

Patient-specific radiomic feature selection with reconstructed healthy persona of knee MR images

Yaxi Chen^{1,2}[0009-0007-5906-899X], Simin Ni³[0009-0007-2780-6118], Shaheer U. Saeed^{2,4}[0000-0002-5004-0663], Aleksandra Ivanova³[0009-0000-4113-8928], Rikin Hargunani⁵[0000-0002-0953-8443], Jie Huang¹[0000-0001-7951-2217], Chaozong Liu^{3,5}[0000-0002-9854-4043], and Yipeng Hu^{2,4}[0000-0003-4902-0486]

¹ Mechanical Engineering Department, University College London, London, UK

² Hawkes Institute, University College London, London, UK

³ Institute of Orthopaedic & Musculoskeletal Science, University College London, Royal National Orthopaedic Hospital, Stanmore, UK

⁴ Department of Medical Physics and Biomedical Engineering, University College London, London, UK

⁵ Royal National Orthopaedic Hospital, Stanmore, UK

Abstract. Classical radiomic features (e.g., entropy, energy) have been designed to describe image appearance and intensity patterns. These features are directly interpretable and readily understood by radiologists. Compared with end-to-end deep learning (DL) models, lower dimensional parametric models that use such radiomic features offer enhanced interpretability but lower comparative performance in clinical tasks. In this study, we propose an approach where a standard logistic regression model performance is substantially improved by learning to select radiomic features for individual patients, from a pool of candidate features. This approach has potentials to maintain the interpretability of such approaches while offering comparable performance to DL. In addition, we also propose to expand the feature pool by generating a patient-specific healthy persona via mask-inpainting using a denoising diffusion model trained on healthy subjects. Such a pathology-free baseline feature set allows not only further opportunity in novel feature discovery but also improved condition classification. We demonstrate our method on multiple clinical tasks of classifying general abnormalities, anterior cruciate ligament tears, and meniscus tears. Experimental results demonstrate that our approach achieved comparable or even superior performance than state-of-the-art DL approaches while offering added interpretability through the use of radiomic features extracted from images and supplemented by generating healthy personas. Example clinical cases are discussed in-depth to demonstrate the interpretability-enabled utilities such as human-explainable feature discovery and patient-specific location/view selection. These findings highlight the potentials of the combination of subject-specific feature selection with generative models in augmenting radiomic analysis for more interpretable decision-making. The codes are available at: <https://github.com/YaxiiC/RadiomicsPersona.git>

Keywords: Knee Joint · Radiomics · Diffusion Models · MRI.

1 Introduction

Anterior cruciate ligament (ACL) tears are the most frequent acute knee injuries[19]. Diagnosis is primarily based on physical examinations for knee instability. For instance, The clinical Lachman test is reported to be 81% sensitive and 81% specific for diagnosing a complete ACL rupture [3]. Meniscus injuries are a common type of damage to cartilage in the knee, arising either from acute trauma or degenerative changes [16]. Physical examination tests diagnose meniscus tears with a 60–70% sensitivity and a 70% specificity [8]. Due to low accuracy in these physical examinations, magnetic resonance (MR) imaging stands as the gold-standard for diagnoses, offering critical characterization of the tear depth, location, pattern, tissue quality, and integrity of any previous meniscal repair.

End-to-end Deep Learning has been employed in most existing automated approaches to MR-based diagnosis of knee injuries. Most previous studies in this area have utilized end-to-end convolutional neural networks (CNNs). For instance, Bien et al. [1] utilized an end-to-end classifier and Tsai et al. [21] applied an a lightweight efficiently layered network (ELNet), and Liu et al. [12] proposed ACLNet, which integrates MRI data with morphological point cloud features of the femur and tibia to improve ACL injury prediction. While methods like class activation mappings [1] have been used to enhance interpretability, a major limitation of these DL models remains their “black box” nature. The lack of transparency and thus human understanding of how image-level features are learned and decisions are made remains an interesting challenge.

