

인공지능 데이터 구축 가이드 라인

- 동의보감 독초 이미지 -

담당 역할	기관명
데이터 구축 총괄	가천대학교 산학협력단
데이터 설계	(유)엔와이텔
데이터 수집	(주)믹스미디어, 가천대학교 산학협력단, (주)동북아생물다양성연구소, (유)엔와이텔
데이터 정제	(유)엔와이텔, 가천대학교 산학협력단
가공(라벨링, 어노테이션)	(유)나인펄스
데이터 검수	(유)나인펄스, (주)아워텍, (유)엔와이텔

구축 가이드라인 작성	// (유)엔와이텔	// 김병철
	// 가천대학교산학협력단	// 서정범
가이드라인 버전 (제작일자)	Ver 1.5 ('21.10.29)	

본 데이터 구축 가이드라인 Ver 1.5에 삽입된 이미지는 사업기간 중 클라우드워커로 참여한 분이 촬영한 이미지를 사용하였으며, 본 과제 특성에 맞게 식물을 재분류 과정에 사용된 식물 명칭, 부위 설명, 식물 관련 용어 등 식물을 표현하는 내용 중 일부는 위키피디아에서 인용하였음.

목 차

1. 데이터 구축 개요	3
2. 문제정의	9
2.1 임무 정의	10
2.2 데이터 구축 유의사항	11
3. 데이터 수집·정제	12
3.1 원시데이터 선정	12
3.2 수집·정제 절차 및 방법	29
3.3 수집·정제 기준	35
3.4 수집·정제 조직	47
3.5 수집·정제 도구	49
4. 데이터 가공	50
4.1 가공 절차 및 방법	50
4.2 가공 기준	51
4.3 가공 규격	61
4.4 가공 조직	66
4.5 가공 도구	66
5. 검수	75
5.1 검수 절차 및 방법	75
5.2 검수 기준	77
5.3 검수 조직	79
5.4 검수 도구	80
5.5 기타 품질관리 활동	81

[부록]

별첨 1 : 식물 잎 유형

1. 데이터 구축 개요

1.1 구축 데이터 요약

<p><구축 데이터 요약> 해마다 독초를 약초로 오인, 섭취하여 발생하는 독초 중독 및 사망 사고를 예방하기 위해, 동의보감 수록 457종 중 독초 60종과 비교식물로 독초와 모양이나 명칭이 유사한 식물 61종을 선정하여 독초판별 AI 모델 학습 및 독초판별 알고리즘 개발을 위한 인공지능 학습용 데이터임.</p>
--

1.1.1 데이터 명세 정보

데이터 이름	동의보감 독초 이미지 데이터	
데이터 활용	원천 데이터	식물 교육 교재 개발 등
		중독사고 예방 책자 제작
		데이터 분류 체계 표준화 수립 기초 연구자료로 활용
데이터 활용	학습 데이터	약용 및 특용작물 인공지능 분류 모델 학습용으로 활용
		식물 성장 예측 모델 연구에 활용
		독초판별 인공지능 서비스 개발에 활용
데이터 구축	획득	독초 60종, 유사 비교식물 61종을 직접 촬영(스마트폰, DSLR 카메라)
	정제	전수 데이터 품질 검증(이미지 품질, 분류 오류, 촬영기준, 전문가 검증)
	가공	공개 저작도구 LableMe를 이용하여 라벨링·어노테이션
	검수	자체 개발한 Label Checker를 이용하여 가공 데이터 검수
	지원	허브넷(herbnet.kr) 플랫폼

1.2 추진 배경 및 필요성

1.2.1 추진배경

- **동의보감(東醫寶鑑)** 1610년(광해군 2년) 허준(許浚)이 펴냄

17세기 동아시아 임상실험 내용을 집대성한 전문 의서로 세계적으로 가치를 인정받고 있으며, 2009년 유네스코 세계기록유산 등재, 2015년 국보 지정. 현재에도 보건의료 가치가 뛰어나 한 의학의 핵심 의서로 활용, 동의보감을 목적에 맞게 해석된 다양한 책이 출간되어 일반인들도 널리 활용하고 있음

- ① **병을 고치기에 앞서 수명을 늘리고, 병이 안 걸리도록 하는 방법을 제시.**

몸을 잘 지키고 병을 예방하는 것이 병 걸린 후 치료하는 것보다 더 낫기 때문.

- ② **처방의 요점만 간추림.** 의서마다 사용하는 처방이 3 다른 것을 요점 위주로 간추려 정리함

- ③ **국산 약초를 쉽게 쓸 수 있도록 약초 이름 한글 표기**

사는 곳 주변에서 나는 독초를 쉽게 쓸 수 있도록 사람이 부르는 독초 이름을 한글로 표기

가. 독초에 의한 중독 및 사망 사고 발생

- 독초를 약초로 잘못 알고 복용한 중독사고가 빈번하게 일어나고 있음, 식품의약품안전처 통계에 따르면 지난 10년(2010년~2019년)간 독초 중독으로 251명 사망.
- **독초 중독사고 및 오인 피해 발생 원인**

- **독초판별 미숙 및 정확하지 않은 정보 활용**

- 독초 중에는 산나물이나 약초와 구분하기 힘들 정도로 형태가 유사한 식물 존재
- 약초에 대한 전문 지식 없는 비전문가의 주관적인 의견으로 게재한 인터넷 정보 활용

- **대표 사례**

- 지리강활과 참당귀, 동의나물과 곰취, 삼지구엽초와 꿩의다리, 칩과 등칫 등

나. 약초를 독초로 오인하여 발생한 사고 사례(동의나물-곰취, 참당귀-개당귀)



중독증상

동의나물의 경우 식용인 곰취, 머위와 유사해 잘못 먹는 사례 발생. 이 식물을 먹으면 구토, 복통, 구강 마비, 저혈압, 쇼크, 호흡장애, 사지 저림, 수포 등의 증상이 나타날 수 있음.

독성성분

쿠마린(comarin)-간손상, 아네모닌(anemonin)-심장마비 유발 등의 독성 물질 함유

**지난 10년간
독초 중독 사망 사고
251건**

출처, 2020년 식약처 통계자료

독초 중독 사고, 약초 오·남용 사고 발생 주요 요인

약초와 독초 판별 능력 부족

부작용 등 주의사항 무시

약초에 대한 맹신으로 인한 과용



중독증상

개당귀(지리강활)의 경우 식·의약품인 왜당귀, 참당귀와 유사해 잘못 먹는 사례 발생. 이 식물은 **항독성 식물**로 안구 충혈, 얼굴이 붉어지고, 경련, **섭취 후 1~2시간 후 사망**

독성성분

에토포시드(etoposide)-급성신부전증 유발, 독소루비신(doxorubicin)-심장손상 등의 독성 물질 함유

1.3.2 추진 필요성

가. 인공지능 연구 개발 특정 분야에 편중

- 자율주행, 사물 인식, 스마트 재배, 스마트 공장 등 산업적 수요가 큰 분야 AI 연구 편중.
- 식물분류 분야 인공지능은 인공지능 학습모델 설계의 어려움, 적합한 데이터 수량 부족, 수익성 담보의 어려움 등의 요인으로 관련 연구와 개발이 다른 분야 대비 상대적으로 더디게 이루어지고 있음.

나. 식물(약초, 독초, 농작물 등) 분류, 판별 등 AI 학습에 필요한 데이터 부족

- 기존 정부 기관이 구축한 데이터는 농업 기술 개발과 식량 생산 증대 연구에 적합하며,
- 인공지능 모델 학습에 적합한 수준의 품질확보가 되어 있지 않으며, 수량 또한 절대적으로 부족하며, 필요 데이터 확보에 따른 많은 인력과 시간, 예산의 한계는 중소 벤처 기업 등이 감당하기 어려운 것이 현실.
- 이에 따라 정부 주도로 인공지능 모델 학습에 필요한 대량의 이미지 데이터와 높은 수준의 품질이 확보된 동의보감 독초 이미지 데이터 구축이 필요.

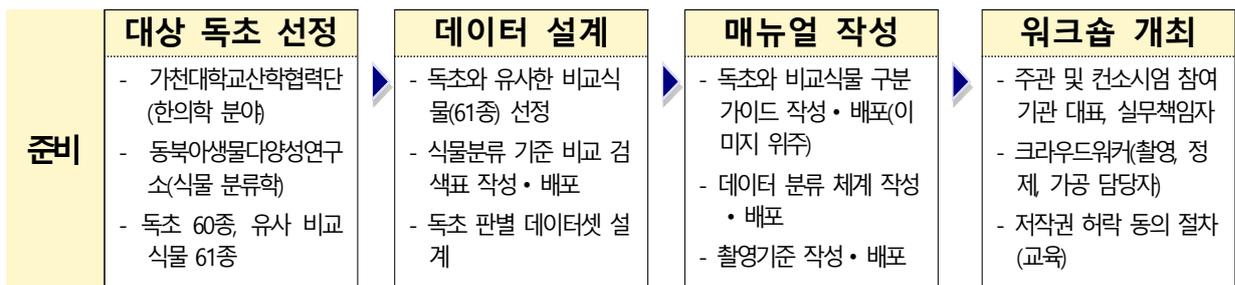
다. 데이터 기반 K-Herb 인공지능 연구 체계 구축

- 우리나라에 자생하는 약초에 대한 임상 정보를 집대성한 책으로 유네스코 세계 기록 유산으로 등재된 '동의보감'을 기준으로 하여, 다량의 국내 자생 약용식물 이미지 데이터 수집하고, 인공지능 학습에 적합하도록 가공 구축하여 폭넓은 인공지능 연구와 기술 개발 필요한 여건을 마련하여,
- 데이터 기반의 인공지능 K-Herb 혁신 필요.

1.4 데이터 구축 절차

- 데이터 구축은 사업 수행계획서와 데이터 구축 계획서상에 명시한 절차와 내용에 따라 수행하고, 각 구축 단계별로 데이터의 품질 검수를 통해 수준이 높은 학습데이터가 될 수 있도록 사업을 수행함.

◦ 구축 절차





1.4.1 데이터 획득

◦ 대상 품목 및 목표

구분	대상 품목	촬영 부위	목표 수량
선정 독초	가죽나무, 가지, 갈대, 감국, 개오동, 곰취, 팽이밥, 까마중, 꿀풀, 넝쿨고사리, 능소화, 단삼, 담쟁이덩굴, 대추나무, 돼지감자, 띠, 마디풀, 마타리, 메밀, 뿔대추나무, 무화과나무, 미나리, 밤나무, 회화나무 등 60종	전초, 잎, 꽃, 열매	각 종당 5,000장 이상
유사 식물	참죽나무, 도깨비가지, 달뿌리풀, 억새, 산국, 꽃개오동, 참취, 선괭이밥, 미국까마중, 배초향, 황고사리, 등골나물, 골등골나물, 대청부채, 물봉선, 개옷나무, 미국실새삼, 세뿔석위, 수박, 석잠풀, 개시호, 난쟁이아욱, 닥풀, 좀담배풀, 왕자귀나무, 가는장구채, 분홍장구채, 당아욱, 맑은대쑥, 털진득찰 등 61종	전초, 잎, 꽃, 열매	각 종당 5,000장 이상

◦ 획득 도구 : DSLR 카메라, 스마트폰을 이용하여 직접 촬영

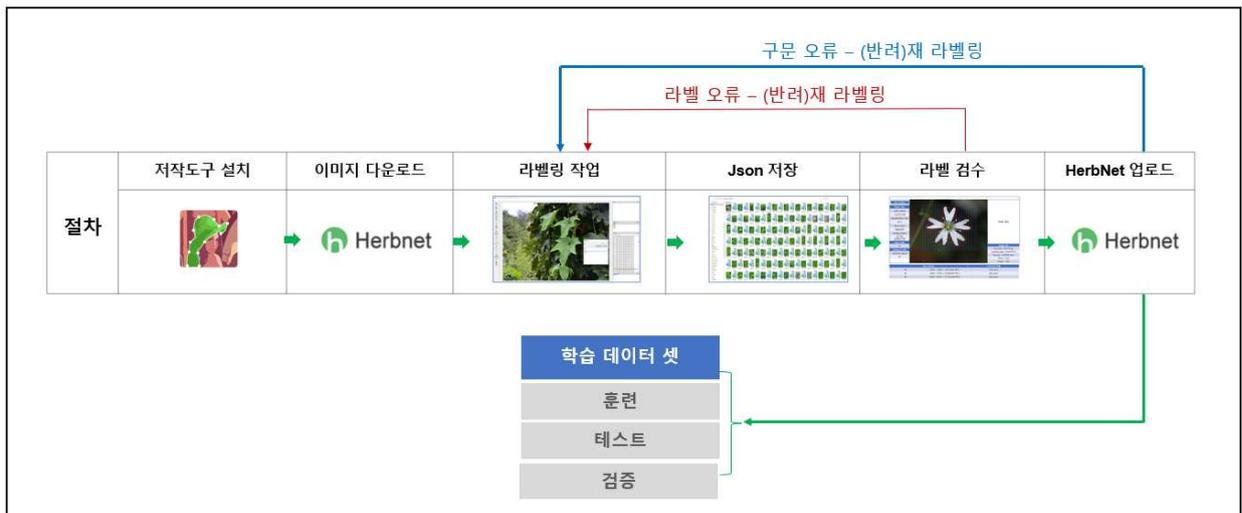
1.4.2 데이터 정제

◦ 데이터 정제 작업은 허브넷 플랫폼(Herbnet)을 통해 온라인에서 수행

단계	1단계	2단계	3단계	4단계
역할	원시 이미지 데이터 품질 검수	데이터 분류 오류 검수	촬영 기준 준수 여부 검수	정제 데이터 품질 검수
정제	흔들림, 초점 흐림, 대상 부위 잘림, 지나친 대상 확대 또는 축소, 원거리 촬영, 다른 사물 개입(손 등), 그림자, 오염(이물질 등) 등의 이미지를 엄선하여 처리함	1차 정제된 데이터 중 분류가 잘못된 데이터를 바로잡는 과정으로 전초, 잎, 꽃, 열매 4부위 중 데이터 분류가 잘못된 경우는 폴더 이동 또는 대상 이동 처리함	각 부위별 촬영 허용 범위에 맞게 촬영되어 있는지 여부를 검수하는 과정으로 전초, 잎, 꽃, 열매 부위별 촬영 허용 기준에 부합되는지 여부를 검수하여 처리함	1,2,3차에 걸쳐 정제 완료된 데이터는 한의학, 식물분류학 전문가의 전수 검수를 통해 반려 또는 승인 처리함
처리	반려, 논의, 승인	폴더 이동 또는 대상 이동	반려, 논의, 승인	반려, 승인

1.4.3 데이터 가공

◦ 데이터 가공 절차



◦ 정제 완료된 데이터(원천)를 인공지능 학습에 적합하도록 공개 저작도구를 활용하여 Bounding box, 폴리곤한 후 어노테이션하여 JSON 파일로 저장

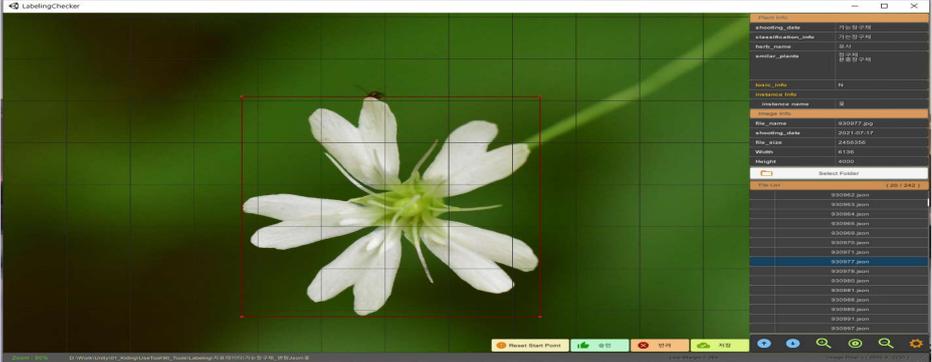
- 라벨링·어노테이션은 공개 저작도구인 LableMe 사용
- 기본적으로 바운딩박스를 하고 일부 품목은 데이터 활용도를 높이기 위해 폴리곤 함
- 정제 완료된 이미지 한 장에서 가공은 최대 2건 이내로 함

1.4.4 라벨링-어노테이션 데이터 검수

검수 절차	<ul style="list-style-type: none"> 라벨링 → 라벨링 참여자 간 교차 확인 → 부적합 시 : 검수자 정정
검수 방법	<ul style="list-style-type: none"> 라벨링 작업 오류 검수 2인 이상 교차검증
검수 조직	<ul style="list-style-type: none"> 라벨링 참여자 및 검수 관리자
검수 기준	<ul style="list-style-type: none"> 표기 오류, 오타 오류, 라벨링 위치 잘못 지정 등 검수 작업자의 실수로 인해 라벨링 위치가 잘못 지정된 데이터 선별 클래스 표기 오류(예, 꽃을 잎으로 잘못 표기) 데이터 선별

◦ 검수 도구

1차 검수		라벨링 과정에서 사용한 공개 저작도구인 Label me 를 이용하여 작업자 검수, 작업자 간 교차 검수를 수행하여 가공 데이터의 품질을 확보
-------	---	--

2차 검수		가공 데이터 검수 툴(Label checker)를 이용하여 가공 데이터를 검수함
-------	---	--

1.4.5 학습데이터 분류 및 저장

- 학습데이터 클래스는 동의보감 독초 및 유사 비교식물 종으로 하고 식물의 4 부위인 전초, 잎, 꽃, 열매를 인스턴스로 분류하여 저장하였음
- 데이터 저장 구조



2. 문제정의

- 사람은 독초나 식물을 판별하기 위해 눈으로 보고(시각), 만져보고(촉각), 냄새 맡고(후각), 맛을 보면서(미각) 종합적이고, 입체적으로 식물의 특징을 파악하고, 자신이 축적된 지식과 경험을 활용하여 종합적으로 판별하며,
- 반면 인공지능은 입력된 학습데이터의 특성을 분석하고 분류하여 독초나 유사 식물을 판별하기 때문에 이미지 데이터 획득 시 식물의 주요 판별 요소인 잎, 꽃, 열매와 식물의 성장을 판단할 수 있는 요소인 전초에 대한 촬영을 입체적으로 할 필요가 있음.

· 독초(식물) 판별 방법 - 사람과 인공지능 비교

구분	사람의 경우	인공지능의 경우
전초	전초 생육단계 모양, 형태 등을 관찰하고, 축적된 지식과 판별 경험을 활용	입력된 전초 이미지 데이터 특성과 특징을 분석
잎	잎의 모양, 털의 존재 여부, 잎끝, 잎 밑, 잎 옆, 잎 맥 등을 관찰하고, 축적된 지식과 판별 경험을 활용	입력된 잎 이미지 데이터 특성과 특징을 분석
꽃	차례, 색, 수술 크기, 개화 시기 등을 관찰하고, 축적된 지식과 판별 경험을 활용	입력된 꽃 이미지 데이터 특성과 특징을 분석
열매	열매 색, 모양, 형태, 착과 시기 등을 관찰하고, 축적된 지식과 판별 경험을 활용	입력된 열매 이미지 데이터 특성과 특징을 분석



분석 방법	시각, 촉각, 후각, 미각을 사용하여 입체적으로 관찰하며 분석	입력된 데이터의 특성과 특징을 분석
판별 방법	분석 결과와 축적된 지식, 판별 경험을 종합하여 최종 판별	분석 결과를 기반으로 학습된 결과와 비교하여 최종 판별 결과 출력

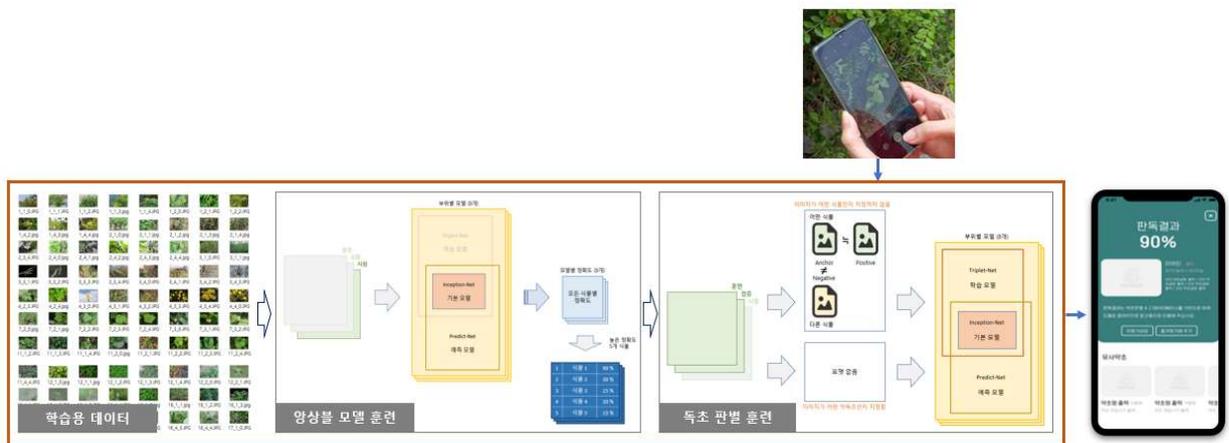
- 표에서 보듯이 사람은 촉각, 미각, 후각, 시각을 이용하여 식물을 자세히 살펴본 후 자신이 가지고 있는 지식과 경험을 종합하여 최종 독초 여부를 판단하지만,
- 인공지능은 입력된 식물 이미지 데이터만을 가지고 특징을 세분화→군집화→재분류 학습 과정을 거친 후 식물을 판별하는 학습모델 훈련을 통해 독초판별 능력을 높이기 때문에,
- 독초 60종과 독초와 유사한 식물 61종을 상호 비교 학습할 수 있도록 하여 정확한 독초 판별이 가능하도록 한국식물분류학회와 대한 본초학회 전문가 자문을 통해 데이터셋을 설계함.

2.2 임무 정의

“독초에 의한 중독사고 예방을 위한 독초판별 인공지능 모델 선정 및 알고리즘 개발”



- 독초를 약초로 잘못 알고 복용한 중독사고가 빈번하게 일어나고 있음, 식품의약품안전처 통계에 따르면 지난 10년(2010년~2019년)간 독초 중독으로 251명 사망.
- 이러한 독초 오인 사고 발생을 줄이고, 독초에 의한 중독사고를 예방을 위한 방법으로 인공지능 기술을 활용하여 약초와 독초를 구분하기 어려운 경우, 현장에서 스마트폰을 이용하여 실시간 독초와 약초 판별이 가능한 서비스를 개발 보급하여 국민의 건강과 생명을 위태롭게 하는 일이 없도록,
- 빈번하게 사고가 발생하는 독초를 중심으로 인공지능 학습데이터를 구축하고 구축된 학습 데이터를 활용하여,
- 독초를 정확하게 분류할 수 있는 인공지능 모델의 학습 및 훈련을 통해 신속한 독초판별이 가능한 알고리즘을 도출하고자 함.



[학습데이터 활용 핵심 임무]

2.4 데이터 구축 유의사항

2.4.1 촬영물에 대한 법적 동의 절차 준수 및 서면 동의 필수

- 촬영 이미지에 대한 저작권 사용 허락 동의 여부 확인 및 저작권 사용 허락 동의에 대한 고지 필수(사용기간, 데이터 공개 방법, 반려 처리된 이미지 활용)
- 촬영 이미지 데이터의 저작권 사용 허락 동의는 반드시 수행하여 데이터 공개 시 저작권 문제가 발생하지 않도록 함.
- 지적 저작권이 있는 식물의 경우는 지적 저작권 활용 동의와 사용기간, 사용 방법, 공개에 대한 동의 등, 관련 문서 작성 필수

2.4.2 독성 있는 식물의 경우 별도 관리

- 유사 비교식물과 선정 독초가 섞이지 않도록 데이터 분류 및 허브넷 플랫폼 등록 시 특히 주의
- 등록 데이터는 전문가(한의학, 식물 분류학)의 전수 검수를 필수 수행하여 독초와 다른 식물들이 섞이지 않도록 관리 철저.

2.4.3 데이터 촬영자 안전사고 예방 교육

- 자연에서 독초와 유사 비교식물을 촬영하게 되어 독충, 뱀 등에 물리지 않도록 촬영 전 주변을 꼼꼼하게 살핀 후 촬영에 임하도록 교육
- 산과 계곡 등 안전사고 발생 소지가 많은 곳에서는 특히 주의 당부
- 코로나-19로 인해 다중 이용 시설은 꼭 필요한 경우가 아니면 입장하지 않도록 하고, 정부의 방역 정책을 충실히 따를 것을 교육
- 특히 코로나-19 확진자와 접촉이 있을 경우는 관리 담당자에게 즉시 통보하여 자가격리 등 빠르게 대응할 수 있도록 교육 철저
- 안전사고 발생 시 담당자에게 즉시 연락을 취할 수 있도록 담당자 연락처 반드시 숙지 교육

3. 데이터 수집·정제

3.1 원시데이터 선정

- 동의보감에 수록된 약초 중에서 실생활에 사용 빈도가 높고, 주변에서 흔히 발견되는 식물 위주로 대상 식물을 선정하였으며, 독초 중 흔히 혼용되어 중독사고를 일으키는 식물을 식물분류학자, 한의학 본초학 전문가의 회의를 통해 독초 60종, 유사 비교식물 61종을 선정하였으며,
- 이렇게 선정된 대상 식물들과 유사한 비교식물들을 비교하기 위해서 식물분류학적인 비교 검색표를 작성하여,
- 비교 검색표에는 모든 대상 식물과 비교식물의 꽃, 잎, 열매, 전초의 특징을 일목요연하게 정리하여 대상 식물과 비교식물을 쉽게 구분할 수 있도록 하였고,
- 이 비교 검색표는 촬영자들이 현장 촬영을 나가기 전에 촬영하려는 식물들의 서로 다른 점을 학습하는 데 기초자료로 활용하거나,
- 정제와 검수 과정에서 촬영된 식물들을 정확히 동정하고 구분하는 데 활용하고,
- AI Hub를 통해 식물분류학적 동정 기준을 메타데이터로 제공함으로써 데이터의 활용성을 높이고자 하였음.

