

Original Article

A Novel Deep Learning Approach with Attention Mechanism for Early Osteoporosis Detection from Knee X-Ray Imaging

Athira O M¹, R Gunasudari²

^{1,2} Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India.

¹Corresponding Author : athira.om.321@outlook.com

Received: 09 April 2025

Revised: 10 May 2025

Accepted: 11 June 2025

Published: 27 June 2025

Abstract - A bone disorder with decreased Bone Mineral Density (BMD), rendering bones weaker and susceptible to fractures, even from minor falls or daily activities, is osteoporosis. An accurate diagnosis is crucial for effective therapy to reduce the fracture risk, yet existing diagnostic procedures are time-consuming. The traditional models relied on manual radiological evaluation and hand-crafted Machine Learning (ML) features to diagnose knee osteoporosis. However, these approaches had limitations in accuracy and efficacy due to radiologists' subjective interpretations and manual feature extraction. Early detection with X-ray imaging allows timely intervention and facilitates proper treatment. This study proposes a novel osteoporosis detection model using a Deep Learning (DL)-based approach enhanced by an attention mechanism to improve classification performance using knee X-ray dataset images. The DenseNet-121 model is the backbone, improving the vanishing gradient problem and ensuring efficient data flow across layers. Channel-wise and spatial attention techniques are utilized with a Convolutional Block Attention Module (CBAM) to refine feature representation. The model demonstrated superior performance in classifying osteoporotic and healthy knees from X-ray images, attaining a remarkable accuracy of 97.43%. This study enhances osteoporosis detection by employing knee X-ray images with excellent prediction and efficient classification, thereby reducing the socioeconomic burden of osteoporosis-related fracture risks and improving patient outcomes.

Keywords - Osteoporosis, Deep Learning, DenseNet 121, Convolutional block attention module, Attention mechanism.

1. Introduction

Osteoporosis is a prevalent bone disease characterized by a progressive loss of bone density and structural degeneration, increasing the likelihood of fractures. It is frequently denoted as a "silent disease" because it appears unrecognized and remains undetected until a fracture occurs, typically in weight-bearing bones such as the backbone and hip. Osteoporosis is an important cause of morbidity and mortality, in addition to the noticeable physical effects of a fracture, such as pain and inconvenience [1]. However, early osteoporosis detection and treatment reduces a person's fracture risk to a minimum. Osteoporosis disrupts the bone growth process, leading to faster bone loss. It is crucial to analyze bone health to prevent osteoporosis. Timely detection of osteoporosis through bone density tests employs X-rays to determine the deposition of calcium and other minerals in the bones. This is crucial for diagnosing osteoporosis and enables early intervention. Such tests reduce the time needed for treatments like bone modelling, which promotes targeted bone growth at specific regions, enhancing bone density and reducing skeletal fragility. Unlike bone remodelling, which sometimes leads to loss, bone modelling effectively focuses on rebuilding bone

structure. Figure 1 illustrates osteoporotic and normal healthy bone.

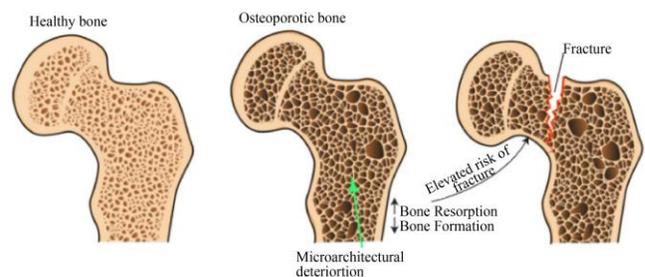


Fig. 1 Healthy bone Vs. Osteoporotic bone

Conventional techniques for diagnosing osteoporosis involve Fracture Risk Assessment Tools (FRAX), Dual-Energy X-ray Absorptiometry (DEXA), and Quantitative Computed Tomography (QCT) [2]. While these techniques offer valuable insights, they pose limitations such as high cost, limited accessibility, radiation exposure, and reliance on subjective interpretation. Furthermore, DXA imaging mainly evaluates BMD, which fails to capture other essential factors



such as bone quality, microarchitecture, and biomechanical characteristics. Existing models confront considerable challenges that limit their widespread clinical application, as most osteoporosis models rely on a single dataset with few diversities, which confines their generalizability to wider populations. Anatomical regions emphasize particular parts, such as the spine and hips, while ignoring other essential areas, such as the knee, which provide valuable insights into bone health. Methodological constraints such as data imbalance, feature selection, and interpretability remain substantial challenges in developing a reliable osteoporosis prediction model. These drawbacks highlight the necessity of sophisticated, automated, and cost-effective methods to improve osteoporosis detection [3].

Artificial Intelligence (AI) and ML techniques have significantly impacted the healthcare industry, enabling the development of automated medical solutions with accurate results [4]. The DL in osteoporosis prediction has attracted much attention, with studies examining various imaging techniques, namely MRI, CT scans, and X-rays, with the intent of predicting fracture risk and identifying early signs of bone loss, which result in better osteoporosis medical care and diagnosis [5, 6]. In order to enhance osteoporosis prediction, this research suggests an effective DL-based model with an attention mechanism using knee X-ray images. The suggested approach intends to enhance the efficacy of the proposed model in real clinical environments. The contributions of the study include:

- To develop a DL-based classification model with an attention mechanism for detecting osteoporosis detection using knee X-ray imaging.
- Assess the efficacy of the suggested classification models, such as DenseNet 121 and CBAM, to ensure generalizability and robustness.
- To evaluate the effectiveness of the suggested model with traditional osteoporosis prediction models.

The remaining portion of the study is organized as follows: A thorough literature review is described in Section 2, emphasizing previous studies and pointing out important research gaps. Section 3 explains the suggested model in detail. Section 4 presents the research findings. The study is concluded in Section 5, which summarises the key findings.

2. Literature Review

Sarhan et al. [7] presented a DL approach utilizing Convolutional Neural Network (CNN) architectures for detecting osteoporosis on X-ray images of the knee. A data sample comprising 1947 knee X-rays was collected and augmented to enhance classification performance, and the authors utilized transfer learning with the pre-trained CNN models for detection. The outcomes demonstrated that the VGG-19 model attained better accuracy, with 92.0% for multiclass and 97.5% for binary classification. The CNN

models also showed a competitive performance. However, the study lacked the integration of clinical factors with image analysis, which reduced the comprehensive diagnostic accuracy.

Jang et al. [8] investigated a DL model from simple hip radiographs to predict osteoporosis. The dataset comprised 1001 proximal femur DXA scans from female patients aged 55 years. The authors employed a deep CNN (DCNN) based on VGG-16, and the DCNN model attained an accuracy of 81.2%. External validation on 117 datasets demonstrated 71.8% accuracy, and Grad-CAM visualizations successfully highlighted relevant bone structures. However, the model had limitations in object detection accuracy, which affected its reliability in real-world scenarios. The study did not consider BMI and mortality rate.

