BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 74(5), 2026, Article number: 1621
DOI: 10.24425/bpasts.2026.1621

Deep learning-based rib segmentation and fracture classification in chest CT images

Osman Doğuş GÜLGÜN [1] *, **Erdem UYSAL** [2], and **Hamza EROL** [2]

[1] Department of Computer Technologies, Vocational High School, Toros University, Mezitli, 33340, Mersin, Turkey
[2] Department of Computer Engineering, Faculty of Engineering, Mersin University, Çiftlikköy, 33100, Mersin, Turkey

Abstract. Rib fractures are among the most common skeletal injuries in blunt thoracic trauma and are associated with serious respiratory complications, especially when multiple fractures are present. Accurate and rapid identification of fracture type and location is critical for treatment planning and prevention of potential complications. However, manual assessment of chest computed tomography (CT) images is time-consuming and prone to error. In this study, we propose a two-stage deep learning framework for automatic rib segmentation and fracture type classification on chest CT images. In the first stage, rib structures are segmented using an enhanced U-Net-based model with a multi-encoder design trained on an expert-annotated chest CT dataset. The results show that the model achieves high segmentation performance in terms of Dice Similarity Coefficient (DSC) and Intersection over Union (IoU). In the second stage, fracture-centered image patches extracted from the RibFrac dataset are classified into four fracture types using an ensemble classifier that exploits the outputs of multiple convolutional neural networks. Experimental results demonstrate that the proposed approach attains an accuracy above 99% in fracture classification, outperforming individual models and recent methods in the literature. Overall, the findings indicate that the proposed two-stage deep learning framework can reliably model rib anatomy and fracture types and has the potential to be integrated into future clinical workflows as a computer-aided decision-support system.

Keywords: rib fracture; chest CT; deep learning; rib segmentation; fracture classification.

1. INTRODUCTION

Rib fractures are among the most common skeletal injuries encountered after motor vehicle collisions, falls from a height, and blunt thoracic trauma; even a single fracture can cause severe pain and respiratory limitation, whereas multiple fractures may lead to life-threatening complications such as pneumothorax, hemothorax, and pulmonary contusion [1]. Therefore, early and accurate identification of the fracture type and its anatomical location is critical for determining the treatment strategy, optimizing the clinical course, and preventing potential complications. Although plain radiography and computed tomography (CT) are most frequently used for the diagnosis of rib fractures, the superimposition of ribs on X-ray images can cause especially small or non-displaced fractures to be overlooked. In contrast, CT delineates the three-dimensional anatomy of the ribs and surrounding soft tissues in detail and thereby facilitates fracture detection; however, the manual assessment of multiple high-resolution slices is time-consuming, fatiguing, and error-prone for radiologists. This has progressively increased the demand for artificial intelligence-based automatic decision support systems that can evaluate rib fractures on CT images rapidly, accurately, and in a reproducible manner.

In recent years, deep learning-based methods have come to the forefront in medical imaging segmentation and classification tasks due to their high accuracy and generalizability. Specifically, convolutional neural networks (CNNs) enriched with multi-scale feature extraction, attention mechanisms, and three-dimensional analysis modules have been widely adopted for the segmentation of rib structures and the automatic detection of fracture foci in chest CT images. Contemporary studies on rib fracture analysis simultaneously propose end-to-end detection and classification frameworks and advanced architectures dedicated to rib and fracture segmentation, thereby enabling the extraction of multi-level information regarding both the presence of lesions and their type and clinical relevance.

In the context of automatic detection and clinically meaningful subtype classification of rib fractures, various deep learning frameworks have been proposed. Marting *et al*. introduced a triple Retina U-Net architecture that combines RetinaNet and U-Net to classify trauma-related rib fractures according to the Chest Wall Injury Society (CWIS) system, achieving an F1 score of 0.83 for fracture detection and 94% accuracy for rib numbering, thereby demonstrating that CWIS-based stratification can be implemented within a single end-to-end framework [2]. Li *et al*. designed a You Only Look Once version 8 (YOLOv8)-based model enhanced with an attention module to assess the severity of acute rib fractures; their approach simultaneously performed fracture localization and grading into mild and severe categories and achieved an F1 score of 0.948 on the development set, indicating that deep learning can be effectively

*e-mail: dogus.gulgun@toros.edu.tr

Manuscript submitted 2025-12-11, revised 2026-05-11, initially accepted for publication 2026-06-09, published in September 2026.

