Fracture-XNet: An Explainable Hybrid Convolutional Neural Network for Automated Bone Fracture Classification in X-ray Images Using Feature Fusion Techniques

Kazi Abdullah Jarif[a,*], Mst. Nadiya Noor[a,*] and M. F. Mridha[b]

[a]Department of Computer Science and Engineering, American International University-Bangladesh, Dhaka, 1229, Bangladesh
[b]Department of Computer Science, American International University-Bangladesh, Dhaka, 1229, Bangladesh

ARTICLE INFO

Keywords:
Bone Fracture Classification
Ensemble Learning
Feature Fusion
Medical X-ray Imaging
Explainable AI

ABSTRACT

Fractures of bones are among the most prevalent types of musculoskeletal disease. Among the most common worldwide determinants of public health morbidity are osteoporotic and other bone fractures. Manual examination of medical X-ray images based on conventional techniques is typically time-consuming and prone to human mistake, with inconsistency between radiologists. Thus, proper detection and precise classification of fractures hold utmost importance for proper diagnosis and successful treatment and ultimately for enhanced patient outcomes. To this extent, the work proposes to achieve a deep learning method through an appropriate combination of various bone fracture classification techniques. It seeks to enhance robustness and precision in the classify and localization of bone fractures. This study propose Fracture-XNet, a novel hybrid model for bone fracture classification by using a combination of two well-known pre-trained models, such as DenseNet121 and Efficient-NetB0, to enhance its performance, also through a feature fusion strategy for robust multi-class bone fracture classification. After utilizing 3 classes of bone fracture datasets, the model achieved a test accuracy of 91.34%, outperforming several established deep learning pre-trained models. The project also attempted to create an explainable AI. Additionally, a web application prototype is suggested in order to implement Fracture-XNet for accessible, real-time imaging of X-ray images. Enhanced performance will offer radiologists an effective tool towards faster and accurate diagnoses, further leading to improved patient care as well as streamlined clinical workflows.

1. Introduction

Bone fractures, the most common form of injury in the world, are induced by trauma, severe trauma, or degenerative diseases such as osteoporosis. Fractures have been increasing with the aging and expansion of the world population. This is one of the major problems worldwide. Every year, several bone fractures occur. A bone fracture refers to any interruption of the continuity of any bone within the body, either complete or partial. In advanced scenarios, a comminuted fracture is a condition in which the bone breaks up into several fragments. A compound or open fracture is a break of a bone that goes through the skin. Fractures rose exponentially in absolute incidence, prevalence, and years lived with disability from 1990 to 2019, according to the Global Burden of Disease Study. Absolute incidence rates by age category were highest in the oldest age categories, with a high proportion of these fractures being fragility fractures. Overall, the rises in the totality have been found to put upward pressure on healthcare spending. According to a statistic cited by the World Health Organization (WHO), in 2019, an estimated 178 million new fractures occurred worldwide, 33.4% more than in 1990. Further statistics cite 455 million prevalent fractures occurred in 2019, 70.1% more than during the period. Again in 2019, fractures made individuals live 25.8 million years with some disability. This may be in the form of mobility limitations, pain, or poor quality of life from fractures. The prevalence was 65.3% since 1990, implying that the years of life lived with disability (YLDs) from fractures have grown notably in the last several decades. A systematic review involving 113 studies approximated the cost for a hip fracture in the hospital to be roughly $10,075. Overall expenditure of social and health care for a single hip fracture at 12 months globally was about $43,669. In Sweden and in the five largest European Union nations, fragility fracture costs are projected to increase by 27% by the year 2030. Some trends have equally been recorded within the remainder of the globe. Fractures are more frequent among elderly persons, especially in women, given conditions like osteoporosis and vulnerability to falling. Falling from standing height accounts for most of the fractures within elderly people. Prevention of fragility fractures through management of osteoporosis and evaluation of risk factors in a timely fashion is, thus, important to all individuals' health and well-being, and especially to older people. Accurate detection and proper classification of the fracture bone can also reduce patient pain and speed up the healing process. Effective classification and analysis of medical imaging, particularly X-rays, allow one to understand the type of fracture. Detecting and accurately classify bone fractures in medical images is important in order to produce timely diagnosis for further treatment and/or an efficient treatment cycle, and ultimately, more positive patient outcomes. Automated bone

*Corresponding author
22-46386-1@student.aiub.edu (K.A. Jarif); 22-46454-1@student.aiub.edu (Mst.N. Noor); firoz.mridha@aiub.edu (M.F. Mridha)
ORCID(s): 0009-0001-3932-3013 (K.A. Jarif); 0000-0001-5738-1631 (M.F. Mridha)

