Multi-label Abnormality Classification using Frontal and Lateral Chest X-ray Images

Zubin Bhuyan
Department of Computer Science, UMass Lowell.

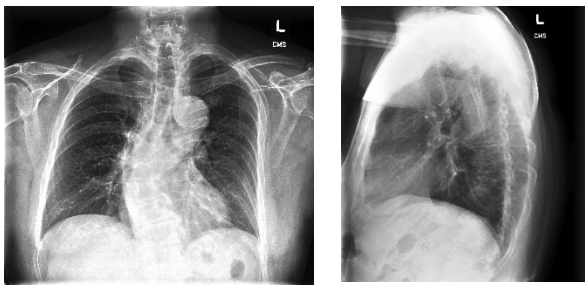
Project report- COMP.5300: Advanced Topics in Deep Learning.

***Abstract*—Deep learning models have been extensively used in image processing and image analysis tasks. One such area is analysis and detection of thoracic diseases in chest x-ray images. The availability of large volume x-ray datasets and successful research efforts and innovations in the field of deep learning hassled to the development of several Computer Aided Diagnosis systems. One of the primary shortcomings of most of these radiograph-based CAD systems is that they are only able to process frontal view x-ray images. In this project report we explore how lateral x-ray images can be incorporated with frontal x-rays so that CNN-based models can best take advantage of two views. To this extent, implement two models which takes two x-ray images, one frontal and one lateral, as input. We compare the results with baseline DenseNet-121 and similar models.**

I. Introduction

X-ray imaging is fast, easy, cheap and is especially helpful in situations of emergency diagnosis and treatment. Chest X-ray (CXR) is the most common radiology examination ordered by physicians, and in recent years, several CXR datasets have been released [1, 2, 3], and this has enabled deep learning researchers and practitioners to engineer and train CNN-based models. The degree to which these models have been successful, is however, another topic which is still open to debate and discussion. Automatic classification and detection of diseases in chest radiographs

The most common orientation of taking x-rays are the frontal view and lateral view. Example images of both these views are shown in Fig. 1; image on the left is the frontal view, and the on the right is the lateral view. These views are also referred to as PA and L views respectively. Most Computer Aided Diagnosis (CAD) approaches make use of the frontal x-ray images to train and detect abnormalities in the chest [4, 5, 8].



(a) Frontal (b) Lateral

Fig. 1. Two views of chest x-ray images from the *CheXpert dataset*.

In this project we propose a method of using both frontal and lateral views to train deep neural models to detect manefestations and abnormalities in lungs. The inclusion of lateral x-rays means inclusion of additional information, which potentially can improve performance of models.

Section II talks about the motivation of this project and justifies the intuition why inclusion of lateral images should improve performance of CNN-based models. Related work is reviewed in Section III. Section IV talks about the proposed approach; datasets, architectures and other implementation details are discussed here. This is followed by Section V where experimental results are analyzed. And finally Section VI concludes this report and discusses future direction in which this work can be further extended.

II. Motivation

Physicians and radiologists order additional lateral view radiographs when there is uncertainty in diagnosis. Most CAD systems however only consider the frontal image when predicting pathologies and detecting manifestations. Additionally, most of the common convolutional neural network architectures, such as ResNet, DenseNet, etc. have been designed to take single image as input. We argue that there is valuable information gain when we include an additional lateral x-ray image to a frontal x-ray, and therefore explore various possibilities of modifying existing architectures to accommodate a second image such that it can benefit from the additional information to the greatest extent.

Our aim is to design models which can take both frontal and lateral chest x-ray images and successfully extract meaningful features from both the images which it uses to make its predictions.

III. Related Work

One of the first research efforts made towards multi-view image analysis using CNN was by Setio et al. in [11]. Here the authors use techniques such as committee-fusion and late-fusion to design a multi-view CNN to reduce false positives in Pulmonary Nodule detection in CT images. They used scans from the LIDC-IDRI dataset and claim that their multi-view ConvNets are suited to for false positive reduction of a CAD system. Havaei, et al. in [12] proposes the Hetero-Modal Image segmentation model (HeMIS) where inputs of different modalities are processed different sequences of layers, point-wise mean/moments are computed and concatenated, and passed on to a single series of convolution layers. A model similar to HeMSI was designed by Bertrand et al. in [13] where they use series of 3 dense blocks to process frontal and lateral images parallelly, following which their outputs are averaged and forwarded to the final dense block. This model is illustrated in Fig. 2.

