SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(4), 1488.
https://doi.org/10.62617/mcb1488

Article

FDMRNet: A classification model for anterior cruciate ligament biomechanical injuries based on FSM and DFFM

Chengbin Luo[1], Bo Liu[2], Long Li[3,*]

[1] School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China
[2] Department of Medical Imaging, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China
[3] Division of Diagnostic Radiology, Department of Medical Imaging, Guangzhou Twelfth People's Hospital, Guangzhou Medical University, Guangzhou 510620, China
*** Corresponding author:** Long Li, radiolilong@hotmail.com

CITATION

Luo C, Liu B, Li L. FDMRNet: A classification model for anterior cruciate ligament biomechanical injuries based on FSM and DFFM. Molecular & Cellular Biomechanics. 2025; 22(4): 1488. https://doi.org/10.62617/mcb1488

ARTICLE INFO

Received: 5 February 2025
Accepted: 25 February 2025
Available online: 10 March 2025

Abstract: The lesion area in magnetic resonance imaging (MRI) of anterior cruciate ligament (ACL) injury is small, the features are difficult to focus, and the multiangle imaging features are scattered, which presents great challenges to clinicians for ACL injury. The ACL plays a critical role in maintaining knee stability. An injury can result in increased laxity, making the knee more vulnerable to further damage. This paper proposes a new neural network model, FDMRNet, which automatically focuses on the area of ACL injury and improves the accuracy of intelligent discrimination of the degree of ACL injury. Understanding the biomechanical effects of ACL injuries is crucial for developing effective rehabilitation protocols aimed at restoring normal knee function and preventing re-injury. First, FDMRNet enhances the focus of lesion features and reduces noise interference through the feature selection module (FSM), thereby improving the lesion localization ability. Secondly, the dimensional feature fusion module (DFFM) is used to fuse multi-angle features, which enhances the accuracy of the fusion representation of multi-angle features. To evaluate the performance of FDMRNet, real datasets from the Guangdong Provincial Armed Police Corps Hospital were used for model training and verification. The experimental results show that compared with the mainstream methods, the AUC (Area Under Curve), accuracy, precision, recall, and f1-score of the proposed model are improved by 2.52%, 3.17%, 5.79%, 4.14% and 4.54% respectively, which fully proves the effectiveness and accuracy of the proposed model in MRI classification of anterior cruciate ligament injury. Recognizing the biomechanical consequences of ACL injuries highlights the importance of accurate diagnosis and effective treatment strategies, which can be significantly enhanced through advanced models like FDMRNet.

Keywords: neural network; classification model; medical imaging; biomechanics; anterior cruciate ligament injury; model optimization

1. Introduction

Severe ACL injuries can lead to complete loss of knee joint function. Diagnosing ACL injuries typically requires a detailed clinical examination by a physician, including observing symptoms, assessing knee joint stability, and conducting imaging studies such as X-rays and MRI to confirm the extent of the injury. Among these, magnetic resonance imaging (MRI) of the knee joint provides the most comprehensive imaging assessment, which is not only cost-effective [1] but also helps predict which patients may require further treatment [2]. Therefore, MRI of the knee joint has become the preferred method for diagnosing ACL injuries.

Deep learning techniques have demonstrated high accuracy in image classification [3–6]. Deep learning models can analyze images quickly, providing a

more comprehensive image analysis than less experienced physicians [7], thereby improving the efficiency and accuracy of identifying ACL injuries. However, integrating deep learning technology with ACL diagnosis presents some challenges. In MRI of ACL injuries, the characteristics of the ACL between patients and normal individuals are relatively subtle, and MRI images from different patients or even different machines may vary. Additionally, MRI of ACL injuries typically includes multiple angles, and the lesion characteristics at different angles are not the same. Therefore, effectively utilizing the characteristics of the lesion area has become a challenge in processing MRI images of ACL injuries with deep learning models. Traditional methods, such as MRNet [8], use multiple neural network models in combination to process images of the ACL from different angles. However, they suffer from a lack of focus on lesion characteristics and the issue of feature dispersion in multi-angle images, resulting in still insufficient accuracy. Investigating a method that can better select compelling features and utilize high-order features is necessary.

This paper proposes the ACL injury classification network FDMRNet, which first uses VGG16 as the backbone network for feature extraction and employs a feature selection module (FSM) to select high-dimensional features to improve the accuracy of focusing on the lesion area, thereby enabling the model to better capture semantic information of the lesion area. The proposed dimension feature fusion module (DFFM) is used to fuse features of different dimensions from the model's output, enhancing the model's feature expression ability and thereby improving the model's performance in classifying ACL injuries. The main contributions of this paper are as follows:

- The proposal of the FSM module adds FSM after the model's backbone network, allowing the model to select high-dimensional features extracted by the backbone network, focusing more on the semantic information of the ACL lesion area and enhancing model performance.
- The proposal of the DFFM module, which, through DFFM, fuses features of different dimensions from the model's output, addresses the issue of feature dispersion in multi-angle ACL MRI images and improves the model's feature representation ability.
- The construction of a dataset for validating the effectiveness of FDMRNet and conducting extensive testing. Experimental results show that FDMRNet's AUC, accuracy, precision, recall, and F1 scores are 2.52%, 3.17%, 5.79%, 4.14%, and 4.45% higher than MRNet's.

2. Related work

2.1. Anterior cruciate ligament injury classification network

For the classification problem of ACL injuries, there are mainly two approaches: using 2D neural networks and 3D neural networks. To test whether deep learning models can accomplish the important task of detecting diseases in MRI, Bien et al. [8] proposed MRNet, which is a convolutional neural network with a backbone consisting of three identical models. These can be various neural networks, such as AlexNet [9] or ResNet [10], each processing images from three different angles. Then, the output features are concatenated in dimension and passed through a fully connected layer to

obtain the final results. Furthermore, when validating with the external dataset made public by Štajduhar et al. [11], it was found that external datasets play an important role in training deep learning models for generalization. However, the MRNet model neglects feature selection and feature fusion, preventing the model from fully utilizing lesion characteristics. Zhang [12] developed a classification CNN based on 3D DenseNet, using 408 MR images to diagnose ACL tears, and compared its diagnostic performance with VGG16 and ResNet. The results were superior to the other two models. In a large-scale study, Germann et al. [13] used arthroscopic surgery reports as the reference standard, and the deep neural network achieved a sensitivity of 99% on an internal dataset of 512 subjects. However, during external validation, the diagnostic performance significantly decreased, indicating that its performance may decline with the increase in MRI examination heterogeneity. Namiri et al. [14] classified ACL injuries using 2D and 3D convolutional neural networks and compared performance metrics. Overall, these networks can assist physicians in classifying ACL injuries. However, since the performance of 2D CNNs is slightly better than that of 3D CNNs, 2D CNNs are closer to practical application scenarios. Nevertheless, when processing MRI images of the ACL, 2D CNNs have the issue of feature dispersion in the features extracted by the backbone network, failing to effectively capture important semantic information from the MRI images. Therefore, there is a need to investigate a new method to improve the problem of feature information loss in traditional methods.

