|  |  |  |
| --- | --- | --- |
| **AUTOMATED DETECTION OF OSTEOPOROSIS AND OSTEOPENIA IN ELDERLY PATIENTS USING KNEE X-RAY IMAGE ANALYSIS** | | |
| GAURAV BHAUMIK  *Vellore Institute of Technology* gaurav.bhaumik2021@vitstudent.ac.in | SARGAM CHAUDHARY *Vellore Institute of Technology* sargam.chaudhary2021@vitstudent.ac.in | SHASHANK AGARWAL *Vellore Institute of Technology* shashank.agarwal2021@vitstudent.ac.in |

***Abstract - This research investigates deep learning-based knee osteoporosis detection using convolutional neural networks (CNNs) and transfer learning. Osteoporosis, characterized by excessive bone loss, increases fracture risks and reduces independence, particularly in the elderly. Traditional diagnostic methods are slow and inefficient, necessitating more accurate approaches. The study merged three datasets containing 1,947 knee X-rays, categorized into Normal, Osteopenia, and Osteoporosis. Several CNN models were tested: Custom CNN (89.1%), VGG19 (94.5%), ResNet-50 (82.1%), DenseNet-121 (93%), and XceptionNet (89.2%). Two hybrid models were developed to enhance detection: one combining VGG19, InceptionResNetV2, and MobileNetV2, achieving 97.5% accuracy, and another integrating DenseNet-121 and EfficientNet, reaching 94.8% accuracy. A key novelty is the interpretability of feature maps, aiding medical professionals in decision-making. The findings highlight the potential of deep learning to provide a scalable and effective solution for early osteoporosis detection, improving clinical outcomes and patient care.***

***Keywords -*** *Knee osteoporosis · Deep learning · Convolutional neural network · CNN · Medical imaging · Image preprocessing*

# Introduction

Osteoporosis, a disease, which often goes undiagnosed until fractures occur, causing pain, disability, and reduced independence in elderly patients. Early detection is crucial for effective management and fracture risk reduction, but current methods are time-consuming. This project presents a novel approach using transfer learning with convolutional neural networks (CNNs) for osteoporosis detection from X-ray images. The approach achieves high accuracy in diagnosis and provides a revealed feature map for medical professionals. The innovation involves a dual strategy: integrating transfer learning from CNN architectures and a dataset collection augmentation mechanism. The study uses 1947 knee X-rays for training and testing. These findings highlight the potential of deep learning with transfer learning in aiding early osteoporosis detection and mitigating fracture risks.

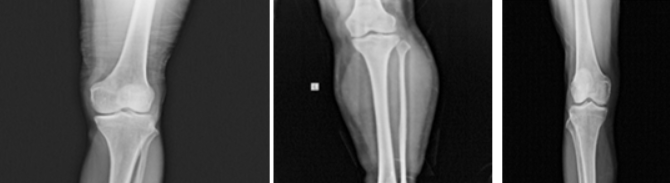
The classification of osteoporosis in the knee has been the subject of numerous studies, each with its own distinct findings and subtleties in methodology. For example, the 2020 study by Wani and Arora revealed an impressive 91% accuracy rate. However, upon closer examination of the con fusion matrix, the actual accuracy was found to be lower. In addition, the models exhibited high validation loss, suggest ing potential limitations in generalization. In contrast, Kumar et al.’s study in 2023 showcased a higher accuracy compared to Wani & Arora’s findings. Similar to this, research by Abubakar et al. in 2022 showed encouraging accuracy in binary classification, but the study lacked reporting on validation loss and had trouble correctly classifying some instances, highlighting the need for improved model robustness. However, the study observed notable drops in model performance during training, suggesting possible instability or overfitting issues. Together, these studies highlight a number of significant gaps in the state of the field. Limited dataset size, poor image quality, and the need for a more complete and reliable convolutional neural network (CNN)-based architecture for osteoporosis identification from X-ray pictures are the main areas of concern. Furthermore, our own research efforts have shown encouraging outcomes, especially in terms of increasing accuracy in binary and multiclass classification problems.

Brief overview of the project:

**1.** Model architecture: We will investigate various deep learning models of convolutional neural networks utilizing transfer learning approach to ensure detection accuracy and efficiency, including VGG-19, Custom CNN, ResNet-50, DenseNet-121 and XceptionNet for accurate knee osteoporosis detection. We also aim to combine two or more models for a better accuracy.

2. The detection aim is binary classes for osteoporosis detection and multiclass for osteoporosis and osteopenia detection. As such, our work serves both directions.

3. Dataset Use: The model is trained and evaluated on an assembled dataset, which is a combination of four independent datasets. Furthermore, after combining the datasets, data augmentation is employed to solve the unbalanced dataset problem

 (a) (b) (c)

X-ray images for (a) normal, (b) osteopenia, (c) osteoporosis knee

Dataset:

The study constructs a custom dataset by combining four publicly available datasets, resulting in a total of 1,947 knee X-ray images: 793 for osteoporosis, 374 for osteopenia, and 780 for normal cases. To address class imbalance, a two-level data augmentation approach is applied. The first level focuses on increasing the number of osteopenia images to balance the dataset, while the second level enhances the entire dataset size to improve deep learning model training.

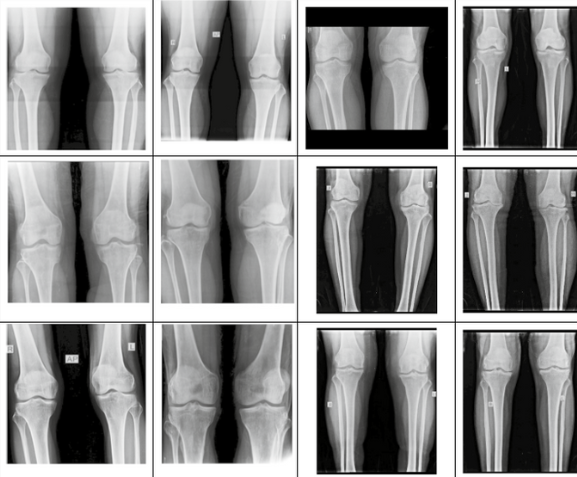
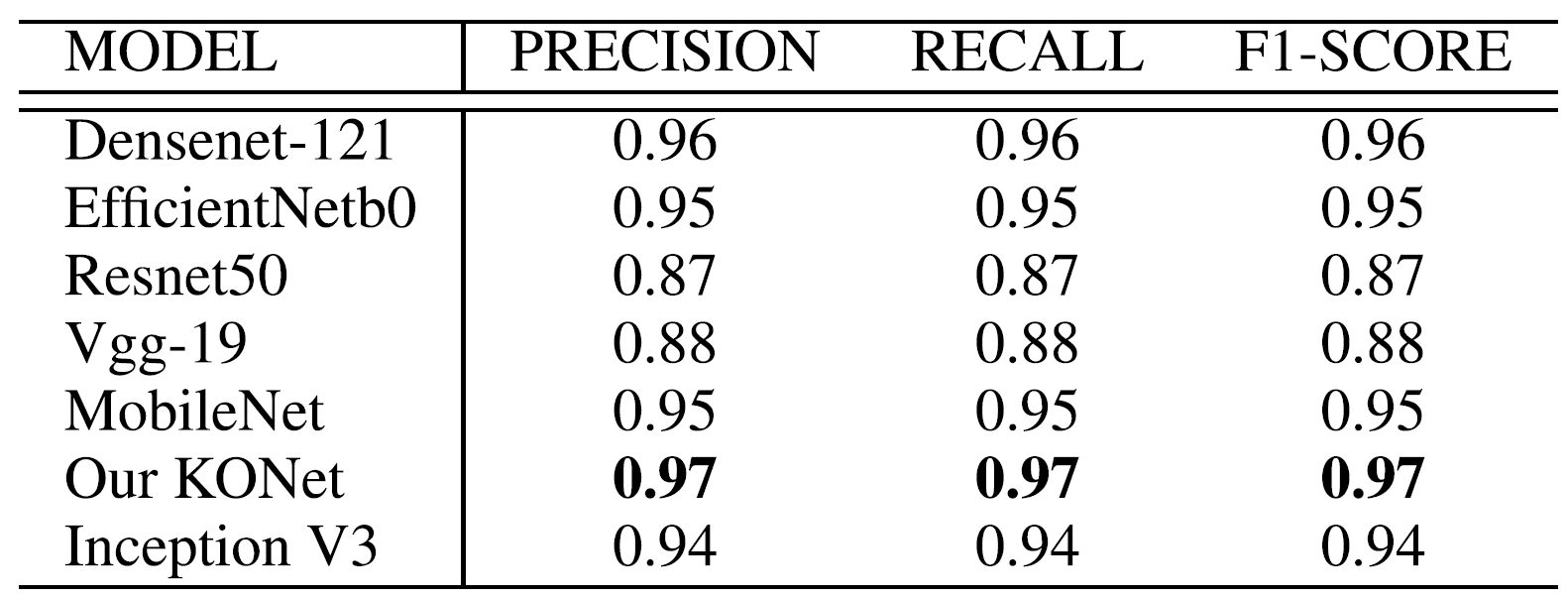


