

PBL 콘텐츠 설계서

PROJECT 1: 뇌 CT 분석



2021 March 30

매직에코



**X**

# 목차

[목차 1](#_Toc71561599)

[백그라운드 스토리 2](#_Toc71561600)

[학습해야 하는 키워드에 대한 설명 2](#_Toc71561601)

[CT 이미지를 분석한 뇌출혈 예측 2](#_Toc71561602)

[프로젝트의 목적 2](#_Toc71561603)

[이미지 데이터 분석 및 학습 2](#_Toc71561604)

[수준 구분 3](#_Toc71561605)

[관련 기술 3](#_Toc71561606)

[Knowledge Skill 3](#_Toc71561607)

[Acquire Skill 3](#_Toc71561608)

[미션 별 목표 및 평가기준 4](#_Toc71561609)

[Mission 2 - CT 이미지 분석 4](#_Toc71561610)

[Mission 2 - 분석한 뇌 CT 이미지를 기반으로 CNN 모델을 개발 5](#_Toc71561611)

[모범 답안 6](#_Toc71561612)

[데이터 파일 설명 6](#_Toc71561613)

[이미지 파일 6](#_Toc71561614)

[labels.csv 6](#_Toc71561615)

[학습 지원 정보 7](#_Toc71561616)

[학습 리소스 7](#_Toc71561617)

[포럼 7](#_Toc71561618)

# 백그라운드 스토리

XX병원은 노쇼 문제를 해결하자 더욱 많은 손님들로 붐비게 되었다. 특히 XX병원의 뇌신경센터는 전국적으로 유명하여 항상 환자로 센터가 가득 찼지만, 노쇼 문제가 해결되자 터져나오는 손님으로 센터가 마비가 될 정도였다. 넘쳐나는 환자들은 뇌신경센터에 인력난을 불러왔고, 병원 운영팀은 인력 보충을 약속했지만 곧바로 전문적인 인력이 보충되기는 말처럼 쉬운 일이 아니었다.

이러한 상황에서 의료사고가 발생했다. 하루에 30명 이상의 환자들을 상대하던 방사선 전문의 A는 뇌졸증 환자를 뇌경색으로 판단하였고 결국 환자는 뇌사상태에 빠졌다. 뇌졸증의 치료는 시작부터 끝까지 모든 순간이 중요하지만 환자의 뇌 컴퓨터단층촬영(CT) 이미지로부터 환자에게 뇌출혈이 있는지 판단하는 과정이 가장 중요하다. 이렇게 출혈을 감지하는 작업은 특별히 훈련된 방사선 전문의들에게도 높은 집중력을 필요로 하는 까다로운 작업이다.

이러한 문제가 발생하자 병원 혁신팀은 뇌출혈을 정확하여 의료사고를 예방하고 의료진들의 피로를 덜어주고자 뇌출혈 판단 모델을 만들기로 결정하였다. 이제까지의 뇌출혈 환자들의 CT사진을 식별하여 뇌출혈 판단 모델을 만들고 해당 모델을 활용하여 새로운 환자의 CT사진만 갖고도 그 환자가 뇌출혈을 갖고 있는지 빠르게 판단하고, 의사들로 하여금 결정을 도울 수 있도록 하였다.

# 학습해야 하는 키워드에 대한 설명

## CT 이미지를 분석한 뇌출혈 예측

CT 검사 시 뇌출혈의 부위는 심각한 정도에 따라 인지하기 어려운 경우도 있으며, 응급상황에서 숙련이 되지 않은 의료진에게는 부담을 준다. 응급상황에서 검사와 동시에 뇌출혈부위를 자동으로 빨리 파악하고 정량적인 정보를 제공하는 보조적인 역할은 필요하며, 컴퓨터를 이용한 자동 검출 및 인식 시스템은 출혈부위 진단에 매우 큰 도움을 줄 수 있다. 본 연구에서는 출혈부위를 문턱치 값 처리, 모폴로지 연산, 원형률 계산을 접목하여 뇌 출혈부위의 자동검출에 성공하였고 검출 후보군에서 잘 못 선정된 영역을 판정하기위한 주성분분석을 이용한 분류기 개발에 성공하였다. 개발된 시스템을 응급상황의 뇌출혈 환자에게 적용한다면 의료진에게 수술계획을 위한 유용한 정보가 될 것으로 사료된다.

프로젝트의 목적

이미지 데이터 분석 및 학습

|  |  |
| --- | --- |
| 구분 | 설명 |
| **1** | 데이터 저장 및 준비 등 데이터 전처리 과정을 경험할 수 있다. |
| **2** | 데이터 전처리 과정 중 이미지 변형에 대해 학습할 수 있다. |
| **3** | 모델링 과정을 통해 CNN을 경험할 수 있다. |

# 수준 구분

* AI/Data분석/Edu4
* AI/모델개발/Edu4
* AI/AIOps/Edu4

# 관련 기술

### Knowledge Skill

* Fundamentals > Python Programming > Python Basics
* Fundamentals > Python Programming > Numpy
* Fundamentals > Python Programming > Pandas
* Data Scientist > Visualization > Matplotlib
* Data Scientist > Visualization > seaborn
* Machine Learning > General
* Machine Learning > Tools > scikit – learn