2.2. Deep feature fusion

To address issues such as model overfitting and information loss, feature fusion modules are commonly used to integrate features from different levels to enhance the model's expressive power. Feature fusion modules can take various forms, such as addition, concatenation, and weighted summation [11]. Precise tracking of transparent objects, like glasses, is crucial in many robotic tasks, especially robot-assisted living tasks. However, traditional tracking algorithms often face performance degradation due to their reliance on general learning functions. Garigapati et al. [15] proposed a new feature fusion module that integrates transparency information into a fixed feature space, making it usable in a broader range of trackers. Deevi et al. [16] proposed an efficient modular RGB-X fusion network that utilizes and fuses pre-trained unimodal models through scene-specific fusion modules, enabling the construction of adaptive network architectures for joint inputs on smaller datasets. Xie et al. [17] proposed a novel fusion method called the Point-based Attentive Cont-conv Fusion (PACF) module, which directly fuses multi-sensor features on 3D points, aiming to effectively merge LiDAR point clouds and RGB images. Chen et al. [18] proposed Freq Fusion, which integrates an Adaptive Low Pass Filter (ALPF) generator, an offset generator, and an Adaptive High Pass Filter (AHPF) generator to address the issue of high-frequency feature interference leading to intra-class inconsistency in dense image prediction tasks. In late fusion, each modality is processed in a separate unimodal convolutional neural network stream, and the scores of each modality are fused at the end. Due to its simplicity, late fusion remains the primary method in many state-of-the-art multimodal applications [19]. However, these feature fusion modules cannot

be easily inserted directly into the ACL classification model. Hence, a new feature fusion module needs to be designed to fully utilize the multi-angle input characteristics of ACL images.

3. Methods

3.1. FDMRNet architecture

The backbone network of FDMRNet is based on VGG16.After the backbone network outputs features, the FSM module performs feature selection to capture adequate semantic information. The selected features are then fused using the DFFM module, which integrates features from various angles of the ACL, followed by the final classification output. The overall architecture of FDMRNet is shown in **Figure 1**.

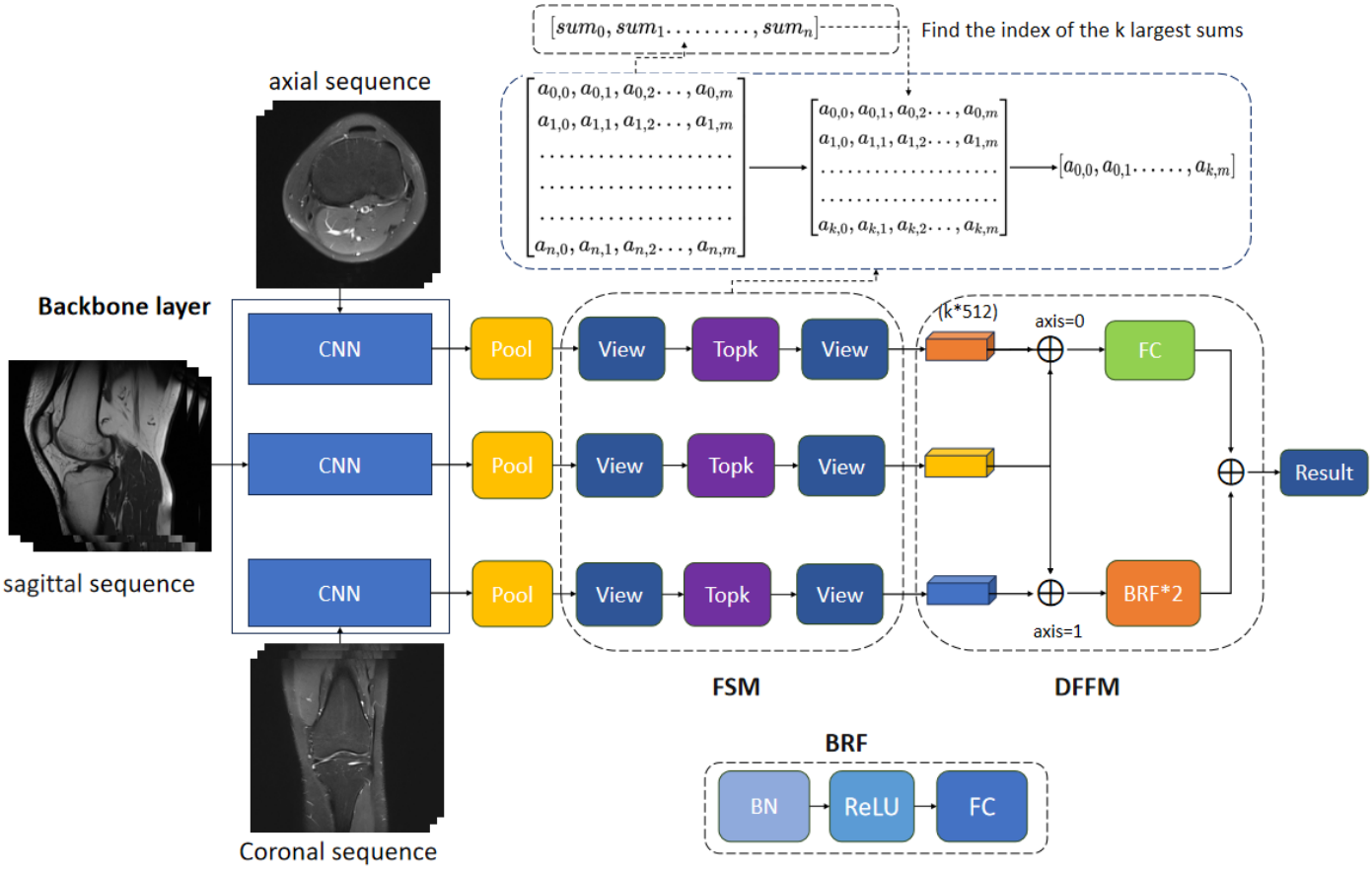

Figure 1. Overview of FDMRNet architecture. The backbone network employs a three-branch convolutional neural network to extract multi-angle features from MRI image sequences of the ACL captured in three orthogonal planes: the axial plane (top branch), sagittal plane (middle branch), and coronal plane (bottom branch).

