

A Hybrid Attention-Driven Deep Learning Model for Osteoporosis Detection in Knees

Ishaq Muhammad

Dept of Information and Communication Engineering
Chosun University
Gwangju, South Korea
ishaq.k@chosun.ac.kr

Bumshik Lee

Dept of Information and Communication Engineering
Chosun University
Gwangju, South Korea
bslee@chosun.ac.kr

Abstract—Osteoporosis is a prevalent bone disease characterized by reduced bone density and increased fracture risk, particularly among older adults and postmenopausal women. Early detection and accurate diagnosis are crucial to mitigate the morbidity and mortality associated with osteoporotic fractures. This paper proposes an advanced deep-learning methodology to enhance the accuracy and efficiency of osteoporosis detection using knee osteoporotic data. Our approach integrates features from pre-trained model, DenseNet169 and a custom designed attention model. The fused feature set from these models is then passed through a fully connected neural network to classify images as either osteoporotic or normal. The results of our proposed model demonstrated significant improvements in classification accuracy, achieving a high accuracy rate on unseen test data. Integrating DenseNet169, and attention mechanisms, effectively captured comprehensive features, resulting in robust and reliable osteoporosis detection. Moreover, the model outperformed current state-of-the-art methods for classification tasks, highlighting its potential for early intervention in clinical settings.

Keywords— *DenseNet169, Attention Model, Knees Dataset, Osteoporosis classification.*

I. INTRODUCTION

Osteoporosis is a pervasive and debilitating bone disease characterized by decreased bone density and increased fracture risk [1]. It primarily affects older adults, particularly postmenopausal women, leading to significant morbidity and mortality [2] [3]. The clinical importance of osteoporosis lies in its potential to cause fractures, particularly in the hip, spine, and wrist, which can result in chronic pain, disability, and even death [4] [5]. The incidence of osteoporosis and associated fractures is predicted to increase as the world's population ages, highlighting the critical need for efficient diagnosis and treatment methods.

Detecting and diagnosing osteoporosis early is crucial to prevent fractures and their associated complications [6]. The gold standard for determining bone mineral density (BMD) is dual-energy X-ray absorptiometry (DXA), although these conventional techniques are frequently constrained by radiation dose, cost, and accessibility. [7]. Additionally, these methods may not fully capture the complex structural changes in bone architecture associated with osteoporosis. Therefore, there is a critical need for more advanced, accurate, and accessible methods for early detection and monitoring of osteoporosis.

This paper aims to explore the use of deep-learning in detecting abnormalities in the knee osteoporotic dataset KOP [8], contributing to the broader effort to enhance osteoporosis detection and treatment. By focusing on the knee, an area particularly vulnerable to osteoporotic fractures, this research

seeks to improve our understanding of osteoporosis progression and develop more targeted intervention strategies. Our contribution is mentioned below.

- We propose a deep-learning-based approach that combines DenseNet169 and a custom-designed Attention Model (AM) for enhanced osteoporosis detection.
- A custom AM is designed, incorporating channel and spatial attention mechanisms to capture detailed spatial and channel-specific information from input images.
- The fusion of features from both models creates a comprehensive representation of input images, leading to more accurate classification.
- The fully connected (FC) neural network is meticulously tuned through extensive experimentation, enhancing the model's overall performance.

II. PROPOSED METHODOLOGY

The proposed methodology aims to leverage advanced deep learning techniques to significantly improve the accuracy and efficiency of osteoporosis detection and classification using KOP. This new approach uses cutting-edge deep learning techniques to overcome the limitations of conventional diagnostic techniques. The overall model architecture is illustrated in Figure 1. We prepare the data by sourcing and augmenting the dataset to enhance model performance. The dataset, obtained from Kaggle, comprises 186 normal and 186 osteoporotic images. Each image is augmented to create five additional versions, resulting in a significantly larger dataset of 1800 training images. The images are divided into training, validation, and testing sets. The augmentation steps include zooming, cropping regions of interest (ROIs), and rotating the images. The testing set, consisting of 72 unseen images, is used for the final evaluation of the model.

A. Feature Extraction

The feature extraction involves the usage of pre-trained model: DenseNet169 [9] and AM.

