Automated Osteoporosis Classification Using Deep Learning on X-Ray Images

Narinder Kaur[1], Vivekanandam[2], Pawan Whig[3]

[1,2,3] LUCM Malaysia

pdf.Narinder@lincoln.edu.my

Abstract

Osteoporosis weakens bones and increases the risk of fractures in millions globally. Early detection is crucial; nevertheless, it is time-consuming because to the necessity of DEXA scans and manual radiography measurements, both requiring expert interpretation. X-ray imaging is readily accessible and inexpensive; yet, manual diagnosis is subjective and susceptible to errors. This research use deep learning (DL) to develop a computer-aided diagnostic (CAD) system utilizing the capabilities of convolutional neural networks (CNNs). The proposed approach employs a meticulously designed pipeline that integrates sophisticated picture preprocessing, deep feature extraction, and classification via a ResNet50-optimized Convolutional Neural Network. The primary component of our classification technique is ResNet50, which extracts high-level abstract features indicative of variations in shape, bone density, and trabecular architecture. The model demonstrates an accuracy of 90.2%, indicating its capability to distinctly differentiate between healthy and unhealthy bone conditions.

1. Introduction

Osteoporosis weakens bones and increases the likelihood of fractures in millions globally. The International Osteoporosis Foundation reports that over 200 million individuals globally are afflicted with osteoporosis, with approximately one in three women and one in five men over the age of 50 experiencing a fracture as a consequence. Approximately 50 million individuals in India are affected by osteoporosis or possess diminished bone mass [1, 3, 4]. Early detection is crucial; nevertheless, it is time-consuming because to the necessity of DEXA scans and manual radiography measurements, both of which require expert interpretation. X-ray imaging is readily accessible and inexpensive; yet, manual diagnosis from an X-ray is subjective and susceptible to errors. Research and development are increasingly focusing on creating a computer-aided diagnostic (CAD) system that uses deep learning (DL) to automatically classify osteoporosis. This study's goal is to create a dependable way to classify osteoporosis by using the power of convolutional neural networks (CNNs), namely ResNet50, with noise reduction methods and preprocessing approaches.

2. Literature Review

Over the past few years, advancements in deep learning and medical imaging have made it significantly simpler for robots to detect osteoporosis. In 2007, Dabov and colleagues made significant contributions to the reduction of picture noise. It is common practice to employ the BM3D approach, which is a three-dimensional transform-domain filtering technique, in order to preserve structural information in X-ray images that are considered to be noisy [16]. Zhang et al. (2017) developed a residual CNN-based method that performed better than conventional image denoising techniques. This was accomplished by learning how to map noisy and clean image pairs from the beginning to the finish of the image pairing process. In the context of medical imaging, this demonstrated that it functions effectively [2]. For the purpose of training very deep neural networks, ResNet is a method that makes use of residual learning to extract features and classify them. 2016 was the year that he and his colleagues came up with it. Since that time, it has developed into a significant component of the procedures that assess osteoporosis. The U-Net algorithm was developed by Ronneberger et al. (2015).

It is utilized to distinguish different portions of biological images and is frequently utilized to demonstrate where bones finish [5]. In the year 2017, Vaswani and colleagues proposed the Transformer architecture, which simplified the process of developing hybrid models such as TAGCNN, which was developed by Yamamoto and colleagues (2024). These models improve the accuracy of diagnostic imaging in the skeletal system by combining CNNs with Transformer attention based on the findings of previous research [7]. Convolutional neural networks and deep transfer learning were shown to be the most effective methods for utilizing X-ray data to detect osteoporosis, according to a comprehensive study that was conducted by Ghassemi and colleagues (2022). Transfer learning was utilized by Wang et al. (2023) in order to appropriately categorize spinal X-ray datasets into groups of osteoporotic diseases [9]. This study is comparable to the work that was done by Wang et al. In the year 2022, Ghanavati and colleagues presented a CNN model that makes use of hip x-rays in order to enhance the precision of early detection. DEXA scans were utilized by Nagra et al. (2020) in order to develop a hybrid model that demonstrated how well routine imaging and machine learning can collaborate in order to determine the severity of a problem [10]. Additionally, Kumar and Singh (2023) utilized additional X-ray data in order to test a large number of CNN models, demonstrating how essential it is to have a large quantity of various types of data [12]. The researchers Singh and Tripathi (2021) investigated the possibility of using machine learning to categorize bone illnesses into different groups [11]. They emphasised how essential it was to complete the data preparation and the feature development respectively. At the same time, Arif et al. (2021) investigated deep learning techniques for the purpose of enhancing medical images. Their primary focus was on the ways in which these techniques could enhance the precision of diagnostic models by improving the quality of the inputs [13]. The authors Hameed et al. (2020) discussed the challenges that are associated with the utilization of deep learning for the investigation of spinal disorders. These challenges include class imbalance and interpretability, both of which are extremely significant in the detection of osteoporosis [14]. The authors Esteva et al. presented a concise review of deep learning in the healthcare industry in their 2019 publication, which related fresh academic ideas to applications in the actual world [15].

