CS 671 Final Project

Ari Perez
Pace University
Seidenberg School of CS
New York, NY
ap79624n@pace.edu

Prudhvi Raj Rekula
Pace University
Seidenberg School of CS
New York, NY
pr45629n@pace.edu

Kanishk Devda
Pace University
Seidenberg School of CS
New York, NY
kd90393n@pace.edu

***Abstract*—This paper presents a methodology for predicting knee osteoarthritis (OA) severity using MRI knee scans and machine learning models. The severity of OA was assessed using the Kellgren-Lawrence (KL) grading scale. We developed an iterative process to extract and refine cartilage thickness measurements from segmented MRI slices using MATLAB. Three distance measurement strategies were explored: centroid-based distances, median of the 3 smallest distances, and normalized ratios. Machine learning models, specifically Multi-Layer Perceptron (MLP) and Random Forest (RF), were trained and evaluated using the Weka platform. Results demonstrated that the normalized distance ratios combined with Random Forest achieved the highest performance, with a ROC Area of 74.0% and 96 correctly classified instances. The study highlights the importance of refined feature extraction and dataset optimization for improving predictive accuracy in medical imaging.**

***Keywords*— Osteoarthritis, Computer Vision, Machine Learning**

I. INTRODUCTION (*HEADING 1*)

Osteoarthritis (OA) is one of the most prevalent musculoskeletal disorders, significantly impacting quality of life, especially in aging populations. The severity of OA is typically measured using the Kellgren-Lawrence (KL) grading scale, which ranges from 0 (healthy) to 4 (severe). Accurate and automated prediction of OA severity from medical imaging, such as MRI scans, could play a critical role in early diagnosis and treatment planning.

The primary biomarker for OA severity is the thickness of the cartilage between the tibia and femur. Manual measurement of this thickness across MRI slices is labor-intensive and subject to variability. Machine learning models offer a promising solution for automating this process, provided the data is properly prepared and features are extracted effectively.

In this study, we processed MRI knee scans for 198 patients, comprising over 31,000 image slices. Using MATLAB libraries, we extracted distance measurements between segmented bone structures (tibia and femur) as a proxy for cartilage thickness. An iterative approach was employed to improve the dataset and refine the measurement strategy:

Initial centroid-based distances between detected boundary objects.

Median of the 3 smallest distances between objects.

Normalized ratios combining centroid and minimum distances.

We trained and evaluated Multi-Layer Perceptron (MLP) and Random Forest (RF) models using Weka, optimizing both the dataset and feature extraction to address issues such as data redundancy and noise. Our results demonstrate that the normalized distance ratios significantly improved performance, with Random Forest achieving a ROC Area of 74.0% and 96 correctly classified instances.

The remainder of this paper is organized as follows: Section II describes the methodology, including data preparation and predictive modeling. Section III presents experimental results, highlighting improvements achieved through iterative refinement. Finally, Section IV provides conclusions and future work.

II. METHOD

A. Data Preparation and Feature Extraction using Matlab

The dataset provided contains MRI knee scans for 198 patients, with each scan consisting of 160 2D slices, resulting in a total of 31,680 images. These images represent segmented cross-sections of the knee joint. The goal was to construct a dataset that could train a machine learning model to predict the severity of knee osteoarthritis (OA) using Kellgren-Lawrence (KL) grades. The KL scale is a 5-point system (0 to 4) where higher grades represent more severe OA.

To identify relevant slices for measurement, we utilized MATLAB libraries to detect "boundary objects," which we assumed to be the tibia and femur. If fewer than two boundary objects were detected in a slice, the slice was marked as invalid (`NaN`).

To extract measurements, we iteratively improved our methodology:

1. **Centroid Distances**: Initially, we measured the distance between the centroids of the two largest boundary objects in each valid slice. (see Fig 1)
2. **Minimum Distance Median**: We then refined the measurement by calculating the median of the 3 smallest distances between the two boundary objects. (see Fig 2)
3. **Normalized Ratios**: Finally, we normalized the minimum distances by dividing them by the centroid-based distances. This approach aimed to account for variations in patient size and improve the robustness of the dataset. Further details on this normalization are provided in the results section.

To handle missing values, we employed linear interpolation. This provided a complete feature vector of distances for each patient, ready for model training.

