This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Symptom Disentanglement in Chest X-ray Images for Fine-Grained Progression Learning

Ye Zhu[1], Jingwen Xu[1], Fei Lyu[1], and Pong C. Yuen[1]

Department of Computer Science, Hong Kong Baptist University
{csyzhu,csjwxu,feilyu,pcyuen}@comp.hkbu.edu.hk

Abstract. Chest radiography is a commonly used diagnostic imaging exam for monitoring disease severity. Machine learning has made significant strides in static tasks (e.g., segmentation or diagnosis) based on a single medical image. However, disease progression monitoring based on longitudinal images remains fairly underexplored, which provides informative clues for early prognosis and timely intervention. In practice, the development of underlying disease typically accompanies with the occurrence and changes of multiple specific symptoms. Inspired by this, we propose a multi-stage framework to model the complex progression from symptom perspective. Specifically, we introduce two consecutive modules namely Symptom Disentangler (SD) and Symptom Progression Learner (SPL) to learn from static diagnosis to dynamic disease development. By explicitly extracting the symptom-specific features from a pair of chest radiographs using a set of learnable symptom-aware embeddings in SD module, the SPL module can leverage these features to obtain the symptom progression features, which will be utilized for the final progression prediction. Experimental results on the public dataset Chest ImaGenome show superior performance compared to current state-of-the-art method. Code is available at: https://github.com/zhuye98/SDPL.git.

Keywords: Symptom disentanglement · Progression learning · Symptom classification

1 Introduction

Chest radiography has significant clinical value as the most common, relatively low-cost diagnostic method in medical practice for diagnosing diseases like Pneumonia, COVID-19 and other lung ailments. Owing to these factors, many AI-based methods for finding detection are proposed and now approaching the performance level of experienced radiologists [3,1]. Although significant strides have been made in AI-assisted segmentation and disease diagnosis, most methods perform medical image analysis based on a single Chest X-ray. Less attention has been paid to monitoring disease progression in a sequence of CXRs, which is necessary for early prediction of adverse outcomes and timely intervention. For example, routine CXRs contain temporal lung changes, which serve as vital indicators of patient outcomes [18]. However, disease progression monitoring

presents a formidable challenge due to the complex and intricate nature of symptoms on CXR, which makes it non-trivial to distinguish significant changes in longitudinal CXR with diverse symptoms. For instance, multiple symptoms can be detected on each CXR while these symptoms exhibit intricate relationships, and each symptom severity can dynamically change in a sequence of CXRs.

Previous works generally use a single CXR study to perform CXR screening, precluding comparison with prior scans [16,6]. Recent works have investigated tracking the progression of disease severity with longitudinal CXRs [2,19]. In [17], geometric correlation maps are used for change detection relying on elaborate feature alignment. In contrast, CheXRelNet [10] proposes to capture anatomical correlation changes in CXRs based on graph attention networks, while CheXRelFormer [15] further enhances anatomical change detection with multi-level feature computation based on a hierarchical vision Transformer. Moreover, [4] creates a joint representation by aligning features for equivalent anatomical region projection in longitudinal CXRs for report generation.

In addition to chest X-rays, the detection of progression between longitudinal patient visits has also been explored across various modalities. For example, the dynamic abnormality detection for knee MRI scans and osteoarthritis in knee radiographs[11,7]. Many works also focus on monitoring the progression of cognitive impairment and Alzheimer's disease using Brain structural MRI [22,12]. Retinopathy in retinal photographs has also gained considerable attention in recent years since some eye diseases are irreversible [21,5,14]. Nonetheless, prior works focus on the visual detail changes in anatomical regions. They suffer from capturing the dynamics and correlations of diverse symptoms, which are regarded as high-level medical semantics and closely related to disease development.

