Temporal Context Matters: Enhancing Single Image Prediction with Disease Progression Representations — Supplementary Material —

Aishik Konwer[1], Xuan Xu[1], Joseph Bae[2], Chao Chen[2], Prateek Prasanna[2]
[1]Department of Computer Science, Stony Brook University
[2]Department of Biomedical Informatics, Stony Brook University
{akonwer, xuaxu}@cs.stonybrook.edu {joseph.bae, Chao.Chen.1, Prateek.Prasanna}@stonybrook.edu

In this supplementary material, we provide additional information to further understand our proposed approach. In section 1, we provide an architectural overview of how to calculate a 512 dimension vector from last the layer of Temporal Convolutional Network (TCN). We provide dataset and preprocessing details in section 2. Finally in section 3, t-SNE plots and additional class activation maps provide insights into OA severity prediction from knee radiographs.

1. Architecture details - TCN output

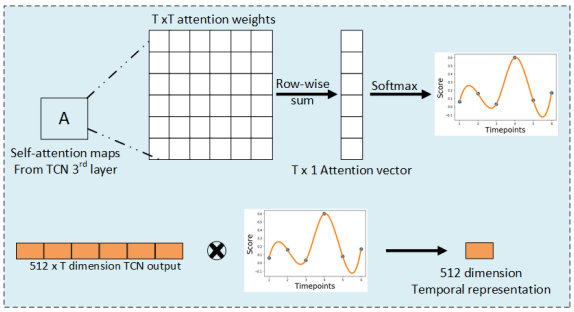

Figure 1. Derivation of temporal representation from last layer of TCN.

A row-wise summation operation is applied on the self-attention weights obtained from the third and final self-attention block in our TCN architecture. This results in a 'T' length attention vector, the softmax of which gives the attention scores. When these attention scores are multiplied with the output of TCN, they generate an optimal 512 dimension representation. The steps are illustrated in Supplementary Fig 1

2. Assets and preprocessing

Chest Radiograph dataset: For snapshot pretraining, we used 28433 chest radiographs (comprising multiple pulmonary diseases). CovidProg dataset, which contained 942 scans from 150 COVID-19 patients, comprised the temporal data. The duration between the CXR scans are variable (1-5 days). The number of timepoints per patient varies from 4 to 16. Out of the total 150 patients, 23 cases were obtained from Newark Beth Israel Medical Center, 77 from Stony Brook University Hospital, and 50 from University Hospitals Cleveland Medical Center.

OA Radiograph dataset: For snapshot pretraining, we used 23008 images. 2474 knee scans from 426 patients comprised the temporal data. The images in the 'train' folder of Kaggle [2] were a fraction of the snapshot cohort used in pretraining the transformer. All the images in 'validation' and 'test' folder [2] were jointly used in finetuning stage.

The experiments were performed in a 5-fold cross validation setting in the finetuning stage where the pretrained transformer model was finetuned on 4 folds and tested on the remaining fold. Details about the data used for each stage can be found in Supplementary Table 1

Stage	COVID	OA
Snapshot	21165 [3] + 7268 [9]	17230 [6] + 5778 train folder [2]
Temporal	942 (CovidProg)	2474 [6]
Finetune	631 (Vent.), 531 (Mort.) [7]	2482 val + test folder [6]

Table 1. Data utilized in different stages

COVID-19 preprocessing: Lung region segmentation was first performed using a Residual UNet model [10]. All chest scans were aligned to the same intensity range through an average histogram matching method.

OA preprocessing: We utilized BoneFinder tool [5] to localize and crop the knee joint landmarks. Following [8], histogram clipping and global contrast normalization were applied to each localized knee joint image.

Samples of COVID and OA images after pre-processing are shown in Supplementary Fig 2 and Supplementary Fig 3, respectively.

