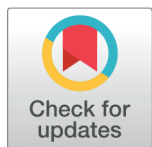


ORIGINAL ARTICLE

 OPEN ACCESS

Received: 18/06/2026

Accepted: 02/07/2026

Published: 11/07/2026

Citation: Chauhan HP, Shah V, Patel H, Mehta P (2026) Automated Musculoskeletal Fracture Analysis Using Hybrid CNN Classification and Attention-Enhanced Segmentation Models in X-ray Imaging. Indian Journal of Science and Technology 19(24): 1671-1691. <https://doi.org/10.17485/IJST/v19i24.925>

* **Corresponding author.**pranavmehta.mech@ddu.ac.in**Funding:** None**Competing Interests:** None

Copyright: © 2026 Chauhan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN

Print: 0974-6846

Electronic: 0974-5645

Automated Musculoskeletal Fracture Analysis Using Hybrid CNN Classification and Attention-Enhanced Segmentation Models in X-ray Imaging

Hemantika Prabhatsinh Chauhan¹, Vipul Shah¹, Himanshukumar Patel¹, Pranav Mehta^{2*}

¹ Department of Instrumentation & Control Engineering, Dharmsinh Desai University, Nadiad-387001, Gujarat, India

² Department of Mechanical engineering, Dharmsinh Desai University, Nadiad-387001, Gujarat, India

Abstract

Objectives: This study aims to build an automated, reliable deep learning system to detect, locate and quantify musculoskeletal fractures from X-ray images. The framework is intended to enhance the diagnostic accuracy and facilitate effective clinical decision-making. **Method:** The proposed system is a complete deep learning system to identify fractures. A VGG19-AlexNet Adaptive Hybrid CNN is used for fracture classification, a transfer-learning model is utilized for fracture detection and an Attention-ResUNet architecture is used for fracture segmentation and localization. The integrated framework allows for an overall evaluation of fractures by classification, detection, segmentation and extraction of quantitative measurements. **Findings:** The performance was tested experimentally on various benchmark datasets and is found to be excellent. The proposed Hybrid CNN was tested on the MURA (WRIST) dataset with an accuracy of 99.63% and an AUC of 99.99%. In the FracAtlas dataset, fracture detection resulted in 91.03% accuracy and 93.46% AUC. The segmentation model attained Dice score of 0.384 and a pixel accuracy of 99.02%, and it allowed for meaningful extraction of the geometric measurements related to fracture. The framework also produces structured diagnostic outputs and enables deployment via app that relates to a FastAPI back end. **Novelty:** The proposed framework combines the classification, detection, localization and quantitative assessment of fractures in a single Hybrid CNN-based architecture, in contrast to conventional fracture analysis approaches that are based on only one task.

Keywords: Medical Image; Musculoskeletal Diseases; X-ray; DL; Convolutional Neural Network; U-Net

1 Introduction

The incidence of musculoskeletal disorders (MSDs) and bone fractures is a major clinical burden worldwide, and the timely and accurate diagnosis of these conditions is vital for avoiding complications and optimizing patient outcomes. Over the past few years, deep learning (DL)⁽¹⁾ and convolutional neural networks (CNNs)⁽²⁾ have advanced greatly in the field of automated radiographic analysis, which involves learning hierarchical features without requiring a lot of manual feature engineering^{(3), (4)}. CNN-based techniques have proved good performance in fracture detection, classification and localization; however, the use of CNNs for overall diagnostic workflow is hindered by the lack of interpretation of the model, the imbalance of classes, the limited region of fracture and the limited integration with clinical workflows^{(5), (6)}.

There has been significant recent research on AI-based fracture analysis. Fouad *et al.*⁽⁷⁾ used PSO-based neural network for musculoskeletal risk prediction, and Singh *et al.*⁽⁸⁾ created a PSO-based neural network "ComDNet-512" to classify finger fractures. Anuradha and Srinivas⁽⁹⁾ proposed a fusion of classification and segmentation networks for the analysis of wrist fractures, while Aldhyani *et al.*⁽¹⁰⁾ showed that transfer learning could be useful for large-scale fracture detection. Abdellatif *et al.*⁽¹¹⁾, explored the potential for clinical validation of AI diagnosis, while Duan *et al.*⁽¹²⁾ Proposed a collaborative diagnostic approach between humans and AI to enhance diagnostic accuracy. Kumar and Anand⁽¹³⁾, studied hierarchical anatomical classification and Hareendranathan *et al.*⁽¹⁴⁾ explored other imaging modalities, and Hinterwimmer *et al.*⁽¹⁵⁾ studied scalable AI data organization. Moreover, Hardalaç *et al.*⁽¹⁶⁾, and Uysal *et al.*⁽¹⁷⁾, Wang and Zhao⁽¹⁸⁾, and Mehr⁽¹⁹⁾ demonstrated the benefits of ensemble learning, object detection, and explainable AI for fracture assessment.

However, most of the current studies are only on specific problems like abnormality classification, fracture detection, localization or fracture segmentation. Very few frameworks offer an integrated pipeline to perform screening, localization, segmentation, quantitative fracture characterization and interpretable clinical reporting all in one. Furthermore, issues like small lesion region, class imbalance, anatomical plausibility, and clinically relevant output generation are not sufficiently solved. This study introduces an end-to-end DL system that addresses these challenges in automated fracture analysis of musculoskeletal radiographs. The proposed framework provides the use of VGG19–AlexNet Adaptive Hybrid CNN for abnormality classification, fracture detection using transfer learning, segmentation using Attention-ResUNet, morphological refinement, and quantitative morphometric analysis. The framework integrates classification, localization, segmentation and fracture characterization in a single and clinically comprehensible workflow in contrast to other methods.

The key contributions of this work are as follows:

- Development of an end-to-end fracture analysis pipeline that mimics the clinical diagnostic workflow from abnormality screening to fracture characterization.
- Integration of a VGG19–AlexNet Adaptive Hybrid CNN to improve robustness and generalization in musculoskeletal abnormality classification.
- Implementation of an Attention-ResUNet framework with post-processing refinement for accurate fracture localization and segmentation.

Generation of quantitative fracture measurements and structured diagnostic outputs to enhance clinical interpretability and decision support.

1.1 Existing Work Gaps

Several promising studies have been showed in the field of classification and segmentation of musculoskeletal abnormalities or fractures, but there are still several significant drawbacks. Existing approaches tend to focus on the prediction accuracy only, and do not provide clinically interpretable quantitative measures such as fracture area, fracture length, fracture orientation, fracture shape descriptors, etc., that are essential for fracture severity assessment and treatment planning. In addition, classification and segmentation are frequently seen as single tasks and not as part of the workflow of a clinical process for screening and identification of abnormalities, fracture localization and characterization. Many also suffer from poor post-processing and have inaccurate results for segmentation output, with no integrated diagnostic reports, severity grading or quantitative explanation of the results visually. To fill these shortcomings, the proposed framework introduces a novel VGG19–AlexNet Adaptive Hybrid CNN for abnormality classification, transfer-learning based fracture detection, Attention-ResUNet for fracture segmentation, morphological refinement, quantitative fracture analysis, interpretable visualization and structured clinical reporting for comprehensive and clinically relevant fracture assessment system.

2 Methodology

The flowchart described in Figure 1 and methodology of the proposed automated MSD radiographic analysis DL framework are shown below. The MURA¹ and FracAtlas² datasets were used in the study. Before training models, preprocessing, data augmentation and class-balancing techniques were employed. For abnormality/fracture classification, a VGG19–AlexNet Adaptive Hybrid CNN was used and for fracture localization and segmentation, an Attention-Enhanced Residual U-Net (Attn-ResUNet) was used. To compare model performance, individual baseline models like VGG19, AlexNet or conventional U-Net architectures were used as internal control models. The accuracy, precision, recall, F1-score and the AUC were used to assess the classification results, and the Dice coefficient, IoU, and the pixel accuracy were used to assess the segmentation results. Paired statistical tests were used to assess the statistical significance of any performance differences at the significance level of $p < 0.05$. The entire framework was implemented using a FastAPI backend with a Flutter mobile app which facilitated automated fracture analysis and a structured clinical reporting system.

2.1 Computational Environment and Dataset Preparation

An NVIDIA GPU, an Intel processor, and 32 GB of RAM were utilized in the research, together with Python, TensorFlow/Keras, and PyTorch. For reproducibility, the fixed random seed method was used, and for preprocessing and evaluation, OpenCV, Albumentations, NumPy, Pandas and Scikit-learn were employed. The MURA dataset contained **10,411** wrist radiographs and the FracAtlas dataset contained **4,083** annotated fracture radiographs. All images and labels were verified for integrity before being separated into training, validation, and testing sets for model development and evaluation.

2.2 Image Preprocessing and Enhancement

The radiographic images are processed to a standardized preprocessing pipeline so that there would be uniform input representation on the classification and segmentation networks. To prevent loss of spatial detail, both the MURA Dataset (classification) and the FracAtlas Dataset (classification and segmentation) images are resized to 224x224 (classification models) and 256x256 (segmentation tasks) pixels. The radiographs are turned into a grayscale to eliminate redundancy and pixel intensity normalization is used to map the values to the range [0,1] as defined in Eq.1:

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}} \quad (1)$$

Subsequently, Contrast Limited Adaptive Histogram Equalization (CLAHE)^(?) is implemented to advance the visibility of micro bone structures, particularly in low-contrast regions, for increasing the image contrast. Gaussian filtering (5×5 kernel) is used to reduce noise, which means that high frequency noise is removed, but edges are not. The Gaussian kernel is well-defined in Eq.2.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}} \quad (2)$$

Where σ controls the degree of smoothing with a clip limit of 2.0 and 8×8 tile grid to improve local contrast and highlight subtle fracture lines. Further denoising is performed using **Discrete Wavelet Transform (DWT)**⁽²⁰⁾, where high-frequency components representing noise are attenuated before image reconstruction. Since most pretrained, convolutional neural networks expect three-channel inputs, grayscale radiographs are converted into pseudo-RGB format by channel replication. For transfer learning models, images are standardized using ImageNet statistics defined as Eq.3:

$$I_{std} = \frac{I_{norm} - \mu}{\sigma} \quad (3)$$

Where $\sigma = [0.229, 0.224, 0.225]$ and $\mu = [0.485, 0.456, 0.406]$. Finally, grayscale images are pretend into three channels to match pretrained CNN input requirements, resulting in tensors of shape (N, 224, 224, 3) for classification and (N, 256, 256, 3) for segmentation.

