

Deep Learning-Based Osteoporosis Classification from Knee Radiographs Using ResNet50, Xception, and InceptionV3 Architectures

Rani Puspita, Asysta Amalia Pasaribu, Neni Nur Laili Ersela Zain

Abstract— Bones serve a crucial function in providing strength to the body and protecting vital human organs. There are at least two bone problems, namely osteopenia and osteoporosis. However, there is a problem with early detection due to high costs and lack of facilities. Therefore, this research is necessary to distinguish between those at risk and those who are not. This research presents a retrospective analysis of a publicly available knee radiograph dataset. This study uses the ResNet50, Xception, and InceptionV3 methods with data acquisition, data preprocessing, and augmentation techniques, then model building, training, validating, testing with ($N = 195$ images) using accuracy, precision, recall, and f1-score measurements using macro-averaging across classes, visualizing them into a confusion matrix and ROC Curve, and obtaining the AUC score results. The results of this study indicate that Xception is the method that achieved the highest accuracy of 88%, macro-precision of 88%, macro-recall of 87%, and macro-F1-score of 87%. However, in the medical world, the AUC score is considered a parameter to consider. The research results indicate that, based on the AUC scores, the three models exhibit comparable performance. The AUC score obtained for the normal class is 0.99, while for the other classes, it ranges from 0.96 to 0.97.

Index Terms— bone classification, deep learning, ResNet50, Xception, InceptionV3

I. INTRODUCTION

A network that adjusts its mass, strength, and structure as pressure changes is called bone. Bones can prevent fractures. However, bone capacity will always decrease over time [1]. Bone tissue is highly metabolically active and is continually regenerated and repaired through a process known as bone remodeling. Bone tissue provides strength to the body

and protects vital organs such as those in the skull and chest. It also protects the internal organs, which contain the bone marrow where blood cells are formed [2]. This is also supported by other research, which states that bone tissue will continue to undergo remodeling over time [3]. However, there are some bone problems, such as osteoporosis and osteopenia. Both of these bone conditions are pressing public health issues and are closely related to aging [4]. Disorders that result in decreased bone mass and bone mineral density (BMD) are called osteoporosis. There are 10 million people who suffer from this disease and 1.5 million who experience quite serious effects. Osteopenia is a common condition affecting nearly 40 million people. It typically occurs in women in their 50s [5]. This is consistent with other research suggesting that osteoporosis also causes bone architecture to deteriorate, increasing the risk of fractures. This condition usually occurs in women and the elderly. In younger people, osteoporosis is usually due to certain medical conditions [6]. A BMD T-score of -2.5 or below is a sign of osteoporosis, while osteopenia is -1 to -2.5 [7]. One symptom is knee pain. Knee pain is a common complaint experienced by many people, impacting lives and futures [8]. Osteoporosis and osteopenia will also cause an increased risk of death, loss of independence, and a large social burden for sufferers. Risk factors for this disease include an unhealthy lifestyle, an abnormal BMI, and other factors. Generally, clinical symptoms of osteoporosis appear after a fracture and can cause profound pain, significantly reducing the sufferer's quality of life [9]. This also occurs in Indonesia, where osteoporosis occurs in approximately 23% of women aged 50-80 and 53% in those aged 70-80. For men, bone density also decreases by 10-20% between the ages of 20-39 and by 40% between the ages of 70-79 [10].

Although osteoporosis and osteopenia carry a very high risk of fracture, early detection remains a challenge. Currently, osteoporosis is diagnosed using dual-energy X-ray absorptiometry (DXA); however, high costs and limited facilities remain the main obstacles. Therefore, research is needed to classify bones as affected by osteoporosis, osteopenia, or normal bone to distinguish between risk factors that require evaluation and those that do not [11]. In this case, the most widely used imaging method that can help researchers detect bone disease is X-ray. X-ray is a common

Manuscript received July 6, 2026; revised August 11, 2026. This work was supported by Computer Science Department, School of Computer Science, Bina Nusantara University.

R. Puspita is with the Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480 (corresponding author: R. Puspita, e-mail: rani.puspita@binus.ac.id).

A. A. Pasaribu is with Department of Statistics, School of Computer Science, Bina Nusantara University, Indonesia 11480 (e-mail: asysta.amalia@binus.ac.id).

N. N. L. E. Zain is with Department of Mathematics, School of Computer Science, Bina Nusantara University, Indonesia 11480 (email: neni.zain@binus.ac.id).

method used in bone diagnosis [12]. X-rays are so widely performed that they can be utilized for screening purposes without incurring additional costs. Currently, Artificial Intelligence is being used in many medical systems to assist in decision-making. It has been widely used as an analytical tool for osteoporosis, hip X-rays, and dental imaging [13]. This aims to provide early diagnosis as a step to prevent bone fractures if someone is diagnosed with osteoporosis [14].

