

MobileNetV2를 활용한 궤양성 대장염 중증도 예측: 온디바이스 적용을 위한 경량화 접근

김범휘, 우동원, 정성문*

한국전자통신연구원, 경북대학교병원 의료인공지능연구센터 *경북대학교 의료정보학교실

bumhwi.kim@etri.re.kr, woodongwon23@gmail.com, *jeongsm00@gmail.com

Ulcerative Colitis Severity Prediction Using MobileNetV2: A Lightweight Approach for On-Device Deployment

Bumhwi Kim, Dongwon Woo, Sungmoon Jeong*

Medical IT Convergence Section, Electronics and Telecommunications Research Institute(ETRI)
Research Center for Artificial Intelligence in Medicine, Kyungpook National University Hospital,

*Department of Medical Informatics, School of medicine, Kyungpook National University

요약

궤양성 대장염(Ulcerative Colitis, UC)의 중증도 평가는 일반적으로 내시경 검사를 기반으로 한 UCEIS(Ulcerative Colitis Endoscopic Index of Severity)를 통해 이루어진다. 그러나 내시경 평가는 침습적이며, 시술 부담과 비용이 수반된다. 선행 연구에서 대변 이미지를 활용하여 UC 중증도를 분류하는 Vision Transformer(ViT) 기반 모델을 개발하였으며, UCEIS 점수의 이진 분류를 통해 유의미한 성능을 보고하였다. 다만, 해당 클래스 구분은 휴리스틱에 기반하였으며, 실제 임상적 분류 체계와 일치하지 않는 한계가 있었다. 본 연구에서는 기존 접근을 두 가지 측면에서 확장하였다. 첫째, 자원 제약 환경에서의 온디바이스 적용 가능성을 확보하기 위해 ViT 대신 경량 합성곱 신경망 구조인 MobileNetV2를 도입하였으며, 이 과정에서 모델 경량화에 따른 성능 저하 가능성을 인지하였다. 둘째, 임상적으로 보다 타당한 분류를 위해 UCEIS 점수를 0-1, 2-4, 5-8의 3클래스로 재구성하였다. 데이터셋의 클래스 불균형 문제를 해결하기 위해 다양한 데이터 증강 기법을 적용하여 일반화 성능 향상을 도모하였다. 본 논문에서는 해당 방법론적 절차와 설계 배경을 기술하고, 비침습적 대변 이미지를 이용한 UC 중증도 예측 모델의 임상적·실용적 의의를 논의한다.

I. 서론

궤양성 대장염(Ulcerative Colitis, UC)은 만성 염증성 장질환으로, 증상의 중증도 평가가 치료 계획 수립에 중요한 역할을 한다. UCEIS(Ulcerative Colitis Endoscopic Index of Severity)는 UC 중증도를 정량화하는 표준 지표이지만, 내시경 검사를 기반으로 하므로 환자에게 침습적이며 비용과 시간이 소요된다. 최근에는 비침습적 방법으로 환자의 상태를 추정하려는 시도가 증가하고 있으며, 특히 대변 이미지를 활용한 인공지능(AI) 기반 접근이 주목받고 있다.

기존 선행 연구에서는 대변 이미지를 기반으로 Vision Transformer(ViT) 모델을 학습하여[1], UCEIS 점수를 0~2와 3~8로 구분하는 이진 분류를 수행하였다[2]. 해당 접근은 혁신적인 아이디어를 기반으로 임상적으로 활용 가능성을 보여주었으나, 다음과 같은 한계가 있었다. 먼저 클래스 구분이 임상 기준이 아닌 휴리스틱에 기반해 설정되어 실제 UCEIS 점수 체계와 차이가 있었다. 또한 ViT 모델은 연산량과 메모리 요구량이 커서 온디바이스 환경에서 실시간 추론을 수행하기 어렵다.

