修士論文概要書

Master’s Thesis Summary

Date of submission: 01/26/2026 (MM/DD/YYYY)

専攻名（専門分野） Department	Computer Science and Communications Engineering	氏 名 Name	Taira Kunitomi	指 導 教 員 Advisor	Hiroshi Watanabe 印 Seal
研究指導名 Research guidance	Research on Audiovisual Information Processing	学籍番号 Student ID number	CD 5124FG03-4		
研究題目 Title	Clinically Prioritized Fusion Incorporating Slice-Level Attention and Continuity Modeling for Robust Knee MRI Injury Classification				

1. Introduction

In recent years, deep learning has shown strong performance in diagnostic imaging [1], particularly for automated knee injury detection using MRI. MRNet [2] is a widely used baseline model for classifying conditions such as anterior cruciate ligament (ACL) and meniscal tears by processing sagittal T2-weighted, coronal T1-weighted, and axial proton density (PD)-weighted sequences with convolutional neural networks.

However, MRNet [2] aggregates features from different sequences by simple averaging, which ignores the heterogeneous diagnostic importance of MRI modalities and the fact that only a subset of slices is clinically informative. T1- and T2-weighted images play a critical role in ligament and soft tissue assessment.

To address these limitations, this study proposes a clinically informed framework that introduces slice-level attention and learnable sequence-level fusion weights initialized according to clinical knowledge, enabling more accurate and interpretable knee injury classification.

2. Related Works

2.1 Deep Learning Approaches for MRI Diagnosis

Deep learning–based approaches have recently shown strong potential in automating MRI interpretation for musculoskeletal injuries by enabling data-driven extraction of complex anatomical and pathological patterns. Among them, U-Net [3] is a widely used segmentation model that employs an encoder–decoder architecture to capture multi-scale contextual information and extract pixel-level pathological features. It has demonstrated high performance across a variety of medical imaging tasks, including brain tumor segmentation and musculoskeletal lesion detection.

For classification-based diagnosis, a representative model is MRNet [2], developed by Bien et al. at Stanford University. MRNet analyzes multiple anatomical views—sagittal, coronal, and axial—to detect clinically important conditions such as ACL tears, meniscal tears, and general abnormalities. Each plane is processed independently using convolutional neural networks to extract slice-level features, which are aggregated to produce plane-level predictions. The final classification score is obtained by combining per-plane probabilities through a logistic regression layer. Figure 1 shows the network architecture and multi-plane aggregation.

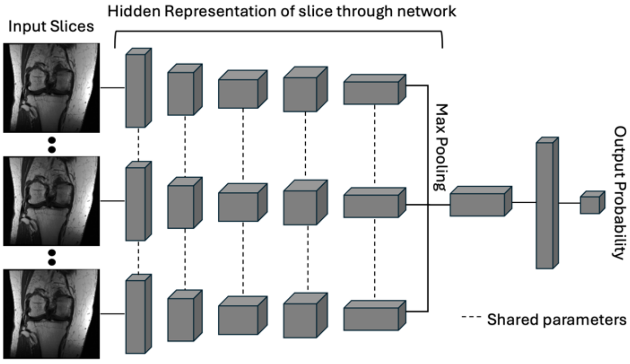


Fig. 1. The model architecture of MRNet [2].

2.2. Public Knee MRI Datasets

Some knee MRI datasets based on functional MRI (fMRI) are primarily intended for image reconstruction research [4]. These datasets often lack clinical labels for injury presence, making them unsuitable for supervised classification tasks.

In contrast, the MRNet dataset [2] includes labeled MRI examinations with annotations for ACL tears, meniscal tears, and abnormalities. Its structured and well-annotated design makes it well suited for training deep learning models aimed at knee injury detection.

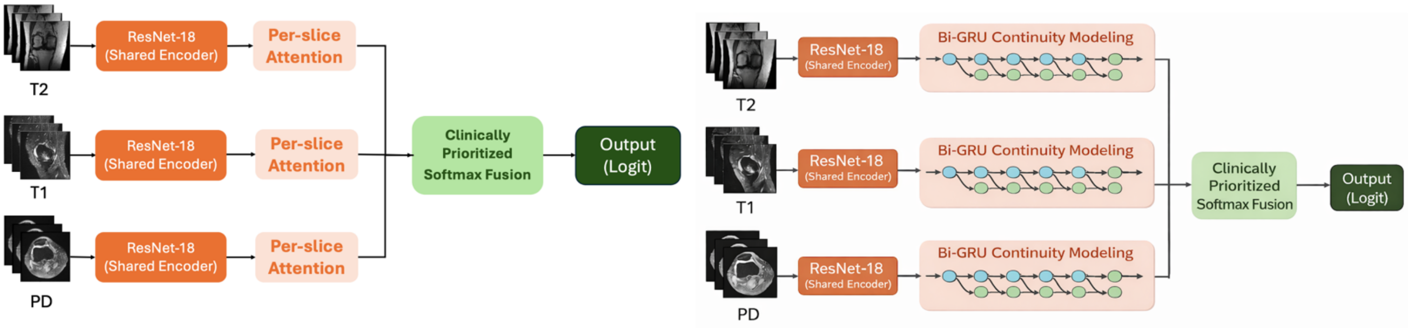
3. Proposed Method

We propose an enhanced MRNet-based framework that incorporates clinically guided sequence-level fusion and slice modeling strategies. The proposed design reflects radiological practice by emphasizing diagnostically important MRI sequences and slices in a task-dependent manner. The overall architecture is illustrated in Figure 2.

3.1. Softmax-weighted Sequence Fusion with Diagnostic Priors

While MRNet simply averages features across MRI sequences, our model introduces learnable fusion weights guided by clinical prior knowledge. Specifically, T1- and T2-weighted images are emphasized for ACL and meniscal tear detection, whereas all sequences are treated as equally informative for abnormality detection.

