Feature-Driven Deep Learning for Osteoporosis Detection

A K M Salman Hosain[1*]

Corresponding author(s). E-mail(s): ahosain@charlotte.com;

1 Introduction

Osteoporosis is a prevalent global bone disorder associated with fractures [1]. Although it is more commonly observed in older individuals, younger populations are also susceptible to this debilitating disease [2]. Osteoporosis disrupts bone homeostasis, resulting in reduced bone mass, compromised quality, and increased risk of fracture. Hormones, cytokines, and growth factors intricately regulate bone health, influencing peak bone mass through coordinated mechanisms. Dysregulation in these processes is implicated in osteoporosis pathogenesis [3–7].

While several studies have successfully applied machine learning and deep learning techniques to medical image classification, the field of osteoporosis detection using knee X-rays has been relatively underexplored. Prior research has primarily focused on transfer learning methods, utilizing models such as AlexNet, VGGNet, and ResNet for classifying osteoporosis based on hip X-rays, but knee X-rays, which are crucial in diagnosing osteoporosis, have received less attention. Moreover, most of these studies did not incorporate explainable AI methods, making the models' decision processes opaque, which limits their practical use in clinical settings. The lack of transparency (the "black box" issue) in deep learning models for medical diagnostics is a significant barrier to their adoption by healthcare professionals. Additionally, earlier works did not emphasize feature-focused data preprocessing techniques, which can significantly improve model performance by highlighting the critical areas in medical images. This research aims to bridge these gaps by employing explainable AI techniques such as LIME (Local Interpretable Model-agnostic Explanations) and Grad-CAM (Gradient-weighted Class Activation Mapping) alongside state-of-the-art CNN architectures to not only improve osteoporosis detection accuracy but also to make the models more interpretable.

The contributions of this work are:

Table 1: Summary of Literature Review

Paper	Method	Dataset	Accuracy	XAI	Disease Detected
Wani et al. [8]	AlexNet, VGGNet-16, ResNet, VGGNet-19 (Transfer Learning)	381 knee X-rays	91.1% (AlexNet)	No	Osteoporosis, Osteopenia
Abubakar et al. [9]	VGG 16 (Transfer Learning)	323 osteoporosis knee X-rays, 323 normal knee X-rays	88%	No	Osteoporosis
Klontzas et al. [10]	VGG-16, Inception-ResNetV2, InceptionV3 (Ensemble Learning)	210 avascular necrosis, 210 transient osteoporosis hip MRIs	97.62% (Inception-ResNet-V2)	No	Transient Osteoporosis
Kumar et al. [11]	Fuzzy rank-based unification of CNN classifiers (Ensemble Learning)	X-Rays from 240 subjects (37 normal, 154 osteopenia, 49 osteoporosis)	93.5%	No	Osteoporosis, Osteopenia
Khanna et al. [12]	Forward Feature Selection, Multi-level ensemble learning stack	1,493 patient records	89%	Yes	Osteoporosis
Suh et al. [13]	Interpretable deep-learning (DL) with XAI (LIME, SHAP)	NHANES, KNHANES datasets	AUC: 0.851 (femoral neck), 0.922 (total femur)	Yes	Osteoporosis (Femoral Neck, Total Femur)
Jang et al. [14]	VGG16 with non-local neural network, Grad-CAM (XAI)	1,001 hip X-rays	81.2% (Accuracy), AUC 0.867	Yes	Osteoporosis

- **Proposed Deep Learning Framework**: Introduce a deep learning pipeline utilizing transfer learning method for detecting osteoporosis from knee X-ray images.
- **Explainable AI (XAI) Implementation**: Address the 'black box' issue of machine learning models by incorporating explainable AI
- **Feature-focused Dataset Creation**: Conduct feature extraction method and propose a dataset for improved quantitative performance parameters
- **Comparative Analysis of Models**: Conduct a comparative analysis of models with original and proposed dataset to showcase performance improvement across various quantitative parameters

2 Methodology

In this work, we propose a novel pipeline to detect osteoporosis from knee X-Ray images employing transfer learning method with EffecientNetB0 [15], and MobileNet [16]. We use XAI architecture, LIME [17] to differentiate the features that our models focus primarily on the classification task. Subsequently, we implement a feature extraction method to isolate the prominent features from the original images, creating a separate dataset, and train our models on this extracted feature dataset to enhance classification accuracy. We utilize various numerical parameters to demonstrate the improvement of model accuracy by adopting our proposed framework. The complete framework is depicted in Fig. 1.

