IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 10, No. 5, October, 2023, pp. 1849-1856
Journal homepage: http://iieta.org/journals/mmep

An Automated System for Osteoarthritis Severity Scoring Using Residual Neural Networks

Aeri Rachmad[1*], Fifin Sonata[2], Juniar Hutagalung[2], Dian Hapsari[3], Muhammad Fuad[4], Eka Mala Sari Rochman[1]

[1] Department of Informatics, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan 69162, Indonesia
[2] Department of Informatics Management, STMIK Triguna Dharma, Medan 20146, Indonesia
[3] Department of Informatics Engineering, Institut Teknologi Adhi Tama Surabaya, Surabaya 60117, Indonesia
[4] Department of Mechatronics, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan 69162, Indonesia

Corresponding Author Email: aery_r@trunojoyo.ac.id

https://doi.org/10.18280/mmep.100538

Received: 2 May 2023
Revised: 12 August 2023
Accepted: 10 September 2023
Available online: 27 October 2023

Keywords:
osteoarthritis, knee image, Residual Neural Network (ResNet-101)

ABSTRACT

Osteoarthritis (OA) is a chronic disease, characterized by progressive deterioration of cartilage tissue and consequent thinning of the cartilage layer within joints. This degradation leads to an increased likelihood of bone collision during movement, typically manifesting in patients as joint pain, knee swelling, stiffness, and difficulties in executing daily activities. The diagnosis of OA often involves the analysis of physical examination results, patient anamnesis, and additional supportive examinations, which are predominantly conducted manually. Addressing these challenges, this study harnesses Convolutional Neural Network (CNN) algorithms, specifically the Residual Neural Network and Mobile Neural Network architectures, to develop an automated system for classifying OA severity. Utilizing a knee image dataset comprised of 8260 records procured from NDA OAI, the model is trained and tested with a data split of 80% and 20% respectively. The Residual Neural Network (ResNet-101) architecture is employed for model training, utilizing Adam optimization with a learning rate set at 0.0001 over 50 epochs. The resulting model yields a training accuracy of 67.65%, and a validation accuracy of 57.06%. This study demonstrates the potential of CNN methods for automated, accurate classification of OA severity using knee imagery, thus offering a promising avenue for enhancing diagnostic efficiency and precision.

1. INTRODUCTION

Osteoarthritis (OA) is a chronic disease characterized by the progressive deterioration of cartilage tissue, leading to the thinning of the cartilage layer in joints [1-3]. This process results in bones rubbing against each other during movement, with common symptoms including joint pain, knee swelling, stiffness, and difficulty in conducting daily activities. Factors such as age, genetics, gender, excess body weight, joint injuries, growth disorders, among others, have been identified as triggers for OA [4, 5]. Consequently, this disease predominantly affects elderly and obese individuals, targeting major weight-bearing joints such as the genu, lumbar, coxa, and cervical.

The diagnosis of OA involves a manual analysis of physical examination results, anamnesis, and various supportive examinations. This manual approach, while necessary, is time-consuming, which can be detrimental given the progressively degenerative nature of OA [4]. Early detection is crucial to mitigate the disease's progression to more severe stages, making it imperative to expedite the diagnostic process. As such, the application of advanced computational techniques, particularly Deep Learning, for the automatic classification of knee X-ray images is seen as a promising solution [2].

The Convolutional Neural Network (CNN) algorithm, a deep learning technique that mimics complex human neural networks, has demonstrated exceptional performance in image-based data classification processes [6-8]. Previous studies, such as Yamashita et al. [9], have attested to the efficacy of the CNN method in medical research, stating its potential to enhance the performance of radiologists and improve patient handling efficiency. Furthermore, Sitaram and Dessai [10] reported a remarkable accuracy of 97.27% in their study on cervical MR image classification using the ResNet-101 CNN.

Motivated by these findings, this study aims to facilitate the diagnosis of OA by utilizing various CNN architectures, namely ResNet-50, ResNet-101, ResNet-152, and MobileNet, for the classification of digital knee images. The comparison of accuracy results across these architectures will aid in determining the optimal CNN architecture for OA severity classification.

2. BACKGROUND STUDY

This study focuses on the classification of Osteoarthritis severity through the examination of knee X-ray images. Image classification refers to the process of classifying image objects into distinct categories based on specific characteristics that define each class. The construction of a highly accurate model for classifying digital image-based data necessitates the use of an effective algorithm. In this study, the Convolutional Neural Network (CNN) method, employing the ResNet and

MobileNet architectures, is utilized for classifying input images. The performance of these architectures is subsequently evaluated using a confusion matrix.

2.1 Convolutional neural networks (CNN)

Deep learning methodologies are frequently employed for pattern recognition, classification, and feature extraction tasks due to their ability to resolve issues in machine learning systems using multilevel concepts [11]. In other words, these methods operate on multiple layers, with each layer performing distinct functions. The CNN algorithm, a component of deep learning, is particularly suited for tasks involving digital image-based data due to its high network depth [12]. CNN algorithms replicate the complex structure of human neural networks, comprising several interconnected layers [13, 14]. Additionally, these algorithms employ a mathematical operation known as convolution to process calculations within the network. The operational structure of this method, comprising multiple network layers, is visualized in Figure 1 below.

