Al-Farouk journal of sciences
Volume (2), Issue (3) (JUNE, 2026)
ISSN: 3135-2359

مجلة الفاروق للعلوم
المجلد (2) ملحق العدد 3 يونيو 2026
ردمد: 3135-2359

Deploying Lightweight AI Diagnostic Tools in Resource-Limited Healthcare Settings: A Case Study in Bone Fracture Detection

Al Mokhtar Abdullah Faraj[1], prof. Ibrahim Ali Almerhag[2]
[1]Ghryan University, Ghryan, Libya
[2] University of Tripoli, College of Information Technology, Department of Computer Networking,
Eng: Al Mokhtar Abdullah Faraj: farajmoktar10@gmail.com

تاريخ الاستلام: 2026/05/21 تاريخ المراجعة 16/06/ 2026 تاريخ القبول:2026/06/28- تاريخ النشر : 10/07/ 2026

Abstract

Bone fractures are among the most common musculoskeletal injuries requiring timely diagnosis, yet accurate X-ray interpretation remains challenging in resource-limited healthcare settings due to radiologist shortages. This study develops and evaluates a Computer-Aided Diagnosis (CAD) system for automated binary fracture detection using Convolutional Neural Networks (CNNs). A publicly available X-ray dataset was rigorously preprocessed and augmented to mitigate overfitting. Multiple architectures, including custom CNNs and transfer learning models (EfficientNetB0, VGG16), were trained and compared. The optimal lightweight model was subsequently integrated into a secure, Flask-based web application featuring real-time inference and automated data privacy protocols. Experimental results demonstrate that transfer learning significantly outperforms custom architectures, with EfficientNetB0 achieving the highest diagnostic performance (93.8% accuracy, 0.96 AUC-ROC, and 92.7% sensitivity). However, the lightweight custom CNN achieved a highly competitive 89.5% accuracy with a rapid CPU inference time of ~120ms, making it ideal for practical deployment. Future studies should focus on multi-class fracture typing, external validation across diverse clinical datasets, and the integration of Explainable AI (XAI) to enhance radiologist trust.

Keywords—Computer-Aided Diagnosis (CAD), Lightweight Deep Learning, Bone Fracture Detection, Resource-Limited Healthcare, Medical Image Analysis, Edge Deployment.

نشر أدوات التشخيص الخفيفة المدعومة بالذكاء الاصطناعي في بيئات الرعاية الصحية محدودة الموارد: دراسة حالة في الكشف عن كسور العظام

المهندس المختار عبدالله فرج [1] ، الأستاذ إبراهيم علي المرهاق [2]
[1]جامعة غريان، غريان، ليبيا
[2] جامعة طرابلس، كلية تقنية المعلومات، قسم شبكات الحاسوب

الملخص

تُعَد كسور العظام من أكثر إصابات الجهاز العضلي الهيكلي شيوعاً والتي تتطلب تشخيصاً دقيقاً وفي الوقت المناسب، إلا أن تفسير صور الأشعة السينية بدقة يظل تحدياً كبيراً في بيئات الرعاية الصحية محدودة الموارد بسبب نقص أخصائي الأشعة. تطور هذه الدراسة وتقيم نظام تشخيص مساعد بالحاسوب (CAD) للكشف الآلي الثنائي عن الكسور باستخدام الشبكات العصبية الالتفافية .(CNNs) تمت معالجة مجموعة بيانات عامة من صور الأشعة مسبقاً وتعزيزها بدقة للتخفيف

