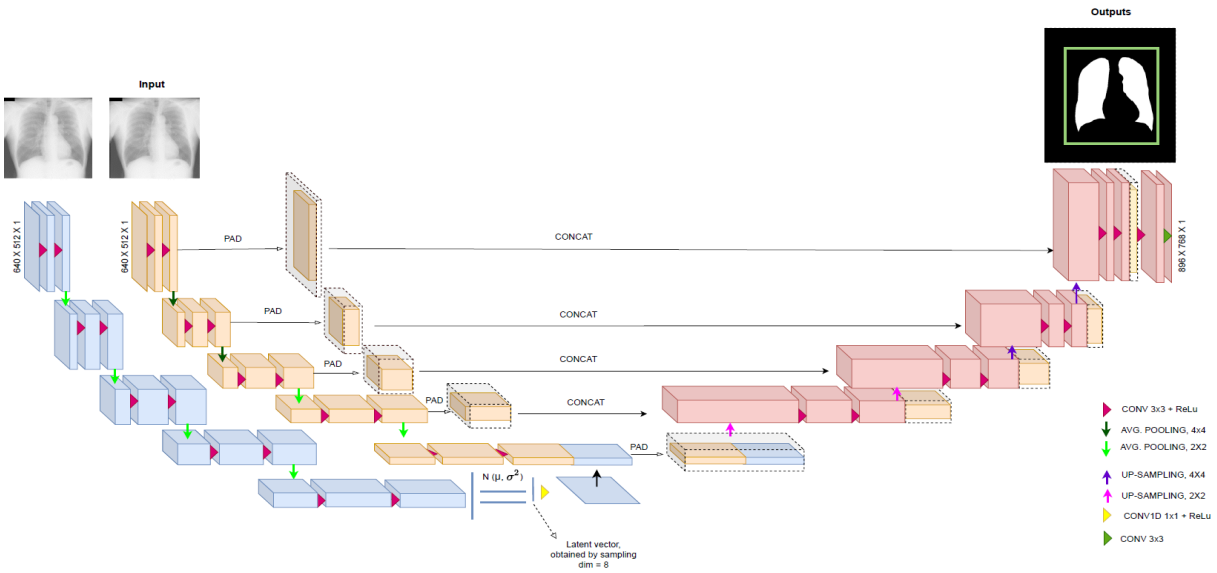

Figure 3.8: Achitecture overview U-Net wider-output with VAE model: the yellow part represents the encoder, the blue part represents the VAE and the red part represents the shared decoder.

3.2.4 CTR estimation

It is possible to calculate the CTR from the contour of the lungs, as Dong et al. [49] did in their work. Starting from the segmented lungs, the maximum diameter of the heart and the maximum diameter of the lungs are extracted, then the CTR is calculated. To extract the maximum diameter of the heart and the lungs, two horizontal lines are positioned on the CXRs as shown in Figure 3.9.

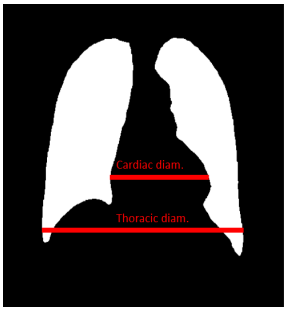

Figure 3.9: Example of cardiac and thoracic diameter identification from a lung mask.

They are respectively:

- *the cardiac diameter* (D_{heart} from equation 2.3): searched above the vertex of the cardiophrenic angle of the right lung (point C from Figure 2.8). It is defined as the maximum horizontal distance between the two lungs. The point that defines the vertex of the cardiophrenic angle is calculated as shown on Figure 3.10, using the

following steps, as Dong et al. [49] did in their work:

1. Find the contour of the right lung mask;
2. Find the convex hull of the right lung mask to obtain points on the right lung boundary;
3. Find the maximum horizontal width in the right lung;
4. From the output points of step 2, select the points whose distance to the rightmost one is at least 2/3 of the width of the right lung;
5. Between these points, select the lowest one.

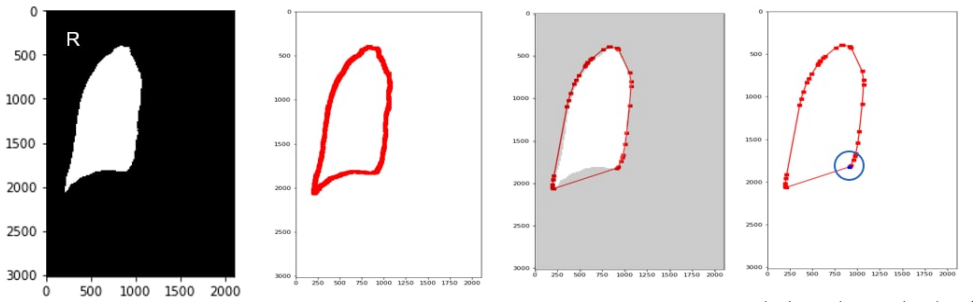

Figure 3.10: Steps to detect the vertex of the cardiophrenic angle.

- *the thoracic diameter*: defined as the maximum distance between the rightmost and the leftmost point on the lung segmentation chosen on the same horizontal line.

The CTR is then calculated as the ratio between the two obtained diameters, as described in equation 2.3.

3.3 Implementation details

Both U-Net with VAE models described in Section 3.2.3 are developed in PyTorch. The same parameters used by Selvan et al. [62] in their proposed methods are chosen: during the training a batch size of 12 is used, with a learning rate of 10^{-4}, with Adam optimizer for a maximum of 200 epochs. The loss function used is composed of two parts: a "reconstruction loss" that helps in the segmentation and a "latent loss" with a regularisation effect that helps with data imputation [62]. Binary cross-entropy loss has been used as the reconstruction loss while KL divergence loss [66] has been used as "latent loss". Convergence is assumed when there is no improvement in validation loss for 20 consecutive epochs. The model with the minimum validation loss is used for testing. In each model, the probabilities of standard augmentation, block masks augmentation and diffuse masks augmentation are settled at 0.1. In case of rotation, the degrees of rotation is randomly chosen between 0 and 15 degree. The probability of clipping augmentation in model 3 is settled to 0.9. The probability of realistic clipping augmentation in models 4 is settled to

0.9. This means that model 3 and 4 are trained mainly on clipped lung CXRs. For the CTR estimation, the convex hull of the right lung mask was obtained using convexHull() function from OpenCV python library.

3.4 Experimental Results and Discussions

The two U-Net models described in Section 3.2.3 were tested both on clipped and not clipped images from JSRT dataset.

To evaluate the segmentation performances of the proposed methods, the Dice similarity coefficient is used. It measures the degree of overlap between the ground truth mask (G) and the predicted segmentation mask (P), and can be written as follows:

$$Dice(G, P) = \frac{2|G \cap P|}{|G| + |P|}. \tag{3.1}$$

From the output of every model, left and right lungs were identified. In this way, it is possible to evaluate separately the performance of the models on the two lungs. The separate analysis was done since they have different anatomical shapes and different variability, as described in Section 2.1.6.

Lung segmentation performance on non-clipped lungs

Firstly, the different models are evaluated on the 247 original images of JSRT dataset (not clipped). The main purpose of this experiment is to see if the performances on non-clipped images will be degraded going from the first model to models trained more and more specifically on clipped lungs. From this experiment, it is also possible to evaluate the differences in performances between the two lungs. The results are visualised in Figure 3.11.

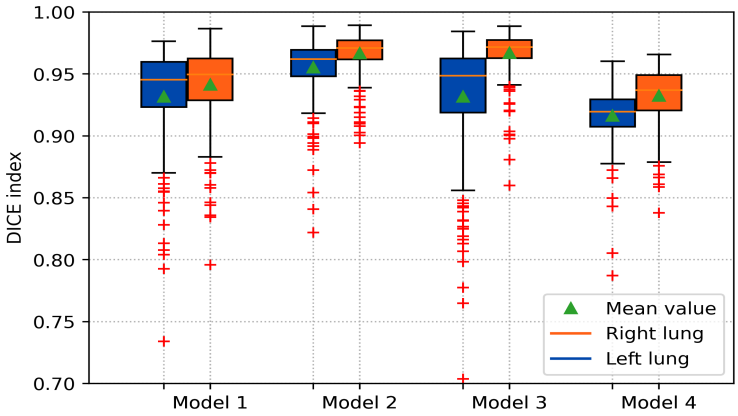

Figure 3.11: Dice coefficient of left (blue) and right (orange) lungs segmentation using different models.

The best mean Dice index is obtained by model number 2. Improvements are shown from using only standard augmentation: this trend replicate the results obtained by Selvan R. et al. [62] in their work.

It is interesting to notice how performance decreases when models more trained on clipped anatomy (models 3 and 4) are evaluated on non-clipped images, but there is not a huge drop: performances never go under a mean value for the Dice index of 0.914.
It is also possible to notice how the mean value of the Dice index of the left lung is slightly lower for all the models. To prove if the differences between left and right Dice indexes are significant, a paired **Z-test** was used: a one-sample Z-test was performed on the differences between right and left Dice indexes. The null hypothesis (H_0) states that the differences between the right and left indexes is zero, while the alternative hypothesis (H_0) states that the index of the left lung is lower than for the right lung. To summarise:

$$H_0 : \mu = 0; H_1 : \mu > 0, \tag{3.2}$$

where μ represents the difference between the mean right dice index and the mean left dice index. The significance level (α) is set at 0.05. It is the probability of the study rejecting the null hypothesis, given that the null hypothesis is true. The probability value (p-value) for each model is calculated and it shows the probability of obtaining a result at least as extreme, given that the null hypothesis is true. The result is statistically significant when $p \leq \alpha$. Results can be visualised in Table 3.2.

Z-test	
Model number	p-value
1	0.0017
2	$3.01*10^{-13}$
3	$1.55*10^{-39}$
4	$4.216*10^{-18}$

Table 3.2: Z-test: p-values for each model tested on non-clipped lungs

For all the models, the probability value was really low and always less than the alpha value, showing that the null hypothesis can be rejected with a degree of confidence of 95%. This difference can be interpreted as a greater difficulty for the models to segment left lung due to the higher variability in its shape, as described in Section 2.1.6.

To understand if the algorithm is learning the presence of anatomical differences between the two lungs or if it is treating them in the same way, another test is conducted.

Dice indexes of left and right segmentation masks of JSRT images are compared to Dice indexes of the same images but horizontally flipped. Model 4 is applied to obtain the segmentation masks. Results are shown in Figure 3.12. The performances are almost the same when the test set is flipped. This means that the algorithm does not learn the anatomical differences between the two lungs.