To tackle the above challenges, our motivation is from symptom-aware disease progression modeling: As the medical manifestation of disease, the occurrence and changes of specific symptoms reflect the development of disease in longitudinal CXRs. Inspired by this, we propose an end-to-end framework that aims to identify and disentangle the symptom-specific features for symptom-level progression modeling. Specifically, we introduce two consecutive modules namely Symptom Disentangler (SD) and Symptom Progression Learner (SPL): (1) The SD module explicitly extracts the symptom-specific features in each chest radiograph with a set of learnable symptom-aware embeddings. (2) The SPL module learns to capture the symptom-level changes by connecting the extracted features from two CXRs. Experimental results on the Chest ImaGenome dataset show superior performance compared to the current state-of-the-art method.

In conclusion, our contributions can be summarized as follows: (1) We propose a novel end-to-end symptom-level progression learning framework that can accurately capture the subtle changes in a pair of chest X-ray images. (2) Our newly introduced Symptom Disentangler can explicitly extract symptom-specific features from chest X-ray images, obviating the requirement for any image registration techniques. This approach fosters an optimal setting for fine-grained disease progression learning. (3) We conduct experiments on a publicly available

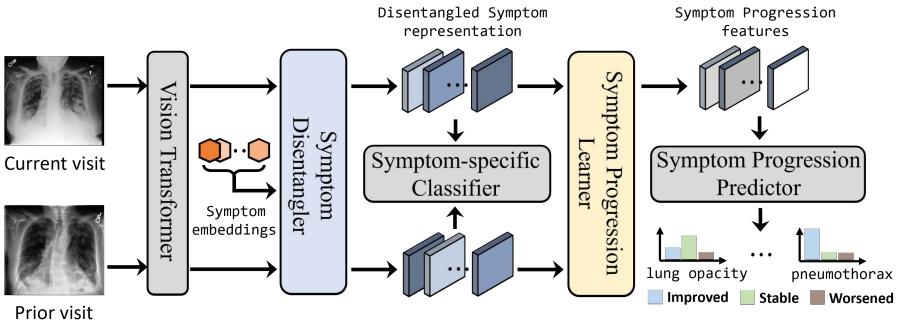

Fig. 1: Overview of the proposed framework: Two CXR images (X, X') are processed by a shared Transformer encoder for tokenization and feature extraction, where X and X' represent the current and the prior image, respectively. These tokenized features are then used by the symptom disentangler to obtain disentangled symptom representations. The symptom progression learner utilizes these symptom-specific features to capture changes between the two CXR images, enabling the extraction of progression information for the final prediction.

dataset Chest ImaGenome, and experimental results demonstrate the effectiveness of our model in predicting the progression status.

2 Proposed Method

Overview. Given a set of current-prior CXR image pairs $D = \{(X, X')_i\}_{i=1}^{N}$ where N is the amount of image pairs, with its record of the progression status $Y_{i,s}^{prog} \in \{0, 1, 2\}$ that indicates whether the symptom s within the image pair is worsened, improved or stable, and a label set $Y_j^{symp} = \{y_j^{symp}\}_{j=1}^{M}$ for symptom classification, where M is the total number of CXR images, $y_j^{symp} = [y_1, \ldots, y_Z], y_Z \in \{0, 1\}$, and $y_Z = 1$ indicates the presence of a corresponding disease symptom s in the images and Z is the number of symptoms in this study, our aim is to accurately predict the progression at a symptom-level between a patient's two visits.

As illustrated in Fig. 1, we introduce two consecutive modules namely Symptom Disentangler (SD) and Symptom Progression Learner (SPL), and a set of learnable symptom-aware embeddings $\mathbf{Q} \in \mathbb{R}^{Z \times 8C}$ that can explicitly extract the symptom representation from CXR images. C is the feature dimension at the first scale and $8C$ is the feature dimension of the learnable embeddings.

