



Dual-Stream Attention-Based Classification Network for Tibial Plateau Fractures via Diffusion Model Augmentation and Segmentation Map Integration

Yi Xie^{1,2} · Zhi-wei Hao³ · Xin-meng Wang⁴ · Hong-lin Wang^{1,2} · Jia-ming Yang^{1,2} · Hong Zhou^{1,2} · Xu-dong Wang^{1,2} · Jia-yao Zhang⁵ · Hui-wen Yang^{2,6} · Peng-ran Liu^{1,2} · Zhe-wei Ye^{1,2}

Received: 20 September 2024 / Revised: 25 October 2024 / Accepted: 4 November 2024 / Published online: 25 February 2025
© The Author(s), under exclusive licence to Huazhong University of Science and Technology 2025

Abstract

Objective This study aimed to explore a novel method that integrates the segmentation guidance classification and the diffusion model augmentation to realize the automatic classification for tibial plateau fractures (TPFs).

Methods YOLOv8n-cls was used to construct a baseline model on the data of 3781 patients from the Orthopedic Trauma Center of Wuhan Union Hospital. Additionally, a segmentation-guided classification approach was proposed. To enhance the dataset, a diffusion model was further demonstrated for data augmentation.

Results The novel method that integrated the segmentation-guided classification and diffusion model augmentation significantly improved the accuracy and robustness of fracture classification. The average accuracy of classification for TPFs rose from 0.844 to 0.896. The comprehensive performance of the dual-stream model was also significantly enhanced after many rounds of training, with both the macro-area under the curve (AUC) and the micro-AUC increasing from 0.94 to 0.97. By utilizing diffusion model augmentation and segmentation map integration, the model demonstrated superior efficacy in identifying Schatzker I, achieving an accuracy of 0.880. It yielded an accuracy of 0.898 for Schatzker II and III and 0.913 for Schatzker IV; for Schatzker V and VI, the accuracy was 0.887; and for intercondylar ridge fracture, the accuracy was 0.923.

Conclusion The dual-stream attention-based classification network, which has been verified by many experiments, exhibited great potential in predicting the classification of TPFs. This method facilitates automatic TPF assessment and may assist surgeons in the rapid formulation of surgical plans.

Keywords Artificial intelligence · YOLOv8 · Tibial plateau fracture · Diffusion model augmentation · Segmentation map

Yi Xie, Zhi-wei Hao and Xin-meng Wang contributed equally to this work.

✉ Hui-wen Yang
xyhw@hust.edu.cn

Peng-ran Liu
lprlprlprwd@163.com

Zhe-wei Ye
yezhewei@hust.edu.cn

¹ Department of Orthopedics Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

² Laboratory of Intelligent Medicine Research, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

³ School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China

⁴ Key Laboratory of Clinical Biochemistry Testing in Universities of Yunnan Province, School of Basic Medical Sciences, Dali University, Dali 671003, China

⁵ Department of Orthopedics Surgery, Fujian Provincial Hospital, Fuzhou 350001, China

⁶ Department of Otorhinolaryngology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

1 Introduction

Timely and accurate diagnosis of tibial plateau fractures (TPFs) is very important for the development of treatment plans [1, 2]. Currently, the diagnosis of TPFs mainly depends on the doctor's experience with the aid of medical imaging techniques. X-rays, magnetic resonance imaging (MRI), and computed tomography (CT) have been commonly used to diagnose TPFs. However, fracture zone detection in these images is complicated and, therefore, the accurate classification of TPFs may be restricted, especially in the initial diagnosis and at the primary stage [3, 4].

In recent years, artificial intelligence (AI) technology, characterized by feature engineering, artificial neural network, and deep learning, has been widely used in the field of orthopedic trauma. It mainly covers clinical applications such as fracture diagnosis, fracture classification, three-dimensional lesion segmentation, surgical planning, and preoperative and postoperative risk prediction [5–8]. It can precisely classify fractures by recognizing and labelling information through multiple classification tasks. Liu et al. [9] adopted the YOLOX-SwinT algorithm to increase the accuracy of the Arbeitsgemeinschaft für Osteosynthesefragen (AO) classification of femoral intertrochanteric fractures by trauma orthopedic surgeons. Zhang et al. [10] explored the influence of deep learning radiomics on the classification of osteoporotic vertebral fractures (OVFs) in X-ray images. Ono et al. [11] constructed a classification model of osteoporotic lumbar fracture on radiographs based on deep learning, and the results demonstrated that AI technology was conducive to the automatic classification of osteoporotic lumbar vertebral fracture (OLVF).

Orthopaedic trauma images encompass a large amount of information, and details such as bone structure, organ boundaries, and soft tissue properties may not be accurately evaluated through visual examination. Deep learning models can assist doctors in locating lesions and the segmentation of medical images has gradually attracted the interest of many researchers. Chung et al. [12] classified proximal humerus fractures on the basis of the convolutional neural network (CNN) algorithm, and the classification accuracy reached 85%. Mutasa et al. [13] employed the classification network to train the CNN recognition algorithm to classify femoral neck fractures accurately in accordance with the Garden classification. For Garden I + II, Garden III + IV, and the three types of differentiation between them and normal X-ray films, the accuracy of the model reached 86%. In addition, in fracture segmentation, U-Net and its variants have been widely used to accurately segment fracture areas and improve diagnostic accuracy [14–18]. Chen et al. proposed a cross-scale residual

network for the segmentation of scaphoid fractures [15]. Zeng et al. designed a dual-branch architecture and fusion strategy for automatic pelvic fracture segmentation [17]. In the field of medical image augmentation, generative adversarial networks (GANs) and conditional diffusion models have been demonstrated for processing image data and improving radiographic image analysis [19]. However, existing studies have not yet combined image segmentation and utilized diffusion models for medical data augmentation in the classification of traumatic fractures [20–23]. In terms of fracture classification, no study has improved the efficiency of the model by combining these two methods, and the current model is generally limited by the amount of data of various types, resulting in insufficient accuracy [24].

In this study, we proposed a novel approach that combines fracture segmentation with classification and data augmentation through a diffusion model to scale up data training to improve training accuracy. We used the medical image segmentation model to segment the fracture image and then stacked the segmentation image with the original image as input to train the classification model to improve the classification accuracy. As a generative model, the diffusion model has shown remarkable advantages in medical image generation and augmentation in recent years. By generating more fracture samples and enhancing the training data of the segmentation model, its performance can be effectively improved. To solve the problem of limited data annotation in the process of segmentation model training, we introduced the diffusion model for data augmentation, and the expanded data yielded good experimental results after training. The contribution of this study is the use of a segmentation map to guide classification, improve the recognition ability of the classification model, and generate more fracture samples through the diffusion model to solve the problem of limited data annotation. Using the diffusion model augmentation and segmentation maps, the dual-stream attention-based classification network not only significantly improved the accuracy of fracture classification but also provided a new solution to the problem of data scarcity in medical image processing.

