Journal of Biomedical Engineering Research 44: 118-124 (2023)
http://dx.doi.org/10.9718/JBER.2023.44.2.118

학 술 논 문

위 내시경 이미지 품질에 따른 병변 검출 모델의 성능 비교 연구

이율희[1] · 김영재[2] · 김광기[2,3]*

[1]가천대학교 간호대학 간호학과, [2]가천대학교 의과대학 의공학교실
[3]가천융합의과학원(GAIHST) 융합의과학과

A Performance Comparison Study of Lesion Detection Model according to Gastroscopy Image Quality

Yul Hee Lee[1], Young Jae Kim[2] and Kwang Gi Kim[2,3]*

[1]Department of Nursing, College of Nursing, Gachon University,
[2]Department of Biomedical Engineering, College of Medicine, Gil Medical Center, Gachon University,
[3]Department of Health Sciences and Technology, Gachon Advanced Institute for Health & Sciences and Technology (GAIHST)
(Manuscript received 20 February 2023 ; revised 26 March 2023 ; accepted 27 March 2023)

Abstract: Many recent studies have reported that the quality of input learning data was vital to the detection of regions of interest. However, due to a lack of research on the quality of learning data on lesion detetcting using gastroscopy, we aimed to quantify the impact of quality difference in endoscopic images to lesion detection models using Image Quality Assessment (IQA) algorithms. Through IQA methods such as BRISQUE (Blind/Referenceless Image Spatial Quality Evaluation), Laplacian Score, and PSNR (Peak Signal-To-Noise) algorithm on 430 sheets of high quality data (HQD) and 430 sheets of low quality data (PQD), we showed that there were significant differences between high and low quality images in lesion detecting through BRISQUE and Laplacian scores ($p<0.05$). The PSNR value showed 10.62±1.76 dB on average, illustrating the lower lesion detection performance of PQD than HQD. In addition, F1-Score of HQD showed higher detection performance at 77.42±3.36% while F1-Score of PQD showed 66.82±9.07%. Through this study, we hope to contribute to future gastroscopy lesion detection assistance systems that involve IQA algorithms by emphasizing the importance of using high quality data over lower quality data.

Key words: Gastroscopy, Image quality assessment algorithm, Deep learning, RetinaNet, Lesion detection

I. 서 론

세계보건기구(World Health Organization, WHO)에 따르면 위암은 2020년 기준 약 109만 건이 발생하였으며 약 77만명이 사망하였다[1-2]. 위암은 조기에 진단하여 치료하는 것이 매우 중요한 것으로 알려져 있으며 위 내시경 검사는 위암 조기 진단에 중요한 진단 방법으로 검사량이 꾸준히 증가하고 있다[3-6]. 하지만 내시경 검사량의 증가에 따라 업무의 피로도가 쌓여 내시경 검사 진단 정확도가 낮아질 위험이 있으며 전문의의 경력에 따라 진단의 편차가 발생할 수 있다는 문제점이 있다.

기존 내시경 검사의 문제를 극복하기 위한 방안으로 컴퓨터 보조 진단 기법(Computer-Aided Diagnosis, CADx)이 있다. CADx 시스템은 의료 영상을 기반으로 진단 의견을 제시하여 전문의의 진단 효율을 상승시킬 수 있는 시스템이며, 최근에는 인공지능을 접목한 다수의 관련 연구들이 보고되고

*Corresponding Author : Kwang Gi Kim
Department of Biomedical Engineering, College of Medicine, Gil Medical Center, Gachon University, Incheon 21565, Republic of Korea
Tel: +82-10-3393-4544
E-mail: kimkg@gachon.ac.kr
이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신산업진흥원의 지원(No.S0252-21-1001, AI정밀의료솔루션(닥터앤서2.0)개발)과 경기도의 경기도 지역협력연구센터 사업의 일환으로 수행하였음[GRRC-가천2020(B02), AI 기반 의료영상 분석].

www.kci.go.kr

있다[7-10]. 2018년 김동현 등은 위 내시경 영상에서 병변을 진단하기 위한 Inception 모델을 제시하였으며, 124개의 이미지에 대한 민감도(Recall)와 특이도(Specificity)는 각각 0.81과 0.68로 나타났다[11]. 2020년 이신애 등은 내시경 영상에서 색상 특징과 질감 특징을 기반으로 정상과 비정상을 분류하기 위한 LS-SVM 모델을 제시하였으며, 335장의 테스트 데이터에 대해 0.85의 ROC(Receiver Operating Characteristic) 곡선의 AUC(Area Under Cover)값을 보였다[12].

기존의 보고된 다수의 연구들은 병변의 검출 또는 진단 성능을 높이기 위해 다양한 아키텍처들을 제시하고, 학습 데이터의 양을 늘리기 위한 방안을 모색한다. 이와 같은 접근은 성능 향상에 분명한 도움이 되지만, 데이터의 품질도 함께 고려할 필요가 있다[13]. 최근에는 딥러닝 학습에서의 데이터 품질과 관련된 연구가 다수 보고되면서, 딥러닝 분야에서 데이터 품질의 중요성이 커지고 있다.