B. Predictive Models

To evaluate the dataset, we used the Weka machine learning platform. The target variable, KL grade, was converted to a nominal attribute to facilitate classification. We experimented with two classifiers:

- **Multi-Layer Perceptron (MLP)**: A neural network model with default settings.
- **Random Forest (RF)**: An ensemble-based decision tree classifier.

Cross-validation with 10 folds was used to assess model performance, and the results were compared based on the ROC Area and the number of correctly classified instances.*Iterative Data Refinement*

To improve model performance and reduce overfitting, we iteratively refined the dataset:

1. **Slice Reduction**: We reduced data granularity by selecting every 3rd and later every 5th slice, starting from slice 27, to minimize redundancy.
2. **Middle Slice Exclusion**: Transition tissues in the middle slices (approximately slices 80-110) were excluded to eliminate noise.
3. **Improved Measurement Methods**: As described in Section II.A, we iteratively refined the distance measurements from centroids to minimum distances and, finally, to normalized ratios.

III. EXPERIMENT RESULTS

A. Initial Results

Using the full dataset (slices 27-137) and measuring centroid distances:

- **MLP**: ROC Area = 60.7%, 71 correctly classified instances.
- **Random Forest**: ROC Area = 58.7%, 64 correctly classified instances.

These results indicated limited predictive capability, prompting further dataset refinement.

B. Reducing Data Granularity

To address overfitting, we reduced the number of slices by selecting every 3rd slice:

- **MLP**: ROC Area = 54.4%, 64 correctly classified instances.
- **Random Forest**: ROC Area = 59.4%, 67 correctly classified instances.

Further reduction to every 5th slice resulted in:

- **MLP**: ROC Area = 55.6%, 69 correctly classified instances.
- **Random Forest**: ROC Area = 59.1%, 64 correctly classified instances.

While Random Forest showed slight improvements, results remained suboptimal.

C. Excluding Transition Tissues (middle slices)

Following the professor's suggestion, we excluded the middle 30 slices (transition tissue regions):

- **MLP**: ROC Area = 55.9%, 77 correctly classified instances.
- **Random Forest**: ROC Area = 56.3%, 55 correctly classified instances.

This yielded marginal improvements in correctly classified instances but little impact on ROC Area. See Table 1 for full results.

D. Improved Distance Measurement

Switching to the median of the 3 smallest distances between objects resulted in noticeable improvements. (see Table 2)

These results demonstrated that focusing on minimum distances provided a more robust measurement for predicting OA severity.

E. Normalized Distance Ratios

Combining the centroid-based and minimum-distance measurements into a ratio further enhanced predictive performance. This normalization accounted for patient size variations, ensuring that cartilage measurements were relative to overall joint size. Random Forest once again outperformed MLP, achieving the best results observed throughout the experiments. (see Table 3)

These results demonstrated that the ratio-based measurement effectively addressed discrepancies due to patient size, leading to a notable increase in both ROC Area and correctly classified instances.

F. Analysis and Interpretation

The combination of refined distance measurements and normalized ratios demonstrated significant improvements in both ROC Area and classification accuracy. The best performance was achieved using the full dataset with the Random Forest classifier, yielding a ROC Area of 74.0% and 96 correctly classified instances. This suggests that leveraging normalized distances effectively accounts for patient size variations, enhancing predictive power.

Further improvements could include exploring additional machine learning models, fine-tuning hyperparameters, and integrating domain-specific features to better predict osteoarthritis severity.

IV. CONCLUSION

In this study, we presented a systematic approach to predicting knee osteoarthritis severity using MRI knee scans and machine learning models. The methodology involved extracting cartilage thickness measurements between the tibia and femur from segmented MRI slices. Through an iterative process, we refined our measurement techniques and optimized the dataset to reduce redundancy and noise.

Three measurement strategies were explored: centroid-based distances, median of the 3 smallest distances, and normalized ratios. The normalized ratios, which accounted for patient size variations, yielded the most significant improvements. Using Random Forest as the classifier, we achieved a ROC Area of 74.0% and 96 correctly classified instances, outperforming other configurations.