2 Materials and Methods

2.1 Data Collection and Ethics

To establish and validate the deep learning model, we collected 3781 knee joint X-ray images from the Orthopedic Trauma Center of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology from 2013 to 2023. Patient data, including age, sex, and the time from injury to operation, were collected. Preoperative and

postoperative imaging data, including X-rays (of the anterior and lateral knee joints), surgical data, fracture reduction approaches, operation duration, and blood loss, were collected. In accordance with the relevant criteria, 2052 patients with symptoms but no fracture of the knee joint injury were included, and 1729 patients were diagnosed with TPF. There were 1087 males and 642 females among the TPF patients. All image data were enumerated, categorized and stored in a standardized, high-quality format. This study was approved by the Medical Ethics Committee of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology (IRB No. 0840). The data collection complied with clinical ethical practices and adhered to the ethical guidelines established in the Helsinki Declaration. Regarding the compliant use of AI in the experiment, we registered this study in the China Clinical Trial Registry (registration number: ChiCTR2300070658). Overall, these measures ensure that research complies with ethical standards and regulatory standards and protects the rights, security, and confidentiality of the participants involved in the research.

2.2 Data Processing

The diagnosis of fractures in the knee joint was typically determined through the combination of anteroposterior and lateral X-ray images. By conducting a comprehensive analysis of the available data, we primarily undertook labelling and further data processing on the anteroposterior images. After the digital imaging and communications in medicine (DICOM) knee image data were converted to joint photographic experts group (JPEG) format, the image data were imported into Labelme software for annotation. The region of interest (ROI) generated by lesions on the tibial platform of trauma patients was manually delineated, and X-ray image omics features were extracted. The Schatzker classification, which is used by orthopedic surgeons for TPF in clinical practice, is more suitable for clinical practice, convenient for doctors to carry out surgical planning and communication, and more convenient for the selection of surgical approaches. The pathological status of the patient's tibial plateau was partitioned and numbered in accordance with the classic Schatzker classification to clearly demonstrate the specific location of displacement and reflect the fracture line and mass. Three senior physicians with more than 10 years of experience were invited to diagnose TPF. Simultaneously, we conducted Schatzker classification for all patients with other doctors who had more than 20 years of experience in traumatic orthopedic joint treatment and analysed the consistency of the classification to reduce labelling errors. When only MR images were available for a patient, at least two senior diagnosing physicians completed the diagnosis. To be more applicable in clinical practice, we further categorized and refined the fracture types. Specifically, Schatzker type I

fractures are classified as class A, Schatzker types II and III as class B, Schatzker type IV as class C, Schatzker types V and VI as class D, and intercondylar spinal fractures as class K. Schatzker type III fractures are frequently associated with type II injuries, and we sorted the two together as class B. Additionally, due to the complex clinical management of bilateral tibial fractures (type V) and combined fractures and dislocations (type VI), which require open surgery and comprehensive measures, they are classified as class D [2, 3, 25]. Finally, owing to the clinical prevalence of tibial plateau intercondylar ridge fractures, we also included this particular type in our intelligent diagnosis. The detailed annotation, classification and corresponding situation of the ROI according to the classic Schatzker classification of TPFs are presented in Fig. 1.

2.3 Baseline Classification Model

In clinical practice, the determination and classification of focal sites can help orthopedic surgeons choose treatment and surgical plans. To the best of our knowledge, previous studies focused on the identification of fractures, and there were no studies on the classification of TPFs. The You Only Look Once (YOLO) series of models are renowned for their single-stage architecture and real-time processing capabilities, facilitating rapid inference while maintaining high accuracy [26–28]. In this study, we chose YOLOv8n-cls as the baseline classification model for the fracture type identification. YOLOv8n-cls is a lightweight classification model of the YOLO series. This model demonstrates excellent performance in image classification and target detection, and its single-stage architecture and efficient computing power are highly compatible with the practical requirements of medical image analysis. The YOLOv8n-cls model adopts a convolutional neural network architecture, encompassing multiple layers of convolutional layers, pooled layers, and fully connected layers. The model input is a fracture image of a fixed size. The image features are extracted through a sequence of convolution operations. The pooling operation reduces the dimension of the feature map while retaining crucial information. Finally, the probability distribution of each fracture type is output via the fully connected layer and the Softmax activation function. This design enables the model to effectively capture significant features in fracture images and accurately classify different types of fractures.

For TPF recognition and classification, the input ports to the model were preprocessed fracture images. These images were normalized, and the standardized format was employed as the input data. The output of the model was a probability distribution for each fracture type, which represented the likelihood of each type of fracture. Specifically, for a given fracture image, the model outputted a vector in which each element represented the probability of the corresponding

