

Conference Paper

A Deep Learning Approach for Hand and Arm Bone Fracture Detection based on ResNet18 and EfficientNetB0 using X-ray images

Ayesha Adnan¹, Ghousia Usman^{1,*}, Usman Ahmad²

¹ Department of Computer Science, Kinnaird College for Women University, Lahore, 54000, Pakistan; ayesha.fitfor@gmail.com, ghousia.usman@kinnaird.edu.pk.

² Department of Computer Science, Lahore College for Women University, Lahore, 54000, Pakistan; usman.ahmad@lcwu.edu.pk.

Corresponding author: Ghousia Usman (e-mail: ghousia.usman@kinnaird.edu.pk).

Submitted: 10-03-2024, **Revised:** 20-06-2024, **Accepted:** 25-06-2024

Abstract

Background: Quick diagnosis and successful treatment depend on the accurate identification of bone fractures. Despite their value, traditional diagnostic techniques like MRI and CT scans are frequently costly, less accessible, and may not be able to identify small fractures. Critical clinical decisions may be delayed due to the time-consuming and human error-prone process of manual X-ray image assessment. Method: This paper offers an automated deep learning-based method for detecting and classifying bone fractures utilizing two advanced convolutional neural network architectures: ResNet18 and EfficientNetB0. The study's dataset includes X-ray images of hands and arms of both normal (4383 images) and fractured (4480 images) patients. To improve image quality and avoid overfitting, preprocessing and data augmentation approaches were applied. Results: With an accuracy of 99%, precision of 99%, and an F1-score of 0.99, EfficientNetB0 outperformed the other evaluated models, demonstrating remarkable classification competencies. Reliability was also demonstrated by the ResNet18 model, which identified fracture patterns with excellent accuracy and robustness of 98%. Conclusion: These findings demonstrate how deep learning may be used to quickly and accurately automate fracture identification. Additionally, even on systems with restricted resources, EfficientNetB0's lightweight design makes it appropriate for real-time clinical applications. This method can help doctors identify fractures early and accurately, decrease errors, and greatly increase diagnostic efficiency.

Keywords: Bone fracture detection; deep learning; efficientnetb0; resnet18; x-ray images.

1. Introduction

With an estimated 2.7 million cases per year in the six EU countries of France, Germany, Italy, Spain, Sweden, and the UK, bone fractures are common these days [1]. During juvenile growth stages, falls and increased physical activity are the main causes of pediatric bone fractures, which are a common medical problem. Boys had almost three times as many fractures as girls, with the left upper extremity being most commonly affected, according to a new study that included 1,020 children. Children between the ages of three and six had the highest incidence, and the main cause was found to be outdoor falls in the spring and summer afternoons [2]. Depending on how much skin remains exposed, bone fractures can be either partial or total, as well as open or closed. Also, diseases like osteoporosis or bone cancer, which weaken bone integrity, can result in fractures [3].

A bone fracture can be identified by sudden pain, obvious misalignment, bruising, swelling, and a warm or red feeling in the affected area [4].

In order to prevent delays in required operations, patients involved in accidents may require immediate medical treatment and close observation. Therapy for these unstable patients might be substantially accelerated by a quick and easy detection of fractures or injuries. It may be crucial to diagnose these injuries as soon as possible in order to ensure their health and prevent further issues [5]. As an immediate and non-invasive approach, X-ray imaging is the main diagnostic tool used to confirm or rule out fractures. Digital X-ray imaging equipment is now widely used in many medical applications due to the combination of radiation and computer image processing technologies [6]. Traditional methods for detecting fractures by manually analyzing X-ray images are time-consuming, error-prone, and tedious [7]. Researchers have evaluated deep learning techniques, particularly CNN models have demonstrated remarkable capabilities in image recognition tasks, including fracture detection in large datasets of labelled X-ray images to detect and classify fractures with high accuracy [8]. In machine learning, the features are not automatically chosen; instead, the user must process the maximum number of features, which is challenging if the input is an image dataset [9].