fracture classification constitutes an important step for a more streamlined diagnosis in addition to correct fracture identification and classification when using medical X-Ray imaging. The aim of fracture classification is to study and identify fractured areas on either X-ray, CT, or MRI images, and to classify automatic and separate out the fractured area from the healthy bone area in order to define and categorize fractures in terms of anatomical location and type of fracture. This allows radiologists in the medical industry to determine if the fracture is simple, compound, or complex. Proper understanding is highly essential to provide the right treatment. This not only accelerates the recovery but also supports the overall care of patients. Thus, there is a colossal variety of bone fracture classification work available for identifying the fracture properly and providing preventative intervention in such a way to achieve improved results. Medical image classification, such as that of bone fractures, has significantly been enhanced over the past several years due to deep learning models, that are, Convolutional Neural Networks (CNNs) and U-Net, DeepLabV3+, and Res-Net architecture. The models are applicable to train specifically to detect and classify fracture in an image. Bone fractures, depending on the type of fracture can be complex, simple and have different appearances. To improve the performance of the entire model, the study will aim to create a more robust bone fracture classification model based on deep learning by combining certain pre-trained models like VGG16, ResNet50, Xception, DenseNet121, MobileNetV2, InceptionV3, EfficientNetB0, NASNetMobile, AlexNet, etc. In addition to helping handle noisy data and minimize overfitting, a combination of multiple weak or strong models performs better than any one model alone. Individual training of each model is done with the use of transfer learning to fine-tune a pretrained network. Approaches like voting, stacking, and averaging are used to combine the predictions of these models to improve the classification accuracy. Large Language Models (LLMs) are used to build an explainable AI for this purpose. The goal is to enhance robustness and accuracy in bone fracture detection from medical X-ray images. Traditional methods are unable to classify fractured areas in medical images precisely because they require precise pixel-level classification. Large data like Mendeley, Kaggle, and some medical X-ray images will be utilized to train models, and these can be fine-tuned on specialized, smaller fracture datasets to detect fractures in X-ray scans. The objective of this project is to build an enhanced bone fracture classification model using the top-performing pretrained models like DenseNet121 and EfficientNetB0, and explainable AI to efficiently and automatically detect bone fractures from medical imaging. This method is designed to improve bone fracture diagnosis accuracy, efficiency, and reliability using advanced machine learning techniques. Further, the integration of Explainable AI (XAI) based on Large Language Models (LLMs) adds an element of transparency to the process of decision-making such that the model's predictions are easier for healthcare professionals to comprehend. This strengthens trust and confidence in the model's prediction because it enables clinicians to understand the reasons why fractures are classified.

The rest of this study is as follows: section 2 literature reviews. The proposed method is discussed in section 3. Section 4 presents the experimental dataset, results and analysis. Lastly, section 5 discusses the impacts, limitations, and future directions of this study, and section **??** concludes the study.

2. Literature Reviews

Bone fracture is the most common illness among all age groups of individuals in the present world. Bone fracture is an illness in which the bone loses its structural stability, typically because of an external situation or force such as osteoporosis, leading to abnormal bone functions and acute pain and discomfort to the patient. Overloading of the bone or accidents is one of the frequent causes of bone fractures. Prompt treatment and diagnosis of fractures are critical, usually involving X-ray imaging to capture the image of the affected area and aid in the process of treatment [1][2]. Over the last few years, computer processing images has been the most efficient technique for the analysis of medical images. This technology is being adopted more and more in the healthcare industry, where it greatly assists medical professionals by providing computerized image analysis, which leads to faster and more accurate diagnoses and decision-making [3]. One of the key uses of digital image processing is in bone fracture classification and segmentation, which is a method that assists in identifying fracture and non-fracture tissues in medical images. X-ray imaging is commonly used to assess the anatomy of bone, and especially of the upper limbs, and is one of the key devices to treat and diagnose fractures of bone [4]. The older system of classification (i.e., longitudinal, transverse, or mixed) was not well comparable with clinical presentation in the pattern of facial nerve weakness and cerebrospinal fluid leakage. It was also not very useful in the prediction of conductive hearing loss and sensorineural hearing [5]. Deep learning, a subset of machine learning, is highly effective in analyzing medical images by learning patterns from large datasets. For instance, while doctors traditionally diagnose bone fractures by visually inspecting X-ray images, deep-learning models can be trained on vast collections of bone scans to accurately perform these diagnostic tasks automatically [6] [7] [8]. Deep learning-based technologies have become powerful devices to diagnose diseases and provide timely treatment [9]. Deep learning, as well as other forms of machine learning, has never been more popular than it is now in recent times. Deep Neural Networks (DNNs) are very well known for their power in classifying images and solving tough issues [10]. One of the studies had several papers reviewed that used different deep learning methods to classify fractures in bones with the intent of bringing out the strength of each approach and obtaining a generalized method of detecting fractures [11]. Another study pointed out that deep learning methods enable old medical professionals to pass their expertise to ground general practitioners, improving patient care to a large extent [12].A recent study 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 which provides a effective result in the field of bone fracture classification [13]. Another study use the various deep learning model like VGG16, ResNet152V2, and DenseNet201, for the detection and diagnosis of bone fractures. They also adupt 97% accuracy during the validation phase[14]. Another study proposed an ensemble model for fracture detection in x-ray images combined with multiple diverse models called leverages MobileNetV2, Vgg16, InceptionV3, and ResNet50, using histogram equalization for preprocessing and a Global Average Pooling layer for feature extraction to improve predictive accuracy and robustness of them. The proposed deep learning model achieve more than 90% accuracy in all performance measurement scales[15]. U-Net consists of an encoder and a decoder. There are many variants of U-Net. One study compared U-Net derivatives such as recurrent residual U-Net, attention U-Net, and attention recurrent residual U-Net to seek differences in segmentation performance. The accuracy of the U-Net models tested ranged from 99.07% to 99.12%, while Dice coefficient values ranged from 88.55% to 89.41%. These values indicated all four U-Net models tested to be effective in segmenting bones in X-ray images to quantify TAOD metrics [16]. Fracture classifier results are compared based on accuracy, training time, and testing time, and linear discriminant analysis (LDA) shows the highest accuracy rate with 88.67% and 0.89 AUC. The new computer-aided diagnosis system (CAD) will reduce the workload of doctors by identifying fractures with high precision [17]. One recent study examined the impact of using different backbone architectures for X-ray image fracture segmentation. The study sought to enhance the well-known DeepLabV3 model by incorporating pre-trained models such as ResNet50, ResNet101, and MobileNetV3 in the encoder to improve feature learning and segmentation performance [18]. An ensemble model combines the outputs of different diverse models to increase overall accuracy and reliability. A model built in one research efficiently automates bone fracture identification in X-ray images of the humerus with superior performance compared to modified deep learning models [19]. Another research aimed at eliminating the limitations of conventional methods and enhancing the effectiveness of fracture classification. The authors proposed a specialized Convolutional Neural Network (CNN) for musculoskeletal radiographic images. To further enhance classification accuracy and reliability, they combined adapted pre-trained models with custom layers, in addition to Ensemble Learning, which leveraged the strengths of multiple models [20]. Lastly, a study showed the application of a DeepLabV3+- based deep learning model for automatic segmentation of fractured tibia and fibula fragments in CT images. The performance evaluation demonstrated a total worldwide accuracy rate of 98.92%, indicating how well the model performs in segmentation tasks for complex fractures [21]. A study examining results 2,292 upper extremity fractures among 2,203 children and adolescents, 26% were of the humerus and 74% of the forearm. Out of the humerus, 61%, and out of the forearm, 80% of the solitary distal fractures were metaphyseal. Out of the adolescents' solitary humerus fractures, they were more epiphyseal and diaphyseal fractures, and out of adolescents' radius fractures, they were more epiphyseal fractures compared to other groups. 47% of double forearm fractures were distal metaphyseal fractures [22]. In a study evaluating the sequentially derived classification system, three sets of 30 to 185 pediatric limb fractures from 2308 fractures from two multicenter trials were independently assessed in blinded format by eight orthopedic surgeons on a total of 5 occasions. Intra- and interobserver reliability and accuracy were calculated [23].