adapted to fracture severity classification [3]. Wang *et al*. reported that an extended U-Net-based system trained on a large patient cohort yielded a sensitivity of 93.3% and a specificity of 98.8% at the rib level, and provided a marked gain in sensitivity compared with experienced radiologists, highlighting the potential of artificial intelligence models as second readers to reduce missed fractures [4]. Similarly, Li *et al*. developed the Fresh Rib Fracture Detection and Positioning System (FRF-DPS), which combines extended U-Net-based rib segmentation with fracture detection and localization modules, achieving an F1 score of 0.923 and 99.7% accuracy in anatomical localization; notably, the system was able to detect small and non-displaced acute fractures with high sensitivity [5]. The Composite Attention Residual U-Net (CAR-U-Net), which integrates attention mechanisms into the U-Net architecture, employs composite channel-wise and spatial attention modules to emphasize fracture regions, thereby improving the F1 score and reducing the false-positive rate compared with conventional U-Net variants [6]. CenterNet-based models using multi-scale heatmap pyramids have been proposed to enhance the detection of small and slender fractures, outperforming popular detection networks such as Faster R-CNN and YOLOv5 in terms of higher mean accuracy and lower false-positive rates [7]. The end-to-end system developed by Yang *et al*. integrates CenterNet-based detection, a bidirectional long short-term memory (LSTM)-enhanced U-Net for segmentation, and cascaded ResNet-34 classifiers within a single pipeline. It achieved an F1 score of 0.827 for fracture detection, 87.6% accuracy in distinguishing acute from chronic fractures, and 95.2% accuracy in differentiating displaced from non-displaced acute fractures, demonstrating that both anatomical localization and clinical subtype classification can be jointly addressed within one unified framework [8].

Another line of research focuses on advanced segmentation architectures that can precisely distinguish rib and fracture structures at the pixel level. Liu *et al*. proposed a UX-Net-based Multi-Stream and Multi-Scale Fusion UXNet (M2SUXNet) model, which combines multi-stream encoders with multi-scale feature fusion. On the RibFrac dataset, their approach achieved a Dice score of 75.34% and a precision of 93.79% for fracture segmentation, demonstrating that the proposed architecture can better capture small and irregularly shaped fractures [9]. Cao *et al*. introduced the Shape-Aware FracNet (SA-FracNet) architecture, which integrates contrastive representation learning on a large number of unlabeled CT images with downstream fracture detection and segmentation tasks; they reported a sensitivity of 92.6% for fracture detection and a Dice score of 75.46% for segmentation, thereby highlighting the performance gains provided by shape-aware learning strategies in small and complex lesions [10]. Jin *et al*. released the RibSeg v2 dataset, consisting of CT scans from 660 patients with expert-validated rib annotations, together with an accompanying method toolkit; they achieved a Dice score of 90.5% and a labeling accuracy of 91.5% for rib segmentation and, at the same time, established a large-scale, standardized benchmark platform for the rib segmentation field [11].

Similarly, comparative studies in the field of bone segmentation indicate that architectural choices are also decisive for the analysis of more complex injuries such as rib fractures. Krawczyk and Starzyński compared FCN, PSPNet, U-Net, and SegNet architectures on pelvic CT images and showed that a ResNet-50-based U-Net achieved an IoU of 92.87% and a Dice score of 96.31% for the bone class, demonstrating that deep convolutional networks provide a robust basis for reliably separating bony structures from surrounding soft tissues [12]. These findings suggest that U-Net-based encoder-decoder architectures constitute a flexible and extensible starting point for tackling more complex segmentation and classification problems related to rib fractures.

The existing literature consistently demonstrates that rib fractures can be automatically analyzed on CT images and classified into clinically meaningful subtypes. However, precise classification of fracture types and accurate segmentation of rib structures remain challenging due to factors such as data heterogeneity, variability in fracture size and shape, and imbalance between lesion and background regions. In this study, we propose a two-stage deep learning-based framework operating on CT images to mitigate these challenges. In the first stage, automatic segmentation of rib structures is performed using CT scans that have been manually masked by an expert physician. In the second stage, a classification module leveraging the multi-class fracture labels provided within the RibFrac dataset is employed to distinguish the type of each fracture. The architectures used in the segmentation and classification stages are task-specific and configured to achieve high accuracy in their respective tasks. In this way, trauma-related rib fractures are analyzed within a unified deep learning framework at both the structural and fracture-type levels. In the following sections, we describe the datasets used, the proposed architectural design, the experimental setup, and the obtained results in detail. We then discuss these findings in comparison with the existing literature.

2. MATERIALS AND METHODS

In this section, we present the technical details of the proposed two-stage deep learning framework. In the first stage, rib structures are segmented; in the second stage, the detected regions are classified according to fracture type. The datasets used, preprocessing steps, model architectures, training procedures, evaluation metrics, and implementation environment are described in dedicated subsections to support the reproducibility of the study.

2.1. Dataset

In this study, two different chest CT datasets were used for rib segmentation and fracture classification. In this subsection, the characteristics of these datasets are presented in detail under separate headings. For both datasets, the image preprocessing steps described in the following sections were applied to render them suitable for training the proposed models.

2.1.1. Chest CT Segmentation dataset

The Chest CT Segmentation dataset consists of 8-bit grayscale chest CT slices with a resolution of 512×512 pixels, accompanied by segmentation masks that annotate anatomical structures such as the lungs, heart, and trachea in separate channels [13]. In this study, a total of 2774 images suitable specifically for rib segmentation were selected, and rib masks were manually generated in Adobe Photoshop by an expert radiologist from Mersin University Faculty of Medicine. This curated subset was used for training and testing the segmentation model.

