

인공지능 기반 IBD Scoring System

IBD Scoring System based on AI

김재승¹, 김현영², 은성종^{3*}

Jae-Seoung Kim¹, Hyun-Young Kim², Sung-Jong Eun^{3*}

요약

염증성 장질환은 크론병과 궤양성 대장염으로 구성되어 있다. 최근 궤양성 대장염 환자의 내시경 수요가 급격하게 증가하고 있으며, 궤양성 대장염은 일생동안 호전과 악화를 반복하는 염증성 장질환이기 때문에 꾸준한 관리가 필요하다. 궤양성 대장염의 내시경적 질병 활성도를 평가하는 지표로 MES를 주로 사용한다. 본 논문에서는 정확도가 우수한 MES 평가 지표를 인공지능 기술을 통해 자동 분석하여 스코어링 해 주는 시스템을 제안한다. 스코어링 자동 분석을 위해 Faster R-CNN 알고리즘을 개선하여 적용하였다. 인공지능 알고리즘의 정확도를 평가하기 위해 MLP, SVM, CNN모델과 본 논문에서 제안하는 방법에 대한 정확도를 비교 평가 하였다. 비교 평가 결과 평균 88%의 정확도를 도출하여 성능을 증명하였다. 이러한 시스템은 임상 현장에서 내시경 시술에 보조적인 수단으로 활용될 수 있다. 내시경 영상을 활용한 본 연구의 성과물은 염증성 장질환 중 하나인 궤양성 대장염의 진료에서 인공지능을 활용하여 진료의 질 향상을 높이는 계기가 될 수 있다. 향후 연구로 다기관 데이터 수집을 통한 데이터의 확장을 통해 정확도를 개선하고자 한다.

핵심어 : 염증성 장질환, 궤양성 대장염, 내시경 점수, 점수 시스템, 영역 기반 합성곱 신경망

Abstract

Inflammatory bowel disease consists of Crohn's disease and ulcerative colitis. Recently, the demand for endoscopy from ulcerative colitis patients is rapidly increasing, and since ulcerative colitis is an inflammatory bowel disease that repeats improvement and worsening throughout life, continuous management is required. MES is mainly used as an index to evaluate the endoscopic disease activity of ulcerative colitis. In this paper, we propose a system that automatically analyzes and scores MES evaluation indicators with excellent accuracy through artificial intelligence technology. Faster R-CNN algorithm was improved and applied for automatic scoring analysis. In order to evaluate the accuracy of the artificial intelligence algorithm, the accuracy of the MLP, SVM, and CNN models and the method proposed in this paper were compared and evaluated. As a result of comparative evaluation, an average accuracy of 88% was derived to prove the performance. Such a system can be used as an auxiliary means for endoscopic procedures in the clinical field. The results of this study using endoscopic imaging can serve as an opportunity to improve

1 Health IT Research center, Gachon University Gil Medical Center, Incheon, Republic of Korea [Researcher]
e-mail: mpmkjs@gilhospital.com

2 Z&K co.ltd, Incheon, Republic of Korea [Researcher]
e-mail: znk1026@kita.net

3 Digital Health Industry Team, National IT Industry Promotion Agency, Jincheon, Korea [Researcher]
e-mail: asclephios@naver.com (Corresponding author)

* 본 연구는 보건복지부의 재원으로 한국보건정보통신원의 의료데이터 중심병원 지원사업 지원에 의하여 이루어진 것임

Received(December 6, 2021), Review Result(1st: December 22, 2021), Accepted(April 13, 2022), Published(April 30, 2022)



© 2022 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

the quality of care by using artificial intelligence in the treatment of ulcerative colitis, one of the inflammatory bowel diseases. Future research aims to improve accuracy through data expansion through multi-institutional data collection.

