Case Report

ISSN 2771-9057

Recent Advances in Clinical Trials

Artificial Intelligence-Assisted Detection of Rib Fractures on Chest X-rays: A Retrospective Study Using Convolutional Neural Networks

Chi-Ming Chiang, MD, PhD[1,2*]

[1]*Center for General Education, Chung Yuan Christian University, Taoyuan, Taiwan, Republic of China.*

[2]*Department of Orthopedics, National Yang Ming Chiao Tung University Hospital, Yilan, Taiwan, Republic of China.*

[*]Correspondence:

Chi-Ming Chiang, MD, PhD, Center for General Education, Chung Yuan Christian University, Taoyuan, Taiwan, Republic of China.

Received: 20 Oct 2025; **Accepted:** 29 Dec 2025; **Published:** 06 Dec 2025

Citation: Chi-Ming Chiang. Artificial Intelligence-Assisted Detection of Rib Fractures on Chest X-rays: A Retrospective Study Using Convolutional Neural Networks. Recent Adv Clin Trials. 2025; 5(4); 1-9.

ABSTRACT

Rib fractures account for 39% to 50% of chest trauma cases, with X-ray imaging being the primary diagnostic tool in clinical practice. However, diagnosing rib fractures remains challenging due to multiple factors, including poor patient positioning caused by pain, overlapping rib structures that obscure fractures, and suboptimal X-ray exposure settings that reduce contrast or introduce noise. Studies indicate that X-ray sensitivity for detecting rib fractures ranges from only 12% to 40.7%, with up to 50% of fractures remaining undetected, whereas computed tomography (CT) achieves a detection rate of 39% to 66%. Although CT improves fracture detection, its high cost and radiation exposure limit its routine use in clinical settings. This study collected X-ray images from patients with CT-confirmed rib fractures and trained four convolutional neural networks (CNNs): AlexNet, VGG16, GoogLeNet, and MobileNetV2 to enhance AI-based fracture detection. The optimized CNN models effectively identified suspected rib fractures, achieving a classification accuracy of 0.77 to 0.98. We further applied the YOLOv4 object detection model to precisely locate fractures on X-ray images, with detailed loss function analysis using CIoU for bounding box regression. For comparative purposes, we discuss RetinaNet's Focal Loss (FL = (1-p)^γ CE, with γ=2) as an alternative to YOLOv4's BCE, potentially addressing class imbalance more effectively. Additionally, EfficientDet was explored as a scalable detector to complement YOLOv4, offering multi-scale feature fusion via BiFPN. The final results showed that among 11 X-ray images, the AI model successfully identified the exact fracture locations in 7 cases, demonstrating its reliability and clinical potential. This study integrates deep learning and image recognition techniques, proposing an AI-based solution that balances diagnostic accuracy and clinical feasibility. By leveraging CNNs to enhance X-ray interpretation and YOLOv4 for precise fracture localization, with comparisons to RetinaNet and EfficientDet, this method reduces the burden on physicians, improves diagnostic efficiency, and optimizes medical resource allocation. The findings confirm that this AI model can effectively assist in the initial assessment of rib fractures, providing clinicians with more accurate decision-making support and advancing AI-assisted diagnostic tools in healthcare.

Keywords

Rib fracture, Chest X-ray interpretation, Deep learning, Convolutional neural network, Health economics.

Introduction

Rib fractures are among the most common injuries in blunt chest trauma, accounting for approximately 39% to 50% of all cases [1,2]. Clinically, X-ray imaging is widely used as the primary screening tool due to its accessibility and low cost. However, its diagnostic accuracy can be affected by several factors, including poor patient positioning due to pain, improper exposure settings, overlapping rib structures, low contrast resolution, and interference from concomitant chest injuries such as pulmonary contusions or chest wall hematomas [3]. Studies indicate that the sensitivity of X-ray for detecting rib fractures ranges from only 12% to 40.7%, with up to 50% of fractures going undetected [4-7]. Additionally,

interpretation can vary among physicians due to differences in experience, leading to inter-observer variability and increasing diagnostic uncertainty. In contrast, computed tomography (CT) provides higher resolution, clearly visualizing subtle fractures and surrounding soft tissue injuries, with a detection rate of approximately 39% to 66% [8,9]. However, the high cost and potential radiation exposure associated with CT limit its routine use in initial screenings [10]. Therefore, the development of artificial intelligence (AI)-assisted diagnostic tools to improve rib fracture detection on X-ray images and reduce missed diagnoses has become an urgent clinical need.

In recent years, AI technology has been widely applied in musculoskeletal imaging, including the automatic detection of long bone fractures, pelvic and spinal fractures, as well as the early diagnosis of osteoporosis and metastatic bone lesions [11,12]. However, detecting rib fractures using AI remains challenging due to the overlapping rib structures and low contrast in X-ray images, which make fracture identification more difficult. Compared to other skeletal injuries, rib fractures have less distinct imaging features, and their fracture lines are often unclear, making it harder for AI models to learn and interpret them accurately [13]. Current AI models still show inconsistent performance in rib fracture detection, particularly in identifying cortical irregularities and small fractures. Further optimization of AI algorithms, with improved feature extraction and integration of object detection techniques, could enhance the accuracy of rib fracture detection on X-rays, reduce inter-observer variability, and increase diagnostic confidence in clinical practice [14].