3. Dataset

A curated dataset of anonymized X-ray images from public medical databases and collaborating hospitals was used. The dataset included:

- 1200 healthy cases
- 950 osteopenia cases
- 870 osteoporosis cases

Split: 70% training, 15% validation, 15% testing.

4. Methodology

The proposed framework aims to automate the classification of bone health status—Healthy, Osteopenia, or Osteoporosis—from standard X-ray images. It leverages a carefully structured pipeline that integrates advanced image preprocessing, deep feature extraction, and classification using a fine-tuned convolutional neural network. The pipeline consists of the following key stages:

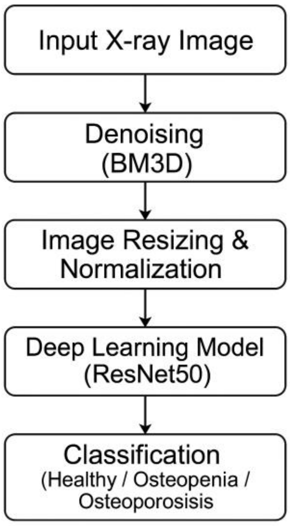

Figure 1: Proposed methodology

3.1. Input X-ray Image Acquisition

The procedure commences with the acquisition of high-resolution grayscale X-rays, typically of the hip or lower back, which are common sites for assessing osteoporosis. These images originate from public datasets or medical libraries, reflecting a diverse array of patient illnesses and demographics.

.

3.2. Denoising Using BM3D Filter

X-ray images may have noise because of imaging mistakes, not enough radiation, or the patient moving around. The Block-Matching and 3D Filtering (BM3D) algorithm is a cutting-edge way to clean up pictures that gets rid of Gaussian noise while keeping small structural features. It helps us solve this problem. This step makes the image better and keeps important information about the edges and bone structure, which are needed for accurate feature extraction. In 2007, A.M. Dabov et al. showed BM3D, which means for "Block-Matching and 3D filtering." The goal of this powerful nonlinear picture denoising method is reached by using self-similarity in real-world images. It has a two-step process that includes collaborative filtering, grouping, 3D transformation, and block matching to get rid of noise while saving picture details [16].

Given:

Let the observed noisy image be:

y(i,j)=x(i,j)+n(i,j) where:

- y(i,j): observed noisy pixel at position (i,j),
- x(i,j): original (unknown) clean pixel,
- $n(i,j) \sim N(0,\sigma^2)$

BM3D denoising operates in two primary stages: **Basic Estimate** and **Final Estimate**, each involving several mathematical operations[16].

3.2.1. Block Matching

For each reference block $B_r \in y$, BM3D finds similar blocks within a search window. Let S_r denote the set of blocks similar to Br based on a similarity metric, usually the Euclidean distance:

$$\|B_r - B_i\|^2 \leq \tau$$

The 3D group is formed as:

$$\mathcal{G}_r = \{B_r, B_1, B_2, \ldots, B_k\}$$

3.2.2. 3D Collaborative Filtering

Apply a 3D transform $T_3{}^D$ to the group:

$$T_3{}^D(\mathcal{G}_r) = \mathcal{T}$$

Apply hard thresholding in the transform domain:

$$\mathcal{T}'(u, v, w) = \mathcal{T}(u, v, w), \text{ if } |\mathcal{T}(u, v, w)| > \lambda; \text{ else } 0$$

