Convolutional neural network architectures and aggregation information models for pulmonary X-ray segmentation *

Antonio Nadal-Martínez
University of the Balearic Islands
antonio.nadal@uib.cat

Lidia Talavera-Martínez, Marc Munar, Manuel González-Hidalgo
SCOPIA Research Group
University of the Balearic Islands (UIB)
Health Research Institute of the Balearic Islands (IdISBa)
Laboratory of Artificial Intelligence Applications (LAIA@UIB)
l.talavera, marc.munar, manuel.gonzalez@uib.es

Abstract

This study explores aggregation and consensus methods to combine lung segmentations from various neural network models in X-ray images, aiming to enhance accuracy and completeness. Through extensive experimentation, the research identifies the most effective aggregation method, with WOWA aggregation and a maximum-based consensus approach outperforming individual models. This underscores the importance of aggregation techniques in optimizing anatomical structure segmentation in medical imaging.

1. Introduction

Segmenting medical images is crucial. It allows precise identification and isolation of regions of interest in medical data. Many imaging techniques like computed tomography, magnetic resonance imaging, and ultrasound offer unique insights into anatomical structures and pathologies. Among these, radiography is widely used due to its cost-effectiveness and diagnostic utility. Despite traditional methods (e.g., thresholding, clustering, region growing, and edge detection) being effective in different scenarios [9, 10, 13, 18], they often struggle with the complexities and variabilities present in medical images [3]. However, CNNs have revolutionized segmentation techniques [11], primarily because of their ability to learn feature representations from extensive datasets automatically. CNN-based methods, including architectures like residual networks [5], fully connected convolutions [4], and UNet-based models [12, 21], have shown significant improvements in segmenting pulmonary regions from radiographic images. However, not all methods offer the same accuracy, and there may be imprecision. So, our main goal is to evaluate the effectiveness of combining results from many CNN models using aggregation and consensus methods to improve pulmonary region segmentation in radiography.

In this study, we selected seven distinct CNN models: UNet [14, 17], UNetPre [12], GSC [6], ERFNet [16], LinkNet [1], ESNet [23], and CGNet [24]. In addition to selecting appropriate CNN architectures, interpreting their outputs is crucial for effective aggregation. To ensure compatibility, we normalized outputs into the range $[0, 1]$. We decided to use aggregation methods, such as Ordered Weighted Averaging (OWA) and Weighted Ordered Weighted Averaging (WOWA), to combine the outputs of multiple CNN models. OWA methods allow the emphasis of specific value ranges, with weights obtained from regular monotone increasing quantifiers [25] enabling flexibility in the aggregation process. WOWA methods [22] extend OWA by adding order relations among prediction models. It does this by using two weight vectors. One is for traditional weighting and the other is to reflect model importance. Conversely, consensus methods [8] determine the optimal aggregation approach. These methods can operate at both image and pixel levels, and improve the combined segmentation accuracy by using the knowledge of multiple models. In this study, we propose two pixel-level consensus methods, which provide more flexible results. The first,

*This work was partially supported by the R+D+i Project PID2020-113870GB-I00-"Desarrollo de herramientas de Soft Computing para la Ayuda al Diagnóstico Clínico y a la Gestión de Emergencias (HESOCODICE)", funded by MCIN/AEI/10.13039/501100011033/.

argmax, selects, at each pixel, the maximum value among the outputs of several CNN models. The second, mean, finds the mean value across all the considered outputs. We aim to improve lung segmentation in radiographic images, which is essential for aiding precise diagnoses. This will be done by using aggregation and consensus techniques to obtain a more accurate and complete segmentation.

2. Methodology

Database We considered publicly annotated databases with frontal chest X-rays and diverse visual traits when constructing the dataset as we aimed to improve the model's ability to generalize. These databases were JSRT [20] with 247 images, Montgomery [7] with 138 images, and the COVID-19 Radiography Database [2] with 2555 images. Examples of these databases are shown in Figure 1.