Keyword : Inflammatory bowel disease, ulcerative colitis, endoscopic score, scoring system, R-CNN

1. 서론

염증성 장질환(Inflammatory bowel disease, IBD)은 원인이 밝혀지지 않는 장의 만성적인 염증을 일으키는 질환을 말하며, 크론병(Crohn's disease, CD)과 궤양성 대장염(ulcerative colitis, UC)으로 구성 되어 있다. 특히 최근 국내에서도 염증성 장질환의 발생이 폭발적으로 증가되고 있다. 크론병 및 궤양성 대장염은 진단 및 치료를 위한 절대적인 특이 마커가 없고, 조직 진단 등이 염증성 장질환의 진단을 확진할 수 없기에, 염증성 장질환의 진단은 임상양상, 조직검사, 내시경검사, 혈액검사 등을 임상적으로 종합적으로 판단하여 진단을 내리게 되고, 이후의 치료 및 관리를 위해서도 많은 인적 물적 자원이 소요되며 적절한 관리를 하지 못하게 되는 경우가 많다.

현재 우리나라의 치료는 질병이 발병한 후에 환자를 치료하는 반응적인 의료 시스템이 주류를 이루고 있으나, 이는 엄청난 비용을 야기한다. 국내 경상의료비 지출 규모는 GDP 대비 6.9%로 OECD 회원국 평균(8.9%) 보다 낮으나, 2009년 대비 2013년 우리나라의 GDP에 대한 경상 의료비는 1.1% 증가하여 같은 기간 OECD 평균 증가 (0.6%) 보다 높은 증가폭을 기록하였다.

염증성 장질환 환자 역시도 이런 치료반응적인 의료 시스템으로 진료를 하고 있으나, 질환의 특성은 환자 개개인에게 적극적인 맞춤형 진료에 필요한 상황이며, 질병의 진단 및 치료가 복잡하고, 많은 진료 시간을 필요로 한다. 최근 전 세계적으로 염증성 장질환의 치료에 관한 연구의 폭발적인 증가 및 많은 치료제의 개발로 인하여 급성 응급실 내원 빈도가 상당부분 줄어들었음에도 이러한 이유로 적절한 시기에 적절한 치료를 받지 못하여 급성 악화를 반복하여 질환의 악화와 의료비의 상승을 가져오게 된다.

최근 궤양성 대장염 환자의 내시경 수요가 급격하게 증가하고 있다. 궤양성 대장염은 일생 동안 호전과 악화를 반복하는 만성 염증성 장질환으로, 임상적 특징은 혈성 설사 및 목통 등이 주증상이다 [1][2].

궤양성 대장염의 내시경적 질병 활성도를 평가하는 지표로 Mayo endoscopic score(MES), ulcerative colitis endoscopic index of severity (UCEIS) 등 다양한 정량 지표가 개발되고 검증되어 임상에서 활용되고 있다 [3]. 특히 MES는 궤양성 대장염 환자를 대상으로 진행되고 있는 다양한 생물학제제 및 소분자제제의 임상시험에서 내시경적 염증 활성도를 객관적으로 평가하고 약제의 내시경 관해 효과를 판정하는 데 있어 가장 많이 활용되고 있는 표준화된 지표이다.

본 연구에서는 정확도가 우수한 MES 평가 지표를 인공지능 기반 모델로 자동 분석하여 스코어

링 하여 임상에서 궤양성 대장염 환자의 내시경 시술에 인공지능을 보조적으로 적용할 수 있는 시스템을 제안한다.

2. 연구 배경

2.1 의료분야 인공지능 적용의 필요성

의료 빅데이터와 인공지능(Artificial intelligence)의 융합으로 의료 패러다임은 의료지능화, 정밀의료로 급속히 확대되고 있다. 방대한 의료데이터가 축적됨에 따라 의료데이터의 생성, 저장, 분석, 활용을 통해 새로운 미래형 의료산업으로 탈바꿈 중이며, 정부, 의료계, 환자 등이 생성하는 의료데이터의 양이 급속도로 증가하고, 의료데이터의 활용은 더욱 빠른 속도로 늘어날 전망이다.

임상에서의 인공지능은 단순히 연구 영역뿐만 아니라 진료 현장에도 빠르게 활용되고 있다. 특히, 소화기 내시경 영상을 활용한 하부 위장관 분야에서의 인공지능은 타 분야에 비하여 선도적으로 발전하고 있으며, 머지 않은 시점에 내시경실에서도 적용되어 임상 진료에 널리 활용될 것으로 예상된다.

