

Classification of Degenerative Arthritis Using Xception Model in Radiographic Images

Sang-min Lee^{1,*} and Namgi Kim²

¹Kyonggi university Department of Computer Science
Suwon, South Korea

²Kyonggi university Department of Computer Science
Suwon, South Korea

*Corresponding Author: Sang-min Lee. Email: d9249@kyonggi.ac.kr, ngkim@kyonggi.ac.kr

Received: XX Month 202X; Accepted: XX Month 202X

Abstract: This paper presents a method for improving the accuracy of degenerative arthritis classification in radiographic images using the existing Xception model after deformation. The proposed model modifies the exit flow of the existing Xception model to generate a larger amount of parameters and then applies the filter of the exit flow by stacking more for effective feature extraction. This presents the possibility of further increasing accuracy through improvements in convolutional neural network models for problems with degenerative arthritis classification. We then present that the Xception model is effective for that domain through the accuracy comparison between deep networks and lightweight models by Layer.

Keywords: Knee osteoarthritis; deep learning; model lightweight; convolutional neural network

1 Introduction

The human body is composed of numerous joints that allow us to move willingly. Articular cartilage, which is the most vital tissue for maintaining normal joint function, sits between these joints to prevent the bones from directly colliding [1].

The cartilage is indispensable for body movement, and because it is continuously used in daily life, the degree of cartilage wear increases with age, causing degenerative arthritis and movement discomfort.

Degenerative arthritis (also called osteoarthritis) is characterized by chronic pain, stiffness, and restrictions in joint range of motion, as well as localized articular cartilage degeneration and joint deformation. Degenerative arthritis progresses faster as more cartilage is used, and thus in Korea, more than 80% of patients over the age of 55 and almost all elderly patients over the age of 75 are diagnosed with knee osteoarthritis through radiation tests. If the cartilage suffers from severe wear and tear, it will cease to function normally, necessitating surgical and non-surgical treatment. Non-surgical treatment includes methods such as physical therapy,



exercise therapy, and medication. Surgical treatments such as knee osteotomy and artificial joint replacement are performed in cases where the pain is too severe to maintain a lifestyle, the pain does not improve even after six months of non-surgical treatment, the structural deformation or instability of the joint is severe, or restrictions in range of motion are severe. However, there is a risk of infection during surgical treatment, which could necessitate re-surgery. Particularly in the case of artificial joint replacement treatment, side effects such as pain and function failure can be accompanied by structural instability after treatment, and the artificial joint's life span is limited to approximately 15 years, necessitating a prudent decision. Owing to the limitations of surgical treatments, there is an increasing demand for oriental medical treatment, which has fewer side effects and is determined based on the individual physical characteristics. Therefore, the scientific and logical evidence for oriental medical diagnosis and treatment of knee osteoarthritis grows, and it is clear that standardized diagnosis and treatment methods that can be used in clinical settings must secure the evidence [2].

The medical community requires a scientific and logical basis in the process of diagnosis. Therefore, it is expected that if the standardized diagnosis method can be used, doctors will be able to use more objective indicators to make faster and more accurate diagnoses, and diseases caused by degenerative arthritis may be prevented owing to the improved method.

The existing procedure for diagnosing osteoarthritis is to first take radiographic images, as shown in Fig. 1, and then to consider the resulting images, which show the gradual reduction of joint spacing and deepening shadows of the bones under the cartilage, as well as the questionnaire administered to patients, when determining the grade of osteoarthritis. Because this procedure allows for the diagnosis of osteoarthritis based on visual changes similar to those shown in Fig. 2, studies to determine the grade of osteoarthritis have been widely conducted using accumulated knee radiology images to train a deep learning model. Although deep learning models based on these visual changes enable information transfer to objective indicators, more extensive research has also been performed aimed at accuracy as the accuracy of osteoarthritis classification needs to be improved for practical use.

In this study, the **K-L grade**, the most used rating index for cartilage wear in knee joints, is used and divided into five stages, as shown in Fig. 1 and Tab. 1. No hospital is allowed to share patients' radiology images without their permission owing to privacy laws. Additionally, each hospital has different radiographic image results and training data distribution. Furthermore, for the model to immediately learn the labeled data from doctors and for the updated model to be readily available, less analyzing time is critical; however, not all hospitals are equipped with high-performance computing devices. Given these considerations, a lightweight model with fewer parameters was used to reduce training time and speed data analysis. Additionally, a model suitable for the domain was found through performance comparisons in the same environment with other lightweight models, and an Xception variant model demonstrating improved accuracy to the osteoarthritis classification problem was proposed.

