Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.2, April 2026, pp. 1191~1199
e-ISSN: 2548-6861

Classifying the Severity of Knee Osteoarthritis on X-ray Images Using NASNet Mobile

Adelia Damayanti [1]*, Dian Candra Rini Novitasari [2]*, Lutfi Hakim [3]*, Musfiroh Musfiroh [4]*
* Matematika, Universitas Islam Negeri Sunan Ampel, Surabaya
adeliadamayanti869@gmail.com [1], diancrini@uinsa.ac.id [2], lutfihakimbungah@gmail.com [3], musfiroh.rohmati22@gmail.com [4]

Article Info

Article history:

Received 2026-01-13
Revised 2026-02-24
Accepted 2026-04-08

Keyword:

CNN,
K-fold Cross Validation,
Knee Osteoarthritis,
NASNet Mobile.

ABSTRACT

Knee osteoarthritis is a joint condition characterized by cartilage breakdown in the knee, most frequently affecting older adults. Early diagnosis of knee osteoarthritis is critical because it can delay disease progression and improve patient quality of life. This study aims to classify the severity of knee osteoarthritis using X-ray images by employing a Convolutional Neural Network (CNN) with the NASNet Mobile model. The dataset consists of 1500 knee X-ray images, divided into three classes: normal, moderate, and severe. The model was trained with various batch sizes and learning rates. The highest accuracy of 92.53% was achieved on the k-fold cross-validation dataset using a batch size of 32 and a learning rate of 0.001, yielding a sensitivity of 92.72% and a specificity of 96.32%.

I. INTRODUCTION

Knee osteoarthritis is a condition that affects the cartilage inside the knee joint. [1]. This condition involves symptoms like pain, stiffness, and deformity due to osteophyte formation, joint space narrowing, and subchondral sclerosis [2]. Knee osteoarthritis is more common in older adults [3]. As the elderly population increases, the number of cases of knee osteoarthritis also increases [4]. Therefore, early diagnosis is essential because it can reduce the progression of knee osteoarthritis in older adults and improve their quality of life [5]. Knee osteoarthritis can be diagnosed using X-ray imaging. X-ray imaging is the best alternative for detecting knee osteoarthritis due to its safety, cost-effectiveness, and wide accessibility [6]. The gold standard for determining the severity of KOA using X-ray analysis is the Kellgren-Lawrence grading system (K-L) [7].

Grading is now largely dependent on physicians' experience and judgment. This grading technique is subjective and may be impacted by various complicating factors, potentially leading to misclassification and impacting clinical decisions as well as patient recovery. Deep learning is becoming more popular in medical fields as computer technology advances and statistical theories are refined. Deep learning, which learns features directly from data, has transformed medical image analysis.

Medical image analysis has advanced, exemplified by the use of deep learning to improve the interpretation of orthopedic radiographs for disease diagnosis. Deep learning has been effectively applied in various orthopedic tasks, such as diagnosing bone tumors [8], detecting mechanical loosening in hip implants [9], and assessing osteoarthritis [10]. Convolutional neural networks (CNNs) are among the most recognized deep learning models, leveraging image convolution to extract feature maps [11]. One CNN architecture is NASNet Mobile. NASNet Mobile is designed to be computationally efficient and optimized for devices with limited resources, like mobile devices. NASNet Mobile attained a top-1 accuracy of 82.7% and a top-5 accuracy of 96.2% on the ImageNet dataset [12]. In addition, NASNet Mobile uses modular cells and blocks that can be customized to build different network architectures. This allows for flexibility in network design and optimization for specific data sets and tasks [13].

NASNet Mobile has demonstrated strong performance across classification, detection, and segmentation tasks. In a study by Adedoja et al., they found that NASNet Mobile achieved 99.31% accuracy in diagnosing plant diseases [14]. In a study on lemon quality classification conducted by Dumen et al., NASNet Mobile was found to provide an accuracy of 96.62% [15]. In addition, NASNet Mobile has demonstrated superior performance compared to other convolutional network architectures. In 2020, Ahsan et al.

compared eight CNN architectures for COVID-19 detection, including VGG16, VGG19, ResNet50, ResNet15V2, Inception ResNetV2, DenseNet201, MobilenetV2, and NASNet Mobile. The study noted that NASNet Mobile outperformed other architectures with an accuracy of 82.94% on CT-scan images and 93.94% on X-ray images [16].

In a subsequent study, Bharati et al. diagnosed COVID-19 using lung CT images by employing two NASNet models, namely NASNet Large and NASNet Mobile. In this study, NASNet Mobile achieved higher accuracy than NASNet Large (82.42% vs. 81.06%) [17]. Yilmaz et al. conducted a study to classify skin cancer using NASNet Mobile, MobileNet, and MobileNetV2. In this study, NASNet Mobile achieved the highest accuracy (82%), followed by MobileNetV2 (81.45%) and MobileNet (80.36%) [18]. In 2023, NASNet Mobile was used by Prabha et al. to classify six Cephalopoda classes and was compared with MobileNetV2 and InceptionV3. The study reported that NASNet Mobile achieved an accuracy of 89.74%, which is comparable to MobileNetV2. Meanwhile, InceptionV3 produced an accuracy of 87.12% [19].

Based on the background description, this study aims to classify the severity of osteoarthritis in knee X-ray images using a deep learning algorithm, specifically the NASNet Mobile CNN model, to support radiologists and orthopedic surgeons with automated image interpretation, thereby enhancing diagnostic accuracy and speed.