Across a range of anatomical locations and imaging modalities, deep learning (DL) techniques have demonstrated remarkable accuracy and efficiency in fracture detection, routinely surpassing traditional methods and, in many situations, even skilled radiologists [10–14]. With reported accuracies typically exceeding 90% and sensitivities reaching 85%, a hybrid YOLO NAS–EfficientDet–DETR3 model attained the maximum accuracy of 98.10% [11]. Models with better detection, classification, and localization performance include DenseNet, ResNet, YOLO versions, and various hybrid architectures [10–13,15–22]. Interpretability, computational efficiency, and clinical application have all been further enhanced by advancements like generative models, explainable AI modules, and attention methods [11,13,15,16]. All things considered, DL-based solutions have a great deal of potential for enhancing diagnostic reliability, reducing effort, and facilitating quicker and more accurate fracture detection in clinical practice [10–22]. This study aims to utilize pre-trained Convolutional Neural Network (CNN) models, specifically ResNet18 and EfficientNetB0, to accurately and efficiently detect hand and arm fractures, providing a faster and more reliable alternative to traditional radiological methods.

We organized this paper into various sections as follows. Section 1 provides an introduction to the topic, a brief overview of the existing literature, and highlights the motivation and objectives of this paper. Section 2 provides a detailed discussion of the methods and materials, including the bone fracture dataset, data pre-processing and augmentation, data splitting, and deep learning models for classification. The results and discussion are given in Sections 3 and 4. And Section 5 concludes the paper.

1.1. Motivation

Since manual diagnosis can be difficult, subjective, and susceptible to mistakes, early and accurate detection of bone fractures in the hand and arm is necessary for appropriate treatment. Even advanced imaging techniques like CT and MRI provide accuracy, but they are frequently more costly and less widely available. This motivates the implementation of deep learning models, like ResNet18 and EfficientNetB0, that provide a scalable, accurate, and effective method for automatic fracture detection using X-ray images.

1.2. Objectives

- To build a deep learning-based automated system for identifying bone fractures in the hand and arm using X-ray images.
- To evaluate ResNet18's and EfficientNetB0's performance in terms of F1-score, recall, accuracy, and precision.
- To prove that lightweight CNN architectures are suitable for use in clinical and resource-limited environments.
- To employ computerized medical imaging technologies to help diagnose fractures more quickly, easily, and effectively.

2. Materials and Methods

Each stage of the implemented methodology is outlined in this section. Figure 1 depicts the entire procedure of the chosen methodology.

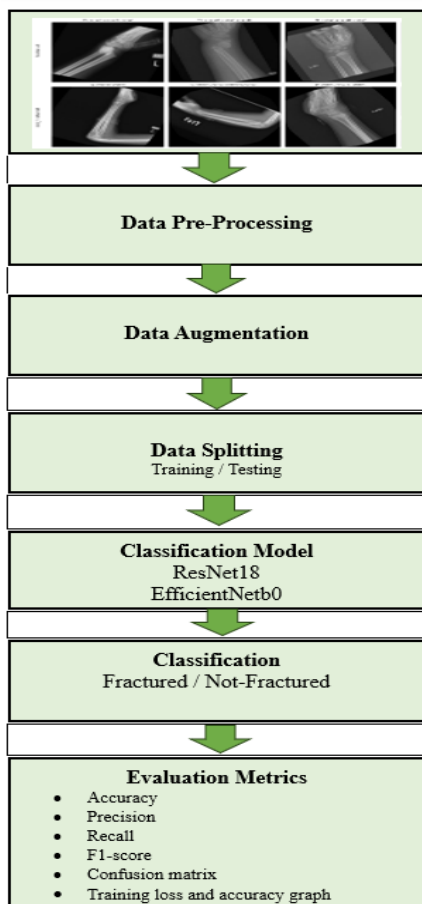


Figure 1. Methodology diagram

2.1. Dataset

The research study's dataset, which included 8,860 radiographic (X-ray) images of human hands and arms, was downloaded from Kaggle. It is divided into two classes: 4,480 images of fractured bones that illustrate different kinds, positions, and severity of fractures, and 4,383 pictures of bones that have not been fractured and exhibit typical anatomical features. It is appropriate for training binary classification models with not much class imbalance constraints due to the almost balanced class distribution. Table 1 shows the number of images in each class along with a description. Figure 2 shows the original images from the dataset.

