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

Antonio Nadal-Martínez[1], Lidia Talavera-Martínez[123], Marc Munar[123], Manuel González-Hidalgo[1234]

[1]University of the Balearic Islands. SCOPIA Research Group, [2]Health Research Institute of the Balearic Islands (IdISBa),
[3]Artificial Intelligence Research Institute of the Balearic Islands (IAIB), [4]Laboratory of Artificial Intelligence Applications (LAIA)

CVPR JUNE 17-21, 2024 SEATTLE, WA

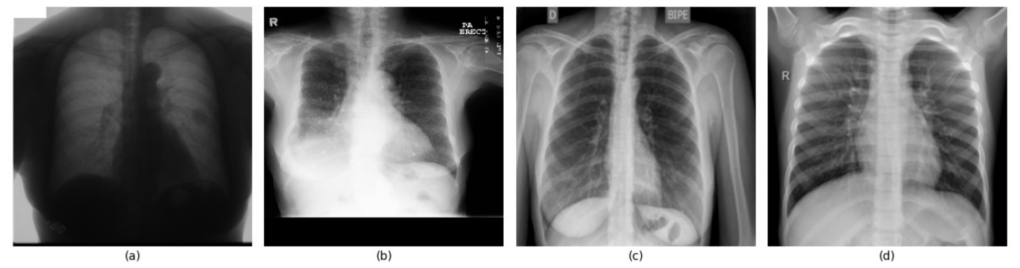
Motivation

- Automatic segmentation plays a vital role in medical processing and analysis.
- Aids specialists for precise identification and isolation of regions of interest in medical imaging.

Are aggregation functions of CNN models and consensus methods effective for improving pulmonary region segmentation in X-ray images?

Methodology

Examples of images from the databases that constitute the dataset used in the study: (a) JSRT, (b) Montgomery, (c) and (d) COVID-19 Radiography Database.



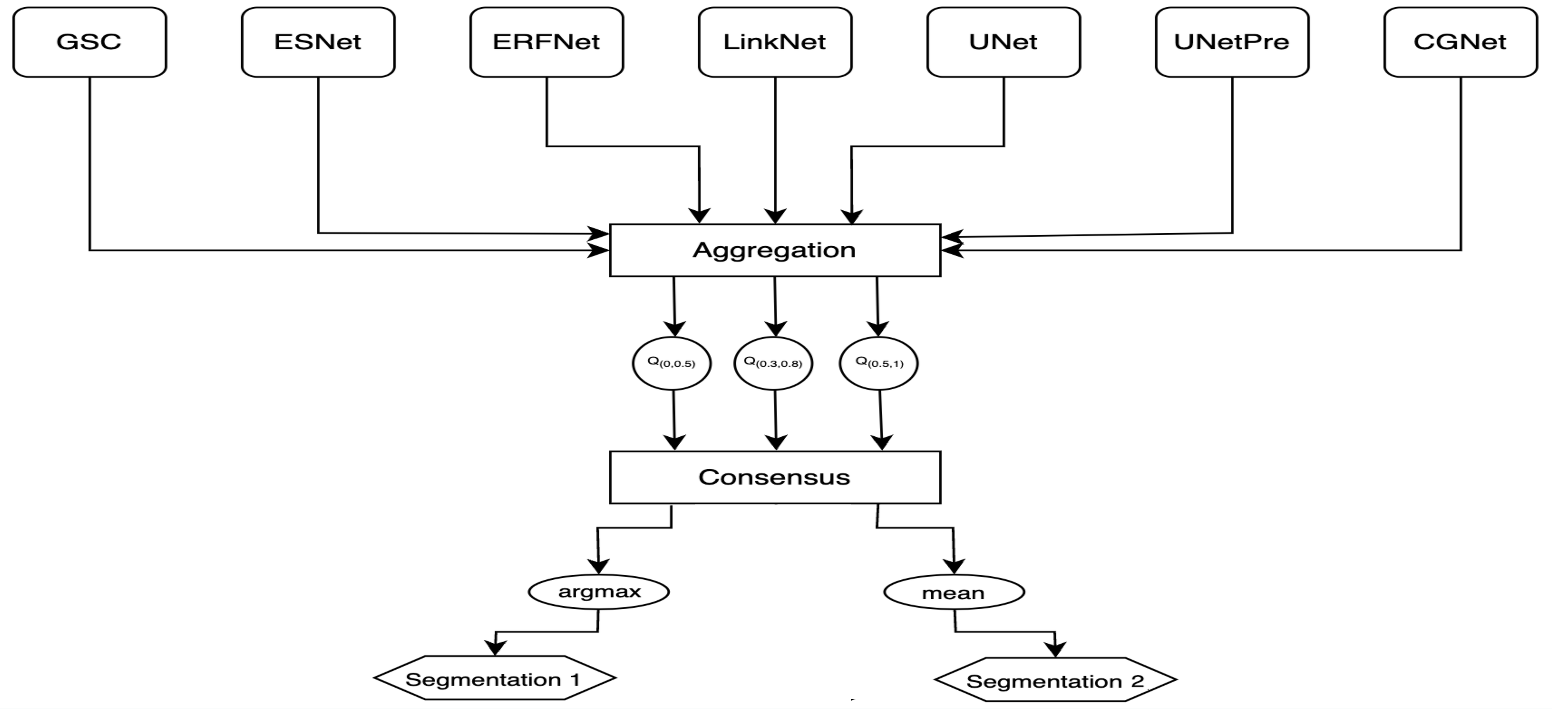
7 CNN models: GSC, ESNet, ERFNet, LinkNet, Unet, UNetPre and CGNet.

