

Leveraging 3D Convolutional Neural Networks for Accurate Recognition and Localization of Ankle Fractures

Hua Wang¹, Jichong Ying², Jianlei Liu², Tianming Yu², Dichao Huang²

¹Department of Medical Imaging, Ningbo No.6 hospital, Ningbo, People's Republic of China; ²Department of Orthopedics, Ningbo No.6 hospital, Ningbo, People's Republic of China

Correspondence: Dichao Huang, Department of Orthopedics, Ningbo No.6 hospital, Ningbo, People's Republic of China, Email hsdichaohuang@163.com

Background: Ankle fractures are common injuries with substantial implications for patient mobility and quality of life. Traditional imaging methods, while standard, have limitations in detecting subtle fractures and distinguishing them from complex bone structures. The advent of 3D Convolutional Neural Networks (3D-CNNs) offers a promising avenue for enhancing the accuracy and reliability of ankle fracture diagnoses.

Methods: In this study, we acquired 1453 high-resolution CT scans and processed them through three distinct 3D-CNN models: 3D-MobileNet, 3D-Resnet101, and 3D-EfficientNetB7. Our approach involved meticulous preprocessing of images, including normalization and resampling, followed by a systematic comparative evaluation of the models based on accuracy, Area Under the Curve (AUC), and recall metrics. Additionally, the integration of Gradient-weighted Class Activation Mapping (Grad-CAM) provided visual interpretability of the models' predictive focus points.

Results: The 3D-EfficientNetB7 model outperformed the other models, achieving an accuracy of 0.91 and an AUC of 0.94 after 20 training epochs. It demonstrated particularly effective in the accurate detection and localization of subtle and complex fractures. Grad-CAM visualizations confirmed the model's focus on clinically relevant areas, aligning with expert assessments and enhancing trust in automated diagnostics. Spatial localization techniques were pivotal in improving interpretability, offering clear visual guidance for pinpointing fracture sites.

Conclusion: Our findings highlight the effectiveness of the 3D-EfficientNetB7 model in diagnosing ankle fractures, supported by robust performance metrics and enhanced visualization tools.

Keywords: 3d convolutional neural networks, ankle fractures, diagnostic accuracy, Grad-CAM, image processing, machine learning, medical imaging, orthopedic diagnostics

Introduction

Background

Ankle fractures account for a considerable proportion of skeletal injuries, representing approximately 9% of all fracture types, and have become an increasingly significant concern in public health arenas.¹ Ankle fractures, common injuries to the ankle joint involving the tibia, fibula, and talus, include several specific types such as pilon fractures, which affect the distal tibial region and its weight-bearing surface; bimalleolar fractures, involving breaks in two of the three ankle malleoli (usually the lateral malleolus of the fibula and medial malleolus of the tibia); trimalleolar fractures, which include all three malleoli and thus are more complex and severe; and isolated fractures of either the tibia or fibula's malleolus. These injuries are often associated with acute pain, substantial swelling, and significant impairment in mobility, thereby severely impacting an individual's quality of life and daily functioning. Recent projections indicate an alarming 25% increase in the incidence of ankle fractures by 2025, accentuating the critical demand for advancements in diagnostic approaches.²

Surgical intervention is the most common treatment method for ankle fractures, aimed at restoring alignment and stability of the joint to prevent long-term complications such as arthritis. Common techniques include Open Reduction and Internal Fixation (ORIF), where bone fragments are realigned using plates and screws; Minimally Invasive Plate Osteosynthesis (MIPO), which involves smaller incisions to minimize soft tissue damage while achieving stable fixation; and External Fixation (EF), typically used for complex fractures or in cases with significant soft tissue injury. According to research by Biz et al³ various surgical techniques including ORIF, MIPO, and EF have been evaluated in the treatment of intra-articular tibial pilon fractures, demonstrating that all techniques can achieve satisfactory outcomes, although the choice of technique may depend on the specifics of the fracture and the overall health condition of the patient.

While traditional diagnostic methods, such as X-rays and computed tomography (CT) scans, remain prevalent in clinical practice, their limitations are well-documented. Studies have shown that certain types of fractures may elude detection through these conventional imaging techniques, leading to a reliance on symptomatology and physical examinations for diagnoses.^{4,5} Additionally, 2D CT images may not always offer clear differentiation of the bone from surrounding soft tissues, making it difficult to discern minute but clinically significant details.⁶

The novelty of this study lies in the application of 3D Convolutional Neural Networks (3D-CNNs) specifically tailored for the detection and localization of ankle fractures, a focus not extensively explored in existing literature. Previous applications of CNN technologies have broadly addressed various orthopedic conditions, but the specific challenges posed by ankle fractures—such as the detection of subtle fracture lines and complex bone structures in a region with high anatomical variability—have not been comprehensively tackled using 3D-CNNs. This research not only fills this gap but also introduces innovative techniques such as the integration of Gradient-weighted Class Activation Mapping (Grad-CAM) for enhanced visual interpretation and spatial localization methods that significantly improve diagnostic accuracy and clinician trust in automated systems.

The transition from 2D to 3D CT imaging has been heralded for its potential to provide a more nuanced view of bone architecture, offering greater clarity on fracture depth and angulation. However, several barriers impede the widespread adoption and utility of 3D imaging in clinical practice. Clinicians, habituated to interpreting 2D images, may find the shift to 3D film interpretation more challenging, and in high-pressure environments, this additional complexity may lead to diagnostic oversights. Missed or incorrect diagnoses carry significant repercussions for patient care, resulting in potential delays in treatment, prolonged disability, and increased healthcare costs.⁷ As such, there is a compelling need for an autonomous, machine-assisted approach that can alleviate the human factors contributing to diagnostic errors.

Several studies have showcased the effectiveness of deep learning in identifying various orthopedic conditions within musculoskeletal imaging domains. Yuan et al developed a Convolutional Neural Network (CNN) tailored for the detection of fractures in prestack seismic data,⁸ suggesting the versatility of CNNs across different fields. Emon et al introduced an innovative CNN utilizing a lazy learning strategy to classify skull fractures,⁹ demonstrating the adaptability of CNNs to various types of medical imaging.