Fig : Example of dataset

# Literature survey

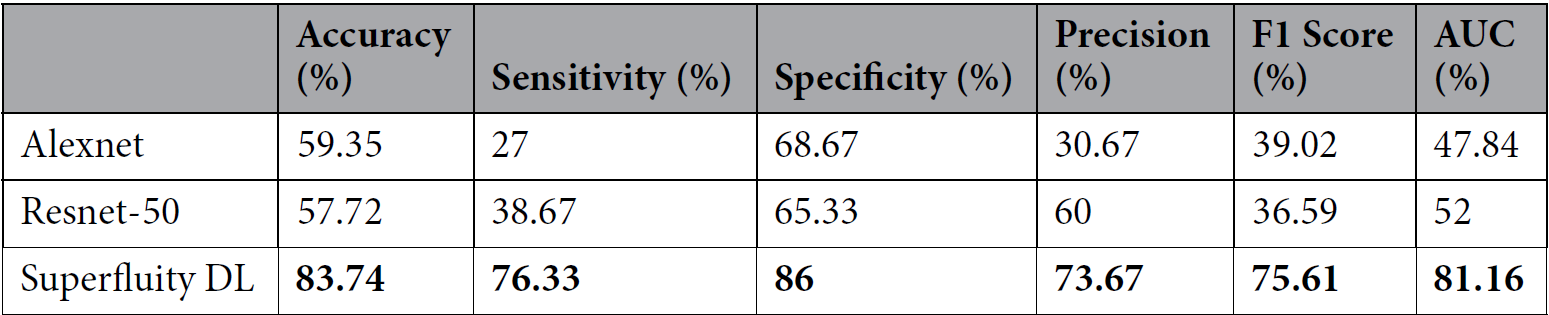
**[1] KONet: Toward a Weighted Ensemble Learning Model for Knee Osteoporosis Classification - M. J. A. Rasool, S. Ahmad, U. Sabina and T. K. Whangbo**

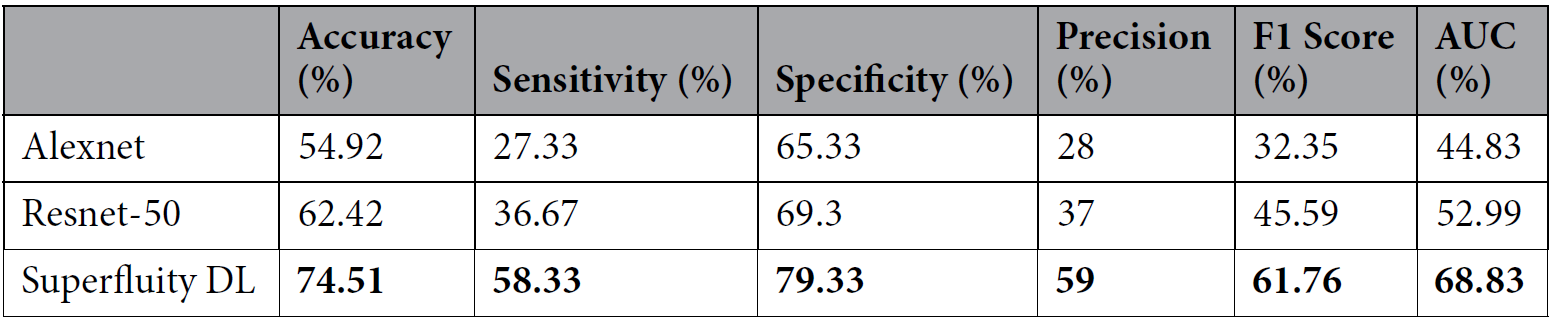
This study presents a robust detection method called KONet, designed to address the challenges in identifying knee osteoporosis, a skeletal disorder characterized by bone tissue degradation and low bone density, which increases the risk of knee fractures. Traditional diagnostic methods like knee radiography often require specialist expertise, and the large volume of X-rays, coupled with subtle variations, can result in misinterpretation. Leveraging recent advancements in deep learning, particularly convolutional neural networks (CNNs), KONet employs a weighted ensemble approach to effectively distinguish between normal and osteoporotic knee conditions, even with minor data variations. Experimental results demonstrate that the proposed model achieves superior accuracy compared to existing methods, significantly outperforming state-of-the-art CNN-based models using transfer learning.



**[2] A new superfluity deep learning model for detecting knee osteoporosis and osteopenia in X-ray images - Naguib, S.M., Saleh, M.K., Hamza, H.M. et al**

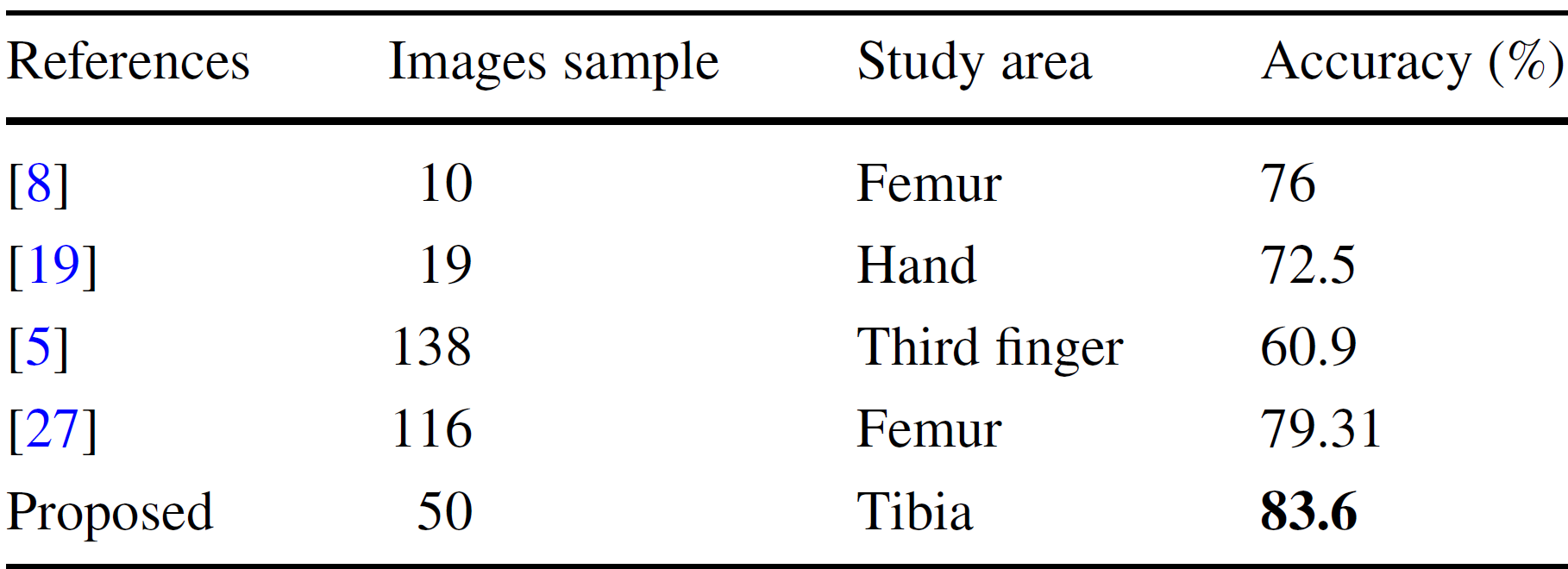
This study proposes a new deep-learning approach using a superfluity mechanism to categorize knee X-ray images into osteoporosis, osteopenia, and normal classes. The superfluity mechanism involves concatenating multiple layers, allowing features to flow into two branches. The proposed Superfluity DL model demonstrated the highest accuracy (85.42% for dataset1 and 79.39% for dataset2). Osteoporosis is a major public health problem, causing a reduction in bone mass and bone architecture disruption, increasing the risk of fracture and compromised bone strength. Early detection of osteoporosis/osteopenia using digital X-ray images would be useful and cost-effective in preventing fractures and other bone complexities. This study contributes to the development of a reliable, automated tool for initial assessment and reducing false positives and false negatives, enhancing the precision with which knee osteoporosis and osteopenia can be detected, allowing healthcare providers to prioritize patient care more effectively.





