

조음교대운동의 음향적 분석과 말명료도 예측을 이용한

마비말장애 중증도 분류

김민서⁰¹ 오준석² 김지환^{2*}

¹서강대학교 인공지능학과

²서강대학교 컴퓨터공학과

minseokim@sogang.ac.kr ohjs@sogang.ac.kr kimjihwan@sogang.ac.kr

Dysarthria Severity Classification Using Diadochokinesis Acoustic Analysis and Speech Intelligibility Prediction

MinSeo Kim⁰¹ Junseok Oh² Ji-Hwan Kim^{2**}

¹Department of Artificial Intelligence, Sogang University

²Department of Computer Science and Engineering, Sogang University

요약

마비말장애 치료를 위한 디지털 치료법 개발에 대한 관심이 증가하고 있는 가운데, 본 논문은 마비말장애의 중증도 예측과 원인 파악을 위해 조음교대운동(DDK)을 분석하는 방법을 제시한다. 오픈소스 소프트웨어인 NeuroSpeech를 활용하여 환자들의 DDK 과제를 분석하고, 규칙성, 속도, 에너지 등 11개의 음향 특성을 측정한다. 또한, ResNeXt101 기반 딥러닝 모델로 환자들의 말명료도를 예측하고, 성별과 나이와 같은 기본 정보와 함께 DDK 과제에서 측정된 음향 특성을 LightGBM 기반의 모델로 입력하여 마비말장애의 중증도를 분류한다. 중요한 특성들을 확인하고 설명하기 위해 SHapley Additive exPlanation(SHAP) 기법을 적용한다. 본 연구에서는 320명의 환자 데이터를 사용하여 모델을 학습시켰고, 말명료도 분류 모델은 테스트셋에 대해 약 65.38%의 정확도를, 중증도 분류 모델은 약 67.84%의 정확도를 보였다. 결과적으로, 초당 멈춤 횟수, 초당 음절 발음 수, 말명료도가 마비말장애 중증도 예측에 큰 영향을 끼친다는 것이 확인되었다.

1. 서론

마비말장애는 말초 신경계 또는 중추 신경계 손상으로 인해 발생하는 일련의 언어 장애로, 특히 뇌졸중을 겪은 사람들 사이에서 흔히 발생한다[1]. 이 장애는 환자들의 일상생활에 큰 영향을 미치며, 지속적인 언어 재활 치료를 통해 일부 완화될 수 있다. 그러나 개별 환자들의 상태와 필요에 따라 적합한 치료 방법이 달라지기 때문에, 개인 맞춤형 치료의 중요성이 갈수록 높아지고 있다. 최근 연구들은 디지털 치료법을 개발하여 이러한 문제에 대응하고 있으나, 그 효과는 제한적이다. 본 연구는 이러한 배경 속에서 마비말장애의 중증도를 보다 정확하게 평가하고, 개인별 맞춤 치료 방안을 도출하는 새로운 방법을 제시한다.

마비말장애를 평가하는 데에는 여러 방법이 있으며, 그 중 모음연장발성, 조음교대운동(DDK), 읽기 과제 등이 주로 사용된다[2]. 특히 DDK는 조음기관의 움직임 속도, 규칙성, 그리고 호흡 및 발성 특성을 분석하는 데 효과적이다. 말명료도(Speech Intelligibility)는 화자가 말

하고자 하는 내용을 청자가 이해한 정도를 의미한다. 말명료도와 마비말장애 중증도 간의 상관관계는 존재하지 않지만, 말명료도 평가 자체만으로는 중증도를 측정할 수 없기에 다양한 관점의 청지각적 평가와 음향학적 평가를 병행해 측정해야 한다[3].

본 논문에서는 DDK 과제를 기반으로 말명료도와 음향학적 특성을 이용하여 마비말장애 중증도를 예측한다. DDK 과제로부터 ResNeXt101[5] 기반 모델을 통해 예측한 말명료도와 NeuroSpeech[4]를 사용해 분석한 음향학적 특성을 사용하여 LightGBM[6] 모델을 기반으로 마비말장애 중증도를 분류한다. 또한, SHapley Additive exPlanation(SHAP)[7]를 적용하여 원인을 설명할 수 있으며 이를 통해 환자에게 중증도 판별의 근거를 제시하는 설명가능한 인공지능 서비스 제공에 일조할 수 있다.

