

인공지능 학습용 데이터 구축·활용 가이드라인

- 60. 정밀농업 노지작물 통합 데이터 구축 -

인공지능 데이터 구축	사업 총괄	전라북도청
	데이터 설계	주식회사 업데이터
	데이터 수집 및 정제	이레아이에스 주식회사
	데이터 가공	주식회사 업데이터
	데이터 검수	국립농업기술원 DSN전주(주)
	클라우드 소싱	캠틱종합기술원 디에스엔주식회사 NH네트웍스
	저작도구 개발	캠틱종합기술원 디에스엔전주 주식회사
	AI모델 개발	주식회사 업데이터
가이드라인 작성	주식회사 업데이터	이여진
	이레아이에스 주식회사	곽종원
	캠틱종합기술원	박영준
	DSN전주(주)	이세종
가이드라인 버전	Ver.2.2 (2021.12.30.)	

목 차

1. 데이터 명세 정보	1
1.1 데이터 정보 요약	1
1.2 데이터 포맷	2
1.3 어노테이션 포맷	3
1.4 데이터 구성	7
1.5 데이터 통계	8
1.6 원시데이터 특성	10
1.7 기타정보	13
 2. 데이터 구축 가이드	 14
2.1 데이터 구축 개요	14
2.2 문제 정의	16
2.3 수집·정제	18
2.4 어노테이션/라벨링	26
2.5 검수	33
2.6 활용	37

1 데이터 명세 정보

1.1 데이터 정보 요약

데이터 이름	데이터102. 정밀농업 농기계 잡초 인식 데이터																																			
활용 분야	2D RGB, 적외선, 토양환경 데이터(토양센서), 생장환경 데이터(환경센서)를 구축하여, 인공지능 기술이 탑재된 농기계로 잡초를 분류 및 제거하는 기술 개발에 기여하고, 최적 제초제 추천 서비스, 방제 방법 추천 서비스 활용 정보 제공																																			
데이터 요약	- 잡초 14종에 대한 총 358,000장의 데이터셋(이미지 데이터 + 센서 데이터) 구축																																			
	<table><tr><th>잡초명</th><th>2D RGB</th></tr><tr><td>강피(논피)</td><td>25,600</td></tr><tr><td>물달개비</td><td>25,600</td></tr><tr><td>올방개</td><td>25,600</td></tr><tr><td>올챙이고랭이</td><td>25,600</td></tr><tr><td>바랭이</td><td>25,600</td></tr><tr><td>쇠비름</td><td>25,600</td></tr><tr><td>깨풀</td><td>25,600</td></tr><tr><td>현명아주</td><td>25,600</td></tr><tr><td>한련초</td><td>25,600</td></tr><tr><td>좁명아주</td><td>25,600</td></tr><tr><td>닭의장풀</td><td>25,600</td></tr><tr><td>개비름</td><td>25,600</td></tr><tr><td>개여뀌</td><td>25,600</td></tr><tr><td>미국가막사리</td><td>25,200</td></tr><tr><td>총 구축 목표량</td><td>358,000</td></tr></table>				잡초명	2D RGB	강피(논피)	25,600	물달개비	25,600	올방개	25,600	올챙이고랭이	25,600	바랭이	25,600	쇠비름	25,600	깨풀	25,600	현명아주	25,600	한련초	25,600	좁명아주	25,600	닭의장풀	25,600	개비름	25,600	개여뀌	25,600	미국가막사리	25,200	총 구축 목표량	358,000
	잡초명	2D RGB																																		
	강피(논피)	25,600																																		
	물달개비	25,600																																		
	올방개	25,600																																		
	올챙이고랭이	25,600																																		
	바랭이	25,600																																		
	쇠비름	25,600																																		
	깨풀	25,600																																		
	현명아주	25,600																																		
	한련초	25,600																																		
	좁명아주	25,600																																		
	닭의장풀	25,600																																		
	개비름	25,600																																		
	개여뀌	25,600																																		
	미국가막사리	25,200																																		
	총 구축 목표량	358,000																																		
	- 적외선 촬영 이미지 데이터 총 89,000장 구축																																			
	- 환경 센서 데이터 120,000 set 구축																																			
	<table><tr><th>번호</th><th>분류</th><th>항목</th><th>센서 데이터 종수</th><th>데이터구축규모</th></tr><tr><td>1</td><td>토양환경 데이터</td><td>토양 EC, 토양온도, 토양 습도, 토양수분장력</td><td>4종</td><td rowspan="2">120,000</td></tr><tr><td>2</td><td>생장환경 데이터</td><td>온도, 습도, 일사량, 일출/일몰시간</td><td>5종</td></tr><tr><td colspan="3">총 구축량</td><td>9종 1set</td><td>120,000</td></tr></table>				번호	분류	항목	센서 데이터 종수	데이터구축규모	1	토양환경 데이터	토양 EC, 토양온도, 토양 습도, 토양수분장력	4종	120,000	2	생장환경 데이터	온도, 습도, 일사량, 일출/일몰시간	5종	총 구축량			9종 1set	120,000													
	번호	분류	항목	센서 데이터 종수	데이터구축규모																															
1	토양환경 데이터	토양 EC, 토양온도, 토양 습도, 토양수분장력	4종	120,000																																
2	생장환경 데이터	온도, 습도, 일사량, 일출/일몰시간	5종																																	
총 구축량			9종 1set	120,000																																
데이터 출처	자체 구축 및 수집 - 완주시 농업기술센터 시험 포장: 총 14종 잡초 데이터 수집																																			
데이터 이력	배포버전	버전 2.2																																		
	개정이력	2021.12.30																																		
	작성자/ 배포자	작성자 : 곽종원 / 배포자 : 임창묵																																		

1.2 데이터 포맷

원시데이터 포맷		
기본 포맷		JPG + CSV
포맷 세부 내용	2D RGB	- 포맷 : JPG - 해상도 : (1080~4800) * (1080~4800)
	적외선	- 포맷 : JPG - 해상도 : 640 * 480
	토양센서	포맷 : CSV
	환경센서	포맷 : CSV

학습용 데이터 포맷	
기본 포맷	JPG + CSV + JSON
포맷 세부 내용	- 이미지 데이터 : JPG - 토양환경 데이터 : CSV - 생장환경 데이터 : CSV - 어노테이션 정보 : JSON - JSON : ECMA-404, ISO/IEC 21778:2017, TTAE.OT-10.0394로 국내외 표준안으로 채택

1.3 어노테이션 포맷

○ 어노테이션 속성 항목 설명

- 인공지능 학습용 데이터셋 구축 공통참조기준을 바탕으로 어노테이션 항목으로 변환
- 이미지 포맷은 JPG 형식을 따름

<어노테이션 항목>

No.	속성명	항목 설명	Type	필수여부	작성예시
1. info-기본 정보					
1-1	description	데이터셋 이름	string	Y	weed_detection
1-2	url	데이터셋 제작자 url	string	Y	https://www.jeonbuk.go.kr/
1-3	version	제작버전	string	Y	v.1.0
1-4	year	제작년도	number	Y	2021
1-5	type	데이터셋 타입	string	Y	jpg
1-6	img_path	이미지데이터 폴더 경로	string	Y	/품종명/img
1-7	label_path	라벨링데이터 폴더 경로	string	Y	/품종명/json
2. collection - 수집 정보					
2-1	obj_num	객체 번호(위치)	string	Y	001~999
2-2	tod_attribute	적외선 촬영 장비 속성	string		FLIR C3X
2-3	tod_temper	객체 표면 온도	number		-20 ~ 300
2-4	soil_ec	토양 전기전도도(EC)	number	Y	0~10
2-5	soil_temper	토양 온도	number	Y	0~60
2-6	soil_humidity	토양 습도	number	Y	0~100
2-7	soil_potential	토양 수분장력	number	Y	0 ~ -1000000
2-8	temperature	온도	number	Y	-40 ~ 124
2-9	humidity	습도	number	Y	0~100
2-10	sunshine	일사량	number	Y	0~2000
2-11	sunrise_time	일출시간	string	Y	04:00~09:59
2-12	sunset_time	일몰시간	string	Y	16:00~22:59
2-13	img_attribute	촬영 장비 속성	string	Y	SM-G965N 등
2-14	img_time	촬영일시	string	Y	YYYY-MM-DD
2-15	img_part	잡초 촬영 범위	string	Y	[whole, part]
2-16	img_dist	촬영 거리(cm)	number		1~1000
2-17	img_angle	촬영 각도	number		0~360

No.	속성명	항목 설명	Type	필수여부	작성예시
3. licenses – 저작권 정보					
3-1	licenses_id	라이선스 고유 번호	number		1
3-2	licenses_name	라이선스 이름	string		weed01
4. images – 이미지 데이터 정보					
4-1	img_file_name	원천데이터 이름(파일명)	string	Y	예: 20210803_RGB_A _0_001_05.jpg
4-2	img_height	세로	number	Y	1080~4800
4-3	img_width	가로	number	Y	1080~4800
5. annotations – 어노테이션 정보					
5-1	segmentation	segmentation 좌표	string	Y	예: [2.14474345,114. 3444656]
5-2	area	어노테이션 면적	number	Y	0~ 23040000
5-3	bbox	bounding box 정보	string	Y	예: [0.0, 0.0, 440.0, 273.0]
5-4	weeds_kind	잡초명	string	Y	[A, B, C, D, E, F, G, H, I, J, K, L, M, N]

- 라벨링 완료 후 사용자의 목적에 맞는 저장 포맷을 선정하고 해당 데이터를 라벨링 완료된 데이터의 항목에 따라 분류하여 저장

○ 최종 데이터셋 JSON 예시

```
{
  "info": {
    "description": "weed_detection",
    "url": "https://www.jeonbuk.go.kr/",
    "version": "v.1.0",
    "year": 2021,
    "type": "jpg",
    "img_path": "/강피/img",
    "label_path": "/강피/json"
  },
  "collection": {
```

```

"obj_num": "012",
"tod_attribute": null,
"tod_temper": null,
"soil_ec": 1.03,
"soil_temper": 22.62,
"soil_humidty": 75.84,
"soil_potential": -0.2,
"temperature": 24.61,
"humidity": 76.93,
"sunshine": 78.0,
"sunrise_time": "05:39",
"sunset_time": "19:17",
"img_attribute": "Galaxy Note20 5G",
"img_time": "2021-08-17",
"img_part": "whole",
"img_dist": 25.0,
"img_angle": 45.0
},
"licenses": {
  "licenses_id": 1,
  "licenses_name": "weed01"
},
"images": {
  "img_file_name": "20210817_RGB_A_012_45_13.jpg",
  "img_height": 1836,
  "img_width": 1836
},
"annotations": {
  "segmentation": [
    835.7142857142857,
    704.5454545454545,
    808.4415584415584,
    661.6883116883117,
    785.0649350649351,
    637.6623376623377,
    764.2857142857142,
    616.8831168831168,
    761.6883116883116,
    603.2467532467532,
    761.6883116883116,
    592.8571428571429,
    767.5324675324675,

