Bone Fracture Classification using Convolutional Neural Networks from X-ray Images

Amal Alshahrani
College of Computing, Department of Computer Science and Artificial Intelligence, Umm Al-Qura University, Saudi Arabia
amshahrani@uqu.edu.sa (corresponding author)

Alaa Alsairafi
College of Computing, Department of Computer Science and Artificial Intelligence, Umm Al-Qura University, Saudi Arabia
s442002182@uqu.edu.sa

Received: 6 June 2024 | Revised: 23 July 2024 and 2 August 2024 | Accepted: 4 August 2024

 DOI: https://doi.org/10.48084/etasr.8050

ABSTRACT

This study investigates a bone fracture classification system using deep learning algorithms to determine the best-performing architecture. The primary focus was on training the YOLOv8 model, renowned for its real-time object detection and image segmentation capabilities, as well as the VGG16 model. CNN architectures, known for their effectiveness in image recognition tasks, were chosen for their proven effectiveness in detecting bone fractures from X-ray images. Hyperparameter tuning was used to improve the system's ability to accurately detect and classify bone fractures. The FracAtlas dataset was utilized, which contains 4,083 X-ray images of fractured and non-fractured human bones. Integrating advanced deep learning techniques aims to assist surgeons with more accurate diagnostics. The performance of the developed system was evaluated against existing methods, showcasing its effectiveness in medical diagnostics and fracture treatment. The methodology employed, including data augmentation, extensive model training, and hyperparameter tuning, significantly improved the accuracy of bone fracture detection and classification, demonstrating the potential of deep learning models in aiding medical professionals with more precise and efficient diagnostics.

Keywords-bone fracture; classification; deep learning; VGG16; YOLOV8; CNN

I. INTRODUCTION

Bone fractures are a common and significant medical issue that often requires a precise and timely diagnosis for effective treatment [1]. Traditionally, fractures are detected and classified by radiologists who analyze X-ray images to identify their presence and type. However, this process can be time-consuming and prone to human error, especially with the increasing volume of medical imaging data. The advent of Deep Learning (DL), a subset of artificial intelligence, has revolutionized the field of medical imaging by offering powerful tools for automated image analysis. DL algorithms, particularly Convolutional Neural Networks (CNNs), have demonstrated exceptional performance in various image classification tasks, making them well-suited for medical applications such as fracture classification [2].

This study focuses on classifying bone fractures from X-ray images of the hand, hip, and shoulder using two state-of-the-art DL techniques: YOLOv8 and VGG16. YOLOv8 is a cutting-edge object detection model known for its real-time detection capabilities and high accuracy [3]. VGG16 is a renowned CNN model known for its depth and simplicity, consisting of 16 layers that can extract intricate features from images [4]. This architecture is particularly effective in recognizing patterns and textures, which are crucial for distinguishing between fractured and non-fractured bones. This study examines the capabilities of YOLOv8 and VGG16 to develop a robust system, capable of accurately classifying bone fractures in X-ray images.

In the realm of medical imaging diagnostics, recent years have witnessed remarkable progress in fracture detection and classification through the application of DL algorithms. In [5], CNNs were used for ankle fracture detection, exploring the effectiveness of training CNN models from scratch with a small dataset. This study collected 298 radiographs of non-fractured bones and 298 radiographs of fractured ankle cases and created single- and multitier models to evaluate the impact of multiple views. The results showed that the ensemble of all five models achieved the best accuracy of 76% when using single radiographic views. Using all three views for a single case, the ensemble of all models resulted in the best metrics, with an accuracy of 81%. In [6], the challenge of enhancing the efficiency and accuracy of bone fracture diagnoses was addressed through an automated system that employed DL techniques. A Deep Neural Network (DNN) was developed to classify bones as fractured or healthy. The model, developed using Python and Keras, incorporated an architecture that included convolution, pooling, flattening, and dense layers, specifically designed for binary classification tasks. This model achieved a classification accuracy of 92.44% through a five-fold cross-validation, surpassing the performance metrics of previous models in the field. Further improvements were suggested, such as the incorporation of additional features or techniques to enhance the accuracy and efficiency of bone fracture detection and classification.