Let $f_1, f_2, f_3 \in R^d$ denote the feature vectors extracted from the three MRI sequences and let $\alpha = softmax(w) \in R^d$ be the learned fusion weights, normalized such that the sum of all elements equals one. These weights control the relative contribution of each sequence to the final fused representation.



(a) (b)

Fig. 2. Architecture of Proposed Method. (a) Slice selection via attention pooling. (b) Anatomical continuity modeling via Bi-GRU. Both models incorporate clinically prioritized sequence-level fusion.

TABLE 1 AUC SCORES FOR EACH MODEL ACROSS DIAGNOSTIC TASKS

Model	ACL Tear	Meniscus Tear	Abnormality
MRNet (Baseline)	0.852	0.856	0.847
Attention + Clinical Prior(Ours)	0.910	0.899	0.877
Bi-GRU + Clinical Prior(Ours)	**0.934**	**0.924**	**0.918**

The final fused feature vector f_{fused} is computed as

$$f_{fused} = \alpha_1 \cdot f_1 + \alpha_2 \cdot f_2 + \alpha_3 \cdot f_3, \qquad (1)$$

where α_1, α_2, and α_3 represent the fusion weights for the sagittal, coronal, and axial sequences, respectively.

In practice, the initial values of the weight vector w are set according to clinical prior knowledge as follows:

- ACL / Meniscal tear detection: (0.45, 0.45, 0.10)
- Abnormality detection: (0.33, 0.33, 0.33)

3.2. Slice Modeling Strategies

To improve diagnostic accuracy, we propose two complementary strategies for identifying salient slices within each MRI sequence.

3.2.1. Slice-Level Attention Pooling

The per-slice attention mechanism assigns adaptive importance weights to individual slices based on their latent diagnostic relevance. Given slice-level feature vectors f_i, the aggregated plane-level representation is computed as

$$F_{att} = \sum_{i=1}^{N} \alpha_i f_i. \qquad (2)$$

This formulation enables the model to emphasize diagnostically informative slices while suppressing redundant regions, resulting in improved sensitivity to localized abnormalities.

3.2.2. Continuity-aware Modeling with Bi-GRU

To capture anatomical consistency across adjacent slices, we further employ a Bi-GRU to model slice features as an ordered sequence. The network reconstructs each slice feature from its contextual representation, and the **continuity score** is defined as

$$c_i = | f_i - \{\widehat{f}\}_\imath |^2, \qquad (3)$$

where f_i and $\hat{f}_i$ denote the original and reconstructed slice features, respectively. Higher continuity scores indicate anatomical irregularities and guide the model to emphasize pathological regions during aggregation.

4. Experiment

We re-implement MRNet under identical training conditions to ensure a fair comparison. Table 2 reports the AUC results for ACL tear, meniscal tear, and abnormality detection. Both proposed variants consistently outperform the baseline across all tasks, with the Bi-GRU–based continuity-aware model achieving the highest performance, followed by the slice-level attention model.

The largest improvement is observed in ACL tear detection, indicating that modeling inter-slice anatomical continuity is particularly effective for capturing subtle structural abnormalities. These results demonstrate the benefit of incorporating clinical priors and anatomical context for knee MRI classification.

5. Conclusion

In this paper, we propose an enhanced MRNet-based framework with clinically guided fusion and slice modeling for knee MRI classification. The proposed method consistently outperforms the baseline, with the Bi-GRU–based model achieving the best performance, highlighting the benefit of integrating clinical priors and anatomical continuity. This framework offers a general and clinically meaningful approach for multi-sequence medical image classification.

References

[1] D. Shen, G. Wu, and H. Suk, “Deep learning in medical image analysis,” Annual Review of Biomedical Engineering, vol. 19, pp. 221–248, Jun. 2017.

[2] N. Bien, P. Rajpurkar, R. L. Ball, J. Irvin, A. Park, E. Jones, M. Bereket, B. N. Patel, M. P. Lungren, and A. Y. Ng, “Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet,” PLOS Medicine, vol. 15, no. 11, e1002699, Nov. 2018.

[3] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), May 2015, pp. 234–241.

[4] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui, “fastMRI: An open dataset and benchmarks for accelerated MRI,” arXiv preprint, arXiv:1811.08839, Nov. 2018.

Clinically Prioritized Fusion Incorporating Slice-Level Attention and Continuity Modeling for Robust Knee MRI Injury Detection

A Thesis Submitted to the Major in Computer Science and Communications Engineering,
School of Fundamental Science and Engineering of Waseda University
in Partial Fulfillment of the Requirements for the Degree of Master of Engineering

Submission Date: January 26th, 2026

Taira Kunitomi
(5124FG03-2)

Advisor: Prof. Hiroshi Watanabe
Research guidance: Research on Audiovisual Information Processing

Acknowledgement

First and foremost, I would like to sincerely thank my esteemed supervisor Professor Hiroshi Watanabe for providing valuable feedback and suggestions throughout my research process. Additionally, I appreciate the experimental environment devices he provided which were greatly beneficial to my research.

Secondly, I appreciate all the members of Watanabe Lab for their insightful conversations and unforgettable time spent together in the lab.

Last but not least, I would like to express my gratitude to my parents who give me unwavering support and respect for my every decision. Without their understanding and belief in me, it would have been impossible for me to complete my study.

Abstract

Accurate diagnosis of anterior cruciate ligament (ACL) and meniscal tears using magnetic resonance imaging (MRI) is essential for timely and effective treatment. However, existing deep learning approaches often aggregate slice and sequence features without explicitly modeling inter-slice continuity, limiting their ability to localize diagnostically relevant regions.

To address this limitation, we propose a continuity-aware evidence fusion framework that explicitly models inter-slice relationships to identify *where to look* before determining *what to trust* across MRI planes. Our method leverages a bidirectional recurrent architecture to quantify slice-level continuity disruption, producing interpretable evidence scores that guide plane-level feature aggregation. These evidence-driven representations are further integrated using clinically informed fusion priors, reflecting radiologists' diagnostic reasoning across coronal, sagittal, and axial views.

