A Robust Deep Learning Framework for Automated Bone Fracture Detection and Classification from X-Ray Images

Rizwan Ullah[1], Samreen Ihsan[2], Muhammad Salman Khan[1], Khalid Khan[3] and Hidayat Ullah[1]

[1]Department of Computer Science, University of Malakand, Chakdara, Pakistan.
[2]School of Software Technology, Dalian University of Technology, Dalian, China.
[3]Institute of Computer Science and Information Technology, University of Agriculture, Peshawar, Pakistan.

***Correspondence:** samreenihsan09@mail.dlut.edu.cn.

Citation | Ullah. R, Ihsan. S, Khan. M. S., Khan. K, Ullah. H, "A Robust Deep Learning Framework for Automated Bone Fracture Detection and Classification from X Ray Images", IJIST, Vol. 8 Issue.6 pp 2678-2694, September 2026

DOI: https://doi.org/10.33411/IJIST/2025

Received | Aug 19, 2026 **Revised |** Sep 02, 2026 **Accepted |** Sep 07, 2026 **Published |** Sep 11, 2026.

Bone fractures are a common medical condition that require accurate and timely diagnosis to prevent complications and ensure appropriate treatment. Manual interpretation of radiographic X-ray images can be challenging, particularly in complex cases, and may be affected by human error and workload. Recent advances in deep learning have demonstrated significant potential for automated fracture detection and classification. This study presents a comparative deep learning framework for automated bone fracture detection from X-ray images using Vision Transformer (ViT-B/16) and ResNet-50 architectures. A Bone Fracture Multi-Region X-ray dataset containing 10,580 grayscale images with a resolution of 224 × 224 pixels was used for training, validation, and testing. The dataset includes radiographic images from multiple anatomical regions, including the upper limbs, lower limbs, lumbar spine, hips, and knees. Appropriate preprocessing and data augmentation techniques were applied to improve data consistency and model generalization. The experimental results demonstrated that both models achieved strong performance in binary fracture classification. ViT-B/16 achieved the best performance, with a validation accuracy of 98.91% and a test accuracy of 100%. It achieved a precision, recall, and F1-score of 1.00, indicating correct classification of all test images. ResNet-50 achieved a test accuracy of 98.62%, with a precision of 0.99, recall of 0.98, and F1-score of 0.99. These results demonstrate that both architectures are effective for automated bone fracture detection, while ViT-B/16 achieved superior classification performance on the evaluated dataset. The findings highlight the potential of transformer-based and convolutional deep learning architectures to support automated bone fracture detection from radiographic X-ray images.

Keywords:
Bone fracture detection; X-ray imaging; Vision Transformer (ViT-B/16); ResNet-50; deep learning; medical image classification.

Introduction:

Bones are a crucial component of the human body structure, facilitating movement and support. Bones not only provide form and support to the body but also protect vital organs. A bone fracture refers to a disruption in the continuity of a bone, generally caused by an accident, heavy force, or medical conditions that decrease bone density. Bone fractures are extensive in the human body, and their correct evaluation is vital in medical care. According to the Global Burden of Disease study, up to 37 million fragility fractures occur annually in individuals aged over 55, equivalent to 70 fractures per minute. Recent developments in artificial intelligence, especially deep learning, have shown potential results in medical imaging, particularly for automating and improving evaluation performance. Automatic fracture detection and classification using deep learning is a popular topic of research lately, since the prevailing practice is to interpret X-ray images by clinical radiologists or orthopedists manually, which may be biased, inconsistent, and error-prone. Automated detection can help to save time in the clinical workflow and reduce costs. CNNs and other deep learning architectures can detect sophisticated structures within comprehensive datasets, enabling accurate identification of fractures that could be overlooked by human evaluation.

Convolutional Neural Networks (CNNs) such as ResNet-50 have demonstrated strong performance in medical image analysis because of their ability to extract hierarchical features from radiographic images. Recently, there has also been an emergence of Vision Transformers (ViTs), which incorporate global context using self-attention mechanisms and have resulted in a better classification performance on vision tasks. Recent studies have further demonstrated the growing role of artificial intelligence and deep learning in automated disease detection. A comprehensive review published in 2023 highlighted the increasing use of deep learning techniques for skeletal fracture detection and emphasized the need for improved model explainability, standardized evaluation, and external validation before clinical deployment [1]. Author [2] conducted a systematic review and meta-analysis of artificial intelligence for fracture detection across different imaging modalities and data types, demonstrating the expanding application of AI-based approaches in musculoskeletal imaging [2].

Similarly, a systematic review and meta-analysis by [3]. reported high pooled diagnostic performance of artificial intelligence systems for fracture detection on plain radiographs, with a sensitivity of 91.43% and specificity of 92.12% across the included studies [3]. These findings indicate the considerable potential of AI-based methods for assisting fracture diagnosis while also highlighting differences in performance among datasets, anatomical regions, and model architectures. Recent research has also explored advanced deep learning architectures beyond conventional CNNs.

Transformer-based approaches, including Vision Transformers, have attracted increasing attention because their self-attention mechanisms enable the modeling of global relationships within medical images. [4] investigated a Vision Transformer-based approach for femur fracture classification and demonstrated the potential of transformer architectures for radiographic fracture analysis. In addition, Tahir et al. developed an ensemble deep-learning model for fracture detection using X-ray images, combining several established architectures and demonstrating the potential benefits of integrating complementary deep learning models [5]. Recent work has further extended deep learning applications to specific clinical fracture patterns, such as atypical femur fractures, demonstrating the continued expansion of AI-based fracture analysis in radiology [6].

Despite these promising developments, the literature also indicates that high performance on a single dataset does not necessarily guarantee robust performance in real-world clinical environments. Differences in image acquisition, patient populations, anatomical regions, dataset composition, and validation strategies can influence model performance

[1][2][3]. Therefore, comparative evaluation of different deep learning architectures on diverse multi-region X-ray datasets remains important for understanding their relative strengths and limitations.

Developed a recurrent neural network (RNN) that uses visual attention to focus on critical areas of an image for precise detection of bone fractures, achieving efficient performance. Author [7] With two-stage cross-validation, our Deep convolutional neural network [DCNN] models provided high accuracy of 90.8 percent in triquetral fractures and 92.5 percent in Segond fractures. These findings show a high detection performance of the models.[8]The negative predictive value (NPV) of the AI model was greater than 99.5 percent on all lesion types, implying the superior accuracy of deep convolutional neural networks (CNNs) to rule out lesions. Author [9] Six studies in total and comprising 33026 radiographs, using a meta-analysis, indicated that CNNs had a high sensitivity of 92-% (95% PI: 80%-97%), and 93-% (95% PI: 76%-98%) specificity, a good indicator of the diagnostic accuracy of CNNs.

Considerable heterogeneity was noted, and the I-squared statistic was 81.9.[10] We took a Multiview application in which a Convolutional neural network [CNN] and a Recurrent neural network [RNN] were combined to analyze three radiographs in serial. This obtained Area under the curve [AUC] = 0.95 and 88% accuracy, which was a better outcome. The three-class binary model returned 86 percent accuracy, and the five-class binary model returned 81 percent accuracy [11].

