

자궁경부 영상에서의 라디오믹스 기반 판독 불가 영상 분류 알고리즘 연구

김고은¹ · 김영재¹ · 주 웅³ · 남계현⁴ · 김수녕⁵ · 김광기^{1,2*}

¹가천대학교 의용생체공학과, ²가천대학교 길의료재단 융합의과학과, ³이화여자대학교 산부인과,
⁴순천향대학교 산부인과, ⁵엔티엘의료재단 R&D센터

A Radiomics-based Unread Cervical Imaging Classification Algorithm

Go Eun Kim¹, Young Jae Kim¹, Woong Ju³, Kye Hyun Nam⁴,
Soonyung Kim⁵ and Kwang Gi Kim^{1,2*}

¹Department of Biomedical Engineering, Gachon University

²Department of Health Sciences and Technology, Gil Medical Center, Gachon Advanced Institute for Health Sciences and Technology (GAIHST), Gachon University

³Department of Obstetrics & Gynecology, Ewha Womans University Seoul Hospital

⁴Department of Obstetrics & Gynecology, Soonchunhyang University, Bucheon Hospital

⁵R&D Center, NTL Medical Institute

(Manuscript 1 September received 27 October 2021 ; revised 2021 ; accepted 29 October 2021)

241

Abstract: Recently, artificial intelligence for diagnosis system of obstetric diseases have been actively studied. Artificial intelligence diagnostic assist systems, which support medical diagnosis benefits of efficiency and accuracy, may experience problems of poor learning accuracy and reliability when inappropriate images are the model's input data. For this reason, before learning, We proposed an algorithm to exclude unread cervical imaging. 2,000 images of read cervical imaging and 257 images of unread cervical imaging were used for this study. Experiments were conducted based on the statistical method Radiomics to extract feature values of the entire images for classification of unread images from the entire images and to obtain a range of read threshold values. The degree to which brightness, blur, and cervical regions were photographed adequately in the image was determined as classification indicators. We compared the classification performance by learning read cervical imaging classified by the algorithm proposed in this paper and unread cervical imaging for deep learning classification model. We evaluate the classification accuracy for unread Cervical imaging of the algorithm by comparing the performance. Images for the algorithm showed higher accuracy of 91.6% on average. It is expected that the algorithm proposed in this paper will improve reliability by effectively excluding unread cervical imaging and ultimately reducing errors in artificial intelligence diagnosis.

Key words: Cervical cancer, Radiomics, Laplacian variance, Euclidean distance, ResNet-50

I. 서 론

*Corresponding Author : Kwang Gi Kim

Department of Biomedical Engineering, Gil Medical Center,
Gachon University, College of Medicine, 21, Namdong-daero 774
beon-gil, Namdong-gu, Incheon, 21565, Republic of Korea.

Tel: +82-32-458-2770

E-mail: kimkg@gachon.ac.kr

이 연구는 중소벤처기업부에서 지원하는 기술 개발 프로그램(S2797147)
과 가천대 길병원(FRD2019-11-02(3))의 지원을 받아 수행되었습니다.

자궁경부암은 여성에게 생기는 악성종양 중 유방암 다음
으로 두 번째로 흔한 질환이며 저개발국으로 갈수록 유병률과
사망률이 높다[1]. 질 확대경 검사를 통한 생검(Colposcopy), 자
궁경관 내 소파술, 자궁경부 세포진 검사(PAP smear) 등의 기
존 자궁경부암 검진 방법은 편리하고 비용이 적다는 장점이

있지만, 임상에서의 훈련도나 진단 방식에 따라 분석 결과가 달라질 수 있다. 특히 널리 사용되고 있는 자궁경부 세포진 검사의 경우, 높은 위음성률이라는 한계가 존재한다[2-3].

이러한 한계를 보완하기 위해 자궁경부 진단 보조 시스템으로서 딥러닝 연구가 다수 보고되고 있다[4]. 2021년 Chandran 등은 자궁경부 분류에 기존의 기계학습 알고리즘과 함께 부스팅 방식의 분류 모델을 결합한 CNN 기반의 앙상블 딥러닝 모델을 이용하여 평균 92.4%의 높은 분류 정확도를 얻었다[5]. 또한 2020년에 자궁경부 컬러 영상의 입력 형태에 따른 분류 성능 비교 연구에서는, 영상의 중심에 맞춰 양쪽을 자르는 방식과 원본 영상의 너비에 맞춰 빈 곳을 채우는 방식으로 입력 형태 차이를 두어, 각각 94%, 83.4%의 분류 성능 차이를 보고하였다[6].

딥러닝 의료 진단 보조 시스템은 충분한 학습데이터의 양과 실험 결과의 신뢰도가 보장되면, 진단의 정확도와 효율성을 높인다. 그러나 인공지능의 학습 방식 특성상, 입력 데이터에 대한 높은 의존도를 가진다. 입력 데이터로 학습에 부적합한 데이터가 사용되었을 경우, 잘못된 진단으로 이어질 수 있으므로 입력 데이터 품질을 검증하는 과정이 필수적으로 선행되어야 한다[7]. 그러나 최근 보고된 연구 논문들에 따르면, 판독 불가능한 데이터를 미리 제외하여 딥러닝의 분류 정확도를 보장하는 연구는 아직 보고된 바가 없다.

이에 본 연구에서는 딥러닝 학습 및 판독에 적합한 자궁경부 영상 분류를 위하여, 학습 전 판독 불가 영상을 미리 제외하는 알고리즘을 제안한다. 라디오믹스(Radiomics)를 기반으로 영상 밝기, 영상의 흐릿한 정도를 화소값으로부터 추출하여 판독 불가 영상의 분류 지표로 정했다. 또한 자궁경부 영역의 주요 판독 영역이 되는 경관 외구(External Orifice)가 적절하게 촬영된 영상과의 유사도를 분류 지표로 하였다. 이때 이미지 유사도 측정 방법은 유클리드 거리(Euclidean Distance)로 하였다. 분류 지표에 대한 임계값을 정한 후, 임계값의 범위에서 벗어나는 영상을 판독 불가 영상으로 제외하였다. 본 논문에서 제안한 알고리즘의 성능 평가를 위해 알고리즘을 적용한 영상과 그렇지 않은 영상을 각각 딥러닝 분류 모델에 학습시킨 후 성능을 비교하였다.

