Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

Bone Fracture Classification in X-ray Images: A Deep Learning Approach Leveraging Transfer Learning

MD Sabbir Ahammed
Department of Computer Science & Engineering
Varendra University, Rajshahi-6204, Bangladesh
mdsa134867@gmail.com

Asif Ahamed
Department of Computer Science & Engineering
Varendra University, Rajshahi-6204, Bangladesh
asifahamedstudent@gmail.com

MD. Humaun Huda
Department of Computer Science & Engineering
Varendra University, Rajshahi-6204, Bangladesh
223311043.cse.vu@gmail.com

Md. Musfiqur Rahman Mridha
Department of Computer Science & Engineering
Varendra University, Rajshahi-6204, Bangladesh
mdmridha100730@gmail.com

Md. Jamil Chaudhary
Department of Computer Science & Engineering
Varendra University, Rajshahi-6204, Bangladesh
jamil@vu.edu.bd

Md. Fatin Nibbrash Nakib
Department of Computer Science & Engineering
Varendra University, Rajshahi-6204, Bangladesh
fatinnibbrash@gmail.com

***Abstract*—Bone fractures are one of the most prevalent concerns in medical diagnostics but are often diagnosed on X-ray imaging for detection. However, interpreting those X-ray images can be susceptible to human error and affect treatment. In that place, artificial intelligence (AI) is coming innovatively to solve this challenge. Our study explores the approach of incorporating the latest SOTA convolutional neural networks (CNNs) for fracture detection using deep learning. We employed pre-trained models such as DenseNet169, DenseNet121, VGG16, VGG19, ResNet50 and ResNet101 trained on larger datasets to extract high-level features from X-ray images. We used classifiers such as Logistic Regression, Random Forest, XGBoost, and a custom feed-forward network (FFN) to analyze these features. Among multiple combinations tested, VGG16 combined with the custom FFN produced the best results, reporting an overall accuracy of 99.37%, Area Under the Curve (ROC-AUC) score of 99.98%, precision of 99.37%, recall of 99.37%, and F1 score of 99.37%. This strategy highlights the potential value of AI technology in improving diagnostic accuracy, providing a fast, reliable tool for medical professionals to use to improve patient care.**

***Index Terms*—Bone Fracture, Deep Learning, Transfer Learning, X-ray Image Classification**

I. Introduction

Bone fractures are a common musculoskeletal injury affecting millions of people worldwide each year. A bone fracture describes the break or discontinuity in the bone structure, typically resulting in physical trauma, stress, or pathological states such as osteoporosis. These fractures can be anything from a hairline fracture to a break with severe involvement of other structures [1] [2]. Fractures are classified into types according to pattern, cause, and location. In 2019, 178 million incident fractures were estimated to have been clinically recognized globally, indicating an increase of 33.4% compared to 1990, largely due to population growth and rising age. Of these cases, 102 million men (56.3 percent) and 76.4 million women (43.7 percent) [3] [4]. Fracture-induced chronic or acute symptoms struck 455 million people in 2019, an alarming 70% increase from 1990. Fractures of the lower leg bones (tibia and fibula) were the most common, followed by the ulna, humerus, and radius, with falls being the leading cause.

This study uses deep learning and convolutional networks such as (DenseNet169, DenseNet121 [6], VGG16, VGG19 [7], ResNet50 and ResNet101 [8]) for fracture detection in X-rays, highlighting each model's strengths. It shows that combining CNNs with classifiers improves diagnostic precision, demonstrating the potential of AI to enhance medical expertise and outcomes [5].

Magnetic resonance imaging, CT, or X-rays are the main diagnostic tools for identifying fractures, whereas X-Rays are the most commonly used because they are fast accessible. X-ray has limitations in identifying small or complex fractures, and additional imaging may be necessary for acomplete pic ture.The analysis of X-ray images by humans is susceptible to errors, especially in cases of compound or minor fractures. Such manual means are limited and automated methods are better suited to enhance diagnostic precision. Advancements in AI and deep learning have shown great promise across many domains, especially in medical imaging for fracture detection. Deep neural networks excel at extracting intricate patterns from visual data, and transfer learning per mits customizing pre-trained models for specific applications. This investigation leverages state-of-the-art convolutional networks like DenseNet169, DenseNet121, VGG16, VGG19, ResNet50 and ResNet101 to glean representations from X-rays, demonstrating each architecture's strengths: VGG is hierarchical, ResNet embraces depth elegantly, DenseNet is compact and efficient. A variety of classifiers then interpret the features, proving simple models can capably classify fractures when coupled with CNNs' feature extraction. The study underlines how joining deep learning and classification amplifies diagnostic precision, favoring doctors and patients. The findings highlight data science's potential to augment medical expertise and outcomes.