Dzierzak et al. [9] presented six pre-trained DCNN architectures for osteoporosis diagnosis from CT images of the lumbar spine, such as Inception-ResNet-V2, VGG-16, ResNet-50, VGG19, Xception, and MobileNetV2. The dataset included CT images from 100 patients. The VGG-16 model demonstrated better classification accuracy. The authors employed data augmentation and transfer learning to enhance model accuracy. The model lacked the automated segmentation algorithm for extracting tissue samples directly from a CT scanner. However, the study limited the analysis of CT images from the spine and other bone structures affected by this disorder.

Fang et al. [10] explored a DL model for osteoporosis detection, using CNNs for automated image analysis. The authors utilized DenseNet-121 for BMD calculation and U-Net for vertebral body segmentation in CT images. The dataset comprised 1449 patients, sourced from three different CT vendors. The QCT was used for BMD evaluation, and manual segmentation served as the ground truth. The U-Net model showed strong segmentation accuracy, and automated BMD calculations demonstrated high correlations with QCT values. However, the system was established from a single centre and trained from a single CT scanner. The study excluded patients with severe scoliosis, which limited its applicability to diverse populations.

Nakamoto et al. [11] explored the DL models for osteoporosis detection using medical imaging. The authors developed a computer-aided screening system using three CNN models, consisting of VGG-16, GoogleNet, and AlexNet, which were trained on labelled radiographs classified by an oral radiologist. The model was evaluated using BMD data from the femoral neck and lumbar spine. The VGG-16 and GoogleNet showed a better accuracy rate of 75%. However, the study relied on the judgement of a single radiologist, which led to inaccuracies in classifying the cortical bone assessment groups.

Sukegawa et al. [12] explored the application of the DL model for classifying osteoporosis using dental panoramic radiographs, which offered a cost-effective screening method. A dataset comprised 778 images from individuals who underwent dental panoramic radiography and skeletal BMD assessments. The model was trained to classify osteoporosis using CNN models, such as ResNet (152, 50, and 18) EfficientNet (B7, B3, and B0), and an ensemble approach with clinical covariates was added. The ResNet-152 and EfficientNet-B7 demonstrated better accuracy. The authors utilized CNN models with a limited data sample, leading to potential overfitting and inadequate generalization. The model lacked the implementation of a Region Of Interest (ROI), and the dataset images were manually cropped for preprocessing, delaying the model's performance in real-world applications. Oh et al. [13] evaluated the DL-based (DL-QCT) solution and analyzed 112 clinical CT scans. The automatically generated BMD values (DL-BMD) for L1 and L2 vertebrae are compared with manual MBD (m-BMD) assessments employing DXA and QCT. The diagnostic performance was evaluated using ROC analysis, where DL-BMD attained an Area Under Curve (AUC) of 0.847 for normal and 0.770 for osteoporosis. The model's reliability was limited due to using a single CT scanner, which primarily focused on patients with normal spine BMD and abnormal pelvis BMD.

A superfluity mechanism was employed by Naguib et al. [14] for categorizing the X-ray images from the knee into osteoporosis categories. The mechanism concatenates various layers, allowing features to flow into distinct branches. The study employed two knee X-ray datasets for model evaluation. The authors utilized the pre-trained models and compared their performance with the Superfluity DL architecture. The approach showed an accuracy of 5.42% for Dataset 1 and 79.39% for Dataset 2, outperforming other pre-trained models. The superfluity mechanism required more computational resources, limiting their clinical application performance. Superfluity affected the model's complexity, and excessive features led to model degradation.

Liu et al. [15] investigated three distinct layers as a hierarchical system to predict osteoporosis using clinical data and CT scans of the lumbar vertebral bodies of 2210 individuals over the age of 40. The authors utilized six ML models for classification, such as Random Forest (RF), Artificial Neural Network (ANN), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, and stacking. The hierarchical LR-based model attained the highest AUC of 0.962 among the three distinct layers in classifying individuals with normal BMD and osteoporosis. The study utilized manual segmentation of the lumbar vertebral bodies, leading to feature extraction variations.

Ou Yang et al. [16] developed a predictive model for osteoporosis screening utilizing ML algorithms, including LR, ANN, SVM, K-Nearest Neighbours (KNN), and RF. The

dataset consisted of 2929 women and 3053 men from a clinical checkup in Taiwan. The system was trained separately for women and men using clinical data as input features. The results showed that the ML models outperformed the conventional Osteoporosis Self-assessment Tool for Asians (OSTA) system, demonstrating their potential for improving osteoporosis screening. The DXA results in the database were categorical rather than continuous, limiting their potential for more precise learning. The medical history was used to record the input features for the study, which reduced the model's performance.

Wani et al. [17] utilized transfer learning with DL-based CNNs, including VGGNet-19, ResNet-18, AlexNet, and VGGNet-16, to categorize X-ray images from the knee into osteoporosis. The dataset comprised 381 medically validated knee X-rays classified based on T-scores from a quantitative ultrasound system. The results demonstrated that AlexNet achieved a better accuracy rate of 91.1%, with a 0.09 error rate, while the VGGNet-19 showed the lowest accuracy rate. The models relied only on image-based analysis without considering clinical factors affecting diagnostic accuracy. Additionally, the authors focused only on knee osteoporosis without exploring its relationship with osteoporosis in other skeletal areas, limiting its applicability as a universal diagnostic system.

Khanna et al. [18] utilized an open-source dataset of 1493 patients and the ML framework to predict osteoporosis risk, which comprised physical examination data, bone density, and blood test results. The model analyzed thirteen feature selection methods to detect the most relevant predictors. The forward feature algorithm was applied, followed by a multi-level ensemble learning-based stacking classifier for risk prediction. Explainable AI (XAI) techniques were implemented to enhance model interpretability and understand the important features. The best-performing pipeline, which included an algorithm for feature selection, achieved an accuracy of 89%, demonstrating its effectiveness in risk prediction. BMD was identified as the most relevant predictor. The study lacked advanced optimization algorithms, reducing the accuracy of the decision-making.

Sebro et al. [19] investigated the feasibility of osteoporosis screening using CT scans of the forearm and wrist with ML technique. A comprehensive analysis was conducted on 196 patients who underwent CT and DEXA scans. The multivariable SVM with a Radial-Bias-Function (RBF) kernel was employed for osteoporosis prediction and achieved an AUC of 0.818, outperforming other models. The result demonstrated that the CT attenuation of multiple bones was more precise in predicting osteoporosis than using a single bone. The study lacked concurrent DEXA studies, indicating insufficiency for osteoporosis screening. All CT scans were analyzed using Siemens scanners, which restricted generalizability.