2.1 Symptom Disentangler

To capture the changes exhibited in patients between two visits on a more granular level, different symptoms' features should be first located and extracted

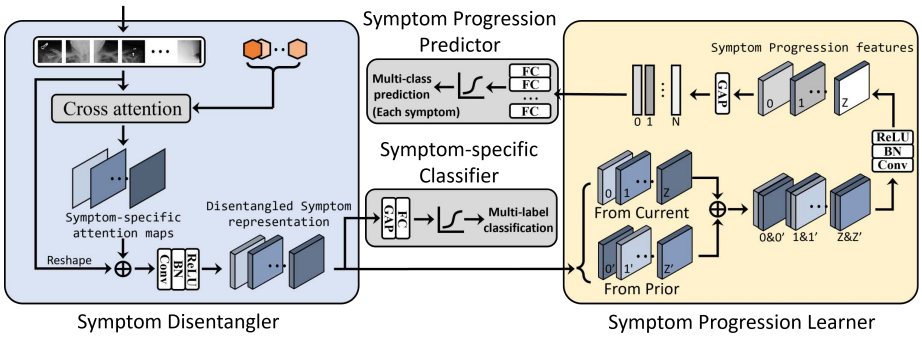

Fig. 2: Details of the Symptom Disentangler and Symptom Progression Learner.

from the CXR images. To this end, we initialize a set of learnable symptom-aware embeddings to interact with the tokenized features in the Cross-Attention (CA) modules, obtaining the corresponding attention maps for each symptom. After applying a series of convolution operations to the obtained attention maps, the latent representation of each symptom is derived and forwarded to the SPL module for symptom-level progression learning.

As depicted in Fig. 2, a pair of current-prior CXR images are first passed to the transformer encoder to get the tokenized features $\{\mathbf{T}, \mathbf{T}'\} \in \mathbb{R}^{8C \times \left(\frac{\mathrm{H}}{32} \times \frac{\mathrm{W}}{32}\right)}$. Along with the initialized learnable embeddings $\mathbf{Q} \in \mathbb{R}^{Z \times 8C}$ as the input queries, these obtained tokenized features are sent to a series of Multi-head Cross Attention (MCA) modules as the input keys and values to obtain the symptom-aware attention maps. For simplicity, we illustrate the process using the tokenized feature T from the current visit as follows,

$$\mathbf{q} = \mathbf{Q}\mathbf{W}_Q, \mathbf{k} = \mathbf{T}\mathbf{W}_K, \mathbf{v} = \mathbf{T}\mathbf{W}_V \tag{1}$$

$$\mathbf{A}_j = \mathrm{CrossAttentionMap}_j\left(\mathbf{q}, \mathbf{k}\right) = \mathbf{q}\mathbf{k}^{\mathrm{T}} \tag{2}$$

$$\mathbf{A}_{symp} = \mathrm{MCA}\left(\mathbf{q}, \mathbf{k}\right) = \mathrm{Concat}\left(\mathbf{A}_1, ..., \mathbf{A}_N\right)\mathbf{W}_O \tag{3}$$

where $\mathbf{W}_Q$, $\mathbf{W}_K$, $\mathbf{W}_V \in \mathbb{R}^{8C \times 8C'}$, $\mathbf{W}_O \in \mathbb{R}^{8C' \times 8C}$ are the learnable parameter for linear projection, the d is the channel dimension $8C$, $8C' = 8C/N$, and N is the number of heads. $\mathbf{A}_j \in \mathbb{R}^{Z \times \left(\frac{\mathrm{H}}{32} \times \frac{\mathrm{W}}{32}\right)}$ is the attention map extracted from j-th head. Then the attention maps are concatenated from different heads. Finally, we can obtain the symptom-specific attention maps $\mathbf{A}_{symp} \in \mathbb{R}^{Z \times N \times \left(\frac{\mathrm{H}}{32} \times \frac{\mathrm{W}}{32}\right)}$ from the multi-head perspective. To enhance information retention from the encoder feature and extract symptom-specific representations, we further concatenate and fuse the original tokenized feature $\mathbf{T}$ with the extracted attention maps via a convolution operation:

$$\begin{aligned} \mathbf{F}_{symp} &= \mathrm{Conv2D}\left(\left[\mathbf{A}_{symp}, \mathbf{T}\right]\right) \\ \mathbf{F}_{symp} &= \mathrm{ReLU}\left(\mathbf{F}_{symp}\right) \end{aligned} \tag{4}$$

where $\mathbf{F}_{symp} \in \mathbb{R}^{Z \times C \times (\frac{\mathrm{H}}{32} \times \frac{\mathrm{W}}{32})}$ is the obtained symptom-specific representations after fusion. Then a classifier head is applied to get the multi-label prediction for the symptom classification task:

$$\hat{y}^{symp} = \text{Classifier}\left(\text{pool}\left(\mathbf{F}_{symp}\right)\right) \tag{5}$$

where pool denotes the global average pooling operation.

