CLIMAT: CLINICALLY-INSPIRED MULTI-AGENT TRANSFORMERS FOR KNEE OSTEOARTHRITIS TRAJECTORY FORECASTING

Huy Hoang Nguyen\textsuperscript{1}, Simo Saarakkala\textsuperscript{1,2}, Matthew B. Blaschko\textsuperscript{3}, Aleksei Tiulpin\textsuperscript{1}

\textsuperscript{1}University of Oulu, Finland, \textsuperscript{2}Oulu University Hospital, Finland, \textsuperscript{3}KU Leuven, Belgium

ABSTRACT

In medical applications, deep learning methods are designed to automate diagnostic tasks. However, a clinically relevant question that practitioners usually face, is how to predict the future trajectory of a disease (prognosis). Current methods for such a problem often require domain knowledge, and are complicated to apply. In this paper, we formulate the prognosis prediction problem as a one-to-many forecasting problem from multimodal data. Inspired by a clinical decision-making process with two agents – a radiologist and a general practitioner, we model a prognosis prediction problem with two transformer-based components that share information between each other. The first block in this model aims to analyze the imaging data, and the second block leverages the internal representations of the first one as inputs, also fusing them with auxiliary patient data. We show the effectiveness of our method in predicting the development of structural knee osteoarthritis changes over time. Our results show that the proposed method outperforms the state-of-the-art baselines in terms of various performance metrics. In addition, we empirically show that the existence of the multi-agent transformers with depths of 2 is sufficient to achieve good performances. Our code is publicly available at \url{https://github.com/MIPT-Oulu/CLIMAT}.

Index Terms— Deep learning, osteoarthritis, prognosis, trajectory forecasting, transformer

1. INTRODUCTION

Clinical diagnosis is made by a treating physician or a general practitioner. These specialists are not radiologists and use their services in decision-making. One of the typical problems that such doctors face, is to make an accurate estimation of the disease trajectory (prognosis) based on patient data, findings from imaging, and auxiliary information, such as blood tests. This is an especially relevant task in the case of degenerative disorders. This paper tackles prognosis prediction in knee osteoarthritis (OA) – the most common musculoskeletal disorder [1].

Among all the joints in the body, OA is mostly prevalent in the knee [2]. OA is characterized by the breakdown of cartilage tissue, the appearance of osteophytes, and the narrowing of joint space [2], which are imaged using X-ray (radiography). The disease severity is graded according to the Kellgren-Lawrence system [3] from 0 (no OA) to 4 (end stage OA) as shown in Suppl. Figure S1. Unfortunately, OA progresses over time (depicted in Figure 1) and no cure has yet been developed for OA. However, prediction of disease evolution at an early stage may enable slowing it down, for example using behavioral interventions [4].

Literature shows that there is a lack of studies on prognosis prediction. From an ML perspective, a more conventional setup is to predict whether the patient has the disease [4, 5, 6]. However, prognosis prediction aims to answer whether and how the disease would evolve over time. Furthermore, in a real life situation, the treating physician makes the prognosis while interacting with a radiologist or other stakeholders who can provide information (e.g. blood tests or radiology reports) about the patient’s condition [7]. We believe that informing prediction model design with this prior knowledge is valuable, and may provide performance benefits.

In this paper, we propose a Clinically-Inspired Multi-Agent Transformers (CLIMAT) framework, which aims to mimic the interaction process between a general practitioner or treating physician and a radiologist. The core novel idea leading to our model design is that a radiologist first analyzes the image, and provides a radiology report to the doctor who makes the prognosis, also taking into account additional data. In our system, a radiologist module, consisting of a feature extractor (convolutional neural network; CNN) and a transformer, analyses the input imaging data and extracts feature vectors per every image superpixel. Subsequently, the