II. 데이터 및 연구 방법

1. 데이터 수집

본 연구에서는 자궁경부 영상진단 장비(엔티엘 헬스케어, Dr. Cervicam)로 촬영된 자궁경부 영상 총 2,257장을 연구 데이터로 수집하였다.

데이터는 기관생명윤리위원회(Institutional Review Board, IRB)의 승인(EUMC 2015-10-004)을 받아, 모두 비식별화되었다[8]. 자궁경부의 명확한 분류를 위해, 자궁경부 암화

단계에서 극단적 병기 단계에 속하는 영상으로 데이터를 구성하였다. 이 과정에서 WHO 분류 시스템을 참고하였다. WHO 분류 시스템에 따르면 자궁경부 암화 단계는 정상(Normal), 비정형 상태(Atypical), 저등급 이형성증(Low grade Squamous Intraepithelial Lesion, LSIL), 고등급 이형성증(High grade Squamous Intraepithelial Lesion HSIL), 침습성 자궁경부암(Invasive cervical cancer)으로 진행되는 자궁경부암 발생 단계를 의미한다[9].

판독 가능한 영상은 정상 등급 병기 단계에 속하는 데이터를 정상 데이터, 고등급 편평 상피 내 병변 단계(HSIL)에 속하는 데이터를 병변 데이터로 각 1,000장씩 총 2,000장을 수집하였다. 판독 불가 영상은 총 257장으로 수집하였다. 영상은 24bit의 JPG 파일의 파일 속성을 가지며, 모델 학습 전 256×256 크기로 조정하였다.

2. 영상 전처리

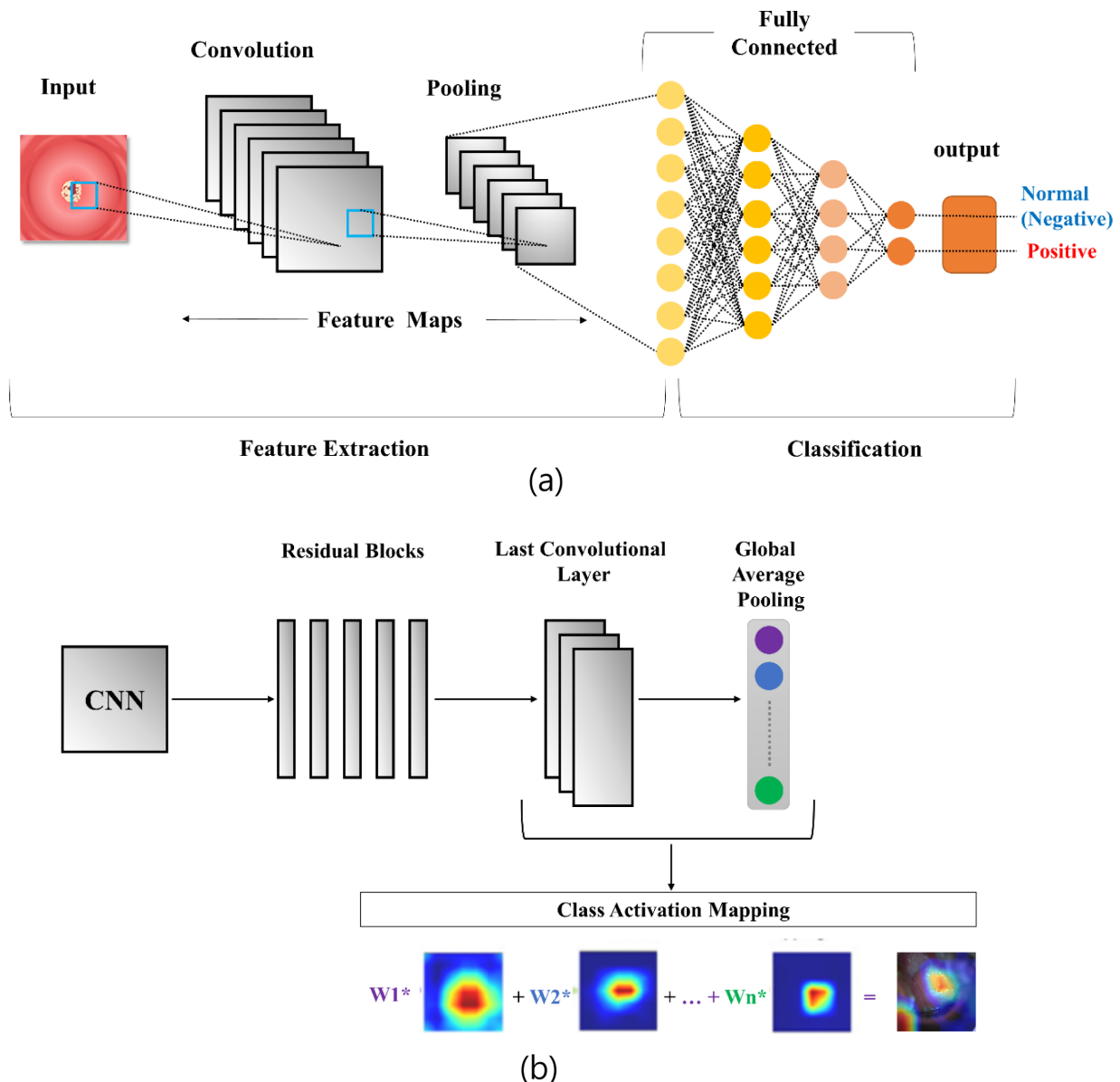
수집한 데이터에서 판독 불가 영상은 전문가에 의해 영상 데이터의 밝기, 빛 반사와 같은 영상 결함 등의 문제로 자궁경부 식별이 불가능하다고 판단되어 판독 불가 영상으로 라벨링되었다. 연구 데이터는 2048×1536 , 1280×960 , 1504×1000 의 다양한 크기와 가로, 세로 비율을 가진다. 일부 데이터의 경우, 이미지의 우측 하단에 수집된 영상 정보에 대한 태그(tag)가 있다. 이는 자궁경부가 아닌 영역을 자궁경부라고 잘못 인식하게 할 가능성이 있으므로 제거해야 한다. 또한 주요 판독 영역이 되는 자궁 경관이 영상 중앙에 있는 영상과 그렇지 않은 영상이 존재한다. 이 경우에 같은 인덱스의 영상이라도 이미지를 자르는 방식(crop)에 따라 화소의 평균 밝기값 등이 다르게 나타날 수 있다[10].