### Acquire Skill

* Data Engineer > Summary of Data Formats
* Data Engineer > Data Discovery
* Data Engineer > Data Integration
* Data Engineer > Data Fusion
* Deep Learning > Tools > Tensorflow
* Deep Learning > Tools > PyTorch

# 미션 별 목표 및 평가기준

## Mission 2 - CT 이미지 분석

**키워드**

numpy hist, image to numpy array, opencv, ImageDataGenerator, train\_test\_split

**설명**

* 구글드라이브에 환자 진료 데이터를 저장하고 구글 드라이브를 구글 Collaboratory에서 사용할 있도록 마운트하라.
* 그리고 데이터들을 다양한 그래프로 출력하여 분석한다. 이미지는 학습을 위해 128, 128 크기로 저장한다.
* 또한 학습을 위해 데이터를 train 80%, validation 10%, test 10%로 나누어라.
* 데이터 변환으로 추가데이터를 확보한다는 것은 train 과 validation 데이터를 가공하여 추가 데이터를 확보한다는 의미이다.
* test 데이터는 가공할 이유가 없으므로 추가데이터 확보는 반드시 데이터를 train 과 test 로 구분하고 난 뒤에 train에 대해서만 가공한다.

**목표**

|  |  |  |
| --- | --- | --- |
| 분류 | 경험 | 난이도 |
| Data분석(데이터 추출 및 가공 역량) | 수집된 데이터를 특정 포맷으로 가공하고 정렬할 수 있다. | 3 |
| Data분석(데이터 추론 역량) | 모집단으로부터 추출된 표본을 바탕으로 통계적 추론이 가능하다 | 4 |
| Data분석(데이터 재가공 및 전산화 역량) | 데이터의 특성을 파악해, 수집된 데이터를 재 가공할 수 있다. | 3 |
| Data분석(데이터 재가공 및 전산화 역량) | 비정형데이터(이미지 등)에 메타데이터(레이블링)를 추가해 데이터셋을 구성할 수 있다. | 3 |

**결과물**

* ImageDateGenerator의 발표 동영상

**평가 기준**

|  |  |
| --- | --- |
| 멘토 평가 항목 | 구분 |
| 1. try except의 사용으로 데이터 엑세스 에러에 대처했는가? | Data Store |
| 2. 구글드라이브에 저장했는가? | Data Store |
| 3. 그래프를 많이 만들어서 데이터 분석을 많이 할수록 높은 점수 부여 | Data Analysis |
| 4. 이미지는 128, 128 크기로 저장했는가? | Data Analysis |
| 5. 데이터를 train, validation, test로 나누었는가? (train\_test\_split을 사용하지 않았어도 감점 요소가 아님) | Data Split |
| 6. ImageDateGenerator를 사용하였는가? | Image Data Generator |
| 7. 발표를 잘할 수록 높은 점수 부여 | Image Data Generator |

**확인 질문**

|  |  |  |
| --- | --- | --- |
| 분류 | 문제 | 정답 |
| Data Analysis | 이미지를 pyplot에서 출력하기 위해서 사용하는 메소드는? | imshow() |
| Data Split | 퀴즈1: train\_test\_split을 사용할 때, 실행 시 마다 같은 학습 데이터를 얻고 싶다. 어떤 인자 값을 어떻게 해야 하는가? | random\_state에 임의의 값을 넣어준다. 시드를 고정하기 때문에 실생 시마다 같은 학습 데이터를 얻는다. |

## Mission 2 - 분석한 뇌 CT 이미지를 기반으로 CNN 모델을 개발

**키워드**

CNN 기초

**설명**

* CNN 모델링에 대해 학습을 진행한다. 이 때, 구현 중심으로 학습한다.
* 여러 알고리즘을 사용하여 CNN모델을 만들어라, 여러 개의 Conv Layer를 사용해 보고 최적의 모델을 개발하라 그리고 만든 모델을 훈련시켜라
* relu, maxpooling, sigmoid, binary\_crossentropy 에 대해서 조사하고 발표하라
* train, validation, test의 accuracy를 구하고 history를 이용하여 모델의 accuracy 와 loss에 대한 그래프를 그리고 평가하라.
* accuracy 와 loss 외의 그래프를 추가하여 결과를 설명하라. (optional)
* 그래프를 해석하고 결과를 설명하는 발표를 하라

**목표**

|  |  |  |
| --- | --- | --- |
| 분류 | 경험 | 난이도 |
| AI모델개발(이미지 관련 프로젝트 수행 역량) | 개와 고양이 사진을 넣고, 개와 고양이를 CNN모델로 분류하기 위해 개와 고양이 이미지를 학습, 검증, 예측 데이터셋으로 나눈다. | 4 |
| AI모델개발(이미지 관련 프로젝트 수행 역량) | 개와 고양이 사진을 넣고, 개와 고양이를 CNN모델로 분류하기 위해 개와 고양이 이미지 데이터를 CNN 모델로 학습한다. | 4 |
| AI모델개발(이미지 관련 프로젝트 수행 역량) | 개와 고양이 사진을 넣고, 개와 고양이를 CNN모델로 분류하기 위해 개와 고양이 이미지를 CNN 모델로 분류한다. | 4 |
| AI모델개발(이미지 관련 프로젝트 수행 역량) | 개와 고양이 사진을 넣고, 개와 고양이를 CNN모델로 분류하기 위해 분류의 정확도를 측정한다. | 4 |

