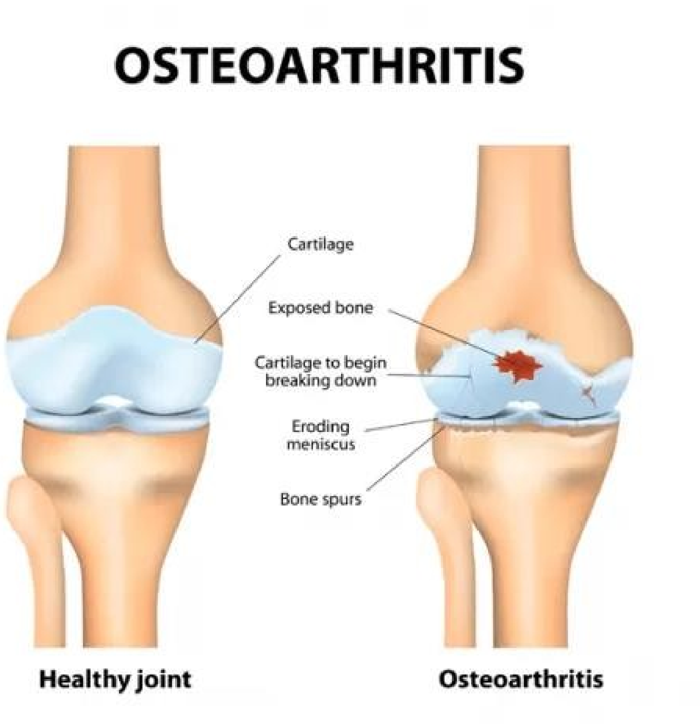
KNEE OSTEOARTHRITIS CLASSIFICATION USING COMPUTER VISION MACHINE LEARNING

Team Members:

Milind Kumar Choudhary
Rahul Singh
Saatvik Chaudhari
Andrew Dinspechin

INTRODUCTION



Osteoarthritis (OA) is a common degenerative joint disease, particularly affecting older adults, leading to pain, disability, and reduced quality of life. Traditionally, knee OA severity is assessed using the manual Kellgren-Lawrence (KL) grading scale, which can be inconsistent and time-consuming. Automated bone distance measurements offer an efficient alternative by indirectly estimating cartilage thickness and streamlining diagnosis. By extracting femur-tibia bone distances from MRI-derived bone masks, this method reduces reliance on manual cartilage analysis and enables integration into machine learning workflows for more consistent and automated OA severity classification.

PROJECT OBJECTIVES

Develop a computer-aided diagnosis system for knee osteoarthritis (OA) severity classification using bone distance measurements as an indirect measure of cartilage thickness.

• **Bone Distance Measurement**: Create an algorithm to measure femur-tibia bone distances from MRI-derived bone masks, accurately capturing relevant bone gaps.

• **Feature Extraction**: Automate the selection of valid MRI slices containing both femur and tibia bones, ensuring robust and consistent feature extraction across cases.

• **Machine Learning Classification**: Use extracted distance vectors as features to train machine learning models (e.g., CNN, SVM) to predict OA severity based on Kellgren-Lawrence (KL) grading.

