INTERNSHIP REPORT

July 2024 to November 2024

A Comparative Study of Deep Learning Models for Bone Fracture Detection in X-Rays: Proposing an Enhanced Model through Improved Architecture

Submitted by

Om Vinayak Kadam
22111217

Under the Guidance of

Prof. Yashwant Ingle

Faculty Mentor and Industry Mentor

Department of Artificial Intelligence and Data Science Engineering
Vishwakarma Institute of Information Technology, Pune

Academic Year 2024-25

DECLARATION

I hereby declare that the project work entitled A Comparative Study of Deep Learning Models for Bone Fracture Detection in X-Rays: Proposing an Enhanced Model through Improved Architecture is an authentic record of my own work carried out at Vishwakarma Institute Of Information Technology as requirements of semester long internship for the award of degree of B.Tech. Department of Artificial Intelligence and Data Science Engineering, Vishwakarma Institute of Information Technology, Pune under the guidance of Prof. Yashwant Ingle, during July 2024 to November 2024 .

Om Vinayak Kadam
22111217
Date: ____________________

Certified that the above statement made by the student is correct to the best of our knowledge and belief.

Prof. Yashwant Ingle

Faculty Mentor and Industry Mentor

I. Abstract

Accurate detection of bone fractures in X-ray images is a vital component of medical diagnostics, where swift and precise identification is essential for effective treatment planning. Recent developments in deep learning technology have shown significant promise in automating this critical task, with Convolutional Neural Networks (CNNs) leading the way. This paper presents a thorough comparative analysis of several deep learning models, including VGG16, ResNet50, InceptionV3, MobileNet, DenseNet, and Perceiver, specifically for bone fracture detection. We meticulously evaluate the performance of these CNN architectures and explore various preprocessing techniques to understand their influence on model performance. Our study reveals the strengths and limitations of each model, providing a detailed assessment of how different approaches impact detection accuracy. Building on these findings, we propose an advanced model that integrates a finely tuned architecture with sophisticated preprocessing methods. This innovative model demonstrates superior accuracy and reliability in detecting bone fractures compared to existing models. The enhanced approach offers a more robust tool for automated analysis of medical images, highlighting the potential for improving diagnostic processes. This research not only sheds light on the development of deep learning models for medical diagnostics but also paves the way for creating more effective and efficient fracture detection systems.

Table Of Contents

II. Introduction

1. Introduction:

The human body comprises 206 bones, each differing in size, shape, and strength. Among these, the auditory ossicles in the ear are the smallest, while the femur is the largest and most robust. Bone fractures are a widespread issue, particularly affecting the lower leg, and they present significant health challenges worldwide, affecting both developing and developed nations. Timely and precise fracture detection is essential for effective treatment and recovery.

Deep learning, a specialized branch of artificial intelligence, has seen increasing application in medical imaging for recognizing patterns and assisting healthcare professionals in diagnosing conditions and selecting appropriate treatments. X-ray imaging, a cornerstone in fracture detection, is valued for its accessibility, cost-effectiveness, and efficiency. Since its discovery by Wilhelm Roentgen in 1895, X-ray technology has undergone substantial advancements, including digital X-ray systems, portable machines, and sophisticated computerized image processing, all of which are integral to contemporary medical diagnostics. Despite its widespread use, manual analysis of X-ray images by radiologists can be susceptible to errors, especially when it comes to identifying subtle fractures. This limitation has spurred the adoption of deep learning methodologies, which are increasingly acknowledged for their prowess in analyzing medical data.

In this research, we focus on the development and assessment of deep learning models for bone fracture detection using a diverse dataset from Kaggle, which includes a variety of X-ray images. The models evaluated in this study encompass GoogleNet, VGG16, Xception, ResNet, DenseNet, and Perceiver. Our comparative analysis reveals that DenseNet performs most effectively in detecting bone fractures. Building on this finding, we propose an enhanced model architecture derived from DenseNet to further improve detection accuracy and reliability. This study aims to advance the creation of more accurate and automated diagnostic tools for clinical applications.

