Bone Fracture Classification Using Neural Networks from X-ray Images

Amal Alsahrani [1], Alaa Alsairafi [1] , Athary Alsowat [1] , Abrar Saigh [1] , Aldanh Almatarfy [1] , Rouaa Redwan [1]

Department of Computer Science and Artificial Intelligence, College of Computers, Umm Al-Qura University, Makkah 21955, Saudi Arabia.

Abstract
In this study, a bone fracture classification system using deep learning algorithms was developed 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. The CNN architecture, known for its effectiveness in image recognition tasks, was chosen for its proven effectiveness in detecting bone fractures from X-ray images. These efforts in model development and hyperparameter tuning significantly enhanced the system's ability to accurately detect and classify bone fractures. The study utilized the FracAtlas dataset, which contains 4,083 X-ray images of fractured and non-fractured human bones, to improve the accuracy and efficiency of fracture detection compared to current methods. By integrating advanced deep learning techniques, the goal was to assist surgeons with more accurate diagnostics. The performance of the developed system was evaluated against existing methodologies, 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.

1. Introduction

Bone fractures are a common and significant medical issue, often requiring precise and timely diagnosis for effective treatment [12]. Traditionally, the detection and classification of fractures are performed by radiologists who analyze X-ray images to identify the presence and type of fractures. 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, a subset of artificial intelligence, has revolutionized the field of medical imaging by offering powerful tools for automated image analysis. Deep learning 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 [13].

In this paper, we focus on classifying bone fractures from X-ray images of the hand, hip, and shoulder using two state-of-the-art deep learning techniques: YOLOv8 and VGG16. YOLOv8 (You Only Look Once, version 8) is a cutting-edge object detection model known for its real-time detection capabilities and high accuracy [14]. VGG16, on the other hand, is a renowned CNN model known for its depth and simplicity, consisting of 16 layers that can extract intricate features from images [15]. Its architecture is particularly effective in recognizing patterns and textures, which are crucial for distinguishing between fractured and non-fractured bones.

By leveraging the capabilities of both YOLOv8 and VGG16, we aim to develop robust systems capable of accurately classifying bone fractures in X-ray images. Following the implementation of these models, we will conduct a comparative analysis to evaluate their performance in terms of accuracy, processing time, and reliability. This comparison will provide valuable insights into the strengths and limitations of each approach, ultimately guiding the selection of the most effective model for clinical use.

The ultimate goal of this project is to assist radiologists by providing reliable second opinions, reducing diagnostic errors, and enhancing the efficiency of the diagnostic process. By improving the accuracy and speed of fracture classification, we aim to facilitate timely and appropriate medical intervention, thereby improving patient outcomes. The remainder of this paper is organized as follows: Section 2: Previous Research, Section 3: Methodology, Section 4: Results, and Section 5: Conclusion

2. Literature Review

In the realm of medical imaging diagnostics, recent years have witnessed remarkable progress in fracture

Manuscript received June 5, 2025
Manuscript revised June 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.6.6

detection and classification through the application of deep learning algorithms.

In 2019, Kitamura et al. [1] discussed the use of convolutional neural networks (CNNs) for ankle fracture detection and explored the effectiveness of training CNN models from scratch with a small dataset. The study collected 298 radiographs of non-fractured bones and 298 radiographs of fractured ankle cases and created single- and multiview models to evaluate the impact of multiple views. Data augmentation techniques were employed during training. The Inception V3, Resnet, and Xception CNN architectures were implemented using the Python programming language with Tensorflow as the framework. The performance of the models was evaluated using metrics such as accuracy, positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity. The results showed that the ensemble of all five models achieved the best accuracy of 76% when using single radiographic views. When utilizing all three views for a single case, the ensemble of all models resulted in the best output metrics with an accuracy of 81%.

The year 2020 witnessed a pivotal moment in fracture diagnosis with D.P. Yadav and S [2]. The paper addressed the challenge of enhancing the efficiency and accuracy of bone fracture diagnoses through an automated system that employed deep learning techniques. They managed the limitations inherent in manual fracture detection methods, which were both time-consuming and prone to errors, by developing a deep neural network (DNN) model capable of classifying bones as either fractured or healthy.

