Journal of Biomedical Engineering Research 43: 102-108 (2022)
http://dx.doi.org/10.9718/JBER.2022.43.2.102

학 술 논 문

딥러닝 기반 후두부 질환 내시경 영상판독 보조기술 개발

정인호[1] · 황영준[1] · 성의숙[2,3]* · 남경원[1,3,4]*

[1]부산대학교 의과대학 의공학협동과정, [2]부산대학교 의과대학 이비인후과학교실
[3]양산부산대학교병원 의생명융합연구원, [4]부산대학교 의과대학 의공학교실

Development of Deep Learning-based Clinical Decision Supporting Technique for Laryngeal Disease using Endoscopic Images

In Ho Jung[1], Young Jun Hwang[1], Eui-Suk Sung[2,3]* and Kyoung Won Nam[1,3,4]*

[1]Interdisciplinary Program in Biomedical Engineering, College of Medicine, Pusan National University, Yangsan, Korea
[2]Department of Otolaryngology-Head and Neck Surgery, College of Medicine, Pusan National University, Yangsan, Korea
[3]Research Institute for Convergence of Biomedical Science and Technology, Pusan National University Yangsan Hospital, Yangsan, Korea
[4]Department of Biomedical Engineering, College of Medicine, Pusan National University, Yangsan, Korea
(Manuscript received 23 December 2021 ; revised 8 April 2022 ; accepted 11 April 2022)

Abstract: Purpose: To propose a deep learning-based clinical decision support technique for laryngeal disease on epiglottis, tongue and vocal cords. **Materials and Methods:** A total of 873 laryngeal endoscopic images were acquired from the PACS database of Pusan N ational University Yangsan Hospital. and VGG16 model was applied with transfer learning and fine-tuning. **Results:** The values of precision, recall, accuracy and F1-score for test dataset were 0.94, 0.97, 0.95 and 0.95 for epiglottis images, 0.91, 1.00, 0.95 and 0.95 for tongue images, and 0.90, 0.64, 0.73 and 0.75 for vocal cord images, respectively. **Conclusion:** Experimental results demonstrated that the proposed model have a potential as a tool for decision-supporting of otolaryngologist during manual inspection of laryngeal endoscopic images.

Key words: Clinical decision support, Endoscopic image, Deep learning, VGG16, Epiglottis, Tongue, vocal cords

I. 서 론

두경부암은 입술, 구강, 코, 부비동, 인두, 후두, 경부식도, 침샘, 갑상선과 경부의 연부조직 등 얼굴과 목 거의 모든 부위에 발생하는 암을 말하며, 수개월 간의 쉰 목소리, 약물치료에도 불구하고 지속되는 입안 궤양, 연하통으로 음식물을 삼킬 수 없는 경우 등이 주요 증상으로 알려져 있다. 전 세계적으로 매년 65 만명 이상의 새로운 두경부암 환자가 발생하며 35 만명 이상이 이와 관련된 질환으로 사망하는 것으로 알려져 있으며, 국민건강보험공단 조사자료에 따르면 국내 두경부암 환자는 2019년 23,691명으로 2015년 19,856명 대비 19.4%가 증가하였다[1]. 두경부암은 장기간에 걸친 흡연과 음주가 가장 큰 원인으로 알려져 있으나 최근에는 혀, 목구멍 편도 부위의 상피나 점막의 손상 부위를 통해 침투하는 인유두종 바이러스 감염으로 인한 발생사례가 크게 증가하고 있는 추세이다[2]. 미국암학회의 통계에 따르면 후두암의 완치율(5-year survival rate)은 1기 90%, 2기 70%, 3기 50%, 4기 40% 정도이며, 이는 후두암을 조기에 발견하고 적절한 치료를 받을경우 환자의 90%

*Corresponding Author : Kyoung Won Nam, Associate Professor
Department of Biomedical Engineering, College of Medicine, Pusan National University
Tel: +82-51-510-8119
E-mail: marmera@gmail.com
*Corresponding Author: Eui-Suk Sung, Assistant Professor
Department of Otolaryngology-Head and Neck Surgery, College of Medicine, Pusan National University
E-mail: sunges77@gmail.com
This study was supported by a 2021 research grant from Pusan National University Yangsan Hospital.

