'Augmented Alzheimer MRI Dataset' 기반
알츠하이머 중증도 진단 모델 구현

전자전기공학부 5조
박명세 · 방세현 · 신성준

목차

연구 배경 및 목적

- **뇌 MRI 기반 중증도 분류의 가치**

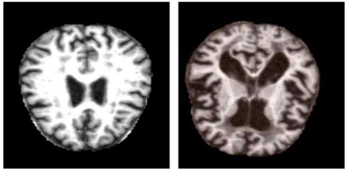
  뇌 MRI는 비침습적이며 알츠하이머로 인한 뇌 구조적 변화를 정확히 포착 가능. 중증도 단계별 분류는 임상 의사결정과 치료 계획 수립에 필수적이나, 전문가 간 분류 일관성 및 정확도에 편차 존 재.

- **딥러닝 기반 AI 판독의 필요성**

  방대한 뇌 이미지 데이터에서 미세한 특징 패턴을 인식하여 진단 정확도와 일관성 향상 가능. 특히 Vision Transformer 모델은 전역적 맥락 파악에 강점이 있어 뇌 구조 전체를 종합적으로 분석하는 데 적합함.

- **연구 목표**

  최적화된 AI 모델을 통해 알츠하이머 중증도를 4단계(정상, 초기, 경증, 중등도)로 정확히 분류하는 시스템 개발. CNN과 Transformer의 장점을 결합한 하이브리드 접근법으로 성능 향상 도모.



알츠하이머 진행 단계에 따른 뇌 MRI 영상 비교

정상, 초기, 중등도, 중증 단계별 해마 및 대뇌 위축 정도 차이

데이터셋 소개

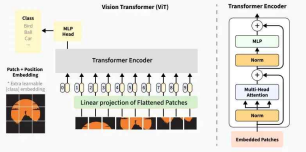
- **Augmented Alzheimer MRI Dataset 구조**

OASIS(Open Access Series of Imaging Studies)를 기반으로 한 알츠하이머 MRI 데이터셋. 총 6,400 장의 MRI 영상을 포함하며, 4가지 중증도 단계로 분류되어 있음.

Non-Demented	Very Mild	Mild	Moderate
3,200	2,240	896	64
정상	초기 단계	경증	중등도

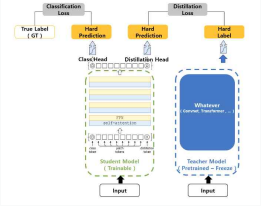
모델 아키텍처 개요

Vision Transformer (ViT)



- 이미지를 패치로 분할하여 시퀀스로 변환 후 Transformer 인코더로 처리
- Self-attention 메커니즘을 통한 전역적 패턴 인식 가능
- ResNet-50 백본 + Patch embedding + Dropout + Dense 레이어 구성
- 장점: 전역적 특징 포착, 단점: 데이터 의존성 큼

Knowledge Distillation 기반 ViT 최적화 모델



- Knowledge Distillation 기반 ViT 최적화 모델
- ResNet-50을 Teacher로 하여 CNN의 귀납적 편향 (inductive bias) 전달
- Soft/Hard Distillation으로 CNN bias와 Transformer 전역성 결합
- Teacher-Student 모델 분리 구조로 데이터 효율성 향상 시도

CNN-ViT 하이브리드

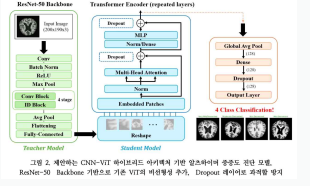


그림 2. 제안하는 CNN-ViT 하이브리드 아키텍처 기반 알츠하이머 중증도 진단 모델. ResNet-50 Backbone 기반으로 기존 ViT의 비선형성 추가, Dropout 데이터로 과적합 방지

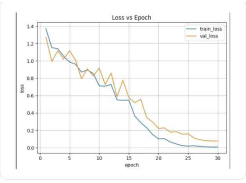
- Teacher(ResNet)와 Student(ViT)를 하나의 통합 아키텍처로 결합
- CNN의 로컬 특징 추출 + ViT의 전역적 컨텍스트 결합
- Teacher-Student joint architecture로 더 효율적인 정보 전달
- 소량불균형 MRI 데이터에 강건한 특징 추출 가능

