

인공지능 데이터 구축 가이드라인

- 데이터셋 명칭 -

| 담당 역할 | 기관명 |
|----------------|---|
| 데이터 구축 총괄 | 서울특별시 보라매병원 |
| 데이터 설계 | 서울특별시 보라매병원 |
| 데이터수집 및 정제 | 서울특별시 보라매병원, 경희대학교 산학협력단, (주)원스글로벌, 부광아이티(주), (주)모비소프트 |
| 가공(라벨링, 어노테이션) | 인피노브(주), (주)더유니파이 |
| 데이터 검수(자체 검수) | 서울특별시 보라매병원, 경희대학교 산학협력단, (주)원스글로벌, 인피노브(주) |

| | | |
|--------------------|------------------------|-------|
| 구축 가이드라인 작성 | 서울특별시 보라매병원 | 이 시 욱 |
| | 서울특별시 보라매병원 | 김 정 현 |
| 가이드라인 버전 (제작일자) | Ver 1.0 (2021. 12. 27) | |

목 차

| | |
|---------------------------|----------|
| 1. 데이터 구축 개요 | 0 |
| 2. 문제 정의 | 0 |
| 2.1 임무 정의 | 0 |
| 2.2 데이터 구축 유의사항 | 0 |
| 3. 데이터 수집·정제 | 0 |
| 3.1 원시데이터 선정 | 0 |
| 3.2 수집·정제 절차 및 방법 | 0 |
| 3.3 수집·정제 기준 | 0 |
| 3.4 수집·정제 조직 | 0 |
| 3.5 수집·정제 도구 | 0 |
| 4. 데이터 가공 | 0 |
| 4.1 가공 절차 및 방법 | 0 |
| 4.2 가공 기준 | 0 |
| 4.3 가공 규격 | 0 |
| 4.4 가공 조직 | 0 |
| 4.5 가공 도구 | 0 |
| 5. 검수 | 0 |
| 5.1 검수 절차 및 방법 | 0 |
| 5.2 검수 기준 | 0 |
| 5.3 검수 조직 | 0 |
| 5.4 검수 도구 | 0 |
| 5.5 기타 품질관리 활동 | 0 |

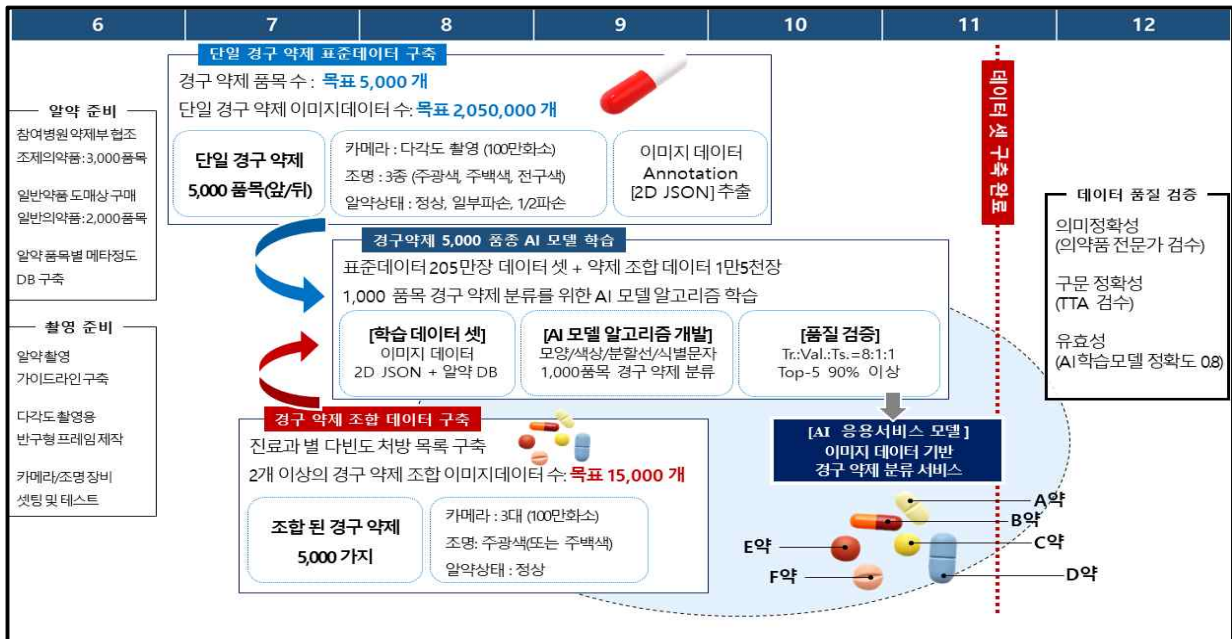
1. 데이터 구축 개요

- 데이터 구축 필요성
 - 경구 약제 이미지 데이터 분석을 통한 경구 약제 식별 시스템은 의약 전문가의 신속한 의약품 통한 응급조치를 가능하게 하며, 소비자관점에서는 복용하고 있는 의약품에 대한 정확한 정보를 전달하여 알약의 식별 불가능으로 발생할 수 있는 약물 오·남용을 예방할 수 있음
 - 이에 경구약제와 같은 의약품 식별 결과를 안내하고 피드백할 수 있는 AI 서비스 모델이 요구되며, 이를 위해 의약품 이미지 자료의 데이터베이스 구축이 필요함
 - 이미지 데이터를 기반으로 낱알식별을 위한 상세정보(크기, 무게, 색상 등) 데이터를 함께 구축하고 함량별 품목기준코드에 따라 의약품에 따른 고유값을 부여하여 약품 식별에 정확도를 높일 필요가 있음
- 데이터 구축 활용 방안
 - 구축된 이미지 데이터를 통해 약품 식별을 위한 AI 학습 모델을 개발
 - 약품식별 AI 서비스 모델은 낱알 상태로 구분하기 힘든 알약을 식별하는데 도움을 줄 수 있기 때문에 정확한 약물 처방이 필요한 의사, 약사 등의 의료전문가와 약품 소비자의 알권리 증진에 도움이 될 것으로 기대됨
 - 다양한 형태의 약물 이미지 데이터를 구축하여 이미지 식별 알고리즘에 관해 관심이 있는 의료진과 연구진들의 AI 알고리즘 개발을 촉진시킬 것으로 기대됨
 - 데이터 활용 방안 예시