3.1.1 원시데이터 수집 대상(121종)

동의보감 독초 60종					유사 비교식물 61종				해상도
번호	식물명	약초 명	독성	형태(포맷)	번호	식물명	독성	형태(포맷)	
1	가죽나무	저근백피	-	이미지(jpg)	1	참죽나무	-	이미지(jpg)	HD, FHD
2	가지	가자	-	이미지(jpg)	2	도깨비가지	○	이미지(jpg)	HD, FHD
3	갈대	노근	-	이미지(jpg)	3	달뿌리풀	-	이미지(jpg)	HD, FHD
					4	역새		이미지(jpg)	HD, FHD
4	감국	감국	-	이미지(jpg)	5	산국	-	이미지(jpg)	HD, FHD
5	개오동	자실	-	이미지(jpg)	6	꽃개오동	-	이미지(jpg)	HD, FHD
6	곰취	호로칠	-	이미지(jpg)	7	참취	○	이미지(jpg)	HD, FHD
7	괭이밥	초장초	-	이미지(jpg)	8	선괭이밥	-	이미지(jpg)	HD, FHD
8	까마중	용규	○	이미지(jpg)	9	미국까마중	○	이미지(jpg)	HD, FHD
9	꿀풀	하고초	-	이미지(jpg)		배초향	-	이미지(jpg)	HD, FHD
10	넉줄고사리	골쇄보	-	이미지(jpg)	10	황고사리	-	이미지(jpg)	HD, FHD
11	능소화	능소화	○	이미지(jpg)	11	미국능소화	○	이미지(jpg)	HD, FHD
12	단삼	단삼	-	이미지(jpg)	12	깨꽃(살비아)	-	이미지(jpg)	HD, FHD
13	담쟁이덩굴	지금	-	이미지(jpg)	13	미국담쟁이덩굴	-	이미지(jpg)	HD, FHD

14	대추나무	대조	-	이미지(jpg)		뿔대추나무	-	이미지(jpg)	HD, FHD
15	돼지감자	똥단지	-	이미지(jpg)	14	해바라기	-	이미지(jpg)	HD, FHD
					15	원추천인국		이미지(jpg)	HD, FHD
16	띠	모근	-	이미지(jpg)	16	물억새	-	이미지(jpg)	HD, FHD
					17	수크령		이미지(jpg)	HD, FHD
17	마디풀	편축	-	이미지(jpg)	18	매듭풀	-	이미지(jpg)	HD, FHD
18	마타리	패장초	-	이미지(jpg)	19	뚝갈	-	이미지(jpg)	HD, FHD
19	메밀	교맥	-	이미지(jpg)	20	서양등골나물	-	이미지(jpg)	HD, FHD
20	뿔대추나무	산조인	-	이미지(jpg)		대추나무	-	이미지(jpg)	HD, FHD
21	무화과나무	무화과	-	이미지(jpg)	21	천선과나무	-	이미지(jpg)	HD, FHD
22	미나리	수근	-	이미지(jpg)	22	독미나리	○	이미지(jpg)	HD, FHD
					23	뿔미나리		이미지(jpg)	HD, FHD
23	밤나무	건울	-	이미지(jpg)		-	-	이미지(jpg)	HD, FHD
24	벌등골나물	패란	-	이미지(jpg)	24	등골나물	-	이미지(jpg)	HD, FHD
					25	골등골나물		이미지(jpg)	HD, FHD
25	범부채	사간	○	이미지(jpg)	26	대청부채	○	이미지(jpg)	HD, FHD
26	봉선화	급성자	-	이미지(jpg)	27	물봉선	-	이미지(jpg)	HD, FHD
27	부처손	권백	-	이미지(jpg)	28	개부처손	-	이미지(jpg)	HD, FHD
28	붉나무	염부목	-	이미지(jpg)	29	개웃나무	○	이미지(jpg)	HD, FHD
29	새삼	토사자	-	이미지(jpg)	30	미국실새삼	-	이미지(jpg)	HD, FHD
30	석위	석위	-	이미지(jpg)	31	세뿔석위	-	이미지(jpg)	HD, FHD
31	수박풀	야서과묘	-	이미지(jpg)	32	수박	-	이미지(jpg)	HD, FHD
32	순비기나무	만형자	-	이미지(jpg)		-	-	이미지(jpg)	HD, FHD
33	쉴싸리	택란	-	이미지(jpg)	33	석잠풀	-	이미지(jpg)	HD, FHD
34	시호	시호	-	이미지(jpg)	34	개시호	-	이미지(jpg)	HD, FHD
35	아욱	규채	-	이미지(jpg)	35	난쟁리아욱,	-	이미지(jpg)	HD, FHD
36	어저귀	경마자	-	이미지(jpg)	36	닥풀	○	이미지(jpg)	HD, FHD
37	여우오줌	향유관	○	이미지(jpg)	37	좀담배풀	○	이미지(jpg)	HD, FHD
					38	담배풀		이미지(jpg)	HD, FHD
38	여주	고과	-	이미지(jpg)	39	오이	-	이미지(jpg)	HD, FHD
					40	수세미오이		이미지(jpg)	HD, FHD

39	연꽃	연화	-	이미지(jpg)	41	수련	-	이미지(jpg)	HD, FHD
40	오갈피나무	오가피	-	이미지(jpg)	42	가시오갈피	-	이미지(jpg)	HD, FHD
41	오동나무	동엽	-	이미지(jpg)	43	벽오동	-	이미지(jpg)	HD, FHD
42	오이풀	지유	-	이미지(jpg)	44	산오이풀	-	이미지(jpg)	HD, FHD
43	용담	용담	-	이미지(jpg)	45	과남풀(큰용담)	-	이미지(jpg)	HD, FHD
44	잇꽃	홍화	-	이미지(jpg)		삼주	-	이미지(jpg)	HD, FHD
45	자귀나무	합환화	-	이미지(jpg)	46	왕자귀나무	-	이미지(jpg)	HD, FHD
46	장구채	왕불유행	-	이미지(jpg)	47	가는장구채	-	이미지(jpg)	HD, FHD
					48	분홍장구채		이미지(jpg)	HD, FHD
47	접시꽃	촉규화	-	이미지(jpg)	49	당아욱	-	이미지(jpg)	HD, FHD
48	제비쑥	모호	-	이미지(jpg)	50	맑은대쑥	-	이미지(jpg)	HD, FHD
49	쥐방울덩굴	마두령	○	이미지(jpg)		등침	○	이미지(jpg)	HD, FHD
50	진득찰	희렴	-	이미지(jpg)	51	털진득찰	-	이미지(jpg)	HD, FHD
51	짚신나물	용아초	-	이미지(jpg)	52	큰땀무	-	이미지(jpg)	HD, FHD
52	쪽	람	-	이미지(jpg)	53	여뀌	○	이미지(jpg)	HD, FHD
					54	개여뀌		이미지(jpg)	HD, FHD
53	큰평의비름	경천	-	이미지(jpg)	55	등근잎평의비름	-	이미지(jpg)	HD, FHD
54	패랭이꽃	우자	-	이미지(jpg)	56	술패랭이꽃	-	이미지(jpg)	HD, FHD
55	피마자	피마자	○	이미지(jpg)		-	-	이미지(jpg)	HD, FHD
56	완두콩	완두	-	이미지(jpg)	57	동부	-	이미지(jpg)	HD, FHD
57	하늘타리	과루인	-	이미지(jpg)	58	가시박	○	이미지(jpg)	HD, FHD
58	한련초	한련초	-	이미지(jpg)	59	털별꽃아재비	-	이미지(jpg)	HD, FHD
59	호장근	호장근	-	이미지(jpg)	60	감절대	-	이미지(jpg)	HD, FHD
60	회화나무	괴화	-	이미지(jpg)	61	아까시나무	-	이미지(jpg)	HD, FHD

- ※ 독초 60종과 유사 비교식물 61종 중 계절적 요인으로 촬영이 어려운 경우, 대상을 변경할 수 있음
- ※ 꽃, 열매, 잎 부위가 없는 품목의 경우, 다른 부위를 추가 촬영하여 품목별 목표인 5,000장 이상을 달성 예정

3.1.2 선정 원시데이터의 특성 및 정의

- 동의보감 독초와 유사 비교식물은 분류학적으로 식물계에 해당하고, 식물은 영양기관과 생식기관을 가지고 있으며
- 뿌리, 줄기, 잎은 식물의 생식과는 관계없이 성장하고 지지하고 유지하는 역할을 담당하는 기관으로 뿌리는 땅속에서 식물체를 지지하며 양분 및 수분을 흡수하고, 줄기는 식물의 뿌리와 다른 기관을 연결하며 식물체를 지지하고 양분 및 수분을 이동시키는 역할을 하며, 잎은 광합성을 통하여 양분을 만들고 호흡 및 증산작용을 하게 되는데 이러한 기관을 영양기관으로 통칭. (출처, 산림청 용어해설)
- 종자식물에서의 꽃이나 열매, 선태식물의 배우자낭이나 포자처럼 후대를 만들어 가는데 관련된 기관. 식물계통분류에 있어서 가장 중요한 특징이 되며, 주요 기관의 하나임.
- 식물의 주요 기관인 잎, 꽃, 열매 부위와 식물 생육단계별 모양과 형태 변화를 관찰 가능한 전초(식물의 지상부) 부위를 나누어 촬영한다.

전초	잎	꽃	열매
			
형식(이미지, jpg)	형식(이미지, jpg)	형식(이미지, jpg)	형식(이미지, jpg)
해상도(HD, FHD)	해상도(HD, FHD)	해상도(HD, FHD)	해상도(HD, FHD)
개체당 24장 이내 촬영	개체당 24장 이내 촬영	개체당 24장 이내 촬영	개체당 24장 이내 촬영

◦ 데이터 분류 및 부위별 정의

대상 품목 121종을 데이터를 식물이나 약초 전문가가 아닌 일반인의 시각을 기준으로 재 분류하여, 데이터의 보편적 편리성을 보장하고자 함

가. 전초 정의 및 분류

- 전초는 통상적으로 식물의 땅 위의 부분 전체를 지칭하지만, 한의학에서는 뿌리를 포함한 모든 부위를 의미한다.
- 본 과제에서는 전초에 대한 정의를 식물 분류학 또는 한의학적 관점이 아닌 순수 인공지능 분야의 관점에서 데이터의 의미를 명확하게 함으로써 데이터 활용 가치를 높이고 데이

터 활용, 촬영 시 혼란을 최소화하고자 한다.

- 식물의 형태에 따라 전초를 **목본, 초본, 덩굴**로 재분류한다.

◦ 전초 분류 예시



1) 목본

◦ 목본 정의

- 목본은 겨울에 지상부가 살아 있으며, 식물 분류학적 기준에 따르면 줄기나 가지가 목질 형질을 가진 식물을 뜻하며, 통상 어린 묘목을 제외하고는 수십 미터 이상 자라는 교목과 그보다는 작게 자라는 관목으로 분류가 된다.
- 사람의 키보다 큰 나무의 경우, 전체 이미지를 확보하기 위해 상당히 먼 거리에서 촬영하게 되므로 나무의 형태는 육안 구분이 가능하지만, 인공지능 학습용 데이터 관점에서는 식물을 판별할 수 없는 데이터가 될 수 있어, 큰 줄기 일부에 잎, 꽃, 열매 등이 포함된 경우, 전초로 인정한다.
- 본 과제에서는 나무의 온전한 모습을 촬영하기보다 나무줄기에 잎, 꽃, 열매 등이 포함된 경우를 전초 범위에 포함하여 나무 모양보다 나무에 잎이 나는 형식, 열매를 맺는 모양, 꽃이 나는 모양 등을 확인할 수 있는 이미지를 데이터로 정의하고자 한다.

◦ 목본(예시)



2) 초본

◦ 초본 정의

- 초본은 겨울에 지상부가 고사하며, 식물학적 분류에 따르면 지상부에 목본 줄기를 가지고 있지 않은 식물이다. 꽃이 피고 열매를 맺으면 생장이 멈추고 식물 전체가 죽는 일년생 식물과 겨울이나 건기에 지상 부위는 죽고, 뿌리나 줄기가 살아남아 다음 해에 이로부터 새순이 솟아올라 자라는 다년생 식물로 구분된다.
- 본 과제에서는 식물의 지상의 모습 중 잎이 나는 형식, 꽃이 피는 모양, 열매 맺는 모양을 관찰할 수 있는 경우 전초 중 초본으로 분류한다.

◦ 초본(예시)



3) 덩굴

◦ 덩굴 정의

- 덩굴은 식물학적 분류에서는 줄기가 가늘고 길게 자라 다른 식물이나, 물체에 걸치거나 감겨서 생활하는 식물로 잎, 꽃차례(화서) 등의 일부가 변형되어 다른 물체를 감으며 생장하는 식물이다. 덩굴 중 목질로 이루어진 등칠편, 등나무 등은 만경식물이라 한다.
- 본 과제에서는 덩굴 식물은 덩굴줄기에서 잎이 나는 모양, 꽃이 피는 모양, 열매가 맺는 모양 등을 확인할 수 있는 이미지 데이터인 경우도, 덩굴로 인정하여 분류한다.

◦ 덩굴(예시)



나. 잎의 정의 및 분류

- 잎(葉, Leaf)은 광합성과 증산작용이 일어나는 식물의 기관으로 잎몸, 잎자루, 턱잎으로 이루어져 있으며, 홑잎과 겹잎으로 나뉜다.
- 식물분류학적 잎의 종류

홑잎		겹잎	
	홑잎(단엽, 單葉, simple leaf)은 잎자루에 잎몸이 하나로 되어 있음		겹잎(복엽, 複葉, compound leaf)은 잎몸이 2개 이상으로 되어 있음

• 겹잎의 종류(8가지)

겹잎 종류	겹잎 모양	겹잎 종류	모양
장상복엽		삼출 복엽	
5출복엽		우상 복엽 (잎 수가 짝수)	 
기수 우상 복엽 (잎 수가 홀수)		2회 3출 복엽	
3회 3출 복엽		조족상 복엽	

- 본 과제에서는 식물 분류학 또는 한의학적인 기준이 아닌 일반인의 사용자 시각을 기준으로 식물의 잎을 재분류함으로써 데이터 사용에 있어 혼란을 방지하고 명확하게 설명 가능한 데이터가 될 수 있도록 잎을 분류하고,
- 잎이 나는 형태적으로 **한 잎, 긴 잎, 여러 잎**으로 재분류한다.
- **잎 분류 방법**



1) 한 잎

◦ 한 잎 정의

- 일반적으로 식물의 줄기에서 잎자루에 잎이 하나씩 나는 형태로 대부분의 식물 잎이 한 잎에 해당하며, 잎이 나는 모양은 어긋나기, 마주나기, 돌려나기 등의 특징을 가지고 있으며
- 한 잎을 촬영할 때는 잎 한 장씩 촬영하거나, 여러 장의 잎을 한꺼번에 촬영하여 줄기에서 잎이 나는 형태, 잎의 성장 모습을 확인할 수 있도록 하며,
- 한 잎 모양 확인 가능한 데이터는 한 잎으로 포함한다.
- 촬영 시 가장 유의할 점은 한 잎을 수평으로 촬영할 경우, 침형(바늘 모양 잎)으로 오인될 소지가 있고, 잎의 형태(잎 옆, 잎 밑, 잎끝, 잎 맥) 확인이 구분되지 않아 인공지능 학습데이터로 활용이 어려움으로 모든 잎은 수평 촬영하지 않는다.

◦ 한 잎(예시)



2) 긴 잎

◦ 긴 잎 정의

- 대상 품목 중 잎이 선형으로 길게 나는 형태를 가진 잎을 긴 잎으로 정의한다.
- 긴 잎은 형태적으로 긴 선형으로 잎이 길게 자라며, 자랄 때는 다른 개체가 섞이지 않고 대부분 같은 종끼리 뭉쳐서 자란다.
- 긴 잎의 가장 큰 특징은 잎과 줄기 구분이 어렵거나, 잎 자체가 줄기인 경우가 많다.
- 따라서 데이터 획득 시 여러 개체가 모여서 나는 긴 잎으로 분류된 품목은 이러한 잎의

특징이 반영되도록 하나의 개체만 촬영하지 않고, 여러 개체를 묶어서 촬영하는 것을 권장한다.

◦ 긴 잎(예시)



3) 여러 잎

◦ 여러 잎 정의

- 여러 잎은 식물 분류학적 기준에는 겹잎에 해당하며, 하나의 잎자루에 2개 이상의 작은 잎이 붙어 나는 것이 특징이며,
- 겹잎을 세분화하여 장상 겹잎, 3 출 겹잎, 우수 우상 복엽 등 식물 분류학적 동정 기준을 적용하지 않고, 모든 겹잎을 여러 잎 하나로 묶음으로써 식물 분류학적 지식이 없는 일반인도 데이터 공개 후 데이터 활용이 쉽도록 여러 잎으로 재분류함.
- 여러 잎 촬영 시 잎을 이루는 개별 작은 잎이 일부 떨어지거나, 벌레가 먹거나, 훼손되었더라도 여러 잎 특징을 나타낸 경우는 여러 잎에 포함한다.

◦ 여러 잎(예시)



다. 꽃 정의 및 분류

◦ 꽃은 속씨식물의 생식기관으로 생식에 관계되는 부분과 관계되지 않는 부분이 모여 있고, 식물을 구별하는 요소 중 하나 임.

• 식물분류학적 꽃차례(花序) 종류

꽃차례	꽃차례 정의
총상	길게 자라나는 꽃대에 꽃자루 있는 꽃들이 달리는 화서
수상	가늘고 긴 꽃대에 꽃자루 없는 꽃들이 이삭 모양으로 촘촘히 붙어 우산 모양의 화서
미상	가늘고 긴 꽃대에 꽃자루 없는 단성화들이 꼬리처럼 아래로 늘어져 달리는 화서
산방	여러 개의 분지점에서 나온 꽃들이 비슷한 높이에 모여 달리는 화서
산형	꽃자루 있는 꽃들이 꽃대 끝 지점에 방사상으로 모여 달리는 화서
두상	꽃자루 없는 꽃들이 넓적한 화탁에 조밀하게 붙어 머리 모양을 이루는 화서
윤산	많은 꽃이 줄기를 둘러싸고 피어나는 꽃차례, 꽃줄기가 없이 꽃이 원줄기에 층을 이루어 바퀴처럼 둥글게 우산 모양으로 펴
육수	육질이 두툼한 꽃대에 꽃자루 없는 꽃들이 조밀하게 모인 화서
취산	맨 위나 안쪽의 꽃이 먼저 피고 그 아래쪽 가지나 곁가지의 꽃들이 피는 화서
배상	암꽃과 퇴화된 수꽃을 둘러싼 몇 개의 소포엽이 술잔 모양을 이루는 화서
권산	꽃들이 한쪽 방향으로 태엽이나 나선 모양으로 말려 있는 화서
은두	주머니처럼 오목한 화탁의 안쪽에 꽃들이 배열돼 밖에서는 보이지 않는 화서
원추	단일화서가 여러 개로 갈라져 전체적으로 원뿔 모양을 이루는 복합화서

◦ 본 과제에서는 꽃차례를 식물분류학적 기준이 아닌 전문 지식이 없는 일반인들이 꽃을 보았을 때를 기준으로 꽃의 형태 또는 꽃이 나는 모양을 기준으로,

◦ 꽃이 나는 모양에 따라 **홀로 피기, 모여 피기, 차례 피기** 3가지로 재분류하여 정의함

◦ 꽃 분류 방법



1) 홀로 피기

◦ 홀로 피기 정의

- 본 과제에서는 일반인의 시각을 기준으로 꽃이 나는 모양이 독립적인 형태를 띠는 경우를 홀로 피기로 정의하였다.
- 독립적인 형태를 띠는 홀로 피기 범위에는 무리를 이루어 꽃이 피는 경우의 꽃이라도 개별적으로 꽃을 구분할 수 있는 경우도 홀로 피기에 포함하며,
- 홀로 피기 꽃은 크기가 큰 경우가 많아 다른 식물 부위와 겹치거나, 꽃들끼리 겹쳐 피는 경우는 개별적으로 촬영하기보다는 홀로 피기 꽃이 겹쳐진 상태로 촬영하여도 됨

◦ 홀로 피기(예시)



2) 모여 피기

◦ 모여 피기 정의

- 작은 꽃 여러 개가 모여 하나의 꽃을 이루는 모양의 경우 모여 피기로 정의하였다.
- 모여 피기 꽃은 작은 꽃들이 모여 하나의 꽃 모양을 이루는 경우가 많고, 식물 분류학적으로는 산형인 경우가 많아,
- 모여 피기 꽃은 작은 꽃들이 모여 있어 촬영 시 모여 피기를 이루는 작은 꽃을 지나치게 확대하여 촬영하는 것을 지양한다.

◦ 모여 피기(예시)



3) 차례 피기

◦ 차례 피기 정의

- 일반인의 시각에서 보면 긴 줄기에 꽃들이 순서대로 피는 모양인 경우를 **차례 피기**로 정의하였다.
- 차례 피기 꽃은 여러 개의 작은 꽃이 긴 줄기에 모여 피는 모양이며, 개화 방법도 아래서 위로 또는 위에서 아래의 차례대로 핀다.
- 차례 피기 꽃 촬영 시 꽃망울과 만개한 꽃이 한 줄기에 있는 경우가 많기에 시간을 두고 여러 번에 나누어 촬영한다.

◦ 차례 피기(예시)



라. 열매 정의 및 분류

1) 식물분류학적 열매 종류와 분류

◦ 열매는 식물의 기관의 하나이다. 수정된 씨방이 발달해서 생긴 것으로, 씨방만으로 생긴 열매를 **참 열매**, 꽃받침 등과 함께 발달한 열매를 **헛열매**라 부르며, 열매는 속씨식물이 수정한 후, 씨방 또는 그 주변의 기관이 자라 이루어진다.

◦ **홀열매와 겹열매(출처, 위키백과)**

홀열매	홀열매(단화과)는 단과라고도 하는데, 한 꽃의 1개 씨방에서 생긴 열매로 가장 흔히 볼 수 있는 형태이다. 열매껍질의 성질에 따라 '건과'와 '액과'로 구분된다. 건과는 성숙 후, 열매껍질이 목질·혁질(革質) 또는 피질(皮質)이 되는 열매이다. 이것은 열매가 성숙한 후 열매껍질이 벌어지는가(裂開), 벌어지지 않는가에 따라 폐과(閉果)와 열 개과(裂開果)로 다시 구분된다.
겹열매	겹열매(복화과, 다화과)는 여러 개의 꽃이 꽃차례를 이룬 채 성숙하여 한 개의 열매처럼 생긴 것으로 오디, 무화과 등이 있다.

◦ **폐과(열매껍질이 터지지 않는 열매)**

구분	열매 정의
건과	각두과(殼斗科)라고도 한다. 과피가 목질 또는 혁질로 그 속에 1개의 씨가 있는데, 여기서는 일반적으로 과피를 씨껍질(종피, 種皮)이라 부른다. 떡갈나무·줄참나무류의 열매 및 밤 등이 이에 속한다
수과	과피는 얇은 막질로 익으면 건조해지며 보통 하나의 방 속에는 1개의 씨가 들어 있는데, 각 방의 크기가 작기 때문에 언뜻 보면 씨로 착각하기 쉽다. 씨방 하위인 꽃에서 볼 수 있으며 꽃받침 또는 꽃대가 열매의 일부를 구성한다. 마디풀과·명아주과·국화과 등에서 볼 수 있다.
영과	과피는 얇은 목질이거나 피질이며 씨껍질과 꼭 붙어 있어, 언뜻 보면 열매가 씨처럼 보인다. 벼과의 열매(현미)가 이에 속한다.
시과	'익과(翼果)'라고도 한다. 건과의 일종으로, 과피의 한쪽 끝이 자라 날개 모양으로 되어 있다. 단풍나무과·느릅나무과·자작나무과에서 볼 수 있다.

◦ **열개과(열매껍질이 터지는-벌어지는 열매)**

구분	열매 정의
골돌과	대과(蓇果)라고도 한다. 여러 개의 심피(씨방)를 가진 꽃(분립 씨방)에 만들어지는 열매로, 각 심피의 봉합선에 따라 벌어지게 된다. 목련과, 미나리아재비과 등 심피가 많은 종류에서 볼 수 있다.

