Deep Neural Networks for Automatic Classification of Knee Osteoarthritis Severity Based on X-ray Images

Rima Tri Wahyuningrum*

Department of Informatics
Engineering, Universitas Trunojoyo
Madura, Bangkalan, Indonesia
rimatriwahyuningrum@trunojoyo.ac.id

Achmad Yasid

Department of Information Systems,
Universitas Trunojoyo Madura,
Bangkalan, Indonesia
ayasid@trunojoyo.ac.id

Gijsbertus Jacob Verkerke

Department of Rehabilitation
Medicine, University of Groningen,
University Medical Center Groningen;
Department of Biomechanical
Engineering, University of Twente,
the Netherlands
g.j.verkerke@umcg.nl

ABSTRACT

Knee Osteoarthritis (KOA) is a type of chronic disease that commonly occurs in older, obese citizens and those with a sedentary lifestyle. This disease causes damage to knee cartilage and causes pain so that the patient's activity is reduced. Radiologists classify the KOA severity based on Joint Space Narrowing (JSN) and the presence or absence of osteophytes into five stages from healthy knee (stage 0) to the worst damage (stage 4). We developed a methodology that aims to accelerate the classification of KOA severity based on information obtained from X-ray images and to reduce the subjectivity of radiologists. This paper describes an automated KOA diagnostic model using hyper-parameter Deep Convolutional Neural Networks (DCNN). Based on our experimental result, it shows the accuracy of the proposed method outperforms other KOA severity classification algorithms, which also discussed deep learning, namely 77.24%. This value is the average result of the accuracy of each fold from each stage of the KOA severity level where we use three-folds cross validation as a method of evaluating system performance. Thus computationally, this method is efficient in automatic diagnosis and has the potential to be a clinician application aid to specify the KOA severity.

CCS CONCEPTS

• **Computing methodologies**; • **Object recognition**; • **Deep learning**, **Neural Networks**;

KEYWORDS

KOA, deep convolutional neural networks, hyper-parameter

ACM Reference Format:

Rima Tri Wahyuningrum*, Achmad Yasid, and Gijsbertus Jacob Verkerke. 2020. Deep Neural Networks for Automatic Classification of Knee Osteoarthritis Severity Based on X-ray Images. In *2020 The 8th International Conference on Information Technology: IoT and Smart City (ICIT 2020)*, December 25–27, 2020, Xi'an, China. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3446999.3447020>

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

ICIT 2020, December 25–27, 2020, Xi'an, China

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-8855-9/20/12...\$15.00

<https://doi.org/10.1145/3446999.3447020>

1 INTRODUCTION

Arthritis is an inflammation that occurs in the joints which causes joint stiffness and pain. This disorder can be experienced by all ages, both old, young and by children. One type of arthritis is osteoarthritis (OA) which is caused by thinning or damaged cartilage. This disease can occur with different etiologies, but results in the same biological, morphological and clinical outcomes. OA usually attacks joints such as knee, hip, spine and ankle. Yet, knee OA (KOA) is the OA which affects most the quality of life and independency of citizens around the world [1]. KOA has become the most common cause of disability for middle age citizens and older. Because there is no known cure for reversing the knee osteoarthritis, hence early diagnosis is very important for the patient to prevent further development by making behavior and lifestyle changes [2].

In determining this early diagnosis a modality is needed to specify KOA severity. Some of the modalities used by experienced clinicians include radiography (X-ray), MRI, CT, ultrasound and nuclear medicine. However, radiography is still the gold standard modality because it is cheap, not invasive and widely used [3], [4]. The evaluation of the KOA severity is generally based on the Kellgren-Lawrence (KL) grading system, which is characterized by measurements of Joints Space Narrowing (JSN), sclerosis of the subchondral bone and osteophyte formation [5]. In the system, the KOA severity is divided into five grades or stages, namely stage 0 indicating healthy knee, stage 1 representing doubtful OA, stage 2 showing mild OA, stage 3 denoting moderate OA and stage 4 signifying severe OA [1], [3], [6], [7]. An example of each KOA stage is shown in Figure 1