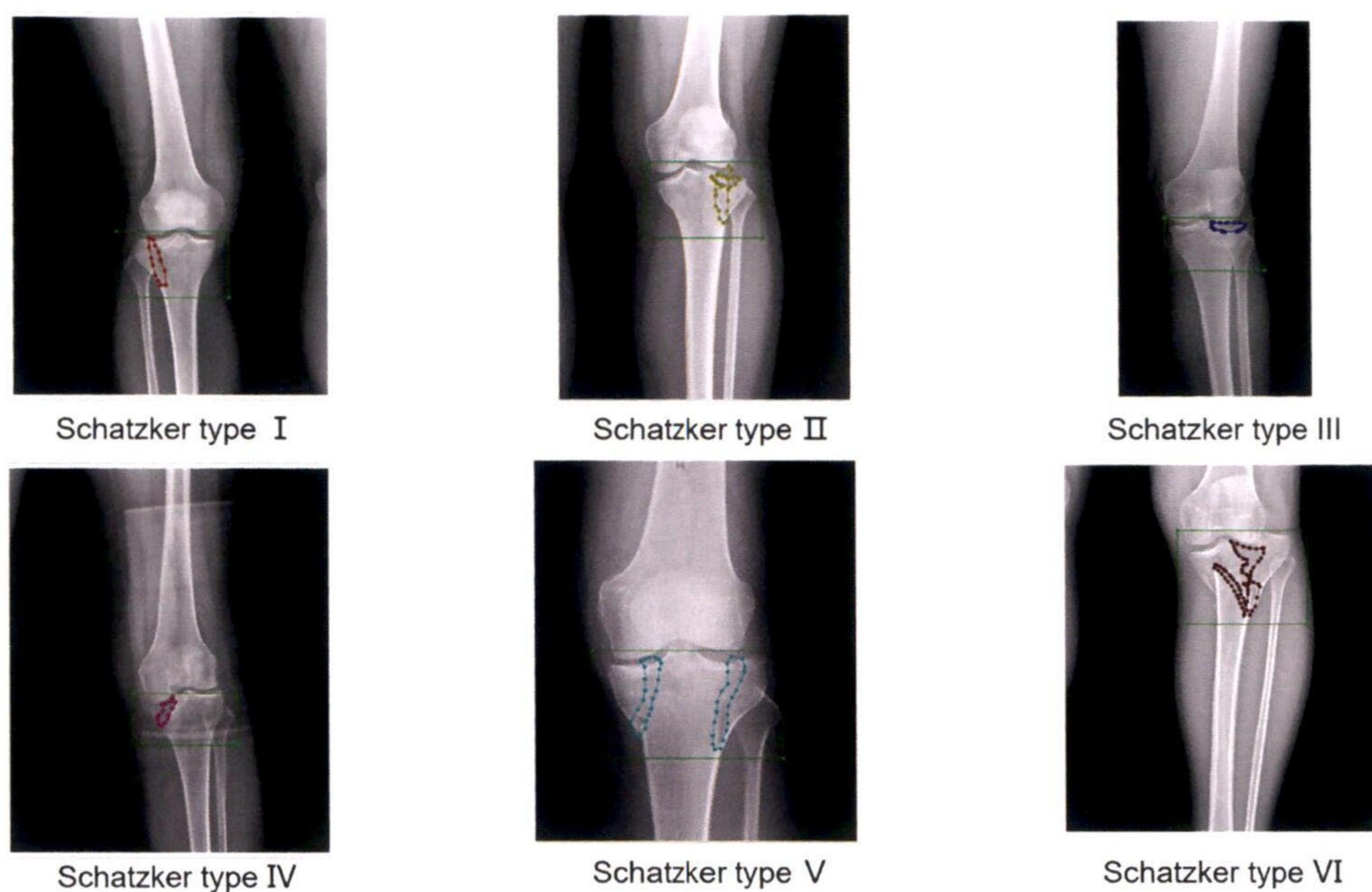


Fig. 1 The detailed annotation, classification and corresponding situation of the ROI according to the classic Schatzker classification of TPFs. The green rectangular box in the image delineates the location of the tibial plateau of the knee joint. The irregularly outlined boxes in various colors denote different categories according to the Schatzker classification: red indicates a split wedge fracture of the lateral tibial plateau; yellow indicates a split wedge depression fracture of

the lateral tibial plateau; blue indicates a pure depression fracture of the lateral tibial plateau; pink indicates a split wedge fracture of the medial tibial plateau; cyan indicates a bicondylar tibial plateau fracture with continuity between the epiphysis and diaphysis; brown indicates a bicondylar fracture with complete dissociation between the epiphysis and diaphysis

fracture type. To ensure the optimal performance of the YOLOv8n-cls model in the fracture type recognition, the following hyperparameters were set: a learning rate of 0.001, a batch size of 32, the optimizer Adam, and 100 training rounds [14, 19]. During the training process, the model employed the cross-entropy loss function to measure the classification error. The Adam optimizer was used to rapidly converge to the optimal solution through adaptive learning rate adjustment. The training dataset was divided into a training set, validation set, and test set at a ratio of 8:1:1, which were used for model training and performance evaluation, respectively. Model assessment was evaluated by accuracy, precision, recall, and F1-scores. With the above setup and training methods, the YOLOv8n-cls baseline model could provide a reliable performance benchmark for fracture type recognition [29, 30].

2.4 Segmentation-Guided Classification Method

To further increase the accuracy and robustness of fracture image classification, we proposed an innovative segmentation-guided classification approach to increase the accuracy and robustness of fracture classification. In segmented networks, previous investigations have demonstrated that U-Net, U-Net++, TransU-Net, and Swin-U-Net are prevalently utilized in the medical field [14–16, 31, 32]. On the

basis of the characteristics of the fracture images, we used SegNeXt as the fracture segmentation model. SegNeXt is an advanced segmentation network equipped with multiscale feature fusion and deep convolution technology that can effectively extract detailed information about the fracture area. It consists of multiple convolution layers and pooling layers, and multiscale feature extraction is performed through the feature pyramid structure to generate a binary segmentation map marking the fracture region (labelled 1) and the normal region (labelled 0). The evaluation metric is the m-Dice coefficient. During training, we employed the Dice loss function to maximize the overlap between the segmentation map and the real annotation.

SegNeXt exhibits excellent performance in various medical image segmentation tasks, and the model was selected due to its outstanding performance and extensive application scope [31]. To effectively incorporate segmentation information into the classification model, a dual-stream classification network was devised. First, the original fracture image and the segmented fracture image were extracted. We processed the original image and the segmented image via two independent convolution branches, each composed of multiple convolution layers and pooling layers, to extract multiscale image features. The original image features provided overall fracture image information, whereas the segmented image features provided detailed information

regarding the fracture area. When the features of the TPF fractures were extracted, the cross-attention mechanism was utilized for feature fusion. The mechanism employed one feature as the query (Q) and the other as the key (K) and value (V) to achieve the information interaction of different feature spaces. In our proposed method, the original image features were utilized as queries, and the segmented image features as keys and values. Simultaneously, the segmented image features were employed as queries, and the original image features as keys and values. Attention weights were obtained by computing the dot product of the query with the key, and these weights were applied to the values to generate the characteristics of the interaction. In the implementation process, we flattened the original image features and the segmented image features into two-dimensional matrices and then conducted dot product operations to calculate the attention weight. These weights were subsequently applied to the value matrix to generate interaction features. Finally, the interactive features of the two directions were weighted and averaged to generate the fused feature map. The advantage of the cross-attention mechanism lied in its ability to leverage the complementary information between different features fully to increase the richness of feature representation and the accuracy of classification.

After integration, we imported the feature maps into the convolution layer to further extract advanced features and then classified them through the fully connected layer to output the probability distributions of the fracture types. Compared with the original YOLOv8n-cls classification model,

we introduced segmentation information and a cross-attention mechanism to enhance the classification model's ability to perceive fracture details. The overall training process of the segmentation-guided classification method is shown in Fig. 2. This segmentation-guided classification method can effectively use segmentation information and improve the accuracy and robustness of fracture classification. The selection and training of the SegNeXt model ensures accurate segmentation of the fracture area and provides detailed feature information for subsequent classification. The cross-attention mechanism can enhance the discriminant ability of the classification model by deeply blending the features of the original image and the segmented image, especially in the processing of complex fracture images.

2.5 Diffusion Model Augmentation

Traditional data image augmentation employs a straightforward copy-and-paste approach to paste fracture line regions onto nonfracture images to generate a new fracture dataset [32]. Despite the fact that this method can rapidly acquire new fracture images and corresponding labels, the pasted fracture line area clearly stands out from the original image, and it is challenging to truly represent the actual fracture circumstances. To address this issue, diverse fusion methods, including alpha fusion and Poisson fusion, were used to merge the pasted fracture line with the original image [33]. Nevertheless, although these methods have enhanced the image quality to a certain extent, the generated image

