

Abstract

본 연구는 피부암 진단을 위해, 이미지 데이터와 CSV 데이터를 결합한 통합 모델을 제안한다. 딥러닝과 부스팅 기반의 머신러닝 기법을 활용하여, CSV 데이터와 이미지 피처를 통합함으로써 높은 예측 정확도를 달성하였다. 기존 연구와 비교한 결과, 본 연구의 접근법이 우수한 성능을 보였으며, 피부암의 조기 진단에 있어 실용성을 입증하였다.

Introduction

피부암은 조기에 발견되지 않을 경우, 치명적인 결과를 초래할 수 있는 질병이다. 조기 발견과 치료는 피부암의 생존율을 크게 향상시킬 수 있지만, 많은 사람들이 적시에 피부암을 발견하지 못하는 경우가 빈번하다. 이러한 문제는 특히 의료 자원이 제한된 지역에서 더욱 심각하게 나타난다.

Aim

위 문제를 해결하기 위해, 저해상도 이미지 (예: 휴대폰 사진)를 활용하여, 피부암을 구별할 수 있는 AI 알고리즘의 개발을 연구 과제로 선정했다. 본 연구에서는 병변의 이미지와 함께, CSV 데이터 (병변의 특징 및 환자 정보)를 통합 학습하여, 진단 정확도를 향상시키는 AI 모델 연구를 목표하였다.

Methods

1) 모델 구조

딥러닝 모델을 활용하여, 이미지 데이터로부터 피처 벡터를 생성한 후, 이를 CSV 데이터와 결합하여, 부스팅 기법을 기반으로 학습하였다.

2) CSV 데이터 피처 엔지니어링

피부암 진단을 위한 'ABCD 규칙'을 기반으로, 병변의 비대칭성, 경계의 불규칙성 등 중요한 특징들을 추가하였다.

3) 이미지 처리

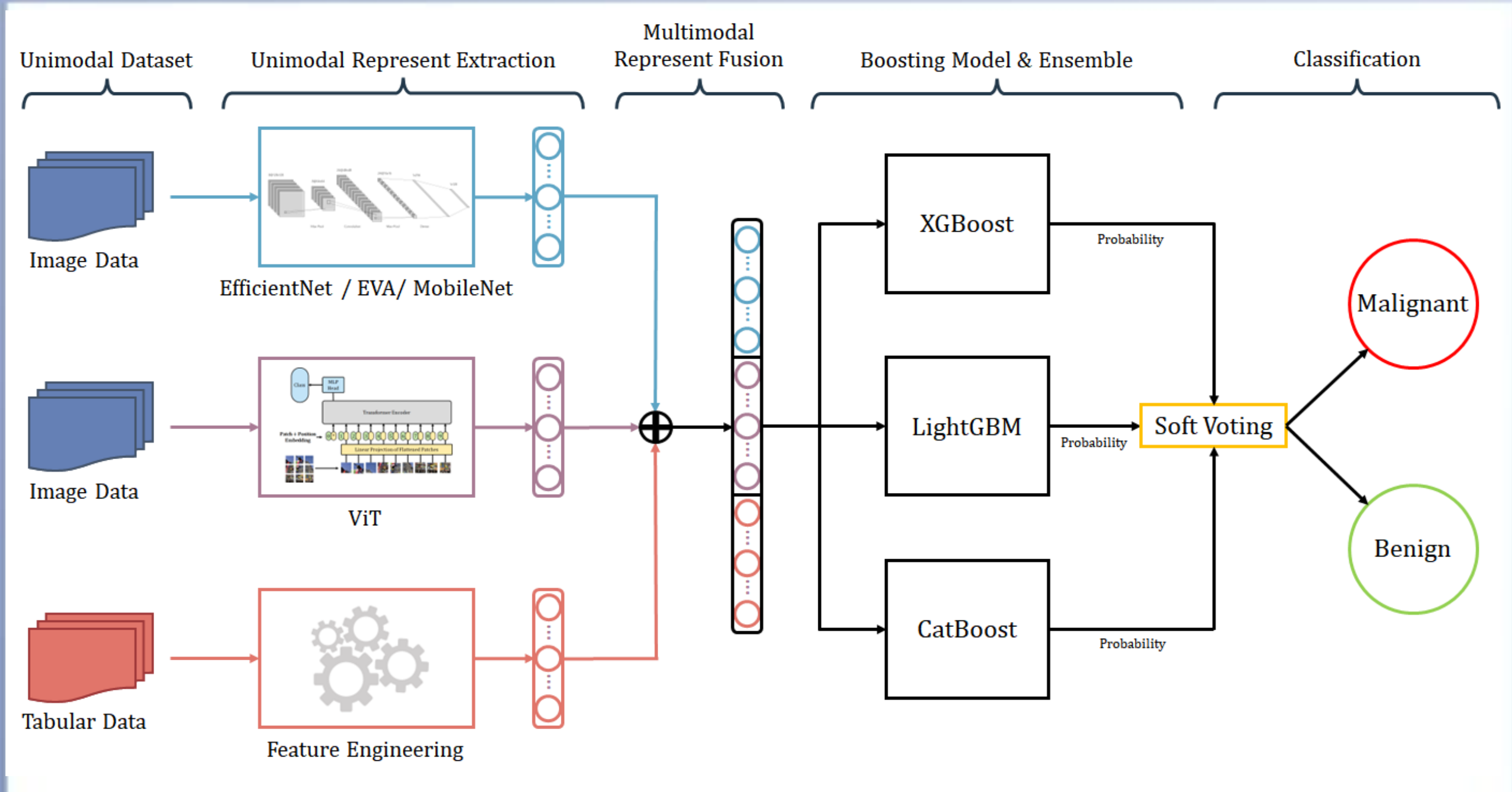
1. 딥러닝 모델이 학습한 예측 결과를 피처로 활용한다. (EfficientNet V1 + EVA-02)
2. 딥러닝 모델의 분류기 이전 레이어에서 추출한 피처를 활용한다. (MobileNet V3 Large + ViT Base)

4) 데이터 불균형

불균형을 지니고 있는 데이터를 해결하기 위해, UnderSampling 및 OverSampling 기법을 활용하였다. 또한, 모델의 일반화 성능을 향상시키기 위해 다양한 이미지 증강 기법들을 적용하였다.

5) 부스팅 기반 학습

XGBoost, LightGBM, CatBoost와 같은 모델을 통해 학습을 진행하였다. 각각의 모델들은 개별적으로 학습되어, Soft Voting으로 앙상블하였다.



Results

Model	pAUC
CNN + FFN Concat	0.151
DL + Boosting	0.183

본 연구는 Qishen Ha의 기존 연구와 비교하여, 우수한 성능을 발휘할 수 있음을 확인하였다.

Conclusion

본 연구는 기존 연구를 바탕으로 피부암 진단의 정확도를 높이기 위해, 의학 정보와 이미지 데이터를 결합하는 새로운 접근법을 제시한다. 환자의 기존 의학 정보를 정형화하고, 피부 병변 이미지를 딥러닝으로 분석하여 얻은 데이터를 부스팅 기반 모델에 함께 학습하였다. 그 결과, 기존의 단일 데이터 방식보다 더욱 정확한 피부암 진단이 가능함을 확인했으며, 이는 의료 정보와 이미지 데이터의 상호 보완적인 특성을 활용하여 진단의 신뢰성을 높이는데 기여할 것으로 기대된다.