2.1.2. RibFrac dataset

The RibFrac dataset is a large medical image collection comprising chest CT scans from 660 patients, with annotations that categorize rib fractures into four classes: Buckle, Displaced, Non-Displaced, and Segmental [14]. The images are provided as three-dimensional ".nii.gz" volumes in which fracture regions are explicitly labeled in the corresponding annotation files. In this study, only the training and validation subsets containing labels were used for training the classification models; the test subset was excluded because ground-truth annotations were not available.

2.2. Data preprocessing

In this section, we summarize the image processing and data augmentation steps applied to both chest CT datasets. These procedures were systematically designed and executed to enhance data quality and improve the performance of the models.

2.2.1. Preprocessing of the Chest CT Segmentation dataset

For training the segmentation model, 2774 images suitable specifically for rib segmentation were selected from the Chest CT Segmentation dataset, and a rib mask for each image was manually created in Adobe Photoshop by an expert radiologist from Mersin University Faculty of Medicine. Among this curated subset, 2219 images were assigned to the training set and 555 to the test set. First, all images were converted to grayscale, and pixel intensities were normalized to span from 0 to 255 using min–max scaling. Next, a piecewise intensity transformation was applied to emphasize bony structures: pixels with intensities between 0 and 50 were multiplied by 0.75, those between 50 and 100 by 0.85, those between 150 and 200 by 1.15, and those between 200 and 255 by 1.25.

To increase data diversity, only the training subset was augmented using horizontal and vertical flipping, random rotation and scaling, brightness and contrast adjustments, gamma correction, and adaptive histogram equalization, expanding the 2219 training images to 5000 samples [15]. The test set was left in its original form without any augmentation. In the last step, all images were converted to three-channel red-green-blue (RGB) format, resized to 256×256 pixels, and normalized using the ImageNet mean and standard deviation. In this way, the model was trained and evaluated on a total of 5555 images. Figure 1 illustrates an example chest CT slice from the Chest CT Segmentation dataset and its corresponding rib mask after the preprocessing steps.

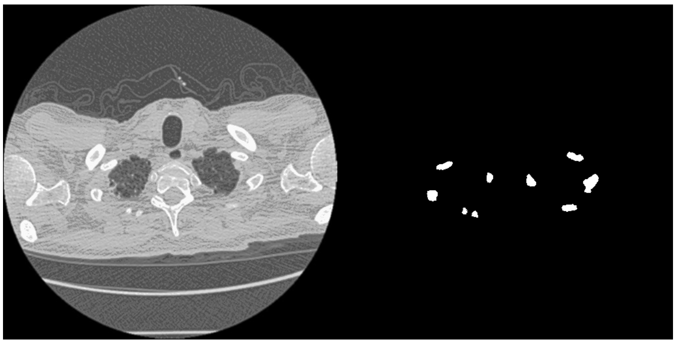

Fig. 1. Example image from the Chest CT Segmentation dataset and its corresponding mask

2.2.2. Preprocessing of the RibFrac dataset

For the classification model, the RibFrac dataset was used, which contains three-dimensional chest CT volumes in ".nii.gz" format and corresponding annotation files that delineate fracture regions for 660 patients. The dataset is organized into four main subsets: two training sets, one validation set, and one test set. Each patient volume contains, on average, 330 axial CT slices. In this study, only the training and validation subsets with available labels were considered. The annotation files provide both the rib-level location of the fracture and the corresponding fracture type. During preprocessing, two-dimensional slices containing fractures were extracted from the volumes using the annotation coordinates, and rib-centered patch images of size 128×128 pixels were generated by cropping around the fracture focus. To enhance image quality, all patches were first converted to grayscale, followed by application of contrast-limited adaptive histogram equalization (CLAHE) for contrast enhancement and unsharp masking to sharpen bone and edge details [16]. Because the raw class distribution was imbalanced, 2000 images from each class (8000 images in total) were reserved for testing, and the remaining training samples were augmented using horizontal/vertical flipping, random rotation and scaling, brightness/contrast adjustment, affine transformations and jittering, yielding a balanced training set with 10 000 images per class across the four classes. To prevent data leakage, the train-test split was performed at the patient level, ensuring that images from the same patient were not included in both sets. These augmentation procedures were applied only to the training data, while the test images were kept in their original form. Example fracture-centered CT patches obtained after preprocessing are shown in Fig. 2.

Fig. 2. Example fracture-centered patch images from the RibFrac dataset

2.3. Model architecture

In this study, we propose a two-stage deep learning architecture for automatic rib segmentation and fracture type classification on chest CT images. In this section, we first describe the segmentation model used to extract rib structures and then summarize the classification model architecture designed to predict the fracture types under separate subsections.