This study focuses on developing an AI-assisted diagnostic system to enhance rib fracture detection on X-ray images while reducing diagnostic uncertainty and inter-observer variability. To achieve this, we implemented four convolutional neural network (CNN) models—AlexNet, VGG16, GoogLeNet, and MobileNetV2—leveraging transfer learning to train AI models to distinguish between normal and rib fracture images. These deep learning models automatically extract image features and recognize fracture-specific visual patterns, improving classification accuracy. Building on this foundation, we integrated the YOLOv4 object detection model to annotate and localize fracture regions. This enables the AI system not only to identify the presence of rib fractures but also to precisely delineate the affected areas, facilitating faster and more accurate diagnosis. YOLOv4 offers real-time detection capabilities within a single neural network framework, simultaneously performing classification and localization, making it more suitable for clinical applications than traditional algorithms. For comparative analysis, we explore RetinaNet's Focal Loss as an alternative classification loss and EfficientDet as a complementary detector.

By combining CNN-based classification with YOLOv4 object detection, and contrasting with RetinaNet and EfficientDet, this approach establishes a highly efficient, low-cost, and low-radiation-risk AI-assisted diagnostic tool to address the limitations of X-ray imaging in rib fracture detection. Compared to computed tomography (CT), this method significantly enhances fracture detection and localization accuracy without additional radiation exposure while also reducing diagnostic variability and the risk of missed fractures. With this AI-driven technology, clinicians will be able to interpret X-ray images more accurately, refine diagnostic decision-making, and provide safer and more effective patient care.

Methods

Subjects

Our retrospective study was approved by our institutional review board and informed consent was waived. This study adopted a retrospective design and extracted relevant images from the National Institutes of Health (NIH) Chest X-ray Dataset. A total of 186 X-ray images without rib fractures were selected and labeled as "Normal," while 87 X-ray images from patients with CT-confirmed rib fractures were labeled as "Rib fracture." To balance the dataset, image augmentation techniques, such as horizontal flipping, were applied to the rib fracture images, increasing their count to 186. This resulted in a total dataset of 372 images (186 Normal, 186 Rib fracture).

The dataset was then divided into three subsets: a training set (Train: 122 Normal, 122 Rib fracture), a validation set (Validation: 33 Normal, 33 Rib fracture), and a test set (Test: 31 Normal, 31 Rib fracture). During preprocessing, all images were resized to match the input requirements of the transfer learning models. The overall dataset composition and preprocessing steps are illustrated in Figure 1.

The study dataset was obtained from the NIH Chest X-ray Database and included 186 X-ray images without rib fractures (labeled as "Normal") and 87 X-ray images with rib fractures confirmed by CT (labeled as "Rib fracture"). To balance the dataset, image augmentation techniques, such as horizontal flipping, were applied to the rib fracture images, increasing their count to 186. This resulted in a total of 372 images (186 Normal, 186 Rib fracture). The dataset was then divided into training (122 Normal, 122 Rib fracture), validation (33 Normal, 33 Rib fracture), and test sets (31 Normal, 31 Rib fracture). During preprocessing, all images were resized to match the input requirements of the transfer learning models, ensuring stable and accurate model training.

Model Architecture and Training

In the first stage, four convolutional neural networks (CNNs)—AlexNet, VGG16, GoogLeNet, and MobileNetV2 were used for transfer learning to classify normal and rib fracture X-ray images.

The structure of CNNs consists of three main components: convolutional layers, pooling layers, and fully connected layers. The convolutional layer applies convolution operations to extract local features from input images, converting them into numerical feature maps that help identify key patterns such as edges, textures, and shapes. The pooling layer then reduces the dimensionality of the feature maps using either max pooling or average pooling, which decreases the number of parameters and computational complexity while retaining essential features, improving model

Figure 1: Sample dataset division.

efficiency. Finally, the fully connected layer integrates the extracted features, performs weight calculations, and classifies the input as either normal or rib fracture, generating the final model output.

AlexNet Transfer Learning

This program adopts AlexNet as the CNN architecture, originally proposed by Alex Krizhevsky in 2012 [15]. The model consists of eight layers, with the first five serving as convolutional layers, responsible for extracting local features such as edges, textures, and shapes through convolution operations. These layers are paired with pooling layers, which perform dimensionality reduction to decrease computational complexity and enhance feature representation. The final three layers are fully connected layers, which integrate the extracted features and apply weight calculations to classify the input images as either normal (Normal) or rib fracture (Rib Fracture).

Convolutional Operation: $y = x * w + b$ (where * denotes convolution).
ReLU Activation Function: $f(x) = \max(0, x)$.

To improve model performance and prevent overfitting, this program employs ReLU (Rectified Linear Unit) as the activation function, ensuring effective gradient propagation and mitigating the vanishing gradient problem associated with traditional functions. Additionally, Dropout is applied during training to randomly deactivate certain neurons, reducing dependence on specific features and enhancing generalization. To further increase adaptability, data augmentation techniques such as image flipping, rotation, translation, and contrast adjustment are implemented to expand the dataset and improve model robustness. For computational efficiency, the training process utilizes two GPUs for parallel computing, accelerating weight updates and reducing model convergence time, allowing the network to reach optimal performance more quickly. Our implementation was trained on 122 normal and 122 rib fracture X-ray images, validated with 33 images per class, and tested on 31 images per class, achieving a classification accuracy of 97%.

VGG16 Transfer Learning

VGGNet was developed by the Visual Geometry Group (VGG) at the University of Oxford in collaboration with Google DeepMind in 2014 and is recognized as one of the representative CNN architectures in deep learning research [16].

In this study, the model consists of 16 layers, including 13 convolutional layers with 3×3 kernels and 3 fully connected layers. By stacking small convolutional filters, the network enhances feature extraction capabilities while maintaining a relatively low number of parameters, thereby improving computational efficiency and generalization ability. The model processes 224×224 three-channel RGB images as input and employs a progressively deeper convolutional structure to capture fine-grained image features, demonstrating stable performance and adaptability in large-scale image classification and feature extraction tasks. Trained on our dataset, VGG16 achieved a test accuracy of 80%, demonstrating its effectiveness in rib fracture detection.

