Magnetic Resonance Imaging Diagnosis of Knee Injuries after Skiing in Adolescents under Deep Learning

Wei Xu[1], Songmei Li[2,*], Guofeng Zhang[3], Qi Zhang[4], Weidong Song[4]

[1]Graduate School, Harbin Sport University, Harbin, 150008, China; 279521603@qq.com

[2]Heilongjiang Ice and Snow Industry Research Institute, Harbin Sport University, Harbin, 150008, China; htxiayuxue2000@126.com

[3]Sports and healthy science college, Mudanjiang Normal University, Mudanjiang, 157011, China; pashiren4795@163.com

[4]The Second Affiliated Hospital, Mudanjiang Medical University, Mudanjiang, 157011, China; chenzhaochao26@163.com; 201431050134@chd.edu.cn

*Corresponding author: Songmei Li, e-mail: htxiayuxue2000@126.com

Abstract: To address the issues of subjectivity, low efficiency, and missed diagnoses of subtle injuries in magnetic resonance imaging (MRI) diagnosis of knee injuries after skiing in adolescents, a precise automatic diagnostic model was constructed to achieve simultaneous segmentation and classification of different types of injuries. A hybrid model combining U-Net++ and DenseNet121 was employed. U-Net++ performed pixel-level segmentation of injured areas, while DenseNet121 classified the injuries based on fused features. The model was trained using a dataset of 309 adolescent MRI scans through a joint loss function and transfer learning strategy. In the segmentation task, the average Dice coefficient (DC) was 0.89, and the intersection over union (IoU) was 0.82. The highest accuracy was achieved for meniscus tears (0.93, 0.87) and anterior cruciate ligament injuries (0.89, 0.82). Cartilage injuries (0.84, 0.77) showed a 6.4% improvement compared to the original U-Net. In the classification task, the average accuracy was 0.90, the F1-score was 0.91, and the area under the ROC curve (AUC) was 0.95. The recall rate for meniscus tears was 0.93, with a precision of 0.94, and the recall rate for cartilage injuries was 0.87. These results were significantly higher than those of SVM + handcrafted features (F1=0.77) and ResNet50 (F1=0.85) ($P < 0.01$). The model can efficiently and accurately perform automatic diagnosis of multiple injury types on MRI scans, outperforming traditional methods. It reduces the rate of missed and incorrect diagnoses, improves diagnostic consistency and efficiency, and holds value for clinical auxiliary applications.

Keywords: skiing; knee injuries; magnetic resonance imaging diagnosis; deep learning

1. Introduction

With the increasing popularity of skiing among adolescents, the risk of related sports injuries has significantly risen, with knee injuries being the most common [1,2]. Common types include anterior cruciate ligament injuries, meniscus tears, cartilage injuries, and medial collateral ligament strains. These injuries not only affect adolescents' sports performance but may also have adverse long-term effects on the musculoskeletal system and potentially hinder the continuation of their sports careers [3,4].

In clinical diagnosis, magnetic resonance imaging (MRI) is a core imaging modality for assessing knee injuries. However, traditional diagnostic methods heavily rely on radiologists' experience, presenting several limitations: diagnostic results are easily influenced by subjective factors, leading to poor consistency among doctors with different levels of experience [5,6]; manual analysis of MRI sequences is time-consuming and inefficient for rapid assessment of sports injuries; subtle injuries, such as microscopic cartilage defects, are prone to being missed in traditional visual interpretation [7,8].

To address these clinical challenges, breakthroughs in deep learning technology in the field of medical image analysis offer a new solution. Deep learning can automatically extract deep features, precisely localize lesions, and efficiently process large amounts of data, effectively compensating for the shortcomings of manual diagnosis [9,10]. In this study, a hybrid model combining U-Net++ and DenseNet121 was selected. U-Net++, as the segmentation network, captures spatial features of subtle injuries through densely connected decoders and deep supervision mechanisms, achieving pixel-level localization of injured areas [11,12]. DenseNet121, as the classification network, relies on feature reuse mechanisms in dense blocks to fully integrate multi-scale semantic information and improve the accuracy of identifying multiple injury types. The combination of these two models aims to construct an efficient MRI image analysis model to achieve automatic diagnosis of multiple knee injuries after skiing in adolescents [13,14], enhancing diagnostic accuracy, consistency, and efficiency, and providing a reliable auxiliary diagnostic tool for clinical practice.

2. Literature Review

2.1 Characteristics of Knee Injuries in Adolescent Skiers

Adolescent skiers have a high incidence of knee injuries, often leading to interruptions in their sports careers. A four-year longitudinal study by Hanimann et al. (2025) [15] found that approximately one-third of competitive alpine skiers aged 15-19 years experienced traumatic knee injuries, with anterior cruciate ligament injuries accounting for nearly half of these cases, and female athletes being at higher risk. The study also identified that a smaller knee flexion angle during landing (i.e., a "stiff landing" strategy) is a significant biomechanical risk factor for anterior cruciate ligament injuries in female athletes.

Different skiing disciplines exhibit varying patterns of knee injuries. Kastner

et al. (2023) [16] reported that in the 2023 Nordic World Ski Championships, knee injuries accounted for 50% of all reported lower limb injuries, with sprains/ligament tears being the most common type. This highlights the high incidence of knee injuries across multiple skiing disciplines, warranting special attention.

2.2 Current Status of MRI Diagnosis and Deep Learning Applications for Knee Injuries

Deep learning technology has shown great potential in the auxiliary diagnosis of knee injuries using MRI. A systematic review by Mead et al. (2025) [17] indicated that models based on convolutional neural networks (e.g., ResNet, VGG) performed exceptionally well in detecting specific lesions such as anterior cruciate ligament injuries, osteoarthritis, and meniscus tears, with accuracy rates approximately 4.5% higher than those of general abnormality detection models. However, the authors also emphasized that current model performance is still limited by the scale and diversity of training data.

To further enhance diagnostic performance, Sun et al. (2025) [18] developed the KneeXNet model, which innovatively combined graph convolutional networks with multi-scale feature fusion technology. This model achieved area under the ROC curve (AUC) values of 0.972 for anterior cruciate ligament injuries and 0.968 for meniscus tears, significantly outperforming traditional deep learning methods. The study also enhanced the transparency of the model's decision-making process through Grad-CAM visualization technology, paving the way for clinical translation.

2.3 Research Gaps in the Diagnosis of Adolescent Sports Injury

There is a significant gap in imaging diagnosis research for adolescent sports injuries. Cerezal et al. (2025) [19] pointed out that the adolescent skeletal system is not fully developed, and sports injuries in this population differ significantly from those in adults in terms of location and characteristics. Despite the continuous increase in adolescent sports participation (especially among female athletes), customized diagnostic solutions for this group remain to be perfected.

Gicquel (2024) [20] further emphasized that in the diagnosis and treatment of knee ligament and meniscus injuries, the age, anatomical features, and injury mechanisms of adolescent patients must be fully considered. Although MRI plays a key role in diagnosis, most existing deep learning models are developed based on adult data and lack targeted optimization for the characteristics of adolescent skeletal development. This limits their accuracy and applicability in diagnosing adolescent skiing injuries, representing an important research gap.

3. Research Methods

3.1 Overall Framework Design of the Model

To meet the core needs of MRI diagnosis for knee injuries in adolescent skiers, this study proposes a hybrid deep learning model combining U-Net++ and

DenseNet121, which achieves joint diagnosis of fine segmentation of injured areas and precise classification of injury types. The overall framework of the model consists of two core modules, which are designed in series to complete the full automation of analysis from MRI image input to diagnostic result output (Figure 1). The design of the model first uses U-Net++ to complete pixel-level localization of injured areas, providing accurate region of interest (ROI) features for the classification task and avoiding interference from background noise. Secondly, the injury region mask output by the segmentation branch is integrated with the raw MRI scan as the input for the classification branch, achieving collaborative optimization of spatial location features and semantic category features. Finally, the segmentation and classification tasks share the underlying feature extraction network, and parameters are updated synchronously through a joint loss function to enhance the overall generalizability of the model.