The common radiograph image diagnosis is very dependent on the radiologist's manual procedure, thus making this diagnosis rather subjectively. In addition, the manual processing method is inefficient, and takes much time [8]. Thus, this approach is not suitable for use in hospitals with large numbers of patients. Hence, there is a need for a computer-assisted diagnostic tool that can reliably and automatically diagnose the KOA severity. In previous research, several methods in Machine Learning have been developed to automatically classify the KOA severity based on digital information obtained from radiograph images [9], [10], [11]. However, these studies have limitations, including the classification accuracy achieved is not optimal and the data used is not too much. Currently, researchers continue to develop new approaches to overcome these limitations, namely Deep Learning. This approach is

Figure 1: An example of each KOA stage of a left knee

part of the latest Machine Learning method that allows learning features directly from data and is able to generate thousands of knee radiographic images that have been assessed by the radiological committee. Several studies have demonstrated success in predicting the development of knee osteoarthritis and automating the KL grade classification using deep neural networks [7], [10], [12], [13], [14], [15].

Antony et al. [10] proposed Sobel horizontal image gradients as the features followed by Support Vector Machine (SVM) to detect the knee joint centers. By using 200 positive and 600 negative training samples, then randomly dividing the samples into 70% as training and 30% as testing data the results obtained with SVM were 95.2% for five-fold cross validation and a test accuracy of 94.2%. In addition, this study also measured the KOA severity automatically using loss regression and CNN. The Mean Square Error (MSE) results showed an accuracy of 57.6% in multiclass classification of the KOA severity (grades 0-4). Hereafter, Antony et al. [12] improved the results of previous CNN-related studies by applying an algorithm to localize knee joints using a Fully Convolutional Network (FCN) and measured knee OA severity thru a network jointly trained for multi-class classification and regression where both networks were trained from scratch. By using 70% of the data as training and 30% of the data as validation, the results of multi-class accuracy were 60.3%. Furthermore, Chen et al. [7] developed fully automatic knee OA severity grading using deep neural networks with a novel ordinal loss. They detected the knee joint using a customized one-stage YOLOv2 network and performed a knee KL assessment using VGG-19 model so that the best classification accuracy was obtained at 69.7%.

Tiulpin et al. [13] developed an efficient neural network architecture that studies highly relevant disease features using ResNet-34 pre-adapted and pre-trained on ImageNet dataset. By using the KOA public database, namely MOST as training data (67%), with Osteoarthritis Initiative (OAI) as validation data (11%) and testing data (22%), the result of the multiclass classification accuracy (grade 0-4) was 66.71%. Meanwhile, Gorriz et al. [14] applied trainable attention modules to detect localization of joints and VGG-16 as learning algorithm to classify knee OA severity. By normalizing the intensity level of X-ray images using histogram equation and

augmentating data by doing horizontal right-left flips more training data were produced. Then, by separating the 70-30 train-test data used and 10% of the training data stored for validation, an accuracy of 64.3% was obtained. Furthermore, Wahyuningrum et al. [15] proposed a modification of the VGG-16 + LSTM with three-fold cross validation, which showed an accuracy of 75.28%. Previously, we have investigated several famous DCNN models in literature for image classification such as VGG family, Res-Net, Dense-Net, and other models [7], [8], [13], [15], [16]. These models have been successfully applied in the classification of the severity of knee OAI. However, our conclusion expressed that the established models still underperforming because the fine detail of the images is low. Therefore, we developed a hyper-parameter models that was optimized for higher accuracy and better generalization of the data [17]. In this study, we present this hyper-parameter model where the find-tuned approach is employed to learning rate, number of epochs, and dropout rate.