꼬투리 열매(협과)	1심피성 씨방으로부터 만들어지며, 익어서 건조하면 봉선을 따라 벌어져 2 조각이 된다. 콩과의 열매가 이에 속한다.
장각과	단각'이라고도 한다. 장각과와 마찬가지로 겨자과에서 볼 수 있으며, 2심피성 씨방에서 생긴 열매로 편평하고 길이가 짧은 것을 말한다. 냉이의 열매가 이에 속한다.
단각과	단각'이라고도 한다. 장각과와 마찬가지로 겨자과에서 볼 수 있으며, 2심피성 씨방에서 생긴 열매로 편평하고 길이가 짧은 것을 말한다. 냉이의 열매가 이에 속한다.
삭과(씨열매)	2개 이상의 심피로 이루어진 씨방이 성숙하여 된 열매로, 많은 방으로 되어 있다. 익으면 심피와 같은 수의 열편(裂片)으로 벌어져 씨가 나온다. 백합·붓꽃·제비·봉선화에서 볼 수 있다.
개과	삭과의 하나로, 익으면 열매가 가로로 벌어져 위쪽이 뚜껑같이 되고 아래쪽은 주발 모양으로 남는다. 채송화·질경이·비름속(屬)·쇠비름·뚜껑별꽃·사리풀 등에서 볼 수 있다.
공개삭과	'공과' 또는 '공삭'이라고도 한다. 삭과의 하나로 열매의 끝에 1개 또는 여러 개로 벌어진 구멍이 있어, 그 곳으로부터 씨가 나온다. 금어초·양귀비의 열매가 이에 속한다.
분리과	'분열과'라고도 한다. 다심피성의 씨방이 성숙한 후에 중축을 남기고 분리한 각 방이 각각 작은 열매가 되는 열매이다. 이때의 작은 열매를 '분과'라고 하는데, 각각 1개의 심피로 이루어져 있다. 미나리과, 차조기과, 아욱과 등의 열매에서 볼 수 있다.

◦ 핵과

'장과' 또는 '육과(肉果)'라고도 한다. 과피는 3층으로 되어 있으며, 종과피나 내과피가 다소 두꺼운 육질이므로 수분이 많고 부드러우며 익어도 벌어지지 않는다.

구분	열매 정의
핵과	외과피는 피질, 종과피는 수분이 많은 다육질이며, 내과피는 매우 단단한 목질이다. 핵과 중에서 1심피성의 씨방으로 되어 있는 예가 매실·복숭아·살구·앵두 등이며, 여러 개의 심피로 이루어진 다수의 핵을 가진 것들은 다정나무·산사나무 등의 열매이다.
액과	액과의 좁은 의미로 사용하고 있다. 종과피, 내과피 모두 수분이 많은 다육질이며, 단단한 씨를 가진다. 포도·토마토가 그 예이다.
호과	외과피가 다소 부드럽고 종과피, 내과피는 수분이 많은 두꺼운 다육질이다.

	수박·참외·오이 등이 그 예이다.
이과	씨방 이외에 꽃턱이나 꽃받침의 밑부분이 다육질로 되어 씨방을 덮어 이루어진 열매로, 헛열매의 한 형이기도 하다. 배가 이에 속한다.

◦ 겹열매

'복과' 또는 '다화과'라고도 하며, 수상화서가 밀집한 꽃의 씨방이 성숙하여 전체가 하나의 열매처럼 된 것을 이른다.

구분	열매 정의
장미과	헛열매의 한 형이다. 종지 모양으로 비대한 꽃턱의 바닥에 다수의 건조과가 생긴다. 언뜻 보아 과피처럼 보이는 부분은 꽃받침이 발달한 것이며, 내부 건조과의 외피가 씨방에서 유래된 참과피이다. 장미·비파나무 등에서 볼 수 있다.
딸기 모양 열매	1개의 꽃에 생긴 다수의 심피가 익어서 된 복화과의 하나로, 꽃턱이 머리 모양으로 비대해져 열매처럼 된다. 참열매는 그 위에 묻혀 있다. 딸기 등이 이에 속한다.
나무딸기 모양 열매	1개의 꽃에 밀생한 다수의 심피가 성숙하여 된 것으로, 마치 여러 개의 작은 액과가 모여 한 열매를 이룬 것처럼 보인다. 명석딸기·산딸기 등이 이에 속한다.
상실과	'육질 집합과'라고도 한다. 많은 꽃이 밀생한 꽃차례가 성숙하여 마치 한 열매 같이 보이는 수분이 많은 육질의 열매이다. 뽕나무 열매인 오디, 파인애플이 그 예이다.
은화과	'무화과'라고도 한다. 씨방이 큰 꽃턱 속에 형성되어 살이 많다. 무화과나무·천선과나무에서 볼 수 있다.

◦ 구과

겉씨식물인 침엽수류의 구화(毬花)가 수정하여 익은 것을 구과(毬果)라고 한다. 구화의 종린(種鱗)은 대포자엽, 즉 심피라고 할 수 있으므로, 그것이 익은 구과를 복화과라고 할 수 있을 것이다. 또한, 심피 외에 포린(包鱗)도 포함되어 있으므로 헛열매로 소나무 등이 이에 속한다. 겉씨식물의 꽃은 씨방이 없으므로 참열매는 없다. 일반적으로 대부분의 씨는 노출되어 있으나, 특이한 예로 주목과의 식물에서는 씨가 다육·다즙질의 씨껍질로 싸여 액과 모양으로 된 헛열매를 만든다.

2) 본 과제에서 정의한 열매 분류 체계

- 본 과제에서는 식물 분류학적 전문 지식이 없는 일반인이 열매를 보았을 때를 기준으로 열매에 대한 분류를 새롭게 정의하였다.
- 새로운 열매 분류 방식은 **홀로 맺기**, **모여 맺기**, **차례 맺기** 3가지로 나누었으며 이는 열매 나는 모양을 일반인이 바라봤을 때를 기준으로 하여 열매를 정의하였다.
- 이는 식물 분류학적 지식이 없는 사람도 쉽게 데이터를 분류하고 관리할 수 있도록 하기 위함이며, 데이터 공개 후 일반인이 데이터 활용과 이용에 편리성을 보장하기 위함이다.

◦ 열매 분류 방법



3) 홀로 맺기

◦ 홀로 맺기 정의

- 홀로 맺기는 일반인이 열매를 바라봤을 때 열매가 독립적으로 맺은 경우를 **홀로 맺기로 정의한다**.
- 홀로 맺기 열매는 열매가 맺기 시작하여 성숙한 열매로 생육이 진행되므로 시기를 두고 촬영해서 홀로 맺는 시기별 열매 모양을 촬영하는 것이 바람직하다.
- 홀로 맺기 열매는 독립적으로 성장하기에 미성숙 열매부터 성숙한 열매를 개별적으로 촬영하되 축소 또는 확대촬영 않는 것을 권장한다.

◦ 홀로 맺기(예시)



4) 모여 맺기

◦ 모여 맺기 정의

- 꽃 분류 방법에서 모여 피기에 해당하는 꽃의 경우 꽃이 진 자리에 열매가 맺히며 여러 개의 열매가 하나의 큰 모양을 이루는 경우를 **모여 맺기로 정의하였다.**
- 모여 맺기 열매는 작은 열매 여러 개가 모여 하나의 형태를 이루므로 작은 열매를 확대 촬영하지 않고 모여 있는 모양이 잘 나타나도록 촬영한다.
- 작은 열매를 확대하여 촬영할 경우, 데이터 왜곡이 발생 될 수 있다.

◦ 모여 맺기(예시)



5) 차례 맺기

◦ 차례 맺기 정의

- 차례 맺기는 한 줄기에 여러 개의 열매가 줄줄이 맺힌 모양을 **차례 맺기로 정의하였다**
- 차례 맺기 열매는 하나의 줄기에 작은 열매가 줄줄이 맺히는 특징을 가지고 있어 차례 맺기 열매 하나에서 열매가 맺는 **모양을 동시 관찰이 가능하다.**
- 차례 맺기 열매 촬영 시 열매가 맺는 모양을 확인할 수 있도록 시기를 두고 여러 번에 나누어 촬영하는 것이 좋다.

◦ 차례 맺기(예시)



3.2 원시 데이터 수집·정제 절차 및 방법

3.2.1 원시 데이터 수집 절차

가. 원시 데이터 수집 절차



1) 원시 데이터 수집 활용 자료

촬영 매뉴얼	구분 가이드	특징 비교표
	<p>촬영</p> <ul style="list-style-type: none"> • 데이터 수집 담당자가 대상 식물 탐색 시 활용 • 자연에서 대상 약초 및 유사 식물 구분 시 활용 • 2021년도 대상 약초 및 유사식물에 대한 기본 학습 자료로 활용 • 촬영 시 대상 약초 특징 요소를 부각 시킬 수 있도록 촬영 계획 수립 시 활용 	

식물분류학자, 한의학 본초 전문가의 자문회의를 통해 중독사고 발생 식물과 생활 주변에서 손쉽게 접할 수 있는 독초 60종과 독초와 모양 형태가 유사한 비교식물 61종을 선정하고 사업 기간 내 촬영이 가능한 품목을 추려 최종 대상으로 확정함

나. 수집 대상 및 방법

1) 수집 대상

- 식물분류학자, 한의학 본초 전문가의 자문회의를 통해 중독사고 발생 식물과 생활 주변에서 손쉽게 접할 수 있는 독초 60종과 독초와 모양 형태가 유사한 비교식물 61종을 선정하고 사업 기간 내 촬영이 가능한 품목을 추려 최종 대상으로 확정함

· 독초(60종)

가죽나무, 가지, 갈대, 감국, 개오동, 곰취, 팽이밥, **까마중**, 꿀풀, 넉줄고사리, 능소화, 단삼, 담쟁이덩굴, 대추나무, 돼지감자, 띠, 마디풀, 마타리, 메밀, 뿔대추나무, 무화과나무, 미나리, 밤나무, 벌등골나물, **범부채**, 봉선화, 부처손, **붉나무**, 새삼, 석위, 수박풀, 순비기나무, 씩사리, 시호, 아욱, 어저귀, **여우오줌**, 여주, 연꽃, 오갈피나무, 오동나무(참오동), 오이풀, 용담, 잇꽃, 자귀나무, 장구채, 접시꽃, 제비쑥, **쥐방울덩굴**, **진득찰**, 짚신나물, 쪽, 큰평의비름, 패랭이꽃, **피마자**, 하늘타리, 한련초, 호장근, **회화나무**

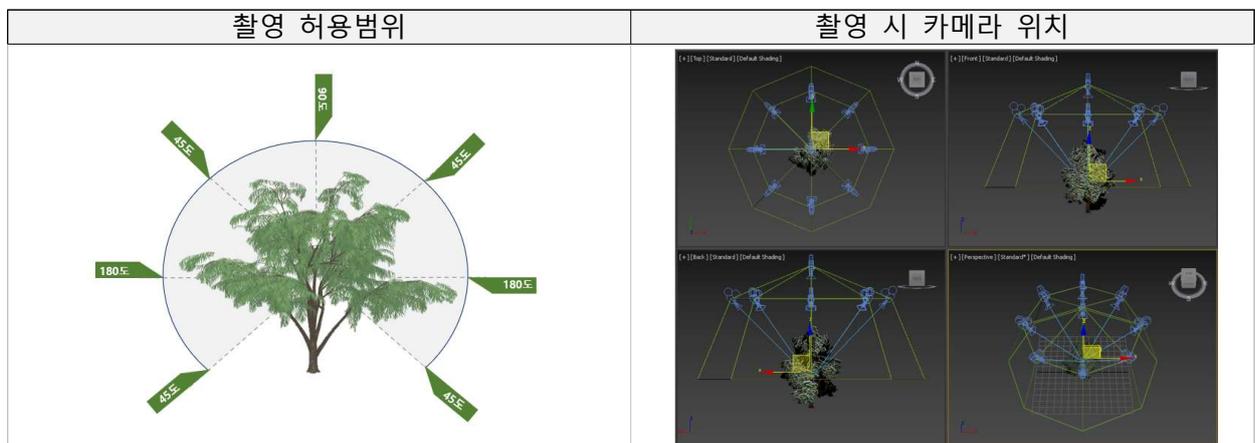
· 유사 비교식물(61종)

참죽나무, **도깨비가지**, 달뿌리풀, 억새, 산국, 꽃개오동, **참취**, 산팽이밥, **미국까마중**, 황고사리, **미국능소화**, 깨꽃(살비아), 미국담쟁이덩굴, 해바라기, 원추천인국, 물억새, 수크령, 매듭풀, 뚝갈, 서양등골나물, 천선과나무, **독미나리**, 뿔미나리, **등골나물**, **골등골나물**, **대청부채**, 물봉선, **개웃나무**, 미국실새삼, 세뿔석위, 수박, 석잠풀, 개시호, 난쟁이아욱, **닥풀**, **좁담배풀**, **담배풀**, 오이, 수세미아욱, 수련, **가시오갈피**, 벽오동, 산오이풀, 과남풀(큰용담), 삼주, 왕자귀나무, 가는장구채, **분홍장구채**, 당아욱, 맑은대쑥, **털진득찰**, 큰뻬무, 여뀌, 둥근잎평의비름, 술패랭이꽃, **가시박**, 감절대, **아까시나무**

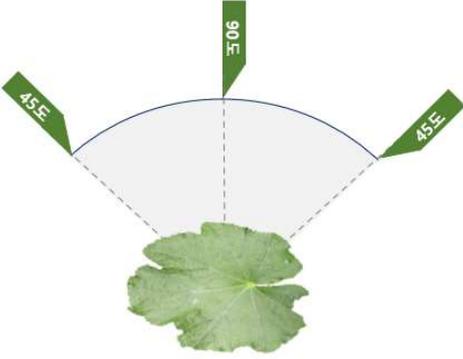
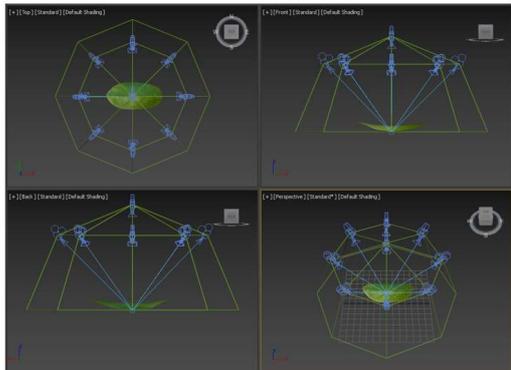
2) 수집 방법

- 이미지 데이터 획득은 식물의 판별 특징이 포함되도록 식물을 전초, 잎, 꽃, 열매 4 부위로 나누고, 획득 데이터가 입체적으로 구성될 수 있도록 대상 부위를 바라보는 면[수평면(180°), 수직면(90°), 상사면(45°), 하사면(45°)]을 기점으로 하여 회전(360°)하며 촬영

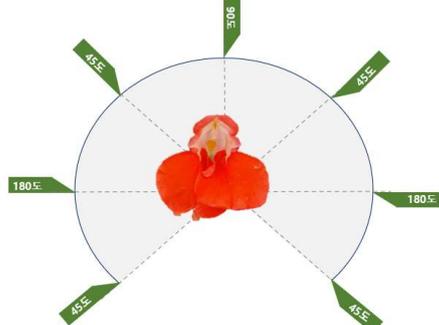
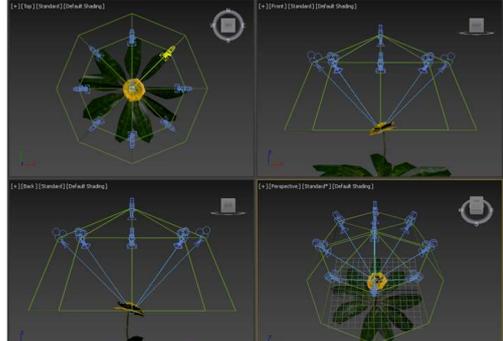
가) 전초 촬영 방법



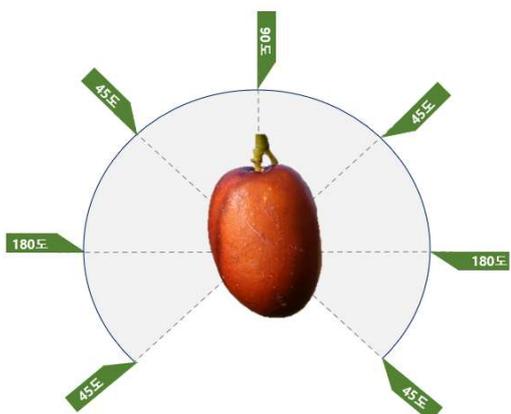
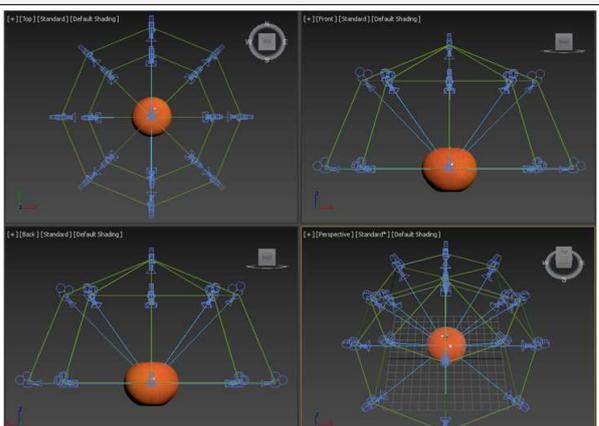
나) 잎 촬영 방법

촬영 허용범위	촬영 시 카메라 위치
	

다) 꽃 촬영 방법

촬영 허용범위	촬영 시 카메라 위치
	

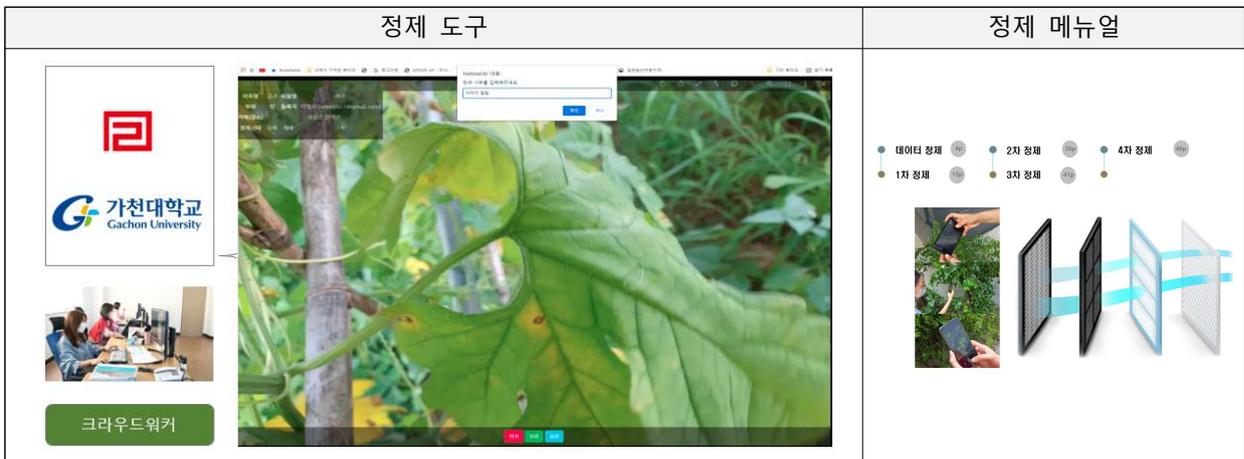
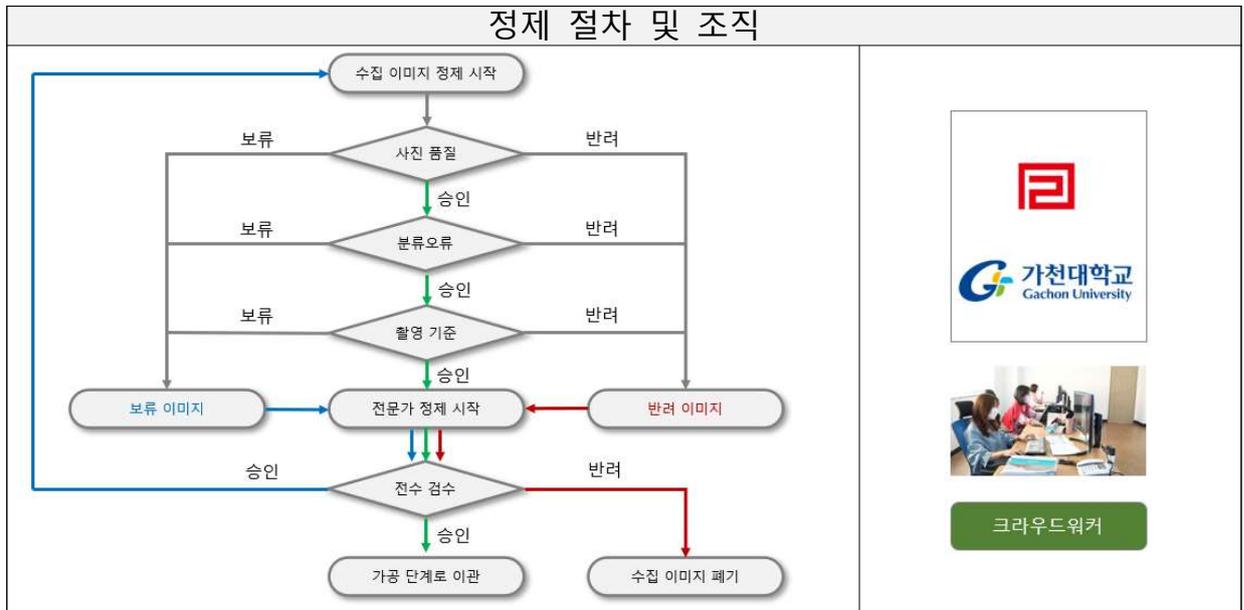
라) 열매 촬영 방법

촬영 허용범위	촬영 시 카메라 위치
	

3.2.2 원시데이터 정제 절차 및 방법

- 데이터 정제는 다양한 촬영 도구, 촬영 환경 및 촬영자에 의해 확보된 원시데이터 중 인공지능 학습용 데이터로서 데이터 품질기준에 부합되는 이미지 데이터를 엄선하는 과정임.

가. 정제 절차



나. 원시데이터 정제 방법

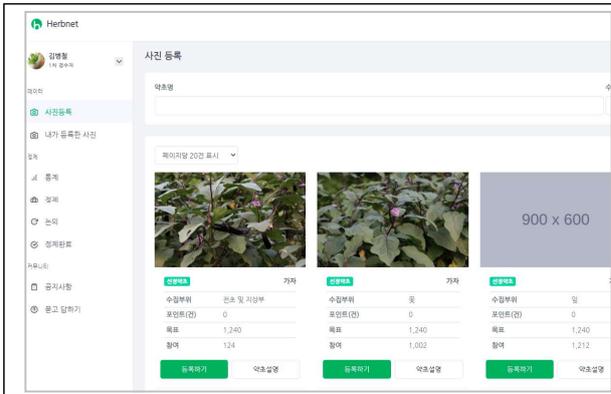
- 데이터 정제는 총 4단계의 과정을 거쳐 이루어지며
 - 촬영 이미지 데이터 품질기준(흔들림, 이미지 잘림 등)에 적합 여부
 - 촬영 데이터의 부위별 정확한 분류 여부
 - 촬영기준 준수 여부를 검증하고
 - 최종 전문가(식물분류학, 한의학)의 전수 검수 후 최종 정제 승인이 이루어짐.

단계	1단계	2단계	3단계	4단계
목적	원시 이미지 데이터 품질 검수	데이터 분류 오류 검수	촬영 기준 충족 여부 검수	정제 이미지 데이터 검수
정제	흔들림, 초점 흐림, 대상 부위 잘림, 지나친 대상 확대 또는 축소, 원거리 촬영, 다른 사물 개입(손 등), 그림자, 오염(이물질 등) 등의 이미지를 엄선하여 처리함	1차 정제된 데이터 중 분류가 잘못된 데이터를 바로잡는 과정으로 전초, 잎, 꽃, 열매 4부위 중 데이터 분류가 잘못된 경우는 폴도 이동 또는 대상 이동 처리함	각 부위별 촬영 허용 범위에 맞게 촬영되어 있는지 여부를 검수하는 과정으로 전초, 잎, 꽃, 열매 부위별 촬영 허용 기준에 부합되는지 여부를 검수하여 처리함	1,2,3차에 걸쳐 정제 완료한 데이터는 한의학, 식물분류학 전문가의 전수 검수를 통해 반려 또는 승인하여 처리함
처리	반려, 논의, 승인	폴더 이동 또는 대상 이동	반려, 논의, 승인	반려, 삭제, 승인
작업	클라우드 워커	클라우드 워커	클라우드 워커	식물분류학자, 한의학자

- 정제 작업은 1차, 2차, 3차 순으로 담당자를 지정하고 품목을 배정한다.
- 정제 방법은 매뉴얼에 따라 수행하고 궁금한 사항은 단체 채팅방을 통해 선임자에게 요청을 익힌다.

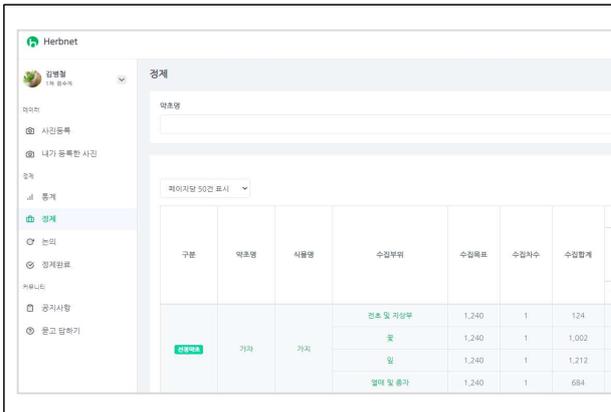
1) 정제 작업 절차

① 정제 작업을 시작한다.