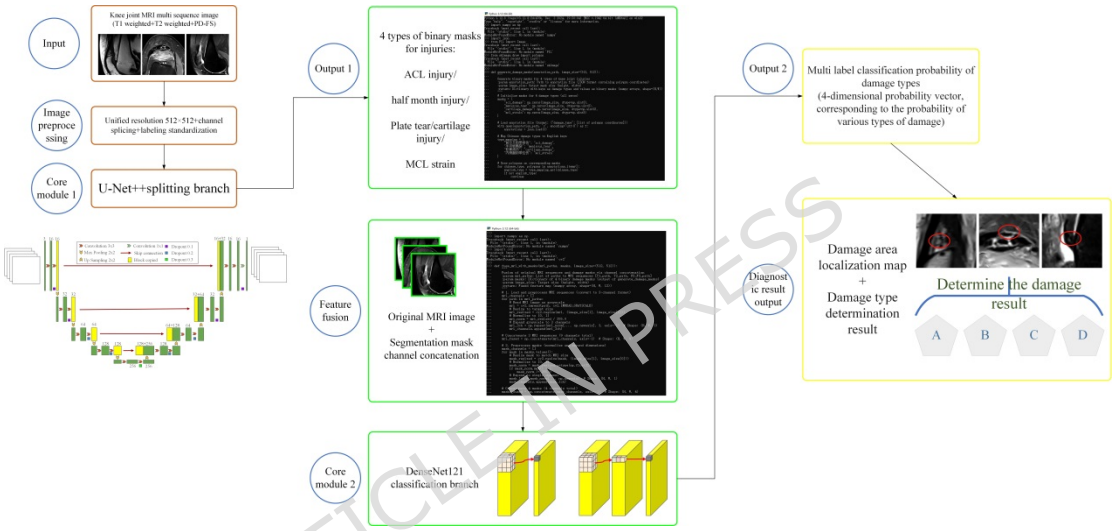

Figure 1. Specific architecture of the model

3.2 Detailed Explanation of Core Algorithm Modules

3.2.1 U-Net++ Segmentation Module

U-Net++ serves as the core network for the segmentation branch. Based on the original U-Net, it introduces densely connected decoder paths and deep supervision mechanisms to address the issues of insufficient precision and gradient vanishing in small lesion segmentation. This makes it suitable for localizing subtle injuries such as meniscus tears and cartilage defects in knee MRI.

(1) Network Structural Features

a. Encoder: Comprising a 4-layer downsampling structure, with each layer consisting of two 3×3 convolutional layers (ReLU activation) + a 2×2 max-pooling layer (stride of 2). This structure progressively extracts multi-scale image features, compressing the spatial dimensions while expanding the receptive field.

b. Decoder: Achieving upsampling through densely connected transposed convolution blocks. Each decoder block establishes dense connections with the

corresponding encoder layer and preceding decoder blocks, fully integrating multi-scale feature information.

c. Deep Supervision: Auxiliary segmentation heads are set at different levels of the decoder, guiding early network training through intermediate loss functions to enhance the segmentation accuracy of small injury regions.

(2) Mathematical Expression of Segmentation Task

It is supposed that the input MRI image is $I \in \mathbb{R}^{H \times W \times C}$ (H denotes the height, W denotes the width, and C denotes the number of channels, and the segmentation module outputs the injury region mask $M \in \mathbb{R}^{H \times W \times K}$ (K = 4, corresponding to binary masks for four types of injuries). The segmentation process can be expressed: $M = f_{U\text{-}Net++}(I; \theta_1)$ (θ_1: network parameters of $U\text{-}Net++$; $f_{U\text{-}Net++}$: mapping relationship of the segmentation network).

3.2.2 DenseNet121 Classification Module

DenseNet121 serves as the backbone network for the classification branch. Through the feature reuse mechanism in dense blocks, it effectively alleviates gradient vanishing and enhances feature propagation, making it suitable for multi-label injury type recognition tasks.

(1) Network Structural Adaptation

Feature Input Layer: The original MRI image is concatenated with the injury region mask output by U-Net++ in the channel dimension, resulting in $I_{fusion} \in \mathbb{R}^{H \times W \times (C+K)}$, a feature that serves as the input to the classification network. The dense blocks and transition layers consist of 4 dense blocks and 3 transition layers. Each convolutional layer in the dense blocks takes the concatenation of all preceding layers' feature maps as input. The transition layers reduce the dimensionality through 1×1 convolutions and compress the feature map size using average pooling. The original fully connected layer of DenseNet121 is removed and replaced with a "global average pooling layer + 2 fully connected layers + Sigmoid activation layer", outputting classification probabilities for four types of injuries.

(2) Mathematical Expression of Classification Task

It is supposed that the input to the classification module is the fused feature map I_{fusion}, and the output is the multi-label classification probability vector $P \in \mathbb{R}^4$ (each element corresponds to the existence probability of one type of injury). The classification process can be expressed: $P = f_{DenseNet121}(I_{fusion}; \theta_2) = \sigma\left(g_{fc}\left(g_{pool}\left(f_{db}(I_{fusion}; \theta_{db})\right)\right)\right)$ (θ_2: classification network parameters; f_{db}: feature extraction function of the dense blocks; g_{pool}: global

average pooling operation; $\mathcal{G}_{fc}$: mapping of the fully connected layer; s Sigmoid activation function).

3.3 Definition and Annotation Standards of Injury Types

Focusing on the high-incidence types of knee injuries in adolescents and combining the actual needs of clinical diagnosis, this study primarily targets four core types of knee injuries. To ensure the consistency and accuracy of data annotation and provide a reliable gold standard for the subsequent training of deep learning models, clear definitions, clinical feature descriptions, and detailed annotation standards have been established for each type of injury.

(1) Anterior Cruciate Ligament Injury: Defined as including both complete and partial tears. On MRI, it is mainly characterized by disruption of ligament continuity, increased internal signal, or irregular shape. During annotation, radiologists are required to precisely delineate the extent of the injury along the entire length of the anterior cruciate ligament.

(2) Meniscus Tear: Covering common types such as longitudinal, transverse, and radial tears. The typical imaging features are high signal intensity within the meniscus that extends to the articular surface. During annotation, it is necessary not only to outline the injured area of the meniscus but also to clearly depict the trajectory of the tear line.

(3) Cartilage Injury: Mainly including surface defects, abnormal thinning, or signal changes in articular cartilage, which are commonly found on the femoral condyles and tibial plateau cartilage. Annotation requires pixel-level precision, i.e., precisely delineating each microscopic region of cartilage injury.

(4) Medial Collateral Ligament Strain: Referring to ligament edema, increased signal, or partial fiber tears without obvious complete disruption of continuity. The annotation range should cover the medial collateral ligament itself and the surrounding tissues with edema or abnormal signals due to strain.

All MRI images were independently annotated by three radiologists with over five years of experience in musculoskeletal MRI diagnosis. Subsequently, the annotation results were subjected to Kappa consistency testing. Only when the consistency coefficient (K value) was greater than or equal to 0.85 was the consensus result determined as the final gold standard. If there were discrepancies in the initial annotations, a group discussion was held until consensus was reached for all doubtful cases.

The standardized operating procedure was adopted for the annotation work in this study. For all MRI images, 5-6 core diagnostic slices were selected from each of the sagittal, coronal, and axial planes for each individual sample, resulting in a total of 15-18 slices per sample as the annotation targets, with a focus on pathological regions with high injury incidence. Pixel-level annotations were performed by three radiologists using the LabelMe 5.1.1 annotation tool. The average annotation time per slice was 3-4 minutes, and the complete annotation process for a single sample took approximately 60-70 minutes. A phased quality

control scheme was developed for the annotation process. In the first phase, pre-annotation was conducted on 50 samples to unify the annotation standards and operational procedures. In the second phase, the samples were divided and annotated in a 1:1:1 ratio, with interim quality control performed after every 50 completed samples. In the third phase, all annotation results were subjected to cross-checking, and in cases of disagreement, immediate discussions were organized within the group.

3.4 Model Input and Output

(1) Input Data Specifications

The model input consists of multi-sequence knee MRI images, including T1-weighted, T2-weighted, and PD-FS sequences. The resolution of each sequence image is uniformly adjusted to 512×512 pixels with 3 channels (converted to RGB format). The multi-sequence images are concatenated across channels to form a 9-channel input (3 sequences × 3 channels), resulting in a final input dimension of 512×512×9.