For the detection of wrist fractures, Hardalaç et al crafted object detection models,¹⁰ while Wei et al a semi-supervised model aimed at localizing thighbone fractures.¹¹ Further honing the accuracy of fracture detection, Krogue et al automated the placement of bounding boxes and successfully classified hip fractures with a deep learning-based model.¹²

The scope of deep learning extends to differentiating between benign and malignant vertebral compression fractures, with Duan et al constructing predictive models,¹³ and Liu et al comparing the efficacy of a Two-Stream Compare and Contrast Network (TSCCN) against traditional radiologist assessments.¹⁴ These advances also permeate osteoporosis research, with Park et al developing a prediction model for osteoporosis risk¹⁵ and Hong et al devising deep learning scores for osteoporosis and vertebral fracture detection using lateral spine radiography.¹⁰

The advent of 3D CNNs marks a significant evolution in deep learning applications, providing a multi-dimensional analysis of medical images that surpasses traditional diagnostic methods.¹⁶ Their capacity to decipher intricate patterns and features, which often elude human experts, makes them particularly effective for complex cases. Notably, in the context of ankle fractures, the comprehensive analysis by 3D CNNs has shown potential in outstripping conventional imaging techniques, which may miss subtle fracture lines or misinterpret bone structure, especially in patients with concurrent conditions like osteoporosis.¹⁷

Uniquely, 3D CNNs bring an element of objectivity and standardization to diagnostic processes, addressing the variability in interpretations often seen among radiologists. Their consistent output promises to mitigate the risk of diagnostic errors, fostering improved patient management strategies as the demand for medical imaging surges.¹⁸ However, there remains a notable gap in literature regarding the application of 3D CNNs specifically for ankle fracture detection, highlighting the novelty and importance of this study.

Research Objectives

The primary aim of our research is to revolutionize the diagnostic process for ankle fractures through the implementation of 3D Convolutional Neural Networks (CNNs). This study is uniquely poised to address the deficiencies and limitations identified in previous research by utilizing a suite of advanced algorithms. We intend to rigorously evaluate 3D-MobileNet, 3D-Resnet101, and 3D-EfficientNetB7, each of which has shown promise in various medical imaging contexts.¹⁹ The utilization of these sophisticated algorithms is anticipated to significantly enhance the precision of ankle fracture diagnosis beyond the capabilities of conventional imaging techniques.

To further this advancement, we prioritize not only the accuracy but also the interpretability of our models. By conducting detailed comparative analyses across the three aforementioned 3D CNN architectures, we aim to determine the most effective model. The selected model will then be augmented with Gradient-weighted Class Activation Mapping (GradCAM) technology. This integration is designed to offer visual explanations for the decisions made by the CNN, thereby providing clinicians with intuitive, transparent insights into the model's diagnostic reasoning.

Methodology

Data Collection

Data collection was a pivotal step in our study, involving the acquisition of 1453 Neuroimaging Informatics Technology Initiative (NIFTI) files from a local hospital. Each file represents a detailed 3D CT scan of a patient's ankle region, with [Figure 1](#) illustrating the depth of data gathered. [Figure 1a](#) provides a visual representation of a 3D reconstruction derived from a CT scan, highlighting the intricate anatomical details captured by our data collection methodology. [Figure 1b](#) complements this by displaying a set of 2D CT images unwrapped from the 3D model, which aids in understanding the transition from a three-dimensional structure to a two-dimensional analysis plane.

To ensure the study had sufficient statistical power, we conducted an a priori sample size calculation based on expected effect sizes from previous studies on fracture detection using CNNs. Assuming a desired power of 0.80, an alpha level of 0.05, and an anticipated effect size (Cohen's *d*) of 0.5, the calculation indicated that a minimum of 1280 scans would be necessary to detect a significant difference in performance metrics. Consequently, we collected a total of 1453 high-resolution CT scans, exceeding the minimum requirement to enhance the robustness of our findings. We

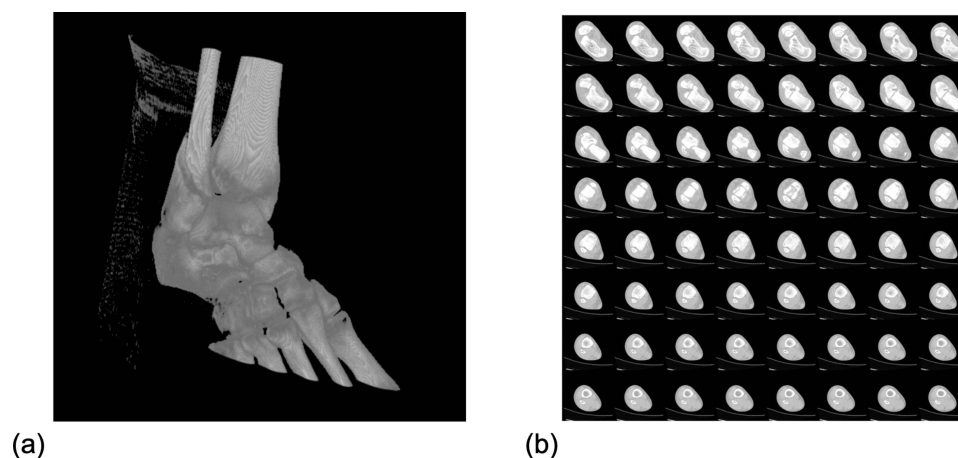


Figure 1 Multidimensional Imaging of Ankle Fractures from CT Scans; (a) 3D Volumetric Reconstruction of the Ankle Region; (b) Sequential 2D CT Slices Derived from the 3D Ankle Model.

initially identified 1600 patients who underwent ankle CT scans at Ningbo No.6 hospital between January 2020 and December 2022.

Inclusion criteria were:

- 1) Patients aged 18 years or older.
- 2) Availability of complete ankle CT scans in NIfTI format.
- 3) No prior ankle surgeries or implants that could affect imaging.

Exclusion criteria included:

- 1) Incomplete imaging data or corrupted files.
- 2) Significant motion artifacts compromising image quality.
- 3) Known systemic bone diseases affecting bone density (eg, severe osteoporosis).

After applying these criteria, 147 scans were excluded, resulting in a final dataset of 1453 scans. These were classified into 820 negative cases (no fracture) and 633 positive cases (fracture present).

To reduce selection bias, we employed a consecutive sampling method, including all eligible patients within the specified timeframe. Observer bias was minimized by using multiple radiologists for image classification, with discrepancies resolved through consensus meetings. Additionally, the radiologists were blinded to the study's objectives to prevent any potential influence on their assessments.

The NIfTI format, traditionally a mainstay in neuroimaging research, has proven invaluable in our analysis due to its robust volumetric data preservation, which is essential for accurate fracture identification.²⁰ The assembled files were then systematically classified into negative (indicative of an unfractured, or “normal”, ankle) and positive (indicative of a fractured ankle) categories, with 820 files classified as negative and 633 as positive.

A panel of seasoned radiologists lent their expertise to the veracity of this categorization, meticulously reviewing each CT scan to confirm or rule out the presence of a fracture. This approach ensures the reliability of our dataset and underscores the diagnostic challenge addressed by our study. The participation of expert radiologists in the classification process further amplifies the reliability and validity of our dataset.

