

알츠하이머 중증도 진단을 위한 데이터 효율적인 CNN-ViT 하이브리드 모델



성현우(전자전기공학부), 임현오(응용통계학과), 정현석(전자전기공학부)

2024 CUA이 중앙대학교 인공지능 학회 하계 컨퍼런스
Proceeding of 2024 Chung-Ang University Artificial Intelligence Summer Conference

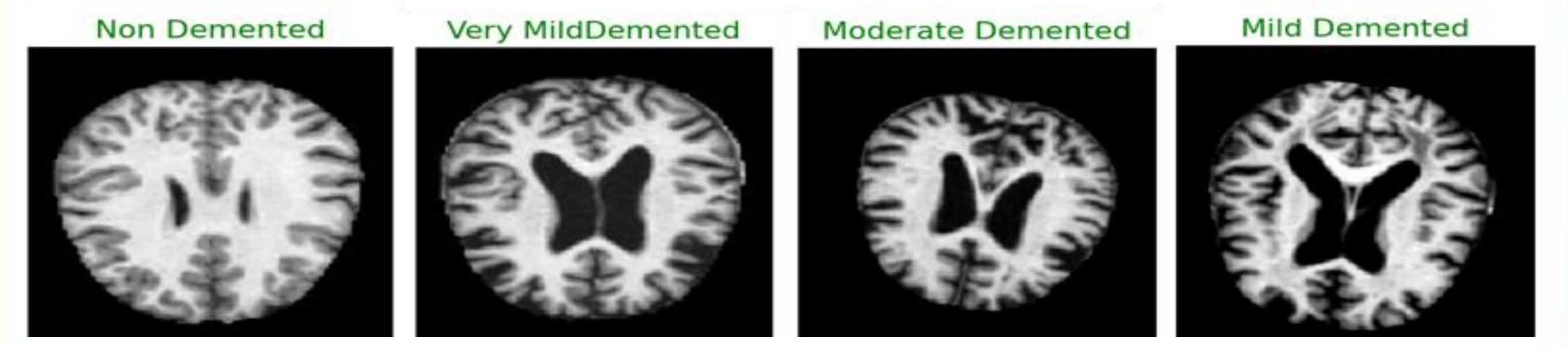
Abstract

본 연구는 알츠하이머 병 중증도 진단을 위해 CNN과 Vision Transformer(ViT)를 결합한 하이브리드 이미지 분류 모델을 제안한다.

MRI 이미지로 구성된 데이터셋을 사용한 실험 결과, 제안된 모델은 86.77%의 정확도와 클래스에 따른 가중화된 0.84의 F1 Score 평균치를 달성했다.

본 모델은 컴퓨팅 비용을 줄이면서도 적은 데이터로도 높은 성능을 유지하며, 의료 분야에서 실용성을 제시한다.

Introduction



메디컬 분야의 어플리케이션을 위해 알츠하이머 중증도 이미지 분류 모델을 설계한다.

ResNet과 ViT를 Base line으로한 융합 아키텍처인 ResNet Back-bone 기반의 ViT 하이브리드 아키텍처를 최적화하여 성능을 비교한다.