Radiomics has long contributed to screening, diagnosis, treatment planning, and follow-up [23]. Hand-crafted radiomics features, e.g. intensity, shape, texture, wavelets, provide specific information regarding region of interest (ROI), while DL features may capture more complex patterns [20]. However, these DL-based radiomics also inherits the interpretability challenge common to most DL methods. Integrating hand-crafted radiomics features and DL can potentially yield more precise and robust prognostic models [4].

Generative models have been extensively applied to medical imaging for tasks such as image quality enhancement, domain transfer, and augmentation of training data [18,5]. For instance, generative adversarial networks (GANs) [6] were employed to reconstruct images from incomplete data and reduce noise in low-dose CT and under-sampled MRI scans. Recently, denoising diffusion probabilistic model (DDPM) [9,15] advanced image generation tasks including those from medical imaging [11]. Generative models have also been applied in anomaly detection and image-to-image translation. For example a localized reconstruction approach used a single-mask inpainting technique to refine both anomaly detection and segmentation outcomes [2]. Besides, exploring generative models in our diagnostic tasks and its interpretation remains limited.

In this work, we propose a DL-based radiomic feature selection framework for musculoskeletal disease classification, incorporating a healthy persona - a synthesized patient-specific healthy baseline for pathological regions. The term persona, inspired by the namesake concept in psychology [10], represents a predicted healthy counterpart of a patient’s anatomy without the imaged manifes-

tation of the pathology of interest. Our approach employs a feature-weighting neural network to predict probabilities of the feature being selected, dynamically adjusting the weighting/selection of radiomic features extracted from both the pathological MRI and its healthy persona. These selected features serve as inputs to a logistic regression model, which performs a downstream clinical task. The feature-weighting neural network and logistic regression function are trained simultaneously supervised solely by the downstream task labels. To construct the healthy persona, we utilize a DDPM trained exclusively on healthy MRI scans, to reconstruct a pathology-free version of the ROI, enabling direct comparisons between the original pathology and its synthetic healthy baseline learned from healthy population. By incorporating this patient-specific persona and radiomic feature selection, this framework provides a more interpretable and clinically improved assessment of disease progression.

To summarize, our main contributions are: 1) We introduce a framework that learns individualized weighting/ selection for patch-based radiomic features, offering more precise disease characterization; 2) We propose a method to reconstruct a healthy persona for each patient’s pathological image to augment radiomic features that are subsequently used to characterize diseases; 3) We achieved comparable or even superior performance compared to traditional end-to-end approaches in knee joint MRI analysis, while offering enhanced interpretability, flexibility, and revealing potential imaging biomarkers; 4) All code is openly shared to ensure reproducibility and advance radiomics-based medical image analysis.

2 Methods

Pool of features: A feature-extraction function $e(\mathbf{x}) = \mathbf{f} = \{f_i\}_{i=1}^F$ computes a fixed pool of F features $\mathbf{f} = \{f_i\}_{i=1}^F$, from an image ROI patch $\mathbf{x} \in \mathbb{R}^{H \times W \times D}$, which corresponds to the location of the pathology, where $H \times W \times D$ denote the height, width, and depth of the patch, respectively.

Feature probability prediction: A feature-weighting neural network predicts feature probabilities:

$$g_\theta(\mathbf{x}) = \mathbf{p} = \{p_i\}_{i=1}^F = P(\mathbf{f}|\mathbf{x}; \theta),$$

where $\mathbf{p} = \{p_i\}_{i=1}^F = P(\mathbf{f}|\mathbf{x}; \theta)$ denotes the feature probabilities given an image patch \mathbf{x} , and θ denotes neural network parameters.