2. 문제 정의

2.1. 임무 정의



| | 항목 | 내용 |
|-----|-----------|--|
| 무엇 | 대상물 | <ul style="list-style-type: none"> 경구 약제 품목 <ul style="list-style-type: none"> 총 5,000품목(조제의약품: 60%, 일반의약품: 40%, (±5%)) 단일 경구 약제 이미지데이터: 2,050,000개 경구 약제 조합 이미지데이터: 15,000개 |
| | 데이터 포함 변수 | <ul style="list-style-type: none"> 경구 약제 메타정보 <ul style="list-style-type: none"> 모양, 색상, 제형, 분할선, 식별문자, 크기 등의 원스글로벌 보유 알약 기존 정보와 식별정보 약제 촬영환경 <ul style="list-style-type: none"> 알약의 상태, 방향, 조명, 카메라 위치 |
| 언제 | 획득 기간 | <ul style="list-style-type: none"> 사업 수행 기간내 완료(21년 8월부터 11월 중순 예정) |
| 어디서 | 획득 장소 | <ul style="list-style-type: none"> 조제의약품 <ul style="list-style-type: none"> 서울시 보라매병원, 강동경희대병원 약제부에서 확보 타 의료기관, 지역약국에서 거래처(도매상)로 낱알 반품된 약품을 구매 일반의약품 <ul style="list-style-type: none"> 국내 일반의약품 전문 도매상으로부터 품목리스트를 확보한 후 다빈도 일반 의약품 중심으로 구매 데이터 획득: 보라매 병원에서 별도로 구축한 경구 약제 촬영용 스튜디오에 다각도 촬영, 3종의 조명 환경 기타 촬영 장비를 설치할 수 있는 반구형 프레임 설치 |
| 누가 | 획득 | <ul style="list-style-type: none"> 서울보라매병원, 강동경희대병원, 원스글로벌 |
| | 정제 | <ul style="list-style-type: none"> 서울보라매병원, 인피노브, 부광아이티 |

| | | |
|-----|--------|--|
| | 가공 | <ul style="list-style-type: none"> 서울보라매병원, 인피노브 |
| | 검수 | <ul style="list-style-type: none"> 서울보라매병원, 인피노브, 원스글로벌 |
| 어떻게 | 획득 방법 | <ul style="list-style-type: none"> 경구 약제 5,000 품목 획득 방안 <ul style="list-style-type: none"> 조제의약품: 보라매병원, 강동경희대병원 보유 품목과 타 의료기관 및 지역 약국별 거래처로 낱알 반품된 약품을 구매하여 촬영 일반의약품: 국내 일반의약품 전문 도매상으로부터 품목리스트를 확보한 후 다빈도 일반의약품 중심으로 구매하여 촬영 경구 약제 촬영 조건 및 환경 <ul style="list-style-type: none"> 100만 화소 이상 카메라 다각도 방향의 카메라 셋팅 가능한 스튜디오 공간 단일 경구 약제 3가지 색상의 배경 적용(1,000종에 적용) 단일 경구 약제 3종의 조명(주광색, 주백색, 전구색) 활용하여 다양한 색감 연출 단일 경구 약제 1개당 앞면과 뒷면 촬영 단일 경구 약제의 상태에 따라 정상, 일부파손, 1/2파손으로 구분하여 촬영 경구 약제 조합은 3개 각도에서 1종의 조명과 1종의 배경에서 촬영 |
| | 샘플 크기 | <ul style="list-style-type: none"> 경구 약제 5,000 품목 <ul style="list-style-type: none"> 단일 경구 약제 이미지데이터: 2,050,000개 경구 약제 조합 이미지데이터: 15,000개 |
| | 데이터 양식 | <ul style="list-style-type: none"> 이미지 데이터: PNG 라벨링 데이터: JSON 약품정보 데이터: JSON |
| 왜 | 구축 목적 | <ul style="list-style-type: none"> 경구 약제와 같은 의약품 식별 결과를 안내하고 피드백할 수 있는 AI 서비스 모델이 요구되며, 이를 위해 의약품 이미지 자료의 데이터베이스 구축 이미지 데이터를 기반으로 낱알식별을 위한 상세정보(크기, 무게, 색상 등) 데이터를 함께 구축하고 함량별 품목기준코드에 따라 의약품에 따른 고유값을 부여하여 약품 식별에 정확도를 높임 |
| | 기대 효과 | <ul style="list-style-type: none"> 약품식별 AI 서비스 모델은 낱알 상태로 구분하기 힘든 알약을 식별하는데 도움을 줄 수 있기 때문에 정확한 약물 처방이 필요한 의사, 약사 등의 의료 전문가와 약품 소비자의 알권리 증진에 도움이 될 것으로 기대됨 다양한 형태의 약물 이미지 데이터를 구축하여 이미지 식별 알고리즘에 관해 관심이 있는 의료진과 연구진들의 AI 알고리즘 개발을 촉진시킬 것으로 기대됨 |

