

Epilepsy Project

- 뇌파 데이터 기반 발작 예측 모델 구현

성공회대학교 2025-상반기 인공지능캡스톤디자인

D팀 : IT융합자율학부 202114110 조예나

목차

ppt

연구 배경 및 목적
데이터셋 및 전처리
모델 구조
실험 방법 및 교차 검증
실험 결과
어려웠던 점
Q&A

notion

연구 배경 및 목적

- 뇌전증(간질) 환자는 갑작스런 발작으로 인해 일상생활에 큰 제약을 받음
- AI와 EEG(뇌파) 신호를 이용하여 실시간 발작 예측 시스템 개발을 목표로 함
- 환자의 안전, 삶의 질 향상에 실질적 기여

데이터셋 및 전처리

공개 EEG 데이터셋(CHB-MIT) 사용

입력: 8개 대표 채널, 80 타임 포인트(약 0.3~0.8초)

슬라이딩 윈도우로 일정 구간씩 잘라 데이터 분할

모델 구조

- 1D CNN + BiLSTM 결합 모델

- CNN: 뇌파 신호의 공간(채널 간) 특징 추출
- BiLSTM: 시간적(연속성) 패턴 학습
- FC Layer: 이진 분류(발작/비발작)

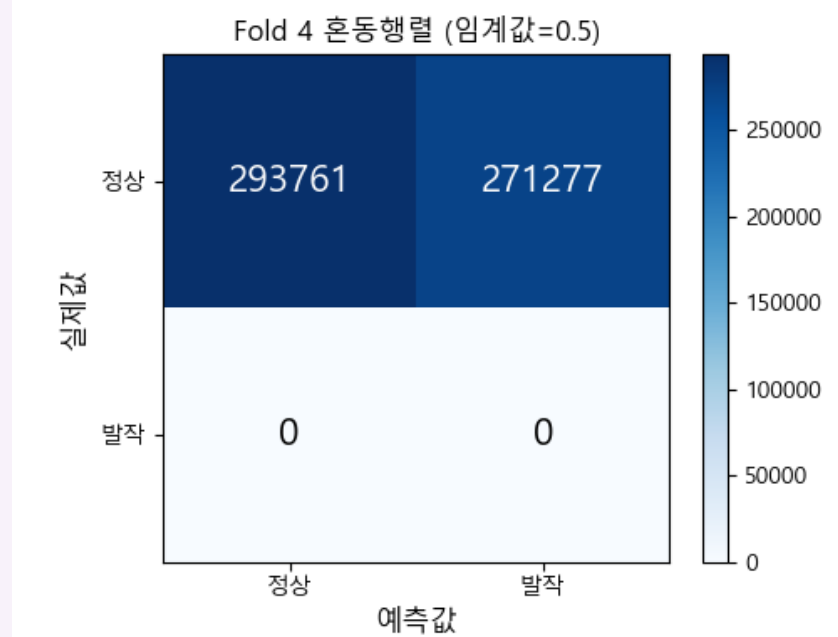
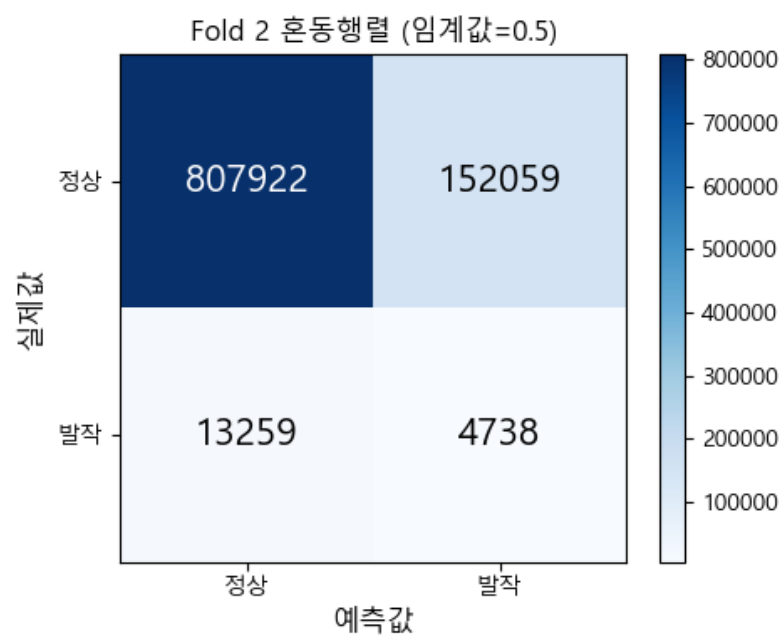
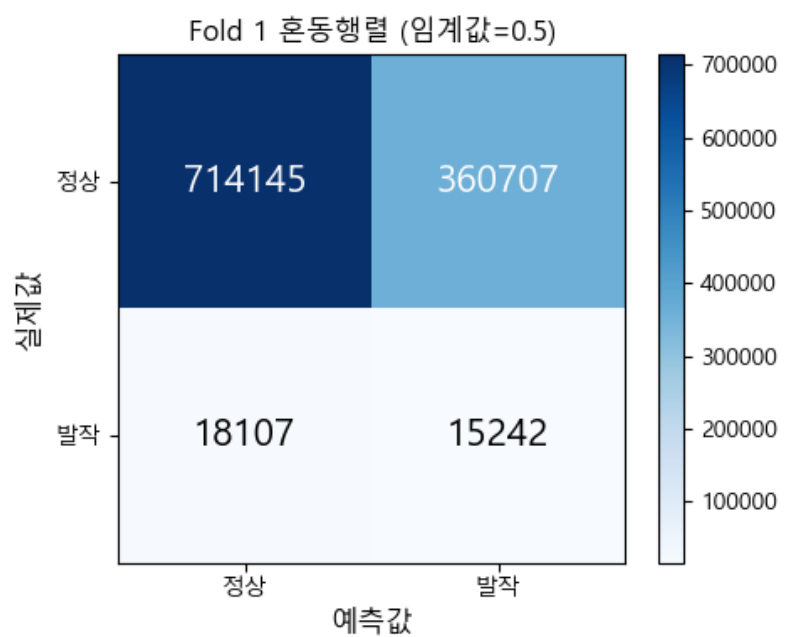
- 경량화 구조 => 이후 하드웨어 구현

