

대상 데이터

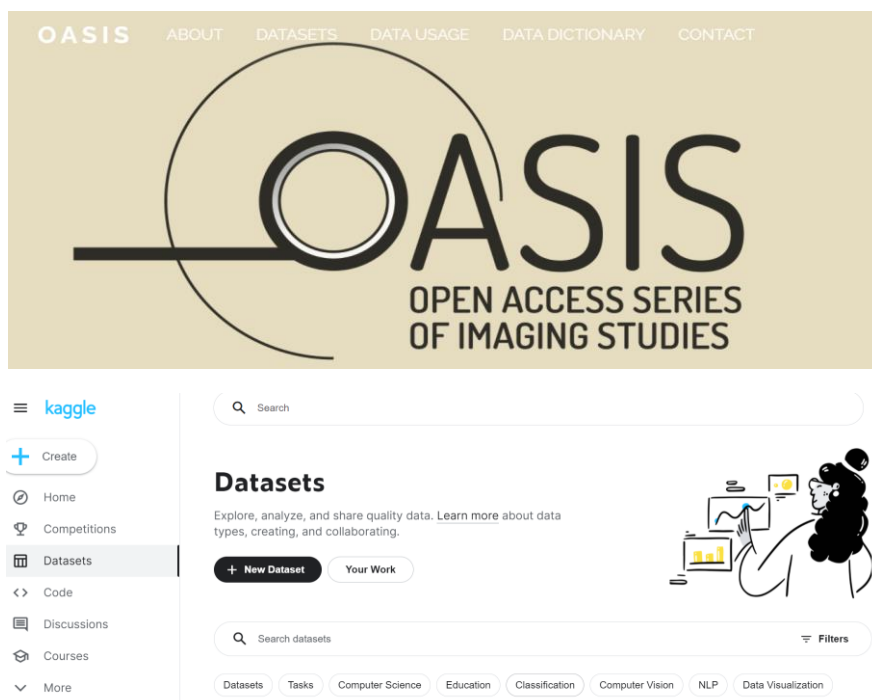
1. MRI data sets of Open Access Series of Imaging Studies (OASIS)

Magnetic Resonance Imaging Comparisons of Demented and Nondemented Adults

2. Images of MRI Segementation

데이터 수집 계획

일반 의료 데이터에는 개인정보 이슈로 인해 권한 문제가 있어 접근이 어렵다. 따라서 누구나 다 사용할 수 있도록 올려놓은 공공 데이터를 사용할 예정이다. 사이트는 OASIS와 kaggle을 사용할 것이다.



분석 방법론

크게 두 가지 방법을 사용하고자 한다. 첫 번째로는 머신러닝 분석 기법을 활용해 조기 진단이 가능하게 하는 알고리즘을 개발할 것이다. 이때 Open Access Series of Imaging Studies의 MRI 데이터를 활용할 것이다. 두 번째로는 Images of MRI Segementation 데이터를 사용해 MRI 이미지를 input으로 넣었을 때 이 이미지가 치매 인지 아닌지를 판단할 수 있는 알고리즘을 개발할 것이다. 개발 방법으로는 컴퓨터 비전(Computer vision) 중 하나인 image classification을 활용할 것이다. 두 가지 방식을 이용해 웹 어플리케이션이나 안드로이드 앱에 배포해 실제로 사용할 수 있게끔 만들 것이다.

1. 머신러닝 기법 활용

일단 첫 번째로 개발을 통해 해결하고자 하는 문제를 정의한다. 필자는 다음의 분석을 통

해 치매를 조기 진단 할 수 있는 알고리즘을 개발하고 이를 웹 어플리케이션이나 안드로이드 앱으로 배포할 예정이다.

다음으로는 데이터에 대한 이해가 필요하다. 가져온 데이터는 어떤 형식으로 구성되어 있는지, 각 행과 열은 어떤 값과 의미를 가지고 있는지 파악해야 한다. 예를 들어 필자가 가져오고 하는 데이터에서 피실험자 중 72명은 연구 기간 내내 '치매가 아닌' 사람으로 분류되었다는 등의 데이터 자체에 대한 정보를 의미한다.

