

## 2023 년 K-IUM 경진대회 데이터셋 및 최종 결과 평가 방법 설명서

안녕하십니까. 본 대회에 참여해주신 모든 참가자분들께 감사 말씀 먼저 드립니다.

아래의 데이터셋 설명 및 최종 결과 평가 방법을 잘 숙지하시어 최종 평가 시 불이익이 없도록 해주십시오. 특히 최종 평가 출력 양식을 어기시는 경우 심사 과정에서 제외됩니다. 결과 출력 양식과 대회 규정을 어기는 경우에는 부득이하게 시상 대상에서 탈락됩니다. 공정한 경쟁을 위해서 이 점을 다시 한번 강조합니다.

### I. 데이터셋 설명서

*아래의 데이터셋은 부산대학교 의료인공지능 사업단 심예지 박사가 작성하였고 부산대학교병원 영상 의학과 황보리 교수의 검수(익명화 확인 및 데이터셋 확인)를 거쳐서 완성되었음.*

#### 1. 데이터셋 개괄적 설명

(1) 목적: 디지털 감산 혈관조영술 표준 영상에서 특정 환자의 뇌동맥류를 발견하는 인공지능 알고리즘을 개발

(2) 개발 모형의 세부 목표

**세부 목표 1:** 특정 환자(Index[아래 참조]의 수 1 개 당 환자 1 명)의 뇌동맥류 존재 여부를 제시

**세부 목표 2:** 뇌동맥류의 위치 정보를 제시 ([test\_set] 디렉토리 내의 샘플 csv 파일 형식으로 제출)

(3) 비식별화 정도: **완전 익명화** (다른 외부의 정보와 결합하더라도 개인의 재식별화 가능성이 없음)

#### 2. 데이터셋의 구성

(1) 디렉토리 [train\_set] → 훈련에 사용하실 수 있는 전체 데이터셋을 포함함.

(2) 파일 설명

● train.csv

■ 변수 Index = 파일번호

- 변수 Aneurysm = 뇌동맥류 여부 (1=TRUE 0=FALSE)
- 이외 변수들 = 뇌동맥류 위치 정보 (1=Presence 0=Absence)
  - ◆ L\_ICA = left internal carotid artery
  - ◆ R\_ICA = right internal carotid artery
  - ◆ L\_PCOM = left posterior communicating artery
  - ◆ R\_PCOM = right posterior communicating artery
  - ◆ L\_AntChor = left anterior choroidal artery
  - ◆ R\_AntChor = right anterior choroidal artery
  - ◆ L\_ACA = left anterior cerebral artery
  - ◆ R\_ACA = right anterior cerebral artery
  - ◆ L\_ACOM = left-sided anterior communicating artery
  - ◆ R\_ACOM = right-sided anterior communicating artery
  - ◆ L\_MCA = left middle cerebral artery
  - ◆ R\_MCA = right middle cerebral artery
  - ◆ L\_VA = left vertebral artery
  - ◆ R\_VA = right vertebral artery
  - ◆ L\_PICA = left posterior inferior cerebellar artery
  - ◆ R\_PICA = right posterior inferior cerebellar artery
  - ◆ L\_SCA = left superior cerebellar artery
  - ◆ R\_SCA = right superior cerebellar artery
  - ◆ BA = basilar artery
  - ◆ L\_PCA = left posterior cerebral artery
  - ◆ R\_PCA = right posterior cerebral artery
  - ◆ 참조 사항: 뇌동맥류가 특정 위치에 2 개 이상 있더라도 위치 정보에는 1 로 표기되어 있음
- [Index]LI-A.jpg 와 [Index]LI-B.jpg
  - Angiography with left internal carotid artery injection
  - 왼쪽 내경동맥에서 조영제를 주입하고 얻은 영상의 A 와 B (서로 다른 각도 촬영)
- [Index]RI-A.jpg 와 [Index]RI-B.jpg
  - 상기 LI-A, LI-B 와 동일하나 오른쪽 내경동맥에서의 검사임
- [Index]LV-A.jpg 와 [Index]LV-B.jpg
  - Angiography with left vertebral artery injection
  - 왼쪽 추골동맥에서 조영제를 주입하고 얻은 영상의 A 와 B
- [Index]RV-A.jpg 와 [Index]RV-B.jpg
  - 마찬가지로 LV-A, LV-B 와 동일하나 오른쪽 추골동맥에서의 검사임
- 위와 같이 각 환자의 검사 당 총 8 개의 영상이 제시되며, 이에 대한 정보는 train.csv 파일을 참조하여 훈련할 수 있게 되어 있음.