**[3] Detection of osteoporosis using image processing methods - Bo, Y., Chen, G., Li, L. et al.**

This paper highlights osteoporosis as a silent, painless illness that increases bone fragility and fracture susceptibility across various areas, emphasizing the importance of early detection for prevention and treatment. While primary osteoporosis is more common in postmenopausal women and elderly men, the disease can affect individuals of any gender or age. Characterized by decreased bone mass, structural weakness, and increased fracture risk, its assessment often relies on bone density measurements, a key indicator for evaluating osteoporosis risk, though not equivalent to the physical density of bone. Diagnostic tools such as bone densitometry, MRI, and MRS provide insights into bone structure and marrow composition but are often costly and inaccessible. To address these challenges, computer-assisted diagnosis (CAD) methods—including image processing, intelligent neural network algorithms, and genetic algorithms—offer solutions for improving image analysis, speeding up diagnoses, and reducing costs and medical errors. These advancements aim to enhance diagnostic accuracy and efficiency while mitigating the limitations of traditional methods.



**[4] Comparative analysis of vision transformers and convolutional neural networks in osteoporosis detection from X-ray images - Sarmadi, A., Razavi, Z.S., van Wijnen, A.J. et al**

This paper investigates the potential of the Vision Transformer (ViT) in medical image analysis, focusing on diagnosing osteoporosis using X-ray radio-images. Compared to traditional convolutional neural networks (CNNs), ViT demonstrates superior performance, highlighting its effectiveness when sufficient training data is available. The study underscores that both ViT and CNNs can offer robust solutions to critical medical diagnostic challenges. Deep learning, a subset of machine learning, leverages artificial neural networks (ANNs) to extract knowledge from extensive datasets and perform complex tasks. Its applications span diverse domains, including computer vision, natural language processing, speech recognition, recommender systems, autonomous vehicles, image generation, machine translation, game playing, and natural language understanding. Beyond these, deep learning is reshaping art, culture, and society by fostering innovation and creating new paradigms. In medical sciences, deep learning holds transformative potential, capable of handling large-scale and high-dimensional data while learning abstract and sophisticated functions, often with reduced dependence on extensive training datasets or predefined assumptions.

**[5] Application of a Neural Network Classifier to Radiofrequency-Based Osteopenia/Osteoporosis Screening - J. Adams, Z. Zhang, G. M. Noetscher, A. Nazarian and S. N. Makarov**

This study introduces a neural network classifier to detect osteoporotic/osteopenic conditions using non-ionizing, very-low-power radiofrequency (RF) signals transmitted through the wrist. Data was collected from 67 participants divided into two groups: Group 1 included 27 osteoporotic/osteopenic individuals (DXA T-score < -1), and Group 2 included 40 healthy individuals without major fracture risk factors. RF spectra (30 kHz–2 GHz) were measured using dual antiphase patch antennas with controlled pressure. Input features included wrist circumference and the normalized transmission coefficient (S21N). A Multi-Layer Perceptron (MLP) neural network was implemented using MATLAB, and leave-one-out cross-validation was performed. Training utilized complex spectra, which combined magnitude and phase for 201 frequency points per wrist. Among tested methods, independently processing left and right wrist spectra, then combining results manually, yielded the best performance. The model achieved a sensitivity of 83% and specificity of 94%. This approach demonstrates potential for quick, cost-effective osteoporosis screening using small, portable devices.

**[6] Using machine learning techniques to predict the risk of osteoporosis based on nationwide chronic disease data - Tu, JB., Liao, WJ., Liu, WC. et al**

This study developed a machine learning (ML) model to predict osteoporosis risk using chronic disease data from 10,000 patient records in the German Disease Analyzer database. Ten chronic conditions, including hypertension, diabetes, COPD, and cancer, were analyzed alongside demographic features like age and gender. Data was split into training (70%) and test (30%) sets using stratified random sampling, and class imbalance was addressed with oversampling. Recursive Feature Elimination (RFE) was used across nine ML algorithms to identify relevant predictors, retaining all eleven features for model development. A stacked ensemble model (Stacker) combining Logistic Regression, AdaBoost, and Gradient Boosting Classifier achieved the best performance. The Stacker model demonstrated superior accuracy, AUC-ROC (0.76), sensitivity (0.722), and specificity (0.664) compared to individual algorithms. SHAP analysis revealed the most influential features: age, gender, lipid metabolism disorders, cancer, and COPD. The model's reliability was validated using cross-validation, confusion matrix, calibration curves, and lift curves. The study underscores the potential of ML in early osteoporosis detection, enabling personalized management strategies. The findings provide a robust framework for leveraging chronic disease data to improve public health outcomes.

**[7] Osteoporosis Prediction Using Machine‑Learned Optical Bone Densitometry Data - Miura, K., Tanaka, S.M., Chotipanich, C. et al**

The study investigates the classification of osteoporosis using a machine learning (ML) model based on bone mineral density (BMD) measurements obtained from dual-energy X-ray absorptiometry (DXA) scans. A total of 203 participants, including 182 patients and 21 hospital staff, underwent BMD assessments of the lumbar spine, femur, and forearm. The optical measurements were taken using a system comprising two convex lenses, annular slits, a photodetector, and a laser diode operating at 850 nm wavelength. The T-scores were calculated to classify participants into osteoporosis (T-score ≤ -2.5) and non-osteoporosis groups. The ML model's parameters were optimized through grid-search with fivefold cross-validation, aiming to minimize the root mean squared error (RMSE). The study also employed ROC analysis to determine the optimal cutoff value for T-scores using Youden's index. Statistical analyses included Student's t-test and Pearson's chi-square test to compare group characteristics. The findings contribute to improving osteoporosis discrimination performance through advanced optical measurement techniques and machine learning methodologies.

**[8] Revolutionizing Osteoporosis and Bone Fracture Diagnostics: The Emergence of Microwave Antenna Technology - Miura, K., Tanaka, S.M., Chotipanich, C. et al**

Osteoporosis, a condition marked by reduced bone density and weakened bone structure, poses a significant health challenge, especially among the elderly, often resulting in fractures and hospitalizations. While established diagnostic tools like DEXA, MRI, QCT, and QUS are widely used, microwave sensing is emerging as a promising alternative. Its non-invasive nature, affordability, and accuracy make it an attractive option for early detection of osteoporosis and fractures. Recent advancements in microwave sensing focus on improved measurement techniques and innovative antenna designs, offering a practical and reliable approach for identifying bone changes and enabling earlier, more effective patient care.

**[9] The clinical use of quantitative ultrasound (QUS) in the detection and management of osteoporosis - D. Hans and M. -a. Krieg**

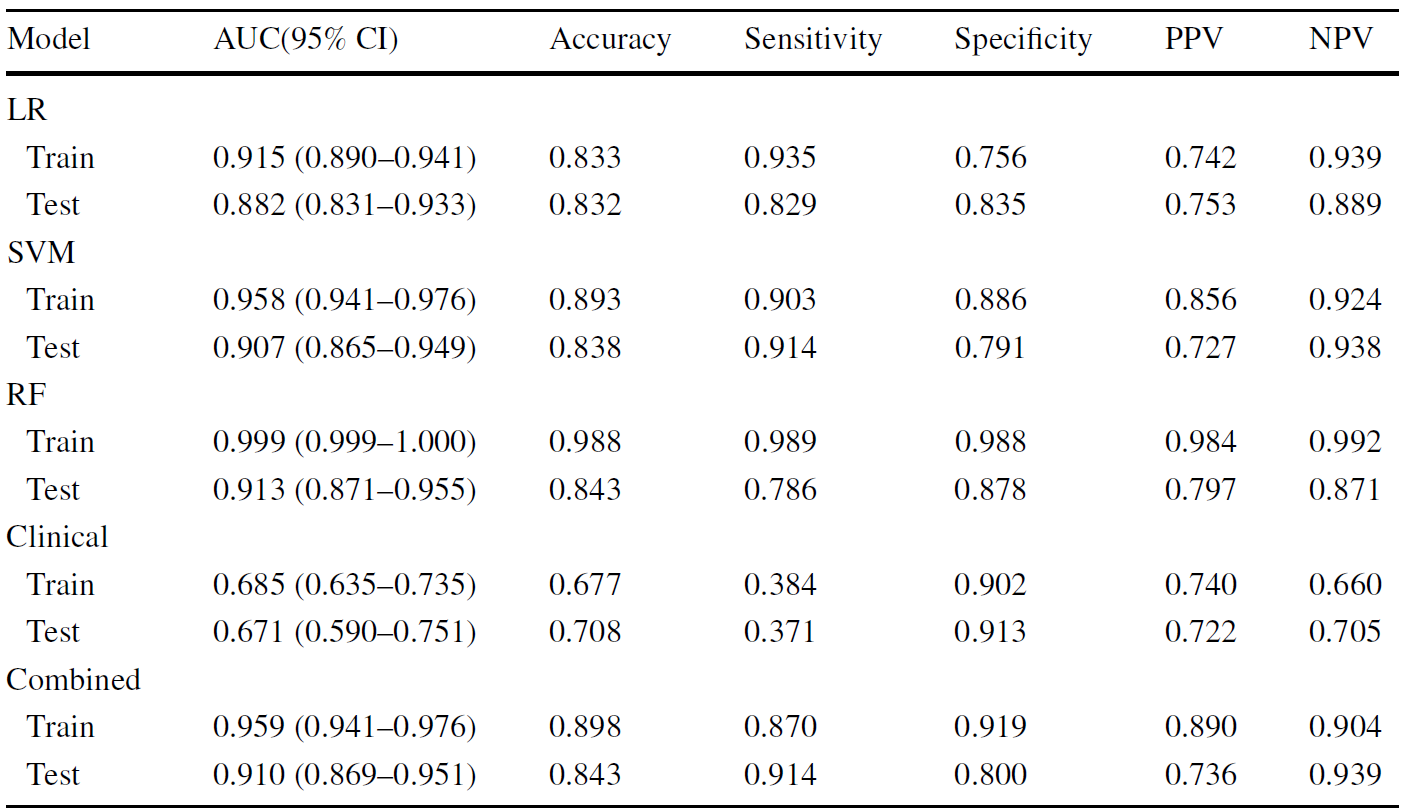
Quantitative ultrasound (QUS) is emerging as an affordable and accessible alternative to dual-energy X-ray absorptiometry (DXA) for evaluating bone mineral density (BMD) in specific situations. This review explores the use of QUS in six areas: assessing fracture risk, diagnosing osteoporosis, initiating and monitoring treatment, identifying cases, and ensuring quality control. Heel QUS, the most studied device, has proven effective for fracture risk assessment, particularly in Caucasian women over 55. While some QUS devices show moderate accuracy in diagnosing osteoporosis, their role in guiding and monitoring treatment remains limited. A practical protocol combining clinical risk factors with heel QUS is suggested for identifying high-risk individuals.