이상이 완치될 수 있다는 것을 의미한다[3]. 일반적으로 두경부 종양에 대한 표준 진단 프로트콜은 1) 내시경을 검사 부위에 삽입하여 측정된 영상을 임상의사가 육안으로 확인하고, 2) 종양이 의심되는 부위가 확인되면 해당 부위에 대한 생검을 실시한 후, 3) 조직 샘플에 대한 정밀 병리검사를 수행하여 해당 종양의 악성 여부 및 세부 특성을 확인하며 종양이 악성일 경우 암으로 판정하는 것이다. 하지만, 중소 의료기관들의 경우 두경부 암 분야 전문의가 원내에 상주하지 않는 경우가 대부분이고, 초기 종양 영상판독의 정확도가 임상의사 개개인의 실력 및 경험 정도에 따라 일정 수준의 편차를 나타낼 수 있어, 비숙련 임상의사의 경우 초기 영상판독 시 오진 발생의 가능성이 상대적으로 높다. 따라서, 임상의사의 개인별 숙련도 및 경험치와 무관하게 두경부 질환 영상판독의 정확도를 일정수준 이상으로 안정적으로 유지하여 오진의 위험을 낮출 수 있도록 도와주는 진단보조기술을 개발할 필요가 있다.

최근 인공지능(AI) 기술을 활용한 진단보조기술이 끊임없이 발전하고 있으나, 그 대부분이 AI 학습에 필요한 균일화, 통일화된 고품질 영상을 대량으로 쉽게 얻을 수 있는 X-Ray, CT, MR 영상을 주 대상으로 하고 있다[4-6]. 구강, 식도, 위, 소장, 대장 등의 인체장기 표면에 발생한 종양, 암 등을 확인하기 위해 가장 많이 활용되는 것은 내시경(Endoscope) 검사이지만, 내시경 영상의 경우 영상의 구도, 각도, 원근, 색상 등이 비정형적이고 병변 이미지의 자유도가 높다는 문제가 있어 상대적으로 AI 기술 적용이 어렵다는 단점이 있어 상대적으로 AI 적용 연구사례가 적다. 또한, 선행연구 사례 대부분이 소화기 내시경(위, 대장/소장) 영상에 딥 러닝(deep learning) 기반 AI 기술을 접목한 사례들이며[7-9], 구강, 식도와 같은 두경부 영역에서의 종양, 암 조기발견을 위한 내시경 기반 AI 진단보조기술 연구는 아직 수행된 사례가 상대적으로 적은 것으로 조사되었다.

본 연구에서는 후두 내시경으로 촬영된 두경부 질환 환자의 내시경 진단영상으로부터 후두덮개(epiglottis), 혀 (tongue), 성대(vocal cord) 부위의 정상·비정상 여부를 자동으로 판별하는 딥 러닝 기반 진단보조기술을 제안하고, 개발 모델의 판별 정확도를 정량적으로 평가하였다.

II. 연구 방법

본 연구에서는 양산부산대학교병원 이비인후과 소속 두경부외과 전문의의 협조와 기관 IRB의 승인절차를 거쳐(No. 05-2019-008) 병원 PACS 데이터베이스로부터 후두덮개, 혀, 성대 부위별로 대표성을 가지는 원본 내시경 영상 873장을 확보하였다: 후두덮개 100장(정상 50장, 비정상 50장), 혀 223장(정상 121장, 비정상 102장), 성대 450장(정상 170장, 비정상 280장). 각 영상에 대한 레이블링은 두경부외과 전문의의 도움으로 후두덮개, 혀, 성대 부분의 악성종양, 양성종양, 폴립, 염증 등을 포함한 영상들을 비정상 영상으로 분류하였으며, 영상에 포함된 환자 식별정보를 사전에 일괄 삭제한 후 각 영상에서 AI 모델 학습에 사용할 주요 관심영역(ROI)을 150 × 150 pixel 크기로 추출(cropping) 하고, 0 - 255의 RGB 계수로 구성된 원본 이미지를 0 - 1 범위로 Rescale 하였다. 비정상 영상에는 악성종양, 양성종양, 폴립, 염증 등을 포함하였으며, 각 부위별 영상 데이터는 6 : 2 : 2의 비율로 분류하여 학습(training dataset), 검증(validation dataset), 평가(test dataset)에 각각 활용하였다.

본 연구에서는 기존의 딥러닝 기반 내시경 영상 분석 연구에서 자주 활용되고 전이학습(Transfer learning) 이 상대적으로 용이한 VGG16 모델을 선택하였으며, NVIDIA Geforce RTX2060, Python 3.7.11, Tensorflow 2.3.0, Keras 2.6.0 및 CUDA 10.1 환경 하에서 모델 개발을 수행하였다. 해당 모델은 총 13개의 Convolution Layer, 5개의 Pooling Layer, 3개의 Fully Connected Layer로 구성되며, Convolution Layer 및 Pooling Layer에서 입력된 이미지의 특징 및 패턴을 추출하여 Feature Map을 생성하면, Fully Connected Layer에서 이를 활용하여 대상을 분류하는 구조로 되어있다(그림 1)[10]. 이러한 Convolutional Neural Network 구조는 인간의 시각정보 처리방식을 모방한 것으로, 입력 이미지와 가까운 층에서는 이미지의 가장자리 부분, 곡선 등과 같은 저수준 특징을 학습하고, 높은 층으로 갈수록 이미지의 특정 부분 형상과 질감과 같은 고수준 특징을 인식하며, 출력층에서는 이미지를 분류하는 복잡한 추론을 수행한다.