Hidjah et al. [20] utilized a DCNN model to detect osteoporosis from dental periapical radiographs. The dataset comprised postmenopausal women with BMD measurements. The methodology involved feature extraction, image acquisition, ROI selection, and classification. The best model was developed using an input image size of 100 to 150 and employing a five-convolution layer. The results demonstrated that larger image blocks provided additional trabecular patterns with an accuracy of 92.50%. The model's generalizability was limited due to its inability to fully capture the complexity of osteoporosis using BMD measurements as the reference standard.

By analyzing balance parameters, Cuaya-Simbro et al. [21] explored computational methods to identify osteoporotic individuals at risk of falling. A study was conducted on 126 community-dwelling older women with osteoporosis. The dataset was analyzed using various ML models, such as Random Forest Classifier (RFC) and Instance-Based k (IBk) KNN. Oversampling methods were applied to address the class imbalance, and the Feature Selection for the Minority Class (FMSC) technique was utilized to identify relevant balance parameters. The best-performing model, the RFC, used oversampling to demonstrate effective risk fall prediction. An imbalanced dataset reduced the effectiveness of the model.

Despite the growing application of ML and DL in thyroid nodule classification, several challenges remain unaddressed, limiting their clinical applicability. Most studies focus only on image-based analysis without considering clinical parameters,

such as BMI, mortality rates, and patient history, which limits diagnostic accuracy and real-world scenarios [8]. Furthermore, the datasets from a single centre lowered model generalizability across diverse populations. The exclusion of specific patient groups, especially those with scoliosis or abnormal bone abnormalities, further restricts model relevance [10]. Another critical concern is the lack of automated segmentation methods for extracting relevant features from medical images. Manual cropping of images increases the risk of unpredictability and human error [12]. The multi-site osteoporosis diagnosis gets less attention, as most models focus on a single skeletal region, rendering it challenging to predict disease progression in other bones [17]. The computational complexity also poses challenges in the real world, especially in clinical settings with limited resources. Furthermore, the lack of advanced optimization techniques limits clinical decision-making and interpretability [18]. Addressing these challenges by combining multi-source data, automating progress, and enhancing model generalizability considerably enhances osteoporosis detection and risk prediction.

3. Materials and Methods

Osteoporosis significantly impacts quality of life, rendering early detection crucial for identifying potential fracture risks. A novel osteoporosis classification model is proposed using a DL model with an attention mechanism to address these challenges. Figure 2 illustrates the proposed classification model for osteoporosis detection by analyzing knee X-ray images.

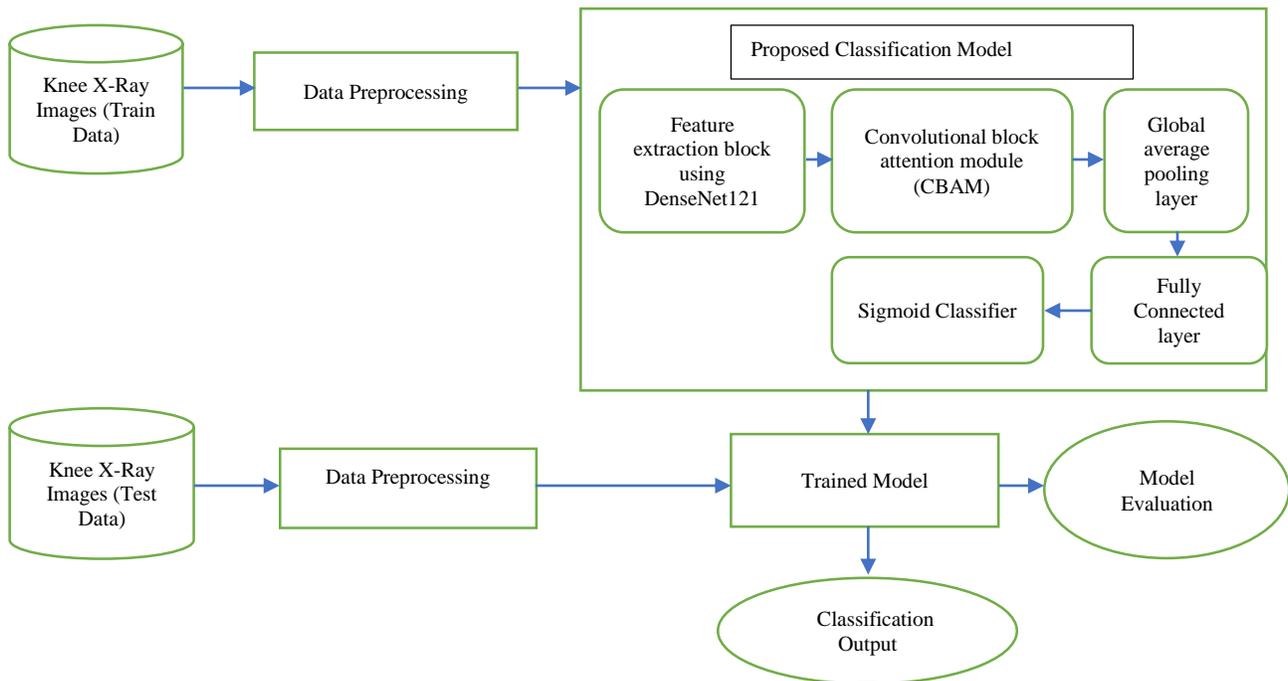


Fig. 2 Layout of the suggested model

The model begins with data preprocessing, where knee X-ray images are resized, normalized, and augmented to ensure consistent input size and enhance the diversity of the training set. The dataset is then split into an 80:20 ratio for training and testing to improve model performance. The preprocessed images are passed through a DenseNet-121-based feature extraction block, capturing high-level and complex patterns within the X-ray images through densely connected convolutional layers. The extracted features are refined by the CBAM, which sequentially applies channel-wise and spatial attention mechanisms to emphasize the images' most informative patterns and regions. These enhanced feature maps are subsequently processed through a Global Average Pooling (GAP) layer, which condenses spatial information into a compact feature vector. This vector is passed to a fully connected layer to enable further feature interaction and representation. Finally, the output is classified by a sigmoid classifier, which assigns each image a probability score indicating whether it is osteoporotic or healthy. The trained model is then assessed using the test images, and its performance is evaluated through standard metrics, generating the final classification output.

3.1. Dataset

The data from the publicly accessible Kaggle repository is used in this study [22]. The dataset comprises 372 knee X-ray images, with 186 knee X-ray images labelled normal (0) and 186 as osteoporotic (1). Figure 3 presents sample data that compares a healthy knee with an osteoporotic knee.



Fig. 3 Sample data of knee X-ray images

3.2. Data Preprocessing and Augmentation

Data preprocessing involves a data mining process involving cleaning and converting raw data for analysis. Subsequently, it improves data quality to enhance model performance and ensure accurate results. This study involves preparing the categorical data into a machine-readable format,

addressing missing values, and scaling numerical features. The dataset images are resized to a standard size of 224×224 pixels for a uniform input size. Additionally, the images are converted to RGB format and processed for binary classification tasks, resulting in standardized pixel values to improve the model training and convergence. The feature data is normalized by applying standardization, which involves scaling it to unit variance and subtracting the mean. This transformation is represented by Equation (1).

$$X_{scaled} = \frac{X - \mu}{\sigma} \quad (1)$$

Where X is the original feature value, σ and μ stand for standard deviation and arithmetic mean of the feature, respectively.