**[10] Detection and Monitoring of Osteoporosis in a Rat Model by Thermoacoustic Tomography - Z. Chi, X. Liang, X. Wang, L. Huang and H. Jiang**

Osteoporosis, a condition marked by reduced bone mineral density (BMD) and poor bone quality, poses a significant health challenge. This study explores the use of thermoacoustic tomography (TAT) as a noninvasive method to detect and monitor osteoporosis over time. Researchers used a rat model, where osteoporosis was induced in four rats through bilateral ovariectomy, while sham-operated rats served as controls. The right tibia of each rat was imaged with TAT at five intervals over 100 days, and the findings were validated using micro-computed tomography (Micro-CT). In sham-operated rats, thermoacoustic signal intensities steadily increased, reflecting normal bone growth, while osteoporotic rats showed fluctuating signals, indicating abnormal bone changes. These results reveal clear differences between healthy and osteoporotic bone, demonstrating TAT's ability to detect osteoporosis. The study highlights TAT as a promising tool for noninvasive, early detection and long-term monitoring of osteoporosis, with potential clinical applications in managing this condition.

**[11] A diagnostic approach integrated multimodal radiomics with machine learning models based on lumbar spine CT and X‑ray for osteoporosis - Cheng, L., Cai, F., Xu, M. et al.**

The paper discusses a study on predicting bone density using radiomics features extracted from medical imaging. It involved retrieving image data from picture archiving and communication systems and clinical data from hospital information systems. Two radiologists performed manual segmentation of the regions of interest (ROI) using 3D slicer and ITK-SNAP software. Radiomic features were extracted using Python and the pyradiomics library, focusing on various matrices and shape-based features. Dimensionality reduction was achieved through intraclass correlation coefficient analysis, z-score transformation, and the minimum redundancy maximum relevance algorithm. The final model combined selected radiomics features with clinical risk factors, evaluated using logistic regression, support vector machine, and random forest algorithms. Performance metrics included AUC, accuracy, sensitivity, specificity, and predictive values. The study concluded that the combined model effectively predicts bone quality, demonstrating its potential for clinical application in osteoporosis screening.

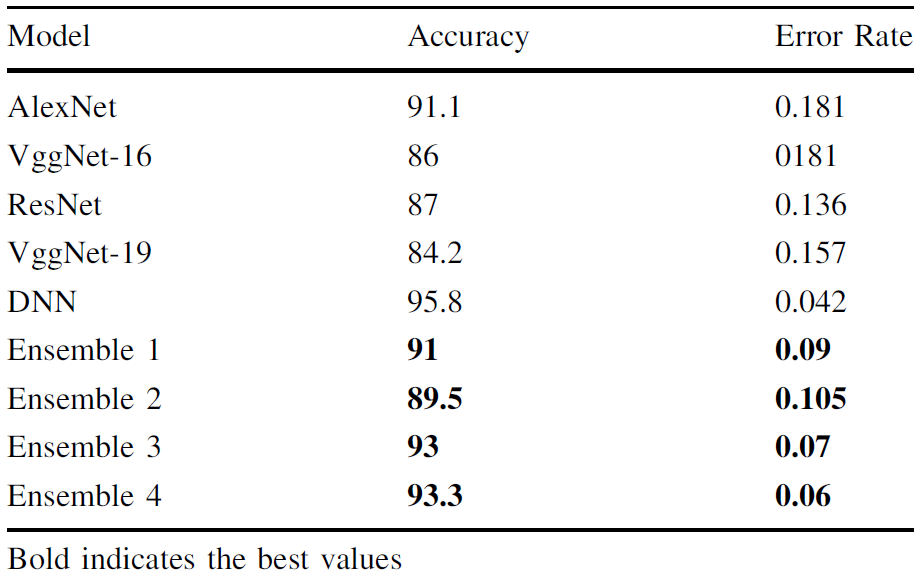


**[12] Interpretable Deep-Learning Approaches for Osteoporosis Risk Screening and Individualized Feature Analysis Using Large Population-Based Data: Model Development and Performance Evaluation - Suh B, Yu H, Kim H, Lee S, Kong S, Kim JW, Choi J**

The paper discusses the development and evaluation of deep learning (DL) models for predicting bone mineral density (BMD) using data from the Korean Health and Nutrition Examination Survey (KNHANES) and the National Health and Nutrition Examination Survey (NHANES). The authors implemented a DL algorithm consisting of three layers, including two dense layers with rectified linear unit activation functions and a final softmax layer. They optimized hyperparameters through five-fold cross-validation, achieving different configurations for NHANES and KNHANES datasets. The NHANES model had dense layers with 128 and 16 nodes for femoral neck BMD, while the KNHANES model had 128 and 32 nodes for total femur BMD. Both models utilized dropout rates and the Adam optimizer with a learning rate of 0.005. The results indicated that the DL models performed effectively in predicting BMD, demonstrating the potential of machine learning techniques in health data analysis.

**[13] Deep ensemble learning for osteoporosis diagnosis from knee X-rays: a preliminary cohort study in Kashmir valley - Wani, I.M., Arora, S. Deep**

The paper discusses a study conducted on bone mineral density (BMD) among participants in Kashmir, where data was collected from a 10-day BMD camp involving 932 individuals. Out of these, 240 participants underwent knee X-rays, and their data included questionnaires and QUS system diagnoses. The study utilized a custom dataset comprising X-ray images and clinical data to enhance osteoporosis diagnosis. Two neural network models were employed: Convolutional Neural Networks (CNN) for image data and Deep Neural Networks (DNN) for clinical data. Models like VggNet-16, VggNet-19, AlexNet, Dense-Net and ResNet-18 are trained with image data. The best classification accuracies achieved by CNNs for image sets are 91.1% achieved by AlexNet and 87% by ResNet-18. The best accuracy achieved by DNN is 95.8. The integration of clinical data with imaging data improved diagnostic accuracy. The proposed ensemble model effectively combined the strengths of both CNN and DNN, leading to better performance in osteoporosis detection. The study highlights the importance of using lower radiation exposure methods for imaging in medical diagnostics.



**[14] Application of deep learning model based on unenhanced chest CT for opportunistic screening of osteoporosis: a multicenter retrospective cohort study - Huang, C., Wu, D., Wang, B. et al.**

The study investigates the use of skeletal muscle index (SMI) as a predictor for osteoporosis and fracture risk in individuals with degenerative spinal disease. It involved 1126 participants across four institutions, with data analyzed using statistical methods such as the Student's t-test and Pearson Chi-square test. Five convolutional neural network (CNN) models (Densent121, Inception\_v3, Googlenet, Resnet50, and VGG16) were constructed using cropped CT images of the T12 vertebrae, focusing on regions of interest (ROI). The models were pre-trained on the ImageNet dataset and trained over 200 epochs with a stochastic gradient descent optimizer. The SMI was calculated by dividing the skeletal muscle area by the square of the patient's height. Results indicated significant differences in age, BMI, and SMI between patients with and without osteoporosis across all institutions. The study concluded that SMI could effectively predict osteoporosis risk, highlighting its potential clinical utility.

