Knee Imaging Biomarkers for Sex and Race: A Deep Learning Analysis of The Osteoarthritis Initiative

Paul H. Yi, MD, Johns Hopkins Radiology; Kevin Z. Xin; Jinchi Wei, David Sing, Paul Tornetta III, Gregory D. Hager, PhD; Jan Fritz, MD

Introduction

Although osseous measurements differ between sexes and races, radiologists cannot reliably distinguish either on bone radiographs.

We evaluated the ability of deep convolutional neural networks (DCNNs) to 1) predict sex and race on knee radiographs, and 2) identify potential osseous imaging biomarkers for sex and race.

Hypothesis

Deep learning systems can distinguish between male and female subjects, as well as between white and black subjects on knee radiographs.

Methods

We obtained 9582 PA knee radiographs from the NIH Osteoarthritis Initiative (OAI) comprised of subjects ages 45-79 (mean 62 years) with osteoarthritis. Sex and race distribution were 58% female and 80% white.

Images were split into training/validation/test sets in 70/10/20% split and used to train, validate, and test ResNet-152 DCNNs to predict sex and race (white vs. non-white). External testing was performed on 150 knee radiographs from two USA centers from two separate regions (mid-atlantic & New England).

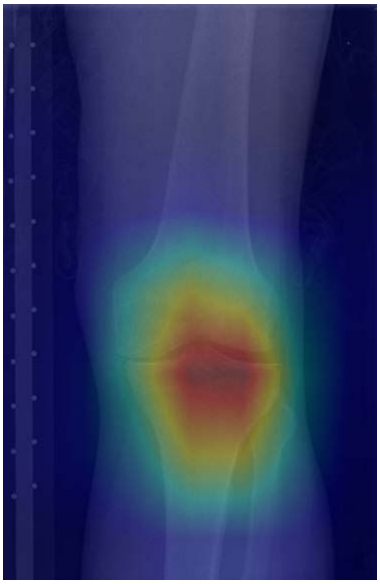
Area under the receiver operating characteristic curve (AUC) were used to evaluate DCNN performance. Class Activation Mapping heatmaps were used to visualize distinguishing features for image classification.

Results

On the OAI test set, the knee DCNNs achieved AUC of 0.99 for sex and 0.98 for race. On the external test set, DCNN performance was similar for sex, albeit decreased, with AUC of 0.91. The external testing for race showed decreased performance with AUC of 0.79.

CAM heatmaps for both the sex and race classifiers demonstrated emphasis of unique osseous regions, such as the central tibial plateau for sex prediction (**Figure 1**) and the medial femoral condyle & intercondylar region of the proximal femur for race prediction (**Figure 2**). The heatmaps were similar in appearance for both the OAI and external test sets.

Figure 1.



CAM demonstrates emphasis of central tibial plateau for correct prediction of male sex.

Figure 2.

CAM heatmap demonstrating emphasis of medial femoral condyle and intercondylar region for correct prediction of white race.

Conclusion
Deep learning can predict sex and race on bone radiographs, a task unable to be performed reliably by human radiologists, and which may aid in forensics investigations. Although there was decrease in performance on external test sets, heatmap analysis suggests discrete osseous imaging biomarkers for sex and race. Identification of such biomarkers could help improve understanding of skeletal pathoanatomy and treatment options, such as design of sex and race-specific implants.

Statement of Impact
The ability to glean demographic information from knee radiographs may aid forensic investigations, as well as help identify novel anatomic landmarks for sex and race, with potential implications for improved understanding of knee disease pathoanatomy and treatment options, such as design of gender and race-specific knee prostheses.

Keywords
deep learning, sex, race, bones, musculoskeletal