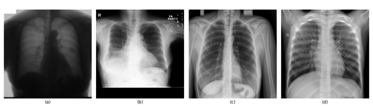


Figure 1. Examples from the dataset used in the study: (a) JSRT, (b) Montgomery, (c) and (d) COVID-19 Radiography Database.

Methodology Initially, see Figure 2, each of the seven considered CNNs generates segmentations for the pulmonary area. Then, a statistical comparison helps to rank the performance of the models. Afterward, we apply aggregation using OWA and WOWA approaches along with two consensus methods (argmax and mean). Finally, through a statistical analysis, we identify the optimal aggregation method for each consensus method and evaluate if aggregation improves individual CNN results.

Experiments Following individual CNN model segmentations, we tested ways of combining them, namely: OWA 1, described in section 1; OWA 2, which uses only the top

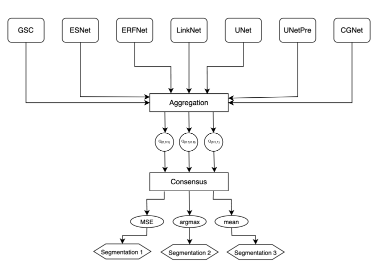


Figure 2. Proposed approach pipeline: Seven CNN networks segment pulmonary regions from X-ray images. Outputs are normalized and combined using aggregation methods (OWA and WOWA) with consensus methods (argmax and mean).

4 models' segmentations; OWA 3, which includes all segmentations except the worst model; and WOWA 1, which assigns weights based on model's performance order. So, with n being the number of models, the i-th model gets the weight $\frac{i}{\sum_{j=1}^{n} j}$; WOWA 2, same as WOWA 1 but without the worst model; WOWA 3, same as WOWA 1 but the best model gets a weight of $\frac{1}{2}$, while the others get $\frac{1}{2(n-1)}$, with n being the number of models used.

Implementation details We divided the database into 80% for training and 20% for testing while maintaining the proportions of images from each database in both sets. We implemented and trained architectures from scratch using Keras. We used a batch size of 16, Adam optimizer with a learning rate of 10^{-4}, and the Dice loss function defined as $D_{\text{loss}}(\boldsymbol{p}, \boldsymbol{q}) = 1 - D(\boldsymbol{p}, \boldsymbol{q})$. Here, we compute $D(\boldsymbol{p}, \boldsymbol{q})$ as

$$D(\boldsymbol{p}, \boldsymbol{q}) = \frac{2 \cdot \sum_{x,y}(p_{x,y} \cdot q_{x,y})}{\sum_{x,y}(p_{x,y}^2 + q_{x,y}^2)},$$

where $p_{x,y}$ and $q_{x,y}$ refer to the value of pixel (x, y) in the prediction $\boldsymbol{p}$ and in the ground truth mask $\boldsymbol{q}$, respectively. The range of each $p_{x,y}$ is the unit interval $[0, 1]$, while $q_{x,y}$ is binary and can only take the values 0 or 1. The training involved early stopping based on loss function monitoring, with weights restored to the best epoch.

Evaluation Qualitative and quantitative analyses evaluated the quality of the results of the different experiments using common performance measures. These include Accuracy, Jaccard index, and Sensitivity. We used Welch's t-test with a significance level α of 0.05 as a statistical test to compare the average performance metrics between experiments. The test assumptions were met due to the large sample size. It ensured a normal distribution of the variable, as the central limit theorem says. Additionally, other test requirements were satisfied as the experiment used independent and randomly selected samples.

We classify the statistical results using the following criterion. If the population mean of the model in the row is better than that of the model in the column: ✓✓. If we can't rule out equality between means, but the sample median in the row is better than in the column:✓. If we can't rule out equal population means, and the sample median in the row is worse than that in the column: ✗ If the population mean of the model in the row is worse than that of the model in the column: ✗✗.