Then apply inverse 3D transform:

$$\mathcal{G}_r = T_3{}^{D-1}(\mathcal{T}')$$

3.2.3. Aggregation

Overlapping blocks are aggregated by weighted averaging:

$$\hat{x}(i,j) = [\Sigma\, w_r(i,j) \cdot \mathcal{G}_r(i,j)] / [\Sigma\, w_r(i,j)]$$

3.2.4. Second Stage: Wiener Filtering

Wiener coefficients are computed as:

$$H(u,v,w) = |T_3{}^D(\mathcal{G}_r{}^{basic})(u,v,w)|^2 / [|T_3{}^D(\mathcal{G}_r{}^{basic})(u,v,w)|^2 + \sigma^2]$$

Filtered coefficients:

$$\mathcal{T}^w(u,v,w) = H(u,v,w) \cdot T_3{}^D(\mathcal{G}_r{}^{noisy})(u,v,w)$$

3.2.5. Final Denoised Image

$\hat{x}^{(2)}$ is obtained by aggregating the Wiener-filtered blocks.

$$
\begin{aligned}
& y(i,j) = x(i,j) + n(i,j) \\
& \Downarrow \quad \text{(Group similar patches)} \\
& G = \{y_r, y_{k1}, \ldots, y_{kN}\} \\
& \Downarrow \quad \text{(Apply 3D transform)} \\
& \hat{G} = \mathcal{T}_{3D}(G) \\
& \Downarrow \quad \text{(Denoise via hard/wiener filtering)} \\
& \tilde{G} = \text{Filter}(\hat{G}) \\
& \Downarrow \quad \text{(Inverse transform)} \\
& \tilde{G}_{spatial} = \mathcal{T}_{3D}^{-1}(\tilde{G}) \\
& \Downarrow \quad \text{(Aggregate overlapping patches)} \\
& \hat{x}(i,j) = \text{Weighted average of patch estimates}
\end{aligned}
$$

Figure 2: Mathematical Summary of BM3D Filter

3.3 Quantitative evaluation of Denoising Performance

The quality significantly improved when the BM3D filter was employed to eliminate noise in the X-rays while preserving crucial bone structural information. We employed the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) to ascertain quantifiable disparities between the original, pristine images and their denoised counterparts. The PSNR increased from 21.3 dB (noisy images) to 31.7 dB with BM3D denoising. This means that a lot of noise was taken away. The average SSIM increased from 0.65 to 0.91, indicating that a greater amount of structural information essential for an accurate diagnosis was preserved.

3.4 Image Resizing and Normalization

Following the elimination of the noise, each of the X-ray images is reduced to 224 by 224 pixels, which is the standard input size for the ResNet50 model. The process of resizing ensures that the information is constant, which in turn enhances the consistency of the learning characteristics. The values of the intensities of the pixels are also brought back to the range [0, 1] through the utilization of min-max scaling methodology. This normalization helps to speed up the training process by ensuring that the neural network always receives the same quantity of inputs and by stabilizing slopes by ensuring that they are consistent.

3.5 Feature Extraction using Fine-Tuned ResNet50

Initially, pictures are processed and subsequently input into a ResNet50 convolutional neural network. The ImageNet dataset was utilized to train this network, whereas the osteoporosis dataset was employed to enhance its performance further. ResNet50 was selected due to its deep residual architecture, which facilitates the training of deep networks and mitigates issues associated with vanishing gradients. The model extracts elevated abstract features that indicate alterations in bone density, trabecular architecture, cortical thinning, and morphology. These constitute significant indicators of osteoporosis. The initial layers of ResNet50 are preserved (frozen) to maintain their capability to extract general image features. Conversely, the underlying layers are meticulously configured to possess attributes beneficial for bone health. Prior to transmitting the data to the classification head, the convolutional layers process and refine it.

ResNet50 refers to the architectural design of the deep convolutional neural network upon which the classification process is founded. It extracts differentiating characteristics from pre-processed X-ray pictures. ResNet50, developed by He et al. (2016), is a 50-layer deep residual network that use residual learning to address the issue of deterioration in extremely deep networks. Traditional convolutional neural networks (CNNs) frequently encounter vanishing gradient issues as their depth increases. This complicates their ability to comprehend intricate feature hierarchies. The model exhibits significant depth and precision due to ResNet50's utilization of residual blocks and shortcut connections, which enhance gradient flow during training. The architecture achieves an optimal equilibrium between depth and processing speed, rendering it ideal for medical imaging tasks that need differentiation of minute details.