2. Motivation:

The motivation for this study arises from the need for precise and rapid detection of bone fractures, a crucial aspect of medical diagnostics. Conventional methods, relying on manual interpretation by radiologists, are prone to human errors, particularly with subtle fractures. Given the advancements in medical imaging and deep learning, there is a compelling opportunity to improve fracture detection accuracy and reliability. By leveraging deep learning models, particularly convolutional neural networks (CNNs), this research aims to automate and enhance diagnostic processes, ultimately reducing the burden on healthcare professionals and improving patient outcomes.

III. Literature Survey

Author	Contributions	Differences
Yu Cao et al. (2015)	Proposed a method using stacked random forests and feature fusion (Schmid texture, Gabor texture, Contextual-Intensity) to detect bone fractures with 81.2% accuracy but low precision of 24.7%.	Used stacked random forests, which differs from CNN-based models explored in later studies.
G. Kitamura et al. (2019)	Proposed a method using stacked random forests and feature fusion (Schmid texture, Gabor texture, Contextual-Intensity) to detect bone fractures with 81.2% accuracy but low precision of 24.7%.	Utilized multi-view ensemble learning; others focused on single view or large datasets for model evaluation.
P. H. S. Kalmet et al. (2020)	Focused on deep learning algorithms (VGG16, Inception V3, ResNet) for fracture diagnosis, demonstrating high performance in emergency situations.	Emphasized emergency applications, whereas later works concentrated on accuracy and dataset size.
Justin Krogue et al. (2020)	Focused on deep learning algorithms (VGG16, Inception V3, ResNet) for fracture diagnosis, demonstrating high performance in emergency situations.	Focused on DenseNet architecture and sub-classification, while others mainly tackled detection accuracy.
W. Abbas et al. (2020)	Proposed an automatic system using Faster-RCNN and VGG-16 for detecting and classifying lower leg bone fractures, achieving promising results.	Combined Faster-RCNN and VGG-16, unlike other studies using only CNN architectures.
D. P. Yadav et al. (2020)	Developed a CNN model with convolution, pooling, and dense layers, achieving 92.44% accuracy in fracture classification, outperforming earlier models.	Focused on basic CNN architectures; others employed more complex or hybrid models like DenseNet or transfer learning.
L. Tanzi et al. (2020)	Used deep learning and transfer learning to classify wrist fractures with a high accuracy of 94.3%, demonstrating the effectiveness of transfer learning on cropped X-ray images.	Focused on transfer learning for wrist fractures, while other studies analyzed different bones using CNN.
N. E. Regnard et al. (2022)	Compared AI with radiologists for detecting fractures, dislocations, and lesions, showing AI's superiority in sensitivity and overall detection.	Unique in comparing AI performance with radiologists, which is absent in most other technical papers.
I. Khatik et al. (2022)	Conducted a systematic review of CNN-based fracture detection models, proposing hybrid pipelines combining CNN techniques for better accuracy.	Emphasized the need for hybrid pipelines, unlike most studies that focus on individual model performance.

A. M. Barhoom et al. (2022)	Applied VGG16 to detect and classify 14 bone abnormalities, achieving precision of 85.96%, recall of 85.82%, and F1-Score of 85.77%.	Focused exclusively on VGG16 for bone abnormalities, unlike others that compare multiple models.
K. Thaiyalnayaki et al. (2023)	Developed a CNN-based system with 99.5% accuracy for bone fracture detection, incorporating image processing, feature extraction, and statistical analysis.	Achieved high classification accuracy with CNN, surpassing the results of earlier studies.
N. Vasker et al. (2023)	Utilized deep learning with data augmentation for real-time femur fracture detection, achieving 92.44% accuracy.	Focused on real-time applications with data augmentation, whereas others used standard datasets for static evaluation.

