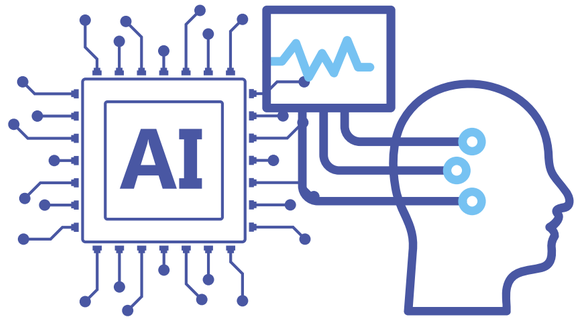


EEG와 Audio 데이터를 이용하여 멀티모달 기법을 통한 우울증 분류 모델 개발

이소호 2022270611
김준협 2022400617

Table of contents

INTRODUCTION

BACKGROUND

1) 최근 우울증 환자 수의 급증

2) 현재 우울증 진단 방법의 문제점:

1. Labor intensive : 현재 DSM-5 등의 진단방식은 주로 전문가와의 임상 면담, 심리 검사, 설문 등을 통해 이루어지며, 이 과정에서 많은 시간과 인력이 소모됨
2. Inaccessible : 충분한 수의 전문인력이 확보되지 않은 지역에서는 접근성의 한계로 인해 진단에 큰 제약이 발생함
3. subjective & unreliable ; 설문이나 면담에 의존하는 주관적 진단 방식은 개인차와 자기 인식의 한계로 인해 일관성과 객관성이 떨어져 신뢰성이 낮음

이에 정신장애 진단을 위해서는 새로운 생리학적 지표를 고려한 모델의 필요성이 대두되고 있음.

목적

1. 기존 방법에 비해 빠르고 정확하며 신뢰성 있는 우울증 진단 모델 개발
2. 여러 개의 생체 데이터에 멀티모달 기법을 적용한 분석 가능성 모색
3. 선행 연구 논문을 참고하여 멀티모달 우울증 분류 모델을 구현해보고 성능이나 방법론적 측면에서 개선 가능성 모색

INTRODUCTION

주요 우울 장애 (Major Depressive Disorder)

지속적인 우울감과 활동력 저하를 특징으로 하는 우울감 상태가 지속 또는 반복적으로 나타나는 정신장애

증상: 흥미나 쾌락의 저하, 불면증, 무력감, 사고력이나 집중력 감퇴, 무가치감 등

본 연구에서 정신장애 분석을 위한 다중 모달 공개 데이터셋 제안

▶ 우을중 환자/정상 환잔의 뇌파(eeg) 및 오디오(Audio) 데이터로 구성

EEG

전두엽 EEG 비대칭성이 우울증 위험의 바이오마커 역할을 할 수 있으며, 이는 향후 부정적 감정을 예측하고 치료 반응을 예측하는 데 있어서 유용성이 밝혀짐

Audio

정신장애가 환자의 음성데이터를 건강한 대조군과 다르게 변화시키며, 음성으로부터 추정된 조음 패턴을 기반으로 우울증 심각도를 추적하는 알고리즘 개발됨

INTRODUCTION: Dataset

EEG 128 Channel resting state

- EEG 신호는 128개의 Ag/AgCl 전극이 장착된 유선 EEG 캡(HydroCel Geodesic Sensor Net, HCGSN)을 사용하여 수집
-전극과 피부의 접촉면은KCL기반 전도성 젤을 도포한 후,임피던스를50kΩ 이하로 보정
- 참가자는 5분 동안 눈을 감고 안정 상태를 유지하며 EEG 신호 기록
- 실험 중 머리나 다리를 움직이지 않으며, 불필요한 눈 움직임, 깜빡임 등 통제

INTRODUCTION: Dataset

Audio

- 실험 중 환경 소음이 60dB를 초과할 경우 녹음이 이루어지지 않음
- 녹음 장비로 Neumann TLM102(마이크로폰) 및 RME FIREFACE UCX(오디오 인터페이스) 사용
- 실험 과제는 약 25분 동안 진행되었으며 질문 순서는 무작위로 배정
- 실험 과제는 총 3가지의 과제로 구성 (면담 과제, 단어 읽기 과제, 그림 설명 과제)

INTRODUCTION: Dataset

Audio

1. 면담 과제 (Interview)
DSM-IV 및 HRSD 척도에서 발췌한 18개의 질문을 포함
예:"휴가를 간다면 여행 계획을 어떻게 세우겠는가?","지금까지 받은 최고의 선물과 그때의 감정을 설명해달라."

2. 단어 읽기 과제(WordsReading)
짧은 이야기와 감정적으로 분류된 세 그룹의 단어(긍정적, 중립적, 부정적)를 읽도록 요청

3. 그림 설명 과제(Picture Description)
중국 얼굴 정서 사진 시스템(CFAPS)에서 선정된 세 개의 얼굴 표정 사진(긍정적, 중립적,부정적) 설명
추가적으로 주제 통각 검사(TAT)의 "우는 여성(crying woman)" 이미지 설명

INTRODUCTION: Reference Paper

모델 구현 Reference


brain sciences

MDPI

Article

Multimodal Fusion of EEG and Audio Spectrogram for Major Depressive Disorder Recognition Using Modified DenseNet121

Musyyab Yousufi, Robertas Damaševičius * and Rytis Maskeliūnas

Centre of Real Time Computer Systems, Kaunas University of Technology, 51368 Kaunas, Lithuania; musyyab.yousufi@ktu.edu (M.Y.)
* Correspondence: robertas.damasevicius@ktu.lt

Abstract: Background/Objectives: This study investigates the classification of Major Depressive Disorder (MDD) using electroencephalography (EEG) Short-Time Fourier-Transform (STFT) spectrograms and audio Mel-spectrogram data of 52 subjects. The objective is to develop a multimodal classification model that integrates audio and EEG data to accurately identify depressive tendencies. Methods: We utilized the Multimodal open dataset for Mental Disorder Analysis (MODMA) and trained a pre-trained Densenet121 model using transfer learning. Features from both the EEG and audio modalities were extracted and concatenated before being passed through the final classification layer. Additionally, an ablation study was conducted on both datasets separately. Results: The proposed multimodal classification model demonstrated superior performance compared to existing methods, achieving an Accuracy of 97.53%, Precision of 98.20%, F1 Score of 97.76%, and Recall of 97.32%. A confusion matrix was also used to evaluate the model's effectiveness. Conclusions: The paper presents a robust multimodal classification approach that outperforms state-of-the-art methods with potential application in clinical diagnostics for depression assessment.