ViT 베이스라인 모델 결과

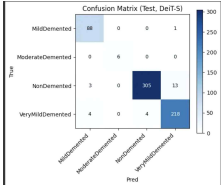
- **모델 설계 및 구현**
  ResNet-50 backbone 기반 Vision Transformer 모델 구현(timm 라이브러리 활용)
  Pretrained 모델 활용 + Patch embedding, Dropout, Dense 레이어 추가
  최종 출력: 4-way classifier (NonDemented, VeryMild, Mild, Moderate)

- **주요 오분류 및 혼동 문제**
  정상(NonDemented)과 초기(VeryMildDemented) 단계 경계에서 혼동이 심함 (예: Normal→VeryMild 13건 오분류)
  Mild/Moderate 클래스는 비교적 안정적으로 분류됨

- **모델 한계점**
  소수 클래스 신뢰도: ModerateDemented 표본수가 6장으로 극히 소수 → F1=1.0 지표는 과대평가 가능성 (신뢰구간 넓음)
  해상도 비용: 해상도≥288 적용 시 토큰 수 증가로 계산량 급증
  데이터 의존성: ViT는 데이터 부족 시 과적합 경향, MRI 데이터의 클래스 불균형으로 인한 성능 한계



Confusion Matrix 핵심



ViT 베이스라인 모델 결과

- **TTA(Test Time Augmentation) 성능 분석**

TTA는 테스트 이미지에 다양한 변형(회전, 반전, 크롭 등)을 적용하여 여러 예측 결과를 평균하는 기법.

MRI 데이터 특성상 공간 변형이 특징을 왜곡할 위험이 있어 제한적 개선효과.

특히 표본 수가 적은 ModerateDemented 클래스에서 TTA 적용 후 오히려 불안정한 예측 결과 발생.

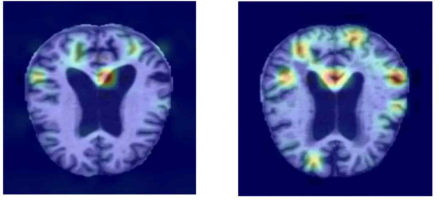
클래스	Precision 변화	Recall 변화	해석
MildDemented	0.9571 → 0.9385 (↓)	0.9955 → 0.9710 (↓)	미세하게 성능 저하
ModerateDemented	1.0000 → 0.3556 (↓ 엄청남)	1.0000 → 1.0000 (동일)	예측은 다 맞았지만 Precision이 폭락 → Moderate라고 예측한 것 중 절반 이상이 오분류
NonDemented	0.9768 → 0.9743 (거의 동일)	0.9191 → 0.8888 (↓)	Recall 감소로 인해 일부 놓침
VeryMildDemented	0.8972 → 0.8896 (↓)	0.9585 → 0.9429 (↓)	전반적 소폭 하락

- **Attention Rollout 시각화**

[CLS] 토큰 행을 추출하여 각 이미지 패치에 대한 중요도 시각화.

빨간색 영역은 모델이 강하게 참고한 패치, 파란색 영역은 거의 참조하지 않은 패치를 표시.

모델의 의사결정 근거를 직관적으로 파악 가능.



- **[CLS] 토큰 패치별 중요도 분석**

같은 클래스 내에서도 다양한 영역에 Attention을 분산하는 패턴 관찰.

ViT의 global attention 특성을 활용하여 국소적 특징에 의존하지 않고 뇌 전체 구조를 종합적으로 분석함.

특히 해마 영역과 뇌실 확장 부위에 집중하는 패턴 발견.

