Learning Low-Rank Feature for Thorax Disease Classification

Yancheng Wang[1] Rajeev Goel[1] Utkarsh Nath[1] Alvin C. Silva[2] Teresa Wu[1]
Yingzhen Yang[1]
[1] School of Computing and Augmented Intelligence, Arizona State University
{ywan1053, rgoel15, unath, teresa.wu, yingzhen.yang}@asu.edu
[2] Mayo Clinic Arizona
silva.alvin@mayo.edu

Introduction

- Recent studies have developed deep neural networks (DNNs), including CNNs and ViTs, for various tasks in medical imaging, such as disease classification and abnormalities detection in anatomy in chest X-ray.
- We study thorax disease classification in this paper.
- **Challenges in the Current Literature for Disease Classification.** Effective and robust extraction of features for the disease areas is crucial for disease classification on radiographic images. Although various neural architectures, such as CNNs and ViTs, and different training techniques, such as self-supervised learning with contrastive/restorative learning, have been employed for disease classification on radiographic images, there have been no principled methods that can effectively reduce the adverse effect of noise and background, or non-disease areas, for disease classification on radiographic images.

Contributions

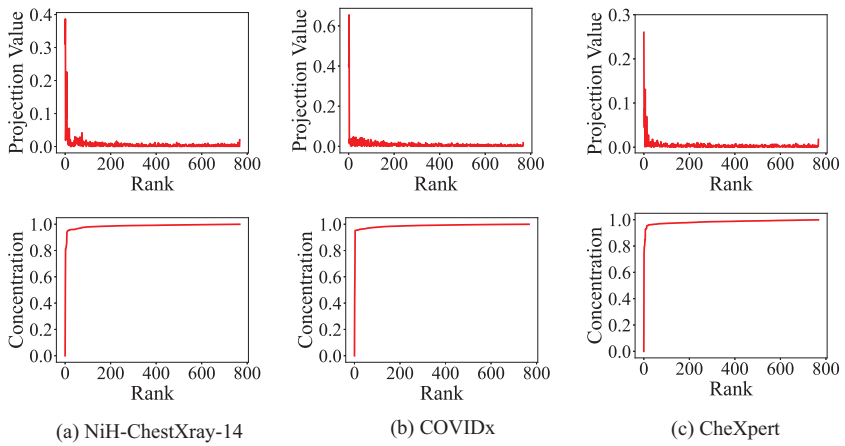

Figure 1: Illustration of LFP.

Eigen-projection (first row) and signal concentration ratio (second row) of Vit-Base on NiH-ChestXray-14, COVIDx, and CheXpert. Please refer to the details about the computation of the eigen-projection in the main paper. The signal concentration ratio for the rank $r = 38$ on NIH ChestX-ray14, COVIDx, and CheXpert are 0.959, 0.964, and 0.962 respectively.

(1) In order to address the aforementioned challenge, we propose a novel Low-Rank Feature Learning (LRFL) method in this paper, which is universally applicable to the training of all neural networks with the application for thorax disease classification. Our LRFL method employs low-rank features for disease classification. The usage of low-rank features is empirically motivated by a Low Frequency Property (LFP) illustrated in Figure 1. That is, the low-rank projection of the ground truth training class labels possesses the majority of the information of the training class labels.

Contributions (Cont'd)

(2) We provide a theoretical analysis showing a sharp generalization bound for the LRFL method, underscoring the substantial benefits of employing low-rank regularization within LRFL. It is worthwhile to mention that the literature has studied low-rank learning using the Truncated Nuclear Norm (TNN) resembling LRFL. Our LRFL method builds upon these foundational principles by incorporating low-rank regularization into the training of neural networks, aiming to improve thorax disease classification by reducing the adverse effects of noise and irrelevant background information. **Different from the conventional low-rank learning methods, our approach introduces a separable approximation to the TNN, facilitating the optimization process and enhancing the generalization ability of the model**.

- Moreover, we have employed a conditional diffusion model trained on COVIDx and CheXpert datasets to generate synthetic images. These synthetic images are then added to their respective training sets to form the augmented training data on which our LRFL models are trained. This approach has further elevated the state-of-the-art mAUC scores achieved by LRFL on both COVIDx and CheXpert datasets.