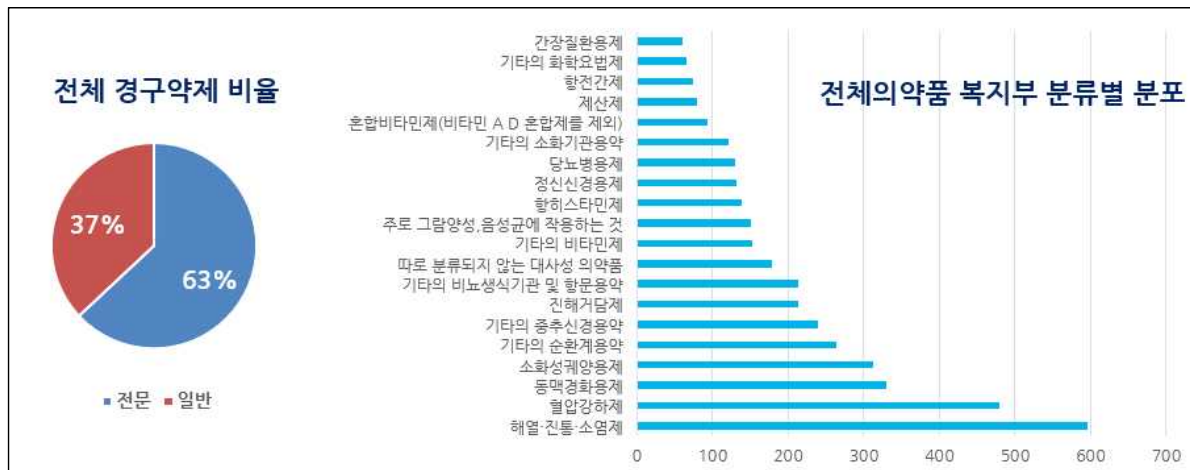
2.2. 데이터 구축 유의사항

- 구축 데이터 배포 위험성: 해당사항 없음
- 법적인 문제: 해당사항 없음
- 개인정보보호: 해당사항 없음

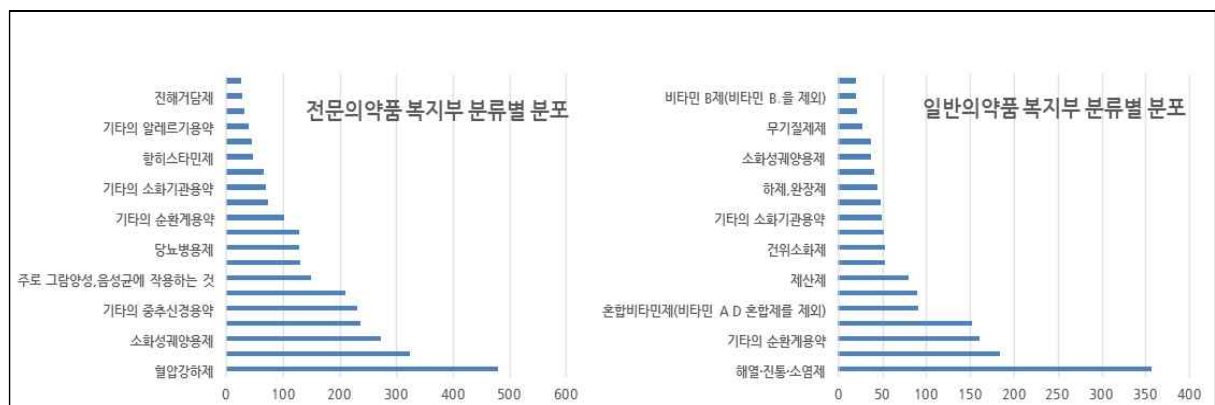
3. 수집·정제

3.1. 원시데이터 선정

- 경구 약제 5000종 리스트 선정
 - 국내 시판중인 경구 약제 중 복지부 분류(효능)별 다빈도 의약품 5,000종을 선정함
 - 경구 약제 5,000종은 전문(조제) 의약품 3,143종(62.8%), 일반 의약품 1,857종(37.2%)의 클래스로 구성하였으며, 계획했던 6:4(±5%)의 다양성을 확보하였음



- 리스트 선정 기준 및 결과
 - 전문/일반의약품 기준: 보건복지부는 2000년 의약분업에 적용할 27,962품목의 의약품을 전문 의약품 17,187품목(61.5%), 일반의약품 10,775품목(38.5%)으로 분류. 전문의약품과 일반의약품의 품목 비율(6:4)을 고려하여 리스트 선정
 - 보건복지부 분류 기준: 보건복지부의 의약품 분류별 약품 개수 확인(2021년 7월 기준). 다빈도 등록 약품을 우선순위로 확보할 약품 개수 선정
 - 결과: 전문의약품 3,143(62.8%)종, 일반의약품 1,857(37.2%)종을 확보하였으며, 보건복지부 의약품 분류의 다빈도 등록 비율을 만족함



- 약품 확보 방법
 - 선정된 리스트를 강동경희대학교 병원과 보라매 병원 약제부에 공유
 - 각 병원에서 관리하는 폐기약을 우선적으로 확보하였으며, 병원의 주거래 원외 약국에서 약품 구매 함
 - 약품 검수 및 라벨링 작업을 진행 후 촬영 스튜디오로 배송

3.2. 수집·정제 절차 및 방법

| | 항목 | 내용 |
|-----|-----------|---|
| 무엇 | 대상물 | <ul style="list-style-type: none"> • 경구 약제 품목 <ul style="list-style-type: none"> - 총 5,000품목(조제의약품: 60%, 일반의약품: 40%, ($\pm 5\%$)) - 단일 경구 약제 이미지데이터: 2,050,000개 - 경구 약제 조합 이미지데이터: 15,000개 |
| | 데이터 포함 변수 | <ul style="list-style-type: none"> • 경구 약제 메타정보 <ul style="list-style-type: none"> - 모양, 색상, 제형, 분할선, 식별문자, 크기 등의 원스글로벌 보유 알약 기존정보와 식별정보 • 약제 촬영환경 <ul style="list-style-type: none"> - 알약의 상태, 방향, 조명, 카메라 위치 |
| 언제 | 획득 기간 | <ul style="list-style-type: none"> • 사업 수행 기간내 완료(21년 8월부터 11월 중순 예정) |
| 어디서 | 획득 장소 | <ul style="list-style-type: none"> • 조제의약품 <ul style="list-style-type: none"> - 서울시 보라매병원, 강동경희대병원 약제부에서 확보 - 타 의료기관, 지역약국에서 거래처(도매상)로 낱알 반품된 약품을 구매 • 일반의약품 <ul style="list-style-type: none"> - 국내 일반의약품 전문 도매상으로부터 품목리스트를 확보한 후 다빈도 일반의약품 중심으로 구매 • 데이터 획득: 보라매 병원에서 별도로 구축한 경구 약제 촬영용 스튜디오에 다각도 촬영, 3종의 조명 환경 기타 촬영 장비를 설치할 수 있는 반구형 프레임 설치 |
| 누가 | 획득 | <ul style="list-style-type: none"> • 서울보라매병원, 강동경희대병원, 원스글로벌 |
| | 정제 | <ul style="list-style-type: none"> • 서울보라매병원, 인피노브, 부광아이티 |
| | 가공 | <ul style="list-style-type: none"> • 서울보라매병원, 인피노브 |
| | 검수 | <ul style="list-style-type: none"> • 서울보라매병원, 인피노브, 원스글로벌 |
| 어떻게 | 획득 방법 | <ul style="list-style-type: none"> • 경구 약제 5,000 품목 획득 방안 <ul style="list-style-type: none"> - 조제의약품: 보라매병원, 강동경희대병원 보유 품목과 타 의료기관 및 지역 약국별 거래처로 낱알 반품된 약품을 구매하여 촬영 - 일반의약품: 국내 일반의약품 전문 도매상으로부터 품목리스트를 확보한 후 다빈도 일반의약품 중심으로 구매하여 촬영 • 경구 약제 촬영 조건 및 환경 |