본 연구에서는 이러한 한계를 보완하기 위해 경량 모델인 MobileNetV2를 적용하여 온디바이스 추론 가능성을 확보하고, 임상 기준에 보다 부합하는 다중 클래스 설정을 도입하며, 다양한 데이터 증강 기법을 적용하여 불균형 문제를 완화하였다. 이를 통해 단순한 분류 성능 개선을 넘어, 실제 임상 적용 가능성과 해석 가능성을 동시에 고려한 모델 개발을 목표로 하였다.

II. 실험 방법

본 연구에 사용된 데이터셋은 궤양성 대장염 환자의 대변 이미지를 기반으로 하며, 각 이미지에는 내시경 평가를 통해 측정된 UCEIS 점수가 라벨로 부여되었다. 원래 임상적 기준에 따른 구분은 0-1, 2-4, 5-6, 7-8로 총 4단계이지만, 표본 수가 제한적이고 특히 고점수 구간(7-8)의 데이터가 매우 희소하여 안정적인 학습이 어려웠다. 이에 따라 0-1을 경증, 2-4를 중등도, 5-8을 중증으로 묶어 3클래스 분류 실험을 진행하였다.

모델 경량화를 위해 MobileNetV2를 활용하였다. MobileNetV2는 inverted residual과 linear bottleneck 구조를 활용하여 연산 효율성을 극대화하며, depthwise separable convolution을 통해 파라미터 수를 크게 줄이면서도 적절한 성능을 유지할 수 있다[3].

본 연구에서는 224×224 이미지를 입력으로 하고, ImageNet 사전 학습 가중치를 초기값으로 사용하였다. 분류기는 Global Average Pooling과 Dropout을 거쳐 Fully Connected Layer와 Softmax로 구성하였으며, Adam 옵티마이저를 사용하고 초기 학습률은 1e-4로 설정하였으며 StepLR 스케줄러를 적용하였다.

데이터 불균형 문제를 완화하기 위해 다양한 데이터 증강 기법을 적용하였다. 기본 변환으로는 수평 및 수직 반전, ±15° 범위 회전, 랜덤 크롭, 밝기·대비·채도 조절을 사용하였다. 여기에 CutMix[4], MixUp[5]과 같은 고급 변환을 적용하여 클래스 간 특징을 혼합하고, Random Erasing을 통해 배경과 형태 변이에 대한 강건성을 강화하였다. 특히 데이터가 적은 클래스에는 증강 횟수를 늘려 클래스별 표본 수를 균형에 가깝게 조정하였다.

성능 평가는 Accuracy, F1-score, 클래스별 민감도와 특이도, 그리고 AUC를 활용하였다. 분석 항목은 모델 경량화에 따른 성능 변화, 다양한 증강 기법 적용 전후의 성능 향상 정도를 포함하였다.

III. 실험 결과

표 1. 모델 별 이진 분류 결과

Model	정확도	민감도	특이도	AUC
ViT	0.932	0.8864	0.9661	0.943
MobileNetV2	0.8932	0.8636	0.9153	0.911

표 1은 ViT와 MobilenetV2의 이진 분류 결과를 나타낸다. MobileNetV2 모델은 전역적 패턴 학습 능력이 우수한 ViT 대비 전반적인 분류 성능은 낮게 나타났지만, 여전히 우수한 성능을 보이는 것을 확인할 수 있었다.

표 2. MobileNetV2 기반 3-class 분류 결과

Model	3-class	3-class with 증강
정확도	0.6125	0.6821

표 2는 MobileNetV2 모델을 활용한 3-class 분류 실험 결과를 나타낸다. 데이터 증강 기법을 적용한 경우, 미적용 대비 약 7% 이상 성능이 향상되었으며, 이를 통해 증강 기법이 클래스 불균형 문제를 일정 부분 완화하는데 효과적임을 확인하였다.