Formulation

- **Training Pipeline.** We utilize the masked MAE technique [1] for the initial pre-training of both CNNs and ViTs following [2], and subsequently fine-tune the pre-trained networks with our LRFL. The full training pipeline has three steps: (1) the **pre-training** step, where we pre-train the networks using the self-supervised restorative learning method, masked MAE [1], on a diverse pre-training dataset that includes ImageNet-1k [3] and a collection of X-rays (0.5M) [2]. (2) the **regular fine-tuning** step, where we fine-tune the pre-trained networks employing cross-entropy loss aimed at image classification on specific target datasets, namely NIH-ChestX-ray [4], COVIDx [5], and CheXpert [6]; (3) the **low-rank feature learning** step, where we fix the backbones of the networks and fine-tune the linear classifier utilizing our novel LRFL method.

Formulation (Cont'd)

- **Notations.** Suppose the training data are given as $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^n$ where $\mathbf{x}_i$ and $\mathbf{y}_i \in \mathbb{R}^C$ are the i-th training data point and its corresponding class label vector respectively, and C is the number of classes. Each element $\mathbf{y}_i$ is binary with $\mathbf{y}_i = 1$ indicating the i-th disease is present in $\mathbf{x}_i$, otherwise $\mathbf{y}_i = 0$. Suppose that the neural network trained by step two of our pipeline generates a feature vector $f_{\mathbf{W}_1(0)}(\mathbf{x}) \in \mathbb{R}^d$ (the output of the layer preceding the final linear/softmax layer of the network) for any input x, and $f_{\mathbf{W}'}(\cdot)$ is the feature extraction function with $\mathbf{W}'$ being the weights of the feature extraction backbone of the network. $\mathbf{W}_1(0)$ denotes the denotes the weights of feature extraction backbone by step two of the pipeline. We can train a neural network by optimizing

$$\min_{\mathbf{W}} L(\mathbf{W}) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{KL}\left(\mathbf{y}_i, \sigma\left(\mathbf{W}_2 f_{\mathbf{W}_1(0)}(\mathbf{x})\right)\right) \tag{1}$$

in the second step of the pipeline, where $\mathbf{W}_1$ is initialized by $\mathbf{W}_1(0)$, $\mathbf{W}_2 \in \mathbb{R}^{C \times d}$, and $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2)$. Here σ is an element-wise sigmoid function, KL stands for the element-wise binary cross-entropy function.

Formulation (Cont'd)

- In the third step of our pipeline, we add a novel approximate truncated nuclear norm $\overline{\|\mathbf{F}\|_T}$ to the trainng loss $L(\mathbf{W})$. Let the Singular Value Decomposition of $\mathbf{F}$ be $\mathbf{F} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, then the approximate TNN is computed by $\overline{\|\mathbf{F}\|_T} = \sum_{i=1}^{n} \left(\sum_{s=T+1}^{d} \sum_{k=1}^{d} \overline{\mathbf{U}}_{si}^\top \mathbf{F}_{ik} \overline{\mathbf{V}}_{ks} \right)$, where $\overline{\mathbf{U}}$ is an approximation to $\mathbf{U}$ and $\overline{\mathbf{V}}$ is an approximation to $\mathbf{V}$.
- The training loss function of LRFL with the approximate truncated nuclear norm $\overline{\|\mathbf{F}\|_T}$ is $\mathcal{L}_{\text{LRFL}}(\mathbf{W}) = \frac{1}{m} \sum_{v_i \in \mathcal{V}_\mathcal{L}} \text{KL}\left(\mathbf{y}_i, \left[\sigma\left(\mathbf{F}\mathbf{W}^{(\text{lin})}\right)\right]_i\right) + \eta\overline{\|\mathbf{F}\|_T}$, which is separable, so that it can be trained by the standard SGD.
- The approximation $\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}$ can be computed as the left and right eigenvectors of the feature $\mathbf{F}$ computed at earlier epochs. In order to save computation and avoiding performing SVD for $\mathbf{F}$ at every epoch, we propose to update $\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}$ only after certain epochs. Please refer to Algorithm 1 of the main paper for more details.

