

어텐션 기반 다중 인스턴스 학습을 이용한 자기공명영상에서 크론병 환자의 염증 활성도 분류

황서현^{1†} 이한상^{1†} 서니은² 임준석² 홍헬렌^{1*}

¹서울여자대학교 소프트웨어융합학과

²연세대학교 의과대학 영상의학교실

[†]공동 제 1 저자

*교신저자

hlhong@swu.ac.kr

Attention-based Multiple Instance Learning for Classification of Inflammatory Activity in Crohn's Disease Patients from MR Enterography

Seohyun Hwang^{1†} Hansang Lee^{1†} Nieun Seo² JoonSeok Lim² Helen Hong^{1*}

Department of Software Convergence, Seoul Women's University

Department of Radiology, Severance Hospital, Yonsei University College of Medicine

요 약

본 연구는 MRE 영상으로부터 크론병 환자의 염증 활성도를 자동으로 평가하기 위해, 파운데이션 모델 DINOv2와 어텐션 기반 다중 인스턴스 학습을 결합한 딥러닝 기법을 제안한다. 제안 기법은 개별 장염증의 MRE 영상 패치를 다중 인스턴스로 처리하고, DINOv2 모델을 특징 추출기로 활용하고 Top-K 어텐션 풀링 전략을 적용하여 핵심 병변에 집중하는 과정을 통해 환자별 sMaRIA 점수를 산출한다. 208명의 환자 데이터에 대한 검증 결과, DINOv2-Small과 어텐션 기반 풀링 및 Top-K 인스턴스 선택(K=3)을 결합한 모델이 77.78%의 가장 우수한 분류 정확도를 달성하였으며, 시각화를 통해 모델이 임상적으로 유의미한 병변에 집중함을 확인하였다.

Abstract

This study proposes a deep learning framework that combines a foundation model, DINOv2, with an attention-based multiple instance learning (MIL) strategy to automatically assess inflammatory activity in Crohn's disease from magnetic resonance enterography (MRE) images. The proposed method treats lesion-level MRE patches as multiple instances and employs DINOv2 as a feature extractor, followed by a Top-K attention pooling strategy to focus on clinically relevant lesions and derive patient-level sMaRIA scores. Experimental validation on a dataset of 208 patients demonstrated that the model integrating DINOv2-Small with attention-based pooling and Top-K instance selection (K = 3) achieved the highest classification accuracy of 77.78%. Qualitative results further confirmed that the model effectively attends to clinically meaningful inflammatory regions.

키워드: 딥러닝, 분류, 다중 인스턴스 학습, 어텐션, 염증성 장 질환

Keywords: Deep Learning, Classification, Multiple Instance Learning, Attention, Inflammatory bowel disease

1. 서론

크론병(Crohn's disease)은 소화관 전 구역에서 발생하는 만성 재발 염증성 장 질환으로, 효과적인 치료 전략 수립과 장기적인

*corresponding author: Helen Hong / Department of Software Convergence, Seoul Women's University (hlhong@swu.ac.kr)

예후 예측을 위해 염증 활성도를 정량적으로 평가하는 것이 필수적이다[1]. 자기공명 장조영술(Magnetic Resonance Enterography, MRE)은 염증성 장질환의 진단 및 치료 반응 모니터링을 위한 핵심 영상 기법으로 사용되고 있으며[2], simplified Magnetic Resonance Index of Activity (sMaRIA)와 같은 정량적 점수 체계를 사용해 염증 활성도를 정량적으로 평가하고 있다[3]. 그러나 MRE 기반 활성도 평가는 고도의 전문 지식을 요구하며, sMaRIA 점수가 장 영역별 활성도를 개별 분석하여 합산하는 구조를 취하고 있어 판독자 간 편차가 크고 오랜 시간이 소모되는 한계를 지닌다[4]. 따라서, 환자 단위 질병 염증 활성도를 보다 일관성 있고 효율적으로 평가하기 위해 자동화된 딥러닝 기반 접근 방식의 개발이 요구된다.

최근 딥러닝을 활용하여 크론병의 염증 활성도를 평가하려는 연구들이 다수 보고되었다. Lamash 등[5]은 소아 크론병 환자 23명의 MRE 영상에서 장 분절을 분할하고 상대적 조영 증강 값 기반의 CNN을 활용하여 활성 병변을 분류하였으며, Van Harten 등[6]은 3D cine MRI에서 CNN 기반의 중심선 분할 및 추적 알고리즘을 통해 장 운동성을 정량화하여 중증 염증을 구분하는 모델을 제안하였다. 또한 Holland 등[7]은 CNN과 어텐션 메커니즘을 결합하여 염증성 장질환 진단 성능을 향상시켰다. 그러나 이러한 선행 연구들은 주로 전반적인 장벽 조영 상태, 장 운동성, 영상 품질 등 전역적인 특성에 기반한 분석에 집중되어 있으며, 환자 내에 산재한 개별 병변들의 기여도를 구조적으로 모델링하지 못하는 공통적인 한계를 가진다.

