Identifying and diagnosing meniscal tears with magnetic resonance imaging (MRI) using a combination of self-monitoring approaches

Mohammad Khanabadi[1], Toktam Khatibi[2], Mohammad Ayati Firoozabadi[3]

[1] Masters Student of Industrial Engineering, Tarbiat Modares University; mohammadkhanabadi@modares.ac.ir
[2] Associate Professor at the Faculty of Industrial and Systems Engineering, Tarbiat Modares University; toktam.khatibi@modares.ac.ir
[3] Associate Professor of Orthopaedics at the School of Medicine, Tehran University of Medical Sciences; dr.mohammad.ayati@gmail.com
* Corresponding author: Toktam Khatibi

ABSTRACT

The meniscus plays an important role in the anatomy of the human body. They are the tissues that make movement possible in the human knee, and by absorbing the extra loads that enter the knee, they try to minimize the load on the knee. they do Many studies show the relationship between meniscus tear and persistent knee pain, decreased function and early osteoarthritis. A problem in meniscus anatomy can lead to other injuries such as arthritis. Effective treatment can reduce the consequences of meniscal tears, improve quality of life, and reduce health care costs. Therefore, accurate diagnosis of meniscus tear is very important. This research introduces a novel self-supervised deep learning framework for the accurate diagnosis of meniscus tears using MRI images. The proposed method addresses the limitations of traditional supervised learning approaches by effectively utilizing unlabeled data to pre-train a convolutional neural network (CNN). The pre-trained model is then fine-tuned on a limited dataset of labeled images to classify meniscus tears. Our experiments demonstrate superior performance compared to existing methods, achieving an accuracy of 93% in diagnosing meniscus tears. The proposed framework offers a promising solution to the challenges of data scarcity in medical image analysis, paving the way for more efficient and accurate diagnosis of knee injuries.

Keywords: Meniscus tear - Deep learning models - Semi-supervised learning- Health data analysis – Magnetic resonance imaging (MRI) images

1. Introduction

The utilization of medical imaging technologies has become an essential part of modern medicine, enabling diagnostic decisions and treatment planning. The importance of medical imaging is exemplified by the consistent rate of growth in medical imaging utilization in modern healthcare . However, as the number of medical images relative to the available radiologists continues to become more disproportionate, the workload for radiologists continues to increase. Studies have shown that an average radiologist now needs to interpret one image every 3–4 s to keep up with clinical workloads . With such an immense cognitive burden placed on radiologists, delays in diagnosis and diagnostic errors are unavoidable . Thus, there is an urgent need to integrate automated systems into the medical imaging workflow, which will improve both efficiency and accuracy of diagnosis.

In recent years, deep learning models have demonstrated diagnostic accuracy comparable to that of human experts in narrow clinical tasks for several medical domains and imaging modalities, including chest and extremity X-rays, computed tomography (CT), magnetic resonance imaging (MRI), whole slide images (WSI), and dermatology images. While deep learning provides promising solutions for improving medical image interpretation, the current success has been largely dominated by supervised learning frameworks, which typically require large-scale labeled datasets to achieve high performance. However, annotating medical imaging datasets requires domain expertise, making large-scale annotations cost-prohibitive and time-consuming, which fundamentally limits building effective medical imaging models across varying clinical use cases[1].

To reduce reliance on labeled data, self-supervised learning (SSL) generates a supervised signal from a corpus of unlabeled data points. Contrastive SSL methods in particular have proven effectiveness in learning input image representations, as they group similar images together and dissimilar images apart

