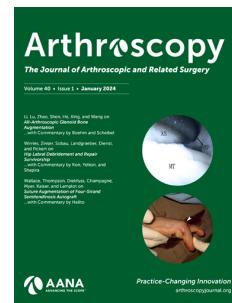


Journal Pre-proof

Transformer-based Multi-label Deep Learning Model is Efficient for Detecting Ankle Lateral and Medial Ligament Injuries on MRI and Improving Clinicians' Diagnostic Accuracy for Rotational Chronic Ankle Instability

Rui Yin, PhD, Hao Chen, PhD, Changjiang Wang, MS, Chaoren Qin, MS, Tianqi Tao, MD, Yunjia Hao, MD, Rui Wu, MD, Yiqiu Jiang, MD, Jianchao Gui, MD



PII: S0749-8063(24)00409-2

DOI: <https://doi.org/10.1016/j.arthro.2024.05.027>

Reference: YJARS 59089

To appear in: *Arthroscopy: The Journal of Arthroscopic and Related Surgery*

Received Date: 10 November 2023

Revised Date: 11 May 2024

Accepted Date: 19 May 2024

Please cite this article as: Yin R, Chen H, Wang C, Qin C, Tao T, Hao Y, Wu R, Jiang Y, Gui J, Transformer-based Multi-label Deep Learning Model is Efficient for Detecting Ankle Lateral and Medial Ligament Injuries on MRI and Improving Clinicians' Diagnostic Accuracy for Rotational Chronic Ankle Instability, *Arthroscopy: The Journal of Arthroscopic and Related Surgery* (2024), doi: <https://doi.org/10.1016/j.arthro.2024.05.027>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier on behalf of the Arthroscopy Association of North America

Transformer-based Multi-label Deep Learning Model is Efficient for Detecting Ankle Lateral and Medial Ligament Injuries on MRI and Improving Clinicians' Diagnostic Accuracy for Rotational Chronic Ankle Instability

Rui Yin, PhD¹, Hao Chen, PhD^{2,3}, Changjiang Wang, MS¹, Chaoren Qin, MS¹, Tianqi Tao, MD¹, Yunjia Hao, MD^{1,4}, Rui Wu, MD^{1,5}, Yiqiu Jiang, MD¹, Jianchao Gui, MD^{1*}

Affiliations:

¹Department of Sports Medicine and Joint Surgery, Nanjing First Hospital, Nanjing Medical University, Nanjing, China

²Department of Clinical Neuroscience, Cambridge University, Cambridge, UK

³School of Computer Science, University of Birmingham, Birmingham, UK

⁴Department of Hand and Foot Microsurgery, Xuzhou Central Hospital

⁵Department of Orthopedics, The Second People's Hospital of Lianyungang

Running title: AnkleNet for CAI Diagnosis

Corresponding Author:

*Jianchao Gui

Professor and chief of the department

Department of Sports Medicine and Joint Surgery, Nanjing First Hospital, Nanjing Medical University

Address: Changle Road 68, Nanjing 210006, China

Telephone number: 025-52271250

E-mail: gui1997@126.com

Funding/Disclosures: This study does not have any prior or duplicate submissions or publications elsewhere of any part of the work. The work was supported by National Natural Science Foundation of China [no. 81672210].

Authors' contribution:

1. Conception and design of the study: Rui Yin, Hao Chen, Jianchao Gui
Acquisition of data: RuiYin, Changjing Wang, Chaoren Qin, Tianqi Tao, Yunjia Hao, Wu
Rui, Yiqiu Jiang
Analysis, or interpretation of data: Rui Yin, Hao Chen, Jianchao Gui
2. Drafting the work: Rui Yin
Revising it critically for important intellectual content: Jianchao Gui
3. Final approval of the version to be published: All authors
4. Agreement to be accountable for all aspects of the work in ensuring that questions related
to the accuracy or integrity of any part of the work are appropriately investigated and
resolved: All authors

Social Media:

Rui Yin

- Twitter: https://twitter.com/chiari_ray

Hao Chen

- LinkedIn: <https://www.linkedin.com/in/hao-chen-cs>

1 **Transformer-based Multi-label Deep Learning Model is Efficient for**
2 **Detecting Ankle Lateral and Medial Ligament Injuries on MRI and**
3 **Improving Clinicians' Diagnostic Accuracy for Rotational Chronic**
4 **Ankle Instability**

5

6 **ABSTRACT**

7 **Purpose:** To develop a deep learning (DL) model that can simultaneously detect lateral and medial
8 collateral ligament injuries of the ankle, aiding in the diagnosis of chronic ankle instability (CAI),
9 and assess its impact on clinicians' diagnostic performance.

10 **Methods:** DL models were developed and external validated on retrospectively collected ankle
11 MRIs between April 2016 and March 2022 respectively at three centers. Included patients were
12 confirmed diagnoses of CAI through arthroscopy, as well as individuals who had undergone MRI
13 and physical examinations that ruled out ligament injuries. DL models were constructed based on
14 a multi-label paradigm. A transformer-based multi-label DL model (AnkleNet) was developed and
15 compared with four convolution neural network (CNN) models. Subsequently, a reader study was
16 conducted to evaluate the impact of model assistance on clinicians when diagnosing challenging
17 cases: identifying rotational CAI (RCAI). Diagnostic performance was assessed using area under
18 the receiver operating characteristic curve (AUC).

19 **Results:** Our transformer-based model achieved AUC of 0.910 and 0.892 for detecting lateral and
20 medial collateral ligament injury, respectively, both of which was significantly higher than that of
21 CNN-based models (all $P < 0.001$). In terms of further CAI diagnosis, it exhibited a macro-average
22 AUC of 0.870 and a balanced accuracy of 0.805. The reader study indicated that incorporation
23 with our model significantly enhanced the diagnostic accuracy of clinicians ($P = 0.042$),
24 particularly junior clinicians, and led to a reduction in diagnostic variability. The code of the model
25 can be accessed at <https://github.com/ChiariRay/AnkleNet>.

26 **Conclusion:** Our transformer-based model was able to detect lateral and medial collateral ligament
27 injuries based on MRI and outperformed CNN-based models, demonstrating a promising
28 performance in diagnosing CAI, especially RCAI patients.

29

30 **Clinical Relevance:** Developing such algorithm can improve the diagnostic performance of
31 clinicians, aiding in identifying patients who would benefit from arthroscopy, such as RCAI
32 patients.

33 **Keywords:** Chronic ankle instability, Deep learning, Artificial intelligence, Transformer,
34 Rotational chronic ankle instability.

35

36 INTRODUCTION

37 Chronic ankle instability (CAI) is a significant health concern, with up to 40% of initial ankle
38 sprain cases developing into CAI, potentially leading to long-term complications^{1,2}. Given that
39 CAI is a significant risk factor contributing to approximately 80% of posttraumatic osteoarthritis
40 in the ankle, it becomes imperative to emphasize early diagnosis and prompt treatment of CAI as
41 a crucial measure for preventing ankle osteoarthritis (AOA)²⁻⁴. However, improved understanding
42 in CAI has led to a focus no longer merely on the lateral collateral ligament of the ankle, and
43 simple diagnosing patient as CAI cannot meet the requirement of clinical practice. Consequently,
44 CAI has been further elucidated into more nuanced patterns⁵⁻⁷. Injuries to the lateral collateral
45 ligament that result in recurrent inversion sprains are termed lateral chronic ankle instability
46 (LCAI); medial collateral ligament (the deltoid ligament) lesions causing medial giving away
47 sensation or pain received the medial chronic ankle instability (MCAI) nomenclature. Additionally,
48 although the primary site of injury typically involves the lateral collateral ligament, approximately
49 40% of patients with LCAI also exhibit partial medial injury^{8,9}. This combined injury, referred to
50 as rotational CAI (RCAI), has garnered increasing attention in recent years^{5,10}. Kim et al.⁷ reported
51 that individuals with RCAI were 3.8 times more likely to express dissatisfaction compared to those
52 with LCAI. Furthermore, Vega et al.¹¹ noted that patients with RCAI may exhibit medial symptoms
53 that would become apparent if they were treated with an isolated repair of the lateral collateral
54 ligament. These findings underscore the significance of detecting injuries to both the lateral and
55 medial collateral ligaments and conducting a comprehensive preoperative evaluation to diagnose
56 CAI in detail.

57 Accurate diagnosis stands as a pivotal determinant for the success of treatment. Nevertheless,
58 achieving a diagnosis for CAI, especially in patients with RCAI, presents a significant challenge.¹¹.
59 The current clinical options to evaluate CAI beyond history, physical examination, and radiographs
60 is magnetic resonance imaging (MRI). However, MRI possesses inherent limitations such as low
61 inter-observer reliability, reduced sensitivity, and the time-intensive nature of the diagnostic
62 process^{9,12,13}. Moreover, while MRI can visualize the deltoid ligament, Yasuda and his colleagues
63 revealed that in 29 out of 30 ankles, MRI failed to depict deep deltoid striation¹⁴. This observation
64 raises concerns regarding the potential neglect of concurrent medial injuries. Taken together, there
65 is an imminent requirement of a decision support tool for detailed CAI diagnosis.

66 Numerous researches have presented evidence that deep learning (DL) support can increase
67 diagnostic performance and reduce diagnostic variability over clinicians working without
68 support^{15–17}. Although studies involving artificial intelligence in the realm of orthopedic surgery
69 and sports medicine remain relatively limited, DL has attracted significant attention within the
70 field^{18–20}. Previous studies have reported that DL models can automated diagnosis for knee MRI,
71 detecting cartilage defects, meniscal tears, and anterior cruciate ligament (ACL) injuries. However,
72 there is a paucity of research investigating the utilization of DL in diagnosing CAI. Assessing ankle
73 joint injuries through MRI is challenging owing to the complex joint structure and involvement of
74 slender ligaments. This necessitates evaluation of multi-plane imaging simultaneously. Moreover,
75 in contrast to knee, ligaments of the ankle joint are relatively dispersed in MRI planes, requiring a
76 DL model to have a better global receptive field. These factors may lead to suboptimal performance
77 when adopting the approach of fine-tuning CNN-based backbones, which has been the
78 predominant method employed in previous studies.