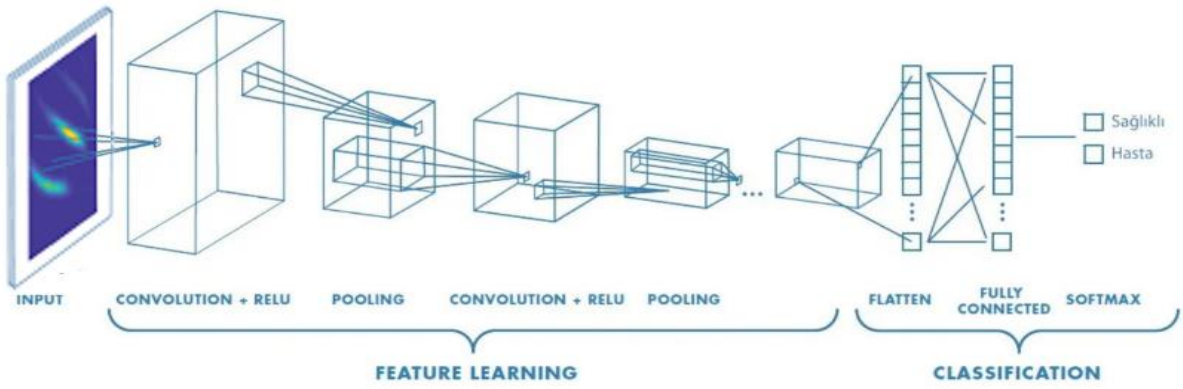


Figure 1. CNN illustration

In Figure 1 above, the basic CNN architecture is divided into 2 main stages, namely Feature Learning and Classification [15]. The feature learning stage is the process of encoding the input image to transform it into a feature or feature that is represented as a value. Meanwhile, the classification stage is the core of the process for categorizing the results of feature learning based on the characteristics of its features to produce the required output. In addition, the CNN architecture has four main layers namely, convolutional layer, pooling layer, flatten layer, and fully connected layer [16]. Each of these layers will be explained as follows:

(1) Convolutional layer

The convolutional layer is the first layer to receive input from image data directly on the network architecture [17]. In this layer, there is a convolution operation that plays a role in the feature extraction process of the input image which is formulated as in Eq. (1) below.

$$s(t) = -(x * w)(t) \quad (1)$$

where,

s(t)=the result of the convolution operation in the form of a function

x=input

w=kernel (weight)

(2) Pooling layer

In the pooling layer, there is a process of reducing the size of the matrix which is carried out after the convolution operation. The output generated during the pooling process is in the form of a matrix with a smaller size compared to the initial matrix [17]. Two types of pooling are commonly used in the CNN method, namely max pooling, and average pooling. Max pooling takes the maximum value from the input matrix, whereas in average pooling the value to be taken is the average value of the initial input matrix.

(3) Flatten layer

The output of the feature map process is still a multidimensional array; therefore, a stage is needed before entering the fully connected layer. This stage is a flattening process or converting the feature map into vector form with the aim that the output of the feature map can be used as input to the fully connected layer [16].

(4) Fully connected layer

In this layer, all activation nerves from the previous layer will be connected to the nerves in the next layer so that it is like a normal artificial neural network and at this layer, the output of the classification process is also produced [18].

Several types of CNN architectures are commonly used in classifying digital images to optimize the accuracy of the built models, as follows:

i. Residual Neural Network (ResNet)

ResNet was first introduced in 2015 as one of the CNN architectures which has several types of layers where the highest level of network depth is 152 layers [19]. The high level of depth in the CNN architecture provides an important role in building the CNN model, which can increase the accuracy of the system. This architecture aims to map the features of the input image data that work by passing through several layers to avoid gradient loss conditions, the working principle is known as the skip connection [20].

ii. Mobile Neural Network (MobileNet)

MobileNet is a simple architecture that uses convolutions where the convolutions can be deeply separated to form compact deep convolutional neural networks. The structure of this architecture is based on a kernel or filter that can be split in depth and this architecture divides the convolution into depthwise and pointwise convolution as illustrated in Figure 2 below [21].

