DETECTION OF OSTEOARTHRITIS AND KL GRADING FOR KNEE OSTEOARTHRITIS USING DEEP LEARNING TECHNIQUES

Dhruva Shaw⋆† *Arittrabha Sengupta*†

⋆† dhruvashaw@ieee.org, Creative Net
† arittrabhasengupta@gmail.com, Lovely Professional University

ABSTRACT

Osteoarthritis (OA) is a prevalent degenerative joint disease, particularly affecting the knees[1]. Early and accurate detection of OA and its severity, often graded using the Kellgren-Lawrence (KL) scale, is crucial for timely intervention and management. This study explores the application of deep learning techniques to automatically detect OA and assign KL grades from knee X-ray images. We propose a novel deep learning architecture that effectively extracts relevant features from X-ray images and classifies them into different KL grades. Our model demonstrates promising results in terms of accuracy and sensitivity, potentially aiding radiologists in making faster and more accurate diagnoses.

***Index Terms*—** Osteoarthritis, Kellgren-Lawrence (KL) Grade, Medical Image Analysis, X-ray Imaging, Healthcare Technology

1. INTRODUCTION

Osteoarthritis (OA) is a degenerative joint disease characterized by the gradual breakdown of cartilage, leading to pain, stiffness, and reduced joint function[1]. Knee osteoarthritis (KOA) is one of the most common forms of OA, significantly impacting the quality of life for millions of people worldwide. Early and accurate diagnosis of KOA is crucial for timely intervention and effective management.

Traditional methods for KOA diagnosis, such as clinical examination and X-ray imaging, often rely on subjective assessments and can be time-consuming. To address these limitations, there has been a growing interest in developing automated methods for KOA detection and grading using advanced imaging techniques and artificial intelligence.

In recent years, deep learning, a subset of artificial intelligence, has emerged as a powerful tool for medical image analysis. Deep learning models, particularly convolutional neural networks (CNNs), have shown remarkable success in various medical image analysis tasks, including disease classification and segmentation. By leveraging the power of deep learning, we aim to develop an automated system for accurate detection and grading of KOA from X-ray images.

This research focuses on two key aspects of KOA diagnosis:

- **General OA detection**: Developing a deep learning model to identify the presence of OA in any joint.
- **KL grade assessment of KOA**: Developing a deep learning model to accurately assign a Kellgren-Lawrence (KL) grade to knee X-ray images, indicating the severity of KOA.

By combining these two models, we aim to provide a comprehensive and accurate assessment of KOA, enabling early intervention and improved patient outcomes.

2. METHODOLOGY

Here there were 2 different types of models where prepared using the transfer learning technique, they are namely as follows:

- **KL grade of OA**: Developing a deep learning model to accurately assign a Kellgren-Lawrence (KL) grade to knee X-ray images, indicating the severity of KOA. Here the data is taken from Kaggle[2] and RoboFlow[3]. Here we developed 4 different models namely:
  - **alexnet** (took least possible time to train with least no of epoch to reach stable state)
  - **densenet201**
  - **googlenet** (the best performing model with the best train time possible)
  - **inceptionresnetv2** (It took the most amount of training time to train)

  Here the model had 5 different classes to train on:
- - ***0*** : Normal[4]
  - ***1*** : Doubtful [4]
  - ***2*** : Minimal [4]
  - ***3*** : Moderate [4]
  - ***4*** : Severe [4]
- **Without KL Grade**: Developing a deep learning model to identify the presence of OA in any joint. From the above best performing model which was **googlenet** again did the transfer learning on 2 different classes namely; ***normal***, ***patient***

3. RESULTS

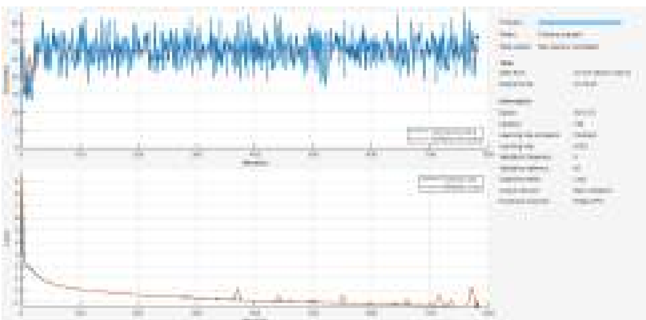
3.1. KL Grade

Here the no preprocessing of the data was done. The source[3][2] had their dataset prepared with all the ROI, thresholding and other techniques.