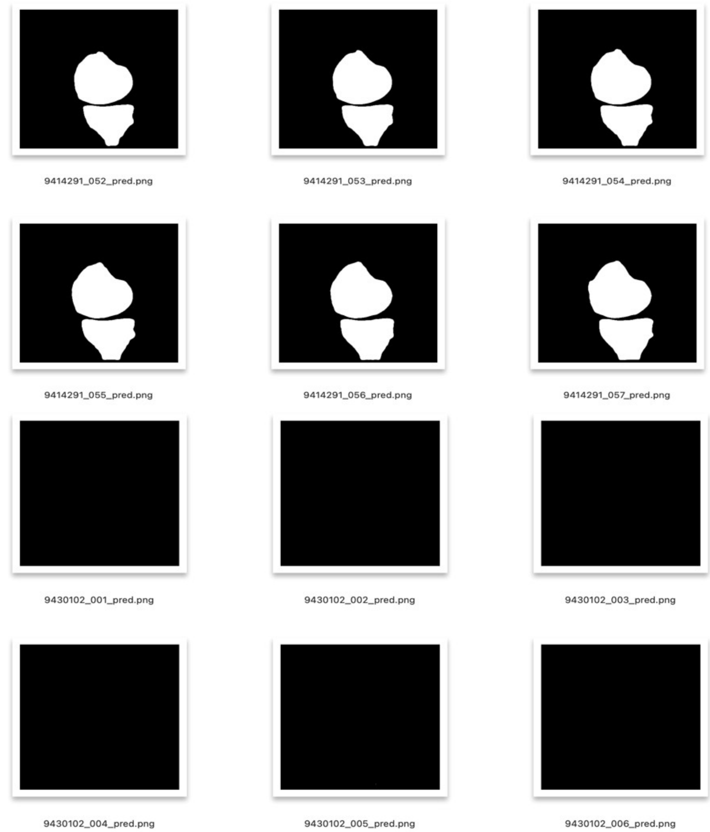
• **Performance Optimization & Evaluation**: Refine algorithms, explore better feature representations, and evaluate model performance using classification accuracy and ROC analysis for improved OA severity predictions.

DATASETS

Dataset Overview: 197 patients, each with 160 MRI slices.

Classification Labels: Based on the **Kellgren-Lawrence (KL)** scale: Classes (“V00XRKL” & “V03KL”)

• Non-OA: KL 0–2
• OA: KL 3–4



STEPS FOLLOWED

STEP 1: Segmentation & Preprocessing
Bone masks were binarized using Otsu's method for adaptive thresholding, followed by morphological closing to smooth contours and eliminate small artifacts.

STEP 2: Feature Extraction
Femur and tibia contours were detected and sorted by area to exclude irrelevant structures. Euclidean distances between the contours were averaged for each slice, generating a feature vector, which was saved to a new .csv file.

STEP 3: Feature Engineering & EDA
The feature vector csv was merged with the KL grades dataset. Exploratory Data Analysis (EDA) was performed to identify patterns, clean the data, and ensure it was ready for modeling.

STEP 4: Classification
The extracted features were used to train a Random Forest Classifier & CNN for OA severity prediction, with performance validated through 10-fold cross-validation.

STEP 5: Optimization
Refine the distance measurement algorithm and experiment with slice selection strategies, such as skipping neighboring slices or excluding middle transition slices, to improve feature representation and model accuracy.

STEP 1: SEGMENTATION & PREPROCESSING

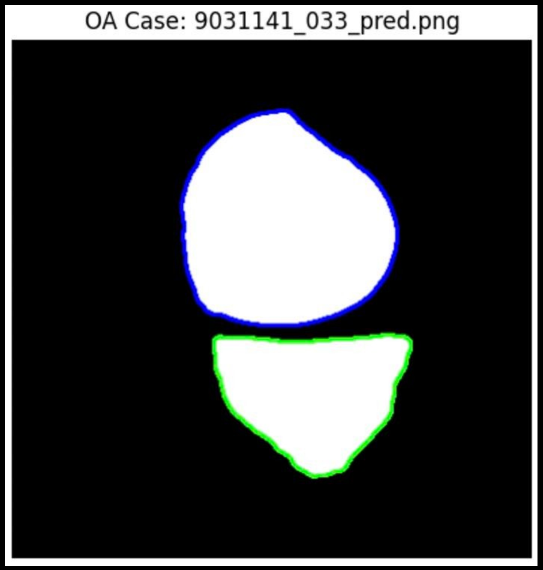
Distance Calculation: Computes minimum and average distances between femur and tibia contours.

Slice Validation: Identifies slices with both femur and tibia bones.