6 aggregation functions based on OWA [1] and WOWA [2] approaches:

- **OWA 1:** weights obtained from regular monotone increasing quantifier;
- **OWA 2:** aggregate only top 4 models' segmentations;
- **OWA 3:** includes all segmentations except the worst-performing model;
- **WOWA 1:** assigns weights based on the model's performance order. The i-th model gets the weight;
- **WOWA 2:** same as WOWA 1 but without the worst model; being n the number of models;
- **WOWA 3:** same as WOWA 1 but the best model gets a weight of $\frac{1}{2}$, while the others get $\frac{1}{2(n-1)}$.

2 consensus methods. At each pixel:

- **Argmax method:** selects the maximum value among the q obtained aggregations**;**
- **Mean method:** obtains the mean value among the q obtained aggregations.



Results

Performance of individual CNN models

		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

⇩

WOWA aggregation weight ratios are determined from the performance classification of the seven networks

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	$\frac{6}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

Which is the best aggregation function when using argmax consensus method?

		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

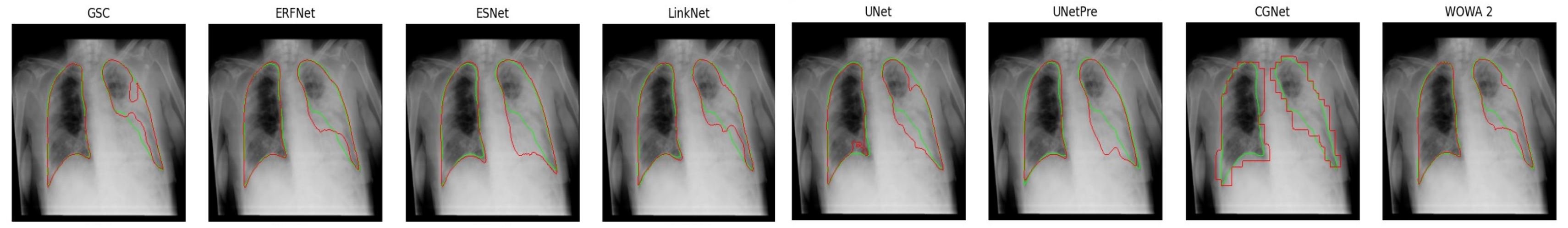
Which is the best aggregation function when using mean-based consensus method?

		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

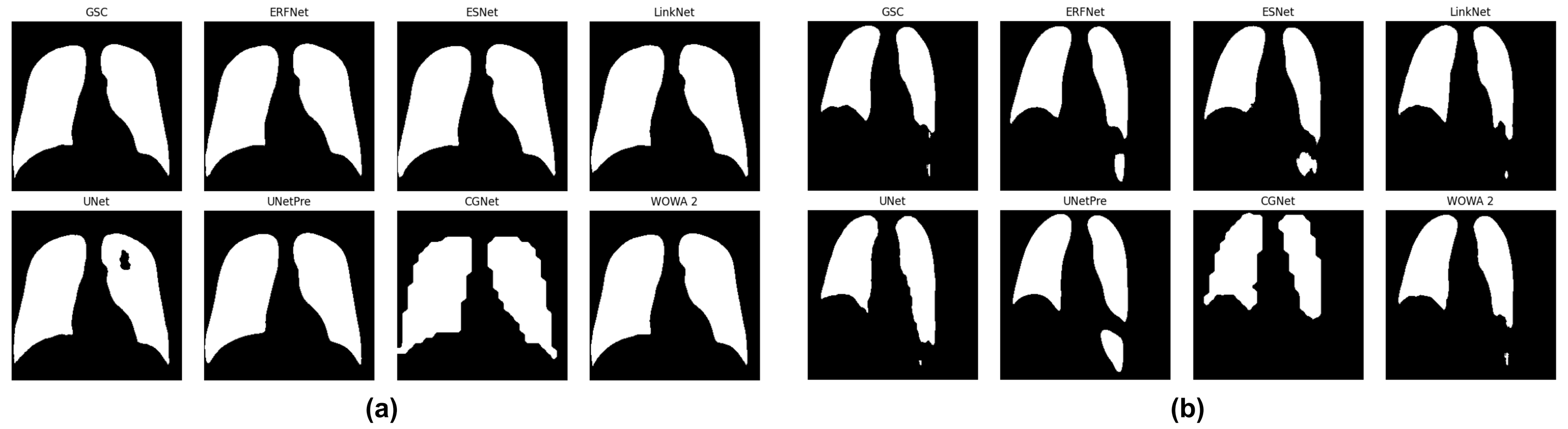
Do aggregation and consensus methods improve individual CNN results?

