DOI: https://doi.org/10.61762/pjstvol9iss1art005

PLATFORM - A Journal of Science & Technology

Received: 21 March 2026, Revised: 28 June 2026, Accepted: 29 June 2026, Published: 30 June 2026, Publisher: UTP Press, Creative Commons: CC BY 4.0

HYPERPARAMETER-OPTIMISED U-NET FOR BONE FRACTURE SEGMENTATION IN MUSCULOSKELETAL X-RAY IMAGES

Ibrahim Muhammad Kurah*, Emelia Akashah Patah Akhir

Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Perak, Malaysia

*E-mail: qurah012@gmail.com

ABSTRACT

Bone fracture segmentation in X-ray images is essential for computer-aided musculoskeletal diagnosis, yet segmentation performance is strongly influenced by model configuration and hyperparameter selection. This study develops a hyperparameter-optimised U-Net model for binary segmentation of bone fracture regions in musculoskeletal X-ray images. The model was trained and evaluated using the FracAtlas dataset, which contains 4,083 annotated X-ray images, supplemented with 327 fracture-specific images obtained from Roboflow. Images and masks were harmonised through grayscale conversion, resizing to 256 x 256 pixels, normalisation, binary mask preparation, and augmentation to improve generalisation. Key hyperparameters, including learning rate, batch size, optimiser, number of epochs, loss function, and augmentation strategy, were empirically tuned using validation performance. The final configuration used a batch size of 8, Adam optimiser, learning rate of 0.001, binary cross-entropy loss, and 200 training epochs. Performance was evaluated using accuracy, validation loss, Intersection over Union (IoU), precision and pixel accuracy. The optimised U-Net achieved a validation accuracy of 99.6%, IoU of 0.995, precision of 0.991 and pixel accuracy of 0.993, demonstrating competitive performance compared with existing musculoskeletal segmentation models reported in the literature, including CMSCNet, 3D U-Net, and modified U-Net variants. The findings show that careful hyperparameter tuning, combined with U-Net encoder-decoder reconstruction and skip connections, can improve fracture-region segmentation in X-ray images. Further validation using larger multi-centre datasets and systematic ablation analysis is recommended before clinical deployment.

***Keywords**: Deep learning, hyperparameter optimisation, medical image segmentation, musculoskeletal X-ray images, U-Net*

INTRODUCTION

Musculoskeletal problems include many ailments that impact bones, muscles, joints, and connective tissues. Conditions such as osteoarthritis, osteoporosis, fractures, and ligament injuries rank among the primary contributors to global disability, profoundly affecting the quality of life and healthcare systems [1]. The Global Burden of Disease Study indicates that musculoskeletal illnesses represent a significant share of disability-adjusted life years (DALYs), particularly among ageing populations and those with sedentary lifestyles [2]. Contemporary medical imaging techniques, including X-rays, computed tomography (CT), and magnetic resonance imaging (MRI), are essential for the diagnosis and management of many disorders [3]. They offer comprehensive anatomical and pathological visualisations that allow doctors to evaluate structural anomalies, strategies for surgical interventions, and track disease advancement. The increasing volume of imaging data presents hurdles for manual interpretation, requiring the advancement of sophisticated computational approaches for precise and efficient image processing.

Deep learning, a kind of machine learning, has become a revolutionary method in the study of medical imaging. Its capacity to autonomously extract characteristics from unprocessed data has transformed tasks, including classification, object detection, and segmentation [4]. Deep learning models have been utilised in musculoskeletal imaging to detect fractures, outline cartilage borders, and assess bone density, demonstrating their versatility and effectiveness [5]-[6]. Segmentation entails identifying and delineating distinct objects or regions within an image. Advanced models like Mask Region-Based Convolutional Neural Network (Mask R-CNN) and versions of U-Net have demonstrated considerable potential in tackling segmentation issues in medical imaging [7]. These models improve diagnostic accuracy and enable quantitative assessments, including bone mineral density and cartilage thickness, essential for tailored treatment strategies [8].

Despite their capabilities, deep learning-based instance segmentation algorithms encounter difficulties in medical imaging, especially in musculoskeletal applications. Anatomical structures display considerable variety in shape, size, and direction among patients, frequently exacerbated by overlapping areas and pathological modifications such as fractures or degenerative changes [9]. Imaging artefacts, noise, and variations in acquisition techniques intensify these challenges, necessitating resilient and flexible models [8]. The efficacy of deep learning models is significantly influenced by hyperparameter configurations, such as learning rate, batch size, layer count, and optimiser selection. These hyperparameters affect the model's convergence dynamics, generalisation performance, and computational efficiency [10]. Inadequate hyperparameter settings may result in overfitting, underfitting, or ineffective learning, regardless of the dataset's size or the architecture's complexity [11]. Methods including grid search, random search, and Bayesian optimisation have been investigated to determine optimal hyperparameter combinations [10]. However, their utilisation in medical imaging, specifically segmentation tasks, is still constrained [12].

Despite the substantial progress of deep learning in medical image analysis, musculoskeletal X-ray fracture segmentation still requires a clearer investigation of how hyperparameter selection affects segmentation performance [13]. Existing studies frequently emphasise architectural changes, whereas parameters such as learning rate, batch size, optimiser, number of epochs, loss function, image size, and augmentation strategy can strongly affect convergence, generalisation, and segmentation reliability [14]. Therefore, this study focuses specifically on U-Net-based binary segmentation of bone fracture regions in X-ray images and evaluates the effect of an empirically selected hyperparameter configuration on segmentation performance.

The specific objectives of this study are to: (1) develop a U-Net-based deep learning model for binary segmentation of bone fracture regions in X-ray images; (2) preprocess and integrate FracAtlas and Roboflow fracture images using consistent resizing, normalisation, and binary mask preparation; (3) empirically optimise key hyperparameters, including learning rate, batch size, optimiser, number of epochs, loss function, and augmentation strategies; (4) evaluate the optimised model using accuracy, pixel accuracy, precision, and Intersection over Union (IoU); and (5) compare the proposed model with existing musculoskeletal image-segmentation approaches reported in the literature.

RELATED WORKS

Recent advancements in deep learning have significantly transformed medical imaging analysis, particularly in fracture detection, anatomical segmentation, and workflow optimisation. This section synthesises key contributions across these domains, highlighting the role of DL in enhancing diagnostic accuracy and clinical efficiency.

Deep learning is a powerful tool for musculoskeletal imaging, especially in bone fracture detection and classification. Studies show its effectiveness in automating diagnostics and aiding clinical decisions. Anttila et al. [15] demonstrated a deep learning tool's high accuracy for detecting enchondromas in hand radiographs, with an ROC curve of 0.95 and an F1-score of 69.5%. Yang et al. [16] developed a system for classifying rib fractures in CT scans, surpassing radiologists with an 87.63% accuracy in distinguishing fracture types and 95.22% in identifying misalignments, highlighting its diagnostic efficiency and workload reduction potential. Ensemble learning has advanced fracture detection accuracy. Tahir et al. [17] proposed a model using multiple architectures like MobileNetV2, VGG16, InceptionV3, and ResNet50 for X-ray-based detection, achieving 92.96% accuracy and a 92.14% F1-score, highlighting the benefit of combining networks for robustness. Similarly, Tanzi et al. [18] introduced a hierarchical DL classification system for proximal femur fractures using a cascade of CNNs aligned with the

AO/OTA classification, increasing diagnostic accuracy by 14% with human expertise, strengthening the importance of CAD systems in practice.

Segmentation has improved with DL innovations. Shan et al. [12] developed a neural network for segmenting knee joint tissues from 3D MRI, outperforming 3D U-Net and 3D V-Net, with a Dice score of 0.939 and lower computational costs. Similarly, Thian et al. [3] introduced CMSCNet for musculoskeletal ultrasound image segmentation, achieving 98.7% accuracy with fewer parameters, enhancing computational efficiency. DL-based fracture detection in paediatrics has been advanced by Ju and Cai [19] with the YOLOv8 framework for wrist trauma in X-rays, achieving an mAP@50 of 0.638 on GRAZPEDWRI-DX. Thian et al. [3] employed an Inception-ResNet Faster R-CNN, showing sensitivity of 98.1% and specificity of 72.9% for wrist radiographs. For rib fractures, Wu et al.'s [20] multicenter CT-based algorithm surpassed radiologists in sensitivity (87.9% vs. 79.7%) with an FROC score of 84.3%. Kalmet et al. [21] confirmed that DL systems reduced diagnosis time by 65.3 seconds when combined with radiologist analysis.