본 연구에서는 판독에 불필요한 영향을 주는 태그를 제거하고, 모든 영상의 가로, 세로 비율을 동일하게 하기 위하여, 가로, 세로 길이 중 짧은 것을 이미지의 한 변으로 한 후 영상 중심에 맞춰 크롭하였다.

3. 태그(tag) 영향 분석(Class Activation Mapping, CAM)

자궁경부 영상의 태그가 모델 학습에 미치는 영향을 분석하기 위해 CAM(Class Activation Mapping)을 영상에 나타내었다.

CNN(Convolutional Neural Network)은 Input, Convolution Layer, Pooling Layer, Fully-connected Layer 단계로 구성되는 모델로, Convolution Layer, Pooling Layer에서 이미지의 특징 맵(feature map)이 생성된다[11]. CAM(Class Activation Mapping)은 특징 맵 생성 과정에서 CNN이 특정 클래스의 이미지를 해당 클래스로 인식하게 한 이미지 내의 영역을 보여주는 가중치 값이다. 자궁경부 영상의 판독 결과를 신뢰하기 위해서는 CAM이 자궁경부 경관 외구에



243

그림 1. 태그 영향 분석에 사용된 모델의 네트워크 구조, (a) CNN, (b) CAM

Fig. 1. Network Structure used in analyzing Tag, (a) CNN Structure, (b) CAM Structure

정확하게 나타나야 한다[12]. 따라서 분류 모델이 자궁경부 영상을 정상 또는 병변으로 예측한 영역인 CAM을 나타내 모델이 자궁경부라고 예측한 영역을 확인하고자 하였다. 태그 유무에 따라 CAM을 나타내, 태그가 모델 학습에 미치는 영향을 확인하였다. 그림 1은 CNN과 CAM의 네트워크 구조를 나타낸 그림이다.

4. 판독 불가 영상 자동 분류 알고리즘

그림 2는 본 논문에서 판독 불가한 영상을 자동 분류하기 위하여 진행한 연구 방법의 흐름도이다. 밝기, 블러, 영역 수치에서의 판독 가능한 값은 판독 가능한 데이터의 값의 전

체 범위를 분모로 하여 0(0%)부터 1(100%)까지 정규화한 값이다.

(1) 알고리즘의 분류 지표

자궁경부 상피세포의 미세한 형태학적 변화는 판독 결과에 주요한 영향을 미치기 때문에[13-14]. 밝기, 화질 등 화소값의 차이가 판독 가능성에 차이를 줄 수 있다. 판독 가능성에 따라 영상을 분류하는 지표는 그림 2와 같이 영상의 밝기, 흐린 정도, 자궁경부의 경관 외구 부위가 적절하게 촬영된 정도로 정하였다. 그림 3의 (a)~(d)는 제안한 알고리즘에 따라 분류된 판독 불가 영상을 나타낸 것이다. 그림 3의 (e)~(f)는

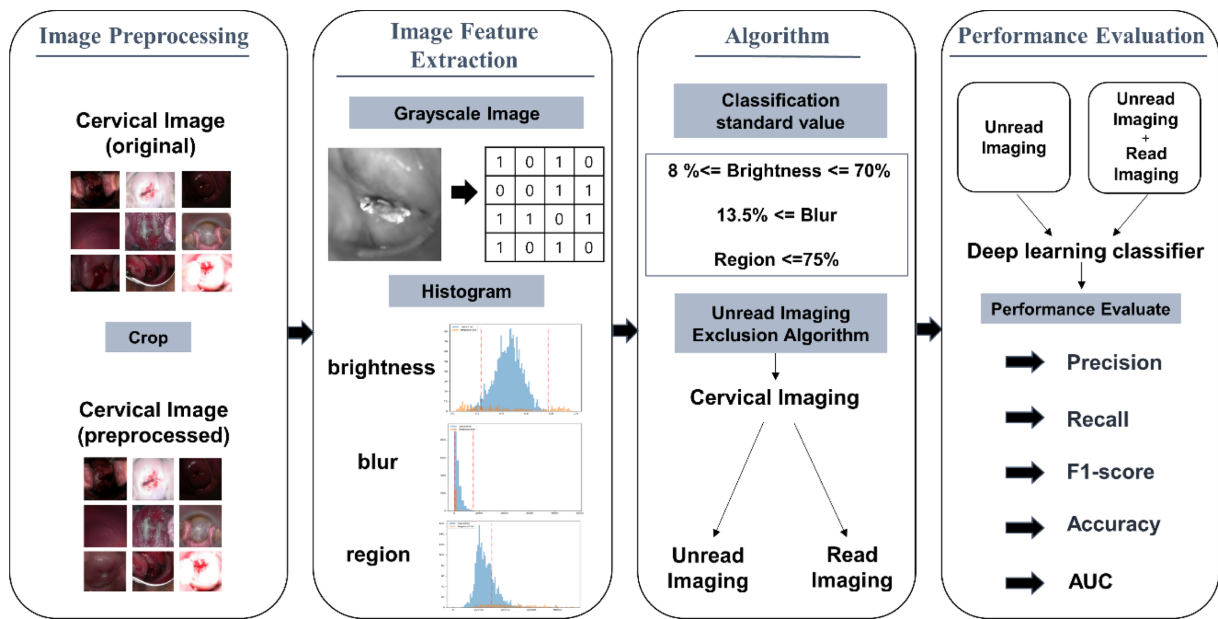


그림 2. 판독 불가 영상을 분류하기 위한 연구방법의 흐름도
Fig. 2. The flowchart illustration of the study method

태그 유무에 따라 CAM이 자궁경부로 예측한 영역을 나타낸 그림이다.