Max pooling layers are applied to the features extracted by the backbone network to reduce spatial dimensions and accelerate model convergence. The FSM contains two view alignment components (before and after feature processing) with a Top-K selection mechanism in between: The view alignment components ensure dimensional consistency, while the Top-K mechanism filters the most discriminative lesion features. The BRF module, sequentially composed of Batch Normalization (BN),

Rectified Linear Unit (ReLU), and Fully Connected (FC) layers, performs feature normalization and non-linear mapping. The DFFM (Dimension Feature Fusion Module) adopts a dual-path strategy: One path concatenates features along the channel dimension (0th axis) followed by a fully connected layer for probability prediction, while the other concatenates features along the spatial dimension (1st axis) processed through two BN-ReLU-FC(BRF) modules before probability generation. Final predictions are obtained by fusing outputs from both paths.

3.2. Features select module

Effective feature selection can improve the prediction performance of the model [20]. The goal of feature selection is to find the optimal feature subset. Feature selection can eliminate irrelevant or redundant features, thereby reducing the number of features and improving the accuracy of the model [21].

The operation of a neural network usually covers three key processes: forward propagation, gradient calculation, and parameter update.

During the forward propagation process, the convolution kernel performs convolution operations with the input feature map to generate an output feature map. This process extracts feature information from the input feature map through convolution operations, which manifests the forward flow of information in the neural network.

Gradient calculation uses the chain rule to backpropagate the error and calculate the gradient of each parameter. In a neural network, the error indicates the difference between the model's prediction result and the actual label. By backpropagating the mistake, the contribution of each parameter to the error can be clarified, thereby providing a basis for a parameter update.

The parameter update adjusts the weight of the convolution kernel according to the calculated gradient direction. In this process, the weights corresponding to the key areas will be updated more. Precisely, the convolution kernel will dynamically adjust its parameters according to the loss function to allow the model to pay more attention to specific areas. For example, it may be the lesion area in medical image analysis.

The core of backpropagation is to calculate the gradient $\frac{\partial J}{\partial W}$ of the loss function $J(\theta)$ to the convolution kernel parameter W, and use the gradient descent method to update the parameter. Its mathematical expression is:

$$W_{\text{new}} = W_{\text{old}} - \eta \times \frac{\partial J}{\partial W},$$

where W_{old} is the current convolution kernel parameter, η is the learning rate, which controls the step size of each parameter update. In the next forward propagation, the updated parameter W_{new} will generate the output feature map X_{output} based on the input feature map X_{input}, and the relationship can be expressed as:

$$X_{\text{output}} = W_{\text{new}} \times X_{\text{input}}.$$

At this time, the input feature map X_{input} is the result after the convolution kernel is filtered, and its value will be more significant in some areas. In a specific classification task, the features of the lesion area are often more critical. This means

that the model will automatically focus more on the lesion area, giving the feature map in this part a more significant value. Based on this, we can further guide the model to focus on key features by letting the model pay attention to the feature matrix with a more significant sum value, thereby improving the model's performance on specific tasks.

When processing MRI images, traditional 2D CNNs fail to focus effectively on the lesion area, which may lead to the negative impact of the loss of some effective semantic information. In order to solve this problem, by adjusting the feature dimension, the method of selecting the optimal feature subset is adopted to extract effective features, thereby enhancing the model's ability to capture effective semantic information.

We take the feature matrix of the data as the entry point for feature selection and propose an FSM, which effectively enables the model to select high-order features extracted by the appropriate backbone network. The module is shown in the FSM in **Figure 1**. Before passing through the FSM module, the backbone network will output the extracted feature map, which will be max-pooled through the Pool module and output the feature matrix X, which will then serve as the input of the FSM.

The FSM module adjusts the dimensions of the input and output vectors. The output X after MaxPool will be adjusted to X_v by the view operation. For example, if the dimension of X is (n, m, 1, 1), it will be adjusted to (n, m) after the view operation, so there will be no loss of information. X_v can be calculated as

$$X_v = view(X),$$

where the feature matrix X_v consists of n rows and m columns, the number of rows represents the number of images at the current angle. In contrast, the number of columns indicates the number of feature maps extracted from a single image by the backbone network. Subsequently, it is necessary to calculate the sum of the elements in each row, denoted as S_i. The sum S_i can be expressed as:

$$S_i = \sum_{j=1}^{m} X_{v_{i,j}},$$

where $X_{v_{i,j}}$ represents the element at the i-th row and j-th column (where $\{\ 0 \le i < n$ and $0 \le j < m\ \}$), indicating the maximum value in the feature map of the j-th channel extracted from the i-th image by the backbone network. In other words, it represents the feature with the most information within that channel. The sum S_i calculated for each row is used with the feature matrix X_v through a Topk operation to compute a new matrix X_v'. The new matrix X_v' can be represented as:

$$X_v' = Topk(X_v, S_i),$$

where $Topk$ can use the sum of each row vector S_i in the X_v matrix to find the row vectors corresponding to the indexes of the first k maximum values, retain the row vectors of the corresponding indexes, and discard all the remaining row vectors to obtain a new matrix X_v' . The dimension of the new tensor X_v' should be (k, m). Regarding feature information, X_v' retains the row vector corresponding to the maximum S_i value equivalent to the maximum channel feature of the feature map of the MRI image with the most feature information. Then, adjust the dimension of X_v'

through View, change the dimension of X_v' to a vector of 1 row and $k \times m$ columns, and obtain a new feature matrix X_{vv}'. X_{vv}' can be calculated as:

$$X_{vv}' = View(X_v').$$

In summary, the $Topk$ operation is used to select the first k most enormous rows and the corresponding row vectors according to the value of S_i, and the remaining row vectors are discarded. This enables the FSM module to accurately retain the features that contain the most information and avoid interference from irrelevant or redundant features. In contrast, some traditional feature selection methods may only screen based on the statistical properties of the features (such as variance, correlation, etc.) and cannot directly focus on high-information feature rows like FSM. Moreover, the FSM module can be well adapted to all neural networks to achieve a plug-and-play effect. At the same time, some traditional feature selection methods may not be well integrated with the neural network's overall architecture and training process.

3.3. Multi-view feature fusion module

Although the FSM module will retain the lesion's feature information as much as possible while removing redundant features, directly using the fully connected layer to process the multi-angle features of the anterior cruciate ligament is not ideal.

The feature fusion module can fully fuse the multi-angle features, but many traditional feature fusion modules only perform simple operations such as feature splicing or weighted summation in a single dimension, which cannot effectively use the characteristics of ACL MRI multi-angle images.