Downstream task: The downstream task uses logistic regression to classify an image into categories (such as classifying an image into disease categories). The model inputs are weighted features $\mathbf{f}^w = \{p_i f_i\}_{i=1}^F$, which are used in the logistic regression model to predict downstream task class probabilities $r_\phi(\mathbf{f}^w) = P(\mathbf{c}|\mathbf{x}, \mathbf{p}; \theta, \phi)$, where the downstream task classification probabilities, given the image \mathbf{x} and feature probabilities \mathbf{p} , are denoted by $P(\mathbf{c}|\mathbf{x}, \mathbf{p}; \theta, \phi)$, where the image features are computed from the image patch $e(\mathbf{x}) = \mathbf{f}$ and the feature probabilities are computed using the feature selection network $g_\theta(\mathbf{x}) = \mathbf{p}$, with parameters θ . The classifier parameters are denoted by ϕ .

Probabilistic interpretation: Using the chain rule of probability, the joint probability of the classes and features, given an image can be expressed as:

$$P(\mathbf{c}, \mathbf{f}|\mathbf{x}; \theta, \phi) = P(\mathbf{c}|\mathbf{f}, \mathbf{x}; \phi) \times P(\mathbf{f}|\mathbf{x}; \theta),$$

where $P(\mathbf{f}|\mathbf{x}; \theta)$ is predicted by the feature selection network $g_\theta(\mathbf{x})$, and $P(\mathbf{c}|\mathbf{f}, \mathbf{x}; \phi)$ is predicted by the logistic regression classifier $r_\phi(\mathbf{f}^w)$.

The goal during training is to model the classification probability given the image, $P(\mathbf{c}|\mathbf{x}; \theta, \phi)$, where feature probabilities act as intermediate variables and we do not need to explicitly model $P(\mathbf{c}, \mathbf{f}|\mathbf{x}; \theta, \phi)$. Since the feature extraction function $\mathbf{f} = e(\mathbf{x})$ is deterministic, the conditional distribution $P(\mathbf{f}|\mathbf{x}; \theta, \phi)$ is concentrated at $\mathbf{f} = e(\mathbf{x})$ (behaving like a Dirac-delta function with probability 1 at $\mathbf{f} = e(\mathbf{x})$). This allows simplification to marginalise out the intermediate variables in two steps: 1) simplify the expression using the chain rule, and 2) collapse the integral due to the deterministic function $\mathbf{f} = e(\mathbf{x})$. This gives:

$$\begin{aligned} \int P(\mathbf{c}, \mathbf{f}|\mathbf{x}; \theta, \phi) d\mathbf{f} &= \int P(\mathbf{c}|\mathbf{f}, \mathbf{x}; \phi) P(\mathbf{f}|\mathbf{x}; \theta) d\mathbf{f}, \\ &= P(\mathbf{c}|\mathbf{f} = e(\mathbf{x}), \mathbf{x}; \theta, \phi), \\ &= P(\mathbf{c}|\mathbf{x}; \theta, \phi) \end{aligned}$$

Training and optimization: Following from the simplification, given a dataset of N training examples $(\mathbf{x}^i, \hat{\mathbf{c}}^i)_{i=1}^N$, where $\hat{\mathbf{c}}^i$ is the ground-truth label for image \mathbf{x}^i , the objective is to find parameters (θ, ϕ) that maximise the likelihood of predicting the correct labels: $\mathcal{L}(\theta, \phi) = \prod_{i=1}^N P(\hat{\mathbf{c}}^i|\mathbf{x}^i; \theta, \phi)$. The negative log-likelihood is used to facilitate the commonly-adopted cross-entropy loss in classification tasks:

$$\mathcal{L}_{loss}(\theta, \phi) = - \sum_{i=1}^N \log P(\hat{\mathbf{c}}^i|\mathbf{x}^i; \theta, \phi)$$

Minimizing this loss using gradient descent ensures that the predicted class probabilities align with the ground-truth labels, i.e. $(\theta^*, \phi^*) = \arg \min_{\theta, \phi} \mathcal{L}_{loss}(\theta, \phi)$.

In summary, the likelihood function captures how well the model’s predicted probabilities align with the ground-truth labels, and minimising loss directly trains the model to improve classification accuracy as well as predict intermediate feature weightings/ probabilities that ought to aid this classification.

