



pISSN 2005-8063
eISSN 2586-5854
2022. 12. 31.
Vol.14 No.4
pp. 45-56

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

<https://doi.org/10.13064/KSSS.2022.14.4.045>



Automatic detection and severity prediction of chronic kidney disease using machine learning classifiers*

Jihyun Mun¹ · Sunhee Kim² · Myeong Ju Kim³ · Jiwon Ryu⁴ · Sejoong Kim^{3,4,5} · Minhwa Chung^{1,**}

¹*Department of Linguistics, Seoul National University, Seoul, Korea*

²*Department of French Language in Education, Seoul National University, Seoul, Korea*

³*Center of Artificial Intelligence in Healthcare, Seoul National University, Seoul, Korea*

⁴*Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Korea*

⁵*Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea*

Abstract

This paper proposes an optimal methodology for automatically diagnosing and predicting the severity of the chronic kidney disease (CKD) using patients' utterances. In patients with CKD, the voice changes due to the weakening of respiratory and laryngeal muscles and vocal fold edema. Previous studies have phonetically analyzed the voices of patients with CKD, but no studies have been conducted to classify the voices of patients. In this paper, the utterances of patients with CKD were classified using the variety of utterance types (sustained vowel, sentence, general sentence), the feature sets [handcrafted features, extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS), CNN extracted features], and the classifiers (SVM, XGBoost). Total of 1,523 utterances which are 3 hours, 26 minutes, and 25 seconds long, are used. F1-score of 0.93 for automatically diagnosing a disease, 0.89 for a 3-classes problem, and 0.84 for a 5-classes problem were achieved. The highest performance was obtained when the combination of general sentence utterances, handcrafted feature set, and XGBoost was used. The result suggests that a general sentence utterance that can reflect all speakers' speech characteristics and an appropriate feature set extracted from there are adequate for the automatic classification of CKD patients' utterances.

Keywords: chronic kidney disease, machine learning, automatic classification

* This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2018-0-01833) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). This work was supported by grant no 13-2022-0008 from the SNUBH Research Fund.

** mchung@snu.ac.kr, Corresponding author

Received 15 November 2022; Revised 6 December 2022; Accepted 6 December 2022

© Copyright 2022 Korean Society of Speech Sciences. 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.

1. 서론

만성콩팥병(chronic kidney disease, CKD)은 신장 기능의 만성적 감소 혹은 신장의 구조적 손상을 특징으로 하는 질병으로, 전 세계적으로 유병률이 높고 계속 증가하는 추세인 질병이다(Webster et al., 2017). 만성콩팥병의 초기에는 증상이 나타나지 않을 수 있는데, 초기 진단 및 치료가 적절하게 이루어지지 않으면 투석과 같은 신대체 요법을 필요로 하게 되기 때문에 초기 진단 및 치료가 중요한 질병이다. 그러나 혈액 혹은 소변 검사와 같은 방식으로만 질병의 진단이 가능하기 때문에 지속적인 건강 검진 혹은 병원 방문을 필요로 하여 초기에 질병을 발견하거나 지속적으로 질병의 진행을 확인하는 데에는 어려움이 존재한다. 따라서 병원 방문 없이도 비침습적이고 반복적으로 질병을 자동으로 진단하고 경과를 포착할 수 있는 보조적인 지표가 필요하다.

만성콩팥병은 심혈관, 신경계, 근골격계, 면역계, 내분비계, 대사계 등 신체의 다양한 기관에 영향을 미치는데, 그중 특히 호흡계에 크게 영향을 미친다. 폐와 신장은 신체의 산-염기 균형을 유지하는 데에 연관되어 있어, 신장의 변화는 호흡계에 장애를 일으킬 수 있다(Kumar & Bhat, 2010). 이로 인한 만성콩팥병 환자의 일반적인 특성으로는 호흡계와 후두 근육의 기능이 손상되고, 건강한 사람에 비해 호흡 근육의 지구력과 강도가 감소한다는 것이 있다(Hassan, 2014). 특히 말기 만성콩팥병 환자의 경우에는 요독 축적, 산-염기 불균형, 체적 과부하 등의 특성으로 인해 폐 기능이 감소하고, 성대 부종이 발생한다(Jung et al., 2014).

이처럼 만성콩팥병으로 인한 호흡계와 후두 근육의 손상이 음성의 변화를 야기함에 따라 만성콩팥병 환자의 음성의 특징에 대한 연구들이 다수 존재한다. 선행연구들에서는 /a/ 모음의 연장 발화에서 음질[jitter, shimmer, harmonics-to-noise ratio(HNR)], 음높이(F0), 공기역학적(최대연장발성시간) 특징을 추출하여 만성콩팥병 환자의 음성의 특성을 분석하고 대조군과 비교하였다(Abd El-gaber et al., 2021; Hassan, 2014; Kumar & Bhat, 2010; Mudawwar et al., 2017; Zaky et al., 2020).

Shimmer의 경우 환자군이 대조군보다 높은 값을 보이고, 최대연장발성시간의 경우 환자군이 대조군보다 낮은 값을 보인다는 공통된 결과가 나타났다. 그러나 jitter, HNR, F0의 경우 선행연구들 간에 상반되는 결과가 나타났다. Hassan(2014), Kumar & Bhat(2010), Zaky et al.(2020)에서는 환자군이 대조군보다 높은 jitter값을 보인다고 밝힌 반면, Abd El-gaber et al.(2021)에서는 환자군이 대조군보다 낮은 jitter값을 보인다고 밝혔다. HNR의 경우 Mudawwar et al.(2017), Zaky et al.(2020)에서는 환자군이 대조군보다 높은 HNR값을 보인다고 밝힌 반면, Abd El-gaber et al.(2021)과 Hassan(2014)에서는 환자군이 대조군보다 낮은 HNR값을 보인다고 밝혔다. F0의 경우 Abd El-gaber et al.(2021), Hassan(2014), Kumar & Bhat(2010), Zaky et al.(2020)에서는 환자군이 대조군보다 높은 F0값을 보인다고 밝혔고, Mudawwar et al.(2017)에서는 환자군이 대조군보다 낮은 F0값을 보인다고 밝

혔다. 이처럼 선행연구들 간에 상반되는 결과가 나타났기 때문에 일부 특징들에 대해 만성콩팥병 환자의 음성 특성을 파악하기에는 어려움이 있었다.

또한 선행연구들에서는 대부분 모음연장발화를 분석하였는데, 이는 문장 기반의 실제 발화 특성을 잘 반영하지 못할 가능성이 존재한다(Moon et al., 2012). 따라서 선행연구들의 결과로는 실제 발화 상황에서의 만성콩팥병 환자의 음성 특성을 파악할 수 없었다. 그리고 선행연구들에서는 만성콩팥병의 단계에 따른 차이를 보고하지 않았고, 환자군과 대조군 간의 차이만 보고하여 질병의 진행에 따른 음성의 변화를 포착할 수 없었다.

기존 선행연구들의 한계점을 보완하기 위해 Mun et al.(2022)에서는 환자군 및 대조군의 모음연장발화, 유성음 문장 발화, 문단 발화를 포함하는 코퍼스를 구축하고, 해당 코퍼스를 사용하여 만성콩팥병 환자의 음성을 분석하였다. Mun et al.(2022)에서는 다양한 발화에서 추출한 특징들 중 환자군과 대조군 간에 유의한 차이가 존재하는 특징을 밝히고, 사구체여과율(estimated glomerular filtration rate, eGFR)과 특징들 간의 상관관계 및 인과관계를 파악하였다.