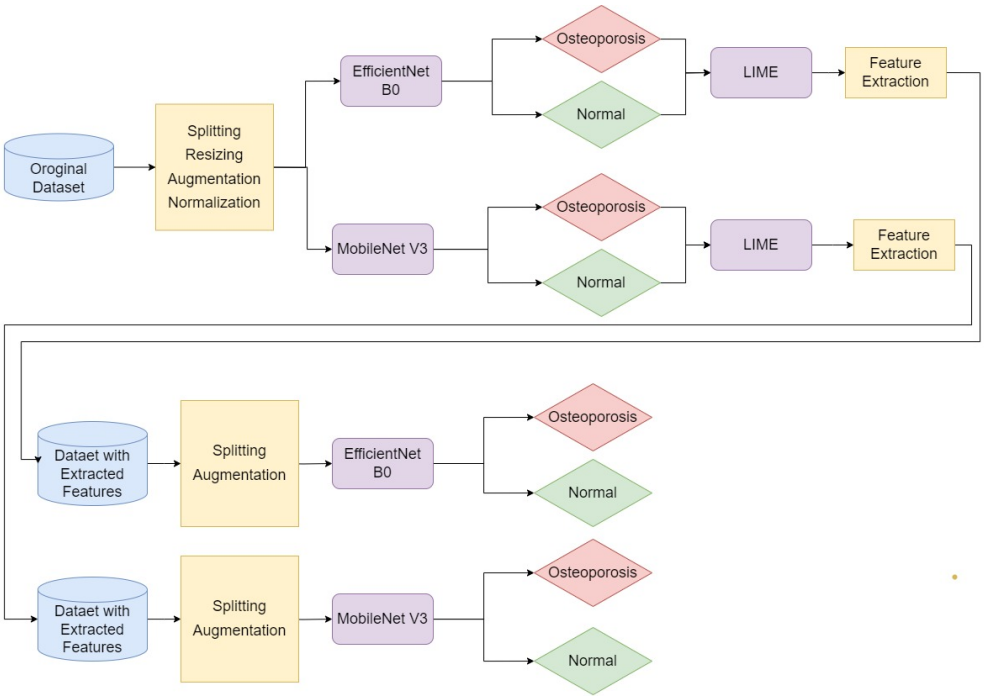


Fig. 1: Proposed framework to detect osteoporosis from Knee X-Ray images.

2.1 Dataset Description

We utilize X-Ray images from available dataset in Kaggle [18]. The dataset contains images of knee radio-graphs of two categories: normal, and osteoporosis. A total of 774 images, distributed equally across two categories. The images are of varying dimensions. Representative samples from the dataset are shown in Fig. 2.

(a) Normal Knee X-Ray

(b) Osteoporosis Knee X-Ray

Fig. 2: Sample images from dataset.

2.1.1 Dataset Pre-processing

We split the dataset into three sets: training, validation, and test by the ration 7:2:1. The training set contains a total of 520 images, the validation set 150 images and the test 74 images.

To ensure uniformity and increase model memory and computational efficiency, we resize all the images to 224×224 pixels. We further augment the images to increase the number of training images. The augmentation parameters we adopt include randomly rotating images by up to 20 degrees, randomly shifting images horizontally by up to 20% of the width, randomly shifting images vertically by up to 20% of the height, randomly applying shear transformations by up to 20%, randomly zooming in/out on images by up to 20%, and randomly flipping images horizontally.

2.2 Model Description

The proposed framework utilizes a pipeline to detect osteoporosis from knee X-Ray images using transfer learning method with EffecientNetB0, and MobileNet.

2.2.1 EfficientNetB0

EfficientNet introduces a Convolutional Neural Network (CNN) architectures by leveraging compound scaling across width, depth, and resolution. This method ensures that the resulting models are efficient in terms of both computational resources and performance, making them highly suitable for deployment in real-world applications where efficiency and accuracy are critical factors [15].

The proposed CNN architecture utilizes the EfficientNetB0 model as a foundational feature extractor for image classification. This model is initialized with weights learned from extensive training on a large-scale image classification dataset, ImageNet, providing a robust starting point for feature extraction in our detection task.

In the feature extraction phase, the top layer of the EfficientNetB0 model is excluded, transforming it into a feature extractor. This modification enables the efficient extraction of high-level features from the input images, which are crucial for capturing the intricate patterns necessary for accurate classification.

Following feature extraction, custom classification layers are added to refine and classify the extracted features. The first layer is a Global Average Pooling layer, which serves to reduce the spatial dimensions while preserving the essential features of the input. This is followed by a Dense layer with 1024 units and a Rectified Linear Unit (ReLU) activation function. The final layer is a Dense layer with a softmax activation function.