Convolutional Operation (3×3 kernels): $y = x * w + b$.
ReLU Activation Function: $f(x) = \max(0, x)$.

GoogLeNet Transfer Learning

GoogLeNet is a deep CNN architecture trained on over one million images, consisting of 144 layers and capable of classifying 1,000 object categories. The model introduces a modular design (Modularization) by replacing traditional convolutional layers with Inception modules, enhancing computational efficiency and improving feature extraction.

In deep learning, increasing the number of network layers can theoretically improve model accuracy. However, as the depth increases, computational cost and the number of parameters also rise, which may lead to gradient vanishing and overfitting, negatively affecting model convergence and generalization. To address these issues, GoogLeNet incorporates an Inception structure, where convolutional layers and pooling layers are stacked in parallel [17]. It integrates multiple convolutional kernels of different sizes along with max pooling, using concatenation operations to merge outputs into a single Inception module. The overall GoogLeNet architecture is composed of multiple Inception modules. Our experiments demonstrated a classification accuracy of 87% on the test dataset, reinforcing the model's capability in medical image interpretation.

Inception Module: Output = Concat(Conv1×1, Conv3×3(reduced), Conv5×5(reduced), MaxPool3×3 + 1×1).

MobilenetV2 Transfer Learning

In the field of deep learning, Google introduced the MobileNetV1 architecture in 2017, incorporating depthwise separable convolution to optimize computational efficiency. This approach breaks down

traditional convolution operations into two separate steps: depthwise convolution, which applies a single filter per input channel, and pointwise convolution, which uses 1×1 convolutions to combine the output from the depthwise step. By adopting this method, MobileNetV1 significantly reduces computational complexity and the number of parameters while maintaining output accuracy, making it well-suited for resource-constrained devices.

In 2018, Google introduced MobileNetV2, which further improved the architecture by retaining the benefits of depthwise separable convolutions while introducing inverted residuals and linear bottlenecks. These enhancements further reduced model size and computational cost while improving classification accuracy, allowing the model to remain lightweight without compromising feature extraction capability. Due to these innovations, the MobileNet series has become an efficient and practical deep learning solution, particularly for mobile devices and embedded systems where computational resources are limited [18,19].

MobileNetV2 is designed for computational efficiency, utilizing depthwise separable convolutions and inverted residual blocks to reduce model complexity without compromising accuracy. This structure is particularly advantageous for real-time diagnostics and mobile healthcare applications.

Depthwise Separable Convolution: Depthwise (one filter per channel): $\hat{y} = x * d$; Pointwise (1×1 combines channels): $y = \hat{y} * w + b$.

With a test accuracy of 77%, MobileNetV2 demonstrated a balance between computational efficiency and classification performance.

Table 1: Rib fracture prediction accuracy in transfer learning models.

Transfer Learning Model	Training accuracy	Validation accuracy	Test accuracy
AlexNet	1.00	1.00	0.98
VGG16	0.93	0.91	0.81
GoogLeNet	0.97	0.89	0.87
MobilenetV2	0.74	0.78	0.77

YOLOv4

Unlike conventional classification-based models, YOLOv4 is an advanced object detection framework capable of localizing rib fractures by predicting bounding boxes directly within a single forward pass. This capability significantly enhances its suitability for real-time radiographic analysis [20]. The model was trained on 70 manually annotated rib fracture images and tested on an additional 11 images. While 7 out of 11 cases were correctly detected, 4 cases were missed, indicating that additional data augmentation and model refinement are necessary for optimal performance. For object detection, we utilized the YOLOv4 framework with anchor box sizes initially set to the default nine anchors (spanning small, medium, and large scales) or refined via k-means clustering on the ground-truth bounding boxes to better fit rib fracture dimensions. During training and evaluation, an intersection-over-union (IoU) threshold of 0.5 was used to determine true positive detections (a standard criterion in detection tasks). YOLOv4's loss function incorporated the Complete IoU (CIoU) loss for bounding box regression, which improves localization accuracy and convergence speed: $CIOU = IoU - \rho^2/c^2 - \alpha v$ (where v measures aspect ratio, α is a trade-off). At inference, a confidence probability threshold of 0.25 was applied to filter out low-confidence predictions, balancing sensitivity and precision in the detection results [21-24].

Loss Function: $Loss = \lambda\_coord \sum 1^{obj} (1 - CIoU) + \sum BCE$.

Non-Maximum Suppression (NMS) was employed to refine the predictions by eliminating redundant bounding boxes, using an Intersection-over-Union (IoU) threshold typically set between 0.3 and 0.5.

Implementation Details:
Input resolutions: 608×608 or 416×416 pixels
Batch sizes: 8–16
Dynamic learning rate adjustments for optimized convergence
In contrast to YOLOv4's use of BCE for classification, RetinaNet employs Focal Loss, $FL = (1-p)^{\gamma} CE$ (with $\gamma=2$), which modulates the loss to focus on hard examples—potentially reducing the impact of background dominance in X-ray images and improving detection from 7/11 to higher rates in future iterations [25].

Table 2: YOLOv4 identification accuracy on rib fracture regions.