II. METHOD

A. Data

The research data is secondary data obtained from the Kaggle website [20]. The dataset is derived from the Osteoarthritis Initiative (OAI), a ten-year observational study of men and women, to understand the prevention and treatment of knee osteoarthritis [21]. This study classified knee X-ray images into three categories according to the Kellgren-Lawrence (KL) grading system: grade 0 (normal), grade 3 (moderate), and grade 4 (severe). A total of 1,251 data points were obtained: 500 in each of the grade 0 (normal) and grade 3 (moderate) classes, and 251 in the grade 4 (severe) class. The knee X-ray image samples are described in Table I.

TABLE I
KNEE X-RAY IMAGE DATA SAMPLE

Class	Data Sample	Description
Normal		Healthy knee image
Moderate		Multiple osteophytes, significant joint space narrowing, and slight sclerosis.
Severe		Large osteophytes, extensive joint constriction, and severe sclerosis.

B. K-fold Cross Validation

K-fold cross-validation is a technique used to estimate a learning model's generalization performance. It involves dividing a dataset into k-section, where each section acts as a test set, and the other sections are used for training [22]. This process is recurrent k-times, with each section serving as a test set once. The model's performance is then evaluated by averaging the results from all iterations. Therefore, this method can yield a more accurate assessment of model performance compared to traditional validation techniques that partition the data into two subsets (training and testing) [10].

C. Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a neural network commonly employed for image processing and classification. CNN is considered the most effective method for these tasks [23]. A CNN consists of three layers: the input layer, the hidden layer, and the output layer. In general, images are represented as pixel matrices, and their pixel values are provided to the input layer along with the weights and biases. The hidden layer can be a convolution layer, a ReLU layer, a pooling layer, or a fully connected layer. The output layer is fully connected. This layer is the result of classifying images into specific classes [24].

1) Input Layer

The input layer includes the input image data. CNNs have different input image size requirements across architectures. The input image size for the NASNet Mobile architecture is $224 \times 224 \times 3$. At this layer, the input image will be normalized to ensure that each feature in the data has the same numerical contribution in the classification process [25].

2) Convolution Layer

Convolution layers are the basic components of CNN architecture that perform feature extraction on images [26]. Convolution is a type of matrix multiplication used for feature extraction in input data involving a filter or kernel matrice [27]. The kernel varies in size across architectures and contains random numbers ranging from -1 to 1. The convolution layer can be calculated using Equation 1.

$$C_y = \left(\sum_{x=1}^{N} I_{x,y} \times K_{x,y} \right) + B_y \quad (1)$$

C_y is the value of the convolutional layer, N is the number of channels, i.e., the number of channels in image data; for example, an RGB image contains three channels. $I_{x,y}$ is the

pixel value in image data that is constructed into a matrix, $K_{x,y}$ is a kernel matrix, and B_y is biased.

3) Batch Normalization

Batch Normalization is a regularization method for neural networks that normalizes feature maps to have zero mean and unit variance across the batch. Additionally, the batch normalization layer helps mitigate vanishing and exploding gradients, improves the network's ability to handle diverse input data, accelerates training, and enhances generalization. It also prevents the problem of data death in the ReLU activation function and aids in weight initialization [28].

The value of batch normalization can be calculated using the following formula.

$$bn = \gamma \frac{C - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (2)$$

Where γ dan β is a learning parameter, C is the input matrix obtained from the convolution layer output, μ is the average value, σ^2 is the variance value of the input matrix C, and ε to improve numerical stability in the form of constant values.

4) Rectified Linear Unit (ReLU)

ReLU is the predominantly employed activation function in CNNs. ReLU changes negative values to zero and retains positive values, as shown in Figure 1 [28].

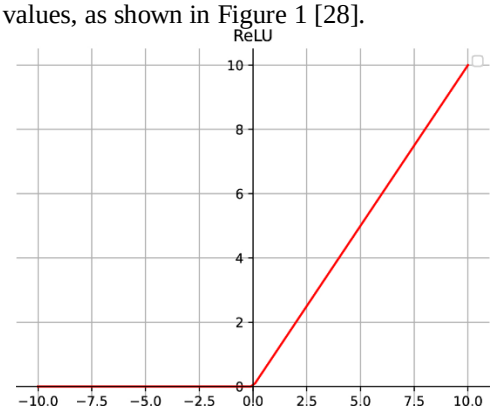

Figure 1. ReLU Graph

The ReLU function can be calculated using the following formula [29].

$$\sigma(x) = \max(0, x) \quad (3)$$

Where x is the input value and $x \in \mathbb{R}$.

5) Pooling Layer

The pooling layer decreases the size of feature maps, speeding up computation by decreasing the number of parameters to update, and helps prevent overfitting [30]. Two frequently used pooling operators are max pooling and average pooling. The process in the pooling layer is described as follows [31].

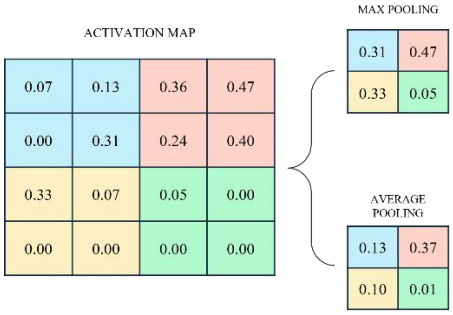

Figure 2. Pooling Process Illustration

An essential part of pooling is down-sampling. In image processing, this is similar to decreasing resolution. The number of filters remains unchanged by pooling. In max pooling, the image is partitioned into rectangular subregions, and only the maximum value in each subregion is returned. Meanwhile, in average-pooling, the average value of each subregion is returned. One of the most common subregion sizes is 2×2 [32].