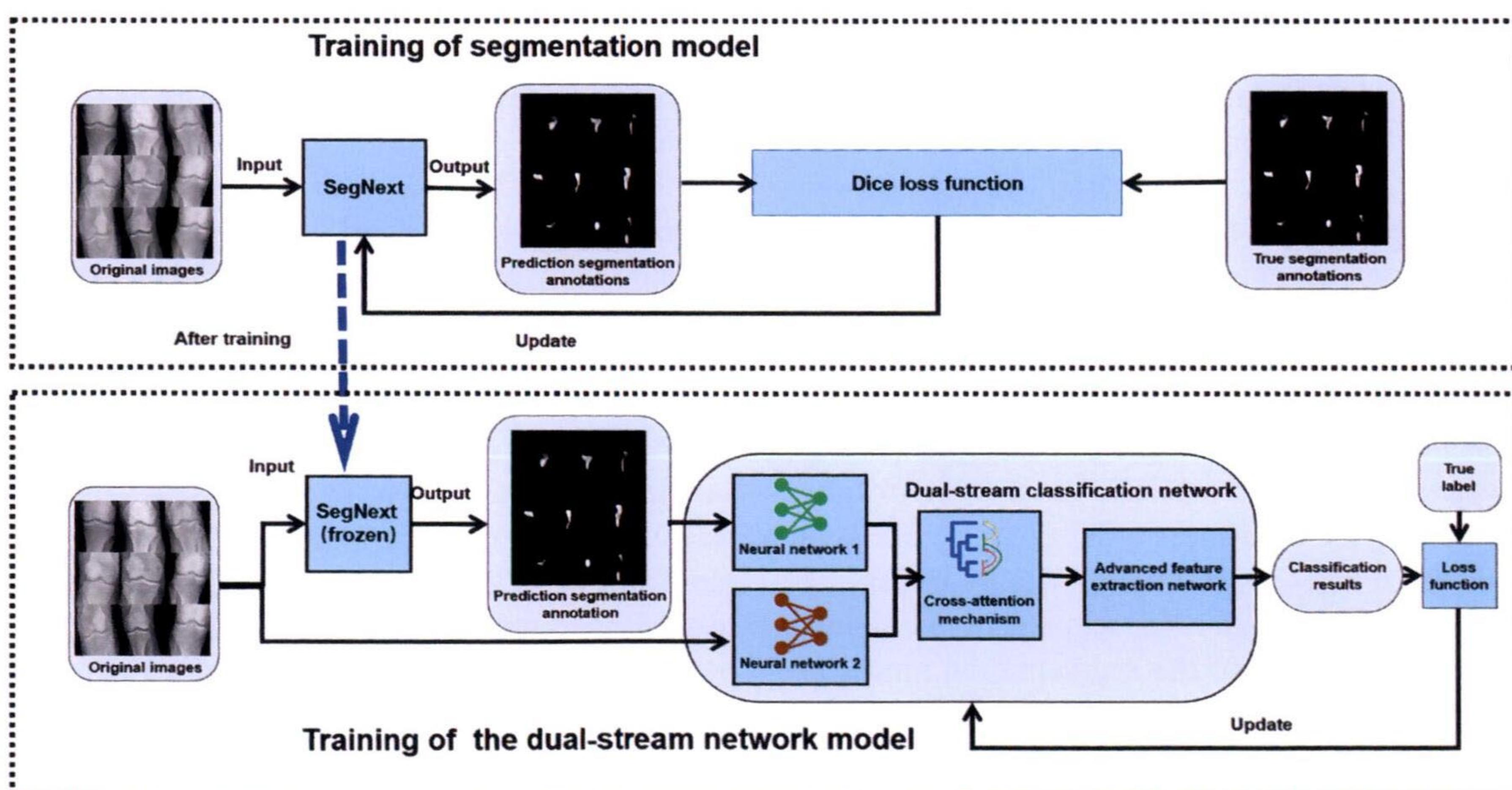


Fig. 2 Training process of the segmentation guide classification method. The cross-attention mechanism in this framework is an attention technique widely used in natural language processing (NLP) that allows models to process sequential data by focusing on differ-

ent locations in the sequence and calculating the correlations between them. This mechanism is particularly suitable for tasks in medical imaging that require understanding the relationships between the various elements in a sequence

still bears traces of artificiality. The simulated fracture image cannot effectively enhance the performance of the model when training the segmentation model and might even have an adverse effect on the learning effect of the model because of the unnatural features [34].

To achieve better data enhancement and model recognition, one approach is to collect a more comprehensive dataset with the goal of capturing a wider range of complexities and variations associated with the TPF. To address the issue of limited data annotation during the training of the segmentation model, we further incorporated the diffusion model for data augmentation. The diffusion model is a generative model that has demonstrated remarkable advantages in medical image generation and augmentation in recent years. We trained a diffusion model using all the fracture images and nonfracture images collectively for training. During the training process, the model learns to generate a new image similar to the distribution of the original image by gradually adding noise to it. After the training, we employed the diffusion model for image restoration. Specifically, for the fracture images initially generated through the copy-paste method, the diffusion model was utilized for inpainting the fracture line area [35, 36]. First, the pasted fracture image was input into the diffusion model, and the entire image was used as a reference to restore the fracture line area through inpainting technology, thereby making it more harmonious with the surrounding tissues. In this process, the diffusion model was gradually denoised, and the learned image distribution generated a natural and realistic fracture area. The fracture area labelling of these images was adjusted in combination with the original labelling to ensure the accurate positioning of the pasted fracture line labelling. The advantage of using the diffusion model for inpainting lies in its ability to leverage information from the entire image, making the transition between the fracture line area and the surrounding tissue more natural and avoiding the conspicuous problems caused by simple pasting methods. The fracture images generated by this approach are visually closer to real medical images and enhance the effectiveness of data augmentation. The use of the additional data generated by the diffusion model can not only increase the number of training samples but also increase the diversity of the samples, increasing the robustness of the model to different types of fracture images. Specifically, with an increasing number and diversity of datasets for segmentation training, the segmentation accuracy of the segmentation model in the fracture region has been significantly improved, providing higher-quality segmentation information for subsequent classification tasks and further enhancing the performance of the classification model [37].

The advantage of applying the diffusion model to inpainting lies in its ability to utilize the information from the entire image to render the transition between the fracture line area

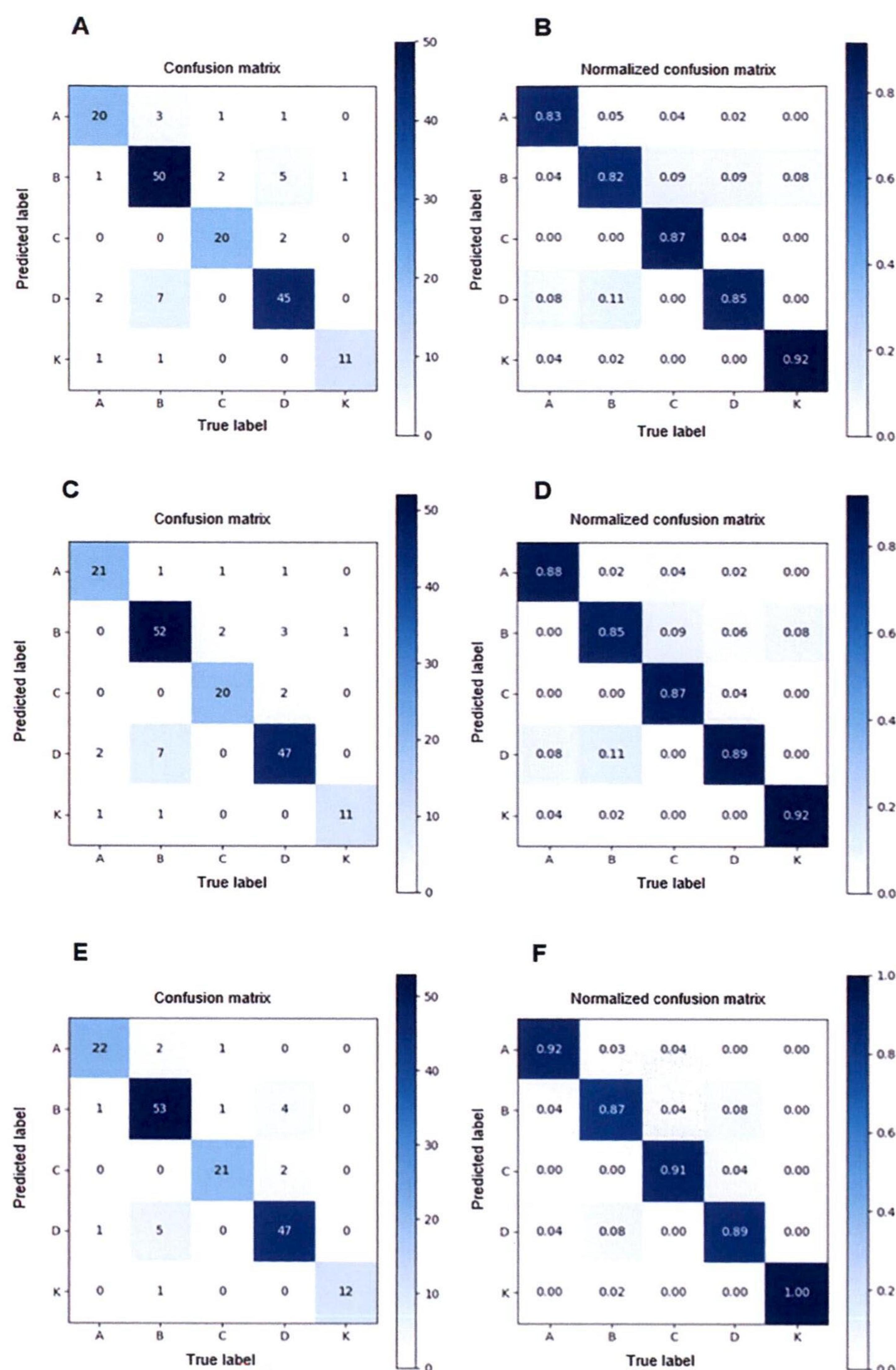
and the surrounding tissue more natural, thereby circumventing the obtrusive issues arising from simplistic pasting approaches. The fracture image produced by this method appears closer to the real medical image visually and enhances the efficacy of data augmentation [38, 39]. The utilization of the additional data generated by the diffusion model can not only increase the quantity of training samples but also increase the diversity of the samples, increasing the robustness of the model to various types of fracture images. Specifically, with an increasing number and diversity of datasets for segmentation training, the segmentation accuracy of the segmentation model in the fracture region has significantly increased, providing higher-quality segmentation information for subsequent classification tasks and further enhancing the performance of the classification model.