Data augmentation enhances model optimization and generalisability by developing a new data sample from the pre-existing data. The data augmentations, including rescaling, shearing, zooming, flipping, and rotation, are used to preprocess and enhance images for training, testing, and validation.

3.3. Model Development

The study suggested a DL model based on DenseNet 121 and CBAM to improve feature extraction and pattern recognition in osteoporosis classification. Combining these two models harnesses their strengths, yielding a more accurate and robust feature representation.

3.3.1. DenseNet-121 Model

The DenseNet, a CNN architecture, is known for its unique architecture. As illustrated in Figure 4, DenseNet-121 employs a distinctive dense connectivity architecture. DenseNet differs from conventional CNN architectures in two important ways. In a dense block structure, every layer is fed forward to every other layer, which is its primary characteristic. In addition, it employs bottleneck layers, allowing for a reduction in the number of parameter numbers without lowering the network's ability to learn. In particular, DenseNet-121 has 121 layers with three primary components: dense blocks, transition layers, and a global average pooling layer. Convolutional layers with dense connections are found in dense blocks, whereas transition layers restrict the number of parameters and lower the dimensionality. While considering all these things, DenseNet's distinctive architecture effectively learns and encodes complex characteristics and provides an effective tool for computer vision tasks like image categorization [23].

DenseNet processes a single X-ray image (Y_0) using a deep neural network (Y_0). Convolution (Conv), pooling, batch normalization (BN), and ReLU are compound functions of operations that use the form $K_l(\cdot)$. The output of the l^{th} layer

is represented as Y_l . The l^{th} layer receives all previous feature maps (y_0, y_1, \dots, y_{l-1}) as inputs, as shown in Equation (2).

$$y_l = K([y_0, y_1, \dots, y_{l-1}]) \quad (2)$$

where, $[y_0, y_1, \dots, y_{l-1}]$ indicates attention to all preceding feature maps of the l^{th} layer. DenseNet addresses the issue of fading gradients by reusing features while decreasing the number of variables. DenseNet -121 employs four dense blocks.

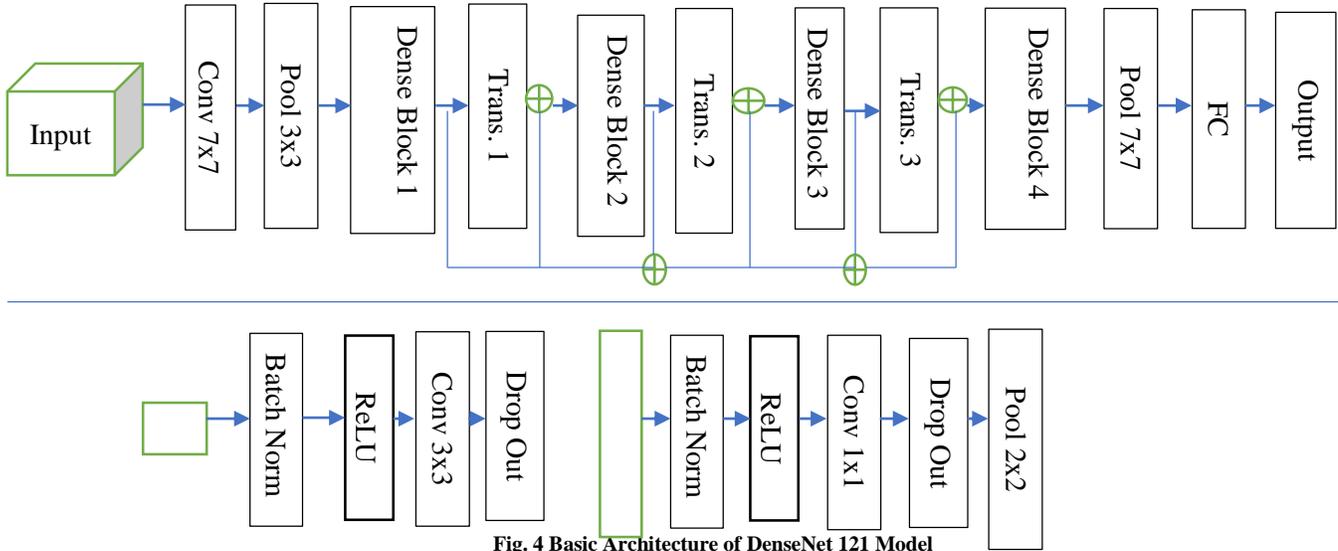


Fig. 4 Basic Architecture of DenseNet 121 Model

The transition layer within each block downsamples the feature maps to produce a 2×2 average pooling layer and a 1×1 convolution layer. Dense blocks consisted of multiple convolutional layers connected sequentially to establish cross-layer connections [24].

3.3.2. Convolutional Block Attention Module

The CBAM enhances feature extraction by sequentially applying the Spatial Attention Module (SAM) and Channel Attention Module (CAM), as shown in Figure 5. The process starts with an input feature, which then passes through the CAM. The SAM then processes the outputs, highlighting significant spatial regions within the feature map. After applying both mechanisms, the extracted feature map is enhanced to capture the essential patterns from the data. The CBAM module enhances feature extraction by adaptively extracting spatial and channel attention mechanisms for highly

effective DL applications. From a transitional feature map, $F \in R^{C \times H \times W}$, CBAM successively concludes a 1D CAM $M_c \in R^{C \times 1 \times 1}$ and a 2D SAM $M_s \in R^{1 \times H \times W}$.

The whole attention mechanism is summarized by Equations (3) and (4).

$$F' = M_c(F) \otimes F \quad (3)$$

$$F'' = M_s(F') \otimes F' \quad (4)$$

Where \otimes signifies element-wise multiplication, and F'' Indicates refined final result. The inter-channel relationship of structures is considered for the CAM. It reduces the spatial dimension of the input feature map to determine the channel attention [25]. Average pooling is employed for aggregating spatial data. Each attention map's computing procedure is shown in Figure 6.