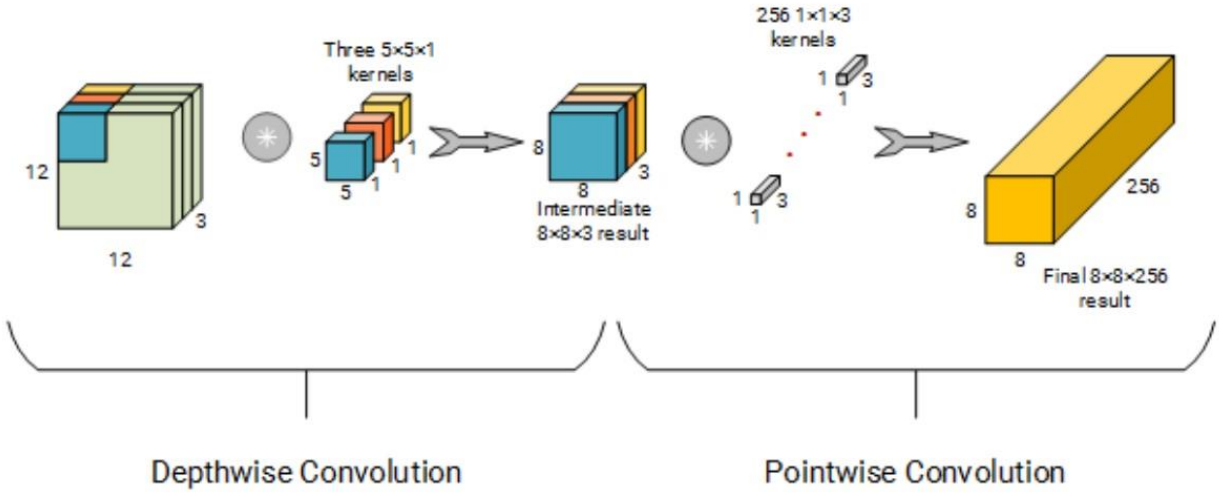


Figure 2. MobileNet architecture

Depthwise convolution is a reduced version of the convolution itself where each channel will go through a separate process. A convolution with a size of height x width x channel feature input map will be divided into several groups whose number depends on the number of channels indicating the depth of the network. Meanwhile, pointwise convolution is the opposite of depthwise convolution where the depth of the network depends on the number of input channels [22].

2.2 Evaluation

The system that carries out the classification process is expected to be able to classify all data sets correctly, so it requires an evaluation of system performance to find out whether the classification results obtained provide good predictions [23]. In this study to determine the performance of the model is done by calculating the accuracy using the confusion matrix and loss calculations.

2.2.1 Accuracy

Calculation of accuracy in this study using the confusion matrix method. The confusion matrix is a tool that can be used to determine the correctness of a system [24]. The confusion matrix contains information from the actual and predicted classification results at the time of text classification [25]. The performance of the system is evaluated using data in a matrix which is illustrated in Figure 3 below.

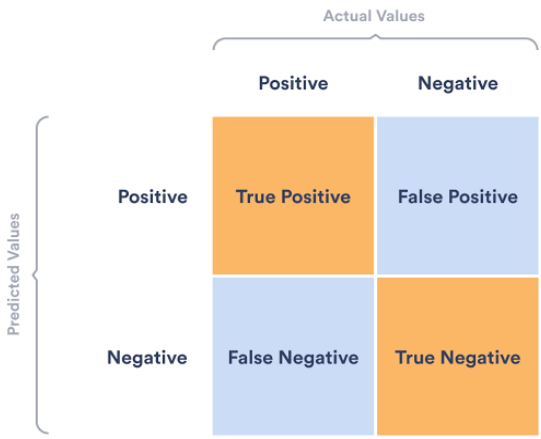


Figure 3. Confusion matrix

Table 1 shows a confusion matrix for evaluating the classification process into positive and negative classes.

Table 1. Target class description

Grade 0	Not detected Osteoarthritis is characterized by no signs of osteoarthritis (green line).
Grade 1	Doubtful Osteoarthritis is characterized by osteophyte formation and joint narrowing of the knee.
Grade 2	Mild Osteoarthritis detected is characterized by the formation of osteophytes (blue) and the possibility of joint space narrowing.
Grade 3	Moderate osteoarthritis was detected, indicated by the presence of many osteophytes (blue) and joint space narrowing sclerosis (purple).
Grade 4	Severe osteoarthritis was detected, which means that there are many enlarged osteophytes, characterized by severe joint space narrowing sclerosis.

Based on the confusion matrix table, the accuracy value can be calculated using the equation below [7].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

where,

TP: All data that is correctly predicted into the positive class,

TN: All data with the original class is positive but the prediction results are negative,

FN: All data that is correctly predicted into the negative class,

FP: All data with the original class is negative but the prediction results are positive.

2.2.2 Loss

The loss calculation plays a role in calculating the amount of data that is lost or not detected during the training and validation process.

3. MAIN RESULTS

3.1 Data collection

The following is an example of the dataset used in this study.