2.3.1. Segmentation model

For automatic extraction of rib structures from CT images, we propose a triple-fusion model based on U-Net++, in which three different encoder architectures are integrated within a single framework [17]. The model takes as input the preprocessed and augmented three-channel chest CT images of size 256×256 pixels obtained from the Chest CT Segmentation dataset described in the previous section and consists of three parallel U-Net++ branches, each producing a binary segmentation mask. The encoder parts of these branches are implemented using ResNet-50, EfficientNet-B4, and DenseNet-121, respectively, thereby combining a deep architecture with residual connections that preserve contextual information, a scalable design that optimizes the parameter-performance trade-off, and a densely connected topology that enhances information flow across layers within a unified structure [18–20]. All encoder networks are initialized with weights pretrained on the ImageNet dataset. The continuous-valued output masks generated by the three U-Net++ branches are multiplied by learnable weighting coefficients $(\mathrm{w}_1, \mathrm{w}_2, \mathrm{w}_3)$ and concatenated along the channel dimension, after which they are passed through a 1×1 convolutional layer to produce a single-channel probability map [21]. In the final stage, this map is passed through a sigmoid activation function, which scales its values to the range from 0 to 1, and a threshold of 0.5 is then applied to obtain the final binary mask.

In this way, the complementary feature representations learned by different encoder architectures are fused at the output level in a trainable manner, yielding a flexible and high-accuracy model for rib segmentation. A schematic overview of the proposed segmentation model is presented in Fig. 3.

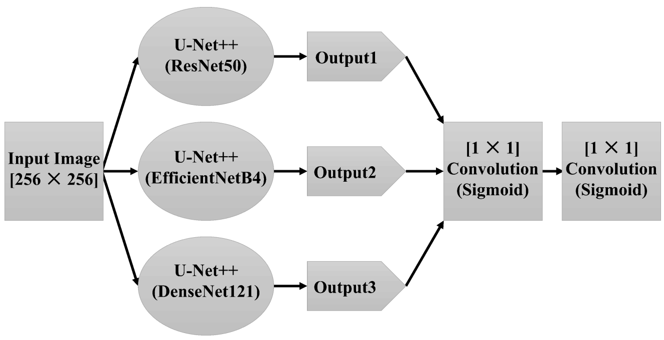

Fig. 3. Overall architecture of the proposed segmentation model

2.3.2. Classification model

To accurately distinguish the four fracture types (Buckle, Displaced, Non-Displaced, Segmental), we adopt a multi-model and ensemble decision-making strategy based on 128×128 fracture-centered patch images extracted from the RibFrac dataset. In addition to the standard augmentations applied during preprocessing, Mixup and CutMix strategies are employed before classification to smooth class boundaries and enhance generalization capability; in this setting, the mixing coefficient for Mixup is set to 0.2, and the parameter controlling the patch area in CutMix is set to 1 [22, 23].

In the classification stage, four different CNN architectures are evaluated: EfficientNet-B1, MobileNetV3-Large, ShuffleNetV2, and ResNet-50. EfficientNet-B1 optimizes the balance between parameter efficiency and accuracy, whereas MobileNetV3-Large offers a lightweight design with low resource consumption, making it suitable for computationally constrained environments [24]. ShuffleNetV2 employs a lightweight architecture based on channel shuffling and efficient information flow, while ResNet-50, with its deep residual structure, mitigates contextual information loss across consecutive convolutional layers [25, 26]. All models are initialized with ImageNet-pretrained weights; their final classification layers are reconfigured to produce four-class outputs, and the feature maps from the last convolutional layers are transformed into fixed-length vectors using Global Average Pooling. These high-dimensional feature vectors are then passed through a three-stage feature selection pipeline to enhance classification performance. First, a Variance Threshold method with a threshold of 1×10^{-6} is applied to remove low-information components [27]. Second, the ANOVA F-test-based SelectKBest method is used to select the top 1000 features with the highest discriminative power [28]. In the last step, Recursive Feature Elimination with Cross-Validation (RFECV), driven by a 200-tree Random Forest classifier with 3-fold cross-validation, is employed to identify the most informative subset, and the minimum number of features is set to 100 [29].

For each CNN architecture, the resulting optimized feature set is classified using a custom four-layer artificial neural network (ANN). Each ANN consists of fully connected layers with 1024, 512, 256, and 128 neurons, respectively; each layer is followed by ReLU activation, batch normalization, and progressively decreasing dropout rates of 40%, 30%, 20%, and 10%, thereby increasing non-linear learning capacity while reducing overfitting risk. Class probabilities obtained from the final layer of each ANN are computed independently and then fused using probability-based soft voting [30]. In this step, the probabilities produced by each ANN are normalized and averaged with equal weights to generate the final class prediction. In this way, the complementary feature-learning capabilities of the different CNN architectures are integrated at the decision level, yielding a more stable and higher-accuracy classification system compared with individual models. A schematic overview of the proposed probability-based ensemble classification architecture is presented in Fig. 4.

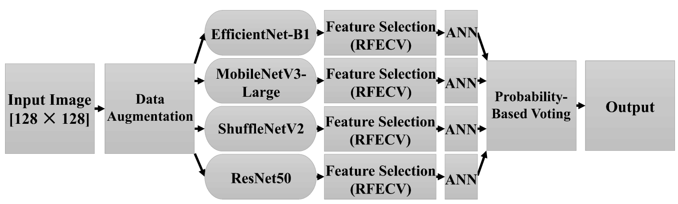

Fig. 4. Overall architecture of the proposed classification model

2.4. Training

In this section, we summarize the training protocols used for the developed segmentation and classification models. The choices of hyperparameters, loss functions, optimization algorithms, and train-test splits are presented in separate subsections, tailored to the specific requirements of the segmentation and classification tasks.