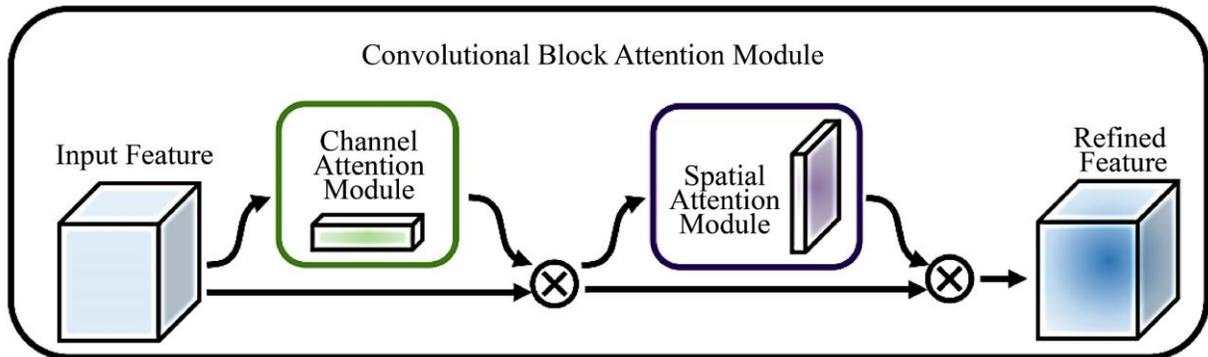


Fig. 5 Visualization of CAM

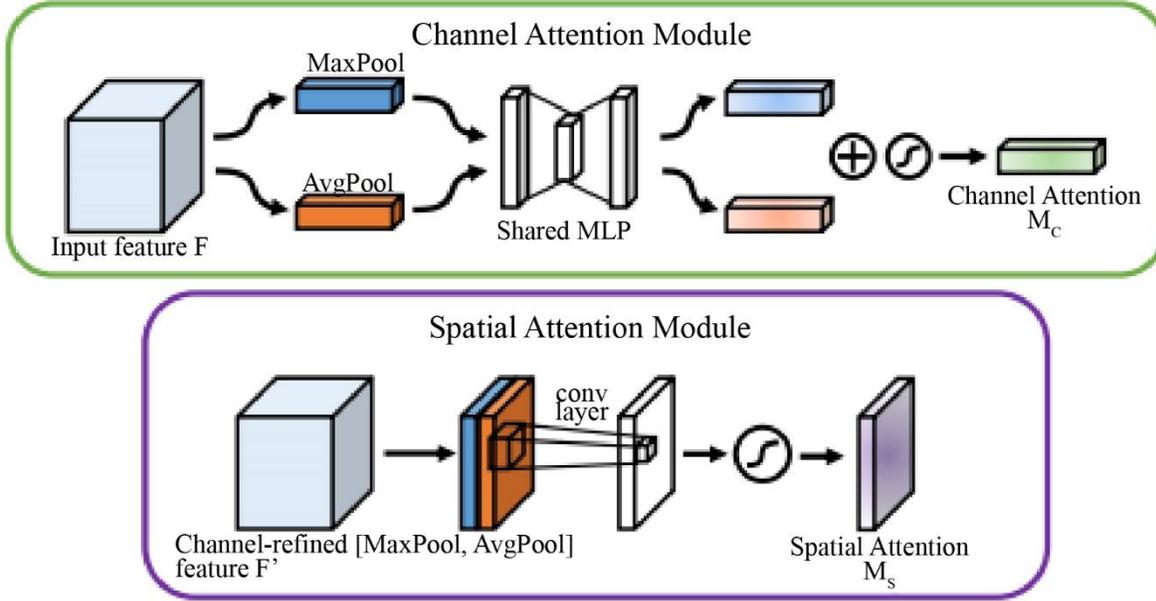


Fig. 6 Layout of each attention sub-module

Two spatial context descriptors, such as F_{avg}^c and F_{max}^c , are generated by combining the spatial data of a feature map through average pooling and max pooling. The CAM is developed by forwarding both descriptors to a shared network. $M_c \in \mathbb{R}^{c \times 1 \times 1}$. The shared network comprises a single Multi-Layer Perceptron (MLP) and one hidden layer. Channel attention is calculated using Equations (5) and (6).

$$M_c(F) = \sigma \left(MLP(AvgPool(F)) + MLP(MaxPool(F)) \right) \quad (5)$$

$$= \sigma \left(W_1 \left(W_0(F_{avg}^c) \right) + W_1 \left(W_0(F_{max}^c) \right) \right) \quad (6)$$

Where σ represents the sigmoid function, $W_0 \in \mathbb{R}^{c/r \times c}$ and $W_1 \in \mathbb{R}^{c/r \times c}$. W_0 and W_1 have the same inputs and W_0 is followed by the ReLU activation function. Equations (7) and (8) represent the spatial attention.

$$M_s(F) = \sigma \left(f^{7 \times 7}([AvgPool(F); MaxPool(F)]) \right) \quad (7)$$

$$= \sigma \left(f^{7 \times 7}([F_{avg}^s; F_{max}^s]) \right) \quad (8)$$

Where $f^{7 \times 7}$ indicates the convolution operation of filter size 7×7 .

3.3.3. Proposed DL Model

The suggested model architecture employs DenseNet 121, a powerful pre-trained CNN known for its efficiency in feature extraction. The model relies on a distinctive dense connectivity pattern where each layer receives input from all previous layers. This ensures maximum data flow and allows the system to capture complex patterns within the data. The base model is configured to exclude the top fully connected

layers, which are typically used for classification in the original DenseNet121 architecture. Instead, the model focuses solely on the convolutional layers, which process the input image data to extract relevant features. The input shape is defined according to the specific task, and the pre-trained ImageNet weights are used to leverage transfer learning. This initialization enables the model to begin with a strong basis, having already acquired general features from a large dataset. Following the DenseNet121 model, a CBAM is used to enhance feature representation.

The CBAM block constitutes a lightweight, powerful attention mechanism that emphasizes the most informative features in the spatial and channel dimensions. The CAM first performs GAP and global max pooling across the spatial dimensions, followed by shared dense layers to capture channel relationships. The outputs from these two pooling operations are subsequently combined into a sigmoid activation and multiplied by the original input tensor to reweight the channel-wise features. This process helps the model focus on the most critical channels contributing to accurate predictions.

Next, the SAM is applied further for spatial feature extraction. The output of the CAM is processed by average pooling and max pooling operations along the channel axis, and the outcomes are concatenated to form a combined feature map. This map highlights the most relevant spatial regions in the feature map, enabling the model to emphasize significant elements of the image. The SAM is multiplied by the input feature map to finalize the attention process. This attention mechanism enables the model to dynamically alter its focus in response to the input data, thereby improving its ability to identify essential patterns.

Algorithm 1: DenseNet 121 - CBAM based Osteoporosis Prediction

Input: Knee X-ray images
 Output: Efficient Osteoporosis Detection and Classification Model.