		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

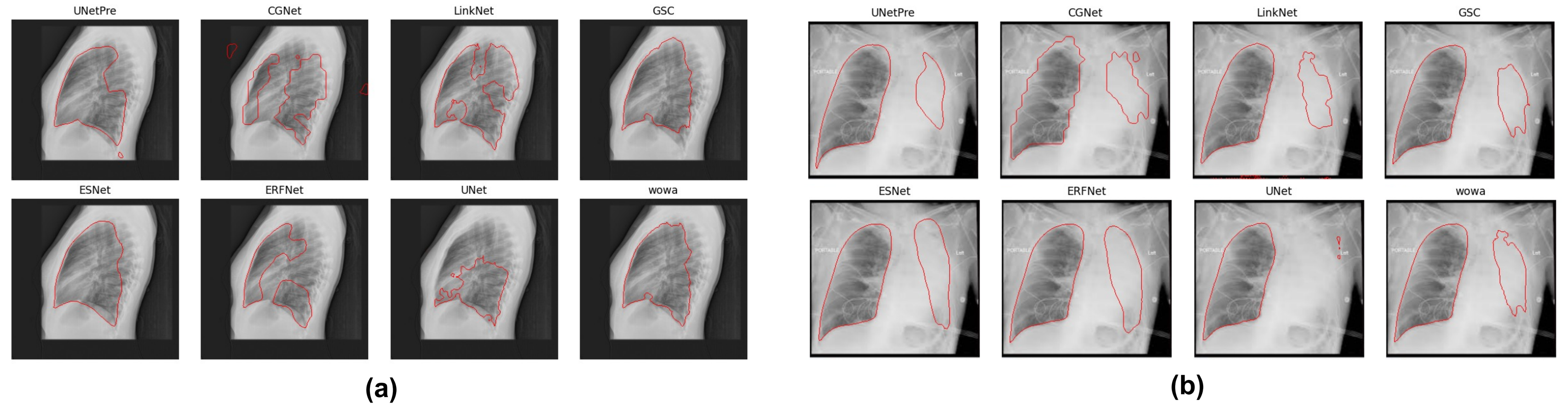
Comparison of expert segmentation (green) and segmentation obtained by individual CNNs and the WOWA 2 aggregation function with argmax consensus method (red).



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



Generalization capability of the individual CNNs and our aggregated method (WOWA 2 + argmax) applied to radiographs with untrained conditions (a) sideways positioning and (b) white lung pathology



Conclusions*

- Individual CNN models: **GSC model exhibited superior performance**, while the **CGNet was the least effective**.
- **WOWA 2** aggregation **with the argmax or mean-based consensus** method had the best performance ↳ **outperforming the GSC-based model**.
- **Aggregation methods have demonstrated their effectiveness in improving lung segmentation in X-ray images**.
- The qualitative evaluation confirmed the **models' ability to generalize across** conditions not included in training data, such as **alternative X-ray positions and "white lung" conditions**.

*All conclusions are supported by statistical analysis.

References

[1] Yager, R. R. (1996). Quantifier guided aggregation using OWA operators. *International journal of intelligent systems*, *11*(1), 49-73.

[2] Torra, V. (1997). The weighted OWA operator. International journal of intelligent systems, 12(2), 153-166.

Acknowledgements

SCOPIA; Universitat de les Illes Balears; IdISBa; European Union European Social Fund; Ministerio de Ciencia e Innovación

Funded by the project: PID2020-113870GB-I00