6) Fully Connected Layer

The last layer of a CNN is the fully connected layer, where each neuron from the previous layer connects to all neurons in this layer to produce the classification result [33]. Operations on fully connected layers can be calculated using Equation 4.

$$q_j = \mathrm{b_j} + \sum r_{i,j} \cdot p_i \quad (4)$$

Where q_j is the value of the fully connected layer j, b_j is biased, $r_{i,j}$ is the network weight, and p_i is the input generated from the learning stage.

D. Neural Architecture Search Network Mobile (NASNet Mobile)

NASNet Mobile is a CNN architecture developed using neural architecture search (NAS), composed of basic building blocks (cells) optimized with an RNN controller [34]. NASNet Mobile was proposed in 2017 by the Google Machine Learning group and optimized for mobile applications. NASNet Mobile has fewer parameters than NASNet Large, namely 5.3 million parameters, making it more efficient and suitable for use on mobile devices or in environments with limited computing resources [13]. The NASNet Mobile architecture is shown in Figure 3.

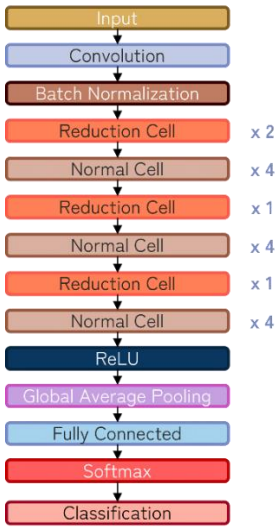

Figure 3. NASNet Mobile Architecture

This architecture features reduction cells and normal cells that produce feature maps of various types and dimensions. Reduction cells produce feature maps by lowering the input's height and width, whereas normal cells generate feature maps equal in size to the input [12].

E. Batch Size

Batch size is the number of data samples processed per iteration. It is a hyperparameter that can be adjusted to enhance model performance. Batch size significantly affects the training process and the model's overall performance [35]. Larger batch sizes can lead to faster convergence during training but require more computational resources. Conversely, smaller batch sizes may result in slower convergence but need fewer computational resources [36].

Mini-batch sizes often include 4, 8, 16, 32, 64, $\cdots$, 2^x. This is because computer memory has a size factor of two. As a result, mini-batch sizes of two can improve memory use and enable computers to allocate memory more efficiently [37].

F. Learning Rate

The learning rate is a hyperparameter in deep learning that controls the size of each step the model takes when minimizing the loss function during training [38]. A higher learning rate enables larger weight updates, thereby accelerating convergence. However, setting a learning rate that is too high can lead the optimization algorithm to overshoot the minimum and prevent convergence. On the other hand, a lower learning rate results in smaller weight updates, which can cause slower convergence but can help the optimization process reach a more stable minimum value [39].

G. Confusion Matrix

This study used a confusion matrix to evaluate the classification performance method [40]. This method entails comparing the prediction matrix to the actual class, encompassing the original data and the predicted classification results [41]. This study uses a confusion matrix to evaluate classification results in more detail.

TABLE II
MULTICLASS CONFUSION MATRIX

Actual Class	**Prediction Class**		
	Normal	**Moderate**	**Severe**
Normal	m_{11}	m_{12}	m_{13}
Moderate	m_{21}	m_{22}	m_{23}
Severe	m_{31}	m_{32}	m_{33}

The classification results in the confusion matrix are interpreted into four terms, which are described as follows [42].

1) TP (True Positive) is the amount of data included in the class m successfully classified as a class m by model. The TP value can be calculated using Equation 5.

$$TP_{all} = \sum_{j=i}^{n} m_{jj} \tag{5}$$

2) FN (False Negative) is the amount of data that should be included in the class m, but classified as a class l by model. The TP value can be determined using Equation 6.

$$FN_i = \sum_{j=1, j\neq i}^{n} m_{ij} \tag{6}$$

3) FP (False Positive) is the amount of data that should be included in the class l, but classified as a class m by model. The TP value can be determined using Equation 7.

$$FP_i = \sum_{j=1, j\neq i}^{n} m_{ji} \tag{7}$$

4) TN (True Negative) is the number of data points included in the class l and the system successfully predicted correctly as a class l. The TP value can be determined using Equation 8.

$$TN_i = \sum_{j=1, j\neq i}^{n} \sum_{k=1, k\neq i}^{n} m_{jk} \tag{8}$$

The TP, FP, TN, and FN values are then used to calculate accuracy, sensitivity, and specificity, as follows.

$$\text{Accuracy} = \frac{TP_{all}}{n_{all}} \times 100 \tag{9}$$

$$\text{Sensitivity} = \frac{\sum_{i=1}^{n} \frac{TP_i}{TP_i + FN_i}}{n} \times 100 \tag{10}$$

$$\text{Specificity} = \frac{\sum_{i=1}^{n} \frac{TN_i}{TN_i + FP_i}}{n} \times 100 \quad (11)$$

Accuracy is the proportion of instances correctly classified by a model [43]. Sensitivity is the percentage of actual positive cases that the model correctly identifies. Specificity is the proportion of negative instances correctly classified by the model [44].