in the feature space. Popular SSL techniques include Bootstrap Your Own Latent (BYOL), Simple Contrastive Learning of Visual Representations (SimCLR), Simple Siamese (SimSiam), and Contrastive Predictive Coding (CPC)[2]. BYOL enhances SSL by focusing on different enhanced views of the same data. It employs a target network that predicts and learns from an online network. SimCLR, on the other hand, attempts to reduce the similarity between distinct data samples while amplifying it across various transformations or perspectives of the same data. In contrast, CPC models SSL as a predictive task, learning to predict future input data representations by framing SSL as a Contrastive Learning for Liver Cancer Tumor Classification 313 contrastive learning problem. SimSiam, introduced by Chen et al., is an SSL strategy that advances conventional methods by creating representations from various transformations of the same image, which boosts the model's resilience to data variations. Through this mechanism, SimSiam could facilitate the extraction of more meaningful and generalized features from images [3]. Hamm et al. expanded the use of CNNs in the diagnosis of liver tumors using multi-phasic magnetic resonance imaging (MRI). They proposed a CNN classifier that was trained to distinguish between six common hepatic lesion types. The successful integration of CNNs with MRI data in this work highlights the adaptability of deep learning techniques and their potential to increase the diagnostic precision of liver cancer classification across multiple imaging modalities [4]. Aligned with these applications, the Inception, Xception, and ResNet152 models stand out among deep learning classifiers in recent years. Inception utilizes factorized convolutions and dimension reductions to optimize computation and accuracy, making it suitable for large-scale applications [5]. To improve feature extraction and classification capabilities, Xception substitutes Inception modules in an inventive way with depth wise separable convolutions [6]. To prevent gradient vanishing or exploding, ResNet152, a more complex variant of the original ResNet, incorporates skip connections [7]. Owing to their unique advantages and proven track records of performance, these models emerge as among the top candidates for classification tasks. Among the few studies that used these CNN classifiers for liver classification, Nakata et al. explored the effectiveness of ensemble CNN models, including Xception, Inception and ResNet152, in classifying ultrasound images of liver tumors. In another similar study, Chen et al. proposed a novel CNN network to classify different types of liver cancer histopathological images, and compared its performance with other CNN classifiers, including ResNet variations [8].

In this study, MRI data from 100 patients was collected at Imam Khomeini Hospital, Tehran. The data was structured such that three-dimensional MRI images were obtained for each patient. However, due to the invisibility of the meniscus in certain image sequences, these sequences were excluded from the analysis. Only frames where the meniscus was clearly visible were utilized to enhance model accuracy and reduce computational complexity.

The primary objective of this research is to identify the optimal encoder for constructing a robust generative adversarial network (GAN). To achieve this, we first extract features from generated images using pre-trained models. Subsequently, we employ traditional machine learning models to build various classifiers on the extracted feature sets from each model. The accuracies of these classifiers are then compared, and the most effective feature extractor is selected

2. Methods

In this research, a supervised learning model was designed to identify the optimal feature extractor or encoder. Subsequently, a contrastive learning model based on the SimSiam architecture was proposed.

