**퍼터커 반복하기**

# 1. 개요

# 1-1. 개발 목적

퍼터커 반복하기는 조음교대운동속도를 검사하는 것이다. 이 검사는 환자의 조음 기관의 구조-기능을 평가하며, 1초당 특정 자음+모음 구조의 일련운동속도(SMR)를 측정한다.

본 프로젝트에서는 단일 음절(퍼/터/커)의 반복 속도를 측정하는 모델과 다중 음절(퍼터커)의 반복 속도를 측정하는 모델을 CNN으로 만들었다.

# 2. 구현

# 2-1. Trainning

학습 모듈의 구성은 다음과 같다.

**데이터 준비**

**Train/Test 분리**

**전처리**

**모델링**

**학습**

검사자료 전달(wav 파일)

90 : 10 비율로 분리

오디오 파일 멜스펙토그램화, 정규화, 패딩

CNN 모델을 커스터마이징해 단위시간당 단일/다중 음정 반복 횟수를 예측

EarlyStopping을 활용해 모델 학습

# 2-1-1. 데이터 준비

모델 공통라이브러리 model\_comm.get\_file\_lst를 데이터 호출용으로 사용한다. CLAP\_D/1 폴더 내 모든 음성파일을 DB에서 조회하여 파일의 경로와 검사 점수, 배점 정보를 가져온다.

# 2-1-2. Train/Test 분리

음성 데이터의 목록은 환자별로 ‘퍼’, ‘터’, ‘커’, ‘퍼터커’ 발성 데이터가 별도 저장되어 있기에, 단일 음성(‘퍼’, ‘터’, ‘커)과 다중 음성(‘퍼터커’)으로 나누었다. 각 데이터는 90%는 학습용으로, 10%는 테스트용으로 사용하도록 분리한다.

# 2-1-3. 전처리

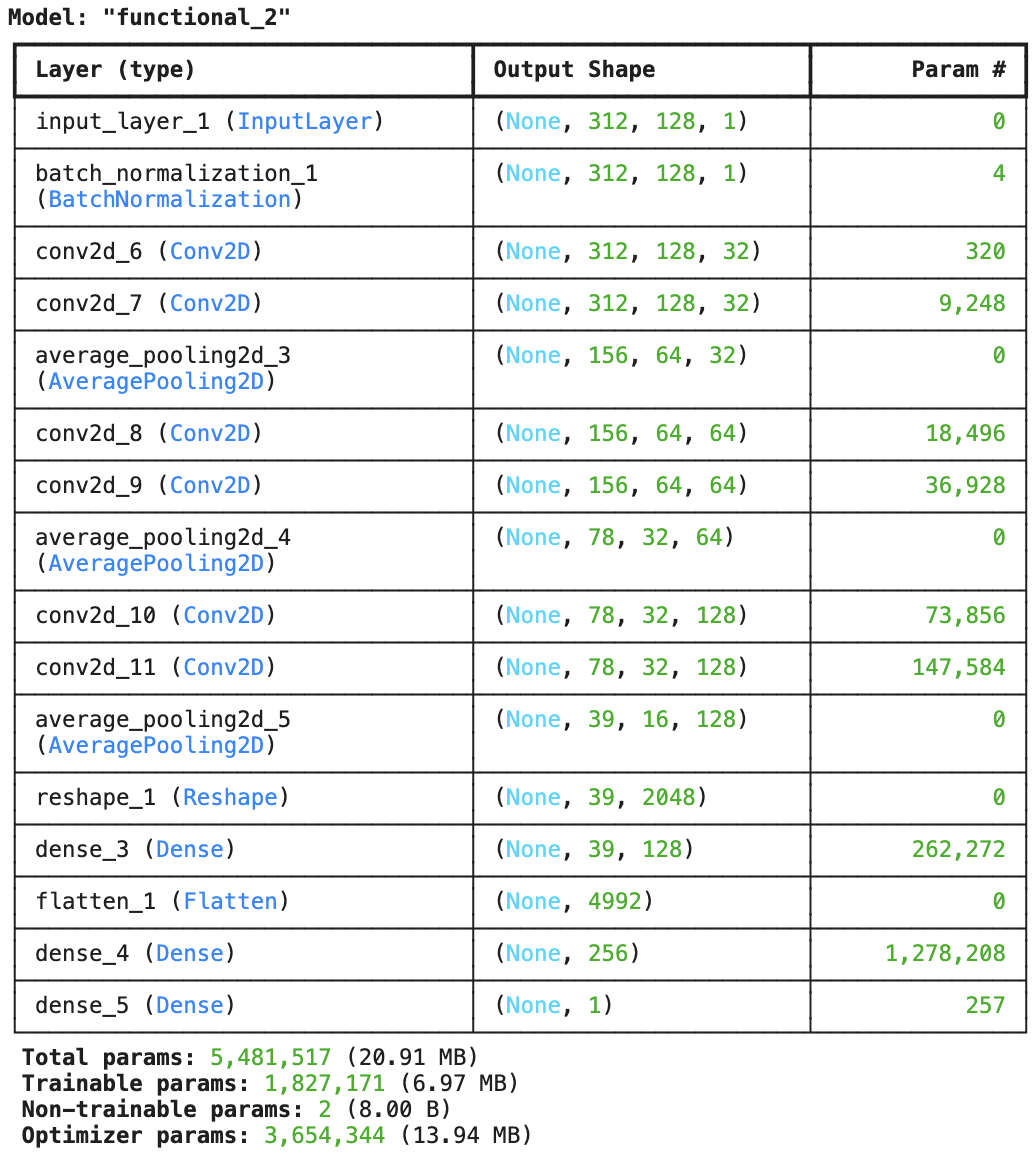
데이터 전처리는 음성 파일을 Mel Spectrogram 변환하고 정규화 및 패딩을 적용한다.