3 Results

3.1 Performance of the Baseline Model

In the initial experiment, we assessed the performance of YOLOv8n-cls as a baseline classification model. To comprehensively evaluate its performance in the fracture classification task, we examined the model via various metrics, such as accuracy, recall, and F1 scores. The experimental results indicated that YOLOv8n-cls performed well on standard datasets. However, its performance declined when dealing with complex fracture images, particularly in the recognition of fine fracture lines. The results for the YOLOv8n-cls strains were further analysed, as depicted in Fig. 3A and 3B. The model achieved an accuracy of 0.80, a recall of 0.833, and an F1 score of 0.816 for class A; an accuracy of 0.847, a recall of 0.820, an F1 score of 0.833 for class B; an accuracy of 0.910, a recall of 0.870, an F1 score of 0.889 for class C; an accuracy of 0.833, a recall of 0.850, an F1 score of 0.841 for class D; and an accuracy of 0.846, a recall of 0.917, and an F1 score of 0.880 for class K. We also generated Receiver Operating Characteristic (ROC) curves for different classifications via YOLOv8n-cls and calculated their corresponding under the curve (AUC) values (Fig. 4A). The AUC values for classes A to K are 0.95, 0.91, 0.97, 0.92, and 0.95, respectively. The overall accuracy of all categories was 0.844, the macro-F1 score was 0.852, the macro-AUC score was 0.94, and the micro-AUC score was 0.94. These results suggested that although YOLOv8n-cls holds certain potential in fracture classification tasks, there is still scope for improvement, particularly in the processing of complex and diverse fracture images.

Fig. 3 Large icons **A** and **B** represent the confusion matrix and normalized confusion matrix results, respectively; **C** and **D** represent the confusion matrix and normalized confusion matrix results of segmentation-guided classification; and **E** and **F** represent the confusion matrix and normalized confusion matrix results further enhanced by diffusion model augmentation. The letters A, B, C, D, and K in the horizontal and vertical coordinates represent the classification types of the TPF (A: Schatzker type I; B: Schatzker type II and III; C: Schatzker type IV; D: Schatzker type V and VI; K: intercondylar ridge fractures)



3.2 Performance of Segmentation-Guided Classification

As an essential approach for facilitating clinical decision-making, medical image segmentation offers a significant reference for precision medicine. Image segmentation of the fracture site can direct the model to perform classification. Through algorithm screening and model validation, we demonstrated the feasibility of segmentation-guided

classification and assessed the performance of segmentation-guided classification. We adopted SegNeXt as the segmentation model, superimposed the segmentation map on the original image, and employed the enhanced classification model for training. The performance of the segmentation model was evaluated via mDice indicators, where 0 indicates normal areas and 1 represents fracture areas. The mDice value of the SegNeXt model reached 0.792 on the data not augmented by the diffusion model, suggesting that

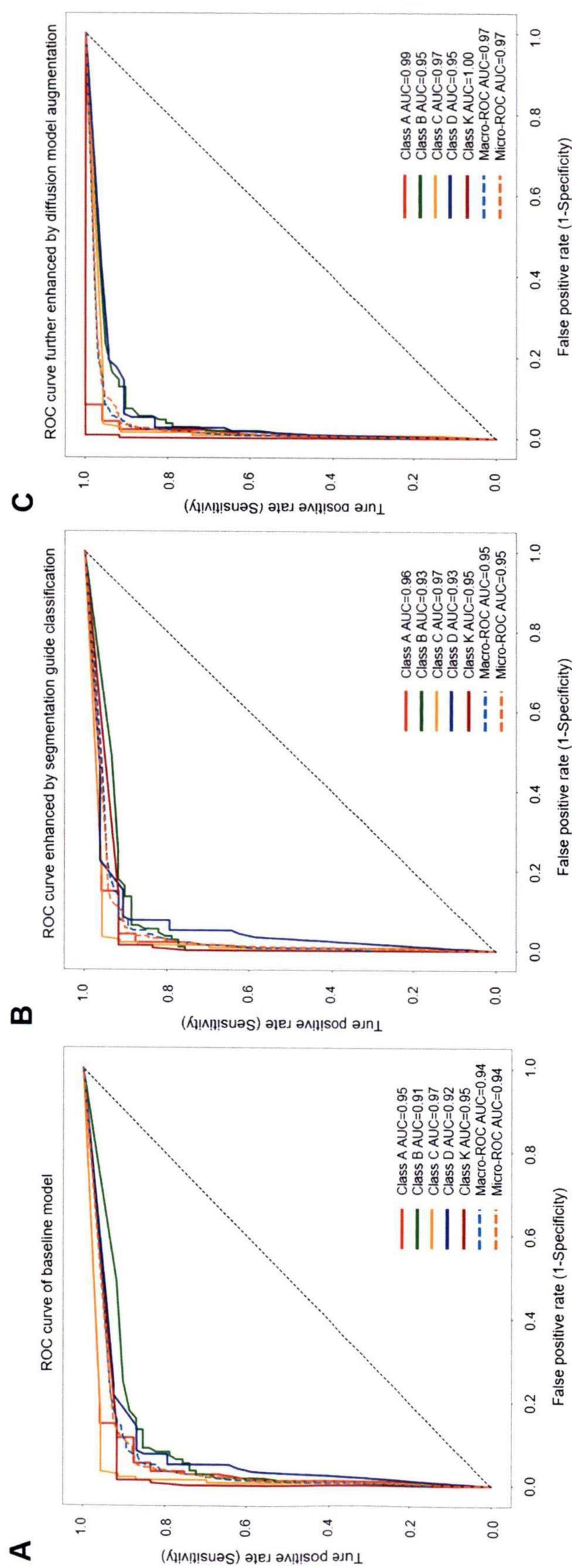


Fig. 4 A, B, and C represent the receiver operating characteristic (ROC) curves of the baseline model, the ROC curve enhanced by segmentation-guided classification, and the ROC curve further enhanced by diffusion model augmentation, respectively. Class A: Schatzker type I, class B: Schatzker type II and type III, class C: Schatzker type IV, class D: Schatzker type V and type VI, class K: intercondylar ridge fractures