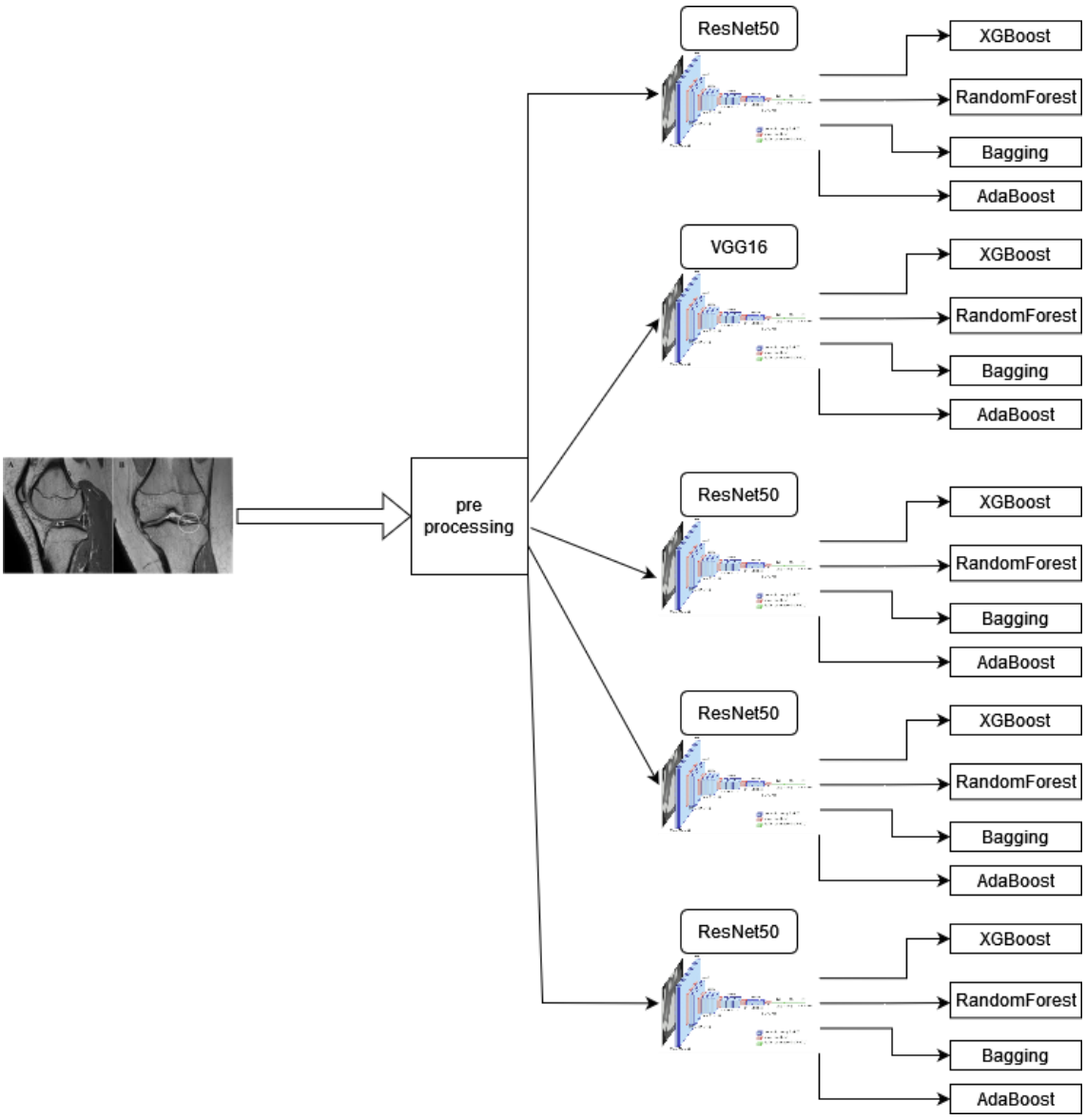


Figure 1- Graphic abstract of the first step

As illustrated in Figure 1, the initial phase of our research aims to identify the optimal encoder for the subsequent development of a robust contrastive learning model. To achieve this, pre-trained models were employed to extract features from the synthetic images. These extracted features were then subjected to traditional machine learning algorithms to train various classifiers. By comparing the performance of these classifiers, the most effective feature extractor was determined.