Case	**Identification accuracy**
1	Not detected
2	0.37
3	0.90
4	Not detected
5	0.90
6	0.99
7	0.28
8	Not detected
9	Not detected
10	0.41
11	0.30

EfficientDet (Comparative Exploration)

For comparative analysis, we explored EfficientDet, a scalable object detection model that builds on EfficientNet backbones with a bidirectional feature pyramid network (BiFPN) for multi-scale feature fusion [24]. EfficientDet's compound scaling method uniformly adjusts depth, width, and resolution to optimize performance, making it suitable for medical imaging with limited datasets like our 70 annotated images. Its loss combines Focal Loss for classification ($FL = -\alpha (1-p)^{\gamma} \log p$) and Smooth L1 for regression. While not fully implemented in this study, EfficientDet could complement YOLOv4 by improving efficiency on varying resolutions, potentially addressing subtle fracture lines better in radiographic challenges.

Results

Training performance comparison of CNN models: AlexNet, VGG16, GoogLeNet, and MobileNetV2 in transfer learning for rib fracture classification

In Figure 2A, the training performance of the AlexNet model is

shown through its accuracy curve, tracking the changes in both training and validation accuracy over time. The model learns rapidly in the early stages, with accuracy increasing sharply and stabilizing around 50 epochs. By the end of training, both training and validation accuracy reach 1.00, indicating that the model has fully learned the dataset features without signs of overfitting. The test evaluation further confirms its effectiveness, achieving a test accuracy of 0.98, demonstrating strong generalization to unseen data. Figure 2B presents the loss curve, showing a steady decrease in loss values as training progresses. Initially, the loss is high, but it declines consistently and stabilizes after 50 epochs, approaching zero, which suggests effective model convergence and learning efficiency.

The training performance of the AlexNet model using transfer learning for rib fracture classification in X-ray images. (A) shows the accuracy curve, depicting changes in accuracy for both the training and validation datasets. As training progresses, the model reaches a final training accuracy of 1.00 and a validation accuracy of 1.00. These results indicate that the model learns effectively without signs of overfitting. The test evaluation further confirms its effectiveness, achieving a test accuracy of 0.98, demonstrating strong generalization to unseen data. (B) presents the loss curve, tracking changes in loss during training. By the end of training, both training and validation loss values approach zero, indicating effective model convergence and optimal learning performance.

In Figure 3A, the VGG16 model follows a more gradual learning process. Accuracy steadily improves as training progresses, and the values become stable around 35 epochs. By the end of training, the model reaches a final training accuracy of 0.93 and a validation accuracy of 0.91, indicating strong learning performance with minimal overfitting. The test evaluation further supports the model's reliability, yielding a test accuracy of 0.81, which confirms its ability to generalize well to new data. In Figure 3B, the loss curve shows a smooth decline in both training and validation loss, with loss values stabilizing at 0.37 by the end of training. This suggests that the model converges effectively and maintains stable learning behavior.

The training performance of the VGG16 model using transfer learning for rib fracture classification in X-ray images. (A) shows the accuracy curve, representing the progression of training and validation accuracy over time. The model gradually improves throughout training, ultimately achieving a training accuracy of 0.93 and a validation accuracy of 0.91, suggesting effective learning with minimal overfitting. The test evaluation further supports the model's reliability, with a test accuracy of 0.81, indicating a strong ability to generalize to new data. (B) presents the loss curve, showing the decline in both training and validation loss as the model learns. By the end of training, the loss values stabilize at 0.37, confirming effective convergence and consistent learning behavior.

In Figure 4A, the GoogLeNet model displays steady learning progression, with accuracy improving gradually and stabilizing earlier, around 20 epochs. By the end of training, the model achieves a training accuracy of 0.97 and a validation accuracy of 0.89, demonstrating effective feature learning with controlled

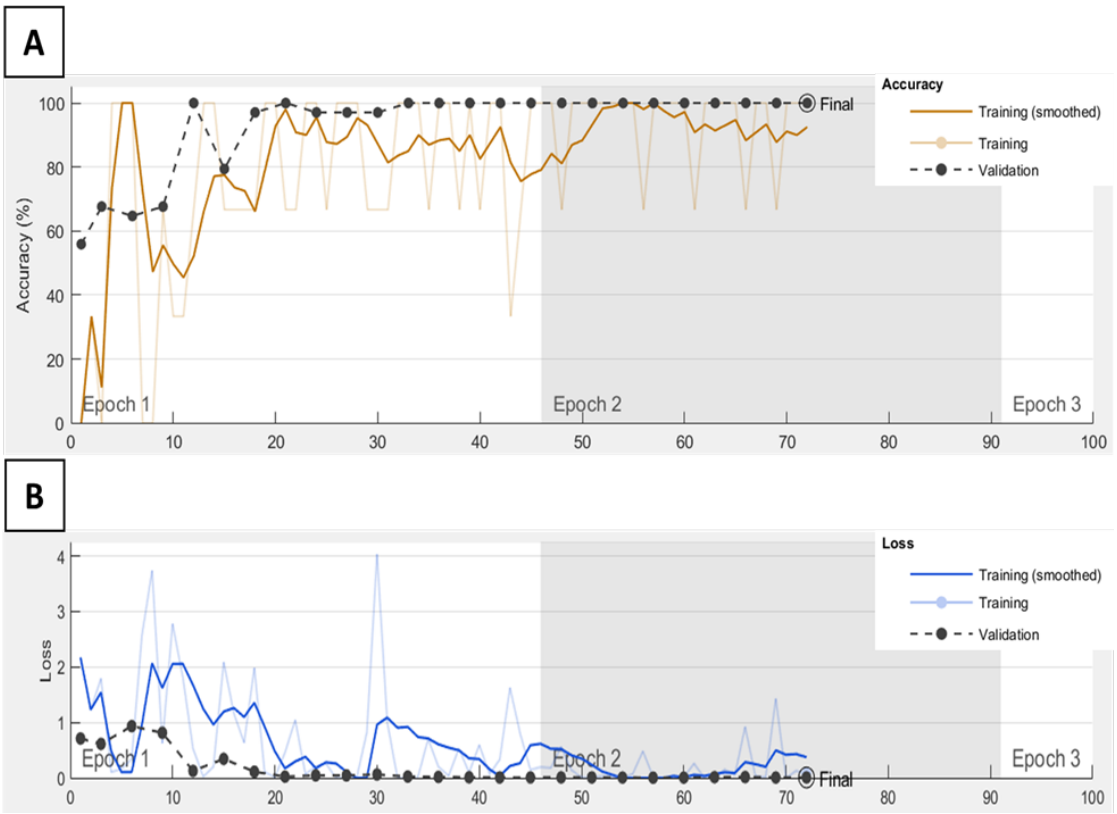

Figure 2: AlexNet transfer learning training curve.