II. Literature Review

Several promising machine learning and deep learning techniques have been explored for automated identification and differentiation of bone fractures. Numerous investigations aim to enhance diagnostic precision while addressing the pitfalls of manual evaluation. A. M. A et al. [11] applied customized Xception model achieved an F1-score of 85.07% and testing accuracy of 85.13% on 42,000 X-rays, demonstrating efficient and reliable performance in diagnosing bone abnormalities. Karanam et al. [12] used advanced CNNs, achieving 94.58% accuracy with InceptionResNetV2 for classifying fractures into types, demonstrating DL's effectiveness. B. Senapati et al. [13] applied Deep learning with CNNs has improved fracture detection, achieving 98% accuracy and 96% F1-score using a custom architecture with residual structures and transfer learning for wrist fracture classification. Shyam Gupta et al. [14] used The EfficientNet-B6 model achieved 96.83% accuracy, 97.70% precision, 96.06% recall, and a 96.86% F1-score for bone fracture classification using the FracAtlas dataset. Uma Devi et al. [16] applied meta-classifier was proposed with MATLAB based meta-classifier using both decision tree & neural network classifiers in which accuracy was obtained 85.00%. Rinisha Bagaria et al. [17], the proposed CNN-based fracture detection system achieved an accuracy of approximately 90%, specificity of 89.87%, and an area under the ROC curve of 0.8088, using 20 epochs. Soumi Ghosh et al. [18], This study presents a deep learning model for bone fracture detection with 97% accuracy and an F1 score of 98%, using heatmaps and performance graphs to aid healthcare workers in India.

III. Materials And Methods

In this section, the methodology of the current study is described. A methodology diagram is provided in Fig. 1.

A. Data Collection and Analysis

"Bone Fracture Multi-Region X-ray Data", a publicly available dataset that has been used in this work for our experiments. The dataset contains a total of 10,580 radiographic (X-ray) images that are classified into two classes, namely, fractured and non-fractured. The number of examples in each class is listed in Table I, and some examples from the dataset are viewed in Fig. 2.

TABLE I: Number of Samples in Each Class

Class	Data Sample
Fractured	4,640
Non-fractured	4,908

B. Data Preprocessing

The preprocessing steps involved resizing the image, applying CLAHE for contrast enhancement, and using a Laplacian filter to sharpen the image.These steps in the preprocessing pipeline are briefly described in this subsection and depicted in Fig. 3.

The pipeline optimizes all images in the dataset, enhancing them for extracting key characteristics.

C. Dataset Split

The photos were split into an 80-20 ratio, with 7,738 images for training and 1,910 for testing, ensuring separate training and test sets to prevent data leakage.

D. Feature Extraction

In this work, we have used several SOTA CNN models as feature extractors. The CNN models, including DenseNet169, DenseNet121, VGG16, VGG19, ResNet50 and ResNet101 pretrained on the ImageNet dataset, have been incorporated to extract discriminating features from the X-ray images.

E. Classification

5 different classifiers: Logistic Regression, Linear Regression, Random Forest, XG-Boost, custom FFN with a global average pooling, 256-node dense layer with ReLU, batch normalization, 30% dropout and sigmoid output were trained. The custom FFN did the best, likely because it could detect more complex patterns.