Valid Slice Detection: Finds the range of valid slices.

Slice Sampling: Selects slices at regular intervals (e.g., every 2nd slice).

Output: Reports valid range, sampled slices, and counts.



OA Case Distance Matrix:

	Image	Minimum Distance	Average Distance
0	9031141_033_pred.png	9.848858	154.480537

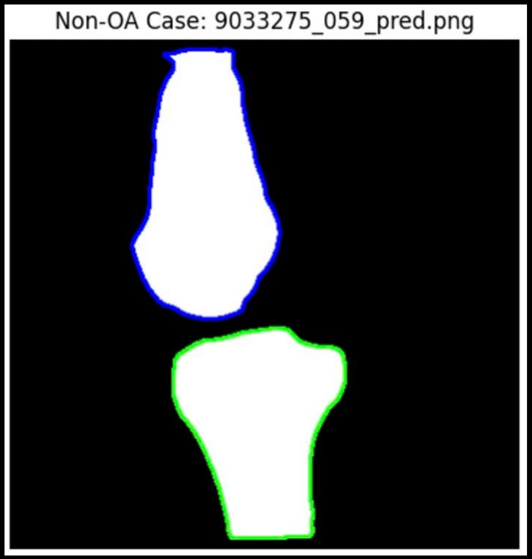
STEP 2: FEATURE EXTRACTION

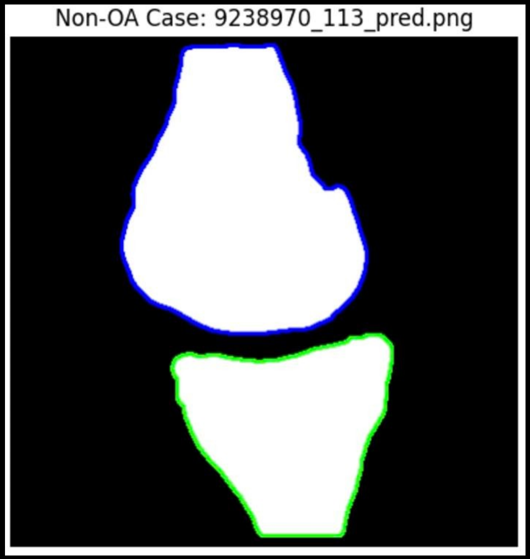
Distance Measurement:
Calculates the average Euclidean distance between femur and tibia contours for each sampled slice.

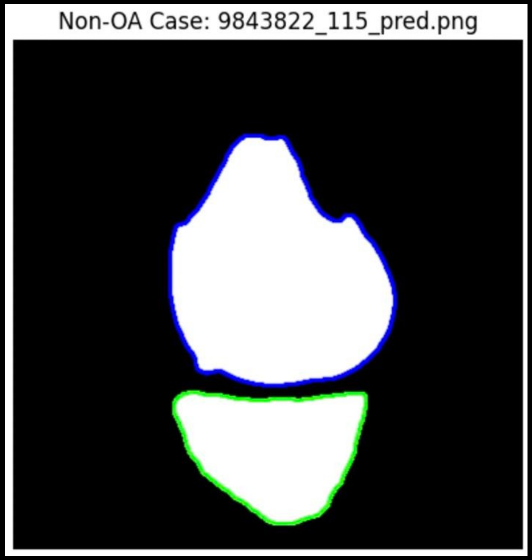
Feature Vector Generation:
Aggregates distances for each patient, pads vectors to uniform length, and saves them as a CSV file.