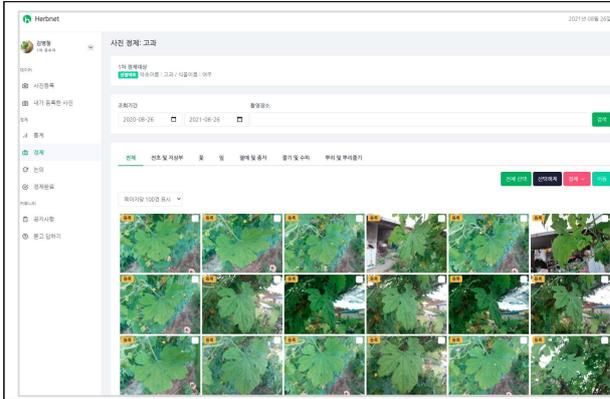
허브넷 로그인 후 정제 메뉴를 선택한다.

② 배정받은 품목을 검색하여 이미지를 선택한다.



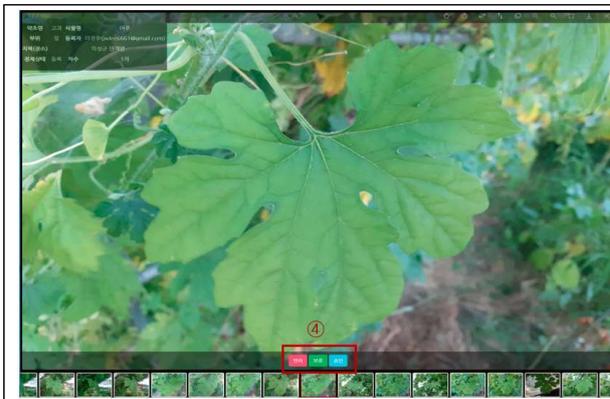
배정받은 품목을 검색하여 이미지를 선택한 후 정제 작업을 시작한다.

③ 작업 대상 이미지를 선택한 후 정제 작업을 수행한다.



정제 작업을 할 이미지를 클릭하여 큰 화면으로 이미지를 보며 정제 작업을 수행한다.

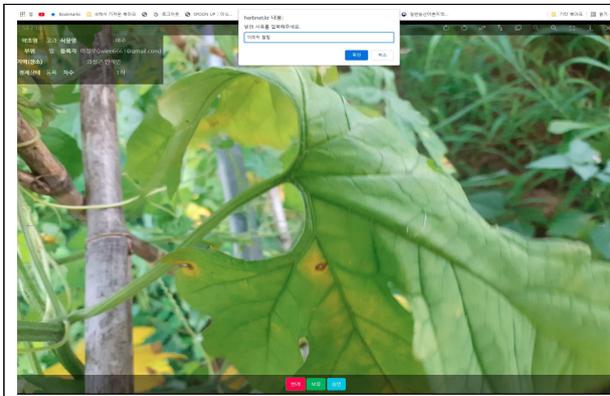
④ 해당 이미지를 정제 매뉴얼에 따라 정제를 수행한다.



정제는 정제 매뉴얼을 숙지한 후 이미지를 보고 정제 기준에 따라 승인, 반려, 보류 처리합니다.

정제 중 궁금한 점은 언제든지 선임자에게 물어보며 수행하도록 사전 교육

④-1 반려 처리



정제 기준에 따라 반려 처리할 경우는 반려 사유를 반드시 입력하여야 한다.

반려 이유는 초점 아웃, 흔들림 ... 등 15개 항목입니다

해당 항목을 찾을 수 없는 경우에는 선임자 또는 책임자에게 문의 후 처리하여야 합니다

⑤ 정제 데이터 전문가 전수 검수 후 정제가 종료된다.

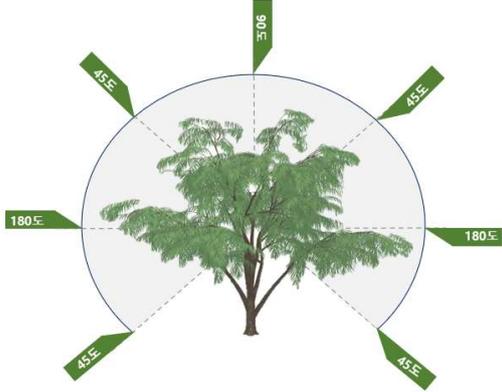
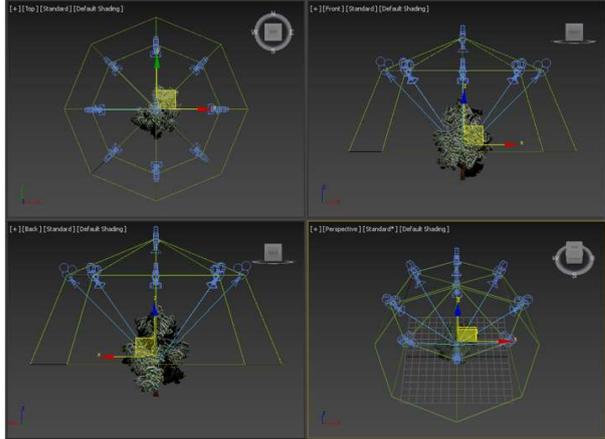


자문위원 또는 식물분류학자, 한의학 본초 전문가 등 동 분야 최고 전문가의 전수 검수를 통해 독초와 다른 식물이 섞이지 않도록 하고, 다른 유사한 품종이 잘못 촬영되었는지를 전문가의 검수 확인 후 최종 정제 작업을 종료한다.

3.3 수집 정제 · 기준

3.3.1 수집 기준

가. 전초 촬영기준(1개체당 24장 이내 촬영)

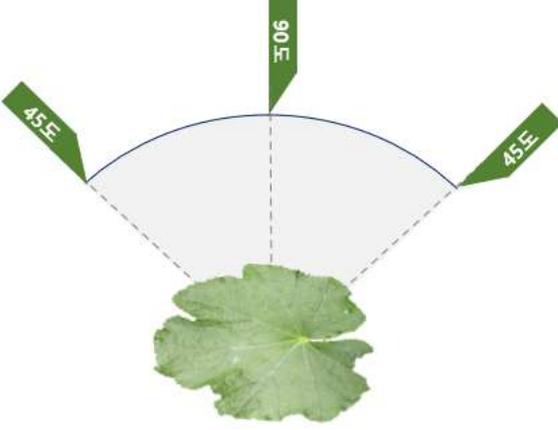
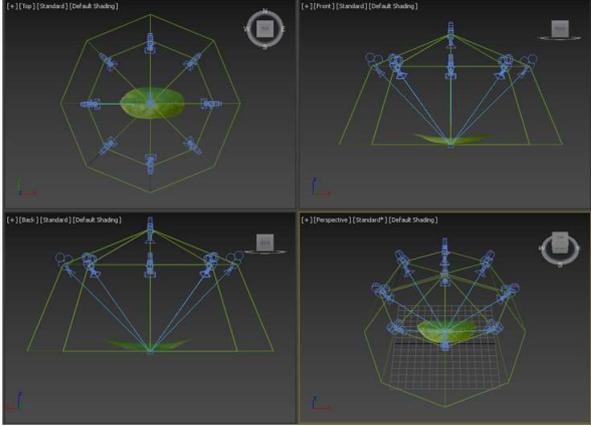
촬영 허용범위	촬영 시 카메라 위치
	

- 1개체당 24장 이내 촬영
- 일반적으로 식물을 바라보는 관점을 고려하여 촬영하고, 앞이 나는 모습 확인이 가능하게 줄기와 잎을 촬영, 먼 거리에서 촬영 삼가
- 여러 종의 식물이 혼재되게 촬영하는 것은 금지하며, 다른 종이 섞이지 않도록 촬영 전에 주변을 정리하여 촬영.
- 대상 부위 또는 개체가 화면에 70% 정도 차지하도록 촬영
- 흔들리거나, 초점이 어긋나지 않도록 촬영
- 햇볕이 강한 한낮에는 촬영을 피해야 함(그림자가 생길 확률이 높고, 그림자로 인해 데이터 왜곡이 발생할 수 있음)
- 이슬이 맺혀 있거나, 비가 갠 뒤 물기가 남은 상태에서 촬영 삼가
- 안전 주의

· 전초 촬영 예시



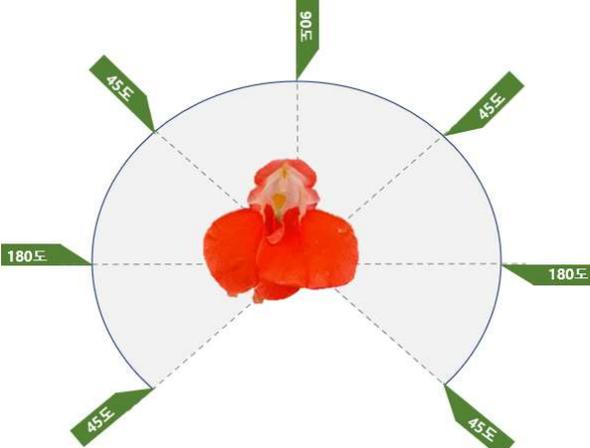
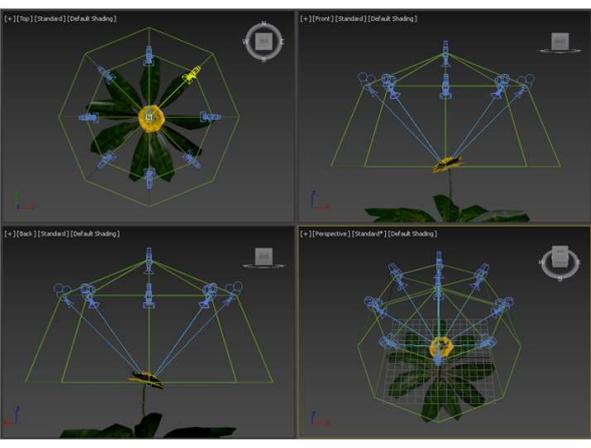
나. 잎 촬영기준(1개체당 24장 이내 촬영)

촬영 허용범위	촬영 시 카메라 위치
	
<ul style="list-style-type: none"> • 1개체당 24장 이내 촬영 • 잎은 식물을 구분하는 주요 요소이며 식물별로 명확한 특징이 있음. • 잎의 형태를 구분할 수 있도록 촬영 • 여러 개의 잎을 같이 촬영, 잎이 겹치거나, 잎에 그림자가 없도록 촬영 • 다른 식물이 섞이지 않도록 주변을 정리한 후 촬영 • 잎의 생육에 따라 잎의 크기가 여러 개가 섞여 있도록 촬영 • 잎맥이 나타나도록 촬영 • 훼손되거나, 오염된 잎 촬영 금지(수평 촬영은 삼가) • 초점, 흔들림, 빛 반사 등이 없도록 촬영 • 잎을 수평 촬영하지 않는다. 수평 촬영할 경우, 침형 또는 선형인 잎과 구분하기 매우 어려움. 	

· 잎 촬영 예시



다. 꽃 촬영기준(1개체당 24장 이내 촬영)

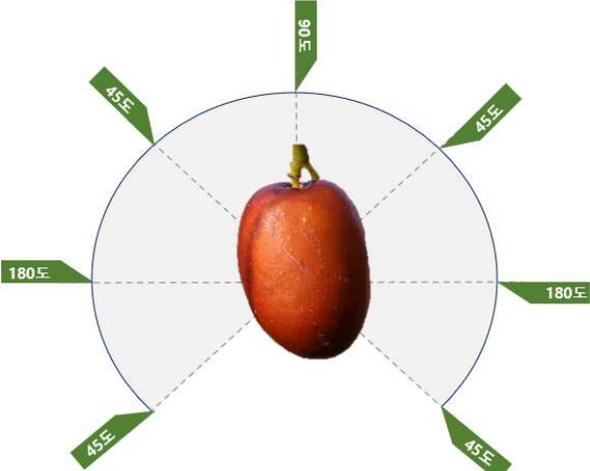
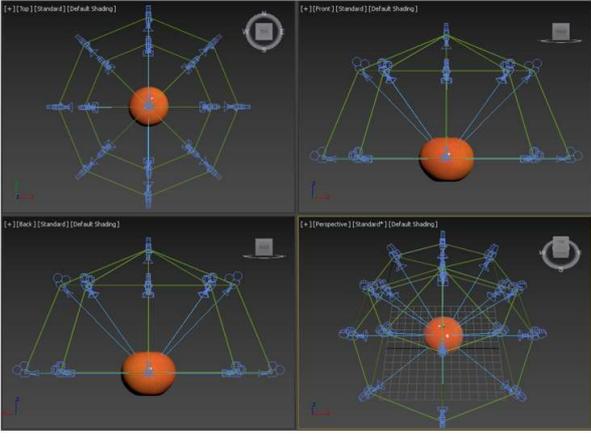
촬영 허용범위	촬영 시 카메라 위치
	

- 1개체당 24장 이내 촬영
- 꽃의 형태가 구분될 수 있도록 촬영
- 여러 개의 꽃을 같이 촬영하여 꽃의 성장 모습 또는 크기를 확인할 수 있게 촬영
- 꽃에 그림자가 생기지 않도록 촬영
- 다른 개체의 꽃과 섞이지 않도록 주변을 정리한 후 촬영
- 꽃의 생육에 따라 크기 별 촬영(만개 꽃, 성장기 꽃, 꽃망울)
- 초점, 흔들림, 빛 반사 등이 없도록 촬영 **허용범위 내에서 촬영**
- 꽃은 모양에 따라 작은 꽃들이 모여서 하나의 모양을 형성하는 경우, 작은 꽃을 지나치게 확대하거나 큰 꽃을 축소 촬영은 지양.

• 꽃 촬영 예시



라. 열매 촬영기준(1개체당 24장 이내 촬영)

촬영 허용범위	촬영 시 카메라 위치
	
<ul style="list-style-type: none"> • 1개체당 24장 이내 촬영 • 열매의 모양을 정확하게 구분할 수 있도록 촬영 • 다른 개체의 열매와 섞이지 않도록 주변을 정리한 후 촬영 • 상하거나, 오염된 열매 촬영 금지 • 초점, 흔들림, 빛 반사 등이 없도록 촬영 • 열매 촬영 시 주의할 점은 작은 열매가 여러 개 모여 하나의 모양을 형성하고 있는 경우에는 모여 있는 모양을 그대로 촬영하며, 이때 작은 열매를 지나치게 확대하거나 큰 열매를 과도하게 축소하여 촬영하지 않아야 함 • 안전 주의 	

• 열매 촬영 예시



3.3.2 원시데이터 정제 기준

가. 1차 정제

구분	정제 방법
1차 정제	<p>촬영된 원시데이터의 촬영 품질에 대한 정제 작업을 수행하는 과정으로</p> <ul style="list-style-type: none"> - 촬영 이미지 데이터 품질은 촬영 매뉴얼에 따라 촬영된 121종, 600,000장에 대한 전체 이미지 데이터를 육안 검수를 원칙으로 하여 정제함. - 촬영 품질이 기준에 충족되지 못한 경우는 반려 처리하고, 전문가의 논의가 필요한 경우에는 논의로 데이터를 이동시킴. - 1차 정제 작업자에 대한 정제 교육, 매뉴얼을 배포하고, - 상시 커뮤니케이션이 가능한 단독방을 개설하여 정제 업무를 효율적으로 수행 - 정제 작업자는 코로나-19 영향과 거주 지역을 고려하여 재택근무가 가능

◦ 1차 정제 기준(이미지 데이터 품질 정제)

번호	용어	이미지 상태	처리
1	초점 아웃	카메라 초점이 맞지 않아 이미지 상태가 흐릿하고 이미지 구분하기 어려운 경우	반려
2	흔들림	촬영 중 바람에 의해 식물이 움직였거나, 촬영 도중 카메라를 움직여서 이미지 데이터가 흔들려 보이는 상태	반려
3	이미지 잘림	대상 부위(꽃, 열매, 잎의 경우)가 화면에서 벗어나게 촬영된 경우. (전초의 경우는 예외로 인정)	반려
4	과대 확대	작은 꽃, 열매, 잎을 지나치게 확대하여 촬영 또는 먼 거리에 있는 대상 식물을 확대하여 촬영한 경우	반려
5	먼 거리	사람보다 큰 식물 전체 모습을 촬영하기 위해 지나치게 멀게 촬영한 이미지 데이터	반려
6	물기, 빛 반사	이슬 등 물기가 있는 상태에서 촬영하여 햇빛 반사가 심하여 지나치게 반짝이게 촬영된 이미지 데이터	반려
7	그림자	촬영 이미지에 다른 부위 또는 사람, 사물 등의 그림자가 포함된 상태로 촬영된 이미지 데이터	반려
8	잎 잘림	잎 나기(한 잎, 여러 잎, 긴 잎) 기준을 지키지 않고 촬영된 경우, 또는 긴 잎의 경우 잎을 잘라서 촬영한 경우 (잎 일부가 잘려있더라도 잎 구분이 가능한 상태이면 승인)	반려
9	병충해	식물의 잎, 꽃, 열매, 전초가 병해 또는 충해를 입은 상태로 촬영된 경우	반려

10	겹침	잎 등의 부위가 다른 잎 또는 부위와 겹쳐 있어 잎, 꽃 등이 온전한 모양을 구분하기 어려운 경우(모양 구분이 가능하면 승인처리 함)	반려
11	훼손	데이터가 훼손된 경우(벌레 먹은 상태 등)	반려
12	시뭇	이미지 데이터가 병에 걸려 시들어 있는 상태로 촬영된 경우(시간의 흐름에 따른 변화 과정은 시뭇으로 보지 않음)	반려
13	오염	흙, 진흙, 먼지 등 대상 식물이 오염되어 있는 상태로 촬영된 경우(촬영 전 깨끗하게 정리 후 촬영 필요)	반려
14	타 개체 개입	노끈, 철조망, 사람, 손 등 다른 사물이 포함되어 데이터 가공 시 사물이 포함될 수 있는 경우	반려
15	대상 아님	동의보감 독초 60종 또는 유사 식물 61종이 아닌 다른 식물을 촬영 또는 식물이 아닌 다른 개체를 촬영한 경우	반려

◦ 1차 정제 예시

번호	용어	이미지 상태 예시
1	초점 아웃	
2	흔들림	
3	이미지 잘림	

<p>4</p>	<p>과대 확대</p>	  <p>원래모습 과대 확대</p>
<p>5</p>	<p>먼 거리</p>	
<p>6</p>	<p>물기, 빛 반사</p>	 <p>감귤대</p>
<p>7</p>	<p>그림자</p>	
<p>8</p>	<p>앞 잘림</p>	
<p>9</p>	<p>병충해</p>	

<p>10</p>	<p>겉침</p>	
<p>11</p>	<p>훼손</p>	
<p>12</p>	<p>시들</p>	
<p>13</p>	<p>오염</p>	
<p>14</p>	<p>타 개체 개입</p>	

나. 2차 정제 기준(이미지 데이터 분류 오류 정제)

구분	정제 방법
2차 정제	1차 정제가 완료된 이미지 데이터의 부위 분류를 검수하는 과정 - 121종의 4 부위 (전초, 잎, 꽃, 열매) 폴더 분류가 잘못된 경우 - 폴더 분류가 잘못된 이미지 데이터를 정확한 부위 폴더로 이동시킴 - 전초 폴더에 꽃, 잎, 열매 등이 포함된 경우 - 잎 폴더에 열매 이미지 데이터가 있는 경우 - 꽃 폴더에 잎 이미지 데이터가 있는 경우 - 잎 폴더에 꽃 또는 열매 이미지 데이터가 있는 경우

◦ 2차 정제 방법

촬영 이미지 데이터 업로드 시 대상 식물 부위(전초, 잎, 꽃, 열매)폴더에 잘못 업로드 한 경우 2차 정제 과정에서 올바른 폴더 위치로 해당 이미지 데이터를 이동 시키는 과정

이동

분류 오류



잎 폴더로 분류 밤나무 데이터

이동

발생

분류 오류 발생

- 분류 오류 발생은 촬영자 촬영 후 이미지 데이터 부위 분류 시 데이터를 다른 폴더에 저장한 경우
- 허브넷 데이터 업로드 시 부위 폴더를 잘못 선택한 상태에서 데이터를 업로드 한 경우
- 정제 작업자의 실수
정제 작업자가 업무 미숙 또는 정제 매뉴얼 미 숙지 상태에서 정제할 경우
- 기타 시스템 오류(극히 예외적인 상황), 소프트웨어 오류로 발생될 수도 있음

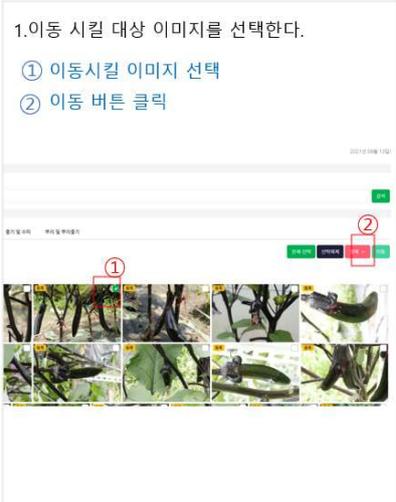


열매 폴더로 이동

◦ 2차 정제 후 데이터 이동 처리(예시)

1. 이동 시킬 대상 이미지를 선택한다.

- 이동시킬 이미지 선택
- 이동 버튼 클릭



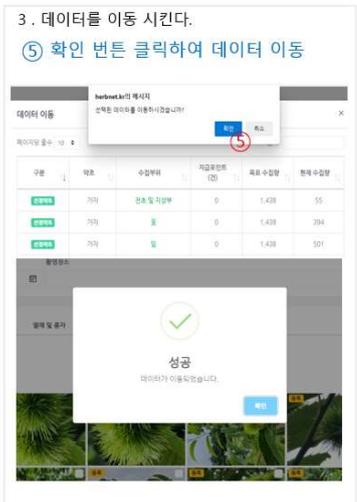
2. 검색 창에 이동시킬 식물을 검색하고 이동하고자 하는 부위를 선택한다.

- 이동시킬 식물 검색
- 이동시킬 부위 선택



3. 데이터를 이동 시킨다.

- 확인 버튼 클릭하여 데이터 이동



다. 3차 정제 기준(촬영기준 준수 여부)

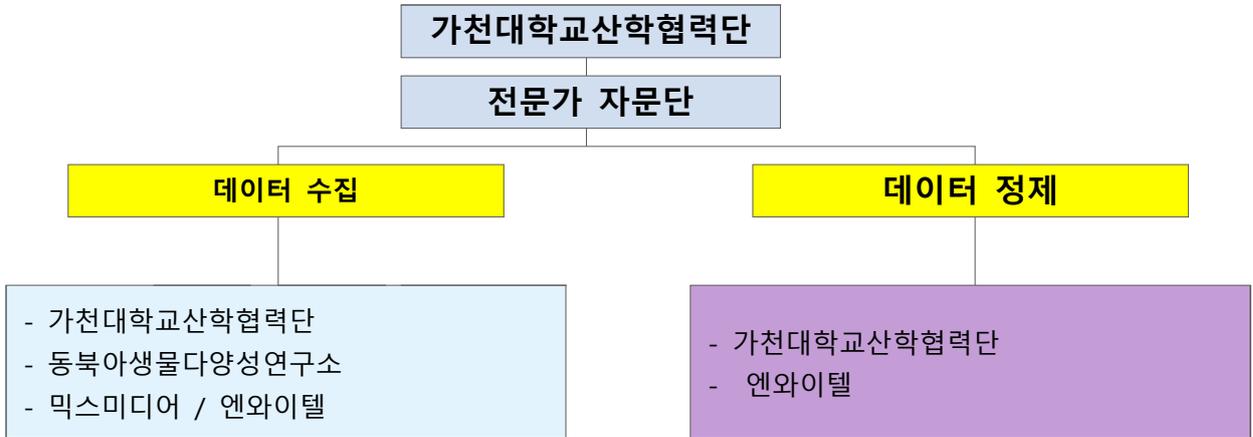
구분	정제 방법
3차 정제	<p>2차 정제가 완료된 이미지 데이터에 대해 촬영 허용 기준 여부를 검수하는 과정으로 식물 또는 한의학 전문가가 아닌 일반인의 시각에서 바라본 식물의 모습을 촬영기준 준수 여부 검수. (촬영기준 참조)</p> <ul style="list-style-type: none"> - 전초(목본, 초본, 덩굴) 촬영 허용 기준 적합 여부 - 잎(한 잎, 긴 잎, 여러 잎) 촬영 허용 기준 적합 여부 - 꽃(홀로 피기, 모여 피기, 차례 피기) 촬영 허용 기준 적합 여부 - 열매(홀로 맺기, 모여 맺기, 차례 맺기) 촬영 허용 기준 적합 여부 - 부위별 촬영 허용 기준 적합 여부를 검수하여 최종 승인 단계로 이관함

라. 4차 정제(전문가 전수 검증)

구분	정제 방법
3차 정제	<p>클라우드워커에 의해 품질 정제가 완료된 이미지 데이터에 대해 전문가(식물, 한의학)의 전수 검수를 통해 정확한 분류 작업이 이루어져 있는지를 검증하고 최종 승인하는 과정</p> <ul style="list-style-type: none"> - 촬영 이미지 데이터 품질 수준 - 부위별 폴더에 적합하게 이미지 데이터가 분류되었는지 검증 - 식물 폴더(상위 폴더)에 정확하게 대상 식물이 저장되었는지 여부 - 독초와 유사 식물 또는 다른 종이 섞이지 않았는지 전수 검수(정제 승인이 완료되면 원천데이터 고유번호 부여) - 전문가의 최종 정제 승인이 완료되면 데이터 가공 단계로 이관(자동 전송)