IV. Result and Analysis

We tried DenseNet169, DenseNet121, VGG16, VGG19, ResNet50 and ResNet101 as features extractors, and found that features from VGG variants worked out best. For classification we used Logistic Regression,Logistic Regression, Random Forest, XGBoost, SVM and custom feed-forward network. We saw that the VGG16 layed ahead of the other models but the combination between VGG16 and the custom network was reserved for our task. In the custom feed-forward network, both the batch-normalization layer and the dropout layer provide regularization that contributes to preventing overfitting, resulting in higher classification accuracy. This regularization effect is evident from training vs. validation accuracy and loss depicted in Fig. 4c. While training the classifier, we have utilized adaptive learning rate by using the Adaptive Moment Estimation (ADAM) optimizer. The effect of using ADAM in training time is illustrated in Fig. 4b. Moreover, the relationship between precision and recall is illustrated in Fig. **??**. The aforementioned combination of VGG16 and a custom feed-forward network achieves classification accuracy of 99.37%, whereas 99.37% of precision, 99.37% recall, 99.37% F1-Score, and 99.98% ROC-AUC score. The ROC curve is illustrated in Fig. 4a. Furthermore, the resulting confusion matrix is illustrated in Fig. 5. Our experimental results are scribed into Table II. Comparison of our work with previous similar works is depicted in Table III.