Keywords: multimodal fusion; EEG; deep learning; depression; speech

Citation: Yousufi, M.; Damaševičius, R.; Maskeliūnas, R. Multimodal Fusion of EEG and Audio Spectrogram for Major Depressive Disorder Recognition Using Modified DenseNet121. *Brain Sci.* **2024**, *14*, 1018. https://doi.org/10.3390/brainsci14101018

Academic Editor: Ateke Goshvarpour

Received: 15 September 2024
Revised: 9 October 2024
Accepted: 10 October 2024
Published: 15 October 2024

1. Introduction

Depression and anxiety are mental disorders that affect a person's ability to perform daily routine tasks effectively with symptoms including sleep disorder, continued sadness, bad eating habits, general cognition problems, suicidal thoughts, etc. [1–3]. Depression and anxiety are often comorbid conditions, with anxiety symptoms frequently preceding the onset of Major Depressive Disorder (MDD). Studies suggest that up to 60% of individuals diagnosed with MDD also experience some form of anxiety disorder. This comorbidity is so common that clinicians often find it challenging to distinguish between the two. Both disorders affect an individual's cognitive and emotional regulation, and their co-occurrence complicates diagnosis and treatment [4,5]. According to the World Health Organization's (WHO) report, more than 350 million people are affected by the disease worldwide [6]. Depression has emerged as the foremost cause of disability and a significant global health concern [7]. A person experiencing minor or mild depression is emotionally struggling, which can negatively influence their relationships with family, coworkers, and their ability to perform tasks effectively [8,9]. Prolonged mild depression can lead to MDD that can cause extreme mood swings, severe physiological issues, and health concerns [10]. Thus, diagnosis and prognosis of depression are crucial for it to be treated timely [11].

Diagnosing MDD involves structured clinical interviews by psychiatrists, focusing on individual moods, thoughts, behaviors, and symptoms. Also, tools like the Hamilton Depression Rating Scale (HDRS), Beck Depression Inventory (BDI), and Depression Rating Scale (DRS) are used [2]. The HDRS is a clinician administered questionnaire designed to assess depression severity, focusing on cognitive and somatic symptoms. Conversely, the BDI allows for self-assessment of mood and physical changes related to depression,

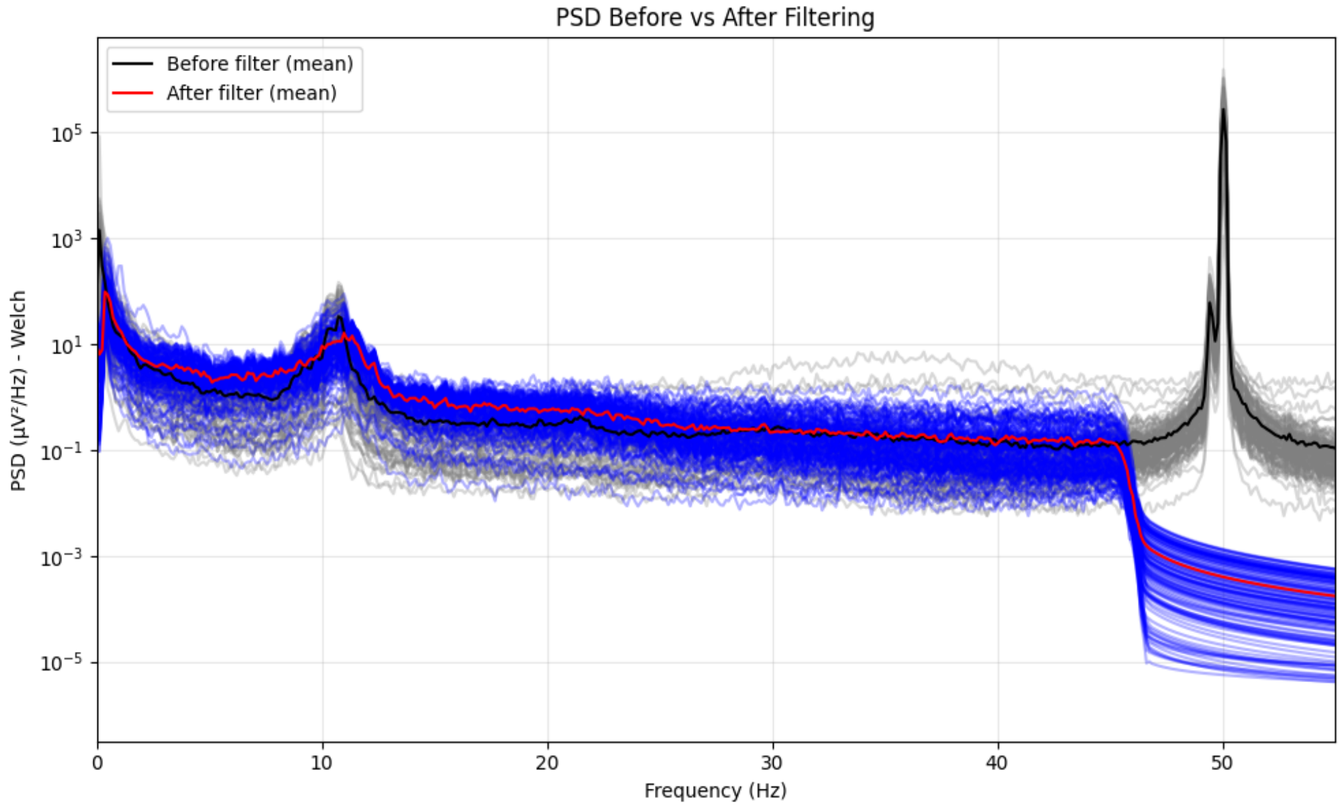
Brain Sci. **2024**, *14*, 1018. https://doi.org/10.3390/brainsci14101018 https://www.mdpi.com/journal/brainsci

Yousufi, M., Damaševičius, R., & Maskeliūnas, R. (2024). Multimodal Fusion of EEG and Audio Spectrogram for Major Depressive Disorder Recognition Using Modified DenseNet121. Brain Sciences, 14(10), 1018.

Dataset(MODMA) Reference

Cai, H., Yuan, Z., Gao, Y., Sun, S., Li, N., Tian, F., ... & Hu, B. (2022). A multi-modal open dataset for mental-disorder analysis. Scientific Data, 9(1), 178.