```

```
1416.8831168831168,  
1457.7922077922078,  
1414.935064935065,  
1214.2857142857142,  
1053.2467532467533,  
1182.4675324675325,  
1166.2337662337661,  
1207.7922077922078,  
1131.168831168831,  
1548.7012987012986,  
1139.6103896103896,  
1559.7402597402597,  
617.5324675324675,  
807.7922077922078,  
646.1038961038961,  
825.3246753246754,  
675.974025974026
```

```
],
```

```
"area": 115037.0,
```

```
"bbox": [
```

```
761.0,
```

```
385.0,
```

```
781.0,
```

```
1214.0
```

```
],
```

```
"weeds_kind": "A"
```

```
}
```

```
}
```

1.4 데이터 구성

○ 선정된 객체 별 명명 규칙

잡초별 코드 정의		
항 목	잡초명	CODE
촬영 잡초	강피(논피)	A
	물달개비	B
	올방개	C
	올챙이고랭이	D
	바랭이	E
	쇠비름	F
	깨풀	G
	흰명아주	H
	한련초	I
	좀명아주	J
	닭의장풀	K
	개비름	L
	개여뀌	M
	미국가막사리	N

○ 데이터 파일 명명 방식 및 사례

구분			
2D	파일명	규칙	날짜_촬영도구_잡초품종코드_객체번호_촬영각도/촬영부위_이미지번호.jpg
		예시	- 전체 촬영 이미지의 경우 20210817_RGB_A_012_45_13.jpg - 부분 촬영 이미지의 경우 20210817_RGB_A_012_X_13.jpg
	라벨링명	클래스 명명규칙	잡초품종코드
		예시	“강피(논피)”의 경우 A
적외선	파일명	규칙	촬영일자_촬영도구_잡초품종코드_0_객체번호_이미지번호.jpg
		예시	20210803_IFR_A_0_001_06.jpg

1.5 데이터 통계

1.5.1 데이터 구축 규모

○ 잡초 14종에 대한 총 358,000장 이상의 데이터셋(이미지 데이터 + 센서 데이터) 구축

번호	잡초명	2D RGB
1	강피(논피)	25,600
2	물달개비	25,600
3	올방개	25,600
4	올챙이고랭이	25,600
5	바랭이	25,600
6	쇠비름	25,600
7	깨풀	25,600
8	흰명아주	25,600
9	한련초	25,600
10	좀명아주	25,600
11	닭의장풀	25,600
12	개비름	25,600
13	개여뀌	25,600
14	미국가막사리	25,200
총 구축 목표량		358,000

○ 적외선 촬영 이미지 데이터 총 89,000장 구축

○ 센서 데이터 구축 규모

번호	분류	항목	센서 데이터 종류	데이터구축규모
1	토양환경 데이터	토양 EC, 토양온도, 토양 습도, 토양수분장력	4종	120,000
2	생장환경 데이터	온도, 습도, 일사량, 일출/일몰시간	5종	
	총 구축량		9종 1set	120,000

1.5.2 데이터 분포

○ 잡초의 종류별 및 생육 전주기 수집을 통하여 데이터의 다양성 확보

- 잡초의 종류 : 14종

번호	잡초명	2D RGB 수량 비율
1	강피(논피)	7.1%
2	물달개비	7.1%
3	올방개	7.1%
4	올챙이고랭이	7.1%
5	바랭이	7.1%
6	쇠비름	7.1%
7	깨풀	7.1%
8	흰명아주	7.1%
9	한련초	7.1%
10	좀명아주	7.1%
11	닭의장풀	7.1%
12	개비름	7.1%
13	개여뀌	7.1%
14	미국가막사리	7.1%
총 구축량 비율		100%

1.6 원시데이터 특성

1.6.1 대상분류

○ 다양한 모니터링 환경에서 적용이 가능하고, AI 모델 학습 및 분류 성능 향상을 기반으로 구축

원시데이터 대상			특성
1	강피(논피)		줄기는 잎집에 싸여 있으며 연한 녹색으로 편평하고 수염뿌리가 있고, 잎은 부드럽고 털이 없음
2	물달개비		줄기는 5-6개씩 뭉쳐 나오고 짧으며 각 1개씩의 잎이 달림. 잎은 넓은 세모진 달걀모양이고, 밑이 둥글거나 약간 심장 모양의 특징
3	올방개		군집생육하는 수생 식물로 근경이 옆으로 길게 뻗어 그 끝에는 괴경(땅속 줄기)이 달림. 괴경의 직경은 5~8mm 정도
4	올챙이고랭이		근경이 짧고, 줄기는 포기져 뭉쳐나고, 50cm 정도 자라고, 굵기는 5mm 정도로, 색조가 우중충함. 둔하거나 약간 모가 진 듯함
5	바랭이		밭에서 흔히 자라는 잡초. 밑부분이 지면으로 번으면서 마디에서 뿌리가 내리고 결가지와 더불어 40~70cm로 곧게 자란다. 잎은 줄 모양이며 분록색 또는 연한 녹색이고, 잎집에는 흔히 털이 있음
6	쇠비름		전체에 털은 없으나 육질이고 뿌리는 흰색이며 줄기는 붉은빛이 도는 갈색으로서 많은 가지가 비스듬히 옆으로 퍼져있음. 잎은 어긋나거나 마주나고, 모양은 달걀을 거꾸로 세운 듯한 모양
7	깨풀		풀 전체에 짧은 털이 나고 곧게 서며 가지를 침. 잎은 어긋나고 잎자루가 있으며 달걀 모양 또는 달걀 모양 바소꼴로 끝이 뾰족한 모양

8	흰명아주		• 앞은 어긋나고 달걀 모양이며 가장자리에 불규칙한 톱니가 있고, 잎 뒷면은 흰가루로 덮여 있음. 꽃은 줄기 끝이나 잎겨드랑이에 수상꽃차례가 모여서 원추꽃차례를 이루고, 꽃잎은 없고 꽃받침은 황록색으로 깊게 5개로 갈라지며 자루와 포가 없음
9	한련초		• 줄기는 서서 자라며, 전체에 짧은 강모가 있고 가지가 어긋남. 앞은 마주나며, 양면에 굳세고 짧은 털이 있으며, 가장자리는 톱니가 약간 있음
10	좀명아주		• 줄기는 곧게 서며 높이 30~60cm이고 윗부분과 잎 뒷면에 흰 가루가 있음. 앞은 어긋나고 바소꼴 모양의 긴 타원형이며 가장자리에 물결 모양의 톱니가 있고 가장 밑의 톱니는 갈래 조각과 비슷하게 큼
11	닭의장풀		• 줄기 밑 부분은 옆으로 비스듬히 자라며 땅을 기고 마디에서 뿌리를 내리며 많은 가지가 갈라짐. 줄기 윗부분은 곧게 서고, 앞은 어긋나고 달걀 모양의 바소꼴이며, 잎 끝은 점점 뾰족해지고 밑 부분은 막질(膜質:얇은 종이처럼 반투명한 것)의 잎집으로 됨
12	개비름		• 전체에 털이 없으며 줄기는 연하고 밑동에서 가지를 많이 냄. 앞은 어긋나고 긴 잎자루가 있으며 녹색이거나 자줏빛을 띤 갈색이고 사각의 달걀 모양
13	개여뀌		• 전체에 털이 없고 줄기는 붉은 자줏빛의 둥근 통 모양으로 곧게 섬. 가지를 많이 내며 마디에서 뿌리를 번으며, 앞은 어긋나고, 나비 넓은 바소꼴. 양면에 털이 나고 잎자루는 짧음
14	미국가막사리		• 줄기는 검은 자줏빛이 돌고 횡단면이 사각형이며 털이 없고, 앞은 마주나고 깃꼴겹잎이며, 작은잎은 3~5개이고 바소꼴이며 가장자리에 톱니가 있고 잎자루가 없음. 꽃은 갈라진 가지 끝에 두상화(頭狀花:꽃대 끝에 꽃자루가 없는 작은 꽃이 많이 모여 피어 머리 모양을 이룬 꽃)를 이루며 달리고 전체가 원추꽃차례를 이룸

1.6.2 제약조건

항목	전제조건	제약조건
이미지 수집	길고 얇은 앞을 가진 잡초의 부분 촬영 진행시 다수의 앞이 함께 촬영되어지고 대상 객체 전체 포커싱 어려움 발생	- 잡초의 부분촬영 진행시 길고 얇은 앞의 경우는 앞의 중간, 끝 등으로 나누어 부분 촬영 진행하여도 학습데이터 용도로 활용 가능 - 부분촬영 진행시 대상 객체(앞)의 수가 3개 이내로 존재할 수 있도록 핸디형 배경판으로 잔여 앞을 가리고 촬영(아웃포커싱 문제 해결)
	적외선카메라 특성상 촬영시 주변 환경 변화요인에 따른 이미지 왜곡 현상 발생됨(객체와 열상영역 단차 발생 등)	적외선카메라 촬영 이미지는 방사율/복사율/주변환경 온도변화 등 외부적 환경요인에 의해서 색상변화가 왜곡될 수 있으나 학습데이터 용도로는 활용 가능함. 따라서, 객체의 중심에서 온도 측정이 되었으나 이미지 왜곡 현상이 발생된 이미지는 수집/정제 검사 시 Pass 처리하기로 함
센서 데이터 수집	- 전세계적인 코로나19 특수상황으로 환경센서 수입 일정 차질 발생(8/9 설치가능) 7월부터 촬영한 이미지데이터는 센서데이터 부재로 인해 학습 데이터셋으로 활용 불가함	노지작물 중 일부는 생육단계 전체의 데이터를 구축하기는 어려운 상황이므로 코로나19 특수상황을 최대한 감안하여, 데이터 구축 목표수량에는 문제 없도록 구축 예정