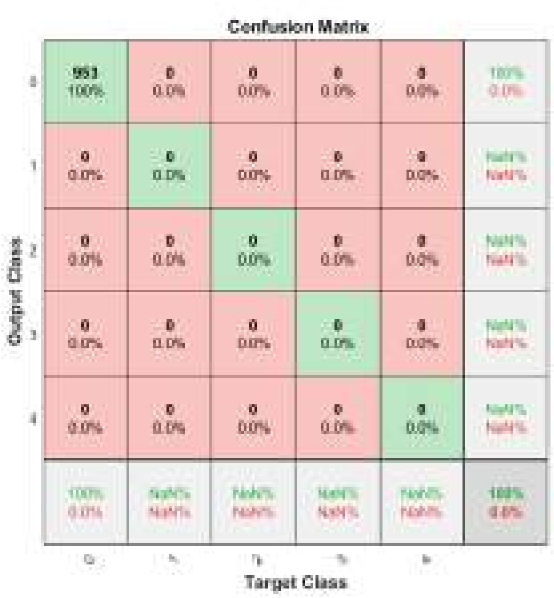
3.1.1. *Alex Net*

1 name	2 classes					3 macroAVG	4 microAVG
"true_positive"	263	0	0	0	0	52.6800	52.6800
"false_positive"	696	0	0	0	0	138	138
"false_negative"	0	253	153	148	138	138	138
"true_negative"	0	700	800	805	817	624.4000	624.4000
"precision"	0.2760	NaN	NaN	NaN	NaN	NaN	0.2760
"sensitivity"	1	0	0	0	0	0.2000	0.2760
"specificity"	0	1	1	1	1	0.8000	0.8190
"accuracy"	0.2760	0.2760	0.2760	0.2760	0.2760	0.2760	0.2760
"F-measure"	0.4326	NaN	NaN	NaN	NaN	NaN	0.2760

(a) Statistics of Interest of Alex Net



(b) Training Graph Alex Net

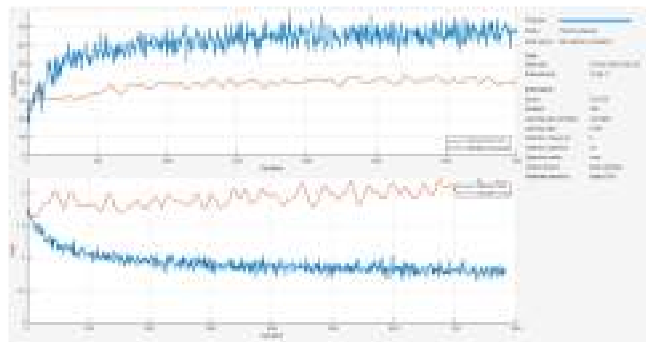


(c) Confusion Matrix Alex Net

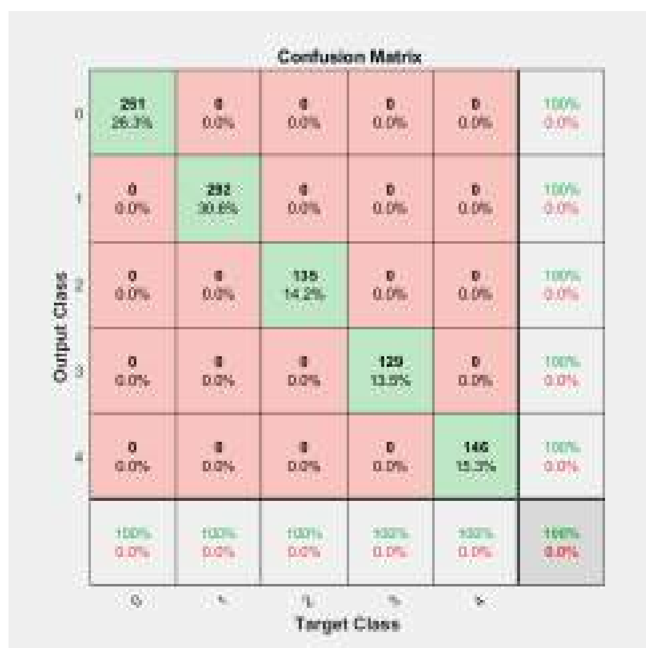
3.1.2. *Dense Net 201*

1 name	2 classes					3 macroAVG	4 microAVG
"true_positive"	172	151	46	19	56	107.8000	107.8000
"false_positive"	79	141	89	55	50	82.8000	82.8000
"false_negative"	81	162	107	74	40	82.8000	82.8000
"true_negative"	611	509	711	750	767	679.6000	679.6000
"precision"	0.6853	0.5171	0.3407	0.5736	0.6575	0.5548	0.5656
"sensitivity"	0.6340	0.5968	0.3007	0.5000	0.7059	0.5515	0.5656
"specificity"	0.8855	0.7986	0.8888	0.9317	0.9300	0.8807	0.8914
"accuracy"	0.5656	0.5656	0.5656	0.5656	0.5656	0.5656	0.5656
"F-measure"	0.6682	0.5541	0.3194	0.5343	0.6809	0.5516	0.5656