2. 관련 연구

2.1 조음교대운동(Diadochokinesis)

조음교대운동 분석에는 ‘교대운동속도(Alternating Motion Rate, AMR)’과제와 ‘일련운동속도(Sequential Motion Rate, SMR)’ 과제가 사용된다. AMR 과제는 /피

* 교신저자(Corresponding Author)

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00621, 대화 기반 설명가능성을 멀티모달로 제공하는 인공지능 기술 개발)

/, /터/, /커/ 와 같이 동일한 음절을 정해진 시간 내에 반복적으로 빠르게 발음하는 과제이고 SMR 과제는 /퍼 터커/와 같은 서로 다른 두 개 이상의 음절을 정해진 시간 내에 반복하여 발음하는 과제이다. 이 과제들을 통해 조음기관 움직임의 속도 및 규칙성을 분석하고 조음의 정확성 및 호흡과 발성의 특성을 분석할 수 있다[2].

2.2 NeuroSpeech

NeuroSpeech[4]는 파킨슨병 환자의 마비말 장애 음성 신호를 모델링 하기 위한 오픈 소스 소프트웨어이다. 이 소프트웨어를 이용해 음성 신호를 발성, 조음, 운율, 명료도, 교대운동에 대해 분석할 수 있으며 사용자가 목적에 맞게 수정 보완하여 사용할 수 있도록 한다. 본 논문에서는 NeruoSpeech를 수정하여 DDK 과제의 음향학적 특성을 분석한다. 표1은 분석에 사용되는 음향학적 특성들이다. 이 특성들을 분석하여 마비말장애 중증도를 예측하는데 활용한다.

feature	설명
FO variability [Semitones]	반음 단위의 기본 주파수의 분산
FO variability [Hz]	HZ 단위의 기본 주파수의 분산
Avg. energy [dB]	에너지의 평균
Energy variability [dB]	에너지의 분산
Max. energy[dB]	에너지의 최댓값
DDK rate	초당 음절 발음 수
DDK regularity	음절 발음 시간의 분산
Avg. duration DDK	음절 발음 시간의 평균
Pause rate [1/s]	초당 쉼 횟수
Avg. duration pause	쉼 시간의 평균
Regularity pause	쉼 시간의 분산

표 1. DDK 과제 분석에 사용되는 음향학적 특성

2.3 말명료도(Speech Intelligibility)

말명료도는 화자가 전하는 말을 청자 입장에서 얼마나 잘 알아들을 수 있는가의 정도를 의미하며 의사소통의 성공 여부를 반영하는 지표이다. 임상에서는 말명료도를 평가하기 위해 전사, 척도법, DME 방법 등을 사용한다 [3]. 본 연구에서는 특정 음절을 반복해서 읽은 음성으로부터 말명료도를 예측해야 하기 때문에 위와 같은 평가 방법이 아닌 ResNeXt101 아키텍처를 사용하여 음성을 입력으로 받아 말명료도를 예측하는 End-to-End 방식을 사용한다. 예측한 말명료도는 NeuroSpeech를 사용해 분석한 특성들과 함께 마비말장애 중증도를 예측하는데 사용한다.

2.4 Light Gradient Boosting Model

LightGBM은 의사 결정 트리 알고리즘에 기반한 알고

리즘으로 특징 선택, 회귀 분석 등 다양한 작업에 활용될 수 있다. 데이터의 일부만으로 빠르게 정보이득을 계산하는 Gradient-based One-Side Sampling(GOSS)과 Exclusive Feature Bundling(EFB)으로 특성 요인들을 줄이는 알고리즘을 활용하여 고차원 특징을 효과적으로 처리하면서 훈련 및 예측 시간을 단축할 수 있다[6]. LightGBM은 기존 부스팅 알고리즘과 비교하여 시간을 줄이고 높은 정확도를 보인다. 본 연구에서는 LightGBM을 기반으로 마비말장애의 중증도를 예측하고 SHAP를 적용하여 중증도를 예측하는데 영향을 준 요인의 중요도에 대해 설명할 수 있다.

3. LightGBM 기반 조음교대운동 음향학 분석과 말명료도 예측을 이용한 마비말장애 중증도 분류

3.1 데이터셋

이 연구에 사용된 데이터셋¹⁾은 마비말장애가 있는 뇌졸중 환자 320명의 1731개 음성 녹음 데이터로 구성되어 있다. 중증도는 Normal, Mild to Moderate, Severe 3개로 평가하며 말명료도는 1~5점 척도로 평가한다. 표 2와 3 각각은 중증도와 말명료도에 따른 데이터셋 구성이다. 중증도와 말명료도에 따른 데이터셋 각각을 8:1:1 비율로 나누어 학습, 검증, 테스트 셋으로 사용했다.