من الإفراط في التجهيز (Overfitting). تم تدريب ومقارنة هياكل متعددة، بما في ذلك شبكات CNN مخصصة ونماذج نقل التعلم (EfficientNetB0 و VGG16). تم دمج النموذج المخصص خفيف الوزن الأمثل لاحقاً في تطبيق ويب آمن قائم على إطار عمل Flask ، يتميز بالاستدلال في الوقت الفعلي وبروتوكولات خصوصية البيانات المؤتمتة. أظهرت النتائج التجريبية أن نقل التعلم يتفوق بشكل كبير على الهياكل المخصصة، حيث حقق نموذج EfficientNetB0 أعلى أداء تشخيصي) دقة 93.8٪، ومنطقة تحت المنحنى AUC–ROC بقيمة 0.96، وحساسية 92.7٪). ومع ذلك، حقق نموذج CNN المخصص خفيف الوزن دقة تنافسية بلغت 89.5٪ مع وقت استدلال سريع على وحدة المعالجة المركزية (CPU) يبلغ حوالي 120 مللي ثانية، مما يجعله مثالياً للنشر العملي. توصي الدراسات المستقبلية بالتركيز على تصنيف أنواع الكسور المتعدد، والتحقق الخارجي عبر مجموعات بيانات سريرية متنوعة، ودمج تقنيات الذكاء الاصطناعي القابل للتفسير (XAI) لتعزيز ثقة أخصائي الأشعة.

الكلمات المفتاحية: التشخيص بمساعدة الحاسوب (CAD)، التعلم العميق خفيف الوزن، الكشف عن كسور العظام، الرعاية الصحية محدودة الموارد، تحليل الصور الطبية، النشر على الحافة.

I. INTRODUCTION

Bone fractures represent one of the most common musculoskeletal injuries globally, requiring timely and accurate diagnosis to ensure effective treatment and optimal patient recovery [1]. Radiographic X-ray imaging remains the primary diagnostic modality for fracture detection due to its widespread availability, low cost, and rapid acquisition [2]. However, the interpretation of X-ray images is highly dependent on the expertise of radiologists and is susceptible to human error, particularly in high-volume clinical settings or regions suffering from acute specialist shortages, such as Libya [3]. Fatigue, subtle fracture appearances, and overlapping anatomical structures can lead to diagnostic oversights, including missed or delayed diagnoses, which may result in severe complications like non-union or malunion [4].

In recent years, artificial intelligence (AI), and particularly deep learning, has demonstrated transformative potential in medical image analysis [5]. Convolutional Neural Networks (CNNs) have achieved remarkable success in automating the detection of abnormalities by learning discriminative features directly from raw pixel data, eliminating the need for manual feature engineering [6]. Despite these advancements, existing fracture detection systems often face significant deployment challenges. Many state-of-the-art models are computationally heavy, acting as "black boxes" that require high-end Graphics Processing Units (GPUs), making them impractical for resource-limited healthcare environments [7]. Furthermore, a critical gap exists between theoretical model development in controlled research environments and the practical, secure deployment of these tools in real-world clinical workflows [8].

To address these challenges, this paper presents the development and deployment of a lightweight, secure Computer-Aided Diagnosis (CAD) system for automated binary bone fracture detection. The study systematically compares custom CNN architectures with transfer learning models, specifically EfficientNetB0 and VGG16, to identify the optimal balance between diagnostic accuracy and computational efficiency. While transfer learning models achieved the highest performance, a lightweight custom CNN was selected for final deployment to ensure compatibility with standard clinical hardware.

The primary contributions of this research are threefold. First, it provides a rigorous empirical comparison of deep learning architectures for musculoskeletal X-ray classification, utilizing aggressive data augmentation to mitigate overfitting. Second, it details the engineering of a robust, Flask-based web application that ensures strict preprocessing consistency between training and deployment phases. Third, it implements essential data privacy protocols, including UUID-based file anonymization and automatic post-inference deletion, ensuring the system is viable for deployment in resource-constrained settings without compromising patient confidentiality. By bridging the gap between deep learning research and practical software engineering, this study demonstrates that lightweight AI tools can effectively serve as reliable "second readers" to augment clinical decision-making.

II. RELATED WORK

The integration of artificial intelligence (AI) into medical imaging has significantly transformed diagnostic radiology, particularly in the domain of musculoskeletal disorders. Fracture detection, a fundamental yet error-prone task in emergency and orthopedic care, has become a focal point for Computer-Aided Diagnosis (CAD) research. This section critically reviews the evolution of CAD systems, the theoretical foundations of deep learning in medical image analysis, and recent advances in automated fracture detection, highlighting the specific research gap this study addresses.