To mitigate the risk of overfitting associated with the small initial dataset of 100 bone X-ray images, they augmented it to 4000 images using a variety of image transformation techniques. The model, developed using Python and Keras, incorporated an architecture that included convolution, pooling, flattening, and dense layers, specifically designed for binary classification tasks. The researchers reported a notable achievement, with the model reaching a classification accuracy of 92.44% through 5-fold cross-validation, thereby surpassing the performance metrics of prior studies in the field. The authors suggested future improvements, such as exploring larger datasets, investigating advanced deep learning architectures, and incorporating additional features or techniques to further enhance the accuracy and efficiency of bone fracture detection and classification.

As machine learning continued to gain traction in medical diagnostics in 2022, the study by Murphy et al. [3] aimed to develop a machine learning method for identifying and classifying hip fractures and compared its performance to that of experienced human observers. The researchers utilized a dataset of 3659 hip radiographs, which were classified by expert clinicians. The results demonstrated that the machine learning method achieved an overall accuracy of 92%, surpassing the accuracy of human experts by 19%. The study highlights the potential of machine learning for improving fracture classification and its impact on patient outcomes and treatment costs. In the study, HardalaÃ§, F et al. [4] provided many object detection models and the accuracy of them. 20 different fracture detection procedures were performed on Gazi University Hospital's dataset of wrist X-ray images. The results achieved from these procedures were examined from different perspectives, and six different ensemble models were developed to further improve the results of detection. Gazi University Hospital's dataset of wrist X-ray images was used along with 542 images collected from the study's hospital, and out of 434 images in the training dataset, 28 were pairs (right and left hands), 187 belonged to the right hand, and 219 belonged to the left hand. And the highest result obtained goes to the dynamic R-CNN model with an accuracy of 77.7%.

In 2023, numerous studies delved into addressing the challenge of fractures, presenting innovative solutions and methodologies in the field of medical imaging diagnostics. Karanam et al. [5] introduced an innovative methodology for detecting and classifying bone fractures through advanced deep learning (DL) techniques. Central to their study is the use of pre-trained deep neural networks, namely ResNeXt101, InceptionResNetV2, Xception, and NASNetLarge, which were adeptly applied to the analysis of 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 have often fallen short. Karanam et al. strategically addressed this need by employing a diverse array of DL models, leading to a notable improvement in the accuracy and efficiency of fracture classification. The dataset was composed of a variety of X-ray images, meticulously selected to represent a wide range of bone fractures. The exact number of images used in their study isn't specified in the details. The X-ray images in their dataset covered different types of fractures, such as simple, complex, and compound fractures. This is exemplified by their highest-performing model, InceptionResNetV2, achieving an impressive accuracy rate of 94.58%.

However, the research is not without its limitations, as it falls short of providing a detailed comparative analysis against existing methodologies, a gap that leaves room for further exploration and validation. Despite this, the study marks a significant stride in the integration of AI in radiological diagnostics, setting a new precedent for the use of deep learning models in the accurate identification of a wide range of fracture types.

Vironicka and Sathiaseelan [6] introduced a pioneering method for detecting fractures in X-ray images, with a particular focus on long-bone fractures. This approach involved modifying the Faster R-CNN deep-learning 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. To achieve this, the study employed complex mathematical techniques, such as the Rotated Discrete Curvature System (RDS) and shape directory, to improve the precision of identifying fractures. The use of a rotated bounding box was especially crucial, as it provided detailed insights into the fracture's orientation and length, streamlining the detection process without the need for additional segmentation or measurement. A dataset comprising 200 X-ray images of long bone fractures was utilized. These images were sourced from the Rajiv Gandhi Government General Hospital in Chennai and were categorized into two groups: 120 images for training the model and 80 for validation. This dataset played a crucial role in evaluating the effectiveness of the modified algorithm. However, while the study made significant strides in medical imaging, it did not provide specific numerical data on the model's accuracy or effectiveness. Despite this, the research made a notable contribution to the field, demonstrating the potential of deep learning models for revolutionizing fracture detection. Importantly, the modified model achieved a high accuracy rate of 96.1%, demonstrating the efficacy of this novel approach in improving fracture detection techniques and laying the groundwork for future advancements in the area.