To analyze images, previous research commonly used Convolutional Neural Networks (CNN). CNNs excel at learning patterns in raw images, making them suitable for image classification tasks such as detecting osteoporosis from knee radiographs. CNNs automatically learn features from existing images [15]. There are many methods in CNN, including ResNet50, Xception, and InceptionV3 [16]. Other studies also show that methods such as InceptionV3, ResNet50, and VGG16 are also included in CNN, and VGG16 has worse accuracy [17]. Xception is a method that is more advanced than Inception. Xception has depthwise convolution layers and pointwise convolution layers. Inception, on the other hand, is a deep learning model and belongs to the GoogLeNet family. For good and efficient accuracy, the InceptionV3 method is usually used [18]. ResNet50 is a CNN model commonly used for image analysis. This model has many layers. This model is suitable for medical image detection [19]. Based on research conducted by Salim et al., Xception is an excellent method. The results showed that Xception produced an accuracy of 97.73%. The steps involved collecting datasets, building the models, and evaluating them mathematically using accuracy, precision, recall, and f-measure [20]. Other research indicates that InceptionV3 also has good accuracy results. This process begins with data collection, model building, data annotation, augmentation, and evaluation. The evaluation results indicate that InceptionV3 achieves an accuracy of over 95% [21]. Similarly, ResNet50 has been used by many previous researchers, with excellent results. For example, research by Elpeltagy et al. found accuracy rates exceeding 97% for several datasets. The steps involved collecting the dataset, preprocessing it, building the models, and evaluating it [22]. Limitations of previous studies include the fact that some of the referenced papers tested the models only in isolation and, for the most part, reported only accuracy results without using ROC-AUC.

Based on the above background, this research aims to classify osteoporosis, osteopenia, and normal bone using the Xception, InceptionV3, and ResNet50 methods. A comparison will then be conducted. In addition to comparing the performance of the three architectures, this study offers guidance on model complexity and accuracy to facilitate implementation in clinical environments with limited computational resources. It also provides a deep learning-based screening framework for the early detection of osteoporosis. Then, this research will provide accurate information regarding various methods in bone objects. This research also aims to contribute, particularly in the utilization of deep learning technology to aid diagnosis and decision-making. Then, this research not only applies a standard deep

learning model to a public dataset, but also to demonstrate that knee radiograph images can be used to assess bone density, making them a cost-effective initial screening tool. Furthermore, this study prioritizes AUC, unlike previous studies that focused solely on standard accuracy evaluations.

II. RELATED WORKS

In this chapter, researchers will explain several related works regarding the classification of osteoporosis, osteopenia, and also several deep learning methods such as Xception, InceptionV3, and also ResNet50.

Based on research conducted by Abubakar, osteoporosis is a bone disease that is highly susceptible to fractures. This occurs because bone mineral density is low and bone tissue structure has also deteriorated. In this case, the researchers used the Visual Geometry Group Deep (VGG-16) method. The steps taken were dataset collection, grayscale conversion, data normalization, data augmentation, building and training a model with an 80:20 ratio, and evaluation. The results achieved an accuracy of around 88% after fine-tuning [23]. This aligns with research conducted by Chen et al., which states that osteoporosis is a systemic disorder that causes low bone mass and an increased risk of fractures. This can impact quality of life and lead to high mortality rates after fracture surgery. Therefore, tools that can detect osteoporosis are needed. They used CNN, and the evaluation results showed 97.2% for sensitivity, 95.6% for specificity, then a range of 95-97% for positive and negative predictive values [24]. Other research indicates that osteoporosis is characterized by low bone mineral density (BMD) and a high risk of fractures. Media imaging has also become increasingly popular recently, and it can be a valuable tool for diagnosing and assessing osteoporosis. The steps taken in this study included preprocessing, segmentation, feature-based (ROI selection) and deep learning, feature selection, and model classification. The output of this study achieved accuracy, sensitivity, and specificity of around 61.8 to 99.4%, 41.0 to 100%, and 31 to 100%. The AUC ranged from 0.582 to 0.994 [25]. Low BMD needs to be detected quickly as a preventive measure. Data were collected directly from 138 women, and 24.1% were osteopenic and 32.85% were postmenopausal. Sequential forward feature selection (SFFS) is a method used in this study. In this study, the accuracy was 92.12% [35]. Other research has shown that many patients with osteopenia and osteoporosis remain undiagnosed. Therefore, this study aims to build a CNN model for primary osteopenia and osteoporosis screening. The methods used include crop size, model building, and fusion module optimization. The output performed well, with an AUC ranging from 0.909 to 0.937. For osteopenia, the highest AUC was 0.785 [26].

Another study is about detecting a disease on radiograph, where researchers also use the Xception, ResNet50, etc. methods. They conducted research by collecting datasets, building models, training, and testing. From here, it can be seen that Xception got an accuracy of 98.34%, InceptionV3 98.86%, and ResNet50 99.32%, followed by precision, recall, and f1-score with similar numbers [28]. Huang et al.

conducted image collection, preprocessing, transfer learning, fine-tuning, model proposal, Grid SearchCV, and performance evaluation using accuracy, recall, and related metrics; the highest accuracy achieved was 96.50% with InceptionV3, and the best-performing model was Xception at 96.78% [29]. Other studies likewise report InceptionV3 and Xception as the best-performing methods, although MobileNet and VGG16 were also tested; following data collection, model creation, dataset splitting, and evaluation, InceptionV3 achieved training accuracy, testing loss, and validation accuracy of around 99%, while Xception achieved 98-99% [30]. Studies suggest that AUC provides a more objective picture of performance than accuracy. AUC is considered more consistent and effective in the evaluation phase [31]. Then supported by other research that AUC is considered a good metric because it evaluates the model's overall performance rather than just the performance at a single cutoff point. It also measures the probability of correctly ranking positive and negative cases [27].