In [7], a novel diagnostic algorithm was proposed for the automatic detection of osteoporosis in mandible Cone-Beam Computed Tomography (CBCT) images. This study utilized 120 CBCT images of women aged 50-85, classified as normal and osteoporotic based on the T-score derived from Dual-Energy X-ray Absorptiometry (DEXA). The algorithm incorporated image processing, feature extraction, and classification using an Artificial Neural Network (ANN) to distinguish healthy and osteoporotic mandibles. This method demonstrated the potential to efficiently predict osteoporosis, reducing the need for additional examinations and aligning with the wide acceptance of CBCT in dentistry. In [8], a diagnostic model for COVID-19 was proposed using advanced DL methods. This study employed wavelet analysis and Mel Frequency Cepstral Coefficients (MFCC) for feature extraction, along with Support Vector Machines (SVM) for classification. The dataset comprised 2400 chest X-ray images, split equally between normal and COVID-19 cases, sourced from public datasets to validate the effectiveness of the model. This approach offered a promising supplement to traditional COVID-19 diagnostic methods, demonstrating high accuracy in rapid diagnosis of the disease.

In [9], a machine learning method was developed to identify and classify hip fractures, and its performance was compared to that of experienced human observers. A dataset of 3659 hip radiographs was used, which were classified by expert clinicians. The results showed that the proposed method achieved an overall accuracy of 92%, surpassing the accuracy of human experts by 19%. This study highlighted the potential of machine learning to improve fracture classification and its impact on patient outcomes and treatment costs. In [10], 20 different fracture detection procedures were compared on the Gazi University Hospital's dataset of wrist X-ray images. The results were examined from different perspectives, and six different ensemble models were developed to improve the detection results. The dynamic R-CNN model achieved the highest results, with an accuracy of 77.7%.

In [11], an innovative method was proposed for detecting and classifying bone fractures using advanced DL techniques. Pre-trained deep neural networks, namely ResNeXt101, InceptionResNetV2, Xception, and NASNetLarge, were used to analyze X-ray images. The impetus behind this approach stems from the critical need in emergency medical settings for accurate and rapid fracture diagnosis, an area where traditional methods often fail. A diverse array of DL models was used, leading to notable improvements in the accuracy and efficiency of fracture classification. The dataset was composed of various X-ray images, meticulously selected to represent a wide range of bone fractures, such as simple, complex, and compound fractures. The InceptionResNetV2 model achieved an impressive accuracy of 94.58%. In [12], a method for detecting fractures in X-ray images was presented, with a particular focus on long-bone fractures. This approach involved modifying the Faster R-CNN DL algorithm and integrating a significant advancement through a rotated bounding box to accurately identify fracture locations. The impetus behind this innovation was to enhance fracture detection accuracy, addressing the diverse types and locations of long-bone fractures. Complex mathematical techniques, such as the Rotated Discrete Curvature System (RDS) and the shape directory, were employed to improve the precision of identifying fractures. The modified model achieved a high accuracy rate of 96.1%, demonstrating its efficacy in improving fracture detection and laying the groundwork for future advancements in the area.

In [13], a model was proposed to address the significant challenge of hip fractures in the elderly. Recognizing the limitations of existing fracture registries, which often rely on inaccurate billing and procedural codes, a deep learning-based solution was proposed that analyzed 18,834 conventional radiographs from 2,919 patients. This model was an ensemble of deep learning architectures, including ResNet, VGG, DenseNet, and EfficientNet, designed to improve hip fracture detection accuracy. This model achieved accuracy between 92% and 100% across various submodules, significantly reducing the time required for image annotation compared to traditional methods. In [14], data augmentation was used to improve the performance of the YOLOv8 algorithm in detecting pediatric wrist traumas. 20,327 X-ray images of pediatric wrist injuries made up the training dataset, and after augmenting the data, they became 28408 images. Compared to the Adam optimizer, the model's accuracy was higher using the SGD optimizer. Future plans called for expanding the application's usage to novice pediatric surgeons in developing nations and launching it on several platforms.

