

온디바이스 적용을 위한 다중 경량화 모델 기반 궤양성 대장염 중증도 예측 연구

Ulcerative Colitis Severity Prediction Using Multiple Lightweight Models for On-Device Deployment

Dongwon Woo¹ · Bumhwi Kim² · Ho-Young Chung³ · Sungmoon Jeong^{4*}

¹Researcher, Research Center for AI-Driven Medical Innovation, Kyungpook National University Hospital, Daegu, 41940, Korea

²Senior Researcher, Electronics and Telecommunications Research Institute(ETRI), Daegu, 42994, Korea

³Professor, Department of Medical Informatics, School of Medicine, Kyungpook National University, Daegu, 41944, Korea

^{4*}Associate Professor, Department of Medical Informatics, School of Medicine, Kyungpook National University, Daegu, 41944, Korea

ABSTRACT

Ulcerative colitis (UC) severity is typically evaluated using the Ulcerative Colitis Endoscopic Index of Severity (UCEIS), but endoscopy is invasive and costly. To explore a non-invasive alternative, previous work proposed a Vision Transformer (ViT)-based model using stool images for binary classification of UCEIS scores. In this study, we extend this approach in two ways. First, to enable practical use in resource-limited and on-device settings, we implement lightweight models—MobileNetV2, EfficientNet-B0, ShuffleNetV2, and GhostNet—and compare their performance and efficiency against ViT. Second, we restructure UCEIS scores into three clinically relevant classes (0-1, 2-4, 5-8) for multi-class classification. This work demonstrates the feasibility and clinical value of stool image-based UC

Keywords : Ulcerative Colitis(UC), Deep learning, UCEIS, Light-weight model

I. 서 론

궤양성 대장염(Ulcerative Colitis, UC)은 재발이 잦은 만성 질환으로, 적절한 치료 전략 수립을 위해 중증도를 정확히 평가하는 것이 필수적입니다. 현재 임상 표준인 내시경 기반의 UCEIS는 정확도가 높지만, 침습성과 높은 비용으로 인해 반복적인 적용에는 한계가 있다[1]. 최근 이러한 부담을 줄이기 위한 비침습적(Non-invasive) 진단 기술의 요구가 증가하면서, 대변 이미지와 인공지능을 활용한 모니터링이 새로운 대안으로 주목받고 있다[2,3]. 다만, Vision Transformer(ViT)를 활용한 선행 연구의 경우 모델의 연산량이 방대하고, 임상 기준과 다른 방식의 이진 분류를 채택하여 실용성에 한계가 있었다[4]. 이를 보완하기 위해 본 연구는 다음과 같은 방향으로 확장되었다. 첫째, 경량화 모델을 적용하여 실제 모바일 환경에서의 운용 효율성을 확보하고 기존 ViT 모델과 성능을 비교하였다. 둘째, 임상 진료 체계와의 정합성을 높이기 위해 UCEIS를 경증(0-1), 중등증(2-4), 중증(5-8)의 3-class 구조로 재구성하였다. 본 연구의 목표는 단순히 분류 정확도를 높이는 데 그치지 않고, 임상 현장에서 활용 가능한 비침습적·경량형 UC 중증도 예측 시스템을 구축하는 데 있다.

II. 본 론

Table. 1 Dataset Collection and Split Summary

	Collection Period	# of images	# of patients
Training	2021.09~2022.09	1,625	229
Validation	2021.09~2022.09	379	57
Test	2022.10~2023.04	1,099	103

2.1 데이터 셋

본 연구는 2021년 9월부터 2023년 4월까지 대구권 4

Received 10 November 2025, Revised 18 November 2025, Accepted 9 January 2026

* Corresponding Author Sungmoon Jeong(E-mail:jeongsm00@gmail.com, Tel:+82-53-200-4524), Associate Professor, Department of Medical Informatics, School of Medicine, Kyungpook National University, Daegu, 41944, Korea

Open Access <http://doi.org/10.6109/jkiice.2026.30.2.371>

print ISSN: 2234-4772 online ISSN: 2288-4165

개 대학병원(경북대, 계명대, 영남대, 대구가톨릭대)에서 궤양성 대장염 환자의 데이터를 수집하였으며, 데이터는 임상 지표인 UCEIS 점수와 내시경 검사일 기준 7일 이내에 스마트폰으로 촬영된 대변 이미지로 구성된다. 데이터셋은 2022년 9월을 기준으로 이전 데이터는 환자 단위 8:2 비율로 학습 및 검증에 활용하였으며, 이후 데이터는 테스트 셋으로 구성하여 모델의 성능을 평가하였다. 표 1은 각 데이터셋의 환자 및 이미지 수를 요약한 것이고, 그림 1은 데이터 예시를 나타낸다.