H. Process Analysis

The following flowchart illustrates the classification of knee osteoarthritis severity using NASNet Mobile.

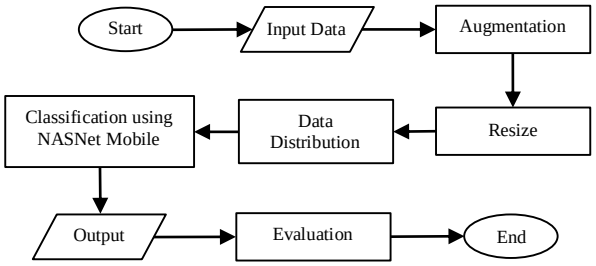

Figure 5. Research Flowchart

1) Input Data

The input data include knee X-ray images classified into three categories: grade 0 (normal), grade 3 (moderate), and grade 4 (severe).

2) Preprocessing

The data will be preprocessed, including augmentation and resizing. Data augmentation is a method in deep learning that enhances the amount and variety of training data by creating synthetic data samples from the original data [45]. Data augmentation can boost deep learning model performance by minimizing overfitting and strengthening the generalization of learned features [46]. This study augmented the grade 4 (severe) class data to increase variation and address the imbalance relative to the other two classes.

Resizing an image involves changing its size to create an image with a different dimension [47]. This study resizes images from 128×128 to 224×224 to match the input size of the NASNet Mobile model. Resizing is used to adjust the image size to the model architecture input and reduce computation time [48]. This study used the bilinear interpolation method for the resizing procedure. Bilinear interpolation resizes images by selecting the four pixels closest to the coordinates to be interpolated.

3) Data Distribution

The preprocessed data will be divided into training and testing sets. This study employs k-fold cross-validation to randomly partition the data.

4) NASNet Mobile

This study applies transfer learning using the NASNet-Mobile architecture that has been pre-trained on the ImageNet dataset. The fine-tuning technique was performed by freezing the first ten layers of the model to retain the basic features, while the remaining layers were retrained using the knee osteoarthritis dataset. Model training in this study uses learning rate and batch size. The learning rates tested are 0.001, 0.0001, and 0.00001, and the batch sizes are 4, 8, 16, 32, and 64. The model trained on the training data will be applied to the test data to evaluate its optimal classification performance.

5) Evaluation

Finally, the evaluation stage uses a confusion matrix to compute accuracy, sensitivity, and specificity, as defined in Equations 5-11.

III. RESULT

A. Preprocessing Data

The first process was to augment the severe knee OA class to balance it with the other classes. The augmentation used was a change in color contrast. The augmentation process was performed once, yielding 502 images in the severe knee OA class. The differences between the pre- and post-augmentation images are shown below.

Pre-augmentation Post-augmentation

Figure 6. Augmentation Result

Next, all data will be resized to 224×224 pixels to adjust the input data size to the NASNet Mobile model. The resized image is shown below.

128×128 224×224

Figure 7. Resize Result

The preprocessed data will be divided into training and test datasets. The training dataset will be utilized to train the NASNet Mobile model. During this process, hyperparameter tuning is conducted to determine the optimal classification parameters model.

B. Model Experiment

During training, the model was trained using 5-fold cross-validation. Several experiments were conducted with varying batch sizes and learning rates. The batch size was tested

starting from 4, 8, 16, 32, and 64. Meanwhile, the learning rates applied were 0.001, 0.0001, and 0.00001. The purpose of this experiment was to identify the optimal model to achieve the highest accuracy, sensitivity, and specificity. The model-experiment results are shown in Table III.

TABLE III
MODEL EXPERIMENT RESULT

Learning Rate	Aspect	Batch Size				
		4	8	16	32	64
0.001	Accuracy	89.60	91.79	91.59	92.53	91.72
	Sensitivity	89.71	91.81	91.71	92.72	91.67
	Spesificity	95.00	96.02	95.98	96.32	95.89
	Time	561m 54s	389m 48s	308m 28s	262m 00s	327m 28s
0.0001	Accuracy	89.80	88.50	85.33	83.53	81.50
	Sensitivity	89.77	88.56	85.94	83.35	81.29
	Spesificity	94.92	94.48	92.64	91.97	90.82
	Time	578m 27s	397m 15s	316m 20s	259m 42s	302m 31s
0.00001	Accuracy	79.93	78.20	76.89	73.93	72.87
	Sensitivity	80.06	77.97	76.44	73.92	72.57
	Spesificity	90.19	89.51	88.62	86.97	86.56
	Time	728m 18s	386m 33s	404m 58s	264m 41s	405m 30s

The test results are presented as two graphs: accuracy and computation time. This study uses the average results across all folds for each learning rate and batch size, as shown in Figure 8.