Table 1
Summary of the accuracy result of pre-trained model for proposed Fracture-XNet architecture:

Pre-Trained Model	Test Accuracy(%)
VGG16	0.7749
ResNet50	0.8312
Xception	0.8355
DenseNet121	0.8485
MobileNetV2	08139
InceptionV3	0.8095
EfficientNetB0	0.8874
NASNetMobile	0.7749
AlexNet	0.8052
VGG19	0.6667
EfficientNetB7	0.5628
Proposed Fracture-XNet	0.9134

3. Methodology

This section highlights the detailed pipeline of our proposed hybrid model, FractureX-Net, which was designed to classify various types of bone fractures using a dual-branch CNN architecture. Through our research, we develop a deep learning model for accurately identifying bone fracture conditions by classifying fractures from X-Ray images. Firsly, our proposed model trained with several pre-trained models from Keras instead of using individual models. To propose our model firstly we ran 11 pre-trained models and found out test accuracy. The pre-trained model test accuracy is given below in Table 1, which is applied for build up our proposed model also thhe methodology overview is shown in Fig 1

3.1. Dataset Analysis and Discussion

To develop our model used larged dataset available from Mandaly[24], which is a reliable image dataset. The datasets is collected from a secondary source containing X-ray images of bone fractures. The dataset includes three different classes named Simple, Compound Fracture and Comminuted Fracture. The Simple class contains X-ray images without any fractures, while the Compound Fracture class