병리적인 음성을 자동으로 진단하는 연구가 다수 존재하고, 진단에 사용되는 모델은 머신러닝 모델과 딥러닝 모델로 나뉜다. 그러나 병리적 음성을 포함하는 코퍼스는 대부분 딥러닝 모델을 적용하기에는 크기가 충분히 크지 않기 때문에, 머신러닝 모델 기반의 방법론이 여전히 널리 사용되고 있다. 머신러닝 모델을 사용하여 병리적 음성을 진단할 때에는 다양한 특징들이 사용되는데, 이때의 특징들은 크게 음향적 특징과 딥러닝 모델로부터 추출한 특징으로 나눌 수 있다. 음향적 특징으로는 Mel-frequency cepstral coefficients(MFCCs), jitter, shimmer, HNR 등의 음질 특징, 음높이 특징 등이 주로 사용된다. MFCCs는 병리적 음성을 진단할 때 가장 많이 사용되는 특징으로, 인간의 음성으로부터 유용한 정보를 추출할 수 있고, 병리적 음성 분류에 효과적인 특징이라고 알려져 있다(Shetty et al., 2018). 음질, 음높이 특징들은 음성 자동 분류에 앞서 병리적 음성의 특징을 파악하기 위해 주로 사용되는 전통적인 특징들이며, 해당 특징들을 사용하여 음성을 분류했을 시에는 해석 가능한 결과를 얻을 수 있다는 이점이 존재한다(Harar et al., 2020).

그러나 위와 같은 음향적 특징들을 사용하여 음성을 분류하기 위해서는 음성의 특성에 맞는 특징들과 분류기를 수동으로 선정하는 데에 어려움이 존재한다. 따라서 병리적 음성을 자동으로 분류하기 위해 표준 음향 특징 목록인 eGeMAPS(extended Geneva Minimalistic Acoustic Parameter Set)가 제안되었고(Eyben et al., 2015), eGeMAPS는 병리적 음성과 호흡계 장애로 인해 나타나는 음성을 분류하는 데에 좋은 성능을 보인다는 실험 결과들이 존재한다(Liu et al., 2018; Triantafyllopoulos et al., 2022). 또한 딥러닝 기반 모델에서 특징들을 자동으로 추출하여 분류 성능을 높이는 연구들이 진행되고 있고(Narendra & Alku, 2020; Omeroglu et al., 2022), 주로 CNN(convolutional neural network)을 사용하여 특징을 추출한다. CNN은 이미지뿐 아니라 오디오를 분류하는 데에도 좋은 성능을 내고 있어(Hershey et

al., 2017), 병리적 음성 분류에도 다수 적용되고 있다.

본 연구에서는 Mun et al.(2022)에서 수집한 다양한 발화 유형과 현재 병리적 음성 분류에서 사용되고 있는 다양한 특징 집합을 사용하여 만성콩팥병 환자의 음성을 진단하고, 중증도를 예측하는 방법을 제안한다. 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안하는 방법을 소개하고, 3장에서는 실험에 사용된 음성 데이터와 진행한 실험에 대해 설명한다. 4장에서는 실험 결과를 요약하며 결과에 대해 논의하고, 5장에서는 본 연구의 의의 및 한계점과 향후 진행할 연구에 대해 소개하며 마무리한다.

2. 방법론

2.1. 만성콩팥병 음성 자동 진단 및 중증도 분류

본 연구에서는 다양한 발화 데이터와 특징 집합을 사용하여 만성콩팥병 음성을 자동으로 진단하고 중증도를 분류하는 방법을 제안한다. 그림 1은 본 연구에서 제안하는 방법을 도식화한 것이다. 먼저 세 가지 문장 유형, 즉 모음연장발화, 유성음 문장 발화, 일반 문장 발화에서 각각 특징을 추출한다. 추출하는 특징으로는 다시 수제 특징, eGeMAPS, 그리고 CNN 추출 특징 집합이 있다. 추출한 특징들을 머신러닝 분류기의 입력값으로 넣어주면 분류기는 입력된 특징들을 바탕으로 만성콩팥병 자동 진단 실험에서는 음성을 대조군과 환자군으로 분류하고, 중증도 예측 실험에서는 음성을 모델이 예측한 만성콩팥병 단계의 환자군 음성으로 분류한다. 분류기의 성능은 weighted F1-score로 평가한다.

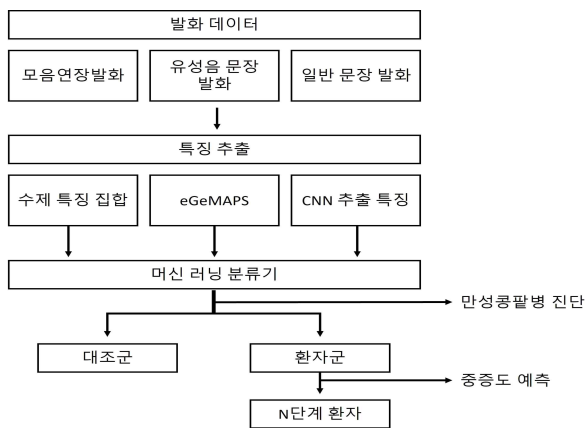


그림 1. 제안하는 방법
Figure 1. Proposed method

2.2. 발화 유형

Mun et al.(2022)에서 구축한 코퍼스 데이터 중 본 연구에서 실험에 사용한 발화 유형은 표 1과 같다. 모음연장발화는 음성 분석에 가장 많이 사용되는 발화 목록이다(Moon et al., 2012). 그 중 /아/ 모음은 가장 높은 제 1 포먼트를 가져 첫 번째 혹은 두 번째 배음값에 크게 영향을 미치지 않는 것으로 알려져 있어

(Ahn, 2000) 음성 분석에 가장 많이 사용된다. 모음연장발화에서는 스펙트럼, 음질, 음높이, 성문(glottal) 특징, 최대연장발성 시간을 추출하였다.

다음으로는 유성음으로만 이루어진 문장 발화를 사용하였다. 모음연장발화는 문장 기반의 실제 발화 특성을 잘 반영하지 못할 수 있고(Moon et al., 2012), Mun et al.(2022)에서 모음연장 발화와 문장 발화에서 다른 음성 분석 결과가 나왔다고 보고했다. 따라서 유성음 문장 발화에서 모음연장발화에서와 동일한 특징들을 추출하였을 때 분류기의 성능에 차이가 존재하는지 확인하고자 한다. 유성음 문장 발화에서는 스펙트럼, 음질, 음높이, 말속도 특징을 추출하였다.

마지막으로 Mun et al.(2022)의 코퍼스에서 수집한 6개의 문장으로 이루어진 문단 발화 중 첫 번째 문장을 사용하였다. 유성음 문장 발화보다 길이가 긴 문장을 사용하여 실제 발화상에서 나타날 것으로 예상되는 호흡 특징을 추출하고자 하였다. 일반 문장 발화에서도 스펙트럼, 음질, 음높이, 말속도 특징을 추출하였다.

표 1. 발화 유형 분류
Table 1. Classification of speech types

발화 유형	내용
모음연장발화	/아/
유성음 문장 발화	오월 오일은 어린이날이에요.
일반 문장 발화	일상이 문득 너무 무덤덤할 때는 여행 같은 특효약이 또 있을까.

2.3. 특징 집합

2.3장에서는 실험에 사용된 세 가지 특징 집합과 그에 속하는 특징들에 대해 설명한다.

2.3.1. 수제 특징 집합

본 연구에서 사용한 수제 특징 집합은 Yeo et al.(2021)에서 제시한 특징 리스트를 토대로 구성하였고, 스펙트럼, 음질, 공기역학적, 성문, 운율 특징이 이에 속한다. 전체 특징 리스트는 표 2에 제시하였다. Yeo et al.(2021)에서는 스펙트럼, 음질, 운율, 발음 특징을 사용하였는데, 그중 만성콩팥병 환자의 음성 특성을 반영할 것으로 예상되는 스펙트럼, 음질, 운율(발화 속도, 음높이) 특징을 본 연구에 사용하였다. Yeo et al.(2021)에서 사용한 발음 및 리듬 특징은 조음과 관련된 것으로, 만성콩팥병 음성과는 무관할 것으로 판단하여 제외하였다. 위 특징들에 더불어 만성콩팥병 음성의 특징을 반영하는 공기역학적, 성문, 심 관련 특징을 추가하여 수제 특징 집합을 구성하였다.