METHOD: EEG PREPROCESSING



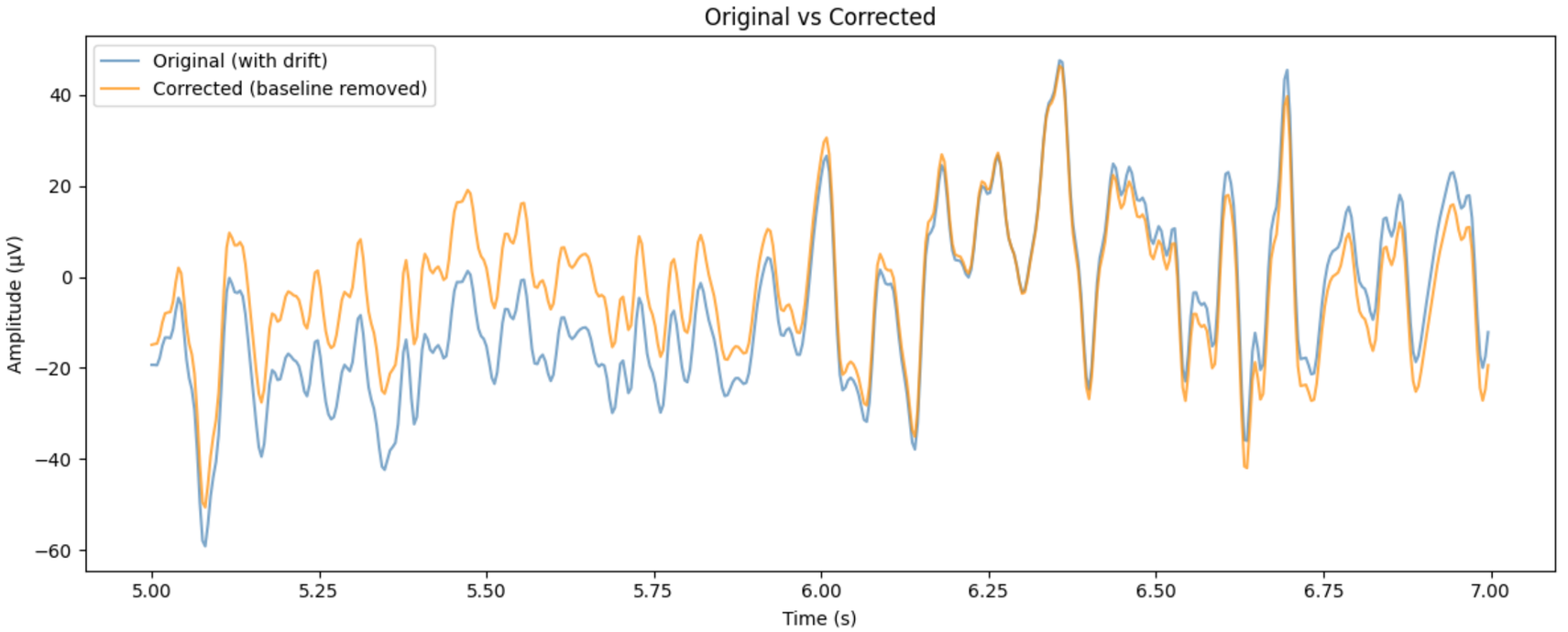
Notch Filter - 50 Hz

- 특정 주파수(예: 전원 잡음)만 선택적으로 제거

Bandpass Filter - 0.4Hz~ 45Hz

- 특정 주파수 대역의 신호만 통과시키고, 나머지 대역의 주파수 신호는 감쇠

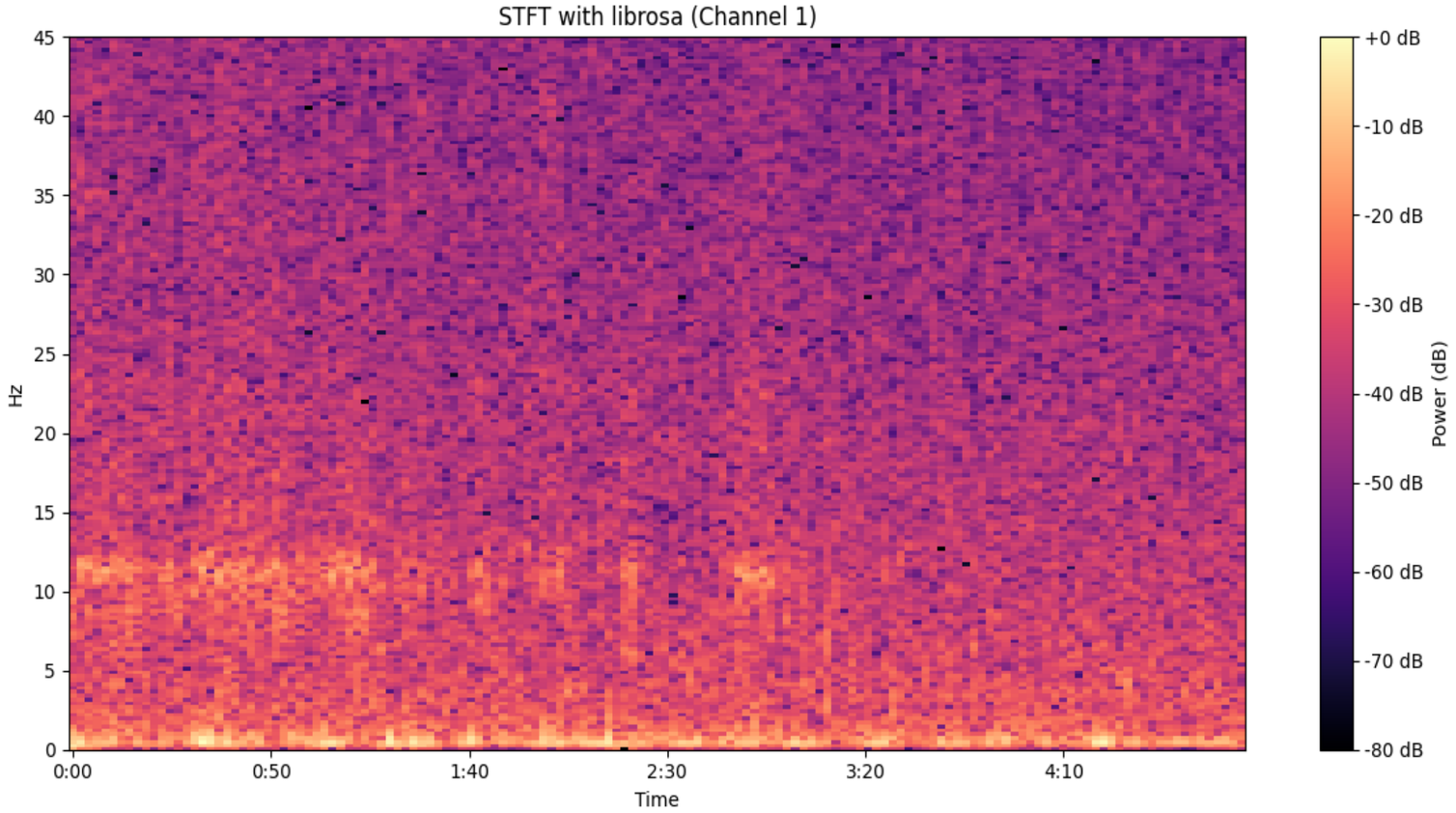
METHOD: EEG PREPROCESSING



Baseline Correction

- 일정 길이의 슬라이딩 윈도우로 로컬 평균 계산해 신호에서 빼 줌으로써 느린 드리프트 성분을 제거

METHOD: EEG PREPROCESSING



STFT

신호를 짧은 구간으로 나누어 각 구간에 푸리에 변환을 적용해 시간에 따른 주파수 변화를 분석

Sampling rate : 250Hz
Window size (n_fft) : 1024
Hop length : 512
Window function : Hann
Frequency cutoff : 0~45 Hz