In [15], an image processing system was developed, using a CNN architecture to diagnose bone fractures through automated analysis of X-ray and CT images. The results highlight the potential of CNN-based approaches for automated bone fracture detection, offering a valuable solution in the field of medical imaging diagnostics. In [16], a comprehensive investigation was presented on the feasibility of utilizing a DL-based decision support system to address the diagnostic challenges associated with musculoskeletal fractures and enhance the detection of fractures in radiographs. The proposed method involved training a DL model using annotated musculoskeletal X-rays, specifically employing the YOLO architecture, and testing its performance on two datasets. The results showed a sensitivity (*Se*) of 0.910 (95% CI: 0.852-0.946) and a specificity (*Sp*) of 0.557 (95% CI: 0.520-0.594), indicating the model's ability to correctly detect fractures. These findings underscore the potential of DL models to improve fracture detection in radiographs. In [17], a sophisticated approach was presented for medical X-ray image classification, called Multi-Versus Optimizer with DL (MVODL-RMXIC). This method integrated Cross Bilateral Filtering (CBF) for effective noise removal, enhancing image quality. Feature extraction was performed using the MixNet architecture, which utilizes mixed depthwise convolutional kernels to capture a broad range of features from images. The Multi-Versus Optimizer (MVO) algorithm was then employed to optimize hyperparameters, improving model performance through advanced optimization techniques inspired by cosmological phenomena. This combination of noise reduction, feature extraction, and optimization significantly enhanced the accuracy and robustness of the image classification framework.

TABLE I. LITERATURE REVIEW

Study	Year	Dataset	Techniques	Accuracy
[1]	2019	596 radiographs of ankles	CNN	81%
[2]	2020	100 X-ray images of different types of human bones	(DNN)	92.44%
[22]	2020	120 CBCT images	ANN	-
[21]	2021	2400 chest X-ray images	SVM	-
[3]	2022	- 429 radiographs of non-fractured bones -2,364 radiographs of fractured bones	CNN	92%
[4]	2022	Gazi University Hospital's dataset + 542 images	Dynamic R-CNN	77.7%
[5]	2023	Range of bone fractures from X-ray images	DL	94.58%
[6]	2023	200 X-ray images of long bone fractures	Faster R-CNN	96.1%
[7]	2023	18,834 conventional radiographs	DL architectures, including ResNet, VGG, DenseNet, and EfficientNet	92%
[8]	2023	20327 X-ray images	YOLOv8	73.4%
[9]	2023	100 cracked and 100 normal bone images	CNN	99.5%
[20]	2023	Medical X-ray images	MVODL-RMXIC	-
[10]	2024	Combination of the MURA and the FracAtlas dataset	YOLOv7	99.5%

Some studies used limited datasets and implemented various algorithms to improve accuracy. These limitations emphasize the importance of having a large dataset and exploring different algorithmic approaches to achieve better accuracy when classifying and identifying fractures. This study aims to assist radiologists by providing reliable second opinions to reduce diagnostic errors and enhance the efficiency of the diagnostic process. By improving the accuracy and speed of fracture classification, the proposed model aims to facilitate timely and appropriate medical intervention, improving patient outcomes.

II. METHODOLOGY

Figure 1 shows the methodology steps, offering a clear and structured overview of the entire process.

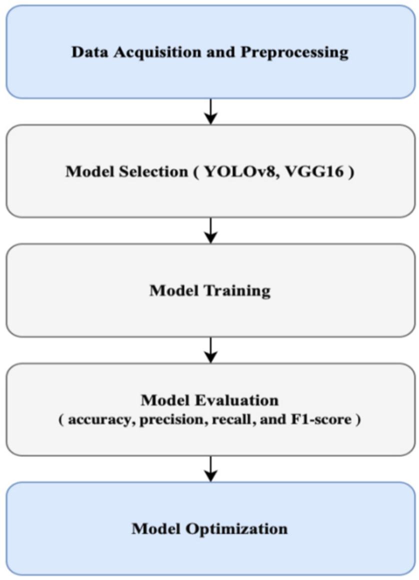

Fig. 1. Methodology.

A. *Data Acquisition and Preprocessing*

This study used the FracAtlas dataset [18], which consists of 4,083 X-ray images showing various types of bone fractures, including hand, leg, shoulder, and hip fractures. This dataset provides annotations for tasks such as classification [19], segmentation, and localization, and supports multiple formats, including COCO, VGG, and YOLO. The dataset was split into training, validation, and testing sets with a ratio of 70:15:15 to ensure proper model evaluation.

B. *Model Development*

Comparing the classification capabilities of YOLOv8 and VGG-16 can provide an evaluation of their effectiveness in accurately identifying and classifying bone fractures from X-ray images. These techniques have balancing strengths in feature extraction and object detection and also permit full-bodied precise fraction cataloging. This comparative analysis is crucial to understanding the potential of these DL models in real-world medical diagnostics and enhancing their accuracy through systematic training and hyperparameter tuning.