1.6.3 속성

원시데이터 속성	
2D RGB	해상도 : 1080~4800 * 1080~4800
적외선	해상도 : 640 * 480

1.7 기타정보

1.7.1 포괄성

구분	데이터 내용
포괄성	- 14종 잡초에 대한 35만장 이상의 데이터셋 구축은 잡초 탐지 및 분류 외에도 잡초에 따른 방제방법 적용 등 다양한 분야에 활용 가능함

1.7.2 독립성

구분	데이터 내용
독립성	- 본 사업을 통해 생성된 인공지능 학습용 데이터의 배포 및 활용에 대한 권한은 NIA에 귀속되어 있음

1.7.3 유의사항

구분	데이터 내용
개인정보 보호	- 잡초 이미지는 특별하게 개인정보 노출 문제가 없지만 AI 학습데이터 중 이미지의 개인정보 문제가 발생하지 않도록 해당사항을 항상 내부적인 토의와 NIA와의 협의 필요
저작권 보호	- 데이터의 각 단계에서 저작권이 노출되거나 저작권자의 허가가 필요한 데이터의 사용 및 공개 가능 여부에 확인을 선행하여 침해 이슈 해결 저작권자의 권리를 침해할 수 있는 데이터는 사전에 배제하는 것을 원칙으로 수행
유해성	- 유해성 여부를 검토하여 사회적 민감 정보 이슈 문제에 대처
기타	- 본 과제를 위하여 각 컨소시엄은 유기적 연락체계를 구성하여 온·오프라인 소통을 하고 있음

2 데이터 구축 가이드

2.1 데이터 구축 개요

- 본 사업은 노지 환경을 고려하여 잡초의 생육 데이터를 수집하여 학습데이터를 구축하는 사업으로써 다양한 취득 센서의 선정과 운용 계획 등 수립 필요
- 따라서, **정밀농업에 필요한 잡초 생육 데이터를 확보하는 것을 목적으로** 지상에서 촬영하는 2D, 적외선 카메라를 중심으로 학습데이터를 구축
- 또한, 토양센서에서 수집하는 토양환경 데이터(토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력) 환경센서에서 수집하는 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터를 세부적으로 제공하여 학습데이터에 파생된 정보를 연계할 수 있도록 하여 활용도 확대
- 학습데이터 활용 확대를 위해 본 사업은 잡초 14종의 데이터 확보

토양/환경 센서 데이터			
토양 센서 데이터	토양 전기전도도	토양 온도	토양 습도 등
환경 센서 데이터	온도	습도	일사량



위치정보+Time 의 시공간 정보

토양/환경 센서 데이터+ 이미지데이터

이미지데이터		
데이터 획득 내용	2D	적외선
수집주기	주 5회	
수집방법	방향과 각도, 대상에 따라 촬영(잡초 별 상이)	

○ 단계별 구축 절차

절차	방법
수집 및 정제	• (수집) 잡초 14종(강피, 물달개비, 올방개, 올챙이고랭이, 바랭이, 쇠비름, 깨풀, 흰명아주, 한련초, 쯤명아주, 닭의장풀, 개비름, 개여뀌, 미국가막사리)에 대해 이미지 데이터 및 센서 데이터 수집 - 이미지 데이터 : 2D RGB, 적외선 - 토양환경 데이터(토양 센서): 토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력 - 성장환경 데이터(환경 센서): 온도, 습도, 일사량 - 데이터 수집 대상지 일출시간과 일몰시간 데이터 - 작업자 2인 교차 검사 • (정제) 수집 데이터의 화질 및 객체 식별 등 검토 - 1차 검사 : 자체 검사 - 2차 검사 : 품질관리 담당자 검사 - 3차 검사 : 제3자 품질검증 용역기관 교차 검증
어노테이션/ 라벨링	• 원천데이터 메타정보 입력 - 이미지, 클래스, 환경 등 정보에 대한 메타데이터 입력 • 각 객체별로 라벨링(클라우드 소싱 병행 라벨링 수행) - Polygon, Segmentation 등 • 라벨링 데이터 검사 - 1차 검사 : 자체 검사 - 2차 검사 : 품질관리 담당자 검사 - 3차 검사 : 제3자 품질검증 용역기관 교차 검증
검수 및 검증	• (검수) 클래스별 라벨링 결과물 검수 - 1차 검사 : 자체 검사 - 2차 검사 : 품질관리 담당자 검사 - 3차 검사 : 데이터 품질 객관성 확보를 위해 제3자 품질검증 용역기관 교차 검증 • (검증) TTA 데이터 검증
데이터셋 구축	• 정밀농업 농기계 잡초 인식 데이터셋(8:1:1) 구축 - 학습 데이터셋 (Train Dataset) 생성 - 검증 데이터셋 (Validation Dataset) 생성 - 시험 데이터셋 (Test Dataset) 생성

2.2 문제 정의

2.2.1 임무 정의

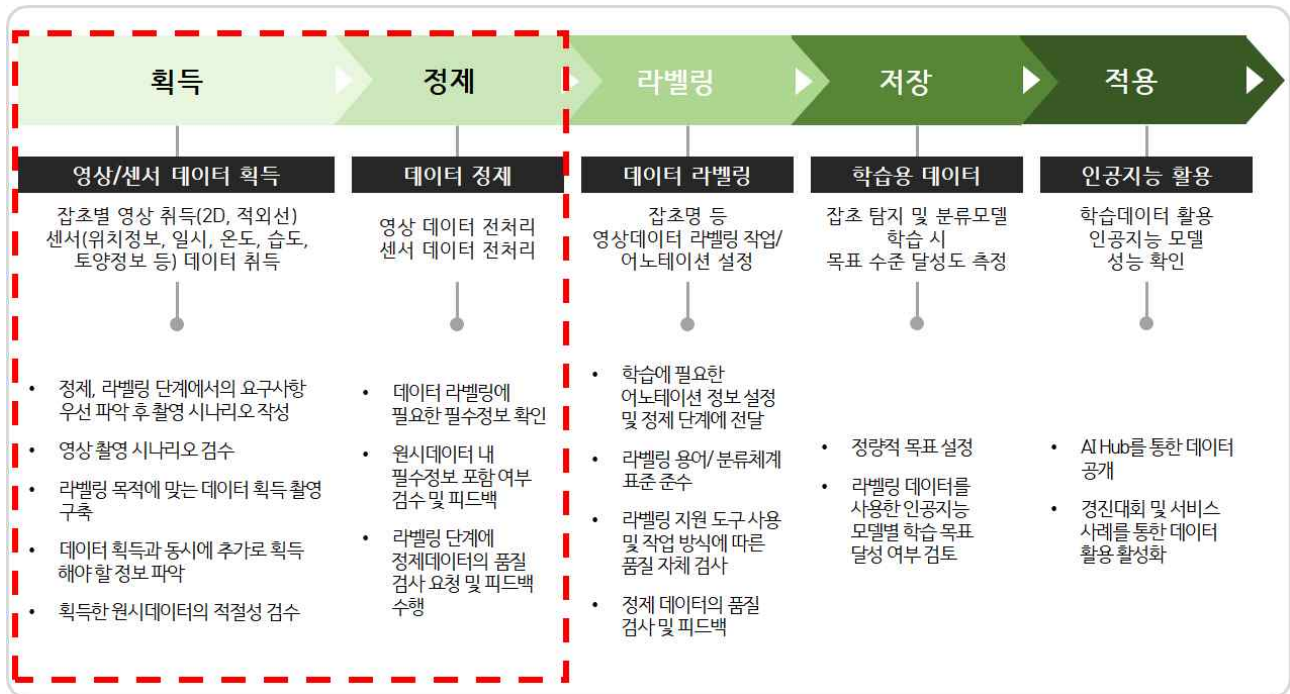
구분	세부 내용																																																			
과제 내용	○ 본 사업은 노지 환경을 고려하여 잡초의 생육 데이터를 수집하여 학습데이터를 구축하는 사업으로써 다양한 취득 센서의 선정과 운용 계획 등 수립 필요 ○ 따라서, 정밀농업에 필요한 잡초 생육 데이터를 확보하는 것을 목적으로 지상에서 촬영하는 2D, 적외선 카메라를 중심으로 학습데이터를 구축 ○ 또한, 토양센서에서 수집하는 토양환경 데이터(토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력) 환경센서에서 수집하는 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터를 세부적으로 제공하여 학습데이터에 파생된 정보를 연계할 수 있도록 하여 활용도 확대 ○ 또한, 학습데이터 활용 확대를 위해 본 사업은 잡초 14종의 생육 주기별 데이터 확보																																																			
과제 목표	○ 잡초 14종에 대한 총 358,000장의 데이터셋(이미지 데이터 + 센서 데이터) 구축 <table><tr><th>잡초명</th><th>2D RGB</th></tr><tr><td>강피(논피)</td><td>25,600</td></tr><tr><td>물달개비</td><td>25,600</td></tr><tr><td>올방개</td><td>25,600</td></tr><tr><td>올챙이고랭이</td><td>25,600</td></tr><tr><td>바랭이</td><td>25,600</td></tr><tr><td>쇠비름</td><td>25,600</td></tr><tr><td>깨풀</td><td>25,600</td></tr><tr><td>흰명아주</td><td>25,600</td></tr><tr><td>한련초</td><td>25,600</td></tr><tr><td>좀명아주</td><td>25,600</td></tr><tr><td>닭의장풀</td><td>25,600</td></tr><tr><td>개비름</td><td>25,600</td></tr><tr><td>개여뀌</td><td>25,600</td></tr><tr><td>미국가막사리</td><td>25,200</td></tr><tr><td>총 구축량</td><td>358,000</td></tr></table> ○ 적외선 촬영 이미지 데이터 총 89,000장 구축 ○ 환경 센서 데이터 120,000 set 구축 <table><tr><th>번호</th><th>분류</th><th>항목</th><th>센서 데이터 종수</th><th>데이터구축규모</th></tr><tr><td>1</td><td>토양환경 데이터</td><td>토양 EC, 토양온도, 토양 습도, 토양수분장력</td><td>4종</td><td rowspan="2">120,000</td></tr><tr><td>2</td><td>생장환경 데이터</td><td>온도, 습도, 일사량, 일출/일몰시간</td><td>5종</td></tr><tr><td colspan="3">총 구축량</td><td>9종 1set</td><td>120,000</td></tr></table>	잡초명	2D RGB	강피(논피)	25,600	물달개비	25,600	올방개	25,600	올챙이고랭이	25,600	바랭이	25,600	쇠비름	25,600	깨풀	25,600	흰명아주	25,600	한련초	25,600	좀명아주	25,600	닭의장풀	25,600	개비름	25,600	개여뀌	25,600	미국가막사리	25,200	총 구축량	358,000	번호	분류	항목	센서 데이터 종수	데이터구축규모	1	토양환경 데이터	토양 EC, 토양온도, 토양 습도, 토양수분장력	4종	120,000	2	생장환경 데이터	온도, 습도, 일사량, 일출/일몰시간	5종	총 구축량			9종 1set	120,000
잡초명	2D RGB																																																			
강피(논피)	25,600																																																			
물달개비	25,600																																																			
올방개	25,600																																																			
올챙이고랭이	25,600																																																			
바랭이	25,600																																																			
쇠비름	25,600																																																			
깨풀	25,600																																																			
흰명아주	25,600																																																			
한련초	25,600																																																			
좀명아주	25,600																																																			
닭의장풀	25,600																																																			
개비름	25,600																																																			
개여뀌	25,600																																																			
미국가막사리	25,200																																																			
총 구축량	358,000																																																			
번호	분류	항목	센서 데이터 종수	데이터구축규모																																																
1	토양환경 데이터	토양 EC, 토양온도, 토양 습도, 토양수분장력	4종	120,000																																																
2	생장환경 데이터	온도, 습도, 일사량, 일출/일몰시간	5종																																																	
총 구축량			9종 1set	120,000																																																
달성 지표	○ 모델 정확도 mAP 기준 0.85 이상																																																			