(a) Statistics of Interest of Dense Net 201



(b) Training Graph of Dense Net 201

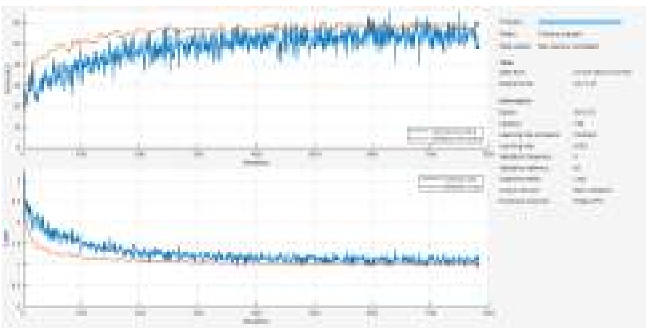


(c) Confusion Matrix of Dense Net 201

3.1.3. *Google Net*

1 name	2 classes					3 macroAVG	4 microAVG
"true_positive"	188	161	51	71	85	107.2080	107.2080
"false_positive"	69	153	97	62	36	83.4080	83.4080
"false_negative"	95	92	102	77	51	83.4080	83.4080
"true_negative"	621	547	793	743	781	679	679
"precision"	0.7085	0.5127	0.3446	0.5338	0.7025	0.5605	0.5624
"sensitivity"	0.6388	0.6364	0.3333	0.4797	0.6250	0.5426	0.5624
"specificity"	0.9080	0.7814	0.8788	0.9230	0.9559	0.8678	0.8986
"accuracy"	0.5624	0.5624	0.5624	0.5624	0.5624	0.5624	0.5624
"F-measure"	0.6720	0.5679	0.3389	0.5053	0.6615	0.5491	0.5624

(a) Statistics of Interest of Google Net



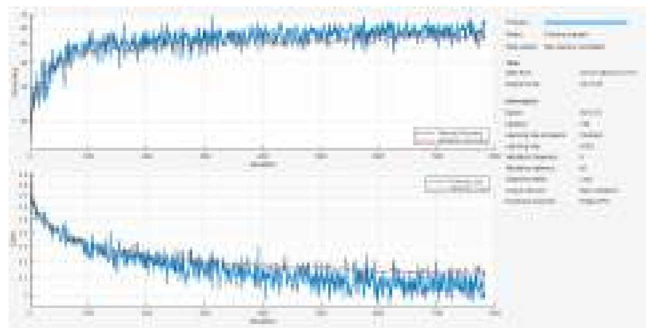
(b) Training Graph of Google Net

(c) Confusion Matrix of Google Net

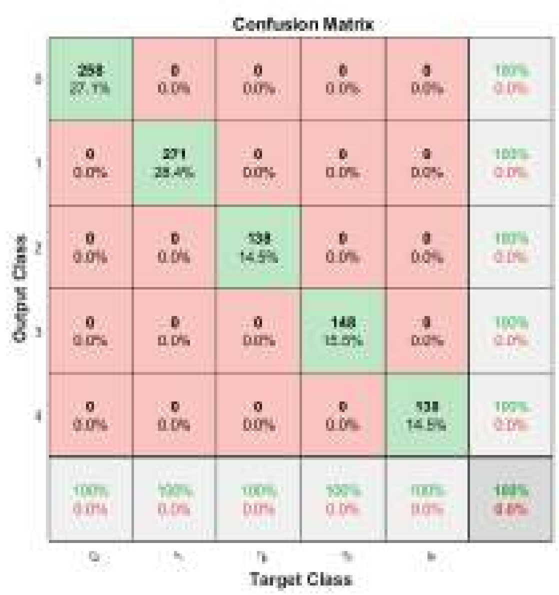
3.1.4. *Inception ResNet V2*

1 name	2 classes					3 macroAVG	4 microAVG
"true_positive"	174	147	54	80	92	109.4000	109.4000
"false_positive"	84	124	84	68	46	81.2000	81.2000
"false_negative"	89	106	98	68	44	81.2000	81.2000
"true_negative"	606	576	718	737	771	681.2000	681.2000
"precision"	0.6744	0.5424	0.3913	0.5405	0.6667	0.5631	0.5740
"sensitivity"	0.6616	0.5810	0.3529	0.5405	0.6765	0.5625	0.5740
"specificity"	0.8783	0.8229	0.8958	0.9155	0.9437	0.8911	0.8935
"accuracy"	0.5740	0.5740	0.5740	0.5740	0.5740	0.5740	0.5740
"F-measure"	0.6679	0.5611	0.3711	0.5405	0.6715	0.5624	0.5740

