



후두암에서의 AI(Artificial Intelligence) 적용

이진춘

부산대학교 의과대학 이비인후과학교실

Application of Artificial Intelligence (AI) in Laryngeal Cancer

Jin Choon Lee

Department of Otorhinolaryngology-Head and Neck Surgery, Pusan National University School of Medicine, Yangsan, Korea

ABSTRACT

The application of artificial intelligence (AI) in laryngeal cancer (LC) has the potential to improve accuracy and efficiency in diagnosis and treatment, as well as aid in predicting patient prognosis and personalized therapy. The application of AI in LC has expanded to include various data modalities, enabling support for screening, diagnosis, therapeutic decision-making, and prognosis. This multidimensional approach utilizes the following data modalities: 1) Videomics: AI algorithms can analyze video recordings of laryngeal endoscopy or surgical procedures to aid in LC diagnosis and treatment planning. 2) Radiomics: AI algorithms can extract and analyze quantitative features from medical imaging data, such as computed tomography (CT) scans and magnetic resonance imaging (MRI), to assist in LC screening and diagnosis. 3) Acoustics: AI-based acoustic analysis techniques can assess vocal characteristics and patterns in LC patients. 4) Clinical data: AI algorithms can integrate clinical data, including patient demographics, medical history, and laboratory results, to support LC diagnosis, staging, and treatment decision-making. 5) Genomics: by integrating genomic data, such as gene expression profiles or DNA sequencing results, with AI algorithms, researchers can identify molecular markers associated with LC. By considering various factors, such as tumor characteristics, stage, and individual patient factors, AI algorithms can provide insights to optimize treatment strategies and improve patient outcomes. However, it is crucial to emphasize that AI should not replace healthcare professionals but rather serve as a supportive tool. Here, published papers on AI applied to the diagnosis, treatment, and prognosis of laryngeal cancer are reviewed and useful information is provided to readers.

KEY WORDS: Laryngeal cancer; Artificial intelligence, Accuracy.

서론

인공지능(artificial intelligence, AI)이란 우리가 평소에 하는 시각적인 지각, 물체 및 단어 인식, 복잡한 의사결정과 같이 보통 인간의 지능이 필요한 작업을 컴퓨터 시스템이 수

행할 수 있게 하는 기술을 말한다. 지난 50년 동안 컴퓨팅 및 데이터 저장 기술의 급속한 발전으로 AI는 과학 소설의 영역에서 현실로 이어졌다. 현재 AI는 일상생활에서 없어서는 안 될 기술로 자리 잡았으며, 여러 가지 분야인 온라인 쇼핑, 모바일 투표 입금, 스마트 개인 비서 등과 같은 기술에서 중요

Received: June 1, 2023 / Revised: June 20, 2023 / Accepted: June 22, 2023

Corresponding author: Jin Choon Lee, Department of Otorhinolaryngology-Head and Neck Surgery, Pusan National University School of Medicine, Yangsan 50612, Korea

Tel: +82-55-360-2652, Fax: +82-55-360-2162, E-mail: ljc0209@hanmail.net

Copyright © 2023. The Busan, Ulsan, Gyeongnam Branch of Korean Society of Otolaryngology-Head and Neck Surgery.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

한 역할을 하고 있다. 또한, AI는 오랫동안 의료서비스 혁신의 잠재적 능력으로 평가되어 왔다. 그러나 현재까지 실제 의료 실무나 환자 간호 분야에서는 그다지 기여를 하지 못했다. 이는 대개 AI가 가져올 기계 오류 가능성으로 인한 인간 감독 부재로 야기되는 여러 가지 문제를 두려워하기 때문이다. 그러나 비용, 질, 접근성 개선 가능성이 높은 AI는 의료의 미래를 형성하는 데 중요한 역할을 할 것이다.¹⁾ AI의 출현은 후두암의 진단, 치료 및 예후 예측에 있어서 정확성과 효율성을 향상시키는 새로운 가능성을 보여주었다. 후두암에서 AI의 적용에 있어 여러 가지 다차원적 접근 방식은 다음과 같은 데이터 형식을 활용한다. 1) 비디오믹스(Videomics): AI 알고리즘은 후두 내시경 또는 수술 절차의 비디오 녹화를 분석하여 후두암의 진단 및 치료 계획을 돕는다. 컴퓨터 비전 및 기계 학습 기술을 내시경 영상 분석에 적용하는 것을 이전에는 "비디오믹스(Videomics)"라고 불렀다.²⁾ 2) 라디오믹스(Radiomics): 디지털 방사선학 이미지에서 양적 특징을 추출하고 이를 분석하여 가설 생성과 검증을 위한 것을 의미한다. AI 알고리즘은 CT(computed tomography) 스캔 및 MRI(magnetic resonance imaging)와 같은 의료 영상 데이터에서 정량적 특징을 추출하고 분석하여 후두암의 검진 및 진단을 지원할 수 있다.³⁾ 3) 음성 자료(acoustic data): AI 기반 음성 분석 기술은 후두암 환자의 음성 특성 및 패턴을 평가할 수 있다. 음성 장애가 종종 성문암의 첫 증상이기 때문에 많은 저자들이 음성 녹음을 기반으로 성문암을 진단하기 위해 기계학습을 시도해 왔다. 음성장애를 가진 후두암 환자의 목소리를 건강한 목소리와 구분하기 위해 추출된 음향적 특징과 불규칙성 측정치를 사용했다.⁴⁾ 4) 임상 데이터(clinical data): AI 알고리즘은 환자 인구 통계, 병력, 실험실 결과를 포함한 임상 데이터를 통합하여 후두암의 진단, 병기 결정 및 치료 의사 결정을 지원할 수 있다.¹⁾ 5) 게노믹스(genomics): 유전자 발현 프로필이나 DNA 시퀀싱 결과와 같은 게놈 데이터를 AI 알고리즘과 통합하여 후두암과 관련된 분자 표지자를 식별할 수 있다.^{1,5)} 앞에서 언급한 여러 가지 정보를 이용한 인공지능을 후두암의 진단 및 치료에 적용해 볼 수 있을 것으로 사료된다. 이에 논문을 검색하고 정리하여 인공지능이 어떻게 후두암에 적용되는지 알아보려고 한다.

