Interpretable Deep Learning Model For Cardiomegaly Detection with Chest X-ray Images

Estela Ribeiro[1,2], **Diego A. C. Cardenas**[1], **Jose E. Krieger**[1,2],
Marco A. Gutierrez[1,2,3]

[1]Heart Institute (InCor) – Clinics Hospital
University of Sao Paulo Medical School (HCFMUSP)
Sao Paulo – SP – Brazil

[2]University of Sao Paulo Medical School (FMUSP)
Sao Paulo – SP – Brazil

[3]Polytechnique School, University of Sao Paulo (POLI USP)
Sao Paulo – SP – Brazil

estela.ribeiro@hc.fm.usp.br, diego.cardona@hc.fm.usp.br

j.krieger@hc.fm.usp.br, marco.gutierrez@incor.usp.br

Abstract. *Cardiomegaly is a medical disorder characterized by an enlargement of the heart. Many works propose to automatically detect cardiomegaly through chest X-rays. However, most of them are based on deep learning models, known for their lack of interpretability. This work propose a deep learning model for the detection of cardiomegaly based on chest x-rays images and the qualitative assessment of three known local explainable methods, i.e., Grad-CAM, LIME and SHAP. Our model achieved Acc, Prec, Se, Spe, F1-score and AUROC of* 91.8±0.7%, 74.0±2.7%, 87.0±5.5%, 92.9±1.2%, 79.8±1.9%, *and* 90.0±0.7%, *respectively. Moreover, except for the SHAP method, our interpretable methods were able to pinpoint the expected location for cardiomegaly. However, Grad-CAM method showed faster computational time than LIME and SHAP.*

1. Introduction

Cardiomegaly, a pathology frequently detected in chest x-rays (CXR), is a medical disorder in which a patient's heart enlarges temporarily or permanently depending on the situation. This enlargement is typically a clinical manifestation of another pathogenic condition, such as chamber dilation, ventricular hypertrophy, or pericardial effusion [Daines et al. 2021], possibly resulting in heart failure, cardiac arrest, and sudden death [Cardenas et al. 2020]. Routine chest radiographs showing cardiomegaly have significant clinical implications, since they help doctors decide whether additional investigation is necessary [Daines et al. 2021].

Deep learning research has the potential to become a standard method for medical analysis and even diagnosis [Hicks et al. 2021]. Convolutional Neural Networks, which use raw data from image pixels as input and abstract the original image data layer-wise to enable CXR evaluation automation, can increase the effectiveness of analysis of large volumes of complex medical exams.

The cardiothoracic ratio (CTR), a comparison of the cardiac and thoracic diameters in CXR images, is a commonly used index that offers predictive information [Alghamdi et al. 2020], and some works have been done to automatically identify the CTR and determine whether a given CXR has cardiomegaly or not [Wu et al. 2022]. However, the CTR must be manually evaluated, which is time-demanding [Junior et al. 2021, Wu et al. 2022]. Moreover, since CXR can contain noises and artifacts, segmentation-based methods used to identify the boundaries of the thorax and heart to measure the CTR can be affected, impairing the results. Other works propose direct detection of cardiomegaly from CXR without the need to compute the CTR [Cardenas et al. 2020, Junior et al. 2021] using Deep Learning, letting the networks automatically extrapolate the CTR and use the global CXR information for decision-making.

Currently, many researchers are focusing on developing techniques that can help understand how deep learning models reach their decisions [Simonyan et al. 2014, Selvaraju et al. 2017, Ribeiro et al. 2016, Lundberg and Lee 2017]. Some of them rely on the analysis of gradients [Selvaraju et al. 2017, Simonyan et al. 2014], which can be difficult to interpret, while others are based on perturbation [Ribeiro et al. 2016, Lundberg and Lee 2017], which can be specific to a particular instance.

In this study, we aim to assess qualitatively the interpretable methods to understand the decisions of a deep learning model for cardiomegaly detection with CXR images. By implementing such a model, low-income hospitals that might not have access to experienced and knowledgeable radiologists might lessen the burden on their medical infrastructure.