현재까지의 인공지능 시스템의 임상 활용은 병변의 발견과 감별진단, 고위험군 선별을 위한 점수 체계 개발, 질환의 예후 또는 치료 반응의 예측 모델 개발, 그리고 장 정결도 평가 등 질 관리를 위한 메트릭스의 평가 등으로 구분되고 있다. 이 중에서 하부 위장관 영역에서의 인공지능에 관한 연구는 대부분 폴립 또는 암의 발견, 폴립의 분류와 감별진단에 집중되고 있다.

2.2 염증성 장질환 서비스 개발의 필요성

당뇨나 고혈압 등 대사질환과 심혈관 질환에서의 만성질환 관리 서비스는 활발히 출시되고 있으나, 염증 성장질환 환자를 위한 건강 관리 서비스는 제한적이다. 이는 운동과 식이에 의해 질환이 관리되는 근거가 강한 대사질환과 심혈관 질환에 비해, 염증성 장질환은 건강 관리를 통한 질환 악화의 예방에 대한 근거가 아직 확실하지 않고, 위 질환들에 비해 환자 수가 많지 않기 때문에 판단 된다. 그러나 염증성 장질환 환자의 증상 완화를 위한 스트레스 관리, 인지행동치료에 대한 근거는 꾸준히 축적되고 있고, 투약을 통한 질환 악화 예방은 근거가 충분한 편임에 비해, 환자들의 투약 순응도가 낮아서 응급실을 방문하는 사례가 임상에서 종종 발견되므로 염증성 장질환 관리를 위한 서비스 및 기술의 개발은 매우 중요하다.

2.3 Mayo Endoscopic Score(MES)

MES는 현재 임상에서 가장 많이 활용되고 있는 궤양성 대장염의 내시경 질병 활성도 평가 지

표이나, 내시경 시술 의사의 경험과 주관적인 판단기준에 영향을 받기 때문에 결과에 대한 inter/intra-observer variability가 비교적 큰 단점이 있어 평가 점수를 좀 더 객관화할 수 있는 표준적인 평가 tool이 요구된다. MES 각 점수별로 객관적이고 특징적인 소견들이 있기 때문에, 이에 대한 기계적인 학습을 통해서 인공지능을 활용한 자동화된 평가 모델을 개발할 수 있다.

해외의 경우, Yao 등 [4]은 궤양성 대장염 환자 51명의 내시경 영상파일을 이용하여 MES를 자동화로 평가하는 CNN 기반의 인공지능 모델을 개발하였으나, 내측 검증 결과 정확도 78%, 외측 검증 결과 정확도 82.8%로 다소 아쉬운 성능을 보였다. 또한, 해당 인공지능 모델은 영상의 해상도에 따라 성능에 큰 차이를 보여 실제 임상에 적용하기 위해서는 개선이 필요하다.

최근 궤양성 대장염 환자의 내시경 수요가 급격하게 증가하고 있고 내시경적 치료 반응의 객관적인 평가가 강조되고 있으나 아직까지 국내에서 개발된 MES 평가용 인공지능 모델은 없다. 이에 해외에서 발표된 인공지능 모델보다 성능이 우수한 인공지능 평가모델을 개발하고 임상에서의 적용 가능성을 평가할 필요성이 대두되고 있다.

3. 인공지능 기반 IBD Scoring System

본 논문에서는 염증성 장질환 중 궤양성 대장염의 내시경 시 의료 영상에서 병변의 정도를 분석하여 전문의의 임상적 진단을 보조하는 수단을 제공하기 위하여 인공지능 기반 스코어링 시스템을 제안한다. 영상 전체에 대한 학습 시, 불필요한 정보가 포함되어 그 정확도가 낮아질 수 있기 때문에, 임상이에 의해 레이블링 된 주요 위치정보를 함께 학습할 수 있도록 Faster R-CNN [5] 기반의 인공지능 모델을 통하여 정확도를 향상시켰다.

3.1 중증도 평가 지표

궤양성 대장염의 중증도는 임상적 중증도와 내시경적 중증도 평가로 구분된다. 임상적 중증도는 주로 Truelove and Witts [6]에 의해 제안된 궤양성 대장염 분류 기준이 주로 활용되며, 내시경적 중증도를 평가하는 지표로는 Mayo endoscopic score(MES), ulcerative colitis endoscopic index of severity (UCEIS) 등 다양한 정량 지표가 개발되고 검증되어 임상에서 활용되고 있다. 특히 MES는 궤양성 대장염 환자를 대상으로 진행되고 있는 다양한 생물학제제 및 소분자제제의 임상시험에서 내시경적 염증 활성도를 객관적으로 평가하고 약제의 내시경 관해 효과를 판정하는 데 있어 가장 많이 활용되고 있는 표준화된 지표이며 [표 1]과 같이 구분된다 [7]. MES에 따라 분류된 경도, 중증도, 중증의 비율은 각각 약 5:4:1의 비율로 그 결과가 보고된다 [8].