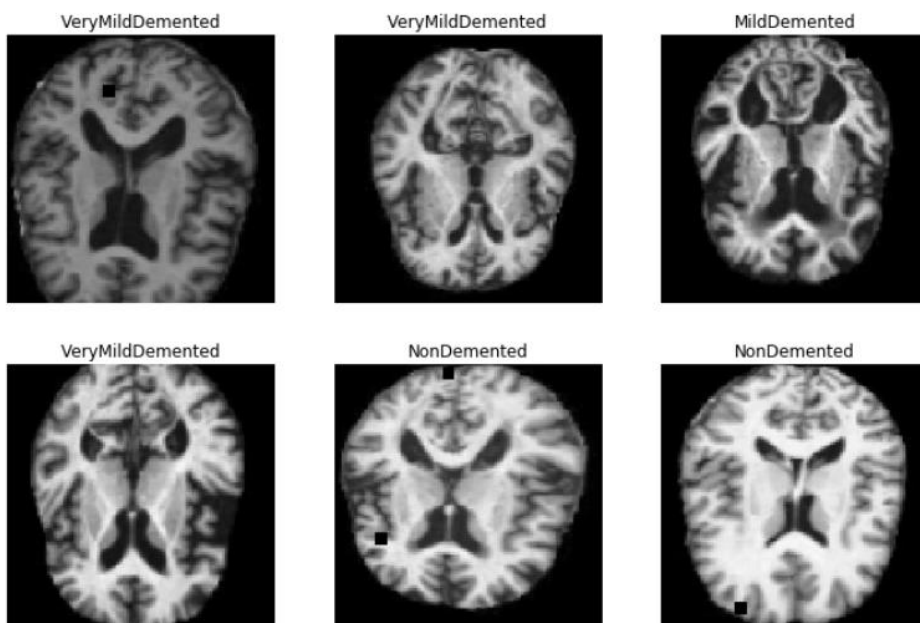
그 다음으로 EDA 즉, Exploratory Data Analysis를 해야하는데, 열을 기준으로 각 열에 대한 행의 값의 통계치를 구하거나 데이터를 불러와 데이터 프레임 형식으로 읽어 다양한 코드를 이용해 데이터 안의 변수들이 서로 어떤 관계를 가지고 있는 데이터 자체는 어떠한지 자세하게 살펴봐야 한다. 위에서 다양한 코드라고 함은 예를 들어, group by를 이용해 확인하고 싶은 형태로 데이터를 그룹핑하고, 그룹핑한 데이터를 이용해 바차트나 매핑 차트를 그려 확인하고자 했던 내용을 시각화한다는 것이다.

이어 확인한 결과를 바탕으로 데이터는 결측치 제거, 데이터셋 분할 등의 전처리 과정을 거쳐야 한다.

이 다음 모델을 생성한다. 이때 다양한 모델을 사용해본 후 정확도가 가장 높은 모델을 과적합이 일어나지 않았다는 조건 하에 채택하면 된다. 모델의 종류로는 logistic regression, support vector machine, decision tree, random forest, adaboost 등이 있다. 이후 생성한 모델들을 accuracy나 AUC 등으로 평가한 후 가장 최적의 모델을 채택한다.

이러한 머신러닝 분석 기법은 특성 분석으로 같은 특성을 가지고 있는 사람들의 치매 조기 진단이 가능한 하지만 이것만으로 정확한 예측을 했다고 단언하기는 어렵다. 이를 보완하고자 MRI 이미지 분류를 통해 좀 더 정확한 예측을 사용자들에게 제공하고자 한다.

2. 딥러닝 – Computer Vision (Image Classification) 활용



다음은 이미지 데이터를 이용해 이를 분류할 수 있는 알고리즘을 개발하고자 한다.

먼저 이미지 데이터이기 때문에 데이터를 시각화한다. 이를 통해 어떤 이미지가 어떤 상태인지 파악한다.

이후 데이터를 one-hot 인코딩의 방법으로 인코딩 작업을 하고 인코딩한 데이터로 데이터 셋 자체가 균형적인지 아닌지에 대해 판단해야 한다. 이를 하는 이유는 균형적이지 않은 데이터셋에서는 우리가 흔히 사용하는 평가지표인 accuracy를 사용할 수 없다. 그렇기 때문에 데이터셋을 확인했을 때 균형적인 데이터셋이면 accuracy, 아니라면 ROC나 AUC를 사용하곤 한다.

평가지표를 결정한 후에는 모델을 생성한다. 이때 어떤 알고리즘을 쓸 것인지 결정하고 과적합 방지를 위해 Dropout과 같은 규제나 early callback stop을 설정해야 한다. 또 callback함수를 사용하여 learning rate에 대한 스케줄러나 모델에 대한 체크포인트 또한 생성해준다. 위에서 생성한 모든 것들을 활용해 모델을 훈련시키면 된다.