The employment of the Computer Aided Design [CAD] system resulted in specialists being more accurate in their decisions by 14% demonstrating the assistance of the CNN in the decision-making process. The authors based their work on Inception V3 pretrained on ImageNet and retrained the top layer in it with the 1,389 labeled radiographs and performed binary classification [12]. This model has a score of 0.95, sensitivity of 0.88, and specificity of 0.90, better than the conventional methods of detecting fractures.[13]introduce deep learning methods to spinal lateral view radiography to improve the detection of osteoporosis and vertebral fractures (VFs).

Using a dataset of 26,299 X-rays, deep convolutional neural network scores (VERTE-X pVF and osteo scores) were developed and bettered clinical models in detecting prevalent VF and osteoporosis. A deep learning system can correctly, automatically, and quickly identify and categorize the fracture of the ribs (accuracy = 87.63), differentiate between new and old fractures (accuracy = 87.63), and identify whether misalignment occurs in fresh fractures or not (accuracy = 95.22) [14] . The authors came up with a deep learning-based approach to fracture detection in X-ray pictures of arm bones with the help of YOLACT++ and the CLAHE algorithm and trained BYOLOv4 based on a limited set of data, using data augmentation methods to improve the work of the detector [15]. Employed deep learning to detect and classify bone fractures and showed the potential of machine learning-based approaches to diagnose fractures automatically [16]. Recentl y, the DL models most appropriate to image recognition have been utilized to recognize fractures in the appendicular skeleton, wrists,11 shoulders,12, and hands and feet [17]. Deep learning-based technology is currently being employed in the field of medical analysis and diagnosis, including brain imaging, breast cancer diagnosis, lung nodule detection, abdominal disease diagnosis, and bone disease diagnosis and measurement[15] [18] .

Machine Learning (ML) and Deep Learning (DL) are increasingly becoming powerful machines in early disease detection, diagnosing, and planning treatment, analyzing large medical datasets (e.g., imaging, genomic data, clinical records, wearable sensor data) to discover hidden patterns that human experts might not have noticed. Convolutional Neural Networks (CNNs) can detect pneumonia in chest X-rays, brain tumors in MRI images, and fractures in bone scans. Deep Learning models predict diabetic retinopathy and glaucoma

based on retinal fundus scans. In this study, we propose a comparative deep learning framework using ResNet-50 and vision Transformer (ViT-B/16) for automated bone fracture detection from radiographic images. The proposed framework aims to investigate the effectiveness of convolutional and transformer-based architectures in distinguishing fractured and non-fractured X-ray images. The models are trained and evaluated using a multi-region bone fracture X-ray dataset, and their performance is assessed using standard classification metrics, including accuracy, precision, recall, and F1-score.

Objectives:

The objectives of this study are:

To develop and evaluate a deep learning-based framework for automated binary classification of fractured and non-fractured X-ray images using Vision Transformer (ViT-B/16) and ResNet-50.

To compare the classification performance of ViT-B/16 and ResNet-50 on the Bone Fracture Multi-Region X-ray dataset.

To evaluate and compare both models using standard performance metrics, including accuracy, precision, recall, and F1-score.

To identify the relative strengths of the transformer-based and convolutional architectures for automated bone fracture detection based on their experimental performance.

Novelty of the Study:

The novelty of this study lies in the comparative evaluation of two state-of-the-art deep learning architectures, Vision Transformer (ViT-B/16) and ResNet-50, for automated binary bone fracture detection using a multi-region radiographic X-ray dataset. Unlike studies that primarily investigate individual deep learning architectures or focus on specific anatomical regions, this study evaluates both convolutional and transformer-based approaches under a common experimental framework using the Bone Fracture Multi-Region X-ray dataset. The comparison provides insights into the effectiveness of global self-attention-based feature learning in ViT-B/16 and hierarchical convolutional feature extraction in ResNet-50 for fracture detection. The experimental results demonstrate that ViT-B/16 achieved a test accuracy of 100%, while ResNet-50 achieved 98.62%, highlighting the relative performance advantage of the transformer-based approach on the evaluated dataset.

Literature Review:

The deep learning models have demonstrated encouraging performances regarding the diagnosis of bone fractures. Fracture classification has been done using many deep learning models e.g., VGG, Inception, Dense Net, Res Net, etc. and these models offer sufficient accuracy in most fields. Numerous researchers applied pre-trained models to their work and obtained encouraging results. The most recent developments in deep learning have made a substantial contribution towards automated detection of bone fractures in radiographic images.

The popularity of Convolutional Neural Networks (CNNs), specifically ResNet-50, is explained by their capacity to produce hierarchical features of the medical images and high accuracy of fracture detection and classification. Experiments have demonstrated that ResNet-50 can be successfully used to distinguish between types of fractures in massive datasets, e.g., the MURA dataset, to enable clinical radiologists with reliable decision support.[19] evaluated the performance of two pretrained DCNNs for the detection of ankle fractures: Inception V3 and Resnet50. The outcome from the Inception V3 neural network outperformed Resnet-50 in overall performance criteria, including specificity, sensitivity, PPV, NPV, accuracy, F-score, and AUC, and most importantly in the detection of occult fractures (98.6 % vs. 95.8 %, respectively. An innovative deep learning algorithm was developed on the Vision Transformer architecture and was used to detect hip fractures with an accuracy of 83%[4] .DP [20] A deep neural network model has been created to categorize fractures and healthy bone.[7] This paper

recommends that deep learning methods, such as VGG16, ResNet152V2, and DenseNet201, be used to detect and diagnose bone fractures.

The dense and VGG16 model had an accuracy rate of 97%. presented a new deep learning-based system, DCFPN, to detect thigh fractures with an average precision of 82.1% and surpassing existing systems, including Faster R-CNN, Cascade R-CNN, and FPN on a new dataset, indicating that deep learning has the potential to transform clinical CAD [21].[22] evidenced the efficacy of a Vit model in distinguishing between non-displaced fractures of the neck of the femur and normal hips by use of both AP and lateral hip radiographs, with an accuracy of 95.8%.[23] proposed CNN algorithm using SFCM (Spatial Fuzzy C-Means). Preprocessing steps are the median filter and the Discrete Wavelet Transforms (DWT). It employs characteristics such as homogeneity, entropy, contrast, correlation coefficient, and energy with an approximation of 78%. Author [24] created an additional set of labels of the hip joint position and trained a preprocessing CNN to cut out the entire radiograph and zoom into this region of interest. Author [18] Conduct an extensive investigation into what sort of patient and hospital process variables DL is capable of identifying in radiographs and whether any of them help the inner workings of a fracture detection model. Author [25] introduces the new framework BoneCLIP-XGBoost, which further develops medical diagnostics and combines information in images and text using modern deep learning and machine learning algorithms.

The model has an excellent diagnostic performance by combining the Vision Transformer (ViT) to interpret X-rays with ClinicalBERT to interpret clinical narratives and optimizing the model with the XGBoost classifier. Author [17]confirmed the practicability and highlighted the importance of introducing a real-time bone fracture detection system using a deep convolutional neural network (CNN) model[26]presented a 2D deep learning and genetic algorithm-based approach towards ”Femoral Neck Fracture Detection In X-ray Images.