Fig. 1: Radiographs of a patient with OA progressed in 8 years. Red arrow indicates joint space narrowing. The disease progressed from Kellgren-Lawrence (KL) grade 0 at the baseline (BL) to 3 in 6 years. At the 8th year, the patient underwent a total knee replacement (TKR) surgery.
Fig. 2: The architecture of CLIMAT consists of three transformers. The transformer D as a radiologist performs a diagnosis for the current stage $y_0$ of a disease from visual features. The combination of the transformers F and P, mimicking a general practitioner, aims to forecast future stages $\hat{y}_{1:T}$ of the disease based on the output states $v_0$ of the transformer D and auxiliary data.

set of superpixels with positional encodings is passed to a transformer that aims to predict the current disease severity stage, characterized by imaging findings. The states of this transformer are fused with auxiliary patient clinical data, and passed to a general practitioner-corresponding transformer module, predicting the disease’s severity trajectory. To summarize, our contributions are the following:

1. We propose CLIMAT, a clinically-inspired transformer-based framework that can learn to forecast disease severity from multimodal data in an end-to-end manner.
2. From a clinical perspective, to our knowledge, we show the first study on predicting a fine-grained prognosis of knee OA directly from raw imaging data and clinical variables.
3. We empirically demonstrate superior performance of our method compared to the state-of-the-art baselines.

2. METHOD

2.1. Overview

We model multi-agent decision-making as follows. A radiologist analyzes a medical image (e.g. a radiograph) of a patient to provide an interpretation with rich visual description and annotations, allowing the diagnosis of the current stage of the disease. Subsequently, the general practitioner relies on the clinical data (e.g. questionnaires or symptomatic assessments), and the provided radiologic interpretation to make a further interpretation if needed to predict the course of the disease in the future.

In Figure 2, we present the workflow of CLIMAT comprising three transformers – namely D, F, and P, which are the abbreviations for Diagnosis, Fusion, and Prognosis respectively. Specifically, the transformer D acts as the radiologist to perform visual reasoning from imaging data and predict the current stage $y_0$ of the knee OA disease. The other two transformers are responsible for data fusion and forecasting. As such, the transformer F aims to extract a context embedding from clinical variables. Subsequently, the transformer P utilizes the combination of the context embedding and the output states of the transformer D to forecast the disease trajectory $\hat{y}_{1:T}$.

2.2. Multi-output-head transformer

A transformer encoder comprises a stack of $L$ multi-head self-attention layers, whose input is a sequence of vectors $\{s_l\}_{l=1}^{N}$ where $s_l \in \mathbb{R}^{1 \times C}$, and $C$ is the feature size. We define a transformer with regard to the number of output heads. As such, a transformer with $K$ output heads ($K \geq 1$) is formulated as

$$h_0 = [E_{[CLS \ 0]}, \ldots, E_{[CLS \ K-1]}, s_1, \ldots, s_N] + E_{[POS]}.$$  

$$z_{l-1} = \text{MSA}(LN(h_{l-1})) + h_{l-1},$$  

$$h_l = \text{MLP}(LN(z_{l-1})) + z_{l-1}, \quad l = \{1, \ldots, L\}$$

where $E_{[CLS \ k]} \in \mathbb{R}^{1 \times C}$ is a learnable token with $k = 0 \ldots K - 1$, and $E_{[POS]} \in \mathbb{R}^{(N+K) \times C}$ is a learnable positional embedding. MLP is a multi-layer perceptron (i.e. a fully-connected network), LN is a layer normalization [8], and MSA(·) is a multi-head self-attention layer [9]. We take the first $T$ representations in the last layer to perform multi-task predictions via non-linear layers. In general, $K$ is chosen such that $T \leq K + N$. We typically set $K$ to 1 or $T$.

2.3. CLIMAT for knee OA trajectory prediction

We firstly extract $H \times W \times C$ visual representations of the input radiograph using a stack of convolutional blocks, and then reshape them to $N \times C$, where $N = HW$. Having in mind the idea of visual reasoning, we treat the reshaped representations as an $N$-length sequence of $1 \times C$ vectors, and pass them through a transformer, which predicts $y_0$. As we convert the image classification problem into a sequence classification one, we include two common ingredients: a sequence start vector, denoted by $E_{[CLS]}^D$, and also the positional embeddings for every super-pixel [9, 10]. Both of these are learnable vectors, and positional embeddings are added to the superpixels of an image. Once the input sequence of superpixels is passed through the transformer D, we take its first element and pass it through a fully connected network, similar to [10].