| | | |
|--|--------|--|
| | | <ul style="list-style-type: none"> - 100만 화소 이상 카메라 - 다각도 방향의 카메라 셋팅 가능한 스튜디오 공간 - 단일 경구 약제 3가지 색상의 배경 적용(1,000종에 적용) - 단일 경구 약제 3종의 조명(주광색, 주백색, 전구색) 활용하여 다양한 색감 연출 - 단일 경구 약제 1개당 앞면과 뒷면 촬영 - 단일 경구 약제의 상태에 따라 정상, 일부파손, 1/2파손으로 구분하여 촬영 - 경구 약제 조합은 3개 각도에서 1종의 조명과 1종의 배경에서 촬영 |
| | 샘플 크기 | <ul style="list-style-type: none"> • 경구 약제 5,000 품목 - 단일 경구 약제 이미지데이터: 2,050,000개 - 경구 약제 조합 이미지데이터: 15,000개 |
| | 데이터 양식 | <ul style="list-style-type: none"> • 이미지 데이터: PNG • 라벨링 데이터: JSON • 약품정보 데이터: JSON |

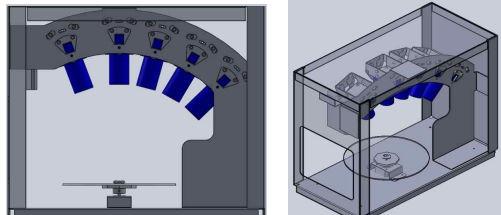

3.3. 수집·정제 기준

- 수집된 경구 약제 이미지데이터를 데이터 수집자가 육안으로 전수 검수하여 마련된 기준에 따라 정제하여 가공 단계로 전달함
- 정제 과정에서 이미지데이터 검사 기준
 - 경구 약제의 모양, 색상, 식별번호, 기호가 명확히 촬영되었는가?
 - 경구 약제에 초점(포커싱)이 잘 맞춰졌는가?
 - 경구 약제 외 불필요한 피사체가 들어있지 않는가?
 - 경구 약제가 조명에 과잉 노출 또는 과부족 되어 촬영되지 않았는가?
 - 경구 약제 캡슐의 경우 약제에 빛 또는 기구물 반사가 나타나지 않았는가?



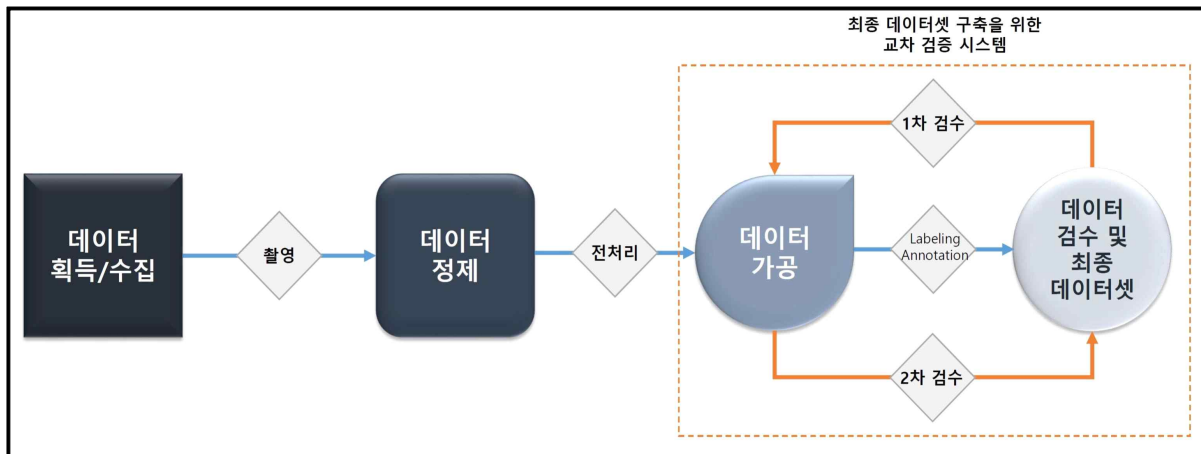
- 정제 과정에서 확인된 오류에 대해서는 모두 재촬영하여 이미지데이터를 다시 수집함

3.4. 수집·정제 도구

| 경구 약제 촬영 환경 (경구 약제 파일럿 테스트 후 박스형 스튜디오로 최종 결정) | | |
|---|--|---|
|  |  |  |
| 돔형 스튜디오 | 박스형 스튜디오 | 캡슐약제용 조명장치 |
| <ul style="list-style-type: none"> • 카메라 : 100만 화소 이상의 카메라로 다각도에서 촬영 • 조명 : 주광색, 주백색, 전구색 이상 3종 • 배경 : 3가지의 서로 다른 배경색 • 17대 카메라 동시 촬영 기반의 돔형 스튜디오와 회전형 테이블 기반의 박스형 스튜디오 구축 및 테스트 후 이미지 결과물과 데이터 획득 프로세스에 용이한 스튜디오 선택 • 약제의 그림자를 최소화하기 위하여 소프트 조명을 사용하거나 간접 조명을 사용하여 그림자를 열게 만들어 줌 • 카메라 위치, 조명, 배경 등 촬영 변수를 충분히 테스트 한 후 AI 모델 학습에 적합한 이미지를 추출할 수 있는 스튜디오를 구축함 • 캡슐약제의 경우 빛반사 저감을 위한 부착형 특수 조명 반사 장치를 활용하여 촬영 | | |

4. 데이터 가공

4.1. 라벨링 절차 및 방법



- 참여기관 인피노브(주)에서 개발한 저작도구를 이용하여 경구 약제 이미지데이터에서 라벨링데이터를 추출함, 검수자의 오류를 줄이고 편리한 작업이 될 수 있도록 개발함
- object detection 학습을 위해 영상에서 Bounding box를 지정하여 라벨링을 수행하고, 약제 메타정보, 촬영정보, bounding box 정보들을 json 형태로 저장
- 라벨링데이터에 대한 검수는 사람에 의해 전수 검수가 진행되며 Bounding box의 정확도는 IoU 0.8 이상, 구문정확성 100%로 진행함