판독 불가 영상의 첫 번째 분류 지표는 영상의 밝기값으로 설정하였다. 육안으로 판독이 어려울 정도의 영상 밝기에 이상이 있는 데이터에 대해 화소값을 추출하여 밝기값을 분석하였다[15]. 영상을 이루는 각 화소의 밝기를 합산하여 화소의 개수로 나누어 각 영상의 평균 밝기값을 계산하였다. 히스토그램 그래프를 사용하여 판독 가능한 영상과 판독 불가 영상 두 데이터 셋의 평균 밝기값을 동시에 시각화한 후, 값의 분포 범위를 비교하였다. 판독 가능한 밝기값의 임계 범위는 t-검정, 히스토그램 분석방법을 사용하여 정의하였다.

두 번째 분류 지표로는 블러(blur)로 설정하였다. 본 연구에서는 영상의 흐린 정도를 블러로 정의하여, 라플라시안 연산자(Laplacian operator)로 해당 값을 측정하였다. 값이 낮을수록, 영상이 흐리며, 높을수록 영상이 선명하다. 라플라시안 연산자는 2차 미분 연산자로, 이미지의 외곽을 검출하고 블러 세기를 측정하는 데 사용된다. Raghav Bansal의 연구에 따르면, 라플라시안 연산자를 이용하였을 때, 높은 정확도로 영상 데이터의 평균적인 블러 세기를 측정하였다[16]. 이러한 연구 결과에 따라 본 연구에서는 라플라시안 연산자를 이용하여 블러 세기를 구하였다. 임계값의 범위는 밝기값과 동일한 방식으로 t-검정과 히스토그램 분석을 사용하여 판독 가능한 데이터의 블러 값 분포에서 상위 13%에 해당하는 수치로 설정하였다. 라플라시안 연산자에 의한 블러의 정도는 상대적이므로, 판독 가능한 데이터의 블러 값 전체범위를 분모로 설정하였을 때, 밀집도가 높게 나타난 13%의 값까지

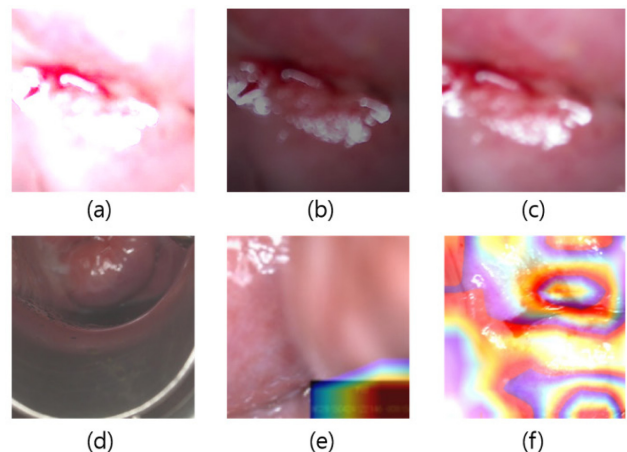


그림 3. 판독 불가 영상과 태그가 있는 영상 예시, (a) 높은 밝기의 영상, (b) 낮은 밝기의 영상, (c) 흐릿한 영상, (d) 자궁경부 영역이 영상에서 적절하게 촬영되지 않은 영상, (e) CAM이 태그를 자궁경부 영역으로 잘못 인식한 영상, (f) 태그가 없는 CAM이 정확히 자궁경부를 예측한 영상

Fig. 3. Examples of Unread Cervical Imaging and Tagged Image, (a) Image of high brightness, (b) Image of low brightness, (c) blurred Image, (d) Image of inappropriately photographed cervical region, (e) Tagged Image of CAM misreading tag into cervical region, (f) Image of CAM reading cervical region right

판독 가능한 범위로 설정하였다.

자궁경부의 경관 외구 영역이 영상의 중심에서 벗어나 판독이 어려운 영상을 세 번째 분류 지표로 정하였다. 유클리드 거리(Euclidean Distance)를 이용하여 판독 불가 영상과 판독 가능한 영상과의 이미지 유사도를 계산하여 값을

분석하였다.

$$d(x,y)=\sqrt{\sum_{i=1}^n(y_i-x_i)^2} \quad (1)$$

식 (1)은 유클리드 거리 계산 식이다. x 와 y 는 직교 좌표계에서 거리를 측정하고자 하는 두 점을, $d(x, y)$ 는 계산된 유클리드 거리를 의미한다. 자궁경부의 경관 외구가 영상의 중앙에 정확하게 위치한 영상을 템플릿(Template) 이미지로 정했다. 템플릿 이미지와 판독 불가 영상과의 이미지 유사도를 유클리드 거리를 이용하여 측정했다. Wang의 연구에 따르면, 유클리드 거리를 이미지 유사도 측정 방법으로 사용하여 딥러닝 모델 정확도를 평가하였을 때, 같은 방법을 적용하지 않은 다른 모델과 비교하였을 때 가장 높은 성능을 낸 것을 입증한 바 있다[17].

유클리드 거리의 값이 작을수록 템플릿 이미지와의 유사도가 높고, 클수록 유사도가 낮으므로, 동일한 이미지에 대해서는 값이 0이다. 앞서 정한 분류 지표의 임계값 설정 방식과 마찬가지로 t-검정과 히스토그램 분석을 사용하여 경험적으로 판독 가능한 범위를 정의하였다. t-검정은 SPSS 통계 분석 프로그램(IBM SPSS 28.0)을 사용하였다[20].

(2) 알고리즘의 분류 단계

본 연구의 알고리즘은 분류 지표에 대한 임계값의 범위를

바탕으로 판독 불가 데이터를 제외하였다. 데이터 분류의 구체적 단계는 다음과 같다.