**[15] Evaluation of deep learning‑based quantitative computed tomography for opportunistic osteoporosis screening - Oh, S., Kang, W.Y., Park, H. et al.**

The study presents a new deep learning (DL) algorithm aimed at automating vertebral segmentation and localizing the L1 and L2 vertebrae in routine clinical CT scans. The methodology involved using a U-Net network architecture with encoder and decoder modules, incorporating various image pre-processing techniques such as window level adjustments and data augmentation to enhance model performance. The algorithm was trained on paired spinal mask data to accurately draw regions of interest (ROIs) for calculating volumetric bone mineral density (BMD). Results indicated that the DL-based automated QCT BMD measurement could effectively screen for osteoporosis, with patient classification into osteoporosis, osteopenia, and normal groups based on central DXA values. The study found that 66.7% of patients fell into the low-BMD category, highlighting the potential of this method to improve osteoporosis detection without relying solely on DXA scans.

# Significance of knee osteoporosis

Accurate measurements of bone density in the knee are crucial for reducing serious consequences. Dual-energy X-ray absorptiometry (DXA) examinations or whole leg models are ineffective for estimating knee mineral density, as BMD over the entire knee area is not of direct clinical significance. In clinical practice, diagnosis of knee osteoarthritis is mainly based on X-ray radiographs. There is a demand for a rapid, non-invasive, and accurate diagnostic tool for knee osteoarthritis. Osteoporosis is a bone disease characterized by decreased bone density and strength, making bones fragile and more prone to fractures. It occurs when bone loss outpaces bone formation, leading to porous and brittle bones. Common in older adults, especially postmenopausal women, osteoporosis increases the risk of fractures in areas like the hip, spine, and wrist. Factors such as aging, hormonal changes, calcium deficiency, and lack of physical activity contribute to its development. Challenges such as knee degeneration, delayed treatment, increased post-fracture hospitalization duration, and major depressive disorder secondary to femoral neck fractures are common, particularly in elderly women. The study aims to develop a rapid diagnostic tool for early detection and support decision-making in clinical practice.

# Transfer learning in deep learning models

Transfer learning is a transformative approach in deep learning that leverages the knowledge embedded in models pre-trained on large, diverse datasets to tackle new tasks where annotated data might be limited or expensive to obtain. This paradigm operates on the premise that the hierarchical features learned from one task—often capturing general patterns such as edges, textures, and abstract shapes in the case of image data, or syntactic and semantic structures in natural language—can be repurposed to serve as a robust foundation for a different, but related, task. By fine-tuning these pre-trained networks on specific target data, researchers can not only accelerate the training process but also improve model performance and generalization, especially in scenarios where starting from scratch would result in suboptimal convergence or overfitting. The utility of transfer learning is evident across a broad range of applications, from computer vision tasks like object detection and medical imaging analysis to natural language processing challenges such as sentiment analysis and machine translation, where models like convolutional neural networks (CNNs) and transformer-based architectures (e.g., BERT and GPT) have demonstrated remarkable adaptability. Despite its many advantages, the process of transferring knowledge is not without challenges; issues such as negative transfer—where the pre-trained features may be misaligned with the target task—necessitate careful consideration of domain similarity and task relevance. Moreover, the development of methods for domain adaptation and multi-task learning continues to evolve, providing frameworks that help mitigate discrepancies between source and target distributions while ensuring that the learned representations remain both robust and task-specific. As deep learning applications become increasingly prevalent in complex real-world environments, transfer learning stands as a critical tool in the researcher’s arsenal, enabling efficient reuse of previously acquired knowledge and fostering advancements in the design of adaptive, high-performing models.

# Osteoporosis diagnose methodologies

• **Dual-Energy X-ray Absorptiometry (DEXA):** Although traditionally used for the hip and spine, specialized DEXA protocols have been adapted to assess bone mineral density in the knee—particularly in the distal femur and proximal tibia—offering a non-invasive, quantitative measure of local bone quality.  
• **Quantitative Computed Tomography (QCT):** QCT provides three-dimensional imaging that differentiates between cortical and trabecular bone, enabling detailed analysis of the knee’s subchondral regions and early detection of microarchitectural deterioration associated with osteoporosis.  
• **Magnetic Resonance Imaging (MRI):** While MRI is predominantly utilized for soft tissue evaluation, its sensitivity to changes in bone marrow composition and microfractures makes it a valuable tool in research settings for identifying osteoporotic changes in the knee.  
• **Quantitative Ultrasound (QUS):** Emerging as a radiation-free alternative, QUS offers portable assessment of bone quality at peripheral sites, including the knee, although its variability in sensitivity currently limits its widespread clinical adoption.  
• **Biochemical Markers of Bone Turnover:** Complementing imaging modalities, the analysis of systemic biomarkers—such as serum osteocalcin and bone-specific alkaline phosphatase—provides dynamic insights into bone remodeling processes, thereby supporting the diagnosis and longitudinal monitoring of osteoporotic changes in the knee region.

# convolutional neural networks (cnn) models

Convolutional Neural Networks (CNNs) consist of three primary layers: **convolutional, pooling, and fully connected**. The convolutional layer extracts features from input images, while pooling reduces spatial dimensionality. Stacking multiple layers allows for higher-level feature maps extraction. Transfer learning in deep learning involves feature extraction and fine-tuning.

Feature extraction removes the top classification layer, while fine-tuning uses pre-trained model weights to adapt general features to specific ones. This study employed pre-trained weights from ImageNet and implemented a transfer learning strategy to mitigate over fitting due to limited trainingdata.

**Normalization:**

Normalization is a crucial step in preparing images for deep learning models. It involves adjusting the pixel intensity values to a consistent range, such as [0, 1] or [-1, 1]. This ensures that features are on the same scale, preventing any single feature from dominating the learning process. Common normalization techniques include Min-Max scaling, where pixel values are scaled to the range [0, 1], and standardization, which transforms values to have a mean of 0 and a standard deviation of 1.

**Resizing:**

Resizing ensures that all images in a dataset are of uniform dimensions, which is a requirement for most deep learning models. Resizing can be done by scaling the image while maintaining its aspect ratio or by cropping and padding to achieve the desired size. Proper resizing preserves the essential features in the image while aligning the input data to the model's architecture. This step is vital for consistent performance and computational efficiency.

**Feature Extraction:**

Our models use pre-trained convolutional layers to extract key features from images, creating feature maps that capture visual patterns. Activation functions and max pooling enhance and reduce these features, allowing them to be used for tasks like image classification. This approach leverages learned knowledge from a large dataset to extract meaningful features effectively.

**Classification:**

The classification process in our models involves extracting features from an input image to assign a label or category. The output is flattened into a vector and passed through a classifier, which makes predictions based on the extracted features. The classifier consists of dense layers and a SoftMax activation function, which adjust weights during training to map extracted features to specific classes. During inference or testing, the input image is passed through the feature extraction layers, and the output of the SoftMax layer represents the probabilities of each class.

# plan of solution

To address the challenge of early osteoporosis detection using deep learning, this study proposes a structured approach leveraging convolutional neural networks (CNNs) and transfer learning. The first step involves dataset preparation, where 1947 knee X-ray images will be preprocessed, augmented, and split into training and testing sets to improve model generalization. Next, multiple pre-trained CNN architectures, including VGG-16, ResNet-50, VGG-19, DenseNet, and XceptionNet, along with a custom-designed CNN, will be fine-tuned for binary and multiclass classification of knee joint X-ray images into normal, osteopenia, and osteoporosis categories. Transfer learning will be employed to harness the feature extraction capabilities of these pre-trained models, reducing the need for extensive dataset size while enhancing accuracy. To further optimize performance, a hybrid model combining two or more architectures will be explored to achieve higher accuracy with reduced computational complexity. Model evaluation will be conducted using standard performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to ensure reliability. Additionally, feature maps generated by CNNs will be analyzed to interpret model decisions and provide diagnostic insights for medical professionals. The final step will involve comparative analysis to determine the most effective approach, balancing accuracy and computational efficiency.

# Methodology

**1. Data Collection and Preprocessing**

The dataset used in this study comprises knee X-ray images classified into three groups: Normal, Osteopenia, and Osteoporosis. To ensure consistency and enhance model performance, the following preprocessing steps are applied:

**1.1 Normalization**

To standardize pixel intensity values and improve learning efficiency, Min-Max scaling is applied, rescaling pixel values to the range [0,1]. This prevents any individual feature from disproportionately influencing the model.

**1.2 Image Resizing**

All images are resized to 224×224 pixels, a standard input size for deep learning models. This ensures uniformity while preserving essential structural details crucial for osteoporosis diagnosis.

**1.3 Data Augmentation**

Given the limited availability of medical image datasets, data augmentation is performed to increase model generalization. Techniques such as horizontal flipping, rotation, zooming, brightness adjustments, and contrast enhancement are used to introduce variations without altering the underlying pathology.