2. Methods

In this section, we describe the dataset, the preprocessing steps, and the deep neural network architecture used for classifying CXR images into Cardiomegaly and Non-Cardiomegaly. Moreover, we describe the interpretable methods used to interpret the model's results and assess qualitatively their interpretability. The general structure of the proposed method is shown in Figure 1.

2.1. Datasets

VinDr-CXR [Nguyen et al. 2020a, Nguyen et al. 2020b] is a dataset of CXR scans obtained retrospectively from two major hospitals in Vietnam. All images are in DICOM format and in postero-anterior (PA) view. It contains 18,000 images manually annotated by a group of radiologists. This dataset is divided into a training set with 15,000 scans independently labeled by three radiologists, and a test set with 3,000 scans labeled by the consensus of five radiologists. It should be stressed that due to the lack of patients ID information, images from the same patient may exist throughout train and test sets.

Besides the VinDr-CXR dataset, we used a private CXR dataset collected from the Picture Archiving and Communication System (PACS) of a tertiary referral hospital in Brazil specialized in cardiology (Heart Institute from Clinics Hospital, University of Sao Paulo), with patients older than 18 years old. All images are in DICOM format in a postero-anterior (PA) view. This dataset contains 230 exams labeled as cardiomegaly and 1003 exams labeled as non-cardiomegaly. This private dataset complies with all relevant ethical regulations and was approved by the Research Ethics Committee.

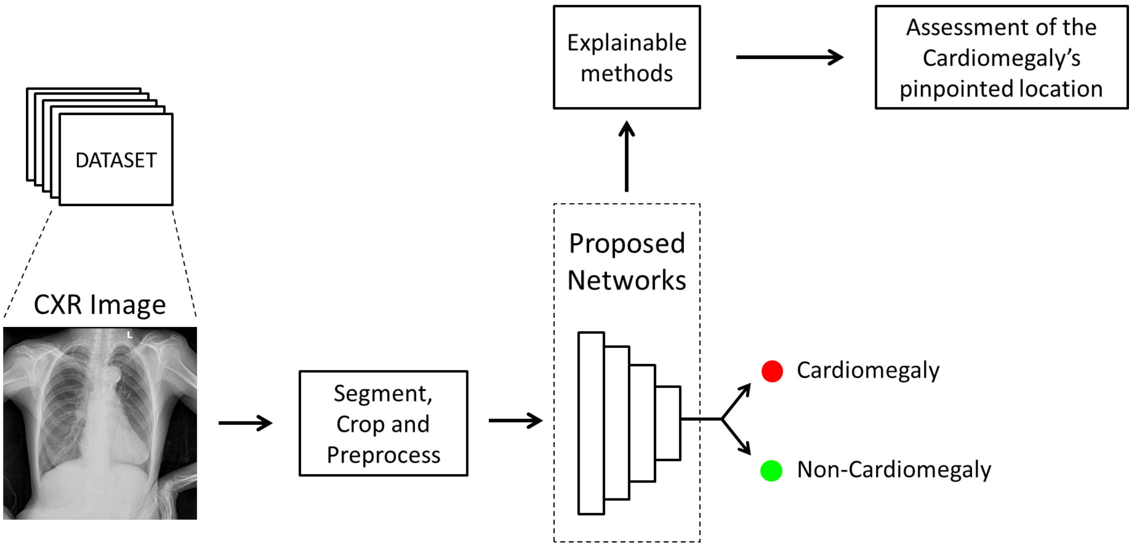
DATASET
CXR Image
Segment, Crop and Preprocess
Proposed Networks
Explainable methods
Assessment of the Cardiomegaly's pinpointed location
Cardiomegaly
Non-Cardiomegaly

Figure 1. General structure of the proposed methodology.

2.2. Data Preprocessing

Our preprocessing steps are based on previous works [Cardenas et al. 2020, Junior et al. 2021] for cardiomegaly classification. We used a previously developed and validated model [Pazhitnykh and Petsiuk 2017] that segments the lungs, to be able to create a binary mask of the chest cavity region, using a fine-tuned UNet-based convolutional neural network. The extreme points of the lung mask were used to create a bounding box to crop the chest cavity, segmenting only this region of interest (ROI), that was used as input in our classification model. We rescaled the intensity of the cropped image with a contrast stretching method, including all intensities that fall within the 1st and 99th percentiles of the image histogram. The image was transformed into a square image using zero padding and resized it to 384x384x3.