3. Results

Quantitative analysis Firstly, we evaluated several CNNs for segmenting the pulmonary region in X-ray images. We summarized our analysis of the results from each

		GSC	ESNet	ERFNet	LinkNet	UNetPre	UNet	CGNet
Accuracy	Mean	**0,9916**	0,9891	0,9892	0,9893	0,9852	0,9873	0,9684
	Std	**0,0074**	0,0092	0,0128	0,01	0,0085	0,0147	0,0133
	Median	**0,9942**	0,9923	0,9926	0,9924	0,9878	0,9917	0,9718
Jaccard	Mean	**0,9652**	0,9559	0,956	0,9559	0,9393	0,9467	0,8753
	Std	**0,0303**	0,035	0,0483	0,0417	0,0358	0,0598	0,0589
	Median	**0,9739**	0,9662	0,9679	0,9663	0,949	0,9646	0,891
Sensitivity	Mean	0,9772	**0,983**	0,9781	0,9762	0,9714	0,9595	0,9376
	Std	0,0247	**0,0171**	0,0285	0,0329	0,0273	0,0574	0,0434
	Median	0,9841	**0,9879**	0,9851	0,9857	0,9795	0,9786	0,9483

Table 1. Central tendency measures and standard deviation calculated over the test set.

CNN in Tables 1 and 2. The analysis revealed that the GSC model statistically outperformed the other models in all metrics, establishing itself as the top-performing model. Conversely, CGNet demonstrated the poorest performance. We classified the seven networks by their performance and used these classifications to set the weights for WOWA aggregations, indicated in Table 3.

Then, we explore if aggregating the segmentations of the seven CNNs is better than each model alone. These results are grouped by consensus method (argmax or mean), and aggregation function (OWA or WOWA). Notably, Tables 5 and 7 show that the WOWA 2 aggregation method was the most effective when using the argmax and mean consensus methods.

Finally, we compare the best aggregation function, WOWA 2, with both the mean and argmax consensus methods with the best-performing individual network, GSC. The results, shown in Tables 8 and 9, demonstrate the superiority of aggregation methods. Specifically, the WOWA 2 with argmax consensus improves segmentation outcomes for all metrics. The findings suggest that combining data can improve the accuracy and reliability of pulmonary region segmentation in medical images, with potential implications for clinical diagnosis and treatment planning.

Qualitative analysis Figure 5, demonstrates the capability of our approach to detect pulmonary regions under diverse circumstances, though further enhancements are possible. Using aggregation methods improves lung segmentation, see Figure 3. Although differences may be subtle, they are significant at the pixel level. Also, Figure 5 shows how the aggregated result fixes flaws in individual CNN models, such as interior holes and isolated regions. In addition, in the absence of specialist segmentations, we visually assessed the generalization capacity of models. We use new unseen data to detect lung regions in X-rays with alternative positions and with the “white lung” condition. Figure 5 shows the power of our approach, although further enhancements are possible.

4. Conclusions

In this study, we tested if aggregation methods improve lung segmentation in X-rays. We trained seven CNN-based