¹<https://www.kaggle.com/datasets/cjinny/mura-v11>

²<https://www.kaggle.com/datasets/mahmudulhasantasin/fracatlas-original-dataset>

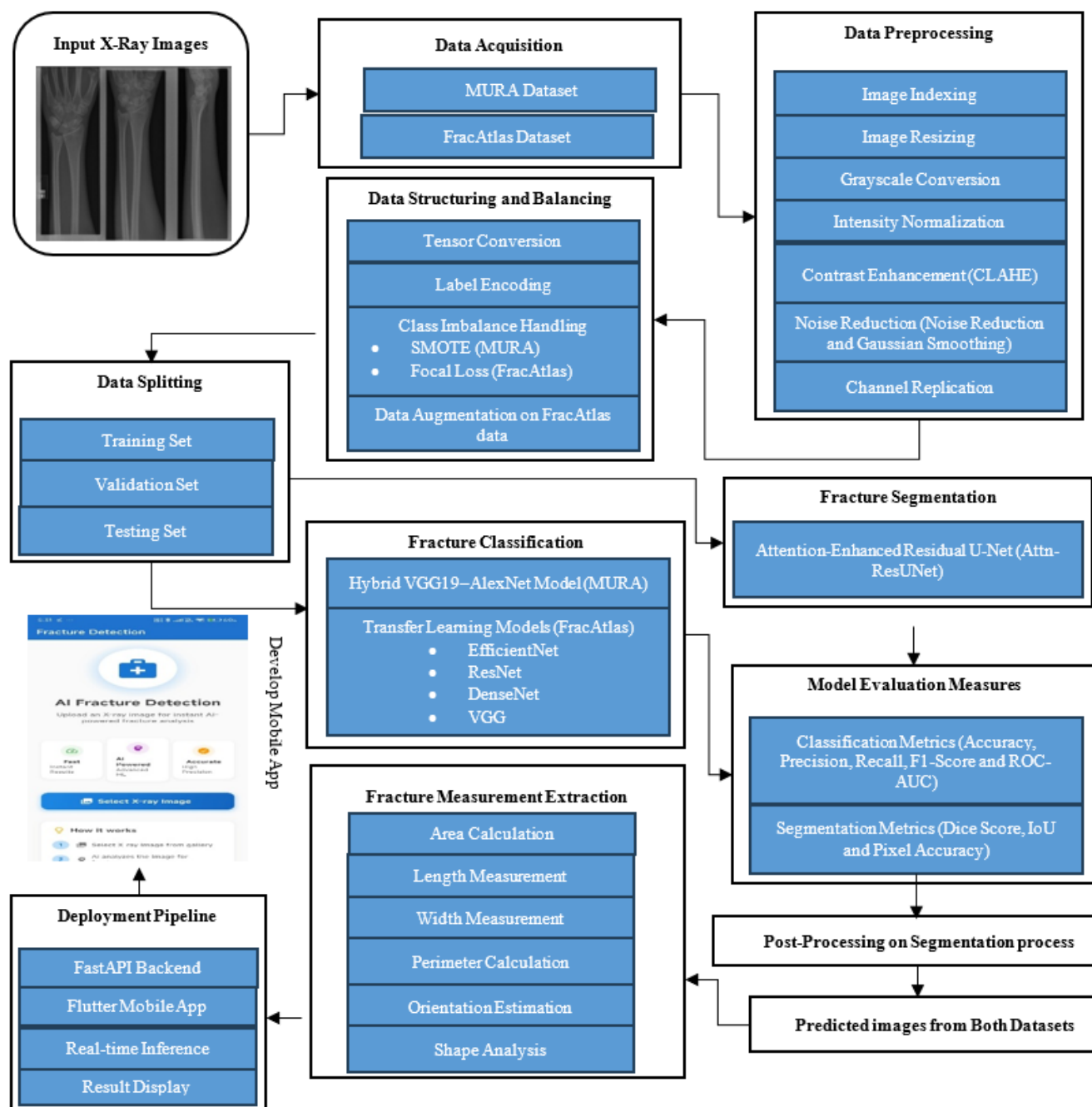


Fig 1. Proposed Flowchart for Medical Image Classification Based on MSD Using Deep Learning Models

2.3 Dataset Structuring and Class Imbalance Handling

After preprocessing, images and labels are organized into tensor structures suitable for deep learning frameworks. Labels are encoded as binary values, where 0 represents normal or non-fractured cases and 1 represents abnormal or fractured cases. The MURA dataset exhibited imbalance between normal and abnormal classes; therefore, Synthetic Minority Oversampling (SMOTE)⁽²¹⁾, is applied to create synthetic samples of the minority class. For a minority class sample x_i , one of its k nearest neighbors x_{zi} from the same class is selected. A new synthetic sample x_{new} is created using linear interpolation defined as Eq.4:

$$X_{new} = X_i + \lambda \times (X_{zi} - X_i) \quad (4)$$

Where $\lambda \in [0, 1]$ is a random number that identify the position of the synthetic point along the line segment between x_i and its neighbor. This process enhances the diversity of minority classes by generating plausible intermediate samples, reducing bias toward the majority class, and enhancing the decision boundaries of the classifier. This increased the dataset size from 10,411 to 18,177 samples, producing an approximately balanced class distribution. For the FracAtlas dataset, imbalance is addressed through Focal Loss (FL)⁽²²⁾, is rather than oversampling. To further emphasize difficult fracture examples, FL is incorporated defined as Eq.5:

$$\mathcal{L}_{FL} = -\alpha(1 - p_t)^\gamma \log(p_t) \quad (5)$$

It balances the importance of the class and misclassified samples (γ focuses learning and α balances class importance and misclassified samples). In this way, this approach enhances the accuracy of fracture detection without introducing any false data.

It should be noted that SMOTE was applied to the extracted feature representation rather than directly generating synthetic radiographic images. Consequently, the method interpolates within the learned feature space and does not create anatomically unrealistic bone structures or artificial fracture patterns in the original X-ray images. Preliminary experiments using class-weighted binary cross-entropy were also performed for comparison. Although both approaches effectively addressed class imbalance, SMOTE yielded slightly higher recall and F1-score for the minority fracture class while maintaining comparable specificity. Therefore, SMOTE was selected as the preferred balancing strategy for the MURA classification model.

2.4 Dataset Augmentation

Data augmentation using Albumentations was applied to the FracAtlas dataset to enhance the generalization and minimize overfitting. The following transformations were applied as part of the augmentation pipeline: ShiftScaleRotate ($p=0.5$; shift=0.0625, scale=0.1, rotation=15°), RandomRotate90 ($p=0.5$), Horizontal Flip ($p=0.5$), and Random Brightness Contrast ($p=0.2$). These transformations helped to model different imaging conditions, patient positioning and orientation, which made the classification and segmentation models more robust.

2.5 Dataset Splitting

To achieve an equal proportion of non-fractured and fractured patients in each subsample, the dataset is divided into subsamples using stratified random sampling. There are 80% train data and 20% test data in the MURA dataset. To put the model through its paces, it is trained using the training set and then tested using the test set. On the other hand, the FracAtlas Dataset split 25% for validation and testing, 70% for training, and 15% for the whole process. Stratification is performed so that annotated fracture masks and fractured and non-fractured samples are represented in each subset proportionally. Table 1 shows the total number of images for both datasets after splitting.

Table 1. Summary of Dataset Splitting with Total Number of Images

Dataset	Subset	Total Images	Fractured	Non-Fractured	Image Shape	Number of Batches
MURA Dataset	Training	14,541	—	—	(224, 224, 3)	—
	Testing	3,636	—	—	(224, 224, 3)	—
FracAtlas Dataset	Training	2,858	502	2,356	(256, 256, 3)*	178
	Validation	612	107	505	(256, 256, 3)*	39
	Testing	613	108	505	(256, 256, 3)*	39