Table 1. Description of classes in dataset

Class	Number of Images	Description
Fractured	4,480	X-ray images showing various types of bone fractures
Not-Fractured	4,383	X-ray images showing normal bones
Total	8,860	Combined total of both classes



Figure 2. Original images from dataset

2.2. Data Pre-Processing and Augmentation

A number of preprocessing techniques were applied to the X-ray dataset used in this study in order to enhance image quality and prepare it for training deep learning models. In order to reduce computing complexity while maintaining crucial structural elements, all images were first transformed to greyscale. For precise fracture line identification, bilateral filtering was used to cut down on noise without obscuring critical edges. Contrast enhancement came next, which helped draw attention to bone structures and accentuate small fractures. To guarantee uniformity throughout the dataset and compatibility with CNN models that had already been trained, images were then rescaled to a standard size. Data augmentation methods including horizontal flipping, vertical flipping, and random affine transformations were employed to enhance model generalisation and avoid overfitting. Figure 3 and Figure 4 shows the images after applying pre-processing and augmentation techniques on original images.

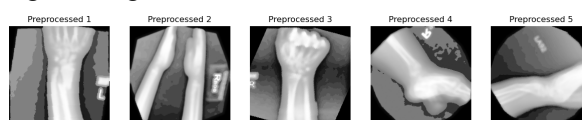


Figure 3. Pre-processed images



Figure 4. Augmented images

2.3. Data Splitting

To provide a sufficient quantity of data for model learning and to allow for performance evaluation, the dataset was split into 70% training and 30% testing. The base structure consisted of pre-trained ResNet and EfficientNet architectures, which enhanced model generalisation and reduced training time. Using an appropriate method, the model was adjusted on the dataset over ten epochs. Using

important performance metrics such as accuracy, precision, recall, F1-score, confusion matrix, and training loss and accuracy graphs, on test set evaluates the model's performance.

2.4. Classification Models

Two pre-trained Convolutional Neural Network (CNN) models, ResNet18 and EfficientNetB0, were used in this study to identify fractures in the hands and arms from X-ray images. These models were chosen due to their apparent effectiveness in image classification challenges and their capacity to extract complex patterns from visual input.

2.4.1. Resnet18

ResNet18 is a deep learning model that belongs to the Residual Network (ResNet) family. It has recently been demonstrated that Deep Residual Networks significantly improve neural network performance [23]. It is well-known for using "skip connections" or residual connections, which aid in deep network training by solving the problem of vanishing gradients. Because of its 18 layers and modest weight, it is effective without compromising performance. The ability of ResNet18 to extract hierarchical features from images is especially useful for spotting minute fracture patterns.

2.4.2. Efficientnetb0

The foundational model of the EfficientNet family, EfficientNetB0, is made to attain excellent accuracy with fewer parameters and less computational expense. It is shown that an EfficientNet model can outperform the most advanced accuracy [24]. It is strong and effective because it balances the model's depth, width, and input resolution using a compound scaling technique. EfficientNetB0 is renowned for its capacity to outperform numerous deeper models while using fewer resources.

The study makes use of transfer learning, which enables the models to retain valuable feature extraction capabilities and adapt to the medical imaging job with fewer training resources, by utilising these models. This method allows for precise fracture identification even with limited medical data and increases training efficiency.

3. Results

The next section describes the results of using deep learning models for bone fracture detection.