Recent advancements in deep learning have led to significant progress in bone fracture detection using various models and techniques. One study proposed a method using stacked random forests combined with feature fusion techniques, achieving an accuracy of 81.2% in fracture detection, although with a relatively low precision of 24.7% [1]. Another investigation focused on using CNN ensembles with multiple views for ankle fracture detection, reporting a notable improvement in accuracy from 76% with a single view to 81% with three views [2]. Deep learning algorithms such as VGG16, Inception V3, and ResNet have been explored for bone fracture diagnosis, showing high potential in emergency scenarios due to their strong performance [3]. A model developed for hip fracture detection using DenseNet achieved an impressive accuracy of 93.7%, along with high sensitivity and specificity, making it a robust tool for clinical use [4]. Additionally, an automatic system using Faster-RCNN and VGG-16 was proposed for detecting and classifying lower leg bone fractures, yielding promising results in both tasks [5]. A CNN model designed with convolution, pooling, and dense layers achieved 92.44% accuracy in bone fracture classification, outperforming previous models and highlighting the importance of effective model architecture [6]. Transfer learning techniques were also utilized to classify wrist fractures, achieving a high accuracy of 94.3%, demonstrating the effectiveness of this approach in medical imaging [7]. Comparative studies have shown that AI can outperform radiologists in sensitivity and overall detection accuracy for fractures and other conditions, underscoring its potential in clinical practice [8]. Several reviews have pointed out the limitations of existing CNN-based models, proposing hybrid pipelines that combine different techniques for improved accuracy and universality [9]. CNNs have also been successfully applied to detect and classify various bone abnormalities, achieving high precision, recall, and F1-Scores [10]. One study reported a remarkable classification accuracy of 99.5% using a CNN-based system that integrated image processing and feature extraction [11]. Finally, data augmentation techniques combined with deep learning achieved a 92.44% accuracy in fracture detection, demonstrating their importance in improving model performance [12].

IV. Details of the Study

1. Objectives:

Comparing Deep Learning Models: The main goal is to compare different deep learning models, such as VGG16, ResNet50, InceptionV3, MobileNet, DenseNet, and Perceiver, to see which one works best for detecting bone fractures in X-ray images.

Improving the Model: Based on the results of the comparison, we aim to create a better model by making improvements to its design, so it becomes more accurate and reliable for medical use.

Effects of Preprocessing: Another goal is to study how different image-preparation techniques affect the performance of these models and to find the best ways to prepare X-ray images for fracture detection.

Advancing Medical Tools: Lastly, this research aims to help develop more efficient and accurate automated tools that can be used by doctors to improve the detection of fractures and enhance patient care in hospitals.

2. Applications for future and current scope:

Automatic Medical Diagnosis

Remote Healthcare (Telemedicine): In areas with limited medical services, this technology can be used to diagnose fractures remotely, allowing doctors to help patients without needing to be on-site.

Emergency Room Support: In emergency rooms, this model can quickly check for fractures, reducing wait times and helping doctors decide on treatments faster.

Mobile Health Devices: The model can be used in portable devices or apps, allowing healthcare workers to detect fractures in the field, like in ambulances or rural areas.

Training and Learning: Medical students and new doctors can use this model to learn about different types of fractures, helping them understand how to diagnose bone injuries using X-rays.

Medical Research: Researchers can study bone fractures using large sets of X-ray images with the help of this model, leading to new discoveries about bone health and recovery

V. Methodology And Results

1. Dataset Description:

The dataset utilized for this project is sourced from Kaggle, created by Vuppala Adithya Sairam, and is titled "Bone Fracture Detection Using X-rays" [i]. This dataset comprises X-ray images of various joints in the upper extremities, including both fractured and non-fractured samples. The dataset is organized to assist in the classification task of detecting fractures in the given X-ray images. The images depict different joints such as elbows, shoulders, and hands, and isolating individual joints is recommended for enhancing the performance of the classifiers. The dataset is comprehensive and provides a valuable resource for training and evaluating deep learning models aimed at detecting bone fractures.

Bone	Fractured	Not Fractured	Total
Train	4480	4383	8863
Valiate	360	240	600

2. Deep Learning Models:

Perceiver Model: The Perceiver architecture combines features from transformers and CNNs, making it powerful and efficient for complex problems with reduced computing power. It uses a two-stage process involving attention operations to process input data into low-dimensional space and encode it into output space. The Perceiver's ability to filter relevant information and use shared weights across modalities makes it highly effective for tasks such as image classification and multimodal tasks.

Xception Model: Xception, short for "Extreme Inception," is based on the Inception model and simplifies the convolution process into pointwise and depthwise convolutions. This design reduces computational overheads while maintaining speed and performance, making it suitable for image classification and other vision-based applications.

VGG16 Model: VGG16 is a convolutional neural network (CNN) with 16 layers, known for its simple yet effective architecture. Developed by the Visual Geometry Group at Oxford, it uses 3×3 convolutional filters, which makes it highly useful in computer vision tasks that require capturing detailed features in images.