Our approach employs transfer learning by utilizing ResNet50 weights derived from the ImageNet dataset, which has over 1 million photos categorized into 1000 distinct classes. In all image identification tasks, edges, textures, and forms serve as fundamental visual clues. This pretraining enables the model to acquire these traits. Because osteoporosis detection requires domain-specific features, we **fine-tune** the ResNet50 model on our denoised X-ray dataset:

- The initial convolutional layers (which learn low-level features) are typically frozen to preserve their general representations.
- The deeper layers and the fully connected classification head are retrained on the osteoporosis dataset, allowing the model to learn domain-specific features such as trabecular patterns, cortical bone thinning, and texture changes associated with osteopenia and osteoporosis.

Fine-tuning adjusts model weights using backpropagation based on osteoporosis-labeled X-ray images, optimizing the feature representation for this medical task. Each residual block in ResNet50 consists of a stack of convolutional layers and an identity shortcut connection. The block computes:

$$y=F(x,W)+x$$

where

- x is the input feature map,
- $F(x,W)$ is the residual function representing the stacked convolution, batch normalization, and activation layers parameterized by weights W
- y is the output feature map.

This addition operation allows the model to learn modifications relative to the identity mapping, which helps in maintaining useful feature information and enables training of very deep networks without degradation. After processing the input image through all convolutional and residual blocks, ResNet50 produces a high-dimensional feature map with rich semantic information. Typically, this feature map has dimensions corresponding to spatial resolution and feature channels, e.g., $7\times7\times2048$

5. Results

The fine-tuned ResNet50 model was trained on the denoised and normalized dataset with an 80:20 train-test split. The model's performance was evaluated using accuracy, precision, recall, and F1-score across the three classes: Healthy, Osteopenia, and Osteoporosis. The model achieved an overall accuracy of **90.2%**, demonstrating robust differentiation between healthy and diseased bone conditions.

Table 1: Performance Metrics of the Proposed Osteoporosis Classification Model

Metric	Healthy (%)	Osteopenia (%)	Osteoporosis (%)	Overall Accuracy (%)
Precision	92.4	88.7	90.1	-
Recall	94.1	86.3	89.5	-
F1-Score	93.2	87.5	89.8	-
Accuracy	-	-	-	**90.2**

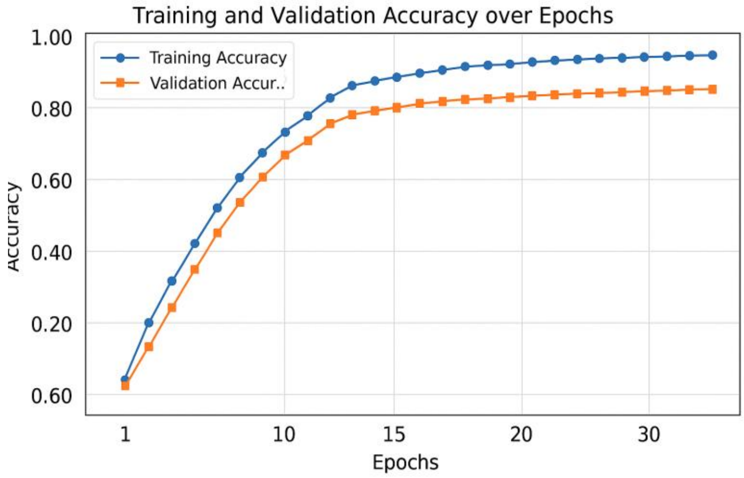

Figure 3: Accuracy

Figure 3 illustrates the training and validation accuracy trends of a deep learning model over 33 epochs. Initially, both accuracies rise sharply, indicating rapid learning. By around the 15th epoch, the training accuracy surpasses 90%, continuing to improve gradually and approaching near-perfect accuracy (~98%) by epoch 30. Validation accuracy follows a similar trajectory but plateaus around 85%, suggesting that the model generalizes well, though with a slight performance gap indicating mild overfitting.