1) *YOLO*

YOLO stands for You Only Look Once and is a widely used algorithm recognized for its outstanding object detection and classification capabilities. Its main goal is to accurately identify and locate objects within an image by predicting bounding boxes and class probabilities. The distinctive approach of YOLO lies in processing the entire image in a single pass, utilizing global context to make predictions, which grants it remarkable speed [20].

YOLOv8 is an advanced object detection algorithm in computer vision. It has revolutionized the field by achieving superior detection accuracy and real-time performance using a single end-to-end neural network. YOLOv8 is widely used in various applications, such as autonomous driving, surveillance systems, and robotics, where rapid and accurate object detection is crucial. Its impressive performance and versatility have made it popular among researchers and practitioners in the computer vision community [21].

2) VGG-16

VGG-16, or Visual Geometry Group 16, is a renowned deep CNN architecture, known for its simplicity and effectiveness in image classification tasks. With 16 layers, including 13 convolutional layers and 3 fully connected layers, VGG-16 captures complex features from input images. Although newer models surpass its performance, VGG-16 remains a popular choice for transfer learning due to its strong feature extraction capabilities and publicly available pre-trained weights [19].

C. Training

This study aimed to evaluate the performance of YOLO in detecting bone fractures. Additionally, the VGG-16 model was employed to perform the same task, but with classification. A comparative analysis can help evaluate the effectiveness of these models in the context of bone fracture detection and classification. The models were trained using the FracAtlas dataset, and several hyperparameters, including epochs varying from 20 to 50 and batch sizes of 32 for YOLOv8 and 16 for VGG16.

D. Evaluation Metrics

This study used training and validation accuracy and loss to compare the performance of the models. In addition, the confusion matrix sheds light on the classification results of VGG16.

III. RESULTS

When comparing the performance of YOLOv8 and VGG16 based on their respective training and testing accuracies, several key differences emerge. Table II shows the best hyperparameter settings for both models. For YOLOv8, key parameters include an input image size of 640, 50 epochs, a batch size of 16, and the use of the SGD optimizer. The initial and final learning rates were both at 0.00001, with a momentum of 0.937 and a weight decay of 0.0005. These settings were crucial in achieving the best performance levels, highlighting the importance of hyperparameter tuning.

TABLE II. HYPERPARAMETER VALUES

Hyperparameter	Value for YOLOv8	Value for VGG16
Input image size	640	224
Epochs	50	40
Batch size	32	32
Optimizer	SDG	Adam
Initial learning rate	0.00001	0.00001
Final learning rate	0.00001	0.00001
Momentum	0.9	-
Weight decay	0.0005	-

The results of YOLOv8 show that the model achieved its peak training accuracy of 81% at 50 epochs, with a corresponding testing accuracy of 80%. This indicates that YOLOv8 performs consistently well at this point, making this number of epochs optimal. However, as the number of epochs increases beyond 50, both training and test accuracies decline, suggesting possible overfitting or other issues affecting the model's performance over extended training periods. Table III shows the accuracies of YOLOv8 and VGG16 for different training epochs, learning rates, and batch sizes.

TABLE III. MODELS RESULTS

Model	Epochs	Learning rate	Batch size	Training accuracy	Testing accuracy
YOLOv8	10	0.00001	32	72.61%	73.88%
	15	0.00001	32	75.44%	76.31%
	20	0.00001	32	78.65%	78.89%
	50	**0.00001**	**32**	**81%**	**80%**
	80	0.00001	32	70.54%	72.33%
	100	0.00001	32	64.83%	62.65%
VGG16	10	0.001	64	37.25%	45%
	20	0.001	32	60.36%	53.96%
	40	0.001	32	64.54%	65.0%
	60	0.001	32	69.62%	69.84%
	10	0.00001	32	70.62%	72.22%
	15	0.00001	32	75.20%	72.22%
	25	0.00001	32	82.37%	72.22%
	35	0.00001	32	84.16%	72.22%
	40	0.00001	32	86.55%	70.63%
	60	0.00001	32	82.97%	69.84%
	25	0.00001	64	73.11%	66.66%
	30	**0.0001**	**32**	**73.71%**	**72.22%**
	40	0.0001	32	83.33%	73.01%
	50	0.0001	32	84.06%	71.42%
	60	0.0001	32	82.17%	71.42%
	60	0.01	32	37.45%	44.44%
	60	0.02	164	38.94%	38.01%
	30	0.02	32	38.94%	38.01%

Figures 2 and 3 demonstrate the loss and accuracy for training and validation for YOLOv8, for different numbers of epochs. Initially, performance improves as the number of epochs increases, indicating that it is learning and capturing more meaningful features from the data. However, after reaching the peak performance at 50 epochs, accuracy starts to decline. This decline suggests that the YOLOv8 may have started to overfit the training data, meaning that it becomes too specialized in recognizing the training examples but fails to generalize well to new, unseen data. Overfitting is a common challenge in machine learning, and it is crucial to monitor the model's performance to prevent it.