Therefore, based on the background and previous research on osteopenia, osteoporosis, and the methods used, researchers will use the Xception, InceptionV3, and ResNet50 methods to detect osteopenia, osteoporosis, and normal knee radiographs. These models will be implemented and evaluated using accuracy, precision, recall, and f1-score. Furthermore, the AUC score will be examined to ensure accurate disease diagnosis. This research contributes by providing a comparison of deep learning architectures. It makes a significant contribution by evaluating metrics using macro-averaging, analyzing error patterns in confusion matrices, and identifying errors associated with suboptimal class performance. This research contributes by providing a comparison of deep learning architectures. It makes a significant contribution by evaluating metrics using macro-averaging, analyzing error patterns in confusion matrices, and identifying errors associated with suboptimal class performance.

III. RESEARCH METHODOLOGY

This study will classify bone diseases, specifically osteopenia, osteoporosis, and normal bone density. The approaches used are Xception, InceptionV3, and ResNet50. Several papers state that Xception, InceptionV3, and ResNet50 are good methods to implement. Therefore, we want to know which method is the most efficient and effective in classifying bones based on knee radiograph images. The research steps undertaken are illustrated in Fig. 1.

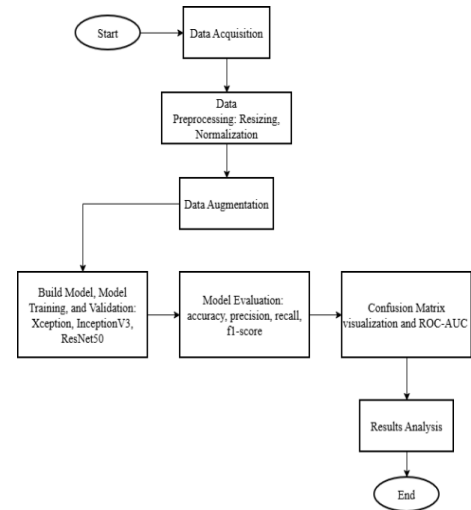


Fig. 1. Research methodology

A. Data Acquisition

This dataset consists of imaging multi-class data in the form of radiographs. This data is collected from Kaggle [8]. There is a total of 1947 image data and each data is 793 knee radiograph image data for osteoporosis, 374 for osteopenia, and 780 for normal. The image labels (Normal, Osteopenia, and Osteoporosis) were adopted directly from the original labels provided by the source public repository without independent re-labeling. Although the dataset is relatively small, this limitation is addressed by using three models and augmentation techniques to ensure good generalization, as well as measures to prevent overfitting. This dataset is anonymous, authentic, and has not been previously augmented. The researchers have ensured that there is no duplicate data. The dataset is divided into three classes: normal, which contains images of knees without signs of osteoporosis; osteopenia, which shows early stages of bone density loss; and osteoporosis, which contains a collection of images showing more advanced bone density degradation. To address the imbalanced dataset, where there are fewer osteopenia classes, a stratified distribution was applied during data division to ensure proportionality. Furthermore, we also conducted an evaluation to ensure bias. The evaluation performance measures used were the f1-value and AUC metrics, which are independent of the threshold, not just the overall accuracy. An example of the dataset can be seen in Fig. 2.



Fig. 2. An example of the dataset

B. Data Preprocessing and Augmentation

In the data preprocessing subsection, the image size becomes 224 x 224 pixels for ResNet50 and Xception, and 299x299 pixels for InceptionV3. The native resolution of each model is maintained, as forcing a reduction in size would result in the loss of important details specific to the model. Then, the data was processed in batches of 16 at each training iteration. This step aims to ensure the data has a consistent and efficient format, thus achieving optimal results during the model training stage. This research was conducted using a random seed of 42. After preprocessing, augmentation was performed. Data augmentation was carried out to increase data variation and reduce the risk of overfitting during training. Data augmentation is performed online. Augmentation was performed using ImageDataGenerator and brightness settings between 0.8-1.2, resulting in brighter or darker images compared to the original. No augmentation was applied during validation and testing to allow for a more objective evaluation of model performance. The data is also divided in a ratio of 80:10:10. Data partitioning was performed proportionally; the ratios of normal, osteopenia, and osteoporosis cases were set to be identical across the training, validation, and testing sets, ensuring a fair and balanced distribution. The dataset after being divided by that ratio, resulted in 1557 for training, 195 for validation, and 195 images for testing. Then, this dataset was split only at the image level, because individual patient identification was not available in the dataset. For further details, Table 1 presents the parameters used in this research.