PACS 데이터베이스에서 추출한 영상의 수가 VGG16 모델을 학습시키는데 충분치 못하므로, GitHub로부터 ImageNet 영상을 활용하여 사전에 학습 완료된 VGG16 모델을 다운로드 한 후, PACS 데이터베이스에서 확보한 후두영상(training dataset) 을 이용하여 모델의 Fully Connected Layer를 추가적으로 학습시키는 전이학습 기법을 적용하였다. 또한, 모델 분류성능을 추가적으로 개선하고 과적합(overfitting) 문제를 방지하기 위해 두 가지의 파라메터 조정방식을 적용하였다: 1) 기본 VGG16 모델의 Block 1부터 Block 5까지의 전체 파라메터 값을 동결시키고, Fully Connected Layer의 파라메터 값만 training dataset을 이용하여 조정(No-Fine-Tuned Model), 2) 기본 VGG16 모델의 Block 1부터 Block 5의 Convolution Layer 1까지의 파라메터 값을 모두 동결시키고, Block 5의 Convolution Layer 2 이후의 파라메터 값만 미세하게 조정(Fine-Tuned Model). Keras에서 제공하는 최적화 도구(Optimizer)는 Adam(1e-4), 손실함수(Loss Function)는 Binary_

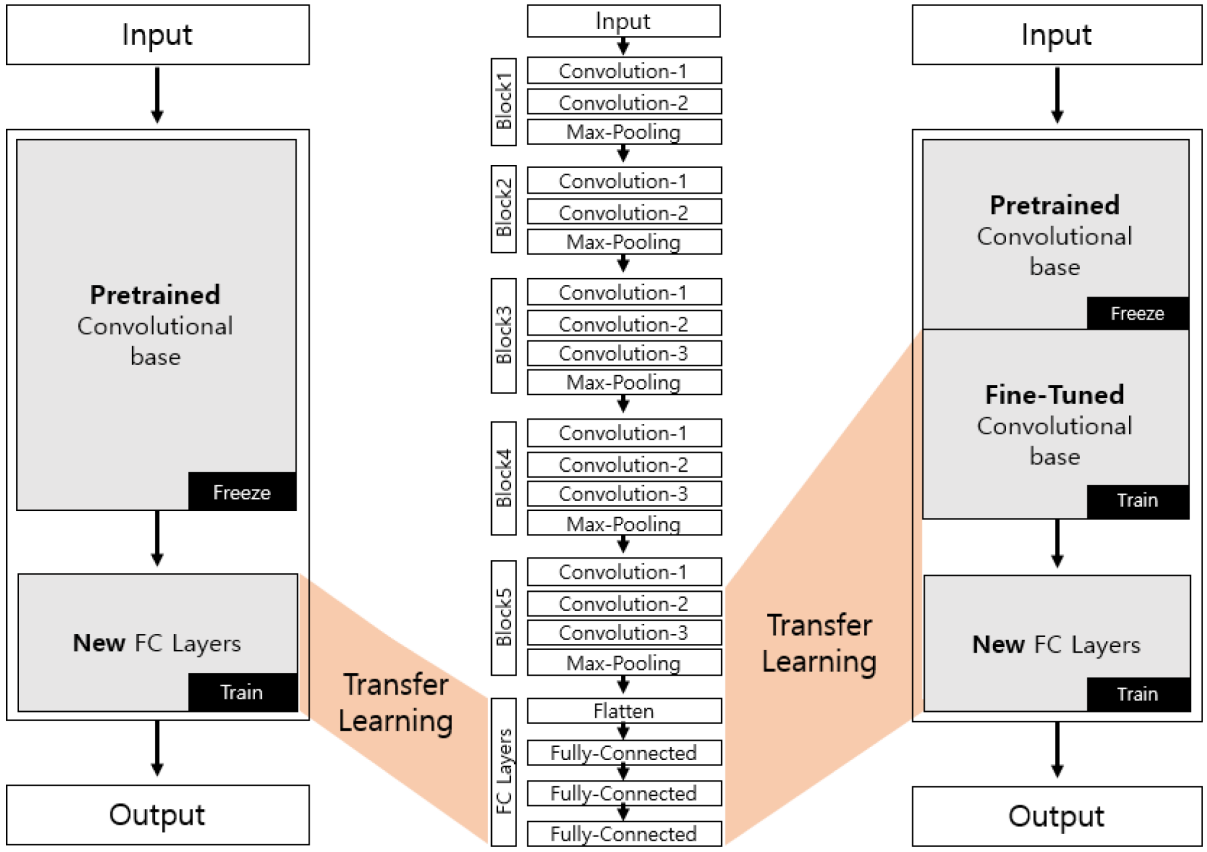


그림 1. 사용된 VGG16 모델의 기본구조 및 두 가지 전이학습 전략 (No-Fine-Tuned 모델, Fine-Tuned 모델) 개요도
Fig. 1. Basic structure of the utilized VGG16 model and two strategies of model transfer learning (No-Fine-Tuned Model and Fine-Tuned Model)