본 논문에서는 MRE 영상에서 염증성 장질환의 염증 활성도 분류를 위한 어텐션 기반 다중 인스턴스 학습 기법을 제안한다. 제안 방법은 각 염증 병변 패치를 인스턴스(instance)로, 환자 1인의 영상 집합을 백(bag)으로 정의하여 다발성 국소 병변의 미세한 정보를 환자 단위의 중증도 평가로 통합한다. 특히 대규모 시각 데이터로 사전 학습된 파운데이션 모델인 DINOv2[8]를 특징 추출기로 활용하고, 어텐션 기반

풀링(Attention Based Pooling)[9]을 통해 활성도 결정에 기여도가 높은 핵심 병변에 강한 가중치를 부여하도록 설계하였다. 또한 Top-K 인스턴스 선택 전략을 도입하여 진단과 무관한 노이즈 인스턴스의 영향을 배제함으로써 분류 성능과 임상적 신뢰성을 동시에 확보하였다. 208명의 환자 데이터에 대한 실험을 통해 제안 방법의 어텐션 기반 풀링 및 Top-K 인스턴스 선택이 불필요한 염증 인스턴스의 간섭을 최소화하고 주요 병변 정보에 집중하여 크론병 염증 활성도 평가를 개선하는 것을 검증하였다.

2. 제안 방법

Figure 1은 제안 방법의 전체적인 구조를 나타낸다. 제안 방법은 (1) DINOv2 기반 인스턴스 임베딩과 (2) Top-K 인스턴스 선택 및 어텐션 기반 풀링과 분류로 구성된다. 첫번째 단계인 DINOv2 기반 인스턴스 임베딩에서는 각 환자로부터 사전 분할된 다수의 장염증 병변 인스턴스에 대해 대규모 파운데이션 모델인 DINOv2를 이용하여 고차원 임베딩 벡터를 추출한다. 두번째 단계인 Top-K 인스턴스 선택 및 어텐션 기반 풀링과 분류에서는 어텐션 네트워크를 통해 인스턴스별 중요도를 학습하고 Top-K 인스턴스 선택을 통해 활성도 결정에 핵심적인 병변들만을 선별하여 환자 수준 단일 임베딩을 제작하고 분류기를 학습한다.

2.1 DINOv2 기반 인스턴스 임베딩

MRE 영상은 장벽 두께 증가와 같은 다양한 염증 징후를 포함하며, 이러한 병변 특징은 비교적 국소적으로 나타나는 경향이 있다. 따라서 환자 전체 영상을 바로 활용하는 전역적 표현 방식보다는 개별 염증 인스턴스 단위에서 의미 있는 국소적 시각 표현 추출을 통해 풍부하고 구분력 있는 특징 표현을 확보하는 것이 보다 효과적이다. 이를 위해 제안 방법에서는 파운데이션 모델인 DINOv2[8]를 특징 추출기로

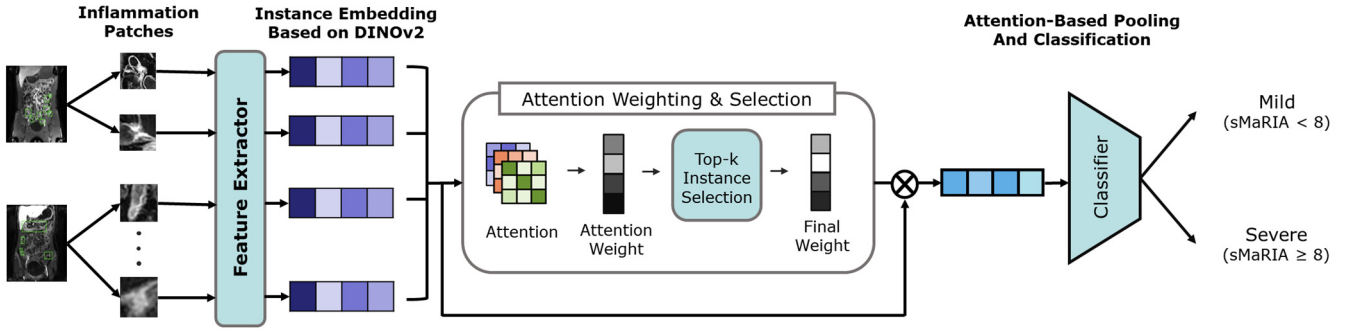


Figure 1: Architecture of the proposed Attention-based MIL method.