TABLE I
RESEARCH PARAMETER

Configuration	Models		
	<i>ResNet50</i>	<i>Xception</i>	<i>InceptionV3</i>
Input Dimension	224x224x3	224x224x3	299x299x3
Data Split Ratio	80:10:10	80:10:10	80:10:10
Data Augmentation	0.8–1.2	0.8–1.2	0.8–1.2
Batch Size	16	16	16
Fully Connected Layer	Dense (128, ReLU) + Dropout	Dense (128, ReLU) + Dropout	Dense (128, ReLU) + Dropout
Optimizer and Learning Rate	Adamax (10^{-3})	Adamax (10^{-3})	Adamax (10^{-3})
Trainable Parameter	23,797,251	21,069,611	22,031,011
Frozen / Fine-tuned Layers	Fine-tuned (all layers)	Fine-tuned (all layers)	Fine-tuned (all layers)
Stopped Epoch	Epoch 10	Epoch 8	Epoch 11
Computational Cost	~46 s/epoch	~45 s/epoch	~46 s/epoch

C. Build and Model Training

At this stage, the researchers used several architectures, including ResNet50, Xception, and InceptionV3. The method for assembling the model will differ for each method, but there are some similarities, including a dropout layer to reduce overfitting. Then, a dense layer with 128 neurons. We use the ReLU activation function. For the dense output layer, the

Softmax function is used to generate the classification probability.

Then, the model is trained with the Adamax optimizer with a learning rate of 0.001 using the multi-class cross-entropy loss function. During training, an early stopping technique was applied to monitor the validation loss. While training is used to train the model, validation is used to evaluate the model during the training process, where the model calculates the loss, accuracy, val_loss, and val_accuracy. To prevent overfitting, an early stopping technique was used to monitor the validation loss with a patience of 3 epochs over a maximum of 15 epochs. We started with ResNet50, followed by Xception and finally InceptionV3, and Fig. 3, Fig. 4, and Fig. 5 show the architectures of ResNet50, Xception, and InceptionV3, respectively.

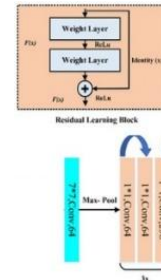


Fig. 3. ResNet50 Architecture [32]

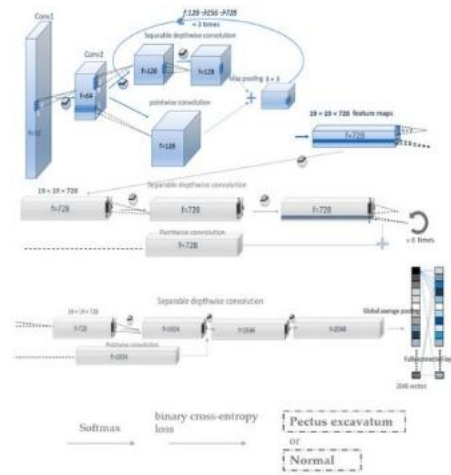


Fig. 4. Xception Architecture [33]

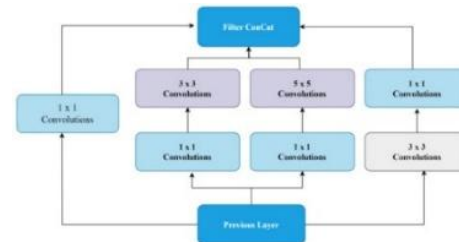


Fig. 5. InceptionV3 Architecture [34]

D. Model Evaluation

At this stage, the trained model will be evaluated using four

methods, namely accuracy, precision, recall, and f1-score to measure its performance. A confusion matrix will also be used to provide a detailed overview of the number of correct and incorrect predictions. Researchers also use AUC, and the Receiver Operating Characteristic (ROC) curve value to assess the model's ability to differentiate between classes.

IV. RESULTS AND DISCUSSIONS

The results of this research will be described in Table 2, Table 3 and Table 4. The attributes consist of the methods used, then the results of accuracy, precision, recall, f1-score per model, and AUC score per class in Table 5.

TABLE II
RESULTS OF RESNET50

ResNet50	Performance Evaluation		
	Precision	Recall	F1-Score
Normal	0.83	0.97	0.89
Osteopenia	0.86	0.77	0.81
Osteoporosis	0.89	0.81	0.84
Accuracy			0.86
Macro Avg	0.86	0.85	0.85
Weighted Avg	0.86	0.86	0.85

TABLE III
RESULTS OF XCEPTION

Xception	Performance Evaluation		
	Precision	Recall	F1-Score
Normal	0.90	0.94	0.92
Osteopenia	0.86	0.79	0.83
Osteoporosis	0.86	0.87	0.86
Accuracy			0.88
Macro Avg	0.88	0.87	0.87
Weighted Avg	0.88	0.88	0.88

TABLE IV
RESULTS OF INCEPTIONV3

InceptionV3	Performance Evaluation		
	Precision	Recall	F1-Score
Normal	0.91	0.89	0.90
Osteopenia	0.90	0.79	0.84
Osteoporosis	0.82	0.91	0.86
Accuracy			0.87
Macro Avg	0.88	0.86	0.87
Weighted Avg	0.88	0.87	0.87

After experimenting with ResNet50, further optimization was required to enhance classification performance, which led to the need for a more complex architecture. Therefore, we conducted research with the Xception and InceptionV3 models. Based on macro average, Table 2 shows that the ResNet50 is the model with the lowest baseline in all

evaluation metrics with accuracy and precision of 86%, recall and f1-score of 85%. Then, Xception method achieved accuracy, precision, recall, and f1-score of 88%, 88%, 87%, and 87%, respectively. This indicates that Xception has excellent capabilities and minimizes prediction errors. As for the InceptionV3 method, its accuracy is 87%, precision 88%, recall 86%, and f1-score 87%. There is a very slight difference with Xception.

