Learning Osteoarthritis Imaging Biomarkers from Bone Surface Spherical Encoding

Authors: Alejandro Morales Martinez, Francesco Caliva, Io Flament, Felix Liu, Jinhee Lee, Peng Cao, Sharmila Majumdar, Valentina Pedoia

Introduction: To learn bone shape features from spherical bone maps of knee MRI images using convolutional neural networks (CNN) and use these features to diagnose and predict osteoarthritis (OA).

Methods: A bone segmentation model was trained with a dataset of 40 manually segmented 3D MRI volumes to segment the Femur, Tibia, and Patella from 47,078 3D MRI volumes. Each of the segmented bone masks was converted to a 3D point cloud and rigidly registered to a reference point cloud to account for rotational variability at scan time. The registered point clouds were then transformed into spherical coordinates and four fusion strategies were performed to merge spherical maps obtained by each bone. The fusion strategies included the following variants: single-bone, early fusion, late fusion, and network ensemble. The Femur, Tibia, and Patella single-bone models consisted of three individual CNNs trained on each single knee bone spherical image. The early fusion model consisted of a CNN trained on the merged three-channel spherical maps of the Femur, Tibia, and Patella. The late fusion model consisted of the concatenation of the last three layers of each trained single-bone model merged into a fully connected layer and trained end to end. There were two network ensemble methods evaluated: majority voting, where the majority prediction from all three individual bone network for each patient was used, and logits averaging, where the average of the probabilities outputted by each of the three single-bone models was used for the prediction. A total of 41,822 bone spherical maps with corresponding Kellgren-Lawrence (KL) grades were used to train an OA diagnosis model for all four fusion strategies to diagnose radiographic OA exclusively using bone shape learned features. An OA incidence model was also trained for all four fusion strategies on a subset of the bone spherical maps to predict future radiographic OA incidence from a healthy MRI scan within a range of eight time points, from 1 year up to 8 years. The OA diagnosis models were pretrained using ImageNet and the OA incidence models were initialized with the learned weights from the corresponding OA diagnosis models. A validation performance comparison was performed for the OA diagnosis and incidence models with a receiver operating characteristic (ROC) curve and the best overall model variant for both tasks was selected.

Results: The logits averaging network ensemble method had the best overall validation performance for both the OA diagnosis and incidence tasks. The OA diagnosis logits averaging model had an area-under-the-curve (AUC) of 0.905 on the test set with a sensitivity and specificity of 0.815 and 0.839. The OA incidence logits averaging models had an AUC ranging from 0.841 to 0.646 on the test set for the range from 1 year to 8 years.

Conclusions: Bone shape was successfully used as a predictive imaging biomarker for OA. This approach is novel in the field of deep learning for musculoskeletal imaging and can be expanded to other OA imaging biomarkers such as cartilage thickness and T2-relaxation time values.