본론

기계 학습(machine learning, ML)은 인간이 학습하는 방식을 모방하는 알고리즘과 대용량 데이터 세트를 결합하는 AI

의 하위 분야이다. 딥 러닝(deep learning)은 뇌의 구조에서 영감을 받은 네트워크를 기반으로 한 ML의 한 형태로, 합성곱 신경망(convolutional neural network, CNN)이라고도 한다. ML 모델은 후두암 진단, 치료 및 예후 예측에 적용되기 위해 비디오 데이터(후두경 비디오 등), 영상 데이터 및 진단-병리학 데이터 및 영상 데이터(방사선 영상, 조직학적 슬라이드, 내시경 영상 등), 음향 데이터(목소리, 호흡 소리 등)와 같은 다양한 데이터 양식을 통해 교육될 수 있다.^{6,7)}

비디오데이터(후두 내시경 영상)를 이용한 AI(artificial intelligence)의 적용

컴퓨터 비전 및 기계 학습 기술을 내시경 영상 분석에 적용하는 것을 이전에 "비디오믹스(Videomics)"라고 불렀다.²⁾ 비디오믹스는 대장 내시경 검사 분야에서 잘 연구되어 왔다.⁸⁾ 최근에는 이와 유사한 분석 기법을 후두 내시경 영상과 같은 다른 영상에도 적용하려는 노력이 있었지만, 후두 내시경 영상의 경우 데이터 세트 부족 및 후두의 해부학적 복잡성 및 동적 특성 때문에 진행이 더 늦어졌다.⁹⁾ 이 연구 분야에서는 주로 후두내시경 비디오에서 추출한 중요 이미지(키프레임)를 기계 학습 모델의 입력 데이터로 사용한다(Fig. 1).

지금까지 가장 큰 후두 내시경 이미지 데이터 세트(2,208 명의 환자에서 14,987개의 이미지)를 사용하여, 경험이 풍부한 이비인후과 전문의를 능가하는 정확도로 후두암, 전암 병변, 양성 병변 및 정상 성대를 분류할 수 있는 모델을 개발하여 보고하였다.¹⁰⁾ 새로운 기술이 등장함에 따라, 기계 학습 기술을 기반으로 중요한 영상과 중요하지 않은 영상을 자동으로 분류하여 데이터 준비 및 수동 키 프레임 추출과 같은 노력을 줄일 수 있는 모델을 개발하는 것이 필요할 것이다. 이것이 후두 내시경 영상을 이용한 AI의 임상 적용에 있어서 핵심적인 과제가 될 것이다.^{11,12)} 2019년 보고에 의하면 딥 합성곱 신경망(deep convolutional neural network, DCNN)은 모든 병변 및 정상 조직에서 후두암 및 전암 병변을 진단하는 데에 있어 민감도 0.731, 특이도 0.922, 곡선 아래의 면적(area under curve, AUC) 0.922, 전체 정확도 0.867을 달성했다. 독립적인 테스트 세트에서 인간 전문가와 비교했을 때, 후두암 및 전암 병변의 감지에 대한 DCNN의 성능은 민감도 0.720, 특이도 0.948, AUC 0.953, 전체 정확도 0.897을 달성했으며, 이는 10-20년의 경력을 가진 경험 많은 인간 전문가와 유사한 수준이었다. 또한, DCNN의 후두암의 진단 전체 정확도는 0.773으로, 10-20년의 경력을 가진 인간 전문가와 유사하며 10년 미만의 경력을 가진 전문가들보다 우수하

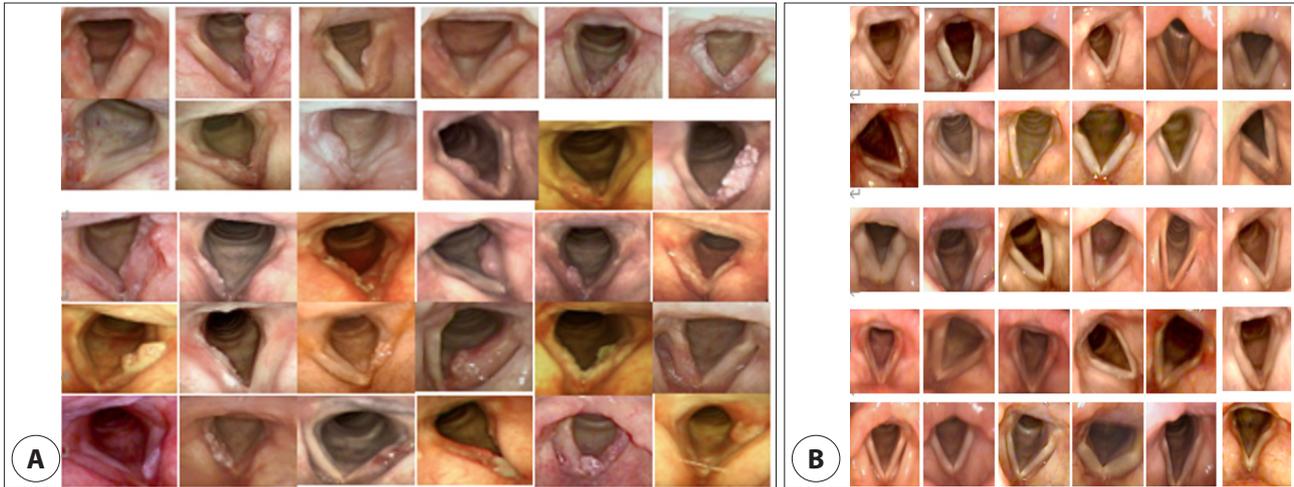


Fig. 1. Laryngoscopic images for videomics. A: Laryngoscopic images of glottic cancer. B: Laryngoscopic images of normal vocal fold.