79 Therefore, we proposed an ankle-specific transformer-based architecture. The purposes of this
80 study were to develop a DL model based on this architecture that can simultaneously detect lateral
81 and medial collateral ligament injuries of the ankle, aiding in the diagnosis of CAI, and assess its
82 impact on clinicians' diagnostic performance. We hypothesize that our transformed-based model
83 could detect the injuries of both lateral and medial collateral ligament accurately, outperforming
84 the general DL models previously used in sports medicine, aiding in the diagnosis of CAI.
85

86 MATERIALS AND METHODS

87 Study population

88 This study was approved by the local institutional review board. We collected data from three
89 medical centers. To create the development set for DL models, ankle MRIs from two centers
90 between April 2016 and March 2022 were retrospectively reviewed. Inclusion criteria were
91 patients who had undergone MRI examination of the ankle joint within three months prior ankle
92 arthroscopy with complete surgical records, while exclusion criteria encompassed rheumatoid
93 arthritis, prior ankle surgery, local ankle fractures, and poor image quality. Additionally,
94 individuals who had undergone outpatient visits and were ruled out for ankle ligament injuries
95 through both MRI and physical examinations were included as normal participants. For model

96 evaluation, MRIs from another medical center were collected with the same inclusion and
 97 exclusion criteria to form the external validation set.

98 In this study, all surgical patients were preoperatively diagnosed with CAI or suspected of
 99 having CAI base on medical history, physical examination, and MRI. Non-RCAI patients
 100 underwent a three-month rehabilitation program supervised by therapists prior to surgery.
 101 Arthroscopy served as the gold standard, and each patient was assigned two labels based on the
 102 arthroscopic examination: one for the lateral collateral ligament (normal/injured) and another for
 103 the medial collateral ligament (normal/injured). The criteria used to determine ligament injury
 104 during arthroscopic examination included: (1) ligament thinning, narrowing, significant decrease
 105 in tension, or discontinuity; (2) absence of ligament or filling of the ligament with fatty tissue or
 106 fibrous scar; or (3) avulsion of the ligament at the attachment. An isolated injury to the lateral
 107 collateral ligament was classified as LCAI, whereas an isolated injury to the medial collateral
 108 ligament was classified as MCAI. Injuries to both lateral and medial collateral ligament was
 109 classified as RCAI. For the enrolled outpatient participants, they may have a history of ankle sprain
 110 accompanied by ankle discomfort. However, they tested negative on the drawer test, and their MRI
 111 exhibited clear and continuous ligaments without significant signal alterations. Therefore, their
 112 labels for the lateral ankle ligament and medial ankle ligament were both classified as normal, and
 113 they were considered as the normal control group.

114 MRI was performed with 1.5 T and 3.0 T MR systems from Philips (Philips Healthcare,
 115 Amsterdam, Netherlands), Siemens (Siemens Healthcare, Forchheim, Germany), and United
 116 Image (United Imaging Intelligence, Shanghai, China) with the ankle coil. For this study, coronal
 117 plane PD-weighted series and axial plane PD-weighted series or T2-weighted series were extracted
 118 from each exam for use in the model.

119 All MRIs were anonymized by pydicom. MRIs in the development set were subdivided into the
 120 training and internal validation sets in an 8:2 ratio. Stratified random sampling was employed to
 121 ensure a consistent label proportion across these sets. Furthermore, examinations from the same
 122 individuals were in the same split to prevent any possible data leakage.

123 Deep learning workflow

124 **Preprocessing.** The flowchart of the DL workflow is depicted in Figure 1A. Prior to feeding
 125 images into the model, several preprocessing steps were applied sequentially to normalize the
 126 dataset, as summarized below. Firstly, each anonymized acquisition was converted to NIFTI

format and resized to 386×386 pixels to facilitate image reading. Since MRI can be affected by local magnetic field variations, the N4 bias field correction was applied²¹. Additionally, a color normalization step was performed, which involved clipping out extreme intensity values in the MRI to eliminate bright artifacts in the images. Furthermore, to enhance the generalization ability of the model, data augmentation was employed. This involved randomly cropping the images to 256×256 pixels, applying rotations and flips, and adding Gaussian noise.

Model architecture. Our transformer-based model, AnkleNet, adopts the architecture of Vision Transformer (ViT)²² to understand and analyze ankle MRI (see Figure 1B). Here's a simpler way to look at it: Think of the model as being a bit like a very attentive student. First, it takes the raw images of ankles and breaks them down into smaller, manageable pieces (image patches) — kind of like taking notes. These “notes” are then passed through a series of steps (we call these transformer blocks) that help the model to piece together a comprehensive understanding of the ankle's structure from different views. Drawing inspiration from how clinicians evaluate MRI, we introduced a special technique named the cross-attention mechanism within these transformer blocks. This technique allows the model to intelligently combine features from different MRI planes — much like how a doctor would compare views from the coronal and the axial plane to get a full evaluation of the ankle joint. This ensures that the model maintains a consistent interpretation of the features from multiple views. At the final stage, the model employs a multilayer perceptron, which functions as a sophisticated decision-maker, to provide multi-label diagnoses based on the analyzed features. Further technical information and training details can be found in Appendix E1.

Model developments and evaluations. To comprehensively evaluated the performance of our transformer-based model, four additional CNN-based models previously utilized in the field of sports medicine were developed for the comparison purpose. These models included MRNet, VGG16, ResNet50, and 3D CNN^{23–26}. Subsequently, the performance of these models and AnkleNet was compared using the external validation set. The model exhibiting the most favorable classification results was selected for further CAI diagnosis.

The diagnosis of CAI is determined based on the multi-label output generated by DL models. In essence, DL models are capable of simultaneously outputting the probabilities of lateral and medial collateral ligament injury. These probabilities are then used to calculate the final

157 probabilities for normal ankle joint, LCAI, MCAI, and RCAI. The category with the highest
158 probability is the ultimate diagnosis.

159 The code used to develop the transformer-based model described herein is publicly available on
160 GitHub (<https://github.com/ChiariRay/AnkleNet>).

161 **Reader study**

162 In clinical practice, identification of potential RCAI within the CAI population holds significant
163 importance due to its direct impact on decision-making. To assess the capability of the DL model
164 in assisting with CAI diagnosis, we conducted a reader study comprising two parts. Among the
165 participants in the reader study, there were two fellowship-trained clinical experts with over 10
166 years of experience in musculoskeletal (MSK) radiology and foot and ankle surgery (J.Y.),
167 respectively. Remaining three participants were junior clinicians with less than five years of
168 experience, consisting of two orthopedists and one radiologist. The orthopaedists were undergoing
169 fellowship training program, specializing in sports medicine, while the radiologist was in the
170 fellowship program of MSK imaging. In part 1, the panel of clinicians was provided with the raw
171 MRI images to determine whether CAI patients were RCAI. In part 2, clinicians from the same
172 panel were given access to the original MRI images, as well as the probabilities of lateral and
173 medial collateral ligament injury, Gradient-weighted Class Activation Mapping (Grad-CAM)
174 heatmaps, and the model's final diagnostic result, all of which are intended to facilitate the
175 diagnostic process. Grad-CAM is a technique used to produce visual explanations for decisions
176 made by DL models, particularly in the context of image classification tasks²⁷. Grad-CAM
177 generates heatmaps that highlight the areas within an image that were most influential in the
178 model's decision-making process. These heatmaps essentially provide a visual representation of
179 the model's focus when predicting a particular class. All RCAI cases in the external validation set
180 were included in the reader study, and an equal number of LCAI patients were randomly selected.
181 The reading experiments were conducted on diagnostic computer monitors, with a two-week gap
182 between the two parts.

183 **Statistical analysis**

184 Statistical analysis was performed with R (version 4.3). The area under the receiver operating
185 characteristic curve (AUC) was computed to assess the discriminative ability of DL models for
186 detecting both lateral and medial collateral ligament injuries. Other performance metrics, such as
187 sensitivity, specificity, and accuracy, were also evaluated. Additionally, one-versus-rest macro-

average AUC and balanced accuracy were documented to illustrate the performance of CAI diagnosis. For all performance metrics, 95% confidence intervals (CIs) were calculated using the bootstrap method. This statistical technique offers a robust way to assess model performance variability and reliability by generating numerous subsets of the original data and evaluating the model on each allowing for a more nuanced understanding of the model's potential effectiveness and stability. The DeLong test was utilized to compare the AUC of our transformer-based model with other models and a one-tailed t-test was used to examine the changes in performance metrics for all clinicians in the two parts of the reader study. Brier score measured the accuracy of probability predictions, indicating how well the model was calibrated (0-1; a score of 0 indicated perfect model calibration, a score of 1 corresponded to poor model calibration)²⁸. Furthermore, inter-observer agreement among the clinicians was evaluated using Fleiss' κ . Statistical significance was defined as $P < 0.05$.

RESULTS

Patient characteristics

In this study, 2267 out of 2573 patients were included in the final analysis (Figure S1). The development set comprised 1,621 individuals, while the external validation set encompassed 646 individuals. No statistically significant differences were observed between the two sets in terms of age, gender, and the proportions of lateral and medial collateral ligament injury. It is noteworthy that a significant class imbalance challenge was encountered in detecting medial collateral ligament injury and diagnosing patient with MCAI. Specifically, MCAI was present in only 3.5% of the development set and 3.3% of the external validation set. Detailed individuals' demographic and ligament injury characteristics are summarized in Table 1.

Performance comparison for detecting ankle ligament injuries

Figure 2 illustrates the ROC curves of DL models. In terms of detecting lateral collateral ligament injury, the AUC values for MRNet, VGG16, ResNet50, 3D CNN, and our transformer-based model were 0.853 [0.821-0.883], 0.838 [0.800-0.874], 0.840 [0.803-0.876], 0.841 [0.804-0.877], and 0.910 [0.884-0.934], respectively. For the detection of medial collateral ligament injury, the AUC values for the same models were 0.810 [0.744-0.866], 0.772 [0.695-0.842], 0.796 [0.723-0.859], 0.784 [0.716-0.843], and 0.892 [0.835-0.936], respectively. The DeLong test demonstrated that our transformer-based model significantly outperformed other models in both lateral (all $P < 0.001$) and medial (all $P < 0.001$) collateral ligament injuries detection.