3.1. Resnet18:

A comprehensive evaluation of the ResNet18 model's performance can be accessed via the confusion matrix in Figure 5. The model made 2,659 predictions in total, based on the matrix. It exhibited a high degree of accuracy in both classes by properly classifying 1,286 normal cases as normal (true negatives) and 1,323 fractured cases as fractured (true positives). Only 26 fractured cases were mistakenly identified as normal (false negatives), whereas only 24 normal cases were mistakenly labelled as fractures (false positives). Particularly in a field with high stakes like medical imaging, these misclassifications are in fact rare. Key performance measures can be calculated based on these values.

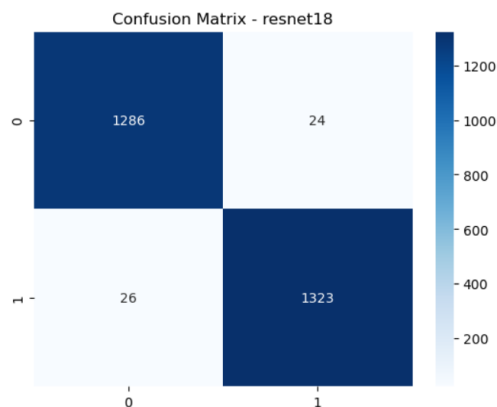


Figure 5. Confusion matrix for ResNet18

The model's overall reliability is represented by its accuracy, which is approximately 98 %. The precision, which measures the percentage of actual fractures that were predicted, is approximately 98 %. About 98 % of real fractures were correctly identified, as indicated by the recall or sensitivity. The F1 score, which maintains a balance between recall and precision, is 98 %. These high scores indicate how well the algorithm recognizes fractures, minimizing errors and missed cases. Table 2 displays the results of evaluation matrix of each class and overall accuracy for ResNet18.

Table 2. Results of evaluation matrices for ResNet18

Class	Precision	Recall	F1 score	Accuracy
0	0.98	0.98	0.98	0.98
1	0.98	0.98	0.98	0.98

The ResNet18 model's training performance across ten epochs for bone fracture detection is depicted in the line graph (Figure 6). Training accuracy (in orange) and training loss (in blue) are the two main measures displayed. The model's initial prediction mistakes are indicated by the relatively substantial training loss, which then steadily and sharply decreases with each epoch until it approaches zero. This means that during training, the model is efficiently learning and reducing its mistakes. At the same time, the training accuracy increases quickly from about 75% to over 98%, maintaining at near-perfect values in subsequent epochs. This fast increase and subsequent stability show that the model picks up accurate bone X-ray image classification.

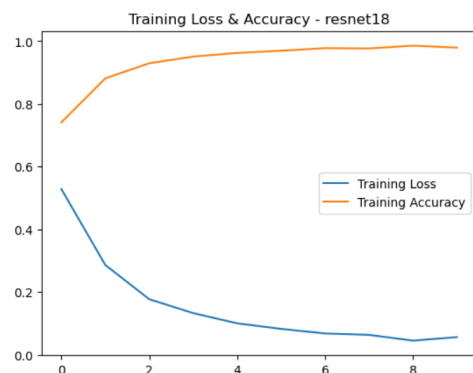


Figure 6. Training Loss and Accuracy graph for ResNet18

3.2. Efficientnetb0:

The EfficientNetB0 model performs outstandingly in a binary classification evaluation, according to the confusion matrix (Figure 7) distinguish between the two classes—normal cases (class 0) and fractured cases (class 1). The model demonstrated a significant capacity to recognise both classes, properly classifying 1,268 cases of class 0 and 1,356 occurrences of class 1. The 35 misclassifications were all false positives, meaning that class 0 instances were mistakenly projected to be class 1. Crucially, the model correctly predicted every class 1 case without missing any, indicating that there were no false negatives.