The model compilation phase involves setting up the model for training by defining the loss function and the optimizer. Categorical crossentropy is selected as the loss function.

The proposed model is depicted in Fig. 3.

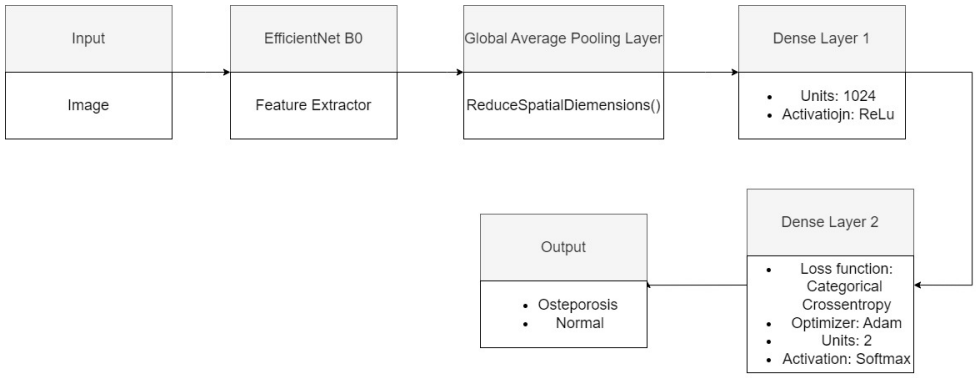


Fig. 3: Proposed model architecture based on EfficientNetB0 as encoder.

2.2.2 MobileNet V3

MobileNet is built on a streamlined architecture that employs separable convolutions to construct lightweight deep neural networks. This architecture splits the standard convolution operation into two distinct layers: a depthwise convolution and a pointwise convolution.

The architecture begins with an Input Layer that accepts input images followed by a Convolutional Layer, which consists of an initial standard convolution layer with a large number of filters, accompanied by Batch Normalization and ReLU activation. The ReLU activation function is applied to introduce non-linearity into the model. The core of MobileNet comprises Depthwise Separable Convolution Blocks. Each block contains a depthwise convolution followed by a pointwise convolution. These blocks significantly reduce the computational load and parameter count, forming the essential building units of MobileNet.

After the convolutional blocks, the architecture includes a Global Average Pooling layer. This layer reduces each feature map to a single value by averaging, thus decreasing the spatial dimensions and further reducing the computational complexity. Finally, a Fully Connected Layer is employed for classification tasks. This dense layer has a number of neurons equal to the number of output classes, completing the architecture and enabling the model to perform its classification functions efficiently.

The proposed model using MobileNet V3 as feature extractor is depicted in Fig. 4

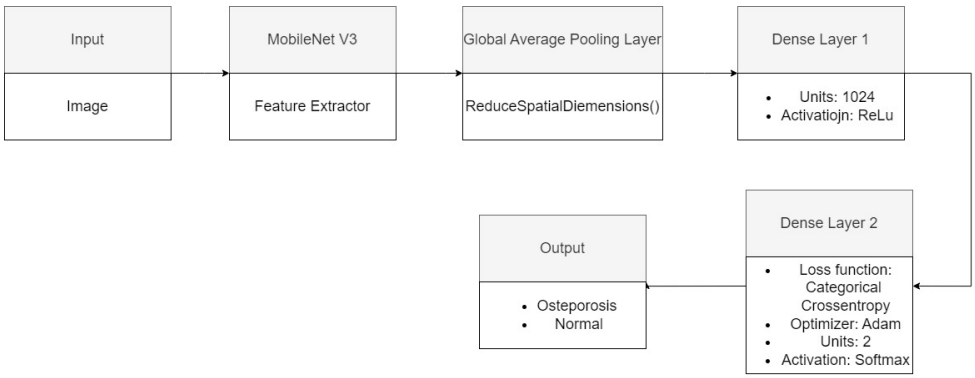


Fig. 4: Proposed model architecture based on MobileNet V3 as encoder.

2.3 Model Training

The models are initially trained on the original dataset, which consist of 520 images, for 30 epochs. Following the training, EfficientNetB0 achieves an accuracy of 87.88%, while MobileNet V3 attains an accuracy of 83.08%. For both models, the 'Adam' optimizer [19] and the 'Categorical Cross Entropy' loss function are utilized.