Traditional Approaches and Limitations

Before the advent of deep learning, fracture detection in X-ray images relied heavily on manual interpretation by radiologists or semi-automated CAD systems based on classical image processing and machine learning techniques. Early CAD systems employed handcrafted features, such as edge detection (e.g., Canny or Sobel filters), texture descriptors, and morphological operations to isolate bone structures and identify discontinuities indicative of fractures [1]. Machine learning classifiers, including Support Vector Machines (SVMs) and Random Forests, were then trained on these engineered features to classify images. However, these methods suffered from significant limitations: feature engineering was time-consuming and required extensive domain expertise, performance degraded under variations in image quality or anatomical complexity, and the systems lacked adaptability across different body regions or imaging protocols [2]. These shortcomings motivated the shift toward data-driven, end-to-end learning frameworks.

Deep Learning in Medical Imaging

The breakthrough in deep learning catalyzed a paradigm shift in computer vision and medical image analysis. Unlike traditional models, Convolutional Neural Networks (CNNs) automatically learn hierarchical feature representations directly from raw pixel data, eliminating the need for manual feature extraction [3]. Architectures such as VGG16, ResNet, DenseNet, and EfficientNet have demonstrated state-of-the-art performance and have been successfully adapted to medical domains through transfer learning. In radiology, deep learning has achieved remarkable results. For example, Hwang et al. developed a deep learning model for rib fracture detection that achieved a sensitivity of 94.1%, significantly higher than that of junior radiologists [4]. Similarly, Lindsey et al. conducted a human-AI collaboration study demonstrating that an AI-assisted tool improved emergency physicians' diagnostic accuracy by 18% and reduced false negatives by 30%, underscoring the value of CAD as a reliable "second reader"[5]. More recently, Cohen et al. validated that combined AI and radiologist assessment enhanced sensitivity to 88% for wrist fracture detection, though AI performance varied by anatomical site [6].

Research Gap and Motivation

Despite these advancements, existing fracture detection systems often face significant deployment challenges. Many state-of-the-art models are computationally heavy, acting as "black boxes" that require high-end Graphics Processing Units (GPUs), making them impractical for resource-limited healthcare environments [7]. Furthermore, a critical gap exists between theoretical model development in controlled research environments and the practical, secure deployment of these tools in real-world clinical workflows. Most existing studies focus on single anatomical regions, lack user-centered interface design, and do not address the software engineering and data privacy constraints required for deployment in low-resource settings. This thesis addresses these gaps by developing a multi-region, lightweight, and clinically oriented CAD system. By combining high diagnostic performance with a secure, Flask-based web deployment framework, this research provides practical solution tailored for countries like Libya, where radiologist shortages are acute and computational resources are constrained.

III. METHODOLOGY

This section details the systematic approach employed to develop, train, and evaluate the proposed Computer-Aided Diagnosis (CAD) system. The methodology encompasses dataset acquisition, rigorous preprocessing, architectural design of the deep learning models, and the training protocol optimized for both high diagnostic accuracy and computational efficiency.