활용하여 염증 인스턴스로부터 이미지 임베딩을 추출한다. DINOv2는 1억 4천만장의 대규모 자연 이미지 데이터셋인 LVD-142M을 기반으로 자기지도학습을 통해 사전 학습된 모델이다. DINOv2는 전역적 이미지 수준의 학습과 Masked Image Modeling 목적 함수를 결합함으로써, 이미지의 전체적인 맥락과 세부적인 국소 정보를 동시에 학습한다. 이를 통해 별도의 미세 조정 과정 없이도 강력한 이미지 특징 추출 성능을 발휘하며, 이미지 분류 및 검색, 분할 등 다양한 분야에서 범용적인 모델로 활용되고 있다.

제안 방법에서 임베딩을 위한 병변 인스턴스는 2차원 MRE 슬라이스 상에서 영상의학 전문가가 수동 분할한 각 염증 병변에 대해 해당 병변이 가장 크게 관찰되는 단일 슬라이스를 선택하고, 분할된 병변 영역을 완전히 포함하는 최소 외접 정사각형을 추출하여 하나의 이미지 패치로 정의한다. 이렇게 생성된 병변 패치는 크기와 밝기 값 정규화 과정을 거친 뒤 DINOv2 인코더를 통해 고정 차원의 임베딩 벡터로 매핑된다. 이렇게 추출된 인스턴스 임베딩은 DINOv2의 강력한 일반화 성능 덕분에 별도의 의료 도메인 미세 조정 없이도 MRE 영상 내의 복잡하고 미세한 병리적 특징을 효과적으로 추출할 수 있으며, 이는 주어진 MRE 데이터에서 개별적인 장염증 병변에 대한 레이블링 없이도 강인하고 효과적인 임베딩 추출을 가능하게 한다.

2.2 Top-K 인스턴스 선택 및 어텐션 기반 풀링과 분류

MRE 영상 데이터는 환자별로 다수의 인스턴스가 포함된 구조를 가지므로, 환자 단위의 진단을 위해서는 개별 인스턴스의 특징 벡터들을 단일한 시각적 표현으로 통합하는

풀링 과정이 필요하다. 그러나 기존의 평균 풀링(Average Pooling) 또는 최댓값 풀링(Max Pooling) 방식은 모든 인스턴스를 동일한 비중으로 처리함으로써 각 인스턴스 별로 상이한 임상적 중요도를 반영하지 못한다는 한계가 있다. 따라서 인스턴스 풀링 시 활성도가 높은 염증 병변의 기여도를 극대화하고자, 어텐션 기반 다중 인스턴스 학습과 Top-K 인스턴스 선택을 결합한 방법론을 제안한다.

먼저, 각 병변 인스턴스의 중요도를 학습적으로 추정하기 위해 다중 퍼셉트론 기반의 어텐션 네트워크를 구성한다. 어텐션 네트워크는 인스턴스 임베딩 벡터 h_i 를 입력받아 선형 계층과 비선형 활성화 함수를 거쳐 최종적인 어텐션 점수(Attention score)를 아래와 같이 출력한다.

$$a'_i = \mathbf{w}^T \cdot \tanh(V_3 \cdot \tanh(V_2 \cdot \tanh(V_1 \cdot h_i^T))) \quad (1)$$

$$a_i = \frac{\exp(a'_i)}{\sum_{j=1}^N \exp(a'_j)} \quad (2)$$

이때 h_i 는 입력 차원 D 를 갖는 인스턴스 임베딩 벡터이며, V_1, V_2, V_3 는 각각 첫 번째, 두 번째, 세 번째 선형 계층의 가중치 행렬을 의미한다. \mathbf{w}^T 는 최종 출력 계층의 가중치 행렬이며 a_i 는 최종 계산된 어텐션 점수를 말한다. 이 어텐션 점수는 해당 병변이 환자 전체 염증 활성도 예측에 기여하는 상대적 중요도를 의미하며, 분류 손실을 최소화하는 방향으로 end-to-end로 학습된다.

Table 1. Dataset statistics for patients and instances

	Label	Train	Validation	Test
Patients	Class 2	53	26	26
	Class 3	37	19	19
Lesions	Class 2	141	54	93
	Class 3	182	66	85

이후, 산출된 어텐션 점수를 바탕으로 노이즈 인스턴스의 영향을 배제하고 결정적인 증거만을 활용하기 위해 Top-K 인스턴스 선택을 적용한다. 제안 방법에서는 K를 3으로 설정하여, 어텐션 점수가 가장 높은 상위 3개의 인스턴스만을 선별하고 이를 어텐션 기반 풀링(Attention Based Pooling) 방식으로 통합하여 384 차원의 환자 단위 벡터를 생성한다. 이 전략은 임상적으로 가장 의심되는 소수 병변을 중심으로 환자의 상태를 판단하는 실제 판독 과정과 구조적으로 대응되며, 배경 구조나 정상 장벽과 같은 노이즈 인스턴스의 영향을 효과적으로 차단한다. 또한 결과적으로 불필요한 정보나 노이즈를 효과적으로 차단하여 활성화 예측 성능을 향상시키는 효과를 갖는다.

