2. 일반혈액검사 데이터 분양:

현재까지 진행 이력

* 최종데이터셋 확인 후, 심의 후 연구 심사통보서 대기 중(금주 혹은 차주 진행 승인 예정).

분양 결과

* • 규모: 총 환자 158,808명, 암 발생 204,377건(환자별 중복 암 발생 건은 데이터 수령 후 식별 예정)
* • 암종별 건수

|  |  |  |
| --- | --- | --- |
| Code | Cancer Type | Case Number |
| C16 | 위암 | 37,503 |
| C18, C19, C20 | 대장암 | 17,290 |
| C22 | 간암 | 21,217 |
| C23, C24 | 담낭 및 기타담도 | 4,787 |
| C25 | 췌장암 | 7,376 |
| C33, C34 | 폐암 | 41,231 |
| C50 | 유방암 | 30,548 |
| C56 | 난소암 | 9,825 |
| C61 | 전립선암 | 9,788 |
| C64 | 신장암 | 4,474 |
| C73 | 갑상선암 | 20,338 |

* 일반 혈액검사와 암 확진일 간의 전후 관계는 해당 데이터 반출 시 확인하겠습니다.

3. 일반혈액검사 데이터 전처리:

현재까지 진행 이력

* 최종데이터셋 확인 후, 심의 후 연구 심사통보서 대기 중.

전처리의 목적

- 데이터 신뢰성 평가 및 모델 성능 최적화

- Feature Selection과 Feature Engineering을 통한 차원 축소 및 파생변수 생성

전처리가 끝났을 때의 산출물

- 최종 분석 데이터 셋

-변수 중요도 평가 결과

4. 암 문진 데이터 분양:

분양 계획:

* CBC 코드와 함께 분양 예정

일정:

- 어떤 일정인지 의미를 모르겠네요.

문진 데이터 항목 수

* 암 관련 문진 데이터 항목 수: 39
* 일반 문진 데이터 항목 수: 19
* 항목 내역:
  + (환자)기본정보: 성별, 생년월일, 혈액형(ABO), 혈액형(Rh) (4)
  + (환자)사망정보: 원내사망일자, 사망장소코드, (가)직접사인, (가)의 원인, (나)의 원인, (다)의 원인, 사망종류코드 (7)
  + (내원)외래: 진료일자, 내원시나이, 초재진구분코드 (3)
  + (진단)진단등록: 진단일, 진단 시 나이, 진단코드, 주상병여부 (4)
  + (진단)암 등록: 암등록초진일, 부위코드(ICD-0-3), 형태코드, 진단코드, 치료형태: 수술여부, 치료형태: 화학여부, 치료형태: 방사선여부, 치료형태: 면역여부, 치료형태: 호르몬여부, 치료형태: 기타, 분화도(≤4개월): Grade, 진단병기(≤4개월): T /N /M /stage /Version (16)
  + (검사) 검체 핵체외포함: 검사코드 (5)
  + (간호기록) 병동 초기평가 성인: 흡연력 유무, 현재흡연유무, 흡연시작연령, 흡연량(갑/일), 흡연기간(년), 음주유무, 음주시작연령, 음주종류코드, 음주량(병/회), 음주횟수(회/월), 음주기간(년), BMI, 과거병력유무 (15)
* 분양이 끝났을 때의 산출물:

최종 활용 가능한 문진 데이터 항목 Screening 결과, 데이터 정의서(encoding 정보 포함),

변수 간의 상호작용(Interaction)이나 복합적인 의미를 갖는 고위험 특성 생성