As the module for predicting the prognosis can utilize other auxiliary modalities, we acquire another transformer –
named F to fuse them. Firstly, we project them into the same $C_0$-dimensional feature space using separate feature extractors $\{FE\}_{m=1}^M$ in Figure 2, consisting of a fully connected layer, a ReLU activation, and a layer normalization [8]. Similar to the transformer D, we include an initial embedding $E^D[C_{CLS}]$ and a positional embedding to derive the input for the transformer F. Finally, we select the first vector $h^F_L[0]$ in the last layer of the transformer F as a context token representing all the modalities.

As soon as the context token $h^F_L[0]$ of length $C_0$ is acquired from the context network, we concatenate $N + 1$ copies of the token into the last states $h^D_L$ of the transformer D. The prognosis transformer (or transformer P) has $K$ embeddings $E^P[C_{CLS}]$. Thus, its input sequence has a length of $K + N + 1 \geq T$. To predict the prognosis $y_1, \ldots, y_T$, we pass the first $T$ elements of the last layer of the transformer P through $T$ distinct feed-forward networks (FFNs), each of which comprises a layer normalization followed by two fully connected layers separated by a GELU activation [11].

2.4. Multi-task learning with missing targets

In practice, each patient commonly has missing annotations throughout follow-up visits. We can handle such an impaired condition with ease by introducing an indicator function to mask out missing targets. Formally, we minimize the following loss

$$L = \sum_{i \in I} \frac{1}{\sum_{t=0}^T \|i\|_t} \sum_{t=0}^T w_i \|i\|_t \ell(f_t(x^i), y^i_t), \quad \text{(4)}$$

where $I$ is the set of sample indices, $(x_i, y_i)$ is a labeled sample, $f$ is our model, $w_i$ is the weight of task $t$, $f_t$ is the output at task $t$, $\|i\|_t$ is an indicator function of sample $i$ at task $t$, and $\ell$ is a cross-entropy loss.

3. EXPERIMENTS

3.1. Data

We conducted experiments on the Osteoarthritis Initiative (OAI), is publicly available at https://nda.nih.gov/oai/. 4,796 participants from 45 to 79 years old participated in the OAI cohort, which consisted of a baseline, and follow-up visits up to 132 months. In the present study, we used all knee images that (i) were annotated for KL grade, (ii) did not include implants, and (iii) were acquired with large imaging cohorts: the baseline, and the 12, 24, 36, 48, 72, and 96-month follow-ups (presented in Suppl. Table S4). We followed [4] to extract two knees regions of interest from each bilateral radiograph and pre-process each of them. Subsequently, each pre-processed knee image was resized to $256 \times 256$. Additionally, we utilized age, sex, body mass index (BMI), history injury, surgery, and total Western Ontario and McMaster Universities Arthritis Index (WOMAC) as clinical variables (Suppl. Table S1).

The OAI dataset includes data from five acquisition centers, which allowed us to utilize the one-center-out cross-validation procedure: data from 4 centers were used for training and validation, and data from the left-out one for testing. We trained and evaluated at 6 major time points: the baseline, and 1, 2, 3, 4, 6, and 8 years in the future. For each training set from a group of 4 centers, we performed a 5-fold cross-validation strategy (Suppl. Table S5).

3.2. Experimental Setup

We trained and evaluated our method and the reference approaches using V100 NVidia GPUs. We implemented all the methods using the Pytorch library, and trained each of them with the same set of configurations. For each problem, we used the Adam optimizer with a learning rate of $1e-4$ and without any weight decay. The list of augmentations is presented in Suppl. Table S2.

