Image Detection of Rare Orthopedic Diseases based on Explainable AI

Qi-Xiang Zhang
Department of Computer Science and Information Engineering, Providence University
200, Sec. 7, Taiwan Blvd., Shalu Dist., Taichung City 433301 Taiwan
g1110198@o365st.pu.edu.tw

Shun-Ping Wang
Division of Hand and Foot Surgery, Taichung Veterans General Hospital
1650, Sec. 4, Taiwan Blvd., Xitun Dist., Taichung City 407219 Taiwan
wsp0120@vghtc.gov.tw

Yu-Wei Chan
Department of Computer Science and Information Management, Providence University
200, Sec. 7, Taiwan Blvd., Shalu Dist., Taichung City 433301 Taiwan
ywchan@gm.pu.edu.tw

Chih-Hung Chang
Department of Computer Science and Communication Engineering, Providence University
200, Sec. 7, Taiwan Blvd., Shalu Dist., Taichung City 433301 Taiwan
ch.chang@gm.pu.edu.tw

Abstract

Image detection has significant application value in medicine, especially in detecting Muller-Weiss Disease (MWD) in orthopedic X-ray images. Traditional manual interpretation methods can be influenced by subjective factors and individual experience, and they can be time-consuming and labor-intensive. In this study, by utilizing advanced object detection models like YOLOv8, we can automatically and accurately identify specific structures and abnormalities in the images, providing real-time feedback, significantly improving physicians' diagnostic accuracy. Furthermore, the use of the Grad-CAM technique to generate heatmaps enhances the interpretability of the model's decisions, helping physicians understand the basis for the model's judgments, further boosting confidence and accuracy in diagnosis. Therefore, image detection plays a critical role in medical image diagnosis, potentially improving diagnostic efficiency and enhancing healthcare quality.

1. Introduction

In recent years, with advancements in technology and the thriving development of artificial intelligence, image detection technology has demonstrated immense application value across various fields. Particularly in medicine, image detection technology is becoming increasingly widespread and holds significant importance in improving the accuracy and efficiency of medical diagnosis. Among these applications, the field of orthopedics, as an integral component of medical image diagnosis, has long faced challenges and demands, seeking faster, automated, and more precise diagnostic methods.

In orthopedic X-ray images, Muller-Weiss Disease (MWD) is a rare yet significant condition involving pathology in the human body's foot joints and skeletal structures. It requires precise evaluation by physicians to provide appropriate treatment recommendations. Due to its infrequent occurrence in clinical practice, there is a risk of delayed or missed diagnosis. If not identified early, it can lead to delayed treatment, worsening the condition, and impacting the patient's ability to walk. However, traditional manual interpretation methods often introduce diagnostic variations due to the physician's personal experience, expertise, and subjective factors. Moreover, these methods can be time-consuming and demanding for clinical practitioners, potentially affecting the accuracy and efficiency of the diagnosis.

To address these challenges, this research focuses on applying image detection technology to tackle the diagnosis of

MWD. Specifically, we have adopted the advanced object detection model YOLOv8. This model delivers high accuracy and significantly improves computational speed while maintaining detection performance. We have achieved substantial success in the test set by training on a large dataset of orthopedic X-ray images. This has elevated the diagnostic rate of physicians from less than eighty percent, as seen with traditional methods, to over ninety percent, markedly enhancing the reliability of diagnosis.

Furthermore, to enhance the interpretability of the model's decision-making process, we have introduced Grad-CAM (Gradient-weighted Class Activation Mapping) technology to generate heatmaps. These heatmaps clearly visualize the key regions the model focuses on during detection. This aids physicians in understanding the rationale behind the model's judgments, boosting their confidence in the diagnostic process.

The primary objective of this research is to apply image detection technology to the detection of MWD in orthopedic X-ray images, to improve the accuracy and efficiency of diagnosis. The outcomes of this study will provide a practical and effective solution to enhance the current landscape of medical image diagnosis. This can positively impact healthcare quality and patients' quality of life.