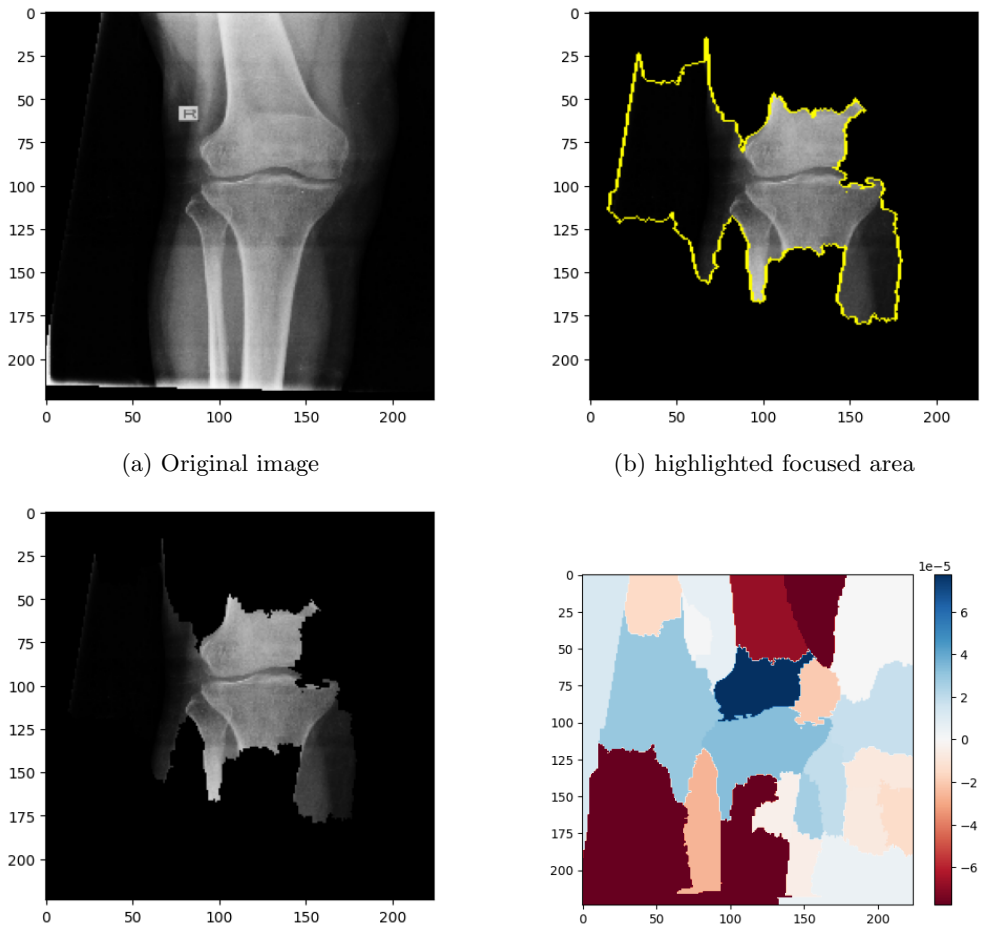
2.4 Explainable AI

Further to improve the models, we adopt two XAI techniques: LIME [17], and Grad-Cam [20]. After classifying the images into two classes with the models, we utilize these XAI architectures to address the black box issue of the deep learning models. It illustrates the regions of the images on which the models primarily focus for classification.

2.4.1 Local Interpretable Model-agnostic Explanations (LIME)

We employ LIME to illustrate the features of the images that the models predominantly utilized for classification. An example of such an image is presented in Fig. 7. This figure shows the region where the model with EfficientNetB0 feature extractor primarily focuses to classify the image.

Same process is followed for MobileNet V3 for an example image in Fig. 6.