219 Table 2 presents the performance metrics for the models. Our transformer-based model
 220 exhibited the highest sensitivity (0.836 [0.780-0.887]) for lateral injury detection and the highest
 221 specificity (0.818 [0.789-0.847]) for medial injury detection. It also obtained the highest accuracy
 222 for both lateral (0.854 [0.828-0.882]) and medial (0.819 [0.788-0.846]) sides. Notably, it achieved
 223 a balanced trade-off between sensitivity and specificity as compared with other models.

224 The calibration curves showed that our model was well calibrated with a Brier score of 0.113
 225 for the detection of lateral collateral ligament injury and 0.051 for the detection of medial collateral
 226 ligament injury (see Figure S2).

227 **Performance of AnkleNet for CAI Diagnosis**

228 Due to the superior discriminative ability in detecting ankle ligament injuries, our transformer-
 229 based model was employed for the subsequent CAI diagnosis, where individuals were classified
 230 into four groups: normal ankle, LCAI, MCAI, and RCAI. The multi-label probability outputs were
 231 transformed into a multi-class representation. For CAI diagnosis, our model achieved a balanced
 232 classification accuracy of 0.805 and a macro-average AUC of 0.870. The accuracy for normal,
 233 LCAI, MCAI, and RCAI were 0.906, 0.871, 0.619, and 0.824, respectively. The relatively low
 234 accuracy of MCAI could be attributed to the imbalanced dataset, which consisted of less than 5%
 235 of MCAI patients in both the development set and external validation set. However, it correctly
 236 diagnosed 56 out of 68 patients with RCAI, which is the most significant class. The confusion
 237 matrix depicting the CAI diagnosis is presented in Figure S3.

238 To interpret the model's predictions, we utilized Grad-CAM to generate an overlaid heatmap,
 239 visualizing the important features detected by the model. Figure 3 shows a case where our model
 240 diagnosed the patient with RCAI. We observed that the model's prediction for lateral collateral
 241 ligament injury focused on the axial plane, specifically highlighting the tear of the ATFL. In
 242 contrast, the prediction for medial collateral ligament injury was associated with the coronal plane,
 243 emphasizing the tear of the deltoid ligament. Arthroscopic examination confirmed the RCAI
 244 diagnosis in this patient.

245 **Reader study**

246 We found that the sensitivity among the clinicians varied significantly: ranging from 0.574 to 0.882.
 247 Although both clinical experts achieved accuracy exceeding 0.8, the average accuracy among the
 248 junior clinicians was around 0.7, which implies that unexperienced practitioners are prone to
 249 neglect the concurrent medial injuries of CAI patient.

Enhanced by model assistance, all clinicians reported improvements (see Figure 4). The average performance metrics for clinicians without model assistance were 0.763 ± 0.078 for accuracy, 0.712 ± 0.129 for sensitivity, and 0.815 ± 0.07 for specificity. However, when provided with model assistance, the average accuracy, sensitivity, and specificity improved to 0.846 ± 0.046 , 0.841 ± 0.057 , and 0.850 ± 0.048 , respectively. Utilizing the one-tailed t-test, we determined that the improvements were statistically significant for accuracy ($p = 0.042$) and sensitivity ($p = 0.045$), though not for specificity ($p = 0.192$). Junior clinicians demonstrated more substantial improvement with model assistance. The improvement in accuracy and sensitivity for the unexperienced clinicians ranged from 12.1% to 17.9% and from 16.7% to 35.9%, respectively. On the other hand, for experts, the improvement was 5.1% to 6.3% in accuracy and 5.0% to 9.3% in sensitivity. It is worth mentioning that the support of our model aided clinicians in evaluating MRI more consistently. While Fleiss' κ was 0.402 for the part 1 of reader study, the agreement between clinicians was improved in part 2, with a κ of 0.685. Table S1 and S2 show the detailed information on individual performance.

264

265 DISCUSSION

Our results revealed that our transformer-based model, exhibited superior performance in the multi-label classification tasks on the ankle MRI, with AUCs for lateral and medial collateral ligament injuries detection of 0.910 and 0.892, respectively. Furthermore, we observed that providing the predictions of the model to clinicians as a decision support tool led to enhanced identification of patients with RCAI, with an average 10.7% increase in accuracy and an average 18.1% increase in sensitivity.

Although significant progress has been made in leveraging DL for automated diagnosis in the field of sports medicine, most of these advancements focused on knee injuries. For instance, Bien et al.²³ developed MRNet, a model designed to identify general abnormalities, as well as ACL and meniscus lesions on sagittal, coronal, and axial planes. They combined predictions from these three planes using logistic regression and achieved an AUC of 0.965 and an accuracy of 0.867 for ACL injury detection. Sharma et al.²⁴ constructed the DL model based on ResNet50 for ACL injury detection, obtaining an AUC of 0.92 on a single MRI plane. In another study, Zhang et al.²⁹ developed DL models for detecting ACL injury using arthroscopy as the gold standard. In their results, the 3D CNN model outperformed than 2D models, achieving an AUC of 0.96. Recently,

281 Wang and colleagues trained a DL model using a dataset consisting of over 20,000 MRIs from
282 multiple centers to diagnose ACL injuries, attaining an impressive AUC of 0.980³⁰. They further
283 reported that with the support of their model, clinicians' diagnostic accuracy improved to 0.96.
284 However, few studies attempted to construct DL models for evaluating ankle injuries. Ni et al.³¹
285 focused on classifying calcaneofibular ligament (CLF) injuries on MRI. They proposed a two-
286 stage approach involving classification followed by segmentation. The authors reported
287 performance with AUCs of 0.94, 0.97, and 0.96 for normal, partial tear, and complete tear,
288 respectively, by concatenating features from both axial and coronal planes. However, their results
289 were presented on a slice-wise basis rather than for the entire patient, which could potentially result
290 in an overestimation of performance. Furthermore, their approach was custom-designed for CLF
291 injuries, necessitating the segmentation of the target ligament as a preliminary step. Therefore,
292 transferring this approach to other ankle-related tasks may pose challenges and require significant
293 modifications to account for different anatomical structures or injury types.

294 In contrast to prior studies that predominantly fine-tuning general CNN-based models, which
295 are characterized by their convolutional layers and local receptive fields, our proposed architecture
296 is tailored specifically for ankle-related tasks and employs a transformer-based design. CNN-based
297 models excel in extracting hierarchical features from images by applying filters that capture local
298 patterns such as edges and textures, which is effective for a wide range of visual tasks³². However,
299 their reliance on local processing can be limiting when dealing with some medical images where
300 global context and the relationship between distant anatomical structures are crucial³³. This is
301 particularly true in ankle MRI, where ligaments like the ATFL and CFL are not confined to a
302 compact area but are dispersed across MRI planes, necessitating a broader perspective for analysis.
303 Our transformer-based model, leveraging the self-attention mechanism, is adept at capturing these
304 global dependencies among image patches, allowing for a comprehensive understanding of long-
305 range interactions and detailed features within the images^{22,33,34}. Moreover, a distinctive feature of
306 our model is the incorporation of a cross-attention mechanism^{35,36}. This mechanism enables it to
307 simultaneously focus on relevant information from multiple planes input. Rather than merely
308 concatenating outputs from distinct planes or employing logistic regression for combining
309 predictions, the cross-attention modules could enhance the model's capacity to understand
310 relationships between different planes, acquire critical features, and filter redundant information.
311 As expected, it outperformed other CNN-based models and balanced the sensitivity and specificity

312 at the same time. For the subsequent CAI diagnosis, our model achieved a macro-average AUC of
313 0.870 and successfully identified 56 of 68 RCAI patients. Nevertheless, the relatively diminished
314 accuracy in diagnosing patients with MCAI cannot be ignored. This outcome is not surprising,
315 considering the highly imbalanced dataset, where MCAI patients constitute only 3.5% and 3.3%
316 of the development and external validation sets, respectively. Even with the implementation of
317 weighted loss during the training stage, the model's ability to acquire distinctive features
318 pertaining to these patients remains constrained, resulting in a predisposition toward other
319 categories. Isolated deltoid ligament injuries are rare, occurring in merely 3 out of 281 recorded
320 ankle injuries^{37,38}. Jeong et al.³⁹ noted that such isolated injuries constituted 8.3% of all deltoid
321 ligament injuries, exclusively affecting the superficial component. To enhance the diagnostic
322 accuracy of models for MCAI, future research needs to expand the dataset with additional
323 instances of isolated deltoid ligament injuries. Moreover, employing advanced data augmentation
324 techniques, like Mixup, to create synthetic MCAI cases might further enhance the model's
325 understanding of MCAI, thereby improving its diagnostic performance⁴⁰.

326 It is important to highlight that DL-based diagnostic models are not designed to replace medical
327 professionals; instead, their role is to augment healthcare practitioners by enhancing the efficiency
328 and accuracy of diagnoses⁴¹⁻⁴³. As our expected, our transformer-based model displayed promise
329 in facilitating the diagnosis of patients with CAI. In the reader study, clinical experts demonstrated
330 a high level of accuracy and sensitivity, whereas it was evident that unexperienced practitioners
331 encountered challenges. Three junior clinicians displayed an average sensitivity of approximately
332 0.6 when only the raw MRI data was accessible. Upon incorporating our model, the performance
333 of all clinicians exhibited improvement, particularly among the junior clinicians. The junior
334 clinicians displayed a substantial increase, ranging from 12.1% to 17.9% for accuracy and 16.7%
335 to 35.9% for sensitivity. Furthermore, this collaboration led to an enhancement in diagnostic
336 consistency, as indicated by an increased κ value of 0.283. Collectively, our transformer-based
337 model could effectively detect both lateral and medial collateral ligament injuries and diagnose
338 CAI patients. Although the accuracy of MCAI is not as good as that of other subtypes of CAI, it
339 has the potential to serve as a decision support tool to assist in RCAI identification, enhance the
340 performance of less experienced clinicians in diagnosing CAI, and improve diagnostic consistency.