표 2. 수제 특징 리스트
Table 2. Handcrafted feature list

특징 종류	특징
스펙트럼	Mel frequency cepstral coefficients (MFCCs)
음질	Jitter, shimmer, harmonics to noise ratio (HNR), voice breaks 개수, voice breaks 정도
공기역학적	최대연장발성시간
성문	H1-H2, H1-A1, H1-A2, H1-A3
운율	음높이
	F0 평균값/표준편차/최솟값/최댓값/중앙값/25분위수/75분위수
말속도	전체 길이, 말 길이, 말 속도, 조음 속도, 쉼의 수, 쉼의 길이

스펙트럼 특징인 MFCCs는 소리의 고유한 특징을 나타내는 수치로, 주로 음성인식, 화자 인식, 음성 합성 등 오디오 도메인의 문제를 해결하는 데 일반적으로 사용되는 특징이며, 의료 목적으로 음성을 평가하는 데에도 사용되고 있다(Benba et al., 2015). 본 연구에서는 Librosa(McFee et al., 2015)를 사용하여 각 발화에서 12차원 MFCCs와 로그 에너지를 추출하였다.

음질 특징으로는 jitter, shimmer, HNR, voice breaks 개수, voice breaks 정도를 사용하였다. Jitter는 시간에 따른 F0값의 변화를 나타내는 지표이고, shimmer는 시간에 따른 진폭의 변화를 나타내는 지표이다. HNR은 잡음에 대한 배음의 비율을 말한다. Jitter, shimmer, HNR은 음성의 특성을 설명하고 병리적인 음성을 진단하는 데에 사용된다(Teixeira et al., 2013). Voice breaks 관련 특징들은 화자의 유성음 유지 능력을 확인할 때 사용되는 특징이다(Yeo et al., 2021).

공기역학적 특징으로는 최대연장발성시간(maximum phonation type, MPT)을 사용하였고, 이는 발화에서 호흡 능력을 객관적으로 측정할 수 있는 특징이고, 숨을 들이마신 후 모음을 최대한 유지하는 능력으로 정의된다(Speyer et al., 2010).

성문 특징으로는 H1-H2, H1-A1, H1-A2, H1-A3를 사용하였고, 이 특징들은 성대의 불완전한 폐쇄 및 발생 중 성대가 열려 있는 주기와 관련되어 있다고 알려져 있다(Lee et al., 2015).

운율 특징 중 음높이 특징으로는 F0의 평균값, 표준편차, 최솟값, 최댓값, 중앙값, 25 분위수, 75 분위수를 사용하였다. 음질, 공기역학적, 성문, 음높이 특징은 모두 Praat(Boersma, 2001)을 사용하여 추출하였다.

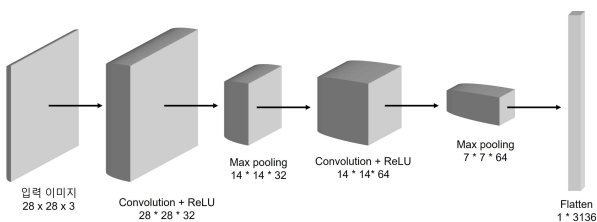


그림 2. 특징 추출에 사용한 CNN 모델
Figure 2. CNN model used for feature extraction

2.3.2. eGeMAPS(extended Geneva Minimalistic Acoustic Parameter Set)

eGeMAPS는 Eyben et al.(2015)에 의해 제안된 병리적 음성을

자동으로 분석하기 위해 구축된 표준 음향 특징 집합을 말하고, 음성 생성의 생리적인 변화를 포착할 수 있는 88개의 음향적 특징들로 구성되어 있다. eGeMAPS에는 주파수(음높이, jitter, F1, F2, F3 값, F1, F2, F3의 bandwidth, MFCC 1-4, spectral flux), 에너지 및 진폭(shimmer, 소리의 크기, HNR), 스펙트럼(alpha ratio, Hammarberg index, spectral slope, F1, F2, F3의 상대적 에너지, H1-H2, H1-A3), 시간(loudness peak의 비율, 유성음 영역의 평균 길이와 표준편차, 1초 당 연속된 유성음 영역의 개수) 관련 특징들이 포함되어 있다. eGeMAPS에는 수제 특징 집합에 포함되지 않은 포먼트의 bandwidth, spectral flux, 소리의 크기, alpha ratio, Hammarberg index, spectral slope, 포먼트의 상대적 에너지, 시간 관련 특징을 포함한다. eGeMAPS는 OpenSmile(Eyben et al., 2010) 툴킷을 사용하여 추출하였다.

2.3.3. CNN 추출 특징

본 연구에서는 딥러닝 모델인 2차원 CNN을 사용하여 특징을 추출하였다. CNN 모델은 이미지를 입력으로 받기 때문에, 음성을 스펙트로그램으로 변환하였다. 이때 사용한 스펙트로그램의 설정은 다음과 같다: n_mels=256, fmax=8,000, hop_length=160, n_fft=512, win_length=400. 특징 추출에 사용한 CNN 모델은 그림 2와 같다. 28×28 크기의 RGB 스펙트로그램을 입력으로 받아 1번째 convolution layer를 지난다. 이때의 convolution layer의 설정은 다음과 같다: input_channels=3, output_channels=32, kernel_size=3, stride=1, padding=1. 그 다음 활성화함수로 ReLU(rectified linear unit)를 사용한다. Convolution과 ReLU layer를 지난 특징에 max pooling layer를 적용하고, 이때의 설정은 다음과 같다: kernel_size=2, stride=2. Max pooling layer를 통과한 특징에 두 번째 convolution과 ReLU 연산을 진행하고, 이때의 convolution layer의 설정은 다음과 같다: input_channels=32, output_channels=64, kernel_size=3, stride=1, padding=1. 두 번째 convolution과 ReLU를 진행한 후 동일한 max pooling layer를 거치고, 마지막으로 전체 특징을 평탄화하여 총 3,136개의 특징을 추출한다.

2.4. 분류기

질병 진단 및 중증도 예측을 위한 머신러닝 분류기로는 SVM(support vector machine)과 XGBoost(Extreme Gradient Boosting)를 사용하였다. SVM은 음성 장애 진단에 가장 많이 사용되는 분류기이고, 고차원, 소규모의 데이터 분류 과제에서 좋은 성능을 보이는 것으로 알려져 있다(Hegde et al., 2019). XGBoost는 적은 수의 데이터로도 현실의 문제를 잘 해결할 수 있고, 특히 범주형 데이터나 적은 양의 데이터셋에서 좋은 성능을 보인다고 알려져 있다(Darouiche et al., 2022). SVM의 하이퍼파라미터인 C와 gamma는 10^{-4} 에서 10^4 사이에서 그리드 서치(grid search)를 통해 최적화하였다. XGBoost의 하이퍼파라미터인 결정 트리의 깊이(3-6개), 생성할 학습자의 수(12, 24, 32개), 학습률(10^{-4} - 10^{-1}), gamma(0.5, 1, 2)도 그리드 서치를 통해 최적화하였다.