(a) Original image

(b) highlighted focused area

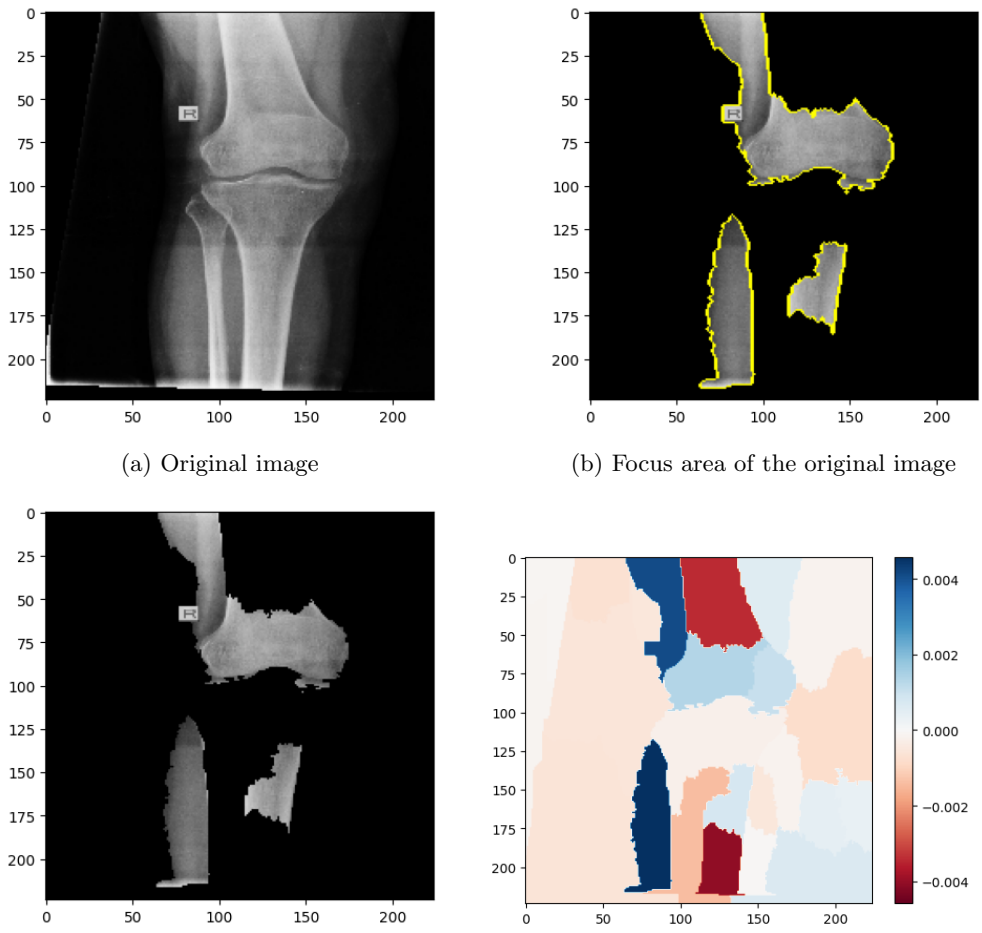
(c) Focus area of the original image

(d) Heatmap

Fig. 5: Focused area of an image with EfficientNetB0.

2.4.2 Gradient-weighted Class Activation Mapping (Grad-Cam)

Fig. 7 depicts the focused regions of the knee joint as highlighted by the models EfficientNetB0 and MobileNetV3 using Grad-CAM (Gradient-weighted Class Activation Mapping). Grad-CAM is a popular interpretability technique that provides visual explanations for deep learning models, allowing us to observe which areas of an image are most influential in the model's decision-making process. The sub-images (a), and (b) demonstrate how both models concentrate on certain areas of the X-ray for the classification task, which in this context is likely related to osteoporosis detection. These visualizations are crucial for validating and understanding the models' behavior in medical imaging.



(a) Original image

(b) Focus area of the original image

(c) highlighted focused area

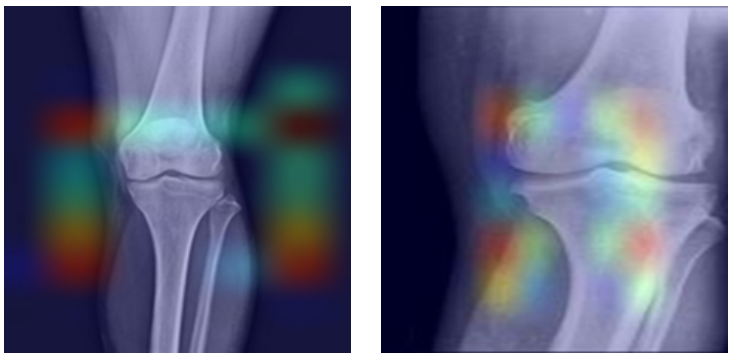
(d) heatmap

Fig. 6: Focused area of an image with MobileNet V3 on the same image as EfficienNet B0.

2.4.3 Dataset with Extracted Feature

We create two separate datasets by extracting key areas from the images for both models. We do this by extracting the mask of LIME multiplying with the original images to generate a new dataset from the original dataset. We follow the same process to resize, split and change statistic parameters for the images of the proposed dataset as the original dataset.

The training images of these datasets are augmented following the same procedure described in 2.1.1. Sample training, validation, and test images from this dataset are depicted in Fig. 8.



(a) Knee region focused by EfficientNet B0

(b) Knee region focused by MobileNet V3

Fig. 7: Focused knee region by the models depicted through GradCam.