Their method also demonstrated the strength of deep learning, as well as genetic algorithms that improved the ability to detect fractures in particular parts of the anatomy.[27] discussed the use of deep learning in bone tumor diagnosis and emphasized that the field of deep learning continues to be explored with regard to new uses and issues. Author [28] created a deep learning convolutional network system to identify acetabular fractures. Author [29] Training an Attention-based transfer learning model to identify bone fractures in X-ray scans[30] proved that deep neural networks using radiograph data with electronic health records greatly enhance the ability to distinguish between atypical and normal femur fractures (AFF) and normal femur fractures (NFF). It has used a deep neural network (DNN) to apply Vision Transformer (ViT) and CNN models to musculoskeletal radiographs of the MURA dataset. Recent peer-reviewed research has further advanced the application of artificial intelligence and deep learning for automated fracture detection and classification [31]. provided a comprehensive review of deep learning approaches for skeletal fracture detection and highlighted the increasing application of deep learning architectures in musculoskeletal imaging [1].

The review also emphasized important challenges associated with dataset quality, model explainability, standardized evaluation, and external validation, which are important considerations for the clinical translation of AI-based fracture detection systems. Author [2] conducted a systematic review and meta-analysis of artificial intelligence in fracture detection using different imaging modalities and data types. Their findings demonstrated the expanding role of AI in fracture diagnosis while also highlighting variations in diagnostic performance according to imaging modality, dataset characteristics, and methodological approaches. Similarly [3]conducted a systematic review and meta-analysis of AI-based fracture detection from plain radiographs.

Their analysis reported pooled sensitivity and specificity of 91.43% and 92.12%, respectively, demonstrating the considerable diagnostic potential of AI systems for fracture detection. However, the study also highlighted variations among the included studies, emphasizing the importance of appropriate validation and generalizability when evaluating AI-based diagnostic systems. Recent studies have also investigated advanced and ensemble deep learning architectures for fracture detection. [5]developed an ensemble deep-learning model for fracture detection using X-ray images by combining multiple established architectures, including MobileNetV2, VGG16, InceptionV3, and ResNet-50. Their findings demonstrated the potential of combining complementary deep learning architectures for improving automated fracture classification. Author [6] developed AFFnet, a deep convolutional neural network for detecting atypical femur fractures from anteroposterior radiographs, demonstrating the application of deep learning to specific and clinically important fracture patterns.

In addition, recent research has investigated deep learning approaches for vertebral compression fracture detection using radiographic images, extending AI applications beyond simple classification toward the identification and localization of fracture abnormalities [32]. The recent literature further emphasizes the importance of evaluating AI systems in clinically relevant settings. Author [33] investigated the potential of a deep learning tool to improve fracture detection by radiologists and emergency physicians using extremity radiographs. These findings indicate that AI systems may serve as decision-support tools that complement clinical expertise rather than simply functioning as standalone classifiers.

Overall, recent developments demonstrate substantial progress in AI-based fracture detection while also identifying continuing challenges related to external validation, dataset diversity, interpretability, and clinical integration. These developments provide further motivation for comparative evaluation of convolutional and transformer-based architectures, such as ResNet-50 and Vision Transformer (ViT-B/16), for automated fracture classification.

Proposed Methodology:

The proposed methodology presents a deep learning-based framework for automated bone fracture detection from X-ray images. The system follows a structured pipeline beginning with dataset acquisition and preprocessing, followed by feature extraction and classification using advanced architectures, including ResNet-50 and Vision Transformer (ViT-B/16). The integration of convolutional and transformer-based models enables both local feature learning and global contextual understanding. Furthermore, standard evaluation metrics are employed to ensure robustness and reliability of the proposed system. The overall workflow of the methodology is illustrated in Figure 1.

The workflow presented in Figure 1 consists of several sequential stages. First, the Bone Fracture Multi-Region X-ray dataset is collected and divided into training, validation, and testing subsets consisting of 9,246, 828, and 506 images, respectively. Second, the X-ray images are preprocessed by resizing them to 224 × 224 pixels, converting them into tensor representations, and normalizing the pixel intensity values using a mean of 0.5 and a standard deviation of 0.5. Third, data augmentation techniques, including horizontal flipping, zooming, and shearing, are applied to improve data diversity and model generalization. Fourth, the preprocessed training images are used to fine-tune the pretrained ResNet-50 and ViT-B/16 models for binary classification of fractured and non-fractured X-ray images. During training, the validation subset is used to monitor model performance. Finally, the trained models are evaluated on the independent testing subset using accuracy, precision, recall, and F1-score to compare their fracture classification performance.

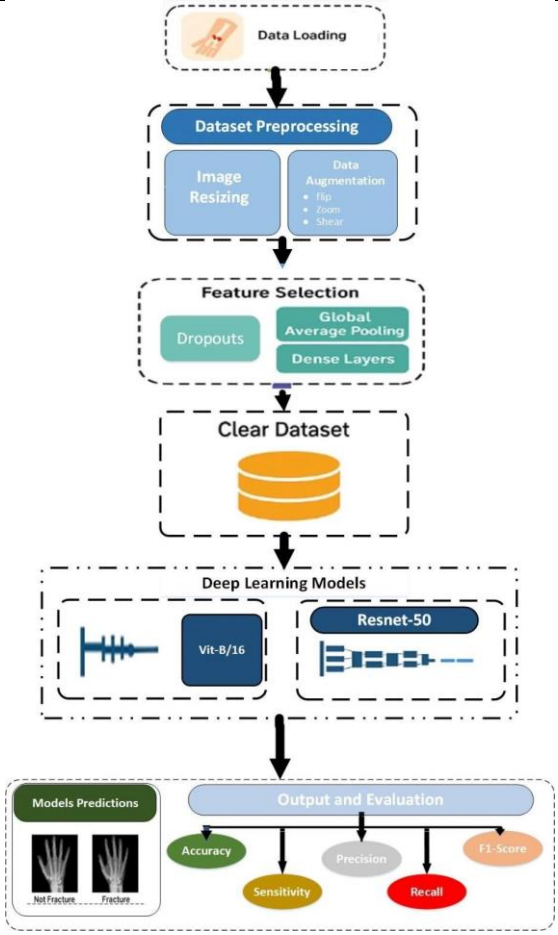

Figure 1. Proposed system framework.

Dataset Collection:

The proposed study utilizes the Bone Fracture Multi-Region X-ray dataset, obtained from the Kaggle repository. The dataset consists of 10,580 grayscale X-ray images with a resolution of 224 × 224 pixels, categorized into fractured and non-fractured classes. The dataset includes multiple anatomical regions such as upper limbs, lower limbs, lumbar spine, hips, and knees, which enhances diversity and improves model generalization. The dataset is divided into training, validation, and testing subsets consisting of 9,246, 828, and 506 images, respectively.

This structured division ensures effective model training and unbiased evaluation. To minimize the risk of data leakage, the training, validation, and testing subsets were maintained separately throughout the experimental procedure. The training subset was used exclusively for model training, the validation subset was used to monitor model performance during training, and the testing subset was kept unseen until the final evaluation of the trained models. Data augmentation was applied during the training process, whereas the validation and testing subsets were used for performance evaluation. The statistical distribution of the dataset is presented in Table 1.

Table 1. Dataset Statistics.

Dataset Split	Number of Images	Description
Training	9,246	Used to train the models
Validation	828	Used to tune the model parameters
Testing	506	Used for final evaluation
Total	10,580	The entire dataset used in this study

Preprocessing:

In deep learning-based medical image analysis, preprocessing is a crucial step to ensure data consistency and improve model performance [34][35][36][37]. In this study, all images are resized to 224 × 224 pixels to match the input requirements of pre-trained models. The images, originally in JPEG or PNG format, are converted into tensor representations for compatibility with the PyTorch framework.