3. 실험

3.1. 데이터

표 3. 실험에 사용한 발화 개수, 시간

Table 3. The number and length of tested utterances

중증도	모음연장발화	문장발화	긴 문장발화
대조군	30개 (10분 9초, 평균 20.3초)	32개 (2분 13초, 평균 4.2초)	31개 (3분 34초, 평균 6.9초)
1단계	57개 (10분 46초, 평균 11.3초)	60개 (3분 39초, 평균 3.6초)	61개 (7분, 평균 6.9초)
2단계	126개 (24분 8초, 평균 11.5초)	124개 (7분 34초, 평균 3.6초)	125개 (14분 32초, 평균 7.0초)
3단계	208개 (39분 17초, 평균 11.3초)	201개 (12분 7초, 평균 3.6초)	200개 (23분 59초, 평균 7.2초)
4단계	91개 (14분 53초, 평균 9.8초)	89개 (5분 31초, 평균 3.7초)	88개 (10분 33초, 평균 7.2초)
전체	512개 (1시간 35분 47초, 평균 11.2초)	506개 (31분 5초, 평균 3.7초)	505개 (59분 33초, 평균 7.1초)

본 연구에서는 Mun et al.(2022)에서 구축한 만성콩팥병 환자 음성 코퍼스를 사용하였다. Mun et al.(2022)의 코퍼스는 만성콩팥병 환자의 음성을 분석하고, 질병을 자동으로 진단하고 중증도를 예측하기 위한 방법론 개발을 위해 구축되었으며, 모음연장발화, 유성음 문장 발화, 문단 발화로 구성된다. 본 연구에서는 그 중 모음연장발화, 유성음 문장 발화, 문단 발화 중 첫 번째 문장 발화를 추출하여 사용하였다. 실험에 사용한 데이터는 표 3과 같고, 총 3시간 6분 25초 분량의 데이터를 사용하였다. 이때의 중증도는 의사의 소견을 바탕으로 만성콩팥병의 중증도를 판단할 때 사용하는 모든 요소들(사구체여과율, 혈뇨, 단백뇨 등)을 기반으로 정의되었다.

그러나 표 3에서 알 수 있듯이, 대조군 발화의 수가 만성콩팥병 환자 발화의 수에 비해 매우 적다. 집단별 데이터 수의 차이가 크게 나타나게 되면 소규모 집단에 속한 샘플은 대규모 집단에 속한 샘플보다 오분류 될 가능성이 높다(Sun et al., 2009). 따라서 본 연구에서는 Mun et al.(2022)에서 제시한 사구체여과율을 기준으로 대조군을 분류하였다. 만성콩팥병 1, 2단계 환자의 경우 3, 4단계의 환자에 비해 질병으로 인해 나타나는 증상이 명확하지 않아 대조군과 음성 특성이 유사할 수 있다는 의사의 소견 하에, 만성콩팥병을 진단할 때 가장 중요하게 여겨지는 요소 중 하나인 사구체여과율(estimated glomerular filtration rate, eGFR)을 기준으로 하여 사구체여과율이 60 이상이면 대조군으로 분류하였고, 이에 따라 실험에 사용된 발화의 수와 발화 길이는 표 4와 같다. 이 과정을 통해 유사한 길이의 대조군(1시간 23분 23초) 발화와 환자군(1시간 46분 20초) 발화로 실험을 진행하였다.

표 4. 사구체 여과율을 기준으로 정의한 집단 별 발화 수, 시간

Table 4. The number and length of utterances of groups defined by eGFR

중증도	모음연장발화	문장발화	긴 문장발화
대조군 eGFR≥60	213개 (45분 3초, 평균 12.7초)	216개 (13분 14초, 평균 3.7초)	217개 (25분 6초, 평균 6.9초)
3단계 30≤eGFR<60	208개 (39분 17초, 평균 11.3초)	201개 (12분 7초, 평균 3.6초)	200개 (23분 59초, 평균 7.2초)
4단계 15≤eGFR<30	91개 (14분 53초, 평균 9.8초)	89개 (5분 31초, 평균 3.7초)	88개 (10분 33초, 평균 7.2초)
전체	512개 (1시간 35분 47초, 평균 11.2초)	506개 (31분 5초, 평균 3.7초)	505개 (59분 33초, 평균 7.1초)

3.2. 분류 실험

3.2.1. 만성콩팥병 자동 진단

만성콩팥병을 자동으로 진단하기 위해 환자군과 대조군을 분류하는 실험을 진행하였다. 표 3을 기준으로 대조군과 환자군(1, 2, 3, 4단계)을 분류하는 실험과 표 4를 기준으로 대조군(대조군, 1, 2단계)과 환자군(3, 4단계)을 분류하는 실험을 진행하였다.

3.2.2. 만성콩팥병 중증도 예측

만성콩팥병의 중증도를 자동으로 예측하기 위해 음성을 모델이 예측한 중증도의 집단으로 분류하는 다중 분류 실험을 진행하였다. 표 3을 기준으로 음성을 대조군, 1단계, 2단계, 3단계, 4단계 환자군으로 분류하는 실험과 표 4를 기준으로 대조군, 3단계, 4단계 환자군으로 분류하는 실험을 수행하였다.

3.2.3. 성능 평가 메트릭

다양한 발화 유형, 특징 집합, 머신러닝 모델의 조합을 사용하여 진행한 분류 실험의 성능을 평가하기 위해 weighted F1-score를 사용하였다. F1-score는 불균형한 데이터에서 모델의 분류 성능을 평가할 때 사용되는 지표이고, 식 1과 같다. 모델이 True or False의 이분류 과제를 진행한다고 했을 때, 정밀도는 모델이 True라고 분류한 것 중에서 실제로 True인 것의 비율을 말하고, 재현율은 실제로 True인 것 중에서 모델이 True라고 예측한 것의 비율을 말한다. 이상적으로 좋은 모델은 정밀도와 재현율이 모두 높게 나타나지만, 실제로는 정밀도와 재현율 간의 트레이드오프가 존재한다. 따라서 정밀도와 재현율 중 더 작은 값에 영향을 많이 받게 하기 위해 두 값의 조화평균인 F1-score를 사용한다. 본 연구의 경우 집단별 데이터 불균형이 크기 때문에, 집단의 크기를 고려하는 weighted F1-score를 사용한다. Weighted F1-score는 각 집단에 대한 F1-score를 계산한 뒤 각 집단별 데이터 비율에 따른 가중 평균을 내어 F1-score를 계산한다.

$$F1score = 2 \times \frac{\text{정밀도} \times \text{재현율}}{\text{정밀도} + \text{재현율}} \quad (1)$$

4. 실험 결과 및 논의

4.1. 만성콩팥병 자동 진단

표 5. 표 3을 기준으로 진행한 자동 진단 실험 결과(weighted F1-score)

Table 5. Automatic diagnosis experimental results based on Table 3 (weighted F1-score)

발화 유형	특징 집합	SVM	XGBoost
모음연장발화	수제 특징	0.93	0.91
	eGeMAPS	0.91	0.91
	CNN 추출 특징	0.91	0.91
문장발화	수제 특징	0.91	0.91
	eGeMAPS	0.88	0.90
	CNN 추출 특징	0.88	0.89
일반 문장 발화	수제 특징	0.93	0.93
	eGeMAPS	0.92	0.91
	CNN 추출 특징	0.91	0.91

SVM, support vector machine; XGBoost, Extreme Gradient Boosting; eGeMAPS, extended Geneva Minimalistic Acoustic Parameter Set.

다양한 발화 목록과 특징 집합, 머신러닝 분류기를 사용하여 대조군과 환자군을 분류하는 과제의 결과는 표 5와 같다. 모음 연장발화로부터 추출한 수제 특징 집합을 사용하여 SVM을 사용하여 분류를 진행했을 때, 일반 문장 발화로부터 추출한 수제 특징 집합을 사용하여 SVM과 XGBoost로 분류를 진행하였을 때 테스트 셋에서 0.93의 가장 높은 F1-score가 나타났다.

대조군을 사구체여과율을 기준으로 분류한 후 대조군과 환자군을 분류하는 과제의 결과는 표 6과 같다. 일반 문장 발화로부터 추출한 수제 특징 집합을 사용하여 XGBoost로 분류를 진행했을 때 테스트 셋에서 0.93의 가장 높은 F1-score가 나타났고, 이때의 혼동 행렬과 특징의 중요도는 각각 그림 3, 그림 4와 같다.