- 조영제를 뇌혈관에서 직접 주입하는 검사에서는 특히 질병이 있는 경우 통상적인 해부학적 구조와 달라진 경우가 있음. 예를들어 왼쪽 내경동맥을 주입하더라도 오른쪽 내경동맥 영역도 함께 나타날 수 있음.
- 이에 대해서는 본 설명서의 마지막 섹션에 있는 참조 사이트들을 방문하여 기본적인 이해를 가지고 개발을 진행하기를 권고함.

## II. 최종 결과 평가 방법 설명서

1. 최종 결과 평가는 사전 공지된 바와 같이 아래와 같은 요건으로 결과를 제시할 수 있는 프로그램에 대해서 평가하게 됨

참가자의 프로그램은 csv 폴더 내부의 파일을 읽어서 같은 디렉토리 내에 위치한 해당 번호의 영상 8 장을 찾아서 확인한 후 아래의 결과값을 규정된 csv 형태로 출력함.

(1) 뇌동맥류 여부를 0 에서 1 사이의 값으로 반환함 (실수값 반환 필수. 지키지 않을 경우 심사에서 제외함)

(2) 뇌동맥류의 위치 정보를 해당 뇌혈관 위치 칼럼에 0(미존재) 또는 1(존재)로 제시 (정수값 반환 필수. 지키지 않을 경우 심사에서 제외함)

2. 평가 지표 (동점자 처리 기준)

(1) C-statistics (AUROC): 위의 II-1.-(1)에서 제출된 실수(real number) 값들을 바탕으로 AUROC 를 계산하여 소수점 이하 4 자리에서 반올림하여 가장 높은 순서로 등수를 확인함.

(2) 위에서 동점자가 있을 경우 II-1.(2)에서 제출된 정수(integer) 값들을 바탕으로 전체 정확도\*를 계산하고 소수점 이하 4 자리에서 반올림하여 평가함.

$$(*) \frac{(true\ positive)+(true\ negative)}{(true\ positive)+(false\ positive)+(true\ negative)+(false\ negative)}$$

(3) 상기 (1), (2)에서도 동점자가 있을 경우 심사위원단이 개발된 모형과 함께 제출된 리포트를 정성적 평가하여 시상 대상자를 결정함.

3. 최종 평가가 이루어질 디렉토리 및 파일 설명

★ 현재 제시된 디렉토리 내부의 파일들은 예시이며, 실제 검증 시 사용하는 데이터셋은 추후 공개될 예정임. 예시 파일의 형식에 맞추어서 csv 파일 형식으로 출력하여야함.

(1) 디렉토리 [train\_set] → 훈련에 사용하실 수 있는 전체 데이터셋을 포함함.