In a study conducted by Oosterhoff et al. [7], an innovative approach that addressed the significant challenge of hip fractures in the elderly was developed. Given the expected rise in hip fracture cases, reaching an estimated 6 million worldwide by 2050, the need for an efficient and accurate registry system was more pressing than ever. Recognizing the limitations of existing fracture registries, which often rely on inaccurate billing and procedural codes, the research team proposed a deep learning-based solution.

They created a sophisticated model 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. Notably, the model achieved accuracy rates between 92% and 100% across various submodules, significantly reducing the time required for image annotation compared to traditional methods.

Moreover, Ju, R.-Y., and Cai, W.P. [8] used data augmentation to improve the model performance of the YOLOv8 algorithm and let surgeons use the provided model for fracture detection in pediatric wrist trauma. X-ray images, which led the researchers to design an application to assist surgeons in diagnosing fractures, reducing the probability of error analysis, and providing more useful information for surgery. Other body component fractures can be identified using the model as a pre-training model. The purpose of the application is to help pediatric surgeons correctly diagnose fractures. 20,327 X-ray images of pediatric wrist injuries make up the training dataset, and after augmenting the data, it became 28408 images. In comparison to the Adam optimizer, the model's accuracy was higher after being trained with the SGD optimizer. Future plans call for expanding the application's usage to novice pediatric surgeons in developing nations and launching it on several platforms. The result obtained after training the model was 73.4%, which is higher than the result of YOLOv8.

K. Thaiyalnayaki et al. [9] successfully tackled the challenge of addressing the need for an accurate and efficient method of diagnosing bone fractures through automated analysis of X-ray and CT images. Their approach involves the development of an image processing system that harnesses the power of a Convolutional Neural Network (CNN) architecture. By utilizing a dataset comprising normal and cracked X-ray images of nearly 100 bones, this system enables rapid and accurate classification of bone fractures. Impressively, their study yielded promising results, achieving a high classification accuracy of 99.5%. These findings highlight the potential of CNN-based approaches for automated bone fracture detection, offering a valuable solution in the field of medical imaging diagnostics. Cross-Center Validation of a Deep Learning Model for Musculoskeletal Fracture Detection in Radiographic Imaging: A Feasibility Study, Robert Hrubý et al. [10] Presented a comprehensive investigation into the feasibility of utilizing a deep learning-based decision support system to address the diagnostic challenges associated with musculoskeletal fractures and enhance fracture detection in radiographic imaging. The methodology involved training a deep learning model using annotated musculoskeletal X-rays, specifically employing the YOLO architecture, and testing its performance on two datasets to evaluate its effectiveness. The attained results showed a sensitivity (Se) of 0.910 (95% CI: 0.852–0.946) and specificity (Sp) of 0.557 (95% CI: 0.520–0.594) on Dataset 1, indicating the model's ability to correctly detect fractures. These findings underscore the potential of the deep learning model for improving fracture detection in radiographic imaging.

Table 1: 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 bone	(DNN)	92.44%

[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	The dataset has an overall image count of 100. Out of the available, the images are categorized as Cracked = 100, Normal bone = 100.	CNN	99.5%.
[10]	2024	Combination of two datasets: the MURA dataset and the FracAtls dataset	YOLOv7	99.5%

After reviewing previous research, it was clear that some studies had limited datasets and could benefit from implementing various algorithms to improve accuracy. These limitations emphasize the importance of having a larger dataset and exploring different algorithmic approaches to achieve better accuracy when classifying and identifying fractures.

3. Methodology

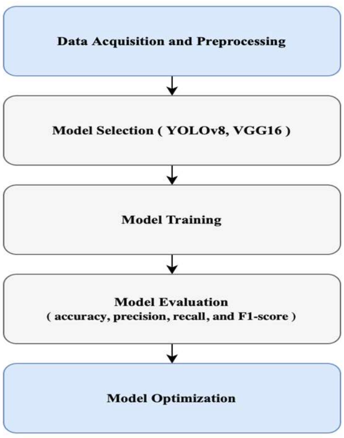

Figure 1: Methodology

This section outlines the systematic process of building a model for our bone fracture detection and classification project. The stages of model development are sequenced as follows: data acquisition and preprocessing, model training, and model evaluation.