GoogleNet Model: Also known as Inception v1, GoogleNet enhances network efficiency by using inception modules that extract hierarchical features from data while regulating the depth of layers. This architecture is ideal for large-scale image recognition tasks due to its ability to handle higher parameters with fewer computations.

DenseNet121 Model: DenseNet121 features a unique architecture with dense connections, allowing each layer to reuse features from previous layers. This results in rich feature propagation and efficient gradient flow, leading to a compact and effective network that excels in handling complex tasks while maintaining manageable network size.

ResNet50 Model: ResNet, or Residual Network, is known for its use of residual connections that bypass certain layers to address the vanishing gradient problem. The architecture's use of residual blocks and identity mappings allows for the training of deeper networks, making ResNet50 highly effective in deep learning tasks with remarkable performance.

Proposed Model: The proposed model is an improved version of DenseNet121, incorporating several enhancements to boost accuracy and performance. Key improvements include the addition of self-attention layers after each dense layer, which helps the model focus on important features and capture long-range dependencies. The model uses the swish activation function instead of ReLU for better performance and includes a learning rate scheduler to gradually reduce the learning rate during training, leading to better convergence. Fine-tuning the model's parameters in the later layers further enhances its suitability for the specific task of bone fracture detection, resulting in improved performance over the base DenseNet121 model.

3. Overview of Model Workflow:

Library Imports: The model development begins with the inclusion of necessary libraries, such as TensorFlow, Keras, OpenCV, among others, which provide the foundational tools for building and processing the model.

Dataset Path Configuration: The directories for the training and testing datasets are specified. This step ensures that the model knows where to locate the data needed for training and validation.

Dataset Preparation: The datasets are then loaded using TensorFlow’s image_dataset_from_directory method. This function streamlines the process of importing and pre-processing the data for model input.

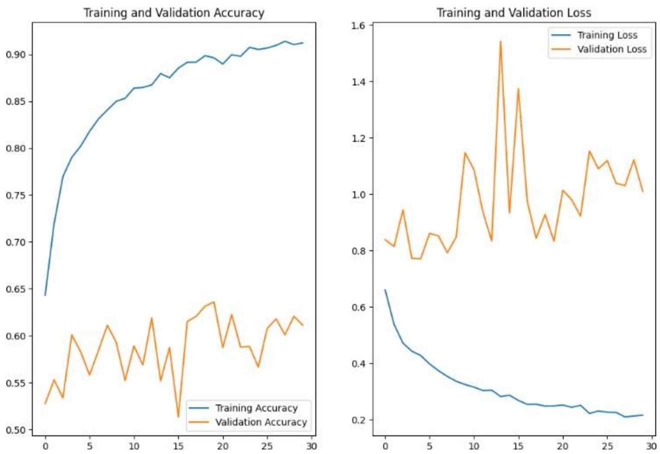
Dataset Visualization: To get an initial understanding of the data, a random sample from the training set is displayed, along with its associated label. This visualization helps in verifying the correctness of the data preprocessing.

Model Construction: The model is built sequentially, using a variety of layers tailored to the specific architecture. These layers typically include Convolutional layers for feature extraction, MaxPooling layers for dimensionality reduction, flatten layers to prepare data for Dense layers, and additional layers such as Dropout and BatchNormalization for regularization and improved training.

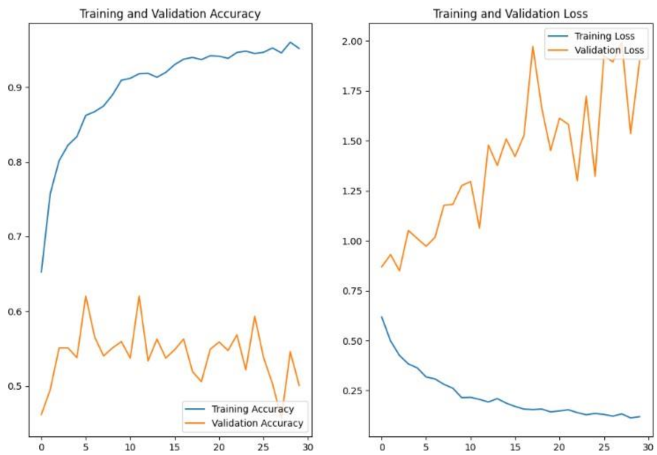
Model Training, Testing, and Evaluation: Post model construction, the model is trained on the prepared dataset. Once trained, the model's performance is evaluated through various metrics. The training process is monitored by plotting graphs such as Training Accuracy vs. Validation Accuracy and Training Loss vs. Validation Loss. Furthermore, performance metrics such as F1 Score, Accuracy, Precision, and Recall are computed to provide a detailed analysis of the model's effectiveness.