2.4.1. Training of the Segmentation model

The proposed triple-fusion U-Net++ model for rib segmentation was trained on the Chest CT Segmentation dataset after applying the preprocessing steps described in the previous section. During training, a composite loss function consisting of Dice Loss, Jaccard Loss, and Focal Loss was employed to maximize the overlap between the predicted and ground-truth masks and to increase robustness to class imbalance [31–33]. The sum of these three components was defined as the overall objective function, and optimization was performed with respect to this combined loss. The model was trained for 25 epochs using mini-batches of size 2, and the Adam optimizer with a learning rate of 1×10^{-4} was adopted. To comprehensively assess segmentation performance, Intersection over Union (IoU) and Dice Similarity Coefficient (DSC) were computed, and the overall detection quality was further quantified using the Mean Average Precision (mAP) metric at thresholds mAP@50 and mAP@75. At the end of each epoch, predicted and ground-truth masks for randomly selected samples were visualized side by side, allowing the training process to be monitored not only through numerical metrics but also via qualitative inspection. In this way, the combination of a multi-component loss function and detailed evaluation strategy enabled a well-balanced training process and contributed to improved segmentation accuracy.

2.4.2. Training of the Classification model

In the fracture-type classification task with four classes, namely Buckle, Displaced, Non-Displaced and Segmental, we employed individual models based on EfficientNet-B1, MobileNetV3-Large, ShuffleNetV2 and ResNet-50, as well as the proposed ensemble classifier that combines their outputs in a probability-based manner. All models were trained under a common training protocol on 128×128 fracture-centered patches extracted from the RibFrac dataset. To strengthen class discrimination, improve sensitivity to minority classes under imbalanced sample distributions, and enhance the generalization capability of the models, a triple composite loss function (Triple Loss) consisting of Cross Entropy Loss, Focal Loss, and Kullback-Leibler Divergence was used [34]. In this formulation, weights of 40%, 40%, and 20% were assigned to Cross Entropy Loss, Focal Loss, and Kullback-Leibler Divergence, respectively. For optimization of the combined loss, the Adam optimizer was used for all CNN models with a learning rate of 2×10^{-4}, and training was performed for 50 epochs using mini-batches of size 32. To further limit overfitting and increase data diversity, Mixup and CutMix strategies were applied during training in addition to the standard augmentation procedures described in the previous sections. The feature maps obtained from the last convolutional layers of the CNN models were processed through Global Average Pooling and the three-stage feature selection pipeline previously detailed, and the resulting optimized feature subsets for each model were classified using the four-layer ANN architectures defined in the same section. Each CNN-ANN combination was trained independently and evaluated on the test set using accuracy, F1 score, precision, and recall metrics. In the final stage, the class probabilities produced by each ANN were normalized and combined via equally weighted probability-based soft voting to obtain the final class predictions. This setup enabled a comparative analysis of both the individual model performance and the overall effectiveness of the ensemble system.

2.5. Implementation environment

This study was implemented in Python on a workstation running the Windows 11 operating system, using Python 3.8 within an Anaconda environment. All deep learning models were developed under the PyTorch framework, and training and testing were performed with graphics processing unit (GPU) acceleration. For the segmentation task, image preprocessing and data augmentation were conducted using the torch, torchvision, albumentations, imgaug, OpenCV, and NumPy libraries, while segmentation masks were manually created in Adobe Photoshop. For training and evaluating the classification models, torch, torchvision, and scikit-learn were employed, whereas matplotlib, tqdm, json, and random were used for visualization and logging of model outputs. The triple composite loss function designed to improve sensitivity to class imbalance was implemented in PyTorch by extending the nn.Module class. Data augmentation pipelines were configured separately for segmentation and classification: albumentations and imgaug-based transformations were applied in the segmentation stage, whereas RandAugment from torchvision.transforms together with customized Mixup and CutMix implementations built using NumPy and PyTorch were used in the classification stage.

3. RESULTS AND DISCUSSION

In this section, we present the test results of the proposed deep learning models for rib segmentation and fracture classification, compare them with recent studies in the literature, and discuss the applicability of the findings to clinical practice.

3.1. Results

In this subsection, we report the test results obtained for the proposed rib segmentation and fracture classification models.

3.1.1. Segmentation model results

The proposed triple-fusion U-Net++ model developed for rib segmentation achieves an Intersection over Union (IoU) of 0.9311 (93.11%), a Dice Similarity Coefficient (DSC) of 0.9576 (95.76%), a mAP@50 of 0.9819 (98.19%), and an mAP@75 of 0.9747 (97.47%) on the test set.