Additionally, we discuss the design considerations integral to the project, providing insights into the decision-making processes and strategies employed to enhance the model's efficacy and efficiency. Each stage is crucial for the successful implementation and operation of the model in a real-world medical setting. Figure 1 in the document visually presents these methodology steps, offering a clear and structured overview of the entire process.

3.1 Data Acquisition and Preprocessing

The dataset used for this study is the FracAtlas dataset [16], which consists of 4,083 X-ray images depicting various types of bone fractures, including hand, leg, shoulder, and hip fractures. This dataset provides annotations for tasks such as classification [11], segmentation, and localization. It supports multiple formats, including COCO, VGG, and YOLO. Also, a dataset was split into training, validation, and testing sets with a ratio of 70:15:15 to ensure the model was properly evaluated.

3.2 Model Development

This section delves into the development of bone fracture detection and classification models, specifically focusing on the YOLOv8 and VGG-16 architectures. We outline the rationale behind selecting these models, the training methodologies employed, and the specific configurations used to optimize their performance. By comparing the classification capabilities of YOLOv8 and VGG-16, we aim to evaluate their effectiveness in accurately identifying and classifying bone fractures from X-ray images. This comparative analysis is crucial for understanding the potential of these deep learning models in real-world medical diagnostics and enhancing their accuracy through systematic training and hyperparameter tuning.

A. YOLO Algorithm

An acronym for "You Only Look Once" (YOLO) 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. YOLO's distinctive approach lies in processing the entire image in a single pass, utilizing global context to make predictions, which grants it remarkable speed [18].

- YOLOv8

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 utilized 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 a popular choice among researchers and practitioners in the computer vision community [18].

B. VGG-16

VGG-16, or Visual Geometry Group 16, is a renowned deep convolutional neural network 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. Despite newer models surpassing 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 Methodology

The primary methodology of this study was to compare the performance of YOLO with previous studies in detecting bone fractures. Additionally, we employed the VGG-16 model to perform the same task, but with the classification of bone fractures. This comparison allowed us to assess and evaluate the effectiveness of both YOLOv8 and VGG-16 in the context of bone fracture detection and classification. The models were trained using the FracAtlas dataset consisting of 4,083 images and a set of hyperparameters that included epochs varying from 20 to 50 and batch sizes of 32 for YOLOv8, and 16 for VGG-16. Below are the tables that show the hyperparameter settings.

3.3 Evaluation Metrics

Our experiments revealed that the YOLOv8 model achieved the highest performance with a testing accuracy of 80% after 50 epochs. The VGG16 model also performed well, achieving a testing accuracy of 73.01% under optimal conditions.

Finally, the methodology adopted in this study, involving data augmentation, extensive model training, and hyperparameter tuning, significantly improved the accuracy of bone fracture detection and classification from X-ray images. The results demonstrate the potential of deep learning models to assist medical professionals with more accurate and efficient diagnostics.

4. Results

When comparing the performance of YOLOv8 and VGG16 based on their respective training and testing accuracies, several key differences emerge. YOLOv8's results show that the model achieves a peak training accuracy of 81% at 50 epochs, with a corresponding test accuracy of 80%. This indicates that YOLOv8 performs consistently well at this point, making it the optimal number of epochs for this model. 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 2: Models Results

Model	Epoch	Learning rate	Batch size	Train Acc	Test Acc
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%

Table 3: Hyperparameters 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 for YOLOv8 as shown in Table 2, are the accuracy scores for different numbers of training epochs which is also shown in Figure 2. Figure 1 demonstrates the training and validation loss with different numbers of epochs. Initially, the model's performance improves as the number of epoch's increases, indicating that it is learning and capturing more meaningful features from the data.