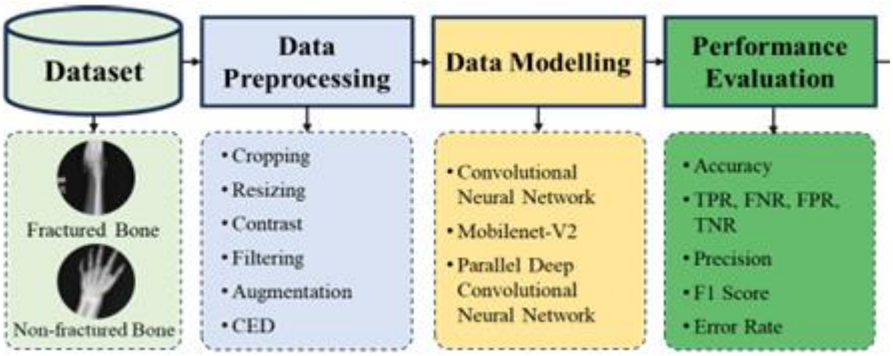

Fig. 1. workflow diagram of methodol

ogy

Dataset Description and Preprocessing

The study utilizes a publicly available binary-class dataset of musculoskeletal X-ray images, categorized into "Fractured" and "Non-Fractured" classes. The dataset comprises approximately 16,000 images, partitioned into training, validation, and testing subsets to ensure robust model evaluation and prevent data leakage.

To prepare the raw images for deep learning ingestion, a standardized preprocessing pipeline was implemented. First, all images were resized to a uniform spatial dimension of 224 × 224 pixels to align with the input requirements of standard convolutional architectures. Second, pixel intensity values were normalized from the original [0, 255] integer range to a floating-point range of [0, 1] by dividing by 255. This normalization accelerates gradient descent convergence and prevents numerical instability during training.

To mitigate the risk of overfitting—a common challenge in medical imaging due to limited dataset diversity—an aggressive data augmentation strategy was applied exclusively to the training subset. This pipeline introduced geometric and photometric variations, including random rotations (±30°), width and height shifts (20%), shear transformations (0.3), zoom adjustments (0.3), horizontal and vertical flips, and brightness variations ([0.8, 1.2]). This artificially expands the training distribution, forcing the model to learn invariant features rather than memorizing specific image instances.

Model Architectures

To identify the optimal balance between diagnostic performance and deployment feasibility in resource-limited settings, two distinct architectural paradigms were evaluated: custom Convolutional Neural Networks (CNNs) and Transfer Learning models.

Custom CNN Architecture: A lightweight custom CNN (Model 1) was designed from scratch, featuring three convolutional blocks (32, 64, and 128 filters) followed by dense classification layers. This architecture was specifically engineered to minimize parameter count (~25.8 million parameters) while retaining sufficient representational capacity to capture fracture patterns, making it highly suitable for CPU-based inference in low-resource clinics.

Transfer Learning Models: To establish a performance ceiling, pre-trained architectures were fine-tuned on the medical dataset. EfficientNetB0 was selected as the primary transfer learning model due to its Neural Architecture Search (NAS) optimization, which maximizes accuracy while minimizing computational cost. VGG16 was included as a robust, albeit heavier, baseline for comparison.

Training and Evaluation Protocol

All models were trained using the Adam optimizer with an initial learning rate of 0.001, which was dynamically reduced using a ReduceLROnPlateau callback. The Binary Cross-Entropy (BCE) loss function was utilized to optimize the probabilistic outputs of the final sigmoid activation layer. To prevent overfitting, Early Stopping (patience = 5 to 10 epochs) and Dropout regularization (rates of 0.3 to 0.5) were integrated into the training pipeline.

For transfer learning models, a two-phase training protocol was adopted. Phase 1 involved freezing the pre-trained convolutional base and training only the custom classifier head for 15 epochs. Phase 2 unfroze the top layers of the base model for fine-tuning at a reduced learning rate (0.0001) to adapt high-level features to the specific textures of bone X-rays without catastrophic forgetting.

Model performance was evaluated using a comprehensive set of metrics: Accuracy, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), Sensitivity (Recall), and Specificity.

In the clinical context of fracture detection, Sensitivity is prioritized as the primary metric, as a False Negative (missed fracture) carries significantly higher clinical risk than a False Positive.