중증도	Normal	Mild to Moderate	Severe
파일 수 (개)	96	1492	143

표 2. 중증도에 따른 데이터셋 구성

말명료도	1	2	3	4	5
파일 수 (개)	59	80	231	609	752

표 3. 말명료도에 따른 데이터셋 구성

3.2 제안 모델

ResNeXt101 아키텍처를 사용하여 DDK 과제 음성의 224차원 멜-스펙트로그램을 입력으로 받아 5점 척도의 말명료도를 예측하는 모델을 제안한다. 이 모델은 시간-주파수 영역에서 2D 컨볼루션을 이용하여 멜-스펙트로그램 데이터에 존재하는 복잡한 패턴을 효율적으로 학습할 수 있다[5]. 데이터 불균형을 완화하기 위해 클래스마다 다른 가중치를 할당하는 weighted categorical

1) 이 연구는 과학기술정보통신부의 재원으로 한국지능정보사회진흥원의 지원을 받아 구축된 "No.2022-데이터-위81, 인공지능 학습을 위한 응급실 임상 대화 및 구음장애인 명령어 데이터 수집 사업"을 활용하여 수행된 연구입니다. 본 연구에 활용된 데이터는 AI 허브(aihub.or.kr)에서 다운로드 받으실 수 있습니다.

cross-entropy를 손실함수로 사용한다.

예측한 말명료도와 NeuroSpeech를 사용하여 분석한 특성들에 성별, 연령과 같은 환자의 기본 정보를 입력으로 받아 마비말장애 환자의 중증도를 LightGBM을 통해 예측하는 방법을 제안한다. 말명료도 예측 모델과 같이 데이터의 불균형을 완화하기 위해 weighted categorical cross-entropy를 손실함수로 사용한다.

4. 실험

말명료도를 예측하는 모델을 학습할 때, SpecAugment[8] 및 speed perturbation과 같은 방법을 사용하여 데이터셋에 변형을 만들어 모델을 더 다양한 시나리오에 노출시킴으로써 모델이 일반화될 수 있도록 했다. 그림 1은 테스트셋에 대한 평가결과를 보여준다. 매크로 평균 65.24%, 마이크로 평균 65.38%의 정확도를 보인다.

중증도를 예측하는 모델을 학습하고 테스트 데이터셋으로 평가한 결과 매크로 평균 69.34%, 마이크로 평균 67.84%의 정확도를 보인다. SHAP를 적용하여 모델을 통해 중증도를 예측할 때, 각 특성들이 끼치는 영향을 분석할 수 있다. 그림 2는 중증도 분류에 사용된 특성들의 중요도를 보여준다. 초당 멈춤 횟수, 초당 음절 발음 수, 말명료도가 큰 영향을 끼쳤다.