the model could segment the fracture area more precisely. In the segmentation-guided classification method, we fused the SegNeXt segmentation results with the original image and utilized the cross-attention mechanism to effectively integrate the segmentation information and original image information. After training, the performance of the improved classification model in the task of fracture classification was notably enhanced, and the specific results were presented. The results of segmentation-guided classification were further analysed, as depicted in Fig. 3C and 3D. The model achieved an accuracy of 0.875, a recall of 0.875, an F1 score of 0.875 for class A; an accuracy of 0.897, a recall of 0.852, an F1 score of 0.874 for class B; an accuracy of 0.910, a recall of 0.870, an F1 score of 0.889 for class C; an accuracy of 0.839, a recall of 0.887, an F1 score of 0.862 for class D; an accuracy of 0.846, a recall of 0.917, and an F1 score of 0.880 for class K. We also generated ROC curves for different classifications and calculated their corresponding AUC values (Fig. 4B). The AUC values for classes A to K are 0.96, 0.93, 0.97, 0.93, and 0.95, respectively. The overall accuracy of all categories was 0.873, the macro-F1 score was 0.876, the macro-AUC was 0.95, and the micro-AUC was 0.95. The segmentation-guided classification method significantly enhanced all the evaluation indicators, particularly the recall rate and F1 score (with specific values provided for these indicators). This finding indicated that by integrating the segmentation information, the accuracy and robustness of our model in the fracture classification task were significantly improved.

3.3 Diffusion Model Augmentation Effect

To achieve more precise fracture classification, we assessed the influence of data augmentation through diffusion models on the performance of the segmentation and classification models. The supplementary data generated by the diffusion model increased the quantity and diversity of training samples, increasing the robustness of the model to various types of fracture images. We contrasted the performance of the segmentation model before and after data augmentation, and the findings indicated that the mean Dice coefficient (mDice) value of the SegNeXt segmentation model increased from 0.792 to 0.824 after the data were augmented with the diffusion model, which significantly enhanced the performance of the segmentation model. Additionally, we compared the performance of the segmentation-guided classification model before and after augmentation. The results of the YOLOv8n-cl model were further analysed, as depicted in Fig. 3E and 3F. The model achieved an accuracy of 0.880, a recall of 0.917, an F1 score of 0.898 for class A; an accuracy of 0.898, a recall of 0.869, an F1 score of 0.883 for class B; an accuracy of 0.913, a recall of 0.913, an F1 score of

0.913 for class C; an accuracy of 0.887, a recall of 0.887, an F1 score of 0.887 for class D; an accuracy of 0.923, a recall of 1.00, and an F1 score of 0.960 for class K. We generated ROC curves for different classifications and calculated their corresponding AUC values (Fig. 4C). The AUC values for classes A to K are 0.99, 0.95, 0.97, 0.95, and 1.0, respectively. The segmentation-guided classification model trained with augmented data demonstrates outstanding performance across all the evaluation indicators, with an accuracy rate of 0.952 and an F1 score of 0.941. The segmentation-guided classification model trained with augmented data performed excellently on all the evaluation indices, and the overall accuracy of all the categories was 0.896, the macro-F1 score was 0.908, the macro-AUC score was 0.97, and the micro-AUC score was 0.97. These results demonstrated that through the introduction of a diffusion model for data augmentation, we not only efficaciously enhanced the accuracy of the segmentation-guided classification model but also further augmented the robustness of the model. The augmentation approach holds extensive application potential in medical image processing and can potentially help improve model performance by enriching the minority class with variations on the basis of the extracted features, particularly in scenarios of scarce data.

3.4 Comparison of the Three-Stage Algorithms

Through systematic enhancements to the baseline model, the performance of our proposed dual-stream attention-based classification network was rigorously validated. Through data analysis, the accuracy increased from 0.844 to 0.896, the macro-F1 score increased from 0.852 to 0.908, and the macro-AUC and micro-AUC both increased from 0.94 to 0.97. The overall efficiency of the model was significantly improved. The performance of classification networks based on dual-stream attention improved significantly, and this enhancement was consistently observed in different metrics. This consistency accounted for substantial advancements in the dual-stream model across several key performance indicators. The numerical differences between the various methods were sufficiently pronounced. For example, the classification accuracy improved significantly from 84.4% in the benchmark model to 89.6% following the implementation of the dual-stream model and data augmentation, resulting in a gap exceeding 5%. This notable increase indicated that these enhancements were not merely coincidental. In terms of overall improvements across key performance indicators, the dual-stream model demonstrated significant gains, particularly after incorporating data augmentation; these metrics exhibited further enhancement. This underscores the comprehensive advantages of the dual-stream model across

different dimensions and highlights its superiority from multiple perspectives. Given that consistent and significant improvements were observed in various indicators with the dual-stream model, these clear numerical distinctions suggested that it outperformed the benchmark model.

4 Discussion

In the diagnosis and treatment of traumatic orthopedics, AI has demonstrated excellent application prospects in the diagnosis and classification of limb fractures, image segmentation of complex pelvic fractures, and spinal trauma diagnosis via deep learning of key site characteristics [40, 41]. However, owing to the difficulty of accessing medical data and the associated privacy and security risk factors, medical models are still useful for practical clinical applications. In the application of precise segmentation algorithms for medical images, the scheme put forward in prior studies has been utilized primarily in research on the scaphoid bone of the hand, the complex pelvis, epiphyseal fractures of the limbs, and calcaneal fractures [15–17, 21]. In this study, we proposed and validated a segmentation-guided classification approach and diffusion model augmentation strategy to increase the robustness of the AI model. By integrating segmentation and classification techniques, we overcome the limitations of traditional methods in processing complex fracture images and alleviate the scarcity of medical image data by generating additional training data via diffusion models. In addition, this method has certain generalizability and holds significant value in clinical applications. Segmentation-guided classification methods and diffusion model augmentation strategies can also be applied to tumor detection and organ segmentation, thereby promoting the advancement of the entire field of medical image analysis. In light of the challenges of insufficient data volume and heightened privacy risks, the proposition of our technical framework will pave the way for the clinical application of AI-assisted analysis of musculoskeletal images.