	PatientID	0	1	2	3
148	9770617	106.360117	108.093939	108.858745	109.235333
149	9031141	69.214327	96.567226	104.193785	108.909473
150	9858252	252.196324	120.895997	115.043638	120.505226
151	9883115	102.546360	112.193758	112.272996	115.005119
152	9637676	150.468964	150.372032	144.498938	169.083962
153	9033275	87.368227	92.914778	97.468372	95.829469
154	9843845	174.720764	117.799205	117.373805	120.511870

Testing of *Non-OA cases*





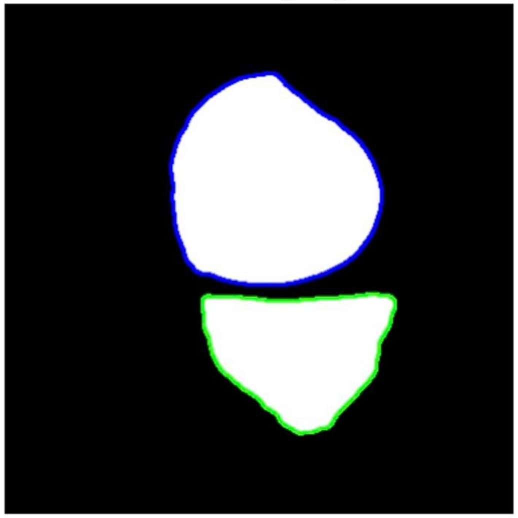


Non-OA Case Distance Matrix:

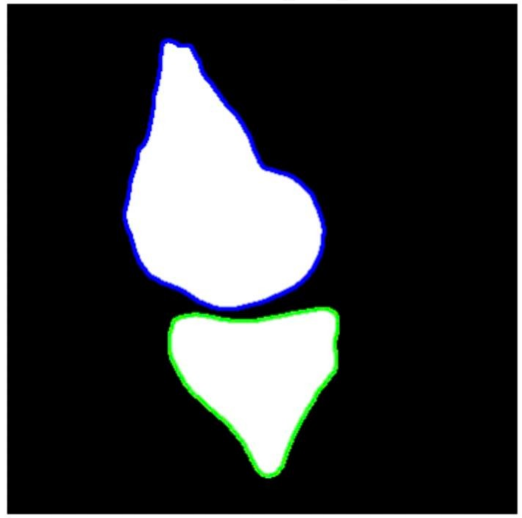
	Image	Minimum Distance	Average Distance
0	9033275_059_pred.png	14.317821	182.379553
1	9238970_113_pred.png	15.297059	192.827893
2	9843822_115_pred.png	11.180340	162.329305

Testing of *OA cases*

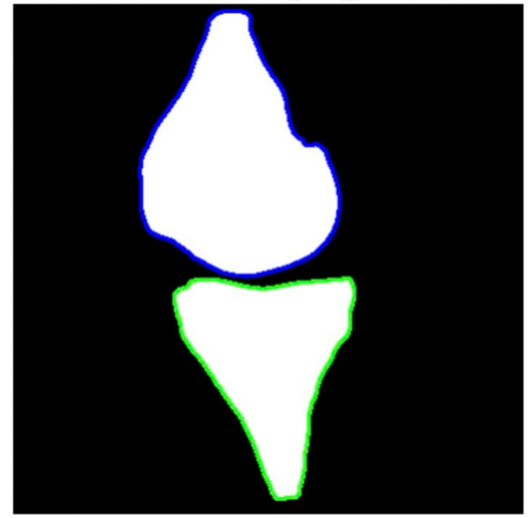
OA Case: 9031141_033_pred.png



OA Case: 9048789_102_pred.png



OA Case: 9917284_062_pred.png



OA Case Distance Matrix:

	Image	Minimum Distance	Average Distance
0	9031141_033_pred.png	9.848858	154.480537
1	9048789_102_pred.png	6.708204	163.054553
2	9917284_062_pred.png	5.385165	184.187346

STEP 3: FEATURE ENGINEERING & EDA

Data Loading: Reads two datasets – KL grade data (df_baseline) and feature vectors (df_features) – into pandas DataFrames.

Column Renaming: Standardizes the column names (cases → Patient_ID and PatientID → Patient_ID) for merging.