2.3. Deep Learning model

The small size of datasets in deep learning for analyzing medical images is a major constraint. Consequently, it is frequently difficult to train a CNN from scratch [Baltruschat et al. 2019]. In this case, Transfer-learning is one solution. Here, we initiated the training with the ImageNet weights [Deng et al. 2009]. All layers were pretrained with the VinDr-CXR dataset images so that the model could learn the patterns of the CXR domain, and posteriorly we did a fine-tuning with our private dataset.

For our experiments, we choose a ResNet 50 v2 architecture, which is a Deep Neural Network widely used in the classification of CXRs. We used the default Keras architecture topology for the network with an input resolution of 384x384x3 before the classification layers.

For the fully connected layers, we used a customized 2-layer perceptron. Our model was trained over 40 epochs by stochastic gradient descent using a 16 batch size per step using RMSprop with a learning rate of 1e-4 and a callback to reduce it by a factor of 0.2 every six epochs in case of no improvement in validation loss, and class weights to balance the dataset. Data augmentation techniques (i.e., image rotation, shift, shear, scale, and flip) oversampled the training set. Furthermore, we used a threshold of 0.5 to define the predicted label.

We used a 5-fold cross-validation method to evaluate our model, and we assessed six different metrics, including Accuracy (Acc), Precision (Prec), Sensitivity (Se), Specificity (Spe), F1-score, and Area Under the Receiver Operating Curve (AUROC). Our experiment was performed using a Foxconn High-Performance Computer (HPC) M100-NHI with an 8 GPU cluster of 16 GB NVIDIA Tesla V100 cards. The methodology was implemented using the Python framework and Keras v2.2.4 with TensorFlow backend v2.3.0.

2.4. Explainable AI methods

We selected the following set of local explainable methods, i.e., Grad-CAM, LIME and SHAP [Molnar 2019]. Here, we examined individual CXR images. Visualizations result in a map with the most significant parts highlighted in red color for Grad-CAM, LIME and SHAP methods.

The Gradient-weighted Class Activation Mapping (Grad-CAM) method [Selvaraju et al. 2017] generates a coarse localization map by using the gradient information flowing into the final convolutional layer, producing a heatmap for a given class label, and assigning important areas of the CXR input image to the prediction. The Local Interpretable Model-Agnostic Explanations (LIME) method [Ribeiro et al. 2016] uses a simpler interpretable surrogate model (e.g. linear regression), applying perturbations on the CXR input, training an interpretable model that mimics the behavior of the original model for an individual CXR input image. The SHapley Additive exPlanations (SHAP) method [Lundberg and Lee 2017], is a game-theoretic approach based on the Shapley values. It can explain the prediction of an CRX input image by computing the contribution of each segment of the image to the prediction. Here we used the Kernel SHAP method, which is very similar to LIME. Moreover, we generated 1000 perturbations for both LIME and Kernel SHAP methods, used a linear regression as the surrogate model, and segmented the images with a quickshift segmentation algorithm.

3. Results

The model used to be interpreted by our explainable methods achieved the results described in Table 1 for the task of detecting cardiomegaly with chest x-ray images.

	Acc	**Prec**	**Se (TPR)**	**Spe (TNR)**	**F1**	**AUROC**
Mean	91.8	74.0	87.0	92.9	79.8	90.0
Std	0.7	2.7	5.5	1.2	1.9	0.7

Table 1. Summary of the obtained results for the Cardiomegaly classification.

Figures 2 and 3 display the results of Grad-CAM, LIME and SHAP methods of an exam labeled as Cardiomegaly. Likewise, Figure 4 displays the results of the explainable methods of an exam labeled as non-Cardiomegaly. Additionally, Table 2 presents the computational time spent by each interpretable algorithm to perform their interpretations.

