Detecting Early Stage Knee Osteoarthritis Using Deep Transfer Learning

Lokesh Meesala
CCSE - Computer Science
001078109
lmeesala@students.kennesaw.edu

Abstract

- Knee osteoarthritis is a prevalent form of the disease, but its early-stage diagnosis can be challenging due to subjective and error-prone interpretation of X-ray images.
- The research aims to develop a deep learning network utilizing Convolutional Neural Networks (CNNs) for multi-class classification of Knee X-ray images into five categories: Normal/Healthy, Doubtful, Minimal, Moderate, and Severe.
- State-of-the-art CNN architectures such as ResNet and VGG Nets are explored to investigate the effectiveness of transfer learning in the classification task.
- To address class imbalance, a selective augmentation technique is employed, and an iterative model training process is utilized for fine-tuning the network.

Introduction

- Osteoarthritis, a common form of arthritis, affects millions globally as cartilage between bones deteriorates.
- Early diagnosis is challenging due to similar symptoms to other diseases, requiring imaging tests like MRI and X-Ray evaluated by radiologists.
- Radiologists use the Kellgren-Lawrence grading system, classifying osteoarthritis into five grades.
- Early detection is crucial for managing symptoms and slowing disease progression with medication or therapy.

My contribution is as follows,

- Transfer learning from ResNet and fine tune it by adding custom layers.
- Selective Augmentation is implemented to deal with the class imbalance problem.
- To improve the fine-tuning results, iterative model training is implemented.
- Achieved an F1 score of 62.3%

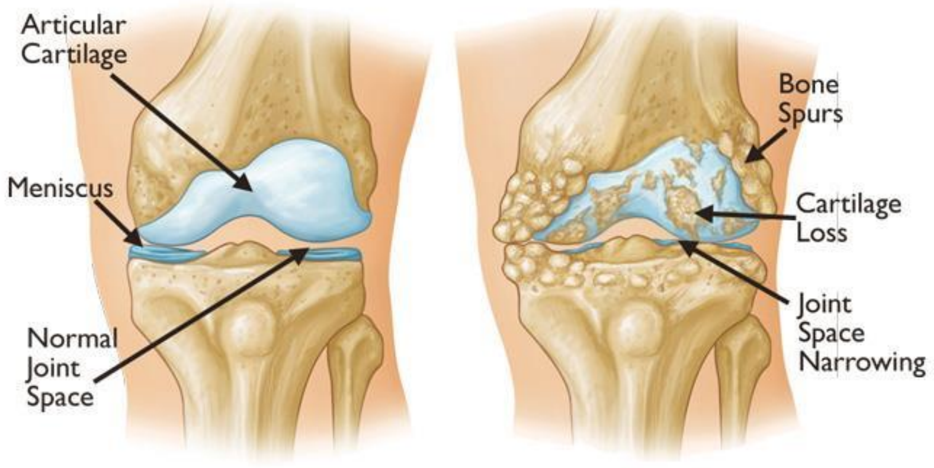


Fig 1: This illustration shows a healthy knee and a knee with osteoarthritis

Related Works

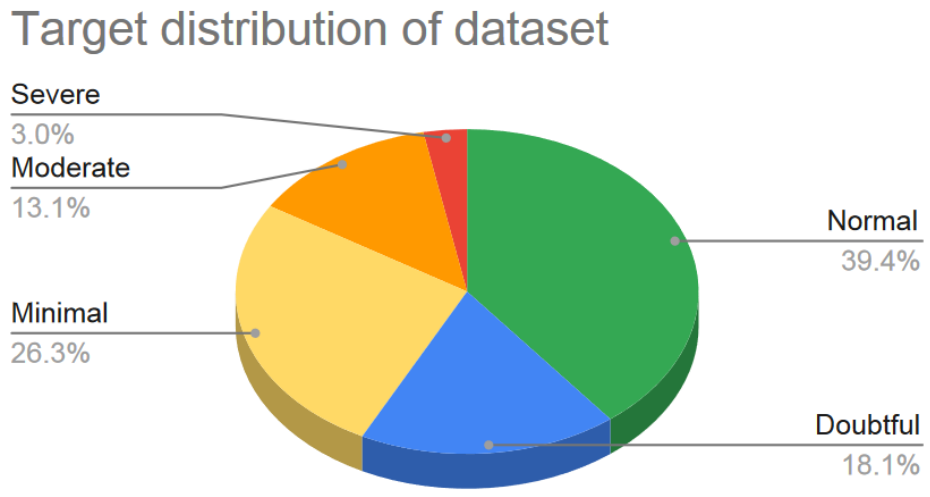
1. The Osteoarthritis Initiative (OAI) is a large-scale study that aims to identify biomarkers for the onset and progression of knee osteoarthritis. It began in 2002 and enrolled over 4,000 participants at high risk or with early signs of the disease, collecting extensive clinical, demographic, imaging, and biomarker data. This dataset is available to researchers worldwide to support further research in the field. [2] Chen, Pingjun (2018) labeled these X-Rays and classified them into five categories.