Proposed 3D-CNN System

Our system architecture initiates with the acquisition of volumetric CT images of the ankle region (Figure 2). These high-resolution images provide a comprehensive 3D view of the anatomy, essential for accurate fracture detection. The primary preprocessing steps involve converting Hounsfield Units to a normalized scale, resampling the images to a standard size, and zero-centering the data to ensure consistency across the dataset. Post preprocessing, the volumetric data is fed into three distinct 3D-CNN models: 3D-EfficientNetB7, 3D-Mobilenet, and 3D-Resnet101. Each model is designed to leverage the spatial hierarchies of features inherent in 3D medical images and is evaluated based on its ability to accurately classify and localize ankle fractures.

Upon evaluation, the best-performing model is then integrated with Gradient-weighted Class Activation Mapping (Grad-CAM) to provide 3D visualizations of the areas contributing to the model's predictions. This aspect is crucial for interpretability, allowing clinicians to understand and trust the AI's decision-making process. The Grad-CAM 3D visualizations highlight the fracture zones within the ankle CT images, giving a transparent overview of the model's diagnostic focus points. The systematic flow from image acquisition to model evaluation and visualization encapsulates a comprehensive approach to fracture diagnosis, aiming to enhance both the accuracy and trustworthiness of the diagnostic process.

Preprocessing

The preprocessing pipeline commenced with the meticulous loading of DICOM files, ensuring that any incomplete metadata was filled in to maintain the integrity and completeness of the data set. Subsequent conversion of pixel values to Hounsfield Units (HU) standardized the radiodensity measurements across all tissues, with a selected range of -1000 to 400 hU to effectively isolate the anatomical structures of interest while excluding extraneous noise and artifacts.²¹ To address the issue of scanner resolution variability, which could potentially skew the diagnostic process, the images were

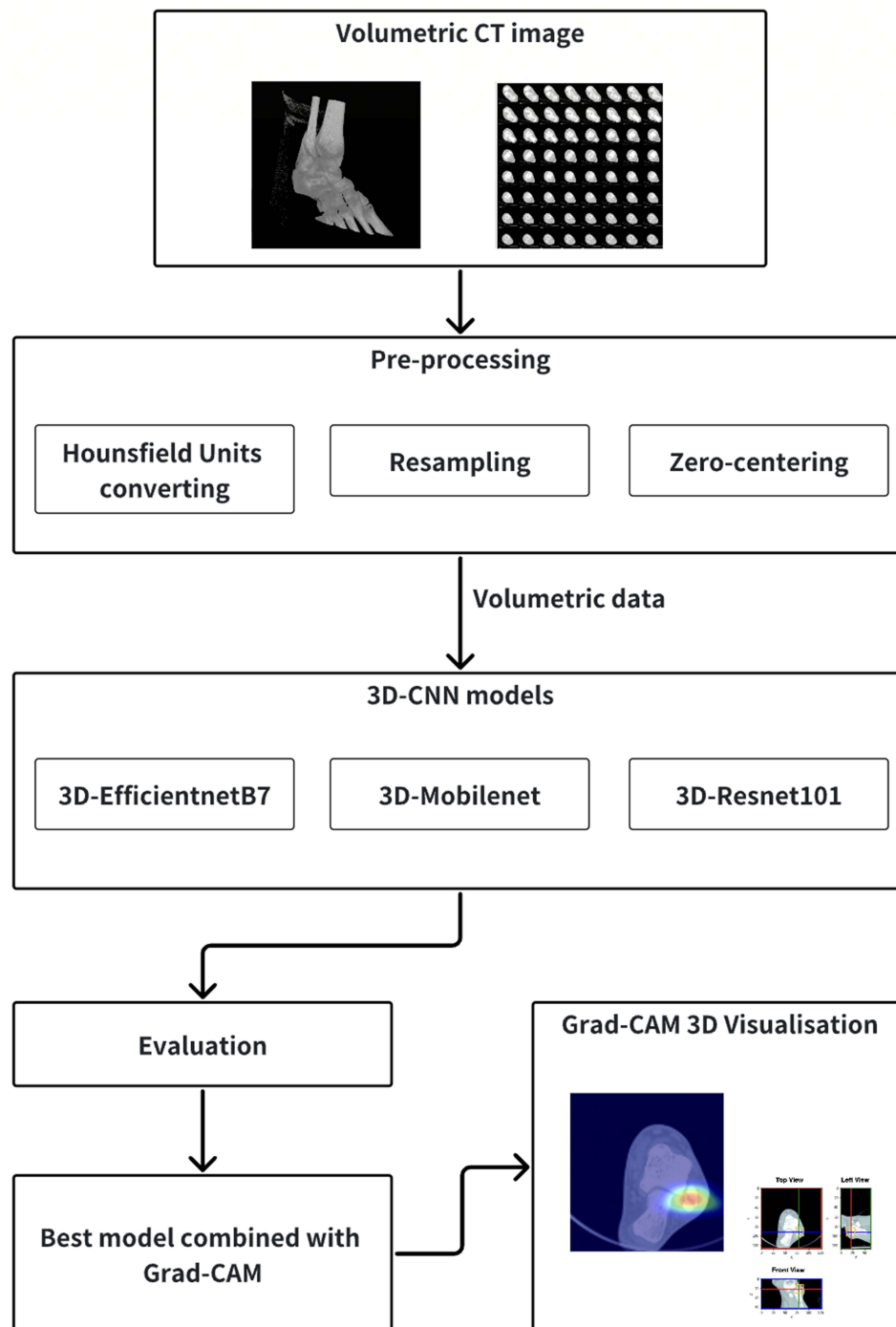


Figure 2 Architecture of the Proposed 3D-CNN System.

resampled to achieve a uniform isomorphic resolution. This standardization is essential for maintaining consistency within our data, as it eliminates discrepancies that could arise from different imaging equipment resolutions.²²

The next phase involved zero-centering the dataset, an important normalization technique where the mean intensity value of the images is adjusted to zero. This process is crucial for the CNN's ability to learn from the data without the interference of intensity bias, thus enhancing the neural network's ability to discern and learn significant features from the images.²³ Lastly, the data was formatted into rank-3 tensors, representing the volumetric nature of the samples in a shape compatible with our 3D CNN models (samples, height, width, depth). An additional single-channel dimension

was appended to facilitate the application of 3D convolution operations. This final transformation of the data structure was vital for enabling the sophisticated analysis required for precise fracture diagnosis.