Details of feature extraction: The feature extractor $e(\mathbf{x})$ first divides the image \mathbf{x} into P subpatches. For each subpatch, m radiomic features are computed in three spatial dimensions, giving a total of $3mP$ features, i.e. $F = 3mP$.

Expanding the pool of features through a healthy persona: First, we train a 3D conditional DDPM model [9] using healthy patient images. The model is trained to reconstruct healthy images from a masked version of the original images. The masks are applied to the center of the image, where all pathological features are considered to be present. At inference, we can then mask a

pathological image and the DDPM can then reconstruct the image, leading to its healthy persona $\mathbf{x}^{persona}$. Further details on datasets and DDPM model are outlined in Sec.3.1. For each image, $3mP$ radiomic features are then computed on the healthy persona reconstructed by the DDPM and added to the pool of features, making the total number of features per-image $F = 6mP$.

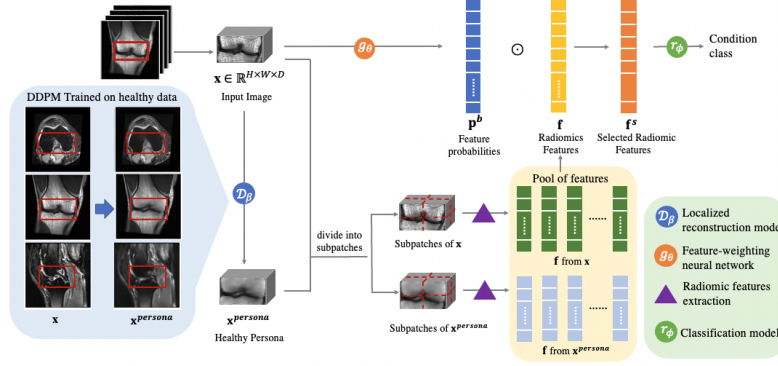


Fig. 1: Overview of the radiomic feature selection framework (details in Sec.2).

Feature selection at inference: At inference, the predicted feature probabilities $g_\theta(\mathbf{x}) = \mathbf{p} = \{p_i\}_{i=1}^F$ are binarized to conduct a hard selection as opposed to the soft weighting during training. The binarized feature probabilities are $\mathbf{p}^b = \{p_i^b\}_{i=1}^F$, where $p_i^b = 1$ if $p_i \geq T$, 0 otherwise, T is an application-specific threshold hyperparameter. These binarized values are then used to select features, resulting in the final set of selected features $\mathbf{f}^s = \{p_i^b f_i\}_{i=1}^F$, which can be used to determine the final predicted disease classification from logistic regression: $r_\phi(\mathbf{f}^s)$.

Note that the expectation of the feature weighting during training is proportional to the selected feature at inference, $\mathbb{E}[\mathbf{f}^w] \propto \mathbf{f}^s$ and $P(\mathbf{f}^w|\mathbf{x};\theta) \propto P(\mathbf{f}^s|\mathbf{x};\theta)$.

3 Experiments and Results

3.1 Dataset and Implementation Details

Dataset and Preprocessing: We use the MRNet dataset [1], which consists of 1,370 3D knee MRI studies, acquired from axial, coronal, and sagittal views. Each case is labeled for the presence of abnormalities, an ACL tear, or a meniscal tear. For our study, we utilize 1,130 cases for training and 120 for validation. All images are preprocessed and resized to $32 \times 128 \times 128$, tested with and without an affine registration (NiftyReg [14] in this study).

DDPM model training: We implement a DDPM3D to reconstruct MRI scans by progressively denoizing corrupted inputs. Our model follows a diffusion process parameterized by a sequence of variance schedules β_t , where we define β_t to

linearly increase from 10^{-4} to 0.02 over $T = 1000$ timesteps. During training, we utilize the healthy subset of the MRNet dataset, which consists of MRI volumes labeled as *abnormal* = 0, comprising 218 patients and a total of 654 scans across three views. Each scan was masked by a central bounding box, with a 50% depth, a 30% height, and a 50% width of the original image size. Preliminary experimentation suggested that the DDPM framework provides qualitatively better reconstruction results compared to both GAN-based models and ViT-based autoencoders.