4. Results:

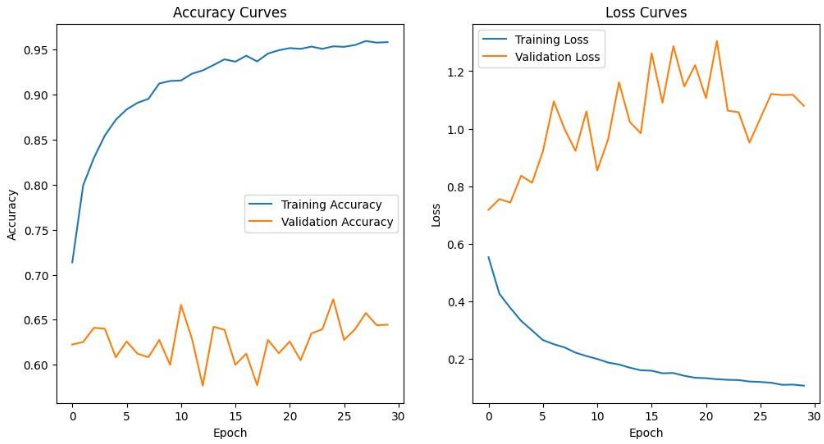
The accuracy and loss plots for each model are as follows:



Above figure represents the accuracy and loss plot of the GoogleNet model, showing how the training and validation accuracy and loss change over the epochs. The accuracy of the GoogleNet model is 91.58%.



Above figure illustrates the accuracy and loss plot of the VGG16 model, where the accuracy is 95.38%.



Above figure presents the accuracy and loss curves for the Xception model, with an accuracy of 95.93%.

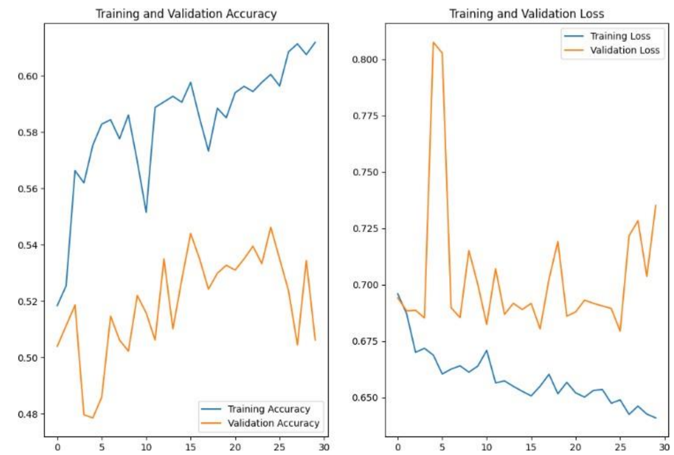


Fig. 4 shows the accuracy and loss plots for the ResNet50 model, which achieved an accuracy of 61.98%.

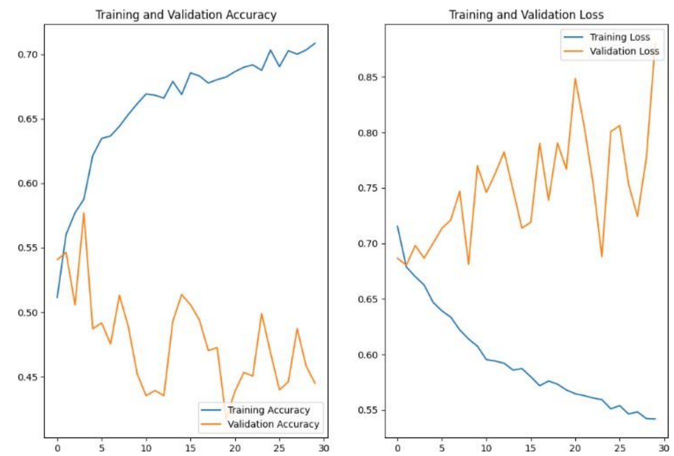


Fig. 5 depicts the accuracy and loss plots for the Perceiver model, which attained an accuracy of 71.52%.