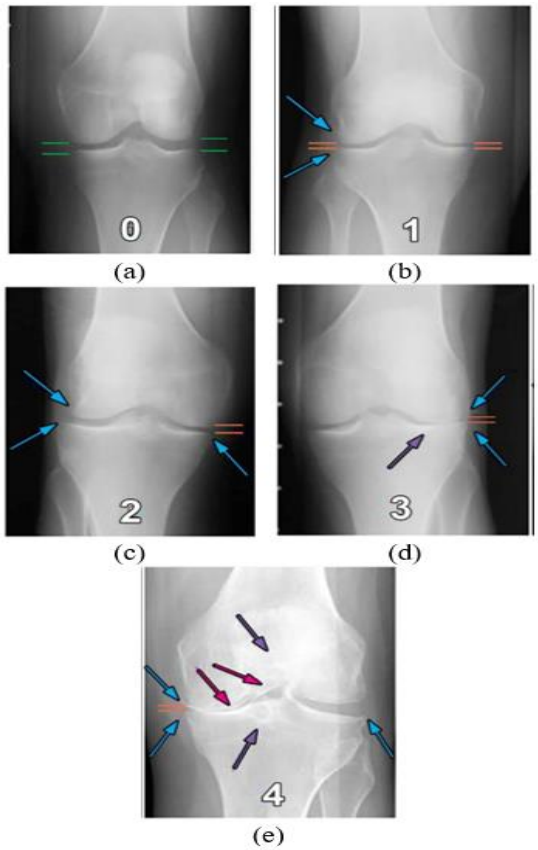


Figure 4. Classification of knee oa severity with the kellgren-lawrence standard

The data used in this study is data collected in the form of digital images of the knee taken from NDA The Osteoarthritis Initiative (OAI) through the NIMH Data Archive website https://nda.nih.gov/oai. which totaled 8260 records with five target classes namely Grade 0 (3253 records), Grade 1 (1495 records), Grade 2 (2175 records), Grade 3 (1086 records), and Grade 4 (251 records). Figure 4 above shows the five target classes in knee OA images, where the severity of the five classes is classified based on the Kellgren-Lawrence standard which is determined based on the degree of osteophytes, narrowing of the space in the joint space, and changes in bone structure. The table below will describe each of these target classes.

3.2 Analysis

This section describes the flow of the research process in classifying digital images of OA of the knee using the Convolutional Neural Network algorithm. The stages of the process will be shown in the Figure 5.

This research includes the following stages:

(1) Data input process

The data input process is the initial process carried out to input the image dataset used in this study, namely in the form of X-Ray images on the knee as many as 8260 records with 5 classes of severity of knee Osteoarthritis.

(2) Data sharing process

The next process is data sharing using the split data method to divide the image data set into training data and test data. The training data plays a role during the process of modeling the classification system and the training data plays a role in evaluating the model. The comparison ratio used in this study was 80% for training data and 20% for test data.

(3) Classification process

At this stage, a learning process is carried out to obtain a classification model using the CNN algorithm with different CNN architectures, namely ResNet-152, ResNet-50, ResNet-101, and MobileNet. Classification models with different architectures will be compared with the accuracy results to obtain optimal classification results.

(4) Outputs

The results of this classification process will categorize the input image into several levels of knee OA severity classes based on the modeling that has been done with the method applied in this study.



Figure 5. The representation of the process flow applied in this study

4. RESULTS AND DISCUSSIONS

In the results and discussion section, the results of the classification that have been carried out on 8260 records of knee OA images will be explained using the CNN method by first dividing the dataset using the data splitting method at a comparison ratio of 80% training data and 20% test data. The parameters used are Adam optimization with a learning rate of 0.0001 and an epoch of 50 iterations. Furthermore, for network architecture, this study uses 4 different types of CNN architectures, and the modeling plots for each architecture are shown in the following Figure 6, Figure 7, Figure 8 and Figure 9.

