

Capstone Three Project Proposal : Problem Identification

Problem statement (Hypothesis formation)

Can a machine learning classification model built using knee X-ray images be used to identify and grade the severity of knee osteoarthritis?

Context

According to the [National Institute of Health](#), osteoarthritis (OA) is a joint disease in which the tissues in the joint break down over time. It is the most common type of arthritis and is more common in older people. Knee osteoarthritis is one of the most common types of OA. The [CDC](#) states that OA affects over 32.5 million adults in the United States. As the population grows and ages, this number is projected to increase. There is no cure for OA, and thus those afflicted must learn strategies to manage their pain and reduce disability. Early identification of OA in patients may help them change their lifestyle in order to slow the progression of OA.

Criteria for success

Grade the severity of the test set of knee images with accuracy.

Scope of solution space

We will focus exclusively on the dataset provided on Kaggle.

Constraints within solution space

1. The knee severity grades are unbalanced in the dataset: the lowest Grade 0 consists of approximately 40% of the entire dataset. Grade 1 consists of ~18%, Grade 2 of ~26%, Grade 3 of ~13% and Grade 4 of only ~3%.
2. There are 9786 files in the dataset, but there appears to be some overlap with one of the folders, so we will likely only use 8260 files.
3. The images in the dataset are very clean, close-up, and centered images of the knee. In reality, knee X-rays are probably not this clean. We could address this issue and the earlier one (limited number of images in the dataset) by augmenting the dataset with synthetically modified images derived from the existing images in the dataset (skewing, rotating, flipping, etc.).

Stakeholders

- Kaggle: provided knee X-ray data
- Medical professionals
- People with knee pain who are wondering if they have knee OA!

Key data sources

From Kaggle: [Knee Osteoarthritis Dataset with Severity Grading](#)

This dataset contains knee X-ray data for both knee joint detection and knee KL grading. The Grade descriptions are as follows:

- Grade 0: Healthy knee image.
- Grade 1 (Doubtful): Doubtful joint narrowing with possible osteophytic lipping
- Grade 2 (Minimal): Definite presence of osteophytes and possible joint space narrowing
- Grade 3 (Moderate): Multiple osteophytes, definite joint space narrowing, with mild sclerosis.
- Grade 4 (Severe): Large osteophytes, significant joint narrowing, and severe sclerosis.

This dataset contains 9786 PNG files in the following subdirectories:

- Auto_test : 1526 files (these files appear to be a subset of the files in the Test directory)
 - 0: 604
 - 1: 275
 - 2: 403
 - 3: 200
 - 4: 44
- Test : 1656 files
 - 0: 639
 - 1: 296
 - 2: 447
 - 3: 223
 - 4: 51
- Train : 5778 files
 - 0: 2286
 - 1: 1046
 - 2: 1516
 - 3: 757
 - 4: 173
- Val : 826 files
 - 0: 328
 - 1: 153
 - 2: 212
 - 3: 106
 - 4: 27