Experiments on the MRNet dataset demonstrate that the proposed framework consistently improves diagnostic performance across ACL tear, meniscal tear, and abnormality detection tasks, while providing enhanced interpretability through explicit modeling of inter-slice continuity.

Keywords: Convolutional neural networks, deep neural networks, medical diagnostic imaging, magnetic resonance imaging.

Contents

Acknowledgement .... *iii*

Abstract .... *iv*

Contents .... *v*

List of Figures .... *vii*

List of Tables .... *viii*

Chapter 1 Introduction *1*

1.1 Introduction **1**

1.2 Thesis Purpose **2**

1.3 Thesis Outline **3**

Chapter 2 Related Works *4*

2.1 Introduction **4**

2.2 Deep Learning Approaches for MRI Diagnoses **4**

2.3 Public Knee MRI Datasets **5**

2.4 Conclusion **7**

Chapter 3 Proposed Method *8*

3.1 Introduction **8**

3.2 Softmax-weighted Sequence Fusion with Diagnostic Priors **9**

3.3 Slice Modelling Strategies **10**

3.3.1 Slice-Level Attention Pooling 11

3.3.2 Inter-Slice Continuity Modeling with Bi-GRU 12

3.4 Conclusion 14

4. Experiments 15

4.1 Introduction 15

4.2 Implementation Details 15

4.2.1 Training 15

4.2.2 Data Loading and Augmentation 15

4.2.3 Optimization and Scheduling 15

4.3. Dataset 16

4.4 Evaluation Metrics 16

4.5 Experimental Results 17

4.6 Ablation Studies 19

4.7 Conclusion 20

Chapter 5 Conclusions and Future Works 21

5.1 Conclusions 21

5.2 Future Works 22

List of Publication 23

Bibliography 24

List of Figures

Fig. 1. The model architecture of MRNet [2]. 5

Fig. 2. Three MRI sequences from the MRNet dataset [2]... 6

Fig. 3. The architecture of Proposed Method. 8

List of Tables

TABLE 1 THE DEMOGRAPHIC INFORMATION OF THE DATASET 6

TABLE 2 DIAGNOSTIC CHARACTERISTICS OF DIFFERENT MRI SEQUENCES 9

TABLE 3 WEIGHTING SCHEMES FOR EACH TASK 10

TABLE 4 AUC SCORES FOR EACH MODEL ACROSS DIAGNOSTIC TASKS.................. 17

TABLE 5 ABLATION STUDY ON ACL TEAR DETECTION (AUC) 19

Chapter 1 Introduction

1.1 Introduction

Knee injuries such as anterior cruciate ligament (ACL) and meniscal tears are prevalent in sports medicine and frequently require magnetic resonance imaging (MRI) for accurate diagnosis [1]. Timely and reliable assessment of these injuries is critical for appropriate treatment planning, rehabilitation, and return-to-play decisions. In recent years, deep learning has demonstrated significant potential in medical image analysis, including the automated detection of musculoskeletal injuries from knee MRI scans. MRI provides excellent soft-tissue contrast without ionizing radiation and enables detailed visualization of ligamentous and cartilaginous structures in the knee [6] [7].

One of the most widely adopted baselines in this domain is MRNet [2], introduced by Stanford University, which performs binary classification of knee injuries using three orthogonal MRI planes: sagittal, coronal, and axial sequences. In MRNet, features are independently extracted from each plane using convolutional neural networks (CNNs), and the final prediction is obtained by uniformly aggregating these plane-level features. While this approach achieves strong baseline performance, it implicitly assumes that all slices and all planes contribute equally to diagnosis.

However, this assumption does not align with clinical practice. Radiologists interpret knee MRI by selectively focusing on diagnostically relevant slices and by weighting information across planes based on the suspected pathology. For example, ACL and meniscal tears are primarily evaluated using sagittal and coronal views, while axial slices often play a secondary role.

Furthermore, even within a single MRI sequence, only a limited subset of slices may contain pathological evidence, whereas many slices provide little or no diagnostic value. Ignoring this variability can dilute critical information and reduce diagnostic robustness.

To address these limitations, recent studies have explored attention mechanisms and slice selection strategies to emphasize informative regions or sequences. While these methods improve performance, most of them treat slice importance as a static or appearance-based property and do not explicitly model the inter-slice continuity inherent to volumetric MRI data. In clinical settings, pathological abnormalities often manifest as localized disruptions across consecutive slices, rather than isolated anomalies in a single slice. Failure to capture such continuity patterns can limit both interpretability and generalization.

1.2 Thesis Purpose

In this work, we propose a multi-view knee MRI classification framework that integrates clinical knowledge into both slice-level evidence modeling and sequence-level feature fusion. Rather than treating slice importance as a static property, the proposed method explicitly models inter-slice continuity to identify diagnostically salient regions, enabling the model to determine where to look by detecting localized disruptions across consecutive slices.

The resulting slice-level evidence is aggregated into plane-level representations, which are further combined using a learnable sequence fusion module guided by clinically informed priors reflecting the diagnostic relevance of each MRI view. By decoupling evidence localization from plane-wise decision fusion, the proposed framework selectively emphasizes meaningful features across both dimensions in a clinically interpretable manner.

1.3 Thesis Outline

The structure of this paper is as follows.

Chapter 1 is the main chapter and describes the background and purpose of this study.

Chapter 2 describes the research related to this study.

Chapter 3 describes the proposed methodology of this study.

Chapter 4 describes the results and discussion of the evaluation experiments of the proposed method.

In Chapter 5, we present the conclusions and future issues of this study.

Chapter 2 Related Works

2.1 Introduction

This section reviews prior work on automated knee injury diagnosis using magnetic resonance imaging (MRI), with a particular focus on deep learning-based multi-plane analysis frameworks.