To extract visual representations of 2D images, we utilized the convolutional blocks of the ResNet18 network [12] pretrained on the ImageNet dataset. We used only 1 [CLS] learnable token ($K=1$) in the transformer P. We used batch size of 128 for the knee OA experiments. For each scalar numerical or categorical input, we used a common feature extraction architecture with a linear layer, a ReLU activation, and the layer normalization [8]. Our baselines were models that had the same feature extraction modules for multimodal data, but utilized different architectures to perform discrete time series forecasting. As such, we compared our method to baselines with the forecasting module using fully-connected network (FCN), GRU [13],
that increasing it 2 and 4 times worsens the performance of prognosis at the early years ($t \leq 4$), but improves it at later years ($t \geq 6$). However, on average, we obtained BAs of 43.47, 43.43, and 43.37 for the depth of 2, 4, and 8, respectively. Thus, the shallow version of the transformer P with only 2 layers achieved the highest BA. Together with Table 1, it indicates that the existence of the modular structure of transformers matters more than the depth of the transformer P.

3.5. Interpretability

Leveraging the modular structure of CLIMAT, we were able to generate groups of attention maps for different input categories. In Figure 4, we present examples of attention maps over a healthy knee X-ray image and the corresponding clinical variables that were extracted from the transformer P and the transformer F, respectively. Figure 4a shows the averages of 4 self-attention maps. Here, we observe that the final prediction was made based on the changes in the intercondylar notch [16], as well as the symptomatic evaluation of the patient. Figure 4b indicates that WOMAC, sex, and history of injury were the top-3 impactful clinical variables according to the transformer F for the future knee OA progression of that particular patient. We present more prediction samples in Suppl. Figure S2.

4. CONCLUSIONS

In this paper, we proposed a novel transformer-based method to forecast a trajectory of a disease’s stage from multimodal data. We applied our method to knee osteoarthritis prognosis prediction problem, and to our knowledge, this is the first study in the realm of OA that tackled this problem. The developed method can be of interest to other fields, where a forecasting of disease course is of interest. The main limitation of this study is that our experiments were conducted on a dataset conducted in the research setting. Evaluation of performance of the method in real clinical setting is still needed to understand the value of the method on the OA treatment process.
5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by Osteoarthritis Initiative (https://nda.nih.gov/oai). A new ethical approval was not required, as the ethical approval and informed consent of the patients were obtained by OAI and published under open access permission group.

Acknowledgments

The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health.

The authors wish to acknowledge CSC – IT Center for Science, Finland, for generous computational resources. We would like to acknowledge the strategic funding of the University of Oulu, Sigrid Juselius Foundation, Finland.

Dr. Claudia Lindner is acknowledged for providing BoneFinder. Phuoc Dat Nguyen is acknowledged for discussions about transformer.

6. REFERENCES

[1] Sion Glyn-Jones, AJR Palmer, R Agricola, AJ Price, TL Vincent, H Weinans, and AJ Carr, “Osteoarthritis,” The Lancet, vol. 386, no. 9991, pp. 376–387, 2015.

[2] Behzad Heidari, “Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part i,” Caspian journal of internal medicine, vol. 2, no. 2, pp. 205, 2011.

[3] JH Kellgren and JS Lawrence, “Radiological assessment of osteo-arthritis,” Annals of the rheumatic diseases, vol. 16, no. 4, pp. 494, 1957.

[4] Aleksei Tiulpin, Stefan Klein, Sita MA Bierma-Zeinstra, Jérôme Thevenot, Esa Raathi, Joyce van Meurs, Edwin HG Oei, and Simo Saarakkala, “Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data,” Scientific reports, vol. 9, no. 1, pp. 1–11, 2019.

[5] Pawel Widera, Paco MJ Welsing, Christoph Ladel, John Loughlin, Floris PFJ Lafeber, Florence Petit Dop, Jonathan Larkin, Harrie Weinans, Ali Mobasher, and Jaume Bacardit, “Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data,” Scientific Reports, vol. 10, no. 1, pp. 1–15, 2020.

[6] B Guan, F Liu, A Haj-Mirzaian, S Demehri, A Samsonov, T Neogi, A Guermazi, and R Kijowski, “Deep learning risk assessment models for predicting progression of radiographic medial joint space loss over a 48-month follow-up period,” Osteoarthritis and cartilage, vol. 28, no. 4, pp. 428–437, 2020.