마지막으로, 풀링된 환자 단위 벡터는 완전 연결 계층에 입력되어 sMaRIA 기준의 활성화 등급을 최종적으로 분류한다. 제안 방법에서는 클래스 불균형 문제에 강인한 분류 성능을 확보하기 위해, Focal Loss를 학습 시 손실 함수로 사용하였다. Focal Loss는 Binary Cross Entropy(BCE) Loss를 기반으로 쉽게 분류되는 샘플의 손실 기여도를 감소시키고, 어려운 샘플에 더 집중하도록 조절항을 추가한 손실 함수이다. BCE Loss는 식 (3)과 같이 정의되며, 이를 기반으로 Focal Loss는 식 (4)와 같이 정의된다.

$$BCE(p_t) = -\log(p_t) \quad (3)$$

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (4)$$

여기서 p_t 는 모델이 실제 정답 레이블에 대해 부여한 예측 확률을 의미한다. α_t 는 클래스 간 빈도 차이에 따른 불균형을 조절하는 가중치 파라미터이며, γ 는 조절 인자로서 분류가 까다로운 샘플에 대한 학습 강도를 높이는 역할을 수행한다.

3. 실험 및 결과

3.1 데이터셋

본 연구에서는 연세대학교 세브란스병원에서 수집된 크론병 환자 208명의 자기공명 장조영술(MRE) 영상을 사용하여 제안 방법의 성능을 검증하였다. 모든 영상 데이터는 숙련된 영상의학 전문의에 의해 염증 병변 영역이 수동으로 분할되었으며, 분할된 병변 영역은 대표 슬라이스의 병변을 포함하는 정방 패치 형태로 변환되어 병변 인스턴스로 정의되었다. 각 환자는 병변의 개수와 분포에 따라 서로 다른 수의 인스턴스를 포함하며, 이러한 다중 병변 구조를 그대로 유지한 상태로 다중 인스턴스 학습 모델의 입력으로 구성하였다.

환자 단위의 레이블은 sMaRIA 점수를 기준으로 정의하였다. 구체적으로 sMaRIA 점수가 8 미만인 경우를 경증(Mild), 8 이상인 경우를 중증(Severe)으로 설정하여 이진 분류 문제로 구성하였으며, sMaRIA 점수가 0으로 염증 소견이 없는 정상 환자는 분석 대상에서 제외하였다. 이에 따라 최종적으로 180명의 환자 데이터가 실험에 사용되었으며, 해당 데이터는 환자 단위로 중복이 발생하지 않도록 무작위 분할하여 훈련, 검증 및 테스트 데이터셋을 2:1:1 비율로 구성하였다. 데이터셋의 환자별 및 인스턴스별 통계는 Table 1에 요약되어 있다.

3.2 실험 환경

모든 실험은 RunPod의 NVIDIA RTX A4500 GPU를 이용해 클라우드 기반 서버 환경에서 Python 3.10 및 PyTorch (2.8.0) 라이브러리를 기반으로 수행되었다. 입력 MRE 영상 인스턴스는 224x224 픽셀로 통일되었으며, 모델의 강건성 확보를 위해 학습 데이터에만 무작위 회전, 이동, 크기 조절 등의 데이터 증강을 적용한 후 ImageNet 평균 및 표준편차를 사용하여 정규화를 진행하였다. 모델 최적화에는 AdamW 옵티마이저와 $8e^{-4}$ 의 weight decay를 사용했으며, 학습률은 ReduceLRonPlateau 스케줄러로 관리하였다

Table 2 : Performance comparison according to Top-K selection and pooling methods.

K	Pooling	Accuracy	Sensitivity	Specificity	AUC
---	---------	----------	-------------	-------------	-----

All Instance	Max Pooling	60.00	68.42	53.85	0.6680
	Avg. Pooling	40.00	31.58	46.15	0.4332
	Attention	73.33	89.47	61.54	0.7712
2	Max Pooling	73.33	84.21	65.38	0.7429
	Avg. Pooling	75.56	89.47	65.38	0.7955
	Attention	75.55	68.42	80.77	0.7530
3	Max Pooling	77.78	84.21	73.07	0.7632
	Avg. Pooling	64.44	78.94	53.85	0.6883
	Attention	77.78	94.74	65.38	0.7551
4	Max Pooling	68.89	78.95	61.54	0.7632
	Avg. Pooling	64.44	84.21	50.00	0.6903
	Attention	66.67	78.95	57.69	0.7409
5	Max Pooling	71.11	84.21	61.53	0.7328
	Avg. Pooling	64.44	63.16	65.38	0.7166
	Attention	66.67	94.74	46.15	0.7490