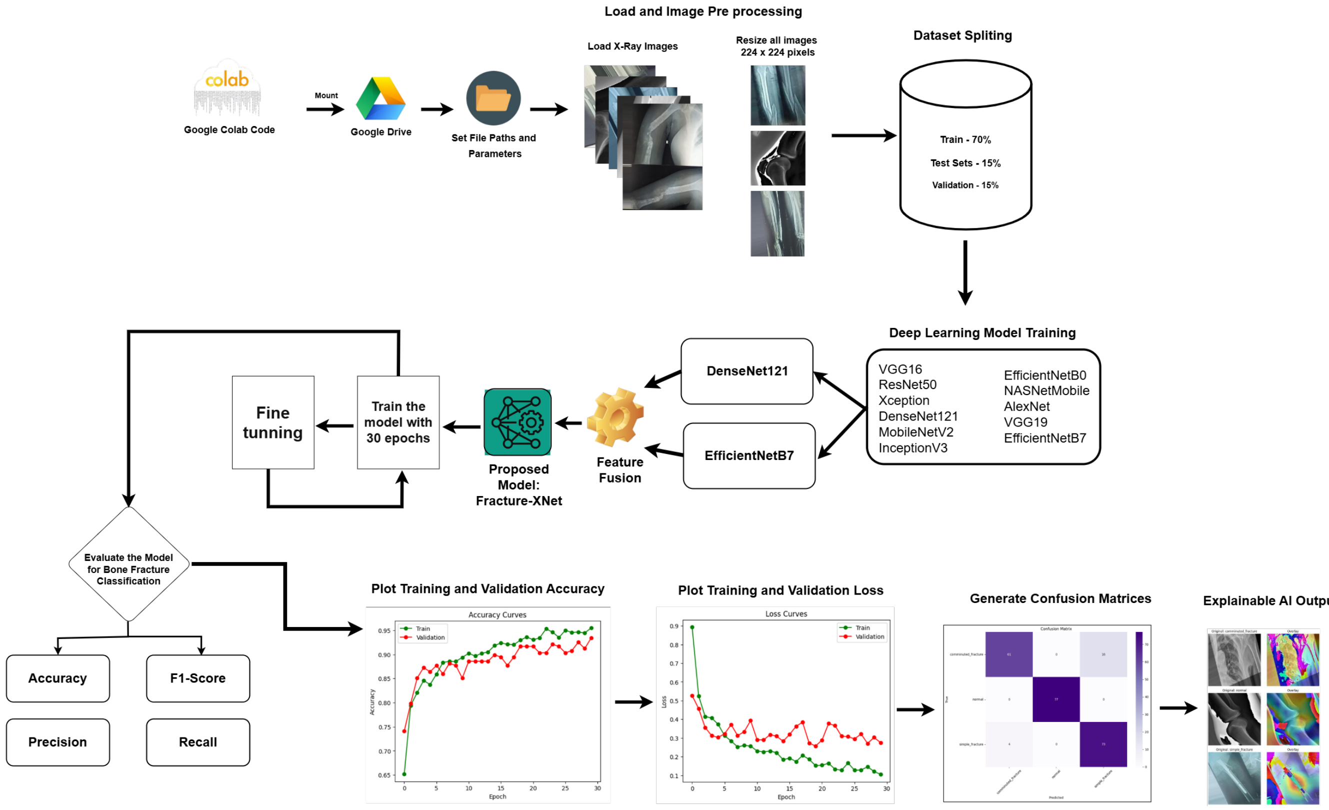

Figure 1: Proposed model Fracture-XNet model architecture overview

Table 2
Original Dataset-Class Breakdown:

Class	Images
Simple (No Fractures)	127
Compound Fractures	1174
Comminuted Fractures	1234
Total images	2535

Table 3
Detailed Augmented Dataset-Class Breakdown:

Class	Images
Simple (No Fractures)	508
Compound Fractures	508
Comminuted Fractures	508
Total images	1524

includes images showing minimal, less complex fractures. The Comminuted Fracture class consists of images depicting more severe and complex types of fractures. The Simple, Compound Fracture and Comminuted Fracture classes contain 127, 1234, and 1174 X-ray images, respectively. The detailed original distribution of the dataset is presented in Table 2.

3.2. Image Data Preprocessing

This stage includes image dataset class balancing, image augmentation, dataset splitting, and image resizing. First of all, to enhance visual diversity, we augmented Simple class images quadrupled to increase images from 127 to 508. Likely, we adjust the Compound Fracture class and the Comminuted Fracture class. From each class we take 508 x-ray images. For this classification, we utilized a total of 1524 images, using 3 classes, which is represent in Table 3 also in this Figure 2. In this study, we resize the image data into 224 x 224 Pixels. For normalization, the pixel values were scaled to a range of [0,1] and dividing them by 255.0.The normalization process is mathematically represented as:

$$I_{\text{normalized}} = \frac{I_{\text{original}}}{255.0} \tag{1}$$

Finally, we split each dataset into 3 subsets: train, validation, and test sets comprising 70%, 15%, and 15% of the original dataset for training and evaluating our proposed model. Out of 1,527 images (both accurate and augmented), 70% are for training, resulting in 1,068 images for classification purposes, 228 images for testing and for validation we utilizes total 228 images. . The detailed distributions of the datasets are presented in Table 4

3.3. Proposed Model: Fracture-XNet

To address the challenges of classify for fracture of bone, we introduce Fracture-XNet model which is hybrid deep learning model that merges the architectural strengths of two

Figure 2: Sample of Bone fracture: a) Simple (no fracture), b) Compound fracture, c) Comminuted fracture

Table 4
Dataset Breakdown:

Class	Images
Training Puporse	1068
Testing Puporse	228
Validation Puporse	228
Total images	1524

top-performing pre-trained model: DenseNet121 and EfficientNetB0, both of test accuracy are 84.85% and 88.74%. The model architecture was implemented in Python and TensorFlow. Figure 3 shows the architecture of our proposed hybrid model, Fracture-XNet. Each input image is simultaneously processed by both branches, with each branches customized for its specific base model. To show the details about proposed model table 5 summarises the experimental settings of the proposed model. Furthermore, a feature fusion strategy was incorporated during training and executed when the test accuracy did not increase for 19 consecutive epochs. In addition, a learning rate reduction strategy has also been incorporated, which reduces the current learning rate by a factor of 0.5 if no increase in the validation accuracy occurs for 19 consecutive epochs. Finally, we split each dataset into 3 subsets: train, validation, and test sets comprising 70%, 15%, and 15% of the original dataset for training and evaluating our proposed model.

3.4. Model Evaluation Metrix

To thoroughly assess the performance of the proposed hybrid CNN model, combining DenseNet121 and EfficientNetB0, a multi-faceted approach was adopted, and an extensive set of evaluation matrices was utilized. This involved analyzing training accuracy and loss curves, generating confusion matrices, and calculating precision, recall and F1-Score.

Training Accuracy and Loss:

During training, the model's learning behavior was evaluated by tracking both training and validation accuracy and loss throughout the 30 epochs. Also this model was optimized by Adam optimizer using a learning rate 0.001 to minimize the categorical cross-entropy loss. At this training process, both the training and validation accuracy and loss were tracked epochs to evaluate how well the model was training.

Table 5
Summary of the model training settings

Attribute	**Value**
Image Size	224×224
Batch Size	32
Initial Learning Rate	0.001
Epoch	30
Attention Head	2
Activation	Softmax
Optimizer	Adam
Loss	Categorical Crossentropy
Rescale Factor (DenseNet121)	1.0/255
Rescale Factor (EffiecientNetB0)	1.0/127.5 - 1

$Acc^{(t)}_{\text{train}}$, is computed as:

$$Acc^{(t)}_{\text{train}} = \frac{\text{Number of correct predictions in epoch } t}{\text{Total number of training samples}} \tag{2}$$

This cross-entropy loss function that was minimized during the training process is expressed as:

$$L^{(t)}_{\text{train}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(\hat{y}^{(t)}_{ic}) \tag{3}$$

where N is number of samples in batch, C stands for number of classes (here utilizes 3 classes), y_{ic} is the ground truth label and $\hat{y}^{(t)}_{ic}$ is the predicted probability for class c at epoch t. Equivalent expressions are applied for computing validation accuracy and loss.