(2) Output Result Definitions

a. Segmentation Output: Four binary segmentation masks corresponding to four types of injuries. Pixels with a value of 1 indicate injured areas, while 0 indicates normal regions. The resolution of the masks is consistent with the input image, at 512×512.

b. Classification Output: A 4-dimensional probability vector, where each element corresponds to the existence probability of one type of injury. A threshold of 0.5 is set; probabilities ≥0.5 are determined as the presence of that injury type, while <0.5 indicate the absence of that injury type.

3.5 Model Training Strategy

(1) Transfer Learning Initialization

The encoders of U-Net++ and DenseNet121 are initialized with weights pre-trained on ImageNet. The classification head and the decoder of U-Net++ are randomly initialized using the He normal distribution to accelerate model convergence and enhance generalizability.

(2) Data Augmentation Strategy

To address the limited quantity of medical imaging data, the following online data augmentation methods are applied in real-time during the training phase.

a. Geometric Transformations: Random rotation (-10° to 10°), translation (±10 pixels), scaling (0.9 to 1.1 times), and horizontal flipping (probability 0.5).

b. Pixel-Level Transformations: Random adjustments to brightness (±10%), contrast (±10%), and Gamma correction (0.8 to 1.2), as well as the addition of Gaussian noise (variance 0.001).

c. Medical Imaging-Specific Augmentations: Random channel reordering of MRI sequences and elastic deformation (with parameters a =100 and s =10) to simulate image variations under different scanning conditions.

(3) Loss Function Design

A joint loss function is employed to optimize both segmentation and classification tasks: $L_{total} = \lambda_1 L_{seg} + \lambda_2 L_{cls}$ (λ_1= 0.6, λ_2= 0.4 are the loss weights; L_{seg}: segmentation loss: L_{cls}: classification loss).

a. Segmentation Loss: A weighted sum of Dice loss and binary cross-entropy (BCE) loss is used to address class imbalance $L_{seg} = 0.5 \times \left(1 - Dice\left(M, M_{gt}\right)\right) + 0.5 \times \mathrm{BCE}\left(M, M_{gt}\right)$. M_{gt} is the gold-standard segmentation mask, and the Dice coefficient (DC) is calculated: $Dice\left(M, M_{gt}\right) = \frac{2\sum_{i,j} M(i,j) \times M_{gt}(i,j)}{\sum_{i,j} M(i,j)^2 + \sum_{i,j} M_{gt}(i,j)^2 + \varepsilon}$. $\varepsilon = 1e-6$ is a smoothing term to avoid division by zero.

b. Classification Loss: Focal loss is used to focus on hard-to-classify samples and address class imbalance: $L_{cls} = -\frac{1}{N}\sum_{k=1}^{4}\sum_{n=1}^{N}\left[y_{nk}(1-p_{nk})^{\gamma}\log(p_{nk}) + (1-y_{nk})p_{nk}^{\gamma}\log(1-p_{nk})\right]$ (N: batch size; y_{nk}: gold-standard label for the k injury type of the sample n (0 or 1); p_{nk}: predicted probability; $\gamma = 2$: focusing parameter.

(4) Training Parameters and Validation Strategy

a. Optimizer: Adam optimizer with an initial learning rate of 1e-4. A learning rate decay strategy is used, reducing the learning rate by a factor of 0.9 every 20 epochs to balance convergence speed and training stability.

b. Training Hyperparameters: Batch size = 16, Epochs = 100. An early stopping strategy is employed with a Patience of 15; training is stopped if the validation loss does not decrease for 15 consecutive epochs.

c. A hierarchical validation strategy was employed in this study. First, the entire set of annotated data was randomly divided into a training set, a validation set, and an independent test set in a ratio of 7:2:1. The independent test set was kept completely uninvolved in model training and tuning and was used only for the final evaluation of the model's generalization ability. Within the training set, a 5-fold cross-validation was performed, where the training set was equally partitioned into 5 subsets. In each iteration, 4 subsets were used as the training samples, and 1 subset served as the internal validation samples for hyperparameter tuning and preliminary validation of model generalization. The

average performance across the 5 iterations was taken as the cross-validation result of the training set to avoid overestimation of model performance caused by a single data split.

3.6 Statistical analysis methods

All quantitative data in this study were presented as mean ± standard deviation. All statistical analyses were performed using Python 3.12 in conjunction with the scipy (version 1.11.4) and scikit-learn (version 1.3.2) libraries. The significance level was set at $\alpha = 0.05$, with $P < 0.05$ considered as indicating a statistically significant difference and $P < 0.01$ considered as indicating an extremely significant statistical difference.

For the segmentation and classification tasks, as the evaluation metrics for different models were all computed based on the same batch of experimental samples, a paired-sample t-test was utilized for comparisons between groups (e.g., Dice coefficient, IoU, accuracy, and F1-score). This was done to assess the differences between the model proposed in this study and a single comparison model. For comparisons involving multiple models, one-way analysis of variance (ANOVA) was first conducted to verify overall differences, followed by post-hoc multiple comparisons using the least significant difference (LSD) method. To compare the area under the receiver operating characteristic curve (AUC) of different models, the DeLong test was employed for statistical validation, ensuring the rationality and accuracy of the comparison results.

For the mean values of all evaluation metrics, the 95% confidence interval (95% CI) was calculated using the normal distribution method to reflect the estimation precision of the metric means. The calculation formula was $\bar{x} \pm 1.96 \times s/\sqrt{n}$ (where n represents the sample size or the number of validation iterations).

The reliability of the manual annotation results was assessed using the Kappa consistency test. A Kappa value ≥ 0.85 was determined to indicate a high level of consistency in the annotation results, which were subsequently used as the gold standard for model training.

4. Experimental Design and Performance Evaluation

4.1 Experimental Materials (Dataset Collection and Processing)

(1) Data Source

The dataset was comprised of knee MRI scans provided by adolescent skiers at XXX Hospital and collaborating hospitals. A total of 309 samples were included, with subjects aged 12-18 years, all presenting with skiing-related knee injuries. All knee MRI images were acquired using a Siemens MAGNETOM Skyra 3.0T superconducting magnetic resonance scanner (Siemens Healthcare, Germany). The scanning was performed on the knee joint in the standard anatomical position, encompassing three orthogonal planes: sagittal, coronal, and axial. The core

scanning sequences included T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), and proton density-weighted imaging with fat suppression (PD-FS). Standardized scanning parameters were applied across all sequences, and the detailed core parameters are presented in Table 1.

Table 1. Core scan sequence parameters of the knee joint MRI

Scan sequence	Weighted method	TR (ms)	TE (ms)	Layer thickness (mm)	Interlayer spacing (mm)	Matrix size	field of vision (FOV, cm)	Number of incentives (NEX)
T1WI	Spin echo (SE)	500-600	10-15	3	0.3	320×256	16×16	1
T2WI	Fast spin echo (FSE)	3000-4000	80-100	3	0.3	320×256	16×16	2
PD-FS	Fast spin echo (FSE)	2500-3000	30-40	3	0.3	320×256	16×16	2

(2) Data Annotation

Data annotation was completed by three senior radiologists with over five years of experience in musculoskeletal MRI diagnosis. The dual gold standards included:

a. Segmentation Labels: Pixel-level annotation was used to delineate the specific regions of four types of injuries: anterior cruciate ligament injury, meniscus tear, cartilage injury, and medial collateral ligament strain, generating corresponding binary masks.

b. Classification Labels: Based on clinical diagnostic conclusions and imaging features, each sample was annotated with multi-label injury types, specifying the presence or absence of the four types of injuries and their specific combinations.

After annotation, the reliability of the results was verified using Kappa consistency testing. When the consistency coefficient $K \geq 0.85$ was achieved, the result was determined as the final gold standard. Discrepancies were resolved through group consultation to reach consensus.

(3) Data Partitioning

The 309 annotated data were randomly allocated to training, validation, and test sets (7:2:1).

a. Training Set (216 cases): parameter learning and feature extraction.

b. Validation Set (62 cases): hyperparameter tuning and early stopping during training to monitor overfitting risk.

c. Test Set (31 cases): evaluation of the final generalization efficacy. The data were never involved in model training to ensure the authenticity of the evaluation results.

