

메타데이터 정보 (다중기입가능)	분야	데이터 유형 ¹⁾	구축 데이터량	원천데이터 형식 ²⁾	라벨링 형식 ³⁾	라벨링 유형 ⁴⁾
	영상이미지	이미지	500,000	jpg	json	polygon, bbox
	데이터 출처 ⁵⁾	데이터 구축년도	구축기관(총괄)	가공기관	검수기관	
	참여 동물병원	2021년	미소정보기술	바이오데이즈	바이오데이즈	
	데이터 문의처	기관명	문의담당자명	전화번호 (유선전화번호기입)	메일주소	
		(주)바이오데이 즈	이종현	031-446-75 00	aes911@naver .com	
	데이터 소개	인공지능 기반 반려동물 질병 판독 기술 개발을 위해 반려동물 2종의 품종, 성별, 나이별, 질환별 학습데이터를 구축				
	주요키워드	반려동물, 반려견, 반려묘, 수의사, 질환 분류, 데이터 수집, 라벨링, 데이터 정제, 데이터 가공, 학습용 데이터 검증				
카테고리 정의서		첨부의 카테고리 정의서 엑셀파일에 데이터카테고리 작성하여 제출(예시참고)				


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)

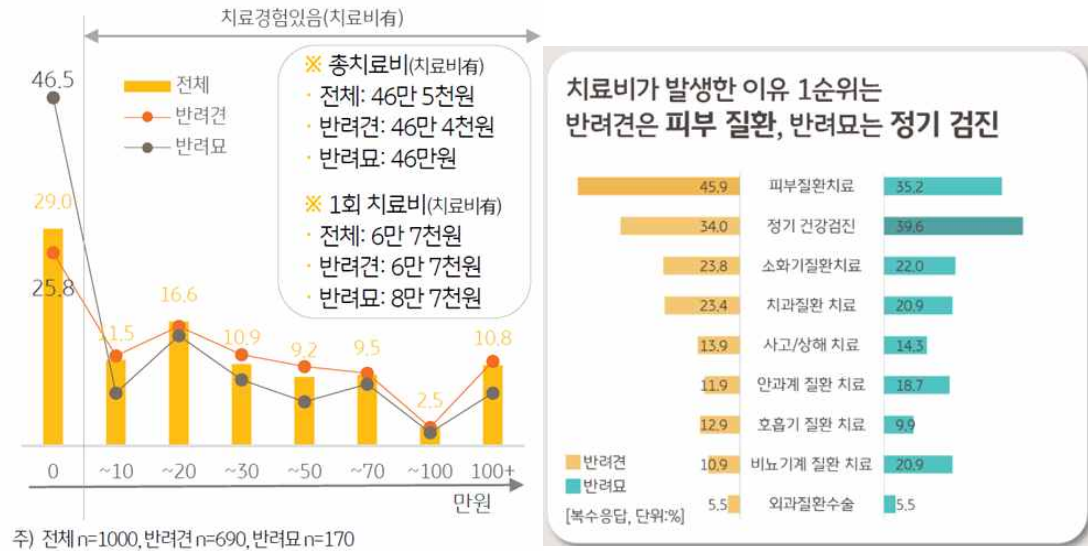
5) 4대 언론기사, 자체 수집,.....

데이터셋명	국문영문	반려동물 피부질환 데이터
		Companion animal skin disease data
구축목적	반려동물의 피부질환 진단을 보조하고 중증도를 판정하는 인공지능 기술개발을 위한 대규모 데이터셋 구축	
활용서비스	구축된 반려동물 피부질환 데이터 셋을 피부질환 진단 및 중증도 판정 알고리즘 개발,검증에 활용	
소개	<p>피부와 비전공 의사에게 학습된 피부질환에 대한 감별질환 목록을 안내하고 우선 순위에 따른 치료방침을 제시하여 조기 치료가 가능한 질병은 신속하게 치료하고, 알러지, 면역매개성, 내분비, 종양성 피부질환과 같이 전문병원 치료가 필요할 경우 해당 2차, 3차 병원을 연계할 수 있는 서비스 제공</p>  <p>[그림 1] 인공지능 학습용 데이터 활용 지원</p>	
데이터셋 통계 (구축 규모 및 분포)	1.데이터 구축 규모 및 분포	
	데이터 명	AI 학습용 반려동물 피부질환 데이터
	데이터 포맷	jpg(이미지 파일), csv(메타데이터 파일)
	활용 분야	반려동물(개, 고양이)의 피부질환 촬영 이미지를 이용한 인공지능 진단 기술개발을 위한 데이터셋 구축
	데이터 요약	<p>피부질환 이환 동물의 개체별 모든 피부병변 발생부위에 대한 정보와 이미지를 함께 확보</p> <p>피부질환 수집 가이드를 기준으로 의무기록에 포함된 품종, 연령, 성별 등의 메타정보를 수집 이미지 획득 시에는 털을 병변부 주변으로 밀거나 스프레이로 물을 분무한 상태에서 촬영을 수행하여 털유무에 따른 영향을 최소화</p>