2019년 이정민 등은 노이즈, 밝기 값, 픽셀 크기 등 다양한 영상 품질에 따른 딥러닝 알고리즘의 성능과 결과를 분석하는 연구를 진행하였다. 그 결과, 딥러닝 폐색 영역 검출에 노이즈, 픽셀 크기, 밝기 값 모두 영향을 미쳤지만 특히 노이즈는 검출 정확성과 검출 비율에 가장 많은 영향을 미치는 요소임을 보고하였다[14]. 뿐만 아니라, 2020년 윤명성 등은 흉부 X선에서 딥러닝을 기반으로 한 기흉 검출 연구를 통해 영상의 밝기 등 픽셀 값 변화에 딥러닝 결과가 민감하게 반응한다는 것을 보고하였다[15]. 이처럼 학습 데이터의 품질이 떨어지면 학습에 안 좋은 영향을 미치기 때문에 품질이 좋지 않은 데이터를 선별하는 과정이 필요하다. 하지만 위내시경 분야에서는 학습 데이터의 품질에 대한 분석 및 병변 검출에 미치는 영향과 관련된 연구는 아직까지 보고된 바 없다. 환자의 생명과 직결될 수 있는 의료 분야에서 학습 데이터의 임상적 신뢰성은 매우 중요한 문제이기 때문에[16], 이미지 품질을 주관적 기준인 육안으로 하는 것이 아닌 컴퓨터 알고리즘을 통한 정량적 평가 수치로 제시하고, 객관적으로 선별할 수 있어야 한다.

이에 본 연구에서는 이미지 품질 평가(Image Quality Assessment, IQA) 알고리즘을 통해 내시경 이미지에서 품질 차이를 정량적으로 수치화하고 품질 차이에 따라 병변 검출 모델에서 검출이 얼마나 차이 나는지 확인하고자 한다.

II. 연구 방법

1. 개발 환경

본 연구는 Intel(R) Core(TM) i7-10700K(Intel, Santa Clara, CA, USA) 및 32 GB RAM의 시스템을 사용하였으며, 64비트 운영체제에서 Python(Version 3.8.12)을 통해 수행되었다. 이미지 전처리 및 딥러닝 학습을 위하여 Tensorflow (Version 2.3.0), Keras(Version 2.3.0), OpenCV(Version 4.7.0.68), image-quality(Version 1.2.7) 라이브러리를 사용하였다. 딥러닝 학습은 1개의 Geforce RTX 2080Ti 24GB (NVIDIA, Santa Clara, CA, USA)를 사용하여 진행하였다.

2. 데이터 수집

본 연구에서는 2004년 9월부터 2022년 9월까지 가천대 길병원을 내원하여 소화기 내시경 검사를 받은 405명의 위 내시경 이미지 데이터 2,490장을 수집하였다. 수집된 데이터는 위 내시경 장비(Olympus GIF-HQ290, Tokyo, Japan)로 해상도 (Dots Per Inch, DPI) 150에서 촬영되었으며, 이미지의 크기는 520*480 픽셀(Pixel)로 모두 동일하다. 사용된 모든 데이터는 기관생명윤리위원회 (Institutional Review Board, IRB)의 허가를 받았다(GBIRB 2021-383). 수집한 이미지 데이터는 전문의에 의해 품질이 좋은 데이터(High Quality Data, HQD) 913장과 품질이 좋지 않은 데이터(Poor Quality Data, PQD) 1,577장으로 분류되었다. PQD 데이터는 Fig. 1과 같이 이미지 데이터의 밝기, 빛 반사, 블러와 같은 영상 결함 등의 문제를 포함한다.

HQD는 정상 데이터 573장, 비정상 데이터 340장으로 이루어져 있다(Fig. 2). PQD는 정상 데이터 311장, 비정상 데이터 1,266장으로 구성하였다(Fig. 1). 참 값(Ground Truth)은

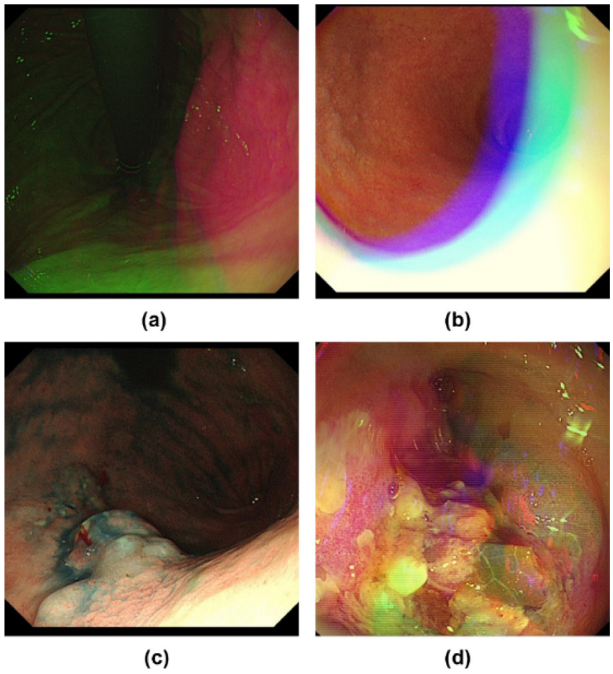


그림 1. 품질이 좋지 않은 데이터의 예시: (a) 정상 어두운 이미지, (b) 정상 빛 반사 이미지, (c) 비정상 흐린 이미지, (d) 비정상 흐린 빛 반사 이미지