Inspired by SimSiam's data projection module [22], this paper proposes the Dimension Feature Fusion Module (DFFM). The DFFM projects the input multi-view features onto two different dimensional spaces for fusion, enabling the model to obtain richer and more comprehensive representations. As a result, the model's performance and generalization ability are improved.

Compared with traditional feature fusion methods, this multidimensional projection approach is more complex and flexible, as it can capture the relationships between features from different perspectives. Moreover, the BN operation in the BRF module of DFFM allows the gradients to flow among batch elements. Additionally, the information about the opposing views in the batch can serve as implicit negative (contrastive) terms [23]. All samples in a batch are associated with each other, thereby achieving the effects of preventing overfitting and feature fusion. The structure of the DFFM is shown in **Figure 1**.

DFFM first concatenates the three-dimensional $k \times m$-dimensional features X_{axial}, $X_{sagittal}$ and $X_{coronal}$ into a $1 \times k \times m \times 3$-dimensional vector X_{axis0}. That is, it adds X_{axial}, $X_{sagittal}$ and $X_{coronal}$ in the 0-th dimension and outputs X_{axis0}, thereby mapping the features of the three scattered angles to the same dimension, X_{axis0} that can be calculated as:

$$X_{axial}, X_{sagittal}, X_{coronal} \in R^{k \times m} \longrightarrow X_{axis0} \in R^{1 \times k \times m \times 3}.$$

Concurrently, the three-dimensional $k \times m$-dimensional features X_{axial}, $X_{sagittal}$ and $X_{coronal}$ of the input are concatenated into a $3 \times k \times m$-dimensional

vector X_{axis1}; that is, the vectors are added in the first dimension, and the three scattered angle features are mapped to different dimensional spaces for vector alignment. X_{axis1} can be calculated as

$$X_{axial}, X_{sagittal}, X_{coronal} \in R^{k\times m} \longrightarrow X_{axis1} \in R^{3\times k\times m}.$$

Then X_{axis0} passes through the fully connected layer and outputs the 1×1 dimensional lesion probability value $f(X_{axis0})$, which can be calculated as:

$$f(X_{axis0}) = W_1 \times X_{axis0} + b_1.$$

Meanwhile, X_{axis1} passes through two BRF layers and outputs a 1×1 dimensional lesion probability value $f(X_{axis1})$, $f(X_{axis1})$ can be calculated as:

$$f(X_{axis1}) = W_3 \times \sigma(BN(W_2 \times \sigma(BN(X_{axis1})) + b_2)) + b_3.$$

Finally, the two lesion probability values $f(X_{axis0})$ and $f(X_{axis1})$ are added together to output the disease diagnosis result, result can be calculated as:

$$result = f(X_{axis0}) + f(X_{axis1}).$$

4. Experimental results and analysis

4.1. Dataset

The experimental dataset is constructed based on the MRI images of the ACL from 725 individuals provided by the Guangdong Provincial Armed Police Corps Hospital. The machines used to capture the MRI images were Siemens MAGNETOM Skyra 3.0T magnetic resonance scanner and GE 1.5T magnetic resonance scanner. The patient labels in this dataset were obtained from arthroscopy, meaning that the patient data was verified as diseased through ACL surgery under arthroscopy, ensuring the accuracy of the patient data. The labels for the experimental dataset include 294 normal individuals and 431 patients with ACL tears. The dataset was randomly divided into three groups with a ratio of 6:2:2 for training, validation, and testing sets, respectively, using five-fold cross-validation.

However, it should be noted that the current dataset only contains samples from military personnel, and the sample homogeneity is high, which limits the generalization ability of the model. In order to solve this problem, comprehensively evaluate the performance of the model, and improve its applicability in different populations, we plan to expand the dataset to personnel of different ages, genders, nationalities, and various occupations in subsequent research to achieve the purpose of diversified samples.

4.2. Experimental environment

All experiments were conducted on a computer with an Intel® Core™ i7-13700 CPU, 64GB of RAM, and the Ubuntu 22.04 64-bit operating system. The computer also had an NVIDIA GeForce RTX 4090 graphics card with 24GB of RAM and NVIDIA CUDA 12.2. Additionally, the experiments in this paper were completed using the Python 3.8.16 programming language within the PyTorch 2.0.1 framework.

In the experiment, the model was trained for 50 epochs each time, with a learning rate of 1×10^{-5}, and the Adam [24] optimizer was used. During the training phase, the input images were uniformly resized to 224×224. To perform effective normalization, we scaled the pixel values to the range of $[0,255]$ and obtained x_{scaled} as follows:

$$x_{\text{scaled}} = \frac{x - \min(x)}{\max(x) - \min(x)} \times 255,$$

here, x represents the pixel value, and $\min(x)$ and $\max(x)$ represent the minimum and maximum pixel values in the image. In this way, the pixel values are linearly scaled to the range of $[0,255]$. Then, we standardized x_{scaled} according to the mean of the data $\text{MEAN} = 58.09$ and the standard deviation $\text{STDDEV} = 49.73$ to obtain $x_{\text{standardized}}$. $x_{\text{standardized}}$ can be expressed as:

$$x_{\text{standardized}} = \frac{x_{\text{scaled}} - \text{MEAN}}{\text{STDDEV}}.$$

In terms of data augmentation, we used a variety of transformation operations. The image was randomly rotated with a probability of 0.2, and the rotation range was ±15° on the *x*, *y*, and *z* axes; the image was randomly scaled with a probability of 0.2, and the scaling ratio was between 0.9 and 1.1; the image was randomly affine transformed with a probability of 0.5, and the translation range was 10 pixels.

The data loader's parameters were set to batch size 1, and the data order was shuffled at the beginning of each training epoch.

In addition, due to the imbalanced data distribution, the ratio of patients to normal people in the data was about 4:6. To solve the problem of data imbalance and classification of difficult and easy samples, the cross-entropy loss function is not used, but Focal Loss [25] is used as the loss function.

The expression for Focal Loss is as follows:

$$FocalLoss(p_t) = -\alpha_t (1 - p_t)^{\gamma} log(p_t),$$

where p_t is the model's prediction probability for the correct class; α_t is a weight factor used to balance positive and negative samples, which can be set according to the actual situation; γ is a hyperparameter that adjusts the focus on easy and hard samples. In this experiment, the γ setting is 2.0, set to the proportion of diseased patients among all patients, which is 0.406.