모델의 성능은 테스트 데이터셋에 대해 정확도(Accuracy), 민감도(Sensitivity), 특이도(Specificity), 그리고 AUC를 계산하여 평가하였다. 제안하는 방법론의 유효성을 다각도로 검증하기 위해, 각 구성 요소의 기여도를 확인하는 비교 실험을 설계하였다. 먼저, 인스턴스 통합 방식의 효과를 검증하기 위해 기존의 최댓값 풀링(Max Pooling), 평균 풀링(Average Pooling) 방식과 제안하는 어텐션 기반 풀링(Attention Based Pooling) 방식의 성능을 비교 분석하였다. 또한, 불필요한 인스턴스를 배제하는 것이 성능 향상에 미치는 영향을 확인하고자, 모든 인스턴스를 사용하는 경우와 Top-K 인스턴스 선택을 통해 상위 K개의 인스턴스만을 선별하여 사용하는 경우를 대조 실험하였다. 또한 최적의 K 값을 찾기 위해 K 값을 2에서 5까지 변화시키며 비교 실험을 수행하였다. 이어 MRE 영상에 최적화된 특징 표현을 탐색하기 위해, 지도학습 기반의 ResNet[10] 계열(ImageNet 및 RadImageNet[11] 사전학습) 모델과 자기지도학습 기반의 DINOv2 모델을 비교 평가하였다. 마지막으로, 제안 모델이 실제로 활성도가 높은 염증 병변에 높은 가중치를 부여하는지 검증하고자, 최고 성능을 달성한

조합을 대상으로 어텐션 시각화 분석을 수행하였다. 이를 통해 모델이 예측 과정에서 어떤 인스턴스를 핵심 정보로 활용하는지 정성적으로 분석하였다.

3.3 실험 결과

Table 2는 제안 방법의 핵심 기여점인 어텐션 기반 풀링(Attention Based Pooling)과 Top-K 인스턴스 선택의 유효성을 검증하기 위해, DINOv2-Small 임베딩을 기반으로 다양한 풀링 방식과 K 값의 변화에 따른 성능을 비교 분석한 표이다. 먼저 인스턴스 선택을 적용하지 않은 경우, 어텐션 기반 풀링(Attention Based Pooling)이 정확도 73.33%, AUC 0.7712를 기록하며 최댓값 풀링(Max Pooling)이나 평균 풀링(Average Pooling)에 비해 일관되게 높은 성능을 보였다. 이러한 경향이 Top-K 인스턴스 선택을 적용했을 때에도 일관적으로 나타나는 것을 통해, 단순한 통계적 대푯값을 취하는 것보다 각 인스턴스의 중요도를 학습하여 가중치를 부여하는 방식이 환자의 활성도 판별에 더 효과적임을 알 수 있다.

Table 3 : Comparison of embedding models.

Embeddings	Accuracy	Sensitivity	Specificity	AUC
ResNet18 (ImageNet)	66.67	63.18	69.23	0.7065
ResNet50 (RadImageNet)	71.11	63.15	76.92	0.7065
DINOv2-S (LVD-142M)	73.33	89.47	61.54	0.7712
DINOv2-B (LVD-142M)	73.33	100	53.84	0.7287

나아가 Top-K 인스턴스 선택 전략의 효과를 분석하기 위해, 서로 다른 K 값에 따른 성능 변화를 비교하였다. 그 결과, K=3일 때 어텐션 기반 풀링(Attention Based Pooling)과 결합한 모델이 정확도 77.78%, 민감도 94.74%를 기록하며 전체 실험 조건 중 가장 높은 성능을 달성하였다. 이는 환자에 포함된 모든 병변을 사용하는 경우보다, 임상적으로 가장 중요도가 높은 소수의 병변만을 선택적으로 활용하는 것이 오히려 더 정확한 진단으로 이어질 수 있음을 의미한다. 반면 K 값을 4 이상으로 증가시킬 경우 정확도와 AUC가 오히려 감소하는 경향이 나타났는데, 이는 상대적으로 중요도가 낮은 병변들이 추가로 포함되면서 풀링 과정에 노이즈가 유입되었기 때문으로 해석된다. 이러한 결과는 크론병과 같이 병변이 다수 존재하는 질환에서, 모든 병변을 균등하게 활용하는 접근보다 핵심 병변 중심의 선택적 풀링이 환자 단위 중증도 분류에 구조적으로 더 적합한 것으로 분석된다.