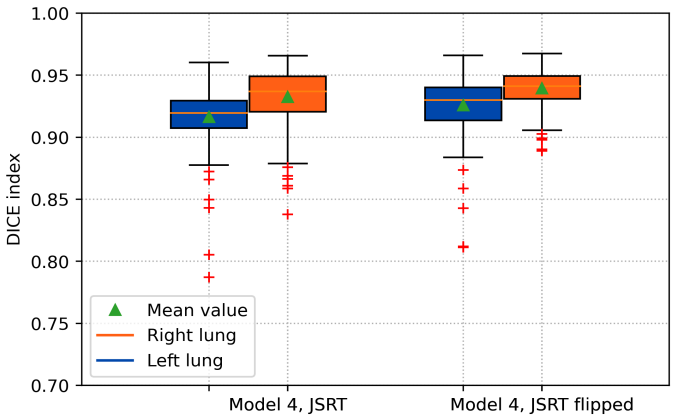

Figure 3.12: Dice indexes of flipped and non-flipped JSRT images.

Lung segmentation performance on clipped lungs

Secondly, the different models are evaluated on the clipped datasets. As discussed in Section 3.2.1, since the U-Net used in models 1, 2 and 3 has the same field of view both in input and in output, padded input images were needed to test the ability to predict also the shape of the clipped part of the lung. For this reason, these models were tested on clipped&padded JSRT dataset, while models 4 is tested on only clipped JSRT dataset. The performances of the different models on clipped lung datasets are visualised in Figure 3.13 and an example of segmented lung masks, using the four different models in shown in Figure 3.14.

To see if the difference between left and right Dice indexes are still significant (as in the lung segmentation performances on non-clipped lungs) a pair Z-test is performed. This statistical test has the same structure of the one presented in the previous experiment with non-clipped lungs, but it is applied on the performance of models on clipped lungs. Results can be visualised in Table 3.3. For models 1, 3 and 4 the probability value is less than the alpha value: the null hypothesis (that indicates that there is no difference between Dice index of the left and the right lung) is rejected. In model no. 2 the null hypothesis is accepted because it is higher than the alpha value set at 0.05. The difference between left and right indexes is not statistically significant in this case.

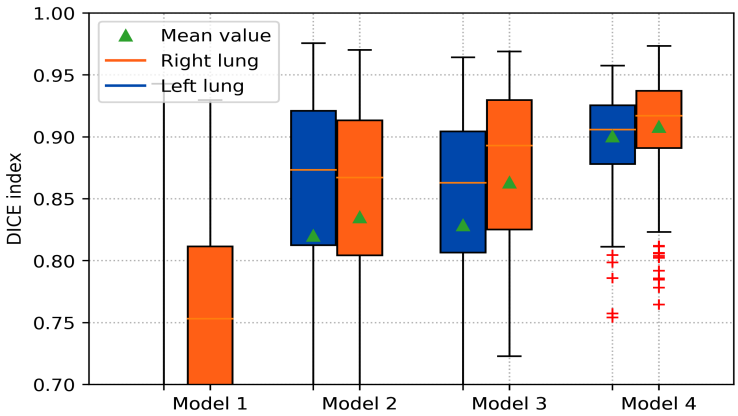

Figure 3.13: Dice coefficient of lungs segmentation using different models

Z-test	
Model number	p-value
1	$1.16*10^{-36}$
2	0.0667
3	$5.29*10^{-6}$
4	0.0140

Table 3.3: Z-test. p-values for each model tested on clipped lungs

From this experiment, it is possible to make several observations:

- Mean performance of model 1 is really low (Mean Dice Left=0.329; Mean Dice right=0.629) and falls outside of the range of the graph in Figure 3.11: most of the time the lung is completely mistaken with the padded region. Shapes of the lungs are usually elongated and far from the original (see Figure 3.14). This model has never seen cases in which part of the lung or part of the image is covered with uniform regions during the training time.

- It is for the same reason that it is possible to observe a huge increment in performance between model 1 and model 2. In model 2 blocking masks were used as augmentation techniques. By simulating high opacity regions, these masks will start to simulate also a first example of missing part of the lung. The model is still susceptible to the presence of a uniform black padding region in the input image: prediction of the lungs were often elongated also in the padded region.

- Performances increase a little when going from model 2 to model 3. In this case, input images are more similar to training samples since it is trained on images that were also padded.

- The best performances are obtained with model number 4. Note that this model was tested on only clipped JSRT (without padding), since it is able to predict a wider field of view. It is possible to assume that the increment in performances is caused by the removal of the padding region because in this way we are preventing some types of segmentation errors:

  – black padding of the image was often misunderstood as lung in the previous models;

  – the presence of an outlined border in the image, where there is a large variation in pixel intensity, often caused the clipped part of the lung not to be recognized.

- The differences between segmentation of left and right seems to decrease.

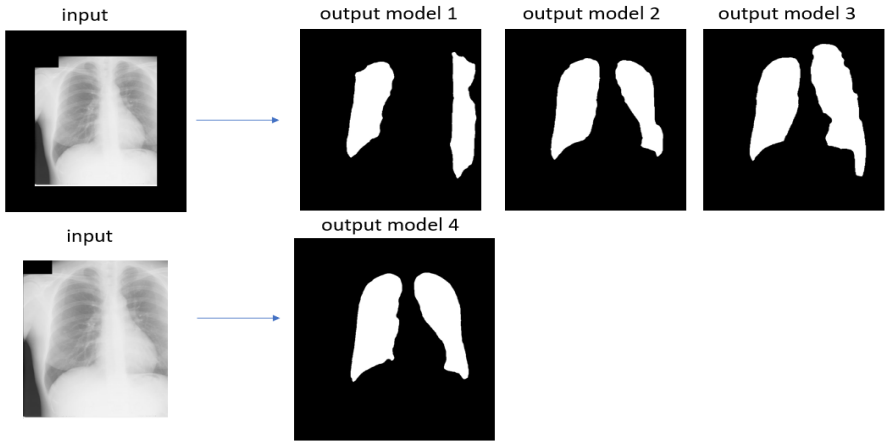

Figure 3.14: Example of segmented lung masks obtained by applying four different models.

CTR measurement performance

Since the objective of this thesis is not the segmentation of the lung itself but the evaluation of the CTR, the CTR calculated from the segmentation of the lung (as described in Section 3.2.4) was evaluated. The ground truth CTR is calculated according to the steps in subsection 3.2.4 based on the lung mask label from the JSRT dataset. These values are compared to the CTR calculation performed based on the predicted lung mask using the best model obtained so far (described in Section 3.2 as model 4). The evaluation is done in terms of Absolute Error (AE):

$$AE = |y_j - \hat{y_j}|, \tag{3.3}$$

by measuring its mean and standard deviation, and Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y_j})^2}. \tag{3.4}$$

The mean value of the AE (MAE) measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average of the absolute differences between prediction and actual observation, where all samples have equal weight. Also the RMSE measures the average magnitude of the error. It is the average of squared differences between prediction and actual observation.

In this context, the MAE shows, on average, how far the prediction of the CTR is from the ground truth. The RMSE has a similar aim but the higher errors between the CTR prediction and the ground truth are amplified because of the squared term: it will highlight methods with the most significant errors. Both metrics can range between 0 and ∞ and are indifferent to the direction of error.

The results obtained from the best method obtained so far (described in Section 3.2 as model 4) are shown in Table 3.4.

	JSRT non-clipped	JSRT clipped
AE (mean± std)	0.038± 0.040	0.058 ± 0.057
RMSE	0.055	0.082

Table 3.4: Evaluation of CTR on JSRT clipped and not-clipped images.

As expected, increased error is observed from the results from clipped JSRT dataset compared to the non-clipped dataset. An example of performance from the best segmentation model on clipped and non-clipped image is shown in Figure 3.15. In this case it is possible to see an absolute error of 0.023 for the non-clipped image and an absolute error of 0.034 for the clipped image.

It is not straightforward to compare these results with the best result found in literature for CTR estimation from lung segmentation prediction since different method might use different test dataset. However, to have an idea of how the proposed method performs, the state-of-art results are shown as follows. The method of Dallal et al. [50] currently shows the best results in literature. The results from their method come from testing 103 images from a private dataset. RMSE and Percentage Error (PE) are evaluated. The PE of a test image j is defined as

$$PE_j = \frac{\hat{y_j} - y_j}{\hat{y_j}} * 100\%. \tag{3.5}$$

The mean and the standard deviation of the PE are calculated on the results of the proposed method both on JSRT clipped and non-clipped images. The results, showed

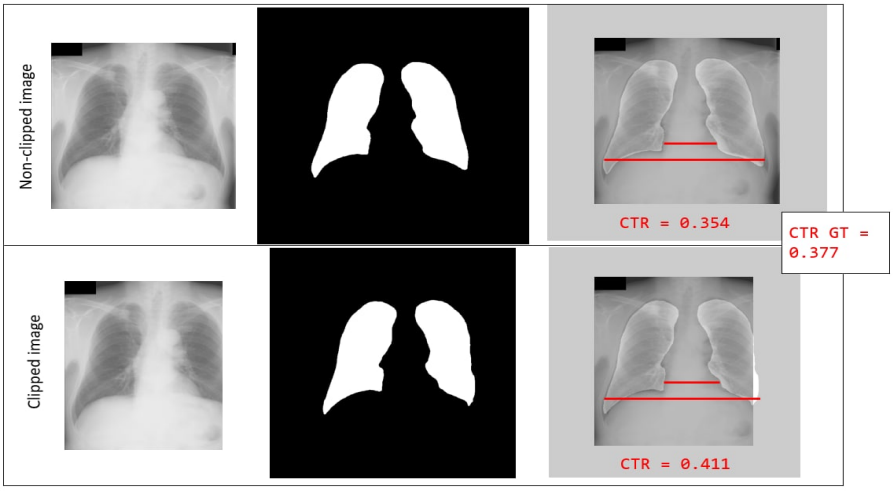

Figure 3.15: Example of performance from the best segmentation model on the same clipped and non-clipped image.

on Table 3.5, are not directly comparable but they show most of the results achieved with the proposed method are in the same order of magnitude. Only the PE on clipped images is slightly above the next order of magnitude, however this result was expected as performance on clipped lungs is lower in general due to the missing information.

	PE	RMSE
Proposed method on non-clipped images	8.5 ± 9.1 %	0.055
Proposed method on clipped images	12.9 ± 13.4 %	0.082
Dallal et al. [50] method	7.9 ± 9.1 %	0.06

Table 3.5: PE (mean ± standard deviation) and RMSE for CTR estimation.

