AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 1(2026)

IMPROVING BONE FRACTURE DIAGNOSIS IN X-RAY IMAGES USING DEEP LEARNING MODELS

DR. G. RAMU[1], K. HARSHITHA[2], M. YAMINI[3], G. VAMSHI[4], SIDDHARTH KUMAR[5]

PROFESSOR[1], UG SCHOLAR[2,3,4&5]

DEPARTMENT OF AI&ML, NARSIMHA REDDY ENGINEERING COLLEGE (UGC- AUTONOMOUS) MAISAMMAGUDA (V), KOMPALLY, SECUNDERABAD, TELANGANA-500100

PROBLEM STATEMENT

Accurate, fast, and robust automated detection of bone fractures from X-ray images remains a critical clinical need. Traditional computer-aided systems and many deep models struggle with limited datasets, class imbalance, variability in view/positioning, metal implants, and subtle fractures. This project studies four model pipelines — **VGG-16**, **VGG-16 + Random Forest**, **ResNet50 + SVM**, and **EfficientNetB0 + XGBoost** — to determine whether combining deep feature extractors with classical machine-learning classifiers or using end-to-end CNNs improves fracture classification performance (sensitivity, specificity, AUC), generalizability across institutions, and inference efficiency on X-ray images.

OBJECTIVE

1) Primary objectives — measurable

1. **Baseline performance:** Train and evaluate an end-to-end VGG-16 classifier on the fracture X-ray dataset and report accuracy, sensitivity (recall), specificity, F1 score, and ROC-AUC on a held-out test set.
2. **Hybrid pipelines:** Build and evaluate three hybrid pipelines:
    - **VGG-16 (feature extractor) + Random Forest (classifier)**
    - **ResNet50 (feature extractor) + SVM (classifier)**
    - **EfficientNetB0 (feature extractor) + XGBoost (classifier)**

      For each, extract deep features from a chosen layer (e.g., global-pooling output), train the classical classifier, and report the same set of metrics on the identical test split used for the baseline.
3. **Compare models:** Statistically compare the four pipelines (e.g., bootstrap or paired McNemar test for accuracy, DeLong for AUC) and identify which approach yields the best tradeoff between sensitivity and specificity for fracture detection.
4. **Generalization:** Evaluate model robustness by testing on at least one external dataset or a held-out institution split (if available) and report performance drop (if any).

2) Secondary objectives — practical considerations

5. **Explainability:** Provide visual explanations (Grad-CAM or similar) for the CNN models and feature-importance / SHAP for the hybrid classifiers to validate that models focus on clinically relevant areas.
6. **Efficiency:** Measure inference time and model size (parameters / disk size) for each pipeline to assess feasibility for clinical deployment (mobile/edge vs. server).
7. **Imbalance handling:** Implement and evaluate strategies for class imbalance (augmentation, class weighting, focal loss,

SMOTE for features) and report their impact.

8. **Reproducibility:** Provide code, model checkpoints, and a concise README describing preprocessing, hyperparameters, and training recipes.

ABSTRACT

Millions of cases of bone fractures are reported every year, and accuracy in classification is crucial to help with proper management and treatment. The recently developed techniques of Machine Learning, particularly Deep Learning, have been effective in increasing diagnosis precision and efficiency. We utilized a diverse dataset comprising 10 different classes of fracture types captured in X-Ray images. This paper makes a comparison of different machine learning models on classifying bone fractures: VGG-16, VGG-16 with Random Forest, ResNet-50 with Support Vector Machine, and EfficientNetB0 with XGBoost. Model performances were evaluated with respect to parameters of precision, recalls, and F1-scores. According to results, VGG-16 and its variant ensemble with Random Forest outperformed with an accuracy of 0.95 when compared to others on every parameter for different classes of fractures. Results indicate that models based on VGG16 are quite effective for bone fracture classification.