To stabilize the training process and ensure efficient convergence, pixel intensity values are normalized using a mean of 0.5 and a standard deviation of 0.5. This normalization reduces variation in feature distribution and improves gradient-based optimization. To enhance robustness and mitigate overfitting, data augmentation techniques such as horizontal flipping, zooming, and shearing are applied. These transformations increase dataset diversity while preserving the semantic structure of X-ray images.

Feature Extraction and Classification:

Feature extraction and classification are performed using deep learning models that automatically learn hierarchical and discriminative representations from X-ray images [38][39][40][41][42]. These models capture both low-level features, such as edges and textures, and high-level structural patterns relevant to fracture detection. Regularization techniques such as dropout are incorporated to prevent overfitting, while pooling operations reduce spatial dimensions and computational complexity. The extracted features are passed to fully connected layers for final classification into fractured and non-fractured categories.

Model Architecture:

The performance of the proposed system largely depends on the selection of appropriate model architectures. In this study, two state-of-the-art models, Vision Transformer (ViT-B/16) and ResNet-50, are employed due to their strong capability in image classification tasks, particularly in medical imaging.

Vision Transformer (ViT-B/16):

The Vision Transformer (ViT-B/16) model is implemented using the *vit_base_patch16_224* configuration and is pre-trained on ImageNet. In this architecture, the input image is divided into non-overlapping patches of size 16 × 16 pixels. Each patch is flattened and projected into a high-dimensional embedding space. Since patching removes spatial continuity, positional encoding is applied to retain spatial information. The embedded patches are processed through multiple transformer encoder layers consisting of multi-head self-attention (MHSA) and feed-forward neural networks, enabling the model to capture long-range dependencies and global contextual relationships within the image. During fine-tuning, the original classification head is replaced with a binary output layer corresponding to fractured and non-fractured classes. The architecture of the Vision Transformer is illustrated in Figure 2.

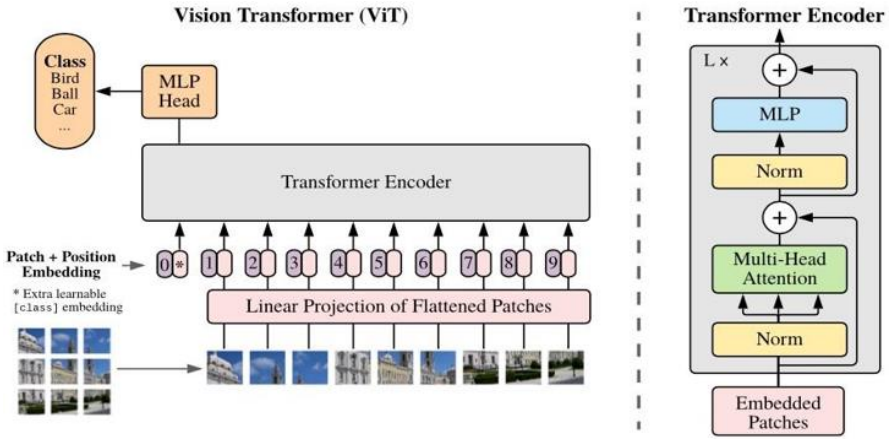

Figure 2. Vision Transformer (ViT-B/16) Architecture

ResNet-50:

ResNet-50 is a deep convolutional neural network consisting of 50 layers and is based on residual learning. The architecture begins with a convolutional layer followed by batch normalization and ReLU activation, along with a max-pooling layer for down-sampling. It includes multiple residual blocks with skip connections that enable efficient gradient flow and address the vanishing gradient problem. These residual connections allow the network to learn identity mappings, making it possible to train very deep architectures effectively. A global average pooling layer is used to reduce feature dimensions before passing them to the final fully connected layer, which is modified for binary classification. ResNet-50 is capable of extracting hierarchical features ranging from simple edges to complex structural patterns, making it highly suitable for fracture detection. The architecture of ResNet-50 is shown in Figure 3.

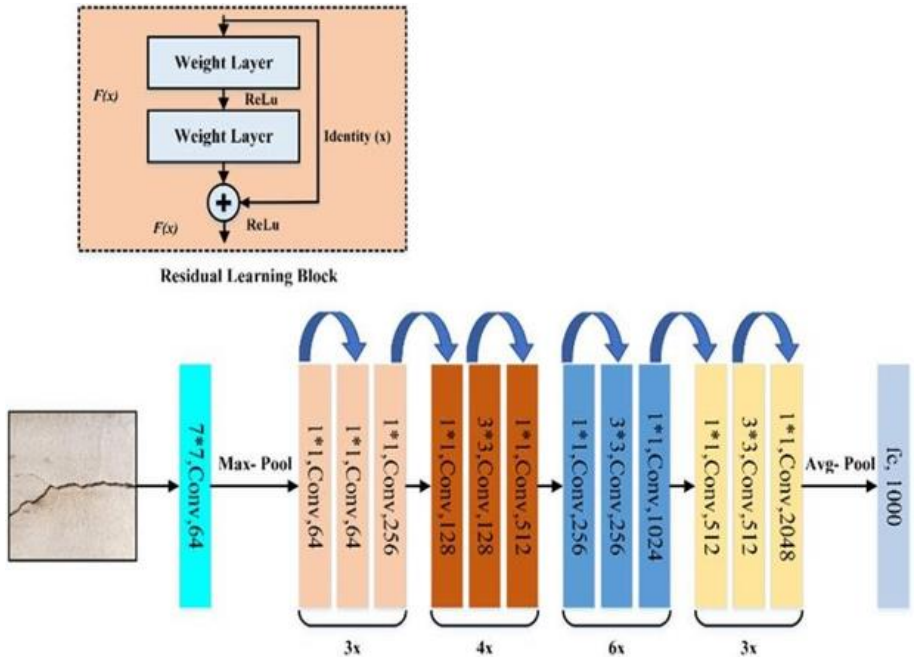

Figure 3. ResNet-50 Architecture Diagram

An in-depth visualization of ResNet-50 architecture is illustrated, including its convolutions initial layer, the following residual blocks with skip dimensions, global average pooling, and a final classification head intended to detect binary fractured. As shown in Figure 3.

Training Configuration:

Both ResNet-50 and ViT-B/16 were fine-tuned using the PyTorch framework with ImageNet-pretrained weights. The models were trained using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 32. Cross-Entropy Loss was used as the loss function for binary fracture classification. The input images were processed at a resolution of 224×224 pixels. Both models were trained for 50 epochs on a CPU, with model performance monitored on the validation set during training. These training parameters were kept consistent for the comparative evaluation of the two architectures.

Performance Evaluation:

The dataset is divided into training, validation, and testing sets as described , and the trained models are evaluated on the testing set to assess their performance. The performance of the proposed system is evaluated in terms of accuracy, precision, recall, and F1-score. The evaluation is based on the confusion matrix, which consists of four possible outcomes obtained after applying the classifier to the input data. True Positive (TP) refers to fractured X-ray images correctly classified as fractured. True Negative (TN) refers to non-fractured images correctly classified as non-fractured. False Positive (FP) refers to non-fractured images

incorrectly classified as fractured. False Negative (FN) refers to fractured images incorrectly classified as non-fractured.