DL has also facilitated vertebral compression fracture detection, Paik et al. [22] employed Mask R-CNN for vertebral compression fracture detection, achieving a mean average precision of 0.58 and high specificity (85.3%) Alzubaidi et al. [23] introduced a trustworthy DL framework for shoulder X-rays, leveraging same-domain transfer learning and feature fusion to attain 99.2% accuracy, surpassing the diagnostic performance of orthopaedic surgeons (79.1%). The integration of DL in radiation therapy and surgical planning has further enhanced the automated segmentation of critical anatomical structures. Wang et al. [24] evaluated a CNN-based workflow for cervical cancer external beam radiation therapy (EBRT), achieving Dice scores of 0.88–0.93 for organs at risk (OARs) while identifying challenges in clinical target volume (CTV) delineation. Vania and Lee [25] proposed a multistage optimisation mask region-based convolutional neural network (MOM-RCNN) for intervertebral disc segmentation, attaining a Dice coefficient of 99% and a sensitivity of 88%. Additionally, Nikan et al. [26] introduced PWD-3DNet for temporal bone CT segmentation, achieving an average Dice score of 86% and demonstrating robustness across low-resolution datasets through augmented training.

Automatic segmentation has enhanced musculoskeletal imaging workflows. Kim et al. [27] utilised DeepLab v3+ for CT image segmentation of fractured tibia and fibula, achieving 98.92% accuracy and reducing processing time. Anttila et al. [28] created a robust DL model for distal radius fractures with an AUC of 0.97. FracNet, introduced by Wei et al. [29], had a 92.9% sensitivity and a Dice coefficient of 71.5% for rib fractures, surpassing human experts. Nguyen et al. [30] developed the classification–integrated gradients explanation (FEC-IGE) framework utilising transfer learning for fracture classification, achieving up to 98.48% accuracy across ten fracture types.

Beyond fracture detection, DL has facilitated the segmentation and classification of musculoskeletal structures. Kim et al. [31] developed a DL model for vertebral segmentation and compression ratio measurement, achieving a Dice similarity coefficient of 0.929 and a strong correlation with manual measurements. Shen et al. [32] explored CNNs and ResNet architectures for classifying cortical and trabecular bone mechanical states from SR-microCT images, achieving over 98% accuracy for cortical bone and 99% for trabecular bone.DL has also contributed to automated muscle and adipose tissue segmentation. Aringhieri et al. [33] applied U-Net variants for segmenting thigh muscles and subcutaneous adipose tissue in muscular dystrophy patients, achieving Dice scores of up to 96.8%. Lin et al. [34] introduced modified U-Net architectures (FFU and AFFU) for muscle segmentation in post-menopausal women, with AFFU achieving the highest Dice scores (0.862 ± 0.048), demonstrating the effectiveness of novel data augmentation techniques.

The application of DL in 3D-printed bone fracture modelling has further reinforced its clinical utility. Bittner-Frank et al. [35] investigated the accuracy of 3D-printed bone fracture models, highlighting the impact of CT imaging and segmentation algorithms on dimensional accuracy. The study confirmed that clinical routine scanners and protocols could produce models with sub-voxel resolution deviations, making them suitable for preoperative planning. These advancements underscore the growing role of DL in musculoskeletal imaging, improving diagnostic accuracy, segmentation efficiency, and clinical workflow automation. Ongoing research continues to refine these models, enhancing their generalisability and real-world applicability in healthcare settings.

In addition to model architecture, hyperparameter optimisation is a decisive factor in medical image segmentation [36]. Hyperparameters such as learning rate, batch size, optimiser, image size, epoch number, loss function, and augmentation strategy influence gradient stability, convergence speed, overfitting, and boundary localisation [37]. Recent optimisation studies have emphasised that dataset-constrained segmentation problems require careful tuning because small medical datasets and class imbalance can cause misleadingly high accuracy while reducing fracture-pixel reliability [38]. However, most musculoskeletal segmentation studies report final model performance without critically explaining how hyperparameter settings were selected or how they affected segmentation behaviour. This study addresses this limitation by explicitly reporting the candidate hyperparameter ranges, the final selected configuration, and the validation-based criteria used to optimise the U-Net model for X-ray fracture segmentation.

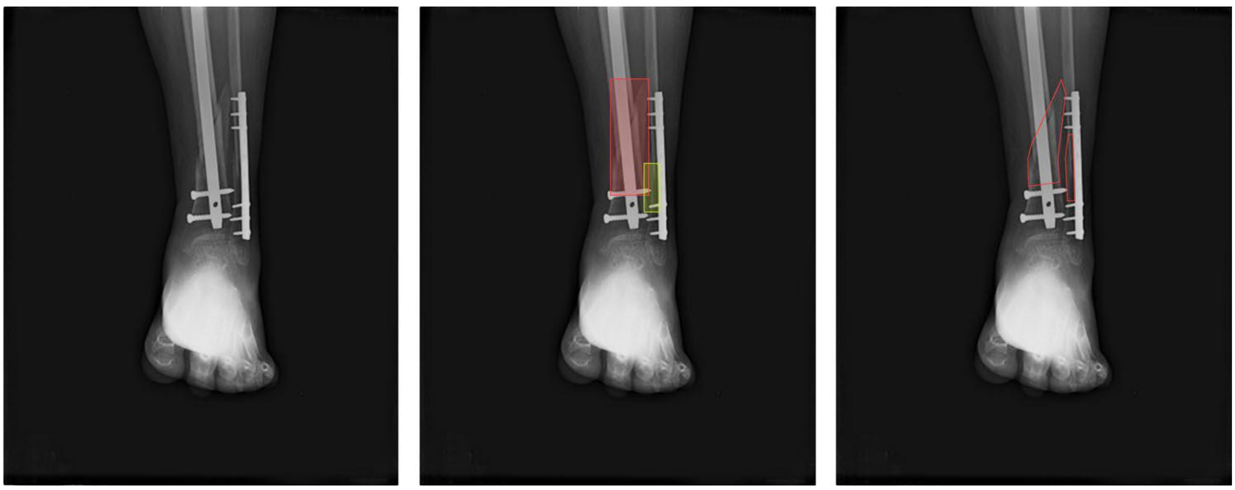
METHODOLOGY

This study presents a comprehensive framework for binary segmentation of bone fracture regions in X-ray images using a U-Net-based deep learning model. The methodology follows a structured workflow that includes data acquisition from FracAtlas and Roboflow, annotation harmonisation, preprocessing of X-ray images and masks, data augmentation, training-validation-test splitting, model training, validation-based hyperparameter tuning, final testing, and performance evaluation. The proposed methodology follows a structured workflow.

Dataset Description

The FracAtlas dataset is publicly available on Figshare. It comprises 4,083 X-ray images collected from three major hospitals in Bangladesh, manually annotated for bone fracture classification, localisation, and segmentation. Annotations were performed by two expert radiologists and an orthopedist using the open-source platform makesense.ai. The dataset includes 717 images containing 922 fracture instances, each annotated with masks and bounding boxes, alongside global classification labels [39]. FracAtlas serves as a valuable resource for advancing machine learning research in bone fracture diagnosis, as illustrated in Figure 1.

In addition to FracAtlas, the dataset used in this study includes 327 additional X-ray fracture images sourced from Roboflow to increase the diversity of fracture appearances. To ensure consistency between the FracAtlas and Roboflow datasets, all images were



(a) (b) (c)

Figure 1 Fully annotated image with global tags, localised fracture regions marked by boxes, and fractured areas highlighted with red borders for segmentation tasks [39] converted to a unified grayscale format and resized to 256 × 256 pixels. Annotation labels were harmonised into a binary segmentation format, where fracture pixels were assigned a value of 1 and background pixels were assigned a value of 0. Where necessary, polygon or mask annotations were converted into binary masks compatible with the U-Net training pipeline. The same preprocessing steps, including normalisation, resizing, and augmentation, were applied to both datasets to reduce distributional differences and maintain labelling consistency.