4.2 Experimental Environment

(1) Hardware Configuration

The experimental hardware platform used high-performance computing devices.

a. GPU: NVIDIA Tesla V100, providing parallel computing capabilities to accelerate model training iterations.

b. CPU: Intel Xeon series processor, ensuring the operation of auxiliary tasks such as data preprocessing and feature transfer.

c. Storage and Memory: Equipped with high-speed SSD and large-capacity memory to meet the rapid reading and processing needs of MRI image data.

(2) Software Environment

The experimental software was built on the Python ecosystem with the following configurations.

a. Programming Language: Python 3.12, providing syntax support and third-party library resources.

b. Deep Learning Framework: TensorFlow 2.5, used for model construction, training, and inference.

c. Data Processing Libraries: OpenCV for MRI image preprocessing and format conversion, scikit-learn for data partitioning, metric calculation, and experimental result statistics.

d. Other Dependency Libraries: NumPy and Pandas for data matrix operations and tabular processing, Matplotlib for visualizing experimental results.

4.3 Parameter Settings

(1) Optimizer Parameters

The Adam optimizer was used for model parameter updates with an initial learning rate of 1e-4. The rate was reduced to 90% of its current value every 20 epochs to balance the convergence speed in the early stages and training stability in the later stages.

(2) Loss Function Configuration

Differentiated loss function designs were used to meet the core needs of segmentation and classification tasks.

a. Segmentation Loss: A weighted sum of Dice loss and BCE loss was adopted, with weights α and $1-\alpha$, respectively, to address class imbalance in medical images and improve segmentation accuracy for small injury regions.

b. Classification Loss: Focal loss was used, with a focusing parameter γ set to 2.0 in the experiment to reduce the weight contribution of easy-to-classify samples and focus on hard-to-classify injury samples, improving class imbalance in multi-label classification tasks.

(3) Training Hyperparameters

a. Batch Size: 16, aiming to balance GPU memory utilization and model training stability, avoiding training fluctuations caused by small batch sizes.

b. Epochs: 100 (maximum), with an early stopping strategy (Patience=15). If the validation loss failed to decrease over 15 consecutive epochs, training was automatically stopped to prevent overfitting.

4.4 Performance Evaluation

To comprehensively and objectively validate the performance of the U-Net++ and DenseNet121 hybrid model in MRI diagnosis of knee injuries after skiing in adolescents, a multi-dimensional evaluation system covering segmentation and classification tasks was constructed. Comparative experiments were conducted, and statistical analysis was used to verify the significance of the results.

4.4.1 Evaluation Metrics and Comparison Schemes

(1) Evaluation Metrics

a. Segmentation Task: DC and intersection over union (IoU) were used to quantify the overlap between model segmentation results and gold-standard annotated regions. $DC = \frac{2|A \cap B|}{|A| + |B|}$ (A: model segmentation result, B: gold-standard annotated region $IoU = \frac{|A \cap B|}{|A \cup B|}$, reflecting the precise coverage ability of the segmented region.

b. Classification Task: accuracy, precision, recall, F1-score, and AUC were used to assess the model's ability to identify multiple injury types.

(2) Comparison Methods

a. Traditional Machine Learning Method: SVM + Handcrafted Features. Texture, shape, and grayscale features were manually extracted and fed into an SVM classifier to complete injury recognition.

b. Benchmark Deep Learning Models:

Original U-Net (baseline for segmentation task).

ResNet50 (baseline for classification task).

U-Net + ResNet50 independent combination (segmentation and classification modules trained independently without feature sharing).

The training of all comparative models was conducted under identical conditions to the proposed model, utilizing the same dataset, hyperparameter settings, and evaluation criteria.

4.4.2 Experimental Results and Statistical Analysis

(1) Segmentation Task Performance Comparison

The results (Table 2) indicated that the proposed model achieved superior segmentation performance for all four types of injuries compared to the original U-Net. The average Dice coefficient (DC) was 0.89±0.03 (95% CI: 0.86 to 0.92), and the average Intersection over Union (IoU) was 0.82±0.04 (95% CI: 0.79 to 0.85), representing an improvement of 7.3% and 8.1%, respectively, over the original U-Net. Among the injury types, the highest segmentation accuracy was observed for meniscal tears, with a DC of 0.93±0.02 (95% CI: 0.91 to 0.95) and an IoU of 0.87±0.03 (95% CI: 0.84 to 0.90). This was attributed to the typical characteristic of meniscal tears on MRI, which present as high signal intensity extending to the articular surface, a feature that could be effectively captured by the dense connection mechanism of the U-Net++ architecture used in the model. Cartilage damage was identified as a challenging task for segmentation, achieving a DC of 0.84±0.04 (95% CI: 0.81 to 0.87) and an IoU of 0.77±0.05 (95% CI: 0.73 to 0.81). Despite the difficulty, this still represented a 6.4% improvement over the original U-Net. This gain was attributed to the deep supervision mechanism, which effectively mitigated the gradient vanishing problem in the segmentation of small lesions.

Statistical analysis indicated that the proposed model had significant differences in DC and IoU for all types of injuries compared to the original U-Net ($P < 0.05$), validating the effectiveness of the improved structure of U-Net++.

Table 2. Comparison of segmentation task performance among different models

Injury Type	Evaluation Metric	Original U-Net	Proposed Model	P-value
Anterior cruciate ligament injury	DC	0.83±0.04 (0.80~0.86)	0.89±0.03 (0.86~0.92)	<0.05
	IoU	0.76±0.05 (0.72~0.80)	0.82±0.04 (0.79~0.85)	<0.05
Meniscus tear	DC	0.86±0.03 (0.83~0.89)	0.93±0.02 (0.91~0.95)	<0.01
	IoU	0.79±0.04 (0.76~0.82)	0.87±0.03 (0.84~0.90)	<0.01
Cartilage damage	DC	0.79±0.05 (0.75~0.83)	0.84±0.04 (0.81~0.87)	<0.05
	IoU	0.72±0.06 (0.67~0.77)	0.77±0.05 (0.73~0.81)	<0.05
Medial collateral ligament strain	DC	0.82±0.04 (0.79~0.85)	0.89±0.03 (0.86~0.92)	<0.05
	IoU	0.75±0.05 (0.71~0.79)	0.82±0.04 (0.79~0.85)	<0.05

Average performance	DC	0.83±0.04 (0.80~0.86)	0.89±0.03 (0.86~0.92)	<0.01
	IoU	0.76±0.05 (0.72~0.80)	0.82±0.04 (0.79~0.85)	<0.01

Note: A *P*-value of <0.01 was considered statistically significant

(2) Comparison of Classification Task Performance

Table 3 demonstrated that the proposed model exhibited excellent overall classification performance, with a mean accuracy of 0.90±0.02 (95% CI: 0.88 to 0.92), precision of 0.92±0.02 (95% CI: 0.90 to 0.94), recall of 0.89±0.03 (95% CI: 0.86 to 0.92), F1-score of 0.91±0.02 (95% CI: 0.89 to 0.93), and Area Under the Curve (AUC) of 0.95±0.01 (95% CI: 0.94 to 0.96). All metrics were significantly superior to those of the comparison models.

Compared with the traditional machine learning method (SVM + handcrafted features), the F1-score of the proposed model increased by 18.6%, and the AUC increased by 15.9%. The main reason is that handcrafted features struggle to capture the deep semantic information of MRI images, while deep learning models can automatically extract multi-scale, multi-level injury features through end-to-end training. Compared with ResNet50, the F1-score of the proposed model increased by 6.5%, thanks to the accurate ROI features provided by the segmentation module, which effectively reduced the interference of background noise on the classification task. Compared with the independent combination of U-Net + ResNet50, the F1-score of the proposed model increased by 4.2%, and the training efficiency improved by 32%, demonstrating that the collaborative optimization strategy of sharing underlying features and using a joint loss function can effectively enhance model performance and training efficiency.

Statistical analysis showed that the proposed model had significant differences in F1-score and AUC compared with all comparative models ($P < 0.01$), validating the rationality of the model design. The ROC curves of different models are illustrated in Figure 2.