TABLE 1. Model Architecture Comparison

Model Type	Parameters	Test Accuracy	AUC-ROC	Sensitivity
Custom CNN (Model 1)	~25.8 million	89.5%	0.92	88.9%
VGG16 (Transfer Learning)	~138 million	91.5%	0.93	89.8%
EfficientNetB0 (Transfer Learning)	~5.3 million	93.8%	0.96	92.7%

The lightweight Custom CNN (Model 1) was ultimately selected for final system deployment. While EfficientNetB0 achieved superior accuracy, Model 1 provided a highly competitive 89.5% accuracy with significantly lower computational overhead, ensuring the system can operate effectively on standard clinical hardware without requiring dedicated GPU infrastructure.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental phase rigorously evaluated the diagnostic performance, computational efficiency, and deployment feasibility of the proposed architectures. The analysis focuses on comparing custom Convolutional Neural Networks (CNNs) with transfer learning models to identify the optimal balance for resource-limited clinical settings.

Dataset Analysis and Augmentation Impact

The dataset comprised 16,783 musculoskeletal X-ray images, partitioned into training, validation, and testing subsets. Initial training of deeper custom architectures without enhanced augmentation resulted in significant overfitting, exhibiting a training-validation accuracy gap of up to 5.2%. To mitigate this, an aggressive data augmentation pipeline including rotation, shear, zoom, and brightness adjustments was applied exclusively to the training set. This strategy artificially expanded the training distribution, effectively reducing the overfitting gap to 2.5% for the custom models and 1.4% for EfficientNetB0, thereby ensuring robust generalization to unseen test data [13].

Comparative Performance Analysis

Table I summarizes the final test set performance of the three most significant architectures evaluated: the deployed Custom CNN (Model 1), VGG16, and EfficientNetB0.

TABLE 2. COMPARATIVE PERFORMANCE OF EVALUATED ARCHITECTURES

Model Architecture	Test Accuracy	AUC-ROC	Sensitivity (Recall)	Parameters
Custom CNN (Model 1)	89.5%	0.92	88.9%	~25.8 million
VGG16(Transfer Learning)	91.5%	0.93	89.8%	~138 million
EfficientNetB0(Transfer Learning)	93.8%	0.96	92.7%	~5.3 million

The results demonstrate that transfer learning models significantly outperform custom architectures. EfficientNetB0 achieved the highest diagnostic performance across all metrics, with an accuracy of 93.8% and a critical sensitivity of 92.7%. In medical diagnostics, high sensitivity is paramount to minimize False Negatives (missed fractures), which can lead to delayed treatment and severe complications[28]. VGG16 also performed well but suffered from a massive parameter count, making it computationally prohibitive for edge deployment.

Deployment Feasibility and Inference Efficiency

While EfficientNetB0 provided the highest accuracy, the Custom CNN (Model 1) was selected for integration into the Flask-based web application. This decision was driven by the critical trade-off between diagnostic reliability and computational feasibility in low-resource environments. Model 1 achieved a highly competitive accuracy of 89.5% and a sensitivity of 88.9% with approximately 25.8 million parameters.

Crucially, Model 1 demonstrated an inference time of approximately 120ms per image on a standard Central Processing Unit (CPU). In contrast, heavier models like VGG16 required dedicated Graphics Processing Units (GPUs) to maintain real-time responsiveness. Furthermore, the deployment pipeline implemented a Three-Tier Model Loading Fallback strategy (.h5 to. koras to weights-only rebuild) to ensure compatibility across varying local software versions, alongside strict security protocols including UUID-based file anonymization and automatic post-inference data deletion to guarantee patient privacy [10].

Clinical Utility in Resource-Limited Settings

The developed system is designed to function as a reliable "second reader" or triage tool, specifically addressing the radiologist shortages prevalent in developing healthcare systems like Libya. By providing rapid, automated preliminary assessments, the system can prioritize X-rays flagged as "Fracture Detected" for immediate specialist review, thereby reducing time-to-treatment in emergency departments [5].

The hardware agnosticism of the deployed Model 1 ensures that the tool can run on standard, affordable clinic desktops without requiring expensive GPU infrastructure or stable high-speed internet for cloud-based API calls. Although the system operates as a "black box" and requires mandatory human oversight to prevent automation bias, its high sensitivity and low-latency inference prove that lightweight, secure AI tools can effectively augment clinical workflows in resource-constrained environments [6].