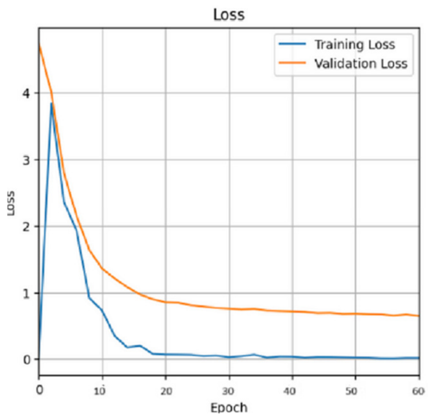

Fig. 2. YOLOv8 validation and training loss.

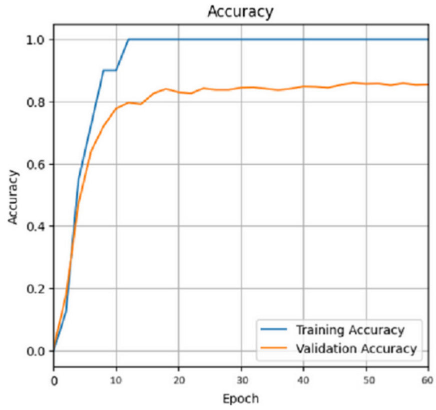

Fig. 3. YOLOv8 validation and training accuracy.

The performance of the VGG16 model varies significantly based on the hyperparameters used during training. The results in Table III highlight the impact of two key hyperparameters: learning rate and batch size. In VGG16, lower learning rates generally led to improved performance. For example, when the learning rate was set to 0.00001, the model achieved higher training and validation accuracy compared to the initial learning rate of 0.001. This suggests that a smaller learning rate allows the model to converge more effectively and learn better representations from the data. The batch size also influenced the model's performance. Smaller batch sizes, such as 32, tend to result in better performance compared to larger batch sizes, such as 64. This indicates that smaller batch sizes allow the model to make more frequent weight updates, which can help it converge faster and potentially achieve better accuracy. Figure 4 shows the training and validation accuracy of VGG16, while Figure 5 shows the confusion matrix of its results. Figures 6, 7, and 8 show some visual results of the predictions made by YOLOv8. Figure 6 shows the classification results for fractured and non-fractured hands, demonstrating the model's ability to accurately distinguish between these two classes. Figure 7 presents the classification results for fractured and non-fractured hips, further highlighting the model's precision in identifying fractures in different types of bones. Figure 8 displays the results for fractured and non-fractured shoulders, completing the set of classifications for the primary bone fracture categories considered in this study. These images underscore the effectiveness of YOLOv8 in accurately detecting and classifying bone fractures from X-ray images, showcasing its potential in medical diagnostics.

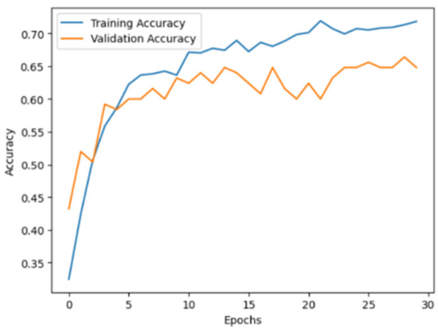

Fig. 4. VGG16 training and validation accuracy.

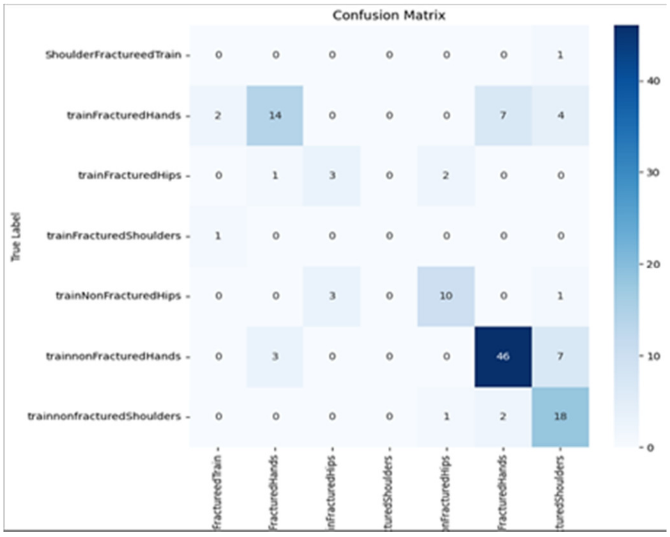

Fig. 5. VGG16 confusion matrix.