2.2 Symptom Progression Learner

As depicted in Fig. 2, the SPL module receives the disentangled symptom-specific representations $\left\{\mathbf{F}_{symp}, \mathbf{F}^{'}_{symp}\right\}$ from both current and prior visits. To further capture the progression information of each symptom, we combine the representation of each symptom to obtain the symptom progression features via concatenation and convolution operations,

$$\begin{aligned} \mathbf{F}_{prog} &= \text{Conv2D}\left(\left[\mathbf{F}_{symp}, \mathbf{F}^{'}_{symp}\right]\right) \\ \mathbf{F}_{prog} &= \text{ReLU}\left(\mathbf{F}_{prog}\right) \end{aligned} \tag{6}$$

where $\mathbf{F}_{prog} = [\mathbf{F}_{p.1}, ..., \mathbf{F}_{p.z}, ...\mathbf{F}_{p.Z}] \in \mathbb{R}^{Z \times C \times (\frac{\mathrm{H}}{32} \times \frac{\mathrm{W}}{32})}$ contains a set of symptom progression feature vectors. Each symptom progression feature $\mathbf{F}_{p.1} \in \mathbb{R}^{1 \times C \times (\frac{H}{32} \times \frac{W}{32})}$ contains the intricate progression information embedded within each pair of symptom-specific features. And finally we obtain the final symptom progression prediction by the corresponding predictor:

$$\hat{y}^{prog}_{z} = \text{Predictor}_{z}\left(\text{pool}\left(\mathbf{F}_{p.z}\right)\right) \tag{7}$$

2.3 The Overall Loss Function

The goal of our framework is to minimize the following combined objective function that contains the symptom classification task and symptom progression prediction task:

$$L = \sum_{i=1}^{N}\left(\sum_{j=1}^{2} L_{BCE}(\hat{y}^{symp}_{i,\,j},\ y^{symp}_{i,\,j}) + \sum_{z=1}^{Z} L_{CE}(\hat{y}^{prog}_{i,\,z}, y^{prog}_{i,\,z})\right), \tag{8}$$

where the N and Z denote the number of CXR image pairs and the number of symptoms, L_{BCE} is the Binary Cross-Entropy loss function, and L_{CE} is the Cross-Entropy loss function.

3 Experiments

3.1 Implementation Details

For a fair comparison with the baseline CheXRelFormer [15], we utilize the same Transformer encoder [20] without pretraining. The model is trained using an

AdamW optimizer [13] with an initial learning rate of 6×10^{-5} and a batch size of 16. The training process takes place on a single V100 GPU for 100 epochs. The number of model parameters of CheXRelFormer is 41.0M while ours is 32.2M, and the computational complexity (FLOPs) of CheXRelFormer is 20.4G while ours is 9.5G.

3.2 Dataset and evaluation metrics

This study conducts experiments on the Chest ImaGenome dataset, following the preprocessing steps described in [15] to obtain pairs of CXR images and corresponding progression labels (i.e., "stable," "improved," or "worsened"). Additionally, we obtain the symptom labels (referred to as "Findings") from the MIMIC-CXR dataset [9] for each extracted CXR image in the Chest ImaGenome. Therefore, unlike [15] that focuses on the cross-image comparison of nine symptoms, our study concentrates on eight unique symptoms to align with the available symptom labels in the MIMIC-CXR dataset. Specifically, the symptoms of each CXR image are labeled by CheXpert labeler [8] with its status (i.e. Positive, Negative, Uncertain and Missing). We convert the Positive and Uncertain as 1, the Negative as 0 and ignore the Missing labeled symptom during training.