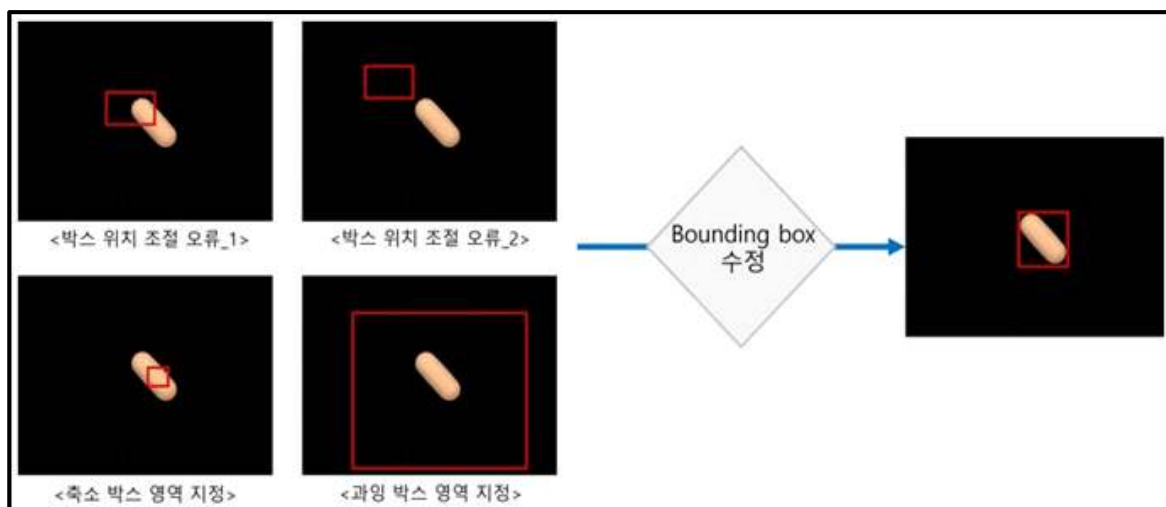


- 라벨링데이터 추출 절차
 - 원천데이터 사진을 전처리하여 얻은 정제 사진 파일을 대상으로 라벨링 작업을 수행

- 데이터 셋에 포함되는 이미지 및 라벨링 텍스트 포맷 정보
- 데이터 저장 구조 및 저장 파일의 명명 규칙 설정
- 데이터 라벨링 작업에는 많은 작업자와 검수자가 필요하므로 이들이 가공/검수 시스템을 사용할 수 있도록 사용자 등록을 하고 툴의 사용법 및 가공/검수 가이드라인 교육을 진행.
- 이하의 라벨링 방법에 따라 라벨링을 진행
- 저작도구로 어노테이션 작업 프로세스
 - 촬영팀에서 받은 원천데이터를 불러온 후 저작도구를 이용하여 경구 약제 코드를 선택하고 자동 어노테이션하여 라벨링데이터를 추출함
 - 어노테이션 결과를 뷰어툴을 이용하여 검수하고, 에러가 확인될 경우 즉시 수정하여 저장

4.2. 라벨링 기준

- Bounding box 검수 및 추출
 - Bounding box 라벨링 시 사진에 촬영된 알약 낱알 전체를 정확히 포함하는 사각형 형태가 되어야 함
 - Bounding box 라벨링 완료 후 해당 분류체계에 따라 명확하게 정의된 라벨을 사용하여 카테고리 지정하며, 카테고리 간 모호성이 없어야 함
 - 모든 라벨링이 완료되면 제출하기 버튼을 클릭하며, 해당 작업물은 '작업완료' 형태로 검수자에게 전달되며 작업자에게는 다음 작업할 이미지가 자동으로 할당됨
- 저작도구 활용한 Bounding box 검수
 - 검수 기준: 클래스 태깅이 정확하고, IoU 0.8 이상
 - 오타깅: 라벨 오류 혹은 IoU 0.8 미만
 - 과태깅: 대상이 아닌 객체에 라벨링
 - 미태깅: 대상 객체 있지만, 라벨링 수행하지 않음