결과 값으로 나온 것을 확인하기 쉽게 시각화하면 모델에 사용한 알고리즘이 적절한지 아닌지, 에포크의 수는 적당한지, 과적합은 일어나지 않았는지 등을 파악하고 모델을 평가할 수 있다.

분석 의의 및 가치

치매(Dementia)란 정상적으로 활동을 할 수 있었던 사람이 뇌에 발생한 여러 질환으로 인해 인지 기능을 상실하고 일상생활을 지속해 나갈 수 없어진 경우의 상태를 의미한다. 인지 기능의 상실은 미국 정신의학회의 DSM-IV 기준에 따르면 기억력 저하와 함께 실어증, 실행증, 실인증, 실행 기능 장애 중 최소한 하나가 동반된 상태로 전반적인 인지 기능의 지속적 저하를 의미한다고 한다.

치매는 일종의 증상이기 때문에 이를 일으키는 원인 질환은 실제로 다양하다고 한다. 치매의 원인질환들은 우리가 흔히 알고 있는 알츠하이머병(Alzheimer)인 퇴행성 뇌질환, 혈관성 치매에 속하는 뇌혈관 질환, 감염성 질환, 중독성 질환, 경련성 질환 등으로 매우 다양하다. 대표적인 것은 알츠하이머형 치매와 혈관성 치매로 약 80-90%를 차지하고 있다. 그 중 알츠하이머형 치매는 중년과 노년기에 발생하는 치매에서 가장 흔한 유형으로 전체 치매 중 약 60-70%를 차지한다고 한다. 이 알츠하이머형 치매가 가장 흔히 있는 치매 질환이다 보니 종종 치매 질환을 나눌 때 알츠하이머형 치매와 비알츠하이머형 치매, 두 가지로 나누기도 한다고 한다.

치매의 원인질환을 알아내는 것이 매우 중요한 이유는 퇴행성 질환을 제외하고는 치료가 가능하거나 조기에 발견하면 더 이상의 진행을 막을 수 있기 때문이라고 한다. 양성 뇌종양이나 갑상선 질환, 신경계 감염, 비타민 부족 등으로 인한 치매는 전체 치매의 10-15%를 차지한다고 하는데, 이 질환들은 적절히 치료하거나 조절하면 완치될 수 있다고 한다. 또한 우리나라에 많은 혈관성 치매는 조기에 발견하면 더 이상의 진행을 막을 수 있고 예방이 가능하므로 원인질환을 알아내거나 조기진단이 매우 중요하다.

어느 질병이나 질환이든 진단 후 관리보다 조기진단을 통해 예방하는 편이 나을 것이다. 특히 가장 흔히 발병하는 알츠하이머형 치매는 현재까지도 근본적인 치료방법이 존재하지 않고 인지기능 개선 및 저하 지연 정도에 그치고 있다. 해서 해당 연구는 치매라는 질환을 조기진단 할 수 있게끔 도움을 줄 수 있는 알고리즘 개발을 진행하려고 한다.

현재 우리나라는 점차 고령화 사회로 진입과 동시에 치매 발병률이 증가하고 있는 상황이다. 이에 따른 의료비 부담 증가로 치매 환자의 조기진단이 매우 중요해진 상황이라고 할 수 있다. 또한 앞서 언급했다싶이, 치매는 조기에 진단이 이루어지지 않고 발병 후 치료를 해야하는 상황이 오게 되면 자체적인 치료 방법이 없을 뿐만 아니라 회복 자체도 쉽지 않아 예방과 조기 검진을 통한 선제적 관리가 최선의 방법이라고 할 수 있다.

제시한 연구를 통해 치매조기진단을 통한 예방과 선제적 관리를 위해 AI학습 모델을 연구해 제공함으로써 현재 우리 사회가 직면에 문제에 긍정적인 기여할 수 있음을 기대한다.

<참고 문헌>

서울아산병원|정신건강의학과[웹사이트].(2021.08.28).URL:

<http://psy.amc.seoul.kr/asan/depts/psy/K/bbsDetail.do?menuId=862&contentId=202060>

전남대학교병원[웹사이트].(2021.08.28).URL:

https://www.cnuh.com/health/disease.cs;WEB_JSESSIONID=1BFB761C16F6FB81D7C88BA9E4B0AD9E?act=view&infold=279&searchKeyword=&searchCondition=&pageIndex=54