The proposed method based on lung segmentation algorithm to estimate CTR achieved good performances. It does not require manually segmented heart mask dataset. However, some limitations are present:

- the calculation of the CTR from horizontal diameters is based on the assumption that the orientation of the CXR image is correct. The obtained results are from a dataset that does not contain high rotation of the patient in CXRs. Performances will expect to drop in case of incorrect positioning of the patient.

- the predicted lung region outside the field of view of the input image should be taken with caution: shape is just assumed by the algorithm based on the shape of the lungs seen in training phase and does not generalize well when new data with different variations, due to disease or acquisition, are present.

- the calculation of CTR is made using only lung masks. It deviates from the original

strict definition of the CTR. For this reason, the idea was to extract the CTR from both the segmentation masks of lung and heart, as mainly seen on literature, to see if better performances can be achieved.

3.5 Conclusion

A method to estimate the CTR from lung segmentation mask is presented in this chapter. The performances of the proposed method are in the same order of magnitude compared to other state-of-art method that computes CTR from lung segmentation. However, the methods in the literature do not consider the case where the lung field in the image is clipped and would likely to fail handling those cases. In contrast, our proposed method takes clipped anatomy into account. Four different models have been developed for this purpose. The best performances for clipped lung images have been obtained using a modified U-Net with VAE architecture that output segmentation mask with a wider field of view than the input. It allows the algorithm to reconstruct the clipped part of the lung. The CTR is then calculated from the lung segmentation. Performance of the proposed method are promising, although only the lung is taken into account. Since the strict definition of the CTR involves also the heart boundary, improvement in CTR estimation is expected when the cardiac diameter is extracted from the segmentation mask of the heart itself instead of from the lung mask. For this reason, the CTR estimation from lung and heart masks is investigated in the next chapter.

Chapter 4

Estimate CTR from Lung and Heart Segmentation

As discussed in Section 2.3, almost all the works regarding CTR estimation are segmentation-based solutions, where the heart and the lung boundary are detected. The calculation of the CTR, consequently, follows its stricter definition as stated in equation 2.3 where it is obtained by calculating the cardiac diameter from heart segmentation mask and the thoracic diameter from the lungs segmentation mask. One issue with this approach is the absence of large public CXRs databases with heart segmentation annotation from radiologists to train a heart segmentation model. A method to overcome this issue is the fine-tuning of the already described lung segmentation model to output also the segmentation of the heart, that way limited number of labeled data with heart segmentation can be used. The segmentation of the heart is expected to improve the segmentation of the lungs since they are bordering each other. The addition of the heart segmentation will thus allow the trained model to have additional information about the mutual position of heart and lung masks.

The objective of this chapter is described in Section 4.1. A description of the methods for heart and lung segmentation and CTR calculation are described in Section 4.2, followed by performance evaluation of the new proposed method in Section 4.3. On Section 4.4 the conclusion can be found.

4.1 Objective

The aim of the work presented in this chapter is to propose a method to estimate the CTR from lung and heart segmentation and to evaluate if the introduction of the heart segmentation for the calculation of CTR can be beneficial when compared to the results from Chapter 3. Moreover, since the actual CTR definition is sensitive to rotation of the patient, a new metric, correlated with the CTR, has been proposed to evaluate the enlargement of the heart in a different and more robust way.

4.2 Methodology

To extract the CTR from CXRs images, a similar approach to the one used in the previous chapter has been followed. It mainly differs in the calculation of the cardiac diameter. The cardiac diameter is calculated now from the segmentation of the heart and not from the segmentation of the lung. The approach can be summarized as follow:

1. Getting the binary segmentation of the lungs and the heart
2. Obtain the maximum cardiac diameter D_{heart} from the heart segmentation mask and the maximum thoracic diameter D_{thorac} from the lung segmentation mask (as described in equation 2.3).

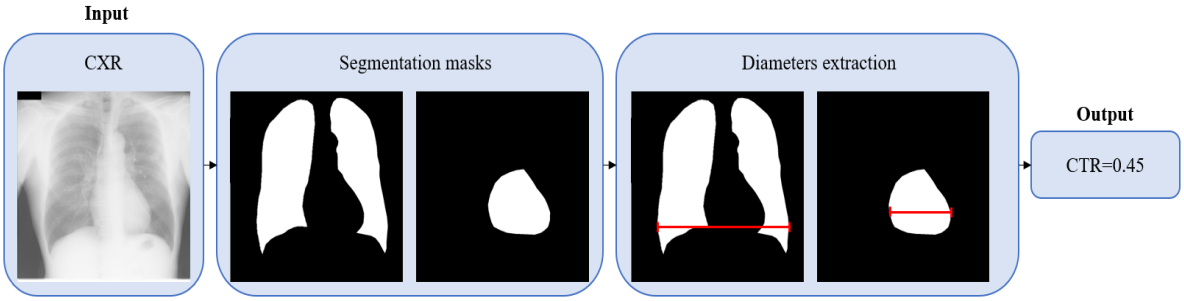

Figure 4.1: Overview of the methodology (estimate CTR from lung and heart segmentation)

To obtain lung and heart segmentation masks, two possible methods are proposed:

A. Lung segmentation masks are obtained applying model 4 described in Section 3.2. A new model has been trained to obtain the heart segmentation. For this purpose, the same U-Net with VAE from Selvan et al. [62] (described in Section 3.2.3) was trained from scratch with CXRs and heart masks. It replicates model 2 of Section 3.2.3, with the difference that it is now trained on heart masks instead of lung masks.

B. The U-Net wider-output with VAE (described in Section 3.2.3 as model 4) was modified to have multi-label outputs. The already trained lung segmentation model was fine-tuned with heart segmentation masks. A more detailed description of this architecture is proposed in Section 4.2.2. With this approach, one U-Net wider output with VAE is used to obtain both the lungs and the heart segmentation.

In model B, the segmentation of the heart is influenced also by the segmentation of the lungs since they are bordering each other. For this reason, the presence of clipped lungs can presumably influence the segmentation of the heart. To evaluate this, model B has been applied both on clipped and non-clipped images.

Subsequently, the CTR (following its definition in equation 2.3), was calculated on the lung and heart masks obtained with the best performing segmentation method. Since this classic definition of CTR is very sensitive to rotation of the patient on the image, a new metric has been proposed to evaluate the anatomical accuracy of the segmentation. This new metric, called in this research "rotational invariant cardio-thoracic ratio" (RI_CTR) is described in section 4.2.3.

4.2.1 Dataset

Wingspan and JSRT dataset (described in Section 2.3.1) are the only public dataset with both lung and heart mask segmentations found in the moment of writing this thesis.

From visual analysis of Wingspan dataset some errors have been found. In the downloaded dataset there were three folders respectively with left lung masks, right lung masks and heart masks. In left lung folder, four "heart shaped" masks have been recognised, while the corresponding masks in the heart folder had "left lung shape". This was presumably due to errors in the way the masks were stored. These detected errors were corrected, putting the images in the correct folder before using the dataset. Furthermore, it is possible to visually notice a difference in how the segmentation of the heart has been performed in annotations of the JSRT and Wingspan datasets. In general, the heart annotated masks of the JSRT are more "circular shaped" when compared to the Wingspan heart annotated masks that are more "triangular shaped" (as shown on Figure 4.2). This could be due to the radiologists using different way of performing the annotation.

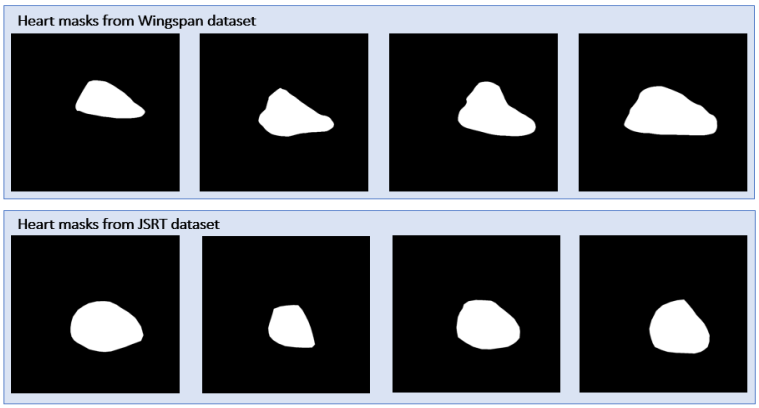

Figure 4.2: Four examples of heart mask from the Wingspan dataset (first row) and four examples of heart mask from the JSRT dataset (second row).

To train model A and to fine-tune model B, the corrected Wingspan dataset was used: from 259 images, the 75% was used for training purpose and 25% for validation purpose. The splitting was done randomly. Similar to the experiment that was presented in Chapter 3, the JSRT dataset was used as test set. In this way, the results for CTR will

be comparable with the results from Chapter 3. To evaluate the influence of clipped lung on heart segmentation in model B, the Clipped JSRT dataset (described in Section 3.2.1) was used.

4.2.2 Multi-label U-Net with VAE Architecture

As stated before, the architecture of model B is a modification of the U-Net wider-output with VAE (described in Section 3.2.3 as model 4). In contrast to that one, now the last layer of the decoder is divided in two parts to allow multiple outputs: the lungs segmentation mask and the heart segmentation mask. An illustration of the architecture is shown on Figure 4.3. The initialisation of the weights comes from model 4 of Section 3.2, which was trained and validated on a total of 704 images with lung masks from Montgomery and Shenzhen datasets. As for models described in Chapter 3, the loss function used is composed of two parts: a "reconstruction loss" that helps in the segmentation and a "latent loss" with a regularisation effect that helps with data imputation [62]. KL divergence loss [66] is still used as "latent loss" while, differently from Chapter 3, a multiple-binary cross entropy loss is used as the "reconstruction loss". The multiple-binary cross-entropy represents the sum of the binary cross-entropy for lung mask prediction task and the binary cross-entropy for heart mask prediction task.