Fig. 1. Examples of Poor Quality Data: (a) normal dark image, (b) normal light-reflected image, (c) abnormal blurred image, (d) abnormal blurred and light-reflected image

www.kci.go.kr

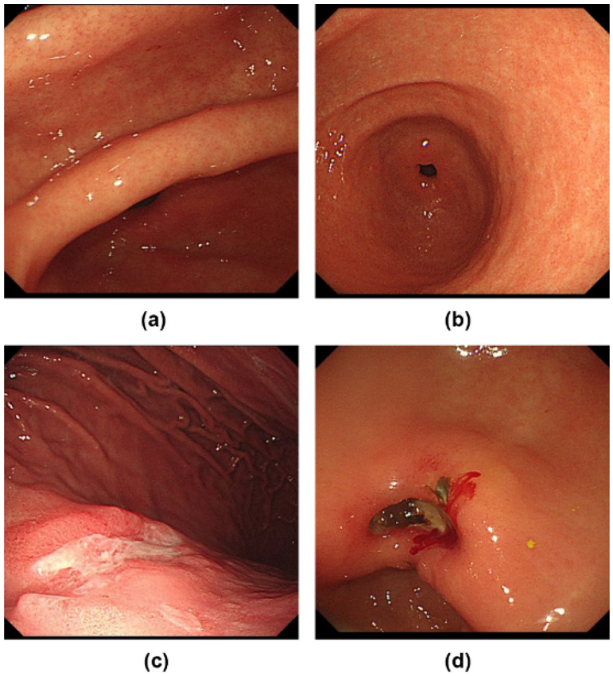


그림 2. 품질이 좋은 데이터의 예시: (a) ,(b) 정상 이미지, (c) 암 이미지, (d) 암 출혈 이미지
Fig. 2. Examples of High Quality Data: (a), (b) normal image, (c) cancer image, (d) cancer bleeding image

전문의가 위 내시경 이미지에서 관심 영역(Region of Interest, ROI)을 지정하였다.

훈련, 검증, 테스트 셋은 무작위로 선정하였으며, 훈련 데이터 셋은 1,224장, 검증 데이터 셋은 406장이고 테스트 셋은 이미지 품질 평가에 사용된 HQD 430장과 PQD 430장으로 구성하였다. 비정상 데이터의 경우 훈련 데이터 셋과 검증 데이터 셋은 각각 888장(암: 320장, 궤양: 287장, 선종: 281장), 294장(암: 106장, 궤양: 95장, 선종: 93장)이며, 정상 데이터의 경우 훈련 데이터 셋과 검증 데이터 셋은 각각 336장, 112장이다.

3. 이미지 품질 평가 알고리즘

위 내시경 이미지 품질에 대한 객관적 평가 방법으로 대표적으로 우수하다고 보고된 브리스큐(Blind/Referenceless Image Spatial Quality Evaluator, BRISQUE), 라플라시안 점수(Laplacian Score), 최대 신호 대 잡음비(Peak Signal-To-Noise Ratio, PSNR)를 사용하였다.

(1) 브리스큐(Blind/Referenceless Image Spatial Quality Evaluator, BRISQUE)

이미지 품질을 분석할 수 있는 BRISQUE라는 지표가 최근 다양한 영상 분야에서 사용되고 있다[17]. BRISQUE는 울림, 흐림 또는 차단과 같은 왜곡 관련 기능을 계산하지 않고 대신 국지적으로 정규화된 휘도 계수를 사용하여 이미지 품질의 총체적 척도인 자연스러움의 가능한 손실을 정량화한다. 2012년 Mittal 등의 연구에 따르면 BRISQUE가 이러한 무참조(Blind/No-Reference, NR) IQA 접근 방식에 대해 매우 경쟁력이 있으며 널리 사용되는 완전 참조(Full-Reference, FR) PSNR 및 구조적 유사성(Structural Similarity Index, SSIM)보다 더 통계적으로 우수하다고 보고하였다. BRISQUE 값은 NR 기준으로 전체 이미지를 분석할 수 있으며, 값이 낮을수록 이미지가 선명하며, 높을수록 흐리다[18].

(2) 라플라시안 점수(Laplacian Score)

Laplacian Score는 2차 미분 연산자로, 이미지의 외곽을 검출하고 블러 세기를 측정하는 데 사용된다[19]. Laplacian Score는 X. He 등이 2006년에 개발하였으며 Locality Perserving Projection과 그 방법의 지역성을 보존하는 힘으로 특징을 평가하는 아이디어를 기반으로 한다[20]. 2018년 Rustam 등의 연구에 따르면 유전자 알고리즘(Genetic Algorithm)과 Laplacian Score를 사용하여 정상과 암 조직 유전자를 분류할 때, Laplacian Score는 86.817%의 정확성을 제공하였다[21]. Laplacian Score는 값이 낮을수록 이미지가 흐리며, 높을수록 선명하다.