Feature Extraction and selection: We extract features using a PyTorch implementation based on the PyRadiomics library [22]. We compute *First-Order Features* and *3D Shape Features*, resulting in a total of 35 feature types. Both the pathological image and its DDPM-generated healthy persona are divided into $2 \times 2 \times 2$ subpatches. For each patient, we extract features from these subpatches in the axial, coronal, and sagittal planes, resulting in a final feature vector of size $3 \times 8 \times 2 \times (35 + 3) = 1824$. We employ a 3DResNet-18 [7], as the feature-weighting neural network to learn patient-specific feature probabilities.

Classification model: During training, the probability-weighted feature vector \mathbf{f}^w is fed into the logistic regression classifier to predict class probabilities for the downstream classification task. After training, the optimal classification thresholds are determined using Youden’s index on the validation set. During inference, we apply a threshold 0.4 for hard feature selection. We also examined the effect of varying the feature selection threshold. As expected, higher thresholds resulted in poorer performance (further details in Tab. 1), and 0.4 was selected for its balance between performance and number of selected features.

Comparison and Ablation Experiments: In summary, we compare our method with end-to-end approaches, including MRNet [1], ELNet [21], and SKID [13] without pre-training, achieving competitive performance with much enhanced interpretability. To evaluate the impact of key components, we conducted ablation studies. Removing the feature-weighting neural network and using raw radiomic features resulted in suboptimal performance, highlighting the importance of adaptive weighting. Different number of subpatch ($1 \times 1 \times 1$ and $3 \times 3 \times 3$) were analyzed, with $2 \times 2 \times 2$ providing the best balance of performance and interpretability. Excluding persona and using only the original ROI reduced classification accuracy, demonstrating the benefit of synthetic healthy baselines. Omitting registration led to decreased performance.

3.2 Results and Interpretation

Our proposed method with $2 \times 2 \times 2$ subpatches outperforms existing models in detecting abnormalities and meniscal tears, achieving 0.90 and 0.82 accuracy, respectively. This is significantly better than ELNet, an end-to-end DL method, in meniscus tear detection (all *p-values* < 0.050 across all evaluated metrics, paired t-test at a significance level $\alpha=0.05$). Increasing the number of subpatches to $3 \times 3 \times 3$ further enhances performance, achieving an accuracy of 0.87 for both ACL and meniscal tear detection. This surpasses the reported accuracy of MRNet and SKID (data unavailable for paired tests). Our method exhibits high

Table 1: Comparison of different configurations and methods. N: Number of subpatches; PFS: Patient-specific features selection; HP: Healthy persona; T: Feature selection threshold; Reg: Registration in preprocessing. * denotes reproduced results