2.2.2 데이터 구축 유의사항

구분	데이터 내용
개인정보 보호	- 정밀농업 잡초 이미지는 특별하게 개인정보 노출 문제가 없지만 AI 학습데이터 중 이미지의 개인정보 문제가 발생하지 않도록 해당사항을 항상 내부적인 토의와 NIA와 협의 필요
저작권 보호	- 데이터의 각 단계에서 저작권이 노출되거나 저작권자의 허가가 필요한 데이터의 사용 및 공개 가능 여부에 확인을 선행하여 침해 이슈 해결 저작권자의 권리를 침해할 수 있는 데이터는 사전에 배제하는 것을 원칙으로 수행
유해성	- 유해성 여부를 검토하여 사회적 민감 정보 이슈 문제에 대처
수집	- 촬영 전용 앱을 통한 환경 통일, FULL HD이상 해상 이미지 수집, 촬영 이미지 모니터링을 통한 지속 관리
정제	- 파일 명명규칙에 따라 파일명이 오류 없이 변환되었는지 지속적으로 모니터링
가공	- 라벨링 작업이 라벨링 기준에 부합되게 이루어지고 있는지 지속적으로 모니터링
기타	- 본 과제를 위하여 각 컨소시엄은 유기적 연락체계를 구성하여 온·오프라인 소통을 하고 있음

2.3 수집 · 정제

2.3.1 데이터 수집·정제 개요



2.3.2 원시데이터 선정

○ 주요 잡초 14종을 선정하고 촬영 기간 및 주기별, 위치별 데이터 취득

잡초 명	촬영 기간 및 촬영 주기	위치	수집 예정량 (장)
강피(논피)	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
물달개비	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
올방개	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
올챙이고랭이	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
바랭이	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
쇠비름	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
깨풀	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
흰명아주	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
한련초	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
좀명아주	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
닭의장풀	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
개비름	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
개여뀌	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600
미국가막사리	8월 ~ 12월 매주 월 ~ 일	완주시 농업기술센터 포장	25,600

○ 주요 잡초별로 이미지 데이터(2D RGB, 적외선), 잡초 재배지점의 토양환경 데이터(토양센서), 잡초 재배 지점의 생장환경 데이터(환경센서) 수집

2.3.3 수집·정제 절차

(1) 개요

- 현장 데이터 수집 담당자가 원시데이터 촬영 이미지 확인 및 분류
 - 자체 검수하여 클라우드 시스템에 잡초별 업로드
- 교차 검수 담당자가 이미지를 고려하여 유의미한 데이터인지 확인하고 파일명 명명
 - 밝기, 선명도 등의 문제로 인해 객체 검출이 불가능한 이미지 배제
 - 촬영 이미지의 화질 및 객체식별 확인 및 정제 후 데이터 수집 담당자에게 피드백
 - 초점 흐림, 뒷배경, 다른 객체 존재, 사람 촬영 등 정제 사유를 파일명에 기재
- 원천데이터 사용 및 폐기 여부를 결정하여 활용 가능한 데이터는 별도 클라우드 저장소에 저장

(2) 데이터 수집방법

잡초명	수집		장소
	방법	수집 데이터	
강피 (논피)	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
물달개비	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
올방개	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
올챙이고랭이	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
바랭이	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장

잡초명	수집		장소
	방법	수집 데이터	
쇠비름	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
깨풀	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
흰명아주	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
한련초	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
좁명아주	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
닭의장풀	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
개비름	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
개여뀌	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장
미국가막사리	직접 촬영, 토양센서, 환경센서	2D 및 적외선 영상 데이터, 토양환경 데이터 (토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력), 생장환경 데이터(온도, 습도, 일사량), 일출시간 및 일몰시간 데이터	완주시 농업기술센터 포장

(3) 데이터 수집 절차

구분	수집 절차																																								
기본 수집 방안	○ 이미지 수집 시 주변의 다른 사물, 작물/잡초 등이 같이 촬영되지 않도록 주변 가림막 (하얀색)을 설치하여 이미지 수집 ○ 객체 앞에 이름표를 작성하여 고정, 이를 통해 이미지별 객체 구분 ○ 촬영 시 Full HD를 기본 설정으로 진행하고 촬영 도중에 해상도를 변경하지 않도록 함																																								
2D 영상 촬영 데이터	○ 전체 이미지 촬영 - 객체는 이미지의 중앙에 위치하도록 촬영 - 객체의 줄기 끝이 잘리지 않고 각 줄기 전체가 모두 나올 수 있도록 촬영 - 수집하고자 하는 객체를 제외한 다른 객체가 나오지 않도록 촬영 - 객체 구분을 위해 풋말 같이 촬영 - 이미지 내 모든 객체의 초점이 명확해지게 촬영 - 이미지의 아웃포커싱 현상을 방지하기 위해 카메라 풍경 모드나 풍경 촬영 앱을 사용하여 촬영 진행(동일한 촬영 장비로 기준 설정) - 전체 이미지 촬영 순서																																								
	촬영방식	구분	라벨표시	촬영수	합계	총합계	기타																																		
	2D	전체 사진	1	17	18	59	수평 : 8방향 x 2각도 = 16장 수직 : 90도 1방향 = 1장																																		
		부분 사진	0	35	35		객체를 식별 할 수 있는 앞 중심으로 촬영																																		
적외선	부분 사진	1	5	6	온도 취득을 위해 앞 위주 촬영																																				
- 여러 줄기로 인해 부분 이미지 촬영에 방해 시, 별도의 장비(흰색 배경판)를 이용하여 촬영 - 이미지 내 모든 객체는 라벨링 할 수 있도록 촬영 - 객체의 개수와 무관하게 명확한 객체 라벨링이 가능하도록 촬영 - 다양한 각도, 방향 등을 다르게 촬영하여 여러 장의 이미지를 확보																																									
<table><tr><th>구분</th><th>0도 (6시)</th><th>45도 (4시 30분)</th><th>90도 (3시)</th><th>135도 (1시 30분)</th><th>180도 (12시)</th><th>225도 (10시 30분)</th><th>270도 (9시)</th><th>315도 (7시 30분)</th></tr><tr><td rowspan="2">① 촬영 순서 1 : 0도~45도 2 : 8방향</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>② 촬영 순서 2 (수직 1장)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table>							구분	0도 (6시)	45도 (4시 30분)	90도 (3시)	135도 (1시 30분)	180도 (12시)	225도 (10시 30분)	270도 (9시)	315도 (7시 30분)	① 촬영 순서 1 : 0도~45도 2 : 8방향																	② 촬영 순서 2 (수직 1장)								
구분	0도 (6시)	45도 (4시 30분)	90도 (3시)	135도 (1시 30분)	180도 (12시)	225도 (10시 30분)	270도 (9시)	315도 (7시 30분)																																	
① 촬영 순서 1 : 0도~45도 2 : 8방향																																									
																																									
② 촬영 순서 2 (수직 1장)																																									
<수생 잡초 예시>																																									

구분	수집 절차								
	구분	0도 (6시)	45도 (4시 30분)	90도 (3시)	135도 (1시 30분)	180도 (12시)	225도 (10시 30분)	270도 (9시)	315도 (7시 30분)
① 촬영 순서 1 : 0도>45도 2 : 8방향									
	촬영 순서 2 (수직 1장)								
<토양 잡초 예시>									
적외선 영상 촬영 데이터	○ 잡초 전체 촬영의 경우 적외선 카메라의 초점 문제로 엽온을 정확하게 측정하기 어렵기 때문에 부분 촬영을 진행하여 보다 정확한 엽온 정보를 취득 ○ 해당 객체의 부위가 겹칠 경우 객체별 온도 파악을 위해서 가림막을 통해 객체 단위로 촬영 적외선 이미지 (부분) <적외선 이미지 예시-개비름, 미국가막사리>								
	- 토양 환경 데이터 측정 항목: 토양 전도도, 토양 온도, 토양 습도, 토양 수준 장력 - 측정 주기: 10분당 데이터 측정 후 LTE 통신으로 서버 전송 - 1일 기준 데이터 그래프								
토양 환경 데이터									