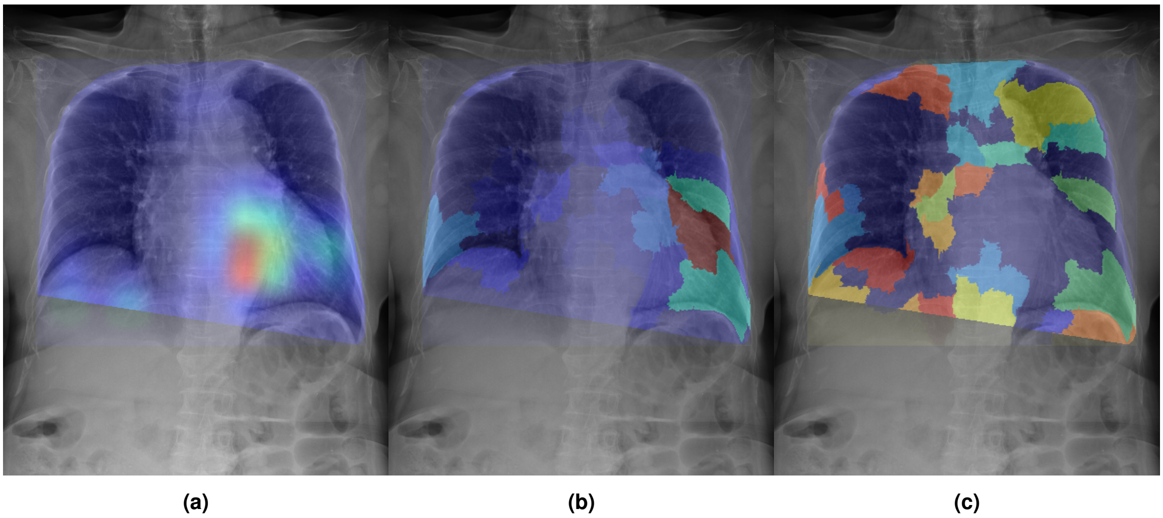

Figure 2. Explainable methods results for an image correctly labeled as cardiomegaly (prediction probability of cardiomegaly: 99.98%): (a) Grad-CAM; (b) LIME; and (c) SHAP.

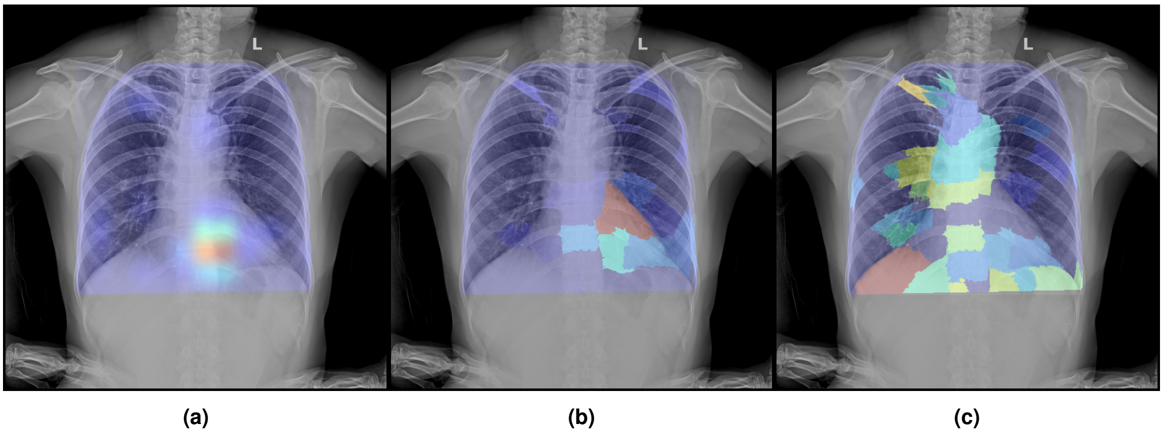

Figure 3. Explainable methods results for an image correctly labeled as cardiomegaly (prediction probability of cardiomegaly: 66.3%): (a) Grad-CAM; (b) LIME; and (c) SHAP.

	Grad-CAM	**LIME**	**SHAP**
Cardiomegaly image	15 seconds	6.74 minutes	3.36 minutes
Non-Cardiomegaly image	16 seconds	6.42 minutes	3.40 minutes

Table 2. Amount of time spent by each algorithm to perform their interpretations.