4.3. Performance indicators

In order to evaluate the classification performance of different models on the anterior cruciate ligament MRI image dataset, the experiment tested five test indicators: area under the ROC curve (AUC), accuracy, precision, recall, and F1-score. The reason for selecting these five indicators is that they can comprehensively evaluate the performance of the model from different dimensions, including the overall performance of the model, the ability to identify positive classes, and the ability to avoid misjudgment and underreporting of cases. The five test indicators are shown below.

$$\text{AUC} = \frac{\sum_{i=1}^{n-1} (\text{FPR}_i - \text{FPR}_{i+1}) \times (\text{TPR}_i + \text{TPR}_{i+1})}{2},$$

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

$$\text{Recall} = \frac{TP}{TP+FN},$$

$$\text{Precision} = \frac{TP}{TP+FP},$$

$$F1-Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}}.$$

TP (True Positives) represents samples correctly classified as positive and indeed positive. FP (False Positives) represents samples misclassified as positive but negative. TN (True Negatives) represents samples correctly classified as negative and indeed negative. FN (False Negatives) represents samples misclassified as negative but positive. Precision measures the proportion of correct positive class predictions. The higher the Precision, the stronger the model's ability to accurately identify positive examples. Recall (also known as Sensitivity) measures the proportion of actual positive classes that are correctly predicted. The higher the Recall, the stronger the model's ability to capture positive examples and the higher the coverage of positive samples in the prediction results. Accuracy can measure the overall performance of the model, and it can intuitively reflect the overall correctness of the model's classification of all samples, not just limited to positive and negative samples. Generally speaking, the higher the Accuracy, the stronger the model's comprehensive judgment ability and robustness. The F1-Score is a metric that combines Precision and Recall to provide an overall measure of model accuracy. AUC (Area Under the ROC Curve) is a metric for measuring the performance of a model at different classification thresholds, represented by the area under the ROC curve. The ROC curve is plotted with the True Positive Rate (the same as Recall) on the y-axis and the False Positive Rate on the x-axis. The range of AUC is from 0 to 1, and the closer the value is to 1, the better the model performance.

4.4. Backbone network

To identify the best backbone network for feature extraction in our dataset, we tested various networks as the backbone of MRNet. We conducted model performance tests and Class Activation Mapping (CAM) experiments on EfficientNet, DenseNet, ResNet, AlexNet, and VGG16. CAM is a model visualization technique that represents the focus of the model on the important regions of the image [26].

In **Table 1**, VGG16 achieved the highest scores with 93.94% AUC, 86.90% accuracy, 88.83% precision, 79.17% recall, and 88.37% F1-Score, outperforming other models when used as the backbone network. Compared to other neural networks, VGG16 effectively captures rich information in images through a large number of convolutional and pooling layers, making it valuable for image classification and recognition tasks. This demonstrates that VGG16 is capable of extracting feature information more effectively when used as the backbone network in our dataset.

Table 1. Experimental comparison of the effects of various backbone networks unit: %.

Backbone Network	AUC	Acc	Precision	Recall	F1-Score
Efficientnet-b0	86.88	78.76	84.97	66.62	79.01
Efficientnet-b1	81.61	73.79	80.97	62.48	74.07
Densenet121	86.72	74.90	76.14	64.55	75.61
Resnet18	76.43	72.55	80.55	58.34	71.79
Resnet34	65.30	63.72	74.34	60.41	66.67
Resnet50	77.12	71.28	87.86	60.41	71.60
Resnet101	85.29	75.03	84.97	70.76	77.27
Alexnet	93.62	78.76	86.90	70.76	80.00
VGG16	93.94	86.90	88.83	79.17	88.37

In **Figure 2**, the focus of the model on the image regions is ranked as follows: red > yellow > blue. In the MRI images of the anterior cruciate ligament, the red dashed line indicates the lesion area. The results show that VGG16 can better focus on the lesion area in our dataset, with the red area containing only a small proportion of non-lesion areas and relatively less yellow area, indicating higher overlap between the model's focus and the lesion area, effectively capturing key information from the lesion area with minimal noise interference. In contrast, AlexNet, ResNet18, ResNet50, ResNet101, DenseNet101, EfficientNet-b0, and EfficientNet-b1 have lower overlap between the red area and the lesion area, or the red area is too large, failing to effectively focus on the lesion area, or focusing on too much noise area. Therefore, it can be concluded that VGG16 performs relatively well in extracting features from this type of data.

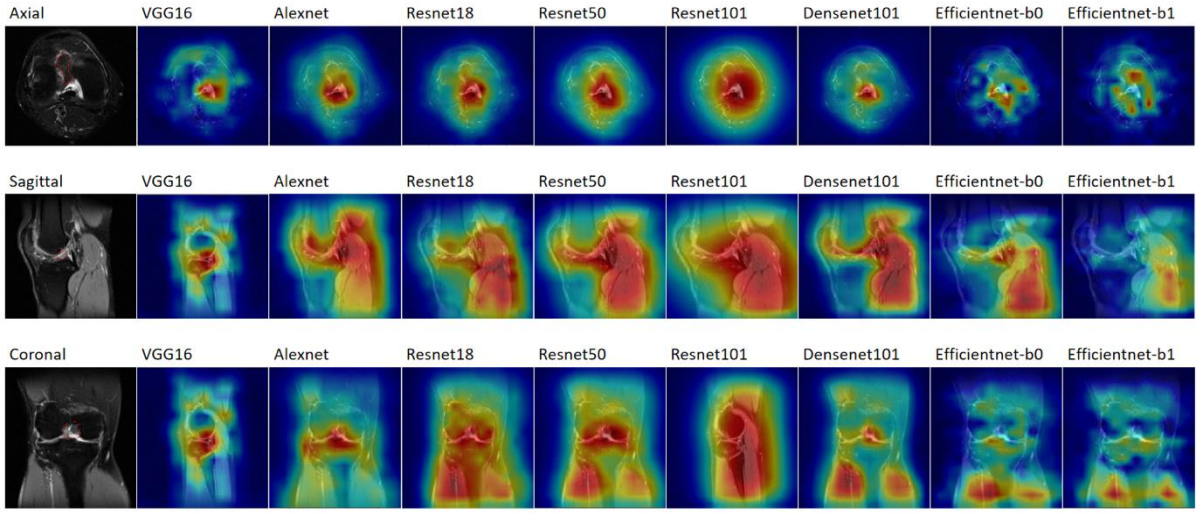

Figure 2. Comparison of class activation mapping. In CAM, the degree of attention that the model pays to regions of interest [26] in the image is represented by color, where red indicates a stronger focus than yellow, and yellow indicates a stronger focus than blue. The red line in the ACL MRI images denotes the area of the lesion.