In general, the dataset comprises 37,695 current-prior CXR image pairs, covering 8 symptoms with a progression label per symptom. In detail, the data is divided into training (70%), validation (10%), and testing (20%) sets. Performance evaluation of the progression prediction is conducted by selecting the best model on the validation set and applying it to the test set, utilizing metrics including Precision, Recall, and F1-score.

3.3 Experimental Results

Comparison with State-of-the-Art Method. Table 1 presents the performance of overall progression prediction using our method and recent baseline (CheXRelFormer). As the results show, our proposed method demonstrates a remarkable performance surpassing the CheXRelFormer across all three progression categories. This surpass is evident in multiple evaluation metrics, including Precision, Recall, and F1-score. Specifically, our method achieves 11.8%,

Table 1: Performance of overall progression prediction.

Methods	**Metrics**	**Symptom Progression**			**Mean**
		Worsened	**Stable**	**Improved**	
CheXRelFormer [MICCAI23]	Precision	0.565	0.433	0.573	0.525
	Recall	0.564	0.431	0.577	0.526
	F1-score	0.525	0.432	0.575	0.526
Ours	Precision	**0.607**	**0.697**	**0.628**	**0.643**
	Recall	**0.686**	**0.506**	**0.708**	**0.636**
	F1-score	**0.644**	**0.587**	**0.665**	**0.633**

Table 2: Performance of symptom progression prediction.

Symptoms	CheXRelFormer[MICCAI23]			Ours		
	Mean Precision	**Mean Recall**	**Mean F1-score**	**Mean Precision**	**Mean Recall**	**Mean F1-score**
Lung Opacity	0.508	0.500	**0.500**	**0.521**	**0.544**	0.483
Pleural Effusion	**0.574**	0.552	0.561	0.569	**0.612**	**0.574**
Atelectasis	0.802	0.481	0.601	**0.852**	**0.925**	**0.890**
Cardiomegaly	0.656	0.508	0.544	**0.683**	**0.727**	**0.670**
Edema	**0.674**	0.542	0.587	0.642	**0.720**	**0.678**
Pneumothorax	0.420	0.402	**0.408**	**0.470**	**0.485**	0.321
Consolidation	**0.817**	0.529	0.527	0.649	**0.666**	**0.656**
Pneumonia	**0.669**	0.582	0.611	0.602	**0.680**	**0.630**

11% and 10.7% performance gains for average Precision, Recall and F1-score, respectively. Furthermore, our method demonstrates a notable performance improvement in the "stable" category, achieving remarkable gains of over 26.4% and 15.5% in Precision and F1-score, respectively. These results strongly indicate that our approach can effectively capture the progression information and exhibit a clearer understanding of the concept of "symptom changes".

Table 2 presents a comprehensive analysis of the progression prediction performance from the symptom's perspective. The results show that our method consistently surpasses the baseline in mean Recall and mean F1-score. This observation highlights the consistently high performance of our method across different symptoms, suggesting that our method has the potential to be a valuable tool for detecting changes in CXR associated with diverse progression dynamics.

Ablation study. In this section, we conduct experiments to investigate the effectiveness of Symptom Disentangle and the symptom classification task by two variants: (1) *w/o Symptom Disentangler* directly concatenates image-level features for progression prediction. (2) *w/o Symptom Classification* excludes the additional symptom label set during training. As Table 3 shows, when conducting progression learning solely based on image-level features, the outcomes display significantly inferior performance compared to utilizing symptom-level information. This emphasizes the importance of Symptom Disentangler in assisting the SPL model in acquiring more fine-grained progression information, thereby enhancing prediction accuracy. Besides, by incorporating additional symptom classification tasks, the performance of the progression prediction experiences a substantial improvement, with the mean F1-score rising from 56.2% to 63.3%. This enhancement can be attributed to the effective enhancement of the SPL module's feature extraction capability through the symptom classification task, enabling the identification of more subtle changes in symptom level.

Age-wise Disease Progression Analysis. We evaluate the performance of our approach across different age groups, as shown in Figure 3(a), and compare

Table 3: Ablation study on Symptom Disentangler and Symptom Classification task.