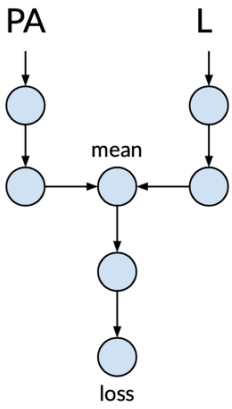


Fig. 2. Modified HeMIS architecture proposed in [13] to process x-ray image pairs.

Kitamura et al [14] designed an ensemble approach to utilize multiple X-rays of different views to detect fractures. Another very interesting model was proposed by Rubin et al., in [9]. They use two DenseNets to process the frontal and lateral images separately. The outputs are then passed through a global average pooling layer. This results are concatenated and fed to a fully connected layer. They used the MIMIC-CXR dataset to train and test their model. The work proposed in this report has resemblance to this model. Fig. 3.

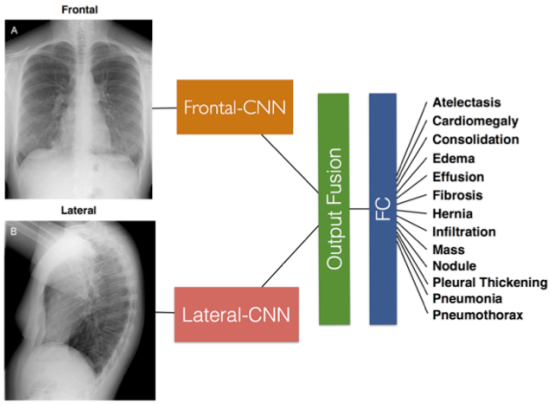


Fig. 3. DualNet proposed by Rubin et al. [9].

Li et al. in [15] proposes lesion detection technique in computed tomography (CT) images with a multi-view Feature Pyramid Network (FPN). The authors claim that their approach effectively combine this multi-view information. They also propose a position-aware attention module in the same paper. A comparative study of multi-view chest x-ray analysis using CNN based model is done in [19]. They also propose a model, called AuxLoss, by modifying the architecture proposed in [9].

IV. Proposed Approach

This project aims to detect abnormalities in chest x-rays by training a series of DenseNets-121 [6] with both frontal and lateral images. We specifically focus on identifying 5 of the 14 pathologies present in the CheXpert dataset. Fig. 4 (a-e) shows sample images of these pathologies.

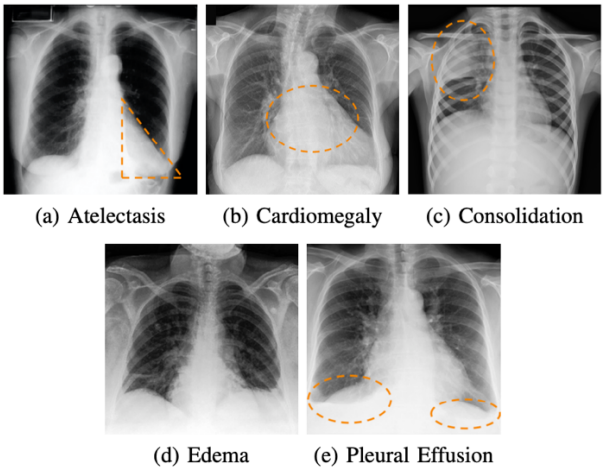


Fig. 4. Sample x-ray images of the 5 pathologies.

A. Dataset manipulation and data transformation

We use the CheXpert [2] dataset for all our training and experiments. This dataset has x-ray records of multiple studies for 65,240 patients. Table 1 shows the distribution of labels. The training set has 191,027 frontal images and 32,387 lateral images. This means that not all the frontal x-ray images have a corresponding lateral image. Moreover, we discovered that some of the studies have only lateral images, and some have three or more than three images.