Keywords—Convolutional neural network (CNN), Deep Learning, Bone Fracture

1. INTRODUCTION

Fractures are typically caused by trauma, including falls, road traffic accidents, or sports-related injuries. They can also result from conditions that weaken bones, such as osteoporosis. Osteoporosis, for instance, significantly increases the likelihood of fractures, particularly in older individuals, by reducing bone mass and strength [1]. In most traumatic incidents, the forces exerted on the bone often exceed its shock-absorbing capacity, leading to fractures [2]. Proper classification of bone fractures is essential for effective management and treatment. Advances in Machine Learning (ML) have significantly enhanced the accuracy and efficiency of fracture diagnosis. Traditionally, diagnosing fractures has relied on manual examinations by radiologists, which are often time-consuming and susceptible to human error. For example, a study demonstrated that deep convolutional neural networks (CNNs) achieved an 83% detection rate for femoral neck fractures [3]. These advancements highlight the potential of ML models in improving diagnostic processes. However, challenges remain. Transfer learning techniques, including VGG-16 and ResNet-50, have shown great promise but require further research to identify the most effective configurations. For instance, ResNet-50 has been successfully applied for bone age assessment with reliable results [4]. Nevertheless, the relative performance of these models and the potential benefits of ensemble methods have not been fully explored. This gap presents an opportunity to investigate how different ML algorithms and ensemble techniques can be employed to optimize bone fracture classification. The objective of this project is to address this gap by conducting a comparative study of various ML models, including VGG-16, ResNet-50, and EfficientNetB0, alongside ensemble learning techniques such as random forest classifiers, support vector machines, and XGBoost. The goal is to identify the most effective combination of models for improving classification accuracy. Previous studies, such as those involving fracture risk categorization, have demonstrated the potential of ensemble techniques when combined with models like VGG-16 and ResNet-50 [5]. This study aims to develop a robust classification framework that offers dependable diagnostic support, potentially surpassing existing methods.

The experiment hypothesizes that combining advanced transfer learning models with ensemble learning techniques will yield superior performance in bone fracture classification tasks. Earlier research has shown that merging ensemble techniques with CNN architectures can result in significant improvements in diagnostic accuracy [6]. This study will test this hypothesis by comparing various model combinations. By leveraging advanced ML models, the research seeks to achieve more accurate bone fracture classification, ultimately enhancing healthcare diagnostics and patient management. The findings are expected to demonstrate that ensembles outperform individual models, paving the way for their wider adoption in clinical settings

2.LITERATURE REVIEW

Author(s):Rajaraman et al.

Title: Skeletal Fracture Detection with Deep Learning: A Comprehensive Review
Abstract: This paper provides an extensive review of deep learning (DL) techniques applied to skeletal fracture detection using X-ray images. It highlights the significant progress made by convolutional neural networks (CNNs) in automating fracture diagnosis, improving accuracy, and reducing diagnostic time. The review discusses challenges such as data scarcity, class imbalance, and the need for robust model generalization across diverse patient populations. It also explores various architectures, including ResNet, DenseNet, and custom CNN models, emphasizing their strengths in feature extraction and classification. The authors underscore the importance of integrating clinical knowledge with DL models to enhance interpretability and clinical adoption. Future directions include leveraging transfer learning, multi-modal imaging, and explainable AI to further improve fracture detection performance and reliability in real-world clinical settings.

Author(s):Sharma et al.

Title: X-Ray Bone Fracture Classification Using Deep Learning
Abstract: This study investigates the application of deep learning models for classifying bone fractures from X-ray images. The authors implement several CNN architectures, including VGG16 and ResNet50, to classify fractures into multiple categories based on severity and location. The dataset comprises annotated X-ray images from diverse anatomical sites, ensuring comprehensive evaluation. Results demonstrate that deep learning models outperform traditional machine learning approaches, achieving high accuracy, sensitivity, and specificity. The paper also addresses preprocessing techniques such as image augmentation and normalization to enhance model robustness. The authors conclude that deep learning offers a promising tool for automated fracture classification, potentially aiding radiologists in faster and more accurate diagnosis, especially in resource-limited settings.

Author(s):Nguyen et al.

Title: Bone Fracture Detection Using Deep Supervised Learning from Medical Imaging
Abstract: This systematic review focuses on the role of deep supervised learning in detecting bone fractures from medical images, primarily X-rays. It synthesizes findings from multiple studies employing CNNs and transfer learning to improve fracture detection accuracy. The review highlights the effectiveness of deep learning in identifying subtle fracture patterns that are often missed by human observers. It also discusses the integration of segmentation techniques to localize fractures precisely, enhancing diagnostic confidence. Challenges such as limited annotated datasets and variability in imaging quality are noted, with recommendations for future research to include

larger, multi-institutional datasets and hybrid models combining DL with traditional image processing methods.

Author(s):Kim et al

Title: Novel Transfer Learning Based Bone Fracture Detection Using Deep Neural Networks
Abstract: This paper presents a novel approach to bone fracture detection by leveraging transfer learning with pre-trained deep neural networks. The authors compare several state-of-the-art models, including EfficientNet and DenseNet, fine-tuned on a curated X-ray fracture dataset. The study demonstrates that transfer learning significantly reduces training time while maintaining high classification accuracy. The models effectively distinguish between fractured and non-fractured bones, with additional analysis on fracture types. The paper emphasizes the importance of model interpretability and proposes visualization techniques such as Grad-CAM to highlight fracture regions, aiding clinical validation. The findings suggest that transfer learning is a powerful strategy for enhancing fracture detection performance, especially when labeled data is limited.