Dataset Preprocessing

Data preprocessing standardised image size, channel format, intensity range, and binary mask structure. All X-ray images were resized to 256 × 256 pixels, converted to grayscale, and normalised using min-max normalisation. The corresponding masks were rescaled to values between 0 and 1 and checked to ensure that the positive class consistently represented fracture pixels. The complete dataset was divided into training, validation, and testing subsets using a 70:15:15 split. Specifically, 70% of the images were used for model training, 15% were used for validation during hyperparameter tuning, and the remaining 15% were reserved for final testing on unseen images.

Experimental Setup

The study used Google Colab Notebooks as the development environment. Python was used for model development, training, and optimisation. Data preprocessing and analysis were conducted using pandas, NumPy, Keras, and OpenCV for image resizing, augmentation, normalisation, and mask preparation. TensorFlow was used for model development, while Matplotlib was used to visualise training curves, validation behaviour, and segmentation outputs.

Model Development

The proposed methodology begins with data preparation, where raw X-ray images and corresponding annotations are organised into a consistent segmentation format. The dataset was split into 70% training, 15% validation, and 15% testing subsets. The training set was used to update model weights; the validation set was used to guide hyperparameter selection and monitor overfitting, and the test set was reserved for final evaluation on unseen images. The preprocessing stage included resizing, grayscale conversion, normalisation, binary mask preparation, and augmentation techniques such as horizontal flipping, rotation, brightness/contrast adjustment, Gaussian blur, and elastic transformation.

Following this, the model compiling and fitting phase configures the deep learning architecture with the appropriate loss functions, metrics, and optimisers. During training, the model learns from the training data. At the same time, hyperparameters such as learning rate, batch size, and the number of epochs are optimised. If training fails, the pipeline loops back to further refine the hyperparameters and model architecture. Once the training is successful, the model progresses to the testing phase, where its performance is evaluated on the test dataset containing unseen images.

Finally, the performance evaluation phase assesses the model using metrics such as accuracy, IoU, and computational efficiency to ensure reliable segmentation results. This iterative and structured approach ensures the model achieves high segmentation accuracy and robustness, making it suitable for clinical applications in bone fracture diagnosis. Figure 2 shows the proposed methodology.

Proposed U-Net Architecture for Bone Fracture Segmentation

In this study, the proposed architecture, illustrated in Figure 3, is a U-Net-based model designed for binary segmentation of bone fracture regions in X-ray images. The encoder path extracts hierarchical features from the input image using convolutional blocks and downsampling operations, while the bottleneck captures the deepest semantic representation. The decoder path reconstructs the segmentation mask through upsampling and convolutional refinement. Skip connections transfer spatial details from encoder layers to their corresponding decoder layers, which helps preserve fine fracture-boundary information and improves localisation of small fracture regions. The final sigmoid layer generates a binary mask in which fracture pixels are separated from the background.

Model Parameter Selection

This study optimised critical hyperparameters to enhance convergence stability, segmentation accuracy, and generalisation performance. Candidate ranges were selected based on common practices in medical image segmentation and then refined empirically using

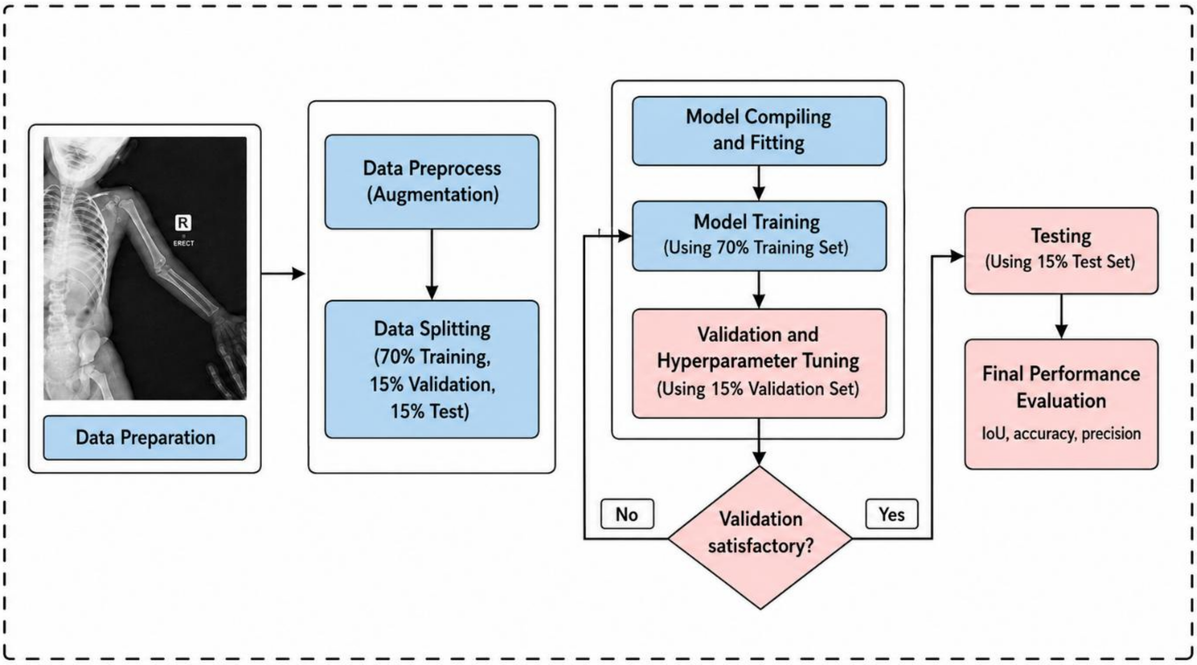


Figure 2 Workflow of dataset preparation, preprocessing, data augmentation, training-validation-test split, validation-based hyperparameter tuning, final testing, and performance evaluation

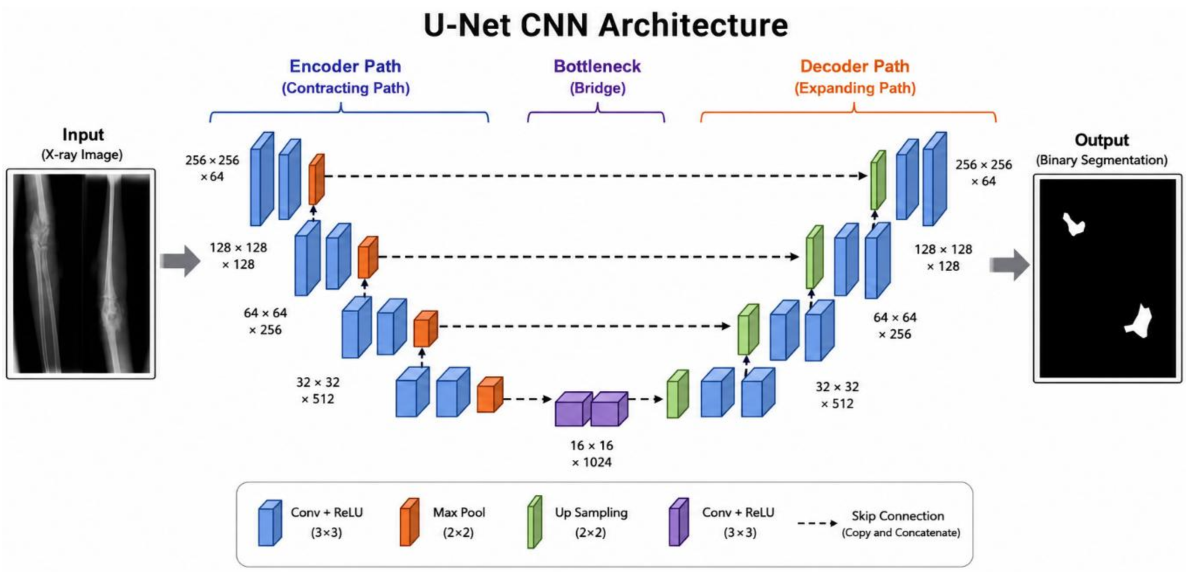


Figure 3 Proposed U-Net architecture for bone fracture segmentation, showing the encoder path, bottleneck, decoder path, skip connections, and final binary segmentation output

validation-set performance. The final configuration was selected based on validation loss, validation accuracy, IoU, and visual stability of predicted masks.