341 As the comprehension of CAI continues to deepen, foot and ankle experts are no longer confined
342 to focus on addressing lateral injuries of the ankle. In addition to deltoid ligament, syndesmosis

343 plays a crucial role in maintaining ankle joint stability as well^{7,44}. Combined injury involving the
344 syndesmosis along with the lateral and medial collateral ligament, has been defined as
345 multidirectional CAI (MD-CAI) recently⁵. While the definition of MD-CAI has yet to gain
346 widespread adoption, an ideal ankle-specific model should have the capability to simultaneously
347 evaluate the syndesmosis, eliminating the need to train separate models for each task as previous
348 studies did. However, as syndesmosis is not routinely assessed during arthroscopy in all centers, it
349 was not utilized as an independent label during the development phase of our model. Nevertheless,
350 our transformer-based model adopts a multi-label paradigm^{45–47} that takes into account such
351 diversity of clinical scenarios. Adjusting the final layer of the model to accommodate the required
352 number of labels can facilitate the creation of a more comprehensive model. Incorporating
353 supplementary labels, such as syndesmosis injury and osteochondral lesions of the talus, to extend
354 the capabilities of the current version of the model is in our future work. Such enhancement would
355 be valuable in assessing a wider range of ankle injuries, thereby benefitting the prevention of AOA.

356 **Limitation**

357 The current study had several limitations to be mentioned. Firstly, the dataset size is relatively
358 limited due to the use of arthroscopy as the gold standard. Secondly, as we mentioned above, the
359 proportion of MCAI patients was small, contributing to the imbalance of the dataset and resulting
360 in the low accuracy of MCAI diagnosis. Indeed, the current model may not effectively assist in the
361 diagnosis of MCAI. It was prone to overlooking isolated medial collateral ligament injuries or
362 misdiagnosing some MCAI patients as LCAI patients.

363

364 **CONCLUSIONS**

365 Our transformer-based model was able to detect lateral and medial collateral ligament injuries
366 based on MRI and outperformed CNN-based models, demonstrating a promising performance in
367 diagnosing CAI, especially RCAI patients.

368

369 **References**

- 370 1. Gribble PA, Bleakley CM, Caulfield BM, et al. Evidence review for the 2016 International
371 Ankle Consortium consensus statement on the prevalence, impact and long-term consequences
372 of lateral ankle sprains. *Br J Sports Med.* 2016;50(24):1496-1505.
- 373 2. Herzog MM, Kerr ZY, Marshall SW, Wikstrom EA. Epidemiology of Ankle Sprains and
374 Chronic Ankle Instability. *J Athl Train.* 2019;54(6):603-610. doi:10.4085/1062-6050-447-17
- 375 3. Valderrabano V, Hintermann B, Horisberger M, Fung TS. Ligamentous Posttraumatic Ankle
376 Osteoarthritis. *Am J Sports Med.* 2006;34(4):612-620. doi:10.1177/0363546505281813
- 377 4. Barg A, Pagenstert GI, Hügle T, et al. Ankle Osteoarthritis. *Foot Ankle Clin.* 2013;18(3):411-
378 426. doi:10.1016/j.fcl.2013.06.001
- 379 5. De Cesar Netto C, Valderrabano V, Mansur NSB. Multidirectional Chronic Ankle Instability.
380 *Foot Ankle Clin.* 2023;28(2):405-426. doi:10.1016/j.fcl.2023.01.012
- 381 6. Mansur NSB, Lemos AVKC, Baumfeld DS, et al. Medial and Lateral Combined Ligament
382 Arthroscopic Repair for Multidirectional Ankle Instability. *Foot Ankle Orthop.*
383 2021;6(1):247301142098615. doi:10.1177/2473011420986150
- 384 7. Kim JS, Young KW, Cho HK, Lim SM, Park YU, Lee KT. Concomitant Syndesmotic
385 Instability and Medial Ankle Instability Are Risk Factors for Unsatisfactory Outcomes in
386 Patients With Chronic Ankle Instability. *Arthrosc J Arthrosc Relat Surg.* 2015;31(8):1548-
387 1556. doi:10.1016/j.artro.2015.02.021
- 388 8. Schäfer D, Hintermann B. Arthroscopic assessment of the chronic unstable ankle joint. *Knee
389 Surg Sports Traumatol Arthrosc.* 1996;4(1):48-52. doi:10.1007/BF01565998
- 390 9. Hintermann B, Boss A, Schäfer D. Arthroscopic Findings in Patients with Chronic Ankle
391 Instability. *Am J Sports Med.* 2002;30(3):402-409. doi:10.1177/03635465020300031601
- 392 10. Valderrabano V, Wiewiorski M, Frigg A, Hintermann B, Leumann A. Chronische Instabilität
393 des oberen Sprunggelenks. *Unfallchirurg.* 2007;110(8):691-700. doi:10.1007/s00113-007-
394 1310-y
- 395 11. Vega J, Allmendinger J, Malagelada F, Guelfi M, Dalmau-Pastor M. Combined arthroscopic
396 all-inside repair of lateral and medial ankle ligaments is an effective treatment for rotational
397 ankle instability. *Knee Surg Sports Traumatol Arthrosc.* 2020;28(1):132-140.
398 doi:10.1007/s00167-017-4736-y
- 399 12. O'Neill PJ, Van Aman SE, Guyton GP. Is MRI Adequate to Detect Lesions in Patients with
400 Ankle Instability? *Clin Orthop.* 2010;468(4):1115-1119. doi:10.1007/s11999-009-1131-0
- 401 13. Cha SD, Kim HS, Chung ST, et al. Intra-articular Lesions in Chronic Lateral Ankle Instability:
402 Comparison of Arthroscopy with Magnetic Resonance Imaging Findings. *Clin Orthop Surg.*
403 2012;4(4):293-299. doi:10.4055/cios.2012.4.4.293

- 404 14. Yasuda T, Shima H, Mori K, Tsujinaka S, Neo M. Simultaneous Reconstruction of the Medial
405 and Lateral Collateral Ligaments for Chronic Combined Ligament Injuries of the Ankle. *Am J
406 Sports Med.* 2017;45(9):2052-2060. doi:10.1177/0363546517700859
- 407 15. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from
408 non-small cell lung cancer histopathology images using deep learning. *Nat Med.*
409 2018;24(10):1559-1567. doi:10.1038/s41591-018-0177-5
- 410 16. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-
411 dimensional deep learning on low-dose chest computed tomography. *Nat Med.*
412 2019;25(6):954-961. doi:10.1038/s41591-019-0447-x
- 413 17. Chen X, Wang X, Zhang K, et al. Recent advances and clinical applications of deep learning
414 in medical image analysis. *Med Image Anal.* 2022;79:102444.
415 doi:10.1016/j.media.2022.102444
- 416 18. Hill BG, Krogue JD, Jevsevar DS, Schilling PL. Deep Learning and Imaging for the
417 Orthopaedic Surgeon: How Machines “Read” Radiographs. *JBJS.* 2022;104(18):1675.
418 doi:10.2106/JBJS.21.01387
- 419 19. Yin R, Chen H, Tao T, et al. Expanding from unilateral to bilateral: A robust deep learning-
420 based approach for predicting radiographic osteoarthritis progression. *Osteoarthritis Cartilage.*
421 2024;32(3):338-347. doi:10.1016/j.joca.2023.11.022
- 422 20. Ramkumar PN, Luu BC, Haeberle HS, Karnuta JM, Nwachukwu BU, Williams RJ. Sports
423 Medicine and Artificial Intelligence: A Primer. *Am J Sports Med.* 2022;50(4):1166-1174.
424 doi:10.1177/03635465211008648
- 425 21. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: Improved N3 Bias Correction. *IEEE Trans
426 Med Imaging.* 2010;29(6):1310-1320. doi:10.1109/TMI.2010.2046908
- 427 22. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers
428 for Image Recognition at Scale. arXiv preprint arXiv:2010.11929, 2020.
429 <http://arxiv.org/abs/2010.11929>
- 430 23. Bien N, Rajpurkar P, Ball RL, et al. Deep-learning-assisted diagnosis for knee magnetic
431 resonance imaging: Development and retrospective validation of MRNet. Saria S, ed. *PLOS
432 Med.* 2018;15(11):e1002699. doi:10.1371/journal.pmed.1002699
- 433 24. Sharma S, Umer M, Bhagat A, Bala J, Rattan P, Rahmani AW. A ResNet50-Based Approach
434 to Detect Multiple Types of Knee Tears Using MRIs. *Math Probl Eng.* 2022;2022:e5248338.
435 doi:10.1155/2022/5248338
- 436 25. Namiri NK, Flament I, Astuto B, et al. Deep Learning for Hierarchical Severity Staging of
437 Anterior Cruciate Ligament Injuries from MRI. *Radiol Artif Intell.* 2020;2(4):e190207.
438 doi:10.1148/ryai.2020190207