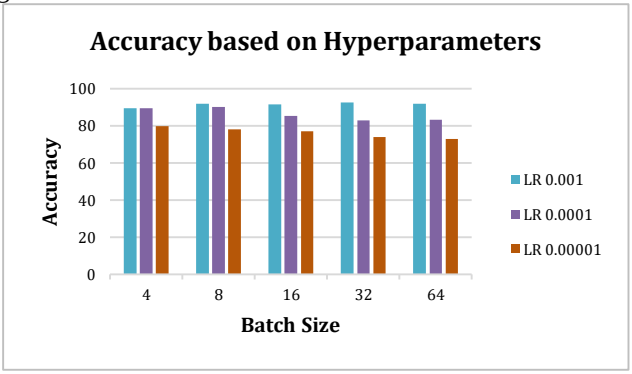

Figure 8. Model Accuracy Graph

The k-fold cross-validation results showed the highest average accuracy of 92.53%, achieved with a learning rate of 0.001 and a batch size of 32. Based on Figure 8, smaller batch sizes and larger learning rates tend to yield higher accuracy. A larger learning rate can help the model escape local minima and reach a better global minimum, thereby improving accuracy when the batch size is small. In addition, the learning rate and batch size affect the required computation time, as shown in Figure 9.

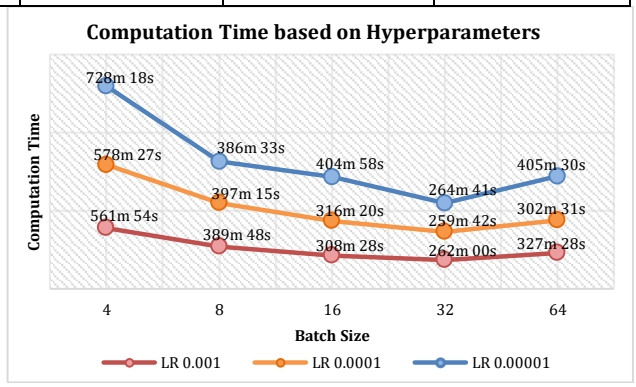

Figure 9. Model Computation Time Graph

Figure 9 shows that a larger batch size requires less computing time. This is because a larger batch size yields more data per iteration, thereby reducing the number of iterations. Thus, the training process becomes faster. However, this does not apply when the batch size is 64, which takes longer. This may be due to insufficient device capacity for a batch size of 64, which requires more memory per iteration.

Based on Figure 9, the batch size of 32 and a learning rate of 0.001 were identified as the most efficient parameters. This combination resulted in the shortest computation time, specifically 4 hours and 22 minutes, while achieving the highest accuracy of 92.53%. Therefore, this parameter set was determined to be the optimal model in this research.

The accuracy, specificity, and sensitivity results were obtained from the average values of the five confusion matrix models. The following is the confusion matrix for each fold shown in Figure 10.

Actual Class	Prediction Class		
	Normal	Moderate	Severe
Normal	93	8	0
Moderate	7	88	5
Severe	0	4	95

1st fold

Actual Class	Prediction Class		
	Normal	Moderate	Severe
Normal	99	7	0
Moderate	1	92	11
Severe	0	1	89

2nd fold

Actual Class	Prediction Class		
	Normal	Moderate	Severe
Normal	97	6	0
Moderate	3	91	14
Severe	0	3	86

3rd fold

Actual Class	Prediction Class		
	Normal	Moderate	Severe
Normal	98	10	0
Moderate	2	82	3
Severe	0	8	97

4th fold

Actual Class	Prediction Class		
	Normal	Moderate	Severe
Normal	97	3	0
Moderate	3	94	10
Severe	0	3	90

5th fold

Figure 10. Confusion Matrix of Optimal Model

Here, the moderate class is the most challenging for the model to identify correctly, followed by the normal class, and then the severe class. Classification errors predominantly occur in classes that are close to each other. Some moderate classes are predicted as normal or severe. This indicates an overlap of visual features at the moderate severity level. The severe class shows the most stable accuracy across folds, indicating that the model performs very well at recognizing severe knee osteoarthritis characteristics. Overall, the consistency of the results across all folds proves the reliability of the NASNet Mobile model in classifying the severity of knee osteoarthritis.

IV. Discussion

This study proposes a NASNet Mobile architecture for classifying the severity of knee osteoarthritis into three classes: grade 0 (normal), grade 3 (moderate), and grade 4 (severe). This model achieved an accuracy of 92.53%, sensitivity of 92.72%, and specificity of 96.32% in 5-fold cross-validation testing.

Most previous studies, such as those conducted by Haseeb et al. (4 classes) [49], Pi et al. (5 classes covering grades 0-4) [50], and Kinger (5 classes) [51], attempted to classify the entire spectrum of the Kellgren-Lawrence scale. These researchers found that deep learning models faced significant challenges in distinguishing visual features in the early stages of the disease, especially in differentiating between grades 1 and 2. By choosing three classes with more distinct visual features, this study effectively avoided this overlap bias, allowing the model to learn features of bone destruction and joint space narrowing more accurately. This aligns with the findings of Ahmed & Mstafa, which indicate that using fewer multiclass labels improves overall performance [52].

Furthermore, NASNet Mobile achieved a 92.53% accuracy, outperforming models from previous studies. For comparison, in a 3-class classification scenario, Rani et al. utilized a custom 12-layer CNN architecture, yielding an accuracy of 78.4% [53]. Similarly, Kinger's fine-tuning of the Inception-V3 model demonstrated a test accuracy of 67% [51]. Haseeb et al. combined EfficientNet-b0 and DenseNet201 for feature extraction, Whale Optimization Algorithm-based optimization, and classification using a Support Vector Machine (SVM), achieving an accuracy of 90.1% [49]. A similar approach was adopted by Ahmed & Mstafa using the Deep Hybrid Learning (DHL) architecture, which combines a CNN and an SVM with binary class labels, achieving an accuracy of 90.8% [52]. On the other hand, Pi et al. used an ensemble method by combining six heavy pre-trained models (such as ResNet-101 and DenseNet-161) and varying image sizes, achieving an accuracy of 76.93% [50].