TABLE I. DISTRIBUTION OF LABELS IN CHEXPERT

Pathology	Positive (%)	Uncertain (%)	Negetive (%)
No finding	16627 (8.86)	0 (0.0)	171014 (91.14)
Enlarged Cardiom.	9020 (4.81)	10148 (5.41)	168473 (89.78)
Cardiomegaly	23002 (12.26)	6597 (3.52)	158042 (84.23)
Lung Lesion	6856 (3.65)	1071 (0.57)	179714 (95.78)
Lung Opacity	92669 (49.39)	4341 (2.31)	90631 (48.3)
Edema	48905 (26.06)	11571 (6.17)	127165 (67.77)
Consoldation	12730 (6.78)	23976 (12.78)	150935 (80.44)
Pneumonia	4576 (2.44)	15658 (8.34)	167407 (89.22)
Atelectasis	29333 (15.63)	29377 (15.66)	128931 (68.71)
Pneumothorax	17313 (9.23)	2663 (1.42)	167665 (89.35)
Pleural Effusion	75696 (40.34)	9419 (5.02)	102526 (54.64)
Pleural Other	2441 (1.3)	1771 (0.94)	183429 (97.76)
Fracture	7270 (3.87)	484 (0.26)	179887 (95.87)
Support Devices	105831 (56.4)	898 (0.48)	80912 (43.12)

In order to resolve these issues we firstly decide to only consider studies which have both frontal and lateral images, and ignore all other studies. Secondly, for studies which have more than two x-ray images, we take all possible combination of frontal-lateral pairs. Finally, we get 33,151 pairs of frontal-lateral pairs, i.e. a total of 66,302 images. Table II is the distribution of labels in this dataset.

TABLE II. LABEL DISTRIBUTION IN PAIR-WISE DATASET

Pathology	Positive	Uncertain	Negative
Atelectasis	3743	3952	25456
Cardiomegaly	3690	1412	28049
Consolidation	1848	3433	27870
Edema	2617	1193	29341
Pleural_Effusion	9536	2095	21520

When compared to the original CheXpert dataset, this dataset has substantially less instances of all the labels. We

therefore augment the dataset by repeating the instances with minor changes. For instances with positive labels of *Consolidation* and *Edema* instances repetition is done thrice, and for all other cases it is done twice. Duplicate instances were flipped horizontally with a probability of 0.5. Apart from this, all images were normalized and randomly rotated by ± 10 degrees.

B. *Proposed System and techniques*

We used DenseNet-121 as our baseline CNN. We implement two models which is able to take two images as input.

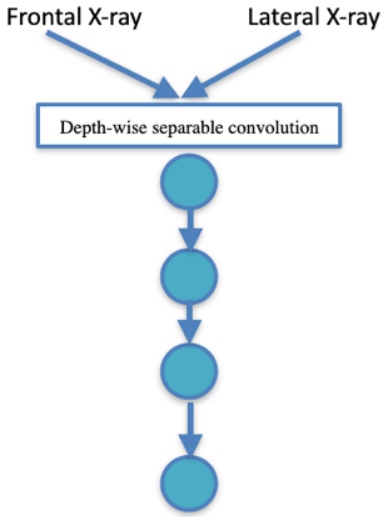


Fig. 5. Modified DenseNet architecture- DenseNet-S.

We call the first approach DenseNet-S (for Stacked). Here we make modification to the initial convolution layer of the DenseNet. The two views are resized to 512x512 and stacked as two channels of an image. The modified input is given as input which is processed by a depth-wise separable convolution layer, resulting in a tensor which is propagated through the 4 blocks of an unaltered DenseNet-121. Fig. 5 shows a graphical illustration of this model.

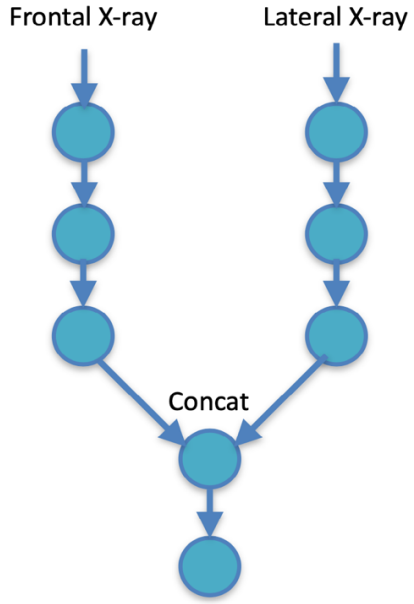


Fig. 6. Modified DenseNet architecture, DenseNet-D.