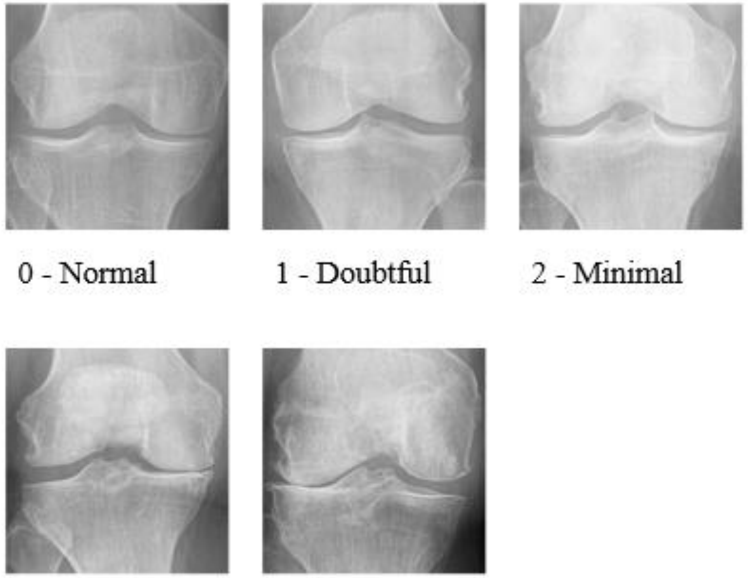
2. Previous studies have explored the use of various machine learning models for knee osteoarthritis classification. [4] Kokkotis, Christos et al. employed Logistic Regression and achieved an accuracy of 77.88% using physical activity indexes, questionnaire data, and self-reported symptoms for binary classification. However, these models rely on challenging data collection methods compared to easily available and non-invasive X-rays, which motivated the use of X-ray datasets in this research.

3. Different deep learning models, such as Fully Convolutional Networks (FCN), Convolutional Neural Networks (CNN), VGG-16, ResNet, and DenseNet, have been applied for multi-class classification in knee osteoarthritis studies. [5] Antony, J. et al. utilized X-ray data and trained FCN and CNN models with joint training on regression and classification, reporting an F1 score of 61%. On the other hand, [6] Wahyuningrum, R. T. et al. (2019) employed a cropped version of the X-ray dataset and trained VGG-16, ResNet, and DenseNet models, achieving an accuracy of 75.28% using CNN for feature extraction and LSTM for severity classification.

Data Description

1. The dataset contains 8260 X-Ray images of size 224x224.
2. Class imbalance is observed, especially for Severe and Moderate, it makes it more difficult for the model to classify them correctly.
3. I have used 5778 X-Ray images for our training purpose, 826 images for validation and 1656 images for testing. i.e. 70% for training, 10% for validation and 20% for testing.