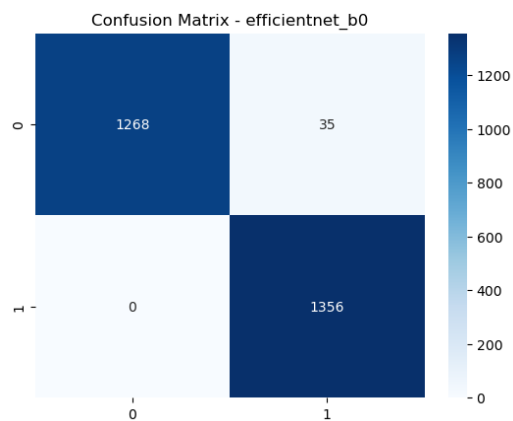


Figure 7. Confusion matrix for EfficientNetB0

On the test dataset, the EfficientNetB0 model demonstrates outstanding classification performance, with an overall accuracy of 99%. For class 0, it obtained a recall of 0.97, which means it correctly identified 97% of actual class 0 samples, and a precision of 1.00, which means all projected cases of class 0 were correct. The model captured all real class 1 instances, achieving a precision of 0.97 and a flawless recall of 1.00. With an F1-score of 0.99, both classes demonstrated a superb balance between recall and precision. All metrics had weighted and macro averages of 0.99, indicating consistent and very accurate model performance throughout the dataset. Results of evaluation matrices of each class and overall accuracy for EfficientNetB0 are shown in Table 3.

Table 3. Results of evaluation matrices of EfficientNetB0

Class	Precision	Recall	F1 score	Accuracy
0	1.00	0.97	0.99	
1	0.97	1.00	0.99	
				0.99

Over ten epochs, the EfficientNetB0 training plot displays a steep learning curve. The model rapidly learnt to minimise errors, as evidenced by the training loss (blue line) settling after a steep decline in the first few epochs, from roughly 0.48 to almost 0.01. Meanwhile, the training accuracy (orange line) increases quickly from about 85% to almost 100%, where it reaches an early plateau and stays steady. Figure 8 Training Loss and Accuracy graph for EfficientNetb0.

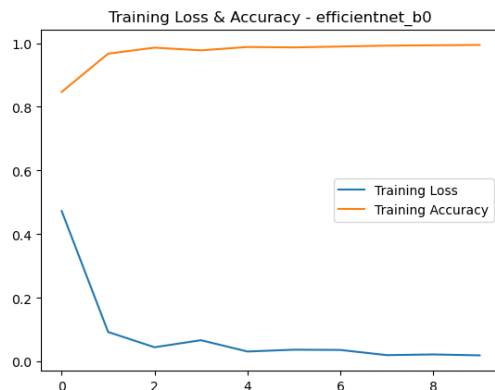


Figure 8. Training Loss and Accuracy graph for EfficientNetB0

This shows that the training data was successfully learnt by the model with almost perfect accuracy and little loss. All things taken into account, the plot shows a well-trained model with quick convergence and consistent performance over training. Strong accuracy for class 0 is also indicated by the low false positive rate. In conclusion, the EfficientNetB0 model has a high-performing and well-balanced classification ability, especially in preventing false negatives while maintaining a high level of overall predictive accuracy.

4. Discussion

Both ResNet18 and EfficientNetB0 were used for this research to identify bone fractures from hand and arm X-ray images, and each model was evaluated using common performance metrics. With a superior accuracy of 99%, precision of 0.99, recall of 0.99, and an F1-score of 0.99, EfficientNetB0 showed exceptional performance. It successfully recognized every fractured case with no false negatives. ResNet18, on the other hand, also performed exceptionally well, displaying stable outcomes during several evaluation runs and helpful classification accuracy. It showed strong generalization and adaptability, making it an effective alternative for fracture detection tasks even though it did not outperform EfficientNet-B0 in precision or recall. Table 4 shows the results of both models.