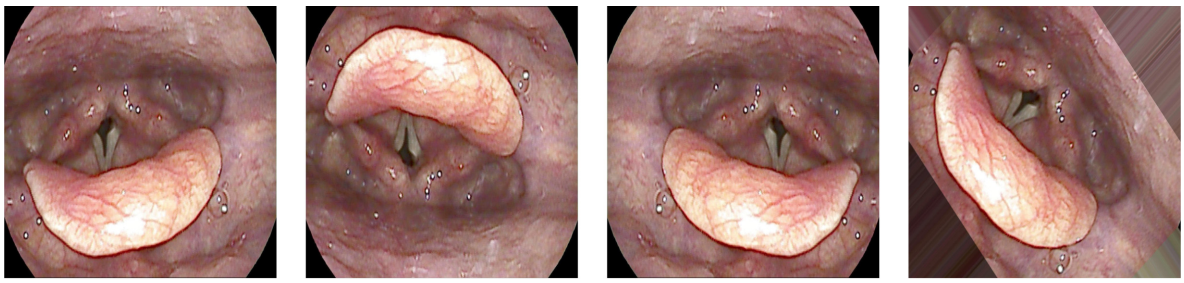


그림 2. 후두덮개 영상증강 사례 (A) Original Image (B) Horizontal_Flip (C) Vertical_Flip (D) 60° Rotation
Fig. 2. Example of epiglottis image augmentation (A) Original Image (B) Horizontal_Flip (C) Vertical_Flip (D) 60° Rotation

crossentropy를 사용하였다.

또한, 모델 추가학습에 사용될 로컬 데이터(training dataset)의 부족을 보완하기 위해 Keras에서 제공하는 ImageDataGenerator 함수를 사용한 영상증강(Data Augmentation)을 모델 학습 시 적용하였다(그림 2): rotation_range = 60, zoom_range = 0.1, horizontal_flip = True, vertical_flip = True, fill_mode = nearest.

개발 모델의 성능평가를 위해 본 연구에서는 혼돈행렬(Confusion Matrix) 분석기법을 활용하였으며, 정상영상을 정상으로 판별한 경우(Normal/Normal)를 True-Negative (TN), 정상영상을 비정상으로 판별한 경우(Normal/Abnormal)를 False-Positive(FP), 비정상영상을 정상으로 판별한 경우(Abnormal/Normal)를 False-Negative(FN), 비정상영상을 비정상으로 판별한 경우(Abnormal/Abnormal) 를

True-Positive(TP)로 각각 정의하였다. 그리고, 이 값들을 기반으로 다음과 같이 정확도(accuracy), 정밀도 (precision), 재현율(recall), F1-값(F1-score)을 각각 계산하였다.

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN)$$
$$\text{Precision} = TP / (TP + FP)$$
$$\text{Recall} = TP / (TP + FN)$$
$$\text{F1-Score} = 2 * (\text{Recall} * \text{Precision}) \cdot (\text{Recall} + \text{Precision})$$

III. 연구 결과

먼저, 최적의 Epoch 값 설정을 위해 Batch Size를 10으로 고정하고, Epoch 값을 10, 20, 30, 40, 50으로 변화시키면서 validation dataset을 이용하여 각 경우에서의 validation loss 및 validation accuracy를 측정한 결과는 표 1과 같다(Fine-Tuned Model). Validation loss를 기준으로 할 경우 후두덮개, 혀, 성대 모두 Epoch = 30일 때 loss 값이 0.0562, 0.2678, 0.6062로 가장 낮았으며, validation accuracy를 기준으로 할 경우에도 후두덮개, 혀, 성대 모두 Epoch = 30일 때 accuracy 값이 95.0%, 92.5%, 83.3%로 가장 높았다. 이러한 평가결과를 바탕으로 이후의 Confusion Matrix Analysis에서는 Epoch = 30에서 산출된 파라메터 값을 활용하였다.