Table 3. Comprehensive performance comparison of classification tasks among different models

Model	Accuracy	Precision	Recall	F1-score	AUC
SVM + handcrafted features	0.78±0.04 (0.75~0.81)	0.79±0.04 (0.76~0.82)	0.76±0.05 (0.72~0.80)	0.77±0.04 (0.74~0.80)	0.82±0.05 (0.78~0.86)
ResNet50	0.85±0.03 (0.82~0.88)	0.87±0.03 (0.84~0.90)	0.84±0.04 (0.81~0.87)	0.85±0.03 (0.82~0.88)	0.90±0.03 (0.87~0.93)
U-Net + ResNet50	0.87±0.0	0.89±0.0	0.86±0.0	0.87±0.0	0.92±0.0

(independent)	3 (0.84~0.90)	2 (0.87~0.91)	3 (0.83~0.89)	3 (0.84~0.90)	2 (0.90~0.94)
Proposed model	0.90±0.02 (0.88~0.92)	0.92±0.02 (0.90~0.94)	0.89±0.03 (0.86~0.92)	0.91±0.02 (0.89~0.93)	0.95±0.01 (0.94~0.96)
Proposed model vs SVM + handcrafted features (P-value)	<0.01	<0.01	<0.01	<0.01	<0.01
Proposed model vs ResNet50 (P-value)	<0.01	<0.01	<0.01	<0.01	<0.01

Note: A *P*-value of <0.01 was considered statistically significant

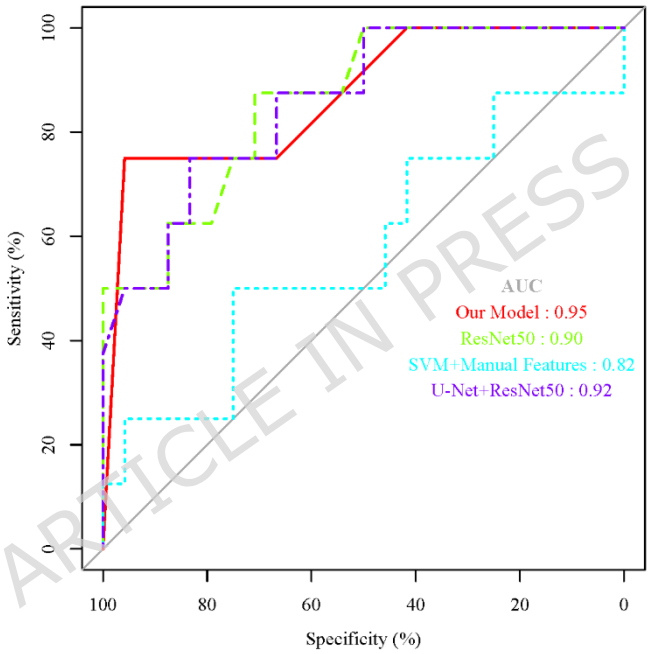

Figure 2. ROC curves of different models

(3) Classification Performance for Different Injury Types

The proposed model demonstrated the best recognition performance for meniscal tears, with a recall of 0.93±0.02 (95% CI: 0.91 to 0.95), a precision of 0.94±0.02 (95% CI: 0.92 to 0.96), and an F1-score of 0.93±0.02 (95% CI: 0.91 to 0.95). This superior performance was attributed to the typical imaging characteristics of this injury type, which could be precisely captured by the model through multi-scale feature fusion. The classification performance for anterior cruciate ligament (ACL) injuries and medial collateral ligament (MCL) sprains was slightly lower, with F1-scores of 0.91±0.02 (95% CI: 0.89 to 0.93) and 0.90±0.03 (95% CI: 0.87 to 0.93), respectively (Figure 3). The model was able to effectively differentiate between complete and partial tears, as well as between ligamentous edema and normal tissue. The classification performance for cartilage damage was

relatively lower, with a recall of 0.87±0.03 (95% CI: 0.84 to 0.90), a precision of 0.89±0.03 (95% CI: 0.86 to 0.92), and an F1-score of 0.88±0.03 (95% CI: 0.85 to 0.91). This was primarily because cartilage damage often manifests as microscopic defects with weak signal characteristics that are susceptible to interference from surrounding tissues. Nevertheless, this level of performance still met the clinical diagnostic requirements regarding false-negative and false-positive rates. The ROC curves for the proposed model's classification of different injury types are presented in Figure 4.

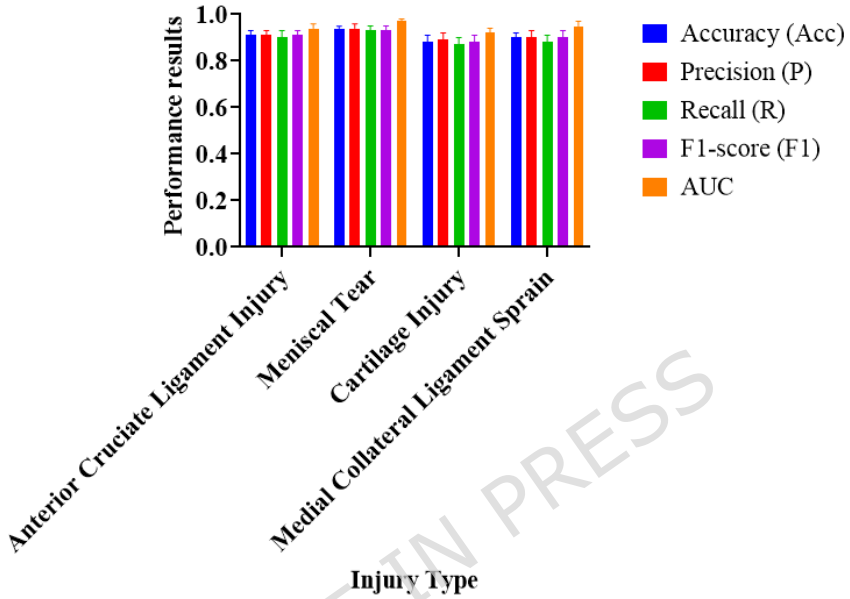

Figure 3. Classification performance of the proposed model for different injury types

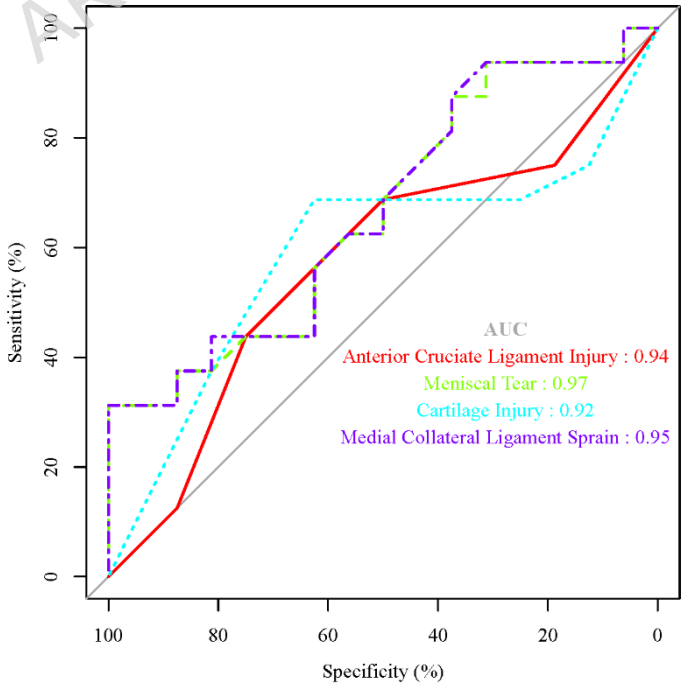

Figure 4. ROC curves for the classification of different injury types by the proposed model

4.4.3 Typical cases

Figure 5 illustrates the segmentation and classification performance for knee joint injuries sustained during skiing in adolescents. For each of the four types of injuries, one typical MRI case was selected, and a comparative visualization was generated showing the original image alongside the manually annotated ground truth (Figure 5).