Confusion Matrix

The confusion matrix represents as fundamental evaluation metric for classification models, particularly multi-class scenarios such as the three-class problem implemented

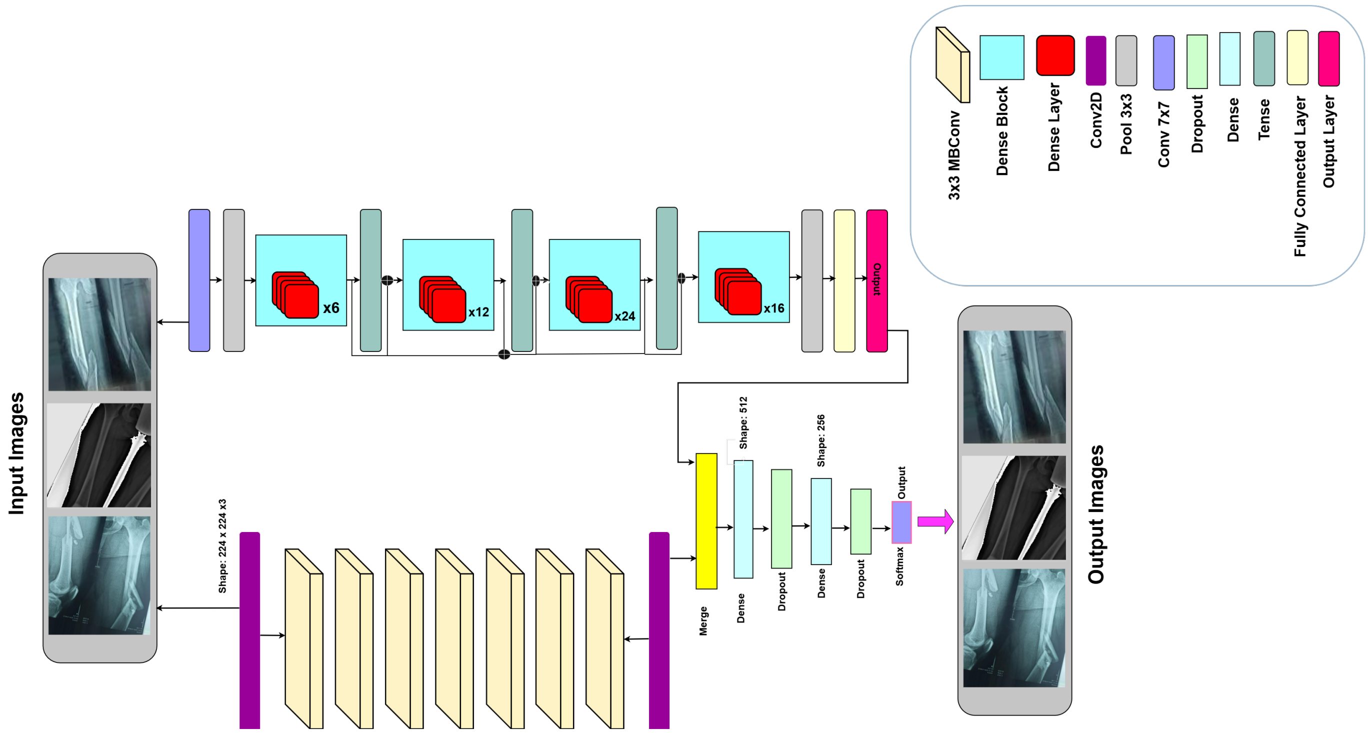

Figure 3: Fracture-XNet Model Architecture

here. It highlights mainly a detailed breakdown of how well the model performs for each individual class. Let $CM \in \mathbb{Z}^{C\times C}$ denote the confusion matrix where C is the number of classes (in this case, uses three classes).

The confusion matrix element CM_{ij} is defined as:

$$\mathrm{CM}_{ij} = \sum_{k=1}^{N} \begin{cases} 1 & \text{if } y_k = i \text{ and } \hat{y}_k = j \\ 0 & \text{otherwise} \end{cases}$$

Precision, Recall, and F1-Score

We evaluated the performance of the proposed model over the following metrics -

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

Here, TP stands for true positive, FP stands for false positive, TN stands for true negative, and FN stands for false negative. The Macro Average of the above-mentioned metrics are calculated as outlined below:

$$M_{\text{macro}} = \frac{1}{N_c} \sum_{i=1}^{N_c} M_i \tag{8}$$

Here, M stands for the respective metric (e.g. $Precision$, $Recall$, etc.), N_c is the number of classes, N_{total} is the total number of instances in the dataset and $support_i$ is the number of true instances for class i. The macro average calculates the metric independently for each class.

4. Results and Discussion

This section highlights the result of propose hybrid CNN model, Fracuture-XNet for classifying three types of bone fracutre. This evaluation matrics of proposed model Fracture-XNet include trainig curves, confusion matrics analysis, precision, recall and F1-Score. This model achieved test accuracy 91.34%, which is achieving substantially better performance than numerous widely recognized deep learning models. Table 6 highlights a comparison between Fracture-XNet and other popular CNN-based models, including VGG16, VGG19, ResNet50, Xception, DenseNet121, MobileNetV2, InceptionV3, EfficientNetB0, NASNetMobile, EfficientNetB7 and Alexnet.