였다.¹⁰⁾ 2015년 보고에 의하면 편평 상피 세포 T1a 후두암 환자 10명, 과각화증이 있는 전암성 병변이 있는 환자 10명, 후두 질환이 없는 10명의 피험자가 초당 4,000회의 이미지를 생성하는 초고속 후두 내시경 검사를 받았다. 음성 진동 그래프를 분석하는 기술인 웨이블릿 기반 음파 진동 분석(wavelet-based phonovibrographic analysis)을 통해 총 15,000개의 비디오 프레임으로 구성된 비디오에서 임상적으로 의미 있는 세 가지 진동 측정 세트를 추출했다. 통계 분석 결과 전암 병변 및 T1a 암종의 존재 시 성대 진동에 크게 영향을 받는 것으로 나타났다. 세 가지 척도를 바탕으로 서로 다른 임상 집단에 대해 개별 분류를 수행하는 지원 벡터 기계 학습 알고리즘(SVM[support vector machine]-learning algorithms)을 사용하여 식별 패턴을 추출했다. 모델의 성능을 평가하기 위한 leave-one-out 교차 검증 전략을 적용함으로써 이 연구의 저자들이 제안된 조치가 전암 병변과 T1a 암(민감도, 100%; 특이도, 100%)을 매우 높은 성능으로 구별한다는 것을 보여줄 수 있었다.¹³⁾ Cho 등은 후두경 화면을 캡처하여 각 이미지의 외전된 성대 영역을 포함하도록 잘라서 총 2,216개의 이미지(정상 899개, 비정상 1,317개)를 훈련(training), 검증(validation) 및 테스트 세트(test sets)에 할당하였다. 그 후 훈련 세트의 증강을 사용하여 합성곱 신경망 6개 레이어(CNN6), VGG16, Inception V3 및 Xception 모델 등 4가지의 모델을 훈련시켰다. 훈련된 모델은 테스트 세트에 적용되었고, 각 모델에 대해 수신자 조작 특성 곡선 및 cutoff값이 얻어졌다. 감도, 특이도, 양성 예측값, 음성 예측값 및 정확도가 계산되었다. 최상의 모델은 비디오 스트림에서 사용되었으며, Grad-CAM을 사용하여 특징의 지점화(localization)를

시도하였다. 모든 훈련된 모델은 수신자 조작 특성(receiver operating characteristics, ROC) 그래프에서 AUC가 높았으며, 정상성의 확률에 대한 가장 구별력 있는 cutoff값은 각각 CNN6, VGG16, Inception V3 및 Xception 모델에 대해 35.6%, 61.8%, 13.5%, 39.7%로 결정되었다. 테스트 세트에서 정상 및 비정상 성대를 선택하는 CNN 모델의 정확도는 각각 82.3%, 99.7%, 99.1%, 83.8%였다. 네 가지 모델 모두 허용 가능한 진단 정확도를 보여주었고, VGG16 및 Inception V3의 성능은 단순한 CNN6 모델과 최근에 발표된 Xception 모델보다 우수했다. 비디오 스트림에서 VGG16 모델, OpenCV 및 Grad-CAM의 조합을 사용한 실시간 분류는 후두경 검사에서 딥 러닝 모델의 잠재적인 임상 적용을 보여주었고 후두암의 진단에도 유용하게 이용될 수 있을 것이다.¹⁴⁾ 2019년 연구에서는 후두 내시경 이미지 데이터 세트에서 종양을 자동으로 감지하기 위해 CNN을 기반으로 일반적으로 사용되는 세 가지 물체 감지 모델을 사용했다. 연구에 사용된 CNN 기반 객체 검출 모델은 Faster R-CNN, YOLOv3 및 SSD(single shot multibox detector)였다. CNN을 기반으로 하는 물체 감지 모델이 후두 종양의 종단 간 감지(end-to-end detection)를 생성하는 데 사용된 것은 이번이 처음이다. 후두 종양의 검출을 위한 end-to-end detection 시스템은 일반적으로 후두경 이미지나 동영상과 같은 입력 데이터를 사용하고, 이를 전처리, 특징 추출 및 분류와 같은 일련의 단계를 거쳐 잠재적인 종양 영역을 식별하고 지역화하는 것을 말한다. 연구 결과 세 가지 방법 모두 좋은 성능을 보였다. 그 중 SSD가 평가 메트릭 측면에서 후두 종양 검출에 더 적합함을 보여주었다.¹⁵⁾

영상 데이터(CT[computed tomography], MRI[magnetic resonance imaging] 등)를 이용한 AI(artificial intelligence)의 적용

Radiomics는 디지털 방사선학 이미지에서 양적 특징을 추출하고 이를 분석하여 가설 생성과 검증을 위한 것을 의미한다(Fig. 2). 성문암 진단을 위해 컴퓨터 단층촬영(CT), 자기공명영상(MRI), 초음파, 광간섭단층촬영(optical coherence tomography)을 이용한 성대 진동 패턴 등 다양한 후두 관련 데이터 세트를 사용할 수 있다.¹⁶⁾ 2019년 Wang 등에 의한 보고에 따르면, 진행된 후두 암에 대한 외과적 의사 결정은 수술 전 T 범주(T3 대 T4)의 식별에 크게 의존하며 이는 외과 의사에게 어려운 일이라고 하였고, T 범주 예측 방사선학(T category prediction radiomics, TCPR) 모델은 후두 암의 수술에 도움이 될 것으로 예상했다. 이 연구에서는 전후두절제술을 받은 국소 진행성 후두암 환자 총 211명이 무작위로 훈련 코호트(n=150)와 검증 코호트(n=61)로 분류되었다. 조영 증강 CT 이미지에서 1,390개의 방사성 특징을 추출했고, 클래스 간 상관 계수 및 최소 절대 수축 및 선택 연산자(LASSO) 분석을 수행하여 병리학적으로 확인된 T 범주와 관련된 기능을 선택했다. 수술 전 T 범주와 관련된 8개의 방사선학적인 특징이 발견되었는데 여기에는 두 개의 1차 특징(`gradient_first order_Skewness`, `lbp.2D_first order_Mean`), 두 개의 형상 특징(`original_shape_LeastAxis`, `original_shape_Sphericity`), 그리고 네 개의 웨이블릿 특징(`wavelet-LLH_first order_Kurtosis`, `wavelet-LLH_glcmlfn`, `wavelet-LLH_first order_Median`, `wavelet-LLL_glcmlmc1`)을 포함하였다. 라디오믹 시그니처(radiomic signature)는 방사성 특징을 가진 support vector machine 알고리즘에 의해 구성되었고, 숙련된 방사선 전문의가 보고한 방사선 서명 및 T 범주를 통합한 노모그램이 개발되었다. 모델의 성능은 곡선 아래 면적으로 평가되었다. 방사선 전문의가