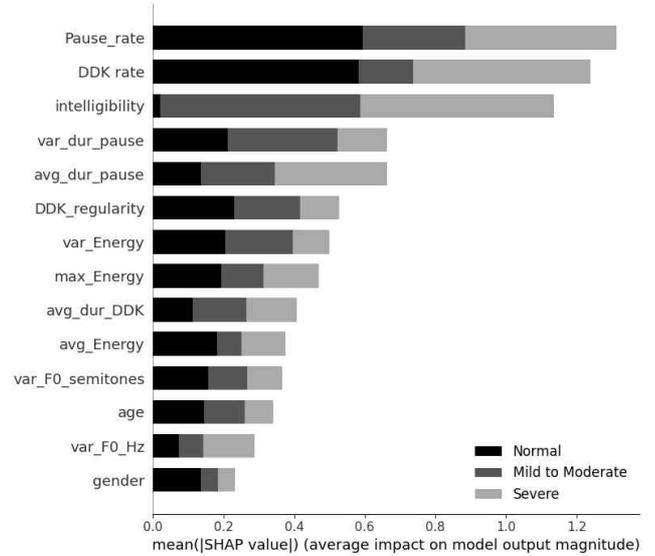


그림 2. 중증도 예측을 위한 특성 별 중요도

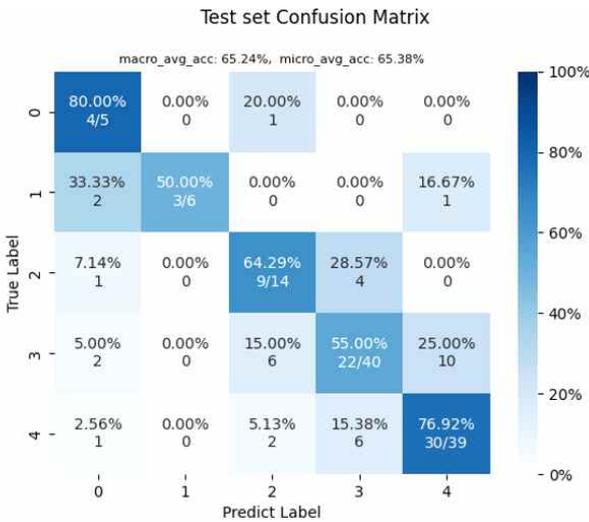


그림 1. 말명료도 예측 모델 평가 결과

SHAP를 사용하여 중증도를 예측하고자 하는 음성별로 어떤 특성이 많은 영향을 주었는지도 확인할 수 있다. 그림 3은 예시 음성의 특성 별 중요도를 보여준다. 예시로 사용한 음성은 중증도가 severe로 예측된 30~40대 남성이 /커/를 반복적으로 발음한 음성이다. 그림 3을 통해 말명료도, 초당 음절 발음 수, 초당 멈춤 횟수가 중증도 예측에 큰 영향을 끼친 것을 확인할 수 있다. 또한, 말명료도가 1로 낮게 측정되고 초당 음절 발음 수가 1.067개로 적기 때문에 중증도가 severe로 예측이 된 것을 알 수 있다.

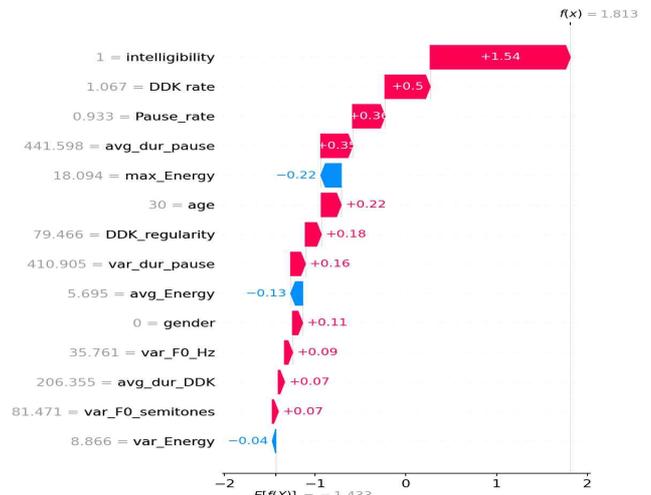


그림 3. 중증도가 Severe로 예측된 30~40대 남성이 /커/를 반복적으로 발음한 음성의 특성 별 중요도

4. 결론

본 논문에서는 마비말장애가 있는 뇌졸중 환자의 음성에서 예측한 말명료도와 분석한 음향학적 특성들을 사용하여 마비말장애 환자의 중증도를 LightGBM을 통해 분류했다. 말명료도를 예측하는 모델은 65.38%, 중증도를 분류하는 모델은 67.84%의 정확도를 보였다. SHAP를 적용하여 중증도를 분류하는데 특성 별 중요도를 확인할 수 있다. 초당 멈춤 횟수, 초당 음절 발음 수, 말명료도가 중증도 예측에 큰 영향을 끼친다는 것이 확인되었다. 또한, 음성별로 중증도 예측에 영향을 준 특성들을 분석하여 환자별 부족한 요인에 대해서도 설명할 수 있다.

중증도 분류 모델은 약 68%의 성능을 보였지만 이 모델만을 사용하여 마비말장애 환자들의 중증도를 예측할 수준의 성능은 아니다. 향후에는 DDK 과제 외의 읽기

과제 등을 추가하여 중증도 분류 모델의 성능을 높일 수 있는 방법에 대해 연구할 예정이다. 또, 말명료도 예측 모델에 다른 아키텍처를 사용하여 성능을 높일 수 있는 방법에 대해 연구할 예정이다.

참고 문헌

- [1] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," *Journal of Speech and Hearing Research*, Vol. 12, No. 2, pp. 246-269, 1969.
- [2] 박희준, "오픈 소스 소프트웨어를 이용한 마비말장애 화자의 일련운동속도 분석," *Journal of Speech-Language & Hearing Disorders*, Vol.29, No. 3, pp. 11-16, 2020
- [3] 이옥분, 한지연, 박상희, "마비말장애 심각도에 따른 음절단위 말명료도와 모음공간," *말소리와 음성과학*, Vol.2 No. 2 pp. 85-92, 2010
- [4] J. R. Orozco-Arroyave *et al.*, "NeuroSpeech: An open-source software for Parkinson's speech analysis," *Digital Signal Processing*, Vol. 77, pp. 207-221, 2018.
- [5] S. Xie *et al.*, "Aggregated Residual Transformations for Deep Neural Networks," *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1492-1500, 2017.
- [6] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," *In Proc. of the Advances in Neural Information Processing Systems*, Vol. 30, 2017.
- [1] S. M. Lundberg, S.-I. Lee. "A unified approach to interpreting model predictions," *In Proc. of the Advances in Neural Information Processing Systems*, Vol.30, pp. 4765-4774, 2017
- [7] D. S. Park *et al.*, "SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition," *in Proc. of the Interspeech* pp. 2613-2617, 2019