3 Results

We conduct a comparative analysis of our models using various quantitative evaluation metrics, including accuracy, precision, recall, and F1 score. Initially, we assess the models' performance using the original dataset.

Subsequently, after generating a second dataset consisting of the focused regions, we conduct a second round of quantitative analysis. The comparative performance of our models between the original and proposed datasets is showcased in Table 2. In both cases, the test sets are used to measure performance.

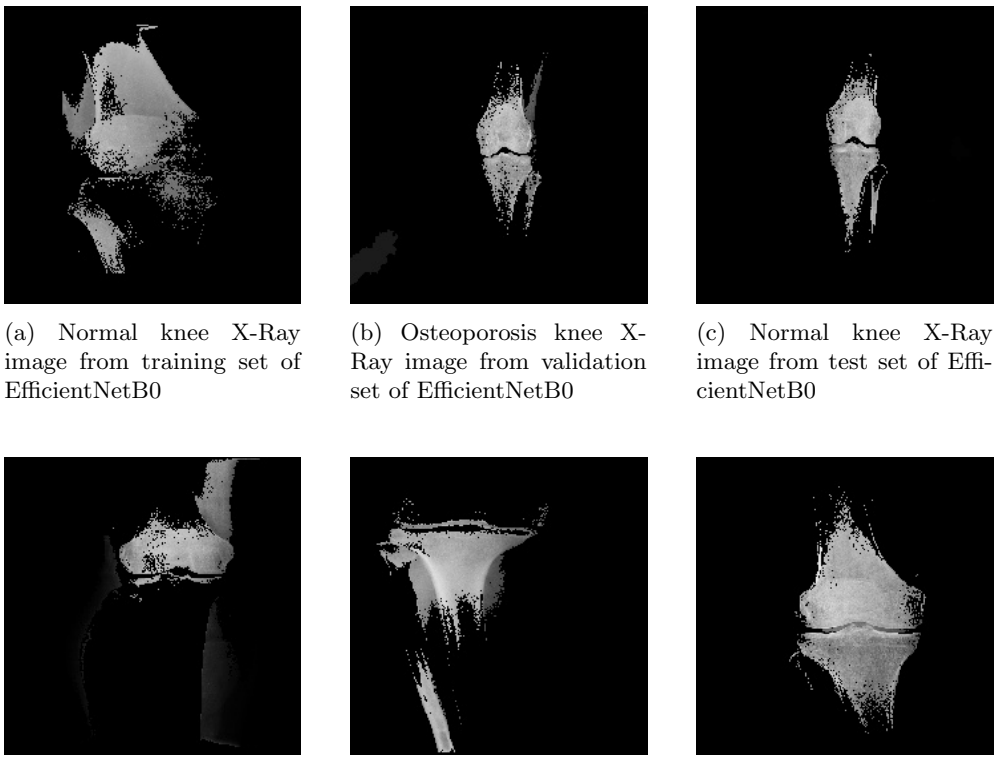
Table 2: Performance metrics for MobileNet v3 and EfficientNetB0 on the original and proposed datasets

Dataset	Model	Accuracy	Precision	Recall	F1 Score
Original	MobileNet v3	0.77	0.7083	0.9189	0.8000
	EfficientNetB0	0.77	0.7632	0.7838	0.7733
New	MobileNet v3	0.8648	0.9354	0.7837	0.8529
	EfficientNetB0	0.9189	0.9444	0.9189	0.9315

The comparative analysis of MobileNet v3 and EfficientNetB0 using various quantitative evaluation metrics on both the original and proposed datasets are shown in Table 2.

4 Discussion

The performance metrics in Table 2 demonstrate a significant improvement for both MobileNet v3 and EfficientNetB0 when evaluated on the new proposed dataset compared to the original dataset. On the original dataset, both models achieved the same



(a) Normal knee X-Ray image from training set of EfficientNetB0

(b) Osteoporosis knee X-Ray image from validation set of EfficientNetB0

(c) Normal knee X-Ray image from test set of EfficientNetB0

(d) Osteoporosis knee X-Ray image from training set of MobileNet V3

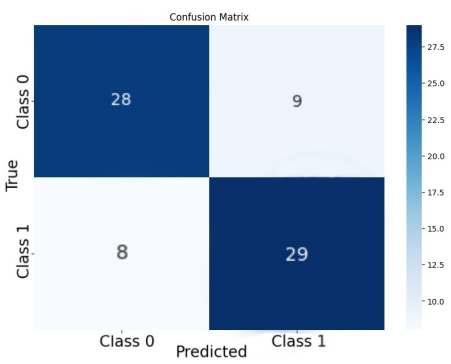
(e) Normal knee X-Ray image from validation set of MobileNet V3