Training Using Pretrained 3D-CNN Models

In our approach, we harnessed pretrained 3D Convolutional Neural Networks (CNNs) — 3D-Mobilenet, 3D-Resnet101, and 3D-EfficientNetB7 — to refine the diagnostic process for ankle fractures. The training of these models was conducted through a volumetric methodology, allowing for the processing of 3D volumes and corresponding sequences of 2D frames, a strategy that captures the complex spatial relationships present within the data's 3D context.²⁴

A 3D CNN model's architecture is delineated through several layers: convolutional layers that apply filters to extract features via convolution operations (as depicted in Figure 3), pooling layers that downsample the feature maps to reduce dimensionality, and fully connected layers that integrate these features for the final classification and prediction tasks.²⁵ Unlike their 2D counterparts, 3D CNNs operate with filters that extend along the depth axis, enabling them to navigate and learn from the data's volumetric depth, thereby capturing the spatial features in all three dimensions (height, width, and depth). Figure 3 illustrates the convolution operation within a 3D CNN, showing how kernels move through the volumetric data. This process involves element-wise multiplication and addition within the volume to produce scalar values that form a new 3D output volume, capturing the essence of the input's spatial structure.²⁶

In our selection of models, 3D-Mobilenet stands out for its lightweight design and depthwise separable convolutions that reduce computational load.²⁷ The 3D-Resnet101 model employs a deep network architecture with 101 layers, integrating residual connections to facilitate the training of such an extensive network.²⁸ The 3D-EfficientNetB7, representing the pinnacle of 3D CNN design, utilizes a compound scaling method to balance the parameters and computational demands during training.²⁹

For optimal model performance, parameters were calibrated with a batch size of 16 and a learning rate set at 0.001. The choice of Adam optimizer reflects its popularity and effectiveness in deep learning applications.³⁰ We employed a binary cross-entropy loss function, aligning with our binary classification objectives. The models underwent a training process spanning 20 epochs, ensuring sufficient exposure to the dataset while preventing overfitting. Upon completion, the best-performing model was determined based on its validation accuracy, signifying its readiness to accurately classify and predict ankle fractures in a clinical setting.

Performance Evaluation

To gauge the efficacy of our 3D Convolutional Neural Network (CNN) models, we employed a suite of metrics. These included accuracy, precision, recall, F1 score, the Receiver Operating Characteristic (ROC) curve, and the Area Under the ROC Curve (AUC). Model accuracy is the ratio of instances correctly predicted to the total in the dataset, providing a snapshot of overall model performance. Precision is defined as the ratio of true positive detections to all instances

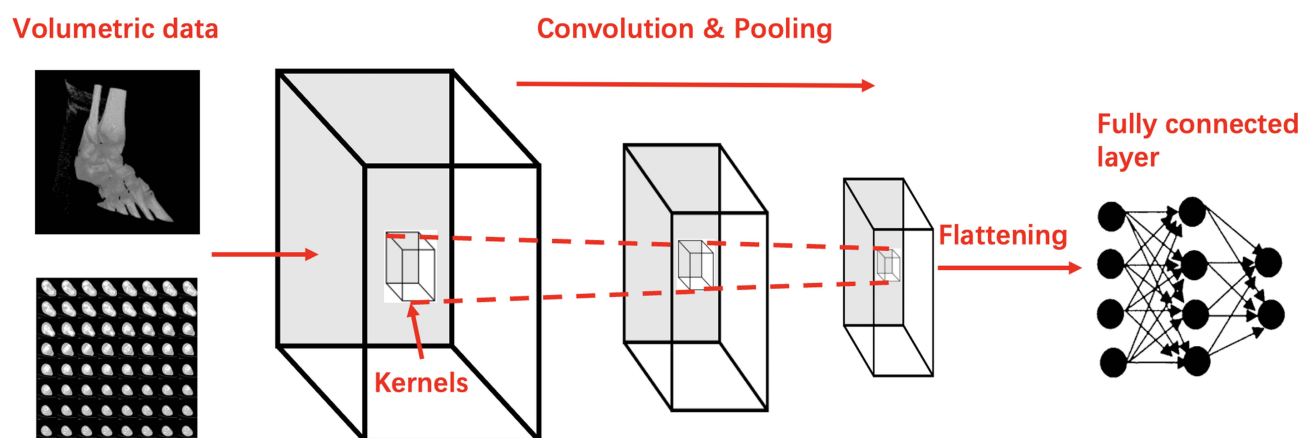


Figure 3 Convolution Operation Workflow in 3D CNNs.

predicted as positive, offering insight into the model's exactness. Recall, or the true positive rate, reflects the model's ability to identify all relevant cases within the dataset. The F1 score, which harmonizes precision and recall, offers a single metric for situations where positive and negative classes are unevenly represented.³¹ The ROC curve illustrates the trade-off between sensitivity (recall) and specificity across various thresholds, serving as a tool for evaluating the model's discriminative ability. The AUC, derived from the ROC curve, quantifies the overall predictive quality of the model, with a score of 0.5 corresponding to no discriminative ability (equivalent to random chance) and a score of 1 representing perfect prediction.³²

Integration of Grad-CAM with 3D-CNN for Enhanced Visualization

Incorporating Gradient-weighted Class Activation Mapping (Grad-CAM) technology into our 3D-CNN framework has been a pivotal advancement in our study. As depicted in Figure 4, the Grad-CAM process begins after the 3D-CNN has processed the input CT scans. It utilizes the gradient information flowing into the final convolutional layer of the neural network to produce a coarse heatmap that highlights the important regions for predicting the classification of fractures. The heatmaps generated by Grad-CAM offer a layer of interpretability by visually emphasizing the critical areas within the 3D CT scans that influence the model's decision-making process.³³ These visual cues are especially valuable for clinicians, providing a transparent and intuitive understanding of the model's analytical focus. To ensure the heatmaps align accurately with the high-resolution CT scans, they undergo a spline interpolation zoom, which scales up the heatmap to match the dimensions of the original images. Subsequently, the refined heatmaps are superimposed on the original CT scans, effectively spotlighting the regions of interest within the three-dimensional anatomical context. For more granular analysis, bounding boxes are algorithmically placed around the high-intensity zones in the heatmaps. This is achieved by setting a specific threshold—determined via Otsu's method³⁴—to discern the most pertinent areas related to the presence of a fracture.

Specifications of Employed 3D CNN Models

Our study utilizes three distinct 3D Convolutional Neural Network models, each selected for their specific architectural strengths and suitability for processing high-resolution CT scans. The first model, 3D-Mobilenet, is designed to be lightweight with 3.2 million parameters, making it particularly efficient for mobile applications. It features depthwise separable convolutions that significantly reduce computational requirements. The second model, 3D-Resnet101, incorporates a deep residual network architecture with 42.5 million parameters. This model uses skip connections to facilitate training across its numerous layers, enhancing its capability to capture complex image features. The third model, 3D-EfficientNetB7, employs a compound scaling method which optimizes depth, width, and resolution, scaling up to