Method	N	PFS	HP	T	Reg	Type	Evaluation Metrics			
							Acc	Sen	Spe	AUC
Ours	2*2*2	✓	✓	0.4	✓	abn	0.90±0.13	0.94±0.10	0.77±0.25	0.85±0.16
						acl	0.81±0.10	0.92±0.08	0.66±0.23	0.80±0.12
						men	0.82±0.11	0.81±0.14	0.88±0.14	0.84±0.11
Ours	3*3*3	✓	✓	0.4	✓	abn	0.84±0.12	0.96±0.08	0.54±0.33	0.75±0.25
						acl	0.87±0.16	0.88±0.13	0.87±0.22	0.87±0.15
						men	0.87±0.23	0.88±0.16	0.87±0.30	0.88±0.20
FST=0.0	2*2*2	✓	✓	0.0	✓	abn	0.90±0.11	0.99±0.03	0.63±0.42	0.81±0.21
						acl	0.84±0.13	0.83±0.23	0.84±0.16	0.83±0.15
						men	0.83±0.16	0.82±0.26	0.83±0.13	0.83±0.17
FST=0.5	2*2*2	✓	✓	0.5	✓	abn	0.76±0.18	0.72±0.22	0.90±0.15	0.81±0.17
						acl	0.75±0.17	0.79±0.18	0.71±0.36	0.75±0.15
						men	0.78±0.13	0.74±0.11	0.85±0.17	0.82±0.09
1patch	1*1*1	✓	✓	0.4	✓	abn	0.70±0.35	0.64±0.46	0.90±0.15	0.78±0.23
						acl	0.73±0.15	0.96±0.06	0.56±0.26	0.76±0.13
						men	0.60±0.08	0.76±0.27	0.50±0.28	0.64±0.05
NoPsa	2*2*2	✓	✗	0.4	✓	abn	0.77±0.13	0.75±0.19	0.87±0.18	0.81±0.10
						acl	0.66±0.18	0.92±0.10	0.45±0.36	0.68±0.15
						men	0.72±0.10	0.80±0.15	0.68±0.10	0.74±0.10
NoFS	2*2*2	✗	✓	0.4	✓	abn	0.81±0.03	0.93±0.05	0.53±0.22	0.68±0.09
						acl	0.77±0.11	0.83±0.12	0.60±0.32	0.72±0.14
						men	0.71±0.05	0.78±0.07	0.66±0.11	0.72±0.03
NoReg	2*2*2	✓	✓	0.4	✗	abn	0.77±0.08	0.57±0.13	0.96±0.04	0.76±0.06
						acl	0.73±0.10	0.40±0.29	0.99±0.03	0.70±0.10
						men	0.67±0.15	0.99±0.03	0.41±0.27	0.70±0.14
ELNet*	-	-	-	-	-	abn	0.80±0.01	0.95±0.01	0.22±0.03	0.73±0.01
						acl	0.70±0.10	0.72±0.26	0.68±0.04	0.73±0.08
						men	0.65±0.06	0.61±0.13	0.63±0.07	0.69±0.03
MRNet	-	-	-	-	-	abn	0.85	0.87	0.71	0.94
						acl	0.87	0.76	0.97	0.92
						men	0.73	0.71	0.74	0.82
SKID	-	-	-	-	-	abn	0.87	0.98	0.49	0.90
						acl	0.80	0.74	0.85	0.89
						men	0.73	0.92	0.57	0.81

sensitivity while maintaining comparable specificity, demonstrating comparable or even superior performance across multiple configurations.

Ablation studies demonstrate that our method significantly outperforms the configurations without registration and without persona in ACL detection, achieving higher accuracy with *p-values* of 0.018 and 0.023, respectively. Furthermore, it achieves a statistically significant improvement in accuracy over the single-patch configuration in meniscus detection (*p-values*=0.043).

ACL tears case studies: For ACL injuries, key features are concentrated in the sagittal view and the medial/posterior regions of the ROI. This aligns with both clinical expectations, as the sagittal view provides the clearest visualization of ACL continuity, and anatomically, the ACL originates from the intercondylar eminence of the tibia and inserts onto the medial aspect of the lateral femoral condyle[17]. The key radiomic features in the medial and posterior subpatches for ACL tears include *Entropy*, *Interquartile Range*, and *Compactness*. Entropy quantifies the complexity and irregularity in the image intensity patterns, typ-

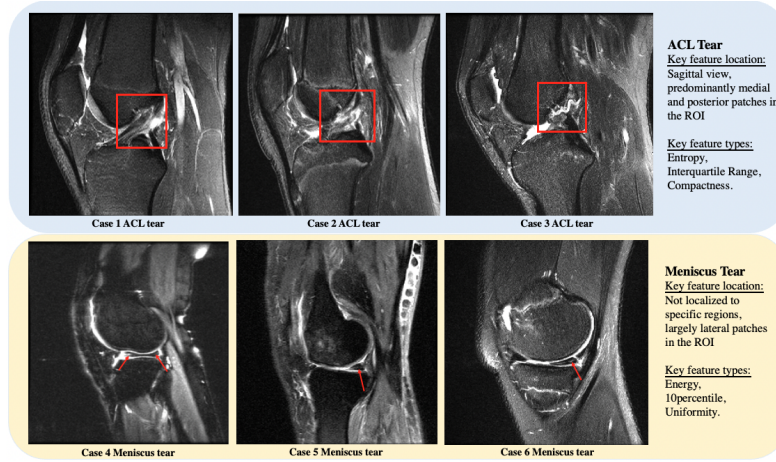


Fig. 2: Representative cases of ACL and Meniscus Tears (details in Sec.3.2).