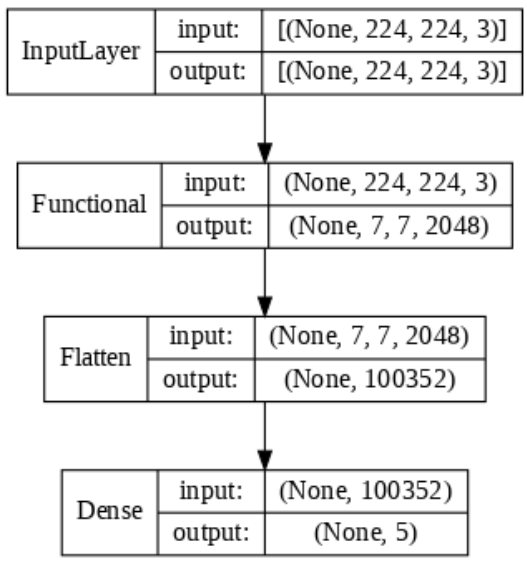


Figure 6. Plot model of ResNet-152

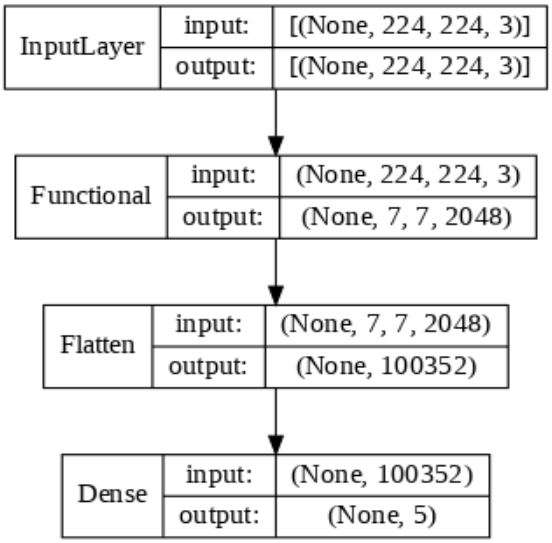


Figure 7. Plot model of ResNet-50

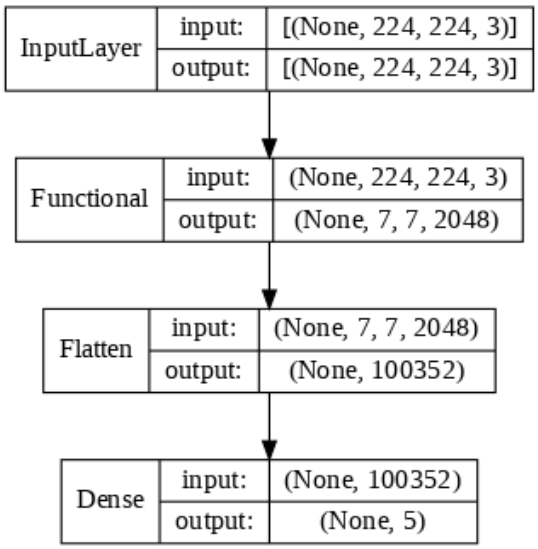


Figure 8. Plot model of ResNet-101

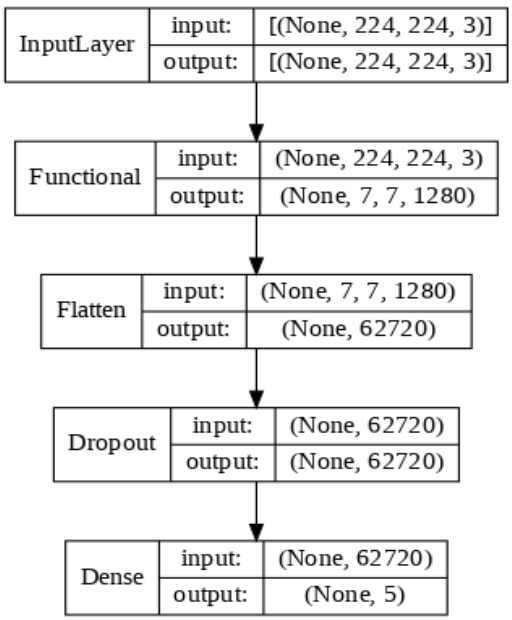


Figure 9. Plot model of MobileNet

Based on the four model architectures as shown in Figures 6, 7, 8, and 9, the accuracy is shown in the Figure 10.

4.1 ResNet-152 architecture accuracy graph

Figure 10 shows the results of the classification accuracy on knee OA images using the ResNet-152 architecture where the blue line representing the accuracy in the training process moves up at each epoch towards 66.50%, but the accuracy in testing (val_accuracy) decreases and only reaches 56.75% which is illustrated by the yellow line on the chart. Meanwhile, Figure 11 is a graph of the loss between the training and testing (validation) processes which shows a significant difference where the loss results in the training process are getting closer to 0, while the loss value in the test data has a stagnant value at a value of 1 since the 3rd epoch 50 so that the accuracy in this testing process has a lower value than during the training process. This can be triggered by a lack of data in the testing process which refers to the data sharing process.

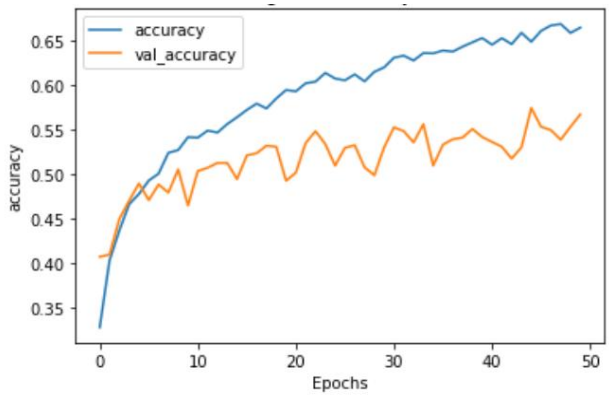


Figure 10. ResNet-152 accuracy chart

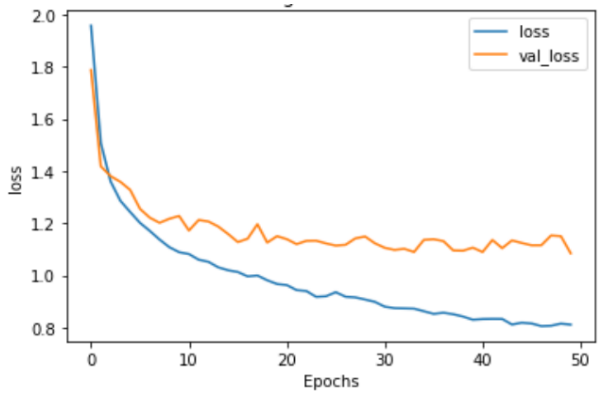


Figure 11. ResNet-152 loss chart

4.2 ResNet-50 architecture accuracy graph

Based on the results of the classification process that has been carried out, the accuracy is shown in Figures 12 and 13. Figure 12 shows a graph of accuracy in training and testing which has quite a large difference in values. Accuracy in the training process reaches 66.52%, while for testing it is only 54.57%. Then, Figure 13 shows the loss value in this experiment is 0.8019 in training and 1.1011 in the training process. So, it can be said that the model built has not been able to recognize the data properly at the time the test was carried out.