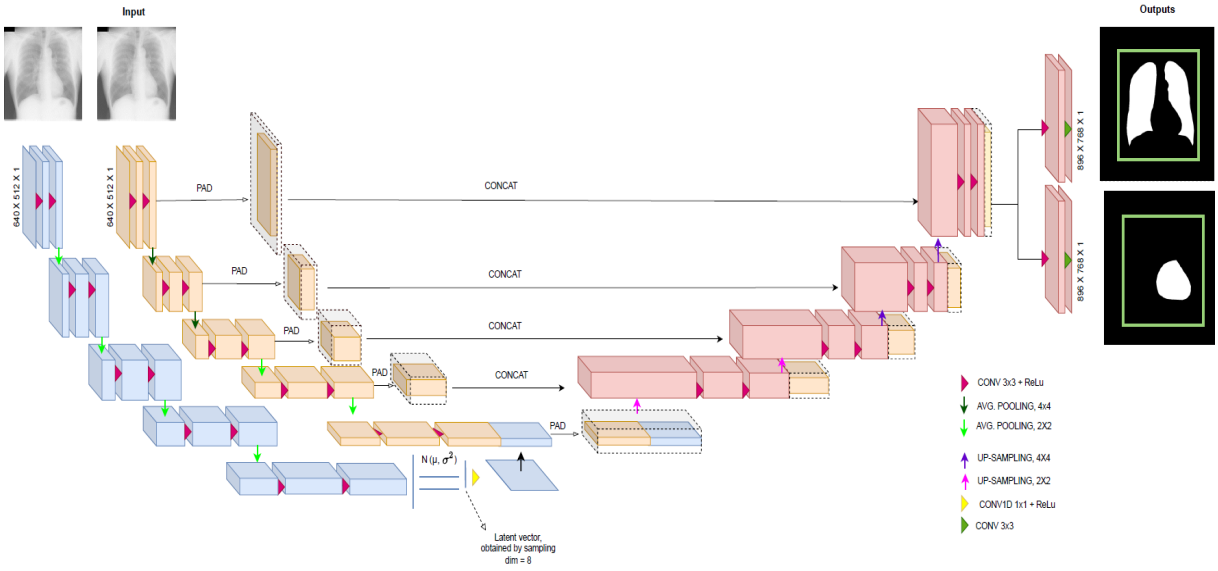

Figure 4.3: Architecture overview of multi-label U-Net with VAE model: the yellow part represents the encoder, the blue part represents the VAE and the red part represents the shared decoder. The green rectangles on the outputs represents the original field of view of the image.

4.2.3 CTR estimation

Starting from the segmentation of the lung and the heart, the CTR is directly calculated from the maximum horizontal width of lung segmentation mask and maximum horizontal width of heart segmentation mask, as done in the majority of works in literature related to this topic, such as [45].

When the segmentation method is tested on JSRT dataset (see Section 4.3), is noticeable how the general outline of a heart mask from the ground truth is different from the one from the model prediction. This is likely due to the differences in radiologist's annotations described in chapter 4.2.1 between Wingspan dataset (used during training phase) and JSRT dataset. This manifests as heart segmentation mask output being more 'triangular' rather than 'circular' causing the CTR to become less accurate. Since the classic definition of CTR is very sensitive to rotation, a different, rotational-invariant, way to evaluate the enlargement of the heart was used: RI_CTR. In RI_CTR, cardiac diameter and thoracic diameter are calculated as follow:

- *cardiac diameter*: the maximum circle inscribed in the heart mask is obtained and the diameter of the circle is used as cardiac diameter,
- *thoracic diameter*: the orientation of the lung mask is derived by finding the mayor axis of the mask and orient it to 0 degrees, and lungs are rotated to remove possible wrong orientation. Maximum horizontal width of the rotated lungs is calculated and used as thoracic diameter.

A visualisation of CTR and RI_CTR is shown on Figure 4.4.

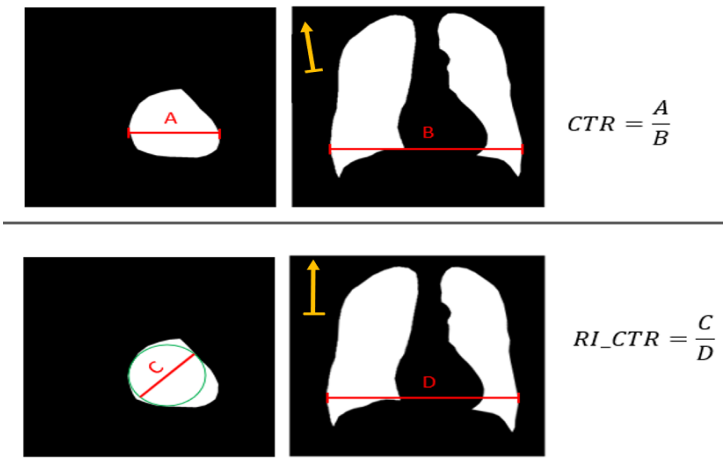

Figure 4.4: Visual illustration of CTR and RI_CTR calculation. The yellow arrows represent the major axis orientation of the lung mask

4.3 Experimental Results and Discussions

Heart segmentation models evaluation

Firstly, the ability to obtain accurate heart segmentation masks from models A and B (described in Section 4.2) was tested on JSRT dataset. The heart segmentation masks were evaluated using the Dice index defined in equation 3.1. Results are visualised in Figure 4.5. As expected, the model based on multi-label segmentation works better. This model seems to be able to derive information on the relative position between the heart and the lungs during training and thus obtain more accurate heart segmentation despite the limited dataset it is trained on. For this reason, all the next experiments are based on model B.

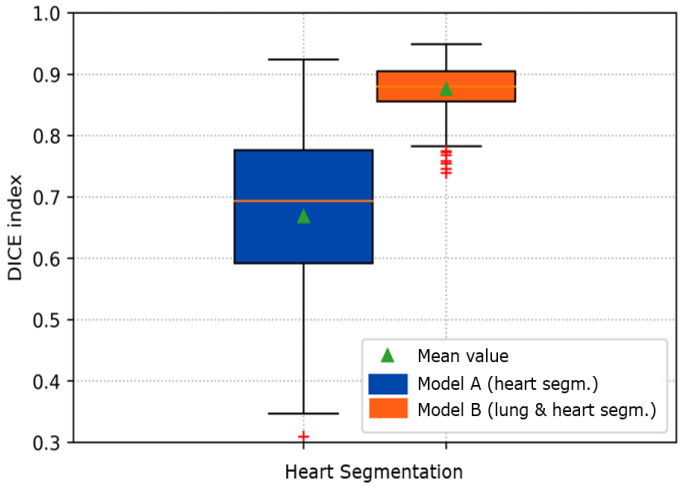

Figure 4.5: Dice indexes of heart segmentation of JSRT dataset using model A and model B.

Heart segmentation performance on clipped and non-clipped lungs

Secondly, model B has been applied also on Clipped JSRT and the heart segmentation performances were compared between clipped and non-clipped datasets. Results, in terms of Dice indexes, are shown on Figure 4.6. Performances on clipped dataset seems to get worse. To prove if the difference is significant, a paired **Z-test** was used: a two-sided Z-test was performed on the differences between Dice indexes from clipped and non-clipped CXRs. The null hypothesis (H_0) states that the differences between them is zero, while the alternative hypothesis (H_1) states that a difference is present:

$$H_0 : \mu = 0; H_1 : \mu \neq 0.$$

The significance level (α) is settled at 0.05. A p-value of $1.23{*}10^{-15}$ was obtained. It is thus possible to demonstrate that there is a statistical difference between Dice index of clipped images and the Dice index of nonclipped images since the p-value is much lower than α value. This difference of performance for the heart masks, shows that the presence of clipped lungs influences also the segmentation of the heart in the multi-label model

described, presumably due to higher errors in segmented lungs.

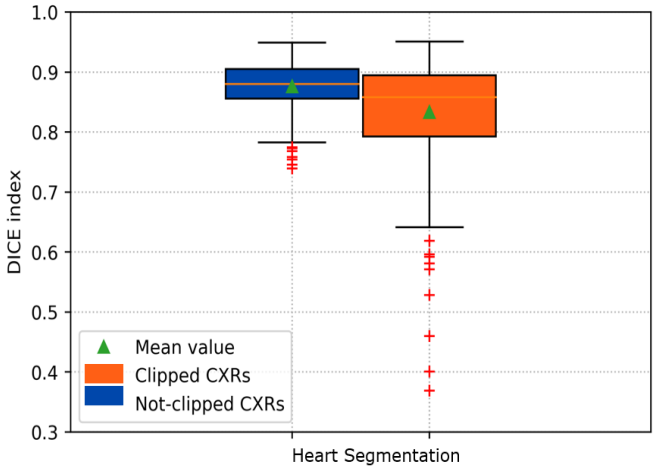

Figure 4.6: Dice indexes of heart segmentation using model B on normal JSRT dataset (blue) and Clipped JSRT dataset (orange).

CTR evaluation

Subsequently, the CTR was calculated from the lung and heart segmentation masks obtained from the proposed multi-label model, model B. The performance on non-clipped images are compared with the CTR performance from only the lung segmentation described in Chapter 3 and reported in Table 3.4. From Table 4.1, it is possible to notice that AE and RMSE get worse when also the heart segmentation is used in the estimation of the CTR. Furthermore, from the calculation of the correlation coefficients it is possible to notice that the method with also heart segmentation yields numbers that correlates worse with the ground truth CTR. This result is likely due to the limited dataset with heart masks used during the training and the differences in the way the heart mask is annotated between training and test datasets, as described in Section 4.2.1.

CTR	AE (mean ± std)	RMSE	Corr. Coeff.
from only **lung** segm.	0.038 ± 0.040	0.055	0.677
from **lung and heart** segm.	0.062 ± 0.059	0.086	0.558

Table 4.1: Performance of different methods used for CTR estimation on JSRT dataset.

Rotation-invariant CTR evaluation

To overcome this issue, the RI_CTR, described in Section 4.2.3, has been calculated from lung and heart segmentation masks. Results have been compared with RI_CTR calculated from ground truth annotation of lung and heart masks of JSRT dataset. Results are listed in Table 4.2. Using this new metric, the correlation coefficient is a lot higher than with

the traditional definition of CTR.

	AE (mean ± std)	RMSE	Corr. Coeff.
CTR	0.062 ± 0.059	0.086	0.558
RI_CTR	0.024 ± 0.021	0.031	0.754

Table 4.2: Performance of CTR and RI_CTR from lung and heart segmentation masks from JSRT dataset.

4.4 Conclusion

A method to estimate the CTR from lung and heart segmentation mask is presented in this chapter. From a small dataset, it was shown that a multi-label classification of lungs and heart improves the segmentation performance of the heart when compared to a segmentation model that has the heart as the only output. However, with this model, clipped anatomy's presence also affects the segmentation of the heart.

The heart segmentation was expected to help the CTR calculation being closer to its clinical definition. The results obtained are not in line with this hypothesis: worse performance was obtained in the estimation of the CTR. The CTR went from a MAE=0.038 when calculated using lung segmentation alone, to a MAE=0.062 when calculated using heart segmentation as well. However, this degradation is presumably mainly due to different ways of annotating the segmentation of the heart between the training and test datasets.

An alternative rotational-invariant method to calculate the CTR, RI_CTR is proposed. This new metric is strongly related to CTR, although not exactly the same. It seems to be less dependent from the different ways of annotating heart masks since the performance of the proposed method using this new metric on JSRT dataset in terms of AE, RMSE and correlation coefficient are higher. However, its clinical relevance is yet to be assessed.