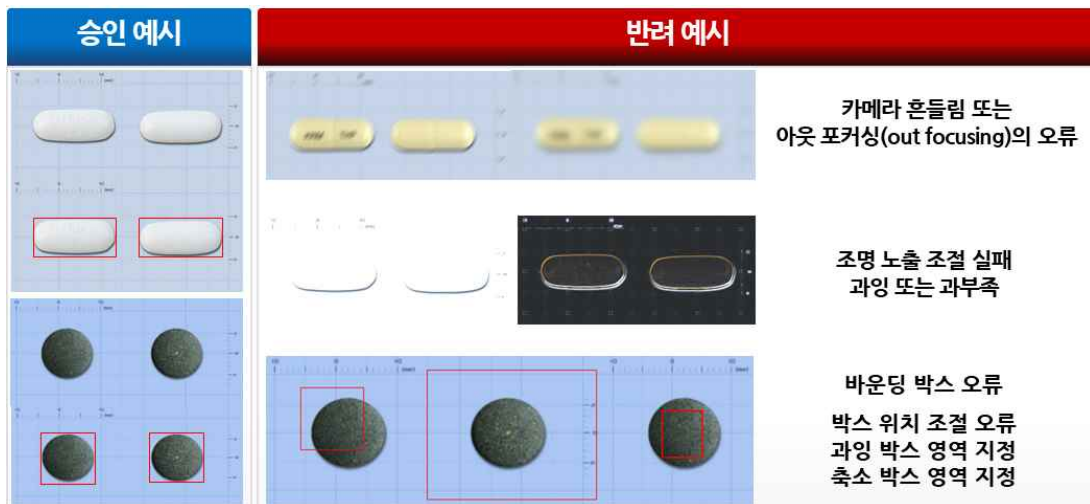
4.3. 라벨링 규칙

| 구분 | 항목명 | 타입 | 필수여부 | 설명 | 범위 |
|------|-----------------------------------|--------|------|------------------|------------|
| 1 | images | Object | M | 약제 이미지정보 | |
| 1-1 | images[].id | Number | M | 약제 이미지식별자 | |
| 1-2 | images[].width | Number | M | 약제 이미지너비 | |
| 1-3 | images[].height | Number | M | 약제 이미지높이 | |
| 1-4 | images[].file_name | String | M | 약제 이미지파일명 | |
| 1-5 | images[].drug_N | String | M | 알약코드 | |
| 1-6 | images[].drug_S | String | M | 알약상태 | |
| 1-7 | images[].back_color | String | M | 촬영배경 | |
| 1-8 | images[].drug_dir | String | M | 알약방향 | |
| 1-9 | images[].light_color | String | M | 촬영조명 | |
| 1-10 | images[].camera_la | Number | M | 카메라위도 | |
| 1-11 | images[].camera_lo | Number | M | 카메라경도 | |
| 1-12 | images[].size | Number | M | 이미지 사이즈 | |
| 1-13 | images[].dl_idx | String | M | 알약 식별자 | |
| 1-14 | images[].dl_mapping_code | String | M | 제품코드 | |
| 1-15 | images[].dl_name | String | M | 제품명 | |
| 1-16 | images[].dl_name_en | String | O | 제품명(영어) | |
| 1-17 | images[].img_key | String | M | 이미지 링크 | |
| 1-18 | images[].dl_material | String | M | 성분명 | |
| 1-19 | images[].dl_material_en | String | O | 성분명(영어) | |
| 1-20 | images[].dl_custom_shape | String | M | 제조 모양 | |
| 1-21 | images[].dl_company | String | M | 제조사명 | |
| 1-22 | images[].dl_company_en | String | O | 제조사명(영어) | |
| 1-23 | images[].di_company_mf | String | M | 위탁제조사명 | |
| 1-24 | images[].di_company_mf_en | String | O | 위탁제조사명(영어) | |
| 1-25 | images[].item_seq | Number | M | 품목기준코드 | |
| 1-26 | images[].di_item_permit_date | Date | O | 허가일자 | YYYY-MM-DD |
| 1-27 | images[].di_class_no | String | M | 약품 분류 | |
| 1-28 | images[].di_etc_otc_code | String | M | 전문의약품/ 일반의약품 | |
| 1-29 | images[].di_edi_code | String | M | EDI 코드 | |
| 1-30 | images[].chart | String | M | 알약 제형 | |
| | images[].drug_shape | String | M | 알약 모양 | |
| | images[].thick | Number | M | 알약 두께 | |
| | images[].leng_long | Number | M | 알약 장축 | |
| | images[].leng_short | Number | M | 알약 단축 | |
| | images[].print_front | String | C | 식별문자_앞 | |
| | images[].print_back | String | C | 식별문자_뒤 | |
| | images[].color_class1 | String | M | 색상 1 | |
| | images[].color_class2 | String | O | 색상 2 | |
| | images[].line_front | String | M | 알약 앞면 분할선 여부 | |
| | images[].line_back | String | M | 알약 뒷면 분할선 여부 | |
| | images[].img_regist_ts | Date | O | 약학정보원 이미지 생성일 | YYYY-MM-DD |
| | images[].form_code_name | String | M | 정제 분류명 | |
| | images[].mark_code_front_an al | String | M | 알약 앞면 마크 형태 | |
| | images[].mark_code_back_an al | String | M | 알약 뒷면 마크 형태 | |
| | images[].mark_code_front_im g | String | M | 알약 앞면 마크 이미지 | |
| | images[].mark_code_back_im g | String | M | 알약 뒷면 마크 이미지 | |

| | | | | | | |
|---|-----|----------------------------|--------|---|----------------|------------|
| | | images[].mark_code_front | String | M | 알약 앞면 마크 코드 | |
| | | images[].mark_code_back | String | M | 알약 뒷면 마크 코드 | |
| | | images[].change_date | Date | O | 변경일자 | YYYY-MM-DD |
| 2 | | type | type | O | json 타입 | |
| 3 | | annotations | Object | M | 라벨링 정보 | |
| | 3-1 | annotations[].area | Number | M | 바운딩박스 면적 | |
| | 3-2 | annotations[].iscrowd | Number | O | 평가 분류 | |
| | 3-3 | annotations[].bbox | List | M | bbox 좌표 | |
| | 3-4 | annotations[].category_id | Number | M | category 아이디 | |
| | 3-5 | annotations[].ignore | Number | O | 무시 여부 | |
| | 3-6 | annotations[].segmentation | List | O | 라벨링 바운딩 박스 | |
| | 3-7 | annotations[].image_id | Number | M | 이미지 아이디 | |
| | 3-8 | annotations[].id | Number | M | 어노테이션 아이디 | |
| 4 | | categories | Object | M | 라이선스 | |
| | 4-1 | categories[].supercategory | String | M | 슈퍼 카테고리 | |
| | 4-2 | categories[].id | Number | M | 카테고리 아이디 | |
| | 4-3 | categories[].name | String | C | 카테고리 명 | |

4.4. 라벨링 도구

- 데이터 라벨링 저작 도구
 - VoTT
 - object detection 학습을 위해 영상에서 Bounding box를 지정하여 라벨링을 수행하고, 그 bounding box 정보들을 json 형태로 저장
 - 사이트 주소 : <https://github.com/microsoft/VoTT>
 - 작업의 효율을 위해 기존 저작 도구를 customizing하거나 새로 개발할 수 있음
- 올바른 라벨링 / 잘못된 라벨링 예시
 - 사진에 촬영된 모든 낱알에 대해 라벨링이 완성되었는가?
 - 라벨링이 불필요한 객체에 라벨링을 하지는 않았는가?
 - 객체의 클래스는 올바르게 지정하였는가?