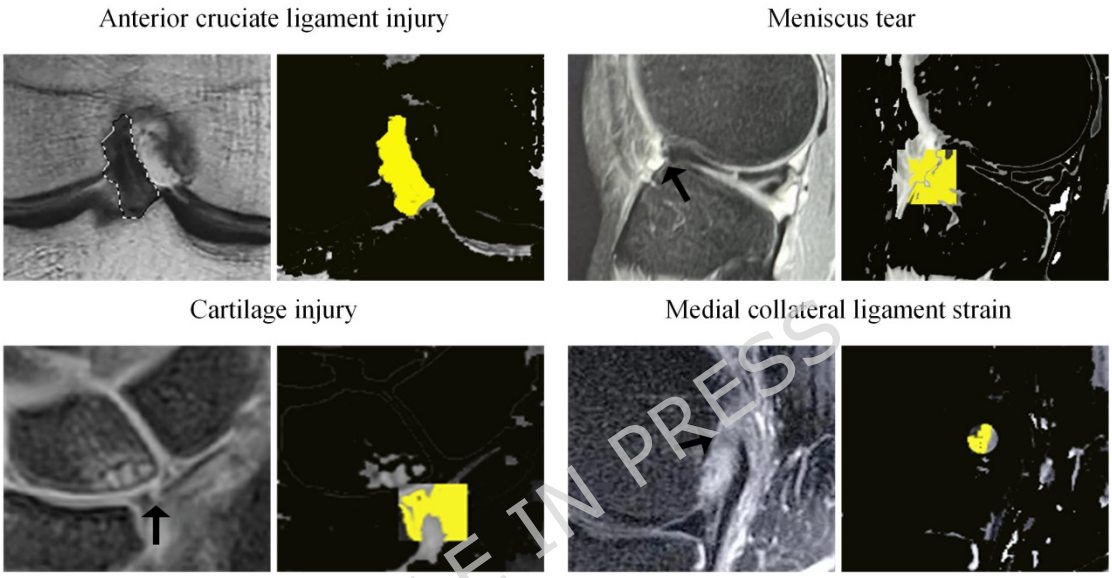

Figure 5 Comparison of typical MRI cases for four types of knee joint injuries
Note: In each image group, the first image shows the original MRI of the knee injury, and the second image presents the segmentation and classification outputs from the model proposed in this paper. The yellow overlay in the figure represents the segmentation mask of the injured area, indicating the specific locations of the knee injury identified by the model.

4.4.4 Results Analysis and Discussion

Based on the quantitative results of segmentation and classification tasks, statistical differences, and comparative experimental conclusions, the performance advantages and core mechanisms of the proposed U-Net++ and DenseNet121 hybrid model are analyzed. The advantages of the proposed model are attributed to the joint architecture design of segmentation, fusion, and classification, as well as targeted structural improvements. The core logic is reflected in three aspects. Firstly, addressing the challenge of small lesion segmentation in medical images, the densely connected decoder and deep supervision mechanism of U-Net++ form a dual optimization [21]. Dense connections enable feature reuse through multi-scale feature pathways, solving the problem of insufficient feature transfer in the original U-Net that leads to missed detections of small lesions. Deep supervision guides early training through

auxiliary segmentation heads in intermediate layers, effectively mitigating gradient vanishing [22]. This was also the key reason for the 6.4% improvement in segmentation accuracy for cartilage damage compared to the original U-Net, and this improvement was statistically significant (independent samples t-test, $P < 0.05$, 95% CI: 0.03 to 0.09). Data showed that the segmentation accuracy for meniscus tears and anterior cruciate ligament injuries (DC=0.93, 0.89) was significantly higher than that for cartilage injuries (DC=0.84). This is because the former has strong imaging features, while the latter is mostly microscopic defects. However, the model still achieves effective capture through deep supervision, validating the adaptability of the structural improvements for weak-feature lesions. Secondly, the feature fusion and sharing mechanism between the segmentation and classification modules breaks through the performance bottleneck of traditional independent segmentation and classification. By concatenating the injury region mask with the original MRI image, the classification module can directly focus on the ROI, avoiding interference from background noise [23]. Shared underlying features reduce parameter redundancy, allowing spatial location features and semantic category features to enhance each other. This was also the core reason for the 4.2% improvement in the F1-score and the 32% increase in training efficiency achieved by the proposed model compared to the independent combined model of U-Net and ResNet50. This difference was highly statistically significant (independent samples t-test, $P < 0.01$, 95% CI: 0.02 to 0.06). Thirdly, the differentiated design of the loss function precisely matches the characteristics of medical imaging data. The segmentation task uses a weighted sum of Dice loss and BCE loss to dynamically address class imbalance [24]. The classification task employs Focal loss to focus on hard-to-classify samples, optimizing the differentiation between partial tears and ligament edema, balancing recall and precision. The proposed model achieved an average F1-score of 0.91±0.02 (95% CI: 0.89 to 0.93), representing a 6.5% improvement compared to the 0.85±0.03 (95% CI: 0.82 to 0.88) achieved by ResNet50. This difference was highly statistically significant (independent samples t-test, $P < 0.01$).

Experimental results showed that the diagnostic performance of the model for the four types of injuries varied: meniscus tears (F1=0.93), followed by anterior cruciate ligament injuries (F1=0.91), medial collateral ligament strains (F1=0.90), and cartilage injuries (F1=0.88). This variation stems from the imaging features, anatomical structures, and clinical pathological characteristics of the injuries themselves. Meniscus tears have the best diagnostic performance because they have clear imaging biomarkers with high signal intensity extending through the articular surface, and their anatomical location is independent with minimal interference from surrounding tissues. The model can quickly capture specific features [25], with recall and precision both exceeding 0.93. Anterior cruciate ligament and medial collateral ligament injuries have slightly lower performance. Both types of injuries present as increased signal, edema, or disrupted continuity, but the anterior cruciate ligament has more consistent imaging features due to its regular course, while the medial collateral ligament is more susceptible to

interference from joint fluid signals [26,27], resulting in a slightly lower recall rate (0.88) compared to the anterior cruciate ligament (0.90). Cartilage injuries have the lowest diagnostic performance. This is because articular cartilage is a microscopic structure with a thickness of only 1-4mm, and the injury signal in MRI images is weak. Additionally, the high water content in adolescent cartilage means that the signal difference between normal and injured areas is small, increasing the difficulty of identification [28,29]. However, the model still achieved a recall rate of 0.87 for cartilage injuries, meeting the core clinical demand to reduce false-negative rates. The recall rate of the proposed model for cartilage damage was 0.87±0.03 (95% CI: 0.84 to 0.90), which satisfied the core clinical requirement of reducing the false-negative rate. Furthermore, this metric was significantly higher than the 0.72±0.04 (95% CI: 0.69 to 0.75) achieved by the traditional SVM with handcrafted features (independent samples t-test, $P < 0.01$).

Deep learning methods offer irreplaceable clinical value in the segmentation of knee cartilage and serve as a core technical approach to address the issue of missed diagnoses of subtle cartilage injuries in adolescents after skiing. The knee cartilage has a thickness of only 1–4 mm, and due to the high water content in adolescent cartilage, the signal difference between normal and injured areas is extremely minimal. Moreover, cartilage injuries often manifest as microscopic defects. Traditional automatic or semi-automatic segmentation methods, which rely heavily on manually designed texture and shape features, are difficult to capture these subtle microscopic features of cartilage damage. This often results in blurred segmentation boundaries and missed detections of small lesions. Furthermore, the quantitative analysis results of such methods are greatly influenced by the selection of manual features, leading to poor diagnostic consistency. In contrast, deep learning methods, through an end-to-end training paradigm, can automatically extract multi-scale, deep semantic features of cartilage damage from MRI images, achieving pixel-level precise segmentation without manual intervention. The reproducibility and consistency of their quantitative analysis results are significantly superior to those of traditional methods. While traditional semi-automatic segmentation methods for cartilage injury typically achieve Dice coefficients in the range of 0.70–0.78, the fusion model proposed in this study improves the Dice coefficient for segmenting cartilage injuries in adolescents after skiing to 0.84, a 6.4% increase over the original U-Net. This fully demonstrates the performance advantage of deep learning in cartilage segmentation. At the same time, the automated segmentation mode of deep learning reduces the segmentation time for a single cartilage image from several minutes with traditional methods to the millisecond level, greatly enhancing clinical diagnostic efficiency and aligning more closely with the clinical need for rapid assessment of skiing-related sports injuries.