Begin:

- Load and preprocess data:
 - Collect dataset: $D = \{(x_i, y_i)\}$, where x_i is a Knee X-ray images and $y_i \in \{0,1\}$ (1: Osteoporosis, 0: Normal).
 - Preprocess:
 - Resize: $x_i \rightarrow R^{(224 \times 224)}$
 - Normalize: $x_i = (x_i - \text{mean}) / \text{std}$
 - Data Augmentation: (Shear, Zoom, Flip, Rotation)
- Define Proposed Classification Model:
 - $\text{base_model} = \text{DenseNet121}(\text{weights}='imagenet', \text{input_shape} =)$
 - $x = \text{base_model. Output}$
 - $x = \text{cbam_block}(x)$
 - $x = \text{GlobalAveragePooling2D}() (x)$
 - $x = \text{Dense}(256, \text{activation}='relu') (x)$
 - $x = \text{Dense}(1, \text{activation}='sigmoid') (x)$
 - $M = \text{Model}(\text{inputs}=\text{base_model. input}, \text{outputs}=x)$
- Model Compilation:
 - $M.\text{compile}(\text{optimizer}=\text{Adam}(), \text{loss}='binary_crossentropy', \text{metrics}=['accuracy'])$
- Model Training:
 - $M.\text{fit}(\text{train_generator}, \text{validation_data}=\text{val_generator}, \text{epochs}=50)$
- Model Evaluation:
 - $\text{metrics} = M.\text{evaluate}(X_{\text{test}}, y_{\text{test}})$

Save the model

End

3.4. Hardware and Software Setup

The system is powered by an Intel Core i7-6850K 3.60 GHz 12-core processor and equipped with an NVIDIA GeForce GTX 1080 Ti GPU with 2760 MB of memory, ensuring high computational performance. Google Collaborator was the DL platform, providing a robust model training and testing environment. The proposed model undergoes training and testing to achieve optimal performance, with its algorithms validated using ground truth data to ensure accurate and reliable results under multiple circumstances. Hyperparameters influence the training of DL models and are set before the training begins. Table 1 showed the list of hyperparameters employed in this study.

Table 1. Hyperparameter specifications

Hyperparameters	Values
Batch Size	16
Optimizer	Adam
Number of Epochs	50
Loss Function	Binary Crossentropy

4. Results and Discussion

The accuracy and loss plots show the system’s learning progress and predictive capabilities over training epochs. Figure 7 illustrates the proposed system’s learning progress and predictive capabilities over 50 training epochs. Initially, the model attained a training accuracy of 75.6%, progressively improving to 83% by epoch 3, demonstrating rapid learning. The accuracy steadily increased as training continued, reaching a final accuracy of 97.43% by epoch 50. This outcome establishes the model’s ability to generalize the osteoporosis detection task effectively. During the early stages of training, the loss was relatively high, approximately 0.7, indicating initial prediction errors. By epoch 10, the loss had decreased below 0.4, reflecting model improvement. This trend continued, with the loss value reducing to 0.1 by epoch 50, indicating efficient learning with minimal errors. These findings demonstrate the model’s ability to classify osteoporosis from knee X-ray images.

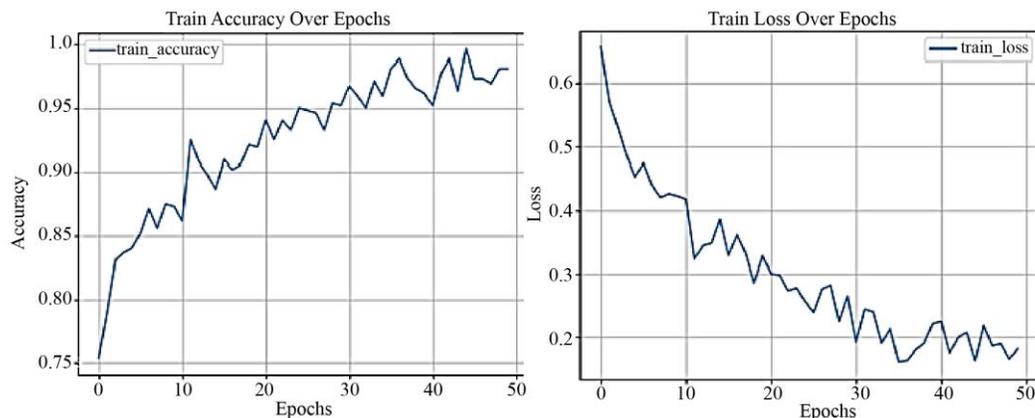


Fig. 7 Graph illustrating the suggested model’s accuracy and loss

To assess the accuracy and robustness of the suggested system, several performance metrics are used based on classification performance, such as True Negative (TN), False Positive (FP), False Negative (FN), and True Positive (TP). These mathematical metrics provide insights into the model’s strength and serve as a benchmark for its classification performance, leading to a comprehensive assessment of the system’s capability in predicting osteoporosis. The evaluation is determined from the following Equations (9) to (12).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

$$F1 - score = 2 \times \frac{precision \times Recall}{Precision + Recall} \tag{12}$$

The performance metrics shown in Figure 8 demonstrated that the system performed well in classifying osteoporosis from knee X-ray images.

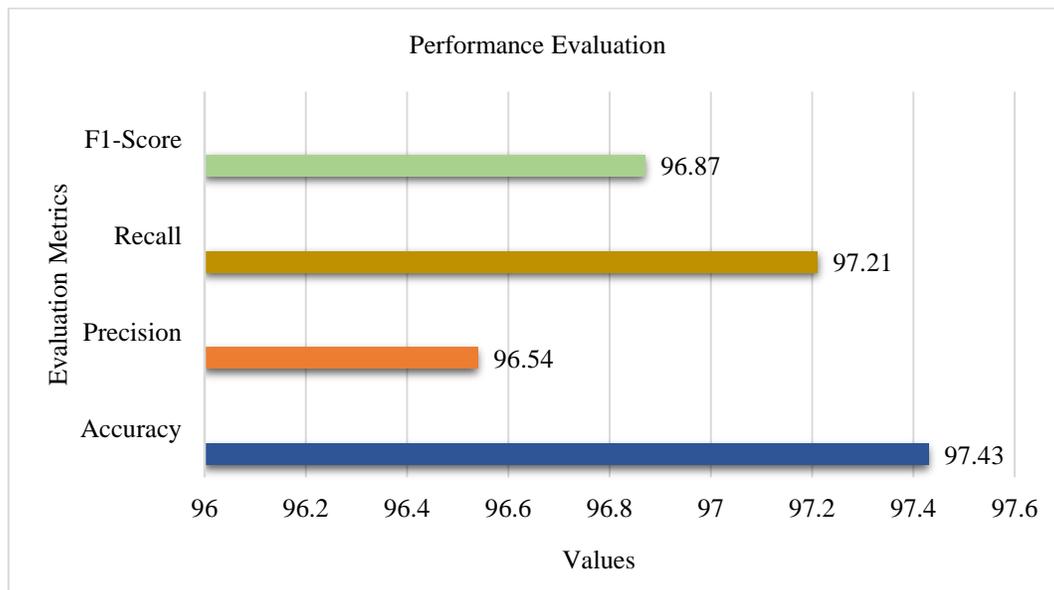


Fig. 8 Performance evaluation of the proposed model

With an accuracy of 97.43%, the proposed classification model suggests that the model effectively predicts osteoporosis and proves to be an excellent tool for precisely classifying most samples. The precision score of 96.54% indicates an excellent prediction with a low rate of FP, ensuring an actual positive prediction. Finally, the recall of 97.21% shows that the model easily identifies 97.21% of the actual positive cases, highlighting its reliability in real-world applications. An F1 score of 96.87% shows the ability of the model to maintain a balance between precision and recall. The confusion matrix is a perfect tool for assessing the effectiveness of the classification approach, representing actual vs. predicted outcomes, and is widely used in medical diagnostics and DL techniques. The efficacy of the proposed classification model is examined using the knee X-ray images, showing a high degree of reliability and accuracy. As shown in Figure 9, the suggested model achieved impressive predictive capabilities and accurately identified 62 typical and 51 osteoporosis cases with few misclassifications. However, the overall performance in classification between osteoporosis and normal remains robust, enhancing its reliability and potential for accurate prediction.

