Automated Deep Learning Based Knee Osteoarthritis Joint Extraction and Classification

Muhammad Sohail Abbas
Department of Data Science and Artificial Engineering,
National University of Computer and Emerging Sciences
(NUCES) Islamabad
Email: sohail.abbas@isb.nu.edu.pk

Sonain Jamil
Department of Computer Science,
Norwegian University of Science and Technology (NTNU),
Gjøvik 2815, Norway
Email: sonainjamil@ieee.org

Atif Khurshid
Department of Data Science and Artificial Engineering,
National University of Computer and Emerging Sciences
(NUCES) Islamabad
Email: atif.khurshid@isb.nu.edu.pk

2023 17th International Conference on Open Source Systems and Technologies (ICOSST) | 979-8-3503-8132-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICOSST60641.2023.10414241

***Abstract*—Knee osteoarthritis (KOA) is a widespread global condition, impacting over 300 million individuals as per the World Health Organization (WHO). Particularly prevalent among older adults, knee OA is a prominent cause of disability. Its occurrence increases with age, especially after 50, and is more frequent in women, particularly post-menopause. Several studies have been carried out so far for automated grading and classification of knee osteoarthritis (KOA), but none of them built strong foundations enough to make this system automated. This study focuses on machine-controlled knee joint extraction and grading classification with improved accuracy and performance. We used the osteoarthritis initiative (OAI) dataset of X-ray images for our study. Initially, a single-stage detector is used for joint extraction of the knee area as the X-ray images contain entire knees with both joints. Enhanced osteoarthritis feature extraction (OAFE) and osteoarthritis dimensionality reduction (OADR) blocks are used for grading classification. We have significantly improved state-of-the-art results. We have acquired joint extraction with a mean average precision (map) of 95.3% and grading classification accuracy of 78.93%. Furthermore, the performance due to the dimensionality reduction block has improved by a huge factor.**

***Index Terms*—Osteoarthritis, Knee-Joint Extraction, Osteoarthritis-Initiative.**

1. Introduction

Osteoarthritis (OA) is of big concern globally these days as a large number of people are being affected by it due to several different reasons. So far, there is no permanent cure for the disease. OA causes severe pain in joints and eventually limits that patient's movement. As OA progresses from grade 1 to grade 4, the patient suffers an immense amount of pain, causing permanent disability. Reasons for OA are very common, and it happens due to lake of routine exercises and genetic problems. Even in Pakistan, every 26.6 people among 1000 suffer from OA [1]. Not only is OA a problem for health, but it has a significant impact on finance worldwide. Billions are being spent on temporary cures for OA patients.

Several different datasets are available for research to make OA grading automated. Still, only one contains few data samples and is collected for a short duration in a single local region, which does not generalize the problem. Osteoarthritis Initiative (OAI) is a dataset collected over 96 months by the National Institute of health [2], containing 4796 subjects ages 45 to 69 years old. There are 2804 female and 1992 male subjects. They use Kellgren and Lawrence (KL) grades and have classified OA grades into five different classes, with grade 0 being health and grade 4 being the worst condition of OA. A sample image of the original dataset is presented in figure 1. Our target is to extract knee joints and classify them into respective grades.

To make the grading process automatic, several authors have worked with different methodologies to improve results significantly and to make this system fully automated. Despite several attempts at automated grading and classification of knee osteoarthritis (KOA) through various studies, none have established a robust framework for full automation. This study has revolutionized the landscape by focusing on automated knee joint extraction and enhanced grading classification, aiming to achieve heightened accuracy and superior performance. Leveraging the osteoarthritis initiative (OAI) dataset of X-ray images, our approach commences with a single-stage detector YOLOv5 for precise knee area extraction from the X-ray images containing complete knee structures. Subsequently, we introduce novel methodologies, the osteoarthritis feature extraction (OAFE) and the osteoarthritis dimensionality reduction (OADR) blocks, to enhance grading classification. Our results present a note-