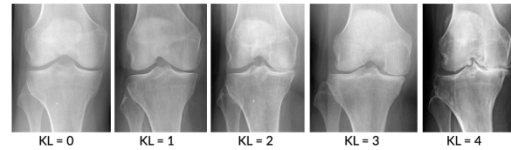


Figure 1: Stage of knee osteoarthritis [3]

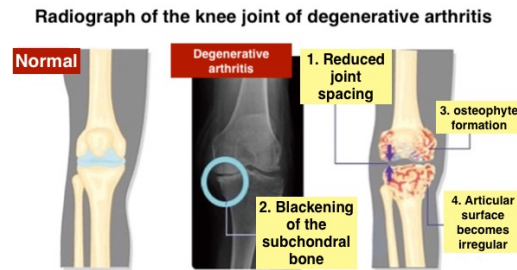


Figure 2: Visual symptoms of osteoarthritis [4]

Table 1: Kellgren Lawrence Rating Indicators [5]

Grade	Description
Grade 0 (none)	definite absence of x-ray changes of osteoarthritis
Grade 1 (doubtful)	doubtful joint space narrowing and possible osteophytic lipping
Grade 2 (minimal)	definite osteophytes and possible joint space narrowing
Grade 3 (moderate)	moderate multiple osteophytes, definite narrowing of joint space and some sclerosis and possible deformity of bone ends
Grade 4 (severe)	large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends

2 Related Work

Deep learning models tend to expand in size continuously for high accuracy, and deep learning models such as VGGNet require increased computation time and proportionately increased training time because they consist of deep layers of the network for high performance. Increased training time may increase in energy consumption. Recent models require more computation and training time owing to the large number of parameters [6].

However, studies using deep network models have demonstrated no significant increase in predictive accuracy, and re-training the models when required in a limited environment or where additional training data are continuously growing has been difficult. To improve this approach, models such as DenseNet have been studied to efficiently use parameters rather than deepen the network, but the accuracy of those studies demonstrated no noticeable improvement in performance, only exceeding around 70%.

Other approaches that have been actively investigated include transforming training data for augmentation or using pre-processing to reduce redundant information in radiology images. In a related study, Yolo-V2 models were used as a Joint Detector in radiographic image joint detection; through pre-processing, cartilage was detected in radiographic images, and only the joint part of the image was cropped from the entire knee radiographic image. VGG-19 models with 1.44 million parameters were used for classification, and the overall test image showed an accuracy of 69.58% [7].

A fully connected network (FCN) was used for detection in a study on reducing parameters, and a classification model with approximately 540,000 parameters was fabricated through an analysis of the number of convolution layers and other parameters, which was the most suitable for knee osteoarthritis classification. The model's performance after training was approximately 63.5% [8].

In a study that used data augmentation and lightweight models, the DenseNet model with an architecture that learns only small portions of an entire image was predicted to show robust performance on the K-L grade classification problem. First, both the DenseNet-169 and Inception-V3 models were used for training, but the Inception-V3 model was excluded owing to poor performance. ImageNet was used as a pre-training of the DenseNet-169 model to solve the problem of lack of training data. Crop, upscale, add noise, flip, and randomize contrast were successively applied to 80% of the training data, and only crop was applied to the remaining 20%. Training based on augmented data showed performances above 71%. Given that the accuracy of radiologists on the same test dataset was above 61%, the study verified that the accuracy of the deep learning model exceeded that of radiologists in osteoarthritis grade diagnosis [9].

SqueezeNet, Xception, MobileNet, and ShuffleNet have all had a significant impact on light-weighting deep learning models. SqueezeNet [10] has a capacity of less than 0.5 MB, allowing for quick training owing to its low computation volume. Fire Module, which is composed of two layers of Squeeze and Expand, has been proposed to achieve this structure. The SqueezeNet model requires 50 times fewer parameters than AlexNet to achieve the same accuracy, making it suitable for real-time updates.

Xception [11] proposed an efficient convolutional layer called a depthwise separable convolution, which reduced the calculation volume to approximately 1/9 of the typical convolutional network model calculations, allowing the training rate to be nine times faster. A depthwise separable convolution enables filtering of each channel to extract the spatial features.

MobileNet [12] offers a lightweight structure that can be operated on mobile devices or applications by properly using the **depthwise separable convolution** structure, and has been published up to version 3 with further improved structures.

ShuffleNet [13] was light-weighted by 'shuffling' **depthwise separable convolution** results. **Channel Shuffle**, a method for solving problems arising from **pointwise group convolution** and **3×3 group convolution**, was proposed to reduce the computation volume of **1×1 pointwise convolution**. Accordingly, the model has a lower computation volume and uses more feature maps in a smaller computation volume.