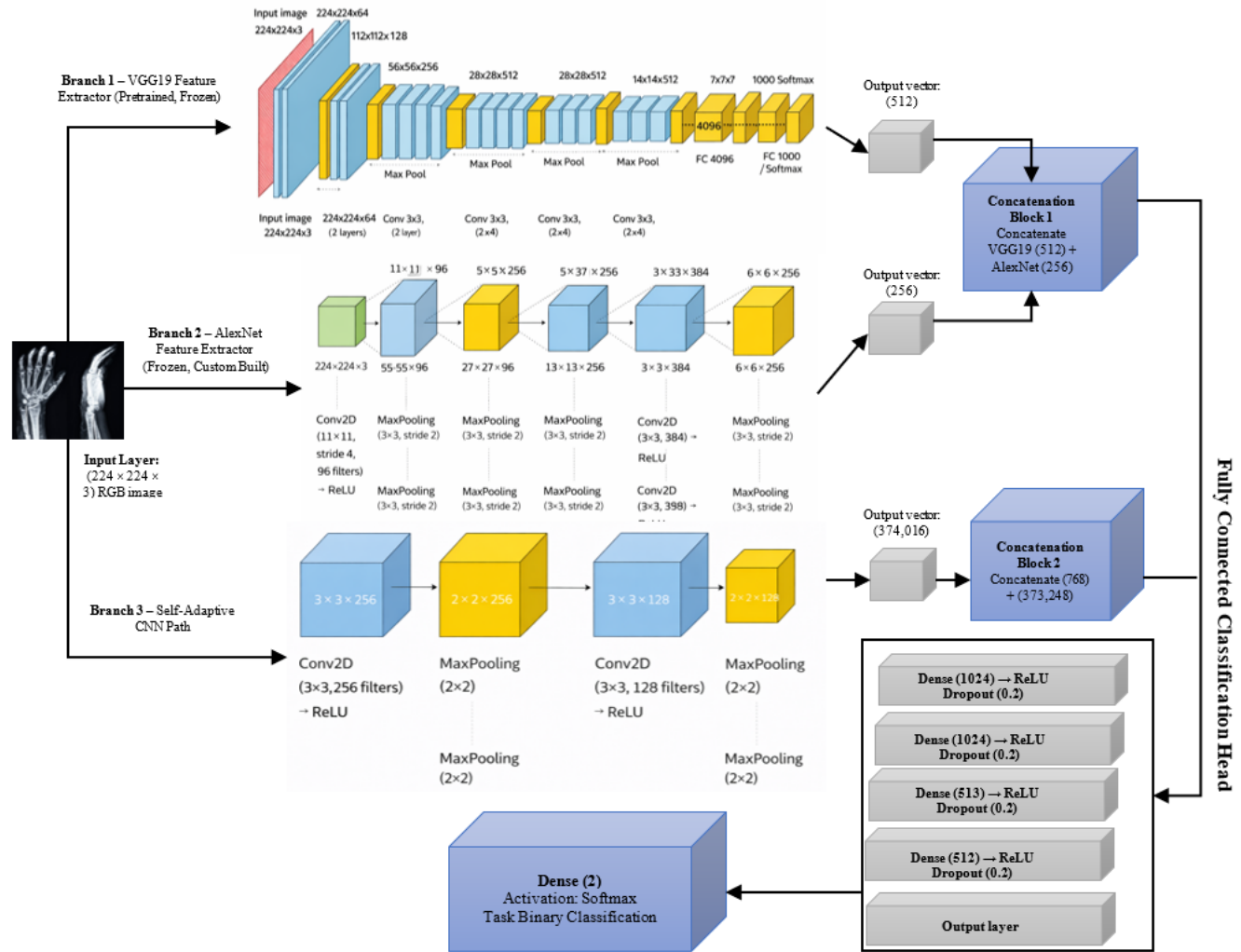


Fig 2. Trained proposed (VGG19_AlexNet_AdaptiveCNN_Hybrid_Model) architecture on the MURA dataset for classification

2.6 VGG19_AlexNet_AdaptiveCNN_Hybrid_Model (MURA Classification)

For fracture classification on the MURA wrist dataset, a VGG19–AlexNet Adaptive Hybrid CNN was developed, as illustrated in Figure 2. The architecture combines pretrained VGG19 and a custom AlexNet as parallel feature extractors to capture complementary visual features, while an adaptive CNN branch learns dataset-specific patterns, improving classification performance.

All the convolutional layers are frozen throughout training of both networks to avoid overfitting and preserve visual representations. Given an input radiograph $X \in \mathbb{R}^{224 \times 224 \times 3}$, deep feature maps are extracted as Eq.6:

$$F_{vgg} = f_{vgg19}(X), F_{alex} = f_{alexnet}(X) \quad (6)$$

where F_V and F_A are high-level feature maps capturing complementary texture and structural information from the radiograph. To reduce spatial dimensionality while preserving discriminative channel information, **Global Average Pooling (GAP)** is applied, it defined as Eq.7:

$$z_{vgg} = GAP(F_{vgg}), z_{alex} = GAP(F_{alex}) \quad (7)$$

These pooled vectors form a compact representation of pretrained deep features. In parallel, a **self-adaptive CNN refinement branch** is introduced to learn task-specific fracture patterns directly from the radiographs. This branch performs sequential

convolutions and pooling defined as **Eq.8 and 9**:

$$H_1 = \sigma (Conv_{3 \times 3.256} (X)), H_2 = MaxPool (H_1), \quad (8)$$

$$H_3 = \sigma (Conv_{3 \times 3.128} (H_2)), H_4 = MaxPool (H_3), \quad (9)$$

followed by flattening to produce a vector defined as Eq.10:

$$h = Flatten (H_4) \quad (10)$$

The pretrained features and adaptive features are fused through concatenation defined as **Eq.11**:

$$z = [g \| h] \quad (11)$$

A **fully connected classifier** comprised of three dense layers with 1024, 1024, and 512 neurons is used to process the fused vector z. Each layer is followed by a Dropout (rate = 0.2) to ensure regularization. The transformations are expressed as **Eq.12 to 14**:

$$u_1 = \sigma (W_1 z + b_1), u_1 = Dropout (u_1, 0.2), \quad (12)$$

$$u_2 = \sigma (W_2 u_1 + b_2), u_2 = Dropout (u_2, 0.2), \quad (13)$$

$$u_3 = \sigma (W_3 u_2 + b_3), u_3 = Dropout (u_3, 0.2), \quad (14)$$

The final classification layer uses Softmax activation to generate normal and fractured class probabilities, as demonstrated in **Eq. 15**:

$$\hat{y} = Softmax (W_o u_3 + b_o), \hat{y} = [p_0, p_1], p_0 + p_1 = 1 \quad (15)$$

Model optimization is performed using the **Adam optimizer** with learning rate $\alpha = 1 \times 10^{-4}$. The training objective minimizes **Binary Cross-Entropy (BCE)** loss defined as **Eq.16**:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (16)$$

The MURA dataset underwent three-fold cross-validation in an effort to enhance generalisability and decrease sampling bias. Using checkpointing, the model's best weights determined by validation accuracy were maintained throughout the model's 5-epoch training with a batch size of 50. By taking this route, the suggested fracture categorization system became more stable and resilient.

2.7 Proposed MURA Classification Algorithm of Hybrid Model

Algorithm 1: MURA classification algorithm of VGG19–AlexNet Adaptive Hybrid CNN model

Input: RGB image $224 \times 224 \times 3$

Branch 1 – VGG19 Feature Extractor (Frozen)

- Load pretrained VGG19 (include_top = False) and freeze all layers
- Extract feature map $7 \times 7 \times 512$
- Apply Global Average Pooling → **512-dim vector**

Branch 2 – AlexNet Feature Extractor (Frozen)

- Sequential Conv + ReLU + MaxPooling layers
- Final feature map $5 \times 5 \times 256$
- Apply Global Average Pooling → **256-dim vector**

Concatenation Block 1

- Merge VGG19 (512) + AlexNet (256)
- Output feature vector → **768**

Branch 3 – Self-Adaptive CNN

- Conv(3×3 , 256) → MaxPool
- Conv(3×3 , 128) → MaxPool
- Feature map $54 \times 54 \times 128$
- Flatten → **373,248-dim vector**

Concatenation Block 2

- Merge 768 + 373,248 → **374,016-dim vector**

Fully Connected Head

- 2 x Dense 1024 → ReLU → Dropout (0.2)
- Dense 512 → ReLU → Dropout (0.2)

Output Layer

- Dense(2) → Softmax
- **Binary Classification Output**

2.8 Fracture Classification via Deep Transfer Learning Models (FracAtlas Dataset)

Several pretrained CNN backbones, including EfficientNet-B0, ResNet50, DenseNet121, and VGG16, were individually evaluated using identical preprocessing, training strategy, and evaluation metrics. Based on validation performance, EfficientNet-B0 achieved the highest classification accuracy, F1-score, and ROC-AUC and was therefore selected as the final fracture classification backbone. Accordingly, all performance results reported for the FracAtlas classification model correspond to this optimized EfficientNet-B0 architecture rather than an ensemble of multiple networks as per Eq. 17:

$$F = f_{backbone}(X), \quad (17)$$

Where F consists of high-level structural and texture representations of bone areas. Since the backbone can generate spatial feature maps, global average pooling is implemented to summarize each channel to a single value to create a concise feature vector z. This vector is then inputted into a custom classifier that has fully connected layers. The transformation by the dense

layers can be summarized as **Eq.18**:

$$h = \phi(W_z + b) \quad (18)$$

Where $\phi(\cdot)$ denotes the ReLU activation function and W and b are learnable parameters. As a measure to mitigate overfitting, a 0.3 dropout rate is implemented between layers. The final neuron outputs a logit ℓ , which is converted into a fracture probability using the sigmoid function defined as **Eq.19**:

$$\hat{y} = \frac{1}{1 + e^{-\ell}} \quad (19)$$

Here, $\hat{y}$ represents the predicted probability that a fracture is present. In the attention-enhanced model, the extracted feature vector is further refined before classification. A small attention network produces a weight vector a , which scales the original features defined as **Eq.20**:

$$z_{att} = \alpha \odot z \quad (20)$$

The element-wise multiplication is denoted with $\odot$. This operation highlights informative features of fracture-related data and inhibits irrelevant background data. This is followed by a dense classifier run on the passed feature vector to get the final probability. The binary cross-entropy loss is minimized by model training and it is a measure of the variance among the predicted and actual labels. Since the dataset has fewer fracture cases, class weights are used to ensure that fracture samples make a greater contribution to the loss. The AdamW algorithm is used to perform optimization, and its learning rate is around 1×10^{-4} .

During initial training, the backbone layers were frozen and only the classifier layers were updated to preserve pretrained features and stabilize learning. After convergence, the backbone was unfrozen for fine-tuning to better adapt to radiographic images. During inference, fracture predictions were generated using a probability threshold of $\tau = 0.5$.

The backbone layer is frozen in the first stages of training where the classifier layers only are trained. This store pretrained visual information, and stabilize education. After convergence has been achieved, the backbone can then be unfrozen so that fine-tuning can be done so that the model can better adjust to the features of radiographic images. The predicted probability is compared with a threshold $\tau = 0.5$ at inference time. If $\hat{y} \geq 0.5$, the image is termed as fractured otherwise, it is termed non-fractured. It is an efficient combination of pretrained visual depictions, task-specific classification layers and optional attention-based refinement executed by using this deep transfer learning model to generate solid fracture detection outcomes.