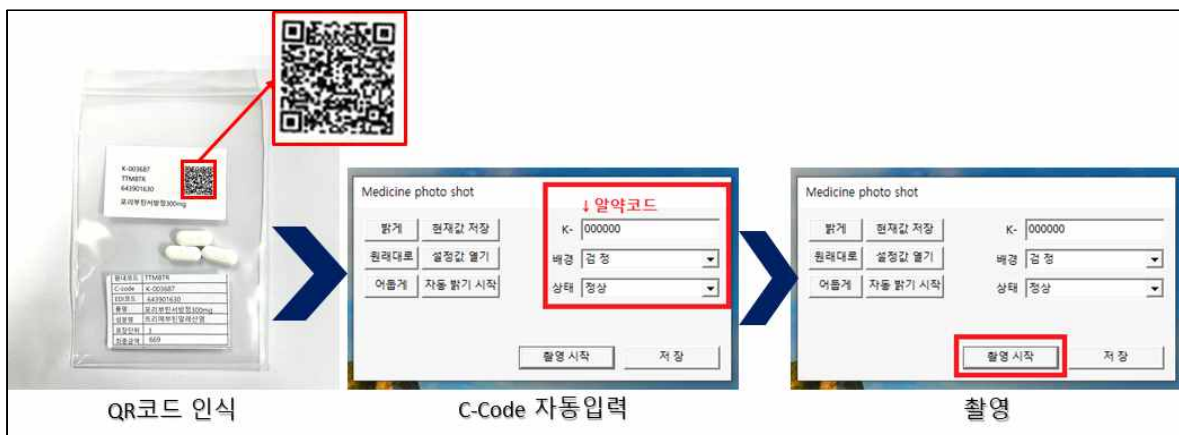


5. 검수

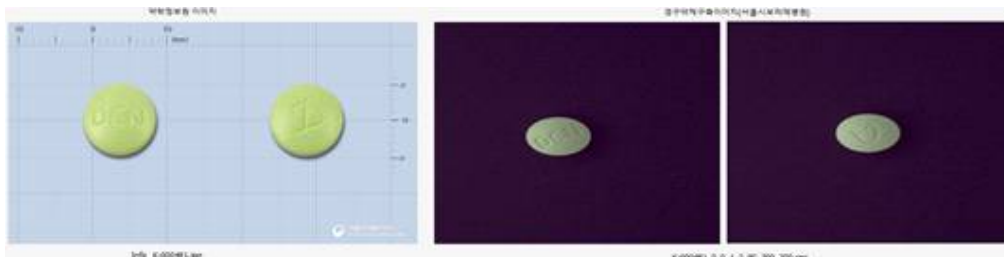
5.1. 검수 절차 및 방법

- 1차 검수: 경구 약제 수집 과정에서 약사에 의해 검수
 - 구매한 경구 약제는 품목, 수량, 제품 상태 등을 1차 검수하여 QR코드를 포함한 라벨지를 생성 후 약사의 실물 확인 후 부착함
 - 보라매병원과 강동경희대병원 약제팀 약사 2인 교차 검수로 진행됨

- 2차 검수: 경구 약제 촬영팀에서 촬영자(수집자)에 의해 검수
 - 촬영할 약품의 QR코드(약제부에서 부착)를 카메라가 인식하여 약품의 고유코드(C-Code)가 자동으로 입력됨
 - 촬영 준비를 마친 후 촬영 시작
 - 촬영자에 의한 약제 관련 데이터 입력 및 개입 최소화(데이터 입력 오류 예방)
 - 촬영한 약품의 이미지 데이터는 규칙화된 파일명 라벨링 되어 컴퓨터에 자동 저장됨
 - 촬영자는 촬영한 이미지 데이터에 대해 전수 검수하여 품질(배경, 포커싱, 흔들림, 경계선)을 확인함
 - 촬영자 검수 후 이상 없을 시 NAS에 업로드하며, 오류 발견 시 재촬영함



- 3차 검수: 외부 약사 4인에 의해 의미정확성에 대한 교차 검수 진행
 - 데이터 구축 및 검증계획서의 목표에 따라 전체 데이터의 10%를 샘플링하여 의미정확성 검증을 수행함
 - 전체 5,000종의 10%인 500종을 샘플링(1개 클래스 당, 전문의약품 300종, 일반의약품 200종)하여 검증 수행함
 - TTA의 권고 사항인 1개 클래스 당 139건(신뢰구간 95% 오차범위±5% 범위, 예상 정확도 90% 상정 기준을 원칙) 이상을 샘플링함
 - 의약품 관련 외부전문가 약사 4인을 섭외하여 전문의약품 300종, 일반의약품 200종에 대해 각각 2인 교차검수 진행함
 - 검사 기준: 원천데이터의 경구약제 이미지와 라벨링데이터의 약제 메타정보(약품명)의 오류 여부를 확인함
 - 검사요소: 경구약제 이미지와 라벨링 정보(메타정보, 약품명)가 일치하는가?
 - 검사 도구: 수집한 원천데이터의 경구약제 이미지와 라벨링 된 메타정보, 대한약학정보원의 이미지를 매칭하여 약제 1종 별로 검수 할 수 있도록 도구 개발함
 - 검사 방법: 개발된 도구와 샘플링한 알약 DB를 USB에 담에 외부검수자에게 전달하여 검수토록 하였음
 - 검사 기간: 2021년 11월 30일, 12월 1, 2, 3, 6, 7일 (총 5일간)
 - 품질(의미정확성) 검증 결과: 의미정확성 오류율 0%



5.2. 검수 기준

| 단계 | 검사 기준 및 항목 |
|----|--|
| 획득 | <ul style="list-style-type: none"> - 숙련된 검수자가 카메라에서 이미지를 확인 - 다음의 조건을 만족하는지 확인 <ul style="list-style-type: none"> · 이미지가 흔들리지 않고 초점이 맞았는가? · 이미지가 규정된 촬영상 조건과 촬영 방식을 준수하였는가? · 수집자가 입력한 영상의 촬영 정보가 올바른가? |
| 정제 | <ul style="list-style-type: none"> - 이미지 데이터 정제 검수 기준 : 다음의 조건을 만족하는 데이터만 사용 <ul style="list-style-type: none"> · 피사체의 모양, 색상, 식별번호, 기호가 명확히 촬영되었는가? |