METHOD: EEG PREPROCESSING

이미지 RESIZE

STFT 파워 스펙트럼을 dB 변환 후 0-255 범위로 정규화해 스펙트로그램 이미지 생성

- 단일 채널을 3채널 (RGB)로 변환
- 244 x 244fh resize
- DenseNet121 입력 규격 일치

TENSOR 변환

RGB Mel-spectogram 이미지를 PyTorch tensor (C x H x W) 형태로 전환

- Tensor 변환 시 픽셀 값을 0-1 범위로 자동 정규화

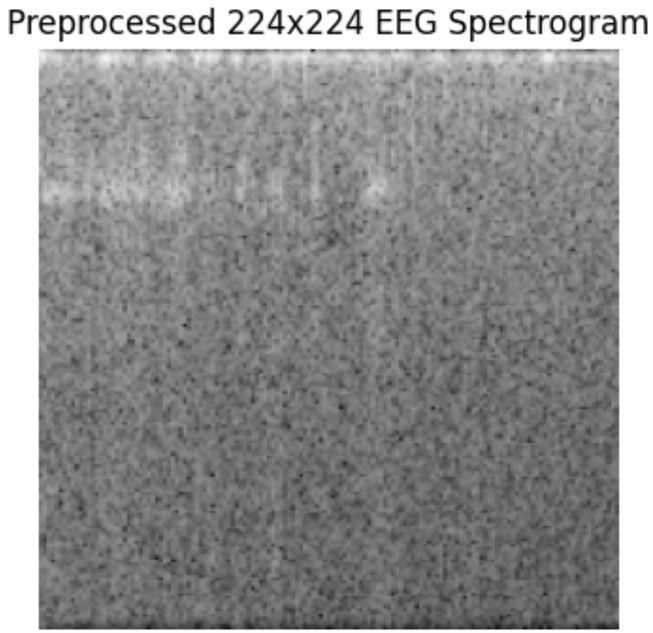
NORMALIZATION

DenseNet121 사전학습 가중치의 입력분포와 일관성을 위해 ImageNet mean/std로 정규화

mean: [0.485, 0.456, 0.406]
std: [0.229, 0.224, 0.225]