데이터 출처		대학병원, 동물병원에서 기 수집한 피부질환 촬영 이미지, 진료를 통해 촬영 된 이미지
데이터 통계	데이터 구축 규모	반려동물 10,000마리 이상의 50만장 이미지
	데이터 분포	반려견 : 7개 질환으로 구성된 총 446,000장 이미지 반려묘 : 4개 질환으로 구성된 총 54,000장 이미지 정상동물: 부위별 피부 사진으로 구성된 250,000장 이미지
기타 정보	대표성	서울, 경기, 대구 등 반려동물 부양률이 높은 지역 동물병원 데이터 활용
	독립성	원시 데이터는 라벨링 데이터와 별개의 데이터 셋으로 제공하여 독립성 유지
	유의사항	비식별화를 통한 법·제도 준수

데이터셋 구성	<ul style="list-style-type: none"> ’ 17년 기준 전국 1,952만 가구 중 29.4%인 574만 가구에서 개 632만마리, 고양이 243만 마리를 기르고 있는 것으로 추정되고 있음(한국농촌경제연구원, 2018)  <p>[그림] 반려동물 연관산업별 시장 성장률(좌) 및 반려동물별 양육 비중(우)</p> <ul style="list-style-type: none"> 연관 산업 규모는 ’ 17년 기준 2조 3,322억원에서 ’ 27년 6조 원 이상으로 전망되며, ’ 14년 기준 1조 5,684억원으로 연평균 14.5%씩 성장하고 있고 수의서비스 분야는 6,551억원으로 ’ 14년 기준 약 42%를 차지함(한국농촌경제연구원, 2018) 치료비가 차지하는 비중으로 보면, 피부질환치료는 반려견에서 45.9%, 반려묘에서 35.2%를 차지하여 가장 높은 것으로 보고되어 있으며, 진료비도 최근 2년간 1회에 67,000원을 차지하여 부담으로 조사되었음

그림III-22 | 지난 2년간 반려동물 총치료비 (단위:%)

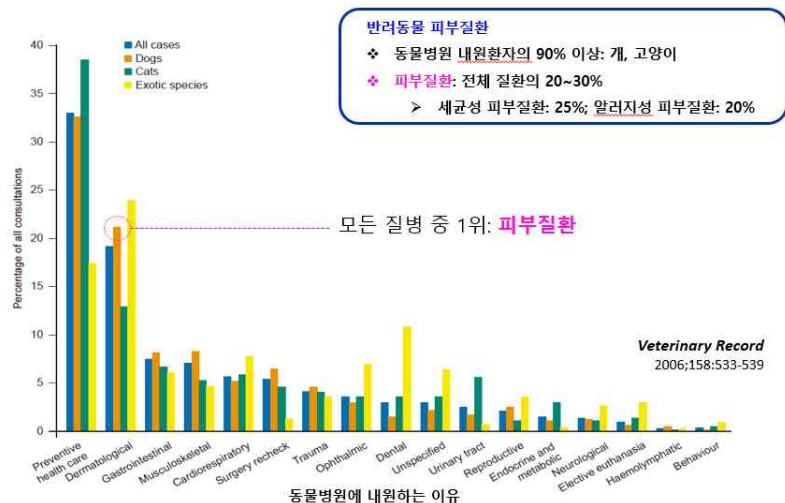


(KB경영연구소의 2021 한국 반려동물보고서)