TABLE V
COMPARISON TABLE OF AUC IN NORMAL, OSTEOPENIA, AND OSTEOPOROSIS CLASSES

Models	AUC Performance		
	Normal	Osteopenia	Osteoporosis
ResNet50	0.99	0.96	0.97
Xception	0.99	0.96	0.97
InceptionV3	0.99	0.97	0.97

Based on the table, all models received an AUC score above 0.95 and were categorized as Excellent. Although the accuracy results differed slightly, this metric is less reliable as a benchmark for decision efficacy. For the normal class, all models received a score of 0.99. Then, for the osteopenia class, the model demonstrated comparable performance. For the osteoporosis class, all models achieved a score of 0.97 because their features are clear. In the medical domain, the AUC score is often considered the primary parameter compared to accuracy due to its ability to measure the stability of a diagnosis across multiple decisions.

Overall, while Xception performed slightly better in overall accuracy, all architectures demonstrated comparable performance, with AUC values reaching 0.96 to 0.97 for the osteopenia class. This research is crucial for early detection to prevent the disease from progressing from the initial phase to more severe osteoporosis. However, it should be noted that while these metrics are promising, the clinical part will require further validation, such as cost analysis, integration testing in real-world hospital workflows, and others. The following can be seen in Fig. 6, Fig. 7, Fig. 8, and Fig. 9 regarding the confusion matrix and ROC curve to provide further insight into the models error distribution.

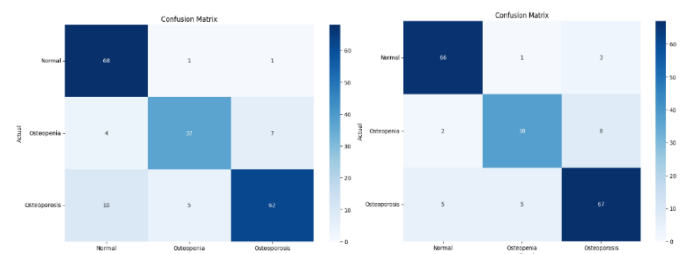


Fig. 6. Confusion Matrix - ResNet50 (left) Confusion Matrix – Xception (right)

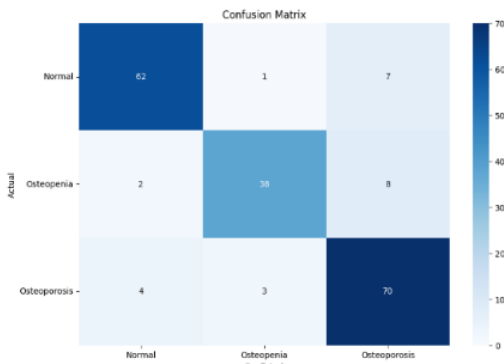


Fig. 7. Confusion Matrix – InceptionV3

Based on the confusion matrix in Fig. 6 and Fig. 7, the model shows a very strong ability in classifying the Normal class. On ResNet, the correct rate is 68 images. Then, Xception shows a fairly balanced performance and is slightly better than ResNet50. Then, there is a challenge in the Osteopenia class, where other models predict many incorrectly. For the osteoporosis class, InceptionV3 also managed to predict 70 correct images, while the others only ranged around 60. ResNet50 only excels in the Normal class. Out of 48 images belonging to the osteopenia category, the Xception and InceptionV3 models misclassified 8 cases as osteoporosis and 2 as normal, whereas ResNet50 misclassified 7 cases as osteoporosis and 4 as normal. Clinically, this indicates a safer outcome for patients, as the tendency toward overdiagnosis prompts earlier medical examinations. This pattern is related to class imbalance. Consequently, models dealing with osteopenia images tend to be biased toward the osteoporosis class, as the bone damage characteristics associated with it are more pronounced in the data.

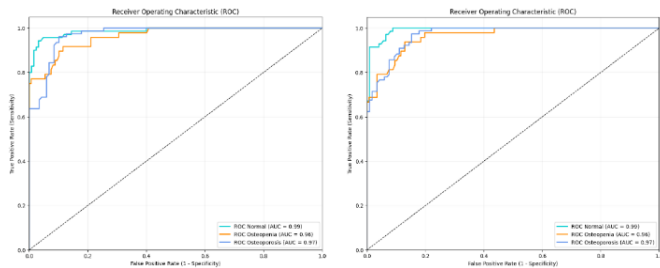


Fig. 8. ROC Curve - ResNet50 (left) ROC Curve – Xception (right)

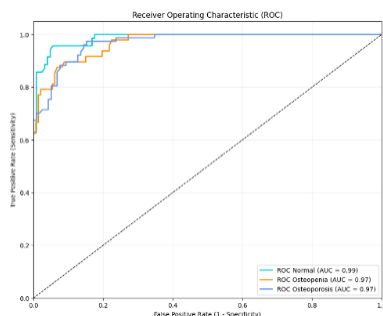


Fig. 9. ROC Curve – InceptionV3

In Fig. 8 and Fig. 9, it can be seen that the three architectures exhibit similar and stable ROC curves. All models achieved an AUC of 0.99 for the Normal class, and 0.96 to 0.97 for the other classes.