Generalization Bound for Low-Rank Feature Learning

- We define the loss function $\ell(\mathsf{NN}_{\mathbf{W}}(\mathbf{x}), \mathbf{y}) := \|\mathsf{NN}_{\mathbf{W}}(\mathbf{x}) - \mathbf{y}\|_2^2$, and the generalization error of the network is the expected risk of the loss ℓ, $L_{\mathcal{D}}(\mathsf{NN}_{\mathbf{W}}) := \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}}\left[\ell(\mathsf{NN}_{\mathbf{W}}(\mathbf{x}), \mathbf{y})\right]$, with $\mathcal{D}$ being the distribution of the data $\mathbf{x}$ and its class label $\mathbf{y}$. The network $\mathsf{NN}_{\mathbf{W}}$ generates a feature $\mathbf{F} \in \mathbb{R}^{n\times d}$ of all the training data with $\mathbf{F}_i = f_{\mathbf{W}_1}^{\top}(\mathbf{x}_i)$ for $i \in [n]$. The kernel gram matrix for the feature $\mathbf{F}$ is $\mathbf{K}_n = \frac{1}{n}\mathbf{F}\mathbf{F}^{\top}$. We let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \ldots \geq \widehat{\lambda}_{\bar{r}} > 0$ where $\bar{r} \leq \min\{n, d\}$ is the rank of $\mathbf{K}_n$. Let $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_d$ be the singular values of $\mathbf{F}$, and $\bar{\mathbf{Y}} = \mathbf{U}^{(\bar{r})}\mathbf{U}^{(\bar{r})^{\top}}\mathbf{Y}$ be the projection of the training label matrix $\mathbf{Y}$ onto the subspace spanned by the top-$\bar{r}$ left eigenvectors of $\mathbf{F}$, where $\mathbf{U}^{(\bar{r})} \in \mathbb{R}^{n\times\bar{r}}$ is formed by the top $\bar{r}$ eigenvectors in $\mathbf{U}$. Then, we have the following theorem giving the sharp generalization error bound for the linear neural network in (1).

Theorem 0.1

For every $x > 0$, with probability at least $1 - \exp(-x)$, after the t-th iteration of gradient descent on the loss $L(\mathbf{W})$ for all $t \geq 1$, we have

$$L_{\mathcal{D}}(\mathsf{NN}_{\mathbf{W}}) \leq \|\mathbf{Y} - \bar{\mathbf{Y}}\|_{\mathsf{F}} + c_1\left(1 - \eta\widehat{\lambda}_r\right)^{2t}\|\mathbf{Y}\|_{\mathsf{F}}^2 + c_2 \min_{h\in[0,r]}\left(\frac{h}{n} + \sqrt{\frac{1}{n}\sum_{i=h+1}^{r}\widehat{\lambda}_i}\right) + \frac{c_3 x}{n}. \tag{2}$$

Generalization Bound for Low-Rank Feature Learning

- The RHS of (2) is the generalization error bound for the linear neural network used in LRFL as step three of the pipeline. Moreover, let $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_d$ be the singular values of $\mathbf{F}$. Due to the fact that $\sqrt{\frac{1}{n}\sum_{i=h+1}^{r}\widehat{\lambda}_i} \leq \frac{1}{n}\sum_{i=h+1}^{r}\sigma_i$, it follows by (2) that

$$L_{\mathcal{D}}(\mathsf{NN}_{\mathbf{W}}) \leq c_1\left(1-\eta\widehat{\lambda}_r\right)^{2t}\|\mathbf{Y}\|_{\mathsf{F}}^2 + c_2\left(\frac{h}{n}+\frac{1}{n}\sum_{i=T+1}^{d}\sigma_i\right)+\frac{c_3 x}{n}, \tag{3}$$

  which holds for all $T \in [0, d]$. (3) motivates the reduction of the truncated nuclear norm of the feature $\mathbf{F}$.