- 439 26. Liu F, Guan B, Zhou Z, et al. Fully Automated Diagnosis of Anterior Cruciate Ligament Tears
 440 on Knee MR Images by Using Deep Learning. *Radiol Artif Intell.* 2019;1(3):180091.
 441 doi:10.1148/ryai.2019180091
- 442 27. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual
 443 Explanations from Deep Networks via Gradient-based Localization. *Int J Comput Vis.*
 444 2020;128(2):336-359. doi:10.1007/s11263-019-01228-7
- 445 28. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol.* 2010;63(8):938-
 446 939.
- 447 29. Zhang L, Li M, Zhou Y, Lu G, Zhou Q. Deep Learning Approach for Anterior Cruciate
 448 Ligament Lesion Detection: Evaluation of Diagnostic Performance Using Arthroscopy as the
 449 Reference Standard. *J Magn Reson Imaging.* 2020;52(6):1745-1752. doi:10.1002/jmri.27266
- 450 30. Wang D yu, Liu S gui, Ding J, et al. A Deep Learning Model Enhances Clinicians' Diagnostic
 451 Accuracy to More Than 96% for Anterior Cruciate Ligament Ruptures on Magnetic Resonance
 452 Imaging. *Arthrosc J Arthrosc Relat Surg.* 2024;40(4):1197-1205.
 453 doi:10.1016/j.arthro.2023.08.010
- 454 31. Ni M, Zhao Y, Wen X, et al. Deep learning-assisted classification of calcaneofibular ligament
 455 injuries in the ankle joint. *Quant Imaging Med Surg.* 2023;13(1):80-93. doi:10.21037/qims-
 456 22-470
- 457 32. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436-444.
 458 doi:10.1038/nature14539
- 459 33. Shamshad F, Khan S, Zamir SW, et al. Transformers in medical imaging: A survey. *Med Image
 460 Anal.* 2023;88:102802. doi:10.1016/j.media.2023.102802
- 461 34. Yuan F, Zhang Z, Fang Z. An effective CNN and Transformer complementary network for
 462 medical image segmentation. *Pattern Recognit.* 2023;136:109228.
- 463 35. Chen CFR, Fan Q, Panda R. CrossViT: Cross-Attention Multi-Scale Vision Transformer for
 464 Image Classification. In: *2021 IEEE/CVF International Conference on Computer Vision
 465 (ICCV).* IEEE; 2021:347-356. doi:10.1109/ICCV48922.2021.00041
- 466 36. Jang J, Hwang D. M3T: three-dimensional Medical image classifier using Multi-plane and
 467 Multi-slice Transformer. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern
 468 Recognition (CVPR).* IEEE; 2022:20686-20697. doi:10.1109/CVPR52688.2022.02006
- 469 37. Crim JR, Beals TC, Nickisch F, Schannen A, Saltzman CL. Deltoid Ligament Abnormalities
 470 in Chronic Lateral Ankle Instability. *Foot Ankle Int.* 2011;32(9):873-878.
 471 doi:10.3113/FAI.2011.0873
- 472 38. McCollum GA, van den Bekerom MPJ, Kerkhoffs GMMJ, Calder JDF, van Dijk CN.
 473 Syndesmosis and deltoid ligament injuries in the athlete. *Knee Surg Sports Traumatol Arthrosc.*
 474 2013;21(6):1328-1337. doi:10.1007/s00167-012-2205-1

- 475 39. Jeong MS, Choi YS, Kim YJ, Kim JS, Young KW, Jung YY. Deltoid ligament in acute ankle
476 injury: MR imaging analysis. *Skeletal Radiol.* 2014;43(5):655-663. doi:10.1007/s00256-014-
477 1842-5
- 478 40. Panfilov E, Tiulpin A, Klein S, Nieminen MT, Saarakkala S. Improving Robustness of Deep
479 Learning Based Knee MRI Segmentation: Mixup and Adversarial Domain Adaptation.
480 Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.
481 <http://arxiv.org/abs/1908.04126>
- 482 41. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med.*
483 2022;28(1):31-38. doi:10.1038/s41591-021-01614-0
- 484 42. Yoon SH, Park S, Jang S, et al. Use of artificial intelligence in triaging of chest radiographs to
485 reduce radiologists' workload. *Eur Radiol.* 2023;34(2):1094-1103. doi:10.1007/s00330-023-
486 10124-1
- 487 43. Sanchez M, Alford K, Krishna V, et al. AI-clinician collaboration via disagreement prediction:
488 A decision pipeline and retrospective analysis of real-world radiologist-AI interactions. *Cell
489 Rep Med.* 2023;4(10):101207. doi:10.1016/j.xcrm.2023.101207
- 490 44. Chun KY, Choi YS, Lee SH, et al. Deltoid Ligament and Tibiofibular Syndesmosis Injury in
491 Chronic Lateral Ankle Instability: Magnetic Resonance Imaging Evaluation at 3T and
492 Comparison with Arthroscopy. *Korean J Radiol.* 2015;16(5):1096.
493 doi:10.3348/kjr.2015.16.5.1096
- 494 45. Bi L, Feng DD, Fulham M, Kim J. Multi-Label classification of multi-modality skin lesion via
495 hyper-connected convolutional neural network. *Pattern Recognit.* 2020;107:107502.
496 doi:10.1016/j.patcog.2020.107502
- 497 46. Chen H, Miao S, Xu D, Hager GD, Harrison AP. Deep hierarchical multi-label classification
498 applied to chest X-ray abnormality taxonomies. *Med Image Anal.* 2020;66:101811.
499 doi:10.1016/j.media.2020.101811
- 500 47. Liu W, Wang H, Shen X, Tsang IW. The Emerging Trends of Multi-Label Learning. *IEEE
501 Trans Pattern Anal Mach Intell.* 2022;44(11):7955-7974. doi:10.1109/TPAMI.2021.3119334
- 502

503 **Figure legends**

504

505 **Figure 1:** The overview of deep learning workflow. **A.** The raw MRI undergo several
 506 preprocessing steps before being fed into the to the deep learning model. The model will
 507 simultaneously output the injury probabilities of both lateral and medial collateral ligament. Finally,
 508 the CAI diagnosis will be generated based on the resulting probabilities from the previous step and
 509 then be provided to the clinicians in an assistive setting offset. **B.** Diagram shows the model
 510 structure.

511 **Figure 2:** The ROC curve of deep learning models. **A.** The ROC curve for detecting the injury
 512 of lateral collateral ligament on the external validation set, which consisted of 48.9% patients with
 513 lateral collateral ligament injury. **B.** The ROC curve for detecting the injury of medial collateral
 514 ligament on the external validation set, which consisted of 13.8% patients with medial collateral
 515 ligament injury.

516 **Figure 3:** Accurate diagnosis of RCAI of the left ankle in a 32-year-old man by AnkleNet. **A.**
 517 Discontinuity and irregular signals (arrow) were found in ATFL of the lateral ankle ligament on
 518 the axial PD-weighted MRI (up) and in the medial deltoid ligament on the coronal PD-weighted
 519 MRI (down), respectively. **B.** The Grad-CAM diagrams demonstrate the AnkleNet focuses on the
 520 regions of injured ATFL on the axial plane (up) and injured deltoid ligaments on the coronal plane
 521 (down). A color close to red indicates that the model pays more attention to this area. **C.** The
 522 diagnosis of RCAI was finally confirmed by arthroscopic examination performed with the patient
 523 in the supine position. The ATFL was found to be ruptured at the lateral malleolus with scar tissue
 524 formation (up) and the superficial anterior deltoid ligament was detached from the medial
 525 malleolus (down) while viewing from the antero-central portal just medial to the anterior tibial
 526 tendon. ATFL, anterior talofibular ligament; LM, lateral malleolus; DEL, deltoid ligament; MM,
 527 medial malleolus, TA, talus.

528 **Figure 4:** Performance of clinicians and AnkleNet in distinguishing the RCAI patients. **A.** The
 529 ROC curve of AnkleNet for the CAI diagnosis, the macro-average AUC is reported. The average
 530 performance of all clinicians with and without the support of the model is plotted as filled and
 531 open star, respectively. Both sensitivity and specificity are increased with model assistance (black
 532 arrow). **B.** A magnified region of the dotted rectangular area of the ROC (as outlined in A.), with

533 individual clinicians represented by open shapes (without model assistance) and filled shapes (with
534 model assistance).

535

536 **Tables**

537

Table 1. The Characteristics of Individuals in the Development and External Validation Dataset

	Development set	External Validation set	<i>p</i> -value
	<i>N</i> =1621	<i>N</i> =646	
Age	34.3 ± 13.6	34.9 ± 12.6	0.371
Gender:			1.000
Male	795 (49.0%)	317 (49.1%)	
Female	826 (51.0%)	329 (50.9%)	
Side:			0.165
Right	852 (52.6%)	318 (49.2%)	
Left	769 (47.4%)	328 (50.8%)	
Lateral collateral ligament injury:			0.764
No	815 (50.3%)	330 (51.1%)	
Yes	806 (49.7%)	316 (48.9%)	
Medial collateral ligament injury:			0.911
No	1393 (85.9%)	557 (86.2%)	
Yes	228 (14.1%)	89 (13.8%)	
CAI:			0.974
Normal	759 (46.8%)	309 (47.8%)	
LCAI	634 (39.1%)	248 (38.4%)	
MCAI	56 (3.5%)	21 (3.3%)	
RCAI	172 (10.6%)	68 (10.5%)	

NOTE: CAI, chronic ankle instability; LCAI, lateral chronic ankle instability; MCAI, medial chronic ankle instability; RCAI, rotational chronic ankle instability.

538

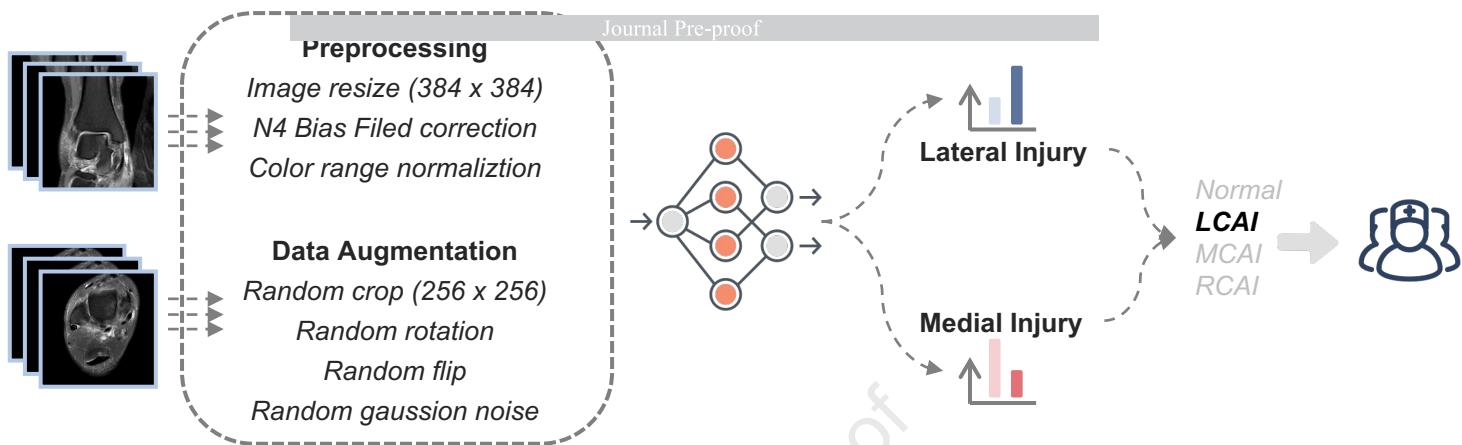
539

Table 2. Performance Comparison among Deep Learning Models in the Ankle Ligament Injury Detection in the External Validation Set