표 6. 표 4를 기준으로 진행한 자동 진단 실험 결과(weighted F1-score)

Table 6. Automatic diagnosis experimental results based on Table 4 (weighted F1-score)

발화 유형	특징 집합	SVM	XGBoost
모음연장발화	수제 특징	0.69	0.62
	eGeMAPS	0.58	0.55
	CNN 추출 특징	0.54	0.48
문장발화	수제 특징	0.57	0.53
	eGeMAPS	0.50	0.54
	CNN 추출 특징	0.50	0.43
일반 문장 발화	수제 특징	0.78	0.93
	eGeMAPS	0.62	0.65
	CNN 추출 특징	0.72	0.69

SVM, support vector machine; XGBoost, Extreme Gradient Boosting; eGeMAPS, extended Geneva Minimalistic Acoustic Parameter Set.

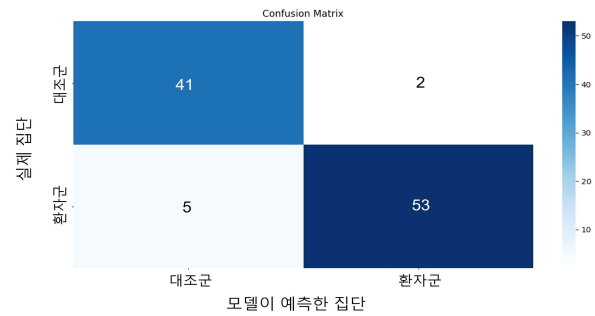


그림 3. 표 4를 기준으로 한 자동 진단 실험의 혼동행렬

Figure 3. Confusion matrix of automatic detection experiment based on Table 4

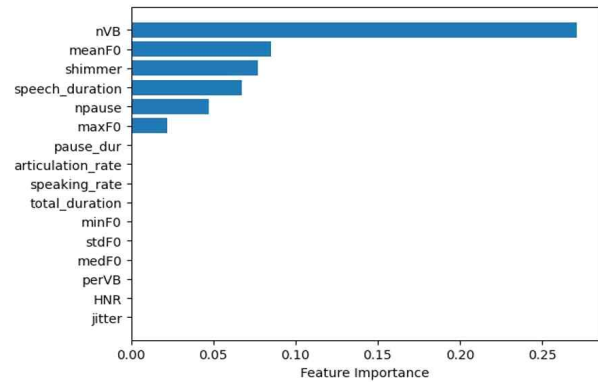


그림 4. 표 4를 기준으로 한 자동 진단 실험의 특징 중요도

Figure 4. Feature importance of automatic detection experiment based on Table 4

4.2. 만성콩팥병 중증도 예측

대조군, 1단계, 2단계, 3단계, 4단계로 발화를 분류하는 다중 분류 과제의 결과는 표 7과 같다. 일반 문장 발화로부터 추출한 수제 특징 집합을 사용하여 XGBoost로 분류를 진행했을 때 테스트 셋에서 0.84의 가장 높은 F1-score가 나타났고, 이때의 혼동 행렬과 특징의 중요도는 그림 5, 그림 6과 같다.

표 7. 표 3을 기준으로 한 중증도 예측 실험 결과(weighted F1-score)

Table 7. Severity prediction experimental results based on Table 3 (weighted F1-score)

발화 유형	특징 집합	SVM	XGBoost
모음연장발화	수제 특징	0.30	0.39
	eGeMAPS	0.36	0.33
	CNN 추출 특징	0.23	0.20
문장발화	수제 특징	0.34	0.26
	eGeMAPS	0.31	0.34
	CNN 추출 특징	0.26	0.27
일반 문장 발화	수제 특징	0.43	0.84
	eGeMAPS	0.31	0.39
	CNN 추출 특징	0.35	0.27

SVM, support vector machine; XGBoost, Extreme Gradient Boosting; eGeMAPS, extended Geneva Minimalistic Acoustic Parameter Set.

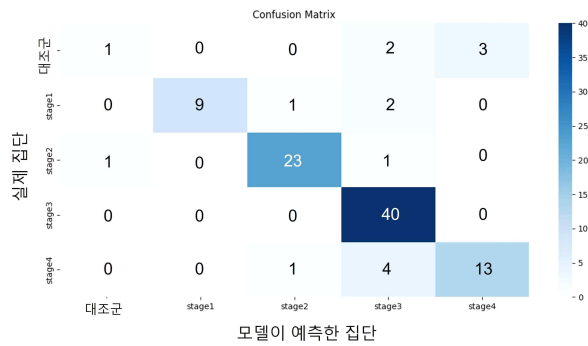


그림 5. 표 3을 기준으로 한 중증도 예측 실험의 혼동 행렬
Figure 5. Confusion matrix of severity prediction experiment based on Table 3

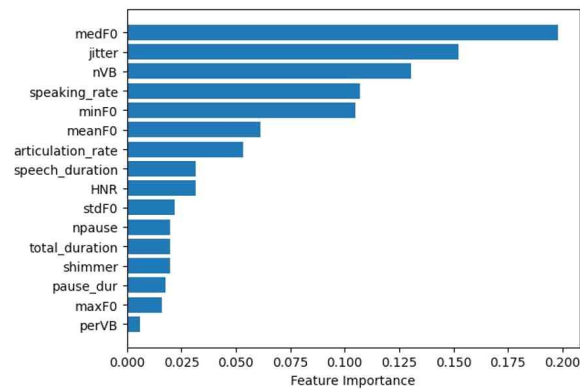


그림 6. 표 3을 기준으로 한 중증도 예측 실험의 특징 중요도
Figure 6. Feature importance of severity prediction experiment based on Table 3

사구체여과율을 기준으로 분류된 대조군과 3단계, 4단계로 발화를 분류하는 다중 분류 과제의 결과는 표 8과 같다. 일반 문장 발화로부터 추출한 수제 특징 집합을 사용하여 XGBoost로 분류를 진행했을 때 0.89의 가장 높은 F1-score가 나타났고, 이때의 혼동 행렬과 특징의 중요도는 그림 7, 그림 8과 같다.

표 8. 표 4를 기준으로 한 중증도 예측 실험 결과(weighted F1-score)
Table 8. Severity prediction experimental results based on Table 4

발화 유형	특징 집합	SVM	XGBoost
모음연장발화	수제 특징	0.50	0.47
	eGeMAPS	0.41	0.33
	CNN 추출 특징	0.42	0.49
문장발화	수제 특징	0.40	0.42
	eGeMAPS	0.42	0.40
	CNN 추출 특징	0.43	0.44
일반 문장 발화	수제 특징	0.60	0.89
	eGeMAPS	0.36	0.40
	CNN 추출 특징	0.46	0.54

SVM, support vector machine; XGBoost, Extreme Gradient Boosting; eGeMAPS, extended Geneva Minimalistic Acoustic Parameter Set.

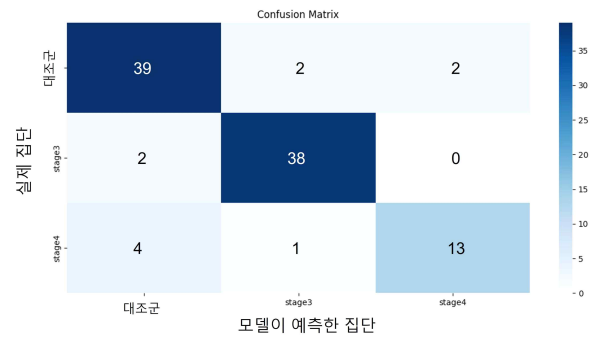


그림 7. 표 4를 기준으로 한 중증도 예측 실험의 혼동 행렬
Figure 7. Confusion matrix of severity prediction experiment based on Table 4