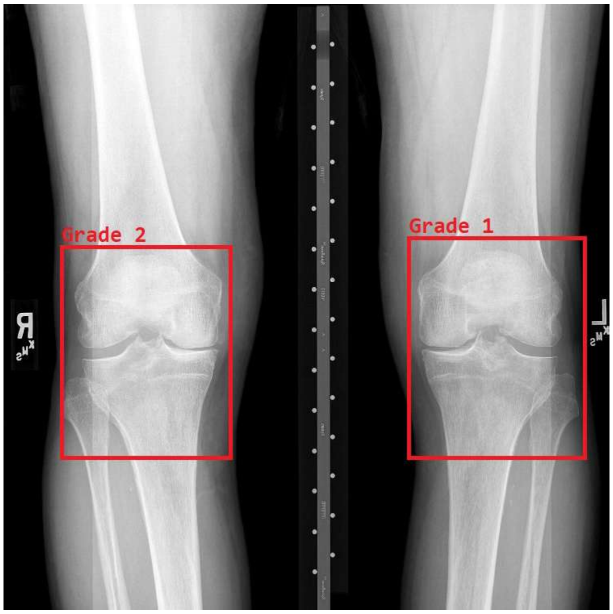

Figure 1. OAI dataset sample image.

worthy advancement, showcasing a joint extraction rate of 95.3% and a grading classification accuracy of 78.93%. Moreover, including the dimensionality reduction block contributes significantly to this enhanced performance.

Related work followed by an explanation of the proposed methodology, experimental results, and the conclusion is presented in the paper next.

2. Related Work

The authors in [3] used a custom network named OsteoHRNet. They used the OAI dataset and classified grades into five classes, achieving an accuracy of 71.74%. However, the system is not fully automated, requiring the extracted knee joints to be tackled manually. This demands human effort and is costly. In [4], the authors proposed a fully automated methodology for extracting knee joints using a convolution neural network (CNN). The method works well in making a cost-effective methodology as it does not require any human effort but only gives a classification accuracy of 69%. Due to these reasons, the model can be used in real-world problems.

The authors in [5] used a dataset with 1856 patients. Although it gives better accuracy on their dataset, it's not well generalized due to the low number of training samples. When testing the same methodology on OAI, it gives 68% of accuracy, but on their dataset, it gives the accuracy of 96%. They used four different models for training on their dataset. The authors in [6] used the same dataset as [4], and their methodology landed an overall accuracy of 70.3% on the OAI dataset as tested; however, they did not mention the accuracy of their proposed methodology for their dataset. The authors in [7] used OAI along with the Rani Channamma University (RCU) dataset. It outperforms state-of-the-art when trained and tested on the RCU dataset, but for OAI, it performs 73%. They used three different combinations of models, including CNNs (VGG-19, ResNet-101) with FFNNs for their methodology. Still, neither can perform significantly on the OAI dataset when classified in 5 KL grades.

In the following, we demonstrate our proposed methodology for segmenting and classifying brain tumors.

3. Proposed Methodology

The proposed framework is presented in Fig. 2. The dataset contains images with left and right knee as shown in 1. First, we must extract the knee joint area and classify both knees separately in their appropriate grades. We cannot classify the X-ray directly as both knees might vary significantly in grading. So, we are using YOLOv5 first to extract knee joints. After that, both knee joints will be applied with preprocessing before they can be fed to our proposed osteoarthritis feature extraction (OAFE) and osteoarthritis dimensionality reduction (OADR) blocks individually. Both blocks are applied three times for fine feature extraction. Extracted features are then passed to fully connected layers where KL grading will be applied to classify them in grades 0 to 4.

3.1. Knee Joint Extraction

On the input image, we apply YOLOv5 to extract knee joints. Initially, we annotated 200 sample images from the dataset to train YOLOv5. Among those 200 images, 40 were used to test the performance of YOLOv5. Once training is complete, perform testing before using the model with the actual proposed methodology.