V. LIMITATIONS

This research has several limitations. Among them is the potential for data leakage, as the data split was performed at the image level while patient identifiers were unavailable in the repository. Furthermore, class labels were derived from a public repository without independent radiological re-verification. Model evaluation was conducted solely on the test dataset, lacking external validation. Model performance is evaluated using fixed training, validation, and test datasets, rather than a cross-validation scheme. Additionally, there is no medically standardized clinical threshold for determining clinical actions. Finally, model performance was evaluated without formal statistical significance testing.

VI. CONCLUSION

Osteopenia and osteoporosis significantly increase the risk of death, loss of independence, and social burden for sufferers. In Indonesia, osteoporosis affects approximately 23% of women aged 50-80, and 53% of women aged 70-80. Therefore, this study consists of three categories of data, including normal, osteoporosis, and osteopenia. Data collected from Kaggle.com and divided into 80% training, 10% validation, and the remaining 10% is testing. The methods used in this study are ResNet50, Xception, and InceptionV3. The steps taken in this study include data acquisition, data preprocessing such as resizing and normalization, data augmentation with an image generator, building, training, and validating models, and evaluating with accuracy, precision, recall, and f1-score, then continued with visualization of confusion matrix and ROC-AUC curve, and ended with results analysis. The results of the study show that in terms of accuracy in general, Xception is higher because it achieved an accuracy score of 88%, precision of 88%, and a recall and f1-score of 87%. However, in medical terms, the AUC score is a parameter that is more considered. Based on the evaluation results, all three architectures demonstrated high and comparable performance, with AUC scores reaching 0.99 for the normal class and ranging from 0.96 to 0.97 for the other classes. This consistency makes the three models promising candidates for supporting initial screening. For future works, implementing Explainable AI (XAI), such as Gradient-weighted Class Activation Mapping (Grad-CAM), will be considered. In addition, more in-depth statistical analysis will be conducted to reinforce the findings.

ACKNOWLEDGMENT

The authors would like to thank Bina Nusantara University for the support provided for this research, as well as the Computer Science Department, School of Computer Science, for the facilities and assistance during the writing process.

DATA AVAILABILITY

The dataset used in this study is publicly available. The dataset used in the research titled: Knee Osteoporosis Diagnosis Based on Deep Learning <https://link.springer.com/article/10.1007/s44196-024-00615-4> (DOI: 10.1007/s44196-024-00615-4), and the dataset link can be accessed in url link as follow:

<https://www.kaggle.com/datasets/mohamedgobara/multi-class-knee-osteoporosis-x-ray-dataset>

AUTHOR CONTRIBUTIONS STATEMENT

R. Puspita: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

A. A. Pasaribu: Formal Analysis, Investigation, Resources, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

N. N. L. E. Zain: Formal Analysis, Investigation, Writing - Review & Editing, Project administration, Funding acquisition.

AI ACKNOWLEDGMENT

The authors used generative AI during the preparation of this manuscript to assist with grammatical correction, language refinement, and text structuring. The AI tools were used as supporting tools and did not replace the authors' intellectual contributions. All AI-generated outputs were carefully reviewed and edited by the authors. The authors take full responsibility for the final content and integrity of this paper.