In the upcoming sections of this paper, we will provide a detailed overview of image detection technology, the YOLOv8 model, and Grad-CAM technology, along with the methodology and experimental results of this research.

2. Related Works

2.1. YOLOv8

YOLOv8 (You Only Look Once version 8) [1] is an object detection model developed by Ultralytics, recognized as one of the most advanced models. It excels in object detection tasks and encompasses multiple functionalities such as classification, instance segmentation, object tracking, and human pose estimation. The YOLO series is renowned for its high accuracy and speed, performing impressively regarding model size variations or detection speed improvements. Even under less favorable hardware conditions, YOLOv8 demonstrates excellent training results.

YOLOv8 distinguishes itself from its predecessors by introducing new convolutional layers and incorporating techniques like anchor-free detection and mosaic augmentation. These enhancements contribute to faster detection speeds for the model.

2.2. Visually Interpretable Thermograms

Convolutional Neural Networks (CNNs) are a familiar network architecture known for their robust capabilities in image recognition. However, despite our knowledge that CNN-based neural networks can perform image recognition even more accurately than humans, we often cannot easily discern what these neural networks are doing. Neural networks are like black boxes, as we only know their inputs and outputs. This leads to skepticism regarding the reliability of CNN-based neural network models.

Zhou [3] proposed Class Activation Mapping (CAM) to address this issue. In CAM, each feature map generated by the last convolutional layer is transformed into a pixel through Global Average Pooling (GAP). This pixel contains information from the entire feature map. Subsequently, the one-dimensional array of pixels is multiplied by weight "w" and passed through softmax to obtain scores for various categories. Finally, CAM displays the interpretability of the "Australian terrier" category as a heatmap.

However, a drawback of CAM is that if the final layer of the neural network does not use Global Average Pooling (GAP), the model's architecture must be modified, and retraining is necessary. Selvaraju et. al. [2] introduced Grad-CAM (Gradient-weighted CAM) to address this issue. Regardless of the neural network architecture used after the convolutional layers, Grad-CAM can be implemented without modifying the model.

The primary weight calculation method in Grad-CAM involves computing the partial derivatives of the neurons with respect to the feature maps of the last convolutional layer, which is done using backpropagation to calculate gradients. This allows for generating interpretable heatmaps without significant architectural changes, making Grad-CAM a versatile and effective tool for understanding neural network decisions.

3. Research Methodology

This research aims to utilize image detection models to determine the presence of Muller-Weiss Disease (MWD). We will employ machine learning and deep learning techniques to detect potential signs of MWD in orthopedic images.

3.1. Research Processes

The research process for this study is depicted in Fig. 1 . Initially, we collaborated with a hospital in the data collection phase to obtain X-ray images of orthopedic foot cases. Subsequently, we engaged expert physicians to assess and categorize these images, and labels were assigned based on their assessments.

Next, we utilized the YOLOv8 model, specifically the YOLOv8m variant, for training and testing the image detection model. We introduced heatmaps for image detection to provide a more unambiguous indication of the criteria for MWD detection. These heatmaps will display hotspot areas in the images that may be related to MWD, aiding our understanding of how the model makes its determinations.