Table 4. Results of both models

Model Name	Precision	Recall	F1 score	Accuracy
RsNet18	0.98	0.98	0.98	0.98
EfficientNetB0	0.99	0.99	0.99	0.99

Deep learning (DL) algorithms have persistently shown outstanding efficiency and accuracy in fracture detection, frequently exceeding both conventional diagnostic methods and, in certain situations, even experienced radiologists. Accuracy reports at times exceed 90%, with the highest recorded accuracies being 98.10% from a hybrid YOLO NAS–EfficientDet–DETR3 model, discussed in literature. Although models can automatically classify fractures, EfficientNetB0 outperformed that benchmark in this study, correctly diagnosing every fracture with no false negatives and achieving 99% accuracy along with excellent precision, recall, and F1-score. This makes it particularly helpful in medical centres where fractures that go undetected could have adverse consequences. Its light-weight design makes it even more appropriate for deployment in resource-constrained scenarios and

real-time applications. Although ResNet18's precision and recall were slightly lower than those of EfficientNetB0, it showed robust, consistent performance with good generalization. These results promote the real-world application of current DL architectures in clinical workflows by proving their ability to combine impressive diagnostic accuracy with computational efficiency.

5. Conclusion

This paper uses two modern CNN architectures, ResNet18 and EfficientNetB0, to demonstrate a deep learning-based method for identifying bone fractures in X-ray pictures of the hands and arms. With 99% accuracy, a precision of 0.99, and an F1-score of 0.99, the EfficientNetB0 model produced outstanding results, whereas ResNet18 also showed great accuracy and dependable performance with 98%. These findings promote early and precise diagnosis by confirming the models' ability to differentiate between normal and fractured cases. Even if the models operate well, future research will concentrate on growing the dataset to cover a wider range of patient profiles and fracture variants. Additionally, clinical validation will be carried out to guarantee that the models can be used practically in medical settings

Abbreviations

The following abbreviations are used in this manuscript:

ADHBF	Automatic Detection of Human Bone Fractures
AI	Artificial Intelligence
AUC	Area Under the Curve
CNN	Convolutional Neural Network
CT	Computed Tomography
DL	Deep Learning
EU	European Union
mAP	Mean Average Precision
MRI	Magnetic Resonance Imaging
MURA	Musculoskeletal Radiographs (Public Dataset)
NAS	Neural Architecture Search
ReLU	Rectified Linear Unit
ResNet	Residual Network
SVM	Support Vector Machine
VGG	Visual Geometry Group
XAI	Explainable Artificial Intelligence
YOLO	You Only Look Once

References

1. L. Tanzi, E. Vezzetti, R. Moreno, A. Aprato, A. Audisio, and A. Massè, "Hierarchical fracture classification of proximal femur X-Ray images using a multistage Deep Learning approach," *Eur. J. Radiol.*, vol. 133, p. 109373, 2020.
2. U. Aygun, "The feature assessment of the bone fractures in 1020 children and review of the literature," *North. Clin. Istanbul.*, vol. 7, no. 5, 2020.
3. A. Ali et al., "DPEBic: detecting essential proteins in gene expressions using encoding and biclustering algorithm," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–8, 2021.
4. S. Thota, P. Kandukuru, M. Sundaram, A. Ali, S. M. Basha, and N. H. Bindu, "Deep Learning based Bone Fracture Detection," in *Proc. 2024 Int. Conf. Smart Syst. Appl. Electr. Sci. (ICSSSES)*, May 2024, pp. 1–7.
5. M. Yaseen, M. Ali, S. Ali, A. Hussain, M. I. Joo, and H. C. Kim, "Cervical spine fracture detection and classification using two-stage deep learning methodology," *IEEE Access*, vol. 12, pp. 72131–72142, 2024.
6. F. Hardalaç et al., "Fracture detection in wrist X-ray images using deep learning-based object detection models," *Sensors*, vol. 22, no. 3, p. 1285, 2022.
7. A. Saif, C. Shahnaz, W.-P. Zhu, and M. O. Ahmad, "Abnormality detection in musculoskeletal radiographs using capsule network," *IEEE Access*, vol. 7, pp. 81494–81503, 2019.
8. B. Guan et al., "Arm fracture detection in X-rays based on improved deep convolutional neural network," *Comput. Electr. Eng.*, vol. 81, p. 106530, 2020.