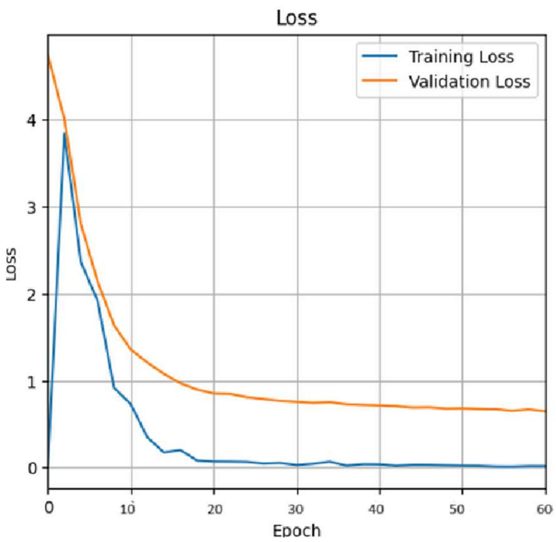

Figure 2 YOLOv8 Validation and Training Loss

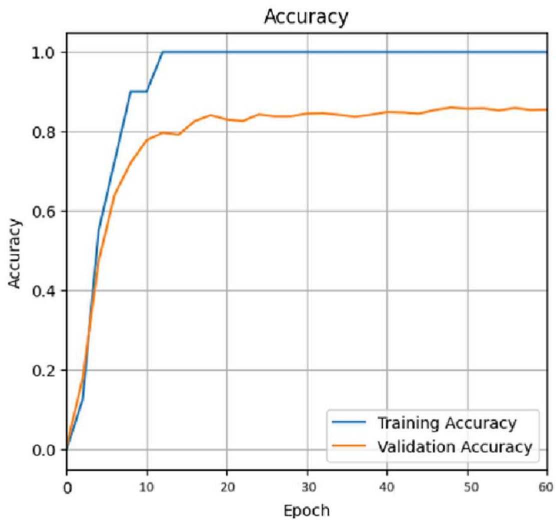

Figure 3 YOLOv8 Validation and Training Accuracy

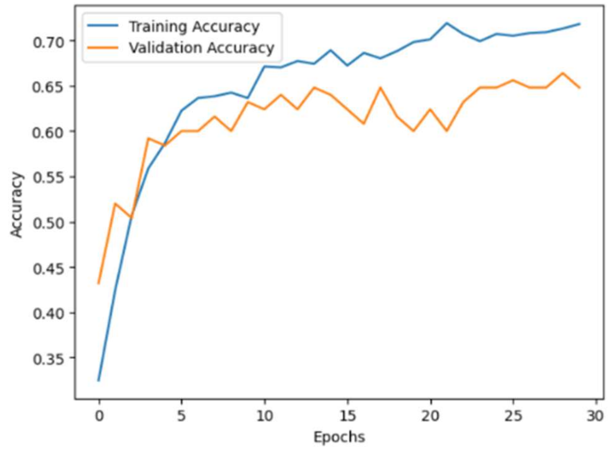

Figure 4 VGG16 Validation and Training Accuracy

The hyperparameters set during the training of the YOLOv8 model are detailed in Table 3. Key parameters include an input image size of 640, 150 epochs, a batch size of 16, and the use of the SGD optimizer. The initial and final learning rates were both set at 0.01, with a momentum of 0.937 and a weight decay of 0.0005. These settings were crucial in achieving the model's performance levels, highlighting the importance of hyperparameter tuning in deep learning model training.

However, after reaching the peak performance at 50 epochs, accuracy starts to decline. This decline suggests that the model may have started to overfit the training data, meaning it becomes too specialized in recognizing the training examples but fails to generalize well to new, unseen data.