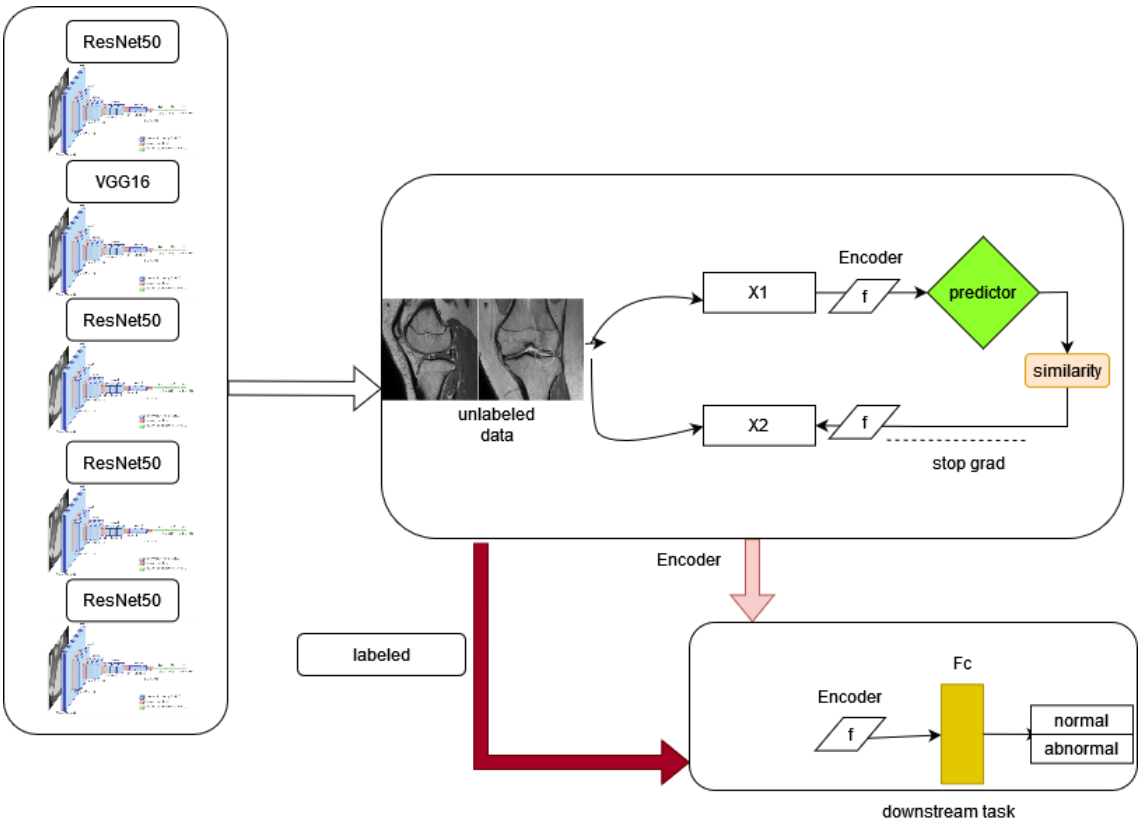


Figure 2- Graphic abstract of the second step

The primary objective of the second stage is to construct a robust self-supervised contrastive learning model. Leveraging the optimal pre-trained models identified in the preceding phase as the backbone encoder, a SimSiam-based architecture was employed. The performance of this self-supervised model was then evaluated and compared.

The SimSiam enabled SSL, allowing the model to discover reliable and generalizable representations before further fine-tuning [9]. SimSiam is chosen in our study for its privileges in computational efficiency, enhanced learning efficacy, and robustness against common learning challenges [10].

To evaluate the performance of the classification models in this study, we employed several metrics: specificity, sensitivity, precision, and accuracy. These metrics enable us to quantitatively assess the models' ability to discriminate between different classes of data. A confusion matrix was utilized to calculate these metrics. This matrix serves as a valuable tool for analyzing the performance of a classification model by tabulating the number of correct and incorrect predictions for each class.

2.1. Preventing Data Leakage

In this study, the patient dataset is structured such that each patient has a dedicated folder containing all their associated images. To prevent data leakage and ensure the integrity of patient-level information, the data was partitioned into training and testing sets while maintaining patient-wise consistency. Specifically, if an image from a particular patient was included in the training set, all images associated with that patient were also assigned to the training set. This approach guarantees that the machine learning model does not inadvertently learn information about patients in the testing set during training, thereby providing a more realistic evaluation of the model's performance on unseen data.

Data leakage refers to the situation where information that should not be available to the machine learning model is inadvertently introduced during the training process. This can lead to an overestimation of the model's performance and a failure to generalize to new, unseen data. In the context of this study, if images from the same patient were randomly assigned to both the training and testing sets, the model could learn patterns specific to that particular patient, resulting in poor performance when evaluated on new patients.