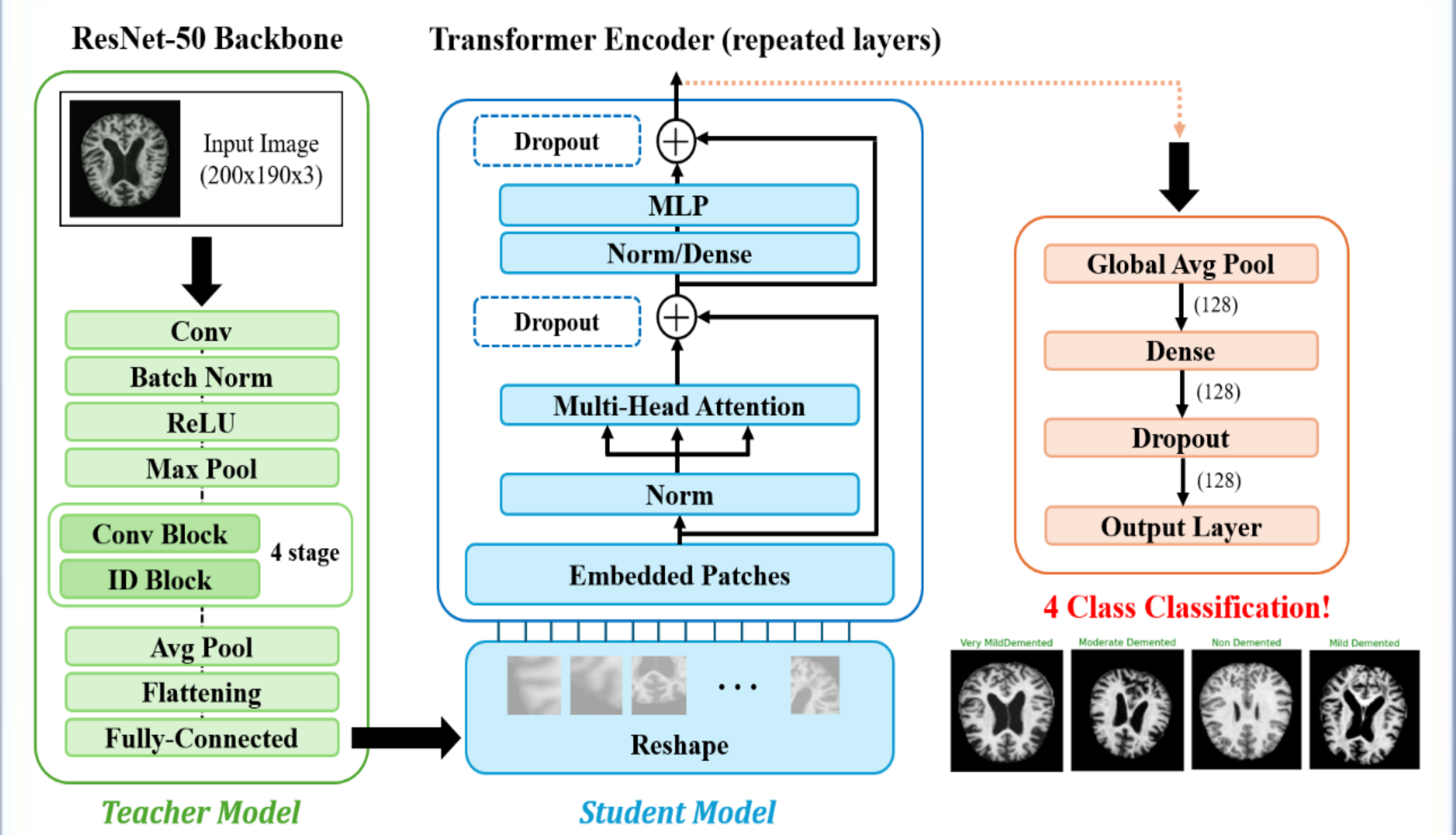
데이터 효율성이 중요한 의료 분야에서 하이브리드 아키텍처를 제안함으로써 성능과 컴퓨팅 리소스, 즉 연산 비용과 메모리 사용량을 최소화한 모델을 구축하여 가능성을 제시하고자 한다.

Aim

알츠하이머 병 중증도 진단 모델: MRI 이미지로 효과적인 진단

CNN-ViT 하이브리드 모델 개발: 기존 CNN과 ViT 아키텍처의 장점을 결합한 새로운 하이브리드 모델을 개발하여 데이터 효율성을 극대화하고, 의료 비용 및 하드웨어 요구 최소화

Methods



기본 DeiT 구조를 유지하며 Teacher 모델은 ResNet-50 Backbone에 기인하며 최종 컨볼루션 레이어의 출력은 Student 모델인 ViT의 입력으로 인가되어 패치 임베딩을 거치게 된다.