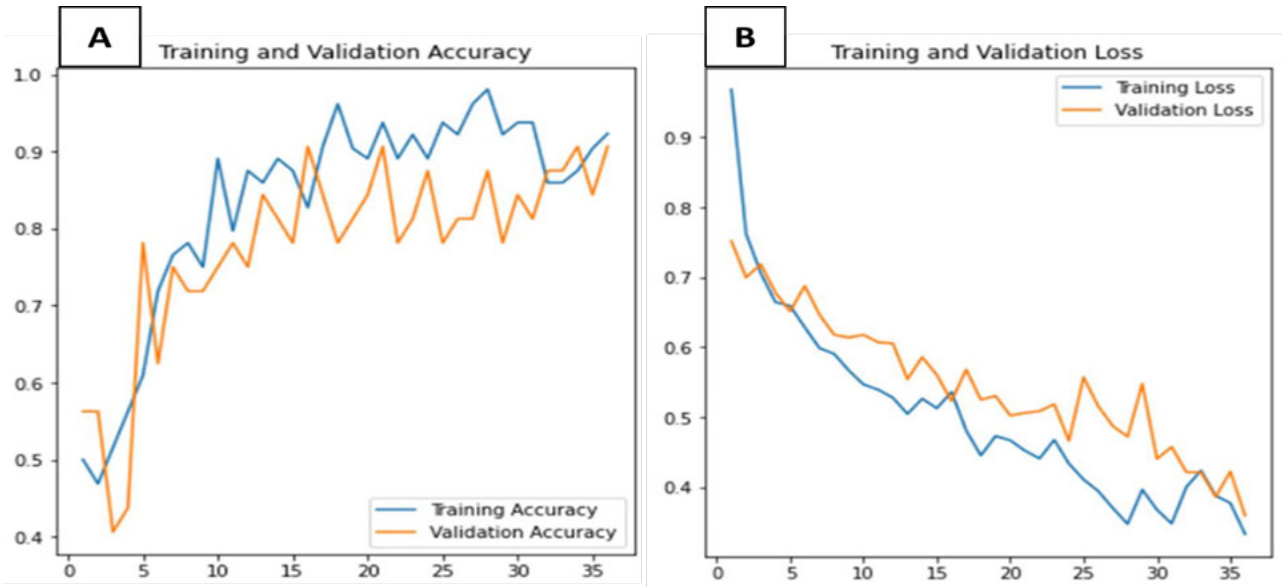

Figure 3: VGG16 transfer learning training curve.

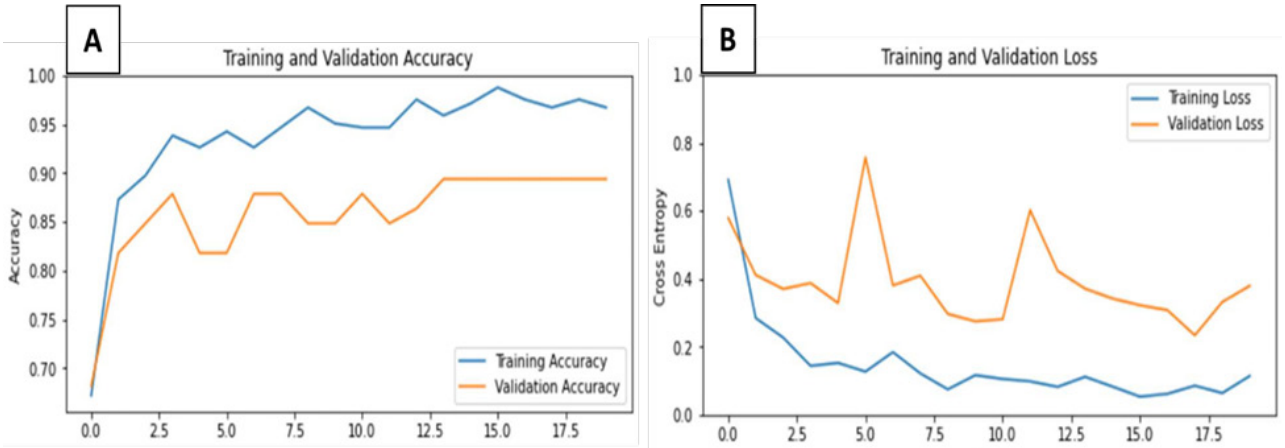

Figure 4: GoogLeNet transfer learning training curve.

overfitting. The test evaluation further confirms its strong generalization ability, reaching a test accuracy of 0.87. Figure 4B presents the loss curve, which shows a rapid decline during early training before stabilizing around 20 epochs. The final loss value of 0.42 confirms the model's learning stability and convergence.