The candidate learning rates included 0.0001, 0.0005, 0.001, and 0.01, while batch sizes of 8, 16, and 32 were considered. The Adam optimiser was selected because it provided stable convergence during training. The final learning rate of 0.001 and batch size of 8 were selected because they produced the most stable validation performance without clear signs of overfitting. Although scheduler-based learning-rate reduction strategies were reviewed as possible optimisation options, the final implemented model used a fixed learning rate of 0.001 rather than a step-size/gamma scheduler. Relevant literature was reviewed to identify commonly recommended hyperparameter ranges, as summarised below:

1. Batch Size: Optimal batch sizes, ranging from 16 to 64, balance gradient estimation noise and memory efficiency. Smaller sizes improve generalisation but require more extended training, while larger sizes converge faster but may risk overfitting [40]-[41].
2. Learning Rate: A learning rate between 0.0001 and 0.01 is recommended for balancing convergence speed and stability. Too high rates can cause divergence, while too low rates lead to slow training and potential local minima [42].
3. Number of Epochs: Training for 50 to 200 epochs is suggested to ensure adequate learning without overfitting. This range provides sufficient exposure to the data while minimising overfitting risks [43].
4. Image Size: For high-resolution remote sensing data, image sizes between 224×224 and 256×256 pixels are effective, balancing detail and computational cost [44].
5. Optimiser: The Adam optimiser is preferred for its adaptive learning rate capabilities, which enhance convergence stability in noisy and sparse gradient scenarios [45].
6. Learning-rate schedule: Scheduler-based step size and gamma strategies were considered during parameter review; however, the final model used a fixed learning rate of 0.001 with the Adam optimiser because this setting achieved stable validation loss and segmentation performance.

Table 1 Model hyperparameter settings

Model Parameters	Value
Model Architecture	U-Net
Input Image Size	256×256
Input Channels	1 (Grayscale)
Batch Size	8 (selected empirically from 8, 16, and 32)
Epochs	200
Optimiser	Adam
Loss Function	Binary Cross-Entropy
Activation Function	Sigmoid (Output Layer)
Image Normalization	Yes (Min-Max Normalisation)
Mask Normalization	Yes (Rescaled to 0-1)
Validation Data	Used
Learning Rate	0.001 (selected empirically from 0.0001, 0.0005, 0.001, and 0.01)
Augmentation	Horizontal flip, random brightness/ contrast, rotation, Gaussian blur, elastic transformation

7. Loss Function: Binary cross-entropy was used as the final loss function for binary fracture-mask segmentation. Dice-based loss was considered during the design stage because it is useful for class-imbalanced segmentation, but the reported final model configuration used binary cross-entropy to maintain consistency with the implementation summarised in Table 1.
8. Augmentation: Data augmentation techniques, including rotations, flips, and scaling, enhance model robustness by simulating real-world conditions and improving generalisation [46].

EVALUATION METRICS

This study evaluates model performance using the following metrics:

1. Accuracy: Measures the proportion of correct predictions but may be unreliable with imbalanced datasets.
2. IoU: Quantifies the overlap between predicted segmentation and ground truth, serving as a critical metric for segmentation accuracy.
3. Precision (Positive Predictive Value): Measures how many of the predicted fracture pixels are actually fractures.
4. Pixel Accuracy: Measures the proportion of correctly classified pixels.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$\text{Pixel Accuracy} = \frac{TP+TN}{\text{Total Pixels}} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{IoU} = \frac{TP}{TP+FP+FN} \tag{4}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative pixels, respectively.

RESULTS AND DISCUSSION

This section assesses the performance of the proposed U-Net model for bone fracture segmentation in X-ray images. The dataset was divided into 70% training, 15% validation, and 15% testing subsets, and the validation subset was used to guide hyperparameter selection. The final model was evaluated on unseen test images using accuracy, validation loss, IoU, pixel accuracy, and precision.

Table 2 Evaluation and performance of the proposed model over the top 10 epochs

Epoch	Loss	Val Loss	Accuracy	Val Accuracy
10	0.0214	0.0194	0.9945	0.9951
29	0.0136	0.0211	0.9948	0.9951
52	0.0115	0.0190	0.9950	0.9955
76	0.0102	0.0185	0.9952	0.9956
88	0.0095	0.0182	0.9954	0.9958
105	0.0092	0.0179	0.9955	0.9960
123	0.0088	0.0175	0.9956	0.9962
140	0.0085	0.0172	0.9957	0.9964
165	0.0081	0.0170	0.9958	0.9965
190	0.0079	0.0168	0.9960	0.9967

The analysis in Figure 4 demonstrated that the proposed model achieved high performance across critical evaluation metrics, including accuracy, validation loss, IoU, pixel accuracy, and precision. The model exhibited consistently high accuracy across the top 10 epochs, with values ranging from 0.9951 in the earlier epochs to 0.9967 at Epoch 190, highlighting its strong generalisation capability. The steady decrease in validation loss, which ultimately reached 0.0168, suggests that the model effectively minimised errors while maintaining a balance between learning and overfitting. Furthermore, the model's performance progressively improved throughout 200 training epochs, as evidenced by the increasing validation accuracy and the declining validation loss. This trend reinforces the model's capacity to generalise well to unseen data, ensuring reliability in practical applications.

The IoU value of 0.995 further emphasises the model's ability to delineate fracture regions within X-ray images. The value is reported as a ratio rather than a percentage. The pixel accuracy of 0.993 indicates that most pixels were correctly classified, while the precision value of 0.991 shows that the model generated very few false-positive fracture predictions. These findings demonstrate the robustness of the model in segmenting fracture regions and suggest that the U-Net architecture can support computer-aided fracture assessment. A full ablation study will be required in future work to isolate the separate effects of learning rate, batch size, optimiser, loss function, augmentation strategy, and number of epochs.

Performance Analysis of the Proposed Model

The proposed deep learning-based model for bone fracture segmentation in X-ray images was rigorously evaluated using multiple performance metrics, including accuracy, precision, validation loss, and IoU. Table 2 provides a detailed assessment of the model's performance across various epochs, demonstrating its capability to tackle complex segmentation tasks.