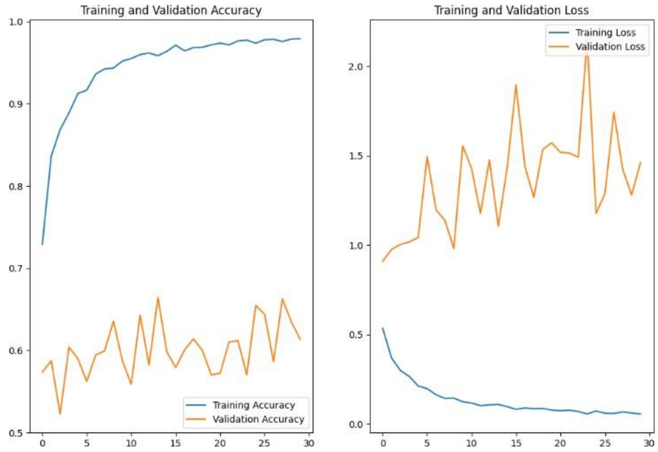


Fig. 6 captures the accuracy and loss plots for the DenseNet121 model, with a notable accuracy of 98.08%.

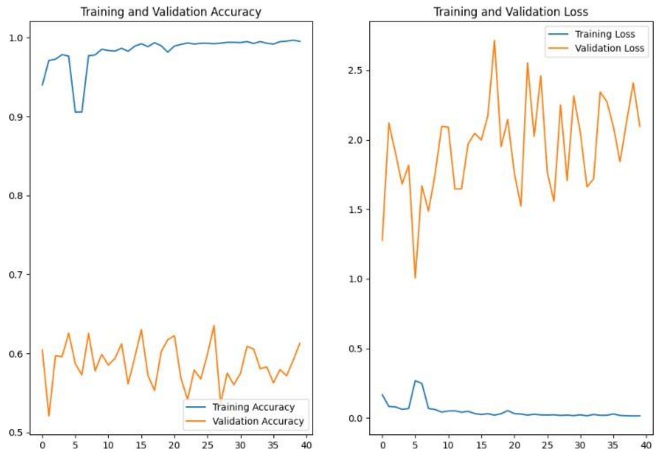


Fig. 7 showcases the accuracy and loss plots for the Proposed Model, which outperforms all other models with a remarkable accuracy of 99.58% and minimal loss.

Table II depicts a comparison between the metrics of all the deep learning models evaluated to figure out the best models suitable to detect bone fractures.

TABLE II:

Model	Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)
ResNet50	52	51.5	51.5	61.98
DenseNet121	51	51	51	98.08
GoogleNet	47	47	47	91.58
VGG16	52	52	52	95.38
Perceiver	48	48	48	71.52
Xception	51	51	51	95.93
Proposed Model	50	50	50	99.58

The Proposed Model demonstrates the highest accuracy among all evaluated models, achieving 99.58%. This model includes several enhancements, such as the incorporation of self-attention layers, the Swish activation function, a learning rate scheduler, and fine-tuning of parameters. These improvements contribute to the superior performance of the proposed model, making it the most effective in detecting bone fractures among the deep learning models considered.

VI. Conclusion

The research highlights the significant potential of deep learning models, particularly Convolutional Neural Networks (CNNs), in improving bone fracture detection using X-ray images. Manual analysis by radiologists can often result in errors, but this study demonstrates how automated detection, powered by CNN architectures like VGG16, ResNet50, DenseNet121, and others, can offer a more reliable solution. Through a comprehensive comparative analysis, it was found that DenseNet121 performed exceptionally well with 98.08% accuracy. However, the development of a proposed model with architectural enhancements, such as self-attention mechanisms and the Swish activation function, outperformed all others, achieving 99.58% accuracy. This demonstrates the ability of deep learning to improve diagnostic accuracy and speed, reducing the burden on healthcare professionals.