4.1. Training and Validation Performance

Figure 4 illustrates the training and validation accuracy and loss curves of proposed model. This loss curve shows that model has learned effectively, with both training and validation accuracy increasing steadily and validation loss

Table 6
Performance Comparison of Different Deep Learning Models with Fracture-XNet:

Model	Accuracy %	Precision %	Recall %	F1-score
Xception	83.55	85.57	84.85	84.70
VGG16	77.49	77.93	77.49	77.61
VGG19	66.67	67.38	66.67	62.18
InceptionV3	80.95	81.11	80.95	80.89
NASNetMobile	77.49	77.55	77.49	76.95
ResNet50	83.12	83.47	83.12	83.00
Xception	83.55	85.57	84.85	84.70
DenseNet121	84.85	85.57	84.85	84.66
MobileNetV2	81.39	81.65	81.39	81.18
EfficientNetB0	88.74	88.99	88.74	88.71
EfficientNetB7	56.28	39.90	56.28	88.71
AlexNet	80.52	81.61	80.52	80.90
Proposed Fracture-XNet	**91.34**	**91.96**	**91.34**	**91.29**

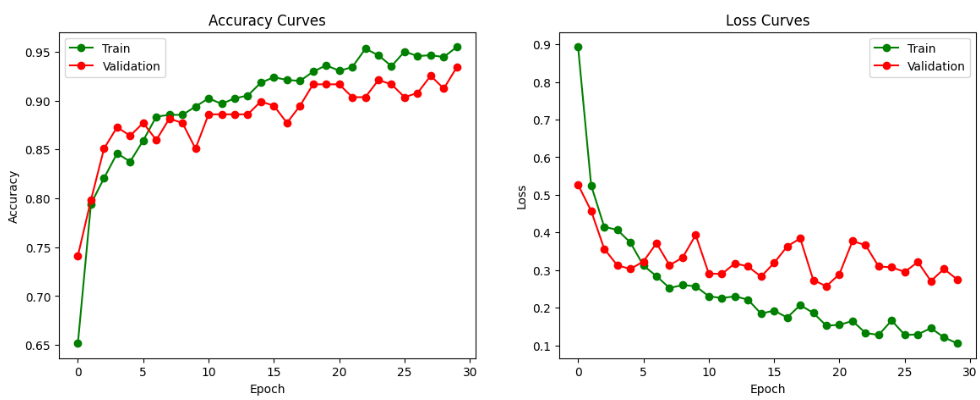

Figure 4: Training and validation accuracy and loss curves

leveling off, indicating suggesting minimal overfitting. The model achieved training accuracy is 91.55% and loss is 91.25%.

4.2. Confusion Matrix

Figure 5 highlights the confusion matrix for the proposed model, Fracture-XNet, elaborates classification performance three classes. The model demonstrated strong performance, accurately classifying 77/77 of normal class, which indicates Simple or No fracture class, 73/77 of simple fracture class, which denotes compound fracture, and 61/77 of comminuted fracture class. The majority of classification errors occurred between visually similar conditions such as, Compound fracture, Simple (No fracture) and Comminuted fracture that the model performs well overall but exhibits slight confusion among closely related classes.

4.3. Ablation Study and Performance Comparison

To gain a deeper insight into the role of each component of in our proposed model, for this performed an extensive ablation study. This study involved systematically alerting parameters, changing the number of layers also add extra dense layer and testing various optimizers to evaluate impact of model's performance. Table 7 illustrates a efficient method by methodically examining the influence of critical architecture and training parameters. The proposed model Fracture-XNet, which combines features from DenseNet121 and EfficientNetB0, consistently suppresses its variation, confirming that our design decisions, substantially improve classification performance.

Explainable AI

Explainable AI used mainly clarify the decision making process of our enhanced model and highlight the key factors that highlight a specific prediction. In this section, we introduce with the well-known XAI Framwork, to examine the predictions made by our proposed model.

This Figure 6 illustrates that it seems to compare original x-ray images with the corresponding Grad-CAM (Gradient-Weighted Class Activation Mapping) visualization. Also, Figure 6 utilizes three types of class, whose label is mainly normal class, which indicates simple fracture label and simple fracture label indicates compound fracture label. In the machine learning technique, Grad-CAM is designed to identify and classify the parts of bone fractures more accurately, which is also significant for the model's decision-making process. The key observation highlights in this Figure 6 for simple class images is that the Fracture-XNet model focuses

Figure 5: Confusion Matrix

Table 7
Ablation Study Results Across Model Variations

Experiment	Optimizer	Activation	LR	Accuracy (%)	Precision	Recall	F1 Score
Extra Dense Layer	Adam	Softmax	1e-4	89.61	90.00	90.00	90.00
Extra Dense Layer	SGD	Softmax	0.001	87.88	89.00	88.00	88.00
Extra Dense Layer	Adam	Softmax	0.001	91.34	91.96	91.34	91.29
Extra Dense Layer	Adam	ReLU	0.0001	90.91	91.00	91.00	91.00
Extra Dense Layer	Adam	Softmax	0.0001	90.48	92.00	90.00	90.00

on the integrity of bone surfaces and joints. For a compound bone fracture, the proposed model highlights the fracture line and the immediate surrounding bone. For comminuted fractures, the proposed model focuses on the fragmented bone region and the broader context of the injury.