REFERENCES

- [1] S. M. Alnasser *et al.*, "Effectiveness of Exercise Loading on Bone Mineral Density and Quality of Life Among People Diagnosed with Osteoporosis, Osteopenia, and at Risk of Osteoporosis — A Systematic Review and Meta-Analysis," pp. 1–27, 2025, doi: 10.3390/jcm14124109.
- [2] A. Wawrzyniak and K. Balawender, "Structural and Metabolic Changes in Bone," 2022, doi: 10.3390/ani12151946.
- [3] R. Florencio-silva, G. Rodrigues, E. Sasso-cerri, M. J. Simões, P. S. Cerri, and B. Cells, "Biology of Bone Tissue : Structure, Function, and Factors That Influence Bone Cells," vol. 2015, 2015, doi: 10.1155/2015/421746.
- [4] Y. Liu, C. Zhang, L. Zhang, X. Ma, and Z. Dong, "Promoting and impeding factors for exercise in patients with osteopenia or osteoporosis: A systematic review," *Int. J. Nurs. Stud. Adv.*, vol. 10, p. 100456, 2026, doi: 10.1016/j.ijnsa.2025.100456.
- [5] J. Huang, Q. Tan, H. Bai, J. Wang, M. Bergman, and Z. Wu, "Bone mineral density, osteopenia and osteoporosis among US adults with cancer," no. January, pp. 653–660, 2022, doi: 10.1093/qjmed/hcac015.
- [6] S. M. Naguib, M. K. Saleh, H. M. Hamza, and K. M. Hosny, "A new superfluity deep learning model for detecting knee osteoporosis and osteopenia in X-ray images," 2025, doi: 10.1038/s41598-024-75549-0 1.
- [7] D. Lee and M. Joo, "Change in Bone Mineral Density in Stroke Patients with Osteoporosis or Osteopenia," 2022, doi: 10.3390/ijerph19158954.
- [8] A. M. Sarhan *et al.*, *Knee Osteoporosis Diagnosis Based on Deep Learning*, vol. 4. Springer Netherlands, 2024, doi: 10.1007/s44196-024-00615-4.
- [9] Y. Fan *et al.*, "Sex- and Age-Specific Prevalence of Osteopenia and Osteoporosis: Sampling Survey," vol. 10, pp. 1–15, 2026, doi: 10.2196/48947.
- [10] B. P. P. Suryana, M. R. Febriliant, R. Rosandi, Sukarlin, and W. Lestari, "Factors Associated with Changes in Bone-Mass Density in Female Patients with Osteopenia Post Menopause," vol. 2, no. 2, pp. 177–186, 2021, doi: 10.21776/ub.crjim.2021.002.02.3.
- [11] C. Qiu, K. Su, Z. Luo, Q. Tian, and L. Zhao, "Developing and comparing deep learning and machine learning algorithms for osteoporosis risk prediction," no. June, pp. 1–13, 2024, doi: 10.3389/frai.2024.1355287.
- [12] M. B. Qureshi *et al.*, "Deep - learning based osteoporosis classification in knee X - rays using transfer - learning approach," pp. 1–16, 2025, doi: 10.1038/s41598-025-24338-4 1.
- [13] M. Shen, *Utilizing Deep Learning for Osteoporosis Diagnosis through Knee X-Ray Analysis*, no. Icaic. Atlantis Press International BV, 2024, doi: 10.2991/978-94-6463-512-6_58.
- [14] S. Feng, S. Lin, Y. Chiang, M. Lu, and Y. Chao, "Deep Learning-Based Hip X-ray Image Analysis for Predicting Osteoporosis," 2024, doi: 10.3390/app14010133.
- [15] A. Siddiqua and R. Hasan, "Computer-Aided Osteoporosis Diagnosis Using Transfer Learning with Enhanced Features from Stacked Deep Learning Modules," vol. 4, pp. 1–20, 2024, doi: 10.48550/arXiv.2412.09330.
- [16] N. Duklan, S. Kumar, H. Maheshwari, R. Singh, S. D. Sharma, and S. Swami, "CNN Architectures for Image Classification: A Comparative Study Using ResNet50V2, ResNet152V2, InceptionV3, Xception, and MobileNetV2," vol. 11, no. 9, pp. 11–21, 2024, doi: 10.14445/23488549/IJECE-V11I9P102.
- [17] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. de la T. Diez, and I. Ashraf, "Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network," *Diagnostics*, vol. 12, no. 5, pp. 1–16, 2022, doi: 10.3390/diagnostics12051280.
- [18] K. Shaheed, Q. Abbas, A. Hussain, and I. Qureshi, "Optimized Xception Learning Model and XgBoost Classifier for Detection of Multiclass Chest Disease from X-ray Images," 2023, doi: 10.3390/diagnostics13152583.
- [19] A. A. Asiri *et al.*, "Brain Tumor Detection and Classification Using Fine-Tuned CNN with ResNet50 and U-Net Model: A Study on TCGA-LGG and TCIA Dataset for MRI Applications," 2023, doi: 10.3390/life13071449.
- [20] F. Salim, F. Saeed, S. Basurra, and S. N. Qasem, "DenseNet-201 and Xception Pre-Trained Deep Learning Models for Fruit Recognition," 2023.
- [21] J. Kim, "Improvement of inceptionv3 model classification performance using chest x-ray images," vol. 22, no. 8, pp. 1–11, 2022, doi: 10.3390/electronics12143132.
- [22] M. Elpeltagy and H. Sallam, "Automatic prediction of COVID- 19 from chest images using modified ResNet50," *Multimed. Tools Appl.*, vol. 80, no. 17, pp. 26451–26463, 2021, doi: 10.1007/s11042-021-10783-6.
- [23] U. B. Abubakar, M. M. Boukar, and S. Adeshina, "Evaluation of Parameter Fine-Tuning with Transfer Learning for Osteoporosis Classification in Knee Radiograph," vol. 13, no. 8, pp. 246–252, 2022, doi: 10.14569/IJACSA.2022.0130829.
- [24] Y. Chen, W. P. Chan, H. Zhang, Z. Tsai, and H. Peng, "Automated osteoporosis classification and T -score prediction using hip radiographs via deep learning algorithm," pp. 1–11, 2024, doi: 10.1177/1759720X241237872.
- [25] W. Ong *et al.*, "Artificial Intelligence Applications for Osteoporosis Classification Using Computed Tomography," pp. 1–26, 2023, doi: 10.3390/bioengineering10121364.
- [26] L. Mao *et al.*, "Deep learning for screening primary osteopenia and osteoporosis using spine radiographs and patient clinical covariates in a Chinese population," no. September, pp. 1–11, 2022, doi: 10.3389/fendo.2022.971877.
- [27] M. A. Rahman and B. M. T. Haque, "Quantitative Benchmarking of Machine Learning Models for Risk Prediction: A Comparative Study Using AUC/F1 Metrics and Robustness Testing," *Applied Science and Technology*, vol. 1, no. 3, pp. 32–60, 2022, doi: 10.63125/9hd4e011.
- [28] A. Shazia, T. Z. Xuan, J. H. Chuah, J. Usman, P. Qian, and K. W. Lai, "A comparative study of multiple neural network for detection of COVID-19 on chest X-ray," vol. 1, pp. 1–16, 2021, doi: 10.1186/s13634-021-00755-1.
- [29] M. Huang and Y. Liao, "A lightweight CNN-based network on COVID-19 detection using X-ray and CT images," *Comput. Biol. Med.*, vol. 146, no. March, p. 105604, 2022, doi: 10.1016/j.compbiomed.2022.105604.
- [30] T. N. Abu-jamie and P. S. S. Abu-naser, "Classification of Sign-Language Using Deep Learning - A Comparison between Inception and Xception models," vol. 6, no. 8, pp. 9–19, 2022.
- [31] C. X. Ling, J. Huang, and H. Zhang, "AUC: a Statistically Consistent and more Discriminating Measure than Accuracy," 2023.
- [32] W. Islam, M. Jones, R. Faiz, N. Sadeghipour, Y. Qiu, and B. Zheng, "Improving Performance of Breast Lesion Classification Using a ResNet50 Model Optimized with a Novel Attention Mechanism,"