2.2. Pre-Processing

MRI images were acquired from the PACS system in JPEG format and resized to a resolution of 512x512 pixels. Subsequently, image sequences where the meniscus was not clearly visible were excluded to minimize errors in the neural network model. Given the large volume of images and to reduce computational costs, the image resolution was further down sampled to 128x128 pixels during the preprocessing stage. Additionally, to ensure pixel values were within a consistent range, pixel normalization was applied using the min-max scaling method, mapping the pixel values to the interval [0, 1].

2.3. Pre-training Models Using SimSiam

In our SimSiam implementation, we leverage positive pairings, which are different augmented versions of the same image. These augmentations include both geometric and color jittering transformations. By contrasting these distinct views of the same image, our model is trained to identify invariant features. This not only simplifies computations but also eliminates the need for maintaining a balance between positive and negative samples. We've also incorporated the stop-gradient mechanism into the model. This step is crucial to prevent the model from learning irrelevant patterns - a common issue seen in most SSL pretraining procedures. The symmetric architecture of our SimSiam implementation ensures equal weight distributions across both pathways. This feature enhances stability and helps generate reliable features.

3. Experimental Results and Discussion

This section evaluates the performance of the first stage of the research. In the initial stage, as previously mentioned, features were extracted from pre-trained image models, followed by classification using traditional machine learning models. It is important to note that 80% of the data was allocated for model training, while the remaining 20% was used for testing. The performance of the models on the test data is presented in the table using the mean accuracy (MAcc) metric.

Adaboost	Bagging	RF	XGBoost	Mean Accuracy
%84	%88	%88	%86	**InceptionResNetV2**
%88	%85	**%90**	%89	**EfficientNetB0**
%86	%85	**%92**	%89	**VGG16**
%85	%88	**%91**	%88	**VGG19**
%83	%85	%90	**%93**	**ResNet50**
%86	%87	%89	**%90**	**ResNet50V2**
%80	%78	%81	%80	**ViTB16(imagenet21k)**

%73	%76	%83	%81	**ViTL16(imagenet21k+2012)**
%81	%78	%84	%83	**ViTL32(imagenet21k+2012)**
%77	%82	%82	%80	**ViT L32 (imagenet21k)**
%77	%78	%82	%81	**Swin Large 384**

Table 1- Performance report of various feature extractors with average accuracy metrics for each category

As can be seen in the table 1, the combination of ResNet50 and XGBoost models show the best performance for feature extraction and class classification. It can also be seen below the performance of the ResNet50 model, which has the best performance:

ResNet50					
AUC	**F-score**	**Recall**	**Precision**	**Accuracy**	_
0.93	**%93**	**%93**	**%93**	**%93**	**XGBoost**
0.90	%90	%90	%91	%90	**RF**
0.85	%85	%85	%86	%85	**Bagging**
0.83	%83	%83	%83	%83	**AdaBoost**

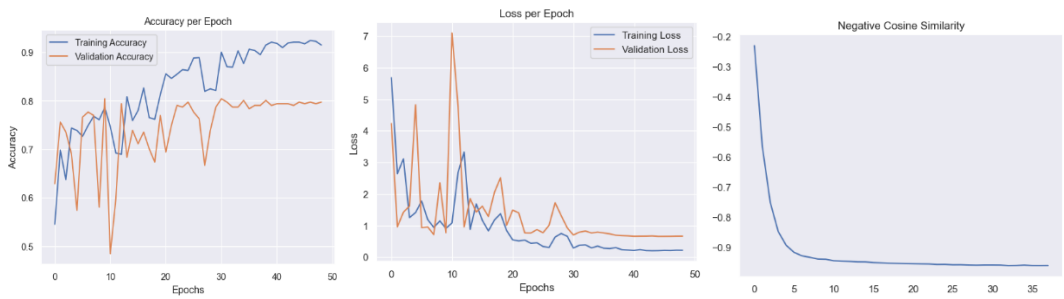
Table 2- Classification model accuracy by features extracted from ResNet50