Fig. 1 Example images of stool

2.2 실험 방법

본 연구에서는 모바일 환경의 실시간 추론을 위해 연산 효율이 검증된 네 가지 경량 모델을 사용하였다. MobileNetV2는 Inverted Residual 구조로 연산 효율을 극대화하며[5], EfficientNet-B0는 Compound Scaling을 통해 적은 파라미터로 높은 정확도를 확보하였다[6]. ShuffleNetV2는 채널 셔플 구조로 최적의 속도-정확도 균형을 제공하며[7], GhostNet은 Ghost Module을 활용하여 연산 중복을 최소화함으로써 효율적인 특징 추출을 수행한다[8].

2.2.1 Image-wise Classification (학습)

이미지 단위 분류에서는 학습 데이터 셋 내 모든 대변 이미지를 개별적인 독립 데이터로 간주하여 학습을 수행하며 그림 2는 이미지 단위 학습 과정의 개요를 나타낸다. 본 연구에서는 ImageNet-21K 사전 학습 가중치를 초기값으로 사용하기 위해 입력 데이터는 224x224 해상도로 정규화 하였다. 최적화 알고리즘으로는 Adam 옵티마이저를 사용하고 초기 학습률은 $1e-3$ 으로 설정하였으며, 20 epoch 주기마다 학습률을 감소시키는 StepLR 스케줄러를 적용하였다. Batch size는 16으로 설정하여 총 100 epoch 동안 학습을 수행하였다. 추가적으로, 충분하지 않은 데이터 수를 보완하기 위해 랜덤 크롭, 회전, CutMix, MixUp과 같은 증강 기법을 도입하였다.

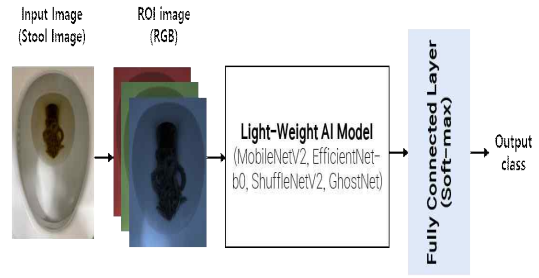


Fig. 2 Image-wise training process

2.2.2 Patient-wise Classification (테스트)

환자 단위 분류는 학습이 완료된 모델을 실제 임상 상황에 가깝게 평가하기 위한 테스트 단계에 해당한다. 테스트 세트에 포함된 각 환자에 대해, 해당 환자가 보유한 모든 대변 이미지에 대해 개별적으로 추론을 수행한다. 이후, 각 이미지에서 산출된 예측 확률값을 평균하여 최종적으로 환자 단위의 증증도를 결정한다. 이와 같은 과정을 통해 단일 이미지 수준의 오분류 영향을 최소화하고, 환자 전체의 상태를 보다 안정적으로 반영할 수 있다. 그림 3은 이러한 환자 단위 테스트 절차를 시각적으로 나타낸 것이다.

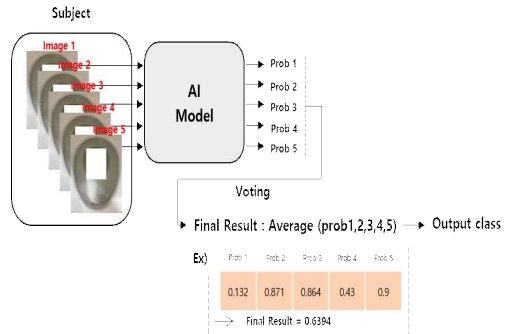


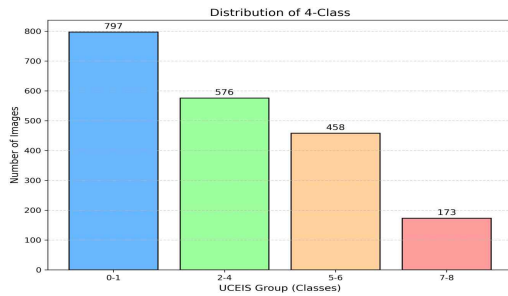
Fig. 3 Patient-wise test process

2.3 실험 결과

본 연구에서는 경량화 모델의 실용성을 평가하고자 이진 분류 및 3-class 분류 실험을 수행하였다. 먼저 선행 연구의 ViT 모델과 이진분류 성능을 비교하고, 파라미터 수, FLOPs 등 경량화 지표를 분석하였다. 또한 임상적 타당성을 강화하기 위해 3-class 분류 실험을 추가하였다. 임상적으로는 UCEIS 점수를 기준으로 0 - 1점 (remission), 2 - 4점(mild), 5 - 6점(moderate), 7 - 8점