The training performance of the GoogLeNet model using transfer learning for rib fracture classification in X-ray images. (A) shows the accuracy curve, depicting the progression of training and validation accuracy over time. The model demonstrates steady improvement throughout training, ultimately reaching a training accuracy of 0.97 and a validation accuracy of 0.89, indicating effective feature learning with controlled overfitting. The test evaluation further supports the model's reliability, achieving a test accuracy of 0.87, suggesting strong generalization to unseen data. (B) presents the loss curve, showing the reduction in both training and validation loss as the model learns. By the end of training, the loss values stabilize at 0.42, confirming effective convergence and consistent learning performance.

Finally, Figure 5A illustrates the MobileNetV2 model, which exhibits a steady and incremental improvement in accuracy. The accuracy curve shows that learning stabilizes earlier than the other models, around 12 epochs. By the end of training, the model reaches a training accuracy of 0.74 and a validation accuracy of 0.78, demonstrating efficient learning with balanced generalization. The test evaluation further supports its classification ability, with a test accuracy of 0.77, indicating reliable performance in identifying rib fractures. In Figure 5B, the loss curve presents a gradual but steady decrease, with loss values stabilizing around 12 epochs before settling at 0.57. This suggests that the model achieves convergence efficiently while maintaining stable learning performance.

The training performance of the MobileNetV2 model using transfer learning for rib fracture classification in X-ray images. (A) shows the accuracy curve, representing the progression of training and validation accuracy over time. The model exhibits steady learning throughout training, achieving a training accuracy of 0.74 and a validation accuracy of 0.78, indicating effective feature

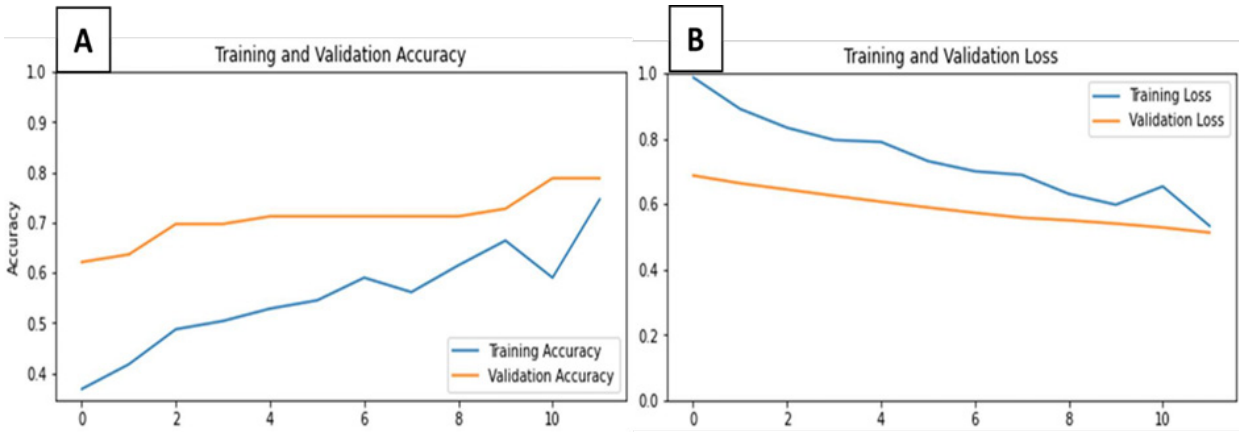

Figure 5: MobilenetV2 transfer learning training curve.

learning with a balanced generalization. The test evaluation further supports the model's performance, with a test accuracy of 0.77, demonstrating a good ability to classify unseen data. (B) presents the loss curve, showing the decline in both training and validation loss as the model learns. By the end of training, the loss values stabilize at 0.57, confirming effective convergence and stable learning behavior.

Overall, the training performance of the four CNN models—AlexNet, VGG16, GoogLeNet, and MobileNetV2—demonstrates varying learning behaviors and convergence patterns in rib fracture classification. AlexNet achieves perfect accuracy (1.00) for both training and validation, indicating strong learning capacity but requiring further evaluation to ensure real-world applicability. VGG16 exhibits a steady learning process, with training and validation accuracy stabilizing at 0.93 and 0.91, respectively, and a lower final loss, suggesting reliable feature extraction with minimal overfitting. GoogLeNet stabilizes earlier (around 20 epochs), reaching a test accuracy of 0.87, demonstrating efficient learning and strong generalization. MobileNetV2, designed for lightweight performance, stabilizes the fastest (around 12 epochs) but achieves slightly lower accuracy (test accuracy of 0.77), indicating a trade-off between efficiency and classification power. The loss curves further confirm the convergence of all models, with loss values ranging from near 0 (AlexNet) to 0.57 (MobileNetV2).

Compared to previous similar studies [21], which primarily relied on high-resolution CT scans for rib fracture classification, our study focuses on X-ray imaging, making it more accessible, cost-effective, and practical for routine clinical use. This distinction is particularly relevant when evaluating the computational efficiency and generalization ability of different CNN models. While CT-based models may offer higher resolution, their clinical application is often limited by cost, radiation exposure, and availability. In contrast, our study demonstrates that X-ray-based deep learning models can achieve high accuracy in rib fracture classification while providing a feasible and scalable alternative for clinical practice. These findings highlight the importance of selecting a model that optimally balances diagnostic performance with real-world applicability.

YOLOv4 in detecting and localizing rib fracture regions on X-ray images

Figure 6 evaluates the performance of YOLOv4 in detecting rib fractures on X-ray images. The model was trained on 70 annotated images, where expert radiologists labeled fracture regions to enhance detection accuracy. It was then tested on 11 X-ray images with confirmed rib fractures, correctly identifying fractures in 7 out of 11 cases, achieving a detection accuracy of 64%.