2.2 Deep Learning Approaches for MRI Diagnoses

Deep learning–based methods are increasingly effective for automating MRI interpretation in musculoskeletal imaging. Among them, U-Net [3], an encoder–decoder segmentation model, demonstrates strong performance in lesion localization tasks such as brain tumour and musculoskeletal abnormality detection.

For classification-based diagnosis, MRNet [2], developed at Stanford University, is a widely used framework for knee injury assessment. The model independently analyzes sagittal, coronal, and axial MRI sequences using convolutional neural networks to extract slice-level features, which are aggregated to predict clinically important conditions such as ACL and meniscal tears. Owing to its standardized design and public availability, MRNet serves as a common baseline in multi-view MRI classification studies.

Figure 1 shows the model architecture of MRNet. Each view is processed by a 2D CNN to extract slice-level features, which are pooled and passed through fully connected layers. The final prediction is obtained by combining outputs from all views.

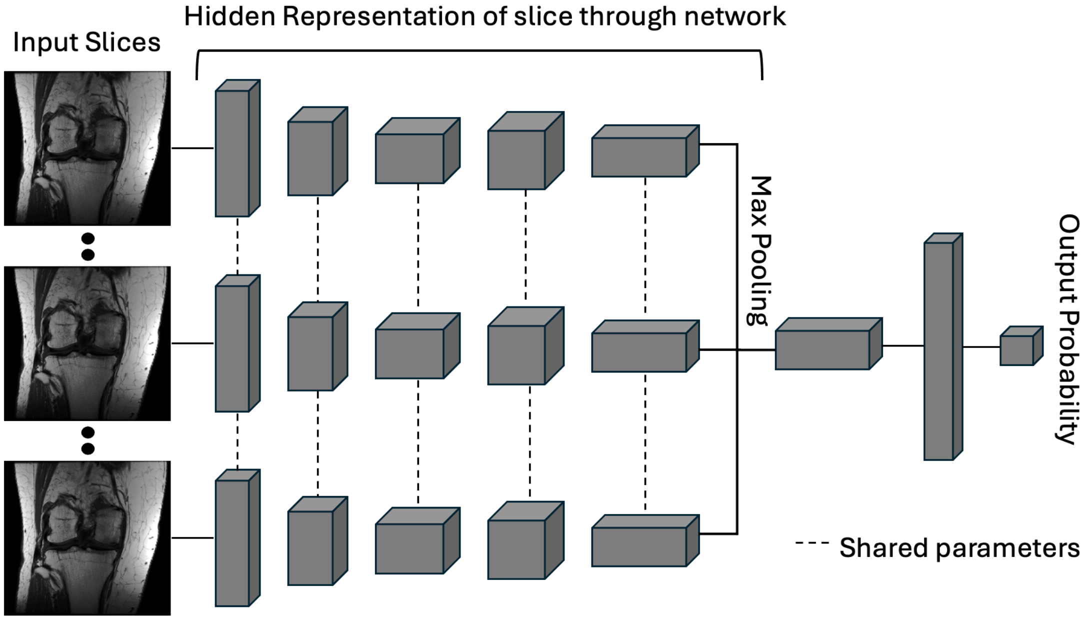


Fig. 1. The model architecture of MRNet [2].

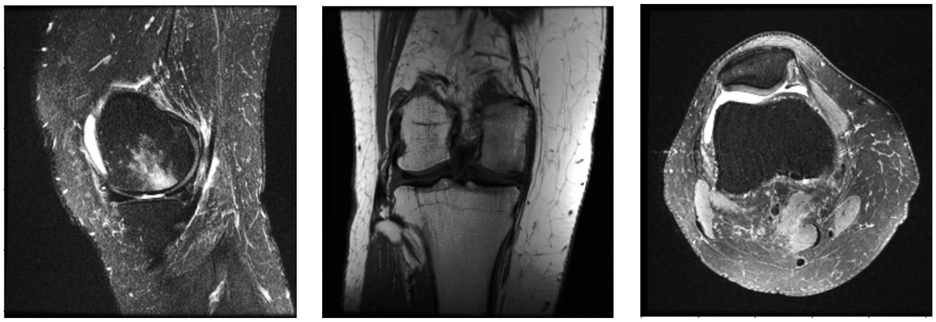
2.3 Public Knee MRI Datasets

Some publicly available knee MRI datasets, such as those based on functional MRI (fMRI), have been primarily developed for tasks in image reconstruction or signal modelling [4]. These datasets typically lack clinical annotations regarding injury presence or type, limiting their utility for supervised learning in diagnostic applications.

In contrast, the MRNet dataset [2] provides a large-scale, clinically annotated collection of knee MRI exams labelled for anterior cruciate ligament (ACL) tears, meniscal tears, and general abnormalities. Each exam includes three orthogonal MRI sequences: sagittal T2-weighted, coronal T1-weighted, and axial proton density (PD)-weighted: capturing complementary diagnostic information across anatomical planes. Representative examples of these MRI sequences are shown in Figure 2, where the sagittal T2-weighted, coronal T1-weighted, and axial PD-weighted images are displayed from left to right.

The dataset encompasses a diverse patient population and imaging variations across sequences. Detailed statistics, including the number of exams, patient demographics, and sequence distribution, are provided in Table 1. These characteristics make MRNet a suitable and widely adopted benchmark for multi-view knee MRI classification.

The MRNet dataset was obtained from the official Stanford ML Group website (https://stanfordmlgroup.github.io/competitions/mrnet/) and used under the terms of the MRNet Dataset Research Use Agreement, which permits non-commercial research use only and prohibits redistribution or commercial use.



(a) Sagittal T2 (b) Coronal T1 (b) Axial PD

Fig. 2. Three MRI sequences from the MRNet dataset [2].

(a) Sagittal T2-weighted. (b) Coronal T1-weighted. (c) Axial PD-weighted views. These orthogonal planes are used as input for classification models in multi-view knee MRI analysis.