# 2-1-4. 모델

본 모델은 전처리된 음성을 배치 정규화하고, AveragePooling을 적용한 CNN 모델에 기반하여 특징을 추출한다.

입력 데이터는 312\*128\*1 차원의 멜스펙토그램 형태를 가진다. 각 차원은 음성 파일의 시간축, 주파수축, 채널의 의미를 갖는다. 먼저 각 데이터에 배치 정규화를 거쳐 학습 안정성과 속도를 높인다. 그리고 CNN으로 특징을 추출하고, 각 시간 프레임마다 특징 벡터를 생성해 시간적 패턴을 학습한다. 그리고 최종적으로 차원을 축소해 예측된 수를 출력한다.

모델의 써머리는 다음과 같다.



EarlyStopping을 사용하여 모델이 더 이상 개선되지 않을 때 자동으로 학습을 멈추고, 가장 성능이 좋았던 시점의 모델을 되돌려주도록 하였다. 이를 통해 과적합을 방지하고 학습 시간을 효율적으로 관리하도록 했다.

# 2-2. Predict

예측 모듈은 다음과 같이 구성하였다.

**전처리**

**모델 로드**

**예측**

**후처리**

입력받은 음성파일을 Mel Spectrogram화, 정규화, 패딩

저장된 학습 모델 로드

학습 모델에 input을 넣어 예측 수행

예측하여 나온 값을 점수로 변환

# 2-2-1. 전처리

환자의 듣고 따라 말하기 파일의 경로 정보를 리스트로 받는다. 음성 파일을 Mel Spectrogram으로 변환 및 정규화, 패딩을 거쳐 일반화한다.

# 2-2-2. 모델 로드 및 점수 예측

학습된 모델(.keras)을 불러와서 예측을 수행한다.

# 3. 예측 실행 결과

학습된 모델을 이용하여 유효한 음성 데이터를 가지고 점수를 예측한 결과를 다음과 같이 산점도로 표현하였다. 모델을 통해 산출한 예측 점수와 실제 점수의 상관계수는 0.815가 나왔다.

텍스트, 라인, 그래프, 도표이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

# 4. 결과 분석 및 개선사항

전반적으로 모델에 기반한 예측 점수와 검사자의 검사 실제 점수가 유사하게 나오고 있다. 그러나 학습에 사용된 데이터의 수가 적고, 특히 이상 환자에 대한 데이터가 부족해 편향이 의심된다.

차후 데이터 추가 수집 또는 생성적 적대 신경망(GAN)을 활용해 성능을 높일 수 있을 것으로 기대한다.