4.4. Discussion

Overall performance, Fracture-XNet shows outstanding performance in categorizing different fracture of bone, types of simple(no fracture), compound fracture and comminuted fracture. Trained this model on balanced dataset, the model effectively prevents both overfitting and underfitting, ensuring reliable generalization across all categories. In this model, with the test accuracy of 91.34%, precision of 91.96%, recall of 91.34% and F1-Score of 91.29%, which is the closely aligned with metrics emphasize the model's consistency and stability. Achieving the highest classification accuracy among widely used 11 pre-trained deep learning models Fracture-XNet, which also delivered the most balanced overall performance. So, in this results confirm Fracture-XNet as highly efficient solution for classifying multiple categories of bone fracture.

5. Web-Based Application Prototype

A prototype of the web application was created implement by HTML, CSS and JavaScript to showcase the functionality of the proposed hybrid model, which is illustrated as in figure 7. This interface provides a means for the users to upload an image file which would be processed and classified by the trained model. Then this model returned the predicted class lavel and confidence scores. This simplicity

Figure 6: Explainable AI

prototype is a demonstration for application of the model in real-model applications.

6. Conclusion

This study proposes a hybrid deep learning model called Fracture-XNet, which is presented for the multiclass classification of bone fracture. The architecture integrates DenseNet-121 and EfficientNetB0, allowing for both feature extraction and accurately classifying the fracture category. When testing three classes of bone fracture datasets, the model demonstrates a test accuracy of 91.34%, surpassing several competitor deep learning models. By accurately identifying and highlighting bone fracture location, the integration of XAI techniques enhances precision. Furthermore, a web-based prototype was implemented, showing that the model has potential for an application in practice. This appendage further enhances the practical feasibility of the proposed system for medical image analysis and aids in supporting its application in clinical decision-making procedures. In this paper, we will improve the diagnosis of bone fractures by presenting an effective and efficient classifier that automatically classifies whether an input X-ray image depicts a bone fracture. More studies are needed to apply the model to classify other bone fractures for accurate and good working capacity, regardless of the variety of fractures. Moreover, the Fracture-XNet model's capability to correctly identify and classify the types of fracture of a bone, including with

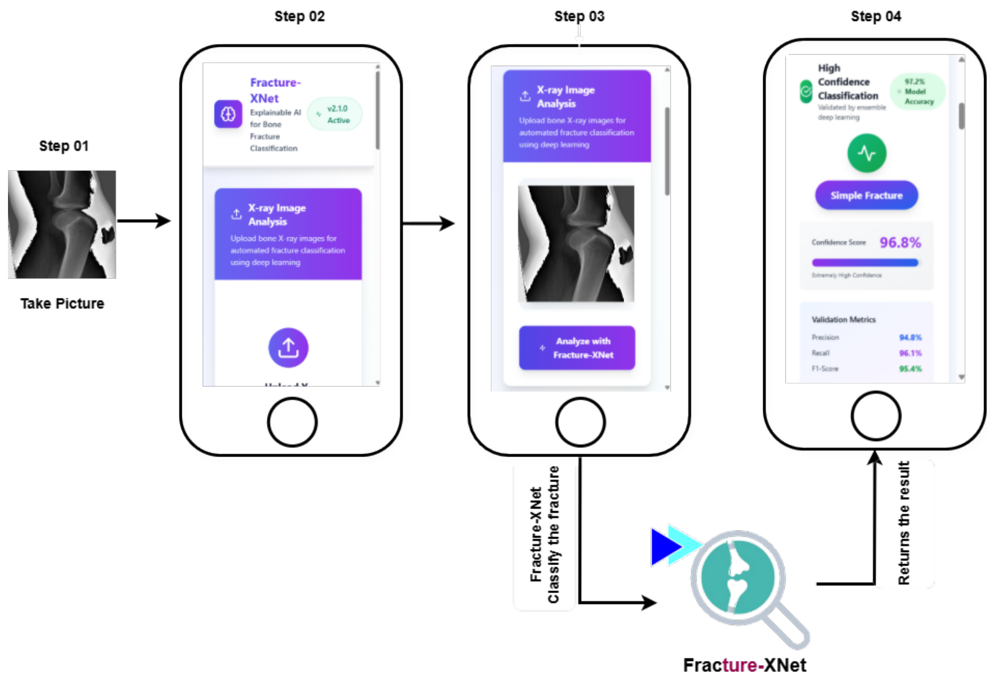

Figure 7: Prototype of proposed web application

complex cases such as comminuted fracture, compound fracture, and simple also, highlights a significant advancement in orthopedic diagnosis.

CRediT authorship contribution statement

Kazi Abdullah Jarif: Conceptualization of this study, Methodology, Investigation, Data Curation & Evaluate Model Accuracy with Fine-tuning. **Mst. Nadiya Noor:** Data Curation, Validation, Writing—Review, Method Diagram Design & Editing. **M. F. Mridha:** Supervision, Project administration.

References

[1] R. Aishwariya, M. K. Geetha, and M. Archana. Computer-aided fracture detection of x-ray images. *IOSR Journal of Computer Engineering*, 2278-2661, 2013.

[2] R. Bagaria, S. Wadhwani, and A. K. Wadhwani. A wavelet transform and neural network based segmentation & classification system for bone fracture detection. *Optik*, 236:166687, 2021.

[3] W. W. Myint, K. S. Tun, H. M. Tun, and H. Myint. Analysis on leg bone fracture detection and classification using x-ray images. *Machine Learning Research*, 3(3):49–59, 2018.