TABLE 1 THE DEMOGRAPHIC INFORMATION OF THE DATASET

Statistics	**Training**	**Validation**
Number of exams	1130	120
Number of patients	1088	111
Number of female patients (%)	480 (42.5%)	50 (41.7%)
Age, mean (SD)	38.3 (16.9)	36.3 (16.9)

2.4 Conclusion

This section reviews representative deep learning–based frameworks for knee MRI diagnosis and publicly available datasets supporting such studies. While multi-plane classification models such as MRNet effectively integrate information from multiple anatomical views and serve as strong baselines, they generally process individual slices independently and rely on simple feature aggregation. Consequently, anatomical continuity across adjacent slices is not explicitly modelled. This limitation motivates the proposed approach, which aims to capture inter-slice anatomical transitions to further improve diagnostic robustness and sensitivity.

Chapter 3 Proposed Method

3.1 Introduction

We propose an enhanced model of MRNet that integrates clinically prioritized multi-plane fusion and inter-slice continuity modelling using Bi-GRU. This allows the model to consider both diagnostic relevance and anatomical consistency across slices. The overall architecture is shown in Figure 3.

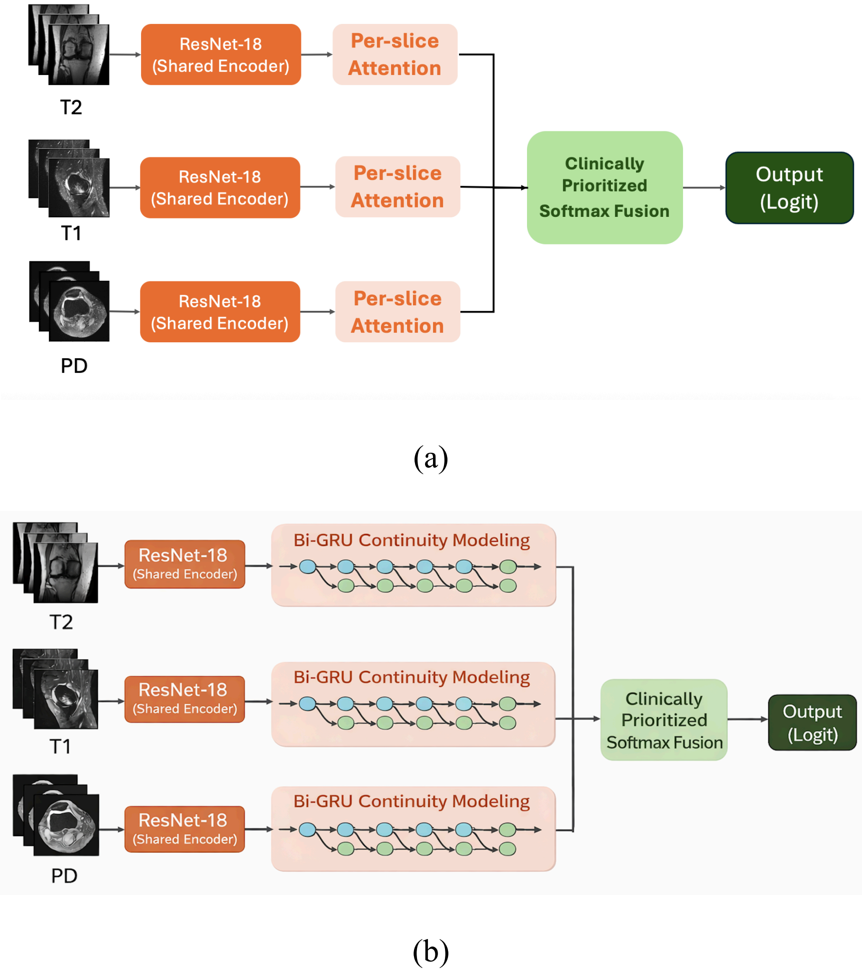


Fig. 3. The architecture of Proposed Method. (a) Slice selection via attention pooling. (b) Anatomical continuity modeling via Bi-GRU. Both models incorporate clinically prioritized sequence-level fusion.

3.2 Softmax-weighted Sequence Fusion with Diagnostic Priors

While MRNet [2] simply averages features from the sagittal, coronal, and axial MRI sequences, our model introduces learnable fusion weights to capture the task-specific diagnostic relevance of each sequence. These weights are initialized using established clinical knowledge and further optimized during training to adapt to different diagnostic tasks.

Each MRI sequence provides complementary diagnostic information. T1-weighted images offer detailed visualization of bone and soft tissue anatomy and are useful for assessing structural abnormalities, T2-weighted images are sensitive to fluid-related pathological changes associated with acute injuries such as ligament tears, and PD-weighted images enhance tissue interface contrast, facilitating the detection of subtle structural irregularities [5].

These complementary roles are summarized in Table 2. By incorporating such modality-specific knowledge into the fusion process, our model emphasizes diagnostically salient features across imaging planes, thereby better reflecting radiological interpretation strategies and improving classification reliability.

TABLE 2 DIAGNOSTIC CHARACTERISTICS OF DIFFERENT MRI SEQUENCES

MRI Sequences	**Diagnostic Strength**
T1-Weighted	Clearly shows anatomical structures of bones and soft tissues. Useful for identifying bone morphology and structural abnormalities.
T2-Weighted	Highlights tissues with high water content (e.g., inflammation, swelling, hematoma). Effective for detecting abnormalities in soft tissues.
PD-Weighted	Emphasizes contrast between bones and soft tissues. Suitable for detecting subtle boundary abnormalities.

Based on these clinical characteristics, we initialize the fusion weights to prioritize sagittal and coronal sequences (T2- and T1-weighted images) for tasks such as ACL and meniscal tear detection, where soft tissue integrity and joint structure are critical. For general abnormality detection, where diverse pathologies may appear across all planes, we adopt a uniform initialization to reflect the balanced diagnostic relevance of each sequence.