The proposed model demonstrates its contribution to early disease classification, helping to slow progression and enhance the community's quality of life. Thus, this model is designed as a decision support system for medical personnel and is not intended to replace the final diagnosis of a radiologist or orthopedic specialist.

There are several limitations inherent in our research. First, the data used three of five grades, namely grade 0 (normal), grade 3 (moderate), and grade 4 (severe). Second, there is no solution yet for overlapping data to improve model performance. Third, the computer equipment is inadequate.

V. Conclusion

The results show that NASNet Mobile performs well, achieving 92.53% accuracy, 92.72% sensitivity, and 96.32% specificity with a batch size of 32 and a learning rate of 0.001. The entire process took 262 minutes. The learning rate and batch size significantly affect accuracy and computation time. A smaller batch size and a larger learning rate tend to improve accuracy by helping the model escape local minima and reach a better global convergence minimum. A larger batch size speeds up training by processing more data per iteration, resulting in fewer total iterations.

Further research may include more detailed multi-grade classification and validation on multi-hospital datasets. In addition, the dataset can be expanded to include five classes according to the Kellgren-Lawrence classification.

References

[1] A. Tiwari, M. Poduval, and V. Bagaria, "Evaluation of Artificial Intelligence Models for Osteoarthritis of The Knee Using Deep Learning Algorithms for Orthopedic Radiographs," *World J. Orthop.*, vol. 13, no. 6, pp. 603–614, 2022, doi: 10.5312/WJO.V13.I6.603.

[2] S. M. Ahmed and R. J. Mstafa, "A Comprehensive Survey on Bone Segmentation Techniques in Knee Osteoarthritis Research: From Conventional Methods to Deep Learning," *Diagnostics*, vol. 12, no. 3, 2022, doi: 10.3390/diagnostics12030611.

[3] I. N. Ackerman, J. L. Kemp, K. A. Y. M. Crossley, A. G. Culvenor,

and R. S. Hinman, “Hip and Knee Osteoarthritis Affects Younger People, Too,” vol. 47, no. 2, pp. 67–79, 2017, doi: 10.2519/jospt.2017.7286.

[4] T. N. Cao, K. N. Huynh, H. T. Tran, and M. D. Nguyen, “Association Between Asymptomatic Hyperuricemia and Knee Osteoarthritis in Older Outpatients,” *Eur. Rev. Med. Pharmacol. Sci.*, vol. 26, no. 18, pp. 6600–6607, 2022, doi: 10.26355/eurrev_202209_29760.

[5] M. Ganesh Kumar and A. Das Goswami, “Automatic Classification of the Severity of Knee Osteoarthritis Using Enhanced Image Sharpening and CNN,” *Appl. Sci.*, vol. 13, no. 3, 2023, doi: 10.3390/app13031658.

[6] V. K. V, V. Kalpana, and G. H. Kumar, “Evaluating the Efficacy of Deep Learning Models for Knee Osteoarthritis Prediction Based on Kellgren-Lawrence Grading System,” *e-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 5, no. July, p. 100266, 2023, doi: 10.1016/j.prime.2023.100266.

[7] H. Zhao, L. Ou, Z. Zhang, L. Zhang, K. Liu, and J. Kuang, “The value of deep learning-based X-ray techniques in detecting and classifying K-L grades of knee osteoarthritis: a systematic review and meta-analysis,” *Eur. Radiol.*, vol. 35, no. 1, pp. 327–340, 2025, doi: 10.1007/s00330-024-10928-9.

[8] B. H. Do, C. Langlotz, and C. F. Beaulieu, “Bone Tumor Diagnosis Using a Naïve Bayesian Model of Demographic and Radiographic Features,” *J. Digit. Imaging*, vol. 30, no. 5, pp. 640–647, 2017, doi: 10.1007/s10278-017-0001-7.

[9] J. Langhorn, A. Borjali, E. Hippensteel, W. Nelson, and B. Raeymaekers, “Microtextured CoCrMo Alloy for Use in Metal-on-Polyethylene Prosthetic Joint Bearings: Multi-Directional Wear and Corrosion Measurements,” *Tribol Int.*, vol. 124, pp. 178–183, 2018, doi: 10.1016/j.triboint.2018.04.007.Microtextured.

[10] Y. Xue, R. Zhang, Y. Deng, K. Chen, and T. Jiang, “A Preliminary Examination of The Diagnostic Value of Deep Learning in Hip Osteoarthritis,” *PLoS One*, vol. 12, no. 6, pp. 1–9, 2017, doi: 10.1371/journal.pone.0178992.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT press., 2016. [Online]. Available: https://scholar.google.com/scholar_lookup?title=Deep+Learning&author=Goodfellow,+I.&author=Bengio,+Y.&author=Courville,+A.&publication_year=2016

[12] D. Saha, M. P. Mangukia, and A. Manickavasagan, “Real-Time Deployment of MobileNetV3 Model in Edge Computing Devices Using RGB Color Images for Varietal Classification of Chickpea,” *Appl. Sci.*, vol. 13, no. 13, pp. 1–16, 2023, doi: 10.3390/app13137804.

[13] K. Radhika, K. Devika, T. Aswathi, P. Sreevidya, V. Sowmya, and K. P. Soman, *Performance Analysis of NASNet on Unconstrained Ear Recognition*, vol. 871. Springer International Publishing, 2020. doi: 10.1007/978-3-030-33820-6_3.