보고한 T 범주는 AUC 0.775(95% CI[confidence interval]: 0.667-0.883)를 달성했고, 라디오믹 시그니처는 0.862(95% CI: 0.772-0.952)의 상당히 높은 AUC를 산출했다. 방사선 전문의가 보고한 라디오믹 시그니처 및 T 범주를 통합한 노모그램의 예측 성능은 AUC가 0.892(95% CI: 0.811-0.974)로 더욱 향상되었다. 결과적으로, 국소적으로 진행된 후두암의 경우 경험이 풍부한 방사선 전문의가 보고한 라디오믹 시그니처와 T 범주를 통합한 TCPR 모델은 개별적으로 정확하게 수술 전 T 범주에 적용될 수 있는 큰 잠재력을 가지고 있었다. TCPR 모델은 전체 후두 절제술 또는 후두 보존 치료에 관한 의사 결정에 도움이 될 수 있을 것으로 사료된다.¹⁷⁾

갑상선 연골 침윤이 있는 후두 및 하인두 편평 세포 암종(laryngeal and hypopharyngeal squamous cell carcinoma, LHSCC)은 T4로 간주되며, 전체 후두 절제술이 필요하다. 그러나 갑상선 연골 침범의 수술 전 진단 정확도는 여전히 낮다. 2020년 보고된 연구는 LHSCC에서 갑상선 연골 침범을 예측할 때 CT 기반 방사능 기능의 가능성이 어느 정도인지를 평가하고자 하였다. 이 연구의 목적은 CT 기반의 라디오믹스 기능이 후두 경부 암으로부터 갑상선 연골 침윤을 예측하는 데 있어 잠재적으로 가지는 역할을 평가하는 것이었고, 앞에서 언급한 바와 같이 라디오믹스는 디지털 방식으로 촬영된 의료 영상에서 정량적인 특징을 추출하고, 이를 가지고 가설을 제시하고 검증하는 데이터 분석 기술을 말한다. 이 연구에서는 병리학적으로 입증된 LHSCC 환자 총 265명이 후향적 연구에 등록되었고(86명은 갑상선 연골 침범, 179명은 침범 없음), 2명의 두경부 영상의학과 전문의가 CT 영상에서 갑상선 연골 침범을 평가하였다. Radiomics 기능은 정맥 위상 조영 증강 CT 이미지에서 추출되었으며(LASSO 및 로지스틱 회귀(LR[logistic regression]) 방법을 치수 축소 및 모델 구성에 사용했다. 또한 SVM(SMOTE(support vector machine-based synthetic minority oversampling) 알고리즘

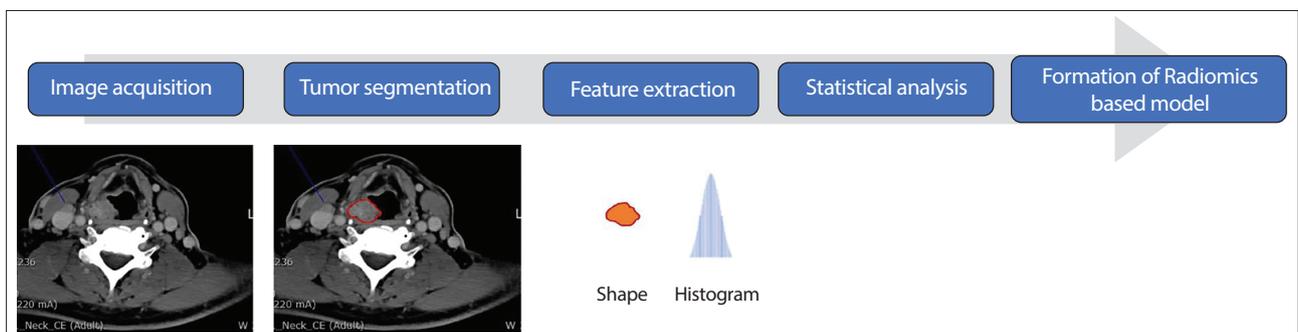


Fig. 2. The workflow of radiomics analysis.

을 채택하여 데이터 세트의 균형을 맞추고 새로운 LR-SVMS-MOTE 모델을 구성했다. 방사선과 전문의와 두 모델의 성능을 ROC(수신기 작동 특성) 곡선으로 평가하고 DeLong 테스트를 사용하여 비교했다. LR-SVMSMOTE 모델, LR 모델, 영상의학과 의사에 대한 LHSCC의 갑상선 연골 침범 예측에서 ROC 그래프의 AUC는 0.905(95% 신뢰 구간[CI]: 0.863-0.937), 0.876(95% CI: 0.830-0.913) 및 0.721(95% CI: 0.663-0.774)이었다. 앞 두 모델의 AUC는 방사선 전문의의 평가보다 높았다(all $p < 0.001$). LR-SVMSMOTE와 LR 모델 사이의 예측 성능에는 큰 차이가 없었다($p=0.05$). CT 방사선 특성에 기반한 모델은 LHSCC에서 갑상선 연골 침범을 예측하는 정확도를 개선하고 LHSCC에서 갑상선 연골 침범을 수술 전 예측하기 위한 잠재적으로 새로운 비침습적 방법을 제공할 수 있었다.¹⁸⁾