첫 번째로, 판독 가능 유무를 확인할 자궁경부 영상을 알고리즘에 입력한다.

두 번째로, 밝기, 블러, 유클리드 거리에 대하여 설정한 임계값을 바탕으로 해당 범위에서 벗어나는 데이터는 모두 판독 불가 데이터로 분류한다. 또한 판독 불가 데이터를 모델의 입력데이터에서 제외할 뿐만 아니라 판독 불가 케이스 중 8가지 경우에 대해서도 데이터를 분류한다. 판독 불가 케이스는 밝기, 블러, 영역 3가지의 분류지표에 대하여 판독 가능한 임계 범위를 벗어나는지, 만족하는지에 따라 두 가지의 경우로 구분되어 총 $2 \times 2 \times 2$ 의 8가지 케이스로 구분되었다.

세 번째로 분류된 판독 불가 데이터는 8가지 원인에 따라 그에 상응하는 메시지가 출력한다. 8가지 분류 지표에 대하여 모두 판독 가능 범위를 만족하는 데이터는 정확한 진단을 위해 전문가에게 직접 판단할 것을 권고하는 메시지를 출력하도록 하였다.

그림 4는 본 논문에서 제안하는 판독 불가 영상 자동 분류 알고리즘에 대한 흐름도이다.

판독 가능 범위를 뜻하는 classification standard value에서 각 지표 값은 판독 가능한 데이터의 값의 범위를 0부터 1까지로 정규화하였을 때 해당하는 값으로 나타낸 것이다.

245

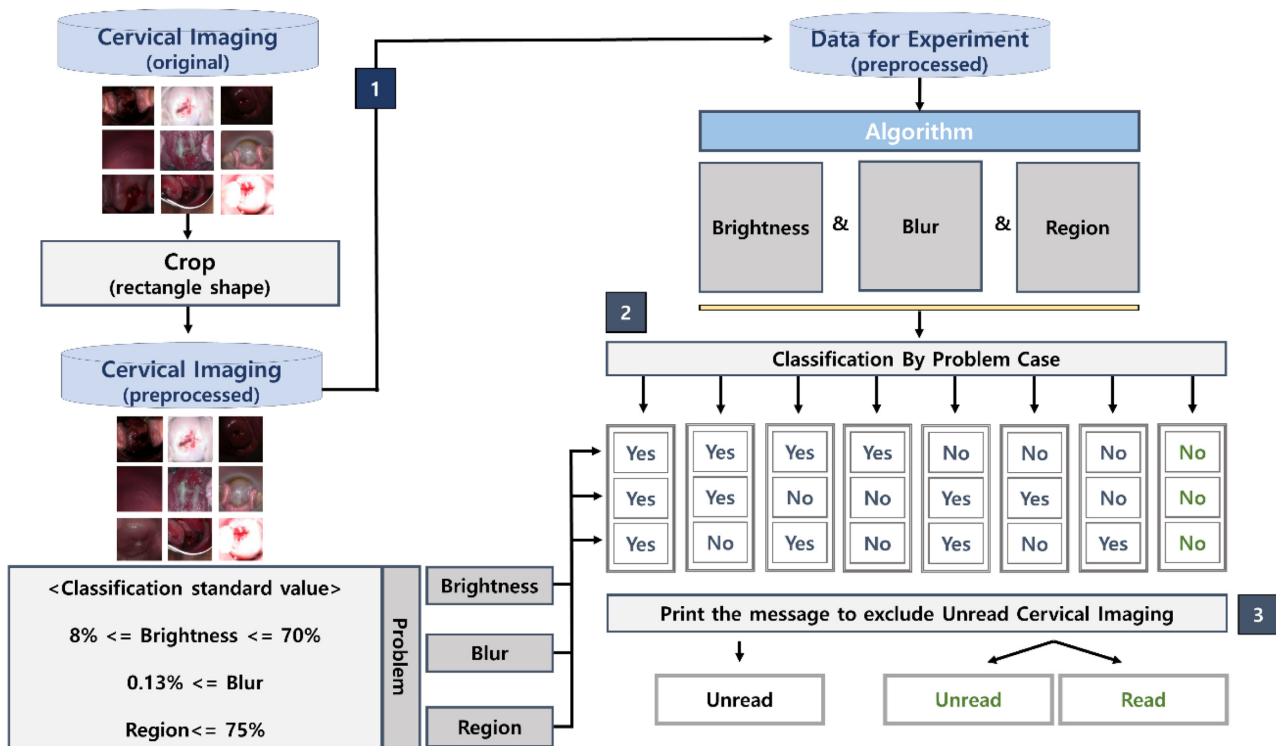


그림 4. 판독 불가 영상 자동 분류 알고리즘의 흐름도

Fig. 4. The flowchart of Automated classification algorithm for Unread Cervical Imaging

밝기(brightness)는 1에 가까울수록 밝고 블러(blur)는 값이 클수록 선명하다. 영역(Region, Euclidean Distance)의 경우 값이 작을수록 자궁경부 판독 영역이 영상에서 잘 나타난다.

(3) 알고리즘의 분류 정확도 평가

본 연구에서 제안한 알고리즘을 통해 판독 가능 유무에 따라 데이터가 정확하게 분류되는지 확인하기 위해, 알고리즘을 적용한 데이터와 적용하지 않은 데이터를 각각 딥러닝 분류 모델에 학습시켜 분류 정확도를 비교하고자 하였다. 학습 모델로는 ResNet-50을 사용하였다. ResNet(Residual neural network)은 지름길(Shortcut) 개념이 적용된 CNN 기반의 분류 모델로, ResNet-50은 50계층으로 구성된 ResNet이다[18]. 기존 CNN 모델과는 다르게 잔차(Residual)가 최소가 되게 학습함으로써 계층이 증가하더라도 학습 에러(Training error)는 감소하는 장점을 가진다[19]. 본 연구에서는 두 차례의 ResNet-50 성능 비교 실험을 진행하였다. 첫 번째 실험에서는 판독 가능한 데이터와 판독 불가 데이터가 분류되지 않은 전체 데이터 셋으로부터, 알고리즘을 거친 데이터만을 선별한 다음, 모델에 학습했다. 두 번째 실험에서는 전체 데이터 셋을 모델에 학습했다. 두 실험의 학습 환경은 배치 사이즈(Batch Size)는 16, 학습 횟수(Epoch)는 300, 학습률(Learning Rate)은 0.0001로 동일하게 설정하였다.