- 4차 산업을 필두로 인공지능 관련 기술과 관련 산업이 급격히 발전하고 있는 가운데, 의료용 빅데이터를 이용한 인공지능 기반의 1차 진단, 판독 등의 의료서비스와 인공지능이 결합한 형태의 융합 비즈니스가 점차 확산

가. 데이터 구축 필요성

- 국외 조사에 따르면 동물병원에 내원하는 환자의 90% 이상이 개와 고양이이며, 내원 이유 중 피부질환은 20~30%를 차지하여 일반건강검진 다음으로 높은 순위를 보이고 있음. 국내 조사에서도 피부질환은 20.1%로 내원사유 중 1위를 나타내 반려동물 질환 중 중요도가 매우 높다고 할 수 있음(농촌진흥청, 2018)



[그림] 동물병원에 내원하는 반려동물 품종 및 질병 빈도

1. 라벨링데이터 구성

<일반 카메라>

구분	영문명	타입	필수여부	국문명	범위	비고
1	metaData	Object		이미지 정보		
1-1	Raw data ID	String	Y	파일명		
1-2	copyrighter	String	Y	저작권자	"(주)미소 정보기술"	
1-3	resolution	String	Y	해상도	"1920x10 80"	
1-4	date	String	Y	촬영 일자	"YYYY-M M-DD"	
1-5	breed	String	Y	품종	말티즈, 포메라니 안,푸들...	
1-6	age	String	Y	나이	1~30	
1-7	gender	String	Y	성별	"F" "M"	
1-8	region	String	Y	촬영위치	"L" "H" "B" "A"	L=다리 H=머리 B=몸통 A=연접부
1-9	camera type	String	Y	촬영장비	"IMG", "CYT"	
1-10	species	String	Y	반려종	"D" "C"	
1-11	lesions	String	Y	증상	"A1", "A2", "A3", "A4", "A5", "A6", "A7"	A1=구진/ 플라크 A2=비듬/ 각질/상피 성잔고리 A3=태선 화/과다색 소침착 A4=농포/ 여드름 A5=미란/ 궤양 A6=결정/ 종괴 A7=무증 상

	1-12	diagnosis	String	Y	질병	"null"	공백
	1-13	Path	String	Y	이미지 폴더명	"유증상", "무증상"	
	1-14	identifier		Y	데이터 구분	"피부질환"	
	1-15	src_path		Y	파일 경로	"/라벨링 데이터/반 려견/피부 /일반카메 라/..."	
	1-16	label_path		Y	데이터셋 경로	"/라벨링 데이터/반 려견/피부 /일반카메 라/..."	
	1-17	type		Y	데이터 타입	"json"	
	1-18	fileformat		Y	이미지 포맷	"jpg"	
	2	inspRejectYn	String		반려처리 여부		
	3	labelingInfo	Object				
	3-1	polygon	Object	Y			
	3-1-1	color	String	Y	라벨링 색상		
	3-1-2	location	Object	Y	라벨링 좌표		
	3-1-3	label	String	Y	라벨링 이름	"A1_구진_ 플라크" "A2_비듬_ 각질_상피 성잔고리" "A3_태선 화_과다색 소침착" "A4_농포_ 여드름" "A5_미란_ 궤양" "A6_결절_ 종괴" "A7_무증 상"	

	3-1-4	type	String	Y	저작도구 종류	"polygon"	
	3-2	box	Object	Y			
	3-2-1	color	String	Y	라벨링 색상		
	3-2-2	location	Object	Y	라벨링 좌표		
	3-2-3	label	String	Y	라벨링 이름	"A1_구진_플라크" "A2_비듬_각질_상피_성잔고리" "A3_태선_화_과다색_소침착" "A4_농포_여드름" "A5_미란_궤양" "A6_결절_종괴" "A7_무증_상"	
	3-2-4	type	String	Y	저작도구 종류	"box"	