표 1. Epoch 값이 10, 20, 30, 40, 50인 경우 후두덮개, 혀, 성대영상에 대한 validation loss, validation accuracy 측정 결과(Fine-Tune Model; Validation dataset)

Table 1. Variations in validation loss and validation accuracy at epoch 10, 20, 30, 40 and 50 for each target organ(Fine-Tuned Model; Validation dataset)

	Epoch	Loss	Accuracy
Epiglottis	10	0.1093	90.0
	20	0.0941	90.0
	30	0.0562	95.0
	40	0.1503	90.0
	50	0.4982	85.0
Tongue	10	0.4118	87.5
	20	1.1663	75.0
	30	0.2678	92.5
	40	0.4061	90.0
	50	0.4925	87.5
Vocal Cord	10	0.6796	72.2
	20	0.7337	78.8
	30	0.6062	83.3
	40	1.4511	72.2
	50	0.8077	77.7

표 2는 Fined-Tuned Model에 입력된 후두덮개, 혀, 성대 입력영상 (test dataset) 중에서 TP, TN, FP, FN 으로 판정된 영상의 예를 나타낸다. 혀의 경우에는 모든 입력영상에 대해 FN이 발생하지 않았다.

표 3은 test dataset을 이용하여 수행한 Confusion Matrix Analysis 결과를 나타낸다(Batch Size = 10, Epoch = 30). 총 194 장의 test dataset에 대해(후두덮개: 정상 30장, 비정상 30장, 혀: 정상 24장, 비정상 20장, 성대: 정상 34장, 비정상 56장), No-Fine-Tuned Model의 경우 30장의 비정상 후두덮개 영상 중 27장은 비정상으로 판정하였으나(TP = 27) 3장은 정상으로 판정하였으며(FN = 3), 30장의 정상 후두덮개 영상 중 27장은 정상으로 판정하였으나(TN = 27) 3장은 비정상으로 판정하였다(FP = 3). 혀 영상에 대해서는 20장의 비정상영상 중에서 19장을 비정상으로 판정하고(TP = 19) 1장을 정상으로 판정하였으며(FN = 1), 24장의 정상영상 중에서 18장을 정상으로 판정하고(TN = 18) 6장을 비정상으로 판정하였다(FP = 6). 성대 영상에 대해서는 56장의 비정상영상 중에서 22장을 비정상으로 판정하고(TP = 22) 34장을 정상으로 판정하였으며(FN = 34), 34장의 정상영상 중에서 29장을 정상으로 판정하고(TN = 29) 5장을 비정상으로 판정하였다(FP = 5). Fine-Tuned Model 의 경우에는 30장의 비정상 후두덮개 영상 중 29장을 비정상으로 판정하고(TP = 29) 1장을 정상으로 판정하였으며(FN = 1), 30장의 정상 후두덮개 영상 중 28장을 정상으로 판정하고(TN = 28) 2장을 비정상으로 판정하였다(FP = 2). 혀 영상의 경우에는 20장의 비정상영상 모두를 비정상으로 판정하고(TP = 20, FN = 0), 24장의 정상영상 중에서 22장을 정상으로 판정하고(TN = 22) 2장을 비정상으로 판정하였다(FP = 2). 성대 영상의 경우에는 56장의 비정상영상 중에서 36장을 비정상으로 판정하고(TP = 36) 20장을 정상으로 판정하였으며(FN = 20), 34장의 정상영상 중에서 30장을 정상으로 판정하고(TN = 30) 4장을 비정상으로 판정하였다(FP = 4).

표 4는 test dataset을 이용하여 No-Fine-Tuned Model과 Fine-Tuned Model에서의 네 가지 성능지표 값을 계산한 결과를 나타낸다(Batch Size = 10, Epoch = 30). 후두덮개 영상의 경우 Fine-Tuned Model에서의 Precision, Recall, Accuracy, F1-score 값이 No-Fine-Tuned Model에서의 값보다 각각 0.04, 0.07, 0.05, 0.05 증가하였으며, 혀 영상의 경우에는 각각 0.15, 0.05, 0.11, 0.11 증가하였으며, 성대 영상의 경우에는 각각 0.09, 0.25, 0.16, 0.22 증가함을 확인할 수 있었다. 영상입력 후 판독결과 도출까지 소요되는 시간은 No-Fine-Tuned Model의 경우 3±2초, Fine-Tuned Model의 경우 4±3초였다.

표 2. 입력데이터 및 모델 판정결과 예시(Fine-Tuned Model)
Table 2. Examples of input image and the results of model output

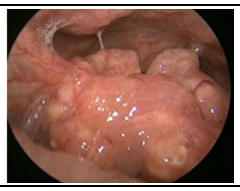
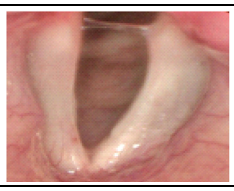
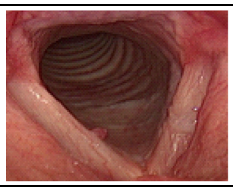
	TP	TN	FP	FN
Epiglottis				
Tongue				None
Vocal Cord				

표 3. 혼돈행렬 분석 결과(No-Fine-Tuned Model / Fine-Tuned Model)
Table 3. Results of Confusion Matrix Analysis(No-Fine-Tuned Model / Fine-Tuned Model Order)