학습 모델의 성능 평가 지표로는 정확도(Accuracy), 정밀도(Precision), 민감도(Recall), F1-score와 AUC를 사용하였다. AUC(area under the curve)는 ROC(receiver operating characteristic) 곡선의 면적을 표현하는 분류 성능 평가 방법이다. 5점 교차 검증(5-fold cross validation)을 통해 각 성능 평가 지표의 평균값을 비교하였다. 성능 지표 계산식은 식 2~5에 나타내었다. 이때 TP(True Positive)는 실제 병변이 있는 자궁경부 영상을 양성으로 분류한 경우인 진양성, FN(False Negative)은 실제 병변이 있는 자궁경부 영상을 음성으로 분류한 경우인 위음성, TN(True Negative)은 정상 자궁경부 영상을 음성으로 분류한 경우인 진음성, FP(False

Negative)는 정상 자궁경부 영상을 양성으로 분류한 경우인 위양성을 의미한다.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{F1-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (5)$$

III. 결 과

1. 판독 불가 영상 자동 분류 알고리즘의 분류 지표 값 분석

표 1의 (a), (b)는 판독 가능한 영상과 판독 불가 영상의 밝기값을 나타낸 표이다. 밝기값의 범위는 0부터 1까지이고, 0에 가까울수록 어둡고 1에 가까울수록 밝다. 표 1에서 볼 수 있듯이, 판독 가능한 데이터 밝기의 평균값은 0.45로 데이터 대부분이 이 조건을 만족하고 최솟값과 최댓값에서 값을 충족하는 데이터의 수가 확연히 줄어드는 것을 볼 수 있다. 반면, 판독 불가 데이터의 밝기값은 특정한 분포 경향 없이 넓게 분포한다. 밝기값의 임계 범위는 통계 분석과 경험적으로 수행하여 판독 가능한 데이터의 사분위 값에서 최소 8%부터 70%까지로 정했다.

그림 5(b)는 x축을 블러 세기, y축을 해당 블러 세기를 가지고 있는 화소 수로 설정하여 영상의 블러값 분포를 나타낸 히스토그램 그래프이다. 그림 5에서 볼 수 있듯이 판독 가능한 영상의 블러값의 경우 최대 767 정도의 수치를 갖는 것에 반해, 판독 불가 영상의 경우 5,000에 가까운 최댓값을 갖고 판독 가능한 영상에 비하여 넓은 범위에 블러값이 분포한다. 블러 세기의 판독 가능한 범위는 통계 분석과 경험적으로 수행하여 사분위 값에서 13.5%부터로 정하였다.

통계 분석과 히스토그램으로 판독 가능한 데이터와 판독 불가 데이터의 유클리드 거리 값의 분포를 나타내, 임계값의 범위는 0%부터 판독 가능한 데이터의 사분위값 중 75%까지로 정

표 1. 판독 가능 여부에 따라 구분된 화소값에 대한 독립 표본 검정 결과값을 나타낸 표

Table 1. Table of independent sample test results for pixel values in Unread Cervical Imaging and Read Cervical Imaging

		Minimum	Average	Maximum	Standard Deviation
Brightness	Read	.09	.45	.78	.11
	Unread	.03	.42	.99	.29
Blur	Read	5.81	141.57	767.66	124.15
	Unread	.13	55.62	4827.15	326.65
Region	Read	.00	12579.19	767.66	3864.90
	Unread	6533.22	26864.09	46311.54	18250.75

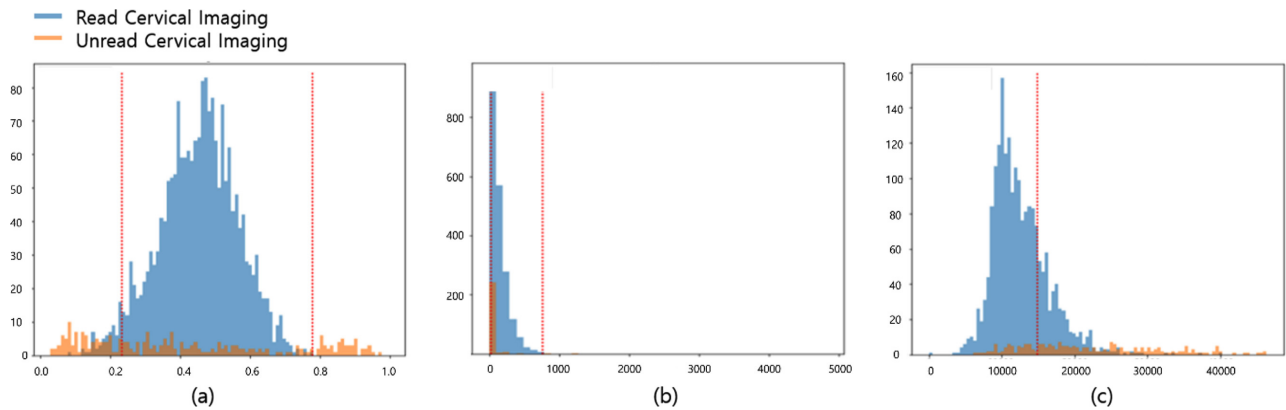


그림 5. 판독 가능한 영상과 판독 불가 영상의 분류 지표에 대한 화소값 분포도를 시각화한 히스토그램, (a) 밝기, (b) 블러, (c) 유클리드 거리
Fig. 5. Histogram of the image pixel distribution of Read Cervical Imaging and Unread Cervical Imaging, (a) Brightness, (b) Blur, (c) Euclidian distance

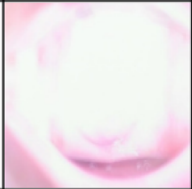
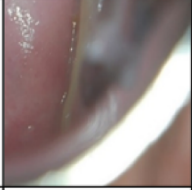
	Unread	Threshold	read	Threshold	Unread
Brightness		8%		70%	
	6%	$\leq$	39%	$\geq$	93%
Blur		13.5%			
	1.38%	$\leq$	15.16%		
Region				75%	
			63.54%	$\geq$	94.76%

그림 6. 판독 가능성에 따라 분류한 자궁경부 영상에서의 화소값 비교

Fig. 6. Comparison of image pixel values of Unread Cervical Imaging and Read Cervical Imaging

했다. t-검정 결과, 밝기($p < 0.001$), 블러($p < 0.001$), 유클리드 거리($p < 0.001$)가 통계적으로 유의한 차이를 보였다.