(f) Osteoporosis knee X-Ray image from test set of MobileNet V3

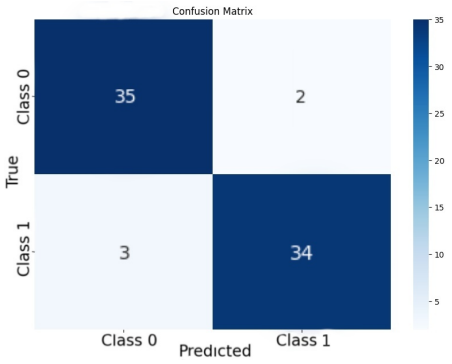
Fig. 8: Sample feature extracted images from proposed dataset.

accuracy of 77%, but EfficientNetB0 outperformed MobileNet v3 in terms of precision (76.32% vs. 70.83%) and F1 Score (77.33% vs. 80.00%), while MobileNet v3 exhibited a higher recall (91.89% vs. 78.38%). On the new dataset, both models showed substantial improvement, with MobileNet v3 achieving an accuracy of 86.48% (a 12.3% increase) and EfficientNetB0 reaching 91.89% (a 14.91% increase). EfficientNetB0 further outperformed MobileNet v3 in precision (94.44% vs. 93.54%), recall (91.89% vs. 78.37%), and F1 Score (93.15% vs. 85.29%), with percentage differences of approximately 0.96%, 17.3%, and 9.22%, respectively. The results underscore the effectiveness of the proposed dataset in enhancing model performance, particularly for EfficientNetB0, which exhibited superior metrics across the board, making it the more robust model for osteoporosis detection.

The confusion matrix shown in Fig. 9a represents the EfficientNetB0 for the original dataset. In the confusion matrix, class 0 represents normal knee X-Ray images, and class 1 represents osteoporosis knee X-Ray images. The EfficientNetB0 correctly classifies 28 normal knee X-Ray images while misclassifying nine normal knee X-Ray

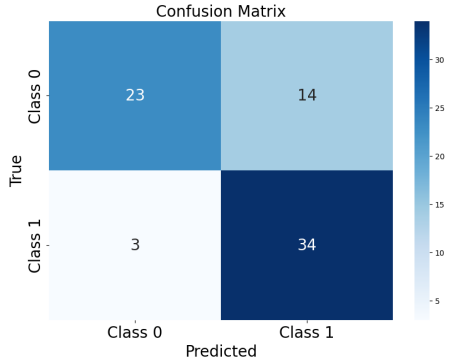


(a) Confusion matrix of EfficientNetB0 on original dataset

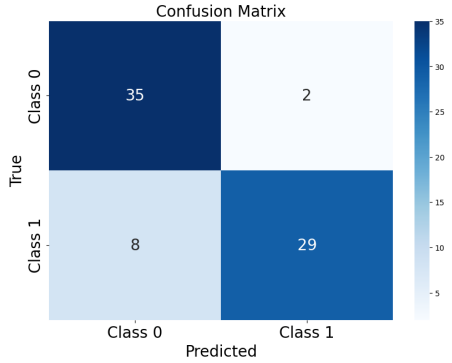


(b) Confusion matrix of EfficientNetB0 on proposed dataset

Fig. 9: Confusion matrix of EfficientNetB0.



(a) Confusion matrix of MobileNet V3 on original dataset



(b) Confusion matrix of MobileNet V3 on proposed dataset

Fig. 10: Confusion matrix of MobileNet V3.

images as osteoporosis images. Additionally, the model correctly classifies 29 osteoporosis knee X-Ray images and misclassifies eight osteoporosis knee X-Ray images as normal knee X-Ray images.

Fig. 9b. represents the confusion matrix for the proposed dataset. Here, EfficientNetB0 correctly classifies 35 normal knee X-Ray images (compared to 28 in the original test set) and misdiagnoses only two normal knee X-Ray images as osteoporosis (compared to nine in the original test set). Additionally, the model correctly classifies 34 osteoporosis knee X-Ray images and misclassifies three osteoporosis knee X-Ray images as normal knee X-Ray images (compared to 29 and eight, respectively, in the original test set).

Fig. 10a. shows the confusion matrix for MobileNet V3 applied to the original dataset. In this matrix, class 0 corresponds to normal knee X-Ray images, while class 1 corresponds to osteoporosis knee X-Ray images. The MobileNet V3 correctly identifies 23 normal knee X-Ray images but misclassifies 14 normal images as osteoporosis. Furthermore, the model accurately classifies 34 osteoporosis knee X-Ray images, with only three being incorrectly identified as normal.