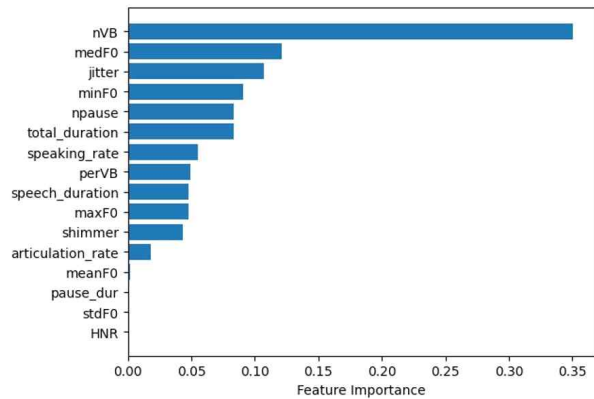
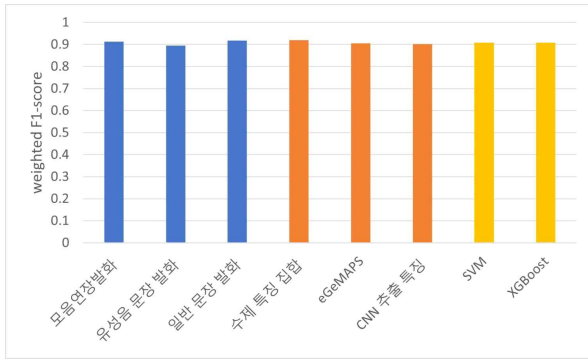


그림 8. 표 4를 기준으로 한 중증도 예측 실험의 특징 중요도
Figure 8. Feature importance of severity prediction experiment based on Table 4

4.3. 논의

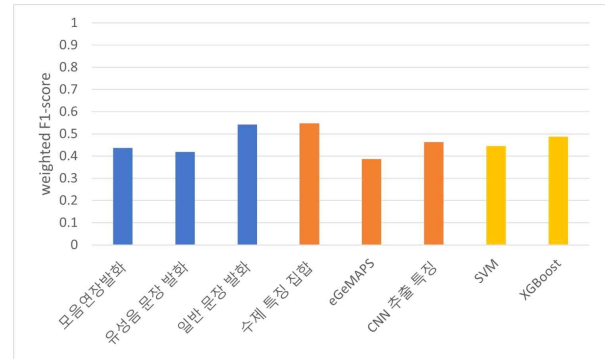
표 3을 기준으로 하여 만성콩팥병을 자동으로 진단하는 분류 실험 결과를 각 발화 유형, 특징 집합, 분류기를 기준으로 정리하여 얻은 평균 weighted F1-score는 그림 9와 같다. 발화 유형에서는 일반 문장 발화, 모음연장발화, 유성음 문장 발화 순으로, 특징 집합에서는 수제 특징 집합, eGeMAPS, CNN 추출 특징 순으로 좋은 성능이 나타났고, 분류기의 사용에 따른 성능 차이는 나타나지 않았다. 표 4를 기준으로 분류 실험을 진행했을 때의 평균 weighted F1-score는 그림 10과 같고, 표 3을 기준으로 진행한 실험에서와 동일하게 발화 유형에서는 일반 문장 발화, 모음연장발화, 유성음 문장 발화 순으로, 특징 집합에서는 수제 특징 집합, eGeMAPS, CNN 추출 특징 순으로 좋은 성능이 나타났고, 분류기의 사용에 따른 성능 차이는 크게 나타나지 않았다.

중증도 예측 실험의 경우에는 자동 진단 실험 결과와 다른 양상이 나타났고, 표 3, 표 4를 기준으로 한 중증도 예측 실험의 평균 weighted F1-score 값은 각각 그림 11, 그림 12와 같다.



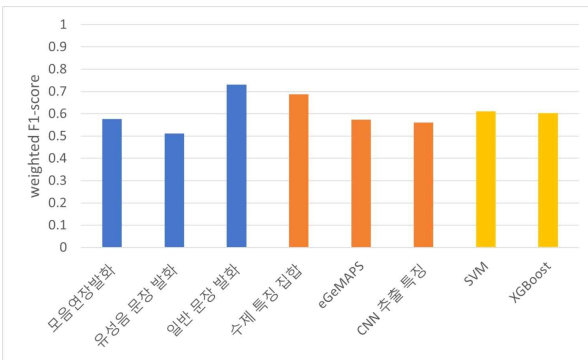
SVM, support vector machine; XGBoost, Extreme Gradient Boosting.

그림 9. 표 3을 기준으로 한 자동 진단 실험의 변수별 평균 성능
Figure 9. Average performance by variables in automatic diagnosis experiments based on Table 3



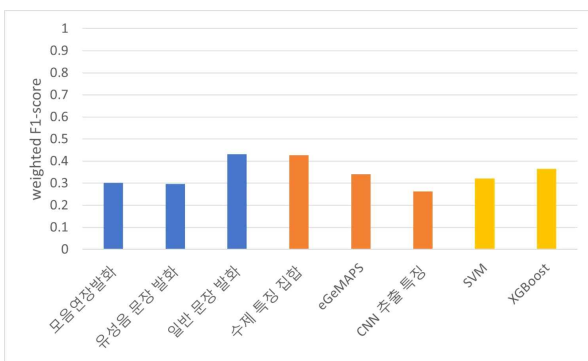
SVM, support vector machine; XGBoost, Extreme Gradient Boosting; eGeMAPS, extended Geneva Minimalistic Acoustic Parameter Set.

그림 12. 표 4를 기준으로 한 중증도 예측 실험의 변수별 평균 성능
Figure 12. Average performance by variables in severity prediction experiment based on Table 4



SVM, support vector machine; XGBoost, Extreme Gradient Boosting; eGeMAPS, extended Geneva Minimalistic Acoustic Parameter Set.

그림 10. 표 4를 기준으로 한 자동 진단 실험의 변수별 평균 성능
Figure 10. Average performance by variables in automatic diagnosis experiment based on Table 3



SVM, support vector machine; XGBoost, Extreme Gradient Boosting.

그림 11. 표 3을 기준으로 한 중증도 예측 실험의 변수별 평균 성능
Figure 11. Average performance by variables in severity prediction experiment based on Table 3

표 3을 기준으로 중증도 예측 실험을 진행했을 때 발화 유형과 특징 집합에 따른 성능은 자동 진단 실험과 동일한 양상으로 나타났고, 분류기를 기준으로 보았을 때에는 XGBoost가 SVM 보다 좋은 성능을 보였다. 표 4를 기준으로 중증도 예측 실험을 진행했을 때 발화 유형에 따른 성능은 위 결과들과 동일하게 일반 문장 발화, 모음연장발화, 유성음 문장 발화 순으로 높게 나타난 반면, 특징 집합을 기준으로 보았을 때에는 CNN 추출 특징이 eGeMAPS보다 좋은 성능을 보였다.

위 결과들로 미루어 보아, 만성콩팥병 환자의 음성을 분류하는 데에는 일반 문장 발화와 수제 특징 집합이 가장 효과적임을 확인할 수 있다. 그리고 자동 진단 실험보다 각 집단에 속한 데이터 수가 더 작았던 중증도 분류 실험에서는 XGBoost가 SVM 보다 좋은 성능을 보였는데, 이는 XGBoost가 적은 양의 데이터셋, 그리고 특히 범주형 데이터에서 좋은 성능을 보이는 특징을 지니 이와 같은 결과가 나타난 것으로 예측할 수 있다.

발화 유형, 특징 집합, 분류기의 조합에 따른 성능을 확인해 보면, 만성콩팥병을 자동으로 진단하기 위해 진행한 실험에서는 표 3, 표 4를 기준으로 했을 때 모두 0.93의 가장 높은 F1-score가 나타났고, 만성콩팥병의 중증도를 예측하기 위해 다분류 실험을 진행한 결과, 표 3을 기준으로 한 실험에서는 0.84, 표 4를 기준으로 한 실험에서는 0.89의 F1-score가 나타났다. 네 가지 실험 모두에서 일반 문장 발화+수제 특징+XGBoost의 조합이 가장 좋은 성능을 도출했다.

그림 4, 그림 6, 그림 8의 특징 중요도를 보면, 일반 문장 발화에서 말속도와 관련된 특징들의 중요도가 높게 나타났다. 만성콩팥병 환자의 경우 호흡계 근력의 약화로 인해 긴 호흡을 필요로 하는 일반 문장 발화에 어려움을 겪을 경우 느린 말속도, 잦은 숨이 나타나 해당 특징들의 중요도가 높게 나타났을 것으로 예상된다. 음높이와 관련된 특징들의 중요도 역시 높게 나타났는데, 이는 환자군과 대조군의 음높이에 차이가 존재한다는 선행 연구 결과와 일치하는 결과로 볼 수 있다.