The DenseNet169, a convolutional neural network (CNN) [11] known for its unique architecture, features dense connectivity where each layer is connected to every other layer in a feed-forward fashion. This dense connectivity mitigates the vanishing gradient problem and improves the flow of information throughout the network. For feature extraction, DenseNet169 is employed by retaining its feature extraction layers and removing the final classification layer. This step produces a fixed-size feature vector of dimension 1664, independent of the original image dimensions.

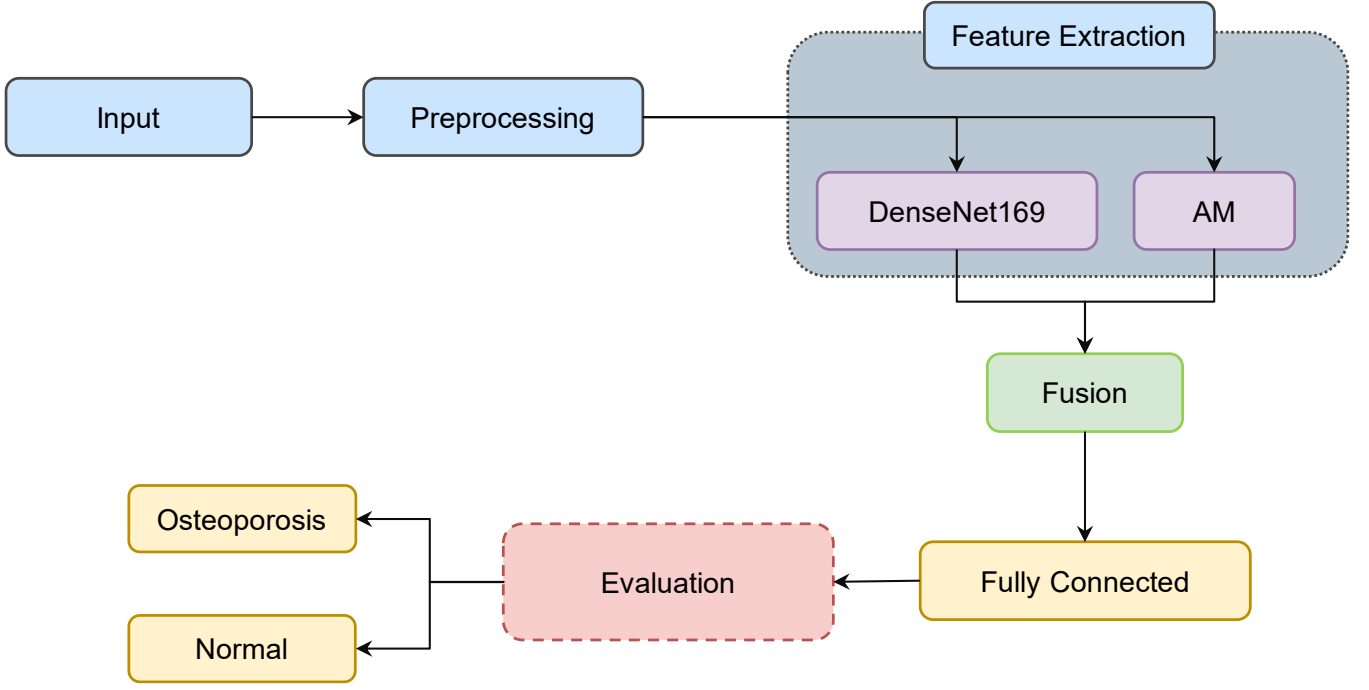


Fig. 1. Overall architecture of the proposed model for osteoporosis detection

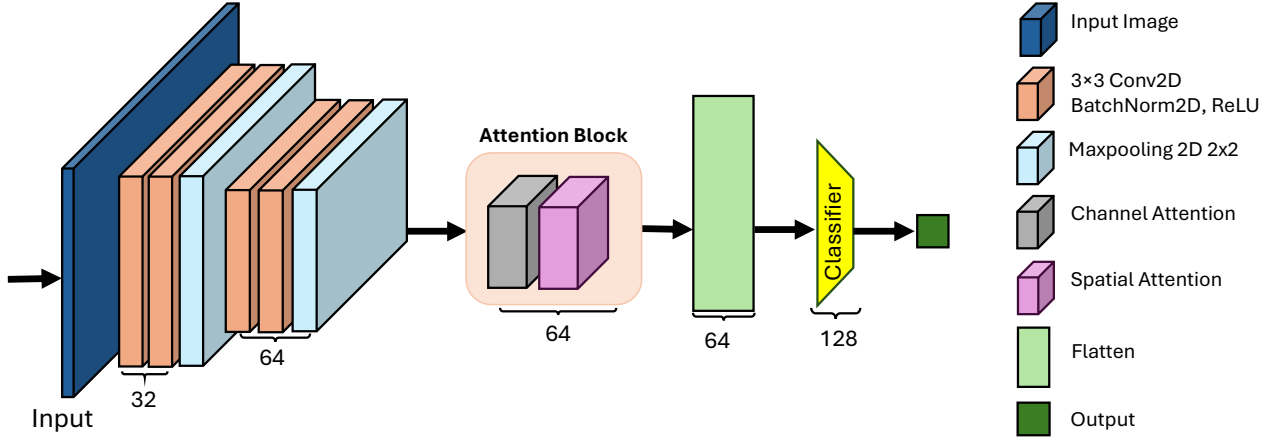


Fig. 2. Architecture of the proposed AM.

The AM block, a core component of our model, is designed to detect osteoporosis by capturing detailed spatial and channel information from knee radiographs. Using multiple convolution layers and attention mechanisms, the model is fine-tuned to highlight critical features indicating osteoporosis, such as cortical thinning, disrupted trabecular patterns, and micro-cracks. The AM employs channel attention to emphasize important channels representing these osteoporotic features and spatial attention to focus on key areas of the image where abnormalities, such as bone porosity or fractures, are likely to appear. Tailored to the specific dataset of knee radiographs, our model is optimized for identifying osteoporotic characteristics. The spatial attention mechanism concatenates pooled feature maps and applies a convolutional layer to ensure the network effectively focuses on regions showing signs of osteoporosis, such as irregular bone structures, cracks, or density loss. Meanwhile, the channel attention mechanism enhances the model's ability to

prioritize features that distinguish osteoporotic bones from normal ones, reinforcing focus on relevant areas while suppressing less informative regions.

