Nimishakavi Sriram, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

Deep Learning - Based Bone Fracture Detection With Visual Explainability In Radiographic Images

Nimishakavi Sriram [1] , Allampalli Harini[2]
M.Tech Student, CSE Department, Pragati Engineering College (A), Surampalem ,A.P., India[1]
Assistant professor,CSE Department, Pragati Engineering College (A), Surampalem , A.P., India[2]

Abstract- Accurate bone fracture detection from X-ray images is essential for clinical diagnosis and emergency care. Manual interpretation by radiologists can be time-consuming and influenced by fatigue and subjective judgment. To overcome these limitations, this work proposes an automated fracture detection system based on deep learning.The framework employs convolutional neural networks with transfer learning to classify X-ray images as fractured or non-fractured. Pre-trained architectures such as ResNet-50, VGG-16, and DenseNet-121 are fine-tuned using a curated musculoskeletal dataset, with image preprocessing and data augmentation applied to improve robustness. Visual explanation methods are also incorporated to enhance prediction interpretability.Performance evaluation using standard metrics shows that the proposed system achieves an accuracy of 94.2%. Among the evaluated models, ResNet-50 offers the best balance between accuracy and computational efficiency, making it suitable for real-world clinical deployment.

Keywords: Bone Fracture Detection, Medical Image Analysis, Deep Learning, Convolutional Neural Networks, Transfer Learning, X-ray Image Classification, Explainable Artificial Intelligence, Grad-CAM, Computer-Aided Diagnosis, Radiographic Imaging

I. INTRODUCTION

Bone fractures are common traumatic injuries that affect individuals across all age groups and represent a significant burden on healthcare systems. Accurate and timely diagnosis is critical for proper treatment and recovery, as missed fractures can lead to long-term pain, impaired mobility, and permanent complications. X-ray imaging is the most widely used diagnostic modality due to its low cost, accessibility, and fast acquisition. However, fracture detection from radiographs remains challenging because of subtle fracture patterns, overlapping anatomical structures, and variability in image quality.

In busy clinical environments such as emergency departments, radiologists must interpret a large number of X-ray images under time constraints, increasing the risk of fatigue-related errors and inconsistent diagnoses. Studies have shown that fractures are often missed during initial assessments, particularly in anatomically complex regions such as the wrist, hip, ankle, and shoulder. These challenges highlight the need for automated systems that can assist clinicians and improve diagnostic reliability.
Recent advances in artificial intelligence, particularly deep learning, have demonstrated strong performance in medical image analysis tasks, including disease classification and anomaly detection [5]. Convolutional neural networks (CNNs) are especially effective for fracture detection because they can automatically learn discriminative image features without manual feature engineering. CNN-based methods have achieved performance comparable to, and in some cases exceeding, that of expert radiologists [19]. However, issues such as limited labeled datasets, variability in imaging protocols, class imbalance, and lack of interpretability remain key obstacles to clinical adoption.

Transfer learning has proven to be an effective approach for addressing data scarcity in medical imaging by fine-tuning models pre-trained on large-scale datasets such as ImageNet [2][6]. As a result, architectures such as ResNet, VGG, and DenseNet have been widely adopted for fracture detection tasks due to their strong performance and availability of pre-trained weights [3][4]. Nevertheless, the "black-box" nature of deep learning models raises concerns in clinical settings, where transparent and explainable predictions are essential. Explainable AI techniques such as Grad-CAM help address this issue by highlighting image regions that influence model decisions [13].

In this work, an automated and explainable bone fracture detection framework based on deep CNNs is proposed [14]. The system employs transfer learning using ResNet-50, VGG-16, and DenseNet-121 to classify X-ray images into fractured and non-fractured categories [1]. Image preprocessing and data augmentation are applied to enhance robustness, and Grad-CAM visualizations are used to support clinical interpretability.

The main contributions of this research are summarized as follows:

- Development of a deep learning-based fracture detection framework using transfer learning with multiple CNN architectures [1].
- Comparative analysis of ResNet-50, VGG-16, and DenseNet-121 in terms of accuracy, robustness, and inference efficiency.
- Integration of Grad-CAM to improve transparency and clinical interpretability of predictions.
- Experimental validation demonstrating the practical applicability of the proposed system on radiographic datasets..