		Predicted Class					
		Epiglottis		Tongue		Vocal Cord	
		Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
Actual Class	Normal	27 / 28 (TN)	3 / 2 (FP)	18 / 22 (TN)	6 / 2 (FP)	29 / 30 (TN)	5 / 4 (FP)
	Abnormal	3 / 1 (FN)	27 / 29 (TP)	1 / 0 (FN)	19 / 20 (TP)	34 / 20 (FN)	22 / 36 (TP)

표 4. 정밀도, 재현율, 정확도, F1-값 계산 결과(No-Fine-Tuned Model / Fine-Tuned Model)
Table 4. Values of precision, recall, accuracy and F1-score(No-Fine-Tuned Model / Fine-Tuned Model Order)

	Precision	Recall	Accuracy	F1-score
Epiglottis	0.90 / 0.94	0.90 / 0.97	0.90 / 0.95	0.90 / 0.95
Tongue	0.76 / 0.91	0.95 / 1.00	0.84 / 0.95	0.84 / 0.95
Vocal Cord	0.81 / 0.90	0.39 / 0.64	0.57 / 0.73	0.53 / 0.75

IV. 고찰 및 결론

2018년 기준 경제협력개발기구(OECD) 국가별 현황자료에 의하면, 우리나라의 활동의사 수는 인구 1,000명당 2.3명(한의사 포함)으로 28개 OECD 회원국 평균인 3.3명보다 낮으며, 인구 5,000만 명을 기준으로 했을 때 OECD 평균에 비해 약 5만 명의 활동의사가 부족하다. 또한, 대부분의 활동의사들이 주거, 교육, 육아, 교통 등 다양한 생활 인프라를 이유로 대도시 근무를 선호하여, 이로 인해 수도권 및 대도시와 지방 중소도시 간 의료서비스 격차 및 불균형이 갈수록 심해지고, 국가 의료복지의 불평등 현상도 심화되고 있어, 이를 보완하기 위한 AI-융합 의료기술의 중요성이 지속적으로 강조되고 있다. 하지만, 대부분의 AI-융합 진단보조기술이 X-Ray, CT, MR 영상을 대상으로 하고 있고, 내시경 영상을 활용한 AI-융합 진단보조기술은 상대적으로 그 수가 미미하다가 최근 3년 이내에 관련 연구들이 지속적으로 증가하는 추세에 있다. 내시경 검사는 다양한 인체장기의 표면에 발생한 종양, 암 등과 같은 이상부위를 육안으로 확인하기 위해 가장 보편적으로 시행되는 검사방법으로 특히 두경부 종양/암의 초기 검진에 유용하나, 아직 국내에서는 위내시경 및 대장내시경을 활용한 AI-융합 연구만 일부 수행된 바 있다. 이는 국가암검진사업을 통해 만 40세 이상

인 국민들을 대상으로 위내시경 및 대장내시경 검사를 기본 시행하는 국내 현실을 고려한 것으로 보인다. 이와 관련된 대표적 국내 연구사례로 ㈜셀바스AI와 강남세브란스병원 김지현 교수팀이 AI 기술을 활용하여 위내시경 영상으로부터 조기 위암이 의심되는 영역을 자동으로 찾고, 종양의 침범 깊이를 예측하는 공동연구를 발표한 바 있으며[7], 서울대학교병원 진은효 교수팀에서 건강검진을 통해 획득된 대장내시경 영상에서 대장용종을 자동 검출하는 AI 모델을 발표한 바 있으며[8], 가톨릭대학교 여의도성모병원 이환희 교수팀에서 소장캡슐내시경으로 촬영된 영상에 AI 기술을 적용하여 출혈성 병변과 궤양성 병변을 96% 이상의 높은 정확도로 분류하는 연구성과를 발표한 바 있다[9]. 다만, 최근에는 국내에서도 딥러닝 기반의 후두 내시경 영상 자동분석을 통해 두경부 질환 진단을 보조하는 AI-융합기술의 발표 사례가 증가하고 있다. 일례로, 부산대학교 남경원 교수팀에서 Mask R-CNN 기반으로 성대 근처에 발생한 종양의 위치를 자동으로 추적하는 연구결과를 발표한 바 있으며[11], 한국항공우주의료원 유태근 교수팀에서 ResNet-50, Inception-V3, MobileNet-V2 등의 CNN 모델을 이용하여 목(throat) 내시경 영상에서 중증 인두염을 판정하는 연구결과를 발표한 바 있으며[12], 서울아산병원 최승호 교수팀에서도 VGG16, Inception-V3, Xcepion, MobileNet-V2, EfficientNet-B0 등의 CNN 모델을 이용하여 후두부 이상을 판정하는 연구결과를 발표한 바 있다[13,14]. 하지만, 본 연구에서와 같이 후두덮개, 혀, 성대의 3개 중요 후두영역에 대한 정상/비정상 여부를 단일 모델로 동시에 판별할 수 있는 딥러닝 기반 연구결과는 아직 국내에서 발표된 바 없다.