그림 6은 판독 가능한 영상과 판독 불가 영상의 화소값을 예시로 비교하여 나타낸 그림이다.

2. 판독 불가 영상 자동 분류 알고리즘의 성능 평가

본 연구에서는, 알고리즘을 적용한 데이터양과 알고리즘을 적

용하지 않은 데이터양을 비슷한 비율로 조정하였다. 알고리즘을 적용한 데이터의 경우 전체 판독 가능한 데이터 2,000장에서 1,357장의 데이터로 선별되었다. 알고리즘을 적용하지 않은 데이터의 경우 알고리즘을 적용하여 선별된 데이터양과 비슷한 비율로 조정하기 위해 판독 가능한 데이터에서 무작위로 추출한 1,000장과 판독 불가 데이터 257장으로 구성하였다.

표 2. 알고리즘 적용 유무에 따른 ResNet-50의 성능평가 비교표

Table 2. Comparative Table of Performance Evaluation of ResNet-50 according to Exclusion Unread Cervical Imaging Algorithm

	Precision(%)	Recall(%)	F1-score(%)	Accuracy(%)
read data	91.59	93.09	92.31	91.61
read data + unread data	72.64	65.04	68.55	78.88

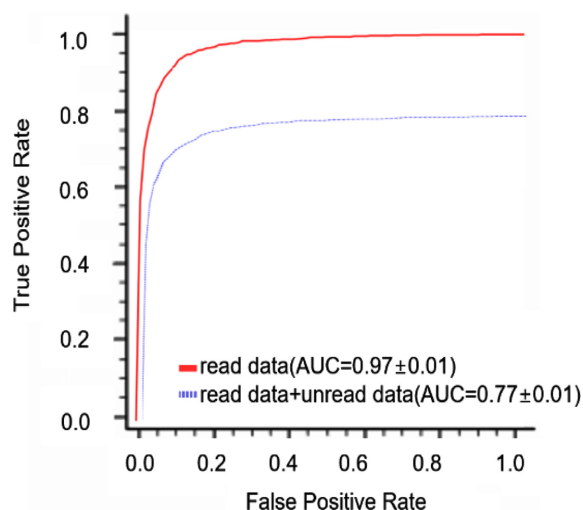


그림 7. 알고리즘 적용 유무에 따른 ResNet-50의 ROC 분석결과
Fig. 7. The results ROC analysis of ResNet-50 according to Exclusion Unread Cervical Imaging Algorithm

248

전체 데이터에서 평균적인 성능은, Precision은 72.64%, Recall은 65.04%, F1-score는 68.55%, Accuracy는 78.88%로 나타났다. 알고리즘을 적용한 데이터에서는, Precision은 91.59%, Recall은 93.09%, F1-score는 92.31%, Accuracy는 91.61%로 나타났다. 이를 통해 본 논문에서 제안한 알고리즘을 적용한 데이터가 accuracy를 포함한 각 성능 평가 지표에서 전체 데이터보다 높은 성능을 내는 것을 확인하였다. 표 2에 알고리즘을 적용하여 판독 가능한 데이터 2,000장을 알고리즘을 통해 한 번 더 정제한 1,357장의 데이터를 read data, 알고리즘을 적용하지 않은 전체 데이터 1,257장을 read data + unread data로 정의하여, 각각 분류 모델에 학습시켰을 때의 분류 성능 결과를 나타냈다.

그림 7에 알고리즘 적용 유무에 따라 분류한 영상의 각 ROC 그래프와 AUC 값을 나타냈다. AUC 값은 전체 데이터에서 0.77을, 판독 가능한 데이터에서 0.97을 나타냈다. 두 ROC 그래프 모두 t-검정 결과, $P < 0.0001$ 의 유의미한 차이를 보였다.

IV. 고찰 및 결론

본 논문에서는 자궁경부 영상에서 밝기, 블러 등의 화소 값을 구하여 판독 가능한 범위에 속하는 화소값의 임계 범

위를 구하였다. 임계값에 따라 알고리즘을 통해 판독 가능한 영상을 선별하였다. 알고리즘에 의해 선별된 영상과 판독 여부에 따라 분류되지 않은 전체 영상을 각각 딥러닝 분류 모델에 학습시켜 분류 성능을 비교하였다. 실험 결과, 본 연구에서 제안한 알고리즘을 적용하였을 때, 평균 91.61%로 높은 분류 성능을 보였고, 전체 데이터에서는 78.88%의 성능을 보였다. 밝기, 화질에 이상이 있는 자궁경부 영상들이 모델 학습 전에 제거되면서, 자궁경부의 병기 단계가 명확하고, 품질이 깨끗한 영상들만 학습되어 비교적 높은 분류 성능을 도출한 것으로 판단된다.

본 연구의 전처리과정에서, 일부 정확하지 않은 케이스가 발견되었는데 그 원인은 자궁경부 영역의 위치가 다양하기 때문으로 파악된다. 이에 대해서는 자궁경부 영역을 따로 검출(detection)하여 판독해야 하는 영역에 맞춰 영상을 자르고 화소값을 구하면, 더 객관적이고 정확한 분류가 가능할 것으로 판단된다.