3.2. Preprocessioning

Preprocessing is being applied after YOLOv5 as for YOLOv5, and the input results are already well generalized. We have applied augmentation of three kinds on extracted knee joints. We are applying contrast, illumination, and rotation. The augmentations were chosen carefully after considering possible problems that can occur in x-ray acquisition. Due to differences in light, contrast can vary in inputs. Similarly, careless procedure, rotation, and illumination can vary, which is why, to generalize our methodology well, we needed to make the above-mentioned augmentations on all training images.

3.3. Osteoarthritis Feature Extraction Block

The Osteoarthritis feature extraction (OAFE) block is shown in Fig. 3. We have applied depth-wise 2D convolutions followed by batch normalization and activation. This small block is followed by another block with a max pooling layer and two convolution layers. Before and after max pooling, features are concatenated to add global context. We

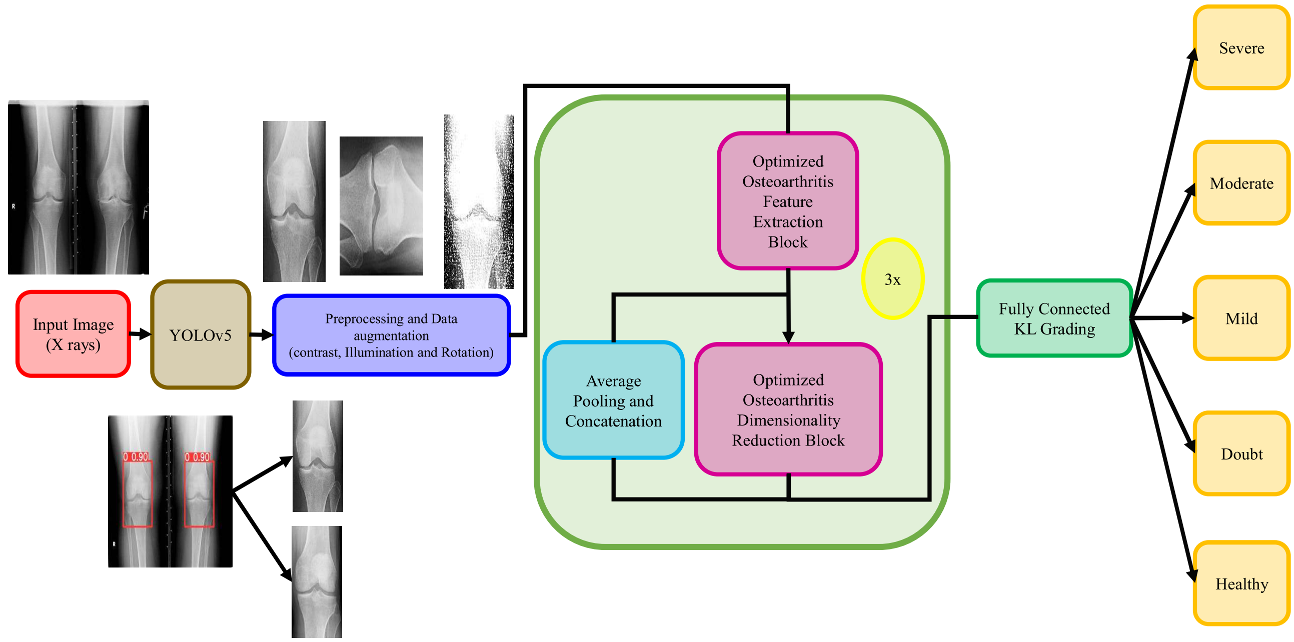

Figure 2. The proposed framework for knee extraction and classification.

are doing zero padding to resolve the spatial dimension's problem. This also helps in the vanishing gradient problem. OAFE block passes the extracted feature to OADR block, which is repeated thrice.

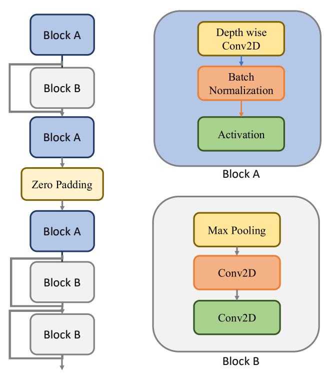

Figure 3. Osteoarthritis feature extraction block.