본 연구에서의 평가결과를 보면 Block 5의 Convolution Layer 2 이하 파라메터 전체를 training dataset으로 전이학습한 경우(Fine-Tuned-Model), Fully Connected Layer의 파라메터만 training dataset으로 전이학습한 경우(No-Fine-Tuned-Model) 에 비해 모델의 성능이 전반적으로 개선됨을 확인할 수 있었다(표 3). Towards Data Science (TDS)에서 제시한 Dataset Size 및 Dataset Similarity 조건에 따른 최적의 Fine-Tuning 가이드라인에 따르면, dataset의 수량이 적고 dataset 간 유사도가 낮을 경우 그림 3의 case 4와 같이 Convolution Layer 후반부와 Fully Connected Layer를 training dataset으로 함께 전이학습하는 것이 보다 효과적인 Fine-Tuning 방법이다[15]. 본 연구에서 사용된 training dataset의 양이 상대적으로 적고 모델의 사전학습에 사용된 ImageNet 데이터베이스가 내시경 의료영상과는 유사성이 별로 높지 않으므로, 본 연구에서는 TDS 가이드라인에 기초하여 그림 1과 같은 Fine-Tuning을 수행하였으며, No-Fine-Tuned Model과 Fine-Tuned Model 간 비교 성능평가 결과는 TDS의 Fine-Tuning 가이드라인의 타당성을 나타내는 사례로 볼 수 있다.

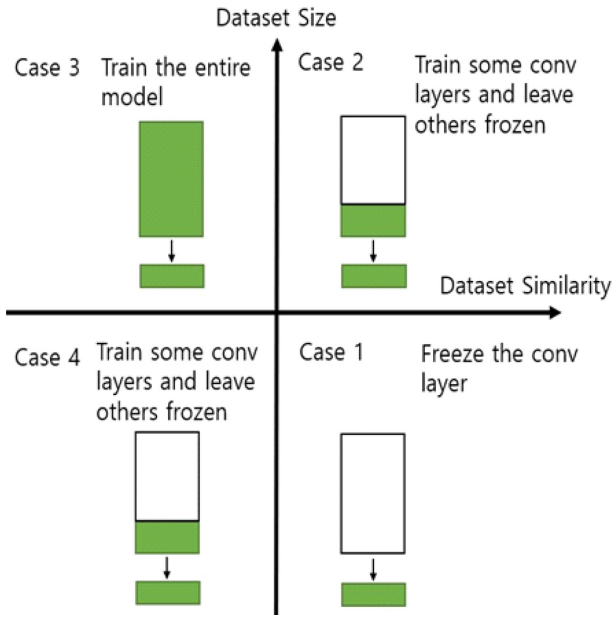


그림 3. TDS "Transfer Learning from pre-trained models"에서 제시한 최적 가이드라인 모식도[14]

Fig. 3. Scheme of optimal fine-tuning guideline from "Transfer Learning from pre-trained model" by TDS[14]

본 연구의 평가결과를 보면 후두덮개 영상 및 혀 영상의 경우 Fine-Tuned Model에서 95% 이상의 판별 정확도를 나타냈으나, 성대 영상의 경우 판별 정확도가 73%로 상대적으로 낮게 나타났다. 특히, 임상현장 적용 시 심각한 문제를 유발할 수 있는 False-Negative(질환영상을 정상영상으로 오진하는 경우)의 사례가 후두덮개(1장), 혀(0장)에 비해 성대(20장)에서 매우 높게 나타났다. 이는 본 연구에서 사용된 성대 영상의 수가 실제 성대 주변에 발생하는 매우 다양한 종양 사례들(cyst, nodule, polyp, leukoplakia, papilloma, Reinke's edema, granulomas 등) 을 모두 반영할 만큼 충분하지 못해 발생한 것으로 판단되며, 향후 다양한 종양 사례들에 대한 성대 영상을 추가로 확보하여 보완 학습을 진행할 경우 성대 영상에 대한 모델의 판별 정확도가 추가적으로 개선될 수 있을 것으로 판단된다. 다만, 본 연구를 위해 병원 PACS 데이터베이스 내 이비인후과 내시경 영상을 최대한 확보하려 노력하였으나, 딥러닝 모델을 충분히 학습시킬 정도의 대용량 영상데이터 확보에는 한계가 있었다. 향후, 추가적인 다기관 임상연구 신청을 통해 부산대학교병원(본원) 및 타 기관 영상자료를 추가로 확보하여 학습에 사용할 수 있다면, 개발 모델의 성능을 보다 높일 수 있을 것으로 사료된다. 또한, 본 연구에서는 VGG16 모델을 사용하였으나, 추후 보다 분류성능이 높은 다양한 CNN 모델들(ResNet, Inception 등)을 추가로 적용하여 후두 내시경 영

상의 판독에 가장 적합한 모델을 확인하기 위한 후속 연구도 수행할 필요가 있다. 다만, 현 시점에서의 평가 결과로 미루어 볼 때, 모델 다변화보다는 다기관 임상연구를 통한 추가 영상확보가 좀 더 시급한 것으로 사료된다.