[7] LBO Jans, JML Bosmans, KL Verstraete, and R Achten, “Optimizing communication between the radiologist and the general practitioner,” JBR-BTR, vol. 96, no. 6, pp. 388–390, 2013.

[8] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton, “Layer normalization,” arXiv preprint arXiv:1607.06450, 2016.

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need,” in Advances in neural information processing systems, 2017, pp. 5998–6008.

[10] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, vol. 1, 2020.

[11] Dan Hendrycks and Kevin Gimpel, “Gaussian error linear units (GELUs),” 2020, arXiv:1606.08415.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[13] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” arXiv preprint arXiv:1409.1259, 2014.

[14] Sepp Hochreiter and Jürgen Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[15] Shi Hu, Egill Fridgeirsson, Guido van Wingen, and Max Wellin, “Transformer-based deep survival analysis,” in Survival Prediction-Algorithms, Challenges and Applications. PMLR, 2021, pp. 132–148.

[16] Heriberto Ojeda León, Carlos E Rodríguez Blanco, Todd B Guthrie, and Oscar J Nordelo Martínez, “Intercondylar notch stenosis in degenerative arthritis of the knee,” Arthroscopy: The Journal of Arthroscopic & Related Surgery, vol. 21, no. 3, pp. 294–302, 2005.
Supplementary materials

Table S1: Input variables for forecasting knee OA severity.

| Group               | Variable name | Data type |
|---------------------|---------------|-----------|
| Raw imaging         | Knee X-ray    | 2D        |
| Clinical Variables  | Age           | Numerical |
|                     | WOMAC         | Numerical |
|                     | Sex           | Categorical |
|                     | Injury        | Categorical |
|                     | Surgery       | Categorical |
|                     | BMI           | Numerical |

Table S2: An ordered list of common transformations. (✓) indicates transformations only used in the training phase.

| Transformation          | Prob. | Parameter |
|-------------------------|-------|-----------|
| Center cropping         | 1     | 700 × 700 |
| Resize                  | 1     | 280 × 280 |
| Gaussian noise (✓)      | 0.5   | 0.3       |
| Rotation (✓)            | 1     | [-10, 10] |
| Random cropping (✓)     | 1     | 256 × 256 |
| Center cropping         | 1     | 256 × 256 |
| Gamma correction (✓)    | 0.5   | [0.5, 1.5]|
| Z-score standardization | 1     |           |

Table S3: Common and specific hyperparameters for the methods.

| Key                        | Value         |
|----------------------------|---------------|
| **Common**                 |               |
| Raw image feature extractor| ResNet18      |
| Number of Conv2D blocks    | 5             |
| Feature length of scalar input | 128     |
| MLP hidden unit            | 256           |
| Dropout rate               | 0.3           |
| **CLIMAT**                 |               |
| Number of [CLS] tokens (K)| 1             |
| Depth of transformer D, F, P | 2            |
| MSA heads of D, F, P       | 4             |
Fig. S1: The Kellgren-Lawrence (KL) system is commonly used to assess the severity of OA. As such, the system classifies OA into 5 grades, which correspond to: no sign of OA, doubtful OA, early OA, moderate OA, and severe OA, respectively.

Table S4: Knee OA target statistics over the 6 primary visits of in the OAI cohort study.

| Visit    | KL 0 | KL 1 | KL 2 | KL 3 | KL 4 | TKR |
|----------|------|------|------|------|------|-----|
| Baseline | 3448 | 1597 | 2374 | 1239 | 295  | 61  |
| 12 months| 3113 | 1445 | 2221 | 1230 | 355  | 76  |
| 24 months| 2893 | 1348 | 2079 | 1172 | 367  | 97  |
| 36 months| 2735 | 1252 | 1986 | 1147 | 377  | 135 |
| 72 months| 1866 | 1007 | 471  | 201  | 26   | 9   |
| 96 months| 1899 | 987  | 488  | 239  | 47   | 15  |

Table S5: Data settings on the OAI dataset across 5 acquisition sites.