The second model we propose a model which processed the frontal and lateral images separately and concatenates the results before the fourth dense-block. This architecture which we call DenseNet-D, is shown in Fig.6.

For DenseNet-D we also a technique called hierarchical multi-label classification [7] which takes advantage of the pathology hierarchy of CheXpert by calculating conditional probability of the labels.

An example of such a hierarchy is shown in Fig. 7. The calculation of probability for Pneumonia, a leaf node in this case, is calculated as shown in Eqn. 1.

$$\begin{aligned} P(Pneumonia) = P(LungOpacity) \\ \cdot P(Consolidation|LungOpacity) \\ \cdot P(Pneumonia|Consolidation) \end{aligned} \quad (1)$$

This also ensures that the probability of child labels is less than or equal to parent label.

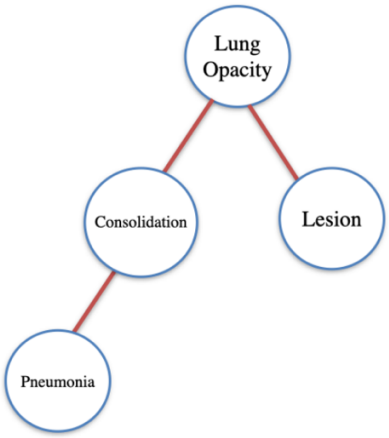


Fig. 7. Example label hierarchy.

The implementation is done in Python 3.7 using PyTorch as the primary deep learning library. We use *U-Ignore* as the training policy described by Irving, et al. in [2].

V. Experimental Results and Discussions

As explained in Section IV-A, we modify the CheXpert dataset before training our models and keep aside 350 pairs of x-ray images as the test set. To ensure that we have sufficient training examples we triplicate the instances of Consolidation and Edema, and we duplicate all other instances. The area under the receiver operating curve (AUC-ROC) is used as the evaluation metric; we calculate the AUC-ROC for each of the 5 classes and take the mean of these values (mean AUC) as the primary measure of performance,- higher the value, better is the performance. All experiments have been performed three times and the outputs are averaged to get the final results.

For DenseNet-S we observe that the performance increase is not as high. Here two input images are stacked and processed in a slightly altered DenseNet-121. This is explained in detail in Section IV-B. We compare the performance of DenseNet-S with an unmodified Densenet-121 which is given only the frontal images, and also with a DenseNet which is given the stacked image, but this model doesnot have a *depth-wise separable convolution layer*. The

average mean AUC performance of all three models is shown in Table III.

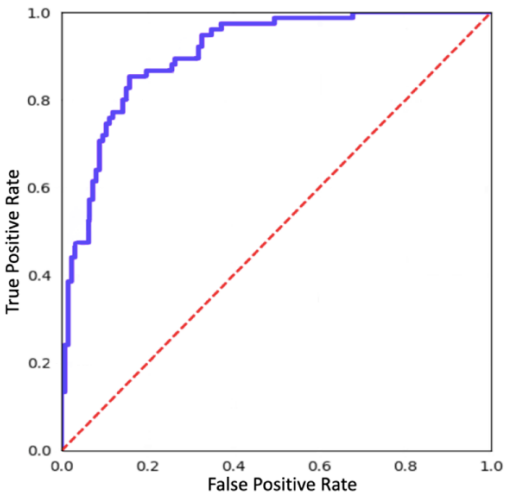
TABLE III. COMPARISION OF MEAN-AUC OF DENSENET-S AND OTHER MODELS

	DenseNet-121	*DenseNet-121 (Stacked input)*	*DenseNet-S*
without preprocessing	0.653	0.698	**0.768**
with preprocessing	0.721	0.733	**0.795**