	Jaccard						
	GSC	ERFNet	LinkNet	ESNet	UNet	UNetPre	CGNet
GSC	-	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
ERFNet	✗✗	-	✓	✓	✓✓	✓✓	✓✓
LinkNet	✗✗	✗	-	✓	✓✓	✓✓	✓✓
ESNet	✗✗	✗	✗	-	✓✓	✓✓	✓✓
UNet	✗✗	✗✗	✗✗	✗✗	-	✓✓	✓✓
UNetPre	✗✗	✗✗	✗✗	✗✗	✗✗	-	✓✓
CGNet	✗✗	✗✗	✗✗	✗✗	✗✗	✗✗	-
	Accuracy						
GSC	-	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
ERFNet	✗✗	-	✗	✓	✓✓	✓✓	✓✓
LinkNet	✗✗	✓	-	✓	✓✓	✓✓	✓✓
ESNet	✗✗	✗	✗	-	✓	✓✓	✓✓
UNet	✗✗	✗✗	✗✗	✗	-	✓✓	✓✓
UNetPre	✗✗	✗✗	✗✗	✗✗	✗✗	-	✓✓
CGNet	✗✗	✗✗	✗✗	✗✗	✗✗	✗✗	-
	Sensitivity						
GSC	-	✗	✓	✗✗	✓✓	✓✓	✓✓
ERFNet	✓	-	✓	✗✗	✓✓	✓✓	✓✓
LinkNet	✗	✗	-	✗✗	✓✓	✓✓	✓✓
ESNet	✓✓	✓✓	✓✓	-	✓✓	✓✓	✓✓
UNet	✗✗	✗✗	✗✗	✗✗	-	✗✗	✓✓
UNetPre	✗✗	✗✗	✗✗	✗✗	✓✓	-	✓✓
CGNet	✗✗	✗✗	✗✗	✗✗	✗✗	✗✗	-

Table 2. Networks classification based on the statistical test performed on the *jaccard index, accuracy and sensitivity*.

WOWA	GSC	ERFNet	LinkNet	ESNet	UNet	UNetPre	CGNet
WOWA 1	$\frac{7}{28}$	$\frac{5}{28}$	$\frac{5}{28}$	$\frac{5}{28}$	$\frac{3}{28}$	$\frac{2}{28}$	$\frac{1}{28}$
WOWA 2	$\frac{6}{21}$	$\frac{4}{21}$	$\frac{4}{21}$	$\frac{4}{21}$	$\frac{2}{21}$	$\frac{1}{21}$	0
WOWA 3	$\frac{6}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

Table 3. WOWA aggregation weight ratios are determined from the prior performance classification of the seven networks, as displayed in the Table 2.

		OWA 1	OWA 2	OWA 3	WOWA 1	WOWA 2	WOWA 3
Accuracy	Mean	0,9917	0,9918	0,9919	**0,992**	**0,992**	0,9918
	Std	**0,0068**	0,0072	**0,0068**	0,0071	0,0071	0,0073
	Median	0,994	0,9943	0,9942	0,9944	**0,9945**	0,9944
Jaccard	Mean	0,9658	0,9663	0,9664	**0,9671**	**0,9671**	0,9659
	Std	**0,0273**	0,0296	0,0277	0,0285	0,0286	0,0299
	Median	0,9731	0,9741	0,9734	**0,9749**	0,9748	**0,9749**
Sensitivity	Mean	0,9807	**0,9812**	0,9804	0,981	0,981	0,9779
	Std	**0,0215**	0,0221	0,0226	0,0216	**0,0215**	0,0244
	Median	0,9872	**0,9881**	0,9873	0,9876	0,9876	0,9846

Table 4. Central tendency and standard deviation values for various aggregation methods using the *argmax* consensus.

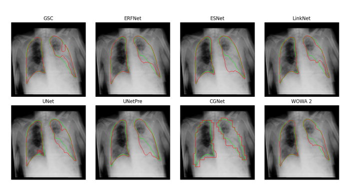


Figure 3. Contours from individual CNNs alongside segmentations from the WOWA 2-argmax aggregation. The annotated region is marked in green, while predictions are in red.

architectures. Later analysis showed that the GSC model exhibited superior average performance, while the CGNet was the least effective. We also assessed many aggregation and consensus functions based on CNNs results. We found that the WOWA 2 aggregation with the argmax consensus method had the best statistically significant average