특징 추출기의 영향을 분석하기 위해 다양한 사전학습 모델을 Table 3와 같이 비교하였다. 실험에서는 자가지도 학습 기반의 DINOv2-Small 모델이 정확도 73.33%, AUC 0.7712로 가장 우수한 성능을 보였다. 반면 ImageNet으로 사전학습된 ResNet18과 RadImageNet으로 사전학습된 ResNet-50 모두 DINOv2-Small 대비 낮은 성능을 기록하였다. 이는 의료 영상 도메인에 특화된 사전학습 모델보다도, 대규모 일반 이미지 데이터에서 자가지도 학습으로 학습된 파운데이션 모델이 병변 표현 학습에 더 효과적일 수 있음을 의미한다. 또한 DINOv2-Base와 DINOv2-Large와 같이 파라미터 수가 더 많은 모델을 사용한 경우, 오히려 성능이 정체되거나 감소하는 현상이 관찰되었는데, 이는 제한된 데이터 규모에서 과도한 모델 복잡도가 일반화 성능을 저해할 수 있는 것으로 분석된다.

Figure 2는 제안 모델이 실제로 어떤 병변에 주의를 집중하는지를 검증하기 위해 각 환자 케이스에서 어텐션 가중치 순위에 따라 해당하는 인스턴스를 바운딩 박스로 시각화한 결과이다. 각 행은 서로 다른 환자 케이스를 나타내며 열은 모델이 산출한 어텐션 가중치 순위에 따라 가장 높은

점수(max score)부터 가장 낮은 점수(min score)까지의 인스턴스를 보여준다. (a)열은 모델이 가장 높은 가중치를 부여한 1순위 인스턴스로 환자 진단에 가장 결정적인 영역을, (b)열과 (c)열은 각각 2순위, 3순위 인스턴스로 1순위 영역을 보완하는 부가적 병변 근거를, (d)열은 최하위 인스턴스로 진단 기여도가 가장 낮은 영역을 나타낸다. (a)열의 인스턴스들은 장벽 두께 등 염증 활성화도 판단과 관련된 병변 영역을 포착하였으며, (b)열 및 (c)열의 인스턴스들 역시 유의미한 염증 소견을 포함하고 있음을 확인하였다. 반면, (d)열의 인스턴스들은 정상 구조나 배경 등 질병 활성화도와 상관관계가 낮은 영역들을 포함하는 것을 볼 수 있다. 이는 제안 방법의 어텐션 점수 산정 과정이 실제 임상적으로 중요한 병변에 선택적으로 주의를 할당함으로써 염증 활성화도 분류에서 높은 성능뿐 아니라 신뢰 가능한 진단을 제공하고 있음을 의미한다.

4. 결론

본 연구는 자기공명 장조영술(MRE) 영상을 기반으로 크론병 환자의 염증 활성도를 자동으로 분류하기 위한 딥러닝 기법을 제안하였다. 제안 방법은 다중 인스턴스 학습 구조를 통해 개별 병변 인스턴스의 중요도를 학습하고, 어텐션 기반 풀링(Attention Based Pooling)과 Top-K 인스턴스 선택 전략을 결합함으로써 환자 단위 진단에 결정적인 핵심 병변에 선택적으로 집중하도록 설계하였다. 실험 결과, 제안 방법은 임상적으로 의미 있는 병변에 높은 가중치를 부여하고 노이즈 인스턴스를 효과적으로 제거함으로써 크론병 염증 활성화도 예측 성능을 안정적으로 향상시켰음을 확인하였다.

향후 연구에서는 교차 검증을 적용하여 제한된 데이터 규모에서의 통계적 유의성을 보완하고, BiomedCLIP[12] 등 최신 의료 파운데이션 모델과의 비교를 통해 MRE 염증 표현에 보다 적합한 임베딩을 탐색하고자 한다. 또한 장 분절별 위치 정보를 반영한 계층적 집계 전략을 도입하여 공간 정보를 모델링에 통합하는 방향을 모색하고자 한다.