**2. Model Architecture and Selection**

**2.1 Convolutional Neural Networks (CNNs)**

CNNs are employed for automatic feature extraction and classification. A custom CNN model is developed with:

* Multiple Convolutional Layers using ReLU activation
* Pooling Layers for feature reduction and spatial invariance
* Fully Connected Layers for final classification

To further improve accuracy, transfer learning is implemented using well-established models, including:

* VGG-16, VGG-19 – Known for capturing hierarchical features
* ResNet-50 – Designed for deep networks with residual learning
* DenseNet-121 – Enhances feature propagation through dense connectivity
* XceptionNet – Employs depthwise separable convolutions for computational efficiency

**2.2 Hybrid Model Approach**

A hybrid ensemble model is formulated by integrating features from multiple architectures. The proposed approach combines:

* VGG-19, InceptionResNetV2, and MobileNetV2, leveraging their individual strengths.
* Grad-CAM visualization is applied to interpret model decisions by highlighting critical image regions influencing predictions.

**3. Training Strategy and Optimization**

**3.1 Transfer Learning & Feature Extraction**

* Pre-trained models on ImageNet are fine-tuned on the osteoporosis dataset.
* Convolutional layers extract high-level features, which are then refined by fully connected layers.

**3.2 Classification Pipeline**

* Extracted features are flattened and passed through Dense Layers with activation functions to distinguish between the three osteoporosis classes.
* A SoftMax activation function is used to generate probability scores for classification.

**3.3 Hyperparameter Tuning**

To optimize model performance, various configurations are tested:

* Batch sizes of 16, 32, and 64
* Learning rates of 0.0001, 0.001, and 0.01
* Dropout values of 0.3 and 0.5 to prevent overfitting

**3.4 Loss Function and Optimization Algorithm**

* Categorical Cross-Entropy Loss is used to handle the multi-class classification problem.
* Adam optimizer is chosen for its adaptive learning capabilities and efficient gradient updates.

**4. Performance Evaluation**

The effectiveness of the proposed model is assessed using the following metrics:

* Accuracy – Measures the percentage of correct predictions.
* Precision & Recall – Evaluates class-wise performance in detecting osteoporosis stages.
* F1-Score – Provides a balanced measure of precision and recall.
* ROC-AUC Score – Assesses the overall discriminatory power of the model.

**5. Deployment & Future Enhancements**

* The trained model can be deployed as a clinical decision-support tool for automated osteoporosis detection.
* Future research will focus on expanding the dataset, cross-validation, and real-time implementation for improved diagnostic reliability.

# Comparitive analysis

1. **Custom CNN for Multiclass Model:**

A custom CNN is a network architecture designed to solve specific tasks or problems. It involves choosing the right number and type of layers based on input data characteristics and task objectives. The proposed model is a deep convolutional neural network (CNN) for image classification, consisting of 13 convolutional layers, 4 max pooling layers, and 3 fully connected layers. Batch normalization is applied after each layer, and the output layer has 3 units with softmax activation.

Fig: Model accuracy for Custom CNN

Fig: Model loss for Custom CNN

Fig: Model loss for Custom CNN

Fig: Model accuracy scores for Custom CNN

Fig: Confusion Matrix for Custom CNN

1. **VGG19 for Binary Model:**

VGG19 is a deep convolutional neural network with 19 layers, designed for image classification and feature extraction tasks. It uses a stack of small 3x3 convolutional filters, combined with max-pooling layers, to progressively capture spatial hierarchies. The model ends with fully connected layers and softmax for classification. Known for its simplicity and effectiveness, VGG19 is widely used in transfer learning

Fig: Model accuracy for VGG19

Fig: Model loss for VGG19

Fig: Model accuracy scores for VGG19

Fig: Confusion Matrix for VGG19

1. **ResNet-50 for Binary Model:**

ResNet-50 represents a convolutional neural network characterized by its depth of 50 layers. It allows for the loading of a pre-trained version, having been trained on an extensive dataset comprising over a million images sourced from the ImageNet database. It uses residual connections, allowing the input to skip layers and be added to the output of later layers, enabling more efficient learning and faster convergence.

Fig: Model accuracy for ResNet-50

Fig: Model loss for ResNet-50

Fig: Model accuracy scores for ResNet-50

Fig: Confusion Matrix for ResNet-50

1. **DenseNet-121 for Binary Model:**

DenseNet-121 is a deep convolutional neural network with 121 layers, designed for efficient feature reuse and gradient flow. Unlike traditional architectures, each layer in DenseNet connects to every other layer, ensuring maximum information flow and reducing redundancy. This design improves parameter efficiency and reduces the risk of overfitting. DenseNet-121 is widely used for image classification and segmentation tasks due to its strong performance and compact model size.

Fig: Model accuracy for DenseNet-121

Fig: Model loss for DenseNet-121

Fig: Model accuracy scores for DenseNet-121

Fig: Confusion Matrix for DenseNet-121

1. **XceptionNet for Multiclass Model:**

XceptionNet (Extreme Inception) is a deep convolutional neural network architecture that improves upon Inception by replacing its modules with depthwise separable convolutions. This modification reduces computational complexity while maintaining high accuracy, especially for image classification and recognition tasks. XceptionNet is widely recognized for its efficiency and strong performance on large-scale datasets, making it a powerful choice for deep learning applications.

Fig: Model accuracy for XceptionNet

Fig: Model accuracy for XceptionNet

Fig: Model loss for XceptionNet

Fig: Model accuracy scores for XceptionNet

Fig: Confusion Matrix for XceptionNet

# prososed system (hybrid model)

**1. VGG19 + InceptionResnetV2 + MobileNetV2**

**Architecture:**

**Feature Extraction:** The input image is resized to a common dimension (e.g., 224×224 or 299×299) and fed into three separate pretrained models (VGG19, InceptionResNetV2, MobileNetV2) where, instead of their classification layers, feature maps are extracted from intermediate layers such as the Global Average Pooling (GAP) layer.

**Feature Fusion:** The feature vectors from the GAP layers of each model are concatenated to form a joint feature representation, enhancing generalization by combining the fine-grained spatial details from VGG19, multi-scale features from InceptionResNetV2, and efficient feature extraction from MobileNetV2.

**Classification Layer:** The fused feature vector is then passed through a fully connected layer with dropout applied to reduce overfitting, followed by a softmax layer that outputs the predicted class labels.

**Why use this model?**

**Enhanced Feature Representation**: By combining the deep hierarchical features of VGG19, the multi-scale capabilities of InceptionResNet, and the efficiency of MobileNetV2, the hybrid model captures a richer set of features.

**Improved Accuracy**: The diversity in feature extraction often results in better performance on complex datasets, as the model is more robust to variations in input data.

**Generalization**: The ensemble approach can help reduce overfitting by leveraging complementary strengths, thereby generalizing better to unseen data.

**Flexibility in Deployment**: Although one branch (MobileNetV2) is lightweight, the combination allows balancing performance and computational efficiency. For example, parts of the network could be selectively pruned or optimized based on deployment needs.

**Layer Breakdown:**

**1. Input Image Preprocessing**

* **Resize & Normalization**: Input images are resized to a fixed dimension suitable for the pre-trained models and normalized to ensure consistent input values across the models.
* **Parallel Paths**: The processed input is fed into three parallel feature extraction paths, each corresponding to one of the pre-trained CNN architectures.

**2. Feature Extraction**  
The architecture leverages three pre-trained models for feature extraction:

* **VGG19**
  + A deep CNN with 19 layers, known for its simplicity and sequential structure.
  + The output of VGG19 is passed through a **Global Average Pooling (GAP) Layer**, which reduces the spatial dimensions to **1×11×1**, retaining only the depth (**512 channels**).
  + The resulting feature vector has a dimensionality of **512**.
* **InceptionResNetV2**
  + A hybrid architecture combining Inception modules and residual connections, optimized for efficient feature extraction.
  + The output of InceptionResNetV2 is passed through a **GAP layer**, reducing spatial dimensions to **1×11×1**, with a depth of **1536 channels**.
  + The resulting feature vector has a dimensionality of **1536**.
* **MobileNetV2**
  + A lightweight CNN optimized for mobile and resource-constrained environments, employing **depthwise separable convolutions**.
  + The output is passed through a **GAP layer**, reducing spatial dimensions to **1×11×1**, with a depth of **1280 channels**.
  + The resulting feature vector has a dimensionality of **1280**.

**3. Feature Fusion**

* The feature vectors from all three models are concatenated to form a single fused feature vector.
* The total dimensionality of the fused vector is:  
  **512 (VGG19) + 1536 (InceptionResNetV2) + 1280 (MobileNetV2) = 3328**.

**4. Fully Connected Layers**

* The fused feature vector is passed through:
  + A **Fully Connected Dense Layer** that maps the high-dimensional features into a lower-dimensional representation while preserving discriminative information.
  + An **activation function** (e.g., ReLU) is applied for non-linearity.
  + A **Dropout Layer** is included to prevent overfitting by randomly

**5. Output Layer**

* After additional dense layers with **ReLU activation** and **dropout**, the final output layer uses a **Softmax Activation Function** for multi-class classification.
* This layer outputs **probabilities corresponding to each class** in the target dataset.