Let f_1, f_2, and f_3 denote the feature vectors extracted from the three MRI sequences, and let α_1, α_2, and α_3 represent the corresponding normalized fusion weights obtained via a softmax function, satisfying $\alpha_1 + \alpha_2 + \alpha_3 = 1$. The final fused feature vector is computed as a weighted linear combination

$$f_{fused} = \alpha_1 \cdot f_1 + \alpha_2 \cdot f_2 + \alpha_3 \cdot f_3. \tag{1}$$

In practice, the initial fusion weights are summarized in Table 3.

TABLE 3 WEIGHTING SCHEMES FOR EACH TASK

Task	Weighting Scheme (T1/T2/PD)
ACL Tear	(0.45/0.45/0.1)
Meniscus Tear	(0.45/0.45/0.1)
Abnormality	(0.33/0.33/0.33)

3.3 Slice Modelling Strategies

In addition to sequence-level fusion, our framework incorporates slice modeling strategies to identify diagnostically important slices within each MRI sequence. We employ slice-level attention pooling to capture local saliency and Bi-GRU–based continuity modeling to account for anatomical consistency across adjacent slices. These strategies enable robust feature aggregation and improve sensitivity to spatially coherent abnormalities.

3.3.1 Slice-Level Attention Pooling

In conventional slice-based MRI classification frameworks such as MRNet, a fixed aggregation method, typically max or average pooling, is employed to compress slice-wise features into a single sequence representation. This approach implicitly assumes that all slices contribute equally to the final prediction. However, in clinical practice, radiologists often focus on a limited subset of diagnostically informative slices, such as those capturing ligament tears or abnormal joint morphology, while ignoring irrelevant or redundant views.

To better reflect this clinical reasoning process, we introduce a slice-level attention pooling mechanism that adaptively weights each slice according to its latent diagnostic relevance.

Let the sequence of slice features be denoted as

$$\{f_1, f_2, \dots, f_N\},\ \ f_i \in R^d, \tag{2}$$

where N is the number of slices and d is the feature dimension.

For each slice feature f_i, a scalar attention score α_i is computed using a shallow feed-forward network with tanh activation

$$a_1 \ = \ W_2 \cdot \tanh(W_2 f_1 + \ b_1) + \ b_2. \tag{3}$$

The attention scores are normalized across the slice dimension using the softmax function

$$\alpha_i = \frac{\exp(a_i)}{\Sigma_{\{j=1\}}^{\{N\}\exp}\left(a_j\right)}. \tag{4}$$

The final sequence-level feature representation F is obtained as a weighted sum of the slice features

$$F \ = \sum_{i=1}^{N} \alpha_i \, f_i. \tag{5}$$

This attention mechanism allows the model to selectively emphasize slices containing pathology, such as torn ligaments or displaced menisci, while suppressing irrelevant frames.

Unlike fixed pooling strategies, this learnable approach enhances sensitivity to subtle abnormalities and improves interpretability by explicitly revealing which slices contribute most to the model's prediction.

Empirically, we observe that attention weights often concentrate on clinically meaningful regions—for example, midsagittal slices showing the intercondylar notch in ACL tear detection—demonstrating that our model aligns well with diagnostic behavior in real-world radiology.

3.3.2 Inter-Slice Continuity Modeling with Bi-GRU

While slice-level attention focuses on selecting diagnostically salient slices, many musculoskeletal injuries manifest as gradual structural changes spanning multiple adjacent slices. For example, anterior cruciate ligament (ACL) ruptures or meniscal tears often appear as continuous anatomical disruptions rather than isolated abnormalities in a single slice.

To capture such structural consistency, we model each MRI sequence as an ordered slice series and introduce an inter-slice continuity modeling module based on a bidirectional gated recurrent unit (Bi-GRU). This module explicitly learns how anatomical structures evolve across slices by incorporating contextual information from both preceding and succeeding slices.

Let the sequence of slice features be denoted as (2).

where N is the number of slices and d is the feature dimension. A forward GRU processes the sequence from the first slice to the last

$$h_i^{\rightarrow} = GRU_{forward\left(f_i, h^{\rightarrow}_{\{i-1\}}\right)}, \tag{6}$$

and a backward GRU processes the sequence in the reverse order

$$h_i = [\, h_i^{\rightarrow};\ h_i^{\leftarrow}]. \tag{7}$$

To derive a sequence-level feature representation, we apply temporal average pooling over all continuity-aware slice representations

$$F = \left(\frac{1}{N}\right)\sum_{i=1}^{N} h_i. \tag{8}$$

This representation encodes both local slice appearance and global anatomical transitions across the entire sequence.

To further quantify anatomical consistency across slices, we introduce a continuity score that measures how well each slice conforms to its surrounding anatomical context as modeled by the Bi-GRU.

After obtaining the continuity-aware hidden representations$h_1, h_2, \ldots, h_N$, we reconstruct each slice representation from its contextual embedding using a linear projection layer

$$\hat{f}_i = W_r h_i + b_r, \tag{9}$$

where $\hat{f}_i$ denotes the reconstructed feature of slice i, and W_r and b_r are trainable parameters.

The continuity scorec_i for slice i is defined as the reconstruction error between the original slice feature f_i and its reconstructed feature $\hat{f}_i$

$$c_i = || f_i - \hat{f}_i ||_2. \tag{10}$$

This score reflects the degree of structural inconsistency between the slice and its anatomical context.

Slices with low continuity scores indicate anatomically consistent transitions, whereas high continuity scores suggest abnormal structural deviations, such as ligament discontinuities or distorted meniscal morphology.

To obtain a sequence-level continuity score, we aggregate slice-level scores by averaging across all slices

$$C = \left(\frac{1}{N}\right)\sum_{i=1}^{N} c_i, \tag{11}$$

here C represents the overall continuity deviation of the MRI sequence.