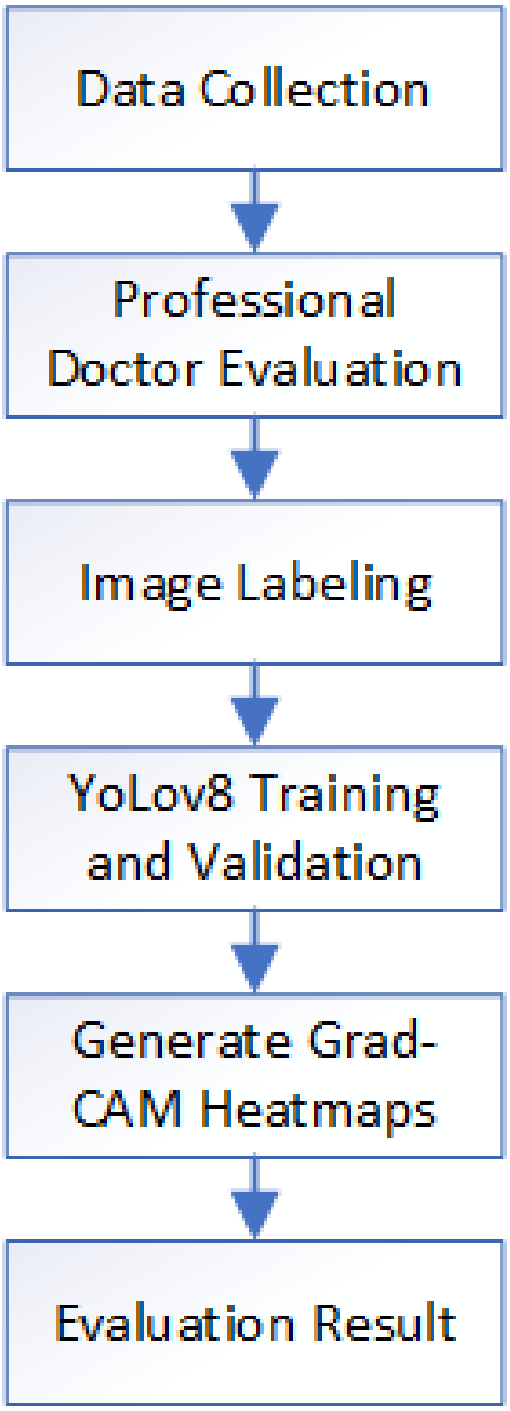

Figure 1. Research process.

Simultaneously, we will highlight the criteria for determining the absence of the disease for comparative analysis. In this regard, we employed the interpretable heatmap model Grad-CAM.

Through these methods, we aim to enhance the accuracy and reliability of MWD detection. This will improve diagnostics and treatment in the medical field and provide earlier opportunities for treatment and care to patients with MWD. Our research findings have the potential to have a positive impact on improving medical diagnosis and healthcare management.

In conclusion, we will evaluate the model's performance using metrics such as accuracy, precision, recall, and mAP (mean Average Precision). Additionally, we will analyze the model's interpretability through heatmaps to better understand the model's decision-making process and criteria.

3.2. Introduction to the Datasets

In the dataset, we have divided the images into two groups: the disease group (Muller Weiss Disease, MWD) and the normal (healthy) group. The determination of the disease group is primarily performed by professional physicians, pre-classified, and then labeled for training the image detection model. Since the number of photos with MWD is extremely limited, with only 256 images available for training. In contrast, there are more photos in the normal group. Still, to enhance the model's learning of the disease group photos, we have reduced the number of photos in the normal group to be closer to the disease group, selecting 294 images. Therefore, the training set comprises a total of 550 photos. In the test set, we will evaluate 18 photos from the disease group and 32 from the normal group, totaling 50 photos.

The design of this dataset aims to ensure that the model can thoroughly learn and identify the features of MWD while maintaining an even distribution of data. Despite the limited number of photos in the disease group, we expect the model to perform well in detecting and classifying the disease through a comprehensive range of data sources in the training set.

It is worth noting that the scale and balance of the dataset are crucial factors influencing model training and performance. Therefore, we have meticulously selected and processed the training set to ensure that it suits our research goals and provides reliable results.

In determining the presence of MWD, it's essential to consider multiple perspectives rather than relying solely on a single image viewpoint. We train our image detection model using images captured from four different viewpoints, allowing the model to learn a broader set of features for classification. As illustrated in Fig. 2, the top-left image represents the ankle AP view, the top-right image represents the ankle lateral view, the bottom-left image represents the standing foot AP view, and the bottom-right image represents the standing foot lateral view, totaling four distinct viewpoints.