Table. 2 2-class classification results

Model	Major Parameter	Mode	Accuracy	Sensitivity	Specificity	AUROC
ViT	Depth:12, Heads: 12, Patch size: 16	image-wise	0.8526	0.76	0.9168	0.904
		Patient-wise	0.932	0.864	0.9661	0.943
MobileNetV2	Depth: 19, factor: 6, Multiplier: 1.0	image-wise	0.8071	0.7178	0.869	0.86
		Patient-wise	0.8932	0.8636	0.9153	0.911
EfficientNet-b0	Depth: 18, Compound Scaling : (1,1,1)	image-wise	0.8007	0.7667	0.8243	0.8728
		Patient-wise	0.8641	0.9091	0.8305	0.918
ShuffleNetV2	Channel size: (116, 232, 464)	image-wise	0.828	0.8311	0.8259	0.9002
		Patient-wise	0.9233	0.8864	0.9492	0.94
GhostNet	Depth: 28 Ghost ratio: 2	image-wise	0.8371	0.7578	0.8921	0.9
		Patient-wise	0.9223	0.8409	0.9831	0.9518

**Fig. 4** 4-class Distribution of training dataset

(severe)의 4-class 구분이 이상적이지만, 그림 4와 같이 실제 데이터에서는 severe 구간 데이터의 부족으로 인한 클래스 불균형 문제가 확인되었다. 이에 따라 본 연구에서는 0 - 1점을 경증, 2 - 4점을 중등증, 5 - 8점을 중증으로 재구성하여 클래스별 불균형 문제를 완화하여 3-class 분류 실험을 수행하였다.

Table. 3 Comparison of lightweightness

Model	# of parameter (M)	FLOPs (MMac)
ViT	85.8	16,870
MobileNetV2	2.23	305.73
EfficientNet-b0	4.01	390.83
ShuffleNetV2	1.26	151.36
GhostNet	3.9	151.84

2.3.1 이진 분류

표 2와 표 3은 모델 별 이진 분류 성능 및 경량화 지표를 나타낸다. 경량화 모델들은 전반적으로 분류 성능은 낮았지만 ViT 대비 파라미터와 연산량이 현저히 적음에도 불구하고 우수한 성능을 유지하였다. 특히 GhostNet은 Ghost Module을 통해 특징 중복을 최소화

함으로써 ViT와 대등한 성능을 보였다. 구조적 효율성 측면에서는 GhostNet과 ShuffleNetV2가 가장 우수했다. 특히 GhostNet은 ViT 대비 약 $1/100$ 수준의 연산량(151.84 MMac)**과 3.9M의 파라미터만으로 높은 정확도를 달성하여, 성능 저하 없이 실시간 추론 가능성과 모델 효율성을 동시에 확보하였다.

Table. 4 3-class classification results

Model		Accuracy	F1-score
MobileNetV2	image-wise	0.596	0.6016
	patient-wise	0.6798	0.6893
EfficientNet-b0	image-wise	0.566	0.5684
	patient-wise	0.5728	0.5943
ShuffleNetV2	image-wise	0.5596	0.566
	patient-wise	0.6408	0.6436
GhostNet	image-wise	0.6433	0.6468
	patient-wise	0.767	0.7336

Table. 5 Confusion Matrix of 3-class classification

			예측		
			경증	중등증	중증
GhostNet	실제	경증	43	5	0
		중등증	7	25	8
		중증	1	3	11
EfficientNet-B0	실제	경증	20	28	0
		중등증	5	27	8
		중증	1	2	12

2.3.2 3-class 분류

표 4는 3-class 분류 성능을 나타내며, Accuracy와 F1-score를 이용하여 정량적으로 평가하였다. 모든 모델에서 환자 단위 예측이 이미지 단위보다 높은 성능을 보였는데, 이는 환자별 결과 평균화 과정에서 예측의 일

관성이 향상된 결과로 해석된다. 클래스 개수 증가로 인해 이전 분류 대비 성능은 다소 하락했으나, GhostNet은 환자 단위에서 Accuracy 0.767, F1-score 0.734를 기록하며 가장 우수한 성능을 나타냈다. 또한, 표 5의 Confusion Matrix를 통해 최저 성능인 EfficientNet-B0와 최고 성능인 GhostNet을 비교한 결과, GhostNet이 전반적으로 균형 잡힌 분류 성능을 유지함을 확인하였다.