(3) 최대 신호 대 잡음비(Peak Signal-To-Noise Ratio, PSNR)

PSNR은 최대 신호 대 잡음비를 의미하며 dB 단위를 사용하고 이미지 손실이 적을수록 높은 값을 가진다[22]. 최근 영상 평가 및 품질 측정에 있어 두루 사용되고 있는 객관적인 측정 방법이다. 이미지 처리 시 두 이미지 간의 차이를 숫자로 나타내는데 픽셀의 0~255 값 중 최댓값인 255에 의미를 두고 두 이미지 간의 차이인 노이즈를 측정한다. 대표적으로 PSNR은 이미지 압축 형식에 대한 평가에서 사용되는데 최대값 255와 원본 이미지, 손상 이미지(압축 이미지)의 MSE 비율에 의해 이미지의 품질 평가가 결정된다. MSE (Mean Squared Error)란 오차의 제곱에 평균을 말한다. 두 이미지가 차이가 없이 동일하다면 MSE가 0이고 PSNR은 50dB 이상 무한하다. 50dB을 원본 이미지로 간주하며 30dB을 기준으로 그 이하의 수치는 상대적으로 품질 수준이 떨어진다고 본다[23].

4. 딥러닝 기반의 병변 검출 알고리즘

이미지 품질 차이에 따라 데이터가 얼마나 정확하게 검출되는지 확인하기 위해, HQD와 PQD를 병변 검출 모델의 테스트 셋으로 사용하여 검출 성능을 비교하고자 하였다. 병변 검출 학습 모델로는 딥러닝 기반 객체 검출 네트워크인 RetinaNet을 사용하였다. RetinaNet은 ResNet-FPN을 Backbone

www.kci.go.kr

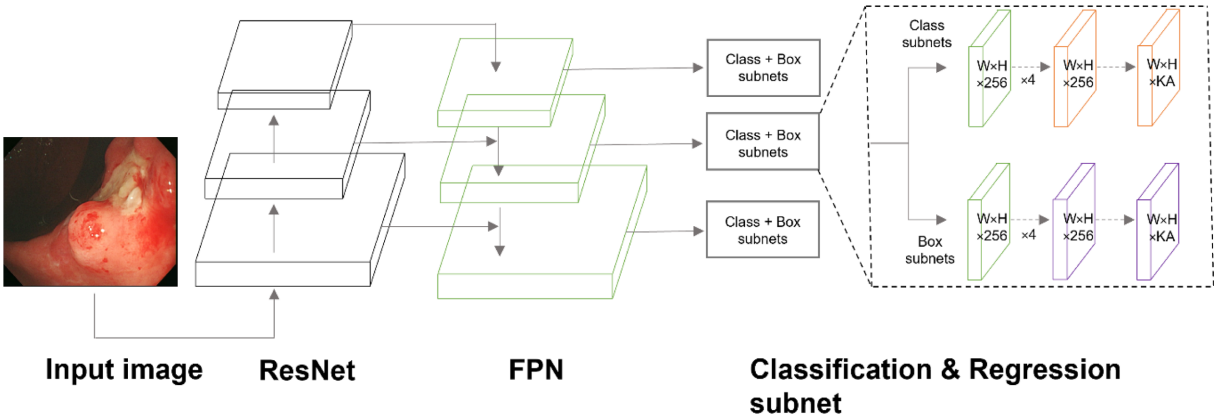


그림 3. RetinaNet의 아키텍처: ResNet-FPN을 백본 네트워크로 사용하고 각 크기에 대한 기능 맵을 사용하여 각 계층에 대한 분류 및 상자 회귀 분석 수행

Fig. 3. Architecture of RetinaNet: Using ResNet-FPN as a Backbone Network, use Feature Maps for each size to perform classification and box regression for each layer

Network로 사용하는데, FPN은 Feature Pyramid Network의 줄임말로 다양한 스케일의 특징 맵을 생성하는 네트워크이다. 일반적인 원 스테이지 디텍터(One Stage Detector) 모델들은 객체와 배경 클래스의 불균형으로 인해 정확도가 낮게 나오는데, RetinaNet은 이를 오분류되는 객체에 대해 더 큰 가중치를 부여하는 Focal Loss 손실 함수를 사용함으로써 이 문제를 해결하였다. RetinaNet은 FPN을 활용하여 작은 크기의 객체를 잘 검출할 수 있다는 장점이 있다[24]. RetinaNet의 Architecture는 Fig. 3을 통해 도식화하였다. 학습 환경은 학습률(Learning Rate) 1e-4, 에폭(Epochs) 200, 배치 사이즈(Batch Size)는 5로 설정하여 진행하였으며 학습에 소요된 시간은 11시간이다.

5. 통계 분석

본 연구에서는 측정된 결과를 검증하기 위하여 BRISQUE와 Laplacian Score 값은 독립 표본 T검정(Independent Samples T-Test), PSNR 값은 단일 표본 T검정(One Sample T-Test)을 이용하여 분석하였다. 검정은 MedCalc 프로그램(Version 19.6.1, MedCalc Software, Ostend, Belgium)을 사용하였다.