구분

수집 절차

샘플 데이터

번호	토양온도	토양습도	토양EC	토양 수분장력	측정일시
1	24.00	3.72	0.04	-37.37	2021-08-29 00:05:00
2	23.95	3.72	0.04	-37.37	2021-08-29 00:15:00
3	23.83	3.72	0.04	-37.37	2021-08-29 00:25:00
4	23.75	3.72	0.04	-37.37	2021-08-29 00:35:00
5	23.67	3.72	0.04	-37.37	2021-08-29 00:45:00

토양 성장환경 데이터 측정 항목: 온도, 습도, 일사량, 일출시간, 일몰 시간

측정 주기: 10분당 데이터 측정 후 LTE 통신으로 서버 전송

일출/일몰 시간은 센서 설치 주소의 위/경도 데이터를 통해 자동 계산

1일 기준 데이터 그래프

생장 환경 데이터

온도(°C)

습도(%)

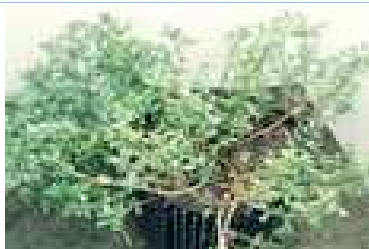
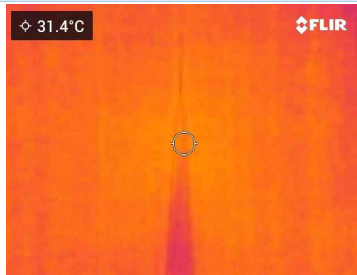
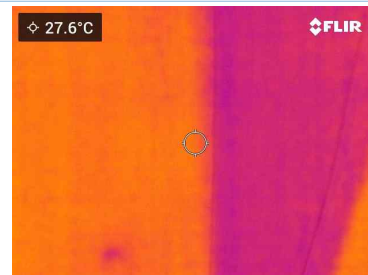
광량(W/m2)

샘플 데이터

항목	일출시간	05:49	일몰시간	19:01
----	------	-------	------	-------

번호	온도	습도	광량	측정일시
1	24.34	82.15	167	2021-08-29 11:45:00
2	24.37	79.11	180	2021-08-29 11:55:00
3	24.71	77.47	260	2021-08-29 12:05:00
4	26.38	71.45	382	2021-08-29 12:15:00
5	29.13	64.43	523	2021-08-29 12:25:00

2.3.4 수집·정제 기준

구분	정제 기준		정제 방식	데이터 형태
2D 정제기준	- 사람의 신체(특히 얼굴, 개인정보)가 등장하지 아니할 것 - 본 사업과 관련 없거나 기존 이미지는 사용 절대 불가 - FHD 미만의 해상도 이미지 폐기 - 초점과 밝기가 명확해 객체의 판별이 가능해야 함(80% 이상) - 객체 전체 촬영 이미지의 경우 객체의 완전한 형태가 확인 가능해야 함		검사도구를 활용한 데이터 해상도 체크 및 육안 검수	.jpg
2D 정제기준 미달 이미지				
	객체 초점 흐림	대상 객체외 타 객체 침범		
				
	30cm 이상 큰이미지(잡초의미 없음)	이미지 중복 촬영		
적외선 정제기준	- 사람의 신체(특히 얼굴, 개인정보)가 등장하지 아니할 것 - 본 사업과 관련 없거나 기존 이미지는 사용 절대 불가 - 객체의 온도를 정확히 파악할 수 있는 센서값이 존재해야 함 - 장비를 통일성 있게 적용하여야 함		검사도구를 활용한 데이터 해상도 체크 및 육안 검수	.jpg
적외선 정제기준 미달 이미지				
	객체 초점 흐림	객체 편향		

2.3.5 수집·정제 조직

○ 세부 업무 분장 현황

업무구분	수행업체
2D/적외선 촬영(잡초)	(주)이레아이에스
정제 및 검수	(사)캠틱종합기술원, (주)업데이터

2.3.6 수집·정제 도구

○ 수집 장비

<스마트폰 카메라 사양>

제조회사	삼성전자 외	등록년월	2020년 02월
디카 분류	스마트폰	유효 화소수	2000만화소 (1920*1080FHD급)
센서종류	Sony IMX374 (CMOS)		
광학 줌	광학줌 (손떨림방지)	최고 렌즈밝기	F/1.8
최단 접사거리	접사:5cm	LCD 화면크기	(6.7인치)
LCD 화면형태	고정형	최고 감도	최고ISO1250
무게	192g	배터리	4,300mAh



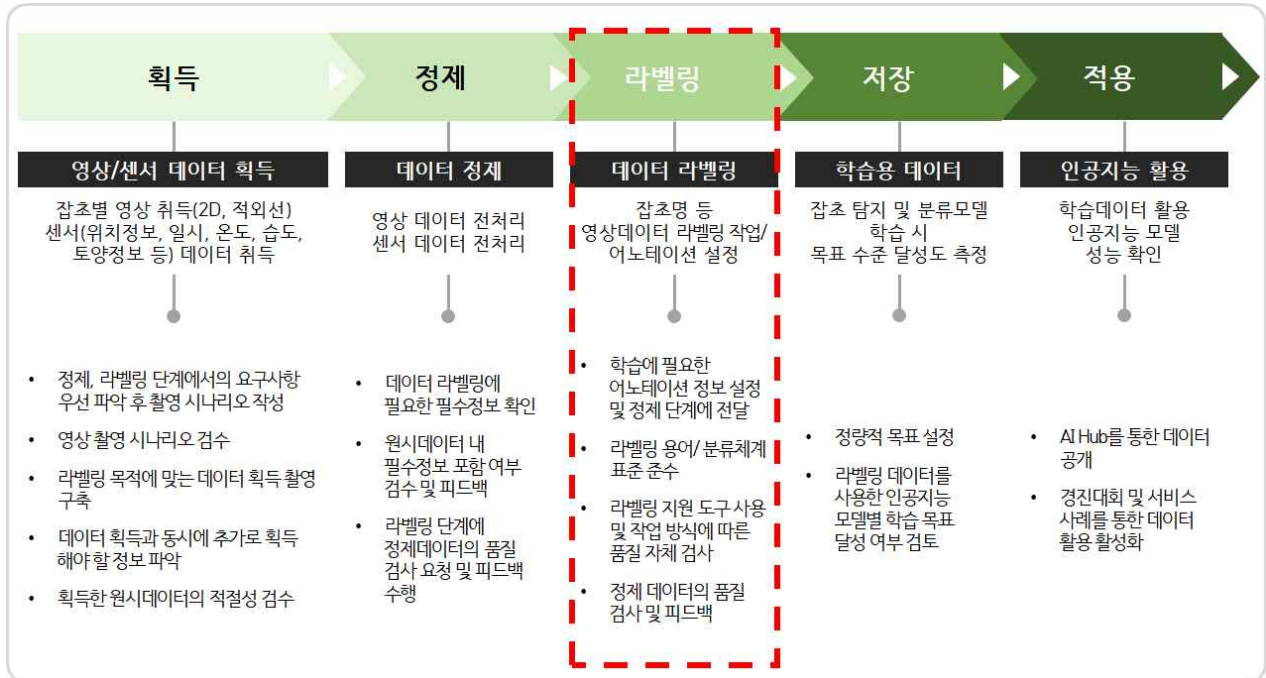
<적외선 카메라>

제조회사	FLIR C3X	등록년월	2020년 09월
디카 분류	콤팩트	유효 화소수	:128x96 pixels(12,288화소)
센서종류	160 x 120 (19,200픽셀)		
대상온도	-20℃ ~ 40℃	실화상 카메라	5MP
열화상 감도	<70mK	LCD 화면크기	3.5인치
최소 초점거리	Thermal:0.1m MSX : 0.3m	터치스크린	LCD 정전식터치
스펙트럼 범위	8-14μm	시야각	54° x 42°
무게	190g	배터리	내장형 리튬이온



2.4 어노테이션/라벨링

2.4.1 어노테이션/라벨링 개요



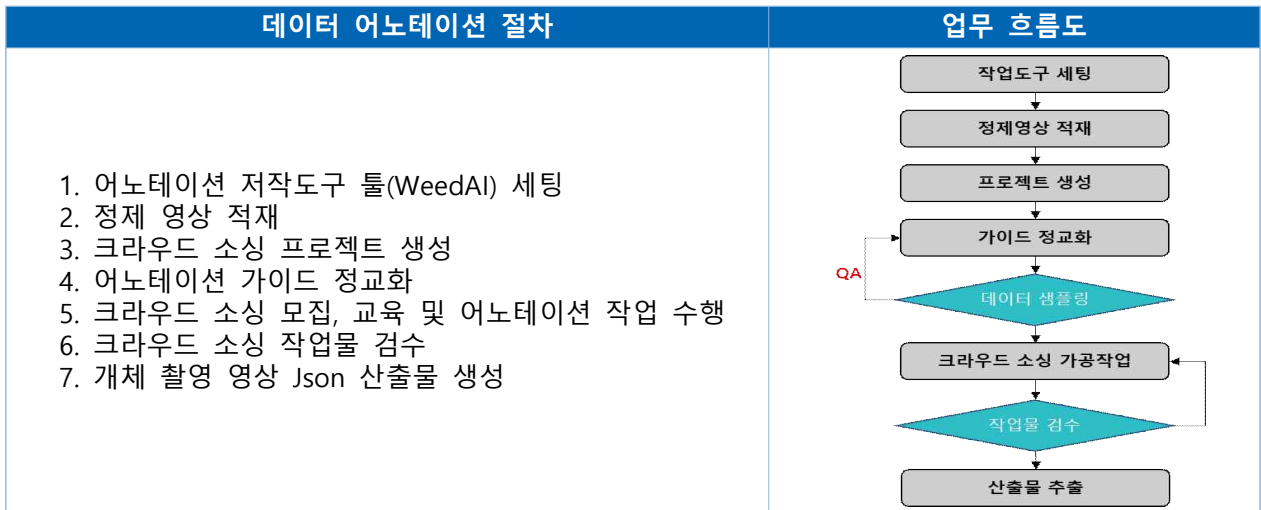
- 완주시 농업기술센터 시험포장에서 촬영한 잡초 데이터에 대하여 데이터 라벨링 및 어노테이션 작업 수행