수제 특징 집합과 달리 eGeMAPS와 CNN에서 추출한 특징 집합을 사용한 경우에는 분류기가 좋은 성능을 보이지 못했다.

eGeMAPS의 경우 총 88개의 다양한 음향 특징을 포함하는데, 만성콩팥병 음성의 특성을 반영하지 못하는 특징들이 포함되어 있었고, 해당 특징들로 인해 성능이 하락했다고 예상할 수 있다. 수제 특징 집합에 포함된 특징들 중에는 eGeMAPS에 존재하는 특징들도 존재하는데, eGeMAPS에 속한 특징들 중 만성콩팥병과 연관이 있을 것으로 보이는 특징들만 선정하여 새로운 특징 집합을 생성한 것이 특징 선택의 효과를 낳아 좋은 성능을 보였을 것이라고 예상된다. CNN에서 추출한 특징의 경우에는 모델로부터 3,136개의 매우 많은 수의 특징이 추출되어, 고차원의 문제를 해결하는 데에 분류기가 어려움을 겪었을 것이라고 예상할 수 있다.

마지막으로 그림 13은 모든 실험들에서 가장 좋은 성능을 보였던 일반 문장 발화+수제 특징 집합+XGBoost의 조합을 사용하여 표 3을 기준으로 자동 진단 실험을 진행하였을 때의 혼동행렬이다. 이 그림에서 분류기가 대부분의 대조군 발화를 환자군으로 분류한 것을 확인할 수 있다. 이는 3.1절에서 언급한 바와 같이 집단 간의 데이터 불균형으로 인한 결과로 볼 수 있다. 표 3을 기준으로 진행한 자동 진단 실험의 경우, 대조군과 환자군 발화의 수가 약 1:16으로, 데이터 수에 심한 불균형이 존재한다고 볼 수 있다. 이 경우에 소규모 집단(대조군)에 속한 샘플은 대규모 집단(환자군)으로 오분류 될 가능성이 매우 높아 위와 같은 결과가 나타난 것으로 보인다. 또한 전체적인 모델의 성능은 0.93의 weighted F1-score로 매우 높은 값이 나타났는데, 집단의 크기를 고려하기 위해 성능 평가 메트릭으로 weighted F1-score를 사용하였으나, 환자군의 데이터 수가 대조군 발화에 비해 16배 가까이 많아 환자군의 F1-score에 가중치가 크게 부여되어 모든 조합에서 분류 성능이 높게 나타난 것으로 보인다. 따라서 데이터 수의 불균형을 해결한 후 해당 과제를 진행해야 신뢰성 있는 결과를 얻을 수 있을 것이다.



그림 13. 표 3을 기준으로 한 자동 진단 실험의 혼동행렬
Figure 13. Confusion matrix of automatic diagnosis experiment based on Table 3

5. 결론

본 연구는 다양한 발화 유형, 특징 집합, 머신러닝 분류기를 사용하여 만성콩팥병을 자동으로 진단하고 중증도를 예측하는

최적의 방법을 탐색하였다. 본 연구는 세 가지 측면에서 의의가 있다. 첫째, 만성콩팥병 환자의 음성을 자동으로 분류한 첫 번째 연구라는 점이다. 만성콩팥병 환자의 음성을 음성학적으로 분석한 연구는 다수 존재했고, 해당 연구들에서 환자군과 대조군 간의 음성의 차이를 밝혔으나, 이를 사용하여 질병을 자동으로 진단하고 중증도를 예측하고자 한 연구는 없었다. 본 연구의 실험 결과를 만성콩팥병 음성 자동 분류의 베이스라인으로 삼을 수 있을 것이다. 둘째, 만성콩팥병 음성을 자동으로 분류하는 최적의 발화 유형, 특징 집합, 머신러닝 분류기의 조합을 발견했다는 것이다. 본 연구에서 수행한 모든 과제에서 일반 문장 발화, 수제 특징 집합, XGBoost의 조합이 가장 좋은 성능을 보였다. 일반 문장 발화는 선행연구들에서 사용되지 않았던 발화 목록이고, 수제 특징 집합은 선행연구들에서 분석에 포함하지 않았던 특징들을 포함한다. 이 결과는 본 연구에서 새롭게 제시하여 적용한 방법이 효과적임을 시사한다. 셋째, 음성을 사용하여 만성콩팥병을 초기에 진단할 수 있는 가능성을 제시하였다. 그림 5에서 모델이 만성콩팥병의 초기 단계인 1, 2단계 음성을 대조군과 잘 분류한 것을 확인할 수 있다. 따라서 징후가 잘 나타나지 않는 초기의 만성콩팥병을 음성이라는 보조적인 지표를 사용하여 진단할 수 있음을 시사한다.

본 연구는 불균형적이고 적은 수의 데이터를 사용했다는 한계가 있다. 데이터 수의 불균형이 존재하여 모델의 성능이 하락하였고, 머신러닝에 적용하기에는 다소 작은 규모의 데이터를 사용하였기 때문에 결과의 일반화를 위해서는 더 큰 규모의 데이터를 사용하여 실험을 진행하여야 할 것이다. 또한 대조군, 1, 2단계 화자의 발화 데이터 수가 다른 단계에 비해 특히 적어 머신러닝 분류에 적합하지 않다고 판단하여 표 4와 같은 새로운 분류 체계를 사용하였는데, 그림 5, 그림 7의 혼동행렬을 확인해보면 기존(표 3)의 분류 체계를 사용하였을 때 오히려 3단계 분류 성능이 더 좋은 것을 볼 수 있다. 논의에서 언급한 바와 같이 1, 2단계 화자의 발화에 질병의 특성이 반영되었을 가능성이 존재하기에, 대조군, 1, 2단계 화자의 발화를 추가적으로 수집하여 기존의 분류 체계를 사용해 질병을 진단하고 중증도를 예측하는 실험을 진행하여야 할 것이다. 따라서 후속 연구에서는 본 연구를 베이스라인으로 사용하여, 더 규모 있는 데이터베이스를 통해 다양한 머신러닝 및 딥러닝 방법론을 적용하여 분류 실험을 진행하고자 한다.

References

- Abd El-gaber, F. M., Sallam, Y., & El Sayed, H. M. E. (2021). Acoustic characteristics of voice in patients with chronic kidney disease. *International Journal of General Medicine*, 14, 2465-2473.
- Ahn, H. K. (2000). The H1*-H2* measure. *Speech Sciences*, 7(2), 85-95.
- Benba, A., Jilbab, A., Hammouch, A., & Sandabad, S. (2015, March). Voiceprints analysis using MFCC and SVM for detecting patients