METHOD: EEG PREPROCESSING

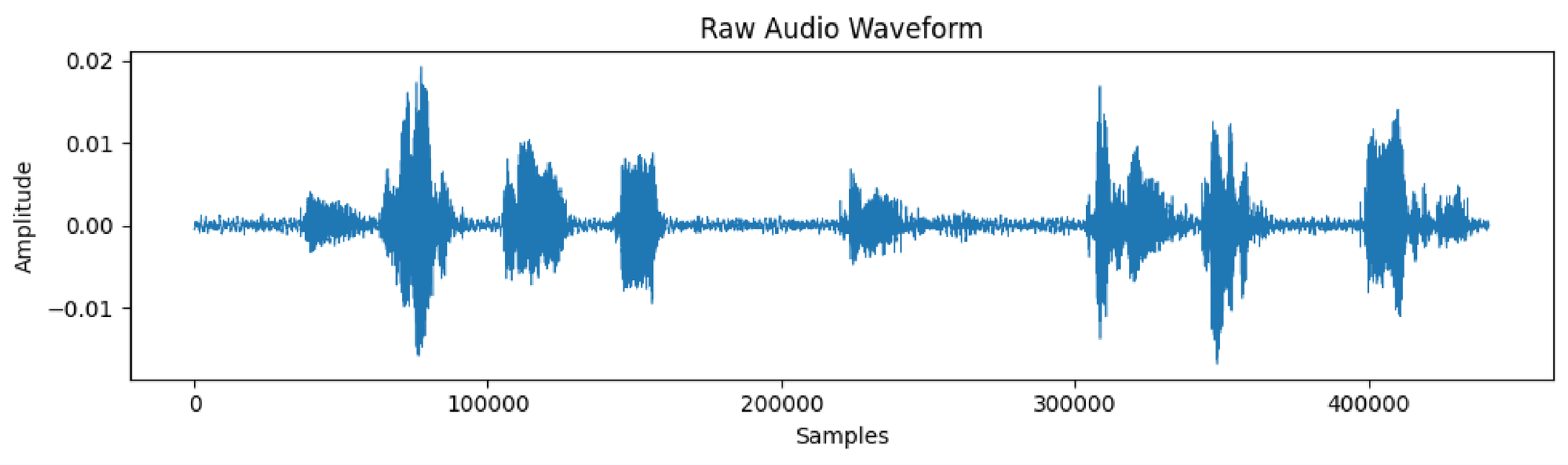
모델 최종 입력형태



모델의 입력구조에 맞게 채널별로 224x224 크기의 RGB 이미지로 변환

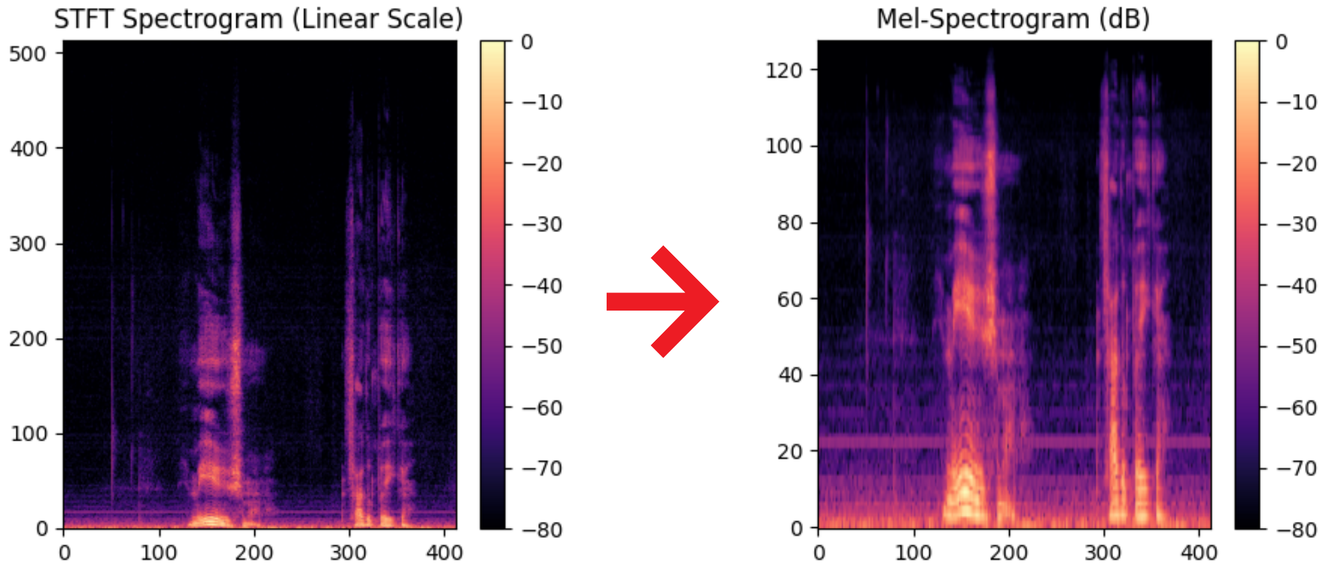
-> tensor 형태로 .pt파일로 저장되어 모델의 입력(Input)으로 사용(shape = [3,224,224])

METHOD: AUDIO PREPROCESSING



Raw Audio data를 특정 주파수 제거 등 별도의 가공 없이 사용

METHOD: AUDIO PREPROCESSING



- 음성 원신호에 STFT(1024 FFT, hop 512)를 적용해 시간–주파수 파워 스펙트럼 생성
- 생성된 STFT Spectrogram에 Mel filter-bank(128 Mel bins)를 적용하여 Mel-spectrogram 생성

Sampling rate : 250Hz
Window size (n_fft) : 1024
Hop length : 512
n_mels : 128
Frequency cutoff : 0-22,050Hz
sr=44100
Mel Filter-Bank 적용

METHOD: AUDIO PREPROCESSING

이미지 RESIZE

Mel-spectogram의 dB 스케일 값을 0-255 범위로 정규화하여 시각화 이미지 생성

- 단일 채널을 3채널 (RGB)로 변환
- 244 x 244fh resize
- DenseNet121 입력 규격 일치

TENSOR 변환

RGB Mel-spectogram 이미지를 PyTorch tensor (C x H x W) 형태로 전환

- Tensor 변환 시 픽셀 값을 0-1 범위로 자동 정규화

NORMALIZATION

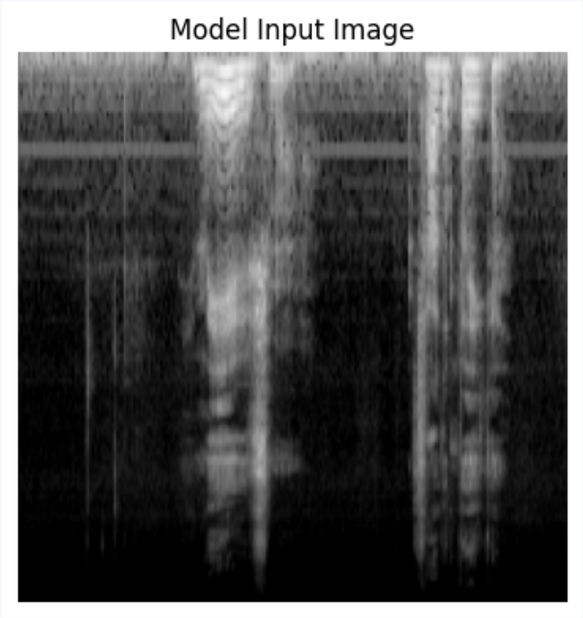
DenseNet121 사전학습 가중치의 입력분포와 일관성을 위해 ImageNet mean/std로 정규화

mean: [0.485, 0.456, 0.406]
std: [0.229, 0.224, 0.225]