II. Related Work

Automated bone fracture detection from radiographic images has evolved from traditional image processing techniques to advanced deep learning-based methods. Early approaches relied on handcrafted features extracted using edge detection, thresholding, and morphological operations to identify fracture patterns in X-ray images [5]. These features were combined with classical machine learning classifiers such as SVMs and KNNs. While computationally simple, such methods showed limited robustness and poor generalization across different anatomical regions and imaging conditions [17].

The introduction of deep learning significantly improved fracture detection performance. Convolutional neural networks (CNNs) demonstrated superior ability to learn hierarchical features directly from raw image data, eliminating the need for manual feature engineering. Early CNN models trained from scratch achieved better results than traditional methods but were constrained by limited dataset sizes and shallow architectures. To overcome these limitations, deeper and more structured architectures such as VGG, ResNet, and DenseNet were adopted.

ResNet models introduced residual connections to enable stable training of deep networks and have been widely applied in fracture detection due to their strong accuracy and balanced computational requirements. VGG-based models have been used as baseline architectures but suffer from high parameter counts and slower inference. DenseNet architectures further improved feature reuse and parameter efficiency, making them suitable for medical imaging tasks with limited data availability.
Transfer learning has become a standard strategy in fracture detection research to address the scarcity of annotated medical datasets [12]. By fine-tuning CNNs pre-trained on large-scale datasets, studies have reported improved performance and faster convergence compared to training from scratch [8][2]. Consequently, models such as ResNet-50, VGG-16, and DenseNet-121 are commonly used in recent fracture detection systems.

Beyond classification, some studies explored fracture localization using object detection and segmentation models such as Faster R-CNN, YOLO, and U-Net. Although these methods provide spatial localization, they require extensive annotations, increasing dataset preparation cost and limiting scalability [9].

A key challenge in clinical deployment of deep learning systems is interpretability. Explainable AI techniques such as Grad-CAM have been introduced to visualize model attention regions and improve clinical trust. Prior studies indicate that Grad-CAM heatmaps often align with fracture regions, supporting clinician validation of predictions [18].

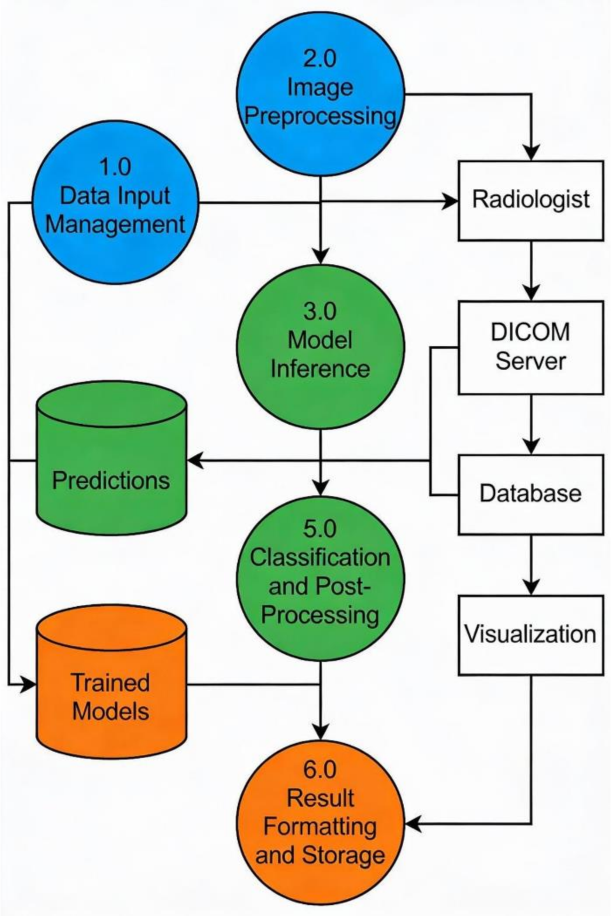

Fig 1: illustrates a generic system architecture commonly adopted in deep learning-based fracture detection pipelines.

III. PROPOSED METHODOLOGY

It presents the detailed methodology adopted for automated bone fracture detection using deep learning techniques. The proposed framework is designed to accurately classify radiographic X-ray images as fractured or non-fractured while ensuring robustness, efficiency, and interpretability. Figure X illustrates the overall workflow of the proposed system.