MES는 내시경적 염증 활성도를 평가하는데 있어, 내시경 시술 중 관찰되는 전체 대장 점막 소견 중에서 가장 염증이 심한 부위를 기준으로 하여 0-3점으로 평가한다.(0, normal or inactive; 1,

mild; 2, moderate; 3, severe)

[표 1] 궤양성 대장염 질환 활동 지수

[Table 1] Ulcerative Colitis Disease Activity Index (UCDAI; range 0~12)

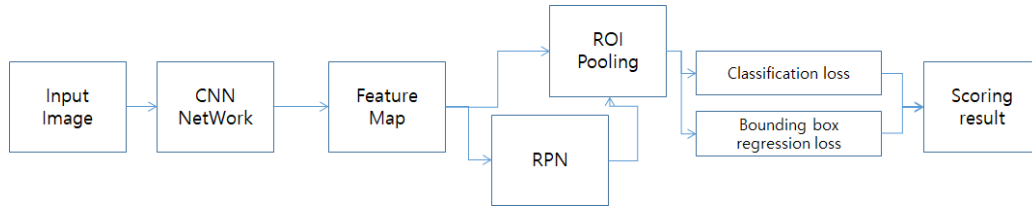
Variables/Score		Criteria
Stool Frequency	0	Normal number of stools
	1	1-2 stools more than normal
	2	3-4 stools more than normal
	3	≥5 stools more than normal
Rectal Bleeding	0	No blood seen
	1	Streaks of blood with stool less than half the time
	2	Obvious blood with stool most of the time
	3	Blood alone passed
Findings at proctosigmoidoscopy	0	Normal or inactive disease
	1	Mild disease(erythema, decreased vascular pattern, mild friability)
	2	Moderate disease(marked erythema, absent vascular pattern, friability, erosions)
	3	Severe disease(spontaneous bleeding, ulceration)
Physician's global assessment	0	Normal
	1	Mild
	2	Moderate
	3	Severe

Mild UC, UCDAI 3~5; Moderate UC, UCDAI 6~10, Severe UC, UCDAI 11~12.

3.2 Faster R-CNN 기반 MES Scoring

본 논문에서는 3.1에서 기술한 MES를 인공지능 기반 모델로 학습하여 자동으로 분류 할 수 있는 방법을 제안한다. 일반적인 인공지능 기반 classification 모델의 경우 CNN모델을 사용하여 영상 전체를 학습하게 되는데, 특히 내시경 영상의 경우 영상 내 불필요한 정보를 포함하게 될 가능성이 높으므로 정확도가 감소하게 된다. 본 논문에서는 다양한 해상도에서도 적응적으로 학습하여 보다 정확한 스코어링 정보를 제공할 수 있도록 영역기반 객체 분류 모델인 Faster R-CNN을 적용하였다.

Faster R-CNN의 기반이 되는 모델인 R-CNN(Region-based CNN) [9]은 영상 내에 검출하고자 하는 객체에 대한 관심영역에 대한 정보별로 CNN 네트워크를 학습시킨다. 초기 R-CNN은 region proposal의 개수만큼 CNN 네트워크가 필요했기 때문에 연산량이 많아 실시간 모델에 적합하지 않았다. 이러한 문제를 해결하기 위하여 2015년 중반, Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun로 구성된 Microsoft Research 팀은 Faster R-CNN을 제안하였다. [그림 1]은 Faster R-CNN 기반의 네트워크 구조를 나타낸다.



[그림 1] Fast R-CNN 기반 네트워크 구조

[Fig. 1] Faster R-CNN-based Network structure

CNN을 통해 추출한 Feature Map은 영상의 세로x가로x채널수의 형태로 출력되며, Feature Map과 RPN을 입력으로 ROI Pooling을 수행하며, 이에 대한 classification loss와 bounding box regression loss를 계산한다. 여기서 classification loss를 계산할 때, MES로 분류하기 위해 4개의 클래스(0, normal or inactive; 1, mild; 2, moderate; 3, severe)에 대한 확률을 각각 계산한다.