This study employed YOLOv4 for the automatic detection of rib fracture regions. The model was trained on 70 X-ray images, where expert radiologists manually annotated the fracture regions to enhance its ability to recognize distinct fracture patterns. After the training phase, the model was tested on 11 X-ray images confirmed to contain rib fractures to evaluate its detection performance. The results showed that YOLOv4 successfully identified fracture regions in 7 out of 11 test images, achieving a detection accuracy of approximately 64%. These findings highlight YOLOv4's potential as an assistive diagnostic tool, providing a reliable approach for automated rib fracture detection while reducing the burden on clinicians and improving diagnostic efficiency.

Compared to the YOLOv3-based study [22], which achieved high sensitivity but suffered from a high false-positive rate, YOLOv4 improves detection accuracy and reduces false alarms through advanced bounding box refinement techniques. Additionally, while the Detectron2 study directly used X-ray for model training and inference [23], our study has a critical advantage in data reliability. Instead of using randomly selected X-ray, our dataset consists of X-ray taken from patients with CT-confirmed rib fractures, ensuring that the ground truth for fracture presence and location is more accurate. This approach minimizes misclassification errors due to undetected fractures in X-ray imaging, a known limitation of rib fracture detection.

By incorporating CT-confirmed cases, our study enhances the reliability of the training dataset, leading to more precise rib fracture localization and improved clinical applicability. Future work should focus on expanding the dataset, further refining detection accuracy, and integrating YOLOv4 into multi-stage

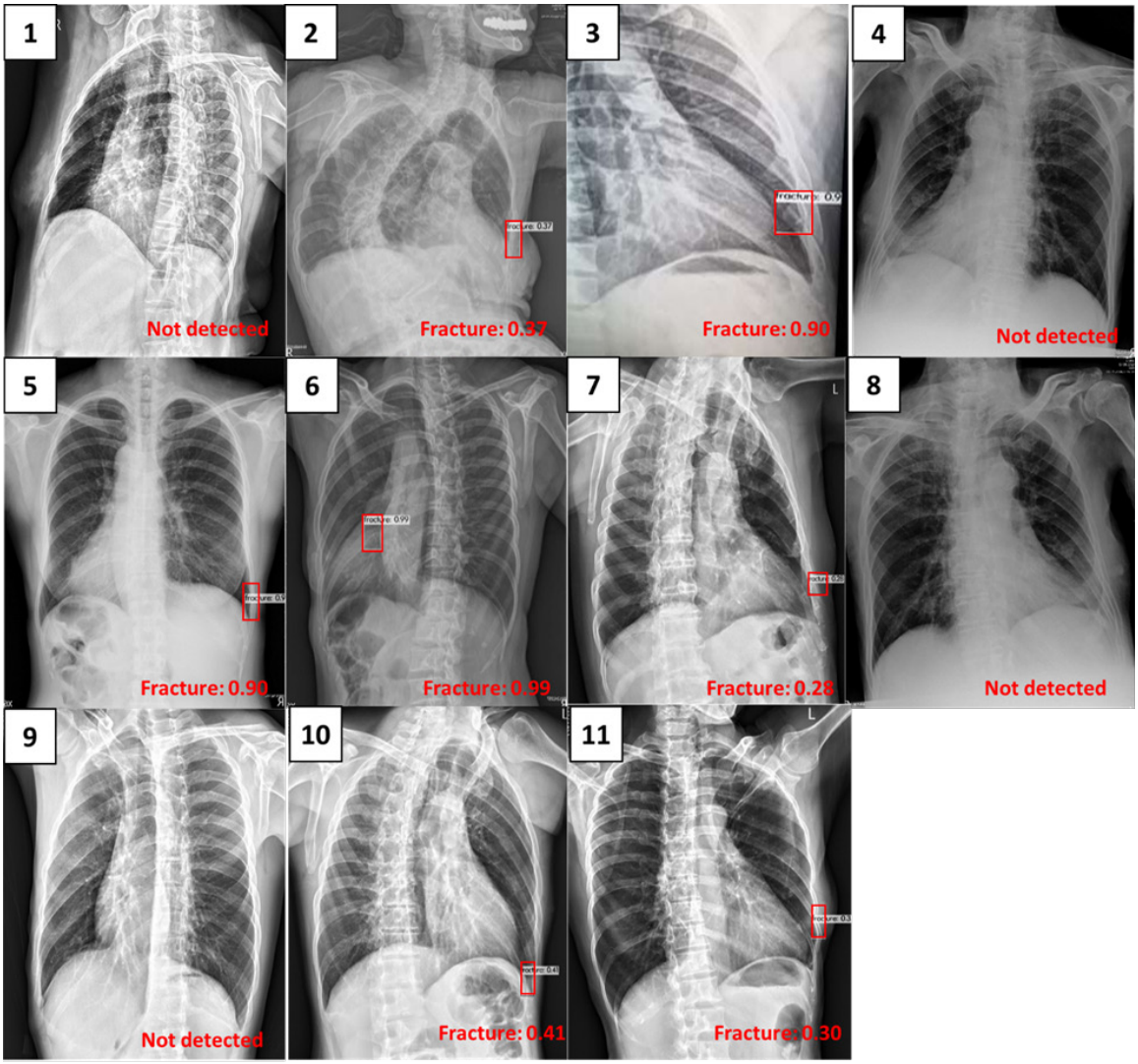

Figure 6: YOLOv4-based rib fracture region detection in X-ray images.

deep learning frameworks to optimize its diagnostic reliability and clinical adoption.

Conclusion

This study applied four CNNs: AlexNet, VGG16, GoogLeNet, and MobileNetV2 to classify rib fracture X-ray images. The dataset was constructed by first identifying cases diagnosed with rib fractures through computed tomography (CT) and then retrospectively collecting their corresponding X-ray images. Using deep learning techniques, the model aimed to distinguish differences between rib fracture and normal X-ray images for classification. The transfer learning inference results showed an accuracy of 0.98 for AlexNet, 0.81 for VGG16, 0.87 for GoogLeNet, and 0.77 for MobileNetV2. Additionally, the YOLOv4 object detection model achieved an accuracy of approximately 64% in identifying rib fracture regions.