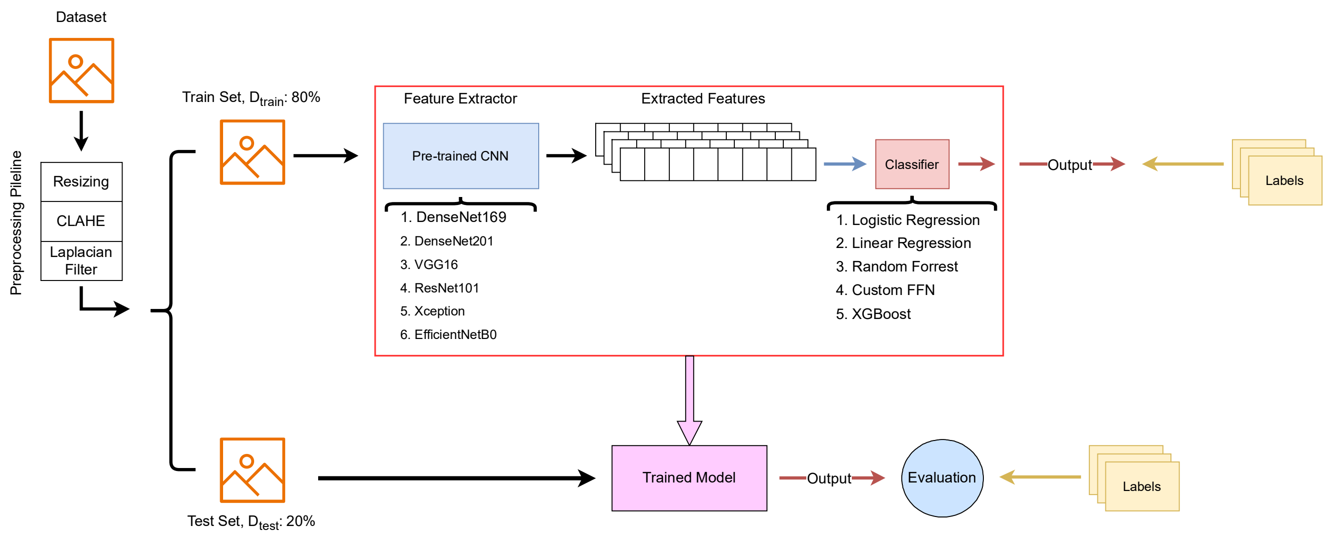

Fig. 1: Proposed Methodology

TABLE II: Performance Comparison of Different Models Used in This Work

Feature Extractor	Classifier	Accuracy(%)	Precision (%) Non-Fractured	Precision (%) Fractured	Precision (%) Macro AVG	Precision (%) Weighted AVG	Recall (%) Non-Fractured	Recall (%) Fractured	Recall (%) Macro AVG	Recall (%) Weighted AVG	F1-Score (%) Non-Fractured	F1-Score (%) Fractured	F1-Score (%) Macro AVG	F1-Score (%) Weighted AVG	ROC-AUC (%)
VGG16	XG-Boost	99.31	**99.45**	99.18	99.32	99.31	99.13	**99.49**	99.31	99.31	99.29	99.39	99.31	99.31	99.96
	Custom FFN	**99.37**	99.24	**99.48**	**99.36**	**99.37**	**99.46**	99.28	**99.37**	**99.37**	**99.35**	**99.38**	**99.37**	**99.37**	**99.98**
ResNet50	Logistic Regression	99.31	99.28	99.38	99.33	99.33	99.38	99.28	99.33	99.33	99.33	99.33	99.33	99.33	99.97
	XG-Boost	99.31	99.45	98.18	99.32	99.31	98.13	99.49	99.31	99.31	99.29	99.33	99.31	99.31	99.96
VGG19	Custom	99.26	99.24	99.28	99.26	99.26	99.24	99.28	99.26	99.26	99.24	99.28	99.26	99.26	99.95
ResNet101	Linear Regression	99.26	99.13	99.38	99.26	99.26	99.35	99.18	99.26	99.26	99.24	99.28	99.26	99.26	99.83
	SVM	99.26	99.13	98.38	99.26	99.26	98.35	99.18	99.26	99.26	99.24	99.28	99.26	99.26	99.78
	XG-Boost	99.16	99.24	98.08	99.16	99.16	98.03	99.28	99.15	99.16	99.13	99.18	99.16	99.16	99.98
DenseNet169	Linear Regression	99.26	99.24	99.28	99.26	99.26	99.24	99.28	99.26	99.26	99.24	99.28	99.26	99.26	99.71
DenseNet121	Custom	99.21	99.35	99.08	99.21	99.21	99.03	99.38	99.21	99.21	99.19	99.23	99.21	99.21	99.98

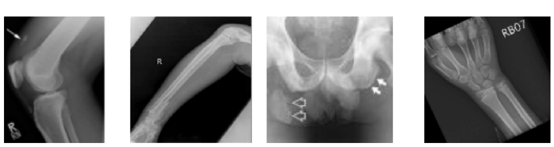

Fig. 2: Sample Images From Dataset

TABLE III: Comparison with Other Works

Model	# of Class	Acc (%)	Pre (%)	Rec (%)	F1 (%)
Xception [11]	2	85.13	85.20	85.13	85.07
InceptionResNetV2 [12]	7	94.58	94.79	98.37	94.68
CNN Model [13]	2	98.0	96.0	99.0	96.0
WFD-C Model [14]	2	96.83	97.70	96.06	96.86
MATLAB-Based Meta-Classifier [16]	2	85.0	76.9	100	70.0
CNN (Proposed) [17]	2	90.0	–	89.87	80.88
MATLAB-Based Meta-Classifier [18]	2	97.0	–	–	98.0
This work— **VGG16** with **custom FFN**	2	**99.37**	**99.37**	**99.37**	**99.37**

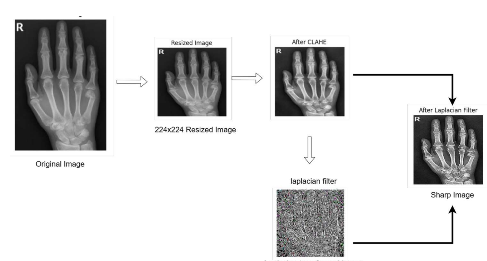

Fig. 3: Image Preprocessing Pipeline

V. LIMITATIONS AND FUTURE DIRECTIONS

While our approach surpasses previous works in classification accuracy and performance metrics, it has some limitations. A key challenge is the use of pre-trained frozen CNNs as fea-

(a) ROC-AUC Curve (b) Learning Rate Vs. Loss (c) Training Vs. Validation

Fig. 4: Performance Curves

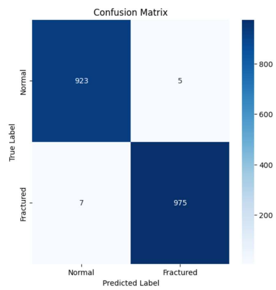

Fig. 5: Confusion Matrix

ture extractors, which limits memory optimizations and may cause issues in memory-constrained real-time applications. Additionally, the model's hyperparameters were not fine-tuned, and the feature extractors retained their pre-trained values from ImageNet instead of being adapted to our specific data. Future work could involve training custom CNN models on our data for memory efficiency and optimal hyperparameter tuning.