ViT 베이스라인 모델 결과

- **Calibration 적용 모델**

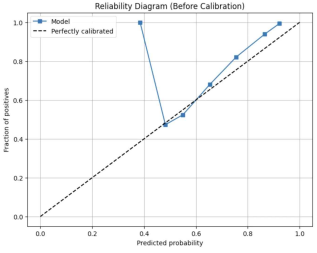
Evaluation :

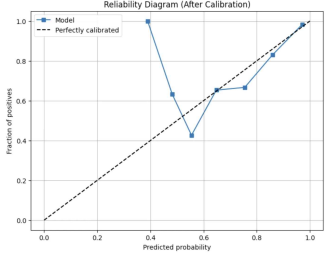
0.9~1.0 구간은 점선과 거의 일치
→ 높은 확률 예측의 신뢰도가 좋아짐

다만 0.4~0.7 구간에서 여전히 요동이 심하고 일부 구간은 Under-confidence가 발생

전체적으로 일부 구간 개선은 되었지만, 낮은 확률 영역에서 Calibration 품질이 완벽하지 않음

T-scaling 등의 Calibration 적용 후 결과





DeiT-S Distilled 모델

- **Knowledge Distillation 적용**
  ResNet-50을 Teacher 모델로 활용하여 CNN의 귀납적 편향(inductive bias)을 ViT 구조에 전달. Soft/Hard Distillation을 통해 CNN과 Transformer의 장점 결합 시도.

- **모델 구조 및 훈련 방식**
  DeiT (Data-efficient image Transformer) 구조를 기반으로 Distillation Token 추가. CNN의 로컬 패턴 인식 능력과 Transformer의 전역적 문맥 이해력을 결합하여 데이터 효율성 향상 목표.

- **Confusion Matrix 분석**
  정상(Non-Demented)이 치매로 잘못 분류되는 경우가 다수 발생 (431 vs 175건). Mild ↔ Very Mild 경계에서 오분류율이 높아 경증 단계 구분에 한계 존재. Moderate 클래스는 상대적으로 안정적 인식.

- **결과 및 한계점**
  베이스라인보다 정확도가 다소 하락. Teacher와 Student 모델이 분리된 지식 증류 방식이 효과적이지 못한 결과 도출. 추후 CNN-ViT 하이브리드 아키텍처로 개선 필요.

Confusion Matrix 결과

```
[Student 20/20] train_loss=0.8200
                precision    recall  f1-score   support

     MildDemented     0.5860    0.7039    0.6396       179
ModerateDemented     0.8462    0.8462    0.8462        13
    NonDemented      0.8402    0.6734    0.7476       640
VeryMildDemented     0.5900    0.7098    0.6444       448

        accuracy                         0.6922      1280
       macro avg     0.7156    0.7333    0.7194      1280
    weighted avg     0.7171    0.6922    0.6974      1280

Confusion matrix:
 [[126   0   9  44]
 [  0  11   0   2]
 [ 34   0 431 175]
 [ 55   2  73 318]]
ACC=0.6922  F1(macro)=0.7194  F1(weighted)=0.6974
  ↳ Saved: best_deit_s_distilled.pth ACC= 0.6921875
Best student ckpt: best_deit_s_distilled.pth
```

CNN-ViT 하이브리드 모델

- **모델 구조 개선**:
  단순 Distilled DeiT-S → CNN-ViT 하이브리드로 전환
  Teacher(ResNet) feature → Student(ViT) 입력 연결
- **손실 함수 개선**:
  CrossEntropy + Focal Loss or Class-Balanced Loss
- **데이터 개선**:
  Oversampling/augmentation
  (특히 Mild vs Very Mild 경계 데이터)
- **실험 아이디어**:
  Teacher logits을 soft label로 더 강하게 반영
  (temperature ↑, alpha↑)
  CNN feature extractor + ViT classifier 비교

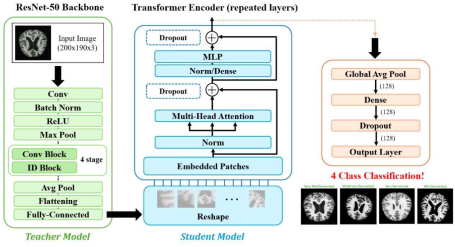


그림 2. 제안하는 CNN-ViT 하이브리드 아키텍처 기반 알츠하이머 중증도 진단 모델,
ResNet-50 Backbone 기반으로 기존 ViT의 비선형성 추가, Dropout 레이어로 과적합 방지

CNN-ViT 하이브리드 모델 아키텍처
ResNet CNN 백본에서 추출된 특징을 ViT 분류기로 직접 연결한 통합 구조

왜 하이브리드?