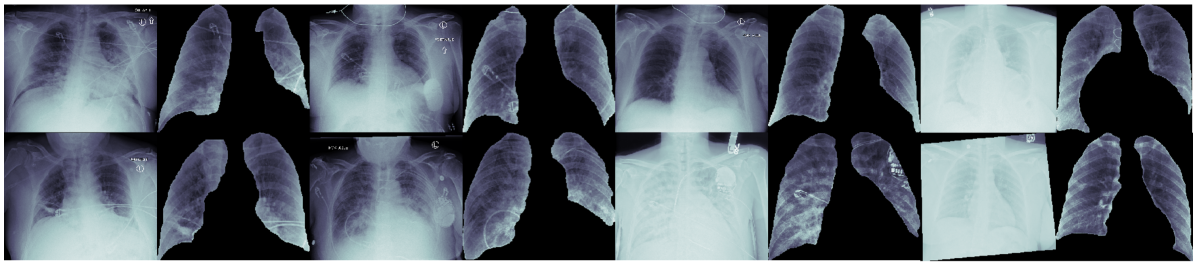

Figure 2. Preprocessed chest scans after applying average histogram matching and lung segmentation

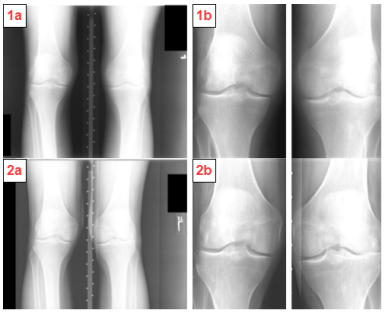

Figure 3. Preprocessed knee radiographs (1b, 2b) generated after joint localization [5] and global contrast normalization [8] on original samples (1a, 2a)

3. Insights from OA severity prediction

Supplementary Fig 4 demonstrates that utilizing temporal representations in our architecture results in better defined clusters for the three severity grades $(0, 1, 2)$ on the t-SNE plot. It may be observed that intermediate grades, such as 1 vs 2, which are more difficult to predict (left) can benefit from the proposed temporal approach (right). Additional CAMs of OA affected knees are compared in Supplementary Fig 5. Each row corresponds to knee radiographs from different severity grades, from 0 to 4. As may be observed, the attention maps from DeepKnee [8] , CNN + Ordinal loss [1] and SE block [4] are very sparse and sometimes react to unnecessary areas (bone texture, joint centre). On the contrary, our method provides more focused attention on the osteophytes and joint narrowing - the two important indicators of osteoarthritis.

4. Recalibration using matching data

In the COVID-19 cohort, we included some matched data. 100 of 150 patients in the CovidProg temporal dataset also have their ventilation status known. We use the 100 patients, take the baseline scans (the first image) of their temporal sequences as matched snapshot images. We evaluated the distance between these matched temporal/snapshot data in the representation space through training. In Supplementary Fig 6, Curve A (blue) shows the average distance (d) between the matched pair of snapshot and temporal representation. d is reduced to only 0.10 after 40 epochs. For reference, we also show d between any snapshot of positive ventilation status (S+) and any temporal sequence of positive ventilation status (Tm+). The result is Curve B (orange). Meanwhile, we also report d between S+ and any temporal sequence with negative ventilation status (Tm-) as in Curve C (green). After 40 epochs, d in A,B,C are 0.10, 0.67 and 4.26, respectively. C>>B>A shows that (1) the matched snapshots and temporal sequences are automatically aligned very well during training, thanks to the MMD loss; (2) generally a positive temporal sequence is aligned much closer to a positive snapshot than a negative snapshot, although not as close as the matched pairs.

5. Longitudinal comparison

We also compared our method with other temporal models, namely CNN+LSTM, CNN+biLSTM and CNN+biLSTM+Attention. It may be observed from Supplementary Table 2 that our approach outperforms all these longitudinal models.

Name	Ventilation
Method	AUC
CNN+LSTM	0.82
CNN+biLSTM	0.83
CNN+biLSTM+Attention	0.85
Ours	**0.88**

Table 2. Comparison with longitudinal methods

6. Limitations

In our temporal analysis, the images are not registered. Registration might result in learning better representations. We aim to address this in the future by using a spatial transformer network as a pre-processing stage before extracting temporal features. Also, due to lack of sufficient temporal data, we did not use transformer architectures to learn disease progression. This research direction can be pursued