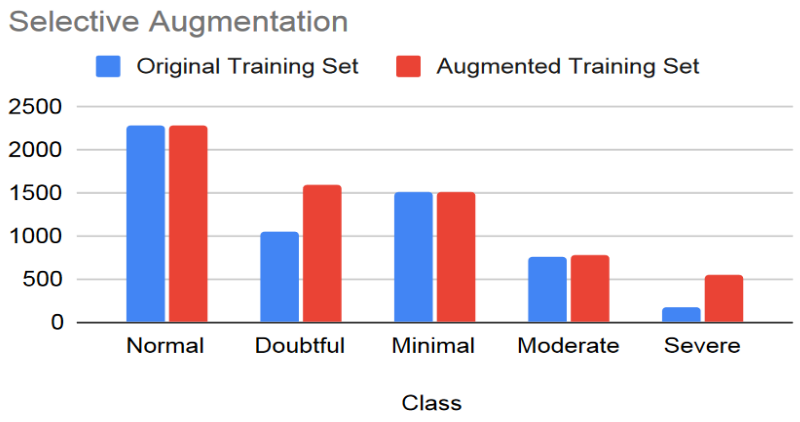


0 - Normal 1 - Doubtful 2 - Minimal

3 - Moderate 4 - Severe

Methodology

- Research begins with collecting knee X-ray images from the OAI database, already labeled by [2] Chen, Pingjun (2018).
- Severity of osteoarthritis is classified into five grades using Kellgren-Lawrence grading system or OARSI atlas.
- Dataset is divided into training, validation, and testing sets using a 70-10-20 split.
- ResNet152 selected as the transfer learning model after analyzing initial results.
- To address class imbalance, selective augmentation is implemented by augmenting underrepresented class images in the training data.



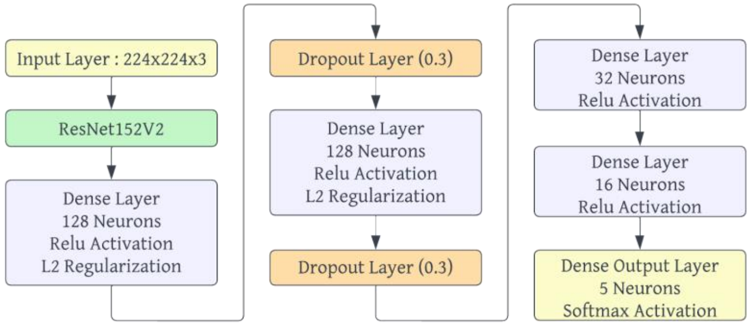


Fig 3: High Level Model Architecture.

Implementation

- After selectively augmenting the data, images and labels are saved in data frames. Keras flow from dataframe is used for creating batches for training and validation, where images undergo further augmentation techniques like rotations, shifting, and flipping.
- The model architecture includes fine-tuning layers with a combination of Dense and Dropout layers, kernel regularization to control overfitting, and a Dense layer with softmax activation as the output layer. The model is compiled using the adam optimizer, categorical cross-entropy loss function, and accuracy metric.
- Training the model involves 100 epochs with an initial learning rate of 0.001. Keras ReduceLROnPlateau and EarlyStopping are used to prevent overfitting and underfitting. Batch size of 64 is selected, and the model undergoes an iterative training approach where layers of the ResNet152 model are unfrozen progressively over 50 epochs.