Based on these outcomes, the performance metrics are defined as follows:

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{F1-Score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

These metrics provide a comprehensive evaluation of the classification performance. Precision measures the correctness of positive predictions, while recall evaluates the ability of the model to correctly identify fracture cases. Accuracy provides an overall measure of correct predictions, while F1-score balances precision and recall, particularly in the presence of class imbalance.

Results and Discussion:

This chapter presents the experimental results obtained from training and evaluating two state-of-the-art deep learning models, Vision Transformer (ViT-B/16) and ResNet-50, on the bone fracture X-ray dataset. Both models were trained for 50 epochs and evaluated using standard performance metrics including accuracy, loss, precision, recall, and F1-score. To provide a more comprehensive evaluation, sensitivity and specificity were also considered, while AUC is discussed where applicable. A comparative analysis is conducted to highlight the strengths and limitations of each model for fracture detection.

Vision Transformer (ViT-B/16) Results:

The Vision Transformer (ViT-B/16) demonstrated excellent learning capability, with a consistent increase in accuracy and a corresponding decrease in loss across training epochs. The model exhibited strong generalization performance on validation and test datasets. The highest validation accuracy achieved was 98.91% at epoch 10, while the final test accuracy reached 100%. The lowest validation loss recorded was 0.0292, and the final test loss was 0.0049, indicating effective convergence and stable optimization. The classification report on the test set shows perfect performance, with precision, recall, and F1-score equal to 1.00 for both fractured and non-fractured classes.

The model correctly classified all 506 test images, consisting of 238 fractured and 268 non-fractured samples. The confusion matrix illustrated in Figure 4 confirms that no misclassifications occurred, with all values lying on the diagonal. The VIT -B/16 achieved perfect classification performance on the evaluated test set, correctly classifying all 506 test images. However, this result should not be interpreted as evidence of perfect clinical performance, and further external validation on independent and diverse clinical datasets is required.

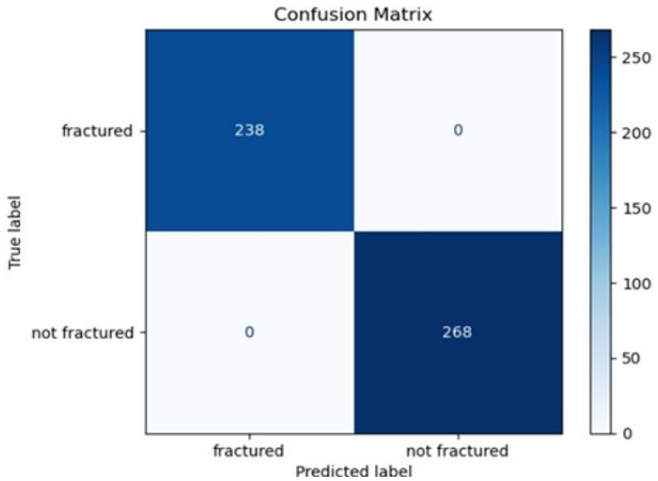

Figure 4. Confusion Matrix – ViT-B/16

The close alignment between the training and validation curves suggests good generalization on the evaluated dataset; however, external validation on independent clinical datasets is required to confirm the model`s generalization capability shown in figure 5. Similarly, the loss curves show a consistent decrease in both training and validation loss, confirming stable optimization throughout the training process, where training loss reduced from 0.4144 to 0.0051 and validation loss decreased to 0.0292. These results highlight the effectiveness of the model in learning meaningful representations from the dataset. However, despite achieving perfect test performance, such results should be interpreted with caution, as real-world clinical scenarios require validation on more diverse and less controlled datasets to ensure robustness and reliability.

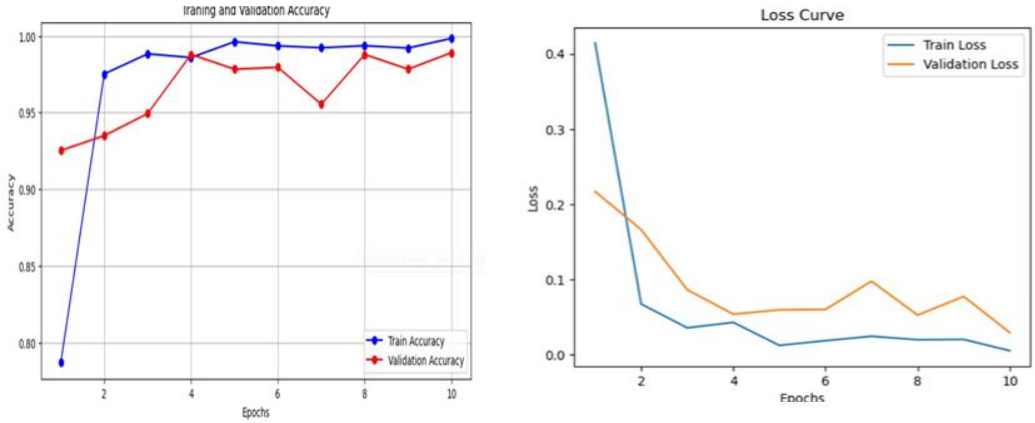

Figure 5. (a) Accuracy Curves(b) Loss Curves – ViT-B/16

ResNet-50 Results:

The ResNet-50 model also demonstrated strong performance, maintaining high accuracy throughout the training and validation phases. The highest validation accuracy reached 100% at epochs 7 and 8, while the final test accuracy was 98.62%. The classification report indicates high precision and recall values for both classes. The fractured class achieved a precision of 1.00 and recall of 0.97, while the non-fractured class achieved a precision of 0.97 and recall of 1.00. The overall F1-score for both classes was 0.99, reflecting balanced performance. The confusion matrix illustrated in Figure 6 shows 231 true positives (TP), 268 true negatives (TN), 0 false positives (FP), and 7 false negatives (FN). Thus, 7 fractured images were incorrectly classified as non-fractured, resulting in a test accuracy of 98.62%. (231 + 268) / 506 = 98.62%. The lower recall for the fractured class (0.97) indicates that ResNet-50 was slightly less sensitive to fracture cases than ViT-B/16.

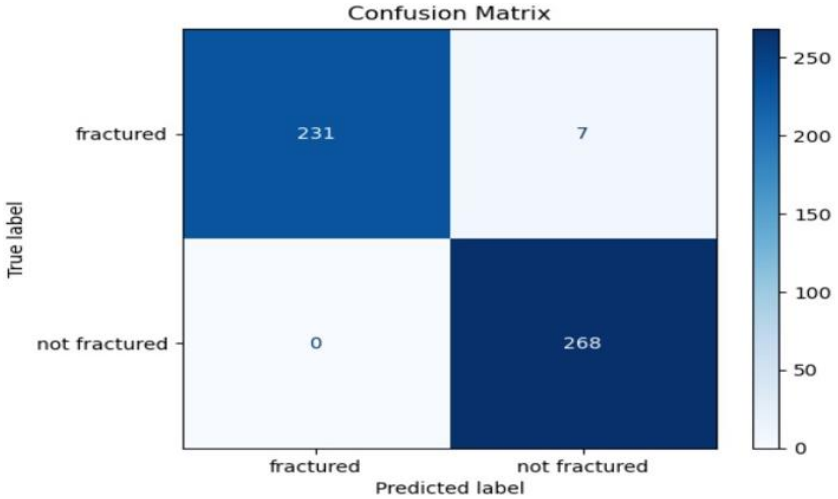

Figure 6. Confusion Matrix – ResNet-50

The training and validation accuracy curves shown in figure 7 demonstrate rapid convergence, with training accuracy increasing steadily and reaching near 100% by the eighth epoch, while validation accuracy also achieved 100%, indicating effective learning of underlying data patterns. The close alignment between the curves suggests strong generalization with minimal overfitting. Similarly, the loss curves show a consistent decrease in both training and validation loss, confirming stable optimization throughout the training process. The training loss reduced significantly during early epochs and stabilized at a low value, while the validation loss closely followed the same trend with minor fluctuations. These results indicate that the ResNet-50 model effectively learned discriminative features for fracture detection. However, slight misclassifications were observed in the fractured class, indicating marginally lower sensitivity compared to the Vision Transformer model.