2.9 Fracture segmentation with Attention U-Net+ Residual (FracAtlas Dataset)

The issue of fracture segmentation is formulated as the medical image segmentation at pixel level, the correct localization of fracture areas in radiographic images. Figure 3 shows the implemented fracture segmentation with Attention U-Net+ Residual architectures. Segmentation provides the spatial delineation of the injury unlike classification, which only predicts the presence or absence of a fracture, thus allows a detailed morphological examination and quantitative measurement. Fractures localization is carried out as a pixel localization segmentation of FracAtlas Radiographs. With an input of an X-ray image (resized $256(3) \times 256(3)$) the model will predict a binary mask that identifies the spatial location of fracture regions. Let the input image be represented as $X \in \mathbb{R}^{256 \times 256 \times 3}$, and the predicted segmentation mask defined as **Eq.21**:

$$\hat{Y} \in [0, 1]^{256 \times 256} \quad (21)$$

The primary architecture that is used in this task is that of Attention U-Net (Proposed algorithm 2), which is a refined variety of the classical U-Net, whose goal is to pay attention to clinically significant regions. An encoder-decoder network is the building block of the system. The encoder path uses convolutional block repeated applications which obtain hierarchical features as max-pooling is implemented progressively on the spatial resolution. The network will be able to learn robust representations of bone structures' edges, contours, and textures with the help of each ConvBlock's two convolutional layers, ReLU activation and Batch Normalization.

The bottleneck layer captures high-level contextual features, while the decoder progressively restores spatial resolution through transposed convolutions. Skip connections preserve fine-grained spatial details essential for detecting subtle fracture lines. To improve localization, attention gates are incorporated into the skip connections, suppressing irrelevant background information and enhancing fracture-related features. The attention mechanism acts as a weighting function, as defined in **Eq. 22**.

$$\tilde{F} = \alpha \bullet F \quad (22)$$

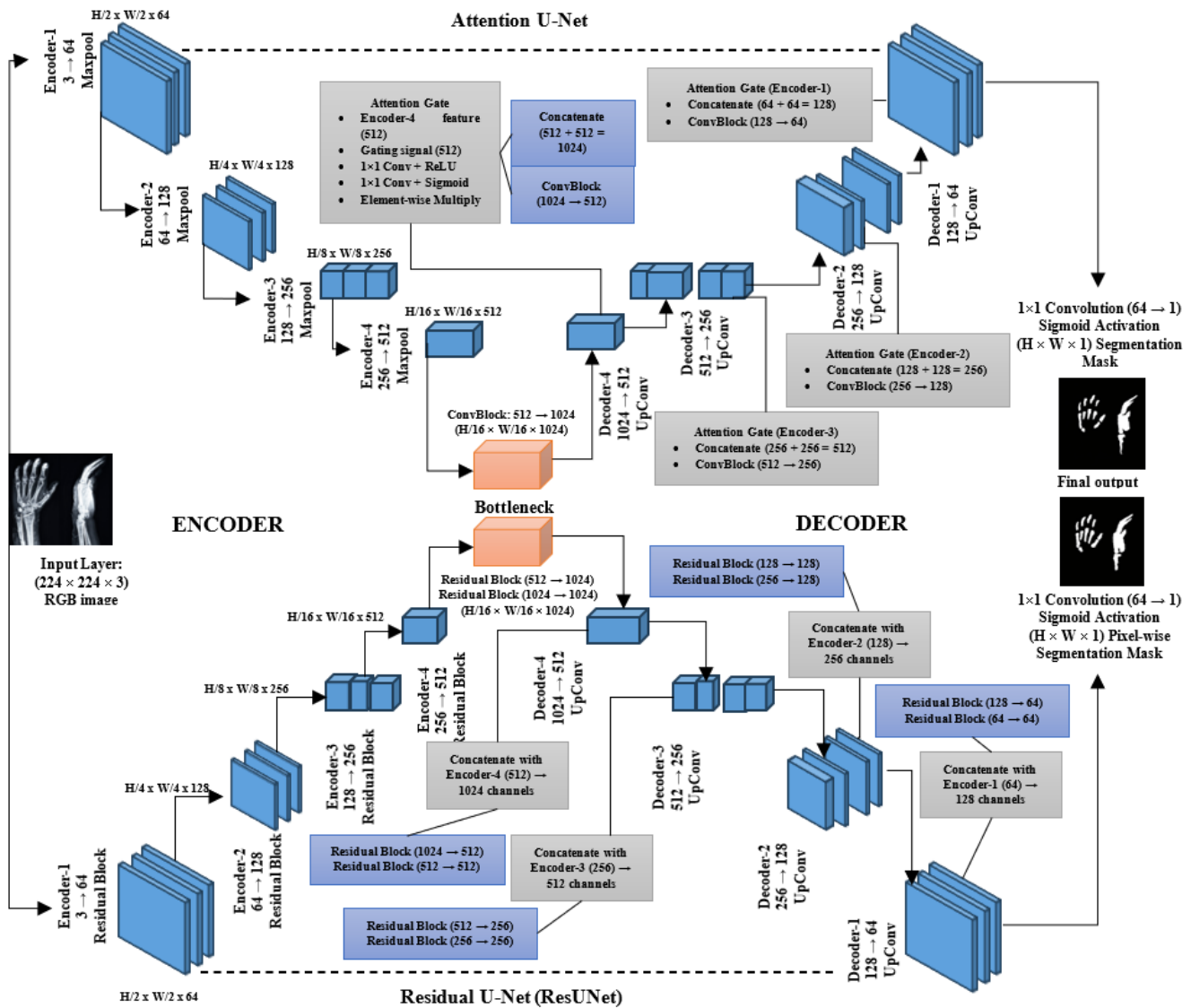


Fig 3. Trained segmentation proposed Attention U-Net+ Residual (FracAtlas Dataset) architecture

Where F represents encoder features and $\alpha \in [0, 1]$ is a learned attention coefficient map. This allows the model to pay attention to tiny discontinuities in bone structure while disregarding the surrounding soft tissue. A second architecture, Residual U-Net (ResUNet), is also implemented to enhance gradient flow and stabilize deep feature learning. Instead of plain convolution blocks, ResUNet (Proposed algorithm 3), uses Residual Blocks, where the input is added to the output of convolutional layers defined as Eq.23:

$$H_{out} = H_{in} + f(H_{in}) \quad (23)$$

This allows more navigation through networks without the vanishing gradient problem. This enhances partitioning of complicated fracture patterns. The last layer of both architectures applies a 1x 1-convolution to map features to single channel probability map and a sigmoid activation, which is defined as Eq.24:

$$\hat{Y}(i, j) = \sigma(z_{i, j}) \quad (24)$$

Where the pixel value of the image is given as the probability that the pixel is in a fracture area. The special segmentation loss functions can be trained to use in imbalanced classes and improving the accuracy of the boundaries. The Dice Loss is a loss that

is region based and is used in medical imaging segmentation tasks. where the pixel value of the image is given as the probability that the pixel is in a fracture area. The special segmentation loss functions can be trained to use in imbalanced classes and improving the accuracy of the boundaries. Dice Loss is a region-based medical picture segmentation loss define as **Eq.25**.

$$Dice = \frac{2|P \cap G|}{|P| + |G|} \quad (25)$$

where P is the anticipated mask and G is the ground truth. Image segmentation makes use of the Jaccard Index, also known as the Intersection over Union (IoU), a region-based evaluation statistic that measures how close the predicted mask is to the ground truth mask.. While Dice focuses on overlap proportionally, IoU measures how much the predicted region overlaps with the true region relative to their total union. IoU (Jaccard) Loss evaluates region agreement defined as **Eq.26**:

$$IoU = \frac{|P \cap G|}{|P \cup G|} \quad (26)$$

To leverage the strengths of multiple metrics, a Combined Segmentation Loss is used defined as **Eq.27**:

$$\mathcal{L}_{total} = 0.5\mathcal{L}_{BCE} + 0.3\mathcal{L}_{Dice} + 0.2\mathcal{L}_{IoU} \quad (27)$$

This weighted loss formulation aims at achieving a balance between the precision of the pixels, the overlap of the regions, and the precision of the boundaries. This segmentation framework is based on an encoder-decoder structure, attention mechanism, residual network and optimized loss function, which can accurately localize fracture regions on radiographic images and extract detailed spatial features for quantitative fracture analysis.

2.10 Post-Processing of Segmentation Masks

Post-processing of segmentation mask refers to improvements and optimizations performed on the segmentation mask results generated by the deep learning model. Most of the time a segmentation model (such as U-Net) will produce a result, usually along with a probability map that provides an indication of how likely each pixel is to be in the fracture region. The network segments the image, and then produces a probability map $P(x,y) \in [0,1]$ of which pixels are in a fracture region. The first step is to use the thresholding given in **Eq.28** to transform this probability map into a binary mask.

$$M(x,y) = \begin{cases} 1, & P(x,y) \geq \tau \\ 0, & P(x,y) < \tau \end{cases} \quad (28)$$

Where τ is a predefined threshold (e.g., 0.5). This operation will split foreground (fracture) and background pixels. Morphological operations are used for the removal of small artifacts and to fill up the minor discontinuities. Morphological closing fills small gaps of fracture regions specified by **Eq.29**:

$$M \bullet B = (M \oplus B) \ominus B \quad (29)$$

Where B is the structuring element, $\ominus$ denotes erosion and $\oplus$ denotes dilation. This helps to maintain continuity of either a thin fracture area or a thin fractured area. Then connected component filtering is applied to eliminate small regions that are anatomically implausible. The area of each connected component C_k is computed as defined as **Eq.30**.

$$A_k = \sum_{(x,y) \in C_k} 1 \quad (30)$$

Any component with an area less than a certain minimum is rejected. Lastly, smoothing can be performed on the jagged boundaries which are defined as **Eq.31** for smoothing:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (31)$$

This sequence progresses will remove noise, mask coherence and is for anatomically consistent fracture segmentation outcomes.