Data Merging: Combines the two DataFrames on the Patient_ID column to align KL grades with corresponding feature vectors.

Data Cleaning: Drops the unnecessary column 'Unnamed: 3' and prints the shape of the final merged DataFrame.

STEP 4: CLASSIFICATION USING ML ALGORITHMS

- **Data Split**: Prepares features and target, encodes labels, and splits data (80-20).

- **Model Training**: Trains a Random Forest Classifier.

- **Validation**: Runs 10-fold cross-validation for performance.

- **Evaluation**: Computes mean cross-validation score and ROC AUC.

```
Cross-validation scores for OA status prediction: [0.65       0.75
 0.6        0.68421053 0.63157895 0.57894737]
Mean CV Accuracy: 0.6544736842105264
Classification Report for OA status prediction:
              precision    recall  f1-score   support

           A       0.50      0.21      0.30        14
           B       0.68      0.88      0.77        26

    accuracy                           0.65        40
   macro avg       0.59      0.55      0.53        40
weighted avg       0.61      0.65      0.60        40

ROC AUC score for OA status prediction: 0.6057692307692308
```

STEP 5: OPTIMIZATION

Feature Vectors Creation: Generated 3 feature vectors with **Step = 1, 2, 3** having sizes (197, 117), (197, 74), and (197, 52), respectively.

Feature Selection: Identified important features (e.g., Step = 1 reduced to size (197, 34)) and filled zeros with row-wise averages.

Model Training: Trained **Random Forest** models with and without "V00XRKL" for the target "V03KL" and tested **CNN** models for comparison using hyperparameters & best hyperparameters.

Approaches & Evaluation: Tested a total of **27 approaches** using all feature vectors and evaluated model accuracies to find optimal results.

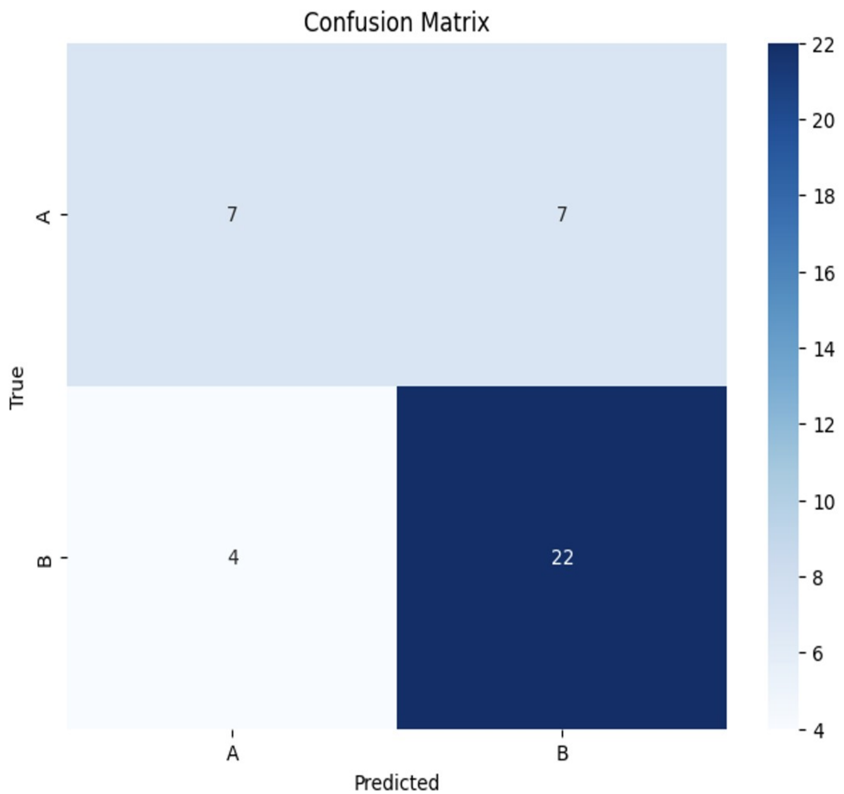
OBSERVATIONS & RESULTS

	STEP = 1	STEP = 2	STEP = 3
Excluding “V00XRKL”	66.55%	67.30%	61.40%
Including “V00XRKL”	76.64%	75.96%	72.52%
Excluding “V00XRKL” (CNN)	52.50%	52.50%	60%
Including “V00XRKL” (CNN)	67.25%	60.23%	60%

We can observe that Step=1, which is processed with feature importance & trained over Random Forest is giving the best results.