딥 러닝 모델 개발 시 과적합 방지 혹은 완화는 모델의 실제 판독 성능에 큰 영향을 미치는 중요한 이슈이다. 보통 모델 학습 단계에서 Epoch 값을 증가시키면 training dataset에 대한 모델의 성능은 증가하지만, Epoch 값이 과도하게 높아질 경우 training dataset 에 대한 성능이 증가함에 반해 validation dataset 및 test dataset에 대한 성능이 악화되는 과적합이 발생하게 된다. 본 연구에서는 과적합 완화를 위해 표 1과 같이 Epoch 값을 최적화하였으나, 향후 추가적인 영상데이터의 확보 외에 drop out, regularization, batch normalization 등의 다양한 기술적 수단을 통해 모델의 실제 성능을 높여나갈 필요가 있다.

결론적으로, 본 연구에서는 후두덮개, 혀, 성대에 발생한 후두질환의 유무를 단일 VGG16 모델을 활용하여 자동 판별하는 후두 내시경 영상 기반 진단보조기술을 제안하고, 제안 모델의 정량적 성능을 평가하였다. 평가결과로 볼 때, 제안된 CNN 모델은 후두덮개와 혀 부위의 이상 여부에 대한 의사들의 내시경 영상 판독 시 판단의 정확도를 높이는 진단보조 스크리닝 용도로 효용성이 있을 것으로 기대된다. 다만, 성대부의 경우 추가적인 데이터 수집 및 모델 보완이 필요한 것으로 판단된다.

References

[1] https://www.nhis.or.kr/nhis/together/wbhaea01600m01.do?mode=view&articleNo=13 8272 (Accessed on Dec 22, 2021)

[2] Jung YS. Human papillomavirus in head and neck cancer: several questions. Korean J Otorhinolaryngol-Head Neck Surg 2014;57(3):143-150.

[3] https://www.amc.seoul.kr/asan/healthinfo/disease/diseaseDetail.do?contentId=33914 (Accessed on Dec 22, 2021)

[4] Sung J, Park S, Lee SM, Bae W, Park B, Jung E, et al. Added value of deep learning–based detection system for multiple major findings on chest radiographs: a randomized crossover study. Radiology 2021;299:450-459.

[5] Han SM, Hwang SI, Lee HJ. The classification of renal cancer in 3-phase CT images using a deep learning method. J Digit Imaging 2019;32(4):638-643.

[6] Suh CH, Shim WH, Kim SJ, Roh JH, Lee JH, Kim MJ, et al. Development and validation of a deep learning–based automatic brain segmentation and classification algorithm for Alzheimer disease using 3D T1-weighted volumetric images. Am J Neuroradiol 2020;41(12):2227-2234.

[7] Yoon HJ, Kim S, Kim JH, Keum JS, Oh SI, Jo J, et al. A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer. J Clin Med 2019;8(9):1310.

[8] Jin EH. Lee D, Bae JH, Kang HY, Kwak MS, Seo JY, et al. Improved accuracy in optical diagnosis of colorectal polyps using convolutional neural networks with visual explanations. Gastroenterology, 2020;158(8):2169-2179.e8.

[9] https://scienceon.kisti.re.kr/srch/selectPORSrchReport.do?cn=TRKO202000004396 (Acc essed on Dec 22, 2021)

[10] Amina B, Nadjia B, Azeddine B. Gastrointestinal image classification based on VGG16 and transfer learning. 2021 International Conference on Information Systems and Advanced Technologies (ICISAT). pp. 1-5.

[11] Kim GH, Sung ES, Nam KW. Automated laryngeal mass detection algorithm for home-based self-screening test based on convolutional neural network. Biomed Eng Online 2021; 20(51):1-10.

[12] Yoo TK, Choi JY, Jang YI, Oh I, Ryu IH. Toward automated severe pharyngitis detection with smartphone camera using deep learning networks. Comput Biol Med 2020;125(103980):1-9.

[13] Cho WK, Choi SH. Comparison of convolutional neural network models for determination of vocal fold normality in laryngoscopic images. J voice 2020;S0892-1997(20):30292-30297.

[14] Cho WK, Lee YJ, Joo HA, Jeong IS, Choi Y, Nam SY, et al. Diagnostic accuracies of laryngeal disease using a convolutional neural network-based image classification system. Laryngoscope 2021;131(11):2558-2566.

[15] https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f12 4751 (Accessed on Dec 22, 2021)