	OWA 1	OWA 2	OWA 3	WOWA 1	WOWA 2	WOWA 3
				Jaccard		
OWA 1	-	✗	✗✗	✗✗	✗✗	✗
OWA 2	✓	-	✗	✗✗	✗✗	✓
OWA 3	✓✓	✓	-	✗✗	✗✗	✓
WOWA 1	✓✓	✓✓	✓✓	-	✗✗	✓✓
WOWA 2	✓✓	✓✓	✓✓	✓✓	-	✓✓
WOWA 3	✓	✗	✗	✗✗	✗✗	-
				Accuracy		
OWA 1	-	✗	✗✗	✗✗	✗✗	✗
OWA 2	✓	-	✗	✗✗	✗✗	✓
OWA 3	✓✓	✓	-	✗✗	✗✗	✓
WOWA 1	✓✓	✓✓	✓✓	-	✗✗	✓✓
WOWA 2	✓✓	✓✓	✓✓	✓✓	-	✓✓
WOWA 3	✓	✗	✗	✗✗	✗✗	-
				Sensitivity		
OWA 1	-	✗	✓✓	✗	✗	✓✓
OWA 2	✓	-	✓✓	✓✓	✓✓	✓✓
OWA 3	✗✗	✗✗	-	✗✗	✗✗	✓✓
WOWA 1	✓	✗✗	✓✓	-	✓	✓✓
WOWA 2	✓	✗✗	✓✓	✗	-	✓✓
WOWA 3	✗✗	✗✗	✗✗	✗✗	✗✗	-

Table 5. Evaluation of aggregation functions using the *argmax* consensus method with respect to the considered metrics.

		OWA 1	OWA 2	OWA 3	WOWA 1	WOWA 2	WOWA 3
Accuracy	**Mean**	0,9917	0,9918	0,9919	**0,992**	**0,992**	0,9917
	Std	**0,0068**	0,0072	**0,0068**	0,007	0,0071	0,0073
	Median	0,994	0,9943	0,9942	0,9944	**0,9945**	0,9943
Jaccard	**Mean**	0,9658	0,9663	0,9664	0,967	**0,9671**	0,9658
	Std	**0,0273**	0,0295	0,0277	0,0285	0,0286	0,0299
	Median	0,9731	0,9739	0,9735	**0,9749**	0,9748	0,9747
Sensitivity	**Mean**	0,9807	**0,9812**	0,9804	0,981	0,981	0,9778
	Std	**0,0215**	0,0221	0,0226	0,0216	**0,0215**	0,0245
	Median	0,9872	**0,9881**	0,9873	0,9876	0,9875	0,9845

Table 6. Central tendency and standard deviation values for various aggregation methods using the *mean*-based consensus.

	OWA 1	OWA 2	OWA 3	WOWA 1	WOWA 2	WOWA 3
				Jaccard		
OWA 1	-	✓	✗✗	✗✗	✗✗	✗
OWA 2	✗	-	✗	✗✗	✗✗	✓
OWA 3	✓✓	✓	-	✗✗	✗✗	✓
WOWA 1	✓✓	✓✓	✓✓	-	✗✗	✓✓
WOWA 2	✓✓	✓✓	✓✓	✓✓	-	✓✓
WOWA 3	✓	✗	✗	✗✗	✗✗	-
				Accuracy		
OWA 1	-	✗	✗✗	✗✗	✗✗	✗
OWA 2	✓	-	✗	✗✗	✗✗	✓
OWA 3	✓✓	✓	-	✗✗	✗✗	✓
WOWA 1	✓✓	✓✓	✓✓	-	✗✗	✓✓
WOWA 2	✓✓	✓✓	✓✓	✓✓	-	✓✓
WOWA 3	✓	✗	✗	✗✗	✗✗	-
				Sensitivity		
OWA 1	-	✗✗	✓✓	✗	✗	✓✓
OWA 2	✓✓	-	✓✓	✓✓	✓✓	✓✓
OWA 3	✗✗	✗✗	-	✗✗	✗✗	✓✓
WOWA 1	✓	✗✗	✓✓	-	✓	✓✓
WOWA 2	✓	✗✗	✓✓	✗	-	✓✓
WOWA 3	✗✗	✗✗	✗✗	✗✗	✗✗	-

Table 7. Evaluation of aggregation functions using the *mean*-based consensus method with respect to the considered metrics.

performance. In all, these findings highlight the power of aggregation. It improves the performance of individual deep learning models. The qualitative evaluation also confirmed the models' ability to generalize across conditions not included in training data, such as alternative X-ray positions and "white lung" conditions.