- with Parkinson's disease. *Proceedings of the 2015 International Conference on Electrical and Information Technologies (ICEIT)* (pp. 300-304). Marrakech, Morocco.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9), 341-345.
- Darouiche, M. S., El Moubtahij, H., Yakhlef, M. B., & Tazi, E. B. (2022, March). An automatic voice disorder detection system based on extreme gradient boosting classifier. *Proceedings of the 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)* (pp. 1-5). Meknes, Morocco.
- Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., ... Truong, K. P. (2015). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2), 190-202.
- Eyben, F., Wöllmer, M., & Schuller, B. (2010, October). Opensmile: The munich versatile and fast open-source audio feature extractor. *MM '10: Proceedings of the 18th ACM International Conference on Multimedia* (pp. 1459-1462). Firenze, Italy.
- Harar, P., Galaz, Z., Alonso-Hernandez, J. B., Mekyska, J., Burget, R., & Smekal, Z. (2020). Towards robust voice pathology detection. *Neural Computing and Applications*, 32(20), 15747-15757.
- Hassan, E. S. (2014). Effect of chronic renal failure on voice: An acoustic and aerodynamic analysis. *The Egyptian Journal of Otolaryngology*, 30(1), 53-57.
- Hegde, S., Shetty, S., Rai, S., & Dodderi, T. (2019). A survey on machine learning approaches for automatic detection of voice disorders. *Journal of Voice*, 33(6), 947.E11-947.E33.
- Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Channing Moore, R., Plakal, M., ... Wilson, K. (2017, March). CNN architectures for large-scale audio classification. *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 131-135). New Orleans, LA.
- Jung, S. Y., Ryu, J. H., Park, H. S., Chung, S. M., Ryu, D. R., & Kim, H. S. (2014). Voice change in end-stage renal disease patients after hemodialysis: Correlation of subjective hoarseness and objective acoustic parameters. *Journal of Voice*, 28(2), 226-230.
- Kumar, R. B., & Bhat, J. S. (2010). Voice in chronic renal failure. *Journal of Voice*, 24(6), 690-693.
- Lee, S. J., Cho, Y., Song, J. Y., Lee, D., Kim, Y., & Kim, H. (2015). Aging effect on Korean female voice: Acoustic and perceptual examinations of breathiness. *Folia Phoniatrica et Logopaedica*, 67(6), 300-307.
- Liu, Y., Lee, T., Law, T., Lee, K., & Ching, P. C. (2018, November). Prediction of voice disorder severity: Contributions from sustained vowels and continuous speech. *Proceedings of the 2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)* (pp. 290-294). Taipei, Taiwan.
- McFee, B., Raffel, C., Liang, L., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto, O. (2015, July). Librosa: Audio and music signal analysis in Python. *Proceedings of the 14th Python in Science Conference* (pp. 18-25). Austin, TX.
- Moon, K. R., Chung, S. M., Park, H. S., & Kim, H. S. (2012). Materials of acoustic analysis: Sustained vowel versus sentence. *Journal of Voice*, 26(5), 563-565.
- Mudawwar, W. A., Alam, E. S., Saredidine, D. S., Turfe, Z. A., & Hamdan, A. H. (2017). Effect of renal failure on voice. *ENT: Ear, Nose & Throat Journal*, 96, 32-36.
- Mun, J., Kim, S., Kim, M. J., Ryu, J., Kim, S., & Chung, M. (2022). A speech corpus for chronic kidney disease. *arXiv*. <https://doi.org/10.48550/arXiv.2211.01705>
- Narendra, N. P., & Alku, P. (2020). Glottal source information for pathological voice detection. *IEEE Access*, 8, 67745-67755.
- Omeroglu, A. N., Mohammed, H. M. A., & Oral, E. A. (2022). Multi-modal voice pathology detection architecture based on deep and handcrafted feature fusion. *Engineering Science and Technology, an International Journal*, 36, 101148.
- Shetty, S., Hegde, S., & Dodderi, T. (2018, February). Classification of healthy and pathological voices using MFCC and ANN. *Proceedings of the 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAEECC)* (pp. 1-5). Bangalore, India.
- Speyer, R., Bogaardt, H. C. A., Passos, V. L., Roodenburg, N. P. H. D., Zumach, A., Heijnen, M. A. M., Baijens, L. W. J., ... Brunings, J. W. (2010). Maximum phonation time: Variability and reliability. *Journal of Voice*, 24(3), 281-284.
- Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(4), 687-719.
- Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis - jitter, shimmer and hnr parameters. *Procedia Technology*, 9, 1112-1122.
- Triantafyllopoulos, A., Fendler, M., Batliner, A., Gerczuk, M., Amiriparian, S., Berghaus, T. M., & Schuller, B. W. (2022, September). Distinguishing between pre- and post-treatment in the speech of patients with chronic obstructive pulmonary disease. *Proceedings of the Interspeech 2022* (pp. 3623-3627), Incheon, Korea.
- Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic kidney disease. *The Lancet*, 389(10075), 1238-1252.
- Yeo, E., Kim, S., & Chung, M. (2021). Automatic severity classification of dysarthria using voice quality, prosody, and pronunciation features. *Phonetics and Speech Sciences*, 13(2), 57-66.

Zaky, E. A., Mamdouh, H., Esmat, O., & Khalaf, Z. (2020). Voice problem in a patient with chronic renal failure. *The Egyptian Journal of Otolaryngology*, 36(1), 1-8.

• **문지현 (Jihyun Mun)**

서울대학교 언어학과 석박통합과정
서울시 관악구 관악로 1
Tel: 02-880-6164
Email: jhhh_1202@snu.ac.kr
관심분야: 음성인식, 음성언어처리, 언어 장애

• **김선희 (Sunhee Kim)**

서울대학교 불어교육과 교수
서울시 관악구 관악로 1
Tel: 02-880-7693
Email: sunhkim@snu.ac.kr
관심분야: 프랑스어 음성학, 음성언어처리

• **김명주 (Myeong Ju Kim)**

서울대학교 AI연구원 연구원
서울특별시 관악구 관악로 1
Tel: 02-880-4258
Email: 99528@snu.ac.kr

• **류지원 (Jiwon Ryu)**

분당서울대학교병원 내과 교수
경기도 성남시 분당구 구미로173번길 82
Email: bboddo5@gmail.com

• **김세중 (Sejoong Kim)**

서울대학교 의과대학 내과 교수
서울특별시 종로구 대학로 103
Tel: 02-1588-3369
Email: sejoong2@snu.ac.kr

• **정민화 (Minhwa Chung)** 교신저자

서울대학교 언어학과 교수
서울시 관악구 관악로 1
Tel: 02-880-9195
Email: mchung@snu.ac.kr
관심분야: 음성인식, 음성언어처리, 컴퓨터 기반 언어교육

머신러닝 분류기를 사용한 만성콩팥병 자동 진단 및 중증도 예측 연구*

문 지 현¹ · 김 선 희² · 김 명 주³ · 류 지 원⁴ · 김 세 중^{3,4,5} · 정 민 화¹

¹서울대학교 언어학과, ²서울대학교 불어교육과, ³서울대학교 AI연구원,

⁴분당서울대학교병원 내과, ⁵서울대학교 의과대학 내과

국문초록

본 논문은 만성콩팥병 환자의 음성을 사용하여 질병을 자동으로 진단하고 중증도를 예측하는 최적의 방법론을 제안한다. 만성콩팥병 환자는 호흡계 근육의 약화와 성대 부종 등으로 인해 음성이 변화하게 된다. 만성콩팥병 환자의 음성을 음성학적으로 분석한 선행 연구는 존재했으나, 환자의 음성을 분류하는 연구는 진행된 바가 없다. 본 논문에서는 모음연장발화, 유성음 문장 발화, 일반 문장 발화의 발화 목록과, 수제 특징 집합, eGeMAPS, CNN 추출 특징의 특징 집합, SVM, XGBoost의 머신러닝 분류기를 사용하여 만성콩팥병 환자의 음성을 분류하였다. 총 3시간 26분 25초 분량의 1,523개 발화가 실험에 사용되었다. 그 결과, 질병을 자동으로 진단하는 데에는 0.93, 중증도를 예측하는 3분류 문제에서는 0.89, 5분류 문제에서는 0.84의 F1-score가 나타났고, 모든 과제에서 일반 문장 발화, 수제 특징 집합, XGBoost의 조합을 사용했을 때 가장 높은 성능이 나타났다. 이는 만성콩팥병 음성 자동 분류에는 화자의 발화 특성을 모두 반영할 수 있는 일반 문장 발화와 거기로부터 추출한 적절한 특징 집합이 효과적임을 시사한다.

핵심어: 만성콩팥병, 머신러닝, 음성 자동 분류

참고문헌

여은정, 김선희, 정민화(2021). 음질, 운율, 발음 특징을 이용한 마비말장애 중증도 자동 분류. *말소리와 음성과학*, 13(2), 57-66.

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음(IITP-2022-2018-0-01833*). 이 논문의 일부는 분당서울대학교병원 연구비 (grant no 13-2022-0008) 지원에 의해 이루어짐.