구분	영문명	타입	필수 여부	국문명	범위	비고
1	metaData	Object		이미지 정보		
1-1	Raw data ID	String	Y	파일명		
1-2	copyrighter	String	Y	저작권자	"(주)미소 정보기술"	
1-3	resolution	String	Y	해상도	"1920x10 80"	
1-4	date	String	Y	촬영 일자	"YYYY-M M-DD"	
1-5	breed	String	Y	품종	말티즈, 포메라니 안,푸들...	
1-6	age	String	Y	나이	1~30	
1-7	gender	String	Y	성별	"F" "M"	
1-8	region	String	Y	촬영위치	"L"	L=다리

						"H" "B" "A"	H=머리 B=몸통 A=연접부		
	1-9	camera type	String	Y	촬영장비	"IMG", "CYT"			
	1-10	species	String	Y	반려종	"D" "C"			
	1-11	diagnosis	String	Y	질병	"C1", "C6"	C1=감염성 피부염 C6=비감염성 피부염		
	1-12	Path	String	Y	이미지 폴더명	"감염성피부염", "비감염성피부염"			
	1-13	identifier			데이터 구분	"피부질환"			
	1-14	src_path			파일 경로	"/라벨링 데이터/반려견/피부 /현미경/ ..."			
	1-15	label_path			데이터셋 경로	"/라벨링 데이터/반려견/피부 /현미경/ ..."			
	1-16	type			데이터 타입	"json"			
	1-17	fileformat			이미지 포맷	"jpg"			
	2	inspRejectYn	String		반려처리 여부				
	3	labelingInfo	Object						
		3-1	polygon	Object	Y				
			3-1-1	color	String	Y	라벨링 색상		
			3-1-2	location	Object	Y	라벨링 좌표		
			3-1-3	label	String	Y	라벨링 이름	"C1_감염성피부염" "C6_비감	

						염성피부 염	
	3-1-4	type	String	Y	저작도구 종류	"polygon"	
	3-2	box	Object	Y			
	3-2-1	color	String	Y	라벨링 색상		
	3-2-2	location	Object	Y	라벨링 좌표		
	3-2-3	label	String	Y	라벨링 이름	"C1_감염 성피부염" "C6_비감 염성피부 염"	
	3-2-4	type	String	Y	저작도구 종류	"box"	

<현미경 카메라>

3 라벨링데이터 실제예시

```

1  {
2    "metaData": {
3      "Raw data ID": "IMG_D_A1_000001.jpg",
4      "copyrighter": "㈜미소정보기술",
5      "resolution": "1920X1080",
6      "date": "2021-09-07",
7      "breed": "포메라니안",
8      "age": "4",
9      "gender": "M",
10     "region": "B",
11     "camera type": "IMG",
12     "species": "D",
13     "lesions": "A1",
14     "diagnosis": "",
15     "Path": "유종상",
16     "identifier": "피부질환",
17     "src_path": "/라벨링데이터/반려견/피부/일반카메라/유종상/A1 구진 플라크",
18     "label_path": "/라벨링데이터/반려견/피부/일반카메라/유종상/A1 구진 플라크",
19     "type": "json",
20     "fileformat": "jpg"
21   },
22   "inspRejectYn": "N",
23   "labelingInfo": [
24     {
25       "polygon": {
26         "color": "#1beaac",
27         "location": [
28           {
29             "x1": 895,
30             "y1": 287,
31             "x2": 1365,
32             "y2": 289,
33             "x3": 1333,
34             "y3": 809,
35             "x4": 946,
36             "y4": 815,
37             "x5": 896,
38             "y5": 290,
39             "x6": 895,
40             "y6": 287
41           ]
42         },
43         "label": "A1 구진 플라크",
44         "type": "polygon"
45       }
46     ]
47   },
48 }

```

데이터셋 구축 수행기관 담당자	주관기관	기관명	책임자명	전화번호 (유선전화번호기입)	메일주소	담당업무
		미소정보기술	강태준	02-2205-0551	tjkang@ misoinfo.co.kr	총괄책임자
	참여기관	기관명	담당업무	기관명	담당업무	
		바이오테이즈	데이터 설계, 수집, 정제, 가공, 검수	서울대학교 산학협력단	데이터 수집	
		충북대학교 산학협력단	데이터 수집	에스에스엘	외부 품질검증 대응	

		가치 랩스	AI 모델 개발			