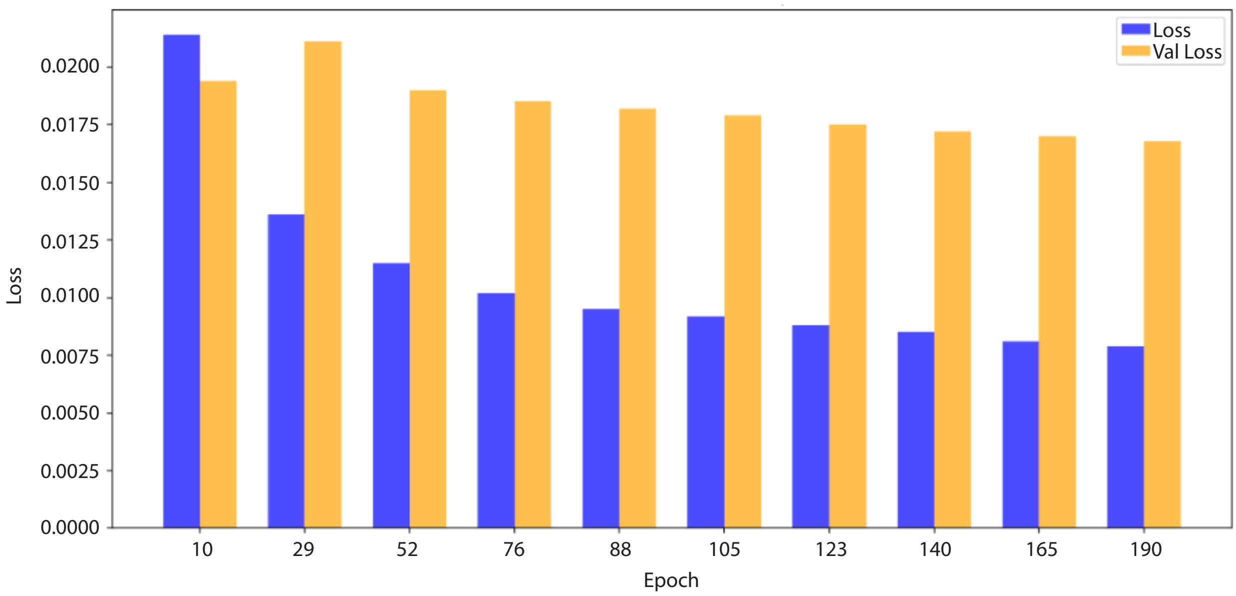


Figure 4 Performance of the proposed model across selected training epochs

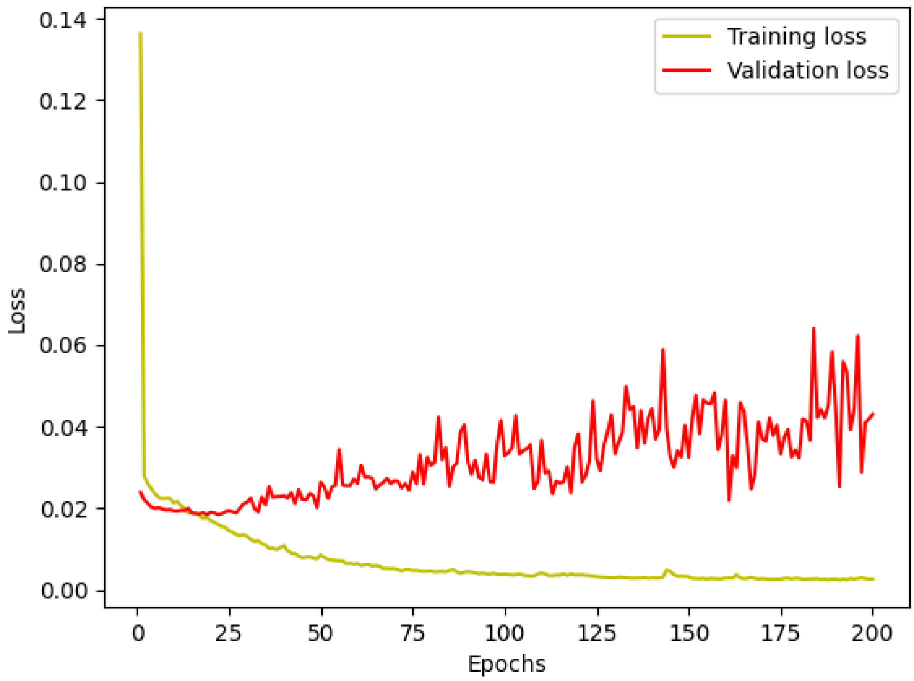


Figure 2 Plot of the training and validation loss of the proposed model

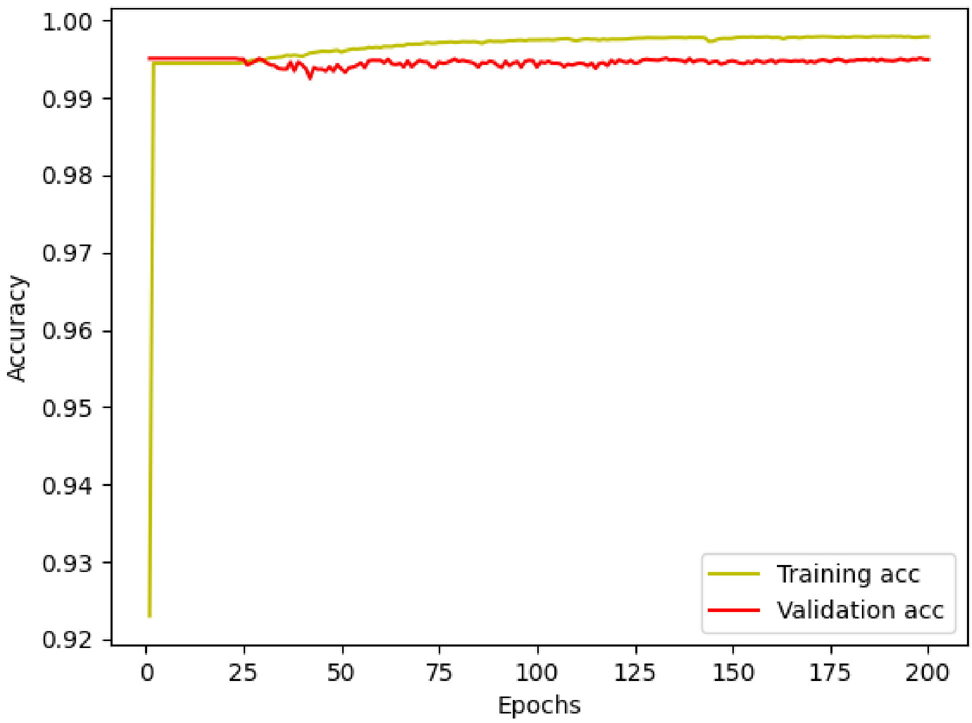


Figure 3 Plot of the training and validation accuracy of the proposed deep learning model

The training process exhibited a steady improvement in both accuracy and validation accuracy, as illustrated in Figure 6. Throughout 200 epochs, accuracy improved from 0.9945 at epoch 10 to 0.9960 at epoch 190, while validation accuracy reached 0.9967. Concurrently, the loss function in Figure 5 shows a consistent decline, with the training loss reducing from 0.0214 to 0.0079 and the validation loss decreasing from 0.0194 to 0.0168. This trend indicates effective model generalisation while minimising overfitting on the validation set.

Table 3 Visual prediction of the input, ground truth and predicted outputs of the model

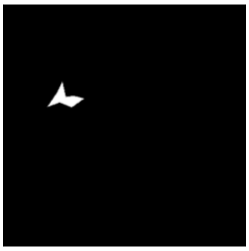
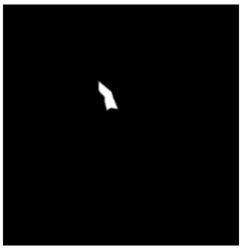
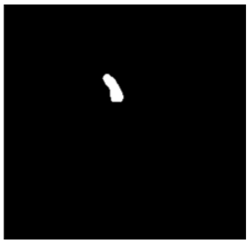
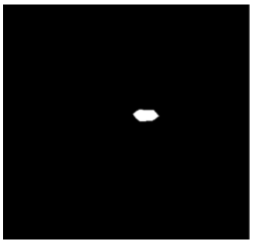
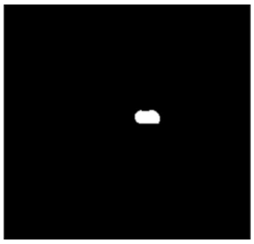
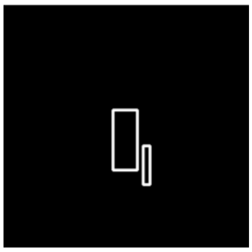
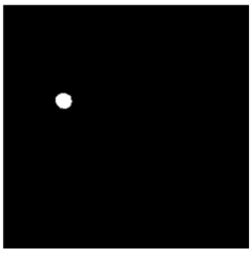
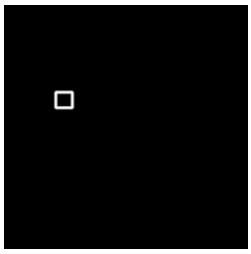
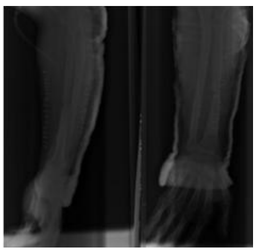
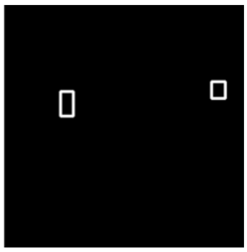
Original Image	Ground Truth	Model Prediction
		
		
		

Table 4 Visualisation of the prediction with boneview-output, with white bounding boxes for fracture regions

Original Image	Ground Truth	Model Prediction
		
		
		