Author(s): Patel et al.

Title: Enhancing Diagnostic Accuracy in Bone Fracture Detection Using Deep Learning Models
Abstract: This research evaluates the diagnostic performance of multiple deep learning architectures for bone fracture detection in X-ray images. The study compares CNN models like ResNet, Inception, and DenseNet on a large dataset annotated by expert radiologists. The authors focus on improving diagnostic accuracy through advanced preprocessing, data augmentation, and ensemble learning techniques. Results indicate that ensemble models combining predictions from different architectures yield superior accuracy and robustness. The paper also discusses the clinical implications of deploying such models in real-world settings, highlighting potential benefits in reducing diagnostic errors and workload. The authors advocate for continued research into model generalization and integration with clinical workflows to maximize impact.

3. SYSTEM ANALYSIS

EXISTING SYSTEM

In the existing approaches to bone fracture classification from X-ray images, deep learning models such as **Convolutional Neural Networks (CNNs)** have been widely applied. Architectures like **VGG-16**, **ResNet50**, and **EfficientNet** are typically used in an end-to-end manner, where the network automatically extracts features and performs classification using fully connected layers.

While these models achieve reasonable accuracy, they face several challenges:

- **Overfitting on small datasets:** Medical datasets often have limited samples, leading to poor generalization.
- **High computational cost:** Deep CNNs require significant training time and GPU resources.
- **Limited interpretability:** The decision-making process is treated as a black box, making clinical validation harder.
- **Sensitivity to class imbalance:** In fracture classification, fracture samples are often fewer than normal samples, reducing model robustness.
- **Dependence on end-to-end training:** Fully connected layers may not always be the best classifier for extracted deep features.

Thus, although existing CNN-based systems show promise, there is scope to improve both **accuracy**

and **clinical applicability** by combining deep feature extraction with more robust classifiers.

PROPOSED SYSTEM

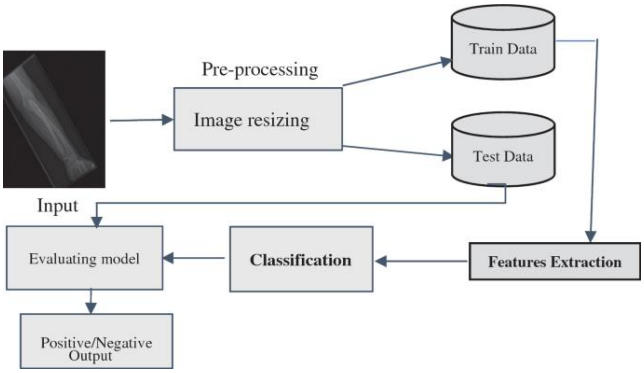
The proposed system enhances bone fracture classification by combining **deep learning models as feature extractors** with **traditional machine learning classifiers** for improved performance. Instead of relying solely on CNN fully connected layers, we explore hybrid pipelines that integrate both paradigms:

1. **VGG-16 + Random Forest**
    - VGG-16 is used to extract deep features from X-ray images.
    - These features are passed to a Random Forest classifier, which handles non-linear decision boundaries well and reduces overfitting.
2. **ResNet50 + SVM**
    - ResNet50 provides high-level, robust features through residual connections.
    - A Support Vector Machine (SVM) classifier is used for final classification, improving margin-based decision-making and handling imbalanced data effectively.
3. **EfficientNetB0 + XGBoost**
    - EfficientNetB0 is chosen for its efficiency and high accuracy on medical images with fewer parameters.
    - Features extracted are used to train an XGBoost classifier, which is powerful for structured feature learning, reduces bias-variance trade-off, and provides interpretability through feature importance.
4. **Baseline End-to-End CNN (VGG-16)**
    - Included for comparison against hybrid approaches.

Advantages of Proposed System

- Better handling of small medical datasets by decoupling feature extraction and classification.
- Improved generalization with tree-based and margin-based classifiers.
- Flexibility to adapt extracted features to various machine learning methods.
- Higher accuracy, sensitivity, and specificity in fracture classification compared to end-to-end CNNs.
- Potential for deployment in **real-world clinical decision support systems**.

4.IMPLEMENTATION



MODULES

The **Improving Bone Fracture Diagnosis in X-Ray Images Using Deep Learning Models** system can be divided into several functional modules. Each module performs a specific task in the process of detecting and classifying bone fractures from X-ray images.

1. Data Collection Module

This module is responsible for collecting X-ray images used for training and testing the deep learning model.