There has been extensive research on improving accuracy, as well as studies on reducing training costs and time by reducing the number of parameters through **network** optimization and using augmentation to solve data shortages.

This study was designed to find an effective lightweight model for the domain to minimize training time and cost by reducing the number of parameters, and **Xception** was selected as the most appropriate lightweight model to perform an experiment among the models that can effectively extract **spatial features** through **depthwise separable convolution**, which is most extensively used for lightweight models.

3 Suggested method

The **depthwise separable convolution** was considered effective in extracting **features** for grade classification through gradual reduction in joint spacing and deepening shadows of bones under cartilage, which are progressive symptoms of osteoarthritis because it allows channel-specific **spatial features** to be extracted.

The accuracy of **Xception** was the best among the results of finding and training a deep learning model suitable for the domain through an accuracy comparison of the models configured in **Keras** and prominent lightweight models. The results are shown in Tab. 3. Accuracy was improved by changing the structure of the best Xception model, and the result of analysis based on the structure of the model indicated that the correct prediction could not be made because of information loss caused by rapidly decreasing features. As a method for extracting a well-established feature, the **convolution network** was built in depth and the **feature map** was gradually reduced through **filters**. The goal was to control the decreasing number of parameters in the end-of-training process. Although we used more parameters than existing models, the model was still relatively smaller in size compared with **VGGNet**, which has many layers, allowing the training to be completed faster and with fewer **parameters**.

An improved **exit flow** over the existing model consists of **kernel sizes 3×3 and 10×10 size**. The **network** is composed of several convolutional network layers, each with a different number of filters. As the number of filters decreases gradually, so does the number of **channels**. Fig. 3 depicts the overall structure of the model.

Fig. 3 shows the modified **Xception** model. The training was divided into two stages: the first stage used **DPhi train** as **training data**, and the second stage used **Kaggle 2018 data** as **validation data**. In the second stage, **Kaggle 2018 data** were divided into two.

The training image used was $224 \times 224 \times 3$, and early stopping was set to **pre-training=imagenet**, **batch size = 8**, and **optimizer = Adam** (learning rate = 0.00001). The training was conducted on **Ubuntu 18.04, V100, TensorFlow 2.6.0, and Keras 2.0**.

The overall performance comparison among models is presented in Tab. 3. **MobileNet, ShuffleNet, SqueezeNet, and Xception**, which greatly influenced the

development of lightweight models, and additional deep learning models that use deep **network layer**s were trained in the same environment. The performance was compared with the deep learning models described in previous studies.