2.11 Proposed Segmentation algorithm of Attention U-Net

Algorithm 2: Segmentation algorithm of Attention U-Net

Input: RGB medical image ($H \times W \times 3$)

Encoder with 4 down-sampling stages

- Each stage: Conv $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Conv $\rightarrow$ BatchNorm $\rightarrow$ ReLU
- Max-Pooling (2×2) after each stage
- Feature channels increase: **64** $\rightarrow$ **128** $\rightarrow$ **256** $\rightarrow$ **512**

Bottleneck: ConvBlock expanding to 1024 channels

Decoder with 4 up-sampling stages

- Transposed Convolution for up-sampling
- **Attention Gates** applied on encoder skip features
- Concatenation of gated encoder features with decoder features
- ConvBlock refinement at each stage

Final Layer: 1×1 Convolution

Output: Single-channel segmentation mask

2.12 Proposed Segmentation algorithm of Residual U-Net (ResUNet)

Algorithm 3: Segmentation algorithm of Residual U-Net (ResUNet)

Input: RGB medical image ($H \times W \times 3$)

Encoder with 4 down-sampling stages

- Each stage uses Residual Blocks with shortcut connections
- Two Conv $\rightarrow$ BatchNorm $\rightarrow$ ReLU operations + Identity Add
- Max-Pooling (2×2) after each stage
- Feature channels increase: **64** $\rightarrow$ **128** $\rightarrow$ **256** $\rightarrow$ **512**

Bridge/Bottleneck: Two stacked Residual Blocks expanding to 1024 channels

Decoder with 4 up-sampling stages

- Transposed Convolution for up-sampling
- Skip-connection concatenation with encoder features
- Two Residual Blocks for feature refinement at each stage

Final Layer: 1×1 Convolution

Output: Single-channel pixel-wise segmentation map

2.13 Model Classification Measures

The proposed framework performance was assessed by classification and segmentation metrics. Fracture detection performance was evaluated using Accuracy, Precision, Recall (Sensitivity), F1 score, ROC–AUC and Cohen’s Kappa for classification. The accuracy is the overall prediction correctness, the Precision is the reliability of fracture prediction, the Recall is the ability to identify true fractures, F1 score is a compromise between Precision and Recall, ROC–AUC is the assessment of class separability, and Cohen’s Kappa is the assessment of the agreement beyond chance. Additionally, a Confusion Matrix was created to evaluate the True Positives, True Negatives, False Positives, and False Negatives. IoU, Dice Coefficient, Pixel Accuracy, Precision and Recall were used to measure the overlap and accuracy between the predicted fracture area and the ground truth annotations

for segmentation. The mathematical expressions for these metrics are given in Eqs. (32)– (36)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (32)$$

$$Precision = \frac{TP}{TP + FP} \quad (33)$$

$$Recall = \frac{TP}{TP + FN} \quad (34)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (35)$$

$$Cohen's\ Kappa = \frac{p_o - p_e}{1 - p_e} \quad (36)$$

These measurements can be used together to provide a comprehensive evaluation of the accuracy of diagnostic detection and anatomical localization, and thus validate the system's clinical reliability and effectiveness.

3 Results and Discussion

3.1 Result Analysis

In this section, the proposed framework is assessed on two datasets, the MURA dataset and FracAtlas dataset. It compares the classification result of MSD abnormality detection and fracture identification with other classification result, and classification result accuracy to locate the fractures precisely. All the tests are run with Pytorch and on both datasets and with Nvidia graphics processors and Intel Xeon E5-2600 v4 3.60GHz central processing units (CPUs). Anatomic variation MURA and fracture variability FracAtlas, are the data-specific attributes that have been discussed. Training behaviour, diagnostic measures, and post-processing performance are used to assess the system's ability to convert deep learning predictions into practical clinical information and quantitative fracture measurements.

3.2 Experiments with MURA dataset

The MURA dataset was used to evaluate the proposed VGG19–AlexNet Adaptive Hybrid CNN for musculoskeletal abnormality classification. After preprocessing and SMOTE-based balancing, the dataset contained 18,177 samples, which were divided into training and testing sets. As shown in Table 2, the model attained excellent performance with a testing accuracy of **99.63%**, F1-score of **99.73%**, Cohen's Kappa of **99.15%**, and ROC-AUC of **99.996%**. Furthermore, Figure 4 demonstrates stable training behavior, with rapid convergence, decreasing loss, and minimal overfitting, confirming the strong generalization capability of the proposed model.

Table 2. Classification performance of the proposed vgg19-alexnet adaptive hybrid cnn on the mura dataset

Metric	Accuracy	Precision	Recall (Sensitivity)	F1 Score	Cohen's Kappa	Specificity	ROC AUC Score
Training Value	0.99907	0.99938	0.99922	0.99930	0.99792	0.99878	0.99999
Testing Value	0.99629	0.99515	0.99939	0.99727	0.99148	0.98976	0.99996

As shown in Figure 4, the model's robustness is also evident in rendering examples of its qualitative output, the ROC curve, and the confusion matrix. The confusion matrix shows that there are 3,282 positive results and 1,547 negative results, with a sensitivity of 0.9994 and a specificity of 0.9898. There are only 16 false positives and 2 false negatives. The ROC curve has an AUC of 0.99996, which implies good discriminatory power over all thresholds. Figure 5 also shows 2 representative radiographs correctly classified "Predicted: Negative" (normal anatomical structures). One of the images depicts a posteroanterior radiograph of the wrist where cortical margins are intact, bone density is normal and joint spaces are preserved, whereas the other image depicts a lateral radiograph of the elbow where the humerus, radius and ulna are aligned properly and no abnormalities. The forecasts were produced with a forecasting time of approximately 5 seconds per step and this was a viable computational capacity. This specificity is of essential relevance since it helps to lower the frequency of false positive tests and the efficiency of the clinical radiology procedure.

