*작성 가이드	• 데이터셋의 국문명, 영문명 모두 필수기재 • 동일 과제에 속한 데이터셋의 소개 내용은 각 데이터셋의 특징을 알 수 있도록 내용을 차별화하여 작성할 것(동일 내용의 반복은 피할 것) • 내용에 삽입된 이미지는 .PNG 포맷(2배 크기)로 저장하여 원본 이미지는 별도 제출할 것 • 내용에 삽입된 표, 그래프 자료는 엑셀 파일에 정리하여 별도 제출 • 내용의 구축기관 정보에 프로젝트 참여 기업 정보는 누락 없이 모두 기재할 것 (주관 수행기관사 책임자명/전화번호/메일주소 등은 연결 가능한 정보로 기재할 것) • 아래 메타테이블에 데이터 관련 문의처 정보는 데이터 이용자들에 대한 기술지원이 가능한 기관의 실무 담당자를 지정하여 정보를 기재할 것. 문의시 담당자와 직접 연결이 가능한 정보로 기재 • 아래 메타테이블에 가공기관, 검수기관은 업무 담당 업체명은 누락 없이 모두 기재할 것 • 전화번호의 경우, 개인 휴대폰 전화번호가 아닌 회사유선번호를 기재해야 함(개인정보) • 한글파일(hwp)로 제출 • 카테고리 정의서는 지정엑셀양식에 데이터카테고리 작성하여 제출

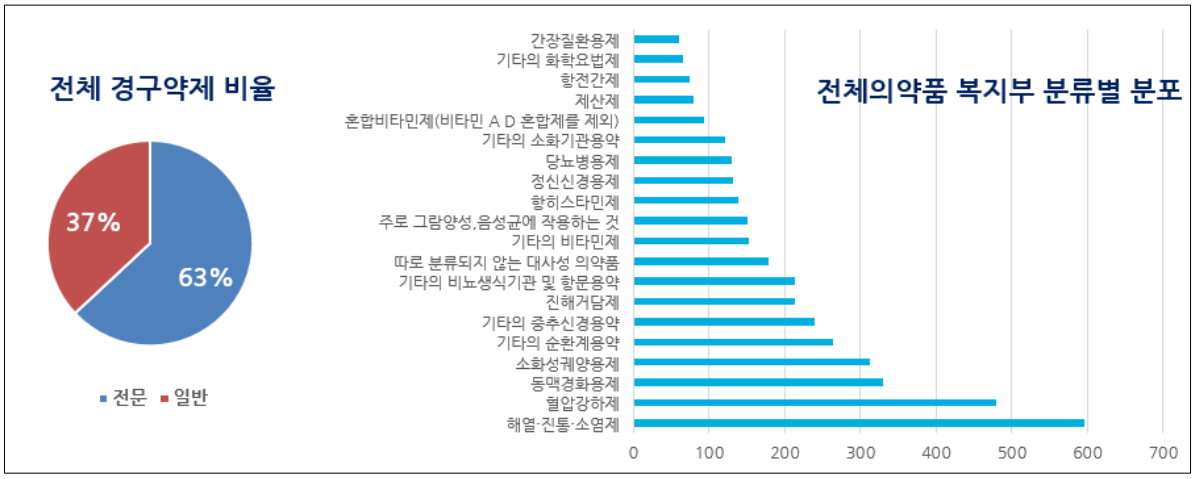
메타테이블 정보 (다중기입가능)	분야	데이터 유형[1]	구축 데이터량	원천데이터 형식[2]	라벨링 형식[3]	라벨링 유형[4]
	NIA 기입	이미지, 텍스트	279만장	PNG	JSON	약제 메타정보, 바운딩박스(이미지)
	데이터 출처[5]	데이터 구축년도	구축기관(총괄)	가공기관	검수기관	
	자체수집	*2021년*	서울특별시 보라매병원	인피노브(주)	서울특별시 보라매병원, 인피노브(주)	
	데이터 문의처	기관명	문의담자명	전화번호 (유선전화번호기입)	메일주소	
		서울특별시 보라매병원	이시욱	02-870-2671	shiuk.lee @gmail.com	
	데이터 소개	약품식별 인공지능 모델 학습을 위한 경구약제 이미지데이터와 라벨링데이터로써, 다양한 각도와 조명, 배경에서 촬영하고 의약품 메타 정보를 제공하는 데이터 셋				
	주요키워드	경구 약제, 알약, 약품식별 인공지능				
카테고리 정의서		전문의약품(3,143종, 62.8%) 일반의약품(1,857종, 37.2%)				

1) 텍스트, 오디오, 이미지, 비디오,
2) txt, jpg,.....
3) json, csv,....
4) 내용요약(텍스트), 번역(자연어), 질의응답(자연어), 바운딩박스(이미지/동영상), 키포인트(이미지/동영상), 세그멘테이션(이미지/동영상), 전자(음성) .....
5) 4대 언론기사, 자체 수집,.,.,.,