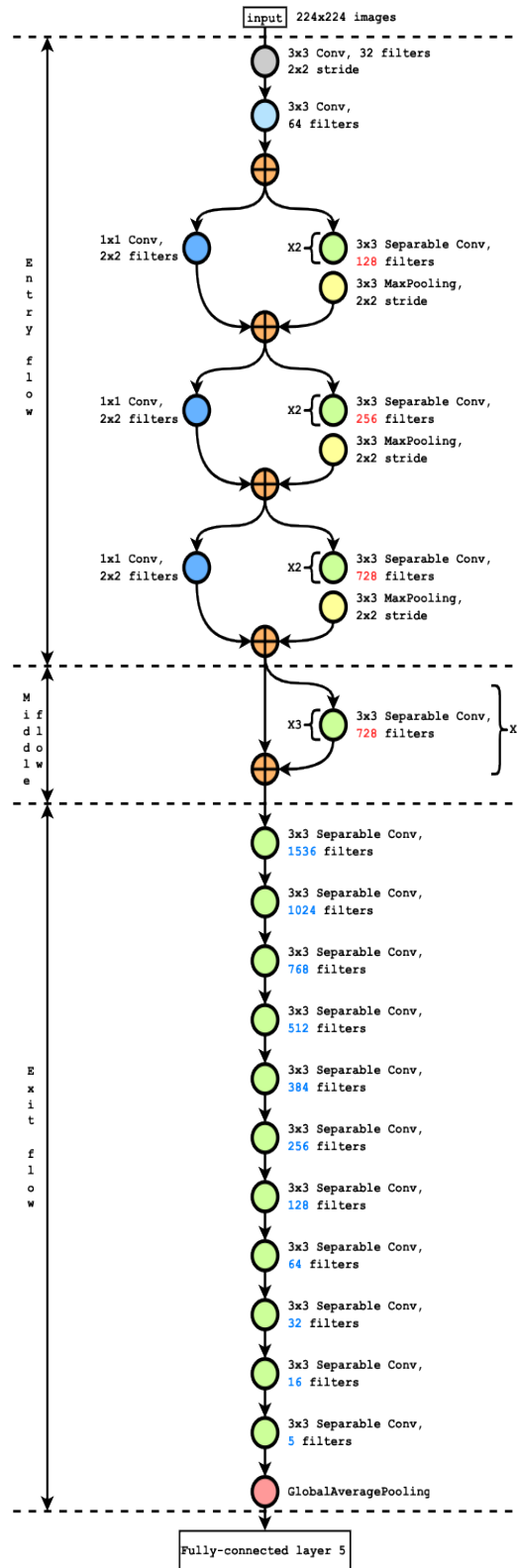


Figure 3: Stage of knee osteoarthritis

4 Experiment

4.1 Experimental method

The detailed composition of the dataset used in this experiment is presented in Tab. 2, and the training and test data provided by DPhi were used for training and testing. The validation data were not provided; however, Kaggle data [14] were used instead.

As shown in Tab. 2, and Figs. 4 and 5, the training and test data imbalance is very severe. To solve this problem, classes with little data were weighted and used as training data.

Table 2: Used dataset configuration

	Gra de 0	Gra de 1	Gra de 2	Gra de 3	Gra de 4
Kaggle 2018 (Fig. 2.)	385 7	177 0	257 8	128 6	295
DPhi train (Fig. 4.)	308 5	141 6	206 2	102 9	236
DPhi test	1958				

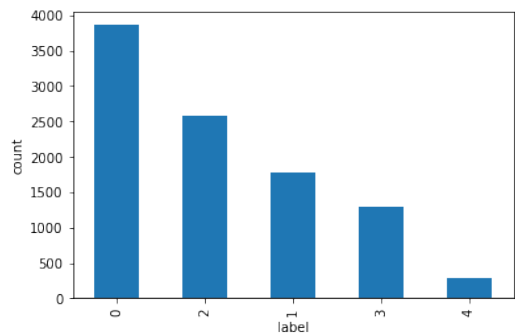


Figure 4: Kaggle 2018 Dataset configuration

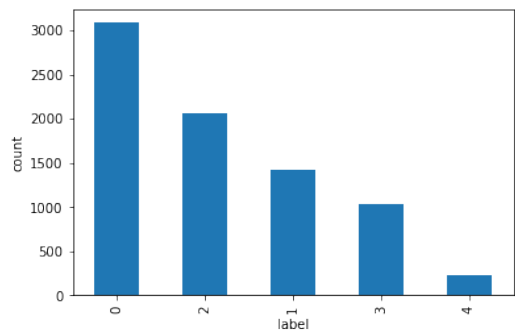


Figure 5: DPhi Train Dataset configuration

4.2 Experimental result

For the proposed Xception model performance evaluation, we participated in the DPhi's "Data Sprint #35: Osteoarthritis Knee X-ray" competition [15] and used the criteria provided in the competition for the prediction accuracy of the test dataset. Additionally, the knee arthritis dataset [6] provided by Kaggle's "Knee Osteoarthritis Dataset with KL Grading - 2018" was used as verification data, and the performance of the model was assessed based on the confusion matrix (Figs. 6 and 7).

The accuracy of the trained model was calculated using DPhi's test dataset, and both experiments were conducted in the same environment. While our proposed model used the highest number of parameters, the Xception model delivered the best performance (72.2%) compared with the models used in other studies or the lightweight models. The modified model used approximately 2.5 times more parameters than the previous model, but demonstrated a significant improvement in accuracy of 79%, which was increased by 7%.

