Knee Osteoarthritis Classification Using Bone Distance

Rahul Singh
Department of Computer Science, Seidenberg School of CSIS, Pace University, New York City, NY, USA
rs63891n@pace.edu

Saatvik Chaudhari
Department of Computer Science, Seidenberg School of CSIS, Pace University, New York City, NY, USA
sc78976n@pace.edu

Andrew Dinspechin
Department of Computer Science, Seidenberg School of CSIS, Pace University, New York City, NY, USA
ad42824n@pace.edu

Milind Kumar Choudhary
Department of Computer Science, Seidenberg School of CSIS, Pace University, New York City, NY, USA
mc95448n@pace.edu

***Abstract*— *Osteoarthritis (OA) is the most prevalent form of arthritis affecting the knee joint, with severity commonly assessed using the Kellgren-Lawrence (KL) scale. Traditional methods for evaluating OA severity rely on direct measurement of cartilage thickness, which is time-consuming and challenging in clinical practice. This study introduces an innovative approach to classify knee OA severity by using femur-tibia bone distances as a proxy biomarker for cartilage thickness. The study analyzed MRI-derived bone mask images from 197 cases, each containing 160 2D slices. A custom algorithm was employed to identify slices with complete femur and tibia structures, automatically excluding irrelevant or incomplete slices. Bone distances were computed from segmented bone masks, generated via U-Net, and further enhanced through adaptive thresholding and morphological closing operations, which improved the segmentation quality. These refined bone masks were used in a Convolutional Neural Network (CNN) framework, which outperformed traditional machine learning models. The updated pipeline achieved an improved ROC AUC score of 76% with a mean cross-validated accuracy of 76% across 10 folds for binary classification, distinguishing non-OA (KL 0–2) from OA (KL 3–4) cases. The proposed method demonstrates the potential to reduce reliance on direct cartilage measurements while achieving significant classification performance improvements. Future work will focus on further refining the segmentation and classification steps to support widespread clinical adoption.*

***Keywords—Knee* Osteoarthritis, *MRI, femur tibia* bone distances**

I. Introduction

Osteoarthritis (OA) is a prevalent degenerative joint disease, disproportionately affecting older adults and contributing significantly to pain, disability, and reduced quality of life. Among various joints, the knee is particularly vulnerable, making the accurate assessment of knee OA severity crucial for timely diagnosis and treatment. Traditionally, severity evaluation relies on the Kellgren-Lawrence (KL) grading scale, a manual method that, while widely used, is subjective, inconsistent, and time-intensive. These limitations highlight the need for automated, objective approaches that can streamline the diagnostic process.

One promising alternative is the use of femur-tibia bone distance measurements as an indirect biomarker for cartilage thickness. Unlike direct cartilage measurement, which often requires specialized expertise and prolonged processing times, bone distance analysis offers a more accessible and efficient approach. By leveraging MRI-derived bone masks, this method facilitates automated severity classification through machine learning workflows, providing consistent and reproducible results.

This study presents a novel pipeline for knee OA severity classification using automated femur-tibia bone distance measurements integrated with a Convolutional Neural Network (CNN) framework. By combining advanced image segmentation techniques with machine learning, the proposed approach demonstrates significant improvements in classification accuracy and reliability, paving the way for scalable clinical applications.

II. Methodology

Custom algorithm identified approximately valid slices per patient by selecting slices with complete femur and tibia structures and excluding irrelevant ones. Followed by adaptive thresholding and morphological closing operations to enhance segmentation, bone contours were detected using external contour methods, with the largest contour identified as the femur and the second largest as the tibia. Distance measurements were restricted to the gap region between the femur and tibia, where Euclidean distances were computed for all boundary points. The average distance across gap-specific points were calculated per slice and aggregated into feature vectors for each patient. These vectors were used to train machine learning models, including CNN and Random Forest classifiers in distinguishing binary classifications of non-OA (KL 0–2) from OA (KL 3–4) cases.

A. Data Acquisition

The dataset consisted of 197 MRI-derived cases, each containing 160 2D slices. The goal was to analyze bone distances using femur and tibia masks segmented from MRI slices.