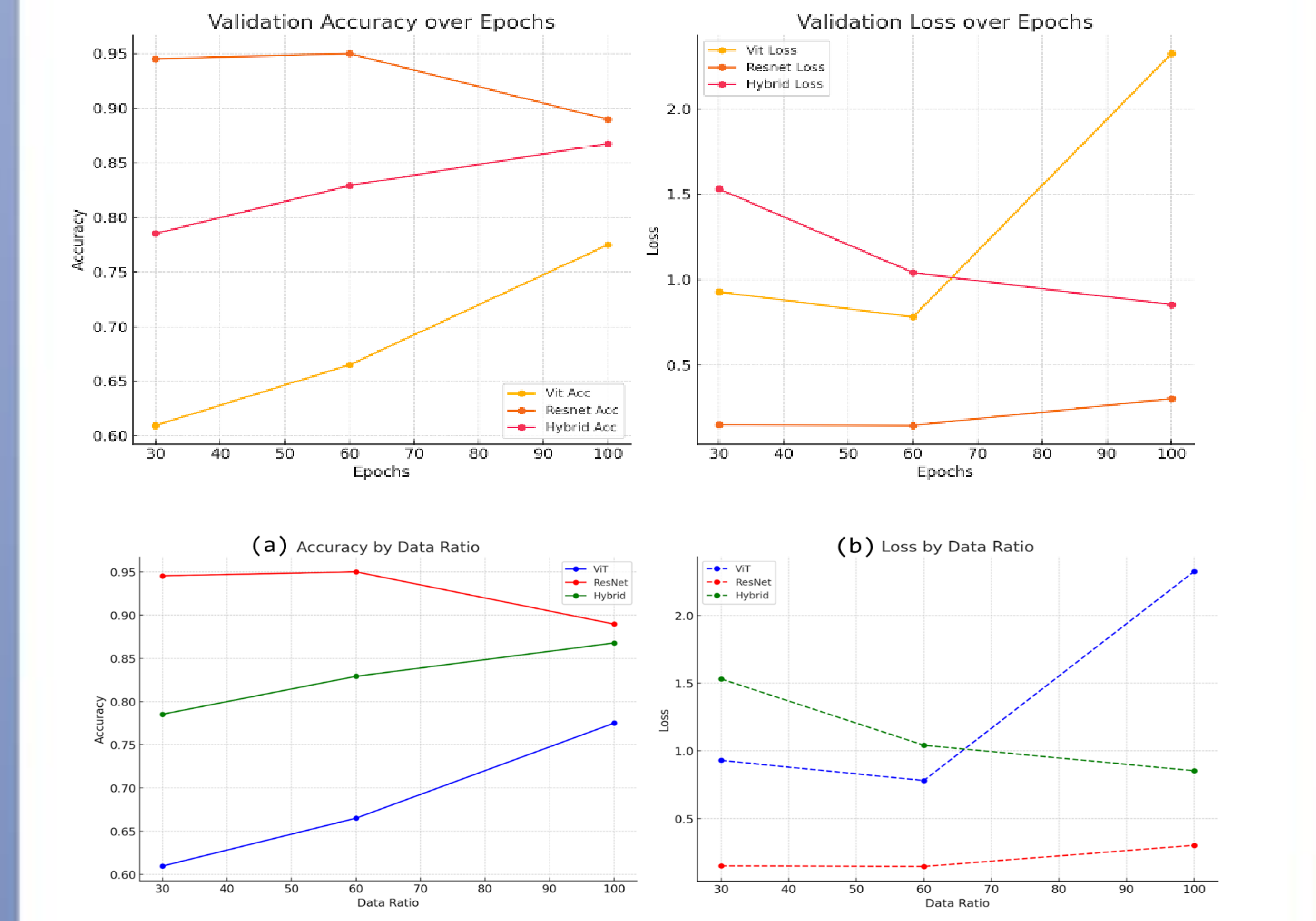
연산량 및 과적합을 줄이기 위한 Dropout 레이어 추가와 더불어 비선형성 강조를 위한 Dense 레이어를 통해 구현했다. 이를 통해 인코딩된 텐서는 최종 Global Average Pooling 이후 알츠하이머 중증도 4가지 클래스에 분류된다.

이로써 기존 CNN 모델의 추가적인 학습 및 가중치 조절이 생략 가능하므로 컴퓨팅 자원을 최소화하여 연산이 가능하며 메모리 사용량 또한 획기적으로 감축할 수 있을 것이 기대된다.

모델 학습 시, 동적 가중치 업데이트 과정에서 Adam Optimizer와 Learning-rate scheduler를 통해 동적으로 학습률을 제어하여 최적화하였다.

CNN-ViT 하이브리드 아키텍처와 더불어 동일 데이터셋으로 학습된 ResNet, ViT 모델의 Validation 데이터에 대한 에폭 당 Accuracy 및 다양한 성능 지표를 분석한다.

Results



Model		Precision	Recall	F1-Score
Hybrid Architecture DeiT	Macro Average	0.86	0.73	0.77
	Weighted Average	0.84	0.84	0.84

Class	Class Count	Precision	Recall	F1-Score
Mild Demented	170	0.77	0.75	0.76
Moderate Demented	13	1.00	0.46	0.63
Non Demented	654	0.89	0.88	0.88
Very Mild Demented	443	0.80	0.83	0.81

Conclusion

해당 연구를 통해 알츠하이머 중증도 진단에 최적화된 이미지 분류 모델을 설계하였다. 적은 하드웨어 리소스와 데이터로 운용 가능한 CNN-ViT 하이브리드 아키텍처는 초기 진단에 응용되어 사회적 순기능을 기대할 수 있으며, 의료 분야에 더욱 최적화된 모델로 발전할 수 있을 것이라 기대한다.

Reference

Alexey, D. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929

Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2021). Training data-efficient image transformers & distillation through attention. In International Conference on Machine Learning (pp. 10347-10357).