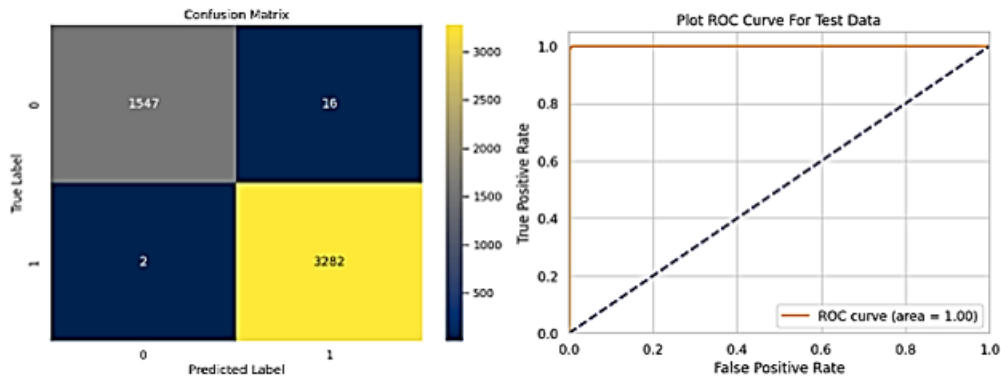


Fig 4. Confusion matrix and ROC curve of proposed hybrid model on MURA dataset

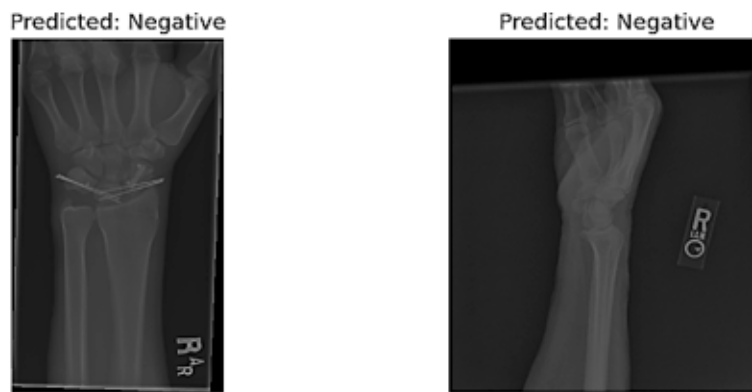


Fig 5. Output images with their predicted labels

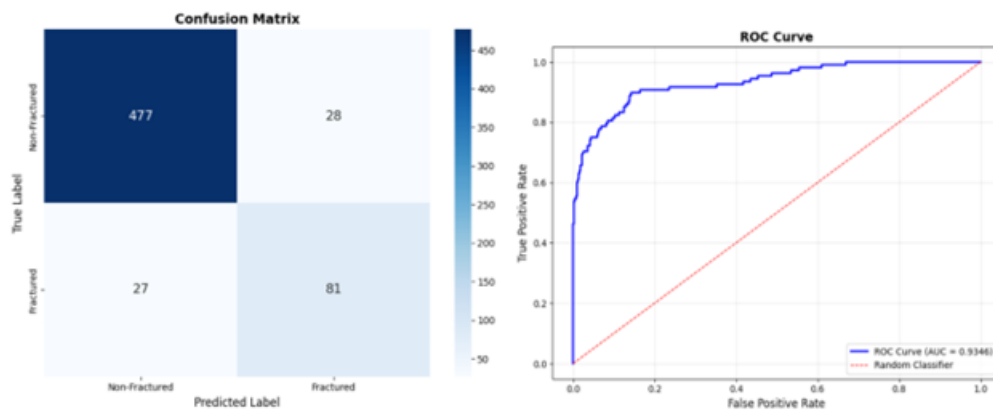
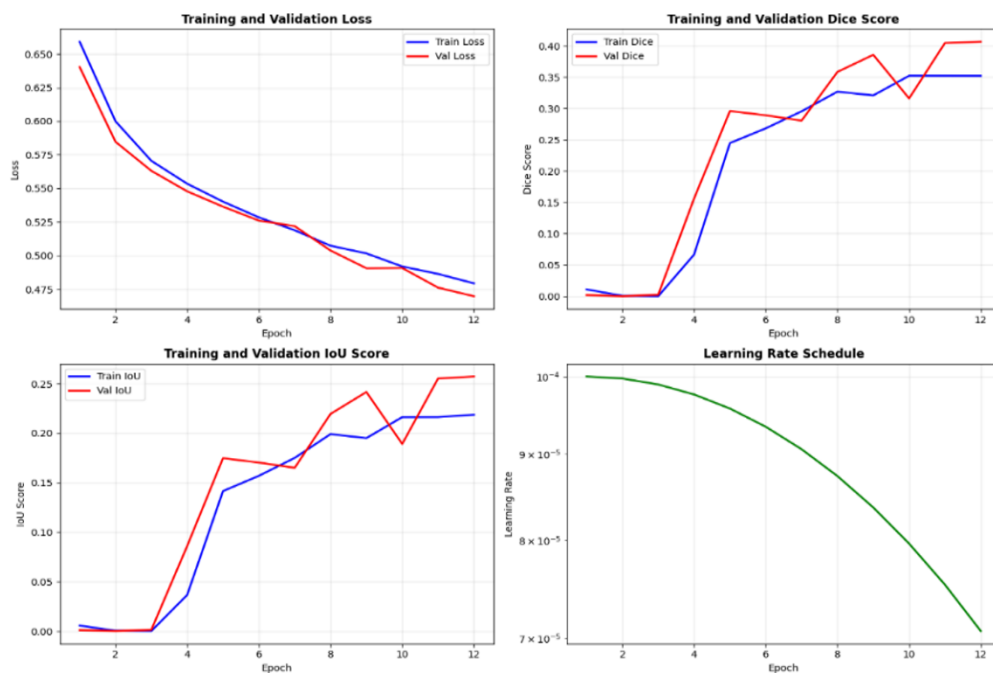
3.3 Experiments with FracAtlas Dataset

The performance of the proposed fracture analysis framework is summarized in Table 3 for the FracAtlas dataset with the transfer learning-based classification model and the Attention-ResUNet segmentation model. The classification model, which is based on the baseline architectures of EfficientNet, ResNet, and DenseNet, achieved an accuracy of 91.03%, precision of 0.7431, recall of 0.7500, F1-score of 0.7465, and an AUC-ROC of 0.9346, which exhibited good performance in terms of fracture detection and class separability. The Attention-ResUNet, in terms of localization, obtained a Dice score of 0.3843 ± 0.2436 and an IoU score of 0.2663 ± 0.1907 . Although the segmentation task is still difficult because of the low size of the fracture areas and the low contrast of the images, the pixel accuracy of the model was high (0.9902). Additionally, the segmentation recall (0.6484) was higher than the segmentation precision (0.3046), signifying that the algorithm was able to identify the regions of fractures with a high rate of success, but also with some false-positive predictions. Overall, the suggested system has excellent fracture detection performance and reasonable localization capabilities. Figure 6 presents the Confusion matrix and ROC curve obtained from the transfer learning-based fracture classification model on the FracAtlas dataset. The confusion matrix shows that most cases of non-fracture and fracture were correctly classified, and the ROC curve had an AUC of 0.9346, which is excellent and has a good discrimination ability and performance in fracture detection.

In fracture segmentation, the Attention-ResUNet model exhibited stable and well-aligned training and validation performance throughout the training process. Over 12 epochs, both training and validation losses gradually reduced from around 0.66 to below 0.50, showing good convergence without much overfitting as shown in Figure 7. Likewise, the Dice Score and Intersection over Union (IoU) also steadily improved up to about 0.35–0.40 and 0.22–0.26, respectively, in the last epochs. These values are considered to be moderate values in respect to segmentation performance, but the segmentation of fracture regions is not easy in radiographs because the fracture regions are small, irregular, and have low contrast. The progressive decrease of the learning rate also played a role in stable optimization and better generalization of the model. In summary, the

Table 3. Performance Summary on the FracAtlas Dataset

Metric	Classification	Segmentation
Accuracy	0.9103	0.9902 (Pixel Accuracy)
Precision	0.7431	0.3046
Recall (Sensitivity)	0.7500	0.6484
F1 / Dice Score	0.7465 (F1 Score)	0.3843 ± 0.2436 (Dice)
IoU Score	—	0.2663 ± 0.1907
AUC-ROC	0.9346	—


Fig 6. Confusion matrix and ROC curve of FracAtlas classification using TL models

Fig 7. Train/Val/Learning Rate curve of FracAtlas Segmentation using Attention with ResUNet

Attention-ResUNet model proved to be capable of accurately localizing fractures, and the improvement of data augmentation, training methods, and network structure could also improve the accuracy of clinical grade fracture segmentation.

Tables 4- 7 provide the quantitative fracture analysis and overall performance of proposed framework. A representative fracture measurement report is provided in Table 4 and consists of fracture area, perimeter, length, width (dimensions of the fracture), aspect ratio, orientation (direction of the fracture), centroid location, and shape descriptors. Table 5 shows that there is substantial variation in fracture size and morphology amongst the different fracture samples. Spatial and structural features, such as orientation, eccentricity, compactness, convexity, centroid coordinates and skeleton length, are summarized in Table 6 and demonstrate the ability of the framework to represent fracture geometry and anatomical alignment. As presented in Table 7, the overall performance of the fracture detection system is excellent with high classification accuracy (Accuracy: 91.03%, AUC: 93.46%), good fracture localization (Dice: 0.384, IoU: 0.266, Pixel Accuracy: 99.02%), good fracture morphometric measurement extraction, and excellent structured visual report generation. All of these findings suggest that the proposed framework is not only capable of detecting and localizing fractures but also offers clinically meaningful quantitative measurements and all-in-one diagnostic reporting.

Table 4. Detailed Fracture Measurement Report

Category	Metric	Value
Detection Summary	Number of Fracture Regions	2
	Total Area	809 pixels (202.25 mm ²)
	Perimeter	150.7 pixels (75.36 mm)
Overall Measurements	Length	159.6 pixels (79.78 mm)
	Width	23.6 pixels (11.81 mm)
	Aspect Ratio	6.76
	Orientation	−5.6°
	Centroid	(121.9, 111.5)
	Eccentricity	0.511
	Compactness	0.448
Shape Metrics	Convexity	0.253
	Skeleton Length	10.50 mm

Table 5. Fracture Size and Geometric Measurements

Sample ID	No. of Regions	Area (pixels)	Area (mm ²)	Length (mm)	Width (mm)	Perimeter (mm)	Aspect Ratio
1	2	809	202.25	79.78	11.81	75.36	6.76
2	2	283	70.75	24.00	7.50	39.92	3.20
3	1	908	227.00	23.42	12.68	65.48	1.85
4	1	1235	308.75	37.27	16.92	110.79	2.20
5	1	783	195.75	19.76	12.33	57.16	1.60
6	1	535	133.75	14.98	10.73	43.63	1.40
7	1	318	79.50	9.50	9.50	31.85	1.00
8	1	108	27.00	7.50	4.00	19.49	1.88
9	2	1222	305.50	78.58	17.65	95.84	4.45
10	1	380	95.00	11.30	9.90	35.80	1.14

3.4 App Development for Fracture Analysis

A mobile-based Fracture Analysis Application was developed to translate the proposed AI framework into a practical clinical support tool. The frontend was developed with Flutter and the inference engine, real-time fracture detection, segmentation and extraction of quantitative measurement, was done in the backend using FastAPI. Users can upload X-ray images via the mobile interface, and the images are securely processed by the models in the backend.

In the case of normal, the application will show a green diagnostic screen which contains the classification result, confidence score, and a brief clinical summary as shown in Figure 8. The user-friendly interface gives a report without fracture, which