Classification labels followed the KL scale:

- Non-OA: KL grades 0–2
- OA: KL grades 3–4

B. Preprocessing

A custom algorithm was employed to select slices with complete femur and tibia structures, automatically excluding irrelevant or incomplete slices. Segmentation of bone masks was performed using a U-Net architecture, enhanced by adaptive thresholding and morphological closing operations to improve segmentation quality.

C. Feature Extraction

Bone distances were calculated as proxy biomarkers for cartilage thickness. Distances were extracted from segmented bone masks, providing a basis for classifying OA severity. Generated a feature vector using aggregated distances for each patient, standardized to uniform length, and saved as a CSV file.

D. Classification Model Framework

The study employed both Random Forest (RF) and Convolutional Neural Network (CNN) models:

- Random Forest (RF): Baseline results were obtained using traditional machine learning methods.
- CNN-Based Framework: Achieved superior performance with cross-validated accuracy and ROC AUC scores.

E. Evaluation Metrics

- The models were evaluated based on the following metrics:
- Binary Classification Accuracy (Non-OA vs. OA)
- ROC AUC Score

III. Results

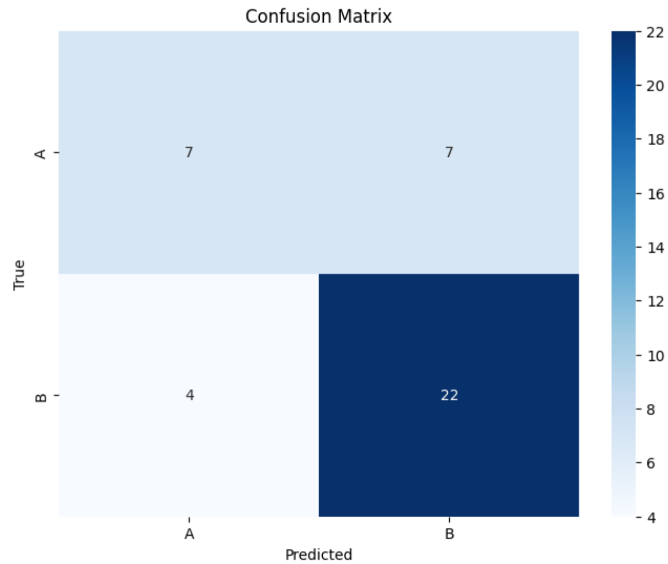
During mid report, the model evaluated using 10-fold cross-validation, achieving a mean accuracy of 65% and a ROC AUC score of 60%. We have kept Step=5 and after trying various classification model, Random Forest was best suited for the case. After fine tuning, refining feature extraction, various trails, testing, evaluating 30 different results for all different approaches, we observed that Step=1, which is processed with feature importance & trained over Random Forest, gives the best results. CNN overfits the data as the dataset is too small for this complex algorithm, which results into low accuracy on testing the trained model over an all together new test dataset. While the feature vectors generated for the step values 2 & 3 performed decent when classified using Random Forest Classifier.

Random Forest (RF) Performance:

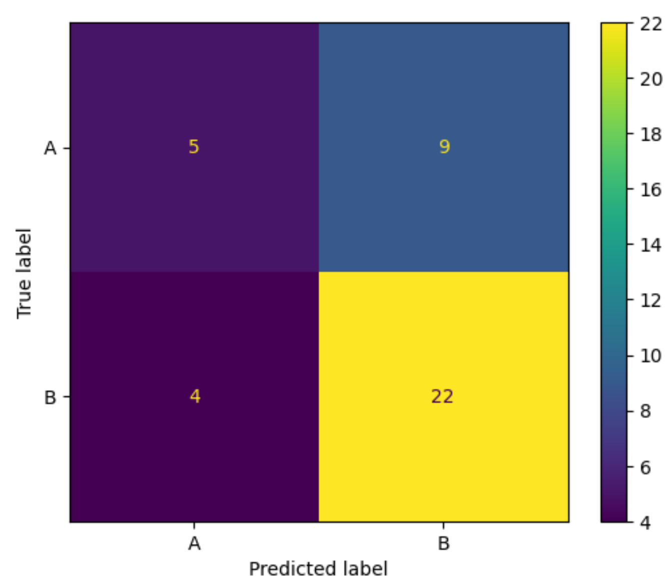
Feature Set	Excluding "V00XRKL" (%)	Including "V00XRKL" (%)
Step 1	66.55	76.64
Step 2	67.30	75.96
Step 3	61.40	72.52