Fig: Architecture diagram for Hybrid Model 1

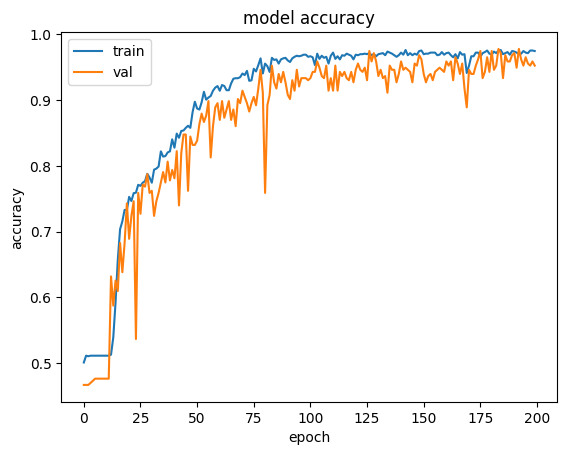
****

Fig: Model accuracy for Hybrid Model 1

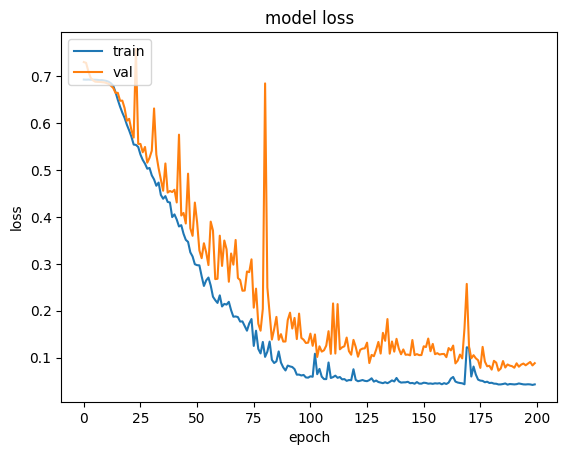


Fig: Model loss for Hybrid Model 1

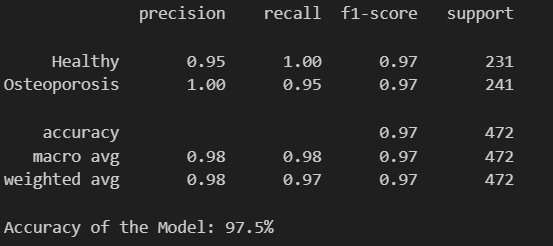
****

Fig: Model accuracy scores for Hybrid Model 1

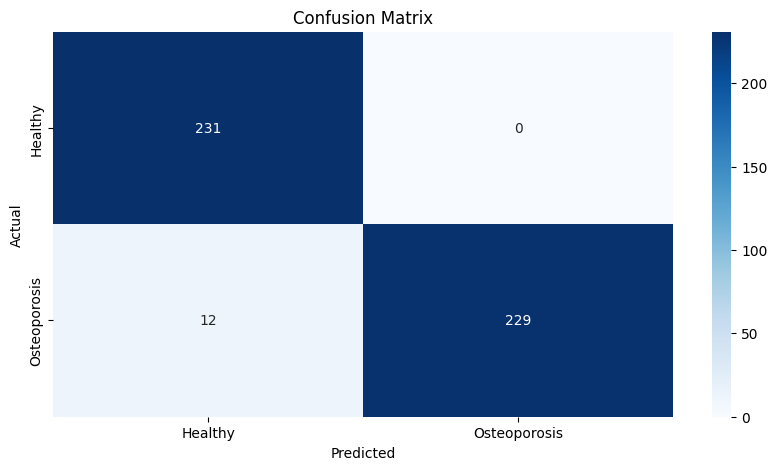
****

Fig: Confusion Matrix for Hybrid Model 1

**2. DenseNet + EfficientNet**

**Architecture:**

**Input Processing**: The input image is uniformly resized and preprocessed (e.g., to 224×224) to meet the requirements of both DenseNet and EfficientNet.

**Parallel Feature Extraction**: In the DenseNet stream, the image is processed through dense blocks and features are extracted from the final pooling or an intermediate layer; simultaneously, the EfficientNet stream processes the same image and extracts features from its global average pooling layer.

**Feature Fusion**: The feature vectors from DenseNet and EfficientNet are concatenated to form a comprehensive representation, though alternative methods like element-wise summation or attention-based fusion can also be used.

**Classification Layer**: The combined feature vector is fed through one or more fully connected layers—with dropout applied to reduce overfitting—before the final softmax (or sigmoid) layer outputs the class probabilities.

Fig: Architecture diagram for Hybrid Model 2

**Why use this model?**

**EfficientNet Component**:

**Efficiency and Scalability**: Quickly extracts primary features with minimal computation. Its scaling strategy ensures that the network can be adapted to various resource budgets.

**Low-level to Mid-level Features**: Captures basic visual patterns efficiently, laying a solid foundation for further feature refinement.

**DenseNet Component**:

**Feature Reuse and Robust Gradient Flow**: With its dense connectivity, it reuses features from multiple stages, ensuring that the network has access to both low-level and high-level information.

**Enhanced Feature Fusion**: The dense blocks help in integrating information across layers, which can lead to better overall representation and improved classification/regression performance.

**Layer Breakdown:**

**1. Shared Stem**

* The input image first passes through a **shared stem** consisting of initial convolutional layers and max pooling.
* This step reduces the spatial dimensions and extracts basic features from the image, preparing it for deeper feature extraction in the subsequent branches.

**2. Branch Split**

* After the shared stem, the network splits into two branches: **DenseNet Branch** and **EfficientNet Branch**.
* Each branch processes the input independently using its respective architecture to extract complementary features.

**3. DenseNet Branch**  
DenseNet employs densely connected blocks and transition layers:

* **Dense Blocks**: Each dense block consists of multiple layers where each layer receives feature maps from all preceding layers via concatenation. This dense connectivity improves gradient flow, promotes feature reuse, and reduces overfitting by incorporating earlier layer contexts.
  + Each layer in a dense block typically performs **batch normalization, ReLU activation, and 3×3 convolution**.
  + The **growth rate (k)** determines how many new features each layer contributes.
* **Transition Layers**: Between dense blocks, transition layers are used to control model complexity.
  + They reduce the number of channels using **1×1 convolution** and perform **downsampling via average pooling**.
  + This step ensures that feature map sizes are manageable across blocks.
* DenseNet's final block outputs high-level features that are passed to the **feature fusion stage**.

**4. EfficientNet Branch**  
EfficientNet utilizes **Mobile Inverted Bottleneck Convolution (MBConv) layers** and **squeeze-and-excitation (SE) blocks**:

* **MBConv Layers**: These layers combine **depthwise separable convolutions** with **inverted residual blocks** to efficiently capture spatial and channel-wise information.
  + The bottleneck design expands channels temporarily before reducing them back to their original size, optimizing computation.
* **Squeeze-and-Excitation Blocks**: SE blocks apply **global average pooling** followed by fully connected layers to compute **attention weights** for different channels.
  + These weights emphasize important features while suppressing irrelevant ones.
* **EfficientNet scales** depth, width, and resolution using **compound scaling coefficients (ϕ)**, ensuring balanced improvements in accuracy and efficiency across different variants.
* The EfficientNet branch outputs refined features after processing through multiple **MBConv blocks**.

**5. Feature Fusion**

* Features extracted from both branches (**DenseNet and EfficientNet**) are concatenated or fused at this stage.
* This fusion combines complementary information from both architectures:
  + **DenseNet** contributes robust feature propagation and reuse.
  + **EfficientNet** enhances efficiency and scalability with optimized representations.
* The fused features are passed to the **global average pooling layer**.

**6. Global Average Pooling**

* **Global average pooling** reduces the spatial dimensions of the fused feature maps to a **single vector per channel**, summarizing spatial information effectively.
* This step prepares the data for classification.

**7. Fully Connected / Softmax Classifier**

* The final prediction is made using a **fully connected layer** followed by a **softmax activation function**.
* The classifier assigns probabilities to predefined categories based on the extracted features.