By explicitly modeling slice dependencies and quantifying structural consistency, the proposed Bi-GRU module enables the network to detect continuity violations that are difficult to capture using independent slice processing or attention-based selection alone.

Compared to slice-level attention, which emphasizes where abnormalities appear, continuity modeling focuses on how anatomical structures evolve across slices. This complementary perspective allows the proposed framework to better reflect radiological reasoning and improves robustness in detecting subtle or spatially distributed knee injuries.

3.4 Conclusion

This section introduced two slice modeling strategies: attention-based slice selection and Bi-GRU-based continuity modeling. While the former emphasizes localized pathological evidence, the latter captures structural transitions across slices. Their combination enables both saliency-aware and structure-aware MRI representation, forming a robust basis for clinically prioritized sequence-level fusion and final classification.

4. Experiments

4.1 Introduction

This section outlines the experimental setup and evaluation protocol for the proposed methods.

4.2 Implementation Details

4.2.1 Training

All experiments were conducted on the MRNet [2] dataset using three MRI planes (coronal, sagittal, axial). To ensure a fair comparison, we trained all methods under identical experimental conditions and used a re-implementation of the original MRNet pipeline as the baseline.

4.2.2 Data Loading and Augmentation

Each MRI exam was loaded with batch size 1 and represented as a set of slices for each plane. During training, we applied data augmentation consisting of random rotation (±12°), random affine translation (up to 3%), and horizontal flipping ($p = 0.1$). Since the backbone encoder is an ImageNet-pretrained ResNet-18, each grayscale slice was replicated to three channels to match the expected input format.

4.2.3 Optimization and Scheduling

We optimized the model using Adam with an initial learning rate of 3×10^(-5) and weight decay of 1×10^(-4). A ReduceLROnPlateau scheduler was used to adapt the learning rate based on the validation loss (patience = 2, factor = 0.5, minimum learning rate = 3×10^(-6)). Training was performed for up to 50 epochs with early stopping (patience = 4) based on validation AUC improvement. Gradient clipping (max norm = 1.0) was applied to stabilize training, and batch normalization layers in the ResNet backbone were frozen to reduce overfitting.

4.3. Dataset

All experiments were conducted using the publicly available MRNet dataset, a benchmark dataset for automated knee MRI diagnosis annotated for anterior cruciate ligament (ACL) tears, meniscal tears, and gene ral abnormalities across the entire sequence.

Each examination consists of three orthogonal MRI sequences: coronal T1-weighted, sagittal T2-weighted, and axial proton density (PD)-weighted images.

We followed the official data split provided by the dataset authors, which divides the dataset into 1,130 training cases, 120 validation cases, and 120 test cases with no patient-level overlap. This protocol was adopted throughout our experiments to ensure reproducibility and fair comparison with prior studies.

4.4 Evaluation Metrics

Following the evaluation protocol of the original MRNet study, model performance was assessed using the area under the receiver operating characteristic curve (AUC). AUC is a threshold-independent metric that quantifies the ability of a classifier to distinguish between positive and negative samples across all possible decision thresholds and is widely adopted in medical image analysis.

Formally, AUC can be interpreted as the probability that a randomly selected positive sample is assigned a higher prediction score than a randomly selected negative sample

$$AUC = P\big(s(x^{+}) > s(x^{-})\big), \tag{12}$$

where s(·) denotes the model's predicted score, x^{+} represents a positive sample, and x^{-} represents a negative sample.

This definition is equivalent to the ROC-based computation implemented in our experiments and enables robust evaluation under class imbalance, while allowing direct comparison with previously reported results on the MRNet benchmark.

4.5 Experimental Results

We report experimental results on the MRNet dataset and compare the proposed methods with the original MRNet baseline. The evaluation focuses on the effectiveness of clinically prioritized fusion, attention-based slice selection, and Bi-GRU-based continuity modeling. Performance is measured using AUC for three diagnostic tasks: ACL tear, meniscal tear, and abnormality detection.

TABLE 4 AUC SCORES FOR EACH MODEL ACROSS DIAGNOSTIC TASKS

Model	ACL Tear	Meniscus Tear	Abnormality
MRNet (Baseline)	0.852	0.856	0.847
Attention + Clinical Prior (Ours)	0.910	0.899	0.877
Bi-GRU + Clinical Prior (Ours)	**0.934**	**0.924**	**0.918**

Table 4 reports the AUC performance of the MRNet baseline and the proposed methods on ACL tear detection, meniscal tear detection, and abnormality detection

By introducing attention-based slice selection with clinically prioritized fusion, the proposed method significantly outperforms the baseline across all tasks, achieving improvements of +0.058, +0.043, and +0.030 AUC for ACL tear, meniscal tear, and abnormality detection, respectively. This demonstrates that incorporating domain knowledge into the sequence-level fusion process effectively enhances the model's ability to focus on diagnostically relevant MRI planes.

More notably, the proposed Bi-GRU-based continuity modeling further improves performance in a consistent and substantial manner, reaching AUC scores of 0.934 for ACL tear detection, 0.924 for meniscal tear detection, and 0.918 for abnormality detection. Compared to the MRNet baseline, this corresponds to relative improvements of +0.082, +0.068, and +0.071, respectively.

These results highlight the importance of explicitly modeling anatomical continuity across adjacent slices. Unlike attention-based slice selection, which emphasizes localized pathological evidence, continuity modeling captures gradual structural disruptions that span multiple slices, such as ligament fiber discontinuity or progressive meniscal deformation. This capability is particularly critical for musculoskeletal MRI interpretation, where abnormalities often manifest as spatially coherent patterns rather than isolated appearances.

Overall, the experimental results indicate that the integration of clinically prioritized fusion and continuity-aware slice modeling provides a robust and generalizable framework for knee injury detection, substantially outperforming the conventional slice-aggregation strategy adopted in MRNet, particularly in its ability to capture subtle and spatially coherent pathological patterns across MRI sequences.