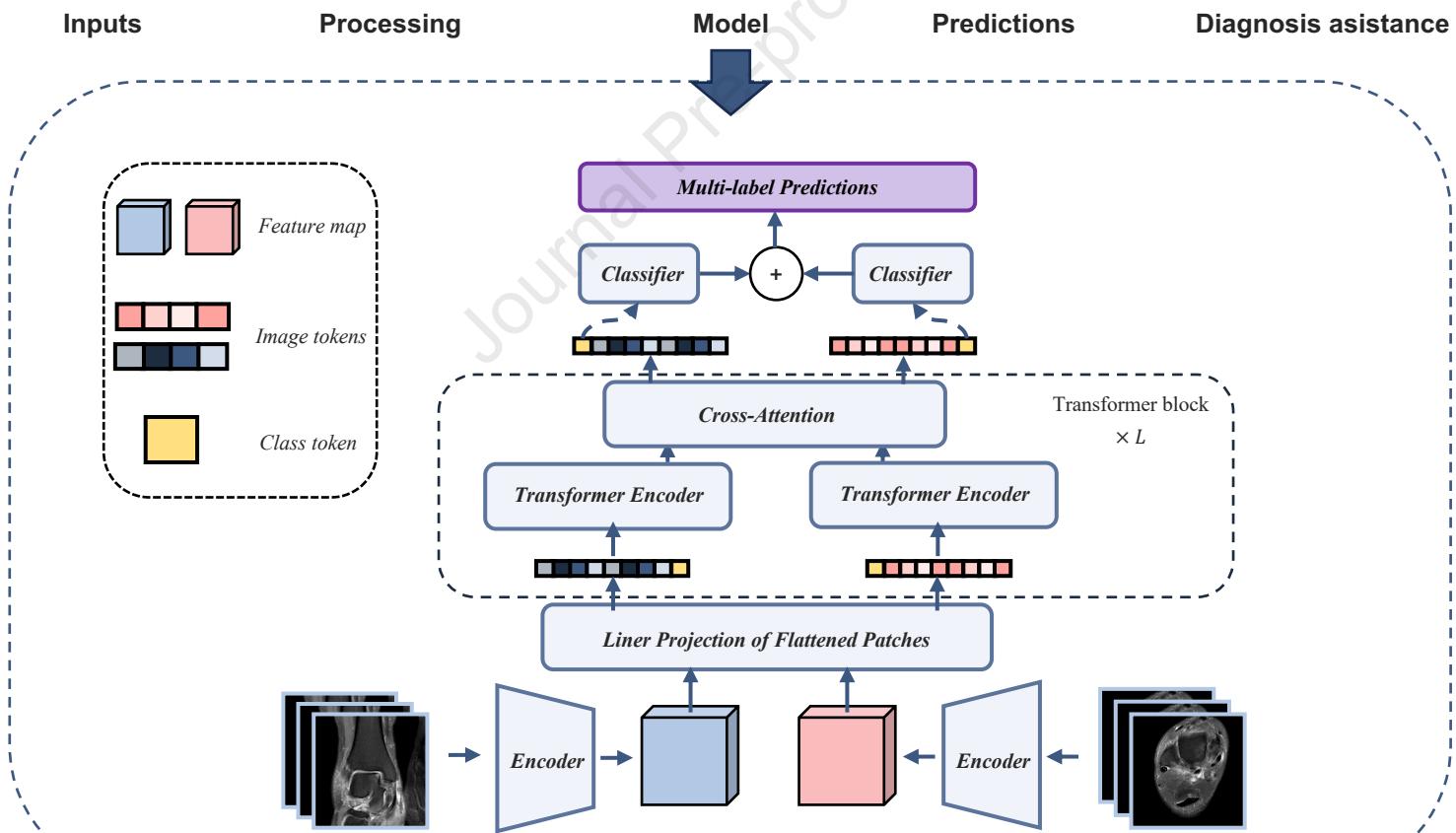
Model	Lateral ligament complex injury			Deltoid ligament injury		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
MRNet	0.812 [0.778-0.838]	0.797 [0.74-0.853]	0.795 [0.759-0.831]	0.793 [0.760-0.821]	0.849 [0.755-0.943]	0.676 [0.638-0.713]
ResNet34	0.796 [0.763-0.825]	0.802 [0.746-0.859]	0.801 [0.767-0.835]	0.690 [0.656-0.726]	0.792 [0.679-0.887]	0.745 [0.711-0.777]
ResNet50	0.801 [0.769-0.828]	0.785 [0.723-0.842]	0.821 [0.789-0.855]	0.794 [0.760-0.822]	0.774 [0.660-0.887]	0.796 [0.762-0.826]
CNN3D	0.838 [0.807-0.863]	0.695 [0.627-0.763]	0.889 [0.861-0.917]	0.749 [0.713-0.778]	0.774 [0.660-0.887]	0.796 [0.764-0.825]
AnkleNet	0.854 [0.828-0.882]	0.836 [0.780-0.887]	0.861 [0.829-0.891]	0.819 [0.788-0.846]	0.830 [0.717-0.925]	0.818 [0.789-0.847]

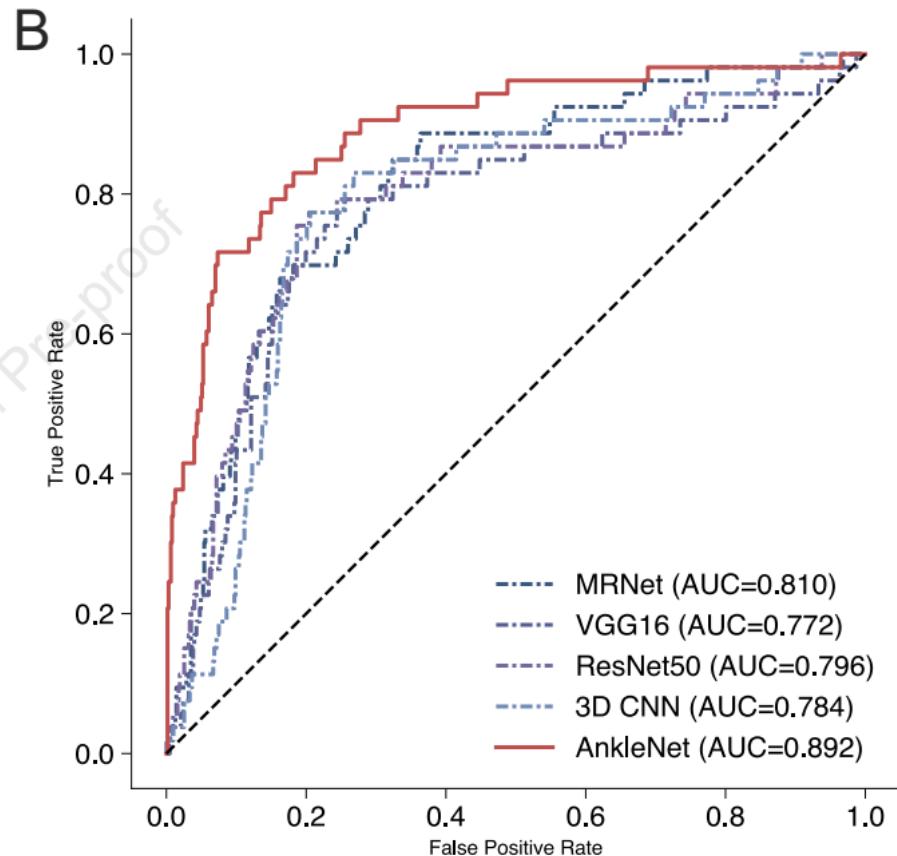
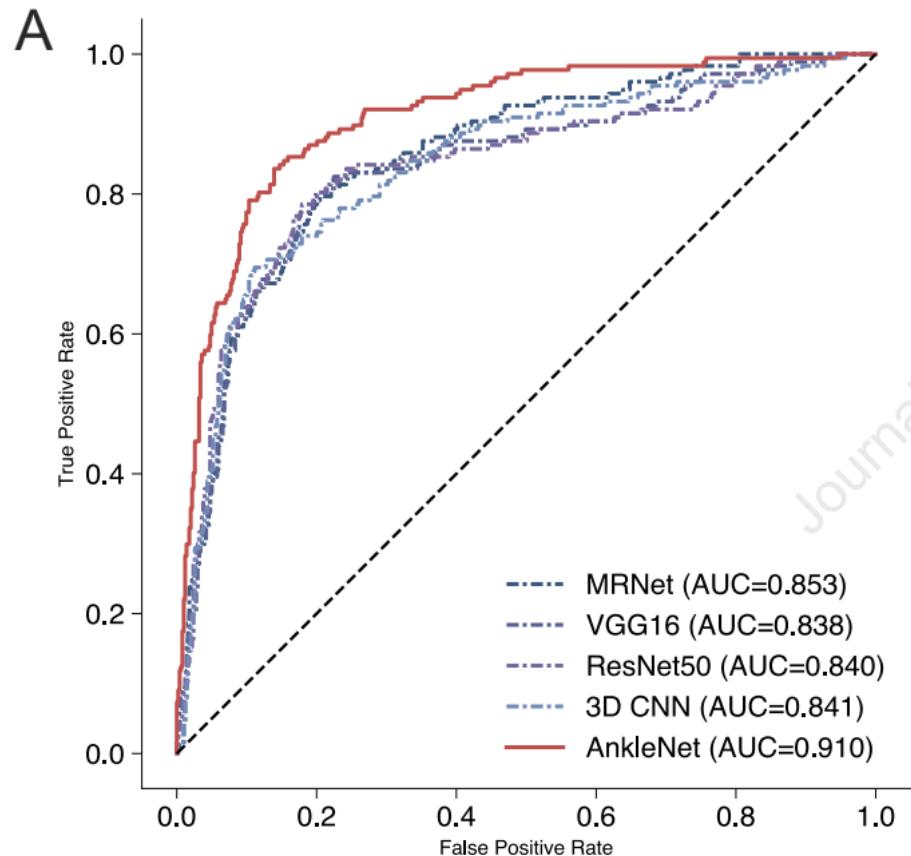
NOTE: Data are presented as value [95% confidential interval].