3.4. Osteoarthritis Dimensionality Reduction Block

The osteoarthritis dimensionality reduction (OADR) block is used to reduce dimensions after every few layers, as our model is very deep. Not reducing dimensions can make the model too complex and lead to model overfitting. In that case, our model will perform well on training data, but for test data, it will not generalize well and will not give good performance. In the OADR block, we used 1x1 convolution followed by batch normalization, relu activation, and a depth-wise convolution. After that, we again applied a 1×1 convolution and batch normalization as shown in Fig. 4. For the sake of global context, we are concatenating the feature from the OAFE block to the output of the OADR block by applying average pooling to that as shown in methodology Fig. 2.

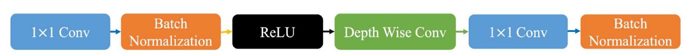

Figure 4. Osteoarthritis dimensionality reduction block.

3.5. Dataset

The knee osteoarthritis initiative (OAI) dataset is classified into five KL grades, with grade 0 being a healthy knee and grade 4 being the worst Knee osteoarthritis (KOA) stage. A sample of extracted knee joints with their respective grades is shown in Fig. 5. There are a total of 8260 x-ray images from 4796 test subjects. The distribution of images for each of the grades for test validation and training set is given below in Table 1.

3.6. Confusion Matrix

A confusion matrix can be used to measure the performance of classification problems in Machine Learning (ML). The following parameters are used in the confusion matrix.

Figure 5. Extracted knee joints for each grade.

TABLE 1. OAI DATASET SUMMARY

Grade	Training	Validation	Test	Total
Grade 0	2286	328	639	3253
Grade 1	1046	153	296	1495
Grade 2	1516	212	447	2175
Grade 3	757	106	223	1086
Grade 4	173	27	51	251
Total	5778	826	1656	8260

1) **True Positive (TP):**TP is the result of the model predicting positive class correctly.
2) **True Negative (TN):** The result TN anticipating the negative class precisely.
3) **False Positive (FP):** FP is concluding positive class inaccurately.
4) **False Negative (FN):** FN results from the model anticipating negative class wrong.
5) **Senstivity:** Sensitivity or True positive rate (TPR) is the ratio that, out of all positive classes, the properly expected positive categories. It must be as high as achievable. Recall is additionally referred by TPR.

$$Recall\ (TPR) = \frac{TP}{TP + FN} \tag{1}$$

6) **Specificity:** Specificity or precision is the ratio out of all negative classes of the expected negative categories.

$$Precision = \frac{TN}{TN + FP} \tag{2}$$

7) **Accuracy:** Accuracy is calculated as the total of two correct predictions divided by the total dataset number.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

4. Experimental Results

Experimental results from knee joint extraction are shown below in Fig. 6. We have acquired that map of 95.3% for detecting knee joints using YOLOv5. For classification, we have acquired the top grading classification accuracy of 78.93% on the OAI dataset. We have shown qualitative results of knee joint detection in Fig. 6.

A confusion matrix is shown in Fig. 7.

We have a specificity of 78.4% and a sensitivity of 79.4%. The proposed methodology not only worked well for true positives but also for true negatives.

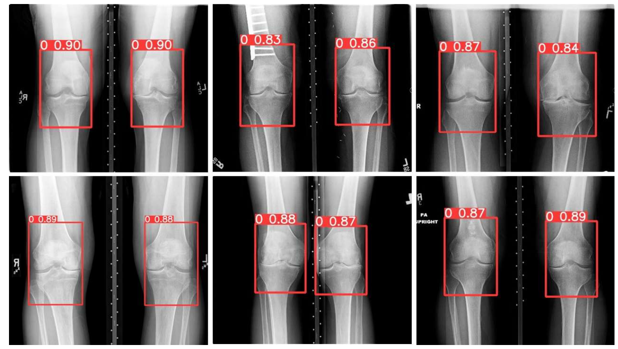

Figure 6. Knee joint detection qualitative results.