- 영상데이터 라벨링을 위한 클래스 분류 기준은 아래와 같이 대상으로 선정된 주요 잡초 14종

잡초명	CODE	잡초명	CODE
강피(논피)	A	흰명아주	H
물달개비	B	한련초	I
울방개	C	좀명아주	J
울챙이고랭이	D	닭의장풀	K
바랭이	E	개비름	L
쇠비름	F	개여뀌	M
깨풀	G	미국가막사리	N

2.4.2 어노테이션/라벨링 절차

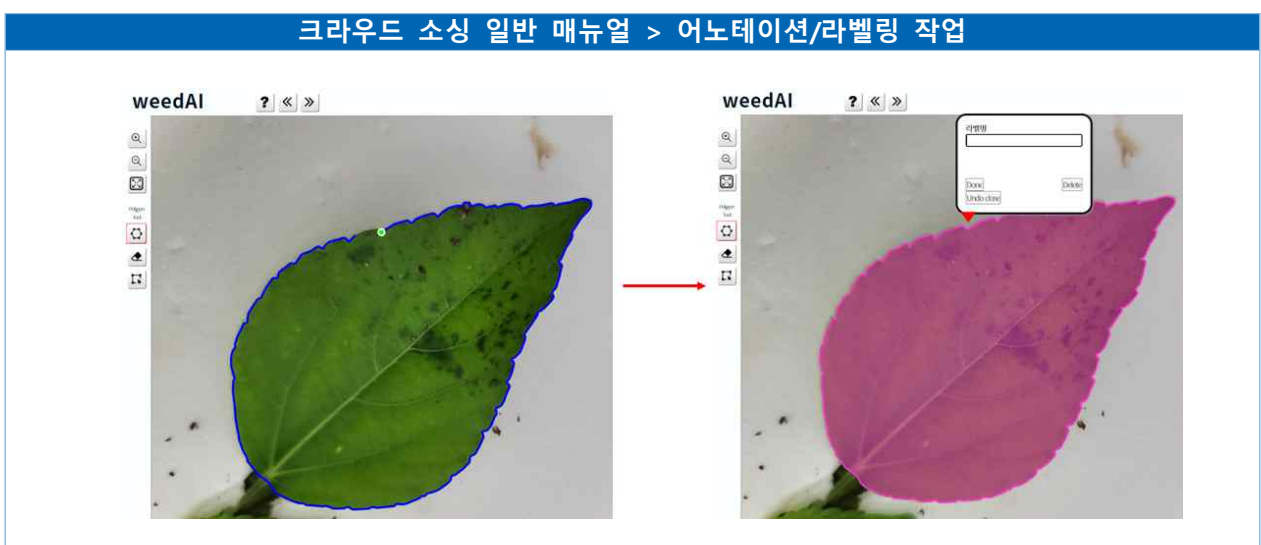
○ 어노테이션 절차



○ 클라우드 소싱 플랫폼



○ 클라우드 어노테이션/라벨링 작업 진행



2.4.3 어노테이션/라벨링 기준

○ 공통 라벨링 기준

- 작업대상 이미지는 촬영 부위에 따라 해당 잡초의 전체/잎을 촬영한 이미지로 구분된다.
- 잡초의 전체를 촬영한 이미지는 한 뿌리의 작물을 하나의 객체로 보며, 하나의 객체는 하나의 폴리곤으로 라벨링 한다.
- 잡초의 잎을 촬영한 이미지는 줄기로 이어진 부분을 하나의 객체로 보며, 하나의 객체는 하나의 폴리곤으로 라벨링 한다.
- 객체의 테두리를 정확하게 라벨링 한다.
- 객체의 테두리를 따라 라벨링 하면서 생기는 공백(배경)에 대해, 별도의 폴리곤을 형성하지 않는다.

라벨링 방식	설명	이미지 예시
Polygon	- 다각형 모양으로 객체의 가시 영역 외곽선을 따라 점을 찍어 그리는 라벨링 방법, 개체 이외의 포함된 빈공간으로 인해 발생하는 오류에 대응할 수 있는 기능 - 객체를 정확하게 인식하기 위해서 사용하며, 사물의 테두리를 따라 그리는 것을 통해 여백 없이 정확히 물체만을 인식하기 위해 사용	

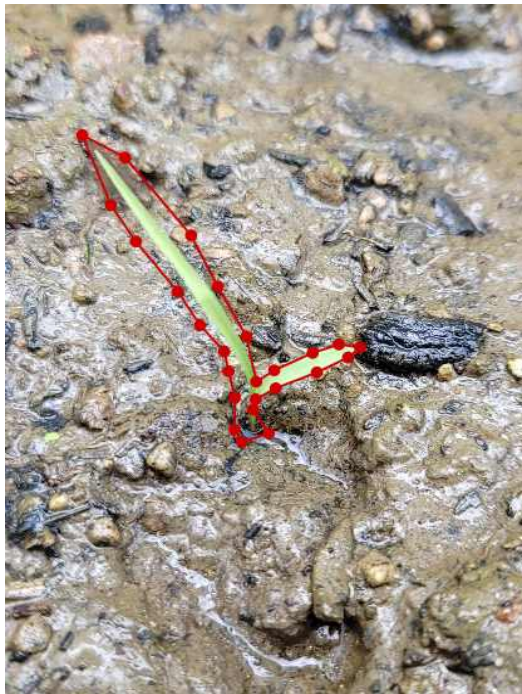
- 주요 잡초별 구체적인 라벨링 사례
- 강피(논피)

허용
이미지
예시



- 폴리곤 라인은 객체의 테두리를 따라 형성

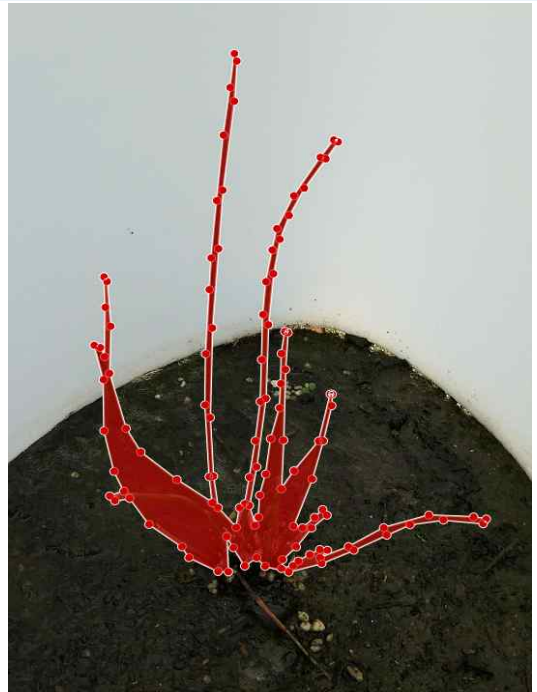
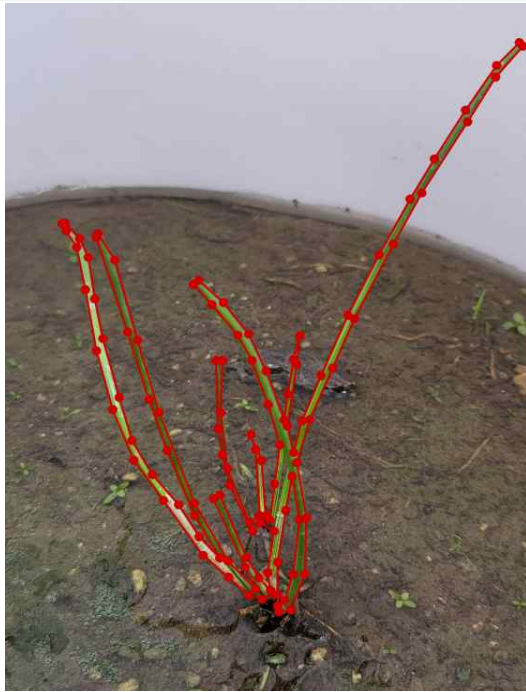
오류
이미지
예시



- 폴리곤 라인이 객체 테두리 바깥쪽으로 벗어나거나 안쪽으로 침범한 경우

- 올방개

허용
이미지
예시

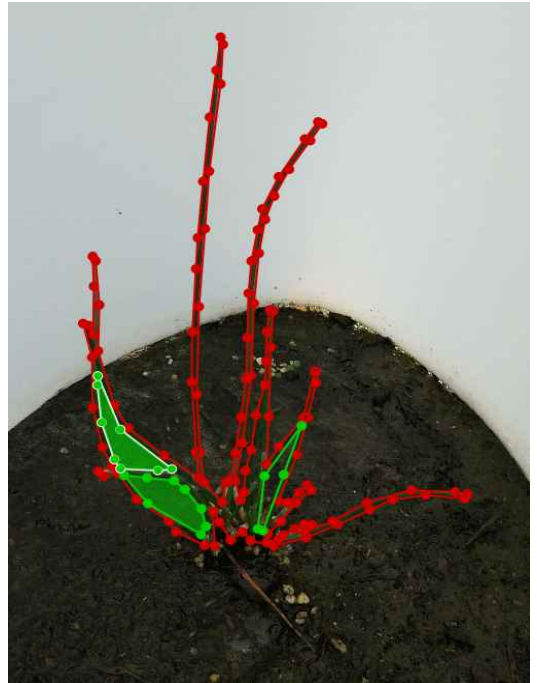


- 폴리곤 라인은 객체의 테두리를 따라 형성

오류
이미지
예시



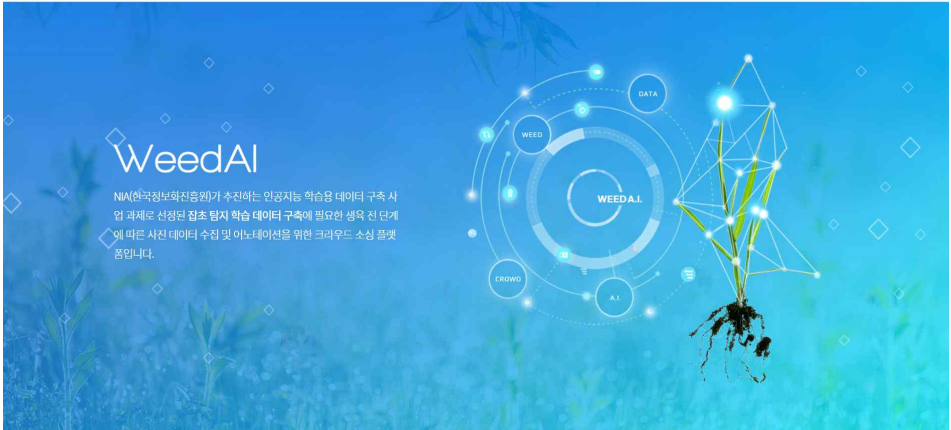
- 객체 내 공백(배경)을 포함하여 라벨링 한 경우



- 객체 내 공백(배경)을 별도로 라벨링 한 경우

2.4.4 어노테이션/라벨링 도구

○ 라벨링 도구

제품명	웹 기반 라벨링 저작도구
작업 화면	회원가입 라벨러 검수자 관리자  <홈페이지>  <작업 화면>
기능	- 객체 어노테이션 및 라벨링 태깅 - 객체 바운딩박스 및 세그멘테이션 - 라벨링 데이터 검수 - 회원 관리 - 작업 관리