The segmentation-guided classification approach enhances the classification accuracy by incorporating the segmentation outcomes of fracture images as supplementary information into the classification model [15–18]. Our method adopted a two-stream network architecture and cross-attention mechanism to integrate the segmentation information with the original image information. One branch of the two-stream network handled the original fracture image, whereas the other processed the segmented fracture image. At the appropriate layer of the network, we introduced the cross-attention mechanism, where one side offers queries and the other provides keys and values to interact and fuse the two kinds of information. The

design enables the network to exploit the segmentation information more effectively and improve the capacity to identify fracture types. The experimental results demonstrated that, in contrast to the conventional baseline classification approach, the segmentation-guided classification method significantly enhanced the classification indices of fractures. The SegNeXt architecture was employed in the segmentation model, and its multiscale feature extraction capacity and efficient convolution operation enabled it to extract fracture lines and other particulars from fracture images of diverse scales and complexities effectively. By integrating these particulars with the original image, the classification model can refer to more beneficial information during the decision-making process, thereby improving the classification accuracy. Specifically, the segmentation-guided classification method outperformed the baseline classification model in terms of accuracy, precision, recall, and F1 score. This finding indicated that the classification model could identify fracture types more precisely with the introduction of segmentation information, particularly in complex and detailed images. The SegNeXt model exploits its potent multiscale feature extraction capabilities to ensure accurate segmentation of the fracture region, thereby providing higher-quality features for subsequent classification.

The diffusion model enriches the training dataset of the segmentation model by generating additional training data, particularly in the fracture region. Although the traditional copy-paste approach can generate new fracture images, it has constraints in dealing with the fusion of the pasted area with the surrounding tissue. We employed a diffusion model for inpainting the pasted fracture area to guarantee that the resulting fracture image was more visually natural and realistic [19, 20]. After training, we utilized the diffusion model to restore the image. For the fracture image initially produced via the copy-paste method, the diffusion model was adopted to inpaint the fracture line area. During this process, the diffusion model is gradually denoised, and the learned image distribution generates a naturally realistic fracture area. The experimental results demonstrated that the diffusion model augmentation strategy significantly enhanced the performance of the segmentation model. A comparison of the performance of the segmentation model before and after augmentation revealed that the data generated by the diffusion model significantly improved the performance of the segmentation model in terms of the mDice index [21–23]. This not only validates the superiority of the diffusion model in generating medical image data but also proves its efficacy in enhancing the performance of segmentation models. The use of the diffusion model for data augmentation not only augments the diversity of the training data but also enhances the model's ability to identify the fracture region, thereby improving the classification performance.

This pioneering study proposes the utilization of medical image segmentation and diffusion model augmentation for fracture classification and enhances the quality of segmentation model-guided classification tasks while increasing data diversity. The classification of fractures is conducive to the formulation of surgical plans. The proposed model significantly shortens the time of diagnosis reports and enhances the working efficiency of imaging and orthopedic doctors.

Although this study demonstrates a novel approach to enhance the classification efficiency of AI models in the process of performing TPF, several limitations need to be noted. First, although our dataset is considerable in size, the diversity and complexity of the data could still pose obstacles to the generalizability of the AI model in real-world clinical applications. More TPF and normal tibial radiographs from different medical centers should be used as additional external datasets to thoroughly evaluate the performance of our model. It is necessary to extend the dataset from an external institution to increase the generalizability and robustness of the approach and make the model applicable to different clinical settings. Second, despite the commendable performance of the strategy in this study, further validation and optimization are needed in clinical applications. Additionally, the model architecture and training strategy adopted in this paper have scope for further enhancement, such as the introduction of more attention mechanisms and the improvement of training methods to further increase the performance of the model. While we have enhanced the diversity of training data through diffusion model augmentation strategies, these synthetic data might not be capable of fully substituting the complexity of real clinical data. In practical applications, the model might encounter images with a distribution that differs from the training data, thereby influencing the classification performance. Furthermore, although the segmentation-guided classification method yields favourable results in experiments, its complexity increases the computational cost and deployment difficulty of the model, and a balance between performance and resource requirements needs to be struck in practical applications. From an academic perspective, our current study merely utilizes X-ray data and lacks the integration of multimodal data. Despite the relatively facile acquisition of this type of image data, there are certain inaccuracies in the identification of fractures behind the TPF and in the middle of the platform. Further collection of more CT and MRI data from knee joint sites is necessary to optimize the TPF diagnostic model more comprehensively [2, 4].

Future research can be enhanced and expanded in the following aspects. First, more diversified datasets could be used to further enhance the generalizability and robustness of the model. For example, the utilization of multimodal data such as CT and MRI data could be increased to broaden the identification of latent fractures and chronic injuries [25, 42].

Second, more advanced architectures of segmentation and classification models can be explored, such as the introduction of new models such as transformers, to further increase classification performance. Additionally, more medical field knowledge and expert annotations can be combined to optimize data augmentation and model training strategies to increase the feasibility and accuracy of the model. Finally, the application and extension of the model to other medical image analysis tasks can be probed to verify its applicability and validity in different tasks. In addition, more controlled clinical trials need to be carried out in multicenter hospitals, and diagnostic comparison studies between AI and experts can be explored to determine the stability of the system and algorithm.

In conclusion, the strategy presented in this study demonstrates outstanding performance in enhancing the robustness of the AI model in fracture image classification and holds significant clinical application value and potential. Through the continuous optimization and expansion of these methods, this study offers a reference for future research. The proposed AI model will be of great value in further promoting the development of imaging omics in the field of orthopedic trauma, providing physicians with a more reliable and efficient tool for clinical practice. This method is capable of generating synthetic images that appear realistic along with accurate segmentation labels. The implications of this advancement are considerable, as it has the potential to significantly increase the accuracy and robustness of deep learning-based models for the diagnosis of fractures in the tibial plateau, which will result in more reliable and effective diagnostic tools and benefit both medical professionals and patients in trauma surgery.

Authors' Contributions All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Funding This work was supported by the National Natural Science Foundation of China (Nos. 81974355 and 82172524), Key Research and Development Program of Hubei Province (No. 2021BEA161), National Innovation Platform Development Program (No. 2020021105012440), Open Project Funding of the Hubei Key Laboratory of Big Data Intelligent Analysis and Application, Hubei University (No. 2024BDIAA03), and Free Innovation Preliminary Research Fund of Wuhan Union Hospital (No. 2024XHYN047).

Data Availability Not applicable.

Declarations

Conflict of Interest All authors declare that there are no competing financial interests.