(a) Statistics of Interest of Inception ResNet V2



(b) Training Graph of Inception ResNet V2



(c) Confusion Matrix of Inception ResNet V2

3.2. Without KL Grade

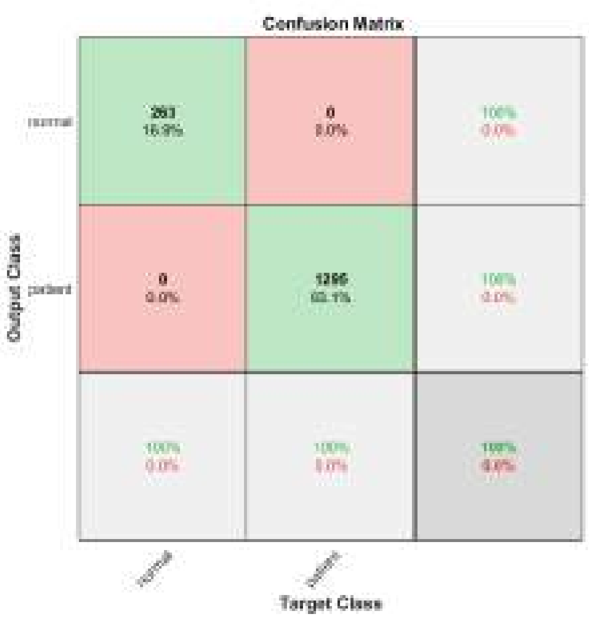
Here the no preprocessing of the data was done. It had 2 classes to recognize ***normal*** and ***patient***. The best model from the KL Grading model was taken (googlenet3.1.3) and then transfer learning was performed on it.

3.2.1. Google Net

1 name	2 classes		3 macroAVG	4 microAVG
"true_positive"	262	1294	778	778
"false_positive"	1	1	1	1
"false_negative"	1	1	1	1
"true_negative"	1294	262	778	778
"precision"	0.9962	0.9992	0.9977	0.9987
"sensitivity"	0.9962	0.9992	0.9977	0.9987
"specificity"	0.9992	0.9962	0.9977	0.9987
"accuracy"	0.9987	0.9987	0.9987	0.9987
"F-measure"	0.9962	0.9992	0.9977	0.9987

(a) Statistics of Interest of Google Net (Without KL Grade)

(b) Training Graph of Google Net (Without KL Grade)



(c) Confusion Matrix of Google Net (Without KL Grade)

4. RESULTS

Our experiments demonstrated that the GoogleNet model achieved the highest accuracy for both knee OA grading and general OA detection tasks.

Specifically, for knee OA grading, the GoogleNet model attained an accuracy of 56.24%. In the case of general OA detection, the same GoogleNet model, fine-tuned through transfer learning, exhibited exceptional performance with an accuracy of 99.87%.

5. ACKNOWLEDGMENTS

We would like to express our sincere gratitude to Dr. Bhaveshkumar C. Dharmani, Professor at Lovely Professional University, for providing the invaluable Osteoarthritis dataset sourced from AIIMS Bhatinda. Additionally, we acknowledge the contribution of the Kaggle and RoboFlow community for providing the Knee OA image dataset, which was instrumental in training and evaluating our Knee OA models.

6. REFERENCES

[1] Baisen Chen, Hongxiang Hong, Yuyu Sun, Chu Chen, Chunshuai Wu, Guanhua Xu, Guofeng Bao, and Zhiming Cui, "Role of macrophage polarization in osteoarthritis (review)," *Experimental and Therapeutic Medicine*, vol. 24, no. 6, Nov 2022.

[2] Shashwat Tiwari, "Knee osteoarthritis dataset with severity grading," *Kaggle.com*, 2018.

[3] Knee, "Knee osteoarthritis dataset object detection dataset (v1, 2023-10-31 7:06pm) by tru projects," *Roboflow*, 2023.

[4] Kaniz Fatema, Md Awlad Hossen Rony, Sami Azam, Md Saddam Hossain Mukta, Asif Karim, Md Zahid Hasan, and Mirjam Jonkman, "Development of an automated optimal distance feature-based decision system for diagnosing knee osteoarthritis using segmented x-ray images," *Heliyon*, vol. 9, no. 11, pp. e21703, Nov 2023.