The ankle AP view provides a top-down perspective of the ankle, while the ankle lateral view offers a side view of the ankle. The standing foot AP view focuses on an over-

Figure 2. Orthopedic imaging perspective.

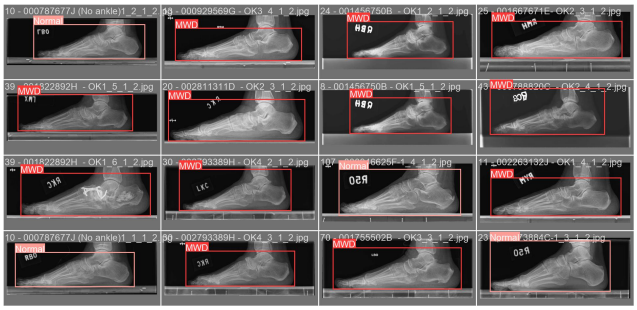

Figure 3. Labeling at verification.

head view of the foot while standing, and the standing foot lateral view presents a side view of the foot while standing. This multi-angle approach helps the model gain a more comprehensive understanding of the condition for accurate classification.

4. Research Result

4.1. Experimental Environment

In terms of the experimental environment, we utilized the following computer hardware and software configurations: The computer hardware comprised an Intel(R) Core(TM) i7-9800X CPU @ 3.80GHz CPU, 64GB of RAM, and a GeForce RTX 2080 Ti GPU. For the operating system, we employed Ubuntu Linux 20.04 LTS. In the software aspect, we used the Anaconda platform as our development environment and conducted model training and experiments using torch version 1.9.0+cu111 and the Python 3.9 programming language. The choice of these hardware and software configurations ensured we had a robust and stable environment to train and evaluate our image detection models efficiently.

During the model training phase, we employed the YOLOv8m model, which has 25.9 million parameters. We set the resolution of the images to 640x640 and chose a batch size of 20 for each training iteration. A larger batch size can enhance training stability but also increase computational demands. Therefore, considering the computing capabilities of the graphics card, we selected 20 as the batch size. The learning rate dictates the speed at which the model updates during each iteration, and setting it correctly can make the loss curve smoother. We initially set the learning rate to 0.01 and utilized a learning rate decay factor of 0.005. The model was trained for 500 iterations, and we used Momentum Stochastic Gradient Descent (Momentum SGD) with a momentum factor of 0.937 for parameter updates. During the training process, we saved the best-performing model on the validation dataset and the model from the final training iteration. To assess the model's performance, we used object detection accuracy as the criterion for selecting the best model. We utilized the validation dataset to validate the model's performance and conducted a detailed analysis of the experimental results.

After completing the training and evaluating the model's performance on the test set, we also employ the Grad-CAM interpretable heatmap model to understand the basis for the MWD diagnosis. With these settings, we aim to achieve efficient detection of MWD and generate interpretable heatmaps, further enhancing the reliability and practicality of our research findings.

4.2. Experimental Result

Fig. 3 represents the original labels, while **??** showcases the predictions results of the model's validation after training. By comparing these two sets of images, it is evident that the outcomes of the training process are quite promising.

Fig. 5 presents several metrics during the training and validation iterations over 500 epochs, including loss, precision, recall, and mAP (mean Average Precision). The solid blue line represents the training and validation results, while the dashed yellow line represents the smoothed results. Regarding loss, there is a rapid decrease in the initial 50 epochs, followed by a relatively stable trend around the 400th epoch. In terms of precision, it reaches a plateau at approximately the 100th epoch. As for recall, there is a sudden drop around the 50th epoch, followed by a gradual in-

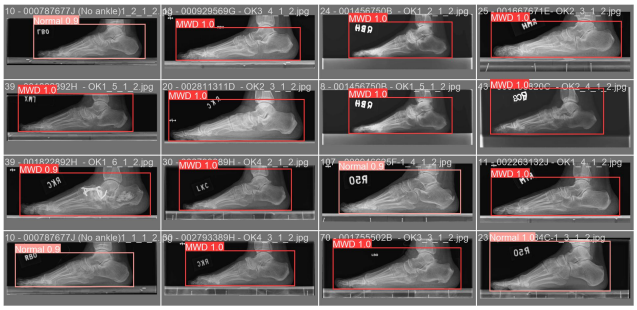

Figure 4. Verification phase model predictions result.