Chapter 5

Clipping detection

The methods proposed in the previous chapters for the automatic calculation of the CTR allow it to be performed even when the lungs are clipped e.g., due to poor patient positioning during image acquisition. Assuming the clinical application of this tool, informing clinicians if the predicted CTR comes from clipped or non-clipped CXRs would be useful information to give an idea about the reliability of the estimated CTR. Moreover, an algorithm that is able to detect clipped lung on CXRs can find other possible applications. Some of them are listed below:

1. It could be used to assess the quality of CXRs, as mentioned in Section 2.2.1. This quality assessment could be used to inform the clinician before its evaluation but could also be used as an internal tool for hospitals to obtain data about the quality of the radiographs that they performed.

2. It could be used as a tool for training of radiologists, by giving direct feedback when clipped anatomy is detected.

3. Since in common clinical practice each CXR is usually retaken in case of clipped anatomy, it can be used to automatically reject such cases directly after the acquisition, saving time for clinicians from having to call the patient back at another time.

4. An automated rejection algorithm would also be useful in research field: in the work of [45] cases of clipped anatomy are manually excluded from their dataset. They stated that an automated rejection would be a useful tool and a good research direction.

There are some methods in the literature that address the issue of clipped anatomy in CXRs. First is by Wu et al. [67], who developed an AI model to interpret CXR based on 72 technical and anatomical core findings, including lungs not fully included in the image. Second is by Kashyap at al. [68], who proposed an automatic deep-learning method to detect "left costophrenic angle not included" and "lungs not fully included" together with

19 other technical deficiencies. Berge et al. [42], instead, address the problem of positioning of the lung area by calculating the distances of the lung region from the borders of the image. When there is no distance between the lung region and the border of the image, this approach may be in a sense a similar topic to clipped anatomy.

The objective of this chapter is described in Section 5.1. A description of the method used to detect clipped lung on CXRs can be found in Section 5.2. The method is then applied to clipped and non-clipped CRXs: details about the experiments and discussions are shown in Section 5.3. On Section 5.4 the conclusion can be found.

5.1 Objective

The aim of the work presented in this chapter is to develop a method to automatically detect the presence of clipped lung in CXRs by exploiting the ability of the U-Net with wider output (described in Section 3.2.3 as model 4). This model is able to predict the shape of the clipped part of the lungs even outside the field of view of the input image. Consequently, instances where the lung lies outside the field of view of the input image can be detected.

5.2 Methodology

The approach used to detect the presence of clipped lungs is shown in Figure 5.1 and can be summarized as follow:

1. Obtain the binary segmentation of the lung with a wider field of view with respect to the input image;

2. Obtain the bounding boxes of both lungs;

3. Calculate the distances of the corresponding bounding boxes from the border of the initial field of view of the image. When the border of the bounding box is outside the field of view the distance is taken as positive, while when the border of the bounding box is inside the field of view the distance is taken as negative;

4. Compare the maximum distance found with a fixed threshold, to classify if the lung is clipped or not.

To obtain the binary segmentation of the lungs, model 4 described in Section 3.2 was used, since it shows the best performance between the lung segmentation models discussed in this thesis. The choice of the threshold used for the classification is discussed later in Section 5.3.

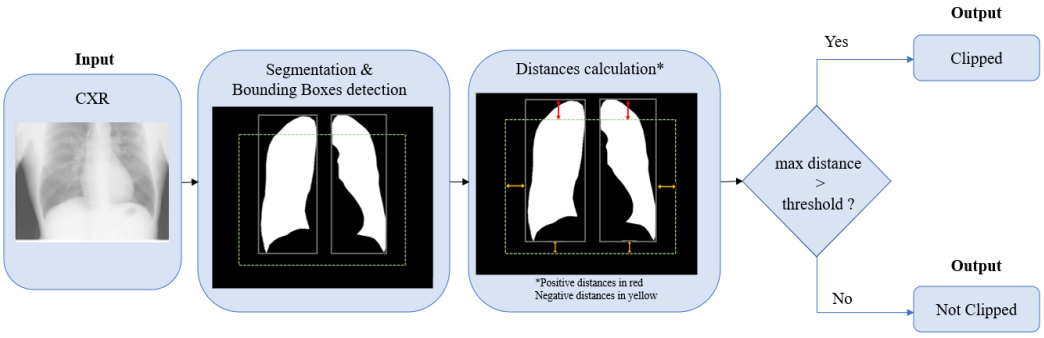

Figure 5.1: Overview of the methodology (clipping detection).

5.2.1 Dataset

The dataset used to train the lung segmentation model is described in Section 3.2 and it corresponds to the 75% of the 704 images from Shenzhen and Montgomery datasets (see Section 2.3.1), resulting in 528 CXRs. Starting from the same dataset, 264 CXRs (50%) have been randomly clipped, using the method described in Section 3.2.2. The remaining 50% is left unchanged. This balanced dataset has been used with training purpose to find the optimal threshold for clipping detection.

The dataset used for testing consists of 50% of non-clipped CXRs from JSRT dataset (described in Section 2.3.2) and 50% of clipped CXRs from Clipped JSRT dataset (described in Section 3.2.1), for a total of 256 CXRs.

5.3 Experimental results and Discussion

Initially, steps 1, 2 and 3 of the proposed method have been applied to a train dataset with 50% of clipped CXRs and 50% non-clipped CXRs (as described in the previous section). It results in a vector of 528 distances, 264 corresponding to clipped images and 264 corresponding to non-clipped images. The complete Receiver Operating Characteristic (ROC) curve which involves all possible classification thresholds [69], has been obtained. The ROC curve is shown in Figure 5.2. It plots the true positive rate (TPR) as a function of the false positive rate (FPR) at different thresholds. The TPR is the rate at which an actual clipped CXR will test as clipped and it is ideally equal to 1. The FPR is the rate at which a non-clipped CXR will test as clipped and it is ideally equal to 0. They are defined as:

$$TPR = \frac{TP}{TP + FN}, \tag{5.1}$$

$$FPR = \frac{FP}{FP + TN}, \tag{5.2}$$

where TP, TN, FP, FN are the number of true positive, true negative, false positive and false negative respectively. Following the most intuitively definition of clipped anatomy and setting the threshold to zero, a TPR=0.875 and an FPR=0.019 are obtained. However,

based on the ROC curve, different thresholds can be considered as optimal. Giving the equal cost to a FN error and a FP error, it was chosen to maximize the accuracy of the model. The accuracy of the model is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{5.3}$$

The threshold that maximizes the accuracy has been found equal to -5 and it is highlighted on the ROC curve. Using this threshold, a TPR =0.98 and a FPR=0.03 are obtained. The new threshold allows the accuracy of the method to increase from 0.93 to 0.98 on the training set. To better quantify the numbers of correct predictions and errors, the confusion matrix corresponding to the application of the two thresholds have been reported in Figure 5.3.

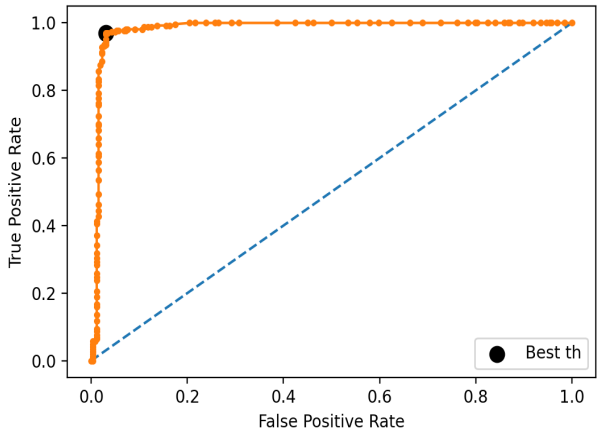

Figure 5.2: ROC curve of the training set of the clipped lung detection method. The threshold that maximizes the accuracy of the method is highlighted.

Actual Class

Th=0	Clipped	Non-Clipped
Clipped	**231**	**5**
Non-Clipped	**33**	**259**

Prediction

Actual Class

Th=-5	Clipped	Non-Clipped
Clipped	**260**	**8**
Non-Clipped	**4**	**256**

Prediction

Figure 5.3: Confusion matrix of the training set with threshold equal to zero (left) and threshold equal to -5 (right).

Subsequently, this method has been tested on 256 CXRs with balanced percentage of clipped and non-clipped images. The consistency of the method can be observed with the Area Under the ROC Curve (AUC). This is a robust method to evaluate the performance of a classifier since it relies on the complete ROC curve. The ROC curve obtained from the application of proposed method on the test set is visualised in Figure 5.4. The corresponding AUC is equal to 0.99. Applying the best threshold, a TPR=1 and a FRP=0.05 have been found. To better quantify the numbers of correct predictions and errors, the confusion matrix of the test set have been reported in Figure 5.5.

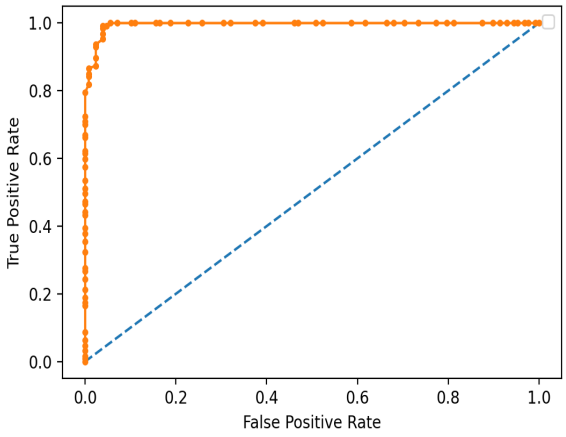

Figure 5.4: ROC curve of the test set for the clipped lung detection method.

Prediction \ Actual Class (Th=-5)	Clipped	Non-Clipped
Clipped	**128**	**7**
Non-Clipped	**0**	**121**

Figure 5.5: Confusion matrix of the test set for a threshold of -5.

5.4 Conclusion

The completeness of the CXR anatomy is normally verified by the clinicians before the radiographic interpretation. Clipped anatomy might make it difficult to interpret the image correctly, mainly for measurement such as CTR which requires the visualisation of the entire lung region. A method to automatically detect the presence of clipped lung in CXR is presented in this chapter. The proposed method was able to recognize all the clipped lung present in the test set, showing an optimal TPR, while 7 CXRs out of 128 non-clipped CXRs have been misclassified as clipped. The AUC presents a very high value, equal to 0.99, which shows good ability of a classifier to distinguish between two classes. In regard to clinical application of automatic CTR estimation in case of clipped anatomy (discussed in Chapters 3 and 4), this method would give useful information for the clinicians to better understand the reliability of the estimated CTR.