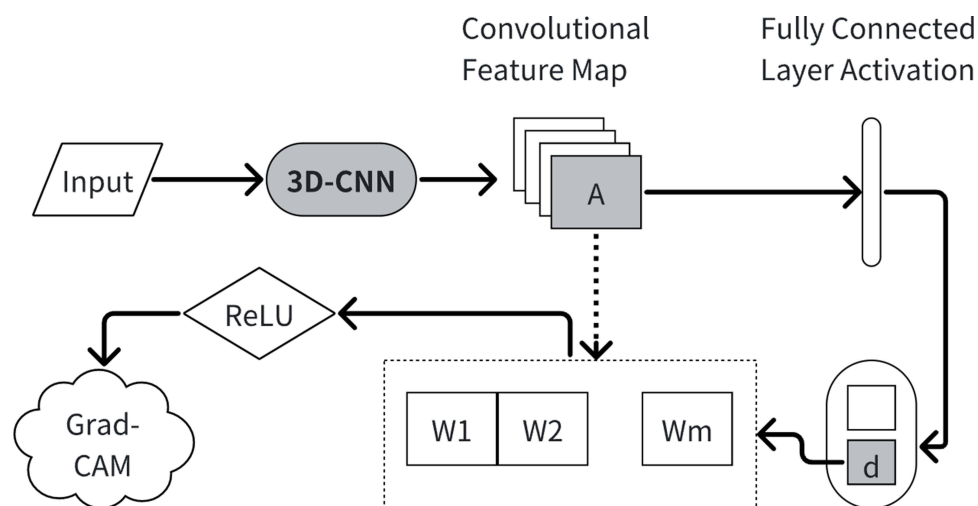


Figure 4 Integration of Grad-CAM with 3D-CNN for Targeted Visualization in Ankle Fracture Detection.

66 million parameters to dynamically adjust to the specific demands of the dataset. All models process an input size of 128x128x128 and have been implemented using the Tensorflow framework, offering robust feature extraction and classification capabilities essential for medical imaging analysis.

Statistical Analysis

We evaluated the performance of the three 3D-CNN models—3D-EfficientNetB7, 3D-Mobilenet, and 3D-Resnet101—using metrics such as accuracy, precision, recall (sensitivity), specificity, F1 score, and the area under the receiver operating characteristic curve (AUC-ROC). To provide uncertainty estimates, we calculated 95% confidence intervals for these metrics using the bootstrap method with 1000 resamples.

Statistical significance between model performances was assessed using the DeLong test for AUC comparisons and the McNemar's test for differences in accuracy and other classification metrics. A p-value of less than 0.05 was considered statistically significant.

Our validation process involved stratified random splitting of the dataset into training (70%), validation (15%), and test (15%) sets, maintaining consistent proportions of positive (fracture) and negative cases across all subsets. We implemented 5-fold cross-validation during model training to enhance generalizability and mitigate overfitting. Hyperparameters were optimized based on validation performance, with the optimal learning rate set at 0.0005, batch size at 16, and the Adam optimizer selected for training.

All statistical analyses were performed using Python (version 3.8) with libraries such as Scikit-learn and SciPy. The code used for analysis is available upon reasonable request to facilitate reproducibility.

Results

Figure 5 presents the numerical evaluation of our study's 3D Convolutional Neural Network (3D-CNN) models, detailing their performance over the training epochs and in classification metrics. As shown in Figure 5a, 3D-EfficientNetB7 achieved the highest accuracy, reaching approximately 0.91 after 20 epochs. In comparison, 3D-Mobilenet and 3D-Resnet101 peaked at lower accuracies of around 0.88 and 0.85, respectively. The comprehensive performance metrics in Figure 5b indicate that 3D-EfficientNetB7 not only attained the top accuracy of 0.91 but also excelled with the highest Area Under the Curve (AUC) at 0.94, suggesting a superior true positive identification rate. 3D-Mobilenet followed with

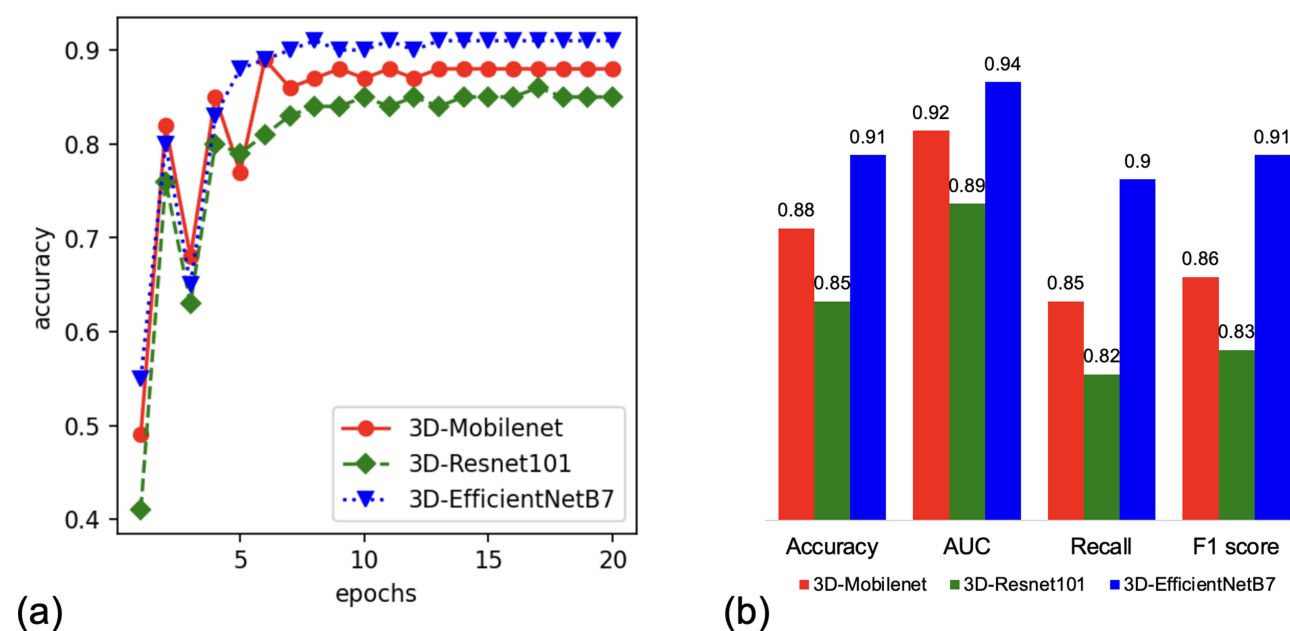


Figure 5 Comparative Performance Analysis of 3D-CNNs for Ankle Fracture Identification; (a) Training Accuracy Evolution Over 20 Epochs; (b) Comparative Evaluation of Model Metrics: Accuracy, AUC, Recall, and F1 Score.