4.6 Ablation Studies

To further assess the contribution of each MRI sequence to model performance, we conducted a targeted ablation study focusing on the ACL tear classification task. Specifically, we compared the proposed learnable fusion strategy with three fixed-weight configurations, each constructed by omitting one sequence through assigning its fusion weight to zero. Table 5 reports the AUC results for each configuration.

TABLE 5 ABLATION STUDY ON ACL TEAR DETECTION (AUC)

Fusion Weights (T1, T2, PD)	AUC	$\Delta\ vs. Ours$
(0.5, 0.5,0.0)	0.878	-0.032
(0.0, 0.5, 0.5)	0.864	-0.046
(0.5, 0.0, 0.5)	0.867	-0.043
Attention + Clinical Priors (Ours)	0.910	-

Among the fixed settings, the configuration with equal weighting of T1- and T2-weighted sequences while excluding PD (0.5, 0.5, 0.0) achieved the highest AUC of 0.878. Nevertheless, this result remained inferior to that of the proposed adaptive fusion method, which achieved an AUC of 0.910. This performance gap highlights the limitations of manually predefined fusion weights and demonstrates the advantage of data-driven optimization in capturing task-specific diagnostic relevance across MRI sequences.

Furthermore, performance consistently deteriorated when any single modality was excluded, indicating that all three sequences: T1, T2, and PD: provide complementary and non-redundant diagnostic information. T1- and T2-weighted images contribute critical anatomical and pathological cues related to ligament integrity, edema, and joint effusion, which are highly relevant for ACL tear detection. Although PD-weighted images are often considered less dominant in clinical assessment, their enhanced contrast at soft tissue–bone interfaces appear to provide auxiliary information that supports discrimination of subtle abnormalities.

Notably, the reduced performance of the (0.5, 0.5, 0.0) configuration suggests that even weakly weighted modalities can supply complementary evidence that improves overall decision making. These findings validate our design choice to adopt an attention-based fusion strategy that dynamically learns task-specific sequence weights, enabling optimal integration of multi-sequence information. Such adaptivity is particularly beneficial for complex musculoskeletal MRI analysis, where pathological patterns may manifest across multiple anatomical planes and contrast mechanisms.

4.7 Conclusion

This section demonstrates that the proposed framework consistently outperforms the baseline MRNet across all evaluation tasks. In particular, the continuity-aware model based on Bi-GRU achieves the highest performance, indicating the effectiveness of modeling anatomical dependencies across slices. These results confirm that incorporating both clinically guided sequence fusion and slice modeling strategies leads to more discriminative and robust representations for knee MRI classification.

Chapter 5 Conclusions and Future Works

5.1 Conclusions

In this study, we propose a clinically informed framework for knee MRI classification that integrates adaptive slice modeling with sequence-level fusion to improve diagnostic accuracy and robustness. The framework combines attention-based slice selection and Bi-GRU–based inter-slice continuity modeling with clinically prioritized fusion weights.

By incorporating clinical prior knowledge: such as emphasizing T1- and T2-weighted sequences for ligament-related injuries: the model learns task-specific fusion strategies consistent with radiological practice. In addition, continuity modeling captures anatomical transitions across adjacent slices via a reconstruction-based score, enabling the detection of gradual pathological changes that are difficult to identify using independent slice analysis.

Experiments on the MRNet dataset demonstrate consistent performance improvements over the baseline across all diagnostic tasks, with the Bi-GRU: based model achieving the highest AUC. An ablation study further confirms the complementary roles of multiple MRI sequences and the advantage of adaptive fusion over fixed weighting schemes.

Overall, the proposed architecture aligns with radiological reasoning by learning where to attend within sequences, how structures evolve across slices, and which modalities are most informative for each task, thereby improving both classification performance and clinical interpretability.

5.2 Future Works

Future research will focus on enhancing the interpretability and generalizability of the proposed framework. Developing visualization techniques for slice-level attention and continuity scores is expected to facilitate deeper understanding of model behavior and improve clinical trust.

In addition, external validation on datasets acquired under heterogeneous imaging conditions will be necessary to establish robustness across domains. Extending the proposed approach to other musculoskeletal regions may further demonstrate its broader applicability.

Finally, investigating the integration of the framework into clinical workflows and its effect on diagnostic accuracy and efficiency represents an important direction toward real-world deployment.

List of Publication

[1] Taira Kunitomi, Taiga Hayami, and Hiroshi Watanabe. “Clinically prioritized attention-based fusion of multi-plane knee MRI for robust injury detection,” in 2025 IEEE 14th Global Conference on Consumer Electronics (GCCE), pp. 485-488, Sep. 2025. DOI:10.1109/GCCE65946.2025.11274947

Bibliography

[1] D. Shen, G. Wu, and H. Suk, “Deep learning in medical image analysis,” *Annual Review of Biomedical Engineering*, vol. 19, pp. 221-248, Jun. 2017.

[2] N. Bien, P. Rajpurkar, R. L. Ball, J. Irvin, A. Park, E. Jones, M. Bereket, B. N. Patel, M. P. Lungren, and A. Y. Ng, “Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet,” *PLOS Medicine*, vol. 15, no. 11, e1002699, Nov. 2018.

[3] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, May 2015, pp. 234-241.

[4] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui, “fastMRI: An open dataset and benchmarks for accelerated MRI,” *arXiv preprint* arXiv:1811.08839, Nov. 2018.

[5] M. P. Recht, R. E. Kramer, P. J. Marcelis, and R. A. Daffner, “MR imaging of the knee: Current status and future directions,” *American Journal of Roentgenology*, vol. 167, no. 3, pp. 593-600, Sep. 1996.

[6] J. R. Haaga, C. F. Lanzieri, and R. C. Gilkeson, CT and MRI of the Whole Body, 5th ed., vols. 1-2. Philadelphia, PA, USA: Elsevier, Mar. 2008.

[7] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *MRI from Picture to Proton*, 3rd ed. Cambridge University Press, Apr. 2017.