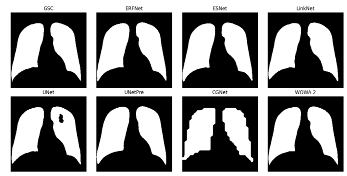
We intend to explore if using this segmentation approach as a prior improves lung disease classification. We will also study how applying techniques like [15] and Grad-CAM [19] help us understand the models' predictions.

		GSC	WOWA 2 argmax	WOWA 2 mean
Accuracy	**Mean**	0,9916	**0,992**	**0,992**
	Std	0,0074	**0,0071**	**0,0071**
	Median	0,9942	**0,9945**	**0,9945**
Jaccard	**Mean**	0,9652	**0,9672**	0,9671
	Std	0,0303	**0,0286**	**0,0286**
	Median	0,9736	**0,9748**	**0,9748**
Sensitivity	**Mean**	0,9772	**0,981**	**0,981**
	Std	0,0247	**0,0215**	**0,0215**
	Median	0,9841	**0,9876**	0,9875

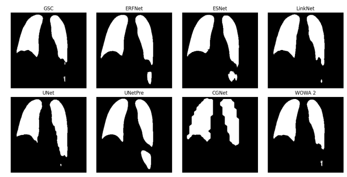
Table 8. Central tendency and standard deviation for GSC, and WOWA 2 aggregation using both argmax and mean consensus.

	Jaccard/Accuracy/Sensitivity		
	GSC	**WOWA 2 argmax**	**WOWA 2 media**
GSC	-	✗✗	✗✗
WOWA 2 argmax	✓✓	-	✓✓
WOWA 2 media	✓✓	✗✗	-

Table 9. Comparison of the best network and optimal aggregation using WOWA 2 with argmax and mean consensus methods.

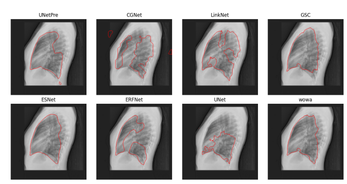


(a)

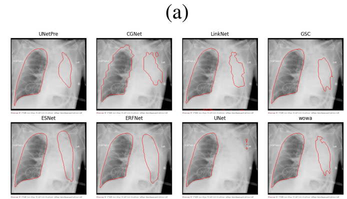


(b)

Figure 4. Lung segmentation examples with (a) internal holes and (b) inaccurately segmented regions outside the lung area.



(a)



(b)

Figure 5. Pulmonary region segmentation examples from seven individual CNNs and our aggregated method applied to radiographs with (a) sideways positioning and (b) "white lung" pathology.

References

[1] Abhishek Chaurasia and Eugenio Culurciello. Linknet: Exploiting encoder representations for efficient semantic seg-

mentation. In *IEEE Visual Communications and Image Processing*, pages 1–4, 2017. 1

[2] Kaggle COVID. Radiography database. *Radiological Society of North America*, 2019. 2

[3] Ahmed Elnakib, Georgy Gimel'farb, Jasjit S Suri, and Ayman El-Baz. Medical image segmentation: a brief survey. *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies: Volume II*, pages 1–39, 2011. 1

[4] Rahul Hooda, Ajay Mittal, and Sanjeev Sofat. An efficient variant of fully-convolutional network for segmenting lung fields from chest radiographs. *Wireless Personal Communications*, 101, 2018. 1

[5] Sangheum Hwang and Sunggyun Park. Accurate lung segmentation via network-wise training of convolutional networks, 2017. 1

[6] M. Jabreel and M. Abdel-Nasser. Promising crack segmentation method based on gated skip connection. *Electronics Letters*, 56, 2020. 1