The contribution of this study lies in advancing automated diagnostic tools for medical imaging, offering an enhanced solution that integrates cutting-edge CNN techniques. The findings not only show the value of deep learning in medical diagnostics but also set the foundation for future advancements, including the integration of more sophisticated models like transformers or real-time clinical applications. The enhanced model's performance, combined with a detailed methodology, underscores its potential to revolutionize diagnostic workflows, providing faster and more accurate fracture detection that can greatly benefit both healthcare providers and patients.

The primary contribution is the development of a highly accurate deep learning model for bone fracture detection. It demonstrates how advanced CNN architectures combined with innovations like self-attention layers can significantly improve diagnostic capabilities in medical imaging This work contributes to medical diagnostics by offering a more reliable and automated system that could reduce the workload of radiologists and improve patient outcomes.

References

[1] Y. Cao, H. Wang, M. Moradi, P. Prasanna, and T. F. Syeda-Mahmood, “Fracture detection in x-ray images through stacked random forests fea ture fusion,” in 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 801–805, IEEE, April 2015.

[2] G. Kitamura, C. Y. Chung, and B. E. Moore, “Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, de novo training, and multiview incorporation,” Journal of Digital Imaging, vol. 32, pp. 672–677, 2019.

[3] P. H. Kalmet, S. Sanduleanu, S. Primakov, G. Wu, A. Jochems, T. Refaaee, A. Ibrahim, L. V. Hulst, P. Lambin, and M. Poeze, “Deep learning in fracture detection: a narrative review,” Acta Orthopaedica, vol. 91, no. 2, pp. 215–220, 2020.

[4] J. D. Krogue, K. V. Cheng, K. M. Hwang, P. Toogood, E. G. Meinberg, E. J. Geiger, M. Zaid, K. C. McGill, R. Patel, J. H. Sohn, and A. Wright, “Automatic hip fracture identification and functional subclassification with deep learning,” Radiology: Artificial Intelligence, vol. 2, no. 2, p. e190023, 2020.

[5] W. Abbas, S. M. Adnan, M. A. Javid, F. Majeed, T. Ahsan, and S. S. Hassan, “Lower leg bone fracture detection and classification using faster RCNN for X-rays images,” in 2020 IEEE 23rd International Multitopic Conference (INMIC), pp. 1–6, IEEE, November 2020.

[6] D. P. Yadav and S. Rathor, ”Bone fracture detection and classification using deep learning approach,” in 2020 International Conference on Power Electronics and IoT Applications in Renewable Energy and its Control (PARC), pp. 282–285, IEEE, February 2020.

[7] L. Tanzi, E. Vezzetti, R. Moreno, and S. Moos, “X-ray bone fracture classification using deep learning: a baseline for designing a reliable approach,” Applied Sciences, vol. 10, no. 4, p. 1507, 2020.

[8] N. E. Regnard, B. Lanseur, J. Ventre, A. Ducarouge, L. Clovis, L. Lassalle, E. Lacave, A. Grandjean, A. Lambert, B. Dallaudi` ere, and A. Feydy, “Assessment of performances of a deep learning algorithm for the detection of limbs and pelvic fractures, dislocations, focal bone lesions, and elbow effusions on trauma X-rays,” European Journal of Radiology, vol. 154, p. 110447, 2022.

[9] I. Khatik and S. Kadam, “A systematic review of bone fracture detec tion models using convolutional neural network approach,” Journal of Pharmaceutical Negative Results, pp. 153–158, 2022.

[10] A. M. Barhoom, M. R. J. Al-Hiealy, and S. S. Abu-Naser, “Bone abnormalities detection and classification using deep learning-VGG16 algorithm,” Journal of Theoretical and Applied Information Technology, vol. 100, no. 20, pp. 6173–6184, 2022.

[11] K. Thaiyalnayaki et al., "Title of paper if known," J. Phys.: Conf. Ser., vol. 2471, p. 012003, 2023.

[12] N. Vasker, S. N. Haider, M. Hasan, and M. S. Uddin, "Deep learning assisted fracture diagnosis: Real-time femur fracture diagnosis and categorization," in 2023 4th International Conference on Big Data Analytics and Practices (IBDAP), pp. 1–6, IEEE, August 2023.

[i] https://www.kaggle.com/datasets/vuppalaadithyasairam/bone-fracture-detection-using-xrays