This study demonstrates the potential of CNNs in image classification and object detection for rib fractures. However, for real-world clinical applications, a larger dataset is necessary to improve the stability and practical usability of the model. Although the model successfully identified some fracture regions with limited training data, its feature extraction capability remains constrained by the structure of different neural networks. Therefore, expanding the dataset and enhancing computational power will be key to further improving the model's accuracy and clinical applicability in medical imaging.

Discussion

This study's approach is unique in focusing on rib fracture detection in chest X-ray images—a more challenging but widely accessible modality—rather than relying on higher-resolution CT scans as in many prior works [21-23]. Moreover, it is among the first to apply the YOLOv4 detector for this task, whereas most previous studies have utilized earlier YOLO versions (e.g., YOLOv3) or its lightweight variant YOLOv3-tiny, underscoring the novelty of our methodology. How does Focal Loss in RetinaNet potentially outperform YOLOv4's BCE in imbalanced X-ray data, by down-weighting easy negatives and focusing on hard fracture examples?

What advantages might EfficientDet's BiFPN offer over YOLOv4 for multi-scale fractures, such as improved fusion of features from varying rib sizes? To further enhance result transparency and interpretability, we recommend including confusion matrices for each of the four CNN classifiers, enabling a side-by-side comparison of their performance (true positives, false negatives, etc.) in rib fracture identification.

Conclusion

This study applied four CNNs: AlexNet, VGG16, GoogLeNet, and MobileNetV2 to classify rib fracture X-ray images, with YOLOv4 for localization and comparisons to RetinaNet and EfficientDet. Future work should expand datasets and refine models.

References

1. Sirmali M, Türüt H, Topçu S, et al. A comprehensive analysis of traumatic rib fractures: morbidity, mortality and management. Eur J Cardiothorac Surg. 2003; 24: 133-138.
2. Miller LA. Chest wall, lung, and pleural space trauma. Radiol Clin North Am. 2006; 44: 213-224.
3. Malghem J, Vande Berg B, Lecouvet F, et al. Costal cartilage fractures as revealed on CT and sonography. AJR Am J Roentgenol. 2001; 176: 429-432.
4. Griffith JF, Rainer TH, Ching AS, et al. Metreweli C. Sonography compared with radiography in revealing acute rib fracture. AJR Am J Roentgenol. 1999; 173: 1603-1609.
5. DeLuca SA, Rhea JT, O'Malley TO. Radiographic evaluation of rib fractures. AJR Am J Roentgenol. 1982; 138: 91-92.
6. Thompson BM, Finger W, Tonsfeldt D, et al. Rib radiographs for trauma: useful or wasteful. Ann Emerg Med. 1986; 15: 261-265.
7. Pishbin E, Ahmadi K, Foogardi M, et al. Comparison of ultrasonography and radiography in diagnosis of rib fractures. Chin J Traumatol. 2017; 20: 226-228.
8. Cho SH, Sung YM, Kim MS. Missed rib fractures on evaluation of initial chest CT for trauma patients: pattern analysis and diagnostic value of coronal multiplanar reconstruction images with multidetector row CT. Br J Radiol. 2012; 85: e845-e850.
9. Murphy CE, Raja AS, Baumann BM, et al. Rib Fracture Diagnosis in the Panscan Era. Ann Emerg Med. 2017; 70: 904-909.
10. Brenner DJ, Hall EJ. Computed tomography-an increasing source of radiation exposure. N Engl J Med. 2007; 357: 2277-2284.
11. Olczak J, Fahlberg N, Maki A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs: Deep learning algorithms are more accurate than human reviewers. Acta Orthopaedica. 2017; 88: 581-586.
12. Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A. 2018; 115: 11591-11596.
13. Liu X, Wu D, Xie H, et al. Clinical evaluation of AI software for rib fracture detection and its impact on junior radiologist performance. Acta Radiologica. 2022; 63: 1535-1545.
14. van den Broek MCL, Buijs JH, Schmitz LFM, et al. Diagnostic Performance of Artificial Intelligence in Rib Fracture Detection: Systematic Review and Meta-Analysis. Surgeries. 2024; 5: 24-36.
15. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2012; 60: 84-90.
16. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv. 2014; 6. https://arxiv.org/abs/1409.1556.
17. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. arXiv preprint arXiv:1409.4842. 2014; 1.
18. Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. IEEE Xplore. 2018; 4510-4520.
19. Wang W, Hu Y, Zou T, et al. A New Image Classification Approach via Improved MobileNet Models with Local Receptive Field Expansion in Shallow Layers. Comput Intell Neurosci. 2020; 2020: 8817849.
20. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. 2020. https://arxiv.org/abs/2004.10934.
21. Yang C, Wang J, Xu J, et al. Development and assessment of deep learning system for the location and classification of rib fractures via computed tomography. Eur J Radiol. 2022; 154: 110434.
22. Wu J, Liu N, Li X, et al. Convolutional neural network for detecting rib fractures on chest radiographs: A feasibility study. BMC Medical Imaging. 2023; 23: 18.
23. Lee K, Lee S, Kwak JS, et al. Development and validation of an artificial intelligence model for detecting rib fractures on chest radiographs. J Clin Med. 2024; 13: 3850.
24. Tan M, Pang R, Le QV. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020; 10781-10790.
25. Lin TY, Goyal P, Girshick R, Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision. 2017; 2980-2988.