2 DATA

The dataset implemented in this study was gained from OAI. This dataset provides radiographic images of bilateral PA fixed flexion knee OA (X-ray and magnetic resonance), but the focus of this study is using X-ray images. All images are converted from DICOM to JPG format. We used 1579 male and female participants, aged between 50-79 years and 45-79, respectively, in whom all images had knee OA severity stages. In detail, the stages for the left and right knees are distributed separately as follows: stage 0-265, stage 1-605, stage 2-297, stage 3-320, and stage 4-92. This is unbalanced so we used a method of data augmentation, which is useful for balancing the data distribution of each class. Some ways to balance the unbalanced data distribution with the augmentation method are geomatric transformation (rotation), flipping, cropping, color space, color space transformation, translation, and noise injection [18].

We used horizontal flipping and rotation 5 degrees clockwise as data augmentation methods randomly. The data augmentation process is only carried out on training data. The overall data after the data augmentation process included 4737 images. In addition, we also carried out oversampling techniques as a way to analyze

Table 1: Distribution of data after augmentation of KOA severity between training and testing data

Method	Train					Test				
	KL0	KL1	KL2	KL3	KL4	KL0	KL1	KL2	KL3	KL4
Fold 1	875	1105	970	1055	205	88	200	105	100	34
Fold 2	885	1086	975	1030	260	85	198	90	96	32
Fold 3	860	1009	990	1045	305	89	202	99	107	31

complex data using cross validation [19]. In this case, to compare with our previous study, we used three-fold cross validation. Table 1 shows the distribution of the knee X-ray images conditioned on the KL assessment scheme after data augmentation.

3 METHODS

In this section, we explain the methods clearly to show the details phase of this study.

3.1 Normalization and Knee Joint Detection

To increase the accuracy of the classification results, normalization of image intensity and dimensions is implemented before entering data into the convolutional neural network. In this study, we used CLAHE as a method for normalizing image intensity and cropping to 400 x 100 pixels at the center of the knee joint on both the right and left knee to normalize the image dimensions. The purpose of cropping is to get a suitable Region of Interest (ROI).

3.2 The Proposed Deep Neural Network Architecture

In this study, Deep Convolutional Neural Networks (DCNN) was employed to characterize narrowing distinctive features of the joint space and to measure osteophyte values from X-ray images. The DCNN image classification technique has two stages, feature extraction and classification. To extract features the soft-max layer was used for classification of training image data in an end-to-end learning mechanism.

Figure 2 shows the process of KOA classification used in this study. Meanwhile, we also created the DCNN architecture as shown in Figure 3. The dataset is divided into training and testing data, because the classification technique is supervised learning. Then the five sets of convolution layers are constructed to get feature extraction networks. Furthermore, each of the convolution layers comprises a Re-LU and max-pooling layer. The detailed network architecture of this method can be seen in Figure 3. It shows the kernel, the number of filters and the pool size of max-pooling layer. Next, for classification networks a flattening layer is followed by a fully connected layer, Re-LU layer, and soft-max layer with five nodes to match KL stage (stage 0, 1, 2, 3, and 4). To avoid overfitting during training, we set a dropout value of 0.1 and recorded the soft-max function to get the probability result of the output layer. In addition, we use Adam as optimizer and Cross-Entropy function to get loss value each epoch with a constant learning speed of 0.001. Every CNN model was trained in 30 epochs with batch sizes of 100.

3.3 Performance Evaluation

The three-fold cross validation method was applied to evaluate the effectiveness of the learned classification model. At first, folds were created by splitting the data set into three. Then the procedure was applied iteratively by holding one-fold as testing data and the two others were trained by deep learning. The distribution composition of three-fold cross validation is shown in Table 1. Furthermore, three test scores were calculated using an average method to report the performance.