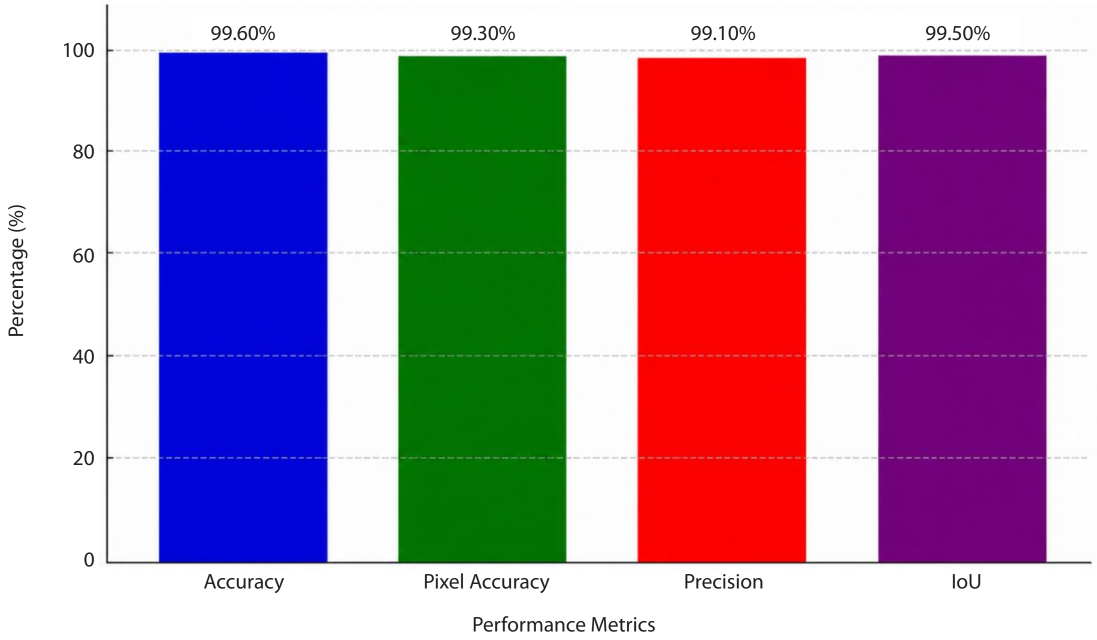


Figure 7 Summary of proposed model performance metrics after hyperparameter-optimised training

IoU values remained consistently high throughout the evaluation, underscoring the model's ability to delineate fracture regions. The highest IoU score achieved was 0.995, reflecting strong overlap between predicted and ground-truth masks. Pixel accuracy reached 0.993, while precision reached 0.991, indicating that the model produced very few false-positive fracture predictions. Because fracture pixels occupy a relatively small proportion of the full X-ray image, accuracy and pixel accuracy were interpreted together with IoU and precision rather than as standalone evidence of clinical reliability. The overall performance metrics after hyperparameter-optimised training are shown in Figure 7.

A qualitative evaluation was performed by comparing the model's segmentation outputs against ground-truth annotations. Table 3 and Table 4 illustrate examples of input X-ray images, corresponding ground-truth masks, and model-predicted segmentations. The visual comparison shows that the model can localise fracture regions and generate segmentation masks that are broadly aligned with expert annotations.

The high accuracy and segmentation quality of the proposed model present significant clinical implications. Automated fracture detection and segmentation can assist radiologists in expediting diagnosis, reducing workload, and improving patient outcomes. The model's integration into computer-aided diagnosis (CAD) systems can facilitate real-time fracture assessment, potentially reducing diagnostic errors and enabling personalised treatment planning. Additionally, its scalability across different fracture types highlights its potential as a robust tool in orthopaedic imaging.

Comparative Analysis with Existing Models

To validate the effectiveness of the proposed model, a comparative analysis was conducted against existing deep learning architectures used in musculoskeletal image segmentation. Table 5 presents studies using efficient neural networks, CMSCNet, 3D U-Net, 2D U-Net, improved Mask R-CNN, and modified U-Net variants. Because these studies use different datasets, anatomical regions, and imaging modalities, the comparison is interpreted as a literature-based performance benchmark rather than a direct head-to-head experimental comparison.

Hyperparameter optimisation contributed to the stability of the final model configuration. The study empirically evaluated candidate settings for batch size, learning rate, optimiser, number of epochs, and augmentation strategy. The final configuration, including batch size 8, learning rate 0.001, Adam optimiser, binary cross-entropy loss, and augmentation, was selected because it produced stable validation loss

Table 5 Proposed model comparison with previous study

Techniques Used	Dataset	Findings	Reference
Efficient neural network with depthwise separable convolution	3D MRI volumes of knee joints	Dice score: 0.939; outperformed 3D U-Net and 3D V-Net	[12]
CMSCNet (U-Net with dense atrous convolution)	Gastrocnemius dataset	Accuracy: 98.7%, IoU: 0.7227	[3]
3D V-Net with a compound loss function	397 QCT subjects	Dice score: 0.9815, Sensitivity: 0.9852, Specificity: 0.9992	[47]
3D U-Net for maxillary and mandibular segmentation	82 CT scans of head and neck cancer patients	Dice similarity coefficient: 0.81–0.91, HD95: 1.61–4.22	[48]
3D U-Net for TMJ segmentation	154 manually segmented CBCT images	IoU: 0.955 (condyles), 0.935 (glenoid fossa); Segmentation time: 3.6 s	[49]
2D U-Net for facial bone segmentation	CT data from 50 patients	Dice coefficient: 0.897±0.077, ASSD: 1.168±1.962 mm	[50]
Improved-Mask R-CNN (iMaskRCNN) for bone and cartilage segmentation	500 knee MRI scans, 97 hip MRI scans	Dice scores: 95–98% (femur), 71–80% (femoral cartilage)	[51]
U-Net variants for thigh muscle segmentation	MRI scans of muscular dystrophy patients	Dice scores: up to 96.8%	[33]
Modified U-Net (FFU, AFFU) for muscle segmentation	MRI scans of postmenopausal women	Dice scores: 0.862 ± 0.048	[34]
Hyperparameter-optimised U-Net	FracAtlas: 4,083 manually annotated musculoskeletal X-ray images, supplemented with 327 fracture-specific Roboflow images, harmonised into binary fracture masks	Precision = 0.991; IoU = 0.995; Validation accuracy = 99.6% Pixel accuracy = 0.993	Proposed Model

and high segmentation overlap. However, since a full ablation study was not conducted, the contribution of each hyperparameter should be examined more systematically in future work.

CONCLUSION

This study developed a hyperparameter-optimised U-Net model for binary segmentation of bone fracture regions in musculoskeletal X-ray images, trained on FracAtlas images supplemented with Roboflow fracture images with consistent preprocessing, normalisation, augmentation, and binary mask preparation. Using a 256 × 256 grayscale input size, batch size of 8, Adam optimiser, learning rate of 0.001, binary cross-entropy loss, and 200 training epochs, the model achieved strong performance: validation accuracy of 99.6%, precision of 0.991, IoU of 0.995, and pixel accuracy of 0.993, reflecting the strength of the U-Net's encoder-decoder architecture and skip connections in localising subtle fracture boundaries with few false positives. Compared with existing musculoskeletal segmentation models, the proposed model showed competitive performance, suggesting potential for computer-aided fracture assessment. However, boundary-level accuracy in complex cases requires further evaluation using boundary-aware metrics and edge-refinement strategies, and the dataset's limited diversity across institutions, scanners, and imaging protocols warrants validation on larger multi-centre datasets, alongside ablation analysis, attention-based U-Net variants, class-balancing strategies, and post-processing to improve clinical robustness.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Petroleum Technology Development Fund (PTDF) for its financial support. The authors also thank Universiti Teknologi PETRONAS, Department of Computer and Information Sciences, for providing the necessary research support and facilities.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article.

ETHICAL CLEARANCE

Ethical clearance was not required for this study because it used publicly available, de-identified secondary datasets, namely FracAtlas and Roboflow, and did not involve direct interaction with human participants, collection of identifiable personal information, or intervention involving patients. The datasets were used in accordance with their respective access and usage conditions.

AUTHORS' CONTRIBUTION

Ibrahim Muhammad Kurah conceptualised the study, developed the methodology, performed data preparation, model implementation, hyperparameter optimisation, analysis, visualisation, and drafted the original manuscript. Emelia Akashah Patah Akhir supervised the study, contributed to the research methodology, reviewed the results, and critically revised the manuscript for intellectual content. Both authors read and approved the final version of the manuscript.

REFERENCES

[1] A. M. Briggs, J. E. Jordan, D. Kopansky-Giles, S. Sharma, L. March, C. H. Schneider, S. Mishrra, J. J. Young, and H. Slater, "The need for adaptable global guidance in health systems strengthening for musculoskeletal health: a qualitative study of international key informants," *Global Health Res. Policy*, vol. 6, no. 1, p. 24, 2021, doi: 10.1186/s41256-021-00201-7.