3.4 데이터 수집·정제 조직

3.4.1 조직도



3.4.2 기관별 역할

구분		수행기관	주요 수행 활동/세부 역할
데이터셋	설계	가천대학교산학협력단 동북아생물다양성연구소 엔와이텔 아워텍 농진청(자문)	<ul style="list-style-type: none"> ○ 독초(60종), 유사 식물(61종) 이미지 및 속성 정보 - 촬영 도구, 촬영지역 촬영 일자 - 컬럼명 표준 기준 적용(외부 자문)
	수집	믹스미디어 동북아생물다양성연구소 가천대학교산학협력단 엔와이텔	<ul style="list-style-type: none"> ○ 대상 식물의 부위별 촬영 - 전초 잎 꽃 열매 ○ 대상 식물의 부위별 입체적 촬영 - 전 후 좌 우부위별 촬영(준)
	정제	엔와이텔 가천대학교산학협력단	<ul style="list-style-type: none"> ○ 이미지 품질 ○ 분류 오류 검수 ○ 촬영기준 검수 ○ 전문가 전수 검수

3.4.3 교육 훈련

장소/일시	일시 : 2021.07.05.(월) 장소 : 가천대학교 한의과대학 501호 시간 : 10:00 ~ 14:00
참석기관	주관기관 : 가천대학교 산학협력단(이영중, 서정범, 정고은 외) 참여기관 : (주)믹스미디어(이상모, 심일보, 이선영 외), (주)아워택(최경웅), (유)나인펄스(노인규), (유)엔와이텔(김병철)
참석대상	촬영 담당자, 업로드 및 검수 담당자, 가공 책임자, 정제 책임자, 인공지능 책임자, 품질관리 책임자
자 료	촬영 매뉴얼, 식물도감, 교육 자료(PPT)
내 용	<ul style="list-style-type: none"> • 10 : 15 ~ 10 : 15 사업 소개 및 인사(가천대학교 이영중 교수) • 10 : 15 ~ 10 : 40 참석자 소개 • 10 : 40 ~ 11 : 00 약초 촬영 준비(믹스미디어 이상모 감독) • 11 : 00 ~ 11 : 20 촬영 이미지 업로드 교육(믹스미디어 이선영) 질의응답 • 11 : 30 ~ 12 : 10 식물 판별 알고리즘 및 허브넷 소개(엔와이텔 김병철) 질의응답 • 12 : 10 ~ 12 : 30 약초 촬영 경험 공유, 촬영 시연(한도훈 작가, 민점호 작가, 김봉희 감독) • 12 : 30 ~ 14 : 00 점심 및 네트워크

구분	논의 및 교육 내용
이미지 데이터 수집 교육	<ul style="list-style-type: none"> • 촬영대상 식물도감 배포 <ul style="list-style-type: none"> - 선정약초의 특징 - 선정약초와 유사 식물 비교 사진 • 촬영기준 및 촬영 방법 교육 <ul style="list-style-type: none"> - 개체당 촬영기준 - 촬영 부위 : 전초, 잎, 꽃, 열매 - 촬영 방법 및 기준 교육 • 수집 대상 : 식물도감과 선정 식물 리스트로 대체
데이터 다양성	<ul style="list-style-type: none"> • 촬영대상 식물의 성장 과정이 나타나도록 대상 식물을 다른 날짜에 촬영 - 식물 성장에 따른 데이터 확보 목적
데이터 품질	<ul style="list-style-type: none"> • 개체별 촬영 기준(수량) 교육 <ul style="list-style-type: none"> ▷ 전초(24장), 잎(24장), 꽃(24장), 열매(24장) ▷ 전초의 경우 목본, 초본, 덩굴의 기준을 달리 적용, 촬영은 4개의 시점(평면, 상, 중, 하)에 8방향에서 촬영 (목본) 가지에 잎, 꽃, 열매가 포함된 것을 전초 정의에 포함,

	<ul style="list-style-type: none"> - 목본 전체가 나오는 데이터는 의미가 없음 - 가지에 꽃, 잎, 열매가 포함되어도 됨 - 작은 목본의 경우에는 지상부 전체 모습을 촬영 <p>(초본) 줄기에 잎, 꽃, 열매 등이 포함된 것을 전초 정의에 포함</p> <ul style="list-style-type: none"> - 줄기에 잎, 꽃, 열매 등이 포함되게 촬영해도 됨 - 뭉쳐나는 품종(예, 갈대, 억새 등)의 경우에는 뭉쳐난 모습 촬영 금지, 분리하여 촬영 요망 <p>(덩굴) 덩굴 전체 모습이 나오게 촬영할 필요는 없음,</p> <ul style="list-style-type: none"> - 덩굴에 잎, 꽃, 열매가 포함된 것을 전초 정의에 포함 <p>▷ 잎의 경우 성장 상태가 나타나도록 촬영-촬영은 3개의 시점(상, 중, 하)에 8방향에서 촬영</p> <p>(각도) 잎 촬영 시 잎을 평면으로 볼 때를 기준으로 앞, 뒤, 좌, 우 각각 45도까지 촬영하는 것을 기준으로 함</p> <p>(크기) 대, 중, 소 각각 크기별 촬영하여 성장 모습을 담아내도록 촬영</p> <p>(겹잎) 겹잎 전체가 나오도록 촬영(겹잎을 날장으로 뜯어 촬영은 금지)</p> <p>(홀잎) 잎의 크기별 촬영(대, 중, 소), 한 장 또는 여러 장이 나오도록 촬영</p> <p>(오염) 잎에 벌레, 물방울 등 다른 이물질이 없도록 촬영, 잎에 물기가 있는 경우도 오염 데이터로 간주</p> <p>▷ 꽃의 경우 성장 상태가 나타나도록 촬영-촬영은 3개의 시점(상, 중, 하)에 8방향에서 촬영</p> <p>(각도) 꽃을 정면에서 바라볼 때를 기준으로 상, 하, 좌, 우 각각 45°까지 촬영하는 것을 기준으로 함</p> <p>(크기) 대, 중, 소 각각 크기별 촬영하여 성장 모습을 담아내도록 촬영</p> <p>(특징) 뭉쳐나는 꽃의 경우에는 뭉쳐나는 모습을 촬영, 작은 꽃(예, 쪽)의 경우는 확대하여 촬영 금지</p> <p>(오염) 꽃에 벌레, 물방울 등 다른 이물질이 없도록 촬영, 꽃에 물기가 있는 경우도 오염 데이터로 간주</p> <p>▷ 열매의 경우 성장 상태가 나타나도록 촬영 - 촬영은 3개 시점(상, 중, 하)에 8방향에서 촬영</p> <p>(크기) 대, 중, 소 각각 크기별 촬영하여 성장 모습을 담아내도록 촬영</p> <p>(특징) 뭉쳐나는 열매와 단일 열매는 구분하여 촬영, 뭉쳐나는 열매는 알알이 촬영 금지</p> <p>(시기) 열매 맺힘 -> 열매 성장 -> 열매 성숙 -> 열매 완숙까지를 볼 수 있도록 촬영</p> <p>(오염) 열매에 벌레, 물방울 등 다른 이물질이 없도록 촬영, 열매에 물기가 있는 경우도 오염 데이터로 간주</p>
	<ul style="list-style-type: none"> • 스마트폰 촬영 설정 교육(저장 형식은 JPG만 가능) ▷ 사진 크기 - 5MB 이하로 설정 권고 • 메타 정보 : 촬영 장소(군, 면까지) • DSLR 촬영 시 속성 정보(기기 모델, 제조사, 화소, 플래시 사용 등)가 포함되어야 함
<p>이미지 저장 방법</p>	<ul style="list-style-type: none"> • 이미지 저장 분류 방법 검토 후 재배포 - 약초 명_촬영 부위_촬영 일자_촬영 장소_촬영자(믹스미디어 안)

	<ul style="list-style-type: none"> • ▷(공통) 독초-식물명-촬영 부위 - 촬영 일자- 촬영장소 - 촬영자 - 예) 야생 촬영 약초_전초_가죽나무_2021.0709_익산시 여산면_홍길순 유사_전초_참죽나무_2021.0709_익산시 여산면_홍길순 - 예) 수목원, 식물원 등 약초_열매_자귀나무_2021.0709_홍릉수목원_홍길순
<p>안전관리</p>	<ul style="list-style-type: none"> • 장마, 폭우, 해충 등에 대한 안전관리 주의 당부 • 코로나19에 따른 정부 지침 준수 당부 • 안전관리 문서 배포(정리 후 추후 배포)



3.5 수집 · 정제 도구

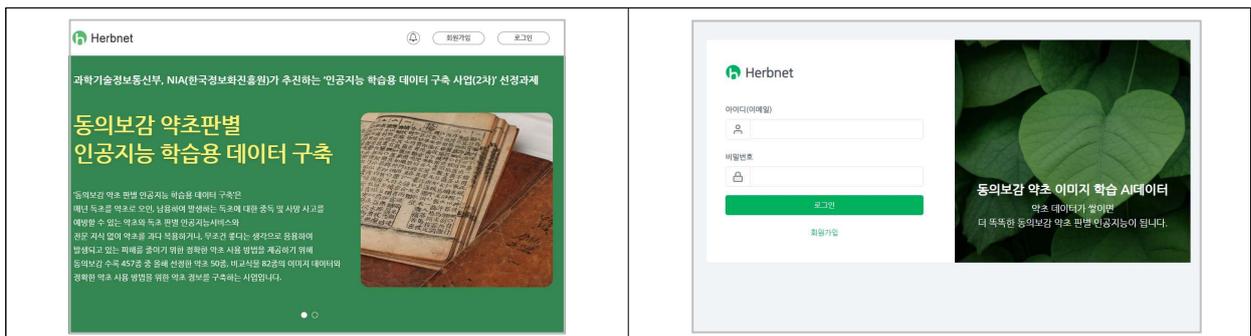
3.5.1 수집 도구



● 촬영자 안전 주의

1. 촬영 대상 식물의 사진 촬영은 촬영자 안전이 확보된 이후 진행하여야 한다.
2. 촬영은 반드시 2인 1조로 하여 촬영을 권장하며, 촬영지역이나 촬영 동선을 가까운 지인에게 미리 알려 두어 위급한 상황 발생 시, 즉시 대응할 수 있도록 해야 한다.
3. 우천 시를 대비하여 가방에는 우비, 방한복, 비상식량 등을 미리 준비하여야 한다.
4. 기상악화 시에는 즉시 촬영을 중지해야 한다.(폭염, 태풍 등)
5. 응급 상황 발생 시 촬영 팀장 또는 담당 **사업책임자에게 즉시 연락**한다.
6. 촬영 중 벌레 물림 사고 및 뱀에 물리는 사고를 대비하여 가까운 보건소 위치를 항상 숙지하고 보건소 전화번호를 미리 알아 둔다. **위급 상황 시 119 이용**한다.
7. 코로나-19에 따라 국가 방역 규칙을 준수하고, 다중이 이용하는 시설은 되도록 입장을 삼간다.
8. 코로나-19 확진자와 접촉한 경우는, 사업 담당자 또는 책임자에게 즉시 보고하고 상응하는 조치가 취해질 수 있도록 하여야 한다.

3.5.2 정제 도구

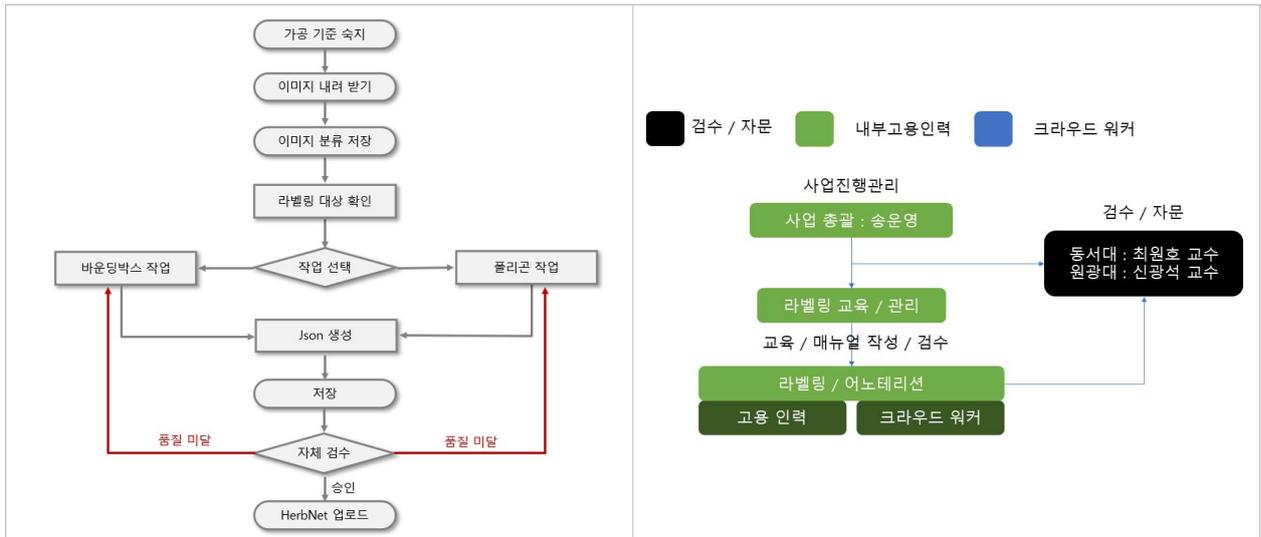


4. 데이터 가공

4.1 라벨링 절차 및 방법

- 데이터 가공은 인공지능 학습에 적합하게 이미지를 라벨링·어노테이션하는 과정으로 가공 대상인 정제 완료된 이미지는 허브넷(Herbnet)에서 내려받아 식물 부위별 특성에 맞게 바운딩 박스하고, 일부 지정된 대상은 폴리곤 하여 작업 함.

가) 가공 절차



가공 도구	가공 메뉴얼
 <p>labelme</p>	 <p>LabelMe는 컴퓨터 비전 연구를 위한 이미지 데이터 베이스를 구축하기 위한 온라인 주석 도구를 제공하는 공개 저작도구로, 본 과제를 통해 획득한 이미지 데이터 가공에 적합하여 저작도구로 사용함.</p>

○ 데이터 가공 프로세스

저작도구 설치	이미지 다운로드	라벨링 작업	json 저장	라벨 검수	허브넷 업로드
					

4.2 데이터 가공 기준

가. 전초 데이터 가공 기준

전초 구분	전초 특징	라벨링 기준
 목본	<ul style="list-style-type: none"> 목본은 식물분류학적 기준에 따르면 줄기나 가지가 목질 형질을 가진 식물을 뜻하며, 통상 묘목을 제외하고 수심 미터 이상 자라는 교목과 그보다는 작게 자라는 관목으로 분류가 된다. 	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 여러 개체가 중첩되어 존재할 경우 여러 개체를 하나로 묶어서 바운딩 박스 함
 초본	<ul style="list-style-type: none"> 초본은 식물학적 분류에 따르면 지상부에 목본 줄기를 가지고 있지 않은 식물로 꽃이 피고 열매를 맺으면 생장이 멈추고 식물 전체가 죽는 일년생 식물과 겨울이나 건기에 지상부위는 죽고 뿌리나 줄기가 살아남아 다음 해에 이로부터 새순이 솟아올라 자라는 다년생 식물로 구분된다. 	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 다른 개체가 섞이지 않고, 대상 품목이 이미지 전체인 경우는 전체를 라벨링 함. 여러 개체가 중첩되어 존재할 경우 여러 개체를 하나로 묶어서 바운딩박스.
 덩굴	<ul style="list-style-type: none"> 덩굴은 식물학적 분류에서는 줄기가 가늘고 길게 자라 다른 식물이나 물체에 걸치거나 감겨서 생활하는 식물로 잎, 꽃차례(화서)등의 일부가 변형되어 다른 물체를 감는 형태로 생장하는 식물이다. 덩굴 중 목질로 이루어진 식물 등침, 등나무 등을 만경식물이라 한다. 	<ul style="list-style-type: none"> 개체를 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 다른 개체가 섞이지 않고, 대상 품목이 이미지 전체인 경우는 전체를 라벨링 함.

나. 잎 데이터 가공 기준

분류	특징	라벨링 기준
 한 잎	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 줄기에서 하나씩 잎이 난 형태로 하나의 잎을 라벨링 하여야 하며, 잎의 특성상 여러 개의 잎이 겹쳐 있을 때는 겹쳐 있는 잎 모두를 묶어서 라벨링 한다. 	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 잎 특성상 여러 개의 잎이 모여 있는 경우는 여러 개체를 묶어서 라벨링 함.

 <p>여러 잎</p>	<ul style="list-style-type: none"> • 줄기 또는 가지에서 여러 장의 잎이 모여 하나의 잎을 이루는 경우를 말하며 여러 잎은 여러 잎 전체 모습을 확인할 수 있도록 라벨링을 권장 • 여러 잎 라벨링 시 여러 잎을 이루는 개별 구성 잎이 떨어졌거나, 벌레가 먹거나, 훼손되어 있더라도 라벨링 가능. 	<ul style="list-style-type: none"> • 개체가 독립적으로 단일 개체를 바운딩 박스 함. • 예외 사항 -여러 잎 중에서 개별 잎 하나를 라벨링하여 잎 모양을 학습할 수 있도록 함.
 <p>긴 잎</p>	<ul style="list-style-type: none"> • 선형으로 길게 나는 형태의 잎을 잎이라고 하며 선형의 피침형태로 나기 때문에 잎과 줄기 구분이 어려워 지상부 전체의 모습을 라벨링하여 긴 잎의 형태가 구분 가능 해야 한다. • 또한 긴잎을 가진 식물의 경우 모여서 자라기 때문에 여러 개체가 모여서 나는 특징이 반영될 수 있도록 라벨링 한다. 	<ul style="list-style-type: none"> • 긴 잎 식물 특성상 여러 개체가 모여서 나는 특징이 반영될 수 있도록 긴 잎 식물이 모여나는 경우에는 전체를 묶어서 라벨링 함. • 긴 잎에 해당하는 대상 품목의 경우 여러 개체가 뭉쳐서 성장하기에 이런 경우는 잎을 분리하여 라벨링 하지 않고 뭉쳐있는 형태 전체를 라벨링 함. • 긴 잎이 누운 상태로 자라는 이미지 경우도 형태 그대로를 라벨링 함.

다. 꽃 데이터 가공 기준

분류	특징	라벨링 기준
 <p>홀로 피기</p>	<ul style="list-style-type: none"> • 식물 분류학적 기준은 꽃 나는 모양에 따라 총상화서, 두상화서 등으로 꽃을 분류하지만 본 과제에서는 일반인의 시각을 기준으로 꽃이 나는 모양이 독립적인 형태를 띄는 경우를 홀로 피기로 정의하였다. 	<ul style="list-style-type: none"> • 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. • 다른 개체가 섞이지 않고, 대상 품목이 이미지 전체인 경우는 전체를 라벨링 함. • 홀로 피기 꽃 중 꽃이 여러 개체가 겹쳐 있는 경우에는 겹친 꽃을 모두 그룹으로 라벨링 한다
 <p>모여 피기</p>	<ul style="list-style-type: none"> • 작은 꽃 여러 개가 모여 하나의 꽃을 이루는 모양의 경우 모여 피기로 정의하였다. 	<ul style="list-style-type: none"> • 꽃이 겹쳐있는 경우 • 모여피기 꽃을 분리하여 촬영된 데이터 경우 독립적으로 라벨링 함.

 <p>차례 피기</p>	<ul style="list-style-type: none"> • 일반인의 시각에서 보면 긴 줄기에 꽃들이 순서대로 피는 모양인 경우를 차례 피기로 정의하였다. 	<ul style="list-style-type: none"> • 차례피기의 경우 꽃 특성사 차례대로 피기 때문에 꽃망울과 꽃이 하나의 이미지에 있는 경우에는 묶어서 라벨링 함. • 벌레나 이물질이 붙어있는 경우도 라벨링 함. • 차례피기 꽃 라벨링 시 꽃망울과 만개한 꽃이 한 줄기에 있는 경우가 많아 꽃의 특성이 반영되도록 라벨링 한다.
--	---	---

라. 열매 데이터 가공 기준

분류	특징	라벨링 기준
 <p>홀로 맺기</p>	<ul style="list-style-type: none"> • 홀로맺기를 독립적으로 맺은 경우를 홀로맺기라 한다. 	<ul style="list-style-type: none"> • 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. • 홀로맺기 열매인 경우에도 여러 개의 열매가 모여있는 열매가 겹쳐 보이는 경우에는 열매를 그룹으로 라벨링 함.
 <p>모여 맺기</p>	<ul style="list-style-type: none"> • 모여 피는 꽃의 경우 꽃이 진 자리에 열매가 맺히며 여러 개의 열매가 하나의 큰 모양을 이루는 경우를 모여 맺기 한다. 	<ul style="list-style-type: none"> • 모여 맺기 열매는 작은 열매 여러 개가 모여 하나의 형태를 이루므로 모여 있는 모양이 잘 나타나도록 라벨링 함. • 모여맺기 열매 하나를 라벨링 하여 모양과 특징을 학습 할 수 있도록 라벨링 함
 <p>차례 맺기</p>	<ul style="list-style-type: none"> • 차례 맺기는 한줄기에 여러 개의 열매가 줄줄이 맺힌 모양을 차례 맺기라 한다. • 차례 맺기 열매는 하나의 줄기에 작은 열매가 줄줄이 맺히는 특징을 가지고 있어 차례 맺기 열매 하나에서 열매가 맺는 모양을 동시에 확인할 수 있다. 	<ul style="list-style-type: none"> • 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. • 차례 맺기 열매 일부를 그룹으로 라벨링 함.

4.2.1 데이터 가공

객체의 외곽 테두리 간격은 바운딩 박스일 경우 50픽셀 이내로 하고, 폴리곤 일 경우에는 외곽 테두리 간격을 100픽셀 이내로 한다. 이는 촬영된 이미지 해상도가 최저 ~ 최고 사양 사이의 차이가 매우 커, 이미지 확대 시 간격이 더 크게 벌어 질 수 있어 사전에 이에 대한 기준을 정함으로써 라벨 작업 후 검수 과정에서 혼란을 예방하고자 함

구분	라벨 테두리 간격(픽셀)	이미지 예시(기준 해상도 2K에서 작업했을 경우)
기준	이미지 해상도 2K 기준	 <p>해상도 2K 테두리 해상도 4K 테두리 해상도 8K 테두리</p>
바운딩 박스	객체와 테두리 사이 간격 50픽셀 이내	
폴리곤	객체와 테두리 사이 간격 100픽셀 이내	

가. 전초 가공

분류	특징	라벨링 규격
목본	<ul style="list-style-type: none"> 목본은 식물분류학적 기준에 따르면 줄기나 가지가 목질 형질을 가진 식물을 뜻하며, 통상 어린 묘목을 제외하고는 수십 미터 이상 자라는 교목과 그보다는 작게 자라는 관목으로 분류가 된다. 	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 여러 개체가 중첩되어 존재할 경우 여러 개체를 하나로 묶어서 바운딩 박스 함
초본	<ul style="list-style-type: none"> 초본은 식물학적 분류에 따르면 지상부에 목본 줄기를 가지고 있지 않은 식물로 꽃이 피고 열매를 맺으면 생장이 멈추고 식물 전체가 죽는 일년생 식물과 겨울이나 건기에 지상부위는 죽고 뿌리나 줄기가 살아남아 다음 해에 이로부터 새순이 솟아올라 자라는 다년생 식물로 구분된다. 	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 다른 개체가 섞이지 않고, 대상 품목이 이미지 전체인 경우는 전체를 라벨링 함. 여러 개체가 중첩되어 존재할 경우 여러 개체를 하나로 묶어서 바운딩박스.
덩굴	<ul style="list-style-type: none"> 덩굴은 식물학적 분류에서는 줄기가 가늘고 길게 자라 다른 식물이나 물체에 걸치거나 감겨서 생활하는 식물로 잎, 꽃차례(화서)등의 일부가 변형되어 다른 물체를 감는 형태로 성장하는 식물이다. 덩굴 중 목질로 이루어진 식물 등척, 등나무 등을 만경 식물이라 한다. 	<ul style="list-style-type: none"> 개체를 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 다른 개체가 섞이지 않고, 대상 품목이 이미지 전체인 경우는 전체를 라벨링 함.

나. 잎 가공

분류	특징	라벨링 기준
한 잎	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. • 줄기에서 하나씩 잎이 난 형태로 하나의 잎을 라벨링 하여야 하며, 잎의 특성상 여러 개의 잎이 겹쳐 있을 때는 겹쳐 있는 잎 모두를 묶어서 라벨링 한다. 	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 잎 특성상 여러 개의 잎이 모여 있는 경우는 여러 개체를 묶어서 라벨링 함.
여러 잎	<ul style="list-style-type: none"> 줄기 또는 가지에서 여러 장의 잎이 모여 하나의 잎을 이루는 경우를 말하며 여러 잎은 여러 잎 전체 모습을 확인할 수 있도록 라벨링을 권장 어려 잎 라벨링 시 여러 잎을 이루는 개별 구성 잎이 떨어졌거나, 벌레가 먹거나, 훼손되어 있더라도 라벨링 가능. 	<ul style="list-style-type: none"> 개체가 독립적으로 단일 개체를 바운딩 박스 함. 예외 사항 - 여러 잎 중에서 개별 잎 하나를 라벨링 하여 잎 모양을 학습할 수 있도록 함.
긴 잎	<ul style="list-style-type: none"> 선형으로 길게 나는 형태의 잎을 잎이라고 하며 선형의 피침형태로 나기 때문에 잎과 줄기 구분이 어려워 지상부 전체의 모습을 라벨링하여 긴 잎의 형태가 구분 가능 해야 한다. 또한 긴잎을 가진 식물의 경우 모여서 자라기 때문에 여러 개체가 모여서 나는 특징이 반영될 수 있도록 라벨링 한다. 	<ul style="list-style-type: none"> 긴 잎 식물 특성상 여러 개체가 모여서 나는 특징이 반영될 수 있도록 긴 잎 식물이 모여나는 경우에는 전체를 묶어서 라벨링 함. 긴 잎에 해당하는 대상 품목의 경우 여러 개체가 뭉쳐서 성장하기에 이런 경우는 잎을 분리하여 라벨링 하지 않고 뭉쳐있는 형태 전체를 라벨링 함. 긴 잎이 누운 상태로 자라는 이미지 경우도 형태 그대로를 라벨링 함.