III. 결 론

본 연구는 대변 이미지를 활용한 비침습적 궤양성 대장염(UC) 중증도 예측 모델의 임상적 가능성을 평가하고, 모바일 환경에 최적화된 경량화 모델의 효율성을 검증하였습니다. 실험 결과, 제안된 경량 모델들은 기존 ViT 대비 현저히 적은 연산량으로도 대등한 성능을 보였으며, 특히 GhostNet은 환자 단위 AUROC 0.9518을 기록하며 성능과 효율성의 최적 균형을 입증하였다. 또한, UCEIS를 3-class(경증, 중등증, 중증)로 재구성하여 임상 현장과의 정합성을 높였으며, 다중 클래스 환경에서도 안정적인 분류 성능을 확인하였다. 이러한 결과는 대변 이미지 기반 UC 중증도 예측이 내시경에 대한 비침습적 대안으로서 잠재력을 지님을 보여준다. 향후 데이터 규모 확대와 다기관 검증을 통해 모델의 일반화 성능을 강화할 예정이며, 본 연구에서 제안한 경량화 접근법은 실시간 진단 보조 및 원격 모니터링 시스템 구축을 위한 핵심 기술이 될 것으로 기대된다.

ACKNOWLEDGEMENTS

This paper was partly supported by the Commercialization Promotion Agency for R&D Outcomes(COMPA) grant funded by the Korean Government(Ministry of Science and ICT). (RS-2023-00304695, RS-2023-00304776) and the Institute of Information & Communications Technology Planning & Evaluation(IITP) - ITRC(Information Technology Research Center) grant funded by the Korea government(MSIT) (IITP-2025-RS-2020-II201808)

REFERENCES

- [1] S. P. L. Travis, D. Schnell, P. Krzeski, M. T. Abreu, D. G. Altman, J. F. Colombel, B. G. Feagan, S. B. Hanauer, M. Lemann, G. R. Lichtenstein, et al., "Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS)," *Gut*, vol. 61, no. 4, pp. 535-542, Apr. 2012. DOI: 10.1136/gutjnl-2011-300486.
- [2] X. Lui, J. Reigle, V. S. Prasath, and J. Dhaliwal, "Artificial intelligence image-based prediction models in IBD exhibit high risk of bias: A systematic review," *Computers in Biology and Medicine*, vol. 171, pp. 108093, Mar. 2024. DOI: 10.1016/j.compbiomed.2024.108093.
- [3] K. Takabayashi, T. Kobayashi, K. Matsuoka, B. G. Levesque, T. Kawamura, K. Tanaka, T. Kadota, R. Bise, S. Uchida, T. Kanai, and H. Ogata, "Artificial intelligence quantifying endoscopic severity of ulcerative colitis in gradation scale," *Digestive Endoscopy*, vol. 36, no. 5, pp. 582-590, May 2024. DOI: 10.1111/den.14677.
- [4] J. W. Lee, D. Woo, K. O. Kim, E. S. Kim, S. K. Kim, H. S. Lee, B. Kang, Y. J. Lee, J. Kim, B. I. Jang, et al., "Deep Learning Model Using Stool Pictures for Predicting Endoscopic Mucosal Inflammation in Patients With Ulcerative Colitis," *Official journal of the American college of Gastroenterology*, vol. 120, no. 1, pp. 213-224, Jan. 2025. DOI: 10.14309/ajg.0000000000002978.
- [5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," in *Proceedings of IEEE conference on computer vision and pattern recognition*, Salt Lake City: USA, pp. 4510-4520, 2018.
- [6] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv: 1905.11946*, May 2019. DOI: 10.48550/arXiv.1905.11946.
- [7] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, Munich: DE, pp. 122-138, 2018. DOI: 10.1007/978-3-030-01264-9_8.
- [8] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in *Proceedings of IEEE/CVF conference on computer vision and pattern recognition*, Seattle: USA, pp. 1577-1586, 2020. DOI: 10.1109/CVPR42600.2020.00165.