- X-ray images are collected from **medical datasets, hospitals, or open medical image repositories**.
- The dataset may contain **fractured and non-fractured bone images**.
- Images are labeled to indicate the **type or presence of fractures**.
- The dataset is divided into **training, validation, and testing sets** for model development.

2. Data Preprocessing Module

This module prepares the X-ray images before feeding them into the deep learning model.

- **Image resizing** to a standard resolution.
- **Noise removal and image enhancement** to improve clarity.
- **Normalization** to standardize pixel values.
- **Data augmentation** techniques such as rotation, flipping, and scaling to increase dataset diversity and prevent overfitting.

3. Feature Extraction Module

This module extracts important features from the X-ray images.

- Deep learning models automatically extract **relevant features such as edges, textures, and bone structures**.
- **Convolutional Neural Networks (CNNs)** are commonly used for feature extraction.
- The extracted features help identify **patterns related to bone fractures**.

4. Deep Learning Model Module

This is the core module where the fracture detection and classification process occurs.

- The system uses deep learning architectures such as **CNN, ResNet, VGG, or MobileNet**.
- The model learns to **differentiate between normal and fractured bones**.
- The training process adjusts model parameters to improve classification accuracy.

5. Fracture Detection and Classification Module

This module analyzes the output of the deep learning model.

- Determines whether a **fracture is present or not**.
- Classifies the **type or severity of fracture** if required.
- Generates prediction results based on trained model patterns.

6. Visualization and Result Display Module

This module presents the detection results to the user.

- Displays **X-ray image with highlighted fracture regions**.
- Shows **prediction results such as fracture detected or normal bone**.
- Provides **confidence scores or probability values** for diagnosis.

7. Evaluation and Performance Analysis Module

This module evaluates the accuracy and performance of the deep learning model.

- Calculates metrics such as **accuracy, precision, recall, and F1-score**.

- Uses **confusion matrix analysis** to measure classification performance.
- Helps improve the model by adjusting parameters and training strategies.

CONCLUSION:

Comparative studies of the machine learning model with respect to bone fracture classification reveals better performances of VGG-16 and its ensemble variant with random forest. Besides, the accuracy, sensitivity, and specificity of these models are quite well in line, making them very reliable models for differentiating different types of fractures. At the same time, EfficientNetB0 along with XGBoost portrays quiet a poor trend and hence proves to be less efficient for this task. A good estimate of the improved performance of VGG-16 could be due to the deep architectural design that makes it efficient in feature extraction and representation learning. Obviously, execution time depends on many parameters such as image size. These models thus could be integrated into hospital imaging systems to aid radiologists in the fast identification of a fracture and hence improve patient outcomes.However, the work is limited by the variation in the data distribution across fracture classes and the inherent variability of fracture patterns, which could impact generalizability. Further, model optimization and improved techniques of ensemble use in medical imaging diagnostics will greatly improve the accuracy and efficiency of future studies.

REFERENCES:

[1] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Inf. Process. Manage., vol. 45, no. 4, pp. 427–437, Jul. 2009.

[2] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and E. Duchesnay, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, Nov. 2011.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.

[4] S. Beyaz, "Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches," Joint Diseases Rel. Surg., vol. 31, no. 2, pp. 175–183, Jun. 2020.

[5] S. Agarwal and S. Agarwal, "Bone age assessment from lateral cephalograms using deep learning algorithms in the Indian population," Indian J. Dental Res., vol. 33, no. 4, pp. 402–407, Oct. 2022.

[6] I. P. Senanayake, I.-Y. Yeo, and G. A. Kuczera, "A random forest-based multi-index classification (RaFMIC) approach to mapping three-decadal inundation dynamics in dryland wetlands using Google Earth engine," Remote Sens., vol. 15, no. 5, p. 1263, Feb. 2023.

[7] N. Papandrianos, E. Papageorgiou, A. Anagnostis, and K. Papageorgiou, "Efficient bone metastasis diagnosis in bone scintigraphy using a fast convolutional neural network architecture," Diagnostics, vol. 10, no. 8, p. 532, Jul. 2020.

[8] A. Ghosh, S. Bose, D. Patton, I. Kumar, V. Khalkhali, M. K. Henry, M. Ouyang, H. Huang, A. Vossough, R. W. Sze, S. Sotardi, and M. Francavilla, "Deep learning-based prediction of rib fracture presence in frontal radiographs of children under two years of age: A proof-of-concept study," Brit. J. Radiol., vol. 96, no. 1145, May 2023, Art. no. 20220778.

[9] J. Hwang, J. Seo, J. Kim, S. Park, Y. Kim, and K. Kim, "Comparison between deep learning and conventional machine learning in classifying iliofemoral deep venous thrombosis upon CT venography," Diagnostics, vol. 12, no. 2, p. 274, Jan. 2022.