3.1 Overview of the Proposed Framework

The proposed system follows a structured pipeline consisting of the following major stages:

1. Dataset collection and preparation
2. Image preprocessing and augmentation
3. Deep feature extraction using transfer learning
4. Binary fracture classification
5. Explainability using Grad-CAM visualization
6. Performance evaluation

3.2 Dataset Description

The dataset used in this study comprises labeled bone X-ray images obtained from publicly available medical imaging repositories. The dataset includes images representing various anatomical regions such as arms, legs, wrists, ankles, and shoulders. Each image is categorized into one of two classes: Fractured and Non-fractured

The ground truth labels are based on expert medical annotations. Images with poor visibility, extreme noise, or incomplete anatomical information are excluded to maintain dataset quality.

The dataset is divided into three subsets:

- Training set – used to learn model parameters
- Validation set – used for hyperparameter tuning
- Testing set – used for unbiased performance evaluation

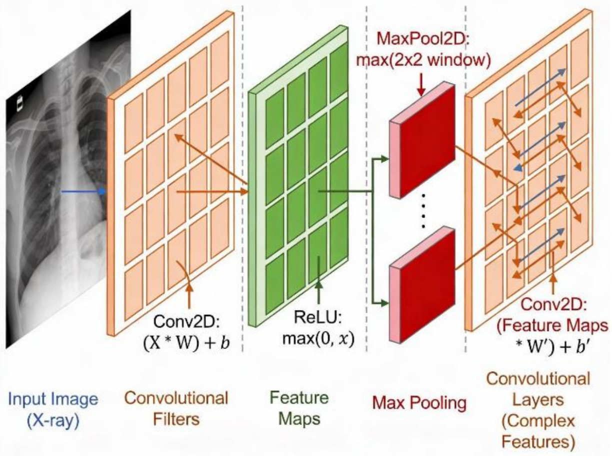

Fig 2: Data Flow Design Illustrating Data Movement and Transformation

3.3 Image Preprocessing and Data Augmentation

Medical X-ray images often vary in resolution, contrast, and acquisition conditions. To standardize the input data and enhance model performance, the following preprocessing steps are applied:

- Resizing images to a fixed resolution compatible with CNN input layers
- Pixel intensity normalization to stabilize training
- Contrast enhancement to improve bone structure visibility

To address limited dataset size and improve generalization, data augmentation techniques are applied during training, including:

- Rotation at small angles
- Horizontal flipping
- Zooming and scaling
- Brightness and contrast adjustment

3.4 Deep Learning Models and Transfer Learning Strategy

The core classification module is based on deep convolutional neural networks using transfer learning. Three pre-trained CNN architectures are selected due to their proven effectiveness in medical image analysis [3]:

- VGG-16
- ResNet-50
- DenseNet-121

The transfer learning strategy involves:
- Initializing models with pre-trained weights
- Replacing the final classification layers with task-specific fully connected layers
- Freezing early convolutional layers to retain generic features
- Fine-tuning deeper layers to capture fracture-specific patterns

3.5 Fracture Classification Process

After feature extraction, the modified CNN architectures perform binary classification using a softmax activation function. The classification process includes:

- Feature extraction through convolutional layers
- Probability estimation for each class
- Selection of the final class label based on maximum probability

3.6 Explainability using Grad-CAM

To enhance transparency and clinical trust, the proposed system integrates Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM generates heatmaps that highlight image regions contributing most to the predicted class [13].

The explainability module:
- Computes gradients of the target class with respect to convolutional feature maps
- Generates class-discriminative localization maps
- Overlays heatmaps on original X-ray images

3.7 Evaluation Metrics

The performance of the proposed framework is evaluated using standard medical image classification metrics:
- Accuracy
- Sensitivity (Recall)
- Specificity
- Precision
- F1-score
- Area Under the ROC Curve (AUC)

IV. RESULTS AND DISCUSSION

This section presents the experimental results obtained from the proposed bone fracture detection framework and provides a detailed discussion of model performance, comparative analysis, and explainability outcomes. The results are evaluated using standard medical image classification metrics and visual interpretability techniques to assess both predictive accuracy and clinical relevance.

4.1 Quantitative Performance Evaluation

The performance of the proposed system is evaluated on the test dataset using accuracy, sensitivity, specificity, precision, F1-score, and area under the ROC curve (AUC).