(2) 파일 설명

- test.csv
  - 변수 Index = 파일번호
  - 변수 Aneurysm = 뇌동맥류 여부는 모두 0 으로 코딩됨.
  - 이외 변수들 = 뇌동맥류 위치 정보도 모두 0 으로 코딩됨.
    - ◆ L\_ICA = left internal carotid artery
    - ◆ R\_ICA = right internal carotid artery
    - ◆ L\_PCOM = left posterior communicating artery
    - ◆ R\_PCOM = right posterior communicating artery
    - ◆ L\_AntChor = left anterior choroidal artery
    - ◆ R\_AntChor = right anterior choroidal artery
    - ◆ L\_ACA = left anterior cerebral artery
    - ◆ R\_ACA = right anterior cerebral artery
    - ◆ L\_ACOM = left-sided anterior communicating artery
    - ◆ R\_ACOM = right-sided anterior communicating artery
    - ◆ L\_MCA = left middle cerebral artery
    - ◆ R\_MCA = right middle cerebral artery
    - ◆ L\_VA = left vertebral artery
    - ◆ R\_VA = right vertebral artery
    - ◆ L\_PICA = left posterior inferior cerebellar artery
    - ◆ R\_PICA = right posterior inferior cerebellar artery
    - ◆ L\_SCA = left superior cerebellar artery
    - ◆ R\_SCA = right superior cerebellar artery
    - ◆ BA = basilar artery
    - ◆ L\_PCA = left posterior cerebral artery
    - ◆ R\_PCA = right posterior cerebral artery
- [Index]LI-A.jpg 와 [Index]LI-B.jpg
  - Angiography with left internal carotid artery injection
  - 왼쪽 내경동맥에서 조영제를 주입하고 얻은 영상의 A 와 B (서로 다른 각도 촬영)
- [Index]RI-A.jpg 와 [Index]RI-B.jpg
  - 상기 LI-A, LI-B 와 동일하나 오른쪽 내경동맥에서의 검사임
- [Index]LV-A.jpg 와 [Index]LV-B.jpg
  - Angiography with left vertebral artery injection
  - 왼쪽 추골동맥에서 조영제를 주입하고 얻은 영상의 A 와 B
- [Index]RV-A.jpg 와 [Index]RV-B.jpg
  - 마찬가지로 LV-A, LV-B 와 동일하나 오른쪽 추골동맥에서의 검사임

- 훈련용 데이터와 마찬가지로 위와 같이 각 환자의 검사 당 총 8 개의 영상이 제시됨.
- `output.csv` // 참가팀 프로그램이 출력해야하는 파일
  - 이 파일에는 상기 `test.csv` 에서 0 으로 코딩되었던 변수들에 참가팀의 인공지능 모형이 제시한 값들이 포함되어 출력되어야 함.
  - **변수 Aneurysm** 의 경우 뇌동맥류 여부에 대한 모형의 결과 값을 **0 과 1 사이의 실수로 출력**하여야 함.
  - **위치 변수들의 경우 0 또는 1 의 정수로 출력**하여야함.
  - 위의 사항들을 위반하는 경우 심사 대상에서 제외함.
- `evaluate.py` 파일을 통해서 참가자가 직접 모형의 AUROC 및 전체 정확도에 대해서 미리 확인할 수 있음.
  - `groundtruth.csv`: `evaluate.py` 스크립트가 사용하는 파일임.

### III. 제출 요구 사항

1. 최종 평가를 위한 제출 사항들은 기 공지된 바와 같이 프로그램 결과물 및 레포트 (기술 문서 포함)임.
  - (1) 프로그램 결과물 (개발된 모델) 압축파일로 업로드
  - (2) 결과물 검증을 위한 레포트 PDF 형식으로 업로드
2. 특히 개발된 모델을 구동하여 위에 설명과 같이 제시된 `output.csv` 파일을 출력할 수 있는 방법에 대한 자세한 명시가 있어야 하며 이를 누락하는 경우 불이익이 발생할 수 있음.

### IV. 데이터셋의 이해를 위한 참조 웹페이지

1. 뇌동맥류에 대한 이해  
<https://www.ninds.nih.gov/health-information/disorders/cerebral-aneurysms>
2. 뇌혈관 해부학에서 중요한 우회로와 관련된 개념의 설명  
<https://www.ncbi.nlm.nih.gov/books/NBK534861/>
3. 혈관조영술은 어떤 검사인가에 대한 설명

<https://www.ncbi.nlm.nih.gov/books/NBK557477/>

위의 설명들을 바탕으로 키워드 서치를 하시면 대략적인 도메인 지식에 대한 이해가 될 것으로 생각합니다.

다소 딱딱하게 불이익이 생길 수 있는 사항들을 설명드려서 안타깝습니다만 공정한 경쟁을 위해서 그리고 우수한 모델을 개발한 뒤에 제시된 요건을 누락하여 억울하게 시상에서 탈락하는 팀을 예방하기 위해서 강조해서 설명드렸습니다.

위의 사항 이외에 궁금하신 점이나 의문 사항 있으실 경우 게시판을 통해서 문의주시면 최대한 빠른 시간 내에 답을 드리겠습니다.