ically increasing when torn ligament fibers appear irregular and heterogeneous (see Fig. 2). Interquartile Range, representing the spread of voxel intensities in the middle half of the patch, captures the intensity heterogeneity indicative of discontinuous fibers. Compactness, quantifying how closely a shape approximates a sphere, reflects structural deformation and may highlight secondary signs such as posterior cruciate ligament laxity and anterior tibial translation.

Meniscus tears case studies: In contrast, key features for meniscus tear are not localized to specific regions, as lesions can occur in either or both menisci. On PD-weighted MRI, these tears characteristically appear as a linear high-intensity signal coursing through the meniscus, typically extending to its superior or inferior surface. Key radiomic features in the lateral subpatches include *Energy*, which measures overall tissue integrity and decreases with disruption of the meniscal substance, *Uniformity*, which detects irregular voxel intensities in torn regions, and the *10percentile* feature, which highlights subtle changes in the lower-intensity range of the meniscal signal. Given the variability in tear location and morphology, patient-specific radiomic feature selection remains essential for robust detection and quantification of meniscal pathology.

Model prediction-informed image view selection: Building on the observation in the model prediction, and expected radiologically, that ACL-related features concentrate in the sagittal view, we refined our framework by extracting features exclusively from sagittal slices while maintaining other configurations. This approach improved the performance to 0.92 accuracy, 0.93 sensitivity, 0.91 specificity, and 0.92 AUC.

4 Discussion and Conclusion

Our study presents a novel framework for knee MRI analysis that integrates generative modeling and logistic regression classification to enable an individualized learning-based radiomic selection, for pathology detection with improved interpretability. By introducing a patient-specific healthy persona, we provide a baseline comparison that facilitates new insights and clinically relevant analysis of pathological regions. Our results demonstrate improvements over conventional end-to-end learning approaches, while maintaining an interpretable feature selection process, is a key advantage of our method.

5 Acknowledgement

This work was supported by the International Alliance for Cancer Early Detection, an alliance between Cancer Research UK [EDDAPA-2024/100014] & [C73666/A31378], Canary Center at Stanford University, the University of Cambridge, OHSU Knight Cancer Institute, University College London and the University of Manchester; and the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

6 Disclosure of Interests

The authors declare that they have no conflicts of interest related to this work.

References

1. Bien, N., Rajpurkar, P., Ball, R.L., Irvin, J., Park, A., Jones, E., Bereket, M., Patel, B.N., Yeom, K.W., Shpanskaya, K., et al.: Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of mrnet. *PLoS medicine* **15**(11), e1002699 (2018)
2. Colussi, M., Ahmetovic, D., Civitarese, G., Bettini, C., Solyman, A., Gualtierotti, R., Peyvandi, F., Mascetti, S.: Loris-weakly-supervised anomaly detection for ultrasound images. In: *International Workshop on Advances in Simplifying Medical Ultrasound*. pp. 198–208. Springer (2024)
3. van Eck, C.F., van den Bekerom, M.P., Fu, F.H., Poolman, R.W., Kerkhoffs, G.M.: Methods to diagnose acute anterior cruciate ligament rupture: a meta-analysis of physical examinations with and without anaesthesia. *Knee Surgery, Sports Traumatology, Arthroscopy* **21**(8), 1895–1903 (2013)
4. Fan, F., Liu, H., Dai, X., Liu, G., Liu, J., Deng, X., Peng, Z., Wang, C., Zhang, K., Chen, H., et al.: Automated bone age assessment from knee joint by integrating deep learning and mri-based radiomics. *International Journal of Legal Medicine* **138**(3), 927–938 (2024)
5. Fu, Y., Li, Y., Saeed, S.U., Clarkson, M.J., Hu, Y.: A recycling training strategy for medical image segmentation with diffusion denoising models. *arXiv preprint arXiv:2308.16355* (2023)