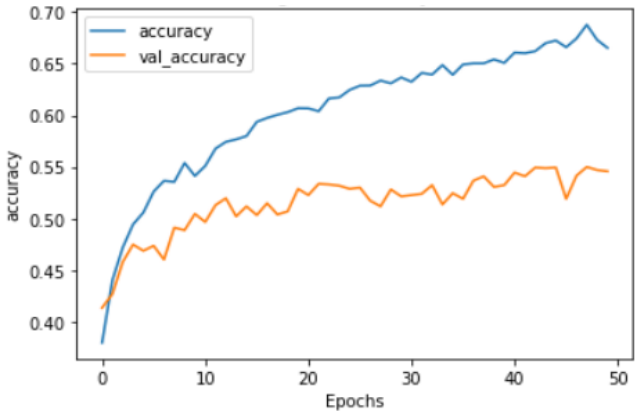


Figure 12. ResNet-50 accuracy chart

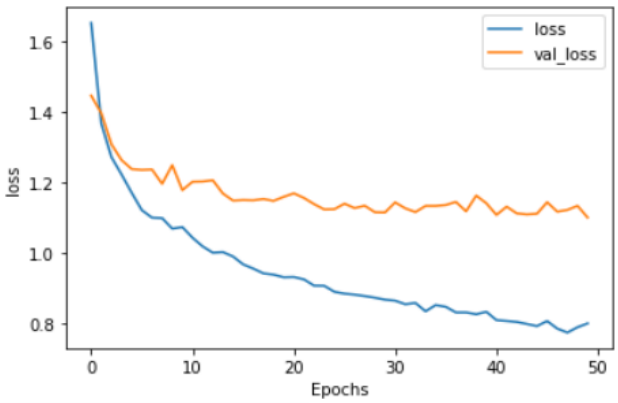


Figure 13. ResNet-50 loss chart

4.3 ResNet-101 architecture accuracy graph

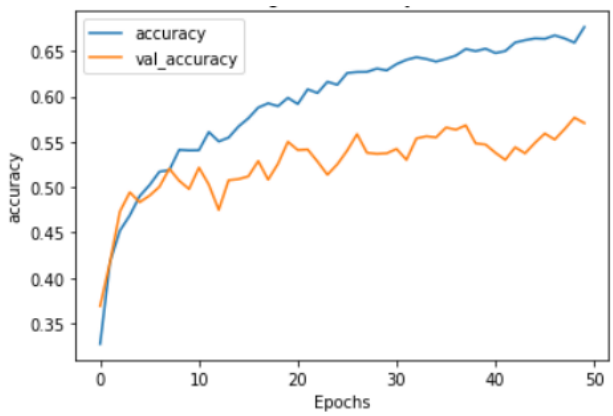


Figure 14. ResNet-101 accuracy chart

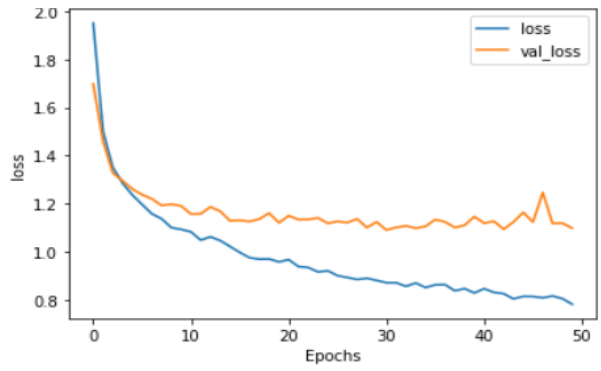


Figure 15. ResNet-101 loss chart

In the accuracy chart above, it is shown that the use of the ResNet-101 architecture provides an accuracy of 67.65% in the training process, and in the testing process, the accuracy decreases as shown in Figure 14 which is 57.06%. Meanwhile, the loss value in the training and testing process is shown in the graph in Figure 15 above.

4.4 MobileNet architecture accuracy graph

Figure 16 shows the accuracy of the training obtained using the MobileNet architecture is 49.01% and the accuracy of the test is 52.57%.

In the application of this architecture, there is a difference with other architectures, namely the accuracy value in the testing process is better than in the training process where which indicates that the system can recognize new input data better after learning is done in the training process. Thus, the loss value obtained is smaller during the testing process as shown in Figure 17.