데이터셋명	국문	약품식별 인공지능 개발을 위한 경구약제 이미지데이터
	영문	Drug(tablet & pill) image data for the development of drug indetification AI
구축목적		국내 유동되고 있는 경구약제 5,000종을 대상으로 약품식별 인공지능 개발 및 학습을 위한 이미지와 라벨링 데이터 셋
활용서비스		• AI 기반 경구 약제 식별 어플리케이션 개발에 활용 • 의료인의 경구 약제 정보 접근성 강화 • DUR(의약품 안전사용서비스 시스템) 연동 후 약제의 이미지 데이터를 함께 제시함으로써 약제 처방의 오·남용을 예방
소개		• 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 • 보건복지부의 의약품 분류별 약품 개수 확인(2021년 7월 기준). 다빈도 등록 약품을 우선순위로 확보할 약품 개수 선정 • 전문의약품 3,143(62.8%)종, 일반의약품 1,857(37.2%)종을 확보 • 확보된 경구 약제는 다양한 각도와 조명, 배경에서 촬영하여 원천데이터를 확보하고, 약제에 대한 메타정보와 촬영 메타정보, 바운딩 박스로 라벨링데이터를 구축함 	
데이터셋 통계 (구축 규모 및 분포)		1. 데이터 구축 규모

구분	구축목표		구축달성	
	1종당 이미지 수	합계	1종당 이미지 수	합계
단일 경구약제 1,000종	1,250	1,250,000	1,296	**1,296,000**
단일 경구약제 4,000종	200	800,000	수량 다양화 (1,296/324/216)	**1,482,975**
경구약제 조합 5,000조합	3	15,000	3	**15,003**
총 합계		2,065,000		**2,793,978**
목표대비 성과율	100%		**135%**	

2. 데이터 분포

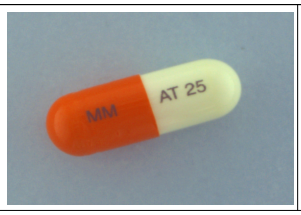


클래스 구분	구축계획	경구 약제 품목(구축달성)	
		종 수량	비율
전문의약품	60%(±5%)	3,143	62.8%
일반의약품	40%(±5%)	1,857	37.2%

데이터셋 구성

1. 대표도면

정제 경구약제



경질 캡슐 경구약제

연질 캡슐 경구약제

d

경구약제 조합 이미지데이터

2. 라벨링데이터 구성

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_anal	String	M	알약 앞면 마크 형태	
	images[].mark_code_back_anal	String	M	알약 뒷면 마크 형태	
	images[].mark_code_front_img	String	M	알약 앞면 마크 이미지	
	images[].mark_code_back_img	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	라벨링정보	
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	

3 라벨링데이터 실제예시


```
"images": [
        {
                "file_name": "K-000255_0_2_0_0_90_260_200.png",
                "width": 976,
                "height": 1280,
                "imgfile": "K-000255_0_2_0_0_90_260_200.png",
                "drug_N": "K-000255",
                "drug_S": "정상알약",
                "back_color": "연회색 배경",
                "drug_dir": "앞면",
                "light_color": "전구색",
                "camera_la": 90,
                "camera_lo": 260,
                "size": 200,
                "item_seq": "197400273",
                "leng_long": "7.3",
                "leng_short": "7.3",
                "thick": "3.3",
                "id": 1,
                "dl_idx": "254",
                "dl_mapping_code": "K-000255",
                "dl_name": "부광살부타몰정 4mg",
                "dl_name_en": "Salbutamol Tab. 4mg Bukwang",
                "img_key": "http://connectdi.com/design/img/drug/1NXA-1cMBWw.jpg",
                "dl_material": "살부타몰황산염",
                "dl_material_en": "Salbutamol Sulfate",
                "dl_custom_shape": "정제, 저작정",
                "dl_company": "부광약품(주)",
                "dl_company_en": "Bukwang Pharm",
                "di_company_mf": "",
                "di_company_mf_en": "",
                "di_item_permit_date": "19740429",
                "di_class_no": "[02220]진해거담제",
                "di_etc_otc_code": "전문의약품",
                "di_edi_code": "642200750,A13100741",
                "chart": "담적색의 정제임.",
                "drug_shape": "원형",
                "print_front": "S",
                "print_back": "",
                "color_class1": "분홍",
                "color_class2": "",
                "line_front": "",
                "line_back": "",
                "img_regist_ts": "20041223",
                "form_code_name": "나정",
                "mark_code_front_anal": "",
                "mark_code_back_anal": "",
                "mark_code_front_img": "",
                "mark_code_back_img": "",
                "mark_code_front": "",
                "mark_code_back": "",
                "change_date": "20190402"
        }
],
"type": "instances",
"annotations": [
        {
                "area": 36465,
                "iscrowd": 0,
                "bbox": [
                        383,
                        550,
                        195,
                        187
                ],
                "category_id": 1,
                "ignore": 0,
                "segmentation": [],
                "id": 1,
                "image_id": 1
        }
],
"categories": [
        {
                "supercategory": "pill",
                "id": 1,
                "name": "Drug"
        }
]
```

		기관명	책임자명	전화번호 (유선전화번호기입)	메일주소	담당업무
데이터셋 구축 수행기관 담당자	주관기관	서울특별시 보라매병원	이시욱	02-870-2671	shiuk.lee @gmail.com	
	참여기관	기관명	담당업무	기관명	담당업무	
		경희대학교 산학협력단	원천데이터 구축, 검수	㈜모비소프트	원천데이터구축	
		㈜원스글로벌	경구약제 메타정보 구축	㈜더유니파이	인공지능 모델 개발	
		부광아이티(주)	원천데이터 구축			
		인피노브(주)	데이터 가공, 라벨렝데이터 구축, 검수			