[14] A. O. Adedoja, P. A. Owolawi, T. Mapayi, and C. Tu, “Intelligent Mobile Plant Disease Diagnostic System Using NASNet-Mobile Deep Learning,” *IAENG Int. J. Comput. Sci.*, vol. 49, no. 1, pp. 216–231, 2022.

[15] S. Dümen, E. K. Yilmaz, K. Adem, and E. Avaroglu, “Achieving High Accuracy in Lemon Quality Classification : A Comparative Study of Deep Learning and Transformer Models,” *Res. Sq.*, pp. 1–20, 2023.

[16] M. M. Ahsan, K. D. Gupta, M. M. Islam, S. Sen, M. L. Rahman, and M. Shakhawat Hossain, “COVID-19 Symptoms Detection Based on NasNetMobile with Explainable AI Using Various Imaging Modalities,” *Mach. Learn. Knowl. Extr.*, vol. 2, no. 4, pp. 490–504, 2020, doi: 10.3390/make2040027.

[17] S. Bharati, P. Podder, M. R. Mondal, and H. N. Gandhi, “Optimized NASNet for Diagnosis of COVID-19 from Lung CT Images,” in *Advances in Intelligent Systems and Computing*, 2021. doi: 10.1007/978-3-030-71187-0_9.

[18] A. Yilmaz, M. Kalebasi, Y. Samoylenko, M. E. Guvenilir, and H. Uvet, “Benchmarking of Lightweight Deep Learning Architectures for Skin Cancer Classification using ISIC 2017 Dataset,” *arXiv*, 2021, [Online]. Available: http://arxiv.org/abs/2110.12270

[19] P. Anantha Prabha, G. Suchitra, and R. Saravanan, “Cephalopods Classification Using Fine Tuned Lightweight Transfer Learning Models,” *Intell. Autom. Soft Comput.*, vol. 35, no. 3, pp. 3065–3079, 2023, doi: 10.32604/iasc.2023.030017.

[20] S. Kumar, “Knee Osteoarthritis Dataset (Preprocessed-128x128).” [Online]. Available: https://www.kaggle.com/datasets/sachinkumar413/knee-osteoarthritis-dataset-preprocessed128x128?resource=download

[21] P. Chen, “Knee Osteoarthritis Severity Grading Dataset,” Mendeley Data.

[22] J. Yoonsuh, “Multiple Predicting K-fold Cross-Validation for Model Selection,” *J. Nonparametr. Stat.*, vol. 30, no. 1, pp. 197–215, 2017, [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10485252.2017.1404598

[23] S. Mohapatra, G. Kumar Pati, M. Mishra, and T. Swarnkar, “Gastrointestinal Abnormality Detection and Classification Using Empirical Wavelet Transform and Deep Convolutional Neural Network from Endoscopic Images,” *Ain Shams Eng. J.*, vol. 14, no. 4, p. 101942, 2023, doi: 10.1016/j.asej.2022.101942.

[24] D. C. R. Novitasari *et al.*, “Image Fundus Classification System for Diabetic Retinopathy Stage Detection Using Hybrid CNN-DELM,” *Big Data Cogn. Comput.*, vol. 6, no. 4, 2022, doi: 10.3390/bdcc6040146.

[25] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance,” *Appl. Soft Comput.*, vol. 97, 2020, doi: 10.1016/j.asoc.2019.105524.

[26] S. N. Fadilah, D. C. R. Novitasari, and L. Hakim, “Pengaruh Reduksi Fitur Pada Klasifikasi Kanker Paru Menggunakan CNN Dengan Arsitektur GoogLeNet,” *J. Fourier*, vol. 12, no. 1, pp. 20–32, 2023, doi: 10.14421/fourier.2023.121.20-32.

[27] D. Z. Haq, “Klasifikasi Citra Kanker Kulit Menggunakan Convolutional Neural Network Model Googlenet,” UIN Sunan Ampel Surabaya, 2021.

[28] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, *A review of convolutional neural networks in computer vision*, vol. 57, no. 4. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10721-6.

[29] Q. Gao, S. Lim, and X. Jia, “Hyperspectral Image Classification Using Convolutional Neural Networks and Multiple Feature Learning,” *Remote Sens.*, vol. 10, no. 2, 2018, doi: 10.3390/rs10020299.

[30] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of A Convolutional Neural Network,” *Proc. 2017 Int. Conf. Eng. Technol. ICET 2017*, pp. 1–6, 2018, doi: 10.1109/ICEngTechnol.2017.8308186.

[31] Q. Sellat, S. K. Bisoy, and R. Priyadarshini, “Semantic Segmentation for Self-Driving Cars using Deep Learning: A Survey,” in *Cognitive Big Data Intelligence with a Metaheuristic Approach*, 2021. doi: 10.1016/B978-0-323-85117-6.00002-9.

[32] M. M. Taye, “Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions,” *Computation*, vol. 11, no. 3, 2023, doi: 10.3390/computation11030052.

[33] F. Sultana, A. Sufian, and P. Dutta, “Advancements in Image Classification Using Convolutional Neural Network,” in *Proceedings - 2018 4th IEEE International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2018*, IEEE, 2018, pp. 122–129. doi: 10.1109/ICRCICN.2018.8718718.