다. 꽃 가공

분류	특징	라벨링 기준
홀로 피기	<ul style="list-style-type: none"> 식물 분류학적 기준은 꽃 나는 모양에 따라 총상화서, 두상화서 등으로 꽃을 분류하지만 본 과제에서는 일반인의 시각을 기준으로 꽃이 나는 모양이 독립적인 형태를 띄는 경우를 홀로 피기로 정의하였다. 	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 다른 개체가 섞이지 않고, 대상 품목이 이미지 전체인 경우는 전체를 라벨링 함. 홀로 피기 꽃 중 꽃이 여러 개체가

		겹쳐 있는 경우에는 겹친 꽃을 모두 그룹으로 라벨링 한다
모여 피기	<ul style="list-style-type: none"> 작은 꽃 여러 개가 모여 하나의 꽃을 이루는 모양의 경우 모여 피기로 정의하였다. 	<ul style="list-style-type: none"> 꽃이 겹쳐있는 경우 모여피기 꽃을 분리하여 촬영된 데이터 경우 독립적으로 라벨링 함.
차레 피기	<ul style="list-style-type: none"> 일반인의 시각에서 보면 긴 줄기에 꽃들이 순서대로 피는 모양인 경우를 차레 피기로 정의하였다. 	<ul style="list-style-type: none"> 차레피기의 경우 꽃 특성사 차레대로 피기 때문에 꽃망울과 꽃이 하나의 이미지에 있는 경우에는 묶어서 라벨링 함. 벌레나 이물질이 붙어있는 경우도 라벨링 함. 차레피기 꽃 라벨링 시 꽃망울과 만개한 꽃이 한 줄기에 있는 경우가 많아 꽃의 특성이 반영되도록 라벨링 한다.

라. 열매 가공

분류	특징	라벨링 기준
홀로 맺기	<ul style="list-style-type: none"> 홀로맺기를 독립적으로 맺은 경우를 홀로맺기라 한다. 	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 홀로맺기 열매인 경우에도 여러 개의 열매가 모여있는 열매가 겹쳐 보이는 경우에는 열매를 그룹으로 라벨링 함.
모여 맺기	<ul style="list-style-type: none"> 모여 피는 꽃의 경우 꽃이 진 자리에 열매가 맺히며 여러 개의 열매가 하나의 큰 모양을 이루는 경우를 모여 맺기 한다. 	<ul style="list-style-type: none"> 모여 맺기 열매는 작은 열매 여러 개가 모여 하나의 형태를 이루므로 모여 있는 모양이 잘 나타나도록 라벨링 함. 모여맺기 열매 하나를 라벨링 하여 모양과 특징을 학습 할 수 있도록 라벨링 함
차레 맺기	<ul style="list-style-type: none"> 차레 맺기는 한줄기에 여러 개의 열매가 줄줄이 맺힌 모양을 차레 맺기라 한다. 차레 맺기 열매는 하나의 줄기에 작은 열매가 줄줄이 맺히는 특징을 가지고 있어 차레 맺기 열매 하나에서 열매가 맺는 모양을 동시에 확인할 수 있다. 	<ul style="list-style-type: none"> 개체가 독립적으로 존재할 경우는 단일 개체를 바운딩 박스 함. 차레 맺기 열매 일부를 그룹으로 라벨링 함.

바. 데이터 가공 예시

◦ 전초 가공 예시

<p>전초 목본→ 이미지 전체를 라벨링</p>	
	
<p>전초 초본→ 여러 개체가 뭉쳐서 난 경우는 여러 개체를 라벨링</p>	
	
<p>전초 덩굴→ 단일 개체를 라벨링</p>	
	

◦ 잎 라벨링 예시

<p>하나의 잎을 라벨링</p>	<p>여러 개의 잎을 묶어서 라벨링</p>
	
<p>여러 잎 라벨링</p>	<p>개별 잎 하나를 라벨링</p>
	
<p>긴 잎 라벨링</p>	<p>긴 잎 여러 개를 묶어 라벨링</p>
	

◦ 꽃 라벨링 예시

<p>홀로 피기 라벨링</p>	<p>꽃을 여러 개 묶어서 라벨링</p>
	
<p>모여 피기 라벨링</p>	<p>개별꽃 하나만 라벨링</p>
	
<p>차례 피기 라벨링</p>	<p>개별꽃 하나만 라벨링</p>
	

◦ 열매 라벨링 예시

단독 또는 묶음 라벨링

식물의 특성상 전초, 꽃, 열매, 잎이 모여서 자라는 특성을 가진 경우에는 개별적으로 라벨링 하기가 어려움으로 여러 객체를 묶어서 그룹으로 라벨링(바운딩 박스 또는 폴리곤)하고 그렇지 않는 경우는 객체별로 라벨링 한다.



4.2.2 폴리곤 대상

◦ 독초 60종 및 유사 비교식물 61종 중 인공지능 판별 정확도를 높이기 위해 유사도가 높은 식물 부위 중 잎, 꽃 열매 부위의 경우에는 학습용 이미지 데이터 목표의 10%인 500장은 폴리곤으로 한다.

◦ 폴리곤 대상

부위	폴리곤 대상선정	대상 식물(독초, 유사 비교식물)
잎	잎 모양이 피침형 또는 긴피침형, 원형, 장타원형중 독초와 유사 비교식물 잎의 유사도가 높은 경우는 목표량의 10%는 폴리곤 함.	갈대, 달뿌리풀, 띪, 물억새, 석위, 세뿔석위, 연꽃, 수련
꽃	꽃차례가 단정화서 또는 단정화서이며 취산화서인 품목 중 독초와 유사 식물 꽃의 경우는 목표량의 10%는 폴리곤 함.	연꽃, 수련, 갈대, 억새, 달뿌리풀, 물억새, 띪, 새삼, 미국실새삼
열매	열매 특성 중 열매 모양이 장과, 구형 또는 협과인 품목 중 독초와 비교식물 잎의 유사도가 높은 경우는 목표량의 10%는 폴리곤 함.	가지. 도깨비가지, 여주, 수세미오이. 하늘타리, 가시박, 수박, 자귀나무. 왕자귀나무. 완두콩. 동부, 회화나무. 아까시나무.오이

4.3 가공 규격

4.3.1 구문 정확성

◦ 라벨은 다음과 같이 구성되며, 항목별 타입과 필수 여부 및 값 범위를 정의한다.

구분	항목명	타입	필수 여부	설명	범위	
1	1	plantinfo	object	클래스 정보		
	1-1	plantinfo.class_name	string	Y	클래스 이름	클래스 이름(121종)
	1-2	plantinfo.classification_info	string	Y	클래스 구분	판별대상, 유사
	1-3	plantinfo.herb_name	string		동의보감 판별대상 이름	동의보감 판별대상 이름
	1-4	plantinfo.similar_plants	list		유사식물	유사식물 이름
	1-5	plantinfo.toxic_info	boolean	Y	독성 여부	Y, N
	1-6	plantinfo.instance_info	object		인스턴스 정보	최소값 : 1, 최대값 : 1
	1-6-1	plantinfo.instance_info.name	string	Y	식물 부위	전초, 꽃, 잎, 열매
2	2	image	object	인스턴스 세부 정보	이미지 정보, 최소값 : 7(필드가 Y인 최소 개수)	
	2-1	image.file_path	string	Y	이미지 경로	../..../원천데이터/가족나무/전초 최소값:1, 최대값 :200
	2-2	image.file_name	string	Y	클래스(식물명)_인스턴스(부위)_file_id	가족나무_전초_000001.jpg 최소값:4, 최대값:100
	2-3	image.shooting_date	string	Y	촬영 날짜	yyyy-mm-dd
	2-4	image.region_name	string		촬영 지역	군(시) 또는 장소명 (예, 홍릉수목원)
	2-5	image.file_format	string	Y	파일 형식	JPG
	2-6	image.file_size	number	Y	파일 크기	이미지 용량(byte) 최소값:1, 최대값:62914560
	2-7	image.width	number	Y	이미지 너비	Pixel 최소값:1, 최대값:99999

	2-8	image.height	number	Y	이미지 높이	Pixel 최소값:1, 최대값:99999
3	3	annotations	array		어노테이션 정보	라벨링 정보
	3-1	annotations[].id	string	Y	라벨링식별자	최소값:1, 최대값:1
	3-2	annotations[].type	string	Y	라벨링 타입	[bbox, polygon]
	3-3	annotations[].bbox	object		바운딩박스 정보	최소값:4, 최대값:4
	3-3-1	bndex_xcrdnt	number		Bounding Box X Coordinate	바운딩박스 X 좌표 최소값:1, 최대값:99999
	3-3-2	bndex_ycrdnt	number		Bounding Box Y Coordinate	바운딩박스 Y 좌표 최소값:1, 최대값:99999
	3-3-3	bndex_width	number		Bounding Box Width	바운딩박스 넓이 최소값:1, 최대값:99999
	3-3-4	bndex_hg	number		Bounding Box Height	바운딩박스 높이 최소값:1, 최대값:99999
	3-4	annotations[].polygon	list		폴리곤 정보	폴리곤 정보
	3-4-1	pyn	number		POLYGON	폴리곤 순번 최소값:1, 최대값:1000
	3-4-2	pyn_xcrdnt	number		POLYGON X Coordinate	폴리곤 X 좌표 최소값:1, 최대값:99999
	3-4-3	pyn_ycrdnt	number		POLYGON Y Coordinate	폴리곤 Y 좌표 최소값:1, 최대값:99999

◦ JSON 예시

```

{
  "plantinfo": {
    "class_name": "아욱",
    "classification_info": "유사",
    "herb_name": "규채",
    "similar_plants": [
      "난쟁리아욱"
    ],
    "toxic_info": "Y",
    "instance_info": {
      "instance_name": "꽃"
    }
  },
  "image": {
    "file_path": "../..../원천데이터/아욱/전초",
    "file_name": "아욱_전초_887778.jpg",
    "shooting_date": "2021-09-30",
    "region_name": "의성군",
    "file_format": "JPG",
    "file_size": 3548726,
    "width": 4000,
    "height": 2252
  },
  "annotations": [
    {
      "id": "1",
      "type": "bbox",
      "bbox": {
        "bndex_xcrdnt": 335,
        "bndex_ycrdnt": 700,
        "bndex_width": 835,
        "bndex_hg": 500
      },
      "polygon": null
    },
    {
      "id": "2",
      "type": "polygon",
      "bbox": null,
      "polygon": [
        {
          "pyn": 1,
          "pyn_xcrdnt": 590,
          "pyn_ycrdnt": 1200
        },
        {
          "pyn": 2,
          "pyn_xcrdnt": 1090,
          "pyn_ycrdnt": 1200
        }
      ]
    }
  ]
}

```

```

        "pyn": 3,
        "pyn_xcrdnt": 1090,
        "pyn_ycrdnt": 1600
    },
    {
        "pyn": 4,
        "pyn_xcrdnt": 590,
        "pyn_ycrdnt": 1600
    }
]
}
]
}

```

○ 구문 규칙

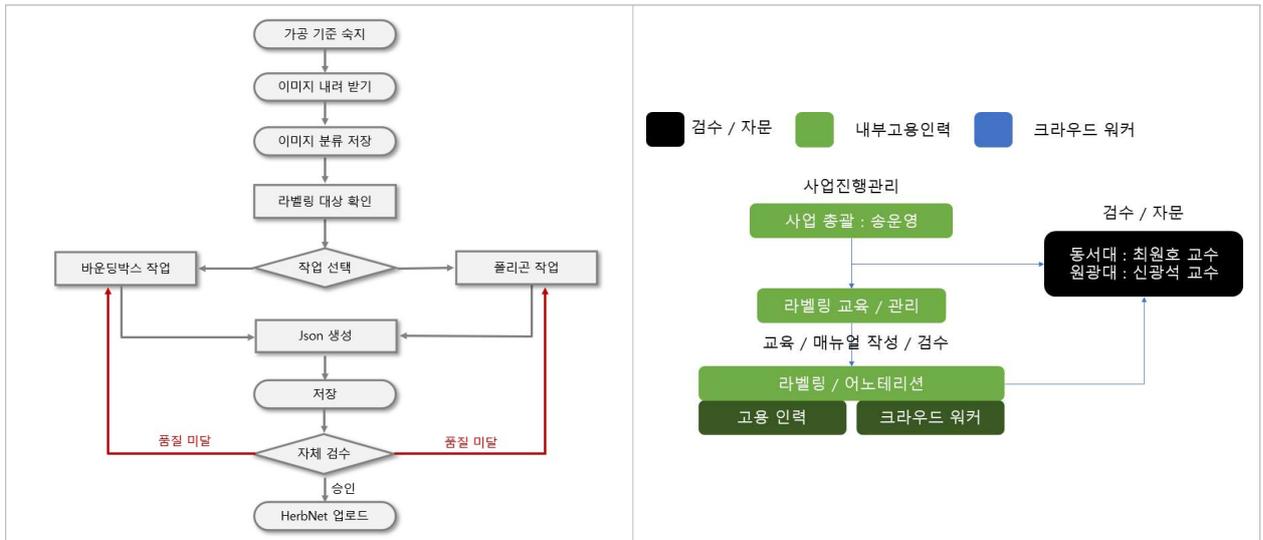
2	2	3	4	5	타입	필수여부	유효값	최소값	최대값	패턴	기타	null타입허용여부	설명
3					object	Y		4				N	클래스 정보
4	class_name				string	Y	"가죽나무","가지","갈대","감국","개오동","곰취","괘이밥","약초","유사"					N	클래스 이름
5	classification_info				string	Y						N	클래스 구분
6	herb_name				string	N	"저근박파","가지","노근","감국","자실","호로철","조강초","용규","하고초","골쇄보","농소화","단삼","지금초","Y					Y	동의보감 독초이름
7	similar_plants				array	N	"가죽나무","가지","갈대","감국","개오동","곰취","괘이밥","깨마중","꿀물","늑골고사리","농소화","단삼","담령<Y					Y	유사식물 항목
8		\$value\$			string	Y						N	유사식물 이름
9	toxic_info				string	Y	"Y","N"					N	독성 여부
10	instance_info				object	Y		1	1			N	인스턴스 정보
11		instance_name			string	Y	"전초","잎","꽃","열매"					N	식물 부위
12					object	Y		7				N	이미지 정보
13	file_path				string	Y		1	200			N	이미지 경로
14	file_name				string	Y		4	100			N	이미지 파일 이름
15	shooting_date				string	Y					date:yyyy-MM-dd	N	촬영 날짜
16	region_name				string	N	"서울경기관", "충청권", "강원권", "경상권", "호남제주권"					Y	[권역정보]서울경기관, 충청권, 강원권, 경상권, 호남-제주권
17	file_format				string	Y	"JPG"					N	파일 포맷
18	file_size				number	Y		1	62914560			N	파일 크기(byte)
19	width				number	Y		1	99999			N	이미지 너비(pixel)
20	height				number	Y		1	99999			N	이미지 높이(pixel)
21					array	Y							라벨링 정보
22	{}				object	Y							
23	id				string	Y		1	1			N	라벨링 식별자
24	type				string	Y	"bbox", "polygon"					N	라벨링 타입
25	bbox				object	N		4	4			Y	바운딩 박스
26		bndex_xcrdnt			number	Y		0	99999			N	바운딩 박스 X 좌표
27		bndex_ycrdnt			number	Y		0	99999			N	바운딩 박스 Y 좌표
28		bndex_width			number	Y		1	99999			N	바운딩 박스 너비
29		bndex_hg			number	Y		1	99999			N	바운딩 박스 높이
30	polygon				array	N							
31		\$value\$			object	Y						N	
32		{}			object	Y						N	
33			pyn		number	Y		1	1000			N	블리크 순번
34			pyn_xcrdnt		number	Y		0	99999			N	블리크 X좌표
35			pyn_ycrdnt		number	Y		0	99999			N	블리크 Y좌표
36													

4.3.2 의미 정확성 검증기준

◦ 라벨링 검증을 위한 클래스별 검수 기준과 오류 유형은 다음과 같다

인스턴스명	검수 기준	기준 설명
전초	라벨 정밀도	121종 식물의 전초 영역을 라벨링한다 오태깅: - 121종 식물 분류 오태깅 - 식물 부위 오태깅 과태깅: - 정의되지 않은 식물 분류 태깅 - 정의되지 않은 식물 부위 태깅
잎	라벨 정밀도	121종 식물의 잎 영역을 라벨링한다 오태깅: - 121종 식물 분류 오태깅 - 식물 부위 오태깅 과태깅: - 정의되지 않은 식물 분류 태깅 - 정의되지 않은 식물 부위 태깅
꽃	라벨 정밀도	121종 식물의 꽃 영역을 라벨링한다 오태깅: - 121종 식물 분류 오태깅 - 식물 부위 오태깅 과태깅: - 정의되지 않은 식물 분류 태깅 - 정의되지 않은 식물 부위 태깅
열매	라벨 정밀도	121종 식물의 열매 영역을 라벨링한다 오태깅: - 121종 식물 분류 오태깅 - 식물 부위 오태깅 과태깅: - 정의되지 않은 식물 분류 태깅 - 정의되지 않은 식물 부위 태깅

4.4 데이터 가공 조직



4.5 데이터 가공 도구

가공 도구	가공 메뉴얼
 <p>labelme</p>	 <p>LabelMe는 컴퓨터 비전 연구를 위한 이미지 데이터 베이스를 구축하기 위한 온라인 주석 도구를 제공하는 공개 저작도구로, 본 과제를 통해 획득한 이미지 데이터 가공에 적합하여 저작도구로 사용함.</p>

4.5.1 데이터 가공 작업



Herbnet

과학기술정보통신부, NIA(한국정보화진흥원)가 추진하는 '인공지능 학습용 데이터 구축 사업(2차)' 선정과제

동의보감 약초판별 인공지능 학습용 데이터 구축

'동의보감 약초 판별 인공지능 학습용 데이터 구축'은 매년 독초를 약초로 오인, 남용하여 발생하는 독초에 대한 중독 및 사망 사고를 예방할 수 있는 약초와 독초 판별 인공지능서비스와 전문 지식 없이 약초를 과다 복용하거나, 무조건 좋다는 생각으로 음용하여 발생되고 있는 피해를 줄이기 위한 정확한 약초 사용 방법을 제공하기 위해 동의보감 수록 457종 중 올해 선정할 약초 50종, 비교식물 82종의 이미지 데이터와 정확한 약초 사용 방법을 위한 약초 정보를 구축하는 사업입니다.

데이터 가공은

허브넷(www.herbnet.kr)을 통해 진행되며

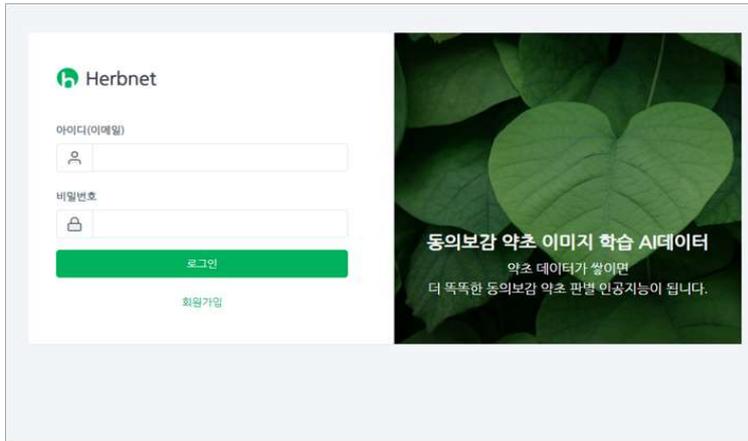
가공 작업에 참여하는 참여자는 허브넷에 가입하여야 한다.

가입 한 후 가공 책임자에게 아이디(허브넷 가입 이메일)를 알려주고 작업을 배정받는다

가공 작업 배정은 품목(독초) → 가공 단계별로 배정된다

가공 방법은 매뉴얼에 따라 수행하면 되고 궁금한 사항은 단체 채팅방을 통해 선임자들에게 요령을 익힌다

② 허브넷에 로그인한다.



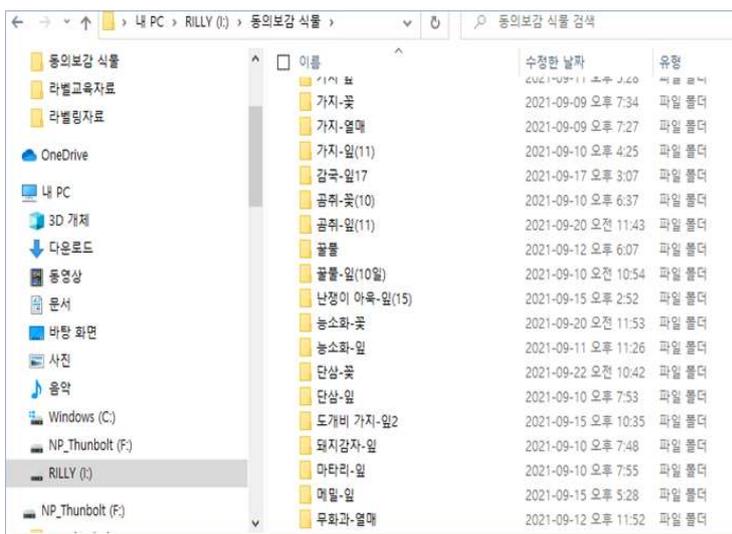
로그인
 허브넷(www.herbnet.kr)에 접속 후
 아이디(가입 이메일)와 비밀번호 입력
 접속 한다

③ 원천 이미지 데이터 다운로드



정제 완료된 이미지 다운로드 받기

④ 다운로드 폴더 지정 후 원천이미지 데이터 저장



다운로드 받은 이미지 폴더 관리하기

③ 다운로드 받은 독초 이미지는
 작업하기 편리하게 식물별로 분류하여
 파일 관리

⑤ 데이터 라벨링 수행(레이블 미실행)



④

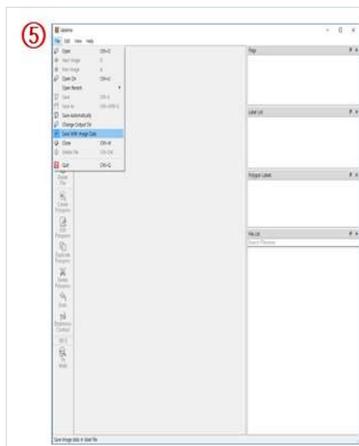
Labelme.exe

Double click this icon

저작도구 실행하기

④ 가공작업을 시작하기 위해 **Labelme.exe**를 클릭

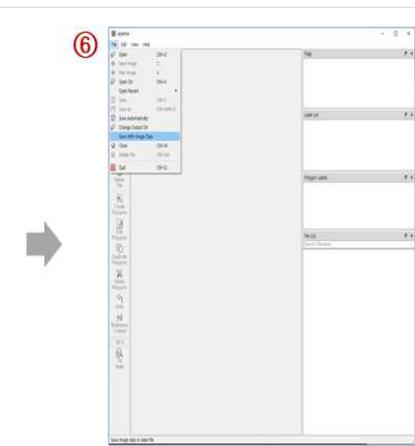
⑥ 데이터 라벨링 작업 수행



⑤

Save with Image Data

➔



⑥

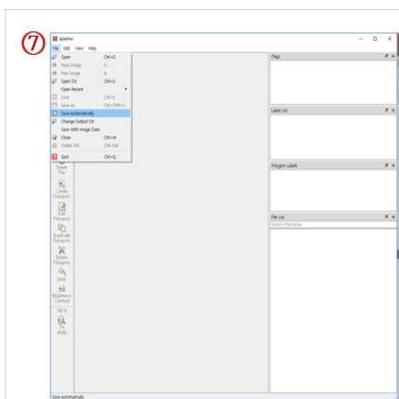
Save with Image Data

주요 항목 체크

⑤ 작업전에 반드시 File에서 **Save with Image Data**

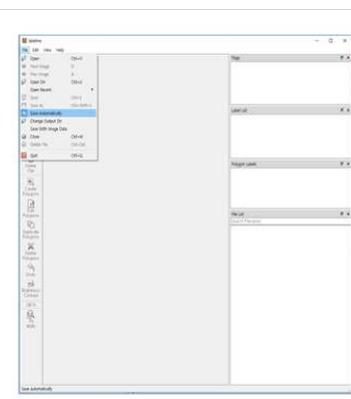
⑥ 클릭하여 해제한다.