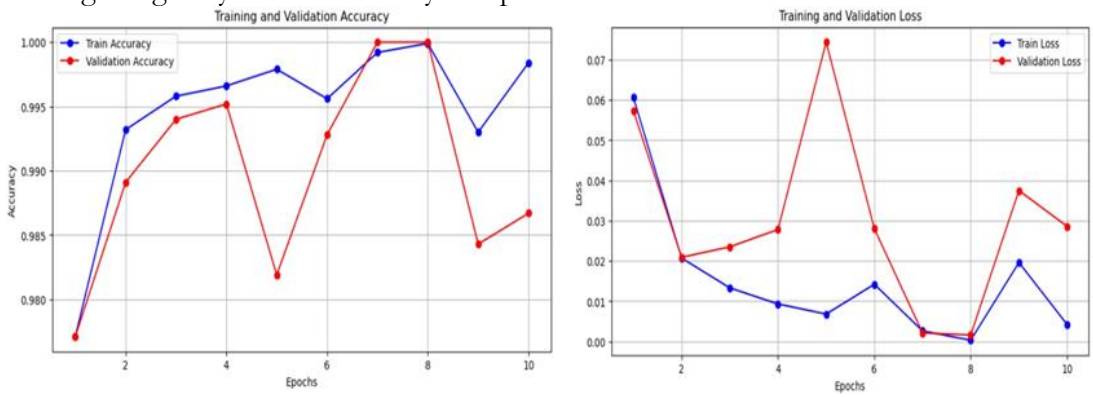

Figure 7. (a) Accuracy Curves(b) Loss Curves – ResNet-50

Comparative Analysis:

The classification results of both models are summarized in Table 2, while Table 3 presents a comparative analysis of their overall performance. In addition to precision, recall, and F1-score, sensitivity and specificity are included to provide a more comprehensive assessment of the models' ability to correctly identify fractured and non-fractured cases.

Table 2. Classification Report for ViT-B/16 and ResNet-50

Model	Class	Precision	Recall	F1-Score	Support
ViT-B/16	Fractured	1.00	1.00	1.00	238
	Not Fractured	1.00	1.00	1.00	268
	Accuracy			1.00	506
	Macro Avg	1.00	1.00	1.00	506
	Weighted Avg	1.00	1.00	1.00	506
ResNet-50	Fractured	1.00	0.97	0.99	238
	Not Fractured	0.97	1.00	0.99	268
	Accuracy			0.99	506
	Macro Avg	0.99	0.99	0.99	506
	Weighted Avg	0.99	0.99	0.99	506

For the binary fracture-detection task, sensitivity is equivalent to recall for the fractured class. Based on the reported classification results, ViT-B/16 achieved 100% sensitivity and 100% specificity. ResNet-50 achieved approximately 97.06% sensitivity and 100% specificity. The sensitivity of ResNet-50 indicates that 231 out of 238 fractured images were correctly identified, while the remaining seven fractured images were misclassified. The specificity of 100% indicates that all 268 non-fractured images were correctly identified by ResNet-50. ViT-B/16 correctly classified all 238 fractured and 268 non-fractured images.

Table 3. Comparative Analysis of ViT-B/16 and ResNet-50

Metric	ViT-B/16	ResNet-50
Test Accuracy	100%	98.62%
Test Loss	0.0049	N/A
Sensitivity (fractured)	100%	97.06%
Specificity (not fractured)	100%	100%
Precision (fractured)	1.00	1.00
Recall (fractured)	1.00	0.97
F1-score (fractured)	1.00	0.99
Precision (not frac.)	1.00	0.97
Recall (not frac.)	1.00	1.00
F1-score (not frac.)	1.00	0.99

The results indicate that ViT-B/16 outperforms ResNet-50 across all evaluation metrics, achieving perfect accuracy, precision, recall, and F1-score. It also achieved 100% sensitivity and specificity on the evaluated test set. This superior performance can be attributed to its self-attention mechanism, which effectively captures global dependencies and subtle fracture patterns in X-ray images. In contrast, ResNet-50 also achieves strong performance with a test accuracy of 98.62% and high F1-scores. Although it exhibits minor misclassifications, particularly in the fractured class, it remains a reliable and efficient model for fracture detection. ResNet-50 provides a practical alternative for fracture detection, particularly where a balance between classification performance and model complexity is required. However, a direct comparison of computational efficiency was not performed in the present study**.** The sensitivity and specificity results further demonstrate that both models performed strongly in distinguishing fractured and non-fractured images. ViT-B/16 achieved 100% sensitivity and specificity, while ResNet-50 achieved 97.06% sensitivity and 100% specificity.

The close agreement between precision, recall, and F1-score also indicates balanced classification performance. However, the unusually high performance, particularly the 100% test accuracy of ViT-B/16, should be interpreted with caution. Such results may partly reflect the characteristics of the dataset and do not necessarily guarantee equivalent performance on unseen clinical data. The possibility of overfitting and dataset-specific learning cannot be completely excluded. The dataset may also have limited variability in patient populations, fracture types, image quality, and imaging conditions, which may affect external generalizability. Therefore, future studies should evaluate the models on larger, more diverse, and independent multi-institutional datasets, with strict patient-level separation to minimize potential data leakage. AUC should also be reported when model probability outputs are available.

Overall, while Vision Transformer provides superior accuracy on the evaluated test dataset**,** ResNet-50 offers a practical balance between performance and computational efficiency. The selection of the model, therefore depends on the specific application requirements and available computational resources. Further external validation is required to confirm the generalizability of the observed performance.

Study Implications, Limitations, and Recommendations:

The findings of this study demonstrate the potential of deep learning models, particularly ViT-B/16 and ResNet-50, for automated bone fracture detection from X-ray images. These models could assist radiologists by providing rapid and consistent fracture screening; however, they should support rather than replace clinical assessment. This study has some limitations, including limited dataset diversity and the possibility of dataset-specific

learning or overfitting. Variations in patient populations, fracture types, image quality, imaging devices, and acquisition protocols may affect model performance in real-world clinical settings.

Therefore, the high performance reported in this study should be interpreted with caution. Future research should focus on larger, more diverse, and multi-institutional datasets and independent external validation to assess model generalizability. Patient-level data separation should also be maintained to minimize potential data leakage. Further work should investigate different fracture types and anatomical regions, explainable AI methods, and prospective clinical evaluation to support reliable clinical deployment.