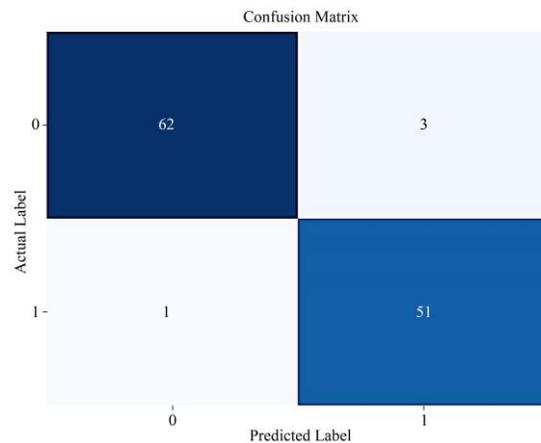


Fig. 9 Confusion matrix of the proposed model

Figure 10 depicts the ROC curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) to demonstrate the system’s performance. The suggested model showed a superior AUC value of 0.97, representing accurate osteoporosis detection while reducing false prediction classification.

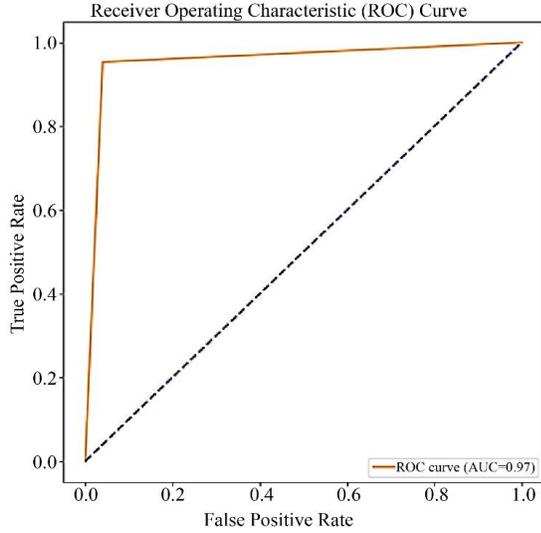


Fig. 10 ROC curve

Figure 11 depicts the predicted classification output, illustrating the model's accuracy and efficacy for detecting osteoporosis using knee X-ray images. Table 2 and Figure 12 compare the proposed model by existing methods. The performance comparison of various models for osteoporosis detection revealed that the suggested DenseNet-121 with CBAM achieved the highest accuracy rate of 97.43%, surpassing existing methods. Traditional DL models such as DNN (81.2%) and AlexNet (91.1%) exhibited lower accuracy due to their limited feature extraction capabilities. Methods such as Deep CNN (90.7%) and CNN (92.5%) enhanced performance; however, they lacked efficient spatial and

channel attention mechanisms, limiting their capacity to focus on important osteoporosis features. The VGG16 model attained 95% accuracy, which improved performance but fell short of the suggested model. DenseNet-121's greater performance with CBAM can be due to its densely connected architecture, which promotes gradient flow, and the CBAM method, which improves feature refinement by emphasizing essential spatial and channel-based information. This combination efficiently addressed redundant features and loss of acceptable resolution, resulting in more accurate osteoporosis detection.



Fig. 11 Classification output

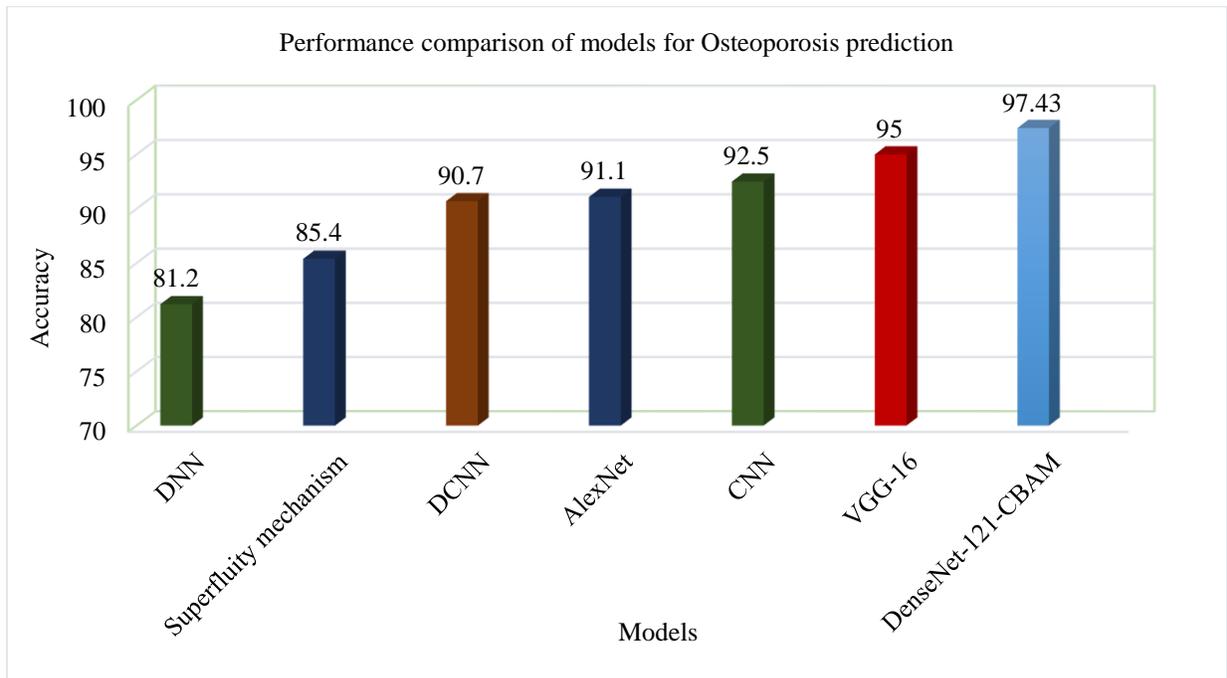


Fig. 12 Visualization of accuracy comparison

Table. 2 Performance comparison with existing models

Author	Methodology	Accuracy (%)
Jang et al. [8]	DNN, VGG16 architecture	81.2 (DNN)
Naguib et al. [14]	Superfluity mechanism, AlexNet and ResNet50.	85.4(superfluity mechanism)
Nakamoto et al. [11]	Deep CNN	90.7
Wani et al. [17]	VGGNet-19, ResNet, VGGNet-16 and AlexNet	91.1% (AlexNet)
Hidjah et al. [20]	CNN	92.5
Dzierżak et al. [9]	VGG 16	95
Proposed model: DenseNet-121 - CBAM		97.43