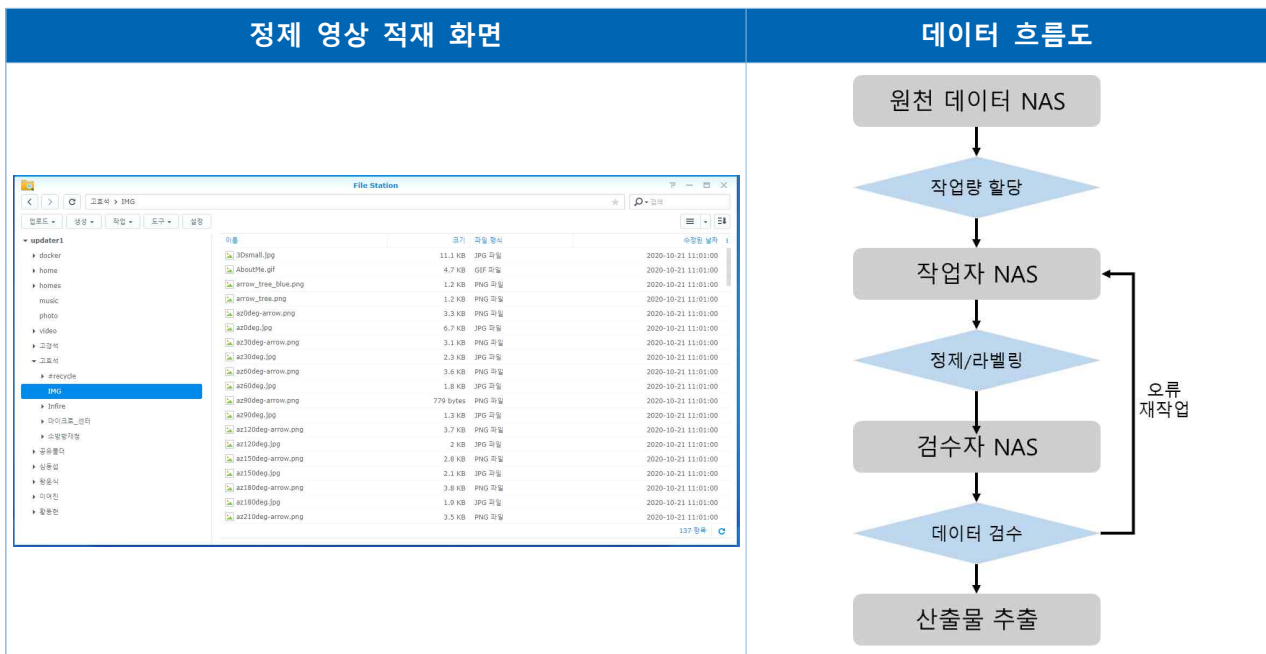
2.4.5 데이터 저장

○ 저장 형식

- 데이터 포맷

원시데이터	주요속성	데이터포맷
이미지 데이터	이미지 라벨링 속성	jpg, json
토양환경 데이터	토양 전기전도도, 토양 온도, 토양 습도, 토양 수분장력	csv, json
생장환경 데이터	온도, 습도, 일사량, 일출시간, 일몰시간	csv, json