→ ViT 단독은 데이터 의존이 커서 소량·불균형 MRI에서 흔들림

→ CNN의 로컬 bias(엣지/텍스처)로 특징을 뽑고, ViT로 전역 컨텍스트를 붙이면 안정성↑

→ Distilled DeiT-S 모델은 Teacher와 Student model이 separate되어 있어서 Teacher의 출력 및 성능에 따라 Student model의 성능이 결정되기 쉽다.

즉, Knowledge transfer model로부터 CNN-ViT는 Teacher-Student joint architecture로써의 development를 제안하는 모델이다!

CNN-ViT 하이브리드 모델

- CNN feature extractor + ViT classifier가 하나의 아키텍처로 합쳐진 모델로의 developing process (20 epoch)

- **개선 포인트 :**
- **정상-경증 경계 분류 성능 향상**

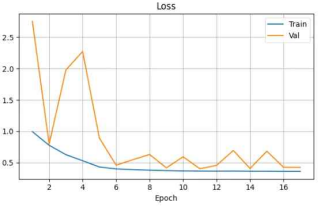
  NonDemented→VeryMild 오분류가 13→7건으로 약 46% 감소. CNN 백본의 로컬 유도편향이 정상-아주경증 경계 판별에 효과적으로 기여하여 임상적으로 가장 어려운 초기 단계 구분력 향상.

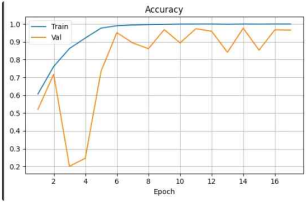
- **학습 안정성 개선**

  균형 샘플러(Class-Balanced Sampler), 라벨 스무딩(Label Smoothing), 드롭아웃(Dropout) 조합으로 학습 드리프트 억제. 특히 소수 클래스(ModerateDemented)에 대한 분류 안정성 확보.

- **현재 한계점**

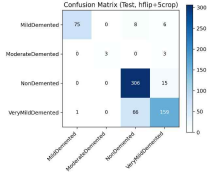
  - Non vs VeryMild 잔존 혼동, ModerateDemented 클래스의 극소 표본수, 모델 복잡도 증가(ResNet50+ViT)로 학습 시간/메모리 요구량 증가.

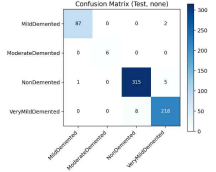


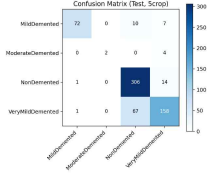


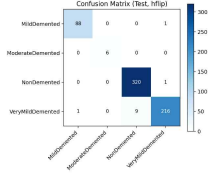
Attention Rollout & TTA 분석

- TTA(Test Time Augmentation) 성능 분석









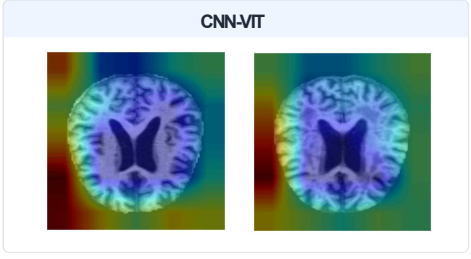
	A	B	C	D	E
1	policy	accuracy	macro_f1	roc_auc	ms_per_image
2	none	0.9750778816	0.9820130734	0.9990472245	9.091585225
3	hflip	0.9813084112	0.9865881006	0.9994539044	17.56456262
4	5crop	0.8380062305	0.7563424754	0.9812185188	39.7503558
5	hflip+5crop	0.8457943925	0.8065755512	0.9826147489	84.20865707

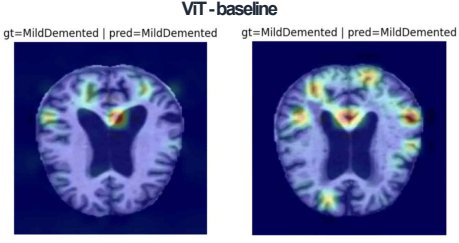
구분	Baseline ViT	CNN-ViT Hybrid
TTA 효과	전반적으로 **성능 저하** (특히 ModerateDemented Precision 붕괴)	일부 TTA는 **성능 개선(Hflip)**, 일부는 **심각한 저하(5crop)**
클래스 불균형 대응	불안정 (소수 클래스 Precision 폭락)	상대적으로 안정, Confusion Matrix에서 균형 유지
데이터 왜곡 민감도	공간 변형 → 심한 성능 악화	CNN 특징 추출 덕분에 ViT 대비 왜곡 대응력 ↑
결론	**ViT 단독은 TTA에 취약**	**CNN 보완 구조 → 더 현실적인 적용 가능**