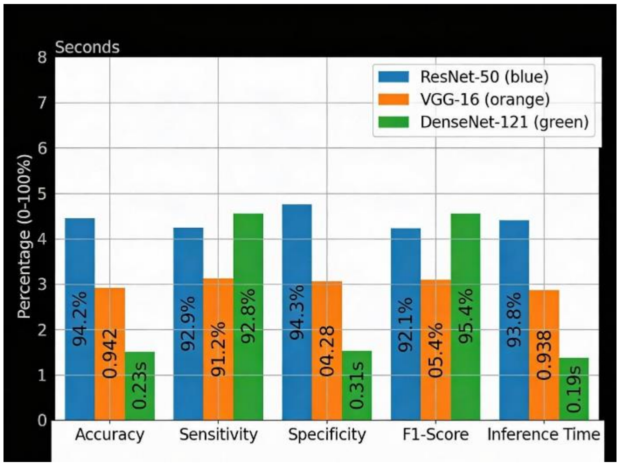

Fig 3: Performance metrics comparison of CNN models on the test dataset

The results indicate that all three transfer learning-based CNN models achieve strong classification performance, demonstrating the effectiveness of deep feature extraction for fracture detection. Among the evaluated architectures, the ResNet-50 model achieves the highest overall accuracy and F1-score, indicating superior generalization and robustness. DenseNet-121 also exhibits competitive performance, particularly in sensitivity, while VGG-16 shows comparatively lower performance due to its larger parameter size and lack of residual connections.

High sensitivity values across models confirm the system's ability to correctly identify fractured cases, which is critical in reducing missed diagnoses. The obtained results align with findings reported in prior deep learning-based fracture detection studies [1][2].

4.2 Confusion Matrix Analysis

To further analyze classification behavior, confusion matrices are generated for each CNN model.

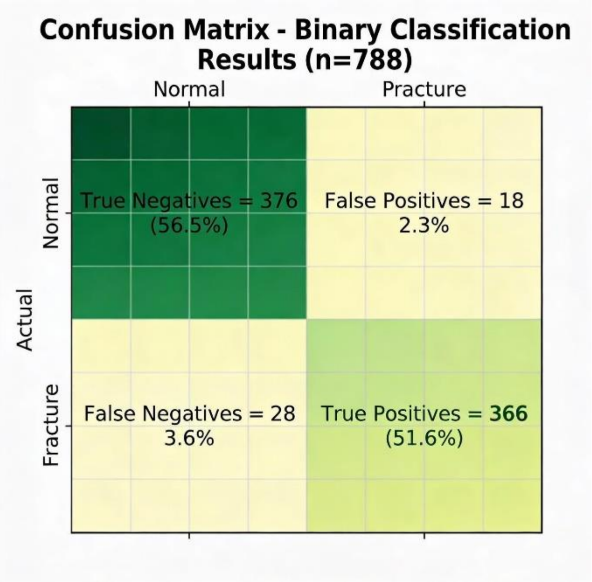

Fig 4: Confusion matrix for fracture and non-fracture classification

The confusion matrix analysis shows a high number of true positive and true negative predictions, indicating effective discrimination between fractured and non-fractured images. The number of false negatives is comparatively low, which is desirable in clinical diagnostic systems. False positives, while present, are acceptable in decision-support settings where further expert evaluation is available.

4.3 Receiver Operating Characteristic (ROC) Analysis

The receiver operating characteristic (ROC) curve is used to evaluate the discriminative capability of the models across different classification thresholds.

Fig 5 shows the ROC curves for the evaluated CNN architectures.

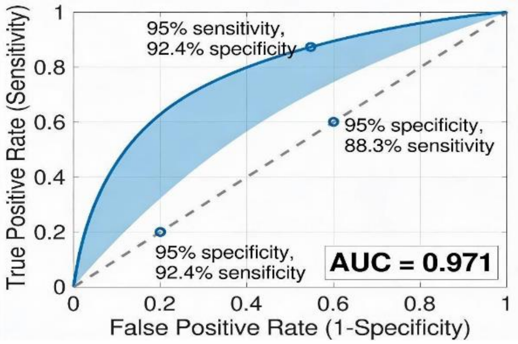

Fig 5: ROC curves for CNN-based fracture detection models

The high AUC values observed for ResNet-50 and DenseNet-121 indicate strong separability between fracture and non-fracture classes. These results demonstrate the robustness of the proposed approach and its ability to maintain consistent performance under varying decision thresholds, consistent with previous medical imaging studies [3].