Fig: Model accuracy for Hybrid Model 2

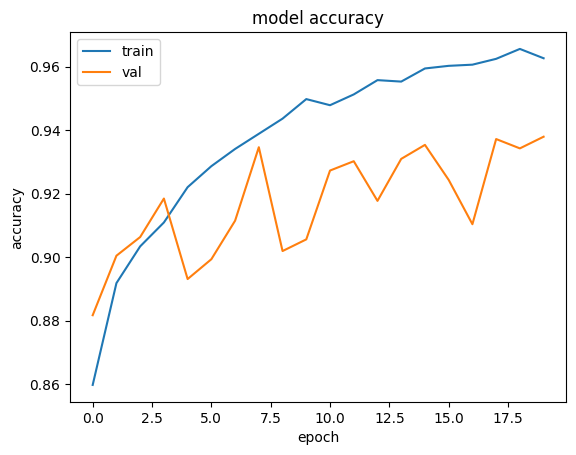


Fig: Model loss for Hybrid Model 2

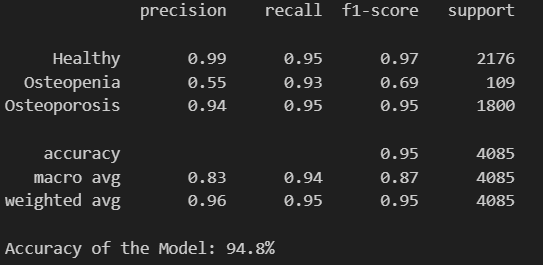


Fig: Model accuracy scores for Hybrid Model 2

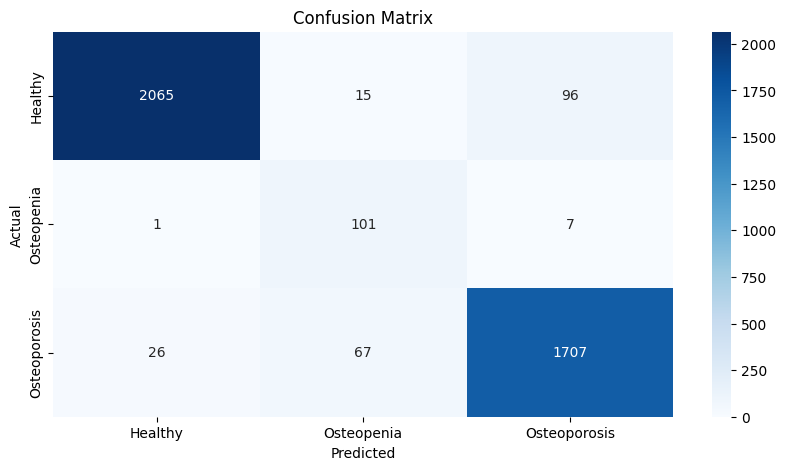


Fig: Confusion Matrix for Hybrid Model 2

# use of grad-cam

Grad-CAM (Gradient-weighted Class Activation Mapping) is a visualization technique that highlights important regions in an input image that influenced a deep learning model’s decision. It computes the gradient of the predicted class score (before softmax) with respect to the last convolutional layer’s feature maps, as this layer retains crucial spatial information. These gradients indicate each spatial location’s importance, and their global average pooling yields weights signifying each feature map’s contribution. A weighted combination of these maps is passed through ReLU to retain only positive influences, producing a heatmap that is upsampled and overlaid on the input, visually highlighting where the network "looked" when making a classification.

Below are two images that uses GRAD-CAM to highlight the important features of the image:

1. **Red/Yellow Areas**: These colors indicate high activation values. In other words, the model finds these regions very important for its prediction.

2. **Green**: Represents intermediate activation values.

3. **Blue**: Indicates low activation values, meaning these areas contributed less to the prediction.

# applications and future improvmenets

Applications:

**Early Osteoporosis Detection** – Helps in diagnosing osteoporosis at an early stage using knee X-ray images.

Automated Diagnosis Assistance – Supports radiologists and doctors by providing AI-generated diagnostic insights.

**Medical Image Classification** – Extends to other bone-related diseases using X-ray or other imaging techniques.

**Telemedicine Integration** – Can be integrated into telemedicine platforms for remote osteoporosis screening.

**Clinical Decision Support** – Enhances clinical decision-making by providing feature maps for explainability.

**AI-Powered Healthcare** – Reduces workload on medical professionals and speeds up diagnosis.

Future Improvements:

**Larger and More Diverse Dataset** – Expanding the dataset with more X-ray images from diverse demographics.

**3D Imaging Analysis** – Incorporating CT or MRI scans for a more comprehensive osteoporosis assessment.

**Hybrid Deep Learning** Models – Developing an optimized hybrid model combining CNNs with transformers

.

**Edge AI Implementation** – Deploying the model on edge devices for real-time osteoporosis screening.

**Explainability and Interpretability** – Improving feature visualization techniques to enhance model transparency.

**Integration with Electronic Health Records (EHR)** – Automating patient record analysis for better clinical insights.

**Personalized Risk Assessment** – Combining patient history, genetic factors, and lifestyle data for better prediction.

**Federated Learning for Privacy** – Training models across multiple hospitals while preserving patient data privacy.

# conclusion

Herein, we presented and tested several deep learning models for automated Normal, Osteopenia, and Osteoporosis classification of knee X-ray images. We initiated the process by thoroughly preprocessing data (normalization, resizing, and augmentation) for uniformity and stable feature extraction. Both customized CNNs and some proven transfer learning networks (VGG19, ResNet-50, DenseNet-121, and XceptionNet) were deployed and tested with a 82.1%-94.5% range of accuracy.

Based on these findings, we developed two hybrid models that capitalize on the complementary strengths of multiple architectures. The highest classification accuracy of 97.5% was attained by the first hybrid model combining VGG19, InceptionResNetV2, and MobileNetV2, indicating the advantage of combining fine-grained spatial information with multi-scale feature extraction and computational efficiency. The second hybrid model, a combination of DenseNet and EfficientNet, also performed well with an accuracy of 94.8%, illustrating its capability for feature reuse and scalable processing.

In conclusion, our results show that hybrid model architectures can significantly improve diagnostic accuracy in medical image analysis, with great promise as clinical decision-support tools for osteoporosis diagnosis. Future research will emphasize increasing the dataset, incorporating cross-validation methods, and optimizing real-time deployment to further enhance diagnostic reliability and generalizability.

# references

[1] M. J. A. Rasool, S. Ahmad, U. Sabina and T. K. Whangbo, “KONet: Toward a Weighted Ensemble Learning Model for Knee Osteoporosis Classification,” in IEEE Access, vol. 12, pp. 5731-5742, 2024

[2] Naguib, S.M., Saleh, M.K., Hamza, H.M. et al. “A new superfluity deep learning model for detecting knee osteoporosis and osteopenia in X-ray images”. Sci Rep 14, 25434 (2024)

[3] Bo, Y., Chen, G., Li, L. et al. “Detection of osteoporosis using image processing methods.” J Opt 53, 2898–2908 (2024)

[4] Sarmadi, A., Razavi, Z.S., van Wijnen, A.J. et al. “Comparative analysis of vision transformers and convolutional neural networks in osteoporosis detection from X-ray images”. Sci Rep 14, 18007 (2024)

[5] J. Adams, Z. Zhang, G. M. Noetscher, A. Nazarian and S. N. Makarov, “Application of a Neural Network Classifier to Radiofrequency-Based Osteopenia/Osteoporosis Screening”, 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2021, pp. 15-18

[6] Tu, JB., Liao, WJ., Liu, WC. et al. “Using machine learning techniques to predict the risk of osteoporosis based on nationwide chronic disease data.” Sci Rep 14, 5245 (2024)

[7] Miura, K., Tanaka, S.M., Chotipanich, C. et al. “Osteoporosis Prediction Using Machine-Learned Optical Bone Densitometry Data.” Ann Biomed Eng 52, 396–405 (2024)

[8] M. Shakhawat Hossen *et al*., “Revolutionizing Osteoporosis and Bone Fracture Diagnostics: The Emergence of Microwave Antenna Technology,” in *IEEE Access*, vol. 12, pp. 160418-160440, 2024

[9] D. Hans and M. -a. Krieg, “The clinical use of quantitative ultrasound (QUS) in the detection and management of osteoporosis”, in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 55, no. 7, pp. 1529-1538, July 2008

[10] Z. Chi, X. Liang, X. Wang, L. Huang and H. Jiang, “Detection and Monitoring of Osteoporosis in a Rat Model by Thermoacoustic Tomography”, in IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology, vol. 4, no. 4, pp. 234-239, Dec. 2020

[11] Cheng, L., Cai, F., Xu, M. et al. “A diagnostic approach integrated multimodal radiomics with machine learning models based on lumbar spine CT and X-ray for osteoporosis.” J Bone Miner Metab 41, 877–889 (2023)

[12] Suh B, Yu H, Kim H, Lee S, Kong S, Kim J, Choi J “Interpretable Deep-Learning Approaches for Osteoporosis Risk Screening and Individualized Feature Analysis Using Large Population-Based Data: Model Development and Performance Evaluation” J Med Internet Res 2023

[13] Wani, I.M., Arora, S. “Deep ensemble learning for osteoporosis diagnosis from knee X-rays: a preliminary cohort study in Kashmir valley.” Neural Comput & Applic 36, 21041–21059 (2024)

[14] Huang, C., Wu, D., Wang, B. et al. “Application of deep learning model based on unenhanced chest CT for opportunistic screening of osteoporosis: a multicenter retrospective cohort study.” Insights Imaging 16, 10 (2025)

[15] Oh, S., Kang, W.Y., Park, H. et al. “Evaluation of deep learning-based quantitative computed tomography for opportunistic osteoporosis screening.” Sci Rep 14, 363 (2024).