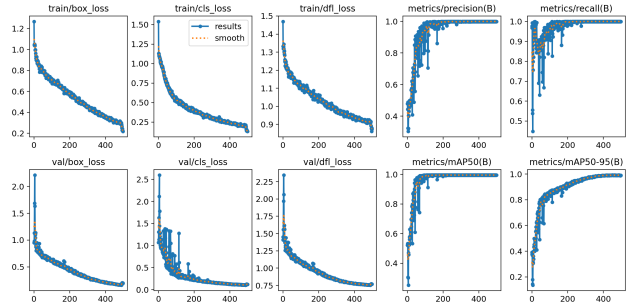

Figure 5. Verification chart overview.

crease after the 100th epoch, reaching a plateau around the 150th epoch. Regarding mAP, mAP50 stabilizes at around the 100th epoch, while mAP95 stabilizes at around the 400th epoch. These changes in metrics reflect the model's performance during training and validation. Despite some fluctuations in the early epochs, continuous training gradually stabilizes the model's performance, ultimately reaching a certain level of performance.

In Fig. 6, the same patient was photographed from four different perspectives. The left half of each view shows the detection results produced by the model, and it can be observed that the model correctly identifies the category and accurately marks the correct position regardless of the viewpoint. The right half of each perspective represents the results of generating heatmaps using Grad-CAM. The heatmap's color intensity, with red indicating higher importance, reveals that regardless of the viewpoint, the model can accurately identify the foot bone's position.

In Fig. 7, the same patient was photographed from four different perspectives. The left half of each view shows the detection results produced by the model, and it can be observed that the model correctly identifies the category (Normal) and accurately marks the correct position regardless of the viewpoint. The right half of each perspective represents the results of generating heatmaps using Grad-CAM. Compared to the MWD patients, the heatmaps for the Normal group show a much smaller area of focus.

Fig. 8 represents the confusion matrix for the test

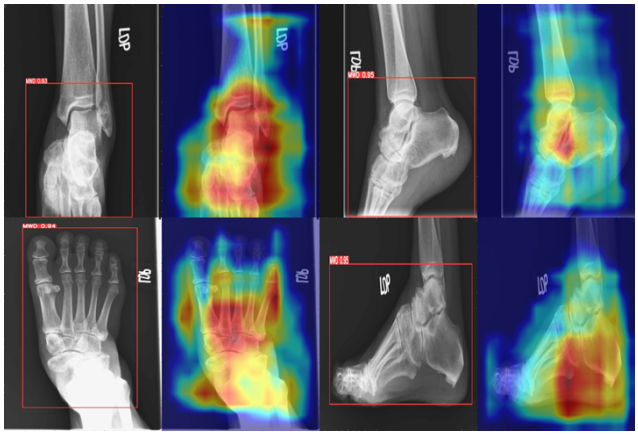

Figure 6. The heatmaps of detection of MWD diseases.

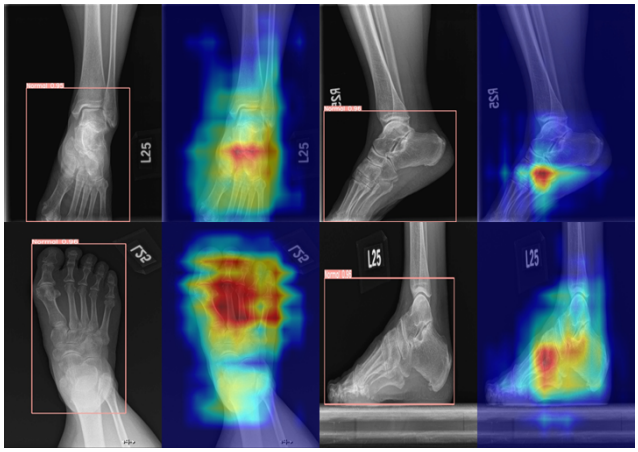

Figure 7. The heatmaps of detection of Normal diseases.