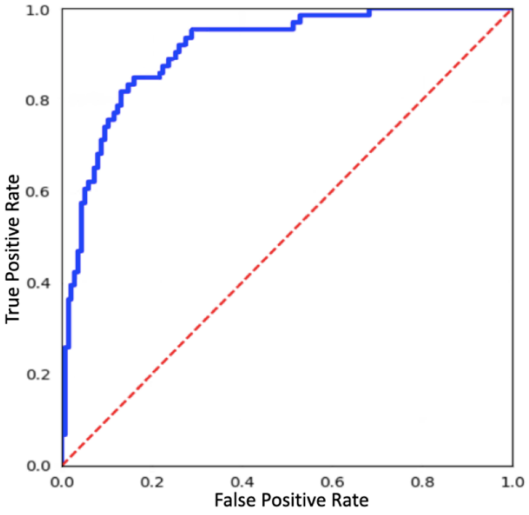
DenseNet-D is the second model we implemented. This model processes the frontal and lateral images separately in two arms of convolution blocks, and combines the output by concatenating and forwards this tensor to a final block. We used exactly the same preprocessing and data augmentation policies here as well. We compare the mean-AUC result of DenseNet-D with the mean-AUC of DenseNet trained on only frontal x-ray images from the full CheXpert dataset. We also performed Conditional Hierarchical Classification with DenseNet-D. This approach gives the highest mean-AUC value. The AUC curve for the five classes is also shown in Fig. 8. Due to time constraint we could only do the experiment with the preprocessed pair-wise image dataset.

TABLE IV. COMPARISION OF MEAN-AUC OF DENSENET-D

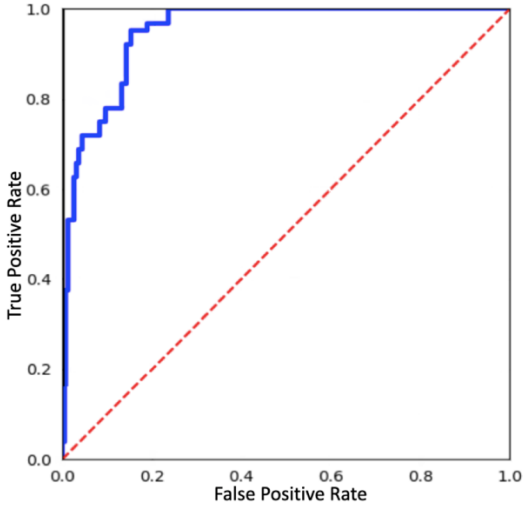
	DenseNet-121 (frontal)	*DenseNet-D (pair-wise)*	*DualNet [9]*	*DenseNet-D (Hierarchical Classification)*
without pre-proc.	0.653	**0.720**	-	-
with pre-proc.	0.721	0.802	0.721	**0.814**



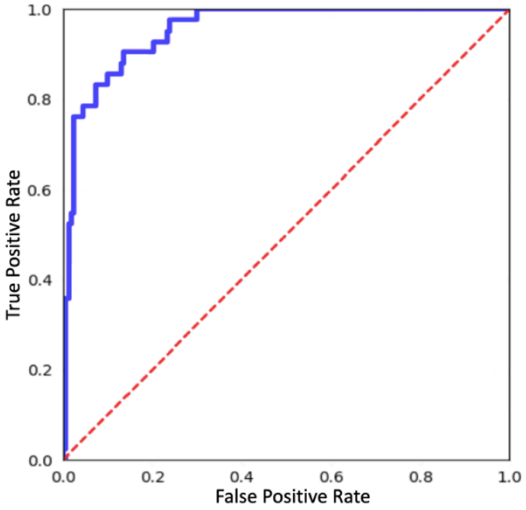
(a) Altelectasis



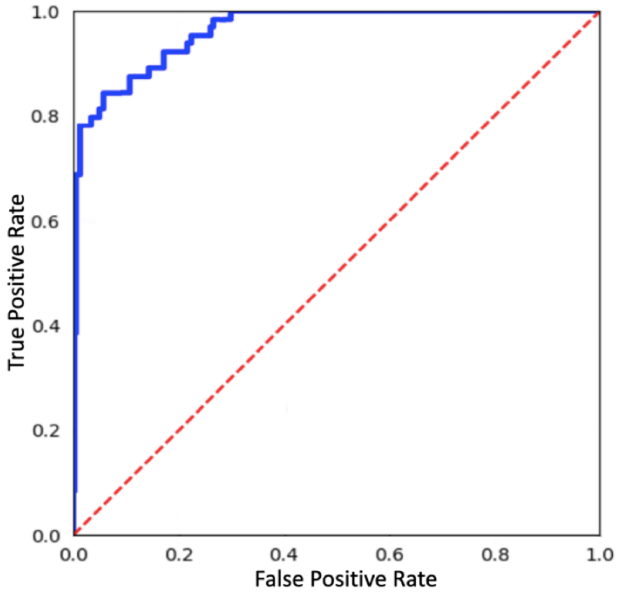
(c) Cardiomegaly



(c) Consolidation



(d) Edema



e) Pleural Effusion

Fig. 8. AUC-ROC for 5 classes classified with DenseNet-D and hierarchical conditional probability.