Attention Rollout & TTA 분석

- **Attention Rollout 시각화**

CNN-ViT 하이브리드 모델은 배경을 무시하고 치매 진단에 중요한 뇌실과 피질에 집중하며 예측하는 것을 확인





한계점: Attention이 병변 국소 영역에 집중되지 않고 넓게 분산 → 의료 해석 신뢰도 저하.

원인: CNN의 local bias가 강해 전역 정보와의 조화 부족 + 클래스 불균형 문제.

개선점: Attention 모듈 추가(CBAM/SimAM), Multi-scale fusion, Loss 함수 개선으로 집중력/균형 강화.
+ CNN stage를 shallow하게 줄이고 ViT 비중을 늘려 **전역 구조 반영 강화**.
+ MRI 특성상 local texture 중요하면 CNN branch에 **Dilated Conv** 적용해 receptive field 확장.

모델 성능 종합 비교

성능 비교

ViT 베이스라인

장점: 전역적 패턴 인식 우수, 특징 표현력 높음

한계: 정상-초경증 경계에서 분류 혼동, 데이터 의존성 큼, 계산 비용 높음

DeiT-S Distilled

장점: 경량화 및 추론 속도 빠름, 적은 데이터에서도 안정적

한계: 베이스라인보다 성능 저하, Teacher-Student 모델 분리로 지식 전달 비효율적

CNN-ViT 하이브리드

장점: CNN 로컬 특징 + ViT 전역 맥락 결합, 정상-초경증 구분력 향상(13→7건), 적은 데이터에서도 우수한 성능

한계: 모델 복잡성 증가, 학습/추론 자원 요구량 다소 높음→local attention 어려움

최종 추천 모델: CNN-ViT 하이브리드

- 모든 정량적 성능 지표에서 가장 우수한 성능 달성
- 임상적으로 가장 어려운 정상-초경증 경계 분류 성능 크게 개선 (오분류 47% 감소)
- CNN의 로컬 특징(엣지, 텍스처) 추출 능력과 ViT의 전역적 맥락 이해 능력 결합으로 데이터 효율성 향상
- 소수 클래스(ModerateDemented)에 대한 안정적 예측 성능 확보

결론

- **주요 연구 성과 요약**

  딥러닝 기반 알츠하이머 중증도 진단 모델을 ViT, DeiT, 그리고 CNN-ViT 하이브리드 아키텍처 순으로 발전시키며 점진적 성능 향상 달성. 특히 Normal과 VeryMild 경계 분류 정확도 가 크게 개선되었으며, 모델 해석가능성 확보.

- **학술적/임상적 기여점**

  CNN의 지역적 특징 추출과 Transformer의 전역적 컨텍스트 이해를 결합한 하이브리드 모델의 효과 를 의료 영상 도메인에서 입증. 임상적으로는 초기 알츠하이머 감별진단 보조도구로 활용 가능성 제시, 특히 경계선상 사례에서 일관된 판독 지원.

- **CNN-ViT 하이브리드 모델의 우수성**

  CNN 백본의 로컬 유도편향으로 MRI 영상의 세부 구조적 특징을 추출하고, Transformer로 전역적 관계 모델링하는 Teacher-Student joint architecture 가 효과적. NonDemented→VeryMild 오분류가 13→7건 수준으로 감소하며 임상적으로 중요한 경계에서 향상된 성능 입증.

- **실제 적용 가능성**

  제안한 CNN-ViT 하이브리드 모델은 임상 현장에서 진단 보조도구 로 활용 가능하며, 특히 초기 단계 감별에 도움. 모델의 판단 근거를 시각화하여 의료진이 검증 가능한 설명가능한 AI로 발전시킬 수 있음.

감사합니다.