5. Conclusion

Osteoporosis is a chronic bone disorder characterized by impaired bone density and a greater chance of fractures, necessitating early detection to prevent severe complications and enhance patient outcomes. Traditional diagnostic methods, including clinical risk factor assessments and DEXA, often struggle with high cost, limited accessibility, and the inability to detect osteoporosis at its initial stage. This research presents an effective DL strategy with an attention mechanism for detecting osteoporosis from X-ray images of the knee by combining DenseNet-121 with the CBAM to enhance feature extraction and improve the model to focus on critical bone structures in knee X-ray images. The model demonstrated high classification performance, achieving an accuracy of 97.43%, precision of 96.54%, recall of 97.21%,

and F1-score of 96.87%, indicating its effectiveness in osteoporosis classification. Future research can explore the integration of multi-modal imaging, such as MRI and CT scans, to improve diagnostic accuracy. Furthermore, using explainable AI techniques can improve model interpretability, allowing healthcare practitioners to understand better and trust automated osteoporosis detection.

Acknowledgements

I want to express my sincere gratitude to all those who contributed to completing this research paper. I extend my heartfelt thanks to my supervisor, family, colleagues and fellow researchers for their encouragement and understanding during the demanding phases of this work.

References

- [1] M.S. LeBoff et al., "The Clinician's Guide to Prevention and Treatment of Osteoporosis," *Osteoporosis International*, vol. 33, pp. 2049-2102, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Md Shakhawat Hossen et al., "Revolutionizing Osteoporosis and Bone Fracture Diagnostics: The Emergence of Microwave Antenna Technology," *IEEE Access*, vol. 12, pp. 160418-160448, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Boyi Li et al., "Effect of Spectral Estimation on Ultrasonic Backscatter Parameters in Measurements of Cancellous Bones," *IEEE Access*, vol. 7, pp. 83034-83045, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Wilson Ong et al., "Artificial Intelligence Applications for Osteoporosis Classification Using Computed Tomography," *Bioengineering*, vol. 10, no. 12, pp. 1-26, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Chaewoo Kim et al., "A Deep Learning Harmonization of Multi-Vendor MRI for Robust Intervertebral Disc Segmentation," *IEEE Access*, vol. 12, pp. 19482-19499, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] M.J. Aashik Rasool et al., "KONet: Toward a Weighted Ensemble Learning Model for Knee Osteoporosis Classification," *IEEE Access*, vol. 12, pp. 5731-5742, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Amany M. Sarhan et al., "Knee Osteoporosis Diagnosis Based on Deep Learning," *International Journal of Computational Intelligence Systems*, vol. 17, pp. 1-33, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Ryoungwoo Jang et al., "Prediction of Osteoporosis from Simple Hip Radiography using Deep Learning Algorithm," *Scientific Reports*, vol. 11, pp. 1-9, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Róża Dzierżak, and Zbigniew Omiotek, "Application of Deep Convolutional Neural Networks in the Diagnosis of Osteoporosis," *Sensors*, vol. 22, no. 21, pp. 1-18, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Yijie Fang et al., "Opportunistic Osteoporosis Screening in Multi-Detector CT Images Using Deep Convolutional Neural Networks," *European Radiology*, vol. 31, pp. 1831-1842, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Takashi Nakamoto, Akira Taguchi, and Naoya Kakimoto, "Osteoporosis Screening Support System from Panoramic Radiographs Using Deep Learning by Convolutional Neural Network," *Dentomaxillofacial Radiology*, vol. 51, no. 6, pp. 1-8, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Shintaro Sukegawa et al., "Identification of Osteoporosis Using Ensemble Deep Learning Model with Panoramic Radiographs and Clinical Covariates," *Scientific Reports*, vol. 12, pp. 1-10, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [13] Sangseok Oh et al., "Evaluation of Deep Learning-Based Quantitative Computed Tomography for Opportunistic Osteoporosis Screening," *Scientific Reports*, vol. 14, pp. 1-9, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Soaad M. Naguib et al., "A New Superfluity Deep Learning Model for Detecting Knee Osteoporosis and Osteopenia in X-Ray Images," *Scientific Reports*, vol. 14, pp. 1-18, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Liyu Liu et al., "A Hierarchical Opportunistic Screening Model for Osteoporosis using Machine Learning Applied to Clinical Data and CT Images," *BMC Bioinformatics*, vol. 23, pp. 1-15, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Wen-Yu Ou Yang et al., "Development of Machine Learning Models for Prediction of Osteoporosis from Clinical Health Examination Data," *International Journal of Environmental Research and Public Health*, vol. 18, no. 14, pp. 1-12, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Insha Majeed Wani, and Sakshi Arora, "Osteoporosis Diagnosis in Knee X-Rays by Transfer Learning Based on Convolution Neural Network," *Multimedia Tools and Applications*, vol. 82, pp. 14193-14217, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Varada Vivek Khanna et al., "A Decision Support System for Osteoporosis Risk Prediction using Machine Learning and Explainable Artificial Intelligence," *Heliyon*, vol. 9, no. 12, pp. 1-19, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Ronnie Sebros, and Cynthia De la Garza-Ramos, "Machine Learning for Opportunistic Screening for Osteoporosis from CT Scans of the Wrist and Forearm," *Diagnostics*, vol. 12, no. 11, pp. 1-17, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Khasnur Hidjah et al., "Periapical Radiograph Texture Features for Osteoporosis Detection Using Deep Convolutional Neural Network," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 223-232, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] German Cuaya-Simbros et al., "Comparing Machine Learning Methods to Improve Fall Risk Detection in Elderly with Osteoporosis from Balance Data," *Journal of Healthcare Engineering*, vol. 2021, no. 1, pp. 1-11, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Osteoporosis Knee X-ray Dataset, Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/stevepython/osteoporosis-knee-xray-dataset/data>
- [23] Saleh A. Albelwi, "Deep Architecture Based on DenseNet-121 Model for Weather Image Recognition," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 1-7, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] Omar Elharrouss et al., "Backbones-Review: Feature Extraction Networks for Deep Learning and Deep Reinforcement Learning Approaches," *Arxiv*, pp. 1-23, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [25] Sanghyun Woo et al., "CBAM: Convolutional Block Attention Module," *Proceedings of the 15th European Conference on Computer Vision*, Munich, Germany, pp. 3-19, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]