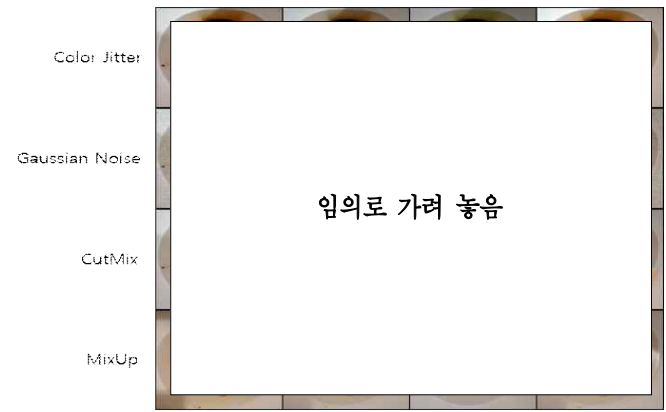


그림 2. 데이터 증강 기법 적용 예시

그림 1은 본 연구에서 적용한 데이터 증강 기법이 실제 영상에 어떻게 반영되는지를 정성적으로 보여준다. 의료 AI 분야에서는 단순한 성능 지표의 향상만으로 모델의 우수성을 평가하기에는 한계가 있다. 성능 개선을 위해 적용한 방법이 어떤 방식으로 영향을 미쳤는지를 시각적으로 해석하는 과정이 중요하다. 이에 본 연구에서는 각 증강 기법이 생성한 영상의 변화를 확인하였으며, 이를 통해 단순한 수치 향상뿐만 아니라 이미지의 시각적 다양성이 확보되어 모델이 보다 풍부한 패턴을 학습할 수 있었음을 확인하였다.

IV. 결론

본 연구에서는 궤양성 대장염 환자의 대변 이미지를 이용하여 UCEIS 기반의 중증도 분류를 수행하는 모델을 개발하였다. 기존 연구에서 사용된 Vision Transformer(ViT) 모델을 경량화하여 온디바이스 환경에서 구현할 수 있도록 MobileNetV2를 적용하였으며, 임상적 기준에 보다 부합하는 3클래스 분류 설정을 도입하였다. 데이터 불균형 문제를 해결하기 위해 다양한 데이터 증강 기법을 활용하였고, 그 효과를 정량적 지표뿐만 아

니라 시각적 해석을 통해 검증하였다.

실험 결과, MobileNetV2는 ViT 대비 전반적인 분류 성능은 다소 낮았으나, 온디바이스 환경에서 구현하기에는 충분한 성능을 보임을 확인하였다. 또한, 임상 기준에 부합하기 위한 3-class 결과도 이진 분류에 비해 성능이 하락되었으나, 다양한 증강 기법의 적용을 통해 클래스 불균형을 완화하고 모델의 일반화 성능을 향상시킬 수 있었다. 특히 의료 AI 분야에서는 단순 성능 지표 외에도 결과에 대한 해석 가능성이 중요하며, 본 연구에서 제시한 시각적 증강 결과 분석은 이러한 관점에서 의미 있는 정보를 제공한다.

향후 연구에서는 데이터셋을 확충하여 임상 기준에 따른 4클래스 분류를 시도하고, 경량 모델의 성능 향상을 위해 지식 증류나 구조 최적화 기법을 적용할 계획이다. 이를 통해 비침습적 대변 이미지 분석 기반의 UC 중증도 예측 모델이 실제 임상 현장에서 보조 진단 도구로 활용될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문(또는 저서, 특허)은(는) 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임(‘학연협력플랫폼구축 시범사업’ RS-2023-00304695(50%), RS-2023-00304776(40%)) 및 한국전자통신연구원 연구운영지원사업 [25ZD1140, 대경권 지역산업 기반 ICT 융합 기술 고도화 지원(의료)](10%)의 지원을 받아 수행된 연구임

참 고 문 헌

[1] Lee, J. W., Woo, D., Kim, K. O., Kim, E. S., Kim, S. K., Lee, H. S., ... & Jeong, S. (2025). Deep learning model using stool pictures for predicting endoscopic mucosal inflammation in patients with ulcerative colitis. Official journal of the American College of Gastroenterology| ACG, 120(1), 213-224

[2] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houslsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929

[3] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[4] Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6023-6032).

[5] Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.