From a clinical practice perspective, the proposed model provides key technical support for the diagnosis of knee injuries in adolescent skiers. In terms of diagnostic efficiency, the model achieves full automation, with analysis from image input to result output taking only 3-5 seconds. This is a significant

improvement compared to the traditional manual diagnosis time of 10-15 minutes per case, meeting the need for rapid assessment of sports injuries [30]. In terms of diagnostic accuracy, the average AUC reached 0.95, with recall rates for all types of injuries exceeding 0.87, effectively reducing the risk of false-negative and false-positive diagnoses. This makes it suitable for primary hospitals or settings lacking experienced radiologists, enhancing diagnostic consistency [31]. In terms of specificity, the model focuses on high-incidence injury types in adolescents, with training data sourced exclusively from this population, fully considering their skeletal and soft tissue development characteristics. Compared to general diagnostic models, it has greater clinical adaptability.

5. Conclusion

5.1 Research Contributions

This study focuses on the MRI diagnostic needs for knee injuries in adolescent skiers and proposes a hybrid deep learning model combining U-Net++ and DenseNet121 for the first time, achieving simultaneous segmentation and classification of four high-incidence injuries, including anterior cruciate ligament injuries and meniscus tears. The model provides a new technical approach for multi-task diagnosis in medical imaging through its joint architecture of segmentation, fusion, and classification, effectively addressing issues such as insufficient segmentation accuracy for small lesions and interference from background noise. It also constructs an end-to-end automated diagnostic system, eliminating the need for human intervention in the entire process. The diagnostic efficiency is improved by over 95% compared to traditional methods, reducing diagnostic variability among physicians. This provides a reliable tool for primary hospitals and rapid assessment of sports injuries, with clear potential for clinical translation.

5.2 Limitations and Future Work

5.2.1 Limitations of the Study

The dataset used in this study was sourced from a single collaborative medical institution. Although the sample size was sufficient to meet the requirements for model training and validation, the study lacked multi-center and multi-device MRI data, and external independent dataset validation was not performed. This was identified as a core factor affecting the model's generalizability. Slight differences exist in MRI scanner models, and scanning parameters across different medical institutions, and factors such as operational protocols for image acquisition and patient positioning can lead to heterogeneity in imaging features. Consequently, models trained on single-center data may experience a slight decrease in segmentation and classification accuracy when applied to image data from other centers. Furthermore, as no external validation cohort was established in this study, the diagnostic performance of the model in real-world multi-center clinical

scenarios could not be objectively assessed. This limitation represents a key issue that should be addressed in future research.

The source of 309 cases is relatively concentrated, lacking multi-center and multi-device data. Additionally, the absence of injury subtypes such as bone marrow edema and tendinitis may affect the model's generalizability. The model is highly dependent on MRI image quality and lacks adaptability to low signal-to-noise ratio or artifact-containing images. The diagnostic value of differentiated information from multi-sequence imaging has not been fully explored.

5.2.2 Future Research Directions

It should collaborate with multiple centers to collect samples, expand the scale and diversity of data, include more injury subtypes and data from different scanning parameters, and enhance model robustness through domain adaptation training. Future studies could introduce attention mechanisms to enhance the capture of weak features and explore 3D structures to fully utilize the spatial dimension information of MRI.

Funding: Research on the Logical Rationale and Practical Pathways of New Quality Productive Forces Empowering the High-Quality Development of the Ice and Snow Sports Industry (Project No. 25156), a Key Research Project on Economy and Society of Heilongjiang Province for 2025.

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author Songmei Li on reasonable request via e-mail htxiayuxue2000@126.com.

Competing Interest

The authors declare no competing financial or non-financial interests.

Author Contributions Declaration

Wei Xu：Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation

Songmei Li：writing—review and editing, visualization, supervision, project administration, funding acquisition

Guofeng Zhang：methodology, software, validation, formal analysis

Qi Zhang：formal analysis, investigation, resources, data curation

Weidong Song: visualization, supervision, project administration

ETHICS STATEMENT

The study was conducted in accordance with the Declaration of Helsinki, the studies involving human participants were reviewed and approved by Graduate School, Harbin Sport University Ethics Committee (Approval Number: 2023.1200232). The participants provided their written informed consent to participate in this study. All methods were performed in accordance with relevant guidelines and regulations.

References

[1] Veldman HD, Jeuken RM, Voskuilen RN, Verlaan L. Operatively treated ischial tuberosity avulsion in an adolescent snowboarder: treatment, patient's perspective, and critical appraisal of current literature on the indication for surgery. BMJ Case Rep. 2025 Jun 30;18(6): e264107. doi: 10.1136/bcr-2024-264107. PMID: 40588300.

[2] Hanimann J, Raschle N, Schmid NE, Bruhin B, Frey WO, Scherr J, de Bruin ED, Spörri J. Jump performance and movement quality in 7- to 15-year-old competitive alpine skiers: a cross-sectional study. Ann Med. 2024 Dec;56(1):2361254. doi: 10.1080/07853890.2024.2361254. Epub 2024 Jun 4. PMID: 38833367; PMCID: PMC11151804.

[3] Liu RKK, Lieu M, Howell DR, Pierpoint LA, Khodaee M, Provance A. Comparing Single-Site Fractures with Multisite Injuries in Pediatric Skiers and Snowboarders. Wilderness Environ Med. 2025 Dec;36(4):542-548. doi: 10.1177/10806032251345771. Epub 2025 Jun 11. PMID: 40495671.

[4] Mugniery Q, Ricard C, Mirtain S, Navarre M, Tanné C. Epidemiology of paediatric winter sports-related injuries in France: The WINTRAUMA-1 retrospective cohort study. Acta Paediatr. 2025 Mar;114(3):603-610. doi: 10.1111/apa.17480. Epub 2024 Nov 1. PMID: 39487605; PMCID: PMC11828721.

[5] Liao J, Yu K. MRI Radiomics-Based Diagnosis of Knee Meniscal Injury. J Comput Assist Tomogr. 2025 Nov-Dec 01;49(6):952-957. doi: 10.1097/RCT.0000000000001759. Epub 2025 Apr 14. PMID: 40249263.

[6] Jing Z, Han Y, Xu Y, Nie X, Sun L, Li D, Li D. MRI measurement analysis of risk factors for popliteal artery injury in knee surgery. J Orthop Surg (Hong Kong). 2025 May-Aug;33(2):10225536251330659. doi: 10.1177/10225536251330659. Epub 2025 May 6. PMID: 40326003.

[7] Wang Y, Ying M, Yang Y, Chen Y, Wang H, Tsai TY, Liu X. Multitask learning for automatic detection of meniscal injury on 3D knee MRI. J Orthop Res. 2025 Mar;43(3):703-713. doi: 10.1002/jor.26024. Epub 2024 Dec 2. PMID: 39620311.

[8] Zhang S, Zheng T, Jin J, Ye C, He R. Diagnostic Value of Pericruciate Fat Pad Measurement by MRI in Patients With Knee Articular Cartilage Injury. Br J Hosp Med (Lond). 2025 Apr 25;86(4):1-13. doi: 10.12968/hmed. 2024.0824. Epub 2025 Apr 15. PMID: 40265547.

[9] Donners R, Vosshenrich J, Segeroth M, Seng M, Fenchel M, Nickel MD, Bach M, Schmaranzer F, Todorski I, Obmann MM, Harder D, Breit HC. Deep Learning Reconstructed New-Generation 0.55 T MRI of the Knee: A Prospective Comparison With Conventional 3 T MRI. Invest Radiol. 2024 Dec 1;59(12):823-830. doi: 10.1097/RLI.0000000000001093. Epub 2024 Jun 11. PMID: 38857414.

[10] Foti G, Spoto F, Spezia A, Romano L, Caia S, Camerani F, Benedetti D, Mignolli T. Deep learning-driven abbreviated knee MRI protocols: diagnostic accuracy in clinical practice. Radiol Med. 2025 Sep;130(9):1460-1471. doi: 10.1007/s11547-025-02038-3. Epub 2025 Jul 4. PMID: 40613973.