| Test site | Phase        | Baseline only | Years from baseline |
|-----------|--------------|---------------|---------------------|
|           |              |               | 1     | 2   | 3     | 4     | 5     | 6     |
| A         | Training/val. | Yes           | 7155  | 6706| 6418  | 6173  | 0     | 3077  | 0     | 3147  |
|           | Test         | Yes           | 1229  | 1195| 1162  | 1075  | 0     | 497   | 0     | 525   |
| B         | Training/val. | Yes           | 6572  | 6196| 5935  | 5671  | 0     | 2797  | 0     | 2851  |
|           | Test         | Yes           | 1812  | 1705| 1645  | 1577  | 0     | 777   | 0     | 821   |
| C         | Training/val. | Yes           | 5902  | 5490| 5289  | 5023  | 0     | 2442  | 0     | 2537  |
|           | Test         | Yes           | 2482  | 2411| 2291  | 2225  | 0     | 1132  | 0     | 1135  |
| D         | Training/val. | Yes           | 6274  | 5951| 5706  | 5459  | 0     | 2636  | 0     | 2698  |
|           | Test         | Yes           | 2110  | 1950| 1874  | 1789  | 0     | 938   | 0     | 974   |
| E         | Training/val. | Yes           | 7633  | 7261| 6972  | 6666  | 0     | 3344  | 0     | 3455  |
|           | Test         | Yes           | 751   | 640 | 608   | 582   | 0     | 230   | 0     | 217   |
Table S6: Detailed performances on the OAI dataset (average and standard errors over 5 random seeds).

| Year | Method | BA (%) ↑ | RMSE ↓ |
|------|--------|---------|--------|
| 1    | FCN    | 53.7±0.2 | 0.67±0.003 |
|      | GRU    | 52.2±0.2 | 0.67±0.003 |
|      | LSTM   | 52.4±0.3 | 0.68±0.002 |
|      | MMTF   | 52.8±0.1 | 0.67±0.003 |
|      | Ours   | **55.3±0.2** | **0.62±0.002** |
| 2    | FCN    | 51.5±0.1 | 0.70±0.002 |
|      | GRU    | 50.8±0.1 | 0.70±0.003 |
|      | LSTM   | 50.9±0.3 | 0.71±0.001 |
|      | MMTF   | 51.4±0.2 | 0.70±0.002 |
|      | Ours   | **53.7±0.2** | **0.64±0.002** |
| 3    | FCN    | 47.7±0.2 | 0.74±0.002 |
|      | GRU    | 48.0±0.2 | 0.76±0.004 |
|      | LSTM   | 47.9±0.1 | 0.76±0.001 |
|      | MMTF   | 47.8±0.2 | 0.75±0.002 |
|      | Ours   | **50.1±0.2** | **0.70±0.002** |
| 4    | FCN    | 44.8±0.2 | 0.78±0.002 |
|      | GRU    | 45.4±0.4 | 0.80±0.003 |
|      | LSTM   | 45.7±0.2 | 0.80±0.002 |
|      | MMTF   | 45.7±0.0 | 0.79±0.002 |
|      | Ours   | **47.5±0.1** | **0.74±0.003** |
| 6    | FCN    | 28.1±0.4 | 0.74±0.003 |
|      | GRU    | **29.2±0.3** | 0.80±0.005 |
|      | LSTM   | 29.0±0.2 | 0.80±0.005 |
|      | MMTF   | 27.4±0.3 | 0.80±0.005 |
|      | Ours   | 28.5±0.4 | **0.73±0.005** |
| 8    | FCN    | 25.9±0.2 | **0.80±0.002** |
|      | GRU    | **27.5±0.3** | 0.88±0.005 |
|      | LSTM   | 26.9±0.3 | 0.88±0.008 |
|      | MMTF   | 26.1±0.2 | 0.88±0.006 |
|      | Ours   | 27.1±0.2 | 0.81±0.002 |
Fig. S2: Selective samples of predictions done by CLIMAT from OAI. Picked imaging features corresponded to known clinical findings – joint space narrowing and osteophytes. Furthermore, other findings (known in the literature) such as the changes in the intercondylar notch and attrition were also picked by the model.