The block comprises four convolutional layers with ReLU activations and max-pooling. The first two layers use 32 filters (3×3), while the latter two use 64 filters (3×3). Max-pooling reduces feature map dimensions, ensuring a balance between computational efficiency and detail retention. Channel attention employs adaptive average and max pooling followed by a shared MLP to emphasize significant channels, while spatial attention combines pooled feature maps, applying a convolutional layer to highlight key regions. Together, these mechanisms refine feature representation for robust detection. Fig 2 illustrates the architecture of AM.

Following the feature fusion step, the FC layer is meticulously designed and tuned for optimal performance. Extensive experimentation led to the following configuration:

TABLE I. SUMMARY OF THE PROPOSED RESULTS COMPARED TO BASELINE MODELS

Models	Sensitivity	Specificity	Precision	FPR	Accuracy	F1 Score	Loss
VGG19 [12]	0.8889	0.7143	0.9211	0.2857	0.7763	0.8046	0.6294
EfficientNetB0 [14]	0.9259	0.7347	0.8774	0.2643	0.8026	0.8276	0.6029
Densenet169 [9]	0.8571	0.8049	0.8684	0.1951	0.8289	0.8354	0.5892
ViT [10]	0.9444	0.6852	0.7782	0.3148	0.7500	0.8043	0.6498
ResNet50[13]	0.8571	0.7083	0.8947	0.2917	0.7632	0.7907	0.6236
Ours	0.8293	0.8710	0.8947	0.1290	0.8472	0.8608	0.5834

The input to the FC layer is a concatenated feature vector of size 1792, combining outputs from DenseNet169 (1664 features), and AM (128 features). The first layer reduces the input size from 1792 to 1024 neurons. It includes batch normalization to stabilize learning, a dropout layer with a rate of 0.5 to avoid overfitting and a ReLU activation function to introduce non-linearity. The second layer further reduces the size from 1024 to 512 neurons, with batch normalization, 0.3 dropout rate, and ReLU activation. The third layer has a dropout rate of 0.2, batch normalization, ReLU activation, and a reduction in size from 512 to 256 neurons. The fourth layer has a dropout rate of 0.1, batch normalization, ReLU activation, and a reduction in size from 256 to 128. The final output layer reduces the size from 128 to 1 neuron, using a sigmoid activation function to produce a binary classification output indicating the presence of osteoporosis or a normal condition. This configuration of layers, channels, and dropout rates was determined to be optimal after extensive experimentation and tuning, ensuring that the model effectively captures and leverages the features extracted from the input images, providing a robust and accurate classification of osteoporosis.