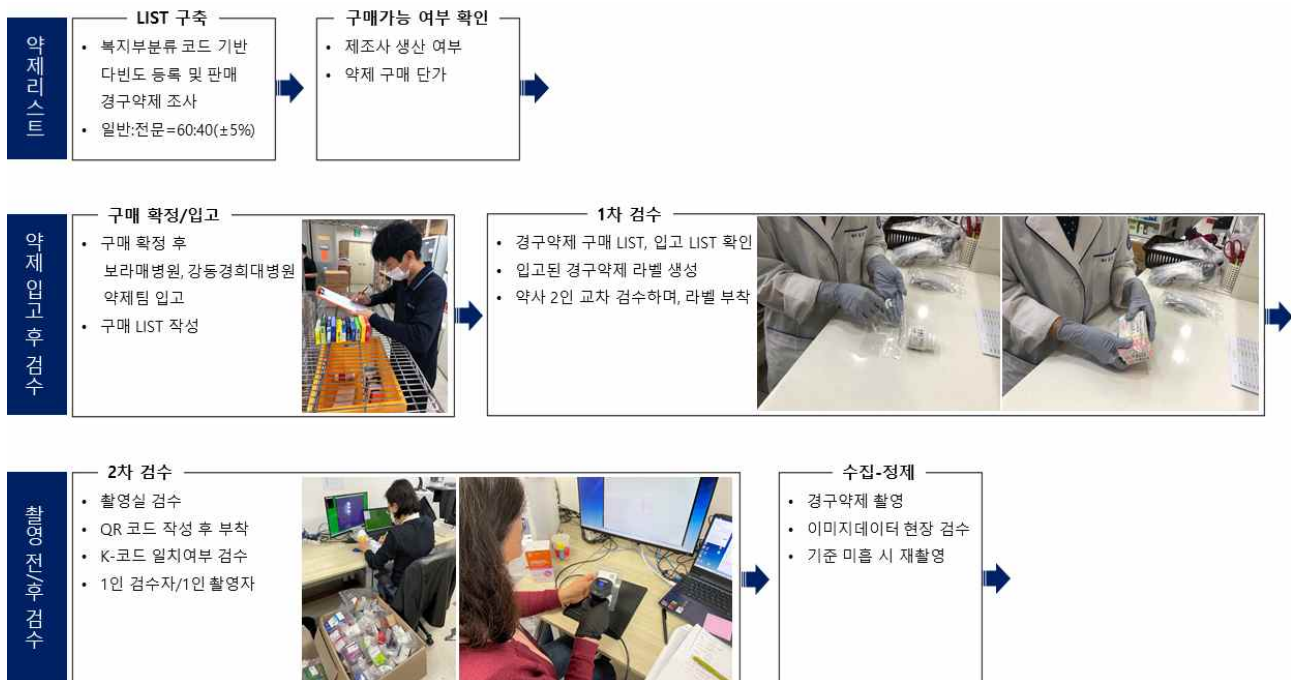
| | |
|-------------|--|
| | <ul style="list-style-type: none"> · 피사체에 그림자가 있는가? · 피사체에 포커싱이 잘 맞춰졌는가? · 불필요한 피사체가 들어있지는 않는가? · 조명에 과잉 노출되지는 않았는가? · 파일명명이 정확하게 되어 있는가? |
| 가공 (라벨링) | <ul style="list-style-type: none"> - 이미지 데이터 라벨링 검수 기준 : 다음의 조건을 만족하는 데이터만 사용 · 이미지 데이터로부터 경구 약제의 Bounding Box가 추출되었는가? · Bounding Box가 과잉 또는 축소 되어 있지는 않는가? · json파일이 구문적 정확성을 만족하는가? · 파일명명이 정확하게 되어 있는가? |
| 검수 | <ul style="list-style-type: none"> - 수집된 모든 데이터가 위의 획득/정제/라벨링 단계의 모든 조건을 만족하는가? - 이전에 수집된 오류 데이터 특성 및 사례에 속하지 않는가? - 중복되어 있는 데이터는 없는가? |

- 1차 검수: 경구 약제 수집 과정에서 약사에 의해 검수
 - 경구 약제 구매 제품과 리스트가 일치하는가?
 - QR코드의 K코드와 일치하는가?
 - 약 봉투의 경구 약제 종류와 라벨지 정보가 일치하는가?
- 2차 검수: 경구 약제 촬영팀에서 촬영자(수집자)에 의해 검수
 - 경구 약제의 모양, 색상, 식별번호, 기호가 명확히 촬영되었는가?
 - 경구 약제에 초점(포커싱)이 잘 맞춰졌는가?
 - 경구 약제 외 불필요한 피사체가 들어있지 않는가?
 - 경구 약제가 조명에 과잉 노출 또는 과부족 되어 촬영되지 않았는가?
 - 경구 약제 캡슐의 경우 약제에 빛 또는 기구물 반사가 나타나지 않았는가?
 - QR코드와 화면의 K 코드가 일치하는가?
- 3차 검수: 외부 약사 4인에 의해 의미정확성에 대한 교차 검수 진행
 - 원천데이터의 경구약제 이미지와 라벨링데이터의 약제 메타정보(약품명)가 일치하는가?
- 가공 과정의 검수: Bounding box 검수
 - 검수 기준: 클래스 태깅이 정확하고, IoU 0.8 이상
 - 오타깅: 라벨 오류 혹은 IoU 0.8 미만
 - 과태깅: 대상이 아닌 객체에 라벨링
 - 미태깅: 대상 객체 있지만, 라벨링 수행하지 않음

5.3. 검수 조직


| 단계명 | 세부 업무 내용 | 담당기관 | 담당자 |
|---------|---|----------------|-----|
| 검수 총괄 | 품질검사 기준 및 범위 마련 품질검사 방법, 일정 협의 품질검사 도구 준비/설치/테스트 | 서울특별시 보라매병원 | 이시욱 |
| 수집 및 정제 | 원시데이터 수집 후 검수 원천데이터 정제 및 저장 이미지데이터 파일 포맷, 기준 검수 최종 검수 및 승인 | 서울특별시 보라매병원 | 이시욱 |
| | | 경희대학교 산학협력단 | 김명천 |
| 가공 | 획득 데이터 검수 라벨링데이터 기준안 마련 라벨링 데이터 구조적 정확성 검수 바운딩박스 검수 | 인피노브(주) | 이경척 |
| 의미정확성 | 원천데이터와 라벨링데이터 정보 일치 여부(외부 약사 4인) | 서울특별시 보라매병원 | 이시욱 |

5.4. 기타 품질관리 활동



이미지 확인

- 전달 받은 알약 이미지 확인 (개수 미달, 사진 오류 검출)
- 데이터 정리 및 검수 DB 입력 진행



가공

- 라벨링데이터 추출
- 알약메타정보, 촬영메타정보, 바운딩박스




라벨링 데이터 교차 검수

- 바운딩박스 정확도 검수
- 검수자 2인 교차 검수 진행




검수 리스트 체크 및 최종 확인

- 검수 리스트 체크 진행
- 검수 결과 오류 발생 시 재가공
- 최종 확인 및 데이터 전달 진행




의미적 정확성 검수

- 내부 전문가(약사) & 외부 전문가(약사-섭외 중)
- 이미지데이터와 라벨링데이터(알약메타정보) 일치 여부 검수
- Tool 개발 중, 11월 셋째 주 시작



"dl_mapping_code": "K-000034",
 "dl_name": "페니라민정 2mg",
 "dl_name_en": "Peniramin Tab.",
 "img_key": "http://connectdi.com/design/img/d
rug/148609556259100153.jpg",
 "dl_material": "클로르페니라민말레산염",
 "dl_material_en": "Chlorpheniramine Maleate",
 "dl_custom_shape": "정제, 저작정",
 "dl_company": "(주)유한양행"



■ 일치

□ 불일치