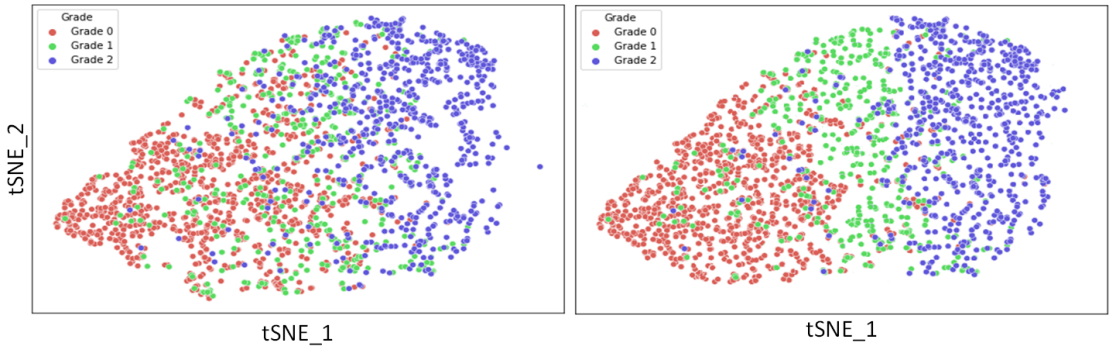

Figure 4. Comparison between t-SNE plots before and after using temporal modeling for severity grades (0,1,2)

with the availability of more temporal cases in future.

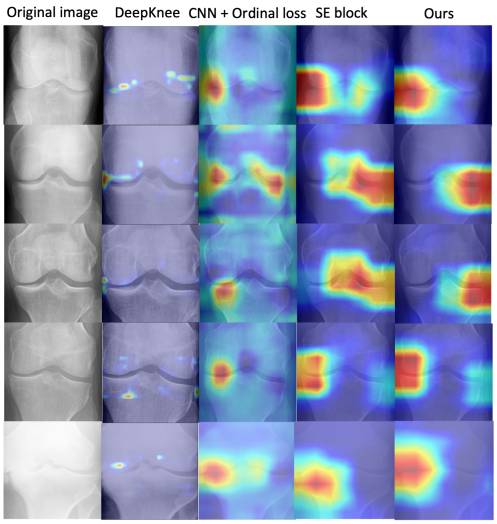

Figure 5. Qualitative comparisons of knee CAMs depicting OA severity grades 0 to 4 (top to bottom)

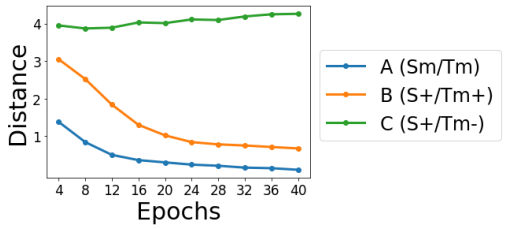

Figure 6. Distance between feature means across training

References

[1] Pingjun Chen, Linlin Gao, Xiaoshuang Shi, Kyle Allen, and Lin Yang. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. *Computerized Medical Imaging and Graphics*, 75:84–92, 2019. 2

[2] Pingjun Chen (2018). Knee osteoarthritis severity grading dataset. Mendeley Data, V1, doi: 10.17632/56rmx5bjcr.1. 1

[3] Muhammad EH Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandakar Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al Emadi, et al. Can ai help in screening viral and covid-19 pneumonia? *IEEE Access*, 8:132665–132676, 2020. 1

[4] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 2

[5] Claudia Lindner, Shankhar Thiagarajah, J Mark Wilkinson, Gillian A Wallis, Timothy F Cootes, arcOGEN Consortium, et al. Fully automatic segmentation of the proximal femur using random forest regression voting. *IEEE transactions on medical imaging*, 32(8):1462–1472, 2013. 1, 2

[6] OAI. Osteoarthritis initiative. https://nda.nih.gov/oai/. 1

[7] Joel Saltz, Mary Saltz, Prateek Prasanna, Richard Moffitt, Janos Hajagos, Erich Bremer, Joseph Balsamo, and Tahsin Kurc. Stony brook university covid-19 positive cases [data set]. the cancer imaging archive. https://doi.org/10.7937/TCIA.BBAG-2923. 1

[8] Aleksei Tiulpin, Jérôme Thevenot, Esa Rahtu, Petri Lehenkari, and Simo Saarakkala. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. *Scientific reports*, 8(1):1–10, 2018. 1, 2

[9] Maria de la Iglesia Vayá, Jose Manuel Saborit, Joaquim Angel Montell, Antonio Pertusa, Aurelia Bustos, Miguel Cazorla, Joaquin Galant, Xavier Barber, Domingo Orozco-Beltrán, Francisco García-García, et al. Bimcv covid-19+: a large annotated dataset of rx and ct images from covid-19 patients. *arXiv preprint arXiv:2006.01174*, 2020. 1

[10] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018. 1