[2] S. Safiri, A. A. Kolahi, M. Cross, C. Hill, E. Smith, K. Carson-Chahhoud, M. A. Mansournia, A. Almasi-Hashiani, A. Ashrafi-Asgarabad, J. Kaufman, M. Sepidarkish, S. K. Shakouri, D. Hoy, A. D. Woolf, L. March, G. Collins, and R. Buchbinder, "Prevalence, deaths, and disability-adjusted life years due to musculoskeletal disorders for 195 countries and territories 1990–2017," *Arthritis Rheumatol.*, vol. 73, no. 4, pp. 702–714, 2021, doi: 10.1002/art.41571.

[3] Y. L. Thian, Y. Li, P. Jagmohan, D. Sia, V. E. Yao Chan, and R. T. Tan, "Convolutional neural networks for automated fracture detection and localization on wrist radiographs," *Radiol. Artif. Intell.*, vol. 1, no. 1, 2019, doi: 10.1148/ryai.2019180001.

[4] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A. L. Boulesteix, D. Deng, and M. Lindauer, "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges," *WIREs Data Min. Knowl. Discov.*, vol. 13, no. 2, 2023, doi: 10.1002/widm.1484.

[5] Y. Chen, Y. Sun, W. Yu, Y. Liu, and H. Hu, "A novel lightweight bilateral segmentation network for detecting oil spills on the sea surface," *Mar. Pollut. Bull.*, vol. 175, p. 113343, Jan. 2022, doi: 10.1016/j.marpolbul.2022.113343.

[6] R. Y. L. Kuo, C. Harrison, T. A. Curran, B. Jones, A. Freethy, D. Cussons, M. Stewart, G. S. Collins, and D. Furniss, "Artificial intelligence in fracture detection: A systematic review and meta-analysis," *Radiology*, vol. 304, no. 1, pp. 50–62, 2022, doi: 10.1148/radiol.211785.

[7] Z. Zhou, G. Zhao, R. Kijowski, and F. Liu, "Deep convolutional neural network for segmentation of knee joint anatomy," *Magn. Reson. Med.*, vol. 80, no. 6, pp. 2759–2770, 2018, doi: 10.1002/mrm.27229.

[8] Y. Zhang, S. Chan, V. Y. Park, K. T. Chang, S. Mehta, M. J. Kim, F. J. Combs, P. Chang, D. Chow, R. Parajuli, R. S. Mehta, C. Y. Lin, S. H. Chien, J. H. Chen, and M. Y. Su, "Automatic detection and segmentation of breast cancer on MRI using Mask R-CNN trained on non–fat-sat images and tested on fat-sat images," *Acad. Radiol.*, vol. 29, pp. S135–S144, 2022, doi: 10.1016/j.acra.2020.12.001.

[9] L. Bonaldi, A. Pretto, C. Pirri, F. Uccheddu, C. G. Fontanella, and C. Stecco, "Deep learning-based medical images segmentation of musculoskeletal anatomical structures: A survey of bottlenecks and strategies," *Bioengineering*, vol. 10, no. 2, 2023, doi: 10.3390/bioengineering10020137.

[10] C. Boyd, G. C. Brown, T. J. Kleinig, W. Mayer, J. Dawson, M. Jenkinson, and E. Bezak, "Hyperparameter selection for dataset-constrained semantic segmentation: Practical machine learning optimization," *J. Appl. Clin. Med. Phys.*, vol. 25, no. 12, 2024, doi: 10.1002/acm2.14542.

[11] C. Xu, P. Coen-Pirani, and X. Jiang, "Empirical study of overfitting in deep learning for predicting breast cancer metastasis," *Cancers*, vol. 15, no. 7, p. 1969, 2023, doi: 10.3390/cancers15071969.

[12] Q.Y.P. Shan, L. Bing, S. Li Goh, K. Hasikin, X. Wu, Y. Chai Hum, Y. Kai Tee, and K. Wee Lai, "An efficient neural network for segmenting multiple joint tissues from knee MRI with hyperparameter optimization: Data from the Osteoarthritis Initiative," *IEEE Access*, vol. 12, pp. 123757–123770, 2024, doi: 10.1109/ACCESS.2024.3454374.

[13] A. Matsushima, T. B. Chen, K. Kimura, M. Sato, S. Y. Hsu, and T. Okamoto, "Proximal femur segmentation and quantification in dual-energy subtraction tomosynthesis: A novel approach to fracture risk assessment," *J. X-Ray Sci. Technol.*, vol. 33, no. 2, pp. 405–419, 2025, doi: 10.1177/08953996241312594.

[14] P. N. Srinivasu, G. L. A. Kumari, S. C. Narahari, S. Ahmed, and A. Alhumam, "Exploring the impact of hyperparameter and data augmentation in YOLO V10 for accurate bone fracture detection from X-ray images," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-93505-4.

[15] T. T. Anttila, S. Aspinen, G. Pierides, V. Haapamäki, M. K. Laitinen, and J. Ryhänen, "Enchondroma detection from hand radiographs with an interactive deep learning segmentation tool—A feasibility study," *J. Clin. Med.*, vol. 12, no. 22, 2023, doi: 10.3390/jcm12227129.

[16] C. Yang, J. Wang, J. Xu, C. Huang, F. Liu, W. Sun, R. Hong, L. Zhang, D. Ma, Z. Li, X. Zhang, J. Cai, and Z. Fu, "Development and assessment of deep learning system for the location and classification of rib fractures via computed tomography," *Eur. J. Radiol.*, vol. 154, p. 110434, May 2022, doi: 10.1016/j.ejrad.2022.110434.

[17] 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," *Clin. Radiol.*, vol. 79, no. 11, pp. e1394–e1402, 2024, doi: 10.1016/j.crad.2024.08.006.

[18] L. Tanzi, E. Vezzetti, R. Moreno, A. Aprato, A. Audisio, and A. Massè, "Hierarchical fracture classification of proximal femur X-ray images using a multistage deep learning approach," *Eur. J. Radiol.*, vol. 133, Feb. 2020, doi: 10.1016/j.ejrad.2020.109373.

[19] R. Y. Ju and W. Cai, "Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-47460-7.

[20] M. Wu, Z. Chai, G. Qian, H. Lin, Q. Wang, L. Wang, and H. Chen, "Development and evaluation of a deep learning algorithm for rib segmentation and fracture detection from multicenter chest CT images," *Radiol. Artif. Intell.*, vol. 3, no. 5, 2021, doi: 10.1148/ryai.2021200248.

[21] P. H. S. Kalmet, S. Sanduleanu, S. Primakov, G. Wu, A. Jochems, T. Refaee, A. Ibrahim, L. v. Hulst, P. Lambin, and M. Poeze, "Deep learning in fracture detection: A narrative review," *Acta Orthop.*, vol. 91, no. 2, pp. 215–220, 2020, doi: 10.1080/17453674.2019.1711323.

[22] S. Paik, J. Park, J. Y. Hong, and S. W. Han, "Deep learning application of vertebral compression fracture detection using Mask R-CNN," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-67017-6.

[23] L. Alzubaidi, A. Salhi, M. A. Fadhel, J. Bai, F. Hollman, K. Italia, R. Pareyon, A. S. Albahri, C. Ouyang, J. Santamaría, K. Cutbush, A. Gupta, A. Abbosh, and Y. Gu, "Trustworthy deep learning framework for the detection of abnormalities in X-ray shoulder images," *PLoS ONE*, vol. 19, no. 3, 2024, doi: 10.1371/journal.pone.0299545.

[24] J. Wang, Y. Chen, H. Xie, L. Luo, and Q. Tang, "Evaluation of auto-segmentation for EBRT planning structures using deep learning-based workflow on cervical cancer," *Sci. Rep.*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-18084-0.

[25] M. Vania and D. Lee, "Intervertebral disc instance segmentation using a multistage optimization mask-RCNN (MOM-RCNN)," *J. Comput. Des. Eng.*, vol. 8, no. 4, pp. 1023–1036, 2021, doi: 10.1093/jcde/qwab030.

[26] S. Nikan, K. Van Osch, M. Bartling, D. G. Allen, S. A. Rohani, B. Connors, S. K. Agrawal, and H. M. Ladak, "PWD-3DNet: A deep learning-based fully-automated segmentation of multiple structures on temporal bone CT scans," *IEEE Trans. Image Process.*, vol. 30, pp. 739–753, 2021, doi: 10.1109/TIP.2020.3038363.