Figure 6 Comparative visualization between the Grad-CAM heatmap and expert's visual judgement of a fractured ankle CT scan. Left: original image; middle: expert's visual judgement of the fracture location; right: Grad-CAM detection of the fracture location.

a notable accuracy of 0.88 and an AUC of 0.92. Meanwhile, 3D-Resnet101 recorded an accuracy of 0.85 and an AUC of 0.89. Recall is a crucial measure for reducing missed diagnoses of fractures, and here, 3D-EfficientNetB7 led with a score of 0.9. 3D-Mobilenet achieved a recall of 0.85, while 3D-Resnet101 was slightly behind at 0.82. In assessing the F1 Score, which combines precision and recall into a single metric, 3D-EfficientNetB7 again ranked highest with 0.91. The scores for 3D-Mobilenet and 3D-Resnet101 were 0.86 and 0.83, respectively. These numerical findings robustly attest to the superior capability of 3D-CNN models in the accurate classification of ankle fractures, with 3D-EfficientNetB7 emerging as the standout model across all evaluated metrics.

Figure 6 presents a compelling visual comparison that illustrates the effectiveness of the Grad-CAM technology in conjunction with our best 3D-CNN model (3D-EfficientNetB7). The left panel of the figure displays the original CT scan of an ankle with a suspected fracture. The central panel highlights the fracture location as determined by an expert radiologist, denoted by a red circle indicating the precise area of concern. This visual judgement by the expert serves as a reference standard for the fracture's location. The right panel of Figure 6 showcases the Grad-CAM heatmap overlay on the same CT scan. The heatmap reveals a concentration of colors in the region corresponding to the expert's identification, with the intensity of the colors representing the model's confidence in the fracture's location. The high degree of overlap between the expert's visual judgement and the Grad-CAM detection underscores the model's accuracy in localizing the fracture. This visual correlation not only reinforces the validity of our 3D-CNN model but also demonstrates the potential of Grad-CAM to enhance the interpretability of the model's decision-making process for medical professionals.

The spatial localization capabilities of our model are further exemplified in Figure 7, which delineates the precise coordinates of a detected fracture within a 3D space. The bounding box, highlighted in orange and centered at the spatial coordinates ($x=76$, $y=90$, $z=22$), is presented from three different perspectives: the top, front, and left views of the ankle CT scan. These orientations provide a comprehensive overview of the fracture's position relative to the entire anatomy of the ankle. The top view demonstrates the fracture's location along the horizontal plane, while the front view offers insight into its depth, and the left view indicates its position along the sagittal plane. The use of bounding boxes is instrumental in enhancing the accuracy of fracture detection, as they encapsulate the region of interest, allowing clinicians to focus on the most relevant area for assessment and treatment planning. This multi-angled visualization technique, as captured in Figure 7, not only augments the detection process but also aids in the communication of the fracture's precise location to the medical team. The clear demarcation of the fracture site through these bounding boxes can significantly streamline the diagnostic workflow, ultimately improving the patient's care pathway.

Discussion

The deployment of 3D Convolutional Neural Networks (3D-CNNs) in radiological diagnostics represents a significant stride in harnessing the capabilities of deep learning for medical applications. Our research demonstrates the potential of 3D-EfficientNetB7, which achieved an accuracy of 0.91 and an AUC of 0.94, to markedly outperform its contemporaries in fracture detection tasks.¹⁶ This is in congruence with findings from other domains, where deeper neural architectures have consistently outperformed shallower counterparts by capturing more nuanced patterns within complex data.¹⁷ The

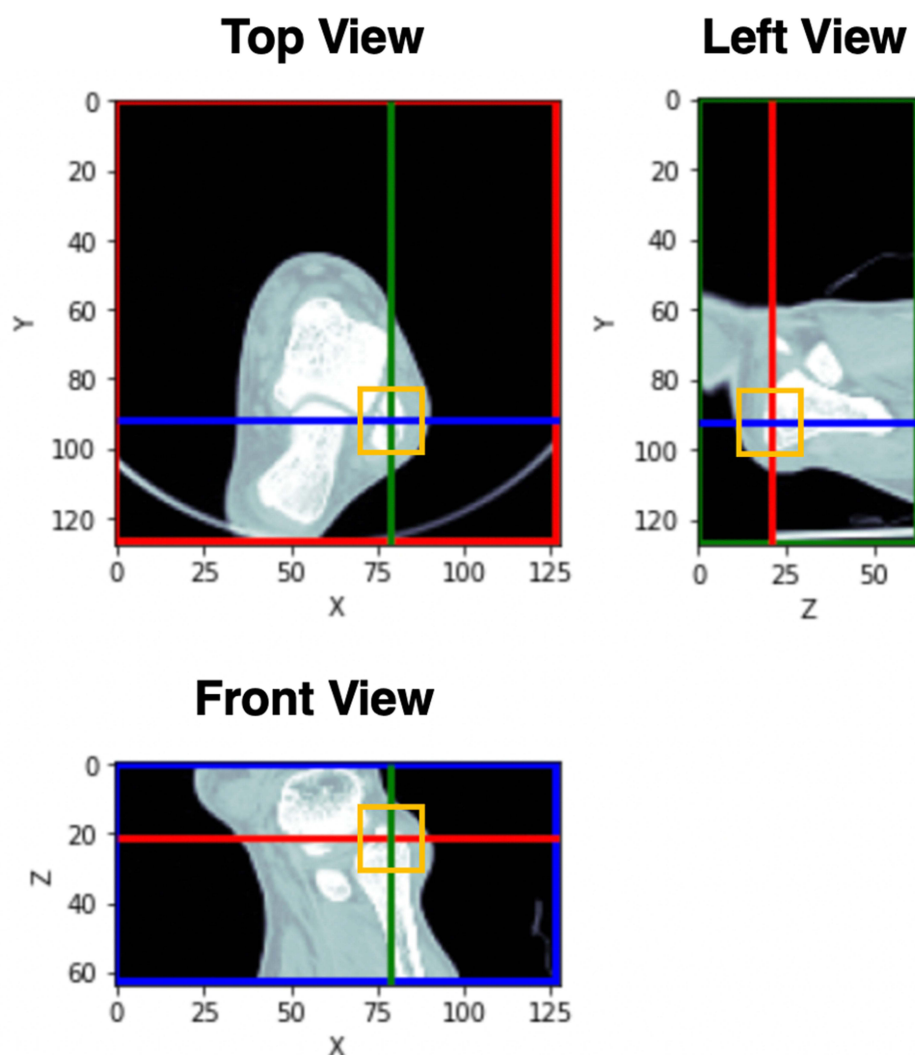


Figure 7 Multi-Angle Visualization of Fracture Detection with Bounding Box Coordinates on Ankle CT Scans.

high recall rate of 0.9 further suggests that the 3D-EfficientNetB7 model is adept at reducing false negatives, which is paramount in clinical practice where the consequences of missing a diagnosis are significant.