Combining deep feature extraction with attention processes improves the model's capacity to identify osteoporosis. The convolutional layers capture features from low to high levels across the radiograph, while the attention mechanisms selectively enhance ROIs. The FC network then integrates and processes these features to produce an accurate classification of osteoporosis presence or absence. By tailoring the model specifically to this dataset, it becomes highly sensitive to the subtle variations in BMD and structures critical for identifying osteoporosis.

III. EXPERIMENTAL RESULTS

In our experiments, we used several performance metrics to evaluate the models. These metrics thoroughly assess the models' effectiveness in detecting osteoporosis. These metrics include Accuracy, F1 score, loss, and precision. Table 2 provides the summary of the performance metrics used to evaluate the performance of our model.

TABLE II. SUMMARY OF THE PERFORMANCE METRICS

Metrics	Quantity
Sensitivity	$TPR = TP / (TP + FN)$
Specificity	$SPC = TN / (FP + TN)$
Precision	$PPV = TP / (TP + FP)$
False Positive Rate	$FPR = FP / (FP + TN)$
Accuracy	$ACC = (TP + TN) / (P + N)$
F1 Score	$F1 = 2TP / (2TP + FP + FN)$

Since this a classification task so our model heavily relies on F1 score, recall and accuracy. Among the models tested, ViT achieved the highest sensitivity at 0.9444, indicating its superior ability to correctly identify positive cases. VGG19, on the other hand, excelled in precision with a score of 0.9211, reflecting its accuracy in predicting true positives. ResNet50 showed balanced performance with a good trade-off between sensitivity and precision, while Densenet169 and EfficientNetB0 resulted in a balance performance throughout. Our proposed model demonstrated a strong overall performance, with a notable accuracy of 0.8472 and the F1 score of 0.8608, our model also exhibited the lowest loss value of 0.5834 and superior specificity of 0.8710 indicating its robustness and reliability in detecting osteoporosis. This balance of metrics highlights the effectiveness of our approach in maintaining both sensitivity and precision while minimizing prediction errors. Table 1 provides the overall results of our models while Fig 3 and 4 shows the overall training and validation accuracy and loss respectively.

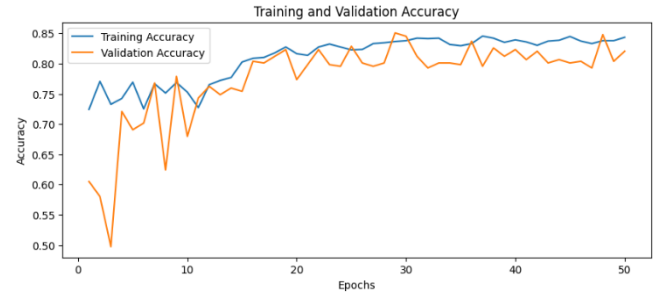


Fig. 3. Training and validation accuracy



Fig. 4. Training and validation loss

A. Confusion Matrix Analysis

In statistics, a True Positive (TP) occurs when a model correctly predicts a positive result for a positive case. A True Negative (TN) occurs when a model accurately forecasts a negative result for a truly negative case. Conversely, a False Negative (FN) occurs when the model predicts a negative outcome for an instance that is actually positive, and a False

Positive (FP) occurs when the model predicts a positive outcome for an instance that is actually negative. The confusion matrix for the suggested model, as seen in Fig. 5, summarizes the performance for the classification of the two classes by illuminating the link between the actual and anticipated values.. The plot shows that TP is 34, FP is 4, FN is 7, and TN is 27.

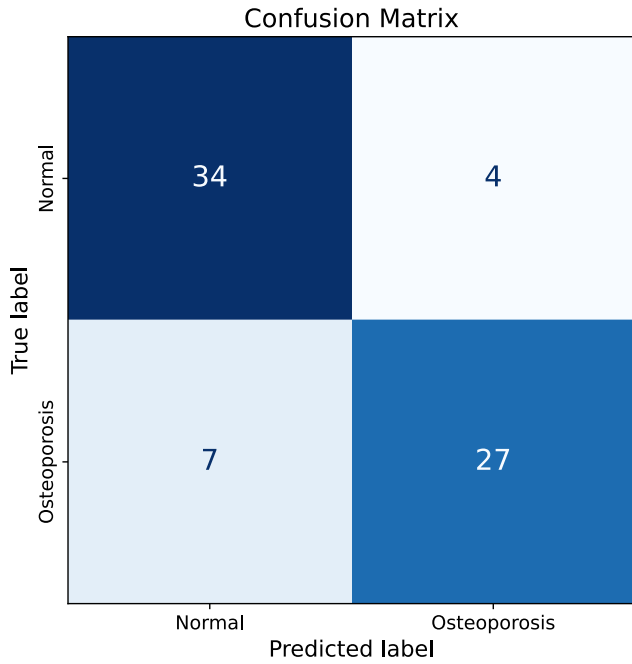


Fig. 5. Confusion matrix of proposed model.