EEG 변환방식과 동일

METHOD: AUDIO PREPROCESSING

모델 최종 입력형태

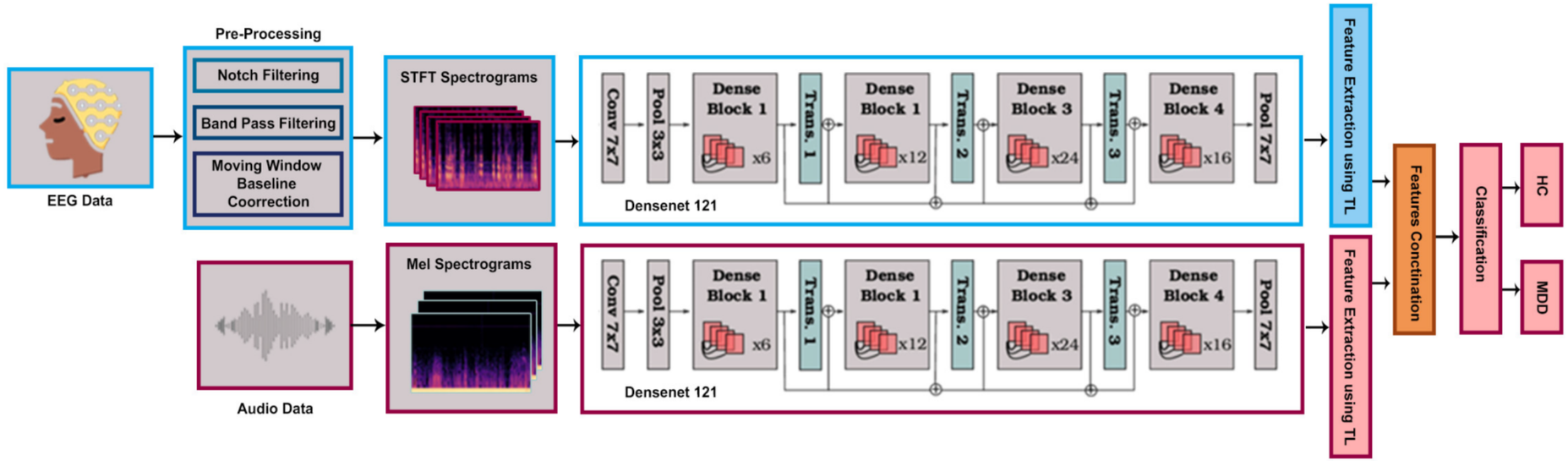


모델의 입력구조에 맞게 Mel-spectogram을 224x224 크기의 RGB 이미지로 변환

-> tensor 형태로 .pt파일로 저장되어 모델의 입력(Input)으로 사용됨

EEG & Audio Multi Modal Modeling

Multi Modal Using Pre-trained Densenet-121 Model



가공된 EEG, Audio data를 DenseNet121에 입력

feature 추출

(멀티모달)두 feature vector concatenate

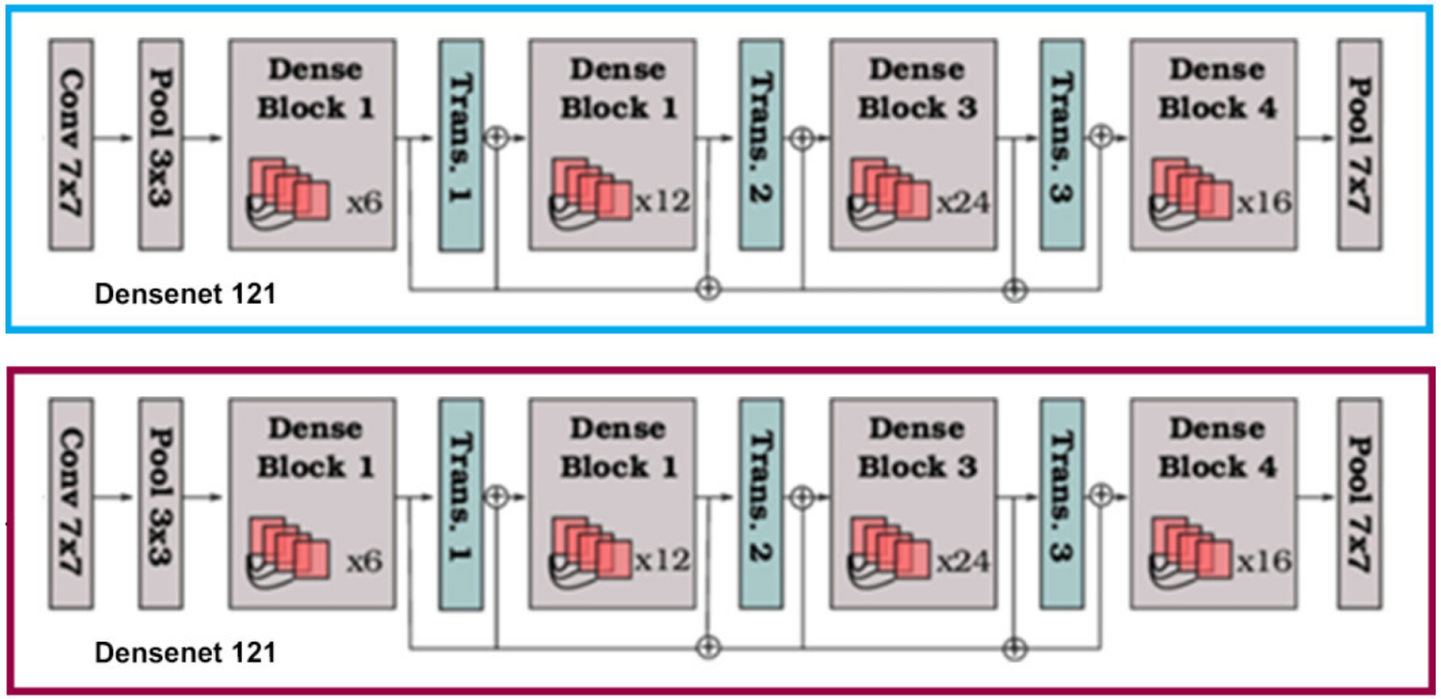
최종 FC layer에서 MDD/HC 분류

Transfer Learning 적용

- ImageNet pretrained DenseNet121 사용
- Convolution 백본 freeze
- 마지막 FC layer만 학습

Densenet-121 Modeling

Pre-trained Densenet-121 Model



입력 형태 :
224×224 RGB Tensor

[Parameter]

Optimizer : Adamax

Learning Rate : 0.001

Batch Size : 16

Dropout : 0.3

Epoch : 100, Early Stopping (최소 15 epoch 보장, loss 기준 stop, patience=3)