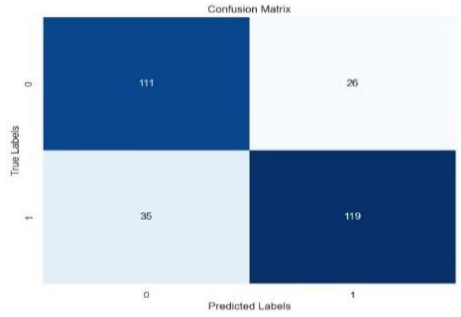
In the second step, the data sets are 80% data for training without labels and 20% data for model testing with labels. By using different models placed in the backbone part of the SimSiam model, it resulted in the results of Table 3.

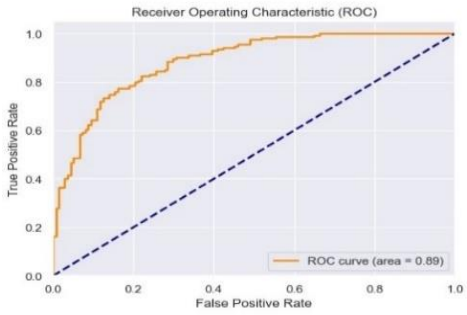
AUC	F-score	Recall	Precision	Accuracy	
0.81	%76	%77	%75	%75	**VGG16**
0.81	%72	%72	%73	%72	**VGG19**
0.91	**%83**	**%83**	**%83**	**%83**	**ResNet50**
0.89	%79	%79	%79	%79	**ResNet50 V2**
0.81	%73	%73	%74	%74	**EfficientNet B0**

Table 3-Evaluation of different models with 20% labeled test data

Figure 3 shows the loss function of the SimSiam model using the ResNet50 encoder when extracting the representation using unlabeled data. The next graph shows the loss function per epoch when evaluated with 20% of the labeled data. Also, the last graph shows the accuracy function for each epoch for evaluation.







4. Conclusion

The primary objective of this research was to develop a model capable of accurately classifying and detecting meniscus tears from MRI images using minimal labeled data. Given the complexity and subjectivity involved in diagnosing meniscus tears, often requiring the expertise of specialized physicians, this research aims to assist medical professionals in diagnosis and patient screening. To achieve this goal, a series of steps were proposed and evaluated in previous sections, ultimately resulting in a model that demonstrated satisfactory accuracy and effectively addressed the research objectives.

References

[1] S. C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, "Self-supervised learning for medical image classification: a systematic review and implementation guidelines," Dec. 01, 2023, *Nature Research*. doi: 10.1038/s41746-023-00811-0.

[2] D. Pup Federico and A. Manfredo, "Applications of Self-Supervised Learning to Biomedical Signals: where are we now," 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/

[3] C. Chen *et al.*, "Classification of multi-differentiated liver cancer pathological images based on deep learning attention mechanism," *BMC Med Inform Decis Mak*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12911-022-01919-1.

[4] C. A. Hamm *et al.*, "Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI," *Eur Radiol*, vol. 29, no. 7, pp. 3338–3347, Jul. 2019, doi: 10.1007/s00330-019-06205-9.

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[6] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Institute of Electrical and Electronics Engineers Inc., Nov. 2017, pp. 1800–1807. doi: 10.1109/CVPR.2017.195.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[8] C. Chen *et al.*, “Classification of multi-differentiated liver cancer pathological images based on deep learning attention mechanism,” *BMC Med Inform Decis Mak*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12911-022-01919-1.

[9] M. N. Alam *et al.*, “Contrastive learning-based pretraining improves representation and transferability of diabetic retinopathy classification models,” *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-33365-y.

[10] X. Chen and K. He, “Exploring Simple Siamese Representation Learning.” [Online]. Available: https://github.com/facebookresearch/simsiam