[7] Stefan Jaeger, Sema Candemir, Sameer Antani, Yì-Xiáng Wáng, Pu-Xuan Lu, and George Thoma. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery*, 4:475–477, 2014. 2

[8] Luis González Jaime, Etienne E. Kerre, Mike Nachtegael, and Humberto Bustince. Consensus image method for unknown noise removal. *Knowledge-Based Systems*, 70:64–77, 2014. 1

[9] Sandra Jardim, João António, and Carlos Mora. Image thresholding approaches for medical image segmentation - short literature review. *Procedia Computer Science*, 219: 1485–1492, 2023. 1

[10] M.C. Jobin Christ. Segmentation of medical image using clustering and watershed algorithms. *American Journal of Applied Sciences*, 8:1349–1352, 2011. 1

[11] Shadi Mahmoodi Khaniabadi, Haidi Ibrahim, Ilyas Ahmad Huqqani, Farzad Mahmoodi Khaniabadi, Harsa Amylia Mat Sakim, and Soo Siang Teoh. Comparative review on traditional and deep learning methods for medical image segmentation. In *2023 IEEE 14th control and system graduate research colloquium (ICSGRC)*, pages 45–50. IEEE, 2023. 1

[12] Wufeng Liu, Jiaxin Luo, Yan Yang, Wenlian Wang, Junkui Deng, and Liang Yu. Automatic lung segmentation in chest x-ray images using improved u-net. *Scientific Reports*, 12: 8649, 2022. 1

[13] Agus Pratondo, Sim Heng Ong, and Chee Kong Chui. Region growing for medical image segmentation using a modified multiple-seed approach on a multi-core cpu computer, 2014. 1

[14] Tawsifur Rahman, Amith Khandakar, Yazan Qiblawey, Anas Tahir, Serkan Kiranyaz, Saad Bin Abul Kashem, Mohammad Tariqul Islam, Somaya Al Maadeed, Susu M. Zughaier, Muhammad Salman Khan, and Muhammad E.H. Chowdhury. Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. *Computers in Biology and Medicine*, 132, 2021. 1

[15] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you": Explaining the predictions of any classifier, 2016. 4

[16] E Romera, J M Álvarez, L M Bergasa, and R Arroyo. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 19:263–272, 2018. 1

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. 1

[18] Nassir H. Salman, Bnar M Ghafour, and Gullanar M Hadi. Medical image segmentation based on edge detection techniques. *Advances in Image and Video Processing*, 3, 2015. 1

[19] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 2019. 4

[20] Junji Shiraishi, Shigehiko Katsuragawa, J Ikezoe, T Matsumoto, T Kobayashi, K.-I Komatsu, M Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. Development of a digital image database for chest radiographs with and without a lung nodule. *AJR. American journal of roentgenology*, 174:71–4, 2000. 2

[21] Lucas O. Teixeira, Rodolfo M. Pereira, Diego Bertolini, Luiz S. Oliveira, Loris Nanni, George D. C. Cavalcanti, and Yandre M. G. Costa. Impact of lung segmentation on the diagnosis and explanation of covid-19 in chest x-ray images. *Sensors*, 21:7116, 2021. 1

[22] Vicenç Torra. The weighted owa operator. *International Journal of Intelligent Systems*, 12(2):153–166, 1997. 1

[23] Yu Wang, Quan Zhou, Jian Xiong, Xiaofu Wu, and Xin Jin. Esnet: An efficient symmetric network for real-time semantic segmentation. In *Pattern Recognition and Computer Vision: Second Chinese Conference*. Springer, 2019. 1

[24] Tianyi Wu, Sheng Tang, Rui Zhang, Juan Cao, and Yongdong Zhang. Cgnet: A light-weight context guided network for semantic segmentation. *IEEE Transactions on Image Processing*, 30:1169–1179, 2020. 1

[25] Ronald R Yager. Quantifier guided aggregation using owa operators. *International Journal of Intelligent Systems*, 11, 1996. 1