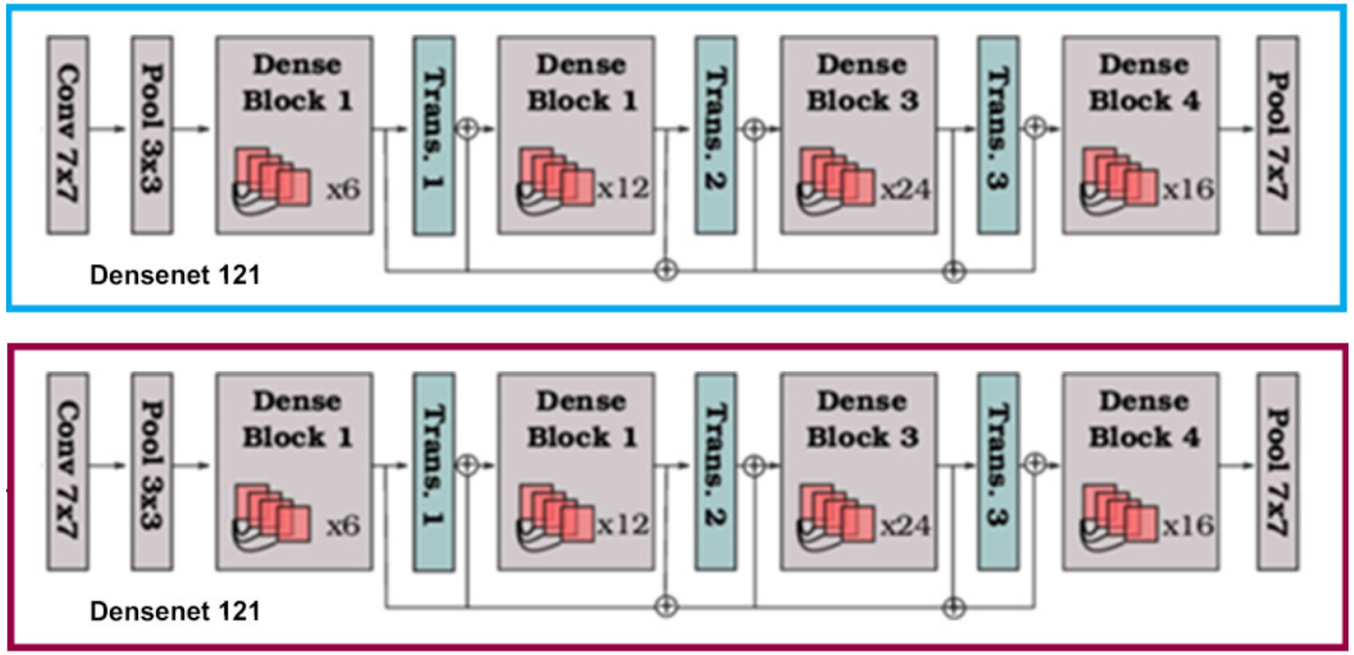
Loss Function : CrossEntropyLoss

Split Ratio – 64:16:20

- (train+validation) : test = 8 : 2
- train : validation = 8 : 2

Densenet-121 Modeling

Pre-trained Densenet-121 Model



평가 방법

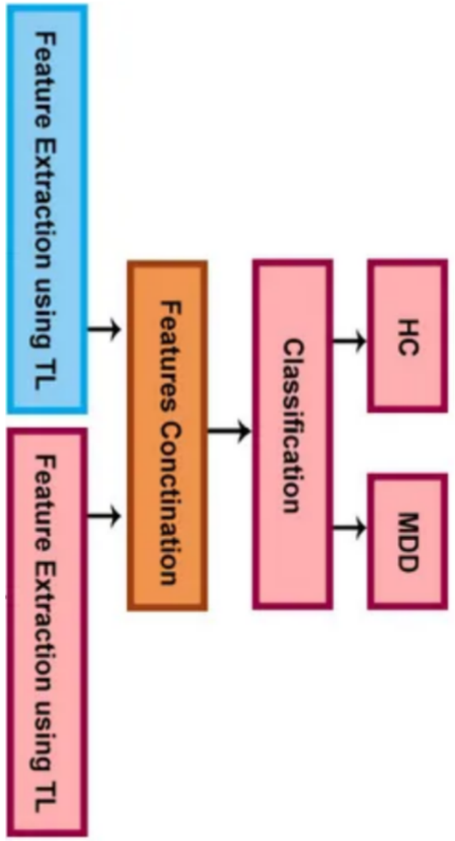
5-Fold Stratified cross-validation

성능 평가 지표

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

Multi Modal Modeling

Multi Modal



평가 방법

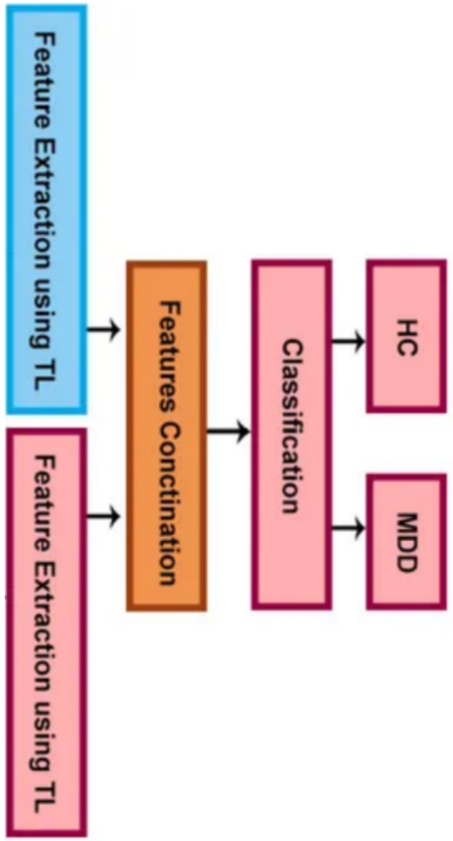
5-Fold Stratified cross-validation

성능 평가 지표

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

Multi Modal Modeling

Multi Modal



입력 형태 :

EEG feature(16(batch), 1024)
Audio feature(16(batch), 1024)

[Parameter]

Optimizer : Adamax

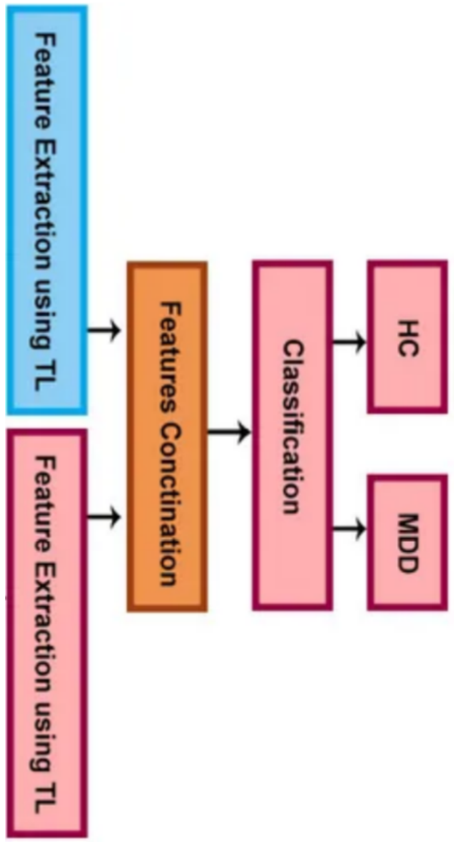
Learning Rate : 0.001

Batch Size : 16

Dropout : 0.3
Optimizer : Adam
Epoch : 100, Early Stopping (최소 15 epoch 보장,
loss 기준 stop, patience=3)
Loss Function : CrossEntropyLoss
Split Ratio – 64:16:20
-(train+validation) : test = 8 : 2
- train : validation = 8 : 2

Multi Modal Modeling

Multi Modal



평가 방법

5-Fold Stratified cross-validation

성능 평가 지표

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

RESULT: EEG, Audio 단일모달

EEG

5-Fold 평균 성능	
Accuracy	0.5816
Precision	0.5518
Recall	0.4657
F1-score	0.4958

AUDIO

5-Fold 평균 성능	
Accuracy	0.7181
Precision	0.7132
Recall	0.5862
F1-score	0.6418

ANALYSIS 1

마지막 Fully Connected Classifier를 제외한 모든 계층이 freeze 되어있으므로
EEG, Audio 패턴을 깊고 정확히 학습하지 못함