VGG16, compared to other neural networks, effectively extracts abundant features from images through many convolutional and pooling layers, which are very

useful for image classification and recognition tasks. Additionally, the VGG16 structure is relatively simple compared to other networks, with fewer layers, and the model converges faster, saving computational resources effectively. From the results in **Table 1**, we can see that using VGG16 as the backbone network achieves higher values in all indicators compared to other models. It demonstrates that the model is a superior classifier in overall performance, disease diagnosis accuracy, precision of case examination, and comprehensiveness of checking cases.

CAM is a model visualization technique that reveals which regions of an image are most important to the model's predictions, allowing us to explore the basis of network predictions further. The heatmap in **Figure 2** shows that when VGG16 is used as the backbone network, the red area significantly overlaps the lesion areas marked by doctors, and the yellow and red regions are relatively small. It indicates that the model's focus areas highly correspond to the lesion regions, effectively capturing the essential semantic information of the lesion area and reducing noise interference. Therefore, VGG16 performs well in feature extraction for such data.

4.5. Model performance comparison

To evaluate the effectiveness of FSM and DFFM, we compared the performance of FDMRNet with MRNet, LeNET-5 [27], 3D DenseNet [12], MR-Transformer [28], and TNIDBL [29], which have performed well in anterior cruciate ligament image classification in recent years, as well as Astroformer [30], Dynamics2 [31], SparseSwin [32], GAC-SNN [33], and ASB-Former-B [34], which have performed well in other classification fields. **Table 2** shows the performance comparison between FDMRNet and other models.

Table 2. Model performance comparison unit: %.

Model	AUC	Acc	Precision	Recall	F1-Score
MRNet (VGG16)	93.94	86.90	88.83	79.17	88.37
LeNET-5	91.48	86.21	86.62	77.10	83.06
3D DenseNet	92.03	85.10	88.41	81.24	84.67
TNIDBL	90.06	84.69	77.79	82.48	79.96
MR-Transformer	94.12	87.31	90.90	80.28	86.72
Astroformer	94.44	88.69	91.03	80.55	89.46
Dynamics2	87.41	78.21	77.24	74.62	75.23
SparseSwin	90.42	85.38	87.59	74.48	82.66
GAC-SNN	88.72	79.03	75.45	80.28	78.38
ASB-Former-B	87.42	80.41	74.31	79.72	76.89
FDMRNet	96.46	90.07	94.62	83.31	92.91

From the results in **Table 2**, we can see that the AUC, Accuracy, Precision, Recall, and F1-Score of FDMRNet are 2.52%, 3.17%, 5.79%, 4.14%, and 4.54% higher than those of MRNet and outperform other models. It can be seen that the addition of FSM and DFFM enhances MRNet's ability to capture important semantic information and feature fusion, thereby improving its generalization performance and comprehensive model performance.

The Precision-Recall Curve is an important tool for evaluating model performance. As can be seen from **Figure 3**, FDMRNet's average precision (AP) value reaches 0.88, and its PR curve completely envelops the curves of other models. This shows that FDMRNet is significantly better than other comparison models in terms of comprehensive precision and recall performance.

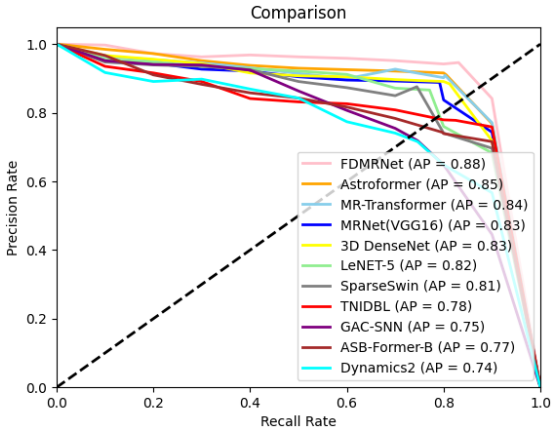

Figure 3. Precision recall graph.

The ROC curve is a core visualization tool for evaluating the performance of a binary classification model. The curve dynamically depicts the corresponding relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) as the classification threshold changes, intuitively reflecting the model's trade-off mechanism between sensitivity and specificity. Combining **Figure 4** and **Table 2**, it can be seen that the ROC curve of FDMRNet shows a significant upper left shift feature, and its area under the curve (AUC) reaches 0.9646, which is an average increase of 2.52% over the baseline model (MRNet) and significantly stronger than other models. Moreover, FDMRNet performs well at each recall rate threshold, which shows that FDMRNet has stronger robustness in controlling the false positive rate.

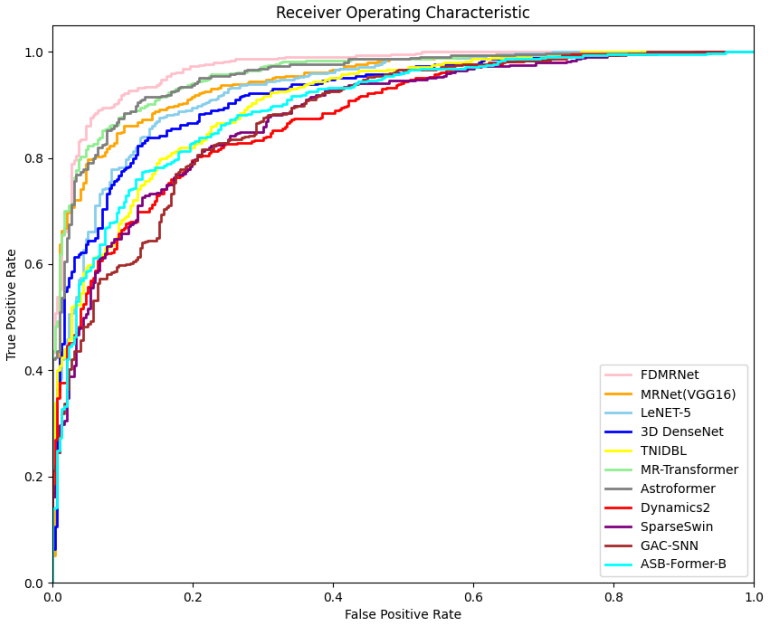

Figure 4. Receiver operating characteristic curve graph.

4.6. Ablation experiment

To assess the effectiveness of the FSM and the impact of the number of selected features on the model's performance, we compared the performance metrics of MRNet with that of MRNet augmented with FSM, using the VGG16 backbone network, which demonstrated the best overall performance. We evaluated the effects of FSM selecting different numbers of features, that is, the number of effective features chosen. The results are shown in **Table 3**. The different parameters of FSM represent the number of selected features. It can be observed that when the parameter $f = 7$, the Accuracy, Recall, and F1-Score of MRNet with FSM added improved by 2.20%, 3.31%, and 2.74% compared to MRNet, respectively. When the parameter $f = 6$, the model achieved the highest AUC of 95.22%, which is 1.28% higher than that of MRNet. When the parameter $f = 5$, the Precision of the model reached 92.28%, 3.45% higher than MRNet. These results validate the effectiveness of FSM in focusing on lesion characteristics and enhancing model performance. However, the feature parameters of FSM need to be adjusted according to the characteristics of the data.