검출 모델의 성능 평가 지표로는 정밀도(Precision), 민감도(Recall), F1-Score, 이미지 1장 당 평균 오검출 개수(False Positives Per Image, FPPI)를 사용하였다. 성능 지표 계산식은 식 (1)~(3)에 나타냈다. 이때 TP(True Positive)는 실제 병변이 있는 위 내시경 이미지를 양성으로 분류한 경우인 진양성, FN(False Negative)은 실제 병변이 있는 위 내시경 이미지를 음성으로 분류한 경우인 위음성, FP(False Positive)는 정상 위 내시경 이미지를 양성으로 분류한 경우인 위양성을 의미한다. 병변 검출 모델의 객관적인 평가를 위해 자유 반응 수신자 조작 특성 곡선(Free Response Operating Characteristic Curve, FROC Curve)을 이용하여 성능을 그래프로 나타냈다. FROC 곡선은 병변의 분류 기준이 되는 임계 값 0~1 사이의 모든 값에 대하여 Recall와 FPPI를 계산해서 만들어내는 곡선이다. FROC 곡선을 통해 구축한 모델의 병변 검출 성능을 확인할 수 있으며, Recall 값이 높아지면 모델의 정확도가 높다고 판단할 수 있다[25-26].

$$\text{Precision}(\%) = \frac{TP}{TP+FP} \times 100 \quad (1)$$

$$\text{Recall}(\%) = \frac{TP}{TP+FN} \times 100 \quad (2)$$

$$\text{F1-Score}(\%) = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

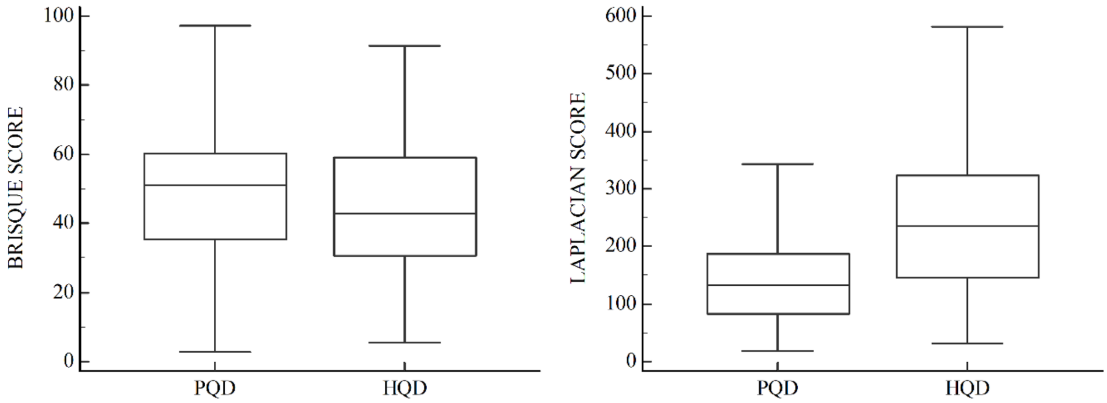
III. 연구 결과

1. BRISQUE, Laplacian Score, PSNR 알고리즘을 사용한 PQD, HQD 그룹 품질 비교 분석

Fig. 4에 BRISQUE, Laplacian Score 알고리즘을 사용하여 PQD, HQD 그룹 이미지 품질을 평가한 값을 그림 상자로 나타냈다. 집단 통계량을 보면 BRISQUE 알고리즘을 활용하여 위 내시경 이미지 품질 평가한 결과, PQD는 48.86±18.16, HQD는 44.08±20.27의 수치를 보였다. Laplacian Score 알고리즘을 활용하여 품질 평가한 결과, PQD는 222.69±300.13, HQD는 280.44±230.16의 수치를 보였다. PSNR 알고리즘을 이용하여 PQD, HQD 그룹 이미지 품질을 비교한 결과, PSNR은 10.62±1.76dB의 수치를 보였다.

독립 표본 T검정을 이용하여 분석한 결과, BRISQUE와 Laplacian Score 값은 유의미한 차이($p<0.05$)가 있었다. 단일 표본 T검정을 이용하여 분석한 결과, PSNR 값은 유의미

www.kci.go.kr



(a) (b)

그림 4. BRISQUE, Laplacian Score 알고리즘을 사용하여 품질 평가 값을 시각화한 그림 상자: (a) BRISQUE (b) Laplacian Score
Fig. 4. Box plot visualizing Poor Quality Data, High Quality Data quality assessment values using BRISQUE, Laplacian Score algorithm: (a) BRISQUE (b) Laplacian Score

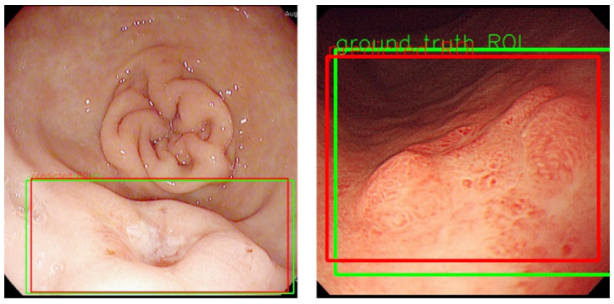
표 1. 알고리즘을 사용한 품질 평가 집단 통계량 및 T검정 결과
Table 1. Quality Assessment Statistics and T Test Results Using Algorithms