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

1. Khan K, Mushtaq M, Rashid M, et al. Management of tibial plateau fractures: a fresh review. *Acta Orthop Belg.* 2023;89(2):265–273.
2. Xie X, Zhan Y, Wang Y, et al. Comparative Analysis of Mechanism-Associated 3-Dimensional Tibial Plateau Fracture Patterns. *J Bone Joint Surg Am.* 2020;102(5):410–418.
3. Kfuri M, Schatzker J. Revisiting the Schatzker classification of tibial plateau fractures. *Injury.* 2018;49(12):2252–2263.
4. Liu PR, Zhang JY, Xue MD, et al. Artificial Intelligence to Diagnose Tibial Plateau Fractures: An Intelligent Assistant for Orthopedic Physicians. *Curr Med Sci.* 2021;41(6):1158–1164.
5. Hill BG, Krogue JD, Jevsevar DS, et al. Deep Learning and Imaging for the Orthopaedic Surgeon: How Machines "Read" Radiographs. *J Bone Joint Surg Am.* 2022;104(18):1675–1686.
6. Yoon AP, Lee YL, Kane RL, et al. Development and Validation of a Deep Learning Model Using Convolutional Neural Networks to Identify Scaphoid Fractures in Radiographs. *JAMA Netw Open.* 2021;4(5):e216096.
7. Senanayake D, Seneviratne S, Imani M, et al. Classification of Fracture Risk in Fallers Using Dual-Energy X-Ray Absorptiometry (DXA) Images and Deep Learning-Based Feature Extraction. *JBMR Plus.* 2023;7(12):e10828.
8. Xu F, Xiong Y, Ye G, et al. Deep learning-based artificial intelligence model for classification of vertebral compression fractures: A multicenter diagnostic study. *Front Endocrinol (Lausanne).* 2023;14:1025749.
9. Liu XS, Nie R, Duan AW, et al. YOLOX-SwinT algorithm improves the accuracy of AO/OTA classification of intertrochanteric fractures by orthopedic trauma surgeons. *Chin J Traumatol.* 2024; 23:1087–1092.
10. Zhang J, Xia L, Liu J, et al. Exploring deep learning radiomics for classifying osteoporotic vertebral fractures in X-ray images. *Front Endocrinol (Lausanne).* 2024;15:1370838.
11. Ono Y, Suzuki N, Sakano R, et al. A Deep Learning-Based Model for Classifying Osteoporotic Lumbar Vertebral Fractures on Radiographs: A Retrospective Model Development and Validation Study. *J Imaging.* 2023;9(9):187.
12. Chung SW, Han SS, Lee JW, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop.* 2018;89(4):468–473.
13. Mutasa S, Varada S, Goel A, et al. Advanced deep learning techniques applied to automated femoral neck fracture detection and classification. *J Digit Imaging.* 2020;33(5):1209–1217.
14. Wang H, Cao P, Yang J, et al. MCA-UNet: multi-scale cross co-attentional U-Net for automatic medical image segmentation. *Health Inf Sci Syst.* 2023;11(1):10.
15. Chen C, Liu B, Zhou K, et al. CSR-Net: Cross-Scale Residual Network for multi-objective scaphoid fracture segmentation. *Comput Biol Med.* 2021;137:104776.
16. Guo J, Mu Y, Xue D, et al. Automatic analysis system of calcaneus radiograph: Rotation-invariant landmark detection for calcaneal angle measurement, fracture identification and fracture region segmentation. *Comput Methods Programs Biomed.* 2021;206:106124.
17. Zeng B, Wang H, Joskowicz L, et al. Fragment distance-guided dual-stream learning for automatic pelvic fracture segmentation. *Comput Med Imaging Graph.* 2024;116:102412.
18. Huang Y, Yang X, Liu L, et al. Segment anything model for medical images? *Med Image Anal.* 2024;92:103061.
19. Chen Y, Yang XH, Wei Z, et al. Generative Adversarial Networks in Medical Image augmentation: A review. *Comput Biol Med.* 2022;144:105382.
20. Kebaili A, Lapuyade-Lahorgue J, Ruan S. Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review. *J Imaging.* 2023;9(4):81.
21. Wu S, Kurugol S, Tsai A. Improving the radiographic image analysis of the classic metaphyseal lesion via conditional diffusion models. *Med Image Anal.* 2024;97:103284.
22. Kazerouni A, Aghdam EK, Heidari M, et al. Diffusion models in medical imaging: A comprehensive survey. *Med Image Anal.* 2023;88:102846.
23. Hao R, Namdar K, Liu L, et al. A Comprehensive Study of Data Augmentation Strategies for Prostate Cancer Detection in Diffusion-Weighted MRI Using Convolutional Neural Networks. *J Digit Imaging.* 2021;34(4):862–876.
24. Abedeen I, Rahman MA, Prottyasha FZ, et al. FracAtlas: A Dataset for Fracture Classification, Localization and Segmentation of Musculoskeletal Radiographs. *Sci Data.* 2023;10(1):521.
25. Yao X, Zhou K, Lv B, et al. 3D mapping and classification of tibial plateau fractures. *Bone Joint Res.* 2020;9(6):258–267.
26. Aly GH, Marey M, El-Sayed SA, et al. YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms. *Comput Methods Programs Biomed.* 2021;200:105823.
27. Jeon YD, Kang MJ, Kuh SU, et al. Deep Learning Model Based on You Only Look Once Algorithm for Detection and Visualization of Fracture Areas in Three-Dimensional Skeletal Images. *Diagnostics (Basel).* 2023;14(1):11.
28. Ju RY, Cai W. Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm. *Sci Rep.* 2023;13(1):20077.
29. Cheng J, Tian S, Yu L, et al. Fully convolutional attention network for biomedical image segmentation. *Artif Intell Med.* 2020;107:101899.
30. Zhuo M, Chen X, Guo J, et al. Deep Learning-Based Segmentation and Risk Stratification for Gastrointestinal Stromal Tumors in Transabdominal Ultrasound Imaging. *J Ultrasound Med.* 2024;43(9):1661–1672.
31. Das N, Das S. Attention-UNet architectures with pretrained backbones for multi-class cardiac MR image segmentation. *Curr Probl Cardiol.* 2024;49(1):102129.
32. Bian X, Wang G, Wu Y, et al. TCI-UNet: transformer-CNN interactive module for medical image segmentation. *Biomed Opt Express.* 2023;14(11):5904–5920.
33. Xu Y, Gong M, Xie S, et al. Semi-Implicit Denoising Diffusion Models (SIDDMs). *Adv Neural Inf Process Syst.* 2023;36:17383–17394.
34. Hung ALY, Zhao K, Zheng H, et al. Med-cDiff: Conditional Medical Image Generation with Diffusion Models. *Bioengineering (Basel).* 2023;10(11):1258.
35. Daher R, Vasconcelos F, Stoyanov D. A Temporal Learning Approach to Inpainting Endoscopic Specularities and Its Effect on Image Correspondence. *Med Image Anal.* 2023;90:102994.
36. Xiao L, Wu J. Image inpainting algorithm based on double curvature-driven diffusion model with P-Laplace operator. *PLoS One.* 2024;19(7):e0305470.
37. Leñena J, Saiz FA, Barandiaran I. Latent Diffusion Models to Enhance the Performance of Visual Defect Segmentation Networks in Steel Surface Inspection. *Sensors (Basel).* 2024;24(18):6016.
38. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science.* 2006;313(5786):504–507.
39. Lin J, Xie X, Wu W, et al. Model transfer from 2D to 3D study for boxing pose estimation. *Front Neurorobot.* 2023;17:1148545.
40. Kuo RYL, Harrison C, Curran TA, et al. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology.* 2022;304(1):50–62.
41. Oakden-Rayner L, Gale W, Bonham TA, et al. Validation and algorithmic audit of a deep learning system for the detection of proximal femoral fractures in patients in the emergency department: a diagnostic accuracy study. *Lancet Digit Health.* 2022;4(5):e351–e358.

42. Qiu Z, Xie Z, Lin H, et al. Learning co-plane attention across MRI sequences for diagnosing twelve types of knee abnormalities. *Nat Communication.* 2024;15:7637.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.