9. M. Nagaraj, V. Prabhakar, and S. Thota, "Classification of mammograms using attention learning for localization of malignancy," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 5, pp. 84–90, May 2019.
10. R. Kumar, S. D. K, and D. P. Mohapatra, "Assessing radiographic findings on finger X-rays using an enhanced deep learning approach," *Int. J. Inf. Technol.*, vol. 16, no. 7, pp. 4279–4288, 2024.
11. S. C. Medaramatla, C. V. Samhitha, S. D. Pande, and S. R. Vinta, "Detection of hand bone fractures in X-ray images using hybrid YOLO NAS," *IEEE Access*, vol. 12, pp. 57661–57673, 2024.
12. V. Kant, "Improving Clinical Outcomes in Fracture Detection with Hybrid CNN Models and Data Augmentation," in *Proc. 2024 2nd Int. Conf. Intell. Cyber Phys. Syst. Internet Things (ICoICI)*, Aug. 2024, pp. 782–787.
13. T. Islam et al., "A novel explainable deep generative model-aided transfer learning CNN for pelvis fracture detection," *Biomed. Signal Process. Control*, vol. 110, p. 107987, 2025.
14. S. Den Hengst et al., "Detection, Classification, and Segmentation of Rib Fractures From CT Data Using Deep Learning Models: A Review of Literature and Pooled Analysis," *J. Thorac. Imaging*, early access, pp. 1–12. [Online]. Available: <https://journals.lww.com/thoracicimaging>
15. S. Wang, J. Fei, Y. Liu, Y. Huang, and L. Li, "Study on the application of deep learning artificial intelligence techniques in the diagnosis of nasal bone fracture," *Int. J. Burns Trauma*, vol. 14, no. 6, p. 125, 2024.
16. P. P. Kale, U. B. Shinde, D. L. Bhuyar, K. T. V. Reddy, and H. B. Mahajan, "Human Body Bone Fracture Identification using Improved Deep Learning Model," in *Proc. 2024 2nd DMIHER Int. Conf. Artif. Intell. Healthc., Educ., Ind. (IDICAIEI)*, Nov. 2024, pp. 1–5.
17. M. R. Mishra, B. R. Parida, T. Samal, L. Parida, and M. R. Kabat, "Binary Classification of Bone Fractures using Transfer Learning: A Comprehensive Approach," in *Proc. 2024 Int. Conf. Intell. Comput. Sustain. Innov. Technol. (IC-SIT)*, Nov. 2024, pp. 1–6.
18. M. Axenus et al., "Automated diagnosis and classification of metacarpal and phalangeal fractures using a convolutional neural network: a retrospective data analysis study," *Acta Orthop.*, vol. 96, p. 13, 2025.
19. J. Chen et al., "Deep learning model for automated detection of fresh and old vertebral fractures on thoracolumbar CT," *Eur. Spine J.*, vol. 34, no. 3, pp. 1177–1186, 2025.
20. M. Yaseen, M. Ali, S. Ali, A. Hussain, M. I. Joo, and H. C. Kim, "Cervical spine fracture detection and classification using two-stage deep learning methodology," *IEEE Access*, vol. 12, pp. 72131–72142, 2024.
21. X. S. Liu, R. Nie, A. W. Duan, L. Yang, X. Li, L. T. Zhang, et al., "YOLOX-SwinT algorithm improves the accuracy of AO/OTA classification of intertrochanteric fractures by orthopedic trauma surgeons," *Chin. J. Traumatol.*, vol. 28, no. 1, pp. 69–75, 2025.
22. 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. 169, p. 111399, 2024.
23. M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," *Appl. Sci.*, vol. 12, no. 18, p. 8972, 2022.
24. M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, May 2019, pp. 6105–6114.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of PAAS and/or the editor(s). PAAS and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.