Grad-CAM visualizations (Figure 6) play a pivotal role in demystifying the decision-making process of 3D-CNNs.³³ By aligning the model's attention with the radiologist's assessment, these visualizations facilitate a more transparent and trust-building approach to AI integration in clinical settings, addressing a common reluctance associated with the adoption of such technologies.³⁵

Emerging imaging techniques, particularly advanced CT and AI, are pivotal across various medical fields. For instance, spectral computed tomography (SCT) integrated with three-material decomposition has enhanced the detection of bone marrow edema, proving invaluable in emergency medical settings.³⁶ In orthopedics, AI-enhanced imaging has successfully been applied to automatically detect vertebral fractures on CT scans. This technology demonstrates high sensitivity and specificity in differentiating between normal and fractured vertebral bodies.³⁷ Additionally, in the field of trauma radiology, the use of dual-layer spectral detector computed tomography (SDCT) and electron density images has significantly improved the accuracy of detecting post-traumatic prevertebral hematomas.³⁸ These developments underscore the significant impact and potential of integrating AI with cutting-edge imaging techniques across diverse medical disciplines.

The localization accuracy provided by the 3D-CNN model (Figure 7) has far-reaching implications for surgical planning. Precise three-dimensional localization is not only vital for diagnosis but also serves as a crucial tool for surgical teams to plan interventions more effectively.³⁹ This multi-perspective visualization capability surpasses the traditional 2D imaging approach, offering a holistic view of the fracture's spatial context.⁴⁰ Such detailed spatial comprehension is essential in orthopedic imaging, where the intricate nature of musculoskeletal injuries demands an exhaustive three-dimensional analysis.⁴¹

While the findings of this study are promising, they are subject to certain limitations that must be considered when interpreting the results. Primarily, the absence of external validation poses a significant limitation. Our models were tested and validated only with internal datasets from a single institution. This lack of external validation may affect the generalizability of our results to other settings and populations. To ensure broader applicability and reliability, future studies should include external validation with diverse datasets to assess the robustness and effectiveness of the 3D-CNN models across various clinical environments. Prospective multi-center trials could establish the generalizability of these findings, while a diverse and extensive dataset could confirm the model's robustness across varying patient demographics and imaging modalities.⁴² As we advance towards the clinical integration of such AI-driven diagnostic tools, it is imperative to maintain a focus on the seamless incorporation into existing workflows, upskilling radiologists to leverage these technologies effectively, and addressing ethical considerations surrounding AI in healthcare.⁴³ Our findings contribute valuable insights into the evolving landscape of medical imaging, highlighting the transformative potential of AI in enhancing diagnostic precision and efficiency. The nuanced detection capabilities of advanced 3D-CNN models, accompanied by interpretative tools like Grad-CAM, present a compelling case for their adoption in modern radiology practices.

Conclusion

Our research has demonstrated the formidable capabilities of 3D Convolutional Neural Networks (3D-CNNs) in the precise identification and localization of ankle fractures, a task that has presented significant challenges in traditional diagnostic methods. The employment of 3D-EfficientNetB7, in particular, has resulted in exceptional accuracy, surpassing its counterparts with significant precision and recall rates. The integration of Gradient-weighted Class Activation Mapping (Grad-CAM) with 3D-CNNs has further augmented the interpretability of diagnostic predictions, providing clinicians with valuable visual aids to corroborate their assessments.

The success of our models in identifying subtle fracture lines and complex bone structures suggests that the implementation of such advanced AI tools in clinical settings could greatly enhance the efficacy and reliability of fracture diagnoses. This, in turn, holds the potential to significantly reduce misdiagnoses and consequently improve patient outcomes by enabling timely and appropriate treatment interventions. In conclusion, this study not only bridges a critical gap in medical imaging but also heralds a new era in diagnostic precision. As we stand on the cusp of a technological revolution in healthcare, the potential of 3D-CNNs in improving diagnostic processes is not just promising—it is transformative.

Patient Consent

We hereby declare that informed consent was obtained from the human subjects involved in this study for the publication of their images. The purpose and nature of the study, as well as the potential risks and benefits, were explained to the participants prior to obtaining their consent.

Data Sharing Statement

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Ethics Approval

This study was conducted in accordance with the principles of the Declaration of Helsinki. The ethical considerations pertaining to this research have been rigorously examined and approved by the Ethics Committee of Ningbo 6. Hospital

(approval number, 2024-25 (L)). The study involving the utilization of medical images and data adheres to the highest standards of ethical conduct and patient confidentiality. We hereby declare that informed consent was obtained from all human subjects involved in this study for the publication of their images. The purpose and nature of the study, as well as the potential risks and benefits, were thoroughly explained to the participants prior to obtaining their consent.

Funding

This research was financially supported by two grants: The Science and Technology Projects in the Field of Agriculture and Social Development in Yinzhou District, Ningbo City, Zhejiang Province, China (Grant No. 2021AS0064), and the Ningbo Public Welfare Research Program Project in Zhejiang Province, China (Grant No. 0211JCGY02005). We gratefully acknowledge these organizations for their generous support and contribution to our research.

Disclosure

This paper has been uploaded to Research Square as a preprint: <https://www.researchsquare.com/article/rs-3583938/v1>.

The authors declare no conflict of interest.