Chapter 6

How CTR changes with age and gender

As mentioned in Section 2.2.2, it is generally accepted that the upper limit of normal heart size correspond to the 50% of thoracic size. However, available literature states there are differences in normal CTR due to age and gender. Brakohiapa et al. [36], in their study analyzes 1047 CXRs of adults aged 21 to 80 years, showing significant age and gender-related differences in cardiac size parameters obtained from routine, frontal chest radiographs. Oberman et al. [70] reported a distinct increase of heart size with age until about 50 years after which the heart size appears stable, in a study of 3985 subjects aged 20 years or more, from Tecumseh, Michigan. These type of population studies are usually difficult to carry out on large scale, because of the need of clear and structured radiologists annotation for each image. For this reason, the automatic calculation of CTR can make this process faster and easily accessible.

It is important to clarify that the studies mentioned before consider only adult patients, even if CXRs is an important and valuable diagnostic tool also for pediatric population. Algorithms for automated CTR calculation trained on adults CXRs, however, may not accurately perform on pediatric case. The first reason is the anatomical shapes of lung and heart in pediatric patients: lungs appear smaller, and the cardiac silhouette is relatively larger, reaching values of CTR that in infants can approach the 0.6 [24]. Moreover, pediatric CXRs are usually noisier when compared to adult CXRs. This noise can be due to mother's handholding, an increased difficulty to obtain a good positioning of the patient [24], and also the tendency of using lower dose for younger patients. Since the work for the CTR estimation method described in the previous chapters does not takes into account variations in pediatrics' radiography, only adult patients are taken into account for this study.

The objective of this chapter is described in Section 6.1. In Section 6.2 the details

about the dataset used are given while in Section 6.3 the discussion of the obtained results can be found.

6.1 Objective

The aim of this study is to investigate the variation in CTR values according to age and gender, on a large dataset. The aim of this research is also to propose an application of the automatic CTR estimation method already discussed in this thesis.

6.2 Dataset

Many public large datasets with CXRs exist and among them the CheXpert [71] dataset was selected for this study. CheXpert is a large public dataset for chest radiograph interpretation, consisting of 224,316 chest radiographs of 65,240 patients. The CXRs and their associated radiology reports were retrospectively collected from Stanford Hospital from acquisitions performed between October 2002 and July 2017 in both inpatient and outpatient centers. A detailed description of the dataset is presented in [72]. Each report was labeled for the presence of 14 observation (12 different pathologies in addition to no finding and support devices) as positive, negative, or uncertain. The labels are extracted from the free text radiology report thanks to an automated rule-based labeler. The patient biological sex and age are available for each image, along with the information whether the image is frontal or lateral. For frontal images, the information on the projection type is reported (AP or PA, described in Section 2.1.4.). The dataset is available in two versions: a large version (440 GB) with DICOM images and a small version (11 GB) with JPEG images. The small version has been used in this research. Each image is downsampled to approximately 390×320 pixels and grayscale downsampled to 256 levels.

In CheXpertCTR, a subsample of the whole CheXpert dataset was selected to be suitable for the population study in this research. It has ideally no enlargement of the heart and images taken on the gold standard CXR projection for the evaluation of CTR. For this reason, only PA images were extracted, and images labeled as positive to cardiomegaly and enlarged cardiomediastinum were excluded. It results in 25,369 CXRs, with 34% female and 66% male. The age of the patients raged from 18 to 90 years, with a mean of 56.5 ± 17.1. Number of images per age and per gender is shown in Figure 6.1.

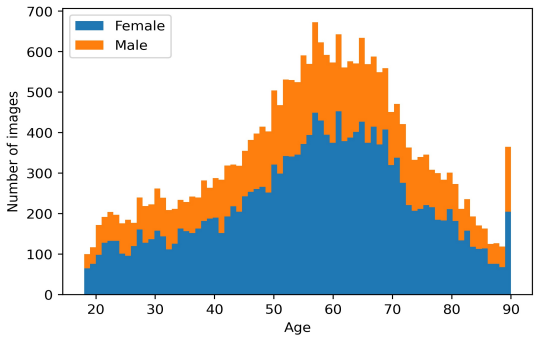

Figure 6.1: Number of CXRs per age and gender in the CheXpertCTR dataset.

6.3 Experimental results and discussion

The best method discussed in this thesis to predict CTR has been applied to the CheXpertCTR dataset. The method is described in Chapter 3, and it is based on a lung segmentation model (U-Net architecture with VAE, able to output a wider field of view compared to the CXR input, called model 4 in Chapter 3). From the lung segmentation mask, cardiac and thoracic diameter are extracted and the CTR is calculated as described in Section 3.2.4. This method showed an AE on the calculation of the CTR of 0.038± 0.040 (as reported in experiments from Section 3.4).

To evaluate how the predicted CTR changes with age, CTR values from CheXpertCTR dataset (described in Section 6.2) have been calculated. One downside of using this dataset is that only the 21% of the selected CXRs presents "No Finding" label marked as true, meaning that no pathology is classified as positive or uncertain. The remaining CXRs have at least one pathology marked as positive or uncertain and could potentially influence the quality of the visualisation of the chest's anatomical structures and thus the quality of the CTR automatic estimation. The robustness of the proposed model, to such variations in fact, has not been tested and could be the scope of future research. However, pathologies directly correlated with an enlargement of the heart region have been excluded from the initial CheXpert dataset, trying to select CXRs with presumably a normal range of CTRs.

The CTR from all the 25,369 CXRs from CheXpertCTR, reported a mean and a standard deviation of 0.498 ± 0.089. A general increase in mean CTR is reported as the age of the patient increases: from 0.448 in 18-year-old patients to 0.562 in 90-year-old patients, showing an increase of the 25%. (as illustrated in Figure 6.2). The increase of CTR appears to be linear. A linear interpolation shows a correlation coefficient of 0.98. This increment agrees with the work of Brakohiapa et al. [73], that showed how CTR increased statistically with age of patients.

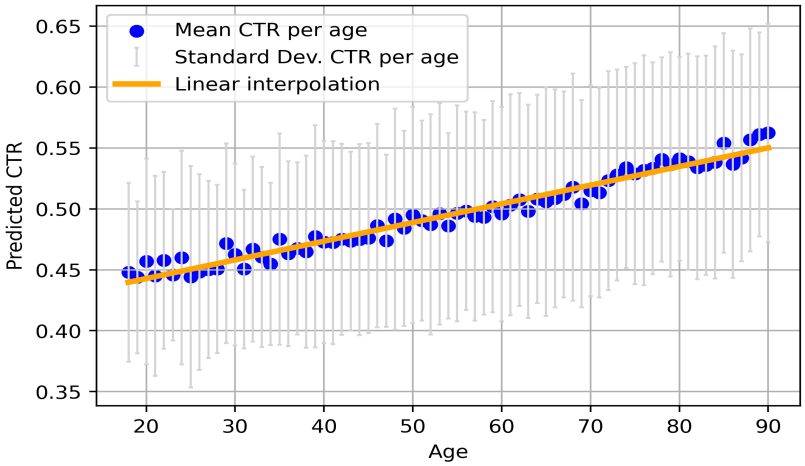

Figure 6.2: Predicted CTR as a function of patient age on CheXpertCTR dataset.

The distributions of male and female CTR values are shown in Figure 6.3. A mean value and a standard deviation of 0.507 ± 0.094 and 0.492 ± 0.085 has been reported for female and male respectively, showing a slightly higher mean CTR for females. A Z-test has been performed to see if there was a statistical difference between the two groups and a p-value of $2.14 * 10^{-36}$ has been found. Since the p-value is much lower than the alpha-value set to 0.5, it is possible to conclude that there is a significant difference in the two distributions. This result agrees with the work of [36] who also reported a statistically significant difference in the overall CTR between female and male. In their work a mean value and a standard deviation of 0.448 ± 0.037 and 0.447 ± 0.037 have been reported for female and male respectively. A slightly higher mean CTR value for female then male has been reported also by [74], who considered the heart size of 306 patients.

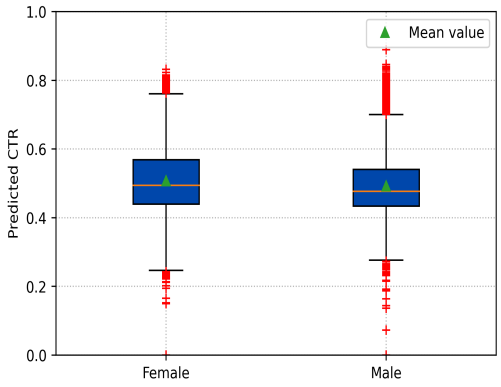

Figure 6.3: Boxplot of predicted CTR on female and male CXRs from CheXpertCTR dataset.

Since a significant difference was reported between the CTR values of male and female, the two groups were considered separately to study the relationship between the CTR value and age. CTR values from males and females reported slightly different trends with age in this work. Figure 6.4 shows how the CTR changes according to age and gender. Both the trends appear linear, and show similar mean values for younger patients while females reach higher mean CTR values as age increases. This trend agrees with the results from the work of Brakohiapa et al. [73]. In their study, the CTR in males increased from the 21-40 years group (0.464 ± 0.040) to the 41-60 years group and increased again minimally in the above 60 years group (0.474 ± 0.048). The CTR in females increased more, from the 21-40 years group (0.0464 ± 0.045) to the above 60 years group (0.0517 ± 0.037).

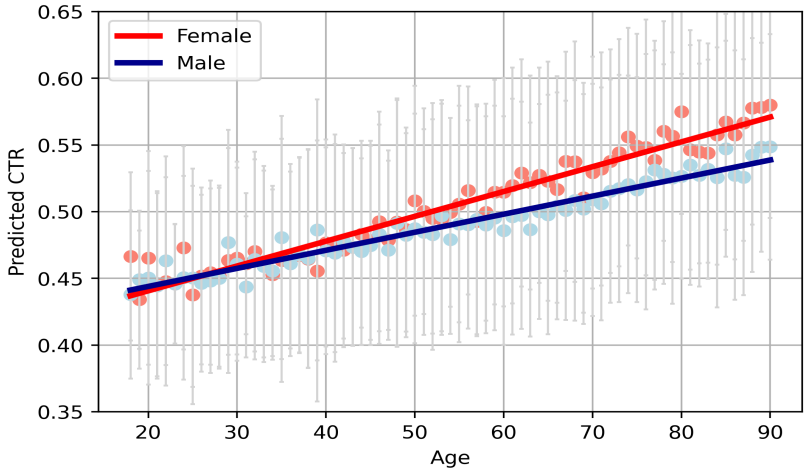

Figure 6.4: Predicted CTR as a function of patient age on female and male CXRs from CheXpertCTR dataset.