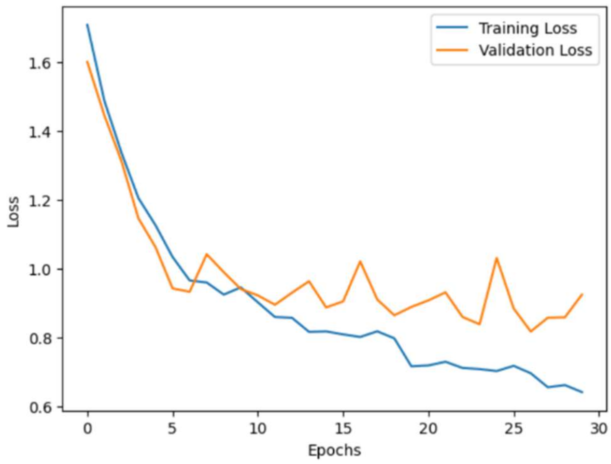

Figure 5 VGG16 Validation and Training loss

Overfitting is a common challenge in machine learning, and it is crucial to monitor the model's performance to prevent it.

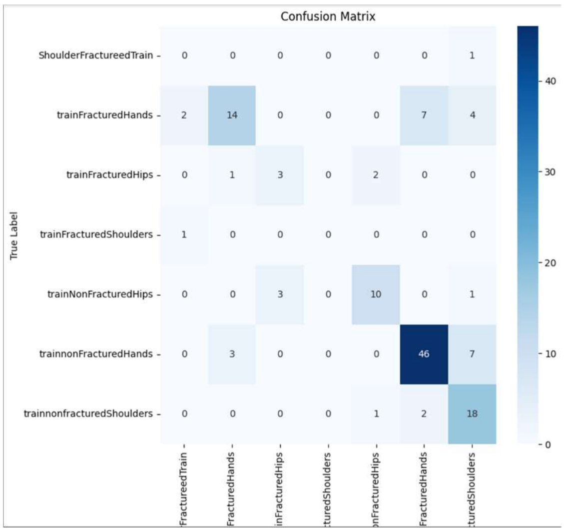

Figure 6: VGG16 Confusion Matrix

The performance of the VGG16 model varies significantly based on the hyperparameters used during training, as shown in Figure 3, Figure 4, and Figure 5.

The results highlight the impact of two key hyperparameters: learning rate and batch size. Learning rate: In the case of VGG16, lower learning rates generally lead to improved performance. For example, when the learning rate is set to 0.00001, the model achieves 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. Batch size: The batch size also influences the model's performance. Smaller batch sizes, such as 32, tend to result in better performance compared to larger batch sizes, like 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.

The visual results of the model's predictions are shown in Figures 7, 8, and 9. Figure 7 shows the classification results for fractured and non-fractured hands, demonstrating the model's ability to accurately distinguish between these two classes. Figure 8 presents the classification outcomes for fractured and non-fractured hips, further highlighting the model's precision in identifying fractures in different types of bones. Finally, Figure 9 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 the YOLOv8 model in accurately detecting and classifying bone fractures from X-ray images, thus showcasing its potential utility in medical diagnostics.

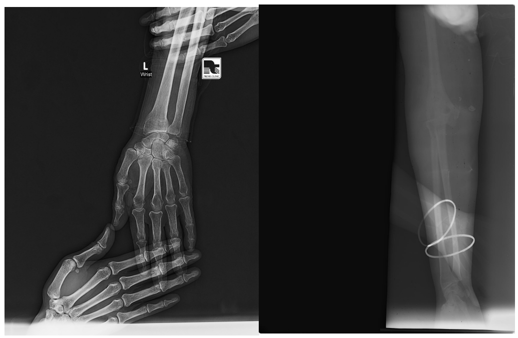

Figure 7 Fractured and Non-Fractured Hands

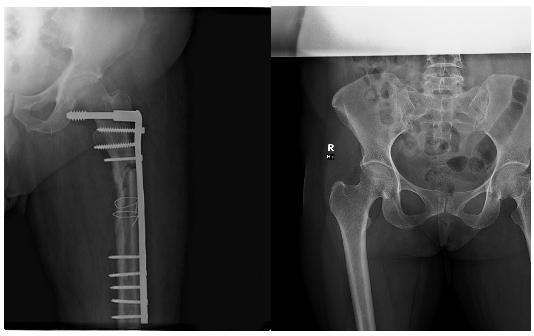

Figure 8 Fractured and Non-Fractured Hips

Figure 9 Fractured and Non-Fractured Shoulders

5. Conclusion

In this paper, a bone fracture classification system using deep learning algorithms has been developed. The dataset, which contains X-ray images of fractured and non-fractured human bones, was used to achieve experimental results. Images of hand, hip, and shoulder bones were collected from the FracAtlas dataset [11]. The total size of the dataset is 1,490. The classification accuracy of YOLOv8 is 80%, which is better than that reported in [8]. As for VGG16, the classification accuracy reached 72.22%.