VI. Conclusion

In this study, we have explored fracture detection using deep learning with various feature extractors, including DenseNet169, DenseNet121, VGG16, VGG19, ResNet50 and ResNet101. All these SOTA CNN models were kept frozen with parameters learnt from ImageNet dataset. The extracted features are discriminated with much simpler classifiers, namely, SVM, Logistic Regression, Random Forest, Linear Regression, XG-Boost, and Custom feed-forward network. Our experiments demonstrate that DenseNet169, when paired with a custom feed-forward network, produces outstanding performance metrices. We have achieved an impressive accuracy of 99.37%, precision of 99.37%, recall of 99.37%, an F1-score of 99.37%, and an exceptional ROC-AUC score of 99.99%. These results highlight the potential of using transfer learning to achieve high classification accuracy in bone fracture detection.

References

[1] "Bone fractures," Cleveland Clinic. [Online]. Available: https://my.clevelandclinic.org/health/diseases/15241-bone-fractures. [Accessed: 06-Dec-2024].

[2] Openmd.com. [Online]. Available: https://openmd.com/define/bone+fracture. [Accessed: 06-Dec-2024].

[3] "Fragility fractures," Who.int. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/fragility-fractures. [Accessed: 06-Dec-2024].

[4] S. Entwisle, "Bone fractures – a systematic analysis from the Global Burden of Disease Study, 2019," APFFA, 25-Nov-2021. [Online]. Available: https://apfracturealliance.org/bone-fractures-a-systematic-analysis-from-the-global-burden-of-disease-study-2019/. [Accessed: 06-Dec-2024].

[5] J. Jung, J. Dai, B. Liu, and Q. Wu, "Artificial intelligence in fracture detection with different image modalities and data types: A systematic review and meta-analysis," PLOS Digit. Health, vol. 3, no. 1, p. e0000438, 2024.

[6] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv [cs.CV], 2014.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[9] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional Neural Networks," arXiv [cs.LG], 2019.

[10] M. E. Sahin, "Image processing and machine learning-based bone fracture detection and classification using X-ray images," *International Journal of Imaging Systems and Technology*, vol. 33, no. 3, pp. 853–865, 2023.

[11] A. M. A. and M. R. J. A.-H. Barhoom, "Deep learning-xception algorithm for upper bone abnormalities classification," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 23, Dec. 2022.

[12] S. R. Karanam, Y. Srinivas, and S. Chakravarty, "A supervised approach to musculoskeletal imaging fracture detection and classification using deep learning algorithms," Centurion University of Technology and Management, Odisha, India. [Online]. Available: http://dx.doi.org/10.24423/cames.682.

[13] B. Senapati, A. Bin Naeem, M. I. Ghafoor, V. Gulaxi, F. Almeida, M. R. Anand, S. Gollapudi, and C. Jaiswal, "Wrist crack classification using deep learning and X-ray imaging," vol. 956, pp. 60–69, 2024, doi: 10.1007/978-3-031-56950-0_6.

[14] S. Gupta and D. S., "Bone fracture classification using transfer learning," arXiv:2406.15958, 2024, doi: 10.48550/arXiv.2406.15958.

[15] K. D. Ahmed and R. Hawezi, "Detection of bone fracture based on machine learning techniques," *Measurement: Sensors*, vol. 27, p. 100723, 2023.

[16] M. Uma Devi, S. Nayak, and A. Gupta, "Bone fracture detection using image processing methods," in *Proc. Int. Conf. Recent Trends in Computing*, Singapore, 2022, pp. 493–502, Springer Nature Singapore.

[17] R. Bagaria, S. Wadhwani, and A. K. Wadhwani, "Bone fracture detection in X-ray images using convolutional neural network," 2022. [Online]. Available: https://doi.org/10.52458/978-93-91842-08-6-43.

[18] S. Ghosh, J. Mallick, M. Chakraborty, and S. Gupta, "Automated bone fracture detection in X-ray imaging to improve orthopaedic diagnostics in healthcare," *Procedia Computer Science*, 2024, doi: 10.1016/j.procs.2024.03.272.