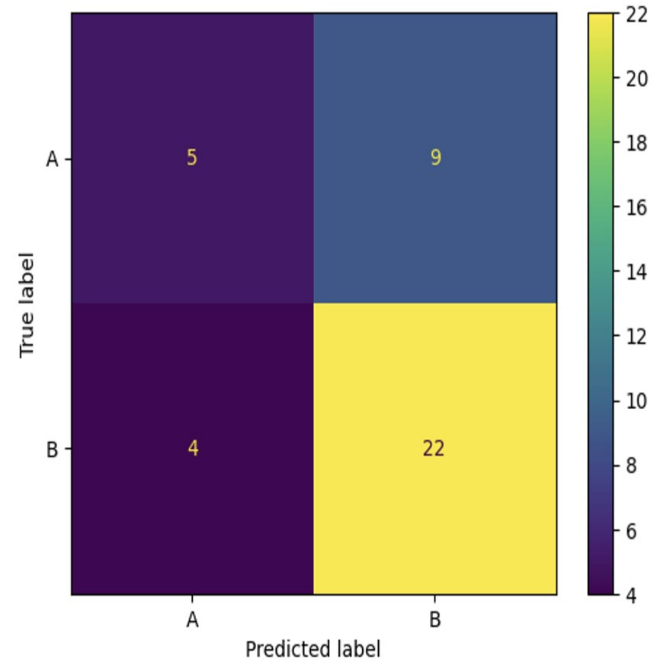
Step = 1

```
Best Parameters: {'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 10, 'n_estimators': 200}
Cross-validation scores: [0.75       0.8        0.8        0.75       0.65       0.5
 0.8        0.89473684 0.84210526 0.84210526]
Mean cross-validation score: 0.7628947368421052
ROC AUC score: 0.7664835164835164
```



Performed feature importance & kept target as "V03Kl" & model trained on the pattern of the patient mentioned by the doctors manually.

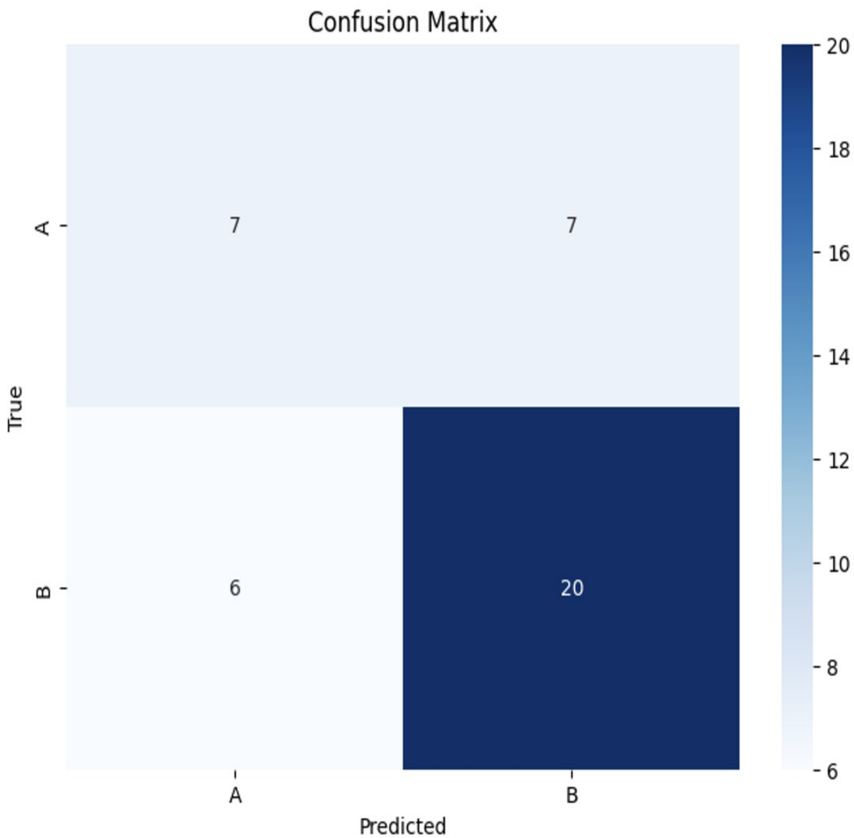
CNN Model: Target "V03KL" including the "V00XRKL"