전체적인 성능 저조

데이터의 패턴을 깊게 학습하기 위해 freeze 되어 있는 layer를 모두 Fine-tuning

ANALYSIS 1 - RESULT

Fine-tuning 적용 결과: EEG

before

5-Fold 평균 성능	
Accuracy	0.5816
Precision	0.5518
Recall	0.4657
F1-score	0.4958

5-Fold 평균 성능	
Accuracy	0.9121
Precision	0.9594
Recall	0.8433
F1-score	0.8951

after

Fine-tuning 적용 결과: Audio

before

5-Fold 평균 성능	
Accuracy	0.7181
Precision	0.7132
Recall	0.5862
F1-score	0.6418

5-Fold 평균 성능	
Accuracy	0.9236
Precision	0.9181
Recall	0.9044
F1-score	0.9108

after

ANALYSIS 2

EEG는 피험자 수가 MDD:HC = 29:24, Audio는 MDD:HC = 29:22
클래스 간 데이터 불균형이 존재

성능 Table을 봤을 때 EEG와 Audio 모두 Precision에 비해 Recall 현저히 낮고,
F1-score가 낮은 값인 점을 봤을 때 클래스 간 불균형을 해결한다면
성능 개선 가능성 있음

EEG는 HC 5명, Audio는 HC 7명을 랜덤으로 제외하는 방식으로 클래스 간 불균형을 해소함 (EEG-24:24, Audio-22:22)

ANALYSIS 2 - RESULT

Class 불균형 해소 적용 결과: EEG

before

5-Fold 평균 성능	
Accuracy	0.9121
Precision	0.9594
Recall	0.8433
F1-score	0.8951

5-Fold 평균 성능	
Accuracy	0.9166
Precision	0.9337
Recall	0.8979
F1-score	0.9149

after

Class 불균형 해소 적용 결과: Audio

before

5-Fold 평균 성능	
Accuracy	0.9080
Precision	0.8989
Recall	0.8888
F1-score	0.8930

5-Fold 평균 성능	
Accuracy	0.9271
Precision	0.9307
Recall	0.9326
F1-score	0.9292

after

ANALYSIS 3

EEG에서 128개의 채널 중 29개의 채널을 선택할 때, reference 논문에서는 측정 부위(측두엽, 두정엽, 전두엽) 외에는 채널 선정 기준에 대한 언급 없음

EEG 국제 표준인 10-10 system과 10-20 system을 참고하여 채널은 선택했으나, 10-10 System의 전극개수는 21개, 10-20 System의 전극개수는 64개이므로 위치, 개수 등 호환되지 않으므로 최적의 성능을 보여주기에는 한계가 있음

EEG를 활용해 우울증을 분류하거나, HydroCel 128채널을 사용하여 정신질환을 분류하는 논문을 참고하여, 일반적으로 사용하거나 성능이 좋은 부위와 전극을 파악하여 채널 재선택

Channel Selection 참고 문헌

- Sun, S., Li, J., Chen, H., Gong, T., Li, X., & Hu, B. (2020). A study of resting-state EEG biomarkers for depression recognition. arXiv preprint arXiv:2002.11039. - HydroCel 128 channel을 이용하여 우울증 인식용 EEG 바이오마커 탐색
- Van Der Vinne, N., Vollebregt, M. A., Van Putten, M. J., & Arns, M. (2017). Frontal alpha asymmetry as a diagnostic marker in depression: Fact or fiction? A meta-analysis. Neuroimage: clinical, 16, 79-87. - 전두엽 알파 비대칭(F3/F4, F7/F8 등)의 우울증 진단 지표 신뢰성을 메타분석으로 평가한 연구로, 전두엽 채널 선택의 타당성을 검증하는 핵심 근거 제시
- Leuchter, A. F., Cook, I. A., Hunter, A. M., Cai, C., & Horvath, S. (2012). Resting-state quantitative electroencephalography reveals increased neurophysiologic connectivity in depression. PloS one, 7(2), e32508. - 우울증 환자에서 전두-측두-두정 영역 간 resting-state EEG 연결성이 증가함을 입증하여, 이 세 영역 중심의 채널 선택이 타당함을 보여주는 대표적 근거 제시

ANALYSIS 3 - RESULT

Channel re-selection 결과: EEG

before

5-Fold 평균 성능	
Accuracy	0.9166
Precision	0.9337
Recall	0.8979
F1-score	0.9149

5-Fold 평균 성능	
Accuracy	0.9331
Precision	0.9071
Recall	0.9727
F1-score	0.9373

after

EEG를 활용해 우울증을 분류하거나, HydroCel 128채널을 사용하여 정신질환을 분류하는 논문을 참고하여, 일반적으로 사용하거나 성능이 좋은 부위와 전극을 파악하여 채널 재선택

전두(13) - Fp1, Fp2, F1, F2, F3, F4, F7, F8, Fz, AF3, AF4, FC3, FC4
측두(8) - FT7, FT8, T7, T8, TP7, TP8, P7, P8
두정(8) - P3, P4, Pz, P1, P2, P5, P6, CPz

FINAL RESULT

FINAL RESULT: EEG

5-Fold 평균 성능	
Accuracy	0.5816
Precision	0.5518
Recall	0.4657
F1-score	0.4958

논문 구현 방식

5-Fold 평균 성능	
Accuracy	0.9331
Precision	0.9071
Recall	0.9727
F1-score	0.9373

Analysis 적용 후

적용한 기법: Fine-tuning, Class 간 불균형 해소, Channel Selection

FINAL RESULT: Audio

5-Fold 평균 성능	
Accuracy	0.7181
Precision	0.7132
Recall	0.5862
F1-score	0.6418

논문 구현 방식

5-Fold 평균 성능	
Accuracy	0.9271
Precision	0.9307
Recall	0.9326
F1-score	0.9292

Analysis 적용 후

적용한 기법: Fine-tuning, Class 간 불균형 해소

FINAL RESULT: Multi-Modal (EEG + Audio)

5-Fold 평균 성능	
Accuracy	0.7101
Precision	0.7176
Recall	0.7145
F1-score	0.7105

논문 구현 방식

5-Fold 평균 성능	
Accuracy	0.9694
Precision	0.9692
Recall	0.9702
F1-score	0.9696

Analysis 적용

Conclusion

- 논문과 동일하게 구현했으나 EEG와 Audio 모두 논문의 성능에 미치지 못하는 저조한 성능 기록
- 논문과 다른 방식이지만 Class 불균형, Fine-tuning, Channel Selection 등을 통해 성능이 현저히 향상
- 기존 연구와 다른 방식과 기법을 지속적으로 시도할 필요성과 개선 가능성을 확인

Future Work

성능 개선 및 새로운 학습 method 탐색

01

- Fully-Connected layer 구성 변경
- 멀티모달 fusion 방식 변경
- Drop out 등 하이퍼파라미터 조정
- 더 적합한 모델 탐색

실시간 EEG Streaming 분석

02

EEG, Audio 외 다른 종류의 data로 MDD 분류 확장 가능성 연구

03

Thank you.