Table 3. Comparison of FSM parameter effects unit: %.

Model	AUC	Acc	Precision	Recall	F1-Score
MRNet (VGG16)	93.94	86.90	88.83	79.17	88.37
MRNet (VGG16) + FSM ($f = 2$)	94.03	87.45	90.07	80.41	89.13
MRNet (VGG16) + FSM ($f = 3$)	94.42	88.69	89.79	80.69	90.72
MRNet (VGG16) + FSM ($f = 4$)	94.07	88.83	90.48	81.52	90.53
MRNet (VGG16) + FSM ($f = 5$)	94.33	85.10	92.28	77.10	87.06
MRNet (VGG16) + FSM ($f = 6$)	95.22	88.97	91.59	82.34	89.89
MRNet (VGG16) + FSM ($f = 7$)	94.84	89.10	91.59	82.48	91.11

To evaluate the impact of the DFFM on model performance, we used VGG16 as the backbone network. We compared the performance of the MRNet model enhanced with the DFFM module against the original MRNet model. The results are shown in **Table 4**.

Table 4. DFFM effect comparison unit: %.

Model	AUC	Acc	Precision	Recall	F1-Score
MRNet (VGG16)	93.94	86.90	88.83	79.17	88.37
MRNet (VGG16) + DDFM	95.48	88.83	93.52	82.62	91.49

From **Table 4**, it can be observed that the model with the added DFFM outperforms the original model across various metrics. Specifically, there were improvements of 1.54% in AUC, 1.93% in Accuracy, 4.69% in precision, 3.45% in Recall, and 3.12% in F1-Score. These results confirm that the DFFM can effectively integrate the high-order features from three perspectives output by the model, leading to a significant enhancement in model performance.

5. Discussion

ACL tears typically refer to injuries or ruptures of the anterior cruciate ligament (ACL) in the knee. Clinical studies in recent years have shown that the number of patients with ACL tears is on the rise, especially among athletes, soldiers, police, and other groups. Patients with ACL tears usually experience symptoms such as knee instability and limited extension and flexion activities. If not treated in time, it will increase the risk of delayed rehabilitation and recurrent sprains. Traditional examination methods rely on doctors manually examining or analyzing patient-specific ACL MRI images for diagnosis. This diagnostic method is highly dependent on the doctor's professional knowledge. However, it is easily affected by the doctor's subjective judgment, and the diagnostic process is time-consuming, which may lead to delayed treatment.

Deep learning methods have brought new opportunities for ACL injury diagnosis in this context. It can automatically analyze image data, improve inspection efficiency, and reduce the time required for manual operation. Analyzing a large amount of data can minimize the impact of subjective judgment, provide more objective results, and help patients take treatment measures as soon as possible. However, there are two problems with traditional deep learning methods when processing ACL MRI images: first, the lesion features with considerable noise are difficult to focus on, and second, the three-view input characteristics of ACL images cause feature diffusion.

In order to solve the above problems, we proposed a targeted improvement plan. In order to solve the problem of the lesion features and noise in ACL MRI images being challenging to focus on, we proposed FSM. FSM is used to select the feature subset containing the richest feature information, direct the model's attention to the lesion area, minimize the noise that may interfere with the model, and enhance the model's ability to focus on the lesion area. On the other hand, selecting features relevant to disease discrimination can simplify the model and help alleviate the overfitting problem.

To solve the feature dispersion problem caused by the model's multi-view input characteristics, we proposed DFFM. This module mainly performs feature fusion, maps the multi-view scattered features to two different vector spaces after filtering by the FSM module, and then converts them into disease probability values, which improves the model's ability to utilize features and fuse multi-view features. By improving the FSM and DFFM modules, our model achieved specific results in the ACL injury MRI image classification task.

However, there is still much room for improvement in our improvements. The aspects in which the model performance can be further improved mainly include the following: First, increasing the amount and diversity of data. Our existing dataset is a pure military dataset with only 725 cases, and the data is relatively homogeneous. Obtaining data sets of different occupational groups and increasing the data will help improve the model's generalization performance to better adapt to the diagnostic needs of different types of patients. Second, the backbone network of the model has not been improved. Other more suitable backbone networks can extract features from ACL MRI images more effectively, improving model performance and generalization ability. Third, the loss function is improved. This dataset has the problem of sample

imbalance. Although the Focal Loss function has been used to solve the sample imbalance problem as much as possible, a better solution is to design a special loss function for this data type.

In addition, ACL injury is closely related to meniscus injury and cartilage injury, and clinical diagnosis often involves multiple types of data, such as patient history, clinical symptom description, physical examination results, and other imaging examinations (such as X-rays, CT, etc.). Future research can be expanded in a broader direction. Specifically, it can conduct in-depth research with the help of multimodal models. Multimodal models can integrate multiple sources of information and can capture injury characteristics from different angles and levels. By collecting a large amount of medical image data, including ACL, meniscus, cartilage and other injuries, covering different modal data such as MRI images, X-rays, and CT, and conducting systematic training, it is expected to build a universal knee injury training model that can identify multiple modal injuries. This universal model is no longer limited to a single ACL injury classification but can accurately diagnose and classify a variety of common knee injuries, such as meniscus injuries, cartilage injuries, and other diseases, as well as multiple image data, including MRI images, X-rays, CT, etc., which will further improve the accuracy and comprehensiveness of diagnosis and provide more substantial support for clinical treatment.

6. Conclusion

In this paper, a new network structure, FDMRNet, was proposed based on MRNet to address the problem of difficulty in focusing on lesion features in ACL MRI images and feature dispersion caused by multi-view input. We designed an FSM module to screen the optimal feature subset of the model so that the model’s attention can be better focused on lesion features. In addition, we proposed a DFFM module to project multi-view model features into different dimensions and then fuse them, thereby giving the model richer data understanding capabilities. We constructed an anterior cruciate ligament MRI image dataset and conducted extensive testing. Experimental results show that compared with mainstream models, FDMRNet achieved the best performance in the ACL classification task. FDMRNet can assist clinicians in ACL diagnosis and has high application and promotion value in clinical diagnosis.