[34] B. Zoph, V. Vasudevan, J. Shlens, and Q. V Le, “Learning Transferable Architectures for Scalable Image Recognition,” *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 8697–8710, 2018.

[35] Y. T. Chen *et al.*, “Deep Learning–based Brain Computed Tomography Image Classification with Hyperparameter Optimization through Transfer Learning for Stroke,” *Diagnostics*, vol. 12, no. 4, 2022, doi: 10.3390/diagnostics12040807.

[36] V. Anand, S. Gupta, A. Altameem, S. R. Nayak, R. C. Poonia, and A. K. J. Saudagar, “An Enhanced Transfer Learning based Classification for Diagnosis of Skin Cancer,” *Diagnostics*, vol. 12,

no. 7, 2022, doi: 10.3390/diagnostics12071628.

[37] T. Putra, Suprapto, and A. Bukhori, "Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia," *J. Ilmu Siber dan Teknol. Digit.*, vol. 1, no. 1, pp. 1–28, 2022, [Online]. Available: https://penerbitgoodwood.com/index.php/jisted/article/view/1509/369

[38] N. Iyer, V. Thejas, N. Kwatra, R. Ramjee, and M. Sivathanu, "Wide-minima Density Hypothesis and the Explore-Exploit Learning Rate Schedule," *J. Mach. Learn. Res.*, vol. 24, pp. 1–37, 2023, [Online]. Available: http://arxiv.org/abs/2003.03977

[39] Y. Wu *et al.*, "Demystifying Learning Rate Policies for High Accuracy Training of Deep Neural Networks," *Proc. - 2019 IEEE Int. Conf. Big Data, Big Data 2019*, pp. 1971–1980, 2019, doi: 10.1109/BigData47090.2019.9006104.

[40] Karsito and Susanti Santi, "Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naive Bayes Di Perumahan Azzura Residence," *J. Teknol. Pelita Bangsa*, vol. 9, pp. 43–48, 2019.

[41] Styawati, N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, [Online]. Available: http://situs.com

[42] K. R. Mahmudah, B. Purnama, F. Indriani, and K. Satou, "Machine Learning Algorithms for Predicting Chronic Obstructive Pulmonary Disease from Gene Expression Data with Class Imbalance," in *12th International Conference on Bioinformatics Models, Methods and Algorithms; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021*, 2021, pp. 148–153. doi: 10.5220/0010316501480153.

[43] O. Sule and S. Viriri, "Enhanced Convolutional Neural Networks for Segmentation of Retinal Blood Vessel Image," *2020 Conf. Inf. Commun. Technol. Soc. ICTAS 2020 - Proc.*, pp. 0–5, 2020, doi: 10.1109/ICTAS47918.2020.233996.

[44] S. Qummar *et al.*, "A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019, doi: 10.1109/ACCESS.2019.2947484.

[45] N. Sakinah, T. Badriyah, and I. Syarif, "Analisis Kinerja Algoritma Mesin Pembelajaran untuk Klarifikasi Penyakit Stroke Menggunakan Citra CT Scan," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 4, p. 833, 2020, doi: 10.25126/jtiik.2020743482.

[46] Y. Xiao *et al.*, "A New Color Augmentation Method For Deep Learning Segmentation Of Histological Images," in *IEEE 16th International Symposium on Biomedical Imaging*, 2019, pp. 886–890.

[47] H. A. Morsy, "Comparison of Commonly Used Non-Adaptive Image Scaling Techniques," *CiiT Int. J. Digit. Image Process.*, vol. 10, no. 9, pp. 177–180, 2018, [Online]. Available: https://www.researchgate.net/publication/344324331_Comparison_of_commonly_used_non-adaptive_image_scaling_techniques

[48] D. Umamaheswari and D. S. Geetha, "Segmentation and Classification of Acute Lymphoblastic Leukemia Cells Tooled with Digital Image Processing and ML Techniques," *Proc. Second Int. Conf. Intell. Comput. Control Syst. (ICICCS 2018)*, 2018, doi: 10.1201/9781439802069-1.

[49] A. Haseeb *et al.*, "Knee Osteoarthritis Classification Using X-Ray Images Based on Optimal Deep Neural Network," *Comput. Syst. Sci. Eng.*, vol. 47, no. 2, pp. 2397–2415, 2023, doi: 10.32604/csse.2023.040529.

[50] S. W. Pi, B. D. Lee, M. S. Lee, and H. J. Lee, "Ensemble deep-learning networks for automated osteoarthritis grading in knee X-ray images," *Sci. Rep.*, vol. 13, no. 1, pp. 1–17, 2023, doi: 10.1038/s41598-023-50210-4.

[51] S. Kinger, "Deep Learning for Automatic Knee Osteoarthritis Severity Grading and Classification," *Indian J. Orthop.*, vol. 58, no. 10, pp. 1458–1473, 2024, doi: 10.1007/s43465-024-01259-4.

[52] S. M. Ahmed and R. J. Mstafa, "Identifying Severity Grading of Knee Osteoarthritis from X-ray Images Using an Efficient Mixture of Deep Learning and Machine Learning Models," *Diagnostics*, vol. 12, no. 12, 2022, doi: 10.3390/diagnostics12122939.

[53] S. Rani *et al.*, "Deep learning to combat knee osteoarthritis and severity assessment by using CNN-based classification," *BMC Musculoskelet. Disord.*, vol. 25, no. 1, 2024, doi: 10.1186/s12891-024-07942-9.