⑦ 데이터 라벨링 작업 수행



⑦

➔



⑦

Save Automatically

➔

Save Automatically

주요 항목 체크

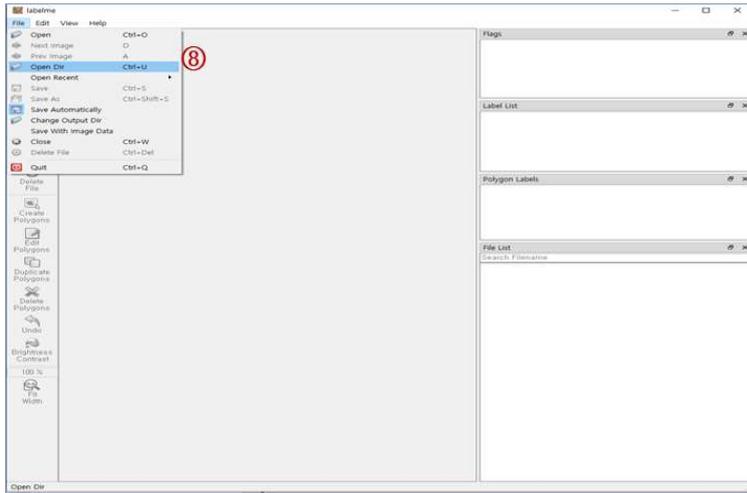
File에서

Save with Image Data를 체크 후

⑦ **save Automatically**클릭 후

작업을 진행해야 **Json 파일이 자동으로 저장됨**

⑧ 데이터 라벨링 작업 수행

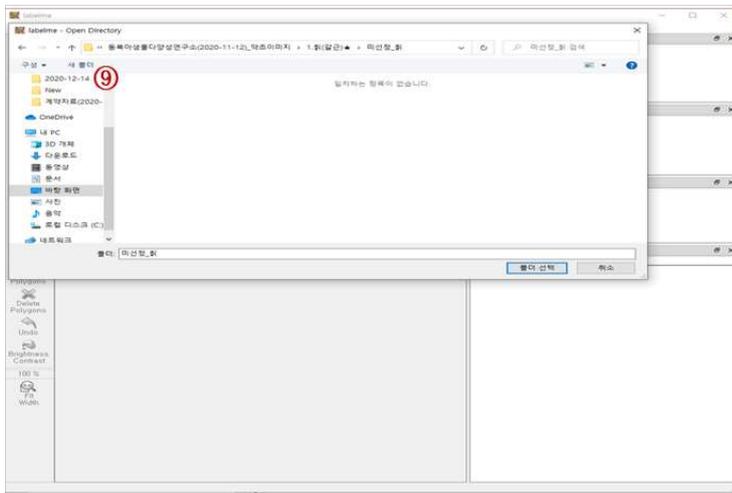


주요 항목 체크

⑧ 앞장에 File에서 Save with Image Data 를 반드시 클릭 Save Automatically를 클릭한 후

Click the Open Dir 선택 하면 다운로드 된 파일을 불러옴.

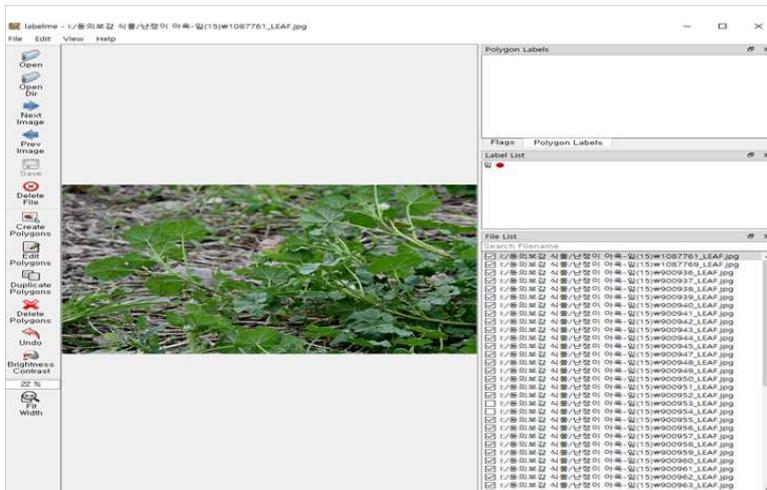
⑨ 데이터 라벨링 작업 수행



다운로드 이미지 파일 열기

⑨ 이미지 디렉토리 및 폴더 버튼 클릭 다음 가공 대상 사진으로 쉽게 이동

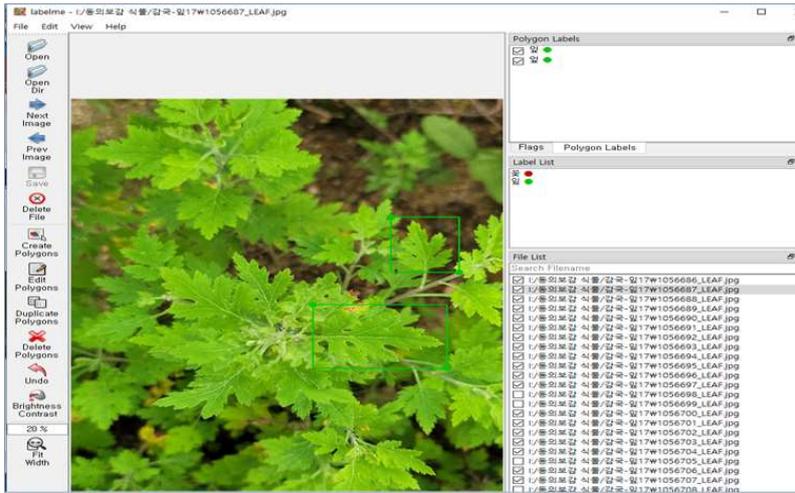
⑩ 데이터 라벨링 작업 수행



파일 불러오기

⑩ 앞의 모든 이미지가 표시됩니다

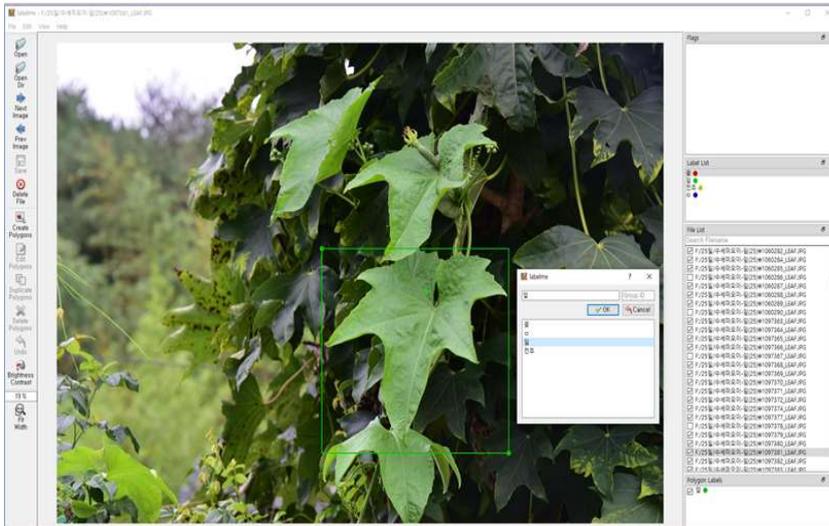
⑪ 데이터 라벨링 작업 수행



이미지 데이터 라벨링

⑫ 사각형이 만들어진 이미지를 확인

⑫ 데이터 라벨링 작업 수행

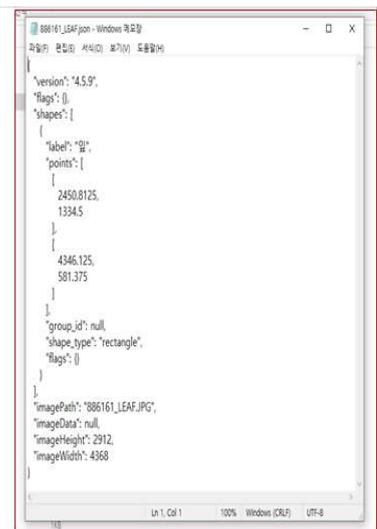


이미지 데이터 라벨링

⑬ 이름 쓰기 : _잎

⑭ OK (확인버튼) 클릭

⑬ Json 파일 저장



4.5.2 데이터 가공 예시

① 목본(예시)



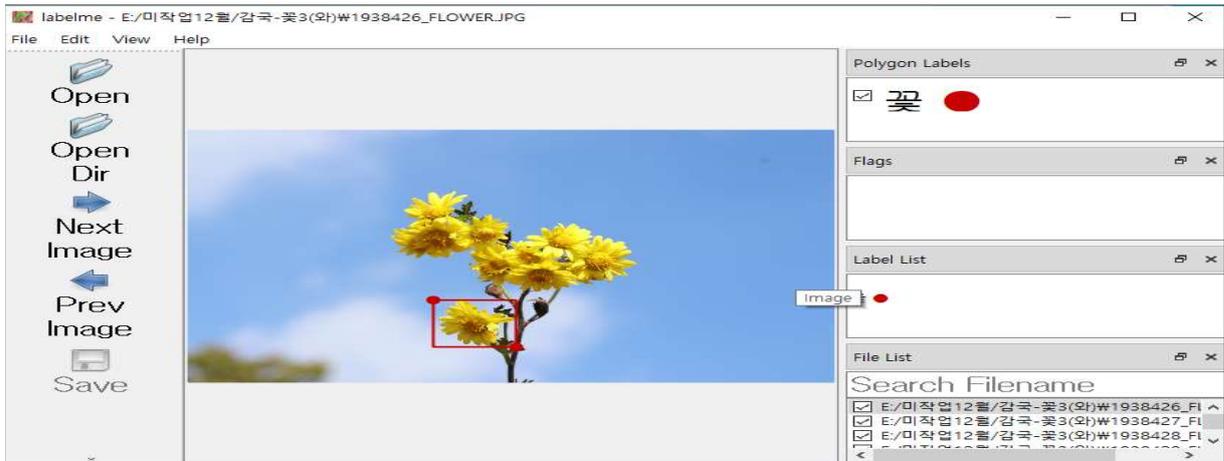
② 초본(예시)



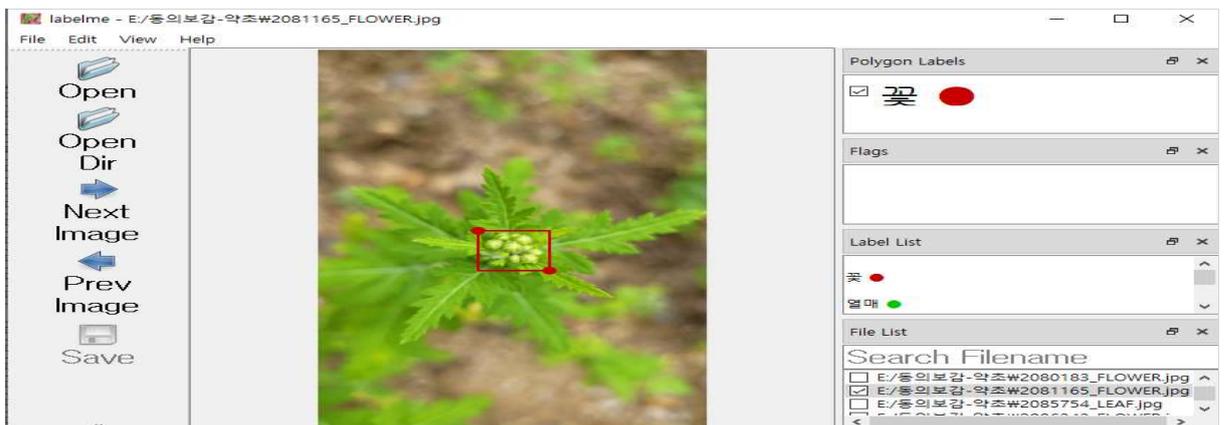
③ 덩굴(예시)



④ 홀로 피기 꽃(예시)



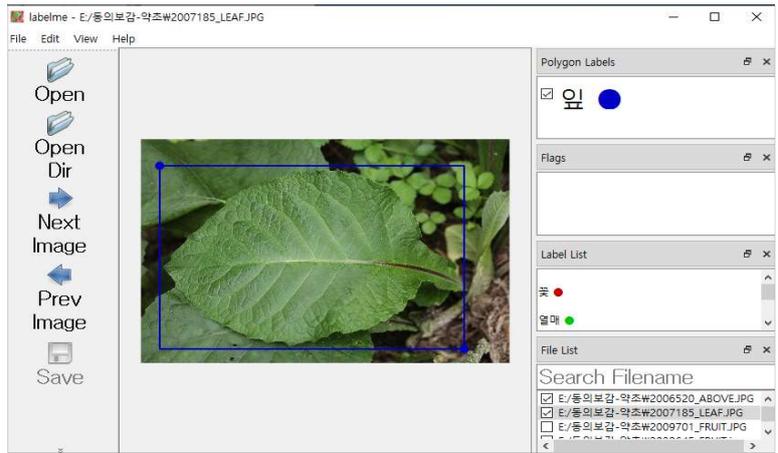
⑤ 홀로 피기 꽃(예시)



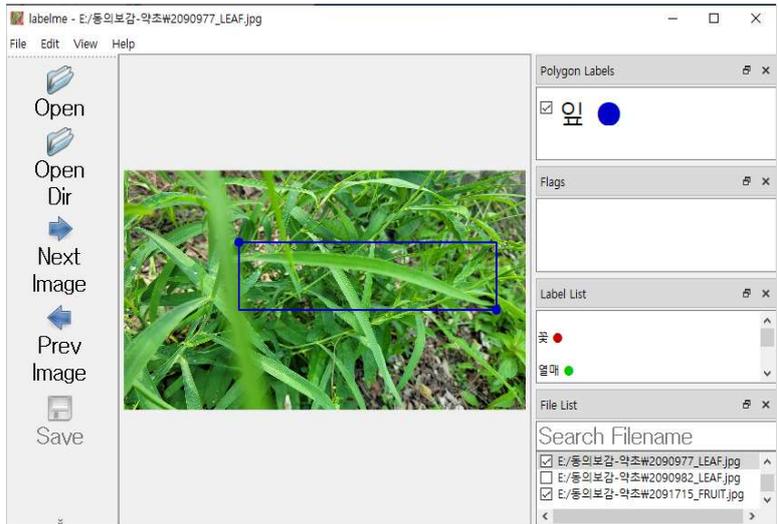
⑥ 차례 피기 꽃(예시)



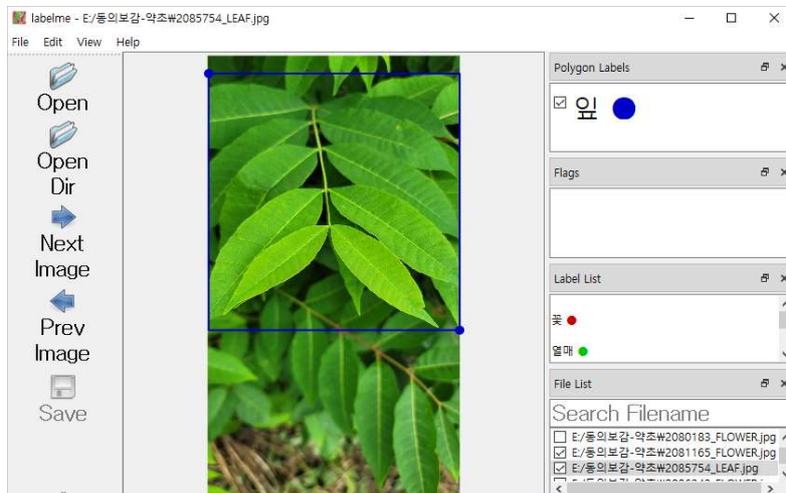
⑦ 한 잎(예시)



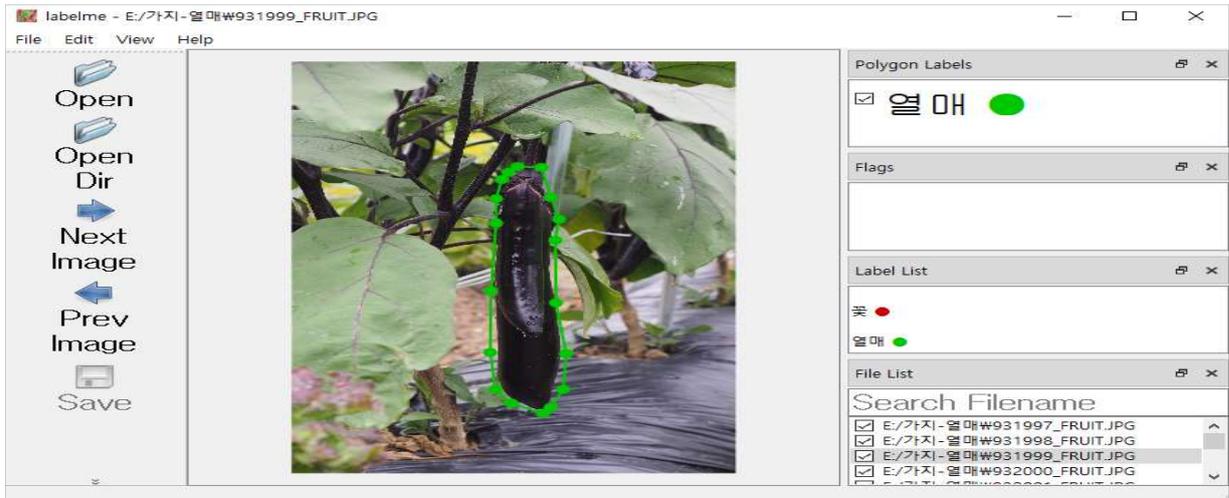
⑧ 긴 잎(예시)



⑨ 여러 잎(예시)



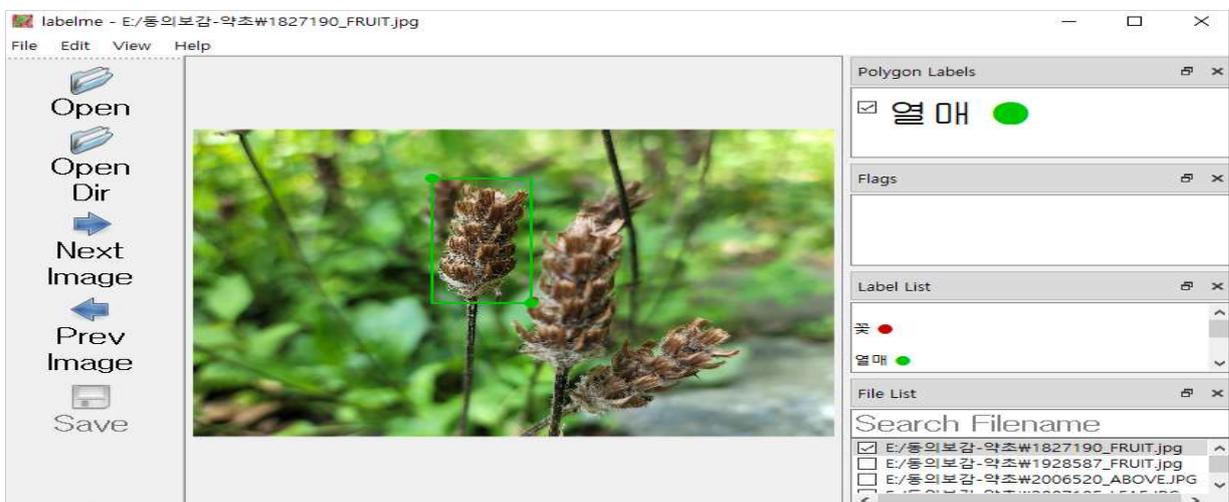
⑩ 홀로 맺기(예시)



⑪ 모여 맺기(예시)



⑩ 차례 맺기(예시)



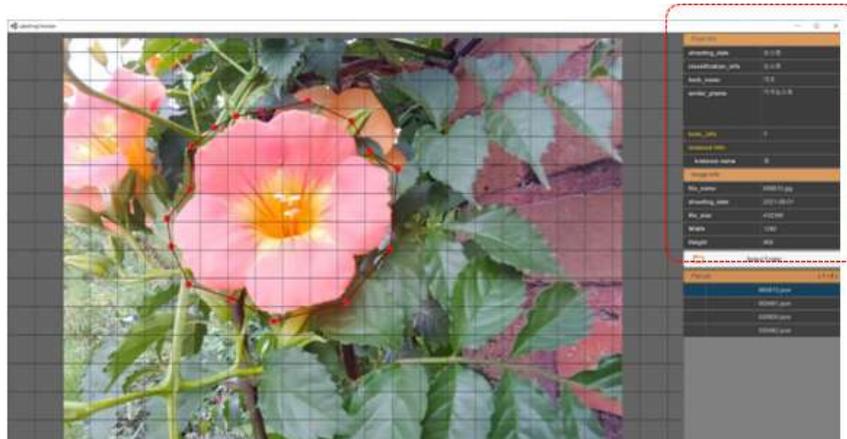
5. 검수

5.1 검수 절차 및 방법



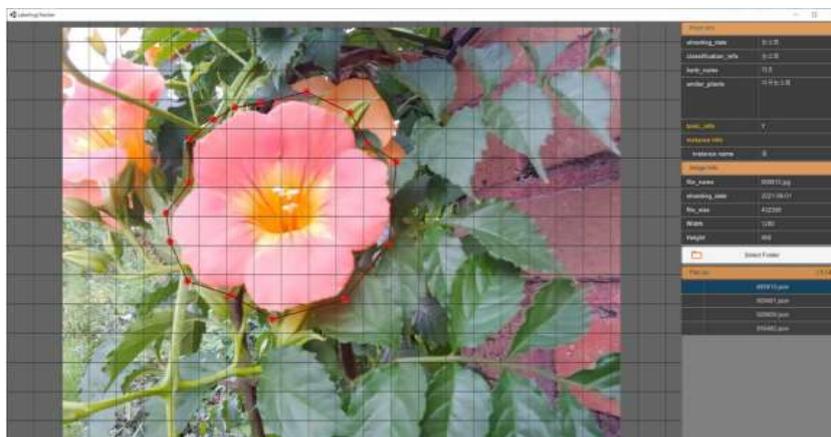
5.1.1 검수 방법

가. 구문 정확성 검수



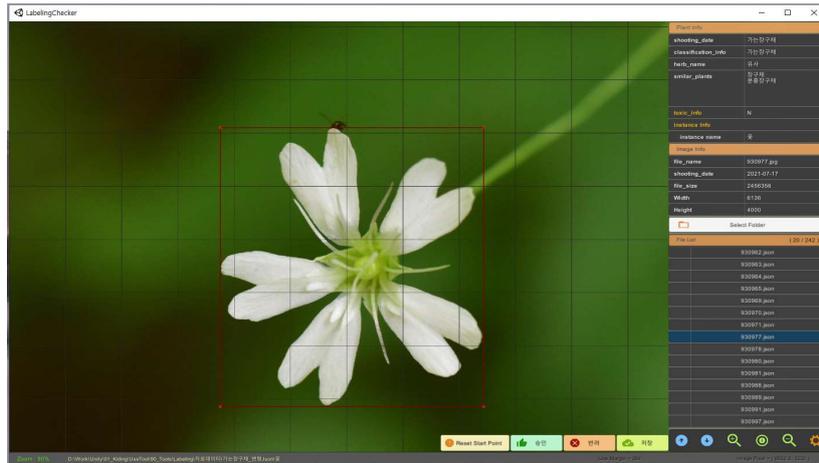
빨간 박스 부분 메타데이터에 기록된 정보와 이미지 육안 확인

나. 의미 정확성(폴리곤) 검수



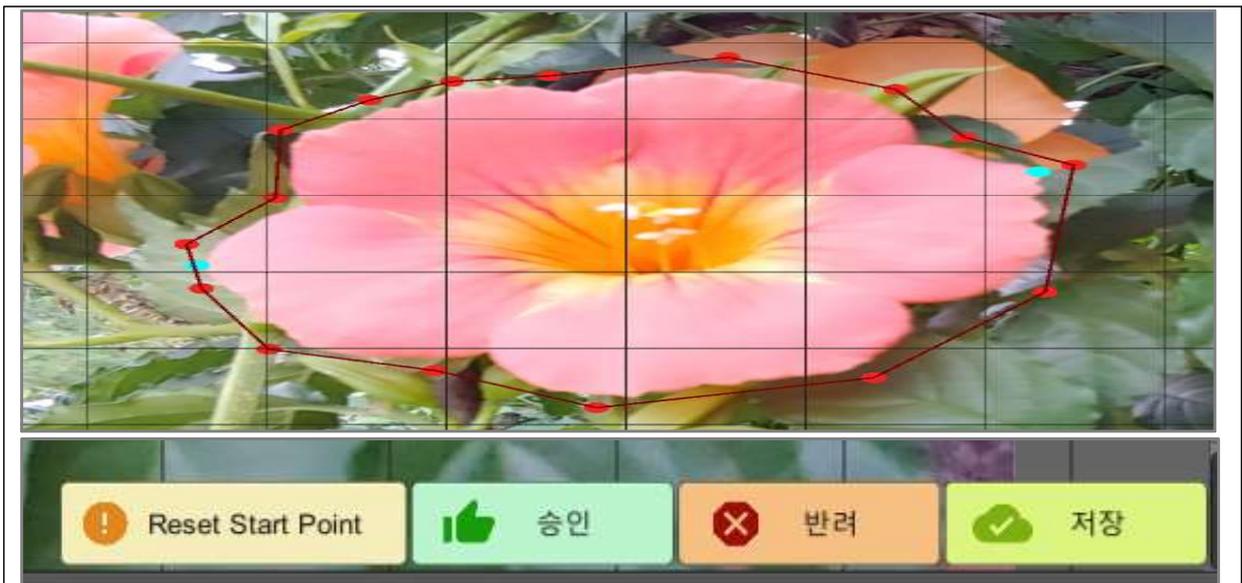
라벨링 대상 객체 테두리가 20% 범위를 벗어나지 않으면 승인

다. 의미 정확성 (바운딩 박스) 검수



라벨링 대상 객체 테두리가 20% 범위를 벗어나지 않으면 승인

라. 검수 결과 처리



- Reset Start Point 버튼 : 점을 찍어서 체크 라인을 다시 그릴 수 있게 초기화한다.
- 승인 버튼: 검수 결과 합격일 경우 승인하기
- 반려 버튼: 검수 결과 불합격일 경우 반려하기
- 저장 버튼: 반려된 정보를 파일에 저장한다.