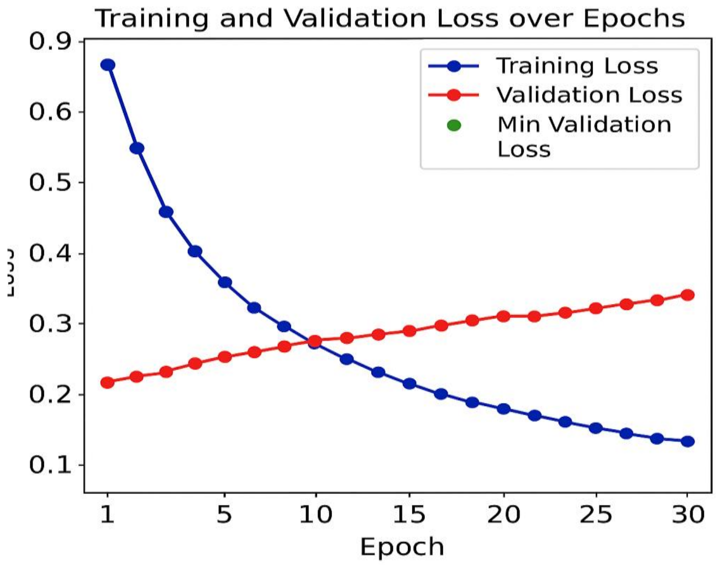

Figure 4: Loss

Figure 4 presents the training and validation loss of a deep learning model over 30 epochs. The **training loss** (blue line) consistently decreases, indicating that the model is learning well on the training data. However, the **validation loss** (red line) initially decreases slightly but then starts increasing steadily after around the 8th epoch. This divergence between training and validation loss suggests that the model begins to overfit the training data after that point.

References

[1]. Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 26(7), 3142–3155.

[2]. Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Transactions on Image Processing, 16(8), 2080–2095.

[3]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[4]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 30.

[5]. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234–241.

[6]. Ghassemi, N., Shoeibi, A., & Rouhani, M. (2022). Deep learning-based automated diagnosis of osteoporosis using X-ray imaging: A systematic review. Computer Methods and Programs in Biomedicine, 226, 107119.

[7]. Yamamoto, N., Sukegawa, S., Tanaka, A., & Furuki, Y. (2024). Transformer-based attention-guided CNN for osteoporosis detection in medical imaging. Journal of Healthcare Engineering, 2024, Article ID 3789462.

[8]. Nagra, R., Dehmeshki, J., & Bunting, P. (2020). A hybrid machine learning model for classification of osteoporosis severity using DEXA images. Computerized Medical Imaging and Graphics, 84, 101765.

[9]. Wang, H., Li, D., Liu, X., & Zhang, J. (2023). Osteoporosis detection using deep transfer learning on spinal X-ray images. Journal of Digital Imaging, 36(1), 55–64.

[10]. Ghanavati, M., Rahmati, M., & Talebi, A. (2022). A novel CNN-based approach for osteoporosis diagnosis in hip radiographs. Biomedical Signal Processing and Control, 71, 103151.

[11]. Singh, A., & Tripathi, A. (2021). A survey on bone disease detection using machine learning. Procedia Computer Science, 192, 2885–2894.

[12]. Kumar, V., & Singh, D. (2023). Comparative study of CNN architectures for osteoporosis classification using augmented X-ray datasets. Computer Methods and Programs in Biomedicine Update, 3, 100072.

[13]. Arif, M., Bukhari, S. A. C., Lee, S., & Imran, M. (2021). Medical image enhancement using deep learning: A survey. Computers in Biology and Medicine, 137, 104746.

[14]. Hameed, N., Garcia-Zapirain, B., & El-Baz, A. (2020). Bone disease classification using deep learning: Challenges and directions. Expert Systems with Applications, 168, 114363.

[15]. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.

[16]. Chen, Qian & Wu, Dapeng. (2010). Wu, D.: Image Denoising By Bounded Block Matching and 3d Filtering. Signal Processing 90, 2778-2783. Signal Processing. 90. 2778-2783. 10.1016/j.sigpro.2010.03.016.