- Tomography*, vol. 8, no. 5, pp. 2411–2425, 2022, doi: 10.3390/tomography8050200.
- [33] Y. Fan *et al.*, "Machine Learning: Using Xception, a Deep Convolutional Neural Network Architecture, to Implement Pectus Excavatum Diagnostic Tool from Frontal-View Chest X-rays," 2023, doi: 10.3390/biomedicines11030760.
 - [34] I. Khalil, A. Mehmood, and H. Kim, "OCTNet: A Modified Multi-Scale Attention Feature Fusion Network with InceptionV3 for Retinal OCT Image Classification," pp. 1–19, 2024, doi: 10.3390/math12193003.
 - [35] D. Balampanos *et al.*, "Interpretable Machine Learning for Osteopenia Detection: A Proof-of-Concept Study Using Bioelectrical Impedance in Perimenopausal Women," *Journal of Functional Morphology and Kinesiology*, 2025, doi: 10.3390/jfmk10030262.

Rani Puspita received the Master of Computer Science degree from Bina Nusantara University, Jakarta, Indonesia, in 2022. Her major field of study is computer science. She is currently a Lecturer with the Computer Science Department, School of Computer Science, Bina Nusantara University. Her research interests include artificial intelligence, computer vision, natural language processing, and data mining. She actively contributes to academic research and publishes her work in Scopus-indexed journals and conferences. Ms. Rani Puspita can be contacted at email: rani.puspita@binus.ac.id.

Asysta Amalia Pasaribu received the Master of Actuarial Science from Institut Teknologi Bandung (ITB), Bandung, Indonesia, in 2021. Her major field of study is applied statistics. Currently, she is Lecturer in Department of Statistics, School of Computer Science, Bina Nusantara University, Indonesia. Her research interests include volatility modeling, small area estimation, Hierarchical Bayesian, and Machine Learning. Ms. Asysta Amalia Pasaribu can be contacted at email: asysta.amalia@binus.ac.id.

Neni Nur Laili Ersela Zain is a Lecturer of Mathematics at the School of Computer Science, Binus University. She earned her Bachelor's degree in Mathematics from Universitas Airlangga and her Master's degree from Sepuluh Nopember Institute of Technology (ITS). Currently, she is pursuing her Doctoral (Ph.D.) degree in Mathematics at Universitas Padjadjaran. Her areas of expertise encompass Computer Vision, Artificial Intelligence, Robust Optimization, and Uncertainty Problems. She actively publishes her research in both nationally accredited (SINTA) and internationally indexed (Scopus) journals. In addition to peer-reviewed articles, she has authored several books, including *Buku Ajar Logika & Algoritma*, *Algoritma Pembelajaran Mesin: Dasar, Teknik, dan Aplikasi*, and *Image Processing dan Artificial Intelligence: Teori, Konsep, dan Implementasi di Berbagai Bidang*. Currently, her ongoing doctoral research is titled "Robust Optimization Model for Two-Photon Microscopy Image Denoising in U-Net Neighbor2Neighbor Networks." Her primary research focus lies in integrating uncertainty optimization into deep learning frameworks for image processing, supported by both mathematical formulation and computational implementation on real-world data. Ms. Neni Nur Laili Ersela Zain can be contacted at email: neni.zain@binus.ac.id.