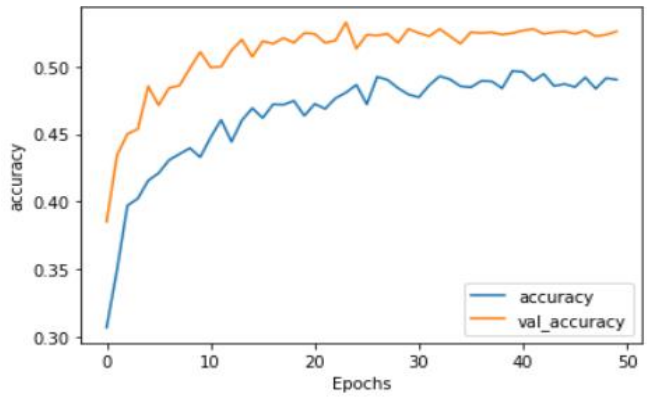


Figure 16. MobileNet accuracy chart

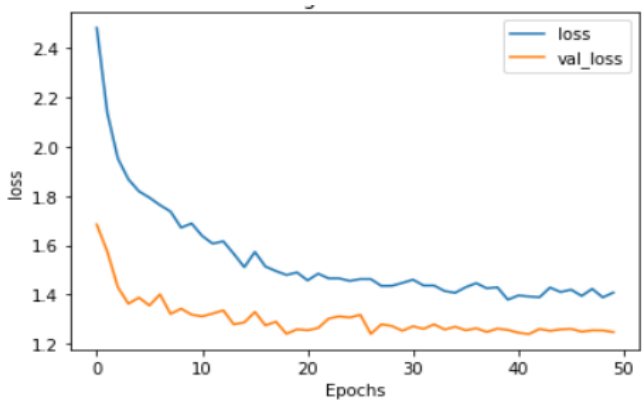


Figure 17. MobileNet loss chart

Based on the four accuracy graphs described above, a comparison can be made in terms of the resulting accuracy and the loss value given in classifying knee OA image data shown in Table 2 below.

Table 2. Comparison of four CNN architectures

CNN Architecture	Accuracy		Loss	
	Training	Validation	Training	Validation
ResNet-152	66.50%	56.75%	0.811	1.085
ResNet-50	66.52%	54.57%	0.801	1.101
ResNet-101	67.65%	57.06%	0.782	1.098
Mobile-Net	49.01%	52.57%	1.407	1.247

The differences in accuracy that resulted for each CNN architecture are shown in Table 2. Because of the table, the ResNet-101 engineering gives better precision results contrasted with different models, to be specific preparation exactness coming to 67.65% and val_accuracy of 57.06%. However, the test's accuracy has increased compared to the training, but not to the same degree as ResNet-101's. This indicates that the MobileNet architecture cannot be ignored. In contrast, the accuracy of the testing process has decreased significantly for the other three architectures in comparison to the accuracy of the training process, allowing for over-fitting during the classification process.

5. CONCLUSIONS

The study aimed to classify the severity of knee osteoarthritis using four distinct Convolutional Neural Network (CNN) architectures: ResNet-50, ResNet-101, ResNet-152, and MobileNet. The dataset comprised 8260 knee OA images, categorized into five target classes. Based on the results and ensuing discussion, the following conclusions can be drawn:

a. The ResNet-50, ResNet-101, and ResNet-152 architectures demonstrated higher accuracy during the training phase compared to the validation phase. The loss graphs for these architectures displayed signs of overfitting, as evidenced by a point where the validation loss increased while the training loss approached zero.

b. In contrast, the MobileNet architecture exhibited higher validation accuracy than training accuracy, suggesting that this architecture more effectively recognized input image objects following the training process. However, the accuracy did not surpass the result obtained by ResNet-101, which achieved a validation accuracy of 67.65% and a training accuracy of 57.06%.

ACKNOWLEDGMENT

We thank the Foundation for Exploration and Local Area Administration, College of Trunojoyo Madura, which has given research financing as far as possible. One of us might want to thank the Workforce of Designing, College of Trunojoyo Madura for permitting specialists to finish this exploration in the Sight and sound research center. We additionally thank the Interactive Media Lab Multimeida group for the fulfillment of this exploration.

REFERENCES

[1] Li, G., Yin, J., Gao, J., Cheng, T.S., Pavlos, N.J., Zhang, C., Zheng, M.H. (2013). Subchondral bone in osteoarthritis: Insight into risk factors and microstructural changes. Arthritis Research & Therapy, 15: 1-12. https://doi.org/10.1186/ar4405

[2] Uthman, I., Raynauld, J.P., Haraoui, B. (2003). Intra-articular therapy in osteoarthritis. Postgraduate Medical Journal, 79(934): 449-453. https://doi.org/10.1136/pmj.79.934.449

[3] Hassanali, S.H., Oyoo, G.O. (2011). Osteoarthritis: A look at pathophysiology and approach to new treatments, a review. East African Orthopaedic Journal, 5(2): 51-57. https://doi.org/10.4314/eaoj.v5i2.72398

[4] Mohammed, A., Alshamarri, T., Adeyeye, T., Lazariu, V., McNutt, L.A., Carpenter, D.O. (2020). A comparison of risk factors for osteo-and rheumatoid arthritis using NHANES data. Preventive Medicine Reports, 20: 101242. https://doi.org/10.1016/j.pmedr.2020.101242

[5] He, Y., Li, Z., Alexander, P.G., Ocasio-Nieves, B.D., Yocum, L., Lin, H., Tuan, R.S. (2020). Pathogenesis of osteoarthritis: Risk factors, regulatory pathways in chondrocytes, and experimental models. Biology, 9(8): 194. https://doi.org/10.3390/biology9080194

[6] Rachmad, A., Chamidah, N., Rulaningtyas, R. (2020). Mycobacterium tuberculosis images classification based on combining of convolutional neural network and support vector machine. Communications in Mathematical Biology and Neuroscience. https://doi.org/10.28919/cmbn/5035

[7] Rachmad, A., Fuad, M., Rochman, E.M.S. (2023). Convolutional neural network-based classification model of corn leaf disease. Mathematical Modelling of Engineering Problems, 10(2): 530-536. https://doi.org/10.18280/mmep.100220

[8] Setiawan, W., Rochman, E.M.S., Satoto, B.D., Rachmad, A. (2022). Machine learning and deep learning for maize leaf disease classification: A review. In Journal of Physics: Conference Series. IOP Publishing, 2406(1): 012019. https://doi.org/10.1088/1742-6596/2406/1/012019

[9] Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. Insights into Imaging, 9: 611-629. https://doi.org/10.1007/s13244-018-0639-9

[10] Sitaram, S., Dessai, A. (2019). Classification of cervical MR images using resnet101. International Journal Research in Engineering, Science and Management, 2(6): 254-257.