4. Discussion

In this study, we develop a deep learning model capable to detect cardiomegaly in chest x-ray images. From it, we demonstrate that well-known methods usually used to interpret black-boxes image-based models such as ours can pinpoint the expected location for cardiomegaly.

Figures 2, 3 and 4 show the explainable methods results for two examples, where Grad-CAM and LIME methods attributed the highest importance to the performance in

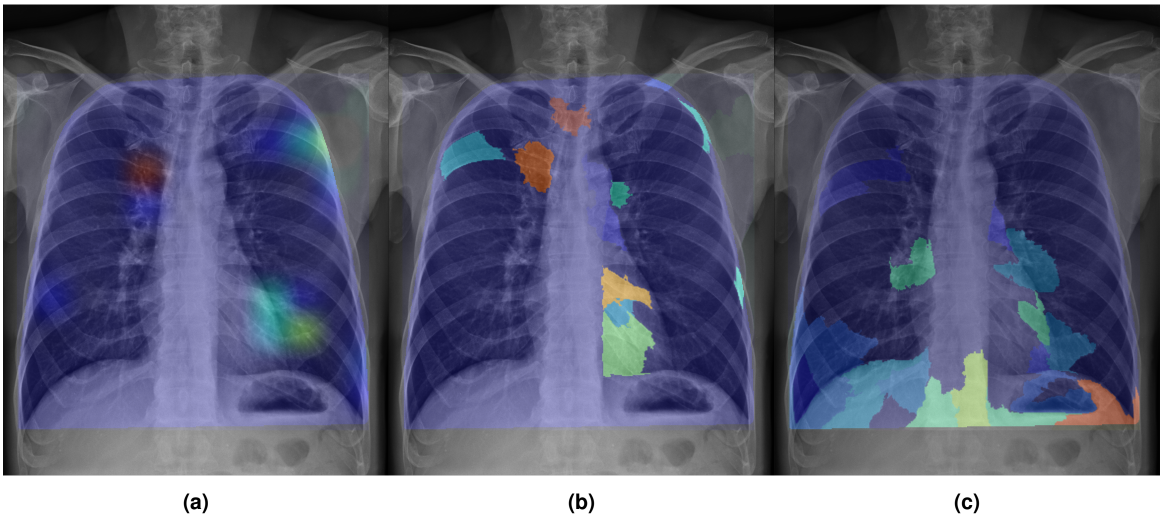

Figure 4. Explainable methods results for an image correctly labeled as non-Cardiomegaly (prediction probability of cardiomegaly: 0.0%): (a) Grad-CAM; (b) LIME; and (c) SHAP.

regions located on the heart, which is expected from a physiologic point of view, since cardiomegaly is related to the size of the heart. Even on an image labeled as non-Cardiomegaly, these two methods still highlight the heart's region. SHAP method, however, didn't indicate the heart as the region most relevant to the performance. It should be emphasized that for all of these methods, relevance refers to how much weight a particular area contributes to the overall prediction.

Differently from Grad-CAM that uses the gradient information to produce a map with important areas to the prediction, LIME and SHAP methods are based on perturbations on the input image, produced by changing the original image, which is highly dependent on the segmentation method applied. Thus, results can be different depending on the segmentation.

Moreover, it should be noted that while Grad-CAM takes just a few seconds to generate its results, the LIME and SHAP methods take minutes, depending on the number of perturbations generated. In a real application, where results are expected to be generated as quickly as possible, both methods may be impractical. Furthermore, future works should add quantitative metrics for the explainable AI methods in order to evaluate the models.

It is indispensable that the models manage to generate explanations about their operation. By adopting such a concept into practice, low-income institutions that might not have access to qualified and experienced radiologists could alleviate the strain on their healthcare system. Furthermore, even the most experienced expert may be susceptible to errors, thus this model can assist in handling the arduous and time-consuming task of interpreting and evaluating CXR images.

5. Conclusion

In this study, we developed a deep learning model with outstanding classification performance that can identify cardiomegaly in chest x-ray images using a transfer learning technique. Additionally, we show that the incorporation of explainable methodologies enables the identification of relevant regions within the CXR images that contribute to the classification and allows the determination of the expected location for cardiomegaly. Such a model can help with the challenging and time-consuming task of deciphering and assessing CXR images.