IV. CONCLUSION

In this study, we developed a deep-learning-based approach for osteoporosis detection using the KOP dataset. Our methodology leverages the feature extraction capabilities of DenseNet169 and a custom-designed AM. This combined approach demonstrated superior performance in classifying osteoporotic and normal conditions. The fusion of DenseNet169's robust feature extraction and the AM's spatial and channel attention mechanisms, which emphasize critical regions and features indicative of osteoporosis, provided comprehensive image representation, resulting in improved classification performance. Our results highlight the effectiveness of integrating pre-trained models with attention-enhanced blocks for medical image analysis, offering a significant improvement over current SOTA techniques. This method holds promise for clinical applications in osteoporosis diagnosis, enabling earlier detection and better-targeted interventions, potentially reducing the morbidity and mortality associated with osteoporotic fractures. Future work may involve further refinement of the model, larger datasets, and its application to other bone-related pathologies for broader clinical adoption.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00217471) and by the National Research Foundation of Korea (NRF) funded by the Korean government under Grant 2022R1I1A3065473.

REFERENCES

- [1] Coll, P. P., Phu, S., Hajjar, S. H., Kirk, B., Duque, G., & Taxel, P. (2021). The prevention of osteoporosis and sarcopenia in older adults. *Journal of the American Geriatrics Society*, 69(5), 1388–1398.
- [2] Rozenberg, S., Al-Daghri, N., Aubertin-Leheudre, M., Brandi, M. L., Cano, A., Collins, P., Cooper, C., Genazzani, A. R., Hillard, T., Kanis, J. A., Kaufman, J. M., Lambrinoudaki, I., Laslop, A., McCloskey, E., Palacios, S., Prieto-Alhambra, D., Reginster, J. Y., Rizzoli, R., Rosano, G., Trémollières, F., & Harvey, N. C. (2020). Is there a role for menopausal hormone therapy in the management of postmenopausal osteoporosis? *Osteoporosis International*, 31(12), 2271–2286.
- [3] Fairhall, N. J., Dyer, S. M., Mak, J. C., Diong, J., Kwok, W. S., & Sherrington, C. (2022). hip fracture surgery in adults. *Cochrane base of Systematic Reviews*, 2022(9). CD001704.
- [4] Papapoulos, S., Bone, H., Cosman, F., Dempster, D. W., McClung, M. R., Nakamura, T., Restrepo, J. F. M., Bouxsein, M. L., Cohn, D., de Papp, A., Massaad, R., & Santora, A. (2021). Incidence of hip and subtrochanteric/femoral shaft fractures in postmenopausal women with osteoporosis in the phase 3 long-term odanacatib fracture trial. *Journal of Bone and Mineral Research*, 36(7), 1225–1234.
- [5] Jiang, Y. W., Xu, X. J., Wang, R., & Chen, C. M. (2022). Radiomics analysis based on lumbar spine CT to detect osteoporosis. *European Radiology*, 32(11), 8019–8026.
- [6] Ahern DP, et al. A meta-analysis of the diagnostic accuracy of Hounsfield units on computed topography relative to dual-energy X-ray absorptiometry for the diagnosis of osteoporosis in the spine surgery population. *Spine Journal* Oct. 01, 2021;21(10):1738–49.
- [7] Albano, D., Messina, C., Vitale, J., & Sconfienza, L. M. (2020). Imaging of sarcopenia: Old evidence and new insights. *European Radiology*, 30(4), 2199–2208.
- [8] Steve Python, "Osteoporosis Knee X-ray Dataset," Available Online: <https://www.kaggle.com/datasets/stevepython/osteoporosis-knee-xray-dataset>.
- [9] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).
- [10] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [11] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12), 6999–7019.
- [12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [14] M. Tan, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.