[11] Botnari A, Kadar M, Puia DR, Patrascu JM, Jr JMP. Automated Segmentation of Knee Menisci Using U-Net Deep Learning Model: Preliminary Results. Maedica (Bucur). 2024 Dec;19(4):690-695. doi: 10.26574/maedica.2024.19.4.690. PMID:

39974461; PMCID: PMC11834842.
[12] Botnari A, Kadar M, Patrascu JM. Considerations on Image Preprocessing Techniques Required by Deep Learning Models. The Case of the Knee MRIs. Maedica (Bucur). 2024 Sep;19(3):526-535. doi: 10.26574/maedica.2024.19.3.526. PMID: 39553362; PMCID: PMC11565144.
[13] Roy C, Roshan M, Goyal N, Rana P, Ghonge NP, Jena A, Vaishya R, Ghosh S. MRI detection and grading of knee osteoarthritis - a pilot study using an AI technique with a novel imaging-based scoring system. Biomater Sci. 2025 Sep 23;13(19):5475-5494. doi: 10.1039/d5bm00470e. PMID: 40889152.
[14] Yasin P, Yimit Y, Abulimiti A, Luan H, Peng C, Yakufu M, Song X. Dual-center study on AI-driven multi-label deep learning for X-ray screening of knee abnormalities. Sci Rep. 2025 Oct 30;15(1):38014. doi: 10.1038/s41598-025-21895-6. PMID: 41168262; PMCID: PMC12575809.
[15] Hanimann J, Fitze DP, Koller T, Schürmann F, Fröhlich S, Feuerriegel GC, Stern C, Scherr J, de Bruin ED, Sutter R, Spörri J. Traumatic knee injuries and career drop-outs in adolescent competitive alpine skiers aged 15-19: a longitudinal 4-year follow-up study examining rates, biomechanical injury risk factors and potential reasons for quitting. Ann Med. 2025 Dec;57(1):2532118. doi: 10.1080/07853890.2025.2532118. Epub 2025 Jul 15. PMID: 40662700; PMCID: PMC12265098.
[16] Kastner T, Dandrieux PE, Fohrmann D, Frohberg F, Turel M, Valtonen M, Edouard P, Spörri J, Wolfarth B, Junge A, Hollander K. Injuries and illnesses during the 54th FIS Nordic World Ski Championships 2023 in Planica: a prospective cohort study. BMJ Open Sport Exerc Med. 2025 Apr 10;11(2):e002156. doi: 10.1136/bmjsem-2024-002156. PMID: 40226334; PMCID: PMC11987098.
[17] Mead K, Cross T, Roger G, Sabharwal R, Singh S, Giannotti N. MRI deep learning models for assisted diagnosis of knee pathologies: a systematic review. Eur Radiol. 2025 May;35(5):2457-2469. doi: 10.1007/s00330-024-11105-8. Epub 2024 Oct 18. PMID: 39422725; PMCID: PMC12021734.
[18] Sun J, Cao Y, Zhou Y, Qi B. Leveraging spatial dependencies and multi-scale features for automated knee injury detection on MRI diagnosis. Front Bioeng Biotechnol. 2025 May 6;13:1590962. doi: 10.3389/fbioe.2025.1590962. PMID: 40395675; PMCID: PMC12088959.
[19] Cerezal A, Roriz D, Canga A, Cerezal L. Imaging of sports injuries in adolescents. Pediatr Radiol. 2025 Apr;55(4):644-659. doi: 10.1007/s00247-024-05991-9. Epub 2024 Jul 12. PMID: 38995428.
[20] Gicquel P. Knee ligament and meniscus injuries in children and teenagers. Orthop Traumatol Surg Res. 2025 Feb;111(1S):104073. doi: 10.1016/j.otsr.2024.104073. Epub 2024 Nov 26. PMID: 39608639.
[21] Pan J, Wu Y, Tang Z, Sun K, Li M, Sun J, Liu J, Tian J, Shen B. Automatic knee osteoarthritis severity grading based on X-ray images using a hierarchical classification method. Arthritis Res Ther. 2024 Nov 18;26(1):203. doi: 10.1186/s13075-024-03416-4. PMID: 39558425; PMCID: PMC11571664.
[22] Zhou L, Nguyen T, Choi S, Yoon J. U-Net-Based Deep Learning Hybrid Model:

Research and Evaluation for Precise Prediction of Spinal Bone Density on Abdominal Radiographs. Bioengineering (Basel). 2025 Apr 3;12(4):385. doi: 10.3390/bioengineering12040385. PMID: 40281745; PMCID: PMC12025265.
[23] Lyu Y, Tian X. MWG-UNet++: Hybrid Transformer U-Net Model for Brain Tumor Segmentation in MRI Scans. Bioengineering (Basel). 2025 Jan 31;12(2):140. doi: 10.3390/bioengineering12020140. PMID: 40001660; PMCID: PMC11852190.
[24] Wang Y, Wen Z, Bao S, Huang D, Wang Y, Yang B, Li Y, Zhou P, Zhang H, Pang H. Diffusion-CSPAM U-Net: A U-Net model integrated with a hybrid attention mechanism and diffusion model for segmentation of computed tomography images of brain metastases. Radiat Oncol. 2025 Apr 5;20(1):50. doi: 10.1186/s13014-025-02622-x. PMID: 40188354; PMCID: PMC11971865.
[25] Pujol N, Giordano AO, Wong SE, Beaufils P, Monllau JC, Arhos EK, Becker R, Della Villa F, Brett Goodloe J, Irrgang JJ, Klugarova J, Klosterman EL, Królikowska A, Krych AJ, LaPrade RF, Manske R, van Melick N, Monson JK, Ostojic M, Paterno MV, Piontek T, Perelli S, Rambaud A, Robinson J, Schmitt LC, Senorski EH, Snaebjornsson T, Tagliero AJ, Benjamin Ma C, Prill R. The formal EU-US Meniscus Rehabilitation 2024 Consensus: An ESSKA-AOSSM-AASPT initiative. Part I-Rehabilitation management after meniscus surgery (meniscectomy, repair and reconstruction). Knee Surg Sports Traumatol Arthrosc. 2025 Aug;33(8):3002-3013. doi: 10.1002/ksa.12674. Epub 2025 May 12. PMID: 40353298; PMCID: PMC12310086.
[26] Toyono S, Suzuki A, Wanezaki Y, Takahara D, Ohe R, Fukushima S, Futakuchi M, Takagi M. Predicting anterior cruciate ligament degeneration using magnetic resonance imaging: Insights from histological evaluation. J Orthop Sci. 2025 Mar;30(2):325-332. doi: 10.1016/j.jos.2024.04.009. Epub 2024 May 21. PMID: 38772763.
[27] Mercurio M, Denami F, Melissaridou D, Corona K, Cerciello S, Laganà D, Gasparini G; IORS; Minici R. Deep Learning Models to Detect Anterior Cruciate Ligament Injury on MRI: A Comprehensive Review. Diagnostics (Basel). 2025 Mar 19;15(6):776. doi: 10.3390/diagnostics15060776. PMID: 40150118; PMCID: PMC11941175.
[28] Zandee van Rilland ED, Fritz RC, Chaudhari AS, Boutin RD. Cartilage Imaging: MRI of Chondral Degeneration and Injury. Clin Sports Med. 2025 Jul;44(3):467-498. doi: 10.1016/j.csm.2024.08.004. Epub 2024 Oct 4. PMID: 40514150.
[29] Miller EY, Lee W, Lowe T, Zhu H, Argote PF, Dresdner D, Kelly J, Frank RM, McCarty E, Bravman J, Stokes D, Emery NC, Neu CP. MRI-derived articular cartilage strains predict patient-reported outcomes six months post-anterior cruciate ligament reconstruction. Sci Rep. 2025 Jul 1;15(1):21426. doi: 10.1038/s41598-025-05306-4. PMID: 40594341; PMCID: PMC12214938.
[30] Wang R, Kou Q, Dou L. LIT-Unet: a lightweight and effective model for medical image segmentation. Radiol Phys Technol. 2024 Dec;17(4):878-887. doi: 10.1007/s12194-024-00844-4. Epub 2024 Sep 20. PMID: 39302610.
[31] Garbaz A, Oukdach Y, Charfi S, El Ansari M, Koutti L, Salihoun M. MLFA-UNet: A multi-level feature assembly UNet for medical image segmentation.

Methods. 2024 Dec;232:52-64. doi: 10.1016/j.ymeth.2024.10.010. Epub 2024 Oct 29. PMID: 39481818.