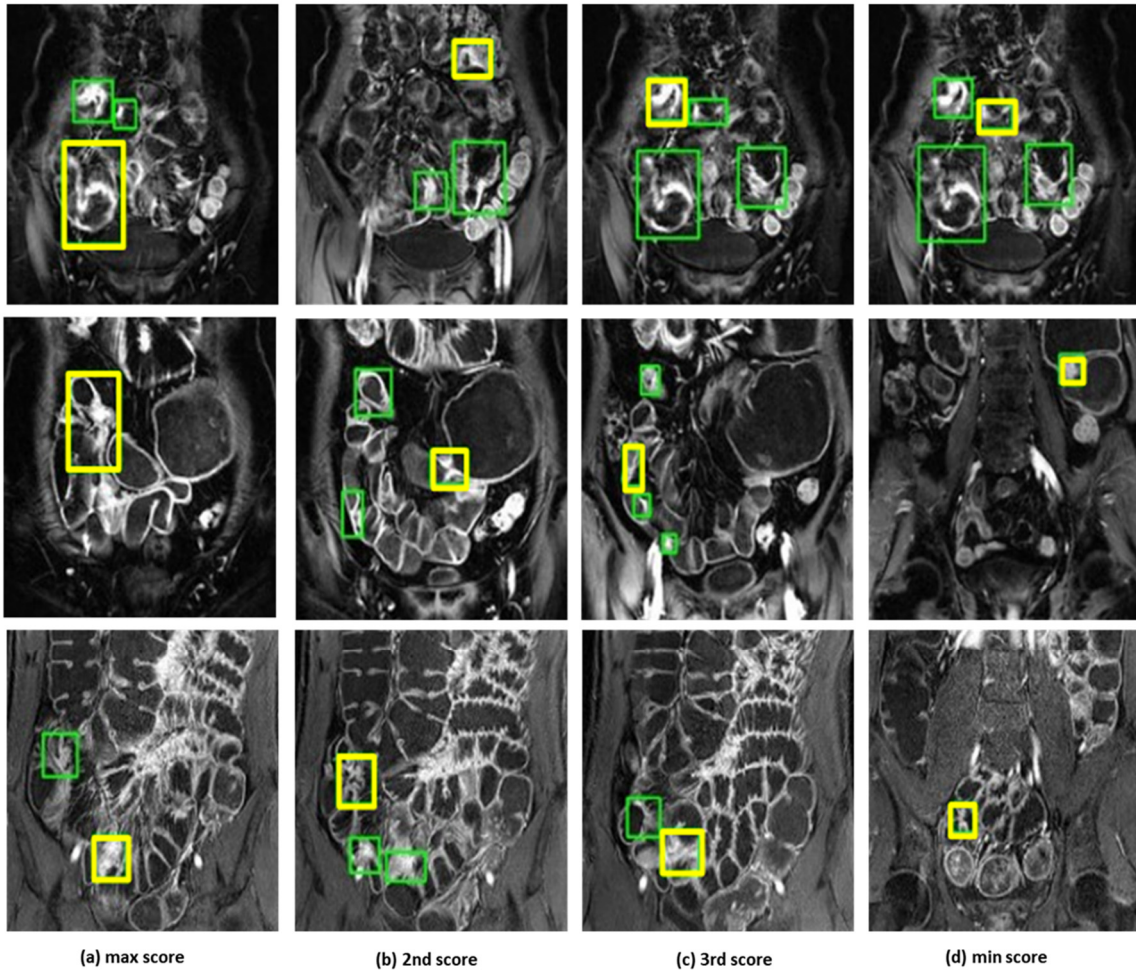


Figure 2 : Visualization of patches ranked by attention score

감사의 글

본 연구는 과학기술정보통신부 재원으로 한국연구재단의 지원(No.RS-2024-00336063), 보건복지부 재원으로 한국보건산업진흥원의 보건의료기술연구개발사업 지원(HI22C1496), 그리고 서울여자대학교 학술연구비(2026)의 지원을 받아 수행되었음.

References

[1] D. C. Baumgart and W. J. Sandborn, "Crohn's disease," *Lancet*, pp. 1590-1605, 2012.

[2] J. Panes, Y. Bouhnik, W. Reinisch, J. Stoker, S. A. Taylor, D. C. Baumgart, S. Danese, S. Halligan, B. Marincek, C. Matos, L. Peyrin-Biroulet, J. Rimola, G. Rogler, G. van Assche, S. Ardizzone, A. Ba-Ssalamah, M. A. Bali, D. Bellini, L. Biancone, F. Castiglione, R. Ehehalt, R. Grassi, T. Kucharzik, F. Maccioni, G. Maconi, F. Magro, J. Martín-Comin, G. Morana, D. Pendsé, S. Sebastian, A.

Signore, D. Tolan, J. A. Tielbeek, D. Weishaupt, B. Wiarda, and A. Laghi, "Imaging techniques for assessment of inflammatory bowel disease: Joint ECCO and ESGAR evidence-based consensus guidelines," *J. Crohn's Colitis*, vol. 7, no. 7, pp. 556-585, 2013..

[3] J. Roseira, A. R. Ventosa, H. Tavares de Sousa, and J. Brito, "The new simplified MARIA score applies beyond clinical trials: A suitable clinical practice tool for Crohn's disease that parallels a simple endoscopic index and fecal calprotectin," *United European Gastroenterol. J.*, vol. 8, no. 10, pp. 1208-1216, 2020.

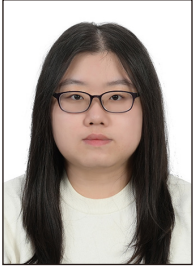
[4] S. Schleder, M. Pawlik, P. Wiggermann, C. Ott, S. Fichtner-Feigl, R. Müller-Wille, C. Stroszczyński, and A. G. Schreyer, "Interobserver agreement in MR enterography for diagnostic assessment in patients with Crohn's disease," *Rofo*, vol. 184, no. 10, pp. 992-997, 2013.