3.1.2. Classification model results

In the four-class fracture-type classification task, models based on EfficientNet-B1, MobileNetV3-Large, ShuffleNetV2, and ResNet-50, together with the proposed ensemble classifier that combines the outputs of these four models in a probability-based manner, were evaluated. Features extracted from the last convolutional layer of each model were classified using the four-layer artificial neural network described in the previous sections, and the resulting class probabilities were fused within the ensemble using a soft-voting scheme. The accuracy, precision, recall, and F1 score values obtained on the test set are summarized in Table 1. As shown in Table 1, the accuracies of the individual CNN models range between 0.9791 and 0.9911, while the proposed ensemble classifier achieves the highest values across all metrics, yielding an accuracy of 0.9935 (corresponding to 99.35%) and an F1 score of 0.9935.

Table 1
Classification results of the proposed models

Model	Accuracy	Precision	Recall	F1 score
EfficientNet-B1	0.9911	0.9911	0.9911	0.9911
MobileNetV3-Large	0.9902	0.9903	0.9903	0.9902
ShuffleNetV2	0.9791	0.9791	0.9791	0.9791
ResNet-50	0.9820	0.9820	0.9820	0.9820
Proposed model	0.9935	0.9935	0.9935	0.9935

3.2. Discussion

The proposed triple-fusion U-Net++ model achieves superior performance compared with recent approaches in the literature, yielding Dice, IoU, mAP@50, and mAP@75 scores of 95.76%, 93.11%, 98.19%, and 97.47%, respectively, on the rib segmentation test set. Learnable fusion of the outputs from three different encoder architectures based on ResNet-50, DenseNet-121, and EfficientNet-B4 enables the joint exploitation of low-level details and high-level contextual features, which in turn contributes to a more consistent delineation of rib contours and fine structural details. In addition, the triple loss function combining Dice, Jaccard, and focal components mitigates issues related to class imbalance and improves sensitivity to minority pixels, thereby enhancing the stability of the model. Table 2 presents the Dice score of the proposed triple-fusion U-Net++ architecture in comparison with the results reported in existing rib segmentation studies.

Table 2
Comparison of the proposed segmentation model with related methods in the literature

Study	Model	Dice (%)
Proposed study	Triple-fusion U-Net++	95.76
Li *et al*. [5]	FRF-DPS	92.30
Jin *et al*. [11]	Two-stage sparse model	90.50
Yang *et al*. [8]	CenterNet + Bi-LSTM U-Net + ResNet-34	82.70
Cao *et al*. [10]	Shape-Aware FracNet	75.46

Furthermore, the proposed ensemble-based classification model combines the class probabilities obtained from four convolutional neural network architectures using a probability-based soft-voting scheme. The RFECV-based feature selection strategy and advanced data augmentation techniques employed in this study substantially improve the classification performance of the model. According to the test results, the model achieves 99.35% accuracy in fracture-type classification, outperforming related approaches reported in the literature. Table 3 presents the performance of the proposed classification model in comparison with recent methods.

Table 3
Comparison of the proposed classification model with related methods in the literature

Study	Model	Accuracy (%)
Proposed study	Ensemble (four CNN + ANN)	99.35
Yang *et al*. [8]	CenterNet + Bi-LSTM U-Net + ResNet-34	95.20
Marting *et al*. [2]	3×Retina U-Net + nnU-Net	95.00
Su *et al*. [7]	CenterNet + heat-map pyramid	89.20

From a segmentation perspective, as shown in Table 2, Li *et al*. report a Dice score of 92.30% with the Fresh Rib Fracture Detection and Positioning System (FRF-DPS), while Jin *et al*., using a two-stage sparse point-cloud–based approach on the RibSeg v2 dataset, achieve a Dice score of 90.50% [5, 11]. In contrast, the Dice score of 95.76% obtained by the proposed triple-fusion U-Net++ model indicates that rib structures can be segmented more finely and consistently compared with these strong methods in the literature.

From a classification perspective, among the studies summarized in Table 3, the end-to-end system proposed by Yang *et al*., which combines CenterNet-based detection, Bi-LSTM U-Net segmentation, and cascaded ResNet-34 classifiers, achieves an accuracy of 95.20% [8]. In comparison, the ensemble-based model proposed in this study, which classifies fractures into four distinct categories and fuses the outputs of four different CNN architectures using probability-based soft voting, attains an accuracy of 99.35%. This result demonstrates a substantial performance gain over existing methods, even under a more complex problem definition.

Overall, the two-stage framework proposed in this study – comprising a triple-fusion U-Net++–based segmentation model and an ensemble-based classification model – provides a powerful solution for the automatic detection and analysis of rib fractures on CT images. The high Dice, IoU, and mAP values obtained in the segmentation task, together with the high accuracy achieved in the classification stage, indicate that the system can reliably model both rib anatomy and fracture types. The combined use of architectural flexibility, multi-encoder fusion, a triple loss function, and an ensemble decision-making strategy highlights the potential of the proposed approach to serve as a decision-support tool for radiologists in clinical evaluation workflows and establishes a solid basis for future studies in AI-assisted medical imaging.

4. FUTURE WORK

The triple-fusion U-Net++–based segmentation model and ensemble-based classification system proposed in this study have yielded promising results for the automatic analysis of rib fractures on CT images. However, the fact that the datasets used were collected from a limited number of centers and that the distribution of fracture types does not fully reflect real clinical populations constitutes an important limitation for the generalizability of the model. In future work, external validation studies are planned to use larger, multi-center CT datasets that encompass a broader range of scanners, imaging protocols, and patient profiles.