[27] H. Kim, Y. D. Jeon, K. B. Park, H. Cha, M. S. Kim, J. You, S. W. Lee, S. H. Shin, Y. G. Chung, S. Bin Kang, W. S. Jang, and D. K. Yoon, "Automatic segmentation of inconstant fractured fragments for tibia/fibula from CT images using deep learning," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-47706-4.

[28] T. T. Anttila, T. V. Karjalainen, T. O. Mäkelä, E. M. Waris, N. C. Lindfors, M. M. Leminen, and J. O. Ryhänen, "Detecting distal radius fractures using a segmentation-based deep learning model," *J. Digit. Imaging*, vol. 36, no. 2, pp. 679–687, 2023, doi: 10.1007/s10278-022-00741-5.

[29] D. Wei, Q. Wu, X. Wang, M. Tian, and B. Li, "Accurate instance segmentation in pediatric elbow radiographs," *Sensors*, vol. 21, no. 23, p. 7966, 2021, doi: 10.3390/s21237966.

[30] T. M. Nguyen, T. Van Tran, and Q. T. Lu, "FEC-IGE: An efficient approach to classify fracture based on convolutional neural networks and integrated gradients explanation," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 6, 2024, doi: 10.14569/IJACSA.2024.01506142.

[31] D. H. Kim, J. G. Jeong, Y. J. Kim, K. G. Kim, and J. Y. Jeon, "Automated vertebral segmentation and measurement of vertebral compression ratio based on deep learning in X-ray images," *J. Digit. Imaging*, vol. 34, no. 4, pp. 853–861, 2021, doi: 10.1007/s10278-021-00471-0.

[32] S. C. Y. Shen, M. Peña Fernández, G. Tozzi, and M. J. Buehler, "Deep learning approach to assess damage mechanics of bone tissue," *J. Mech. Behav. Biomed. Mater.*, vol. 123, 2021, doi: 10.1016/j.jmbbm.2021.104761.

[33] G. Aringhieri, G. Astrea, D. Marfisi, S. C. Fanni, G. Marinella, R. Pasquariello, G. Ricci, F. Sansone, M. Sperti, A. Tonacci, F. Torri, S. Matà, G. Siciliano, E. Neri, F. M. Santorelli, and R. Conte, "Convolutional neural network-based automated segmentation of skeletal muscle and subcutaneous adipose tissue on thigh MRI in muscular dystrophy patients," *J. Funct. Morphol. Kinesiol.*, vol. 9, no. 3, p. 123, 2024, doi: 10.3390/jfmk9030123.

[34] Z. Lin, W. H. Henson, L. Dowling, J. Walsh, E. Dall'Ara, and L. Guo, "Automatic segmentation of skeletal muscles from MR images using modified U-Net and a novel data augmentation approach," *Front. Bioeng. Biotechnol.*, vol. 12, 2024, doi: 10.3389/fbioe.2024.1355735.

[35] M. Bittner-Frank, A. Strassl, E. Unger, L. Hirtler, B. Eckhart, M. Koenigshofer, A. Stoegner, A. Nia, D. Popp, F. Kainberger, R. Windhager, F. Moscato, and E. Benca, "Accuracy analysis of 3D bone fracture models: Effects of computed tomography (CT) imaging and image segmentation," *J. Imaging Inform. Med.*, vol. 37, no. 4, pp. 1889–1901, 2024, doi: 10.1007/s10278-024-00998-y.

[36] A. Mortazi, V. Cicek, E. Keles, and U. Bagci, "Selecting the best optimizers for deep learning–based medical image segmentation," *Front. Radiol.*, vol. 3, 2023, doi: 10.3389/fradi.2023.1175473.

[37] P. K. Mondol, M. A. Islam Mozumder, H. Cheol Kim, M. Hassan Ali Al-Onaizan, D. S. M. Hassan, M. Al-Bahri, and M. S. A. Muthanna, "A ResNet-50–UNet hybrid with whale optimization algorithm for accurate liver tumor segmentation," *Diagnostics*, vol. 15, no. 23, 2025, doi: 10.3390/diagnostics15232975.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, vol. 9351, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[39] I. Abedeen, M. A. Rahman, F. Z. Prottyasha, T. Ahmed, T. M. Chowdhury, and S. Shatabda, "FracAtlas: A dataset for fracture classification, localization and segmentation of musculoskeletal radiographs," *Sci. Data*, vol. 10, no. 1, pp. 1–8, 2023, doi: 10.1038/s41597-023-02432-4.

[40] L. Gao, Y. Zhang, J. Han, and J. Callan, "Scaling deep contrastive learning batch size under memory limited setup," in *Proc. 6th Workshop Represent. Learn. NLP (RepL4NLP)*, 2021, pp. 316–321, doi: 10.18653/v1/2021.repl4nlp-1.31.

[41] T. Takase, "Dynamic batch size tuning based on stopping criterion for neural network training," *Neurocomputing*, vol. 429, pp. 1–11, 2021, doi: 10.1016/j.neucom.2020.11.054.

[42] C. Yu, X. Qi, H. Ma, X. He, C. Wang, and Y. Zhao, "LLR: Learning learning rates by LSTM for training neural networks," *Neurocomputing*, vol. 394, pp. 41–50, 2020, doi: 10.1016/j.neucom.2020.01.106.

[43] C. Ying, B. Li, and B. Li, "Tempo: Improving training performance in cross-silo federated learning," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2022, pp. 4358–4362, doi: 10.1109/ICASSP43922.2022.9747880.

[44] L. Liu, Z. Zou, and Z. Shi, "Hyperspectral remote sensing image synthesis based on implicit neural spectral mixing models," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, doi: 10.1109/TGRS.2022.3232705.

[45] R. Abdulkadirov, P. Lyakhov, and N. Nagornov, "Survey of optimization algorithms in modern neural networks," *Mathematics*, vol. 11, no. 11, 2023, doi: 10.3390/math11112466.

[46] P. Sengupta, A. Mehta, and P. S. Rana, "Enhancing performance of deep learning models with a novel data augmentation approach," in *Proc. 14th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, 2023, doi: 10.1109/ICCCNT56998.2023.10308298.

[47] C. Zhao, J. H. Keyak, J. Tang, T. S. Kaneko, S. Khosla, S. Amin, E. J. Atkinson, L.-J. Zhao, M. J. Serou, C. Zhang, H. Shen, H.-W. Deng, and W. Zhou, "A deep learning-based method for automatic segmentation of proximal femur from quantitative computed tomography images," 2020.

[48] L. Melerowitz, S. Sreenivasa, M. Nachbar, A. Stsefanenka, M. Beck, C. Senger, N. Predescu, S. Ullah Akram, V. Budach, D. Zips, M. Heiland, S. Nahles, and C. Stromberger, "Design and evaluation of a deep learning-based automatic segmentation of maxillary and mandibular substructures using a 3D U-Net," *Clin. Transl. Radiat. Oncol.*, vol. 47, 2024, doi: 10.1016/j.ctro.2024.100780.

[49] S. Vinayahalingam, B. Berends, F. Baan, D. A. Moin, R. van Luijn, S. Bergé, and T. Xi, "Deep learning for automated segmentation of the temporomandibular joint," *J. Dent.*, vol. 132, 2023, doi: 10.1016/j.jdent.2023.104475.

[50] D. Morita, A. Kawarazaki, J. Koimizu, S. Tsujiko, M. Soufi, Y. Otake, Y. Sato, and T. Numajiri, "Automatic orbital segmentation using deep learning-based 2D U-Net and accuracy evaluation: A retrospective study," *J. Craniomaxillofac. Surg.*, vol. 51, no. 10, pp. 609–613, 2023, doi: 10.1016/j.jcms.2023.09.003.

[51] B. Felfeliyan, A. Hareendranathan, G. Kuntze, J. L. Jaremko, and J. L. Ronsky, "Improved-Mask R-CNN: Towards an accurate generic MSK MRI instance segmentation platform (data from the Osteoarthritis Initiative)," *Comput. Med. Imaging Graph.*, vol. 97, 2022, doi: 10.1016/j.compmedimag.2022.102056.