Methods	Metrics	Disease Progression			Mean
		Worsened	Stable	Improved	
w/o Symptom Disentangler	F1-score	0.558	0.428	0.575	0.522
w/o Symptom Classification	F1-score	0.519	**0.592**	0.576	0.562
Ours	F1-score	**0.644**	0.587	**0.665**	**0.633**

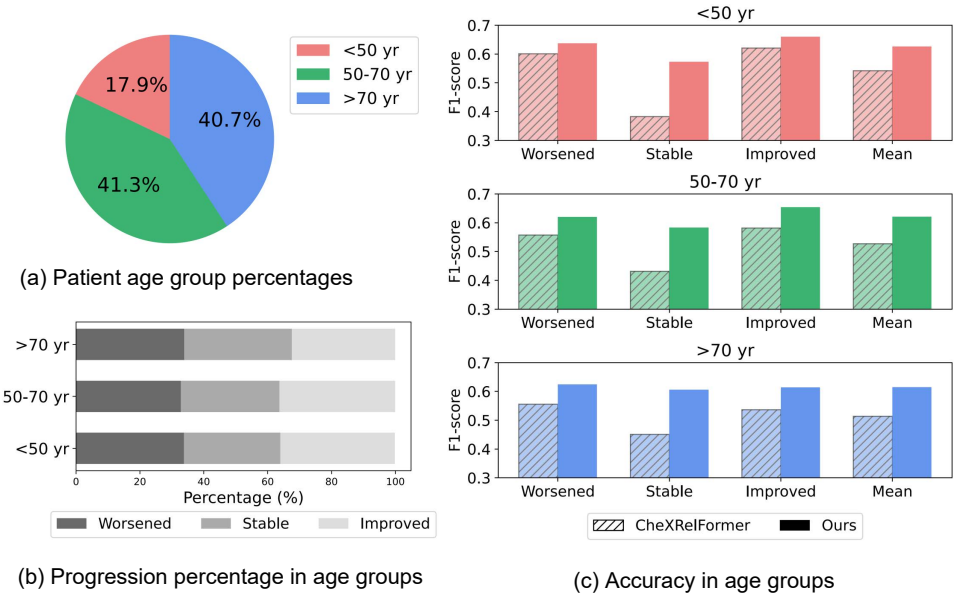

Fig. 3: Disease progression monitoring in different age groups.

it with the CheXRelFormer. Three age groups are categorized: < 50, $50 - 70$, > 70 years. Figure 3(b) shows the diverse percentages of three disease progression categories for different age groups. Based on the observations derived from Figure 3(c), it becomes evident that our method showcases a stable performance in all age groups, consistently outperforming the baseline by a margin, especially in the categories "stable".

4 Conclusion

In this paper, we propose a symptom-level progression learning framework for analyzing the changes within pairs of current-prior CXR images. To achieve this, we introduce a novel Symptom Disentangler and a Symptom Progression Learner. The former module first explicitly locates and extracts symptom-specific features, and the latter module utilizes the extracted symptom features to capture

the subtle changes for accurate progression prediction. Experimental results on the Chest ImaGenome dataset demonstrate the effectiveness of our proposed method. Future work may explore leveraging the complex progression information of multiple symptoms for precise outcome prediction.