From a model development perspective, future work will focus on designing hybrid architectures that more effectively capture both the topological relationships of rib anatomy and the volumetric context of CT data by integrating Transformer-based components, graph neural networks, and three-dimensional segmentation models. From a clinical integration perspective, interpretability approaches that visualize the decision-making process and quantify predictive uncertainty will be incorporated. In this regard, methods such as attention maps and explainable feature-attribution maps will be explored, and the system will be integrated into PACS-based workflows as a second reader to evaluate its impact on radiologist performance and reporting time in routine clinical practice. Additionally, the source code and pretrained model weights developed in this study will be made available from the corresponding author upon reasonable request.

REFERENCES

[1] S. Aydin, S. Kahraman Aydin, B. Gulmez, S.G. Gunes, O. Kavurmaci, and O.F. Dadas, "Rib fracture characteristics increasing the risk of hemothorax: a multicenter study," *Sci. Rep.*, vol. 14, no. 1, p. 29827, 2024, doi: 10.1038/s41598-024-79548-z.

[2] V. Marting, N. Borren, M.R. van Diepen, E.M.M. van Lieshout, M.M. E. Wijffels, and T. van Walsum, "A deep learning-based approach to automated rib fracture detection and CWIS classification," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 20, no. 7, pp. 1381–1389, 2025, doi: 10.1007/s11548-025-03390-5.

[3] T. Li *et al.*, "Intelligent detection and grading diagnosis of fresh rib fractures based on deep learning," *BMC Med. Imaging*, vol. 25, no. 1, p. 98, 2025, doi: 10.1186/s12880-025-01641-0.

[4] S. Wang, D. Wu, L. Ye, Z. Chen, Y. Zhan, and Y. Li, "Assessment of automatic rib fracture detection on chest CT using a deep learning algorithm," *Eur. Radiol.*, vol. 33, no. 3, pp. 1824–1834, 2022, doi: 10.1007/s00330-022-09156-w.

[5] N. Li *et al.*, "An automatic fresh rib fracture detection and positioning system using deep learning," *Brit. J. Radiol.*, vol. 96, no. 1146, p. 20221006, 2023, doi: 10.1259/bjr.20221006.

[6] X. Wang and Y. Wang, "Composite attention residual U-Net for rib fracture detection," *Entropy*, vol. 25, no. 3, p. 466, 2023, doi: 10.3390/e25030466.

[7] Y. Su, X. Zhang, H. Shangguan, and R. Li, "Rib fracture detection in chest CT image based on a CenterNet network with heatmap pyramid structure," *Signal Image Video Process.*, vol. 17, no. 5, pp. 2343–2350, 2023, doi: 10.1007/s11760-022-02451-5.

[8] C. Yang *et al.*, "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, 2022, doi: 10.1016/j.ejrad.2022.110434.

[9] Y. Liu, L. Zhang, and Z. Jiang, "Multi-stream and multi-scale fusion rib fracture segmentation network based on UXNet," *Quant. Imaging Med. Surg.*, vol. 15, no. 1, pp. 230–248, 2025, doi: 10.21037/qims-24-1356.

[10] Z. Cao, L. Xu, D.Z. Chen, H. Gao, and J. Wu, "A robust shape-aware rib fracture detection and segmentation framework with contrastive learning," *IEEE Trans. Multimedia*, vol. 25, pp. 1584–1591, 2023, doi: 10.1109/TMM.2023.3263074.

[11] L. Jin *et al.*, "RibSeg v2: a large-scale benchmark for rib labeling and anatomical centerline extraction," *IEEE Trans. Med. Imaging*, vol. 43, no. 1, pp. 570–581, 2024, doi: 10.1109/TMI.2023.3313627.

[12] Z. Krawczyk and J. Starzyński, "Segmentation of bone structures with the use of deep learning techniques," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 69, no. 3, p. e136751, 2021, doi: 10.24425/bpasts.2021.136751.

[13] M. Polo. "Chest CT Segmentation." Kaggle. [Online]. Available: https://www.kaggle.com/datasets/polomarco/chest-ct-segmentation [Accessed: 19. Jan. 2025].

[14] J. Yang *et al.*, "Deep rib fracture instance segmentation and classification from CT on the RibFrac Challenge," *IEEE Trans. Med. Imaging*, vol. 44, no. 8, pp. 3410–3427, 2025, doi: 10.1109/TMI.2025.3565514.

[15] G. Alwakid, W. Gouda, and M. Humayun, "Deep learning-based prediction of diabetic retinopathy using CLAHE and ESRGAN for enhancement," *Healthcare*, vol. 11, no. 6, p. 863, 2023, doi: 10.3390/healthcare11060863.

[16] H. Avci and J. Karakaya, "A novel medical image enhancement algorithm for breast cancer detection on mammography images using machine learning," *Diagnostics*, vol. 13, no. 3, p. 348, 2023, doi: 10.3390/diagnostics13030348.