실험 방법 및 교차 검증

- 4-Fold 교차 검증: 전체 환자 ID를 4개로 나누어 각 Fold마다 테스트셋 교체
- Fold별 Accuracy, AUC, F1-score 등 성능 측정

실험 결과

- 정확도 약 63% (Fold 평균)
- F1-score, AUC 등 다양한 지표도 함께 평가



어려웠던 점

<데이터 전처리와 입력 구조 설계>

원본 EEG 데이터는 수십 채널, 수십만 타임포인트로 매우 복잡함
모든 채널을 다 쓰면 연산 부담이 크고 실제 적용이 어려워서 대표 8채널만 선정하는 과정에서 고민이 많았음
타임포인트(윈도우 크기)도 너무 짧거나 길면 정보 손실/노이즈/계산비용 증가로 이어져서, 실험적으로 최적의 윈도우(80포인트)를 찾는 과정이 어려웠음

<딥러닝 모델 구조 선택>

CNN, LSTM, BiLSTM 등 여러 구조를 실험했으나, 시간-공간적 특성을 모두 잘 반영하는 구조를 찾기 위해 많은 시행착오를 겪음
특히 CNN-LSTM에서 입력 shape, 차원 변환, 파이토치 텐서 처리에서 자주 오류가 났고, 모델이 학습되지 않거나 loss가 터지는 경험도 있었음

어려웠던 점

<데이터 불균형 문제>

발작(1) 샘플이 비발작(0) 샘플에 비해 매우 적어서, 모델이 비발작에 치우친 예측만 하기도 함
SMOTE 등 오버샘플링 방법, 손실함수 pos_weight 조정 등 다양한 방식으로 데이터 불균형을 보정하려 했으나, 완벽하게 해결하기 어려웠음

<평가지표의 다양성과 해석>

교수님/멘토님 피드백을 받고서야, 단순 정확도가 아닌 F1-score, Precision, Recall, 혼동행렬 등 다양한 지표를 같이 봐야 함을 깨달음
각 지표가 의미하는 바(특히 혼동행렬의 FN/FP, 임상적 해석)가 익숙하지 않아서 한동안 발표/보고서에서 부족하게 설명할 뻔함

어려웠던 점

<실험 자동화 및 교차 검증>

K-fold/Fold별 데이터 분리, 평가 자동화 코드 작성에서 반복적으로 실수가 발생했음
파일 저장 경로, fold별 환자 할당 실수 등으로 재실험/디버깅 시간이 오래 걸렸음

<결과 해석의 한계>

정확도만 높게 나오는 상황이 오히려 "진짜 발작을 못 잡는 것 아닌가?"라는 불안감으로 이어짐
실험 결과가 논문 수준보다 낮거나, 기대만큼 F1-score가 오르지 않아 심리적으로 위축됐던 적도 있음

어려웠던 점

평가지표 및 혼동행렬 설명

- Accuracy: 전체 예측 중 정답 비율
- Precision: 발작 예측 중 실제 발작 비율(오진 적은지)
- Recall: 실제 발작 중 예측 성공 비율(놓침 적은지)
- F1-score: Precision과 Recall의 조화 평균
- Confusion Matrix(혼동행렬):
 - 실제값/예측값이 교차하는 표
 - TP(진짜 발작), FP(오진), FN(놓침), TN(비발작)

결론 및 한계점

- 정확도만으로는 실제 임상적 성능 평가에 한계
- 다양한 지표를 종합적으로 보고 해석 필요
- 향후 계획: 더 많은 채널·윈도우 조합, 실제 임상적용 실험

한 줄 요약

"정확도만 보는 게 아니라 다양한 지표와 혼동행렬까지 함께 확인
해야 임상적으로 의미 있는 EEG 발작 예측 모델이 된다!"

Q&A

<https://www.notion.so/Epilepsy-Project-208c18065ca480648d29df5cd95e32e6>

<https://github.com/yena-yena/Epilepsy-Project>