6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in neural information processing systems* **27** (2014)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
8. Hegedus, E.J., Cook, C., Hasselblad, V., Goode, A., Mccrory, D.C.: Physical examination tests for assessing a torn meniscus in the knee: a systematic review with meta-analysis. *journal of orthopaedic & sports physical therapy* **37**(9), 541–550 (2007)
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840–6851 (2020)
10. Jung, C.G.: *Two essays on analytical psychology*. Routledge (2014)
11. Kather, J.N., Ghaffari Laleh, N., Foersch, S., Truhn, D.: Medical domain knowledge in domain-agnostic generative ai. *NPJ digital medicine* **5**(1), 90 (2022)
12. Liu, C., Yu, X., Wang, D., Jiang, T.: Aclnet: A deep learning model for acl rupture classification combined with bone morphology. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 57–67. Springer (2024)
13. Manna, S., Bhattacharya, S., Pal, U.: Self-supervised representation learning for knee injury diagnosis from magnetic resonance data. *IEEE Transactions on Artificial Intelligence* (2023)
14. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., et al.: Fast free-form deformation using graphics processing units. *Comput Methods Programs Biomed* **98**(3), 278–284 (2010)
15. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: Meila, M., Zhang, T. (eds.) *Proceedings of the 38th International Conference on Machine Learning*. *Proceedings of Machine Learning Research*, vol. 139, pp. 8162–8171. PMLR (18–24 Jul 2021), <https://proceedings.mlr.press/v139/nichol21a.html>
16. Noble, J., Hamblen, D.L.: The pathology of the degenerate meniscus lesion. *The Journal of Bone & Joint Surgery British Volume* **57**(2), 180–186 (1975)
17. Petersen, W., Zantop, T.: Anatomy of the anterior cruciate ligament with regard to its two bundles. *Clinical Orthopaedics and Related Research* (1976-2007) **454**, 35–47 (2007)
18. Saeed, S.U., Syer, T., Yan, W., Yang, Q., Emberton, M., Punwani, S., Clarkson, M.J., Barratt, D., Hu, Y.: Bi-parametric prostate mr image synthesis using pathology and sequence-conditioned stable diffusion. In: Oguz, I., Noble, J., Li, X., Styner, M., Baumgartner, C., Rusu, M., Heinmann, T., Kontos, D., Landman, B., Dawant, B. (eds.) *Medical Imaging with Deep Learning*. *Proceedings of Machine Learning Research*, vol. 227, pp. 814–828. PMLR (10–12 Jul 2024), <https://proceedings.mlr.press/v227/saeed24a.html>
19. Salzler, M., Nwachukwu, B.U., Rosas, S., Nguyen, C., Law, T.Y., Eberle, T., McCormick, F.: State-of-the-art anterior cruciate ligament tears: a primer for primary care physicians. *The Physician And Sportsmedicine* **43**(2), 169–177 (2015)
20. Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M.M., Greenspan, H., Klang, E.: Convolutional neural networks for radiologic images: a radiologist’s guide. *Radiology* **290**(3), 590–606 (2019)
21. Tsai, C.H., Kiryati, N., Konen, E., Eshed, I., Mayer, A.: Knee injury detection using mri with efficiently-layered network (elnet). In: *Medical Imaging with Deep Learning*. pp. 784–794. PMLR (2020)

22. Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. *Cancer research* **77**(21), e104–e107 (2017)
23. Zhang, W., Guo, Y., Jin, Q.: Radiomics and its feature selection: A review. *Symmetry* **15**(10), 1834 (2023)