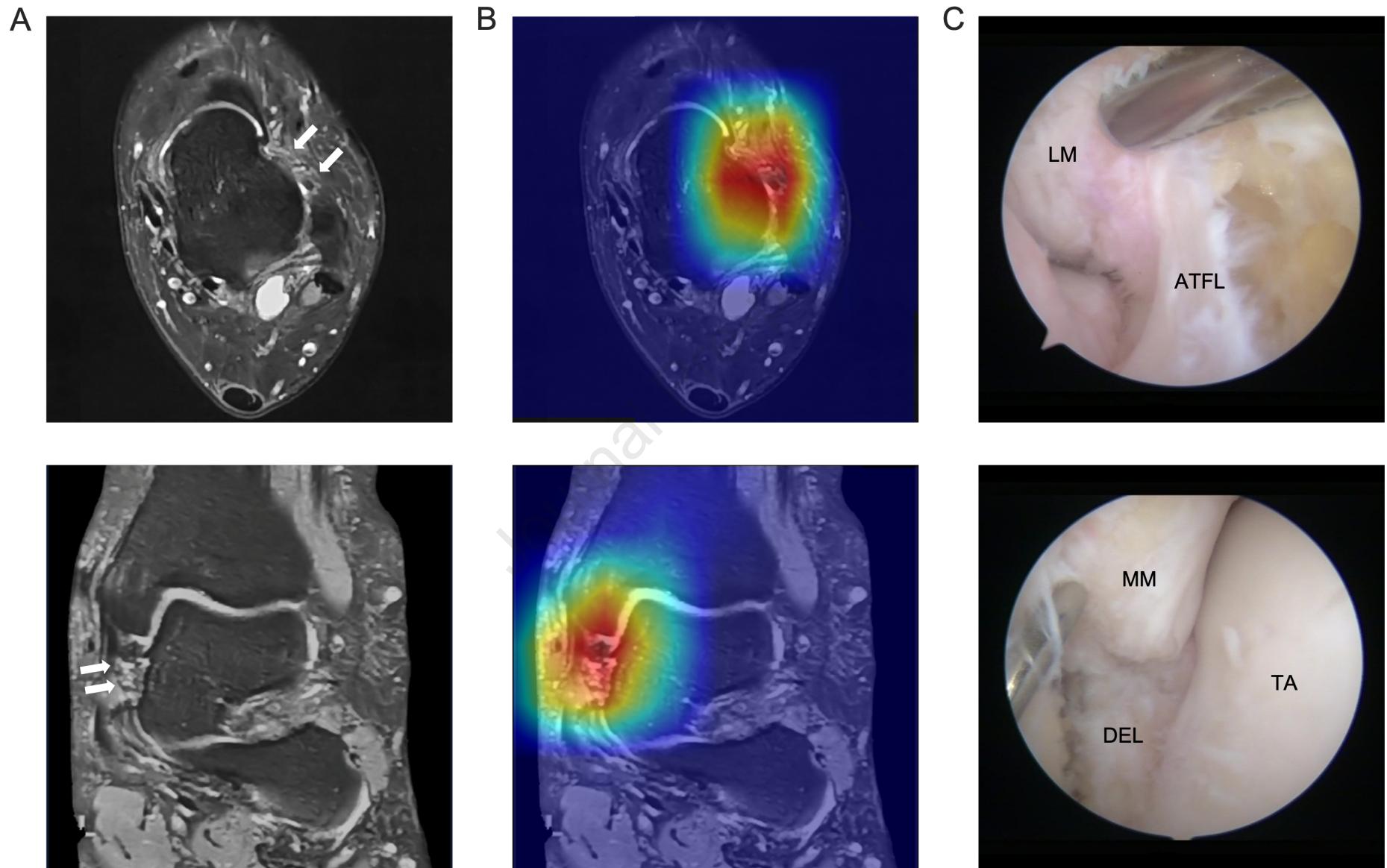
A

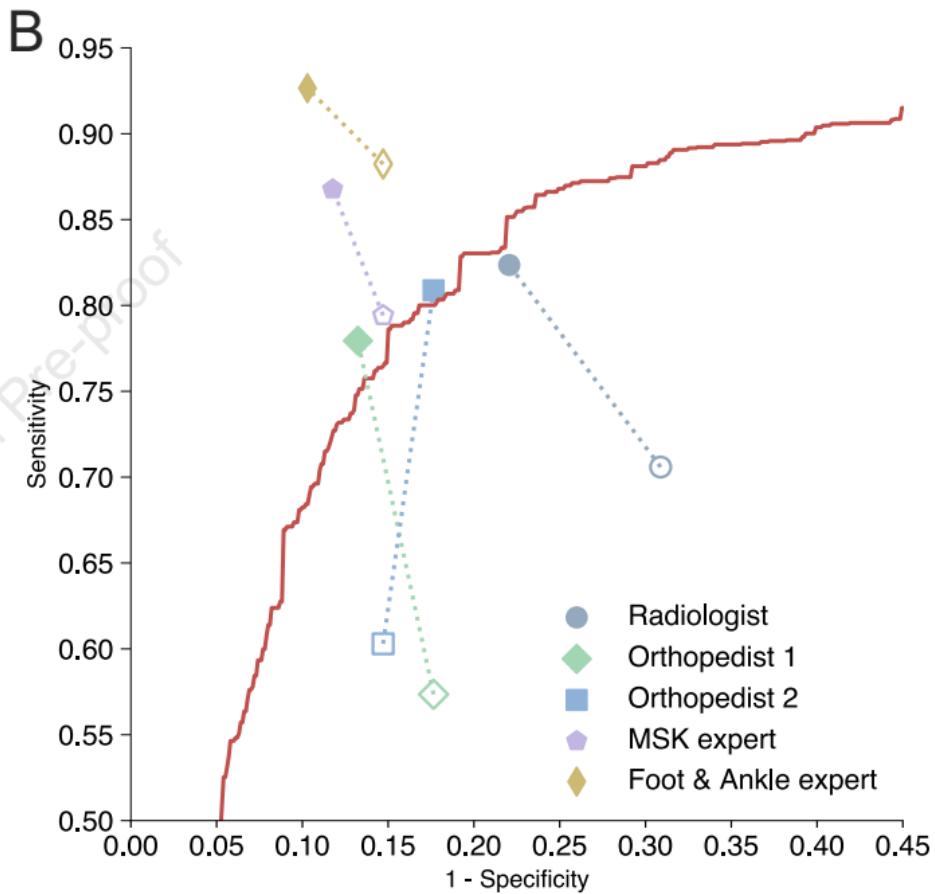
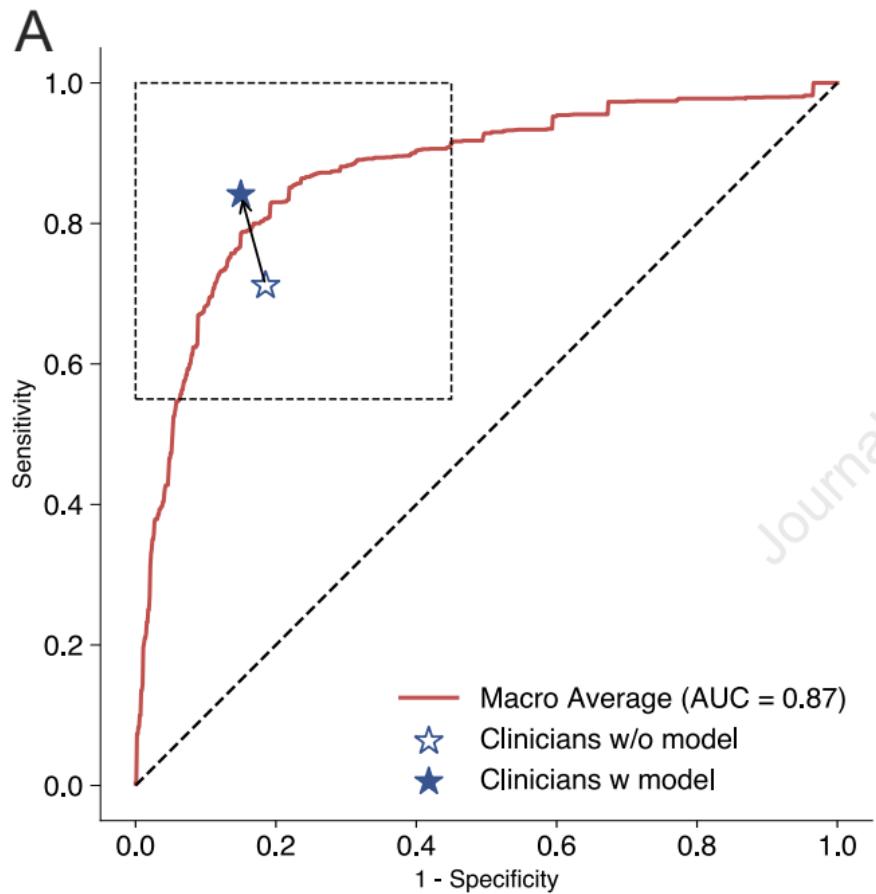


B









Appendix E1

Model Architecture

The AnkleNet, which follows a two-view approach, encompasses a sophisticated learning schema that can be dissected into multiple distinct learning sections (Figure 2). The efficacy of the local-global learning schema in acquiring meaningful features has been extensively demonstrated in prior research^{1,2}. Drawing inspiration from this, our initial endeavor involves employing a Convolution model to extract features at a local level, resulting in condensed local feature maps. Within these maps, each individual pixel element establishes connections solely with its adjacent region, facilitating localized information processing.