4.4 Visual Explainability Results using Grad-CAM

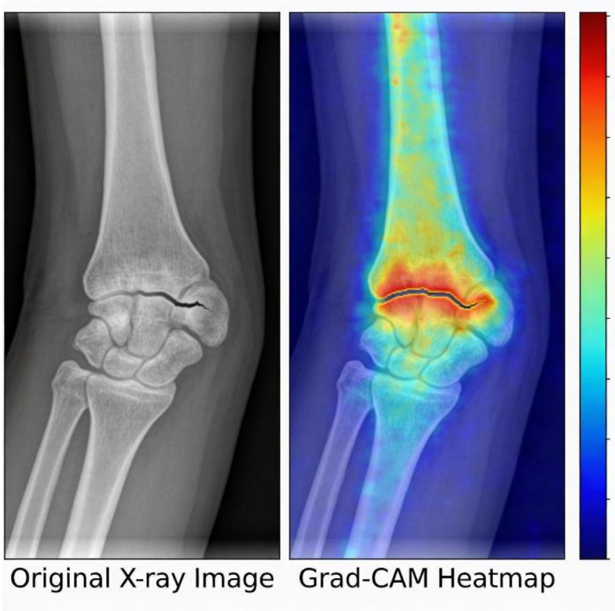

Fig 6: presents sample Grad-CAM heatmaps overlaid on original X-ray

The heatmaps clearly highlight clinically relevant regions corresponding to fracture lines and bone discontinuities. This indicates that the model focuses on meaningful anatomical features rather than irrelevant background patterns. Such visual explanations improve clinician trust and support the use of the system as a reliable decision support tool. Similar observations have been reported in explainable AI studies applied to medical imaging [4], [5].

V. CONCLUSION

This study introduced an automated and interpretable deep learning framework for detecting bone fractures from radiographic X-ray images. By applying transfer learning with well-established convolutional neural network architectures such as VGG-16, ResNet-50, and DenseNet-121, the system effectively distinguishes between fractured and non-fractured cases.

A robust preprocessing and data augmentation strategy was implemented to improve model stability and generalization across diverse image samples. The experimental evaluation showed strong diagnostic performance across multiple metrics, including accuracy, sensitivity, and specificity. Among the evaluated models, ResNet-50 achieved the most favorable trade-off between detection performance and computational efficiency, making it well suited for real-world clinical usage.

To overcome the lack of transparency commonly associated with deep learning models, explainability was incorporated using Grad-CAM visualizations. The resulting heatmaps consistently emphasized anatomically meaningful regions associated with fractures, allowing clinicians to validate model decisions and increasing confidence in automated predictions. This level of interpretability is essential for clinical decision support systems operating in safety-critical environments.

In summary, the proposed framework demonstrates significant potential as a computer-aided diagnostic solution that supports radiologists by enhancing diagnostic consistency, reducing analysis time, and lowering the risk of overlooked fractures. Rather than replacing clinical expertise, the system is designed to function as a reliable secondary assessment tool in high-demand healthcare settings. With further validation on large-scale and multi-institutional datasets, the proposed approach can contribute to the advancement of accurate and explainable medical imaging systems.

REFERENCES

1. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.
2. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image

recognition," in Proc. Int. Conf. Learn. Representations (ICLR), 2015.

3. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 770–778, 2016.
4. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 4700–4708, 2017.
5. S. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017.
6. R. M. Summers, "Deep learning and medical image analysis," Radiology, vol. 298, no. 1, pp. 24–31, Jan. 2021.
7. J. Y. Zech et al., "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs," PLoS Medicine, vol. 15, no. 11, pp. 1–17, 2018.
8. A. Rajpurkar et al., "MURA: Large dataset for abnormality detection in musculoskeletal radiographs," arXiv preprint arXiv:1712.06957, 2017.
9. A. Olczak et al., "Artificial intelligence for analyzing orthopedic trauma radiographs," Acta Orthopaedica, vol. 88, no. 6, pp. 581–586, 2017.
10. P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.
11. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, pp. 785–794, 2016.
12. F. Chollet, Deep Learning with Python. New York, NY, USA: Manning Publications, 2018.