Conclusion:

This study presented a deep learning-based framework for automated bone fracture detection from radiographic X-ray images using two state-of-the-art architectures, Vision Transformer (ViT-B/16) and ResNet-50. A Bone Fracture Multi-Region X-ray dataset containing 10,580 grayscale images from multiple anatomical regions was used for training and evaluation. Appropriate preprocessing and data augmentation techniques were applied to improve data consistency, model learning, and generalization. The experimental results demonstrated that both models achieved strong performance in binary fracture classification. However, ViT-B/16 achieved the best overall performance, obtaining a validation accuracy of 98.91% and a test accuracy of 100%, with a test loss of 0.0049. It correctly classified all 506 test images and achieved precision, recall, and F1-score of 1.00 for both fractured and non-fractured classes.

In comparison, ResNet-50 achieved a test accuracy of 98.62% and an overall F1-score of 0.99, with seven fractured images incorrectly classified as non-fractured. The superior performance of ViT-B/16 may be attributed to its self-attention mechanism, which enables the model to capture global relationships and contextual patterns across radiographic images. ResNet-50, while slightly less accurate, demonstrated competitive performance and provides a practical alternative for fracture detection. Overall, the findings demonstrate the potential of deep learning, particularly transformer-based architectures, to support automated and reliable bone fracture detection from X-ray images. However, the 100% test accuracy achieved by ViT-B/16 should be interpreted cautiously, as the evaluation was conducted on a single dataset. Further validation using larger, diverse, multi-institutional, and clinically validated datasets is required to assess the robustness and real-world applicability of the proposed approach.

References:

[1] P. G. Su Z, Adam A, Nasrudin MF, Ayob M, "Skeletal Fracture Detection with Deep Learning: A Comprehensive Review. Diagnostics.," *Diagnostics (Basel).*, 2023, doi: 10.3390/diagnostics13203245.

[2] W. Q. Jung J, Dai J, Liu B, "Artificial intelligence in fracture detection with different image modalities and data types: A systematic review and meta-analysis," *PLOS Digit Heal.*, 2024, doi: 10.1371/journal.pdig.0000438.

[3] K. M. Nowroozi A, Salehi MA, Shobeiri P, Agahi S, Momtazmanesh S, Kaviani P, . "Artificial intelligence diagnostic accuracy in fracture detection from plain radiographs and comparing it with clinicians: A systematic review and meta-analysis.," *Clin Radiol*, 2024, doi: 10.1016/j.crad.2024.04.009.

[4] V. E. Tanzi L, Audisio A, Cirrincione G, Aprato A, "Vision transformer for femur fracture classification.," *Injury*, 2022, doi: 10.1016/j.injury.2022.04.013.

[5] A. R. Tahir A, Saadia A, Khan K, Gul A, Qahmash A, "Akram, R.N. Enhancing diagnosis: ensemble deep-learning model for fracture detection using X-ray images," *Clin Radiol.*, 2024, doi: 10.1016/j.crad.2024.08.006.

[6] E. P. Nguyen HH, Le DT, Shore-Lorenti C, Chen C, Schilcher J, Eklund A, Zebaze R, Milat F, Sztal-Mazer S, Girgis CM, Clifton-Bligh R, Cai J, . "AFFnet—a deep convolutional neural

network for the detection of atypical femur fractures from anteroposterior," *Bone.*, 2024, doi: 10.1016/j.bone.2024.117215.

[7] T. H. Hz. A. A. M. A. Aldhyani and Sultan Ahmad., "Diagnosis and detection of bone fracture in radiographic images using deep learning approaches," *Front. Med.*, 2025, DOI:10.3389/fmed.2024.1506686.

[8] Y. P. Ren M, "Deep learning detection of subtle fractures using staged algorithms to mimic radiologist search pattern," *Skelet. Radiol.*, 2021, doi: 10.1007/s00256-021-03739-2.

[9] F. A. Regnard NE, Lanseur B, Ventre J, Ducarouge A, Clovis L, Lassalle L, Lacave E, Grandjean A, Lambert A, Dallaudière B, "Assessment of performances of a deep learning algorithm for the detection of limbs and pelvic fractures, dislocations, focal bone lesions, and elbow effusions on trauma X-rays," *Eur J Radiol.*, 2022, doi: 10.1016/j.ejrad.2022.110447.

[10] V. Hansen, J. Jensen, M. W. Kusk, O. Gerke, H. B. Tromborg, and S. Lysdahlgaard, "Deep learning performance compared to healthcare experts in detecting wrist fractures from radiographs: A systematic review and meta-analysis," *Eur. J. Radiol.*, vol. 174, p. 111399, May 2024, doi: 10.1016/J.EJRAD.2024.111399.

[11] A. A. Rayan JC, Reddy N, Kan JH, Zhang W, "Binomial Classification of Pediatric Elbow Fractures Using a Deep Learning Multiview Approach Emulating Radiologist Decision Making," *Radiol Artif Intell*, 2019, doi: 10.1148/ryai.2019180015.

[12] M. A. Tanzi L, Vezzetti E, Moreno R, Aprato A, Audisio A, "Hierarchical fracture classification of proximal femur X-Ray images using a multistage Deep Learning approach.," *Eur J Radiol*, 2020, doi: 10.1016/j.ejrad.2020.109373.

[13] P. M. Kalmet PHS, Sanduleanu S, Primakov S, Wu G, Jochems A, Refaee T, Ibrahim A, Hulst LV, Lambin P, "Deep learning in fracture detection: a narrative review.," *Acta Orthop.*, 2020, doi: 10.1080/17453674.2019.1711323.

[14] M.-H. Hsieh, C.-Y. Chang, C.-Y. Chang, or hasn't claimed this research yet. This person is not on ResearchGate, and S.-M. Hsu, . "Accurate detection of fresh and old vertebral compression fractures on CT images using ensemble YOLOR," *Multimed. Tools Appl.*, 2024, DOI:10.1007/s11042-024-20355-z.

[15] F. Z. Yang C, Wang J, Xu J, Huang C, Liu F, Sun W, Hong R, Zhang L, Ma D, Li Z, Zhang X, Cai J, "Development and assessment of deep learning system for the location and classification of rib fractures via computed tomography," *Eur J Radiol.*, 2024, doi: 10.1016/j.ejrad.2022.110434.

[16] R. A. G. P. A. Dibo and D. V. Dylov, "Deep learning-based bone pathology localization and classification in wrist X-ray images.," *Anal. Images, Soc. Networks Texts*, 2024, doi: DOI:10.1007/978-3-031-54534-4_14.

[17] N. Vasker, S. N. Haider, M. Hasan, and M. S. Uddin, "Deep Learning-assisted Fracture Diagnosis: Real-time Femur Fracture Diagnosis and Categorization," *2023 4th Int. Conf. Big Data Anal. Pract. IBDAP 2023*, 2023, doi: 10.1109/IBDAP58581.2023.10271972.

[18] L. et al. Badgeley, M.A., Zech, J.R., Oakden-Rayner, "Deep learning predicts hip fracture using confounding patient and healthcare variables," *npj Digit. Med.*, 2019, doi: https://doi.org/10.1038/s41746-019-0105-1.