4 EXPERIMENTAL RESULTS AND DISCUSSION

In the section, the experimental results that have been gained with different DCNN architectures are presented. The proposed DCNN model has been applied in the training and testing dataset to classify five stages of the KOA severity level. Figure 4 shows the accuracy of the classification results obtained from our DCNN architecture using the three-fold cross validation evaluation method. In addition, we also calculate the average accuracy of each stage for each fold. Finally, we calculate the average result for each fold, which is 77.24%. From the graph in Figure 4 illustrates that the mean of the three folds shows stages 0 and 4 are higher than the other stages. This happens because the X-ray image on that stage is clearly visible. In stage 0, a healthy knee is indicated by a clearly loose knee joint, meaning that the distance between the femur and tibia is wide and there is no osteophyte. The opposite is stage 4, the worst knee OA, which shows a very narrow knee joint that has osteophytes so that the patient experiences pain and difficulties in moving.

Based on the experimental results, it shows that the method proposed in this manuscript outperforms the current method in terms of classification accuracy for KOA severity as shown in Table 2. Table 2 provides a comparison of the accuracy, loss functions, training and testing time (second) of the proposed method with the state-of-the-art KOA deep learning methods. Even though the loss function for our method is slightly higher than the previous method, this is not a problem. Because what is needed in this study is the exactitude of the classification accuracy results. In addition, the time required for training and testing data in this study was shorter than in previous studies. Thus, the methodology developed has the potential to be a useful tool in KOA clinical trials and can provide better quantitative information to clinicians in a systematic means.

5 CONCLUSION

Deep learning is an emerging trend in computer science and health care. This method is part of machine learning that is able to ensure the effectiveness and availability of feature maps with a greater number of hidden layers so that the classification accuracy results are better. In this study, we have developed DCNN by using hyper-parameter with fine-tuned approach to classify the KOA severity based on X-ray images. For several system tests, it turns out that the hyper-parameter that produces the best accuracy is the five sets of convolutional layers where each layer consists of Re-LU and max pooling with a kernel filter and several parameters as outlined in the network architecture design that has been built. The average accuracy result with three-fold cross validation system performance

Figure 2: The process of KOA classification based on hyper-parameter model

Figure 3: Network architecture of our proposed DCNN model

Table 2: Experimental results from several deep learning methods for classification of KOA severity

Method	Accuracy (%)	Loss Function	Training Time (second)	Testing Time (second)
Antony et.al [12]	63.40	0.66	423.67	15.74
Tiulpin et.al [13]	66.71	0.48	150.67	3.8
Lim et.al [16]	71.97	0.20	266.67	14.63
Chen et.al [7]	69.70	0.34	364.67	25.43
Wahyuningrum et.al [15]	75.28	0.09	510	39.37
Proposed method	77.24	0.15	457	28.53

Figure 4: The KOA severity accuracy results for each stage and the average accuracy of each fold

evaluation is 77.24%. This score outperformed the accuracy in the same study. In the future research, we plan to append some semi-supervised learning methods to subtract the model's reliance on high-quality training data.

ACKNOWLEDGMENTS

The authors would like to thank the OAI to supply data in this study. In addition, this work has been supported by Grant Sponsor: International Collaborative Research, DIPA Universitas Trunojoyo Madura in 2020 with Grant no: 156/UN6.4.1/PT.01.03/2020.