※ 벌레 먹은 객체, 객체에 벌레가 있거나, 사람이 등장하거나, 차량이 등장한 경우는 반려.

5.2 검수 기준

대상 부위	검수 기준	비고
전초	<ul style="list-style-type: none"> • 전초에서 꽃, 잎, 열매를 라벨링 한 경우도 전초로 인정함. • 전초에서 잎, 꽃, 열매만을 분리하여 라벨링 한 경우도 허용범위에 포함되어 오타킹으로 분류하지 않고 전초를 라벨링 할 때 일부 부위가 잘린 경우도 오타킹으로 분류하지 않음 • 자연환경에서 자란 전초의 열매. 잎이 다 있는 경우는 전체를 라벨링 함 → 꽃, 잎, 열매가 없는 경우는 해당 개체만 라벨링 함. • 자연환경에서 식물이 성장하면서 서로 겹쳐있는 이미지는 원본 이미지로 여러 개체를 라벨링 할 수 있음. 	승인
잎	<ul style="list-style-type: none"> • 잎 여러 장이 겹쳐 있는 경우, 묶어서 라벨링 한 경우 • 긴 잎의 경우는 전초와 잎 구분이 어려운 경우 전초 또는 잎으로 라벨링 한 경우. • 잎은 잎 모양을 기준으로 식물을 분류하기에 잎의 90% 이상을 라벨링 한 경우, 모양을 확인하는데 영향이 제한적이라 오타킹 아님. • 여러 잎인데 개별 잎 하나를 라벨링하고 잎 모양이 가능한 경우 • 잎의 성장 과정에 따라 단풍이 든 잎은 잎 훼손 아님. • 한 장의 이미지에서 잎 여러 개를 라벨링 한 경우 • 한 장의 이미지에서 여러 잎 여러 개를 라벨링 한 경우, 잎 옆 일부가 잘려있어도 잎 구분이 가능한 경우 • 한 잎으로 한 잎 성장 특성상 잎이 겹쳐 나는 경우, 겹쳐 나는 잎 일부를 포함하여 라벨링 한 경우 • 잎 모양을 확인 가능한 수준에서 라벨링 됨 경우 • 잎의 끝 일부가 잘리거나, 여러 잎 일부만 라벨링 된 경우 	승인

<p>꽃</p>	<ul style="list-style-type: none"> • 차례 피기의 경우 꽃 특성상 차례대로 피기 때문에 꽃망울과 꽃이 하나의 이미지에 있는 경우에는 묶어서 라벨링 함. • 모여 피기 꽃은 지는 것과 열매가 맺는 과정이 동시에 이루어지는 특징이 있어, 열매랑 꽃을 묶어 라벨링 한 경우 • 홀로 피기 꽃의 경우 여러 개의 꽃이 동시에 피기 때문에 묶어 라벨링 한 경우 • 모여 피기 꽃은 작은 꽃이 피어 하나의 모양을 이룸으로 일부 꽃이 잘릴 경우 • 벌레나 이물질이 붙어있는 경우 → 반려 처리 • 한 장에 최대 5개 이내의 라벨링 한 경우 • 같은 종의 꽃을 겹치게 라벨링 한 경우 • 대추나무꽃은 작은 개체가 특징으로 작은 꽃을 라벨링 한 경우 • 꽃 모양을 확인 가능한 범위에서(꽃이 90% 이상 포함된 경우)라벨링 한 경우 • 피미자 꽃처럼 꽃과 열매가 섞여 있는 경우, 꽃과 열매를 묶어 라벨링 한 경우 • 모여 피기 꽃 여러 개를 묶어서 라벨링 한 경우 • 모여 피기 여러 개의 꽃 중 일부 꽃만 라벨링 한 경우 • 여뀌, 개여뀌, 쪽의 경우 꽃 모양이 매우 작아 꽃만을 라벨링 한 경우 	<p>승인</p>
<p>열매</p>	<ul style="list-style-type: none"> • 차례 피기의 경우 꽃 특성상 차례대로 피기 때문에 꽃망울과 꽃이 하나의 이미지에 있는 경우에는 묶어서 라벨링 함. • 벌레나 이물질이 붙어있는 경우 라벨링 함. • 작은 열매가 모여 있는 경우 묶음으로 전체를 라벨링 • 열매 여러 개가 겹쳐 있는 경우 • 열매가 잎을 포함하여 라벨링 한 경우 • 깨꽃 열매는 꽃순이 빠지고 꽃 모양이 퇴색하며 열매를 맺는 특성을 가지고 있어 처음 열매는 꽃과 구분이 어려움 꽃과 열매를 묶어서 라벨링 함. • 열매가 떨어지기 직전에는 열매 모양이 흩어지는 경우가 있어 이런 열매를 라벨링 한 경우 가능함 	<p>승인</p>

	<ul style="list-style-type: none"> • 착과 후 열매에 털이 있는 경우 털 일부가 잘리게 라벨링 한 경우 • 물봉선 열매는 꽃이 떨어지면서 열매가 맺기 시작하고, 열매가 겹쳐 있게 라벨링 한 경우 • 돼지감자 열매는 꽃잎이 떨어지고 열매가 맺기 시작함 꽃잎이 떨어진 경우, 열매로 라벨링 한 경우 • 완두, 동부 등 협과 열매는 열매 형태를 따라 바운딩 박스 또는 폴리곤 • 열매 일부가 잘렸지만 열매 모양 확인 가능한 범위인 열매가 90% 이상 포함되게 라벨링 한 경우 • 열매가 겹친 경우, 각각 열매를 라벨링 한 경우 • 차례 맺기 열매의 경우 열매만 라벨링 한 경우 • 열매가 뭉쳐있지만 하나의 열매만 라벨링 한 경우 	
--	--	--

5.3 검수 조직



◦ 검수 결과

구분	클래스 검수	검수 수량	인스턴스 검수	검수 수량
가공 데이터	121	747,869	꽃	218,471
			열매	164,391
			잎	236,953
			전초	128,105
학습 데이터	121	807,068	꽃	240,194
			열매	179,003
			잎	259,766
			전초	128,105

품질 특성	검증 항목	측정지표	정량 목표	품질 검수 결과
구문 정확성	구문 정확도	정확도(%)	99.9% 이상	품질 목표 달성
의미 정확성	전초	정밀도(%)	99% 이상	품질 목표 달성
	꽃	정밀도(%)	99% 이상	품질 목표 달성
	잎	정밀도(%)	99% 이상	품질 목표 달성
	열매	정밀도(%)	99% 이상	품질 목표 달성

5.4 검수 도구

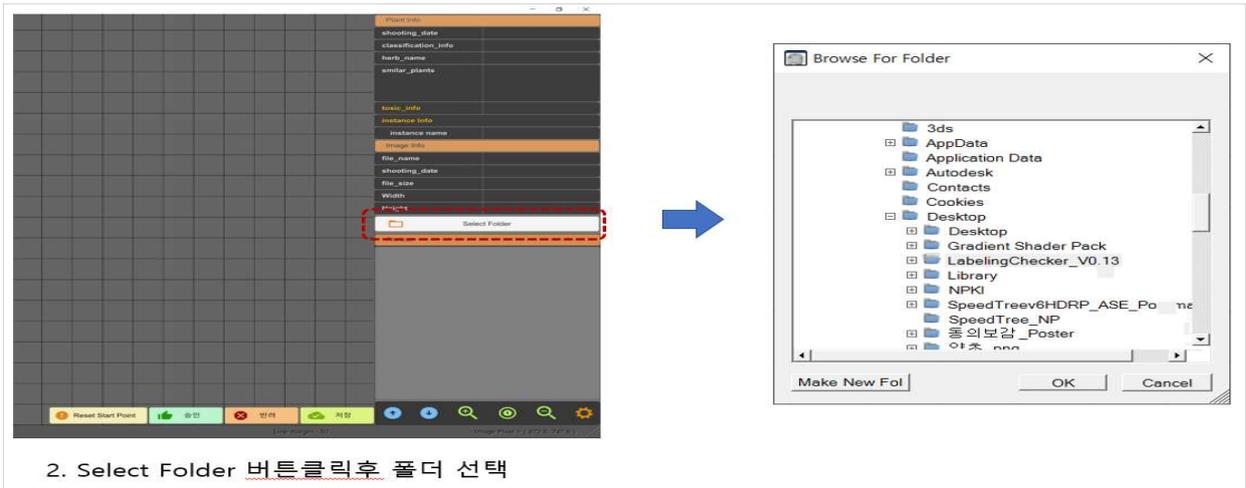
가. 검수 툴(TOOL) 설치 방법

데이터 검수는 자체 개발한 검수 도구인 Labeling Checker를 이용하여 전수 검수를 수행함.

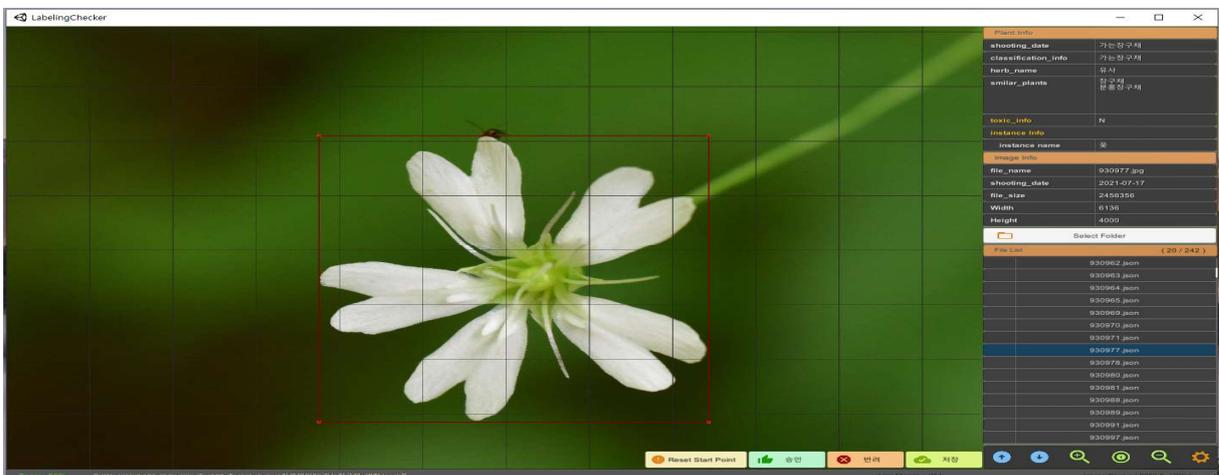


나. 검수 방법

Labeling Checker 실행 후 검수 대상 파일 열기



1) 라벨링 데이터 검수 수행



2) 라벨링 데이터 검수 결과 처리

데이터 가공 기준에 맞지 않게 라벨링 된 파일을 반려 처리함.



The screenshot shows a software interface for labeling data. At the top, there is a grid overlay on a photograph of a pink flower. A red polygonal bounding box is drawn around the flower. Below the image is a control panel with four buttons: 'Reset Start Point' (yellow with an exclamation mark icon), '승인' (green with a thumbs-up icon), '반려' (orange with a red X icon), and '저장' (light green with a checkmark icon).

- Reset Start Point 버튼 : 점을 찍어서 체크 라인을 다시 그릴 수 있게 초기화한다.
- 승인 버튼: 검수 결과 합격일 경우 승인하기
- 반려 버튼: 검수 결과 불합격일 경우 반려하기
- 저장 버튼: 반려된 정보를 파일에 저장한다.

5.5 기타 품질관리 활동

5.5.1 외부 전문기관에 의뢰하여 품질 검증작업 수행.

가. 품질검증 방법

전체 AI 학습용 데이터셋 중 현재까지 구축을 완료한 데이터 중 3%(18,634건)를 무작위 샘플링 하여 취득한 데이터를 대상으로 외부기관에 의뢰하여 품질검증 수행.

나. 품질검증 내용

대상 사업은 AI 기술 · 응용 서비스 개발에 활용 가치가 높은 AI 학습용 데이터 구축 · 개방 및 AI 응용 서비스 개발을 목적으로 추진되는 사업의 하나로 동의보감 독초 판별 이미지 데이터 600,000장의 AI 학습데이터 구축을 목표로 추진되는 사업임.

외부기관(3자) 품질검증은 제출된 AI 학습용 데이터셋에서 약 3% 범위의 무작위 샘플링을 통해 식별된 데이터에 대해 “데이터 구축 가이드라인”의 구축 방법에 따라 구문 및 의미적 정확성이 목표를 충족하는지를 외부기관의 객관적 시각으로 검증.

다. 품질검증 수행

가공 완료 데이터(747,595건, 목표 600,000건 대비 초과 구축)를 기준으로 무작위 샘플링한 18,634건(목표 가공량 대상 샘플링 비율 : 약 3.07%)에 대해 가천대학교 산학협력단

컨소시엄이 수립한 품질 특성(구문 정확성과 의미 정확성)의 측정지표별 목표를 달성하고 있는지에 대해 점검한 결과, 목표 품질 범위에서 적정하게 관리 중인 것으로 판단.

품질특성	항목명	측정 지표	정량 목표	제 3자 품질검증 결과
구문정확성	구문 정확도	정확도(%)	99.9% 이상	품질수준 100.0% 유지
의미 정확성	전초	정밀도(%)	99% 이상	품질수준 99,80% 유지
	잎	정밀도(%)	99% 이상	품질수준 99,79% 유지
	꽃	정밀도(%)	99% 이상	품질수준 99,22% 유지
	열매	정밀도(%)	99% 이상	품질수준 99,08% 유지

1) 구문 정확성 검증

구문 정확성은 18,634건의 데이터를 대상으로 품질검증합의서에 명시된 구문 형식을 적정하게 준수하고 있는지에 대해서 점검한 결과 정량 품질 목표 범위에서 적정하게 관리 중인 것으로 판단하였음.

다만, annotations[].id(라벨링 식별자)의 경우, 식별자 역할이 아닌 구분의 의미("1", "2") 값으로 채워져 있어 적정한 값인지 확인이 필요함.

구문 정확성	점검 대상	오류율 계산식	오류건수	점검 결과
데이터 구조	레이블링 데이터	$\text{오류율}(\%) = \frac{\text{오류건수(데이터수)}}{\text{검사총건수(데이터수)}} \times 100(\%)$	0	100.0%
데이터 값	레이블링 데이터	$\text{오류율}(\%) = \frac{\text{오류건수(데이터수)}}{\text{검사총건수(데이터수)}} \times 100(\%)$	0	100.0%
데이터 형식	레이블링 데이터	$\text{오류율}(\%) = \frac{\text{오류건수(데이터수)}}{\text{검사총건수(데이터수)}} \times 100(\%)$	0	100.0%
평균 품질				100.0%

2) 의미 정확성 검증

의미 정확성은 121종 식물 데이터 18,634건의 데이터와 17,732건의 바운딩박스, 902건의 폴리곤 세그먼테이션 레이블링 결과를 대상으로 품질검증합의서의 의미 정확성 검증기준에 따라 레이블링을 적정하게 구축하고 있는지를 점검한 결과 정량 품질 목표 범위에서 적정하게 관리 중인 것으로 판단.

의미 정확성		오류율 계산식	오류건수	점검 결과
전초	정밀도	$\text{정밀도}_{\text{바운딩박스}} = \frac{\text{바운딩박스개수} - \text{분류오류개수} - \text{위치오류개수}}{\text{바운딩박스생성개수}}$	10	99.80%
잎	정밀도	$\text{정밀도}_{\text{바운딩박스}} = \frac{\text{바운딩박스개수} - \text{분류오류개수} - \text{위치오류개수}}{\text{바운딩박스생성개수}}$	10	99.79%
꽃	정밀도	$\text{정밀도}_{\text{바운딩박스}} = \frac{\text{바운딩박스개수} - \text{분류오류개수} - \text{위치오류개수}}{\text{바운딩박스생성개수}}$	35	99.22%
열매	정밀도	$\text{정밀도}_{\text{바운딩박스}} = \frac{\text{바운딩박스개수} - \text{분류오류개수} - \text{위치오류개수}}{\text{바운딩박스생성개수}}$	41	99.08%

라. 품질검증 결과

1) 품질검증 내용

품질 특성	검증 항목	측정 지표	정량 목표	외부기관 품질검증 결과
구문 정확성	구문 정확도	정확도(%)	99.9% 이상	품질수준 100% 유지
의미정확성	전초	정밀도(%)	99% 이상	품질수준 100% 유지
	꽃	정밀도(%)	99% 이상	품질수준 100% 유지
	잎	정밀도(%)	99% 이상	품질수준 100% 유지
	열매	정밀도(%)	99% 이상	품질수준 100% 유지

가) 구문 정확성

구문 정확성	점검 대상	오류율 계산식	오류 건수	외부기관 품질검증 결과
데이터 구조	레이블링 데이터	$오류율(%) = \frac{오류 건수(데이터수)}{검사총건수(데이터수)} \times 100$	0	100%
데이터 값	레이블링 데이터	$오류율(%) = \frac{오류 건수(데이터수)}{검사총건수(데이터수)} \times 100$	0	100%
데이터 형식	레이블링 데이터	$오류율(%) = \frac{오류 건수(데이터수)}{검사총건수(데이터수)} \times 100$	0	100%

나) 의미 정확성

의미 정확성		오류율 계산식	오류 건수	점검 결과
전초	정밀도	$정밀도바운딩박스 = \frac{바운딩박스 개수 - 분류 오류 개수 - 위치 오류 개수}{바운딩박스 생성 개수}$	10	99.80%
잎	정밀도	$정밀도바운딩박스 = \frac{바운딩박스 개수 - 분류 오류 개수 - 위치 오류 개수}{바운딩박스 생성 개수}$	10	99.79%
꽃	정밀도	$정밀도바운딩박스 = \frac{바운딩박스 개수 - 분류 오류 개수 - 위치 오류 개수}{바운딩박스 생성 개수}$	35	99.2%
열매	정밀도	$정밀도바운딩박스 = \frac{바운딩박스 개수 - 분류 오류 개수 - 위치 오류 개수}{바운딩박스 생성 개수}$	41	99.08%

마. 외부기관 품질검증 결과 의견

1) 구문 정확성 검증 결과

- 샘플링 데이터는 입력 유효값을 모두 충족하고 있음
- 그러나 annotations[].id(라벨링 식별자)의 경우 식별자로서 역할이 아닌 구분의 의미("1", "2") 값으로 채워져 있어 적절한 값인지 확인이 필요함

2) 의미 정확성 검증 결과

- 일부 식물에서 오 태깅, 과 태깅 등 결함이 파악되었으나 목표 범위 내에서 관리 중임
- 태깅 오류는 아니지만, 품질검증 합의서에 의미 정확성 검증기준에 부합하지 않은 62건의 오류가 발견되었음
- 학습데이터 품질이 확보될 수 있도록 검토 후 반영하는 것이 필요함

[부록] #별첨 - 1 식물 잎 유형

1. 잎의 유형

식물 잎의 유형은 잎몸이 한 장으로 이루어진 홑잎과 잎몸이 2개 이상으로 이루어진 겹잎으로 나눔

- 단엽 : 잎이 작게 갈라져 있지 않고, 하나로 되어 있는 잎을 말함
- 복엽 : 잎이 2개 이상의 작은 잎으로 이루어진 잎의 모양을 말함.

[표 1-1] 잎의 유형, 출처 - 허브넷	
단엽	복엽
	

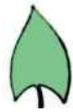
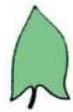
[표 1-2] 복엽의 유형, 출처 - 개인 블로그(한울)				
구분	잎 유형	모양	잎 유형	모양
복엽	장상복엽		삼출 복엽	
	5출 복엽		우상 복엽 잎 수(짝수)	 
	기수 우상복엽 잎 수(홀수)		2회 3출복엽	
	3회 3출 복엽		조족상 복엽	

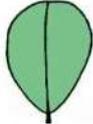
2. 잎의 나기

[표 1-3] 잎의 나기(차례) / 이미지 출처(wildlife.kr)		
잎 나기	정의	모양
마주나기 (대생-對生)	한 개의 마디에 잎이 두 장씩 마주 붙어 남	
		
		
		
어긋나기 (호생-互生)	한 개의 마디에 잎이 한 장씩 어긋나게 붙어 남	
		
		
		
돌려나기 (윤생-輪生)	한 개의 마디에 세장 이상씩 붙어 남	
		
		
		
뭉쳐나기 (총생-叢生)	여러 개의 잎이 짧은 줄기에 뭉쳐서 남	
		
		
		

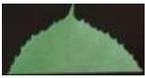
3. 잎의 모양

[표 1-2] 잎의 모양 - 출처(꽃다비-개인 블로그)				
번호	생김새	영어	한글	모양
1	원형	orbicular	동그란 모양	
2	아원형	roundish	동그스런 모양	
3	난형	ovate	달걀 모양	
4	도란형	obovate	거꾸로 세운 달걀모양	
4	타원형	elliptical	타원 모양	
5	장타원형	oblong	길쭉한 타원모양	
6	광타원형	oval	넓은 타원모양	
7	피침형	lanceolate	창 모양	
8	도피침형	oblanceolate	거꾸로 된 창 모양	
9	침형	acicular	바늘모양	

				
10	설형	lingulate	혀모양	
11	선형	linear	직선모양	
12	주걱형	spatulate	주걱 모양	
13	신장형	reniform	신장 모양	
14	심장형	cordate	심장 모양	
15	도심장형	obcordate	거꾸로된 심장모양	
16	신월형	crescent	초승달 모양	
17	능형	rhomboid	마름모 모양	
18	화살형	sagittate	화살촉 모양	
19	심장상 화살형		심장형태 화살촉 모양	
20	극형	hastate	창 끝 모양	

21	삼각형	deltoid	세모 모양	
22	제금형	pandurate	바이올린 모양	
23	민들레형	runcinate	밑으로 향한 톱니 모양	
24	썰기형	wedge shap	썰기 모양	

4. 잎의 끝 모양

[표 1-4] 잎 끝 모양 / 출처(꽃다비-개인 블로그)		
잎 끝	정의	모양
점첨두	점점 길게 뾰족해진 잎 끝	
급첨두	잎맥만이 자라서 잎 끝이 가시와 같이 뾰족한 잎	
첨두	잎몸의 끝이 뾰족한 모양보다 짧으며, 그 이루는 각도가 45~90도를 표현한 것	
원두	원형인 잎 끝	
평두	자른 것처럼 밋밋한 잎 끝	
예두	뾰족한 잎 끝	
둔두	둔한 잎의 끝	

5. 잎의 아랫부분 모양

[표 1-5] 잎 기부(시작) / 이미지 출처(수원 꽃집, 산림청 블로그, 농수산대학-권영휴)		
잎 기부(시작)	정의	모양
유저	엽신의 양쪽 가장자리 밑이 합쳐지지 않고, 잎자루의 날개처럼 된 기부	
설저	잎 아랫부분이 혀 모양	
예저	잎 아랫부분이 좁아지면서 뾰족해지는 모양	
원저	잎 아랫부분이 둥근 모양	
평저	잎 아랫부분이 약 180도 각으로 평평한 모양	
순저	잎 아랫부분이 방패 모양	
이저	잎 아랫부분이 사람의 귀 모양	
심장저	잎 아랫부분이 심장 모양	
왜저	잎 아랫부분이 좌우 양측이 대칭을 이루지 않고 한쪽이 일그러진 모양	
둔저	잎 아랫부분의 양쪽 가장자리가 90도 이상의 각으로 합쳐져 둔한 모양	

6. 잎의 가장자리(엽연)

[표 1-6] 잎 가장자리 모양 / 출처(꽃다비-개인 블로그, 설명-산림청 용어집)		
잎 가장자리	정의	모양
치아상 거치	크기가 비슷한 뾰족한 톱니 모양의 잎 가장자리 밖을 향하여 뾰족하게 뺀 커다란 톱니 같은 모양	
침상 거치	바늘처럼 가늘고 끝이 뾰족한 모양의 톱니 모양 가장자리	
예 거치	밖으로 향하여 둥글게 뾰족하게 뺀 톱니 모양 가장자리	
복 거치	톱니 같은 거치가 크게 있고 다시 큰 톱니에 작은 톱니가 있어 이중을 거치가 발달한 모양	
파상 거치	잎 가장자리가 물결 모양 가장자리	
둔거치	둔한 톱니 모양 가장자리	
전연	톱니가 없어 밋밋한 가장자리	
반곡	뒤로 젖혀진 가장자리	
심열	깊은 v자 톱니 모양의 잎 가장자리	

장상열	손바닥을 편 모양으로 찢어진 뾰족한 톱니가 있는 것	
천열	앞 가장자리의 갈라진 아랫부분이 원형이고 깊이가 주맥까지의 1/2이 아닌 것	
전열	주맥까지 또는 완전히 갈라진 앞 가장자리	

7. 잎맥의 모양

[표 1-7] 잎 맥의 종류 / 출처(개인 블로그 - 꽃다비, 산림청 - 용어집)		
잎 맥	정의	모양
우상맥	주맥에서 측맥이 좌우로 갈라져 나와 전체적으로 깃털 모양으로 된 맥이며, 많은 쌍떡잎 식물의 잎이 이러한 맥을 가지고 있음	
평행맥 (나란히 맥)	나란히 맥, 잎자루에서 잎끝까지 서로 평행하게 있는 잎맥	
출맥	잎 끝에서 어긋난 잎맥	
차상맥	엽맥이 두 개씩 갈라지는 맥	
망상맥 (그물맥)	측맥이 잎 끝을 향해 가다가 세맥이 되어 흩어지는 맥	
장상맥	손가락 개수만큼의 주맥이 발달한 손바닥 닮은 맥	