Our results highlight the importance of refined feature extraction and targeted dataset optimization in improving predictive accuracy for medical imaging tasks. Future work will focus on exploring additional machine learning models, fine-tuning hyperparameters, and incorporating domain-specific features to further enhance performance. Additionally, validating the methodology on larger and more diverse datasets will help assess its generalizability for clinical applications.

The findings of this study demonstrate the potential for automated, machine learning-based methods to assist in the early diagnosis and assessment of osteoarthritis severity, ultimately supporting clinicians in improving patient care.

A. Figures and Tables

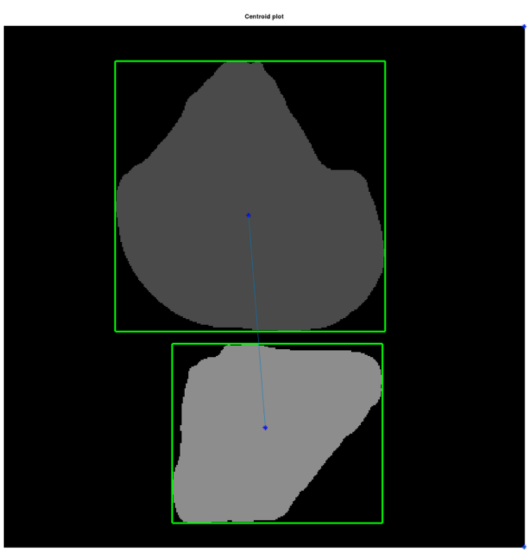

Fig 1. Measuring distance between centroids. Blue line is measured distance.

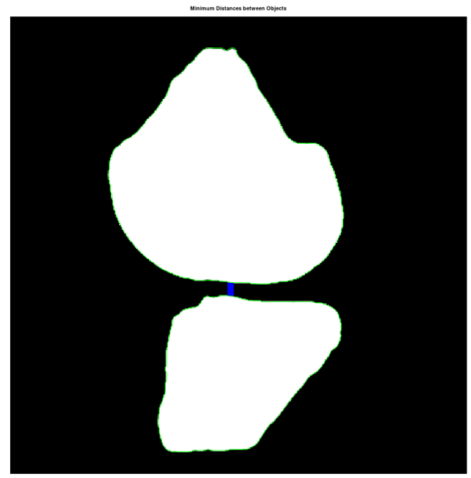

Fig 2. Measuring smallest distance between objects. Blue line is measured distance.

TABLE I. CENTROID DISTANCE RESULTS SUMMARY

Configuration	*Classifier*	*Correct Instances*	*ROC Area*
Full dataset[a], all slices	MLP	71	60.7%
Full dataset[a], all slices	RF	64	58.7%
Every 3rd Slice	MLP	57	54.4%
Every 3rd Slice	RF	67	59.4%
Every 5th Slice	MLP	69	55.6%
Every 5th Slice	RF	64	59.1%
Every 5th Slice, exluding middle 30	MLP	77	55.9%
Every 5th Slice, exluding middle 30	RF	55	56.3%

TABLE II. MINIMUM DISTANCE MEDIAN RESULTS SUMMARY

Configuration	*Classifier*	*Correct Instances*	*ROC Area*
Full dataset[a], all slices	MLP	74	63.9%
Full dataset[a], all slices	RF	87	71.9%
Every 3rd Slice	MLP	64	62.9%
Every 3rd Slice	RF	86	72.7%
Every 5th Slice	MLP	63	62.8%
Every 5th Slice	RF	85	70.5%
Every 5th Slice, exluding middle 30	MLP	64	59.7%
Every 5th Slice, exluding middle 30	RF	72	64.4%

TABLE III. NORMALIZED RATIOS RESULTS SUMMARY

Configuration	*Classifier*	*Correct Instances*	*ROC Area*
Full dataset[a], all slices	MLP	70	63.8%
Full dataset[a], all slices	RF	96	74.0%
Every 3rd Slice	MLP	60	59.4%
Every 3rd Slice	RF	87	73.0%
Every 5th Slice	MLP	71	65.6%
Every 5th Slice	RF	86	72.2%
Every 5th Slice, exluding middle 30	MLP	76	63.9%
Every 5th Slice, exluding middle 30	RF	63	62.7%

a. The full dataset contains slices 27 - 137. As mentioned in II.A, we removed all slices below 27 and above 137. Slices 27 - 137 were never used for any experiments.