Subsequently, the aforementioned local feature maps are partitioned into patches. These patch-based representations, together with their embedded spatial information, are then inputted into a Vision Transformer^{3,4}. The Transformer architecture endeavors to acquire a comprehensive understanding of the relevant feature alignment across patches in a global manner within each MRI plane. Through this process, every element within the patches engages in cooperative learning within other patches, thereby constructing a robust global relation mapping. Furthermore, the alignment between the two views, namely the coronal plane and axial plane, is achieved by means of a cross-attention mechanism^{5–7}. This mechanism establishes an information tunnel that facilitates the exchange of relevant information between the features of the coronal and axial plane, consequently ensuring view-consistency regarding the ankle's characteristics across different views.

Following the alignment of global features and cross-view features, these refined representations are then forwarded to a classifier, which leverages the acquired knowledge to predict the lateral and medial collateral ligament injury.

Training Details

AnkleNet employs the preceding three conv layers (conv1, layer1, layer2) within pretrained ResNet34 as the encoder. The model was configured with four transformer blocks ($L = 4$, see Figure 2), with both the number of heads in multi-header attention (transformer encoder) and cross-attention was set to 8. The dimension of the transformer encoder was set at 2048, and a dropout

rate of 0.1 was applied. The batch size was set as 8, the maximum epoch was 50, the learning rate was 0.0003, the optimizer was AdamW, and the weight decay of 0.0005 was applied.

All models were implemented using PyTorch (verison1.8) and trained on an NVIDIA RTX 4090 and an AMD EPYC 7443 24-core processor. For testing purposes, we select the checkpoint with the highest mean AUC achieved on the internal validation set during training.

Code Availability

Code will be available on <https://github.com/ChiariRay/AnkleNet> after peer review.

Reference

1. Peng Z, Huang W, Gu S, et al. Conformer: Local Features Coupling Global Representations for Visual Recognition. Published online May 9, 2021. Accessed October 17, 2023. <http://arxiv.org/abs/2105.03889>
2. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. Published online March 2, 2022. doi:10.48550/arXiv.2201.03545
3. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Published online June 3, 2021. Accessed June 3, 2023. <http://arxiv.org/abs/2010.11929>
4. Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In: ; 2021:10012-10022. Accessed October 13, 2023. https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.html
5. Chen CF (Richard), Fan Q, Panda R. CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification. In: ; 2021:357-366. Accessed April 12, 2023. https://openaccess.thecvf.com/content/ICCV2021/html/Chen_CrossViT_Cross-Attention_Multi-Scale_Vision_Transformer_for_Image_Classification_ICCV_2021_paper.html
6. Hung ALY, Zheng H, Miao Q, Raman SS, Terzopoulos D, Sung K. CAT-Net: A Cross-Slice Attention Transformer Model for Prostate Zonal Segmentation in MRI. *IEEE Transactions on Medical Imaging*. 2023;42(1):291-303. doi:10.1109/TMI.2022.3211764
7. Lin J, Lin J, Lu C, et al. CKD-TransBTS: Clinical Knowledge-Driven Hybrid Transformer with Modality-Correlated Cross-Attention for Brain Tumor Segmentation. *IEEE Trans Med Imaging*. 2023;PP. doi:10.1109/TMI.2023.3250474

Supplementary Figure Legends

Figure S1: Flowchart of retrospective inclusion process.

Figure S2: Calibration curves of AnkleNet on (A) the detection of lateral collateral ligament injury and (B) medial collateral ligament injury.

Figure S3: Confusion matrix of AnkleNet for CAI diagnosis in the external validation set.

Supplementary Tables

Table S1. Average Performance of Clinicians in Identifying RCAI from CAI patients

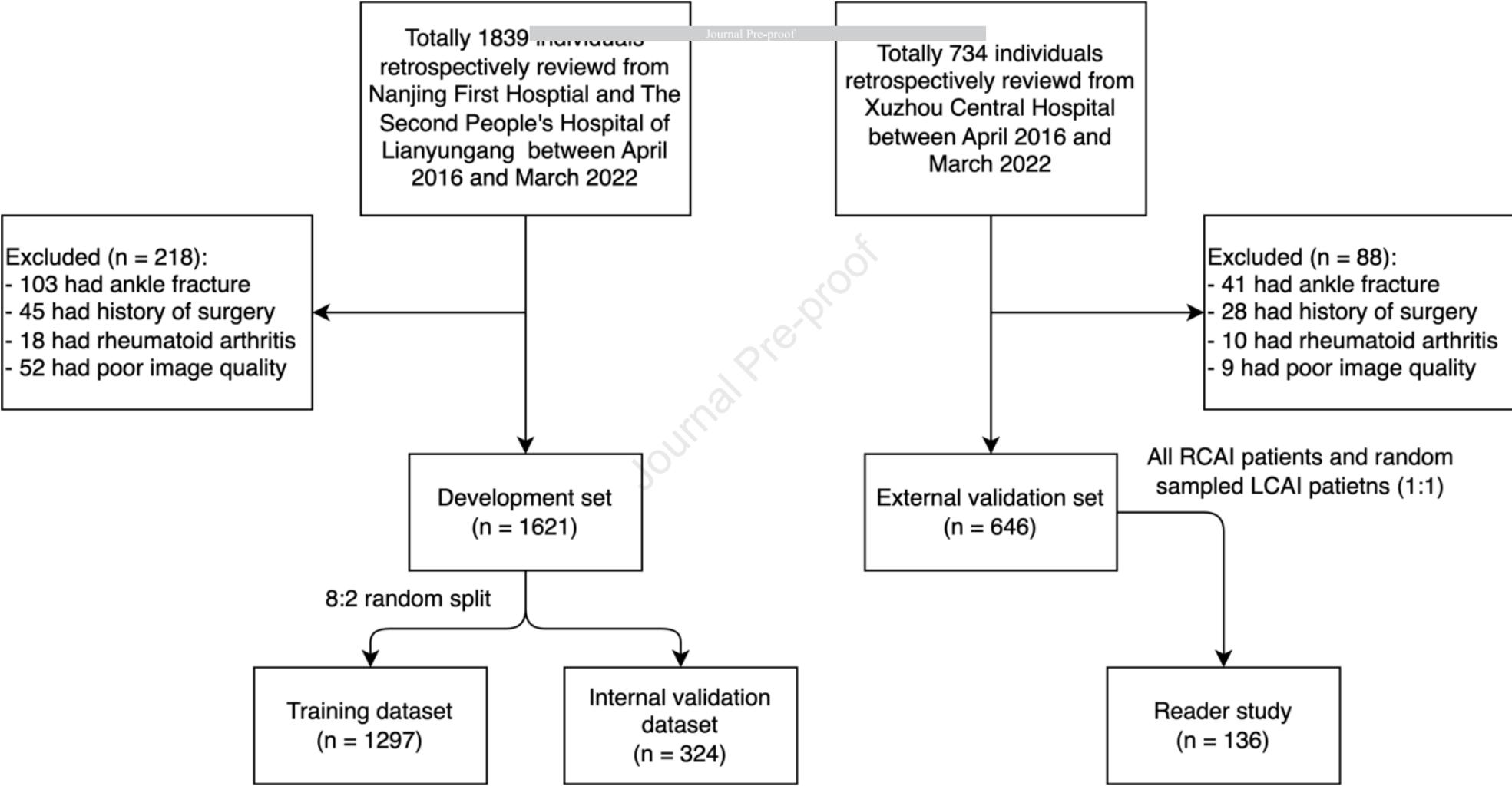
	Accuracy	Sensitivity	Specificity	Fleiss' Kappa
Without model	0.763 ± 0.078	0.712 ± 0.129	0.815 ± 0.07	0.402
With model	0.846 ± 0.046	0.841 ± 0.057	0.850 ± 0.048	0.685
Difference	0.082 ± 0.034[†]	0.129 ± 0.075[‡]	0.035 ± 0.042	0.283

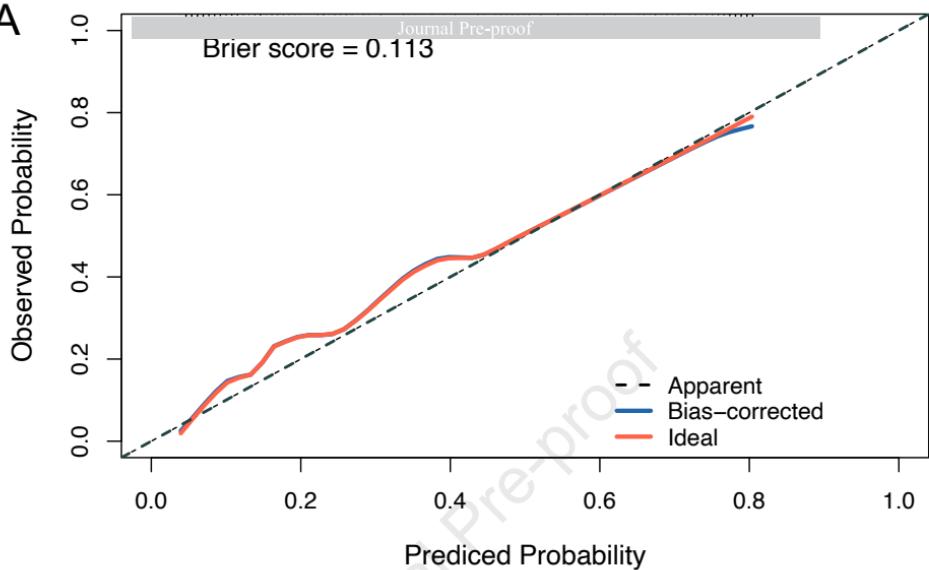
NOTE: [†] A one-tail *t* test shows significant improvement, *p-value* = 0.042; [‡] A one-tail *t* test shows significant improvement, *p-value* = 0.045.

Table S2. Comparison of Individual Unassisted and Assisted Clinicians in RCAI Identification

	Without model assistance			With model assistance		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Radiologist	0.699	0.706	0.691	0.801	0.824	0.779
Orthopedist 1	0.699	0.574	0.824	0.824	0.779	0.868
Orthopedist 2	0.728	0.603	0.853	0.816	0.809	0.824
MSK expert	0.824	0.794	0.853	0.875	0.868	0.882
Foot & Ankle expert	0.868	0.882	0.853	0.912	0.926	0.897

NOTE: MSK, musculoskeletal.



A**B**