[17] S. Anari, S. Sadeghi, G. Sheikhi, R. Ranjbarzadeh, and M. Bendechache, "Explainable attention based breast tumor segmentation using a combination of UNet, ResNet, DenseNet, and EfficientNet models," *Sci. Rep.*, vol. 15, no. 1, p. 1027, 2025, doi: 10.1038/s41598-024-84504-y.

[18] N. Behar and M. Shrivastava, "ResNet50-based effective model for breast cancer classification using histopathology images," *Comput. Model. Eng. Sci.*, vol. 130, no. 2, pp. 823–839, 2022, doi: 10.32604/cmes.2022.017030.

[19] W. Liu, J. Luo, Y. Yang, W. Wang, J. Deng, and L. Yu, "Automatic lung segmentation in chest X-ray images using improved U-Net," *Sci. Rep.*, vol. 12, no. 1, p. 8649, 2022, doi: 10.1038/s41598-022-12743-y.

[20] N. Cinar, A. Ozcan, and M. Kaya, "A hybrid DenseNet121-UNet model for brain tumor segmentation from MR images," *Biomed. Signal Process. Control*, vol. 76, p. 103647, 2022, doi: 10.1016/j.bspc.2022.103647.

[21] H. Wang, P. Cao, J. Yang, and O. Zaiane, "MCA-UNet: multi-scale cross co-attentional U-Net for automatic medical image segmentation," *Health Inf. Sci. Syst.*, vol. 11, no. 1, p. 10, 2023, doi: 10.1007/s13755-022-00209-4.

[22] P. Rana, A. Sowmya, E. Meijering, and Y. Song, "Data augmentation with improved regularisation and sampling for imbalanced blood cell image classification," *Sci. Rep.*, vol. 12, no. 1, p. 18101, 2022, doi: 10.1038/s41598-022-22882-x.

[23] D. Sun, F. Dornaika, and J. Charafeddine, "LCAMix: local-and-contour aware grid mixing based data augmentation for medical image segmentation," *Inf. Fusion*, vol. 110, p. 102484, 2024, doi: 10.1016/j.inffus.2024.102484.

[24] J. Di, W. Guo, J. Liu, L. Ren, and J. Lian, "AMMNet: a multimodal medical image fusion method based on an attention mechanism and MobileNetV3," *Biomed. Signal Process. Control*, vol. 96, p. 106561, 2024, doi: 10.1016/j.bspc.2024.106561.

[25] C. Liu, J. Yang, Y. Liu, Y. Zhang, S. Liu, T. Chaikovska, and C. Liu, "A cervical lesion recognition method based on ShuffleNetV2-CA," *Inf. Dyn. Appl.*, vol. 2, no. 2, pp. 77–89, 2023, doi: 10.56578/ida020203.

[26] N.L. Yadav, S. Singh, R. Kumar, and D.K. Nishad, "Transfer learning with fuzzy decision support for multi-class lung disease classification: performance analysis of pre-trained CNN models," *Sci. Rep.*, vol. 15, no. 1, p. 35127, 2025, doi: 10.1038/s41598-025-19114-3.

[27] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Gradient boosting feature selection with machine learning classifiers for intrusion detection on power grids," *IEEE Trans. Netw. Serv. Manage.*, vol. 18, no. 1, pp. 1104–1116, 2021, doi: 10.1109/TNSM.2020.3032618.

[28] A. Hage Chehade, N. Abdallah, J.-M. Marion, M. Oueidat, and P. Chauvet, "Lung and colon cancer classification using medical imaging: a feature engineering approach," *Phys. Eng. Sci. Med.*, vol. 45, no. 3, pp. 729–746, 2022, doi: 10.1007/s13246-022-01139-x.

[29] N.S. Azman *et al.*, "Support vector machine–recursive feature elimination for feature selection on multi-omics lung cancer data," *Prog. Microbes Mol. Biol.*, vol. 6, no. 1, p. a0000327, 2023, doi: 10.36877/pmmb.a0000327.

[30] S.H. Ahn *et al.*, "Uncertainty quantification in automated detection of vertebral metastasis using ensemble Monte Carlo dropout," *J. Imaging Inform. Med.*, vol. 38, no. 5, pp. 2700–2715, 2024, doi: 10.1007/s10278-024-01369-3.

[31] Q. Ming and X. Xiao, "Towards accurate medical image segmentation with gradient-optimized Dice loss," *IEEE Signal Process. Lett.*, vol. 31, pp. 191–195, 2024, doi: 10.1109/LSP.2023.3329437.

[32] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Res. Notes*, vol. 15, no. 1, p. 210, 2022, doi: 10.1186/s13104-022-06096-y.

[33] Y. Nie, P. Sommella, M. Carratù, M. O'Nils, and J. Lundgren, "A deep CNN transformer hybrid model for skin lesion classification of dermoscopic images using focal loss," *Diagnostics*, vol. 13, no. 1, p. 72, 2022, doi: 10.3390/diagnostics13010072.

[34] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified focal loss: generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Comput. Med. Imaging Graph.*, vol. 95, p. 102026, 2022, doi: 10.1016/j.compmedimag.2021.102026.