**결과물**

* CNN 발표 동영상 & 결과 발표 동영상
* accuracy와 loss에 대한 그래프와 evaluate 결과에 대한 스크린 샷
* 코드

**평가 기준**

|  |  |
| --- | --- |
| 멘토 평가 항목 | 구분 |
| 1. 여러 개의 Conv Layer를 시도해 봤는가? (학생 코드 관찰 필요) | Data Analysis |
| 2. 발표를 잘할 수록 높은 점수 부여 | Data Analysis |
| 3. 스크린샷에 accuracy 그래프, model loss 그래프, eavluste 결과가 포함되었는가? (각각 점수 부여) | Estimation |
| 4. test 데이터로 모델을 평가하여 높은 정확도 순으로 높은 점수 부여 | Estimation |
| 5. 결과 발표를 잘한 순으로 높은 점수 부여 | Estimation |
| 6. 2개 이상의 모델을 만들어서 비교하여 좋은 모델을 선택한 경우 높은 추가 점수 부여. 많은 모델을 비교할 수록 높은 점수 부여 (코드 관찰이 필요 - 멘토 )  추가 평가 | Estimation |

**확인 질문**

|  |  |  |
| --- | --- | --- |
| 분류 | 문제 | 정답 |
| Estimation | accuracy 그래프에서 train과 test의 그래프가 차이가 커지는 것은 무엇을 의미하는가? | 오버피팅이 증가함을 의미한다. |

# 모범 답안

<https://github.com/edu10make/Project-X_/blob/main/Project1-Medical%20Analysis/002_Head%20CT-Hemorrhage/002_Head%20CT-Hemorrhage.ipynb>

# 데이터 파일 설명

## 이미지 파일

* 정상 뇌 사진 100장
* 뇌출혈 뇌 사진 100장

## labels.csv

뇌출혈 여부 데이터

# 학습 지원 정보

## 학습 리소스

* <https://en.wikipedia.org/wiki/Convolution_neural_network>  
  : Convolution Neural Network(합성곱 신경망)에 대한 일반적인 이해 (위키피디아)
* <https://en.wikipedia.org/wiki/Scikit-learn>  
  : scikit-learn에 대한 일반적인 이해 (위키피디아)
* <https://en.wikipedia.org/wiki/CT_scan>  
  : CT scan에 대한 일반적인 이해 (위키피디아)
* [Tensorflow KR 페이스북 그룹](<https://www.facebook.com/groups/TensorFlowKR>) : tensorflow를 학습하거나, tensorflow, 다른 Machine Learning 도구에 관한 연구, 응용 제작을 하는 분들의 모임. 질의 응답이 매우 활발함
* •[Machine Learning Yearning](<https://www.deeplearning.ai/machine-learning-yearning/>) : 앤드유 응교수의 ML 프로젝트를 진행할 때 고려해야 할 우선순위, 성능에 대한 고려 사항 (딥러닝 기술 자체를 다룬 책은 아니라 필독). PDF 버전을 인터넷에서 쉽게 구할 수 있음
* •[모두를 위한 딥러닝]([https://www.inflearn.com/course/기본적인-머신러닝-딥러닝-강좌#](https://www.inflearn.com/course/기본적인-머신러닝-딥러닝-강좌%23)) : 머신 러닝과 딥러닝에 대해 더 이해하고 본인들의 문제를 Tensorflow를 이용하여 풀 수 있게 도와주는 동영상 강의. 이 강좌는 수학이나 컴퓨터 공학적인 지식이 없어도 어렵지 않게 볼 수 있음.
* <https://www.youtube.com/playlist?list=PLS8gIc2q83OjStGjdTF2LZtc0vefCAbnX&ab_channel=NeoWizard>  
  : 머신 러닝, 딥러닝
* [https://www.youtube.com/watch?v=AHdGc4nR5f4&list=PLS8gIc2q83OhM0RTktKDitgZGX5dHo7Vs&ab\_channel=NeoWizard](https://www.youtube.com/watch?v=AHdGc4nR5f4&list=PLS8gIc2q83OhM0RTktKDitgZGX5dHo7Vs&ab_channel=NeoWizard%20)   
  : TensorFlow

# 포럼

* 학습 리소스: <https://github.com/edu10make/Project-X_>
* 토론과 Q&A: https://prj.innoaca.kr/forum/project/2/forum
* 코드리뷰: https://prj.innoaca.kr/forum/project/2/codereview
* 동료평가: https://prj.innoaca.kr/forum/project/2/evaluation
* 멘토: https://prj.innoaca.kr/forum/project/2/mentor
* 동료목록: https://prj.innoaca.kr/forum/project/2/peer