6.4 Conclusion

An automatic CTR calculation method has been applied on a large CXRs dataset to evaluate how this measurement, linked with the cardiac enlargement, changes with age and gender. Tested on 25,369 CXRs, an increment from 0.448 to 0.562 on mean CTR value from 18 to 90-year-old has been found, with an overall mean value of 0.498. A significant difference between male and female CTR values has been reported. The difference between the mean CTR value for males and females becomes more important as age increases, showing higher values of mean CTR for female patients. The conclusion obtained on this study reflects the observations of preliminary studies on age-dependency of CTR measurements by supporting the hypothesis that the proposed method for CTR evaluation could be reliable for a large dataset study. Moreover, this study shows the that a general threshold of 0.5 to detect cardiomegaly would not be consistent with normal variations between different ages and genders that the CTR can have.

Chapter 7

Conclusion

CXR has shown to be one of the most complex imaging modalities to interpret, with an high inter-reader variability which could have been caused by varying years of experience of the clinicians. For this reason, many CAD systems have been found in the literature to help clinician's evaluation by showing directly the detected disease. However, such systems can cause an over-reliance on the technology and the tendency to blindly accept the CAD system output.

In order to cope with this problem, methods to perform objective measurement from CXRs have been studied, focusing on the estimation of the CTR, which correlates the size of the heart to the size of the chest and can be an indicator of cardiomegaly when the value is too high. Before doing any evaluation, a quality check of the image is always done by radiologists to assure that the whole lungs are in the field of view of the image, otherwise the radiograph is usually rejected and repeated. The extraction of CTR information when clipped anatomy occurs, is also investigated.

Similar to most of the work in the literature, segmentation-based approaches are proposed in this work. Two segmentation-based approaches were developed. The first segmentation-based solution assumes the possibility to calculate the CTR using only the lungs segmentation. It showed the possibility to calculate the CTR with errors in the same order of magnitude as the state-of-art method for CTR estimation from lung segmentation prediction in the literature, with the advantage that it also deals with clipped anatomy. Subsequently, the heart segmentation has been added in the model to calculate the CTR following its medical definition. However, it showed a decrease of performance for the CTR evaluation due to the variation found in the labeling of the training and testing dataset.

As previously mentioned, quality check is an important step prior to computing CTR from a CXR. Based on the first segmentation based solution, a method to automatically detect the presence of clipped lung in CXR have been proposed showing promising results.

This quality check can be presented as additional information to the clinicians to warn them when the CTR was computed from CXR with clipped anatomy. Consequently, a new metric is proposed as an alternative to the classical CTR. The new metric, RI_CTR is a rotation-invariant version of the CTR where the heart size is measured by the diameter of the largest inscribed circle of the heart mask and the thorax diameter is taken from the largest span of the major axis aligned lung masks. It seems to be less dependent from the different ways of annotating heart masks since the performance of the proposed method in terms of AE, RMSE and correlation coefficient are higher. However, its clinical relevance is yet to be assessed.

As a practical application, the proposed automatic evaluation of CTR has been applied on a large dataset to study the dependency of CTR measurement on age and gender. An increment on mean CTR value from 18 to 90-year-old has been observed, starting from 0.445 to 0.562. A statistical difference between male and female CTR values has been reported. This difference becomes more important when age increases, showing higher vales of CTR for female patients.

7.1 Future works

One possible step on the future work of this subject could be improving the calculation of the CTR from lung and heart segmentation, since it theoretically better reflects the evaluation carried out by clinicians. To do so, it would be necessary to obtain training and testing set that are annotated in a consistent way or improve on the generalization of the heart segmentation methods. The RI_CTR proposed in Chapter 4 calculated from heart and lung segmentations, is a metric strongly related to CTR, that seems to be less dependent on heart annotation by having a high correlation coefficient when calculated between the ground truth and the predicted segmentation masks. A study on its clinical relevance would be a possible research direction. Moreover, the proposed CTR estimation from lung segmentation method has some limitation that can be addressed in future works. First of all, it is applicable only to adult CXRs due to the availability of only adult CXR datasets with lung and heart segmentation. Secondly, the proposed algorithm considers only the case of clipped anatomy but many other quality issues (such as performed at fully inspiration, correct patient orientation, positioning of the shoulder blades the lung field, etc.) need to be checked and could be taken into account by the algorithm before assessing the CTR from a CXR (as listed in Section 2.2.1).

Bibliography

[1] Ronan. CC BY-SA 3.0, via Wikimedia Commons.

[2] Your practice online, https://www.ypo.education/medical-tests/x-ray-t291/video/.

[3] D. G. Lloyd-Jones, "Chest x-ray anatomy." Salisbury NHS Foundation Trust UK - www.radiologymasterclass.co.uk.

[4] Case courtesy of Assoc Prof Frank Gaillard, Radiopaedia.org, rID: 7142.

[5] R. Muthiah, "Rheumatic aortic valve disease with mitral stenosis—a case report," *Case Reports in Clinical Medicine*, vol. 05, pp. 268–295, 01 2016.

[6] A. J. Nicol, P. H. Navsaria, S. Beningfield, and D. Kahn, "A straight left heart border: A new radiological sign of a hemopericardium," *World Journal of Surgery*, vol. 38, no. 1, p. 211–214, 2013.

[7] J. Lee, H. Ahn, and C. Yoon, "Severe sinus bradycardia requiring cardiac pacing in a lupus patient who was successfully treated using methylprednisolone pulse therapy," *The Korean Journal of Medicine*, vol. 94, pp. 225–229, 04 2019.

[8] K. Truszkiewicz, R. Poreba, and P. Gać, "Radiological cardiothoracic ratio in evidence-based medicine," *Journal of Clinical Medicine*, vol. 10, no. 9, p. 2016, 2021.

[9] A. Mittal, R. Hooda, and S. Sofat, "Lung field segmentation in chest radiographs: a historical review, current status, and expectations from deep learning," *IET Image Process.*, vol. 11, pp. 937–952, 2017.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[11] C. Barsanti, F. Lenzarini, and C. Kusmic, "Diagnostic and prognostic utility of non-invasive imaging in diabetes management," *World journal of diabetes*, vol. 6, no. 6, p. 792, 2015.

[12] E. Bercovich and M. C. Javitt, "Medical imaging: from roentgen to the digital revolution, and beyond," *Rambam Maimonides medical journal*, vol. 9, no. 4, 2018.

[13] E. C. Lin, "Radiation risk from medical imaging," in *Mayo Clinic Proceedings*, vol. 85, pp. 1142–1146, Elsevier, 2010.

[14] J. G. Jones, C. N. Mills, M. A. Mogensen, and C. I. Lee, "Radiation dose from medical imaging: a primer for emergency physicians," *Western Journal of Emergency Medicine*, vol. 13, no. 2, p. 202, 2012.

[15] M. Nauenberg, "Max planck and the birth of the quantum hypothesis," *American Journal of Physics*, vol. 84, no. 9, pp. 709–720, 2016.

[16] D. Chang, "Physical interpretation of planck's constant based on the maxwell theory," *Chinese Physics B*, vol. 26, p. 040301, 04 2017.

[17] D. Calloway, "Beer-lambert law," *Journal of Chemical Education*, vol. 74, no. 7, p. 744, 1997.

[18] A. Tafti and D. W. Byerly, "X-ray image acquisition," in *StatPearls [Internet]*, StatPearls Publishing, 2021.

[19] S. V. Musolino, J. DeFranco, and R. Schlueck, "The alara principle in the context of a radiological or nuclear emergency," *Health physics*, vol. 94, no. 2, pp. 109–111, 2008.

[20] C. Schaefer-Prokop, U. Neitzel, H. Venema, M. Uffmann, and M. Prokop, "Digital chest radiography: An update on modern technology, dose containment and control of image quality," *European radiology*, vol. 18, pp. 1818–30, 05 2008.

[21] M.-A. Jones, J., "Chest radiograph," *Reference article, Radiopaedia.org.*

[22] S. Candemir, S. Rajaraman, G. Thoma, and S. Antani, "Deep learning for grading cardiomegaly severity in chest x-rays: an investigation," in *2018 IEEE Life Sciences Conference (LSC)*, pp. 109–113, IEEE, 2018.

[23] A. C. of Radiology, "Understanding radiation risk from imaging tests. radiological society of north america. children and radiation safety.,"

[24] S. Candemir and S. Antani, "A review on lung boundary detection in chest x-rays," *International journal of computer assisted radiology and surgery*, vol. 14, no. 4, pp. 563–576, 2019.

[25] C. M. Walsh-Kelly, M. D. Melzer-Lange, H. M. Hennes, P. Lye, M. Hegenbarth, J. Sty, and R. Starshak, "Clinical impact of radiograph misinterpretation in a pediatric ed and the effect of physician training level," *The American journal of emergency medicine*, vol. 13, no. 3, pp. 262–264, 1995.

[26] D. G. Lloyd-Jones, "Chest x-ray quality projection." Salisbury NHS Foundation Trust UK - www.radiologymasterclass.co.uk.

[27] R. S. Crausman, "The abcs of chest x-ray film interpretation," *Chest*, vol. 113, no. 1, p. 256, 1998.

[28] B.-D. Ryu, Y., "Chest radiograph assessment using abcdefghi. reference article, radiopaedia.org.."

[29] Physiopedia, "Chest x-rays — physiopedia,," 2020.

[30] Physiopedia, "Lung anatomy — physiopedia,," 2022.

[31] "Cardiac silhouette: https://radiopaedia.org/articles/cardiac-silhouette."

[32] P. D. Maldjian and M. Saric, "Approach to dextrocardia in adults: Review," *American Journal of Roentgenology*, vol. 188, no. 6_supplement, pp. S39–S49, 2007.

[33] Z. Li, Z. Hou, C. Chen, Z. Hao, Y. An, S. Liang, and B. Lu, "Automatic cardiothoracic ratio calculation with deep learning," *IEEE Access*, vol. 7, pp. 37749–37756, 2019.

[34] A. L. Goldberger, Z. D. Goldberger, and A. Shvilkin, "Chapter 6 - atrial and ventricular enlargement," in *Goldberger's Clinical Electrocardiography (Eighth Edition)* (A. L. Goldberger, Z. D. Goldberger, and A. Shvilkin, eds.), pp. 44–53, Philadelphia: W.B. Saunders, eighth edition ed., 2013.

[35] "Enlarged heart." Mayo Foundation for Medical Education and Research (MFMER) - www.mayoclinic.org.

[36] E. K. Brakohiapa, B. O. Botwe, and B. D. Sarkodie, "Gender and age differences in cardiac size parameters of ghanaian adults: Can one parameter fit all? part two," *Ethiopian Journal of Health Sciences*, vol. 31, no. 3, 2021.