	Group	N	Mean ± SD (95% CI)	p
BRISQUE	PQD	430	48.86 ± 18.16 (47.15~50.56)	<0.05
	HQD	430	44.08 ± 20.27 (42.89~46.73)	
Laplacian Score	PQD	430	222.69 ± 300.13 (194.50~250.88)	<0.05
	HQD	430	280.44 ± 230.16 (258.62~302.25)	
PSNR(dB)		860	10.62 ± 1.76 (10.45~10.79)	<0.05

하게(p<0.05) 나타났다(Table 1).

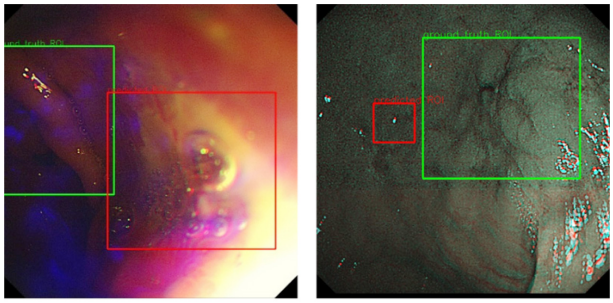
2. 품질에 따른 병변 검출 성능 비교 분석

Fig. 5는 HQD에서 병변을 검출한 결과이고, Fig. 6는 PQD에서 병변 검출을 실패한 결과이다. 실제 관심영역은 초록색, 예측된 관심영역은 빨간색으로 표시하였다.



(a) (b)

그림 5. 품질이 좋은 데이터 병변 검출 결과 예시: 초록색 상자=참값 ROI, 빨간색 상자=예측된 ROI
Fig. 5. Examples of High Quality Data lesion detection result: The green box is the Ground Truth ROI and the red box is the predicted ROI



(a) (b)

그림 6. 품질이 좋지 않은 데이터 병변 검출 결과 예시: 초록색 상자=참값 ROI, 빨간색 상자=예측된 ROI
Fig. 6. Examples of Poor Quality Data lesion detection result: The green box is the Ground Truth ROI and the red box is the predicted ROI

PQD에서 검출 성능은 Precision 92.48±4.98%, Recall 53.80±11.77%, F1-Score 66.82±9.07%, FPPI 0.02±0.02로 나타났다. 반면, HQD에서는 Precision 85.08±9.57%, Recall 72.14±5.48%, F1-Score 77.42±3.36%, FPPI 0.07±0.06로 나타났다.

독립 표본 T검정을 이용하여 분석한 결과, 각 성능 평가 항

www.kci.go.kr

표 2. 품질에 따른 RetinaNet의 성능 평가 비교 표
Table 2. Comparison table of RetinaNet's performance assessment by data quality

	Precision(%)	Recall(%)	F1-score(%)	FPPI	p
PQD	92.48 ± 4.98	53.80 ± 11.77	66.82 ± 9.07	0.02 ± 0.02	<0.05
HQD	85.08 ± 9.57	72.14 ± 5.48	77.42 ± 3.36	0.07 ± 0.06	

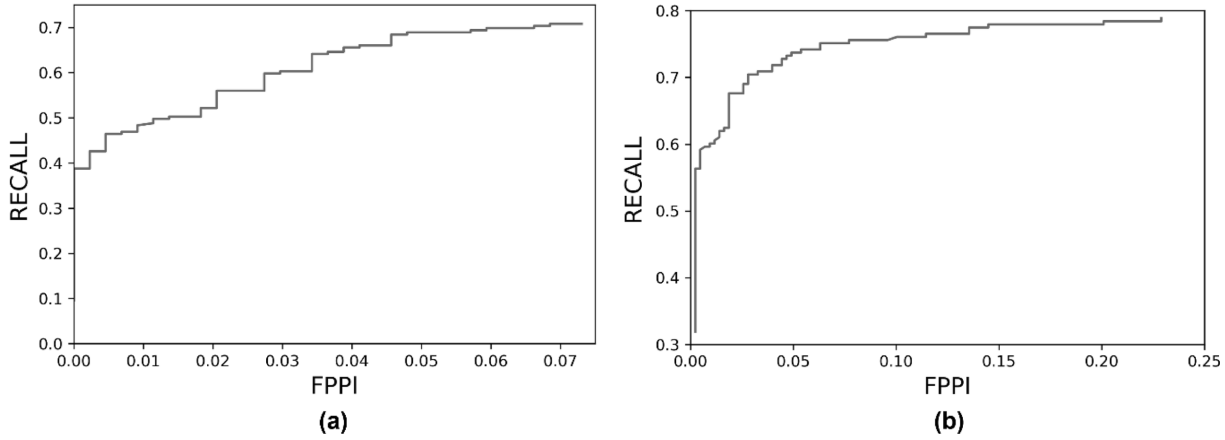


그림 7. 품질에 따른 병변 검출 모델 FROC Curve: (a) 품질이 좋지 않은 데이터 (b) 품질이 좋은 데이터
Fig. 7. Lesion Detection Model FROC Curve Based on Data Quality: (a) Poor Quality Data (b) High Quality Data

목 별로 PQD와 HQD 간의 유의미한 차이(p<0.05)가 있었다(Table 2). 이를 통해 HQD가 Recall, F1-score 성능 평가 지표에서 PQD보다 높은 성능을 보이는 것을 확인하였다. 검출 모델의 객관적인 평가를 위해 Fig. 7에 품질에 따른 병변 검출 모델 FROC Curve를 나타냈다.