[19] G. D. Ashkani-Esfahani S, Mojahed Yazdi R, Bhimani R, Kerkhoffs GM, Maas M, DiGiovanni CW, Lubberts B, "Detection of ankle fractures using deep learning algorithms.," *Foot Ankle Surg*, 2022, doi: doi: 10.1016/j.fas.2022.05.005.

[20] D. P. Yadav and S. Rathor., "Bone fracture detection and classification using deep learning approach.," *Int. Conf. Power Electron. IoT Appl. Renew. Energy its Control*, 2020, doi: doi: 10.1109/PARC49193.2020.236611.

[21] B. G. A, J. Y. B, G. Z. A, and X. W. A, "Thigh fracture detection using deep learning method

based on new dilated convolutional feature pyramid network," *Pattern Recognit. Lett.*, 2019, doi: 10.1016/j.patrec.2019.06.015.

[22] W. L. Wang LX, Zhu ZH, Chen QC, Jiang WB, Wang YZ, Sun NK, Hu BS, Rui G, "Development and validation of a deep-learning model for the detection of non-displaced femoral neck fractures with anteroposterior and lateral hip radiographs," *Quant Imaging Med Surg*, 2024, doi: doi: 10.21037/qims-23-814.

[23] S. S. S. Y. P. K. H. Y. P. B. Poojitha, "Bone fracture detection system using CNN algorithm," *2026 Int. Conf. Electron. Comput. Commun. Control Technol.*, 2024, doi: 10.1109/ICCS45141.2019.9065305.

[24] L. J. P. William Gale, Luke Oakden-Rayner, Gustavo Carneiro, Andrew P. Bradley, "Detecting hip fractures with radiologist-level performance using deep neural networks.," *arXiv:1711.06504 [cs.CV]*, 2017, doi: https://doi.org/10.48550/arXiv.1711.06504.

[25] H. C. and H. Z. Z. Su, Y. Zhou, J. Zhou, "BoneCLIP-XGBoost: A Multimodal Approach for Bone Fracture Diagnosis," *IEEE Access*, vol. 12, 2024, doi: doi: 10.1109/ACCESS.2024.3479698.

[26] S. E. Beyaz S, Açıcı K, "Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches.," *Jt Dis Relat Surg*, 2020, doi: doi: 10.5606/ehc.2020.72163.

[27] T. C. Zhou X, Wang H, Feng C, Xu R, He Y, Li L, "Emerging applications of deep learning in bone tumors: current advances and challenges," *Front Oncol.*, 2022, doi: doi: 10.3389/fonc.2022.908873.

[28] K. M. Erne F, Dehncke D, Herath SC, Springer F, Pfeifer N, Eggeling R, "Deep learning in the detection of rare fractures—Development of a 'deep learning convolutional network' model for detecting acetabular fractures," *Z Orthop Unfall.*, 2023, doi: doi: 10.1055/a-1577-4645.

[29] A. F. A. Sayeda Sanzida Ferdous Ruhi, Fokrun Nahar, "A novel approach towards the classification of bone fracture from musculoskeletal radiography images using attention based transfer learning.," *Int. Conf. Comput. Inf. Technol.*, 2024, doi: https://doi.org/10.48550/arXiv.2410.14833.

[30] E. A. Schilcher J, Nilsson A, Andlid O, . "Fusion of electronic health records and radiographic images for a multimodal deep learning prediction model of atypical femur fractures.," *Comput Biol Med.*, 2024, doi: doi: 10.1016/j.compbiomed.2023.107704.

[31] M. ul A. F. D. G. Islam, "An improved technique for bone fracture classification using visual transformers and convolutional neural networks," *SSRN Electron. J.*, 2025, doi: DOI:10.2139/ssrn.5191555.

[32] H. S. Paik S, Park J, Hong JY, "Deep learning application of vertebral compression fracture detection using Mask R-CNN," *Sci Rep.*, 2024, doi: doi: 10.1038/s41598-024-67017-6.

[33] F. N. Fu T, Viswanathan V, Attia A, Zerbib-Attal E, Kosaraju V, Barger R, Vidal J, Bittencourt LK, "Assessing the Potential of a Deep Learning Tool to Improve Fracture Detection by Radiologists and Emergency Physicians on Extremity Radiographs.," *Acad Radiol.*, 2024, doi: doi: 10.1016/j.acra.2023.10.042.

[34] G. Z. Khan and Ullah, Inam., "Efficient technique for monkeypox skin disease classification with clinical data using pre-trained models'," *J. Innov. Image Process.*, vol. 5, no. 2, pp. 192–213, 2023, doi: 10.36548/jiip.2023.2.009.

[35] G. Zaman Khan, I. Ali Shah, Farhatullah, M. A. Hassan, H. Junaid and F. Sardar., "Intelligent Systems for Early Malaria Disease Detection in Patient Cells Using Transfer Learning Approaches," *Int. Conf. Comput. Math. Eng. Technol. (iCoMET), Sukkur, Pakistan,* pp. 1–6, 2023, doi: 10.1109/iCoMET57998.2023.10099260.

[36] G. Zaman Khan *et al.*, "An Efficient Deep Learning Model based Diagnosis System for

Lung Cancer Disease," *2023 4th Int. Conf. Comput. Math. Eng. Technol. Sustain. Technol. Socio-Economic Dev. iCoMET 2023*, 2023, doi: 10.1109/ICOMET57998.2023.10099357.

[37] U. Ahmad, U. Bin Ahmad, G. Z. Khan, S. Shahbaz, and I. Ullah, "Automated Detection and Classification of Plasmodium Parasites in Blood Smears Using the RFCN Deep Learning Model," *2025 4th Int. Conf. Commun. Comput. Digit. Syst. C-CODE 2025*, 2025, doi: 10.1109/C-CODE67372.2025.11204078.

[38] I. A. Khan *et al.*, "A Comprehensive Review on 3D Volumetric CT Liver Segmentation: Techniques, Challenges, Trends, and Future Research Directions," *ICCK Trans. Appl. Intell. Cybern.*, vol. 1, no. 1, pp. 5–35, 2026, doi: 10.62762/TAIC.2025.965486.

[39] Zaman, R.; Shah, I.A.; Ullah, N.; Khan, G.Z., "A Robust Deep Learning-Based Approach for Detection of Breast Cancer from Histopathological Images," *Eng. Proc*, vol. 56, no. 131, 2023, doi: https://doi.org/10.3390/ASEC2023-16598.

[40] S. H. Badshah *et al.*, "An Efficient Classification Techniques for Brain Tumor Using Features Extraction and Statistic Methods, with Machine Learning Algorithms," *Lect. Notes Data Eng. Commun. Technol.*, vol. 158, pp. 757–775, 2023, doi: 10.1007/978-3-031-24475-9_62/SAVE-RESEARCH.

[41] A.-S. Y. G. Z. K. Kindi and S. Ulhaq, "'Improved neural network-based system for early and accurate diagnosis of alzheimer disease', Journal of Computer Science and Technology Studies," *reseach gate*, 2025, doi: doi:10.1007/s13369-02307973-9.

[42] Sajad Ulhaq, G. Z. Khan, Imran Ulhaq, I. Ullah, and F. Rabbi, "Epilepsy Seizures Classification with EEG Signals: A Machine Learning Approach," *J. Comput. Sci. Technol. Stud.*, vol. 5, no. 3, pp. 25–34, Aug. 2023, doi: 10.32996/JCSTS.2023.5.3.3.

Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.