REFERENCES

- [1] M. J. Lespasio, N. S. Piuze, M. E. Husni, G. F. Muschler, A. Guarino, and M. A. Mont, "Knee Osteoarthritis: A Primer," *Perm. J.*, vol. 21, pp. 1–7, 2017, doi: 10.7812/TPP/16-183.
- [2] X. Ji and H. Zhang, "Current Strategies for the Treatment of Early Stage Osteoarthritis," *Front. Mech. Eng.*, vol. 5, no. September, pp. 57–64, 2019, doi: 10.3389/fmech.2019.00057.
- [3] D. Hayashi, F. W. Roemer, and A. Guermazi, "Imaging for osteoarthritis," *Ann. Phys. Rehabil. Med.*, vol. 59, no. 3, pp. 161–169, 2016, doi: 10.1016/j.rehab.2015.12.003.
- [4] A. Tiulpin and S. Saarakkala, "Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks," *Osteoarthr. Cartil.*, vol. 28, no. April, p. S308, 2020, doi: 10.1016/j.joca.2020.02.480.
- [5] L. Shamir *et al.*, "Knee X-ray image analysis method for automated detection of osteoarthritis," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 2, pp. 407–415, 2009, doi: 10.1109/TBME.2008.2006025.
- [6] J. H. Kellgren and J. S. Lawrence, "RADIOLOGICAL ASSESSMENT OF OSTEOARTHRITIS," *Ann. Rheum. Dis.*, vol. 16, no. 4, pp. 494–502, 1957, doi: http://dx.doi.org/10.1136/ard.16.4.494.
- [7] P. Chen, L. Gao, X. Shi, K. Allen, and L. Yang, "Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss," *Comput. Med. Imaging Graph.*, vol. 75, pp. 84–92, 2019, doi: 10.1016/j.compmedimag.2019.06.002.
- [8] K. A. Thomas *et al.*, "Automated Classification of Radiographic Knee Osteoarthritis Severity Using Deep Neural Networks," *Radiol. Artif. Intell.*, vol. 2, no. 2, p. e190065, 2020, doi: 10.1148/ryai.2020190065.
- [9] L. Anifah, I. K. E. Purnama, M. Hariadi, and M. H. Purnomo, "Osteoarthritis Classification Using Self Organizing Map Based on Gabor Kernel and Contrast-Limited Adaptive Histogram Equalization," *Open Biomed. Eng. J.*, vol. 7, no. 1, pp. 18–28, 2013, doi: 10.2174/1874120701307010018.
- [10] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran, "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks," *Proc. - Int. Conf. Pattern Recognit.*, vol. 23, pp. 1195–1200, 2016, doi: 10.1109/ICPR.2016.7899799.
- [11] L. Minciullo, P. A. Bromiley, D. T. Felson, and T. F. Coates, "Indecisive Trees for Classification and Prediction of Knee Osteoarthritis," in *Lecture Notes in Computer Science (Machine Learning in Medical Imaging)*, no. September, 2017, pp. 283–290.
- [12] J. Antony, K. McGuinness, K. Moran, and N. E. O'Connor, "Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks," *Lect. Notes Comput. Sci. Int. Conf. Mach. Learn. Data Mining, MLDM 2017, New York, USA*, vol. 10358, no. 2013, pp. 376–390, 2017, doi: 10.1007/978-3-319-62416-7.
- [13] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, "Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-20132-7.
- [14] M. Górriz, J. Antony, K. McGuinness, X. Giró-i-Nieto, and N. E. O'Connor, "Assessing Knee OA Severity with CNN attention-based end-to-end architectures," *Proc. Mach. Learn. Res.*, vol. 102, pp. 197–214, 2019, [Online]. Available: http://arxiv.org/abs/1908.08856.
- [15] R. T. Wahyuningrum, L. Anifah, I. K. Eddy Purnama, and M. Hery Purnomo, "A New Approach to Classify Knee Osteoarthritis Severity from Radiographic Images based on CNN-LSTM Method," *2019 IEEE 10th Int. Conf. Aware. Sci. Technol. iCAST 2019 - Proc.*, pp. 528–533, 2019, doi: 10.1109/ICAwST.2019.8923284.
- [16] J. Lim, J. Kim, and S. Cheon, "A deep neural network-based method for early detection of osteoarthritis using statistical data," *Int. J. Environ. Res. Public Health*, vol. 16, no. 7, p. 1281, 2019, doi: 10.3390/ijerph16071281.
- [17] M. Claesen and B. De Moor, "Hyperparameter Search in Machine Learning," in *MIC 2015: The XI Metaheuristics International Conference*, 2015, pp. 10–14, [Online]. Available: http://arxiv.org/abs/1502.02127.
- [18] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, pp. 60–108, 2019, doi: 10.1186/s40537-019-0197-0.
- [19] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araujo, and J. Santos, "Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [Research Frontier]," *IEEE Comput. Intell. Mag.*, vol. 13, no. 4, pp. 59–76, 2018, doi: 10.1109/MCI.2018.2866730.