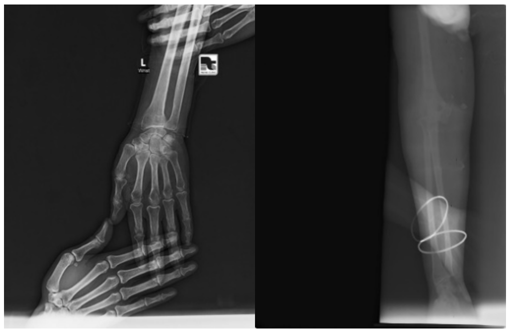

Fig. 6. Fractured and non-fractured hands.

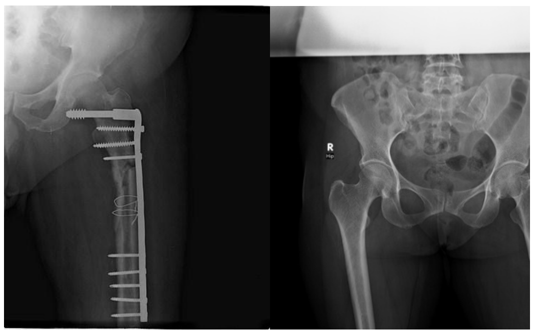

Fig. 7. Fractured and non-fractured hips.

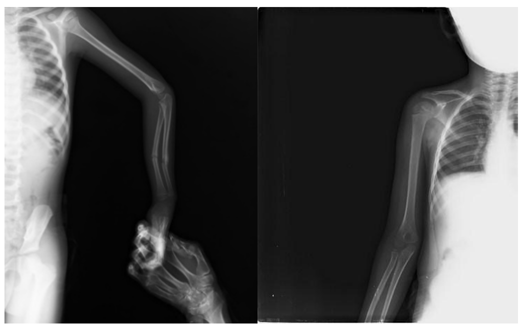

Fig. 8. Fractured and non-fractured shoulders.

IV. CONCLUSION

This study developed a bone fracture detection system using DL, specifically focusing on the YOLOv8 and VGG16 models, unlike previous studies that specifically focused on a specific type of bone fracture. These results suggest significant advancements in accuracy compared to similar studies. After training for 50 epochs, the YOLOv8 model exhibited remarkable performance, achieving an 81% training accuracy and an 80% testing accuracy. Its ability to perform real-time object detection and image segmentation makes it highly suitable for medical diagnostics. In comparison, the VGG16 model achieved a maximum testing accuracy of 73.01%. Although this is lower than the performance of YOLOv8, VGG16 still demonstrated the ability to identify fractures effectively. The choice of hyperparameters significantly influences the performance of these models. Optimal results were achieved with smaller learning rates and batch sizes. In general, YOLOv8 outperformed VGG16 in accuracy and consistency, making it a superior choice for real-time medical diagnostics. However, VGG16, despite its lower accuracy, can benefit from larger datasets and further hyperparameter tuning to enhance its effectiveness.

REFERENCES

[1] S. Sharma, "Artificial intelligence for fracture diagnosis in orthopedic X-rays: current developments and future potential," *SICOT-J*, vol. 9, Art. no. 21, https://doi.org/10.1051/sicotj/2023018.

[2] P. K. Mall *et al.*, "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities," *Healthcare Analytics*, vol. 4, Dec. 2023, Art. no. 100216, https://doi.org/10.1016/j.health.2023.100216.

[3] Ultralytics, "YOLOv8 Documantation." https://docs.ultralytics.com/.

[4] N. Kumar, A. Hashmi, M. Gupta, and A. Kundu, "Automatic Diagnosis of Covid-19 Related Pneumonia from CXR and CT-Scan Images," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 7993–7997, Feb. 2022, https://doi.org/10.48084/etasr.4613.