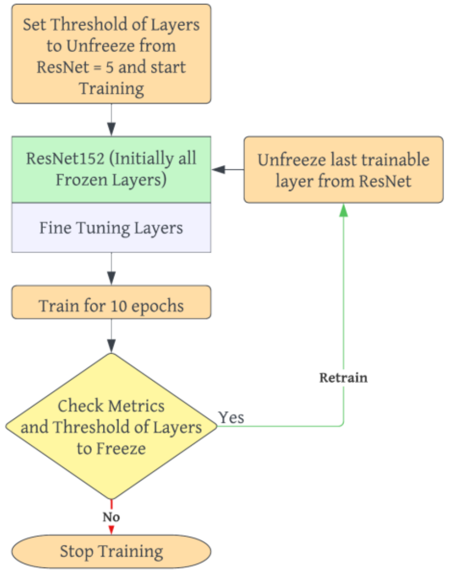


Fig 7: Iterative Model Training

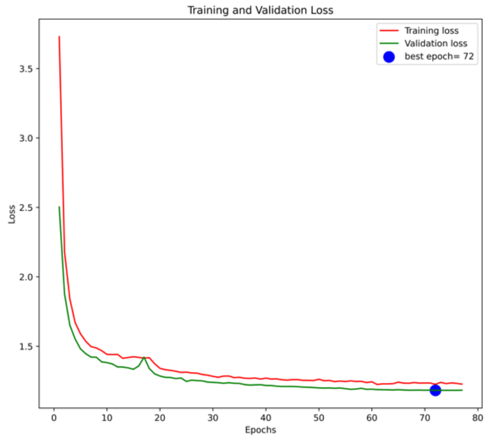


Fig 4: Training and Validation Loss across epochs

Results

- Selective Augmentation significantly improved results, but exceeding a threshold of 0.9 reduced the f1 score by approximately 3.5%. This discrepancy may arise from the uneven distribution of disease progression in reality.
- In comparison with [5] Antony, J. et al, my implementation gave a slightly better F1 Score of 62.3% and Precision of 79.1% and lower Recall Score of 51.9%. My model compromised Recall for a better Precision score.
- Confusion matrices show misclassifications predominantly in target grades 1 and 3. Models struggle to distinguish subtle differences despite selective augmentation. Consideration for 3 classes and relabeling mislabeled images under expert supervision could enhance model performance.

Implementation	Model	F1 Score	Precision Score	Recall Score
Current	ResNet152 With Selective Augmentation (0.7)	**0.623**	0.791	0.519
[5] Antony, J. et al	CNN with joint training on regression and classification	0.61	0.61	0.63

Fig: Comparison with related research paper.

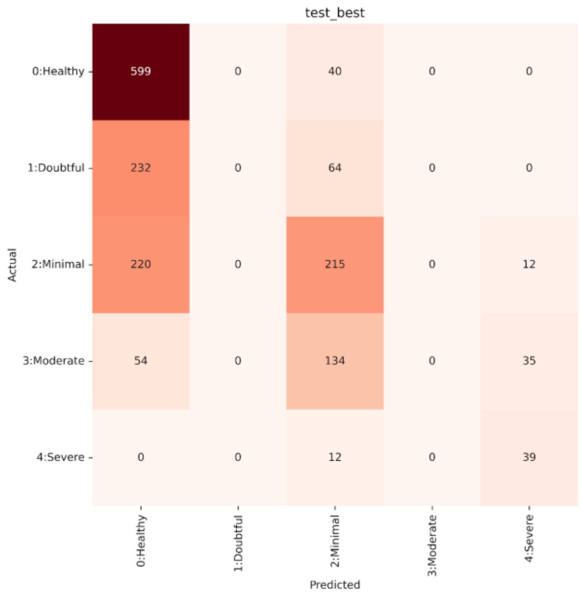


Fig: Confusion Matrix for Test Predictions

References

1. Kellgren-Lawrence Scale for radiographic classification of osteoarthritis - mcmaster textbook of internal medicine. (n.d.). Retrieved March 6, 2023, from https://empendium.com/mcmtextbook/table/031_0728
2. Chen, Pingjun (2018), “Knee Osteoarthritis Severity Grading Dataset”, Mendeley Data, V1, doi: 10.17632/56rmx5bjcr.1
3. Osteoarthritis - orthoinfo - aaos. OrthoInfo. (n.d.). Retrieved April 28, 2023, from https://orthoinfo.aaos.org/en/diseases--conditions/osteoarthritis/

4. Kokkotis, Christos et al. (2020) “A Machine Learning workflow for Diagnosis of Knee Osteoarthritis with a focus on post-hoc explainability”, en 2020 11th International Conference on Information, Intelligence, Systems and Applications IISA. IEEE.

5. Antony, J. et al. (2017) “Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks,” arXiv [cs.CV]. Available at: http://arxiv.org/abs/1703.09856
6. Wahyuningrum, R. T. et al. (2019). A New Approach to Classify Knee Osteoarthritis Severity from Radiographic Images based on CNN-LSTM Method. In 2019 IEEE 10th International Conference on Awareness Science and Technology, iCAST 2019 - Proceedings [8923284] (2019 IEEE 10th International Conference on Awareness Science and Technology, iCAST 2019 - Proceedings). Institute of Electrical and Electronics Engineers Inc.
7. Mayo Foundation for Medical Education and Research. (2021, June 16). Osteoarthritis. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/osteoarthritis/symptoms-causes/syc-20351925
8. Schiratti, J.-B. (2021, October 18). A deep learning method for predicting knee osteoarthritis radiographic progression from MRI - Arthritis Research & Therapy. BioMed Central. Retrieved March 6, 2023, from https://arthritis-research.biomedcentral.com/articles/10.1186/s13075-021-02634-4

THANK YOU