Fig. 10b. illustrates the confusion matrix for the proposed dataset. For the proposed dataset, MobileNet V3 correctly classifies 32 normal knee X-Ray images (up from 23 in the original test set) and misclassifies only five normal knee X-Ray images as osteoporosis (compared to 14 in the original test set). Additionally, the model correctly identifies 29 osteoporosis knee X-Ray images and misclassifies eight osteoporosis images as normal (compared to 34 correct and three incorrect classifications in the original test set).

5 Conclusion

In conclusion, while both MobileNet v3 and EfficientNetB0 exhibited competent performance on the original dataset, the refined dataset allowed both models to achieve better results. EfficientNetB0 demonstrated superior performance across all metrics, making it the preferable choice for this classification task. These findings underscore the importance of both model selection and data preprocessing in enhancing the performance of deep learning models in medical imaging tasks.

6 Individual Contribution

The code and report of this whole project was done by A K M Salman Hosain.

References

[1] Johnell, O., Kanis, J.: An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporosis international **17**, 1726–1733 (2006)

[2] Clynes, M.A., Harvey, N.C., Curtis, E.M., Fuggle, N.R., Dennison, E.M., Cooper, C.: The epidemiology of osteoporosis. British medical bulletin **133**(1), 105–117 (2020)

[3] Rosen, C.J.: The epidemiology and pathogenesis of osteoporosis (2015)

[4] Li, H., Xiao, Z., Quarles, L.D., Li, W.: Osteoporosis: mechanism, molecular target and current status on drug development. Current medicinal chemistry **28**(8), 1489–1507 (2021)

[5] Gao, Y., Patil, S., Jia, J.: The development of molecular biology of osteoporosis. International journal of molecular sciences **22**(15), 8182 (2021)

[6] Aibar-Almazán, A., Voltes-Martínez, A., Castellote-Caballero, Y., Afanador-Restrepo, D.F., Carcelén-Fraile, M.d.C., López-Ruiz, E.: Current status of the diagnosis and management of osteoporosis. International journal of molecular sciences **23**(16), 9465 (2022)

[7] Adejuyigbe, B., Kallini, J., Chiou, D., Kallini, J.R.: Osteoporosis: molecular pathology, diagnostics, and therapeutics. International journal of molecular sciences **24**(19), 14583 (2023)

[8] Wani, I.M., Arora, S.: Osteoporosis diagnosis in knee x-rays by transfer learning based on convolution neural network. Multimedia Tools and Applications **82**(9), 14193–14217 (2023)

[9] Abubakar, U.B., Boukar, M.M., Adeshina, S.: Evaluation of parameter fine-tuning with transfer learning for osteoporosis classification in knee radiograph. International Journal of Advanced Computer Science and Applications **13**(8) (2022)

[10] Klontzas, M.E., Stathis, I., Spanakis, K., Zibis, A.H., Marias, K., Karantanas, A.H.: Deep learning for the differential diagnosis between transient osteoporosis and avascular necrosis of the hip. Diagnostics **12**(8), 1870 (2022)

[11] Kumar, S., Goswami, P., Batra, S.: Fuzzy rank-based ensemble model for accurate diagnosis of osteoporosis in knee radiographs. International Journal of Advanced Computer Science and Applications **14**(4) (2023)

[12] Khanna, V.V., Chadaga, K., Sampathila, N., Chadaga, R., Prabhu, S., Swathi, K., Jagdale, A.S., Bhat, D.: A decision support system for osteoporosis risk prediction using machine learning and explainable artificial intelligence. Heliyon **9**(12) (2023)

[13] Suh, B., Yu, H., Kim, H., Lee, S., Kong, S., Kim, J.-W., Choi, J.: Interpretable deep-learning approaches for osteoporosis risk screening and individualized feature analysis using large population-based data: Model development and performance evaluation. Journal of medical Internet research **25**, 40179 (2023)

[14] Jang, R., Choi, J.H., Kim, N., Chang, J.S., Yoon, P.W., Kim, C.-H.: Prediction of osteoporosis from simple hip radiography using deep learning algorithm. Scientific reports **11**(1), 19997 (2021)

[15] Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114 (2019). PMLR

[16] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)

[17] Ribeiro, M.T., Singh, S., Guestrin, C.: " why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144 (2016)

[18] Osteoporosis — kaggle.com. https://www.kaggle.com/datasets/mrmann007/osteoporosis. [Accessed 15-07-2024]

[19] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

[20] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)