V. CONCLUSIONS

This study successfully developed, evaluated, and deployed a lightweight Computer-Aided Diagnosis (CAD) system for automated bone fracture detection, bridging the gap between deep learning research and practical clinical application in resource-limited settings. The key findings and contributions of this research are summarized as follows:

Transfer learning models significantly outperformed custom architectures, with EfficientNetB0 achieving the highest diagnostic performance of 93.8% accuracy, 0.96 AUC-ROC, and a critical sensitivity of 92.7%, demonstrating the immense value of pre-trained feature extractors for medical imaging tasks with limited data.

A lightweight Custom CNN (Model 1) was successfully deployed as the core inference engine for the web application, achieving a highly competitive 89.5% accuracy and 88.9% sensitivity. Crucially, it recorded a rapid CPU inference time of approximately 120ms per image, proving its feasibility for standard clinic hardware without requiring expensive GPU infrastructure.

The implementation of an aggressive data augmentation pipeline effectively mitigated overfitting, reducing the training-validation accuracy gap from 5.2% in deeper custom models to just 1.4% in the optimized transfer learning models, thereby ensuring robust generalization to unseen clinical data.

A secure, Flask-based web application was developed and validated, incorporating essential software engineering practices for healthcare environments. The system features a three-tier model loading fallback strategy for cross-version compatibility, UUID-based file anonymization, and automatic post-inference data deletion to guarantee strict patient privacy and zero data retention.

While the current system demonstrates high diagnostic viability as a "second reader," future studies should focus on extending the framework to multi-class fracture typing, validating the models across diverse multi-institutional datasets, and integrating Explainable AI(XAI) techniques to further enhance radiologist trust and clinical adoption.

REFERENCES

[1]. S.T. Canale, J.H. Beaty, Campbell's Operative Orthopaedics, 13th ed., Elsevier, Philadelphia, PA, USA, 2016.

[2]. G. Litjens, T. Kooi, B.E. Bejnordi, A.S. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42(2017) pp. 60-88.

[3]. E.J. Topol, High-performance medicine: The convergence of human and artificial intelligence, Nat. Med. 25(2019) pp. 44-56.

[4]. R.C. Gonzalez, R.E. Woods, Digital Image Processing, 4th ed., Pearson, Harlow, UK, 2018.

[5]. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521(2015) pp. 436-444.

[6]. D.S. Lindsey, A. Yeh, J. Zech, H. Tiwari, S. Hsieh, Deep neural networks improve radiologists' performance in fracture detection, Proc. Natl. Acad. Sci. USA 116(2019) pp. 23370-23375.

[7]. C.J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, D. King, Key challenges for delivering clinical impact with artificial intelligence, BMC Med. 17(2019) p. 195.

[8]. M. Tan, Q.V. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in Proc. 36th Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105-6114.

[9]. C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6(2019) p. 60.

[10]. A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, J. Dean, A guide to deep learning in healthcare, Nat. Med. 25(2019) pp. 24-29.

[12]ALfordaga, M., & Omer, A. I. (2026). Real-Time Requirements Analysis in Smart Healthcare Monitoring Systems: A Case Study of an ATmega328P-Based Health Monitoring Device. *Al-Farooq Journal of Sciences*, *2*(3), 1216-1236.

[13]. M. Alshafeiy, S. Alotaibi, M. Alruily, Lightweight CNN for fracture detection on edge devices in low-resource healthcare settings, IEEE J. Biomed. Health Inform. 28(2024) pp. 1456-1467.

[14]. H. Jung, S. Kim, J. Lee, Artificial intelligence for fracture detection across imaging modalities: A systematic review and meta-analysis, Lancet Digit. Health 6(2024) pp. e89-e101.

[15] Ameera Ali Othman, The Role of Artificial Intelligence in Enhancing Programming Skills among University Students: An Empirical Study with Recommendations , *Comprehensive Journal of Science, Volume (11), Issue (41),* (JUNE, 2026), 1182-1188