[5] E. A. Murphy *et al.*, "Machine learning outperforms clinical experts in classification of hip fractures," *Scientific Reports*, vol. 12, no. 1, Feb. 2022, Art. no. 2058, https://doi.org/10.1038/s41598-022-06018-9.

[6] D. P. Yadav and S. Rathor, "Bone Fracture Detection and Classification using Deep Learning Approach," in *2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)*, Mathura, India, Feb. 2020, pp. 282–285, https://doi.org/10.1109/PARC49193.2020.236611.

[7] R. F. A. Marar, D. M. Uliyan, and H. A. Al-Sewadi, "Mandible Bone Osteoporosis Detection using Cone-beam Computed Tomography," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 6027–6033, Aug. 2020, https://doi.org/10.48084/etasr.3637.

[8] H. A. Owida, A. Al-Ghraibah, and M. Altayeb, "Classification of Chest X-Ray Images using Wavelet and MFCC Features and Support Vector Machine Classifier," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7296–7301, Aug. 2021, https://doi.org/10.48084/etasr.4123.

[9] Implemented with a Small Sample, De Novo Training, and Multiview Incorporation," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 672–677, Aug. 2019, https://doi.org/10.1007/s10278-018-0167-7.

[10] F. Hardalaç *et al.*, "Fracture Detection in Wrist X-ray Images Using Deep Learning-Based Object Detection Models," *Sensors*, vol. 22, no. 3, Jan. 2022, Art. no. 1285, https://doi.org/10.3390/s22031285.

[11] S. R. Karanam, Y. Srinivas, and S. Chakravarty, "A Supervised Approach to Musculoskeletal Imaging Fracture Detection and Classification Using Deep Learning Algorithms," *Computer Assisted Methods in Engineering and Science*, vol. 30, no. 3, pp. 369–385, Mar. 2023, https://doi.org/10.24423/cames.682.

[12] S. Vironicka and J. G. R. Sathiaseelan, "Classification of Long-Bone Fractures Using Modified Faster RCNN for X-Ray Images," *Indian Journal Of Science And Technology*, vol. 16, no. 1, pp. 56–65, Jan. 2023, https://doi.org/10.17485/IJST/v16i1.1690.

[13] J. H. F. Oosterhoff *et al.*, "A deep learning approach using an ensemble model to autocreate an image-based hip fracture registry," *OTA International*, vol. 7, no. 1S, Jan. 2024, Art. no. e283, https://doi.org/10.1097/OI9.0000000000000283.

[14] R. Y. Ju and W. Cai, "Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm," *Scientific Reports*, vol. 13, no. 1, Nov. 2023, Art. no. 20077, https://doi.org/10.1038/s41598-023-47460-7.

[15] K. Thaiyalnayaki, L. Kavyaa, and J. Sugumar, "Automated Bone Fracture Detection Using Convolutional Neural Network," *Journal of Physics: Conference Series*, vol. 2471, no. 1, Dec. 2023, Art. no. 012003, https://doi.org/10.1088/1742-6596/2471/1/012003.

[16] R. Hrubý, D. Kvak, J. Dandár, A. Atakhanova, M. Misař, and D. Dufek, "Cross-Center Validation of Deep Learning Model for Musculoskeletal Fracture Detection in Radiographic Imaging: A Feasibility Study." medRxiv, Art. no. 2024.01.17.24301244, Jan. 17, 2024, https://doi.org/10.1101/2024.01.17.24301244.

[17] T. Kumar and R. Ponnusamy, "Robust Medical X-Ray Image Classification by Deep Learning with Multi-Versus Optimizer," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 111406–11411, Aug. 2023, https://doi.org/10.48084/etasr.6127.

[18] A. Faramawy, "notebook86c6300cc2." https://www.kaggle.com/code/abdelazizfaramawy/notebook86c6300cc2/noteb.

[19] I. Abedeen, M. A. Rahman, F. Z. Prottyasha, A. Tasnim, S. Shatabda, and T. M. Chowdhury, "Fracture Classification Dataset." Kaggle, https://doi.org/10.34740/KAGGLE/DSV/7718956.

[20] A. Al-Shahrani, W. Al-Amoudi, R. Bazaraah, A. Al-Sharief, and H. Farouquee, "An Image Processing-based and Deep Learning Model to Classify Brain Cancer," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15433–15438, Aug. 2024, https://doi.org/10.48084/etasr.7803.

[21] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, https://doi.org/10.1007/s13244-018-0639-9.