Experiments

- We evaluate the LRFL models for thorax disease classification on CheXpert, COVIDx, and NIH ChestX-ray14.

Table 1: Performance comparisons between LRFL models and SOTA baselines on CheXpert. The best result is highlighted in bold, and the second-best result is underlined. This convention is followed by all the tables in this paper. DN represents DenseNet.

Method	Architecture	Rank	Atelectasis	Cardiomegaly	Consolidation	Edema	Effusion	mAUC (%)
Irvin et al. [6]		-	81.8	82.8	<u>93.8</u>	93.4	92.8	88.9
Pham et al. [7]	DN121	-	82.5	85.5	93.7	93.0	92.3	89.4
Kang et al. [8]	DN121	-	82.1	85.9	**94.4**	89.2	93.6	89.0
MoCo v2 [2]	DN121	-	78.5	77.9	92.5	92.8	92.7	88.7
ViT-S [2]	ViT-S/16	-	<u>83.5</u>	81.8	93.5	<u>94.0</u>	93.2	89.2
ViT-S-LR (Ours)	ViT-S/16	0.05r	**83.7**	<u>86.3</u>	90.9	93.7	93.1	<u>89.6</u>
ViT-B [2]	ViT-B/16	-	82.7	83.5	92.5	93.8	**94.1**	89.3
ViT-B-LR (Ours)	ViT-B/16	0.05r	81.6	85.4	93.4	**94.6**	<u>93.9</u>	**89.8**

Experiments (Cont'd)

Table 2: Performance comparisons between LRFL models and SOTA baselines on COVIDx (in accuracy). DN represents DenseNet.

Method	Architecture	Rank	Covid-19 Sensitivity	Accuracy
COVIDNet-CXR Small [9]	-	-	87.1	92.6
COVIDNet-CXR Large [9]	-	-	96.8	94.4
MoCo v2 [2]	DN121	-	94.5	94.0
DN121 [2]	DN121	-	97.0	93.5
ViT-S [2]	ViT-S/16	-	94.5	95.2
ViT-S-LR (Ours)	ViT-S/16	0.01r	<u>97.5</u>	<u>96.8</u>
ViT-B [2]	ViT-B/16	-	95.5	95.3
ViT-B-LR (Ours)	ViT-B/16	0.003r	**98.5**	**97.0**

Table 3: Performance comparison of baseline models and LRFL models on the CheXpert and COVIDx datasets, with and without synthetic data. n denotes the number of training images in the respective dataset.

Method	Architecture	CheXpert			COVIDx		
		Rank	# Synthetic Images	mAUC (%)	Rank	# Synthetic Images	Accuracy (%)
ViT-S [2]	ViT-S/16	-	-	89.2	-	-	95.2
ViT-S-LR (Ours)	ViT-S/16	0.05r	-	89.6	0.01r	-	96.8
ViT-S (Ours)	ViT-S/16	-	$0.2n$	89.3	-	$1.0n$	97.0
ViT-S-LR (Ours)	ViT-S/16	0.05r	$0.2n$	89.7	0.01r	$1.0n$	<u>97.3</u>
ViT-B [2]	ViT-B/16	-	-	89.3	-	-	95.3
ViT-B-LR (Ours)	ViT-B/16	0.025r	-	89.8	0.003r	-	97.0
ViT-B (Ours)	ViT-B/16	-	$0.25n$	<u>89.9</u>	-	$1.0n$	97.0
ViT-B-LR (Ours)	ViT-B/16	0.025r	$0.25n$	**90.4**	0.003r	$1.0n$	**97.5**

Experiments (Cont'd)

Table 4: Performance comparisons between LRFL models and SOTA baselines on NIH ChestX-ray14. RN, DN, and SwinT represent ResNet, DenseNet, and Swin Transformer.