Ethics Statement

This research was approved by the Internal Review Board (IRB), registration CAAE 45070821.3.0000.0068, as part of the Machine Learning in Cardiovascular Medicine Project.

Acknowledgements

This study was supported by São Paulo Research Foundation (FAPESP) – grant n^o 2021/12935-0. Zerbini Foundation, and Foxconn Brazil, as part of the research project "Machine Learning in Cardiovascular Medicine".

References

Alghamdi, S. S., Abdelaziz, I., Albadri, M., Alyanbaawi, S., Aljondi, R., and Tajaldeen, A. (2020). Study of cardiomegaly using chest x-ray. *Journal of Radiation Research and Applied Sciences*, 13(1):460–467.

Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., and Saalbach, A. (2019). Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. *Scientific Reports*, 9(1):6381.

Cardenas, D., Junior, J. F., Moreno, R., Rebelo, M., Krieger, J., and Gutierrez, M. (2020). Multicenter validation of convolutional neural networks for automated detection of cardiomegaly on chest radiographs. In *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 179–190, Porto Alegre, RS, Brasil. SBC.

Daines, B., Rao, S., Hosseini, O., Prieto, S., Abdelmalek, J., Elmassry, M., Sethi, P., Test, V., and Nugent, K. (2021). The clinical associations with cardiomegaly in patients undergoing evaluation for pulmonary hypertension. *Journal of Community Hospital Internal Medicine Perspectives*, 11(6):787–792.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.

Hicks, S. A., Isaksen, J. L., Thambawita, V., Ghouse, J., Ahlberg, G., Linneberg, A., Grarup, N., Strümke, I., Ellervik, C., Olesen, M. S., Hansen, T., Graff, C., Holstein-Rathlou, N.-H., Halvorsen, P., Maleckar, M. M., Riegler, M. A., and Kanters, J. K. (2021). Explaining deep neural networks for knowledge discovery in electrocardiogram analysis. *Scientific Reports*, 11(1):10949.

Junior, J. R. F., Cardenas, D. A. C., Moreno, R. A., de Fátima de Sá Rebelo, M., Krieger, J. E., and Gutierrez, M. A. (2021). A general fully automated deep-learning method to detect cardiomegaly in chest x-rays. In Mazurowski, M. A. and Drukker, K., editors, *Medical Imaging 2021: Computer-Aided Diagnosis*, volume 11597, pages 537 – 542. International Society for Optics and Photonics, SPIE.

Lundberg, S. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. arXiv: 1705.07874.

Molnar, C. (2019). *Interpretable Machine Learning*.

Nguyen, H. Q., Lam, K., Le, L. T., Pham, H. H., Tran, D. Q., Nguyen, D. B., Le, D. D., Pham, C. M., Tong, H. T. T., Dinh, D. H., Do, C. D., Doan, L. T., Nguyen, C. N., Nguyen, B. T., Nguyen, Q. V., Hoang, A. D., Phan, H. N., Nguyen, A. T., Ho, P. H., Ngo, D. T., Nguyen, N. T., Nguyen, N. T., Dao, M., and Vu, V. (2020a). Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations.

Nguyen, H. Q., Lam, K., Le, L. T., Pham, H. H., Tran, D. Q., Nguyen, D. B., Le, D. D., Pham, C. M., Tong, H. T. T., Dinh, D. H., Do, C. D., Doan, L. T., Nguyen, C. N., Nguyen, B. T., Nguyen, Q. V., Hoang, A. D., Phan, H. N., Nguyen, A. T., Ho, P. H., Ngo, D. T., Nguyen, N. T., Nguyen, N. T., Dao, M., and Vu, V. (2020b). Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations (version 1.0.0).

Pazhitnykh, I. and Petsiuk, V. (2017). Lung segmentation 2d.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. arXiv: 1602.04938.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626.

Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv: 1312.6034.

Wu, J.-X., Pai, C.-C., Kan, C.-D., Chen, P.-Y., Chen, W.-L., and Lin, C.-H. (2022). Chest x-ray image analysis with combining 2d and 1d convolutional neural network based classifier for rapid cardiomegaly screening. *IEEE Access*, 10:47824–47836.