음성 자료(acoustic data)를 이용한 AI(artificial intelligence)의 적용

2018년 Tsui 등의 연구에서는 발성 장애 증상에 대한 설문지를 사용해서 인구통계학적 특징과 증상적 특징의 차별적 표현이 전산화된 분류에 적용될 수 있는지 여부를 조사했다. 성문부위 종양 100명, 음성의상성 병변 508명, 일측성 성대마비 153명을 모집하여 통계 분석 결과 인구통계학적 변수와 증상 변수의 분포가 상당히 다른 것으로 나타났다. 의사결정 트리, 선형 판별 분석, K-최근접 이웃, 지원 벡터 머신 및 인공 신경망을 포함한 기계 학습 알고리즘을 적용하여 음성 장애를 분류했다. 결과적으로 인구통계학적 특성이 종양 및 음성의상성 병변을 감지하는 데 더 효과적인 반면, 증상은 성대마비를 감지하는 데 유용하다는 것을 보여주었다. 인구통계학적 변수와 증상 변수를 결합했을 때 인공 신경망은 $83 \pm 1.58\%$ 의 가장 높은 정확도를 달성한 반면 다른 알고리즘은 74%-82.6% 범위의 정확도를 달성했다. 의사결정나무 분석 결과 성별, 연령, 흡연 여부, 발성장애의 갑작스러운 발병, 10개 항목의 음성 장애 지수 점수가 분류에 중요한 특성인 것으로 나타났다. 결론적으로 이 연구에서는 성문 부위 종양, 음성의상성 병변, 성대마비 사이에 인구통계학적 특징과 증상적 특징에서 유의미한 차이를 보였다. 이러한 기능은 기계 학습 알고리즘을 통해 음성 장애의 자동 분류를 용이하게 할 수 있었다.¹⁹⁾ 음성 변화는 후두암의 초기 징후일 수 있다. 2020년 보고된 연구에서는 자동 음성 신호 분석을 사용하여 후두암 환자와 건강한 피험자를 구별할 수 있는지 여부를 조사했다. PRAAT(Speech Analysis in Phonetics) 소프트웨어 패

키지를 사용하여 음성의 특징을 추출하고 /a:/ 모음 소리의 음성 샘플에서 Mel-frequency cepstral 계수를 계산했다. 제안된 방법은 SVM, XGBoost(extreme gradient boosting), LGBM(light gradient boosted machine), ANN(artificial neural network), 1D-CNN(1D convolutional neural network), 2차원 컨볼루션 신경망(2D-CNN)이었다. 이용된 인공 지능의 성능은 정확도, 민감도 및 특이성 측면에서 평가되었다. 결과는 인간의 성능과 비교되었는데 총 4명의 자원봉사자 중 2명은 후두를 전문으로 전공한 전문의로 동일한 파일을 평가했다. 1D-CNN은 85%의 가장 높은 정확도와 78%와 93%의 민감도 및 특이도를 보였다. 두 명의 전문의는 69.9%의 정확도를 달성했지만, 민감도 수준은 44%였다. AI를 이용한 음성 신호의 자동 분석은 4명의 지원자가 수행한 것보다 진단의 정확도가 더 높게 나타나 건강한 대상과 후두암 대상을 구별할 수 있었다.²⁰⁾ 2022년 보고된 연구에서는 상대적으로 작은 데이터세트에 대한 분류 정확도를 향상시키기 위해 의사결정 트리 앙상블 학습 방법과 컨볼루션 신경망(CNN) 알고리즘의 결과를 비교하여 성문암 진단에 대한 정확도를 향상하고자 하였다. 본 병원의 데이터세트를 사용하여 분류기를 구축하고 분원 병원의 데이터세트를 사용하여 생성된 모델의 분류기 성능을 검증하였다. 우선 성문암의 진단을 위해 성문 부위를 촬영한 후두경 이미지와 음성 데이터를 사용하여 딥러닝 기반의 CNN 모델을 구축하고 분류하였다(Fig. 3). CNN 분류 알고리즘을 통해 확률을 사용하여 의사결정 트리 앙상블 학습을 수행하여 분류 정확도를 얻었고, 이 과정에서 분류 및 회귀 트리(CART) 방법을 사용하였다. 그런 다음, 성문 이미지와 음성 의사결정 트리 분류기를 융합하여 의사결정 트리 앙상블 학습의 분류 정확도를 CNN 개별 분류기와 비교하였다. 그 결과 본 병원 훈련 데이터세트를 사용하여 구축된 성문 이미지와 음성 분류 모델에서 얻은 분류 정확도를 각각 81.03%와 99.18%이었다. 그러나 분원의 외부 데이터세트를 사용할 때에는 CNN 분류기의 정확도는 음성에서 73.88%, 성문 이미지에서 68.92%로 감소했다(Fig. 4). 이 문제를 해결하기 위해 성문 이미지와 음성의 의사결정 트리 앙상블 학습을 사용하여 동일한 개인의 성문 이미지와 음성 데이터를 통합하여 분류 정확도를 개선했다(Fig. 5). 개인화된 성문 이미지와 음성 의사결정 트리 모델의 분류 정확도는 각각 87.88%와 89.06%이며, 성문 이미지와 음성 의사결정 트리 결과를 융합한 분류 정확도는 95.31%를 나타냈다. 결론적으로 작은 데이터세트에도 불구하고 분류 정확도를 향상시키기 위해서는 여러 분류기를 훈련하는 의사결정 트리 앙상블 학습이 유용하다는 것을 시사