[4] K. Ghoti, U. Baid, and S. Talbar. Mura: bone fracture segmentation using a u-net deep learning in x-ray images. In *Techno-Societal 2020: Proceedings of the 3rd International Conference on Advanced Technologies for Societal Applications—Volume 1*, pages 519–531, Cham, 2021. Springer International Publishing.

[5] Stacey L. Ishman and David R. Friedland. Temporal bone fractures: traditional classification and clinical relevance. *The Laryngoscope*, 114(10):1734–1741, 2004.

[6] Luke Oakden-Rayner. Exploring large-scale public medical image datasets. *Academic Radiology*, 27(1):106–112, 2020.

[7] Andre Klein, Jan Warszawski, Jens Hillengaß, and Klaus H Maier-Hein. Automatic bone segmentation in whole-body ct images. *International Journal of Computer Assisted Radiology and Surgery*, 14:21–29, 2019.

[8] Markus T Löffler, Anjany Sekuboyina, Alexander Jacob, Adrian L Grau, Andreas Scharr, Moustafa El Husseini, et al. A vertebral segmentation dataset with fracture grading. *Radiology: Artificial Intelligence*, 2(4):e190138, 2020.

[9] W. Abbas, S. M. Adnan, M. A. Javid, W. Ahmad, and F. Ali. Analysis of tibia-fibula bone fracture using deep learning technique from x-ray images. *International Journal for Multiscale Computational Engineering*, 19(1), 2021.

[10] S. R. Karanam, Y. Srinivas, and S. Chakravarty. A systematic approach to diagnosis and categorization of bone fractures in x-ray imagery. *International Journal of Healthcare Management*, pages 1–12, 2022.

[11] 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*, 10(4):1507, 2020.

[12] R. Lindsey, A. Daluiski, S. Chopra, A. Lachapelle, M. Mozer, S. Sicular, and H. Potter. Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences*, 115(45):11591–11596, 2018.

[13] S. Torne, D. K. Shetty, K. Makkithaya, P. Hegde, M. Sudhi, P. K. Pullela, and P. PS. Vgg-16, vgg-16 with random forest, resnet50 with svm, efficientnetb0 with xgboost—enhancing bone fracture classification in x-ray using deep learning models. *IEEE Access*, 2025.

[14] Tareq Aldhyani, Zakariya A. Ahmed, Bader M. Alsharbi, Suhail Ahmad, Mohammed H. Al-Adhaileh, Ahmad H. Kamal, and Javed Nazeer. Diagnosis and detection of bone fracture in radiographic images using deep learning approaches. *Frontiers in Medicine*, 11:1506686, 2025.

[15] A. Tahir, A. Saadia, K. Khan, A. Gul, A. Qahmash, and R. N. Akram. Enhancing diagnosis: ensemble deep-learning model for fracture detection using x-ray images. *Clinical Radiology*, 79(11):e1394–e1402, 2024.

[16] Jong Hyuk Shim, Woo Sun Kim, Ki Gyoung Kim, Gi Taek Yee, Yeo Ju Kim, and Taek Sang Jeong. Evaluation of u-net models in automated cervical spine and cranial bone segmentation using x-ray images for traumatic atlanto-occipital dislocation diagnosis. *Scientific*

Reports, 12(1):21438, 2022.

[17] Mustafa Emre Sahin. Image processing and machine learning-based bone fracture detection and classification using X-ray images. *International Journal of Imaging Systems and Technology*, 33(3):853–865, 2023.

[18] Serkan Turk, Orhan Bingol, Ahmet Coskuncay, and Tuncay Aydin. The impact of implementing backbone architectures on fracture segmentation in X-ray images. *Engineering Science and Technology, an International Journal*, 59:101883, 2024.

[19] Amna Tahir, Amber Saadia, Kashif Khan, Ayesha Gul, Abdullah Qahmash, and Raja Naeem Akram. Enhancing diagnosis: ensemble deep-learning model for fracture detection using X-ray images. *Clinical Radiology*, 79(11):e1394–e1402, 2024.

[20] Sakib Shahriar Bhuiyan, Lamia Tasnim Tarin, Md. Shafiul Niaz, Md. Nasim Dolon, Afsana Afroz, and Rafeed Rahman. Fracture classification in musculoskeletal radiographs using custom CNN and ensemble learning. In *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, pages 764–769. IEEE, May 2024.

[21] Hyungseob Kim, Yong Deok Jeon, Kyung Bin Park, Hyunseok Cha, Min Su Kim, Jisun You, others, and Dae Kyung Yoon. Automatic segmentation of inconstant fractured fragments for tibia/fibula from CT images using deep learning. *Scientific Reports*, 13(1):20431, 2023.

[22] Alexander Joeris, Nicolas Lutz, Andreas Blumenthal, Theddy Slongo, and Laurent Audigé. The AO pediatric comprehensive classification of long bone fractures (PCCF) part I: location and morphology of 2,292 upper extremity fractures in children and adolescents. *Acta Orthopaedica*, 88(2):123–128, 2017.

[23] Dominik Schneidmüller, Christoph Röder, Ralf Kraus, Ingo Marzi, Michael Kaiser, Dietmar Dietrich, and Lutz von Laer. Development and validation of a paediatric long-bone fracture classification: A prospective multicentre study in 13 european paediatric trauma centres. *BMC Musculoskeletal Disorders*, 12:1–13, 2011.

[24] Fahim Faisal Talha, Md Soroar Jahan, and Mayen Uddin Mojumdar. X-ray Bone Fracture Dataset: Comprehensive Imaging for Fracture Classification and Medical Research, 2024.