References

1. Court-Brown CM, Caesar B. Epidemiology of adult fractures: a review. *Injury*. 2006;37(8):691–697. doi:10.1016/j.injury.2006.04.130
2. Riggs BL, Melton LJ 3rd. The worldwide problem of osteoporosis: insights afforded by epidemiology. *Bone*. 1995;17(5 Suppl):505S–511S. doi:10.1016/8756-3282(95)00258-4
3. Biz C, Angelini A, Zamperetti M, et al. Medium-long-term radiographic and clinical outcomes after surgical treatment of intra-articular tibial pilon fractures by three different techniques. *Biomed Res Int*. 2018;2018:6054021. doi:10.1155/2018/6054021
4. Tollefson B, Nichols J, Fromang S, Summers RL. Validation of the sonographic Ottawa foot and ankle rules (SOFAR) study in a large urban trauma center. *J Miss State Med Assoc*. 2016;57(2):35–38.
5. Miller AN, Prasarn ML, Dyke JP, et al. Quantitative assessment of the vascularity of the talus with gadolinium-enhanced magnetic resonance imaging. *J Bone Joint Surg Am*. 2011;93(12):1116–1121. doi:10.2106/JBJS.J.00693
6. Choksi P, Jepsen KJ, Clines GA. The challenges of diagnosing osteoporosis and the limitations of currently available tools. *Clin Diabetes Endocrinol*. 2018;4:12. doi:10.1186/s40842-018-0062-7
7. Chan HJ, Woods M, Stella D. Three-dimensional computed craniofacial tomography (3D-CT): potential uses and limitations. *Aust Orthod J*. 2007;23(1):55–64.
8. Yuan ZY, Jiang YX, Li JJ, Huang HD. A convolutional neural network for prestack fracture detection. *arXiv preprint*. 2021;2107.01466. doi:10.48550/arXiv.2107.01466
9. Emon MM, Ornob TR, Rahman M. Classifications of skull fractures using CT scan images via CNN with lazy learning approach. *arXiv preprint*. 2022;2203.10786. doi:10.48550/arXiv.2203.10786
10. Hardalaç F, Uysal F, Peker O, et al. Fracture detection in wrist X-ray images using deep learning-based object detection models. *Sensors*. 2022;22(3):1285. doi:10.3390/s22031285
11. Wei JM, Yao JK, Zhang GS, et al. semi-supervised object detection based on single-stage detector for thighbone fracture localization. *arXiv preprint*. 2022;2210.10998. doi:10.48550/arXiv.2210.10998
12. Krogue JD, Cheng KYV, Hwang KM, et al. Automatic hip fracture identification and functional subclassification with deep learning. *Radiol Artif Intell*. 2020;2(2):e190023. doi:10.1148/ryai.2020190023
13. Duan S, Hua Y, Cao G, et al. Differential diagnosis of benign and malignant vertebral compression fractures: comparison and correlation of radiomics and deep learning frameworks based on spinal CT and clinical characteristics. *Eur J Radiol*. 2023;165:110899. doi:10.1016/j.ejrad.2023.110899
14. Liu B, Jin Y, Feng S, et al. Benign vs malignant vertebral compression fractures with MRI: a comparison between automatic deep learning network and radiologist's assessment. *Eur Radiol*. 2023;33(7):5060–5068. doi:10.1007/s00330-023-09713-x
15. Wu X, Park S. A prediction model for osteoporosis risk using a machine-learning approach and its validation in a large cohort. *J Korean Med Sci*. 2023;38(21):e162. doi:10.3346/jkms.2023.38.e162
16. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88. doi:10.1016/j.media.2017.07.005
17. Huang G, Liu Z, Maaten LVD, et al. Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; Honolulu, Hawaii, USA.4700–4708. doi:10.1109/CVPR.2017.243.
18. Choy G, Khalilzadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. *Radiology*. 2018;288(2):318–328. doi:10.1148/radiol.2018171820
19. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500–510. doi:10.1038/s41568-018-0016-5
20. Cox RW, Ashburner J, Breman H, et al. A (sort of) new image data format standard: niFTI-1. In: 10th Annual Meeting of the Organization for Human Brain Mapping. Budapest, Hungary. 2004;1.
21. Kalender WA, Seissler W, Klotz E, et al. Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation. *Radiology*. 1990;176(1):181–183. doi:10.1148/radiology.176.1.2353088
22. Zitova B, Flusser J. Image registration methods: a survey. *Image Vis Comput*. 2003;21(11):977–1000. doi:10.1016/S0262-8856(03)00137-9

23. Cun YL, Bottou L, Orr G, et al. Efficient backprop, neural networks: tricks of the trade. *Lect Notes Comput Sci.* 1998;1524:5–50.
24. Ji SW, Xu W, Wang M, et al. 3D convolutional neural networks for human action recognition. *IEEE Trans Pattern Anal Mach Intell.* 2012;35(1):221–231. doi:10.1109/TPAMI.2012.59
25. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM.* 2017;60(6):84–90. doi:10.1145/3065386
26. Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3d convolutional networks. In: IEEE International Conference on Computer Vision; 2015; Santiago, Chile. 4489–4497. doi:10.1109/ICCV.2015.510.
27. Howar AG, Zhu ML, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint.* 2017:1704.04861. doi:10.48550/arXiv.1704.04861.
28. He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; Las Vegas, Nevada, USA. 770–778. doi:10.1109/CVPR.2016.90.
29. Tan MM, Le QV. Efficientnet: rethinking model scaling for convolutional neural networks. *arXiv preprint.* 2019;1905.11946. doi:10.48550/arXiv.1905.11946
30. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv preprint.* 2014;1412.6980. doi:10.48550/arXiv.1412.6980
31. Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint.* 2020;2010.16061. doi:10.48550/arXiv.2010.16061
32. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29–36. doi:10.1148/radiology.143.1.7063747
33. Selvaraju RR, Cogswell M, Das A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization. In: IEEE International Conference on Computer Vision; 2017; Venice, Italy. 618–626. doi:10.1109/ICCV.2017.74.
34. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern.* 1979;9(1):62–66. doi:10.1109/TSMC.1979.4310076
35. Tizhoosh HR, Pantanowitz L. Artificial intelligence and digital pathology: challenges and opportunities. *J Pathol Inform.* 2018;9:38. doi:10.4103/jpi.jpi_53_18
36. Schierenbeck M, Grözing M, Reichardt B, et al. Detecting bone marrow edema of the extremities on spectral computed tomography using a three-material decomposition. *Diagnostics (Basel).* 2023;13:2745. doi:10.3390/diagnostics13172745
37. Polzer C, Yilmaz E, Meyer C, et al. AI-based automated detection and stability analysis of traumatic vertebral body fractures on computed tomography. *Eur J Radiol.* 2024;173:111364. doi:10.1016/j.ejrad.2024.111364
38. Sedaghat S, Langguth P, Larsen N, et al. Diagnostic accuracy of dual-layer spectral CT using electron density images to detect post-traumatic prevertebral hematoma of the cervical spine. *Rofo.* 2021;193(12):1445–1450. doi:10.1055/a-1529-7010
39. Yu VY, Tran A, Nguyen D, et al. The development and verification of a highly accurate collision prediction model for automated noncoplanar plan delivery. *Med Phys.* 2015;42(11):6457–6467. doi:10.1118/1.4932631
40. Roth HR, Lu L, Liu JM, et al. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Trans Med Imaging.* 2015;35(5):1170–1181. doi:10.1109/TMI.2015.2482920
41. Yan K, Wang XS, Lu L, et al. Deep lesion graphs in the wild: relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018; Seattle, Washington, USA. 9261–9270. doi:10.1109/CVPR.2018.00965.
42. Johnson AEW, Pollard TJ, Greenbaum NR, et al. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. *arXiv preprint.* 2019:1901.07042. doi:10.48550/arXiv.1901.07042.
43. Chartrand G, Cheng PM, Vorontsov E, et al. Deep Learning: a primer for radiologists. *Radiographics.* 2017;37(7):2113–2131. doi:10.1148/rg.2017170077

Therapeutics and Clinical Risk Management

Dovepress

Publish your work in this journal

Therapeutics and Clinical Risk Management is an international, peer-reviewed journal of clinical therapeutics and risk management, focusing on concise rapid reporting of clinical studies in all therapeutic areas, outcomes, safety, and programs for the effective, safe, and sustained use of medicines. This journal is indexed on PubMed Central, CAS, EMBASE, Scopus and the Elsevier Bibliographic databases. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/therapeutics-and-clinical-risk-management-journal>