Predicting Knee Osteoarthritis

Background and Objective
Osteoarthritis (OA) is a leading cause of disability worldwide. OA severity quantified using the Kellgren Lawrence (KL) grading scale.

Osteoarthritis Initiative (OAI) Dataset
- Patient data (4794 patients) at 0, 12, 24, 36, 48, 72, and 96 month time points
- Bilateral knee radiographs (DICOM images)
- Bilateral Radiological KL Grades

Data Pre-Processing and Labeling
- Excluded all knees with OA at baseline (KL ≥ 2)
- Classified knees as progressors (y=1) if KL ≥ 2 at later time point
- Total number of progressors = 1586 knees
- Total number of non-progressors = 7540 knees
- Converted DICOM images to PNG
- Split images into right and left knees
- Mirrored left knee and normalized images
- Paired images with associated labels
- Data rescaling/cropping to identify knee region

Data Augmentation
- Brought data from multiple time points to augment data set
- Sampled X-ray images from all KL ≤ 1 knees

Model Results & Software Flow
Binary Cross-Entropy Loss: \( \text{CE}(\hat{y}, y) = -y \log \hat{y} - (1 - y) \log (1 - \hat{y}) \)

Future Work
- Increasing the size of the data set would improve results
- Visualizing the features learned by these networks could help identify clinically correctable problems
- Include patient demographic data to give model more features
- Performance could be improved by using magnetic resonance images (MRI)

Experiments
- Batch Normalization very important to model
- SGD better performance than Adam optimizer
- Lower alpha without BN reduce overfitting
- Regularization had little effect
- Dropout at 90% had negative effect but helped at 90%
- Data augmentation using multiple time points increased precision and recall

Discussion
- This study showed proof of concept for the use of deep learning to detect features of “healthy” knee radiographs that are predictive of OA that current medical techniques have failed to identify
- Using only KL = 1 knee radiographs
- 70% precision/recall on the test set
- Deep learning detecting features predictive of OA that may be undetectable to the human eye
- Increasing dataset size improved results
- Still overfitting on our training data set
- More data will be useful for mitigating the high variance
- It may not be possible to achieve accuracy much higher than this using only knee radiographs
- Radiographs contain only information on bony structures
- Soft tissue information from cartilage and other structures in the knee may be important in predicting knee OA

References