Method	Architecture	Pre-training	Rank	mAUC
Wang et al. [4]	RN50	ImageNet-1K	-	74.5
Li et al. [10]	RN50		-	75.5
Yao et al. [11]	RN&DN		-	76.1
Wang et al. [12]	R152		-	78.8
Ma et al. [13]	R101		-	79.4
Tang et al. [14]	RN50		-	80.3
Baltruschat et al. [15]	RN50		-	80.6
Guendel et al. [16]	DN121		-	80.7
Guan et al. [17]	DN121		-	81.6
Seyyed et al. [18]	DN121		-	81.2
Ma et al. [19]	DN121($\times 2$)		-	81.7
Hermoza et al. [20]	DN121		-	82.1
Kim et al. [21]	DN121		-	82.2
Haghighi et al. [22]	DN121		-	81.7
Liu et al. [23]	DN121		-	81.8
Taslimi et al. [24]	SwinT		-	81.0
MoCo v2 [2]	DN121	X-rays (0.3M)	-	80.6
MAE [2]	DN121		-	81.2
RN-50 [2]	RN50	ImageNet-1K	-	81.8
RN-50-LR (Ours)	RN50		0.05r	82.2
DN-121 [2]	DN121	ImageNet-1K	-	82.0
DN-121-LR (Ours)	DN121		0.05r	82.4
ViT-S [2]	ViT-S/16	X-rays (0.3M)	-	82.3
ViT-S-LR (Ours)	ViT-S/16		0.05r	82.7
ViT-B [2]	ViT-B/16	X-rays (0.5M)	-	<u>83.0</u>
ViT-B-LR (Ours)	ViT-B/16		0.05r	**83.4**

Experiments (Cont'd): Grad-CAM Visualization

- To study how LRFL improves the performance of base models in disease detection, we use the Grad-CAM to visualize the parts in the input images that are responsible for the predictions of the base models and low-rank models.

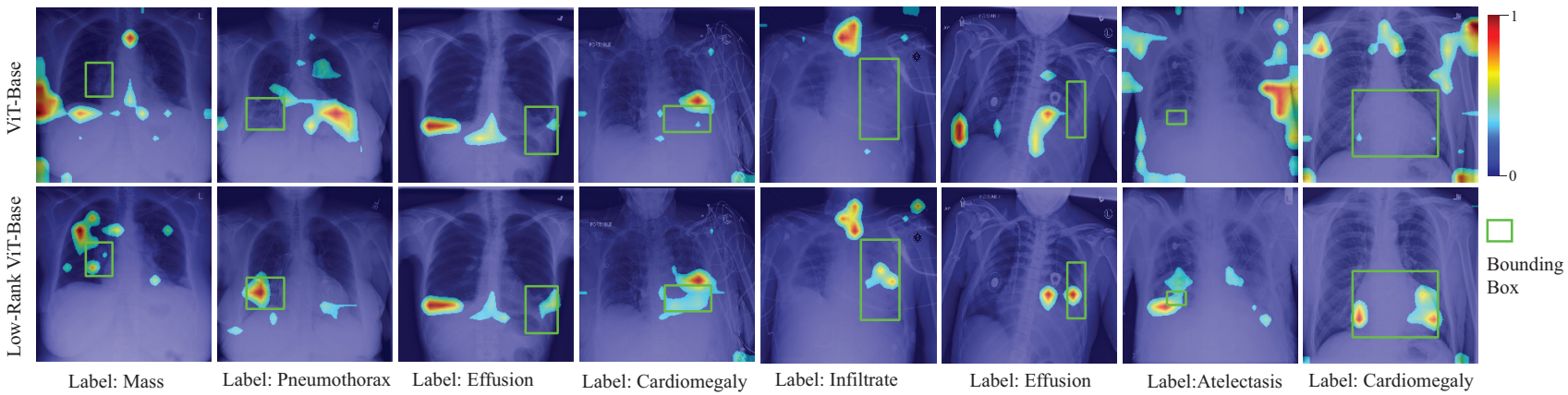

Figure 2: Robust Grad-CAM [25] visualization results on NIH ChestX-ray 14. The figures in the first row are the visualization results of ViT-Base, and the figures in the second row are the visualization results of Low-Rank ViT-Base.