Acknowledgments. This work is supported by Hong Kong Research Grants Council General Research Fund under Grant RGC/HKBU12200122.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ait Nasser, A., Akhloufi, M.A.: A review of recent advances in deep learning models for chest disease detection using radiography. Diagnostics **13**(1), 159 (2023)
2. Bannur, S., Hyland, S., Liu, Q., Perez-Garcia, F., Ilse, M., Castro, D.C., Boecking, B., Sharma, H., Bouzid, K., Thieme, A., et al.: Learning to exploit temporal structure for biomedical vision-language processing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15016–15027 (2023)
3. Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K.G., Murphy, K.: Deep learning for chest x-ray analysis: A survey. Medical Image Analysis **72**, 102125 (2021)
4. Dalla Serra, F., Wang, C., Deligianni, F., Dalton, J., O'Neil, A.Q.: Controllable chest x-ray report generation from longitudinal representations. In: The 2023 Conference on Empirical Methods in Natural Language Processing (2023)
5. Emre, T., Chakravarty, A., Rivail, A., Riedl, S., Schmidt-Erfurth, U., Bogunović, H.: Tinc: Temporally informed non-contrastive learning for disease progression modeling in retinal oct volumes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 625–634. Springer (2022)
6. Hou, B., Kaissis, G., Summers, R.M., Kainz, B.: Ratchet: Medical transformer for chest x-ray diagnosis and reporting. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24. pp. 293–303. Springer (2021)
7. Huang, C., Xu, Z., Shen, Z., Luo, T., Li, T., Nissman, D., Nelson, A., Golightly, Y., Niethammer, M., Zhu, H.: Dadp: dynamic abnormality detection and progression for longitudinal knee magnetic resonance images from the osteoarthritis initiative. Medical image analysis **77**, 102343 (2022)
8. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Thirty-Third AAAI Conference on Artificial Intelligence (2019)
9. Johnson, A.E., Pollard, T.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., Peng, Y., Lu, Z., Mark, R.G., Berkowitz, S.J., Horng, S.: Mimic-cxr-jpg, a large publicly available database of labeled chest radiographs. arXiv preprint arXiv:1901.07042 (2019)
10. Karwande, G., Mbakwe, A.B., Wu, J.T., Celi, L.A., Moradi, M., Lourentzou, I.: Chexrelnet: An anatomy-aware model for tracking longitudinal relationships between chest x-rays. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 581–591. Springer (2022)

11. Li, M.D., Chang, K., Bearce, B., Chang, C.Y., Huang, A.J., Campbell, J.P., Brown, J.M., Singh, P., Hoebel, K.V., Erdoğmuş, D., et al.: Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging. NPJ digital medicine **3**(1), 48 (2020)
12. Liang, W., Zhang, K., Cao, P., Zhao, P., Liu, X., Yang, J., Zaiane, O.R.: Modeling alzheimers' disease progression from multi-task and self-supervised learning perspective with brain networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 310–319. Springer (2023)
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
14. Luo, Y., Shi, M., Tian, Y., Elze, T., Wang, M.: Harvard glaucoma detection and progression: A multimodal multitask dataset and generalization-reinforced semi-supervised learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 20471–20482 (2023)
15. Mbakwe, A.B., Wang, L., Moradi, M., Lourentzou, I.: Hierarchical vision transformers for disease progression detection in chest x-ray images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 685–695. Springer (2023)
16. Müller, P., Meissen, F., Brandt, J., Kaissis, G., Rueckert, D.: Anatomy-driven pathology detection on chest x-rays. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 57–66. Springer (2023)
17. Oh, D.Y., Kim, J., Lee, K.J.: Longitudinal change detection on chest x-rays using geometric correlation maps. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. pp. 748–756. Springer (2019)
18. Rousan, L.A., Elobeid, E., Karrar, M., Khader, Y.: Chest x-ray findings and temporal lung changes in patients with covid-19 pneumonia. BMC Pulmonary Medicine **20**(1), 1–9 (2020)
19. Signoroni, A., Savardi, M., Benini, S., Adami, N., Leonardi, R., Gibellini, P., Vaccher, F., Ravanelli, M., Borghesi, A., Maroldi, R., et al.: Bs-net: Learning covid-19 pneumonia severity on a large chest x-ray dataset. Medical Image Analysis **71**, 102046 (2021)
20. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 568–578 (2021)
21. Wu, B., Ren, S., Li, J., Sun, X., Li, S.M., Wang, Y.: Forecasting irreversible disease via progression learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8117–8125 (2021)
22. Zhang, L., Wu, J., Wang, L., Wang, L., Steffens, D.C., Qiu, S., Potter, G.G., Liu, M.: Brain anatomy-guided mri analysis for assessing clinical progression of cognitive impairment with structural mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 109–119. Springer (2023)