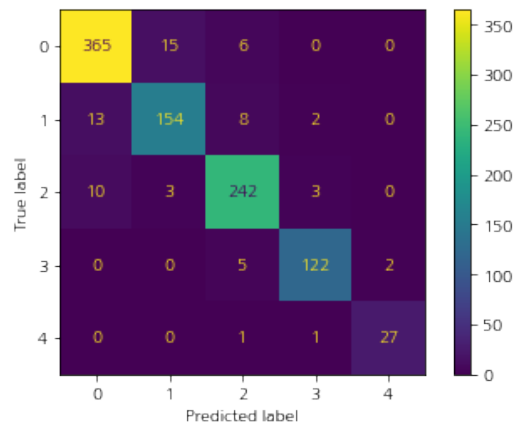


Figure 6: Xception validation data result, at Kaggle 2018 dataset

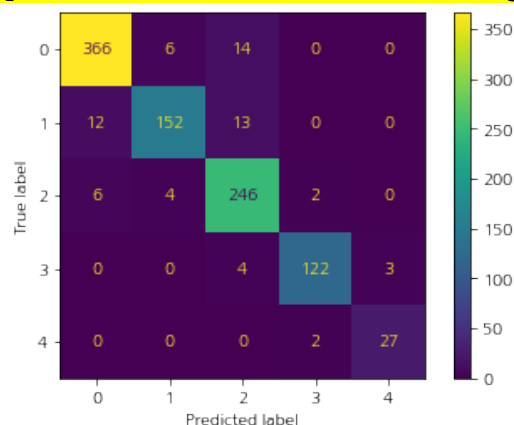


Figure 7: Improved Xception validation data result, at Kaggle 2018 dataset

Table 3: Accuracy result

Model	Parameter	Accuracy
-------	-----------	----------

DenseNet121	6,958,981	55.46
DenseNet169	12,492,805	55.76
DenseNet201	18,102,533	66.85
EfficientNetB0	4,013,953	39.68
EfficientNetB1	6,519,589	39.17
EfficientNetB2	6,519,589	38.51
EfficientNetB3	10,703,917	38.51
ResNet50	23,544,837	65.17
ResNet101	42,563,077	50.56
ResNet152	58,229,765	40.60
SqueezeNet [7]	737,989	39.43
Xception [8]	20,899,127	72.22
MobileNet [9]	3,212,101	40.50
MobileNetV2	2,230,277	40.60
MobileNetV3S mall	1,522,981	40.60
MobileNetV3La rge	4,208,437	44.18
ShuffleNet [10]	918,125	43.05
Ours*	56,755,255	79.62

5 Conclusions

This study proposed to change the exit flow of the Xception model to classify osteoarthritis grades. The proposed model is more accurate than other training models and the pre-modification model, with a 79.62% accuracy rate. Future studies must include a mathematical calculation analysis of the improved models or network improvements. Additionally, different knee joint radiology image datasets may be used to analyze the performance of the proposed Xception model more clearly and quantitatively.

Acknowledgement: Acknowledgement, Funding statement, Conflicts of Interest and References headings should be left justified, bold, not numbered, with the first letter capitalized. Sections following those headings are formatted as the main text. Authors should thank those who contributed to the article but cannot include themselves.

Funding Statement: This research was conducted with the support of the Korea Research Foundation, and funded by the Ministry of Education of Korea in 2020.

(No. 2020R1A6A1A03040583)

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] https://www.snubh.org/dh/main/index.do?DP_CD=JRC&MENU_ID=002008024
- [2] S. Oh, E. Jang, Y. Oh, W. Kang, E. Lee, et al., "An exploratory study on the reliability of degenerative knee arthritis dialectics and the correlation between knee ROM, VAS, and WOMAC," *Journal of Korean Medicine Rehabilitation*, vol. 29, no. 3, pp. 113-128, 2019.
- [3] N. Bayramoglu, M. T. Nieminen, and S. Saarakkala, "A lightweight CNN and joint shape-joint space (JS2) descriptor for radiological osteoarthritis detection," in Papież, B., Namburete, A., Yaqub, M., Noble, J. (eds) *Medical Image Understanding and Analysis. MIUA 2020. Communications in Computer and Information Science*, vol 1248. Cham, Switzerland: Springer, pp. 331-345, 2020.
- [4] <https://www.100ssd.co.kr/news/articleView.html?idxno=70447>
- [5] V. Pai and H. Knipe, "Kellgren and Lawrence system for classification of osteoarthritis," Reference article, Radiopaedia.org. <https://radiopaedia.org/articles/kellgren-and-lawrence-system-for-classification-of-osteoarthritis>
- [6] K. Lee, E. Kim, and Intelligent Infrastructure Technology Research Group, "Deep Learning Model Lightweight Technology Analysis," *Korea Institute of Science and Technology Information*, 2020.11
- [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," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 84-92, 2019
- [8] 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," in Perner, P. (ed) *Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science, MLDM 2017*, vol. 10358, Cham, Switzerland: Springer, pp. 376-390, 2017.
- [9] K. A. Thomas, Ł. Kidziński, E. Halilaj, S. L. Fleming, G. R. Venkataraman, et al., "Automated classification of radiographic knee osteoarthritis severity using deep neural networks," *Radiology: Artificial Intelligence*, vol. 2, no. 2, e190065, 2020
- [10] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, et al., "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," ICLR 2017
- [11] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800-1807, 2017.
- [12] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017 CVPR
- [13] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6848-6856, 2018.
- [14] Knee osteoarthritis dataset with KL grading - 2018, <https://www.kaggle.com/datasets/tommyngx/kneeo>
- [15] Data spring #35: Osteoarthritis knee X-ray, <https://dphi.tech/challenges/data-sprint-35-osteoarthritis-knee-x-ray/81/overview/about>