IV. 고찰 및 결론

본 연구에서는 IQA 알고리즘을 통해 내시경 이미지에서 품질 차이를 정량적으로 수치화하고 품질 차이에 따라 병변 검출 모델에서 검출이 얼마나 차이가 나는지 확인하였다. 위 내시경 이미지의 품질 비교를 위해 IQA 방법인 BRISQUE, Laplacian Score, PSNR 알고리즘을 사용하여 PQD와 HQD의 품질을 수치화하였다. 그 결과, PQD와 HQD 간의 BRISQUE와 Laplacian Score 모두 유의미한 차이(p<0.05)를 보였다. 의료 영상은 적은 품질 손실에서도 심각한 오진을 발생시킬 수 있으므로 PSNR 값이 30dB 이상의 품질 수준이 나와야 한다[22]. 본 연구 결과에서 PSNR은 10.62±1.76dB로 PQD는 HQD에 비해 손실이 심하며 병변 검출 성능 저하 결과를 야기하였다.

품질에 따른 병변 검출 성능을 평가하기 위해 딥러닝 모델인 RetinaNet을 통해 위 내시경 이미지를 학습하여 검출 모델을 만들었다. 검증 과정에서 HQD의 경우 F1-Score 95.91±4.91%, PQD의 경우 F1-Score 74.02±1.76%를 보였다. 모델의 테스트 결과는 검증보다 다소 낮았지만 HQD의 경우 F1-Score 77.42±3.36%로 높은 검출 성능을 보였고, PQD의 경우 F1-Score 66.82±9.07%의 성능을 보이며 유의미한 차이(p<0.05)를 나타냈다. 이와 같은 연구 결과를 기반으로 노이즈가 적고 밝기 값이 적절한 품질이 좋은 위 내시경 이미지를 선별한다면 CADx 시스템의 병변 검출 성능을 향상시킬 수 있을 것으로 예상된다.

본 연구는 BRISQUE, Laplacian Score, PSNR 등 다양한 이미지 품질 평가 알고리즘을 사용하여 품질을 정량적으로 수치화했다는 점에서 의의가 있다. 그리고 이미지 품질 평가에서 유의미한 차이가 있었던 두 개의 데이터 집단을 가지고 병변 검출 모델에서 성능을 확인함으로써 실제로 위 내시경 데이터의 품질이 병변 검출에 영향을 미친다는 결과를 보인 첫 시도이다. 그럼에도 불구하고, 본 연구의 한계점은 이미지 품질 평가 연구에서 BRISQUE, Laplacian Score 알고리즘이 서로 상반된 의견을 표하거나 수치 상 품질이 좋은 이미지더라도 육안으로 확인 시 품질이 좋지 않다고 의심되는 데이터가 포함되었을 수 있다는 점이다. 또한, PSNR 알고리즘은 동일한 위치에서 촬영된 두 영상을 비교해야 하는데 PQD와 HQD는 내시경의 움직임으로 인해 완전히 동일하지 않을 수 있다는 한계가 있다. 실제 임상에서 동일하게 촬영하기는 쉽지 않으며 비교를 위해 한 영상을 인위적으로 압축하는 것보다 완전히 동일하지 않더라도 PSNR 값의 경향성을 보는 것이 낫다고 판단하여 연구를 진행하였다. 따라서 향후 전문의와 논의하여 PSNR 값 비교가 가능한 데이터를 수집하여 추가

www.kci.go.kr

연구를 진행해야 할 것으로 보인다. 품질에 따른 병변 검출 성능 비교 연구 결과에서 PQD에 비해 HQD의 FPPI가 높게, Precision이 낮게 나타났다는 한계가 있었다. HQD의 경우 PQD에 비해 선명한 해부학적 구조와 경계가 병변 검출에 도움이 되지만 오검출 발생에도 영향을 미쳤을 가능성이 있다. 또한, 병변 검출 모델의 테스트 셋으로 사용한 HQD와 PQD의 이미지가 서로 동일하지 않기 때문에 품질 외의 외부적 요소가 성능 평가에 개입하여 HQD의 FPPI, Precision 결과에 영향을 미쳤을 것으로 파악된다. 본 연구에서 내시경 이미지 품질에 따라 병변 검출에 영향을 미친다는 것은 보였으나 검출에 영향을 미치는 데이터를 선별할 품질 기준은 제시하지 못하였다. 따라서 추가 연구를 통해 품질에 따른 데이터 선별 기준을 마련하고 BRISQUE와 Laplacian Score의 정량적 수치를 이용하여 전문의 육안으로 PQD와 HQD 데이터를 분류 및 평가해야 할 것으로 보인다.