[5] Y. Lamash, S. Kurugol, M. Freiman, J. M. Perez-Rossello, M. J. Callahan, A. Bousvaros, and S. K. Warfield, "Curved planar reformatting and convolutional neural network-based segmentation of the small bowel for visualization and quantitative assessment of pediatric Crohn's disease from MRI," *J. Magn. Reson. Imaging*, vol. 49, no. 6, pp. 1565-1576, 2019.

[6] L. D. van Harten, C. S. de Jonge, K. J. Beek, J. Stoker, and I. Išgum, "Untangling and segmenting the small intestine in 3D cine-MRI

- using deep learning," *Med. Image Anal.*, vol. 78, Art. no. 102386, 2022.
- [7] R. Holland, U. Patel, P. Lung, E. Chotzoglou, and B. Kainz, "Automatic detection of bowel disease with residual networks," *Proc. PRIME-MICCAI Workshop*, pp. 151–159, 2019.
- [8] M. Oquab, T. Darcet, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jégou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, "DINOv2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.
- [9] M. Ilse, J. Tomczak, and M. Welling, "Attention-based Deep Multiple Instance Learning," *International Conference on Machine Learning (ICML)*, pp. 2127-2136, 2018.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016
- [11] X. Mei, Z. Liu, P. M. Robson, B. Marinelli, M. Huang, A. Doshi, A. Jacobi, C. Cao, K. E. Link, T. Yang, Y. Wang, H. Greenspan, T. Deyer, Z. A. Fayad, and Y. Yang, "RadImageNet: An open radiologic deep learning research dataset for effective transfer learning," *Radiology: Artificial Intelligence*, vol. 4, no. 5, Art. no. e210315, 2022.
- [12] S. Zhang, Y. Xu, N. Usuyama, H. Xu, J. Bagga, R. Tinn, S. Preston, R. Rao, M. Wei, N. Valluri, C. Wong, A. Tupini, Y. Wang, M. Mazzola, S. Shukla, L. Liden, J. Gao, A. Crabtree, B. Piening, C. Bifulco, M. P. Lungren, T. Naumann, S. Wang, and H. Poon, "A multimodal biomedical foundation model trained from fifteen million image–text pairs," *NEJM AI*, vol. 2, no. 1, Art. no. AIoa2400640, 2025.

〈저자소개〉



황서현

- 2026년 2월 서울여자대학교 화학생명환경과학부 생명환경공학전공 졸업(학사)
- 2026년 3월~현재 서울여자대학교 컴퓨터학과 석사과정
- 관심분야 : 인공지능, 딥러닝, 의료영상처리 및 분석
- <https://orcid.org/0009-0008-5524-6880>



임준석

- 1994년 02월 연세대학교 의과대학 의학과 졸업(학사)
- 1998년 02월 연세대학교 의과대학 진단방사선과 졸업(석사)
- 2007년 02월 연세대학교 의과대학 진단방사선과 졸업(박사)
- 2005년 03월~현재 연세대학교 의과대학 영상의학교실 교수
- 관심분야 : 복부영상의학, 디지털헬스, 인공지능
- <https://orcid.org/0000-0002-0334-5042>



이한상

- 2011년 2월 한국과학기술원 전기 및 전자공학과 졸업(학사)
- 2013년 2월 한국과학기술원 전기 및 전자공학과 졸업(석사)
- 2019년 2월 한국과학기술원 전기 및 전자공학과 졸업(박사)
- 2019년 3월~2025년 2월 한국과학기술원 정보전자연구소 연수연구원
- 2025년 3월~현재 서울여자대학교 미래산업융합연구소 연구교수
- 관심분야 : 인공지능, 딥러닝, 컴퓨터 비전, 의료영상처리 및 분석
- <https://orcid.org/0000-0003-0917-1589>



홍헬렌

- 1994년 2월 이화여자대학교 전자계산학과 졸업(학사)
- 1996년 2월 이화여자대학교 전자계산학과 졸업(석사)
- 2001년 8월 이화여자대학교 컴퓨터학과 졸업(박사)
- 2001년 9월~2003년 7월 서울대학교 컴퓨터공학부 BK 조교수
- 2006년 3월~현재 서울여자대학교 소프트웨어융합학과 교수
- 관심분야 : 인공지능, 딥러닝, 의료영상처리 및 분석
- <https://orcid.org/0000-0001-5044-7909>



서니은

- 2008년 02월 울산대학교 의학과 졸업(학사)
- 2012년 02월 울산대학교 의학과 졸업(석사)
- 2016년 02월 울산대학교 의학과 졸업(박사)
- 2015년 03월~현재 연세대학교 의과대학 영상의학과 교수
- 관심분야 : 복부영상의학, 위장관질환, 자기공명영상, 영상분석
- <https://orcid.org/0000-0001-8745-6454>