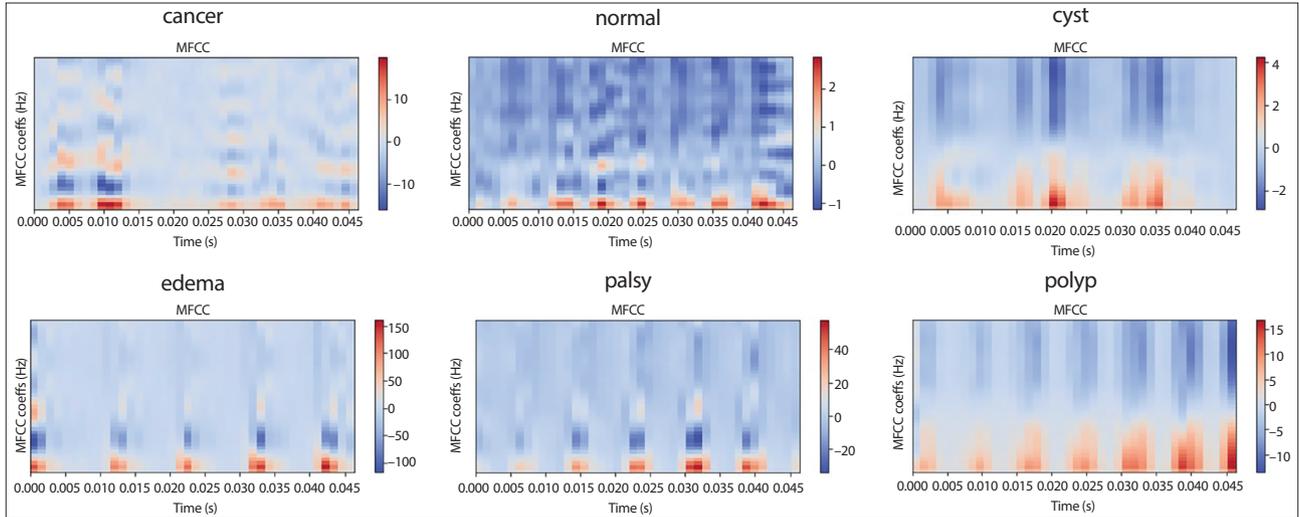


Fig. 3. Acoustic data for diagnosis of laryngeal cancer using AI. AI: artificial intelligence.

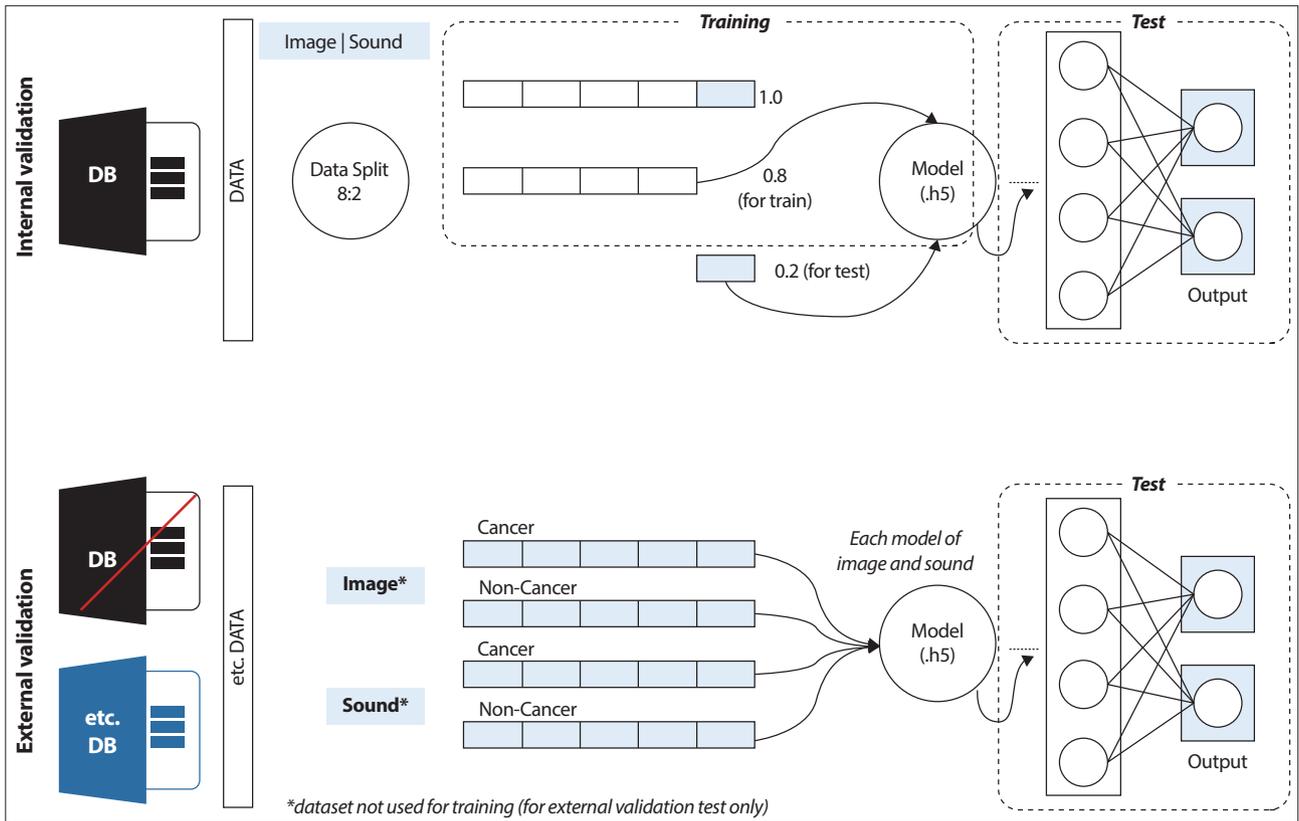


Fig. 4. The schematic layout of external validation for evaluating the accuracy of a CNN classifier. CNN: convolutional neural network.