[11] Zhu, J.H., Munjal, R., Sivaram, A., Paul, S.R., Tian, J., Jolivet, G. (2022). Flow regime detection using gamma-ray-based multiphase flowmeter: A machine learning approach. International Journal of Computational Methods and Experimental Measurements, 10(1): 26-37 https://doi.org/10.2495/CMEM-V10-N1-26-37

[12] Coşkun, M., Uçar, A., Yildirim, Ö., Demir, Y. (2017). Face recognition based on convolutional neural network. In 2017 International Conference on Modern Electrical and Energy Systems (MEES), IEEE, 376-379. https://doi.org/10.1109/MEES.2017.8248937

[13] Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J. (2021). 1D convolutional neural networks and applications: A survey. Mechanical Systems and Signal Processing, 151: 107398. https://doi.org/10.1016/j.ymssp.2020.107398

[14] Anton, A., Nissa, N.F., Janiati, A., Cahya, N., Astuti, P. (2021). Application of deep learning using convolutional neural network (CNN) method for Womenâ€™ s skin classification. Scientific Journal of Informatics, 8(1): 144-153. https://doi.org/10.15294/sji.v8i1.26888

[15] Hema, M.S., Sharma, N., Sowjanya, Y., Santoshini, C., Durga, R.S., Akhila, V. (2021). Plant disease prediction using convolutional neural network. EMITTER International Journal of Engineering Technology, 9(2): 283-293. https://doi.org/10.24003/emitter.v9i2.640

[16] He, C., Kang, H., Yao, T., Li, X. (2019). An effective classifier based on convolutional neural network and

regularized extreme learning machine. Mathematical Biosciences and Engineering, 16(6): 8309-8321. https://doi.org/10.3934/mbe.2019420

[17] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778.

[18] Zhai, J., Shen, W., Singh, I., Wanyama, T., Gao, Z. (2020). A review of the evolution of deep learning architectures and comparison of their performances for histopathologic cancer detection. Procedia Manufacturing, 46: 683-689. https://doi.org/10.1016/j.promfg.2020.03.097

[19] Michele, A., Colin, V., Santika, D.D. (2019). Mobilenet convolutional neural networks and support vector machines for palmprint recognition. Procedia Computer Science, 157: 110-117. https://doi.org/10.1016/j.procs.2019.08.147

[20] Zaki, S.Z.M., Zulkifley, M.A., Stofa, M.M., Kamari, N.A.M., Mohamed, N.A. (2020). Classification of tomato leaf diseases using MobileNet v2. IAES International Journal of Artificial Intelligence, 9(2): 290-296. https://doi.org/10.11591/ijai.v9.i2.pp290-296

[21] Mehdiyev, N., Enke, D., Fettke, P., Loos, P. (2016). Evaluating forecasting methods by considering different accuracy measures. Procedia Computer Science, 95: 264-271. https://doi.org/10.1016/j.procs.2016.09.332

[22] Rachmad, A., Syarief, M., Rifka, S., Sonata, F., Setiawan, W., Rochman, E.M.S. (2022). Corn leaf disease classification using local binary patterns (LBP) feature extraction. In Journal of Physics: Conference Series. IOP Publishing, 2406(1): 012020. https://doi.org/10.1088/1742-6596/2406/1/012020

[23] Damayanti, F., Muntasa, A., Herawati, S., Yusuf, M., Rachmad, A. (2020). Identification of Madura tobacco leaf disease using gray-level Co-occurrence matrix, color moments and Naïve Bayes. In Journal of Physics: Conference Series. IOP Publishing, 1477(5): 052054. https://doi.org/10.1088/1742-6596/1477/5/052054

[24] Solihin, F., Syarief, M., Rochman, E.M.S., Rachmad, A. (2023). Comparison of support vector machine (SVM), k-nearest neighbor (K-NN), and stochastic gradient descent (SGD) for classifying corn leaf disease based on histogram of oriented gradients (HOG). Feature Extraction. Elinvo (Electronics, Informatics, and Vocational Education), 8(1). https://doi.org/10.21831/elinvo.v8i1.55759

[25] Ubaidillah, A., Rochman, E.M.S., Fatah, D.A., Rachmad, A. (2022). Classification of corn diseases using random forest, neural network, and naive bayes methods. In Journal of Physics: Conference Series. IOP Publishing, 2406(1): 012023. https://doi.org/10.1088/1742-6596/2406/1/012023