VI. Conclusion and Future Work

In project we have implemented and compared two models which take both frontal and lateral chest x-ray images for detecting torax diseases. The results indicate that the inclusion of lateral images can improve the overall performance. It is also observed that DenseNet-D performs better than DenseNet-S. It is our belief that the performance of these models are restricted by the limited number of examples in the pair-wise dataset. Also, the hierarchical conditional probability seems to improve the mean-AUC values.

In future work, we plan to implement self-learning and noisy-student learning policy [10]. Instead of averaging performance three times, a more robust testing scheme with ensemble of the same models is also planned. We also plan to incorporate additional image pairs to the dataset so that the results are comparable to networks which are trained with about quarter million images.

References

[1] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. (2017). ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3462-3471.

[2] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D., Halabi, S., Sandberg, J., Jones, R., Larson, D., Langlotz, C., Patel, B.N., Lungren, M., & Ng, A. (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. *AAAI*.

[3] Johnson, A.E., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M., Deng, C., Mark, R., & Horng, S. (2019). MIMIC-CXR: A large publicly available database of labeled chest radiographs. *ArXiv, abs/1901.07042*.

[4] Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D., & Lyman, K. (2017). Learning to diagnose from scratch by exploiting dependencies among labels. *ArXiv, abs/1710.10501*.

[5] Cohen, J.P., Bertin, P., & Frappier, V. (2019). Chester: A Web Delivered Locally Computed Chest X-Ray Disease Prediction System. *ArXiv, abs/1901.11210*.

[6] Huang, G., Liu, Z., & Weinberger, K.Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261-2269.

[7] Chen, H., Miao, S., Xu, D., Hager, G., & Harrison, A.P. (2019). Deep Hierarchical Multi-label Classification of Chest X-ray Images. *MIDL*.

[8] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M., & Ng, A. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *ArXiv, abs/1711.05225*.

[9] Rubin, J., Sanghavi, D., Zhao, C., Lee, K., Qadir, A., & Xu-Wilson, M. (2018). Large Scale Automated Reading of Frontal and Lateral Chest X-Rays using Dual Convolutional Neural Networks. *ArXiv, abs/1804.07839*.

[10] Xie, Q., Hovy, E., Luong, M., & Le, Q.V. (2020). Self-Training With Noisy Student Improves ImageNet Classification. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684-10695.

[11] Setio, A.A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., Riel, S.J., Wille, M., Naqibullah, M., Sánchez, C., & Ginneken, B. (2016). Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. *IEEE Transactions on Medical Imaging, 35*, 1160-1169.

[12] Havaei, M., Guizard, N., Chapados, N., & Bengio, Y. (2016). HeMIS: Hetero-Modal Image Segmentation. *MICCAI*.

[13] Bertrand, H., Hashir, M., & Cohen, J.P. (2019). Do Lateral Views Help Automated Chest X-ray Predictions? *ArXiv, abs/1904.08534*.

[14] Kitamura, G., Chung, C.Y., & Moore, B.E. (2019). Ankle Fracture Detection Utilizing a Convolutional Neural Network Ensemble Implemented with a Small Sample, De Novo Training, and Multiview Incorporation. *Journal of Digital Imaging, 32*, 672 - 677.

[15] Li, Z., Zhang, S., Zhang, J., Huang, K., Wang, Y., & Yu, Y. (2019). MVP-Net: Multi-view FPN with Position-aware Attention for Deep Universal Lesion Detection. *MICCAI*.

[16] Hashir, M., Bertrand, H., & Cohen, J.P. (2020). Quantifying the Value of Lateral Views in Deep Learning for Chest X-rays. *MIDL*.