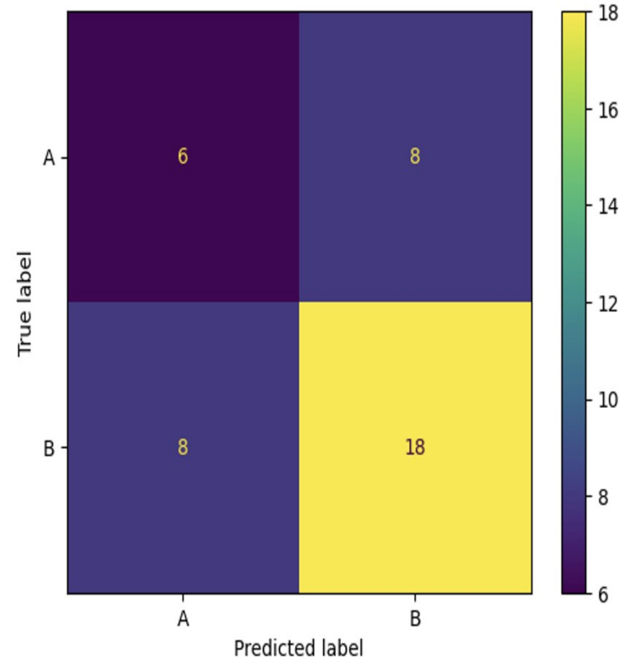
	precision	recall	f1-score	support
0	0.56	0.36	0.43	14
1	0.71	0.85	0.77	26
accuracy			0.68	40
macro avg	0.63	0.60	0.60	40
weighted avg	0.66	0.68	0.65	40

Step = 2

```
Cross-validation scores: [0.75       0.8        0.9        0.75       0.7        0.45
 0.85       0.78947368 0.78947368 0.73684211]
Mean cross-validation score: 0.7515789473684211
ROC AUC score: 0.7596153846153847
```



CNN



Predicted label

```
              precision    recall  f1-score   support

           0       0.43      0.43      0.43        14
           1       0.69      0.69      0.69        26

    accuracy                           0.60        40
   macro avg       0.56      0.56      0.56        40
weighted avg       0.60      0.60      0.60        40
```

Target “V03KL” and feature dataset includes “V00XRKL”

CONCLUSION

- Successfully developed a machine learning pipeline to classify knee OA severity using femur-tibia bone distances extracted from MRI-derived bone masks.
- Created and optimized 3 feature vectors, applied feature selection, and preprocessed data to improve model performance.
- Trained Random Forest and CNN models, testing multiple approaches to achieve reliable accuracies through cross-validation.
- The optimized models demonstrated effective OA classification, providing a robust and automated solution for knee osteoarthritis diagnosis.

THANK YOU

- Milind Kumar Choudhary
- Rahul Singh
- Saatvik Chaudhari
- Andrew Dinspechin

- milindkumar.choudhary@pace.edu