CNN Performance:

Feature Set	Excluding "V00XRKL" (%)	Including "V00XRKL" (%)
Step 1	52.50	67.25
Step 2	52.50	60.23
Step 3	60.00	60.00



Confusion Matrix; for Step = 1 (RF)



Confusion Matrix; for Step = 1 (CNN)

IV. Conclusion

This study successfully developed an automated machine learning pipeline for classifying knee osteoarthritis (OA) severity using femur-tibia bone distances extracted from MRI-derived bone masks. By creating and optimizing three feature vectors, applying rigorous feature selection, and preprocessing data, the pipeline enhanced model performance and reliability.

The evaluation included training and testing both Random Forest and Convolutional Neural Network (CNN) models, with the CNN-based approach achieving superior results through cross-validation. The final optimized models demonstrated effective binary classification of OA severity, offering a robust and scalable solution for automating knee OA diagnosis.

These findings emphasize the potential of automated bone distance measurements as a reliable biomarker, reducing reliance on direct cartilage analysis and addressing the limitations of traditional KL grading methods. Future work will aim to further refine segmentation and classification processes and explore the pipeline's application across larger, more diverse datasets to support widespread clinical adoption.

V. Acknowledgement

We would like to express our deepest gratitude to our teammates Milind Kumar Choudhary, Rahul Singh, Saatvik Chaudhari, and Andrew Dinspechin for their unwavering dedication and collaborative efforts throughout this project. Their expertise and commitment have been invaluable in bringing this research to fruition.

We extend our heartfelt thanks to Dr. Juan Shan for her exceptional mentorship and guidance. Her insights, encouragement, and support in resolving research challenges and addressing doubts were instrumental in shaping the success of this project. We are truly grateful for her time and effort in mentoring us.

VI. Team Members Task Distibution

We have distributed the task among each member cognitively, and every step is also assisted by other team members.

TASK	Members
Step1: Distance Measurement Algorithm	Rahul Singh, Andrew Dinspechin
Step2: Valid Slice Algorithm	Saatvik Chaudhari, Milind Kumar Choudhary
Step3: Feature Vector Generation	Andrew Dinspechin, Saatvik Chaudhari
Step4: Machine Learning Processing	Milind Kumar Choudhary, Rahul Singh
Step5: Fine Tuning	All

VII. References

[1] R. Tiulpin, M. Thevenot, M. Odena, M. R. G. M. van Ginneken, and M. A. Viergever, "Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach," *Sci. Rep.*, vol. 8, no. 1, p. 1727, 2018

[2] S. Antony, S. S. S. R. Anjan, and S. S. S. R. Anjan, "Automatic Classification of Knee Osteoarthritis Severity Using Convolutional Neural Networks," *Appl. Sci.*, vol. 13, no. 3, p. 1658, 2023

[3] M. Thomee, M. S. S. R. Anjan, and S. S. S. R. Anjan, "Application of Machine Learning in Predicting Osteoarthritis Progression," *J. Clin. Med.*, vol. 13, no. 3, p. 1658, 2023

[4] J. Liu, Y. Zhang, and Y. Wang, "Bone Surface Distance as an Indicator for Osteoarthritis Diagnosis Using Deep Learning," *J. Med. Imaging*, vol. 7, no. 3, p. 035501, 2020

[5] R. Ponnusamy, M. Zhang, Y. Wang, X. Sun, M. Chowdhury, J. B. Driban, T. McAlindon, and J. Shan, "Automatic segmentation of bone marrow lesions on MRI using a deep learning method," Bioengineering 2024, vol. 11, no. 4: 374.

[6] R. Almajalid, J. Shan, M. Zhang, G. Stonis, and M. Zhang, "Knee Bone Segmentation on Three-Dimensional MRI," 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), New York City, NY, USA, 2019

[7] A. D. Desai, A. M. Schmidt, E. B. Rubin, C. M. Sandino, M. S. Black, V. Mazzoli, K. J. Stevens, R. Boutin, C.Ré, G. E. Gold, B. Hargreaves, and A. Chaudhari, "SKM-TEA: A Dataset for Accelerated MRI Reconstruction with Dense Image Labels for Quantitative Clinical Evaluation," in Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 2021 (NeurIPS Datasets and Benchmarks 2021)