Author contributions: Conceptualization, CL; methodology, CL; software, CL; validation, CL, BL and LL; investigation, CL; data curation, CL and BL; writing—original draft preparation, CL; writing—review and editing, CL, BL and LL; supervision, LL; project administration, CL; funding acquisition, LL. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported partly by the National Natural Science Foundation of the High-level Talent Scientific Research Project of Guangzhou Twelfth People’s Hospital under Grant2022-GCC-1, and the Guangzhou Municipal Science and Technology Program under Grant 2024A0310497.

Ethical approval: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Medical Ethics Committee of the General Hospital of the People's Liberation Army General Hospital and Medical School of NAME OF

INSTITUTE (Ethics Approval No. S2023-022-01).This study was a retrospective study, did not involve diagnostic and therapeutic intervention, and the data were anonymized, so the ethics committees of the research center and other sub-centers waived patient informed consent.

Conflict of interest: The authors declare no conflict of interest.

References

1. Oei EHG, Nikken JJ, et al. Costs and effectiveness of a brief MRI examination of patients with acute knee injury. European Radiology. 2008; 19(2): 409-418. doi: 10.1007/s00330-008-1162-z
2. Oei EHG, Nikken JJ, Ginai AZ, et al. Acute Knee Trauma: Value of a Short Dedicated Extremity MR Imaging Examination for Prediction of Subsequent Treatment. Radiology. 2005; 234(1): 125-133. doi: 10.1148/radiol.2341031062
3. Srivastava S, Sharma G. OmniVec: Learning robust representations with cross modal sharing. arXiv; 2024.
4. Kabir H. Reduction of class activation uncertainty with background information. ArXiv; 2023.
5. Singh M, Duval Q, Alwala KV, et al. The effectiveness of MAE pre-pretraining for billion-scale pretraining. ArXiv; 2023.
6. Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale. ArXiv; 2020.
7. Li M, Jiang Y, Zhang Y, et al. Medical image analysis using deep learning algorithms. Frontiers in Public Health. 2023; 11. doi: 10.3389/fpubh.2023.1273253
8. Bien N, Rajpurkar P, Ball RL, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. Saria S, ed. PLOS Medicine. 2018; 15(11): e1002699. doi: 10.1371/journal.pmed.1002699
9. Krizhevsky A, Sutskever I, and Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems. 2012; 25.
10. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016.
11. Štajduhar I, Mamula M, Miletić D, et al. Semi-automated detection of anterior cruciate ligament injury from MRI. Computer Methods and Programs in Biomedicine. 2017; 140: 151-164. doi: 10.1016/j.cmpb.2016.12.006
12. Zhang L, Li M, Zhou Y, et al. Deep Learning Approach for Anterior Cruciate Ligament Lesion Detection: Evaluation of Diagnostic Performance Using Arthroscopy as the Reference Standard. Journal of Magnetic Resonance Imaging. 2020; 52(6): 1745-1752. doi: 10.1002/jmri.27266
13. Germann C, Marbach G, Civardi F, et al. Deep Convolutional Neural Network–Based Diagnosis of Anterior Cruciate Ligament Tears. Investigative Radiology. 2020; 55(8): 499-506. doi: 10.1097/rli.0000000000000664
14. Namiri NK, Flament I, Astuto B, et al. Hierarchical severity staging of anterior cruciate ligament injuries using deep learning with MRI images. ArXiv; 2020.
15. Garigapati K, Blasch E, Wei J, et al. Transparent Object Tracking with Enhanced Fusion Module. In: Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2023.
16. Deevi SA, Lee C, Gan L, et al. RGB-X Object Detection via Scene-Specific Fusion Modules. In: Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2024.
17. Xie L, Xiang C, Yu Z, et al. PI-RCNN: An Efficient Multi-Sensor 3D Object Detector with Point-Based Attentive Cont-Conv Fusion Module. Proceedings of the AAAI Conference on Artificial Intelligence. 2020; 34(07): 12460-12467. doi: 10.1609/aaai.v34i07.6933
18. Chen L, Fu Y, Gu L, et al. Frequency-Aware Feature Fusion for Dense Image Prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2024; 46(12): 10763-10780. doi: 10.1109/tpami.2024.3449959
19. Joze HRV, Shaban A, Iuzzolino ML, and Koishida K. MMTM: Multimodal transfer module for CNN fusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020.
20. Pudjihartono N, Fadason T, Kempa-Liehr AW, et al. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. Frontiers in Bioinformatics. 2022; 2. doi: 10.3389/fbinf.2022.927312
21. Halilaj E, Rajagopal A, Fiterau M, et al. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. Journal of Biomechanics. 2018; 81: 1-11. doi: 10.1016/j.jbiomech.2018.09.009

22. Chen X, He K. Exploring Simple Siamese Representation Learning. In: Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021.
23. Richemond PH, Grill J, Altché F, et al. Byol works even without batch statistics. ArXiv; 2020.
24. Kingma DP. Adam: A method for stochastic optimization. ArXiv; 2014.
25. Lin TY, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); 2017.
26. Tang D, Chen J, Ren L, et al. Reviewing CAM-Based Deep Explainable Methods in Healthcare. Applied Sciences. 2024; 14(10): 4124. doi: 10.3390/app14104124
27. Javed Awan M, Mohd Rahim M, Salim N, et al. Efficient Detection of Knee Anterior Cruciate Ligament from Magnetic Resonance Imaging Using Deep Learning Approach. Diagnostics. 2021; 11(1): 105. doi: 10.3390/diagnostics11010105
28. Zhang C, Chen S, Cigdem O, et al. MR-Transformer: Vision Transformer for Total Knee Replacement Prediction Using Magnetic Resonance Imaging. ArXiv; 2024.
29. Sethi S, Reddy S, Sakarvadia M, et al. Toward non-invasive diagnosis of Bankart lesions with deep learning. ArXiv; 2024.
30. Dagli R. Astroformer: More data might not be all you need for classification. ArXiv; 2023.
31. Phong NH, Santos A, Ribeiro B. PSO-Convolutional Neural Networks With Heterogeneous Learning Rate. IEEE Access. 2022; 10: 89970-89988. doi: 10.1109/access.2022.3201142
32. Pinasthika K, Laksono BSP, Irsal RBP, et al. SparseSwin: Swin transformer with sparse transformer block. Neurocomputing. 2024; 580: 127433. doi: 10.1016/j.neucom.2024.127433
33. Qiu X, Zhu RJ, Chou Y, et al. Gated Attention Coding for Training High-Performance and Efficient Spiking Neural Networks. Proceedings of the AAAI Conference on Artificial Intelligence. 2024; 38(1): 601-610. doi: 10.1609/aaai.v38i1.27816
34. Su Z, Chen J, Pang L, et al. Adaptive Split-Fusion Transformer. In: Proceedings of the 2023 IEEE International Conference on Multimedia and Expo (ICME); 2023.