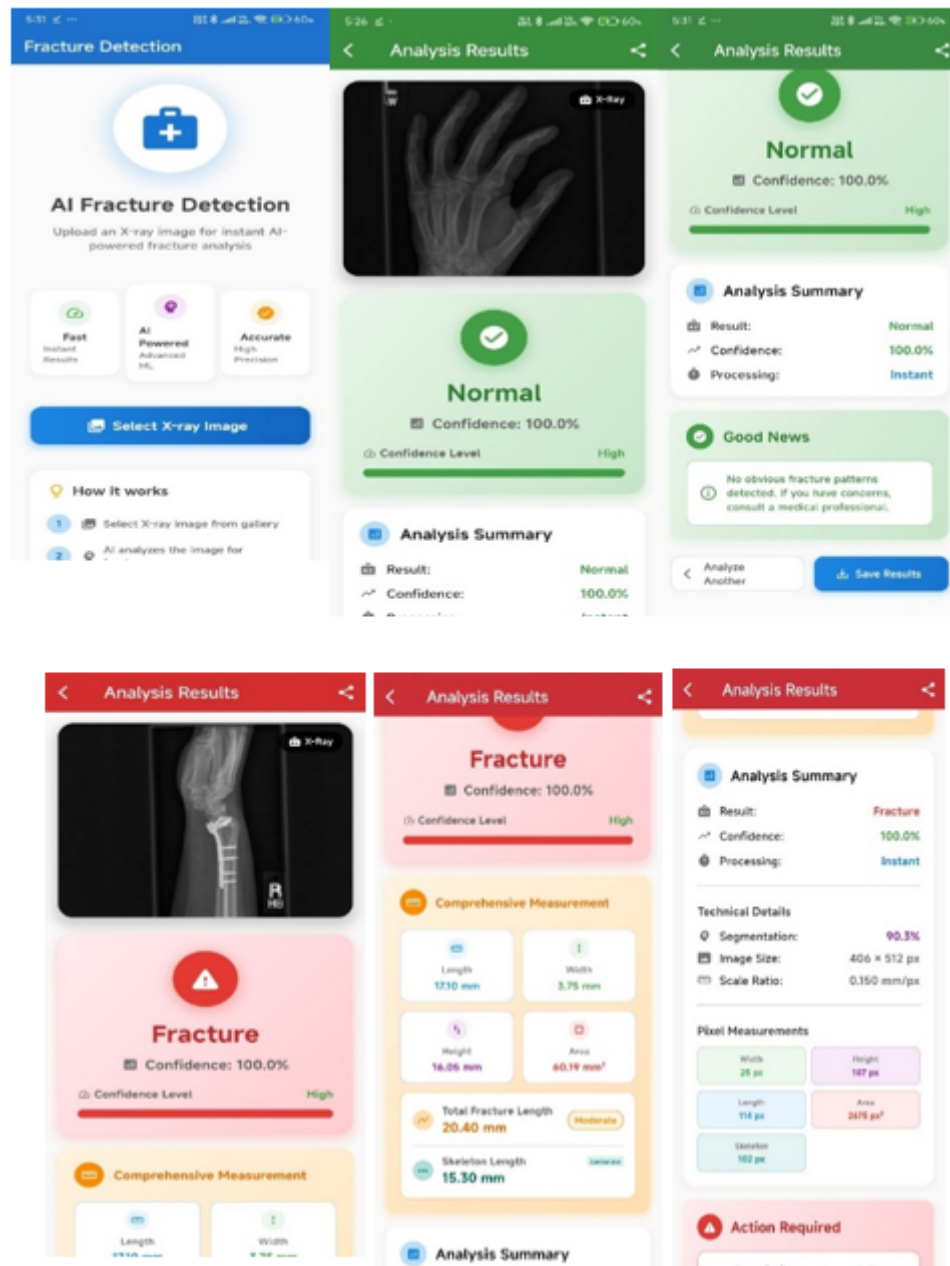


Fig 8. Mobile application interfaces for automated fracture analysis showing normal-case diagnosis, fracture detection, quantitative measurements, and structured clinical reporting

Table 6. Fracture Location, Shape and Structural Metrics

Sample ID	Orientation (°)	Eccentricity	Compactness	Convexity	Centroid X	Centroid Y	Skeleton Length (mm)
1	-5.63	0.511	0.448	0.253	121.88	111.52	10.5
2	-88.01	0.527	0.558	0.505	162.64	148.50	1.5
3	-44.31	0.868	0.665	0.954	26.15	131.24	26.5
4	-95.79	0.878	0.316	0.694	136.34	117.47	46.0
5	-79.18	0.816	0.753	0.970	125.70	116.88	15.0
6	-19.88	0.713	0.883	1.042	101.97	120.44	6.5
7	-116.57	0.263	0.985	1.074	52.15	185.67	1.5
8	-18.76	0.878	0.894	1.174	105.52	36.07	5.5
9	5.97	0.787	0.418	0.325	140.66	94.99	26.0
10	-101.33	0.625	0.932	1.061	165.57	111.08	5.0

Table 7. Final Selected Classification Backbone: EfficientNet-B0 (Best Performing Individual Model)

Component	Model / Output	Key Results
Classification	EfficientNet-B0	Accuracy: 91.03%, F1: 74.65%, AUC: 93.46%
Segmentation	U-Net + ResNet34	Dice: 0.384, IoU: 0.266, Pixel Acc: 99.02%
Fracture Measurements	Morphological Analysis	Area, Length, Width, Orientation, Shape metrics
Outputs Generated	Visual & Reports	Mask overlays, annotated images, text reports, CSV summaries
	Classification Model File	classification_model.pth
Deployment Files	Segmentation Model File	segmentation_model.pth
	Configuration Files	model_config.json, pipeline_config.json

enables clinical reassurance. The segmentation and measurement pipeline is automatically triggered for fracture-positive cases. The application creates fracture masks and provides quantitative parameters for fractures such as fracture length, width, area, orientation, and centroid location. Results are displayed in an intuitive, red alert interface that displays fracture visualization, confidence indicators, quantitative measurements, technical image details and structured diagnostic reporting. The automated fracture detection, segmentation, quantitative analysis, and clinical reporting are seamlessly integrated within the Flutter–FastAPI framework. The deployment illustrates the real-world utility of the proposed system for clinical support applications, telemedicine and health research applications.

This study introduces a comprehensive, clinically diverse, selective design of deep learning for the task of automated fracture detection, localization and quantitative analysis of radiographic images. The system provided is a hierarchical fracture detection, attention-based segmentation and morphometric measure extraction in one pipeline, rather than the traditional approaches of fracture detection with a classification process and segmenting with a segmentation process. The hybrid VGG19AlexNet Adaptive CNN demonstrated almost perfect performance on the MURA wrist dataset (accuracy: 0.9963), proving that it can be trained to learn the discriminative features of the MSD. The transfer-learning classifier was able to achieve high accuracy (0.9103) on more difficult FracAtlas data, showing good generalization between the dynamics of anatomical regions and imaging conditions. Despite moderate segmentation Dice and IoU scores, the model had been able to localize fracture regions sufficiently to follow by downstream quantitative analysis, and as such reminded the clinical importance of using both detection and structural assessment rather than relying on classification alone.

The emphasis on interpretability and application to translation is an important contribution of the work. The clinical parameters that are obtained from the post-segmentation morphological analysis are fracture length, width, orientation, area and skeletal continuity. The measurements give objective parameters for use in decision making in orthopedic cases, fracture severity grading, and long-term monitoring. Also, the implementation of the AI pipeline in a future analysis application based on Flutter and FastAPI demonstrates the likelihood of real-time clinical use. The proposed hybrid classification model generates a better recall accuracy, and F1 score when compared to the existing CNN and ensemble-based classifiers, which demonstrate higher sensitivity of fracture detection. Several past studies (Table 8) yielded high-quality results on filtered datasets, but they lacked quantitative fracture characterization and did not support infrastructure for implementation. This is being addressed by the current system, which offers high diagnostic performance and easy to explain outputs, as well as automated reporting, enhancing the role of computer-aided musculoskeletal imaging.

Table 8. Comparison among proposed and existing baseline models for fracture classification

Reference	Model	Accu- racy	Preci- sion	Recall	F1 Score	Cohen's Kappa	AUC Score	App
Our Proposed Model	VGG19–AlexNet Adaptive Hybrid CNN (Proposed)	0.996	0.995	0.999	0.997	0.991	0.990	×
He, Wang & Zhao ⁽¹⁸⁾	MSCNN-GCN	0.952	N/A	N/A	0.969	0.936	N/A	×
	DenseNet169	0.951	N/A	N/A	0.968	0.931	N/A	×
	Res+Dense	0.920	0.950	0.820	N/A	0.820	0.970	×
	Calibrated Model	0.920	0.960	0.870	N/A	0.820	0.970	×
Singh et al. ⁽⁸⁾	ResNet	0.900	0.940	0.810	N/A	0.800	0.970	×
	ConvNet	0.890	0.920	0.750	N/A	0.760	0.930	×
	DenseNet	0.890	0.920	0.800	N/A	0.770	0.950	×
Chawla & Kapoor ⁽²³⁾	ComDNet-512	0.894	0.860	0.940	0.890	0.788	0.890	×
Dahal, Thapa & Panth ⁽²⁴⁾	Ensemble	0.885	0.871	0.857	0.892	0.770	0.927	×
Liang & Gu	HRD-Model	0.867	0.823	0.906	0.812	0.767	0.940	×
	ResNet152	0.852	0.913	0.742	0.819	0.696	N/A	×
	ResNet101	0.848	0.901	0.744	0.815	0.688	N/A	×
	VGG16	0.846	0.917	0.721	0.808	0.682	N/A	×
	ResNeXt50	0.844	0.893	0.742	0.810	0.680	N/A	×
Mehr ⁽¹⁹⁾	VGG19	0.844	0.907	0.727	0.807	0.679	N/A	×
	ResNet34	0.842	0.905	0.725	0.805	0.675	N/A	×
	MobileNetV2	0.837	0.896	0.721	0.799	0.665	N/A	×
	ResNet50	0.838	0.910	0.710	0.798	0.666	N/A	×
	VGG-16 + ViT	0.828	0.780	0.880	N/A	N/A	N/A	×
Chawla & Kapoor ⁽²³⁾	Inception V3	0.787	0.854	0.632	0.727	0.558	N/A	×
	InceptionResNetV2	0.776	0.840	0.720	0.780	0.550	N/A	×
	DenseNet201	0.765	0.840	0.690	0.760	0.530	N/A	×

Although the proposed Attention-ResUNet achieved satisfactory localization performance, the moderate Dice and IoU scores indicate that precise delineation of fracture boundaries remains challenging due to extremely thin fracture lines, low image contrast, irregular anatomical structures, and annotation variability within the FracAtlas dataset. Future research may improve boundary refinement through deep supervision, where auxiliary loss functions are applied at multiple decoder levels to encourage accurate multi-scale feature learning. In addition, boundary-aware loss functions and edge-guided refinement modules could further improve localization of fine fracture contours. Recently, transformer-based segmentation models, including Vision Transformers (ViTs), Swin Transformers, and hybrid CNN-transformer architectures, have demonstrated superior capability in modeling long-range contextual dependencies while preserving global anatomical information. Integrating transformer attention with the proposed framework may substantially improve fracture boundary localization, particularly for subtle cortical discontinuities and complex fracture morphologies.

Limitations: Although the proposed framework achieved strong fracture detection and analysis performance, several limitations remain. The moderate Dice and IoU scores reflect the difficulty of accurately segmenting small, irregular, and low-contrast fracture regions. It is based on publicly available datasets that do not necessarily reflect the full clinical variability. Furthermore, synthetic class balancing might not fully reflect the natural fracture distribution and multicenter clinical validation is still needed. Future work will focus on larger multicenter datasets, transformer-based attention mechanisms, uncertainty-aware prediction, self-supervised learning, advanced data augmentation, and PACS/DICOM integration to improve localization accuracy, scalability, and clinical deployment readiness.