References

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

[2] Yadav, D. P., & Rathor, S. (2020). Bone Fracture Detection and Classification using Deep Learning Approach. 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), IEEE.

[3] G. Kitamura, C. Y. Chung, and B. E. Moore, "Ankle Fracture Detection Utilizing a Convolutional Neural Network Ensemble Implemented with a Small Sample, De Novo Training, and Multiview Incorporation," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 672–677, Apr. 2019, doi: https://doi.org/10.1007/s10278-018-0167-7.

[4] Hardalaç, F.; Uysal, F.; Peker, O.; Çiçeklidag˘, M.; Tolunay, T.; Tokgöz, N.; Kutbay, U.; Demirciler, B.; Mert, F.(2022) Fracture Detection in Wrist X-ray Images Using Deep Learning-Based Object Detection Models. Sensors. https://doi.org/10.3390/s22031285

[5] 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, doi: https://doi.org/10.24423/cames.682.

[6] 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, doi: https://doi.org/10.17485/ijst/v16i1.1690.

[7] Oosterhoff, J. H. F., Jeon, S., Akhbari, B., Shin, D., Tobert, D. G., Do, S., & Ashkani-Esfahani, S. (2023). A deep learning approach using an ensemble model to auto-create an image-based hip fracture registry. OTA International, e283. http://dx.doi.org/10.1097/OI9.0000000000000283

[8] Yam Ju, R.-Y., & Cai, W. (2023). Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm. Scientific Reports, 13(1). https://doi.org/10.1038/41598-023-47460-7

[9] Krishnan, K.Thaiyalnayaki & Kavyaa, L & Sugumar, Joshua. (2023). Automated Bone Fracture Detection Using Convolutional 012003. 10.1088/1742-6596/2471/1/012003.

[10] Hruby, Robert & Kvak, Daniel & Dandár, Jakub & Atakhanova, Anora & Misař, Matěj & Dufek, Daniel. (2024). Cross-Center Validation of Deep Learning Model for Musculoskeletal Fracture Detection in Radiographic Imaging: A Feasibility Study. 10.1101/2024.01.17.24301244.

[11] Fracture Classification Dataset," *www.kaggle.com*. https://www.kaggle.com/datasets/akshayramakrishnan28/fracture-classification-dataset/data

[12] S. Sharma, "Artificial intelligence for fracture diagnosis in orthopedic X-rays: current developments and future potential," *SICOT-J*, vol. 9, pp. 21–21, Jan. 2023, doi: https://doi.org/10.1051/sicotj/2023018.

[13] 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, p. 100216, Dec. 2023, doi: https://doi.org/10.1016/j.health.2023.100216.

[14] G. Jocher, "YOLOv8 Documentation," *docs.ultralytics.com*, May 18, 2020. https://docs.ultralytics.com/

[15] "VGG-16 | CNN model," *GeeksforGeeks*, Feb. 26, 2020. https://www.geeksforgeeks.org/vgg-16-cnn-model/

[16]"FracAtlas," *www.kaggle.com*. https://www.kaggle.com/datasets/abdelazizfaramawy/fracatlas

[17] 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, Jun. 2018, doi: https://doi.org/10.1007/s13244-018-0639-9.

[18] "Under the Hood: YOLOv8 Architecture Explained," *Keylabs: latest news and updates*, Dec. 20, 2023. https://keylabs.ai/blog/under-the-hood-yolov8-architecture-explained/

[19] "VGG-16 | CNN model," *GeeksforGeeks*, Feb. 26, 2020. https://www.geeksforgeeks.org/vgg-16-cnn-model/