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

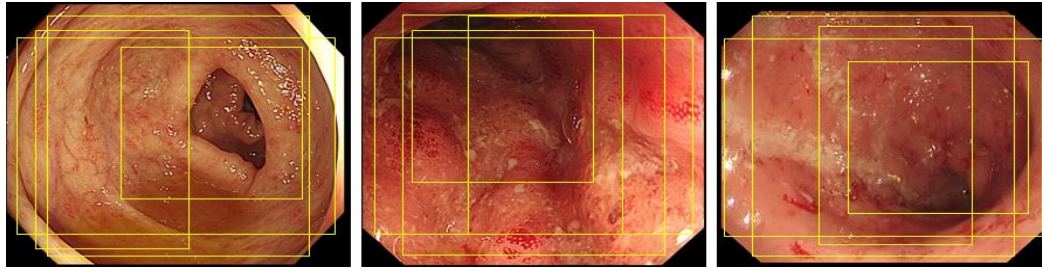
RPN을 학습시키기 위한 연산식은 (식 1)과 같다. RPN은 classification loss와 bounding box regression loss를 결합한 형태를 보인다. 여기서 i 는 1개의 anchor를 가리킨다. feature map에 대하여 미리 정의한 박스들을 사용하여 탐색을 수행하는데, 사전에 정의한 박스를 anchor라 부른다. 본 논문에서는 1개의 영상에 임상의가 레이블링한 주요 관심 영역인 4개의 영역을 anchor로 재정의 하여 사용하였다. RoI pooling은 K. He 등 [10]이 제안한 SPP(Spatial Pyramid Pooling)기법으로 SPP를 CNN 특징맵에 적용하였다.

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases} \quad (2)$$

여기서 Classification loss는 log loss를 통해서 계산되며, regression loss의 경우 (식 2)와 같이 smoothL1 함수를 사용한다.

4. 성능 평가

궤양성 대장염의 중증도 분류 모델에 대한 성능 평가를 위해 중증도별 내시경 영상을 사용하였다. 모든 영상에 대한 중증도 결과는 소화기 내과 전문의의 검증을 통해 분류하였으며, 총 500명의 환자를 대상으로 데이터를 수집하였다.



[그림 2] 수집 데이터 샘플 및 관심영역

[Fig. 2] Collect data samples and region of interesting

데이터는 [그림 2]와 같이 중증도별 영상을 클래스별로 구분하여 학습 및 평가에 사용하였다(1, mild; 2, moderate; 3, severe). [그림 2]에서 노란색 박스 영역은 전문의에 의해 그려진 주요 관심 영역으로, 본 논문에서는 해당 영역을 RPN 네트워크 학습을 위한 anchor로 활용한다.

본 논문에서 제안하는 인공지능 모델의 성능평가를 위해 MLP, SVM, CNN 알고리즘과의 비교 평가를 수행하였다. [표 2]에서 Accuracy는 사전에 중증도별 분류한 테스트 데이터에 대한 분류 정확도를 측정하였다.

[표 2] 알고리즘 성능 비교

[Table 2] Comparison of Algorithm Performance

Con	MLP	SVM	CNN	Proposed
Accuracy(%)	70.7%	72%	81.5%	88%

5. 결론

본 논문에서는 MES를 인공지능 기술 기반으로 자동으로 분류하여 임상에게 보조적인 수단으로 활용 될 수 있는 시스템을 제안하였다.

빅데이터 처리 방법의 발전과 인공지능 기술 개발로 인하여 의료현장에서의 진료 패러다임이 바뀔 것으로 예상되고 있으며, 인공지능 기술로 다양한 형태와 방대한 규모의 의료 빅데이터 분석을 통해 의료의 질 향상과 의료비 절감을 기대할 수 있다.

특히, 위장관 질환에서의 인공지능 시스템은 병변의 발견과 진단을 보조하는 역할 뿐만 아니라 질병 활성도와 치료 반응을 자동으로 평가하기 위하여 활발히 연구되고 있으며, 일부는 진료 현장에서 바로 활용할 수 있는 단계까지 발전하였다.