dataset's detection results. In the matrix, if the detection model is not boxed, it is classified as background. If it is boxed, it is classified as either MWD or Normal. On the left side of the matrix are the actual categories, with 18 images of MWD 32 images of Normal, and no background images. Below the matrix are the predicted categories by the model. The model correctly classified a total of 45 images.

Tab. 1 presents the evaluation metrics for the YOLOv8m model on the test dataset. The model achieves an accuracy of over 90%, with an F1-score of 89.23% and a mean Average Precision (mAP) of 92.31%. Overall, the model demonstrates a good level of generalization.

Model	Accuracy	F1-Score	mAP
YOLOv8m	0.9	0.8923	0.9231

Table 1. Test set evaluation metrics.

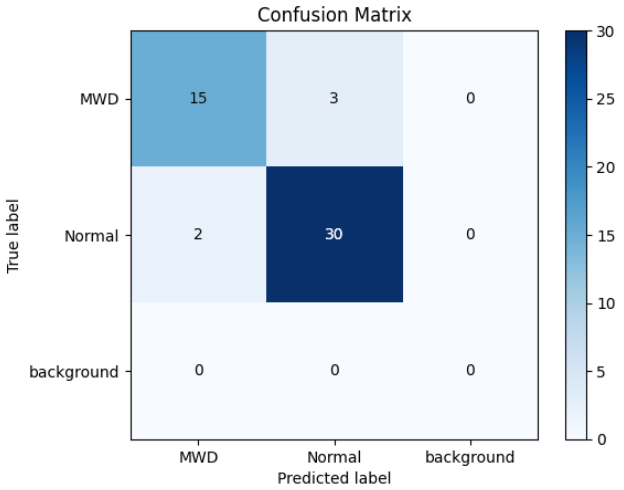

Figure 8. Test set confusion matrix.

5. Conclusions

This study introduces an MWD detection model based on YOLOv8, which has achieved satisfactory results on the test dataset. This indicates that our model possesses high accuracy and generalization capabilities. By utilizing Grad-CAM to generate heatmaps, we can visualize the model's decision-making process, making it easier for physicians and medical students to understand the basis of the model's judgments and enhancing the interpretability of medical image interpretation.

The findings of this study hold significant value in the field of medicine and have the potential to play a crucial role in clinical practice. For diseases with variations in diagnosis between novice and experienced physicians, our model can serve as an objective and accurate assistive tool, enabling healthcare professionals to make diagnoses more swiftly and accurately. Additionally, it can enhance the learning efficiency of medical students during their training, making them more proficient in observing and interpreting medical images.

However, we also acknowledge that there are some limitations to this study. The current model may exhibit misclassifications in specific situations, mainly when dealing with lower-quality images or subtle pathological changes. Therefore, in the future, we plan to improve the model further to enhance its adaptability and robustness under various conditions. Additionally, we aim to expand the training dataset to encompass a broader range of MWD diseases and image qualities to bolster the model's generalization capabilities further.

In summary, this study provides a practical and interpretable solution for automatically detecting MWD diseases, opening up new possibilities in medical imaging. The research outcomes have the potential to significantly improve the accuracy and efficiency of medical diagnoses, advance medical education, and enhance the quality of patient care. We look forward to the insights and inspirations this study may bring to future medical imaging and clinical practice research.

References

[1] Glenn Jocher. Ultralytics yolov8. https://github.com/ultralytics/ultralytics, 2023. 2

[2] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 2

[3] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016. 2