Novelty of the Proposed Framework: The novelty of this work is the creation of an integrated and clinically oriented framework for fracture analysis that encompasses abnormality classification, fracture localization, quantitative morphometric analysis and structured reporting in a single pipeline. In contrast to most of the previous works which focus on either the classification of the fracture or the segmentation, the proposed framework combines hierarchical abnormality screening based

on a VGG19–AlexNet Adaptive Hybrid CNN with abnormality segmentation based on Attention-ResUNet and automated geometric measurement extraction. Another unique feature is the ability to provide clinically relevant outputs such as fracture area, length, width, orientation, centroid location, shape descriptors, probability heatmaps, and structured diagnostic reports. Furthermore, the complete system when deployed into the mobile application created using the Flutter and FastAPI frameworks demonstrates the potential of this proposed solution for clinical application, beyond the algorithmic evaluation. The proposed framework is special in that it merges detection, localization, quantitative evaluation and contextualized reporting for deployment.

Discussion: The experimental results show the effectiveness of the proposed framework for the automated fracture analysis. The VGG19–AlexNet Adaptive Hybrid CNN achieved a classification accuracy of 99.63% and AUC of 0.9999 on the MURA dataset, while the transfer learning model attained 91.03% accuracy and 0.9346 AUC on the FracAtlas dataset. The Attention-ResUNet provided reliable fracture localization despite the challenges of segmenting small and low-contrast fracture regions. Unlike existing studies that focus mainly on fracture detection, the proposed framework additionally performs quantitative fracture characterization, interpretable visualization, and structured reporting. The successful deployment of the system in a mobile application further demonstrates its potential for clinical decision support, telemedicine, and real-time fracture assessment.

Compared with recent AI-based fracture analysis studies, the proposed framework demonstrates several important advantages. Recent investigations have shown that transfer learning and deep CNN architectures provide high fracture detection accuracy but remain largely focused on image-level classification without comprehensive quantitative characterization^{(10), (11), (6)}. Similarly, recent wrist fracture detection studies have demonstrated effective fracture classification; however, these approaches do not integrate pixel-level localization with automated morphometric measurements^{(9), (12)}.

4 Conclusion

This study aimed to establish and clinically interpret a new DL framework for automated fracture detection, localization, and quantitative analysis from X-ray images. The proposed framework was implemented in the three components: VGG19–AlexNet Adaptive Hybrid CNN for musculoskeletal abnormality classification, Attention-ResUNet model for fracture segmentation, and automated morphometric analysis module for fracture characterization and structured reporting, which successfully achieved this objective. This is one of the novelties of the work as it combines classification, localization, extraction of quantitative measurements, and mobile-based deployment within one end-to-end diagnostic pipeline, surpassing traditional methods that deal with classification or segmentation alone. Experimental results have shown good performance with the hybrid classifier outperforming with 0.9963 accuracy and 0.9973 F1-score for the MURA set, and the FracAtlas classifier with an accuracy of 0.9103 and an AUC of 0.9346. In addition, the effectiveness of the Attention-ResUNet segmentation model was demonstrated by obtaining a Dice score of 0.384 and a pixel accuracy of 0.9902, which allowed for reliable localization of fractures and downstream extraction of clinically relevant measurements such as fracture area, length, width, orientation, centroid location, and shape descriptors. The successful implementation of the framework in a mobile application (Flutter–FastAPI) further highlights its applicability in clinical support, telemedicine, and healthcare research. Multi-center clinical validation, finding the location of fractures through the use of a transformer, uncertainty-aware prediction mechanisms, integration of the PACS/DICOM and optimization of the deployment in real-time will be the focus of future work. Future intelligent fracture-analysis systems should integrate multiple AI technologies, such as explainable AI, quantitative morphometric evaluation, and clinical workflow integration, to improve reliability, interpretability, and decision support in orthopedics.

References

- 1) Galić I, Habijan M, Leventić H, Romić K. Machine Learning Empowering Personalized Medicine: A Comprehensive Review of Medical Image Analysis Methods. *Electronics*. 2023;12(21). <https://doi.org/10.3390/electronics12214411>.
- 2) Alzubaidi L, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. 2021;8(1):53. <https://doi.org/10.1186/s40537-021-00444-8>.
- 3) El-Shafai W, Algarni AD, Banby GME, El-Samie FEA, Soliman NE. Classification framework for COVID-19 diagnosis based on deep cnn models. *Intell Autom Soft Comput*. 2022;31(3):1561–1575. <https://doi.org/10.32604/IASC.2022.020386>.
- 4) Jones RM, et al. Assessment of a deep-learning system for fracture detection in musculoskeletal radiographs. *npj Digit Med*. 2020;3(1):1–6. <https://doi.org/10.1038/s41746-020-00352-w>.
- 5) Debs P, Fayad LM. The promise and limitations of artificial intelligence in musculoskeletal imaging. *Front Radiol*. 2023;3. <https://doi.org/10.3389/fradi.2023.1242902>.
- 6) Laçi H, Sevrani K, Iqbal S. Deep learning approaches for classification tasks in medical X-ray, MRI, and ultrasound images: a scoping review. *BMC Med Imaging*. 2025;25(1):156–173. <https://doi.org/10.1186/s12880-025-01701-5>.
- 7) Fouad DM, Mahfouz MM, Mohamed MM, Elzanaty MY, El-Hafeez TA. Classification of musculoskeletal pain using machine learning. *Sci Rep*. 2025;15(1):27158. <https://doi.org/10.1038/s41598-025-12049-9>.

- 8) Singh G. Hybrid Deep Learning Approach for Automatic Detection in Musculoskeletal Radiographs Gurpreet. 2022.
- 9) Rafi SM, Anuradha T, Srinivas KSVVR. Wrist Fracture Detection Using Deep Learning. IEEE. 2025. <https://doi.org/10.1109/SCEEC64059.2025.10940086>.
- 10) Aldhyani T, et al. Diagnosis and detection of bone fracture in radiographic images using deep learning approaches. *Front Med.* 2025;11:1–15. <https://doi.org/10.3389/fmed.2024.1506686>.
- 11) Abdellatif NAM, El-Rawy AS, Abdellatif AR, Al-Shatouri MA. Assessment of artificial intelligence-aided x-ray in diagnosis of bone fractures in emergency setting. *Egypt J Radiol Nucl Med.* 2025;56(1):153. <https://doi.org/10.1186/s43055-025-01575-1>.
- 12) Duan G, Zhang S, Shang Y, Kong W. Research on X-ray Diagnosis Model of Musculoskeletal Diseases Based on Deep Learning. *Appl Sci.* 2024;14(8). <https://doi.org/10.3390/app14083451>.
- 13) Singh G, Kumar P, Anand D. Hybrid Deep Learning Model for Classification and Prediction of Abnormalities in Upper and Lower Extremities of Musculoskeletal Radiographs. *SN Comput Sci.* 2024;6(1):32. <https://doi.org/10.1007/s42979-024-03575-w>.
- 14) Hareendranathan AR, Tripathi A, Panicker MR, Zhang J, Boora N, Jaremko J. Deep Learning Approach for Automatic Wrist Fracture Detection Using Ultrasound Bone Probability Maps. *SN Compr Clin Med.* 2023;5(1). <https://doi.org/10.1007/s42399-023-01608-8>.
- 15) Hinterwimmer F, et al. SAM-X: sorting algorithm for musculoskeletal x-ray radiography. *Eur Radiol.* 2022;33(3):1537–1544. <https://doi.org/10.1007/s00330-022-09184-6>.
- 16) Hardalaç F, et al. Fracture Detection in Wrist X-ray Images Using Deep Learning-Based Object Detection Models. *Sensors.* 2022. <https://doi.org/10.3390/s22031285>.
- 17) Uysal F, Hardalaç F, Peker O, Tolunay T, Tokgöz N. Classification of Shoulder X-ray Images with Deep Learning Ensemble Models. *Appl Sci.* 2021;11(6):2723. <https://doi.org/10.3390/app11062723>.
- 18) He M, Wang X, Zhao Y. A calibrated deep learning ensemble for abnormality detection in musculoskeletal radiographs. *Sci Rep.* 2021;11(1):9097. <https://doi.org/10.1038/s41598-021-88578-w>.
- 19) Mehr G. Automating Abnormality Detection in Musculoskeletal Radiographs through Deep Learning. *arXiv.* 2020 Oct.
- 20) Mostafiz R, Uddin MS, Alam NA, Reza MM, Rahman MM. Covid-19 detection in chest X-ray through random forest classifier using a hybridization of deep CNN and DWT optimized features. *J King Saud Univ - Comput Inf Sci.* 2022;34(6):3226–3235. <https://doi.org/10.1016/j.jksuci.2020.12.010>.
- 21) Mahmudah KR, Indriani F, Takemori-Sakai Y, Iwata Y, Wada T, Satou K. Classification of Imbalanced Data Represented as Binary Features. *Appl Sci.* 2021;11(17):7825. <https://doi.org/10.3390/app11177825>.
- 22) Khattab R, Abdelmaksoud IR, Abdelrazek S. Automated detection of COVID-19 and pneumonia diseases using data mining and transfer learning algorithms with focal loss from chest X-ray images. *Appl Soft Comput.* 2024;162:111806. <https://doi.org/10.1016/j.asoc.2024.111806>.
- 23) Chawla N, Kapoor N. Musculoskeletal Abnormality Detection in Humerus Radiographs Using Deep Learning. *Rev d'Intelligence Artif.* 2020;34(2):209–214. <https://doi.org/10.18280/ria.340212>.
- 24) Dahal S, Thapa R, Panth S. A Hybrid Deep Learning Model for Musculoskeletal Abnormality Detection. *Proceedings of 15th IOE Graduate Conference.* 2024;15:13–18.