염증성 장질환 진료에서의 인공지능의 활용은 임상 정보와 유전체 정보를 융합한 플랫폼을 구축하고, 예측, 진단, 치료 및 예방에 도움이 될 수 있는 전방위적인 지능형 소프트웨어를 개발한다면 의료 질 향상 및 의료비용 절감에 큰 효과가 있을 것으로 판단 된다. 내시경 영상을 활용한 본

연구의 성과물은 염증성 장질환 중 하나인 궤양성 대장염의 진료에서 인공지능을 활용하여 진료의 질 향상을 높이는 계기가 될 수 있다.

개발된 인공지능 시스템의 보편화를 위해서는 외부 검증을 위한 양질의 대규모 데이터베이스를 구축해야 하며, 비용 효과적인 측면과 임상 프로세스의 활용적 측면이 좀 더 과학적으로 입증되어야 한다. MES는 궤양성 대장염 환자의 내시경적 질병 활성도 평가에서 임상시험뿐 아니라 치료 반응 및 약제 급여조건 여부에 대한 객관적인 평가 지표로 널리 활용되고 있다. 따라서 본 논문에서 제안된 인공지능을 이용한 MES 자동화 평가 도구는 연구 측면 뿐만 아니라 국내 내시경실에서 널리 적용될 수 있는 가능성이 높다. 본 논문에서 제안하는 시스템을 실제 의료기관에 적용하기 위한 향후 연구 방향으로, 다기관 빅데이터 수집을 통하여 다양한 기기 및 영상 환경에 강인한 모델로의 개선 및 실시간 시스템에 적합하도록 인공지능 모델의 개선과 하이퍼 파라미터의 조정에 대한 보완을 수행하고자 한다.

References

- [1] A. Kornbluth, D. B. Sachar, and Practice Parameters Committee of the American College of Gastroenterology, "Ulcerative colitis practice guidelines in adults: American College Of Gastroenterology, Practice Parameters Committee", The American journal of gastroenterology, vol. 105, January 2010, pp. 501-523; doi: 10.1038/ajg.2009.727.
- [2] M. Feldman, L. S. Friedman, L. J. Brandt, Sleisenger and Fordtran's gastrointestinal and liver disease: pathophysiology, diagnosis, management, Saunders, 2006.
- [3] M. F. Neurath, S. P. L. Travis, "Mucosal healing in inflammatory bowel diseases: a systematic review", Gut, vol. 61, Nov 2012, pp. 1619-1635, doi: 10.1136/gutjnl-2012-302830.
- [4] H. Yao, K. Najarian, J. Gryak, S. Bishu, M. D. Rice, A. K. Walgee, H. J. Wilkins, R. W. Stidham, "Fully automated endoscopic disease activity assessment in ulcerative colitis", Gastrointest Endosc, vol. 93, March 2021, pp. 728-736, doi: 10.1016/j.gie.2020.08.011.
- [5] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, June 2017, pp. 1137-1149, doi: 10.1109/TPAMI.2016.2577031.
- [6] S. C. Truelove, L. J. Witts, "Cortisone in ulcerative colitis: final report on a therapeutic trial", British Medical Journal, vol. 2, October 1955, pp. 1041-1048, doi: 10.1136/bmj.2.4947.1041.
- [7] K. W. Schroeder, W. J. Tremaine, D. M. Ilstrup, "Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis: a randomized study", The New England Journal of Medicine, vol. 317, December 1987, pp. 1625-1629, doi: 10.1056/NEJM198712243172603.
- [8] Y. M. Kim, S. H. Park, S. K. Yang, J. W. Choi, S. H. Kim, J. S. Byeon, S. J. Myung, Y. K. Cho, C. S. Yu, J. W. Choi, B. Kim, K. D. Choi, J. H. Kim, "Clinical characteristics and long-term course of ulcerative colitis in Korea", Intestinal Research, vol. 4, June 2006, pp. 12-21.

- [9] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, USA, pp. 580-587, doi: 10.48550/arXiv.1311.2524.
- [10] K. He, X. Zhang, S. Ren, J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, September 2015, pp. 1904-1916, doi: 10.1109/TPAMI.2015.2389824.