Key Takeaways

- Low-rank feature learning improves the generalization of DNNs for disease classification, and it also reduces the adverse effects of the background and noise.
- A novel approximate TNN is proposed to improve the efficiency and the scalability of low-rank feature learning for DNNs by the standard SGD.
- Sharp generalization error bound for low-rank feature learning is proved.

Thank you!


K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 000–16 009.


J. Xiao, Y. Bai, A. Yuille, and Z. Zhou, "Delving into masked autoencoders for multi-label thorax disease classification," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 3588–3600.


A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.


X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.


M. Pavlova, T. Tuinstra, H. Aboutalebi, A. Zhao, H. Gunraj, and A. Wong, "Covidx cxr-3: a large-scale, open-source benchmark dataset of chest x-ray images for computer-aided covid-19 diagnostics," *arXiv preprint arXiv:2206.03671*, 2022.


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, 2019, pp. 590–597.


H. H. Pham, T. T. Le, D. Q. Tran, D. T. Ngo, and H. Q. Nguyen, "Interpreting chest x-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels," *Neurocomputing*, vol. 437, pp. 186–194, 2021.


M. Kang, Y. Lu, A. L. Yuille, and Z. Zhou, "Label-assemble: Leveraging multiple datasets with partial labels," *In Submission: Thirty-Sixth Conference on Neural Information Processing Systems*, 2021. [Online]. Available: https://arxiv.org/pdf/2109.12265.pdf


L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, no. 1, p. 19549, Nov 2020.


Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L.-J. Li, and L. Fei-Fei, "Thoracic disease identification and localization with limited supervision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8290–8299.


L. Yao, J. Prosky, E. Poblenz, B. Covington, and K. Lyman, "Weakly supervised medical diagnosis and localization from multiple resolutions," *arXiv preprint arXiv:1803.07703*, 2018.


H. Wang, H. Jia, L. Lu, and Y. Xia, "Thorax-net: an attention regularized deep neural network for classification of thoracic diseases on chest radiography," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 475–485, 2019.


Y. Ma, Q. Zhou, X. Chen, H. Lu, and Y. Zhao, "Multi-attention network for thoracic disease classification and localization," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1378–1382.


Y. Tang, X. Wang, A. P. Harrison, L. Lu, J. Xiao, and R. M. Summers, "Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2018, pp. 249–258.


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.


S. Guendel, S. Grbic, B. Georgescu, S. Liu, A. Maier, and D. Comaniciu, "Learning to recognize abnormalities in chest x-rays with location-aware dense networks," in *Iberoamerican Congress on Pattern Recognition*. Springer, 2018, pp. 757–765.


Q. Guan and Y. Huang, "Multi-label chest x-ray image classification via category-wise residual attention learning," *Pattern Recognition Letters*, 2018.


L. Seyyed-Kalantari, G. Liu, M. McDermott, I. Y. Chen, and M. Ghassemi, "Chexclusion: Fairness gaps in deep chest x-ray classifiers," in *BIOCOMPUTING 2021: Proceedings of the Pacific Symposium*. World Scientific, 2020, pp. 232–243.


C. Ma, H. Wang, and S. C. H. Hoi, "Multi-label thoracic disease image classification with cross-attention networks," 2020.

R. Hermoza, G. Maicas, J. C. Nascimento, and G. Carneiro, "Region proposals for saliency map refinement for weakly-supervised disease localisation and classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 539–549.

E. Kim, S. Kim, M. Seo, and S. Yoon, "Xprotonet: Diagnosis in chest radiography with global and local explanations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 15 719–15 728.

F. Haghighi, M. R. H. Taher, M. B. Gotway, and J. Liang, "Dira: Discriminative, restorative, and adversarial learning for self-supervised medical image analysis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20 824–20 834.

F. Liu, Y. Tian, Y. Chen, Y. Liu, V. Belagiannis, and G. Carneiro, "Acpl: Anti-curriculum pseudo-labelling for semi-supervised medical image classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20 697–20 706.

S. Taslimi, S. Taslimi, N. Fathi, M. Salehi, and M. H. Rohban, "Swinchex: Multi-label classification on chest x-ray images with transformers," *arXiv preprint arXiv:2206.04246*, 2022.

R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.