한다. 인공지능 분석에는 대량의 데이터가 필수적이지만, 다양한 입력 데이터를 결합하여 통합적인 접근 방식을 취할 때 진단 분류 정확도가 높아질 수 있다는 결론을 얻었다. 결국 이 연구에서는 성분 영상 이미지와 음성의 특징을 융합하여 서로 보완적인 역할을 하여 정확도가 개선된 결과를 얻었다.¹⁶⁾

임상적 자료(clinical data)를 이용한 AI(artificial intelligence)의 적용

증상과 인구 통계 데이터를 분석하여 음성이상성 양성 병변, 악성 후두 병변 및 성대 마비를 구분하기 위해 기계 학습 알고리즘을 사용했다. 그들은 100명의 후두 악성종양 환

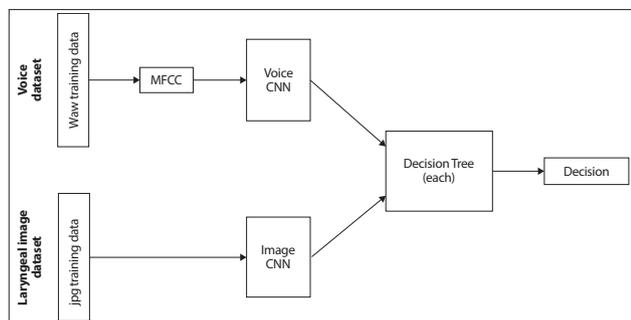


Fig. 5. Structure of basic classifier. MFCC: Mel-frequency cepstral, CNN: convolutional neural network.

자, 509명의 음성외상성 병변 환자 및 153명의 일측성 성대마비 환자로 구성된 단일 기관의 개인 데이터 세트를 사용했다. 그들의 결과는 인구 통계 데이터를 사용하여 양성 및 음성 병변을 진단하는 데 높은 임상 정확성을 보였으며, 증상 변수는 성대마비를 진단하는 데 더 유용했다.¹⁹⁾ Howard 등은 수술 후 절제연이 양성이거나 피막 외침범이 있는 중간위험(intermediate)의 편평세포 두경부암(후두암 포함) 환자들 중 어떤 환자가 보조 화학요법을 통해 이익을 얻을 수 있는지를 결정하기 위해 기계 학습 생존 모델을 사용한 다른 대규모 연구를 수행했다. 후두암을 포함한 두경부암에 대한 구체적인 데이터 세트는 자국의 국립암 데이터베이스(National Cancer Database)에서 얻었다. 인구 통계 데이터, 암 특이 요인 및 치료 특이 요인을 후두암 환자 5,631명에 대해 수집되었다. 기계 학습 생존 모델은 방사선 치료만을 받는 경우와 화학방사선 치료의 생존 이익을 예측하기 위해 훈련되었으며, 이러한 예측에 기초하여 생존 이익이 있는 환자 하위 그룹과 보조 화학요법에서 생존 이익이 없는 환자 하위 그룹을 식별할 수 있었다.²¹⁾

게노믹스(genomics)를 이용한 AI(artificial intelligence)의 적용

TP53, CDKN2A, PIK3CA, KMT2D, CCND1 등과 같은 유전적 변화는 후두암에서 발견될 수 있는 유전자 무효화 또는 과발현을 일으킨다.⁶⁾ 2017년도에 보고된 연구는 대규모 데이터 분석 방법을 사용하여 후두암에서 재발을 예측하기 위한 유전자 세트를 식별하는 것을 목표로 수행되었다. 후두암에 대한 두 개의 유전자 발현 프로파일 데이터(GSE27020 및 GSE25727)를 공공 데이터베이스에서 다운로드하였고, 종양 재발과 관련된 유전자인 정보 유전자를 Cox 회귀 분석을 통해 식별하였다. 그런 다음 정보 유전자로 구성된 단백질-단백질 상호작용(protein-protein interaction, PPI) 네트워크를

구축하였다. 이후 특정 PPI 네트워크 내의 유전자를 기반으로 재발된 후두암 샘플을 분류하기 위해 최적화된 서포트 벡터 머신(SVM) 분류기를 구축하였다. 또한, SVM 분류기의 효율성은 다른 두 개의 독립적인 데이터 세트로 검증되었다. 이 연구에서는 결과적으로 GSE27020 및 GSE25757 데이터셋으로부터 총 331개의 정보 유전자가 얻어졌다. 후두암 재발과 관련된 특정 PPI 네트워크가 구성되었으며, 이 네트워크는 정보 유전자와 중요한 비-정보 유전자로 구성되었다. 특정 PPI 네트워크에서 상위 10개 유전자는 APP, NTRK1, TP53, PTEN, FN1, ELAVL1, HSP90AA1, XPO1, LDHA 및 CDK2이며, betweenness centrality값에 따라 순위가 매겨졌다. 상위 80개 유전자를 포함한 최적화된 SVM 분류기는 후두암 샘플에서 재발 사례를 분류하는 데 100%의 정확도를 보였다. SVM 분류기의 재발 샘플 예측 효율성은 다른 독립적인 두 개의 데이터셋에서 검증되었으며, 97.47%의 정확도를 보였다. 최적화된 SVM 분류기의 정보 유전자는 종양 진행과 관련된 여러 경로에 다양하게 품고 있었다. 결론적으로 이 연구에서는 80개의 유전자 세트가 후두암 재발을 예측하는 바이오 표지자로 식별되었으며, 이는 재발 위험이 다른 환자들에게 다른 치료 방법을 결정하는 데 잠재적으로 유용하게 적용될 수 있을 것으로 사료된다.²²⁾

결론

후두암에서 AI의 적용은 초기 탐지, 분류 및 예후 예측, 치료 계획 및 개인화된 치료, 임상 결정 지원 등 다양한 측면에서 유용하다. 후두암의 초기 단계에서는 증상이 미미할 수 있어서 진단이 어려울 수 있으나, AI는 후두암 환자의 음성 데이터 혹은 영상 데이터를 분석하여 종양의 특성을 식별하고, 종양의 크기, 위치, 전이 등을 예측할 수 있다. 이를 통해 환자의 예후를 예측하고, 적절한 치료 전략을 제안할 수 있다.