[37] M. S. Brown, L. S. Wilson, B. D. Doust, R. W. Gill, and C. Sun, "Knowledge-based method for segmentation and analysis of lung boundaries in chest x-ray images," *Computerized medical imaging and graphics*, vol. 22, no. 6, pp. 463–477, 1998.

[38] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. `http://www.deeplearningbook.org`.

[39] C. Qin, D. Yao, Y. Shi, and Z. Song, "Computer-aided detection in chest radiography based on artificial intelligence: a survey," *Biomedical engineering online*, vol. 17, no. 1, pp. 1–23, 2018.

[40] M. Sujan, C. Baber, P. Salomon, R. Pool, N. Chozos, C. Aceves-González, M. Cooke, C. Escobar-Galino, C. Flashman, G. Frau, R. Hawkins, H. Hughes, G. Mejia, G. Kaya, B. Kirby, I. C. Landa-Avila, K. Laudanski, P. Lewis, F. Magrabi, and S. White, "Human factors and ergonomics in healthcare ai," 09 2021.

[41] E. Obikili and I. Okoye, "Transverse cardiac diameter in frontal chest radiographs of a normal adult nigerian population," *Nigerian Journal of Medicine*, vol. 14, no. 3, pp. 295–298, 2005.

[42] J. von Berg, S. Krönke, A. Gooßen, D. Bystrov, M. Brück, T. Harder, N. Wieberneit, and S. Young, "Robust chest x-ray quality assessment using convolutional neural networks and atlas regularization," in *Medical Imaging 2020: Image Processing*, vol. 11313, pp. 391–398, SPIE, 2020.

[43] J. H. E. Carmichael, "European guidelines on quality criteriafor diagnostic radiographic images," Officer for Official Publication of the European Communities, 1996.

[44] J. Whaley, B. Pressman, J. Wilson, L. Bravo, J. Sehnert, and D. Foos, "Investigation of the variability in the assessment of digital chest x-ray image quality," *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology*, vol. 26, 08 2012.

[45] E. Sogancioglu, K. Murphy, E. Calli, E. Scholten, S. Schalekamp, and B. Ginneken, "Cardiomegaly detection on chest radiographs: Segmentation versus classification," *IEEE Access*, vol. PP, pp. 1–1, 05 2020.

[46] T. Gupte, M. Niljikar, M. Gawali, V. Kulkarni, A. Kharat, and A. Pant, "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)*, pp. 1–6, IEEE, 2021.

[47] H. Becker, W. Nettleton, P. Meyers, J. Sweeney, and C. Nice, "Digital computer determination of a medical diagnostic index directly from chest x-ray images," *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 67–72, 1964.

[48] M. Arsalan, M. Owais, T. Mahmood, J. Choi, and K. R. Park, "Artificial intelligence-based diagnosis of cardiac and related diseases," *Journal of Clinical Medicine*, vol. 9, no. 3, p. 871, 2020.

[49] N. Dong, M. C. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. P. Xing, "Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio," in *MICCAI*, 2018.

[50] A. H. Dallal, C. Agarwal, M. R. Arbabshirani, A. Patel, and G. Moore, "Automatic estimation of heart boundaries and cardiothoracic ratio from chest x-ray images," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, pp. 134–143, SPIE, 2017.

[51] I. Chamveha, T. Promwiset, T. Tongdee, P. Saiviroonporn, and W. Chaisangmongkon, "Automated cardiothoracic ratio calculation and cardiomegaly detection using deep learning approach," *arXiv preprint arXiv:2002.07468*, 2020.

[52] C. Balakrishna, S. Dadashzadeh, and S. Soltaninejad, "Automatic detection of lumen and media in the ivus images using u-net with vgg16 encoder," *arXiv preprint arXiv:1806.07554*, 2018.

[53] M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, "Abnormality detection and localization in chest x-rays using deep convolutional neural networks," *arXiv preprint arXiv:1705.09850*, 2017.

[54] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, "Comparison of deep learning approaches for multi-label chest x-ray classification," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[55] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, 2014.

[56] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 577–590, 2014.

[57] S. Jaeger, A. Karargyris, S. Candemir, J. Siegelman, L. Folio, S. Antani, and G. Thoma, "Automatic screening for tuberculosis in chest radiographs: a survey," *Quantitative imaging in medicine and surgery*, vol. 3, no. 2, p. 89, 2013.

[58] S. Stirenko, Y. Kochura, O. Alienin, O. Rokovyi, Y. Gordienko, P. Gang, and W. Zeng, "Chest x-ray analysis of tuberculosis by deep learning with segmentation and augmentation," pp. 422–428, 2018.

[59] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000. PMID: 10628457.

[60] B. van Ginneken, M. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006.

[61] Q. Que, Z. Tang, R. Wang, Z. Zeng, J. Wang, M. Chua, T. S. Gee, X. Yang, and B. Veeravalli, "Cardioxnet: automated detection for cardiomegaly based on deep learning," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 612–615, IEEE, 2018.

[62] R. Selvan, E. B. Dam, N. S. Detlefsen, S. Rischel, K. Sheng, M. Nielsen, and A. Pai, "Lung segmentation from chest x-rays using variational data imputation," 2020.

[63] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013.

[64] J. Subramanian and R. Simon, "Overfitting in prediction models–is it a problem only in high dimensions?," *Contemporary clinical trials*, vol. 36, no. 2, pp. 636–641, 2013.

[65] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.

[66] J. M. Joyce, *Kullback-Leibler Divergence*, pp. 720–722. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[67] J. T. Wu, K. C. Wong, Y. Gur, N. Ansari, A. Karargyris, A. Sharma, M. Morris, B. Saboury, H. Ahmad, O. Boyko, *et al.*, "Comparison of chest radiograph interpretations by artificial intelligence algorithm vs radiology residents," *JAMA network open*, vol. 3, no. 10, pp. e2022779–e2022779, 2020.

[68] S. Kashyap, M. Moradi, A. Karargyris, J. Wu, M. Morris, B. Saboury, E. Siegel, and T. Syeda-Mahmood, "Artificial intelligence for point of care radiograph quality assessment," 02 2019.

[69] F. Melo, *Area under the ROC Curve*. New York, NY: Springer New York, 2013.

[70] A. Oberman, A. R. Myers, T. M. Karunas, and F. H. Epstein, "Heart size of adults in a natural population-tecumseh, michigan: Variation by sex, age, height, and weight," *Circulation*, vol. 35, no. 4, pp. 724–733, 1967.

[71] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 590–597, 2019.

[72] C. Garbin, P. Rajpurkar, J. Irvin, M. P. Lungren, and O. Marques, "Structured dataset documentation: a datasheet for chexpert," *arXiv preprint arXiv:2105.03020*, 2021.

[73] E. K. K. Brakohiapa, B. O. Botwe, B. D. Sarkodie, E. K. Ofori, and J. Coleman, "Radiographic determination of cardiomegaly using cardiothoracic ratio and transverse cardiac diameter: Can one size fit all? part one," *The Pan African Medical Journal*, vol. 27, 2017.

[74] G. Oladipo, P. Okoh, E. Kelly, C. Arimie, and B. Leko, "Normal heart sizes of nigerians within rivers state using cardiothoracic ratio," *Scientia Africana*, vol. 11, no. 2, 2012.

[75] A. ALLEA, "The european code of conduct for research integrity, revised edition," *Berlin2017 [Available from: http://www. allea. org/wp-content/uploads/2017/04/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017. pdf*, 2017.

[76] A. Hleg, "Ethics guidelines for trustworthy ai," *B-1049 Brussels*, 2019.

Appendix A

Ethics and ethical implications of the work

Ethic refers to the moral principles that govern the conduct of certain activities or a person's behavior. One possible application of ethic is research ethic: it involves the application of fundamental ethical principles to research activities. Research ethic firstly includes ethical principle such as the protection of the right, dignity and welfare of anyone involved in the research. Moreover, the research must be conducted in a transparent and independent manner. Ethical considerations fall into the research of integrity (RI). Research integrity is a framework that discusses the attitude of researchers according to appropriate ethical, legal and professional frameworks, obligations and standards. The European Code for Research Integrity [75] or ALLEA code provides full guidance for researchers, describing an approach for conducting good scientific work.

The main ethical implication of this work regards ethical principle on AI. At European level the "EU ethic guidelines for trustworthy AI" [76], prepared by the High-Level Expert Group on Artificial Intelligence (AI HLEG), have been recognized as the guiding ethics principle on AI. The most important and applicable key requirements listed in the EU ethic guidelines are taken into account and discussed.

The primarily application of the automatic CTR calculation method discussed in this thesis, would be a decision support system for clinicians. The output of the algorithm would influence human decision-making process and consequently someone's health or well-being. To follow the ethic guidelines, the AI system must support human decision making by enabling users to make informed autonomous decisions. This refers to the principle of **human agency and oversight** and it has been taken into account since the initial idea of the project: the system output the CTR value, without inferring any autonomous decision regard the presence or absence of related diseases, such as cardiomegaly. Moreover, for a system to be trustworthy it should be able to explain why it behaved a

certain way and provided a given output. This is still an open challenge for system based on neural network such as the one proposed. However, the proposed method for estimating the CTR is based on the calculation of both the cardiac and thoracic diameter on the CXR image: it is possible, in a practical application of the method, to show to the clinician both the diameters as lines superimposed on the newly acquired CXR image. The correctness of the proposed heart and thoracic diameters can be checked by the clinician. This will increase the degree of interaction between user and AI application by support situation awareness and potentially reduce unwarranted over-reliance on AI. A system to detect cases in which the CTR has been estimated from clipped lung CXRs has been proposed, with the aim of informing the clinician about the reliability of the CTR estimation. This information would also increase the ability of the clinician to make informed autonomous decisions.

Privacy and data protection throughout the system's life-cycle is a second element to consider, following the ethic guidelines. However, the presented work does not involve primary data collection. The datasets used for training, validation and testing purpose were downloaded from publicly available datasets. Any reference to the datasets can be found in the thesis.

All the information about the method of training the algorithm, including which input data was selected, together with the information about the data used to test and validate can be found in the thesis, with the aim of carry out a **transparent** research.

Lastly, the **environmental impact** of the AI system development is another aspect that should be considered following the European guidelines. It has not been directly measured but some considerations can be done. The heaviest step should be the development of the segmentation model since it involves neural network training. As discussed in this thesis, the proposed segmentation-based method is a modified version of the work by Selvan et al. [62]. The environmental impact of the proposed method can be assumed to be almost comparable to the environmental impact of their work. They reported an average CO2 footprint of developing and training their baseline and proposed model of around 7.3 Kg (equivalently about 60 km traveled by a car).