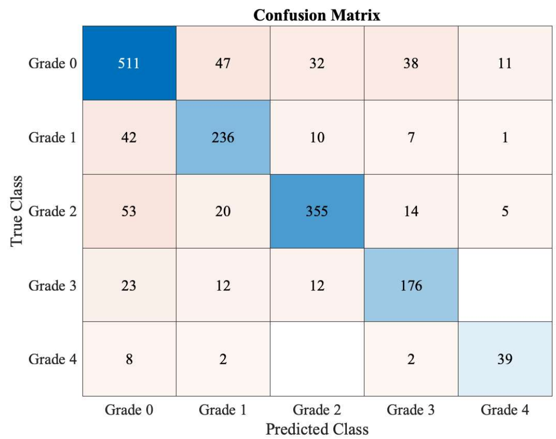

Figure 7. Confusion matrix for test data.

Results with comparisons to other algorithms are given in Table 2.

TABLE 2. COMPARATIVE RESULTS

Paper	Fully Automated	Dataset	Accuracy on OAI
[3]	No	OAI	71.74%
[4]	Yes	OAI	69%
[5]	No	Local	68%
[6]	No	Local	70.3%
[7]	No	RCU	73%
Proposed Blocks	**Yes**	**OAI**	**78.93%**

The next section will present the conclusion of all the work.

5. Conclusion

We proposed a fully automated knee osteoarthritis grading methodology for the osteoarthritis initiative dataset. We proposed a framework in which first knee areas are extracted using a stage detector YOLOv5 to detect knee joints to fully automate the process instead of manually. Then, we used two blocks for feature extraction and dimensionality reduction on extracted knee joints and graded them using

KL grading. We acquired an overall map for knee joint detection of over 95% and classification accuracy of 78.93%, which is a significant improvement compared to state-of-the-art methodologies.

References

[1] F. Saeed, A. Humayun, S. M. Fatima, V. Junaid, H. Imtiaz, M. Zehra, A. Zahid, A. Channa, A. I. Meherally, Z. Z. Shah, and A. Hoosseny, “The pressing need to raise awareness about osteoarthritis care among elderly females in pakistan: a cross-sectional study,” *Cureus*, vol. 11, no. 8, Aug 2019.

[2] “National Database for Autism Research,” https://nda.nih.gov/.

[3] N. Gayatri, M. M. Naveed, S. Nagaraju, S. C. Anumandla, J. Chandupatla, and H. Priya, “Knee osteoarthritis grading using osteo hr net model,” *Knee*, vol. 13, no. 1, 2023.

[4] Y. Wang, X. Wang, T. Gao, L. Du, and W. Liu, “An automatic knee osteoarthritis diagnosis method based on deep learning: data from the osteoarthritis initiative,” *Journal of Healthcare Engineering*, vol. 2021, pp. 1–10, 2021.

[5] W. Li, Z. Xiao, J. Liu, J. Feng, D. Zhu, J. Liao, W. Yu *et al.*, “Deep learning-assisted knee osteoarthritis automatic grading on plain radiographs: the value of multiview x-ray images and prior knowledge,” *Quantitative Imaging in Medicine and Surgery*, vol. 13, no. 6, p. 3587, 2023.

[6] C. W. Grote, M. J. Mackay, Q. Lu, X. Liu, A. R. Meyer, and J. Wang, “A whole-joint histopathologic grading system for murine knee osteoarthritis,” *Journal of Orthopaedic Research®*, vol. 41, no. 7, pp. 1407–1418, 2023.

[7] A. Khalid, E. M. Senan, K. Al-Wagih, M. M. A. Al-Azzam, and Z. M. Alkhraisha, “Hybrid techniques of x-ray analysis to predict knee osteoarthritis grades based on fusion features of cnn and handcrafted,” *Diagnostics*, vol. 13, no. 9, p. 1609, 2023.