예를 들어, 종양의 위치와 크기에 따라 수술, 방사선 치료, 화학 요법 등의 치료 방법을 최적화하거나, 환자의 개인적인 특성과 반응을 고려하여 치료 일정을 조정할 수 있다. 또한, 환자의 예후를 예측하고 치료 결과를 모니터링하는 데에도 활용할 수 있다. 하지만 AI의 적용은 의료 전문가와의 협력과 검증이 필요하며, 신중한 접근과 윤리적 고려가 함께 이루어져야 할 것으로 사료된다. 즉, AI는 의료진의 의사결정을 보조하는 도구로 활용될 수 있지만, 최종적인 진단과 치료 결정은 의료 전문가의 판단에 근거해야 한다.

Acknowledgements

Not applicable.

Funding Information

This study was supported by a 2-year Research Grant of Pusan National University.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

ORCID

Jin Choon Lee, <https://orcid.org/0000-0002-4056-8763>

Author Contribution

The article is prepared by a single author.

Ethics Approval

Not applicable.

References

1. Bensoussan Y, Vanstrum EB, Johns MM 3rd, Rameau A. Artificial intelligence and laryngeal cancer: from screening to prognosis: a state of the art review. *Otolaryngol Head Neck Surg* 2023;168(3):319-29.
2. Paderno A, Holsinger FC, Piazza C. Videomics: bringing deep learning to diagnostic endoscopy. *Curr Opin Otolaryngol Head Neck Surg* 2021;29(2):143-8.
3. Koçak B, Durmaz EŞ, AteŞ E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol* 2019;25:485-95.
4. Ezzine K, Ben Hamida A, Ben Messaoud Z, Frikha M. Towards a computer tool for automatic detection of laryngeal cancer. *Proceedings of the 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP): 2016 Mar 21-23: Monastir.*
5. André F, Arnedos M, Baras AS, Baselga J, Bedard PL, Berger MF, et al. AACR project GENIE: powering precision medicine through an international consortium. *Cancer Discov* 2017;7(8):818-31.
6. Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P. Machine learning and integrative analysis of biomedical big data. *Genes* 2019;10(2):87.
7. Crowson MG, Ranisau J, Eskander A, Babier A, Xu B, Kahmke RR, et al. A contemporary review of machine learning in otolaryngology 2nd International Conference on Adv 2020;130(1):45-51.
8. Yamada M, Saito Y, Imaoka H, Saiko M, Yamada S, Kondo H, et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci Rep* 2019;9(1):14465.
9. Yao P, Usman M, Chen YH, German A, Andreadis K, Mages K, et al. Applications of artificial intelligence to office laryngoscopy: a scoping review. *Laryngoscope* 2022;132(10):1993-2016.
10. Xiong H, Lin P, Yu JG, Ye J, Xiao L, Tao Y, et al. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine* 2019;48:92-9.
11. Ren J, Jing X, Wang J, Ren X, Xu Y, Yang Q, et al. Automatic recognition of laryngoscopic images using a deep-learning technique. *Laryngoscope* 2020;130(11):E686-93.
12. Yao P, Witte D, Gimonet H, German A, Andreadis K, Cheng M, et al. Automatic classification of informative laryngoscopic images using deep learning. *Laryngoscope Investig Otolaryngol* 2022;7(2):460-6.
13. Unger J, Lohscheller J, Reiter M, Eder K, Betz CS, Schuster M. A noninvasive procedure for early-stage discrimination of malignant and precancerous vocal fold lesions based on laryngeal dynamics analysis. *Cancer Res* 2015;75(1):31-9.
14. Cho WK, Choi SH. Comparison of convolutional neural network models for determination of vocal fold normality in laryngoscopic images. *J Voice* 2022;36(5):590-8.
15. Yan P, Li S, Zhou Z, Liu Q, Wu J, Ren Q, et al. Automated detection of glottic laryngeal carcinoma in laryngoscopic images from a multicentre database using a convolutional neural network. *Clin Otolaryngol* 2023;48(3):436-41.
16. Kwon I, Wang SG, Shin SC, Cheon YI, Lee BJ, Lee JC,

- et al. Diagnosis of early glottic cancer using laryngeal image and voice based on ensemble learning of convolutional neural network classifiers. *J Voice*. Forthcoming 2022. <https://doi.org/10.1016/j.jvoice.2022.07.007>
17. Wang F, Zhang B, Wu X, Liu L, Fang J, Chen Q, et al. Radiomic nomogram improves preoperative T category accuracy in locally advanced laryngeal carcinoma. *Front Oncol* 2019;9:1064.
 18. Guo R, Guo J, Zhang L, Qu X, Dai S, Peng R, et al. CT-based radiomics features in the prediction of thyroid cartilage invasion from laryngeal and hypopharyngeal squamous cell carcinoma. *Cancer Imaging* 2020;20(1):81.
 19. Tsui SY, Tsao Y, Lin CW, Fang SH, Lin FC, Wang CT. Demographic and symptomatic features of voice disorders and their potential application in classification using machine learning algorithms. *Folia Phoniatr Logop* 2018;70(3-4):174-82.
 20. Kim H, Jeon J, Han YJ, Joo Y, Lee J, Lee S, et al. Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy. *J Clin Med* 2020;9(11):3415.
 21. Howard FM, Kochanny S, Koshy M, Spiotto M, Pearson AT. Machine learning-guided adjuvant treatment of head and neck cancer. *JAMA Netw Open* 2020;3(11):e2025881.
 22. Yang B, Guo Q, Wang F, Cai K, Bao X, Chu J. A 80-gene set potentially predicts the relapse in laryngeal carcinoma optimized by support vector machine. *Cancer Biomark Sect Dis Markers* 2017;19(1):65-73.