또한 판독 가능한 임계 범위설정을 위해 t-검정과 히스토그램 분석법 등의 통계적 분석을 사용하여 범위를 지정하는 과정에서 일부 한계가 있었다. 자궁경부 영상의 판독 가능 여부를 결정하는 요인이 매우 다양하고 일관적이지 않아, 데이터 제외의 정확도가 떨어졌을 가능성이 있다. 본 연구에서 제안한 알고리즘은 세 가지 분류 기준에 대해 한가지라도 판독 가능한 범위에 속하지 않으면, 판독 불가 데이터로 제외하였다. 이때 본 논문에서 제안한 분류 지표에 대해서는 판독 가능한 임계값을 모두 만족하지만 다른 판독 불가 케이스의 영상이 존재할 수 있다. 본 논문에서 분류 지표로 사용되지 않은 다른 화소값에 대해 이상이 있다면, 실제로는 판독이 불가능한 영상이 판독 가능한 영상으로 분류되었을 수 있다. 즉 자궁경부 영상 촬영 기기의 촬영 방식 또는 기기 자체의 결함으로 인한 노이즈, 빛 반사, 영상의 컬러 이상 등의 문제에 대한 가능성은 배제하였다. 따라서 기존의 영상처리기법을 사용하여 자궁경부를 분류한 연구사례의 정확도보다 낮은 수치로, 본 알고리즘의 최종 성능이 도출된 것으로 판단된다. 향후 연구에서, 판독 불가 케이스를 충분히 수집하고, 자궁경부 영상 분류에 오류를 일으키는 요소에 관하여 추가적 연구를 진행한다면 알고리즘의 정확도가 향상될 것으로 판단된다. 또한 판독 불가 케이스 뿐만 아니라, 수집량에 대해서도 판독 가능한 데이터양의 비율과 동일하게

조정한다면, 알고리즘의 신뢰도와 정확도가 향상될 것이다.

앞서 설명한 연구 제한점을 추후의 연구를 통해 보완한다면 본 연구에서 제안한 알고리즘의 활용으로, 딥러닝 기반 자궁경부 영상 분류 및 진단 보조시스템 연구의 정확도가 더욱 향상될 것으로 기대된다.

Reference

- [1] Hyun-Ju Choi, Tae-Yun Kim, Patrik Malm, Ewert Bengtsson, Heung-Kook Choi. Study on evaluating the significance of 3D nuclear texture features for diagnosis of cervical cancer. Korean Society of Computer Information. 2011;16(10):83-86.
- [2] <http://www.amc.seoul.kr/asan/healthinfo/disease/diseaseDetail.do?contentId=31818>, accessed on 2014.
- [3] https://www.cancer.go.kr/lay1/program/S1T211C223/cancer/view.do?cancer_seq=4877, accessed on Nov. 17, 2019.
- [4] Alyafeai Z, Ghouti L. A fully-automated deep learning pipeline for cervical cancer classification. Expert Systems with Applications. 2020;83-86
- [5] Chandran V, Sumithra MG, Karthick A, George T, Deivakani M, Elakkiya B, Subramaniam U, Manoharan S. Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images. BioMed Research International. 2021;83-86
- [6] Yu Jin Seol, Young Jae Kim, Kye Hyun Nam, Kwang Gi Kim. Comparison on the Deep Learning Performance of a Field of View. Journal of Korea Multimedia Society. 2020; 23(7):812-818.
- [7] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, Yongjun Wang. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230-243.
- [8] Youngtak Cho, Kiok Ahn. Design and Implementation of Automated Detection System of Personal Identification Information for Surgical Video De-Identification. Journal of convergence security. 2019;19(5):75-84.
- [9] Korean society of Obstetrics and Gynecology (<https://www.ksog.org/public/index.php?sub=4&third=4>).
- [10] http://www.samsunghospital.com/home/healthInfo/content/contentList.do?CONT_CLS_CD=001020001&ST=DIS&TAB=DIS_NM&SW=%EC%9E%90%EA%B6%81%EA%B2%BD%EB%B6%80%EC%95%94, accessed on 2015.
- [11] Song E-H, Kim J-J, Myung N-H, Park C-H. Spiral Brush in PapSure Test for Cervical Cancer Screening. Korean Journal of Gynecologic Oncology and Colposcopy. 2002;13(4):313-326.
- [12] Hee HS, Chern YJ, Classify And Visualize The Fatty Liver Using Class Activation Maps And CNN, Korea Institute Of Communication Sciences, 2020;18(10):83-86.
- [13] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba. Learning Deep Features for Discriminative Localization. Computer Vision and Pattern Recognition. 2016; 2921-2929.
- [14] <https://acsjournals.onlinelibrary.wiley.com/doi/full/10.3322/caac.21628>, accessed on Jul. 30, 2020.
- [15] Sangchul Kim, J. Nang. An Analysis of Luminance Histogram and Correlation of Motion Vector for Unsuitable Frames for Frame Rate Up Conversion. Korean Institute of Information Scientists and Engineers. 2016;22(10):532-536.
- [16] Raghav Bansal, Gaurav Raj, Tanupriya Choudhury. Blur image detection using Laplacian operator and Open-CV. IEEE. 2017.
- [17] Wang J, Song Y, Leung T, Rosenberg C, Wang J, Philbin J, Chen B, Wu Y. Learning Fine-grained Image Similarity with Deep Ranking. IEEE. 2014;1,386-1,393
- [18] Kim MK. Feature Extraction on a Periocular Region and Person Authentication Using a ResNet Model. Journal of Korea Multimedia Society. 2019;22(12):1,347-1,355.
- [19] Kang H-J. Efficient Fixed-Point Representation for ResNet-50 Convolutional Neural Network. Journal of the Korea Institute of Information and Communication Engineering. 2018;22(1):1-8.
- [20] Jeon S-S, Son K-Y, Lee J-H, Oh J-S, Son S-H. A Comparison Analysis on the Sales Price of Apartments according to G-SEED by Using T-test. Journal of the Autumn Academic Presentation Conference. 2019;19(2):207-208.