앞서 설명한 연구 한계점을 추후의 연구를 통해 보완하고 본 논문에서 제안한 BRISQUE, Laplacian Score, PSNR 등 IQA 알고리즘의 활용으로 품질에 따른 데이터 선별 기준을 세워 학습에 좋지 않은 데이터를 선별한다면 딥러닝 기반 위내시경 병변 검출 보조 시스템의 정확도가 더욱 향상될 것으로 기대된다.

References

[1] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Pieros M, Bray F. Global cancer observatory: cancer today. *Lyon, France: International Agency for Research on Cancer*. 2018;3(20):1-2.

[2] https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed on 3 Feb 2022. cancer.

[3] Choi ES, Jeon WH, Sohn HJ. Clinical study of early gastric carcinoma. *Korean J Gastroenterol*. 1995;27:31-38.

[4] Park IS, Lee YC, Kim WH, Noh SH, Lee KS, Kim H. Clinicopathologic characteristics of early gastric cancer in Korea. *Yonsei Medical Journal.* 2000;41(5):607-614.

[5] Song HT, Kim CD, Ryu HS, Hyun JH. *A clinical study of early gastric cancer.* Korean J Gastroenterol. 1994;26:789-799.

[6] National Cancer Center. *Cancer Facts and Figures 2014*. 2014; 18-27.

[7] Lee SA, Cho HC. A New Image Augmentation Method for Improving Computer-aided Diagnosis System Performance. *The transactions of The Korean Institute of Electrical Engineers*. 2021;70(1):102-107.

[8] Belle A, Kon MA, Najarian K. Biomedical informatics for computer-aided decision support systems: a survey. The Scientific World Journal. 2013;2013:1-8.

[9] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*. 2017;2(4):230-243.

[10] Miller RA. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. *Journal of the American Medical Informatics Association.* 1994;1(1):8-27.

[11] Kim DH, Cho HC. Deep Learning based Computer-aided Diagnosis System for Gastric Lesion using Endoscope. *The transactions of The Korean Institute of Electrical Engineers*. 2018;67(7):928-933.

[12] Lee SA, Cho HC, Cho HC. Computer-aided Diagnosis System for Abnormalities Classification in Gastric Endoscopy Images using Machine Learning. *The Transactions of the Korean Institute of Electrical Engineers*. 2020;69(1):107-113.

[13] Park SH, Cho YB. An Effect of Genetic Algorithm for Creating a Dataset. *Journal of Digital Contents Society*. 2019;20(1): 127-133.

[14] Lee JM, Ham GW, Bae KH, Park HK. Detection Ability of Occlusion Object in Deep Learning Algorithm depending on Image Qualities. *Journal of the Korean Association of Geographic Information Studies*. 2019;22(3):82-98.

[15] Yoon MS, Oh JH, Kang BG, Im JU, Kwon GT, Han DG. Effects of Image Extensions in Chest X-ray Deep Learning-Based Pneumothorax Detection Studies. *Green Collection of the Korean Society of Emergency Medicine's Academic Conference*. 2020;2020(2):75.

[16] Kim YJ, Kim KG. Development of an Optimized Deep Learning Model for Medical Imaging. *Journal of the Korean Society of Radiology*. 2020;81(6):1274-1289.

[17] Cho YK, Lee Y, Lee K. Quality Evaluation of Ultrasonographic Equipment Using an ATS-539 Multipurpose Phantom in Veterinary Medicine. *Journal of Veterinary Clinics.* 2022;39(3):114-120.

[18] Mittal A, Moorthy AK, Bovik AC. No-reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing*. 2012;21(12):4695-4708.

[19] Bansal R, Raj G, Choudhury T. Blur image detection using Laplacian operator and Open-CV. in *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*. 2016;63-67.

[20] He X, Cai D, Niyogi P. Laplacian score for feature selection, in proceeding of Advances in Neural Information Processing Systems. 2006;18:507-514.

[21] Rustam Z, Primasari I, Widya, D. Classification of cancer data based on support vectors machines with feature selection using genetic algorithm and laplacian score. In *AIP Conference Proceedings*. 2018;2023(1):020234.

[22] Kang KS, Lee, J-H. PSNR appraisal of MRI image. *Journal of the Korean Society of Radiology*. 2009;3(4):13-21.

[23] Jung HK. Fundamentals of multimedia chapter 8, Lossy compression Algorithms. Hallym University. 2008.

[24] Lin TY, Goyal P, Girshick R, He K, Dollr P. Focal loss for dense object detection. in *Proceedings of the IEEE international conference on computer vision*. 2017;2980-2988.

[25] Lee JH. Radiological feature estimation and risk assessment of mammographic mass and density by computer-aided analysis. Department of Medical Science The Graduate School, Yonsei University. 2004;3-50.

[26] Cheng HD, Cui M. Mass lesion detection with a fuzzy neural network. *Pattern Recognition*. 2004;37(6):1189-1200.

www.kci.go.kr