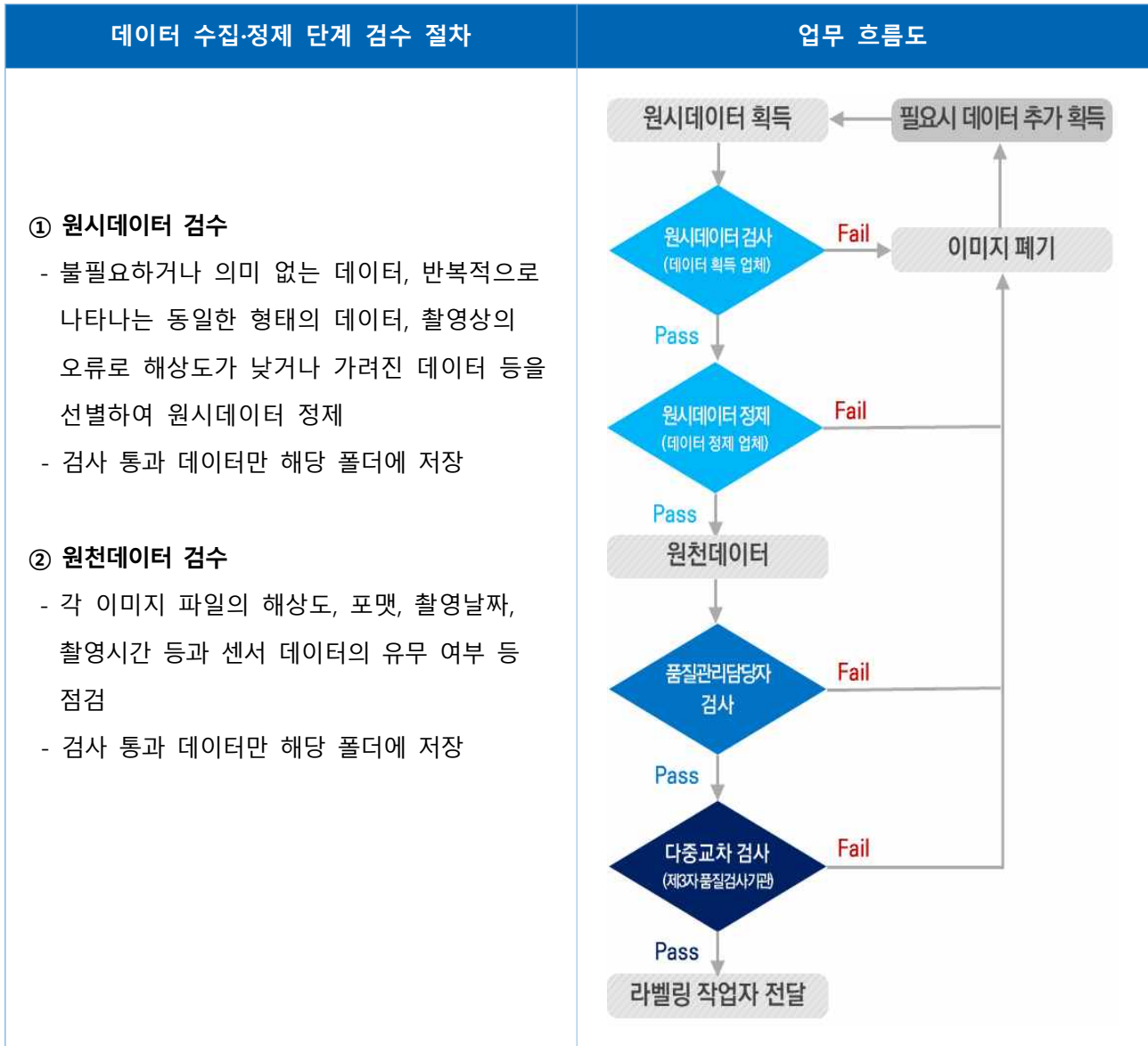
- 데이터 적재



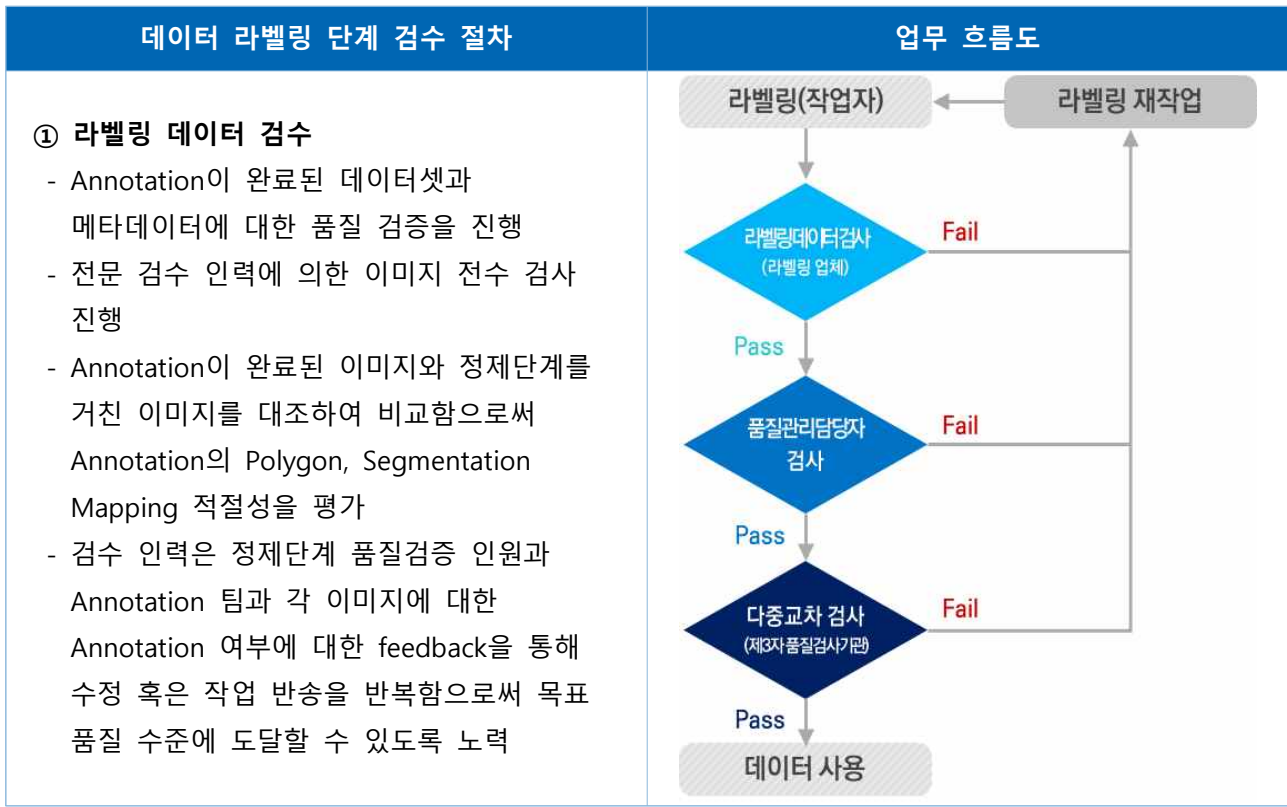
2.5 검수

2.5.1 검수 절차

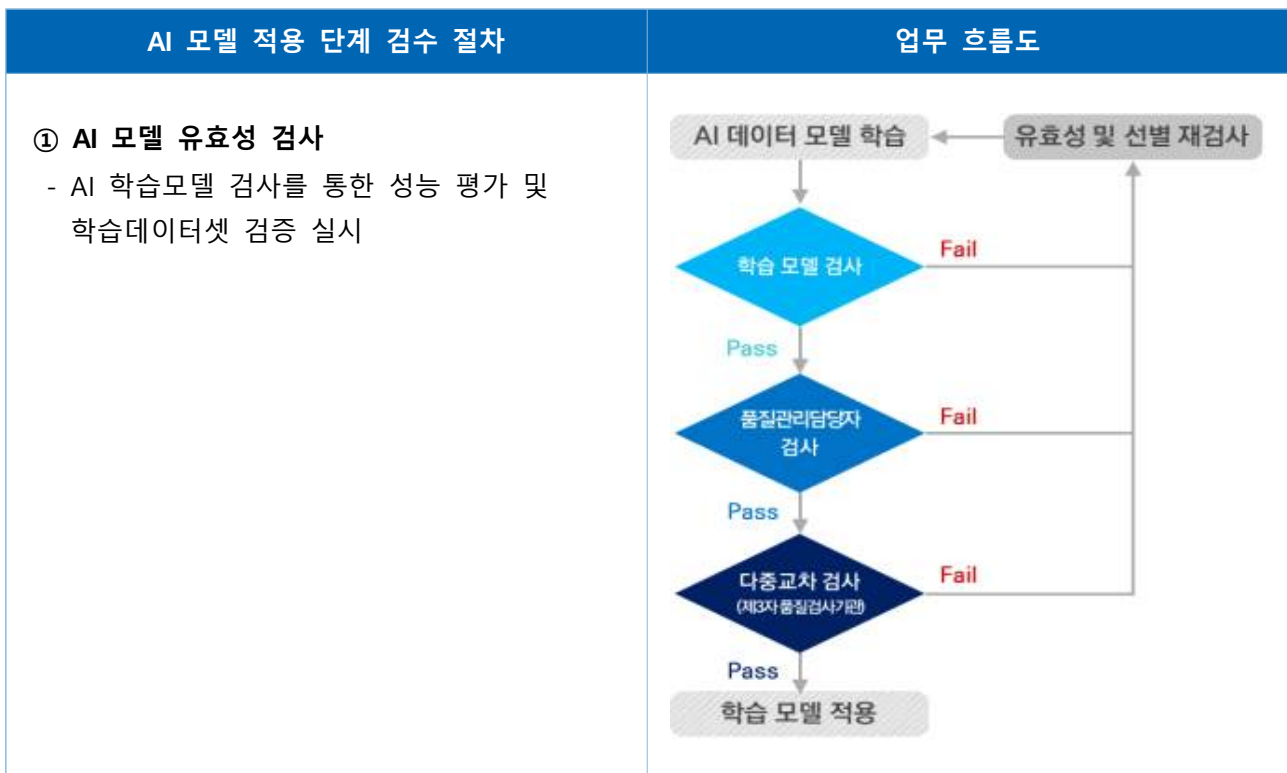
○ 데이터 수집·정제 단계 검수



○ 데이터 라벨링 단계 검수



○ AI 모델 적용 단계 검수



2.5.2 검수 기준

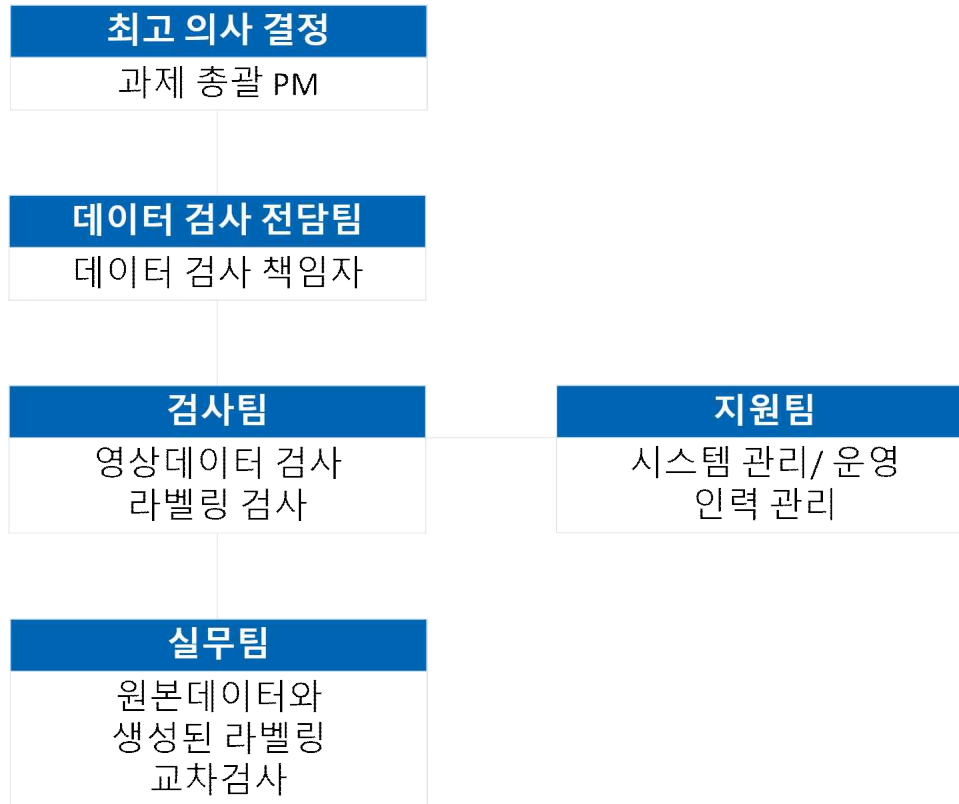
○ 데이터별 검수 기준

구분		기준 (조건)	주요 확인내용 (관리자)
지상 촬영 영상 데이터 검사	촬영 원본 검수	영상 품질 검수 (해상도, 촬영시기 등)	영상데이터별 200만 화소 & 적시성 (이레아이에스)
	처리 영상 검수	부정합 처리 영상 검수 (초점, 블러링 등)	품질(90%) 미달 데이터셋 재작업 (이레아이에스)
		잡초 미촬영 영상 검수	데이터 식별 가능성 (이레아이에스)
라벨링 데이터 검수	메타데이터 검수	데이터 라벨링 정확도 (ID 및 속성값, 지번, 좌표 연동 등)	속성 데이터의 정확도 검토 (업데이터)
		메타데이터 무결성 정확도 (개체, 참조 및 범위의 무결성)	이미지 및 속성 데이터의 정확성 및 일관성 검토 (업데이터)
		메타데이터와 이미지 일치성 검토 (ID 및 속성값)	학습데이터 검증셋 (업데이터)

○ 검수 기준에 따른 검사요소

검사요소		진단내용
원시 이미지 데이터	해상도	객체식별이 가능한 해상도인가?
	적시성	이미지 수집 시기는 언제인가?
학습용 데이터	정확성	사용 목적에 부합하는가?
	완전성	필요한 데이터 식별 수준인가?
	일관성	수집기준에 부합하는가?

2.5.3 검수 조직



※ 데이터 검사 전담팀, 검사팀, 지원팀, 실무팀 등으로 구성

※ 해당 분야의 도메인 전문가, 데이터 요청자, 외부 업체를 통한 검사 병행

2.6 활용

2.6.1 활용 모델

2.6.1.1 모델 학습

○ 인공지능 학습모델 선정기준 및 고려사항

구분	고려 사항	설명
1	적합성	• 주요 잡초 학습 데이터셋 구축 목적에 적합한 학습모델인가
2	활용성	• 정밀농업 및 스마트팜에서 분야 및 농업인에서 활용성이 높은 학습 모델인가
3	실현가능성	• 구축된 학습데이터셋을 활용하여 실제 농사에 적용하고 실현가능성이 높은 모델인가
4	선정 절차	1) 선정기준에 적합한 후보 리스트업 2) 내부 연구원 대상 설문조사를 통해 우선순위 도출 3) 1 cycle 학습모델 개발 4) 성능평가 5) 최종 학습모델 선정

○ 인공지능 학습 모델 후보군 적합성 검토 결과

- 데이터명 : 데이터102. 정밀농업 농기계 잡초 인식 데이터

후보 학습모델	적합성	활용성	실현 가능성	선정 여부
인공지능 기반 잡초 탐지 및 분류 모델 개발	상	상	상	선정
인공지능 기반 잡초 생육 분류 예측 모델	중	중	상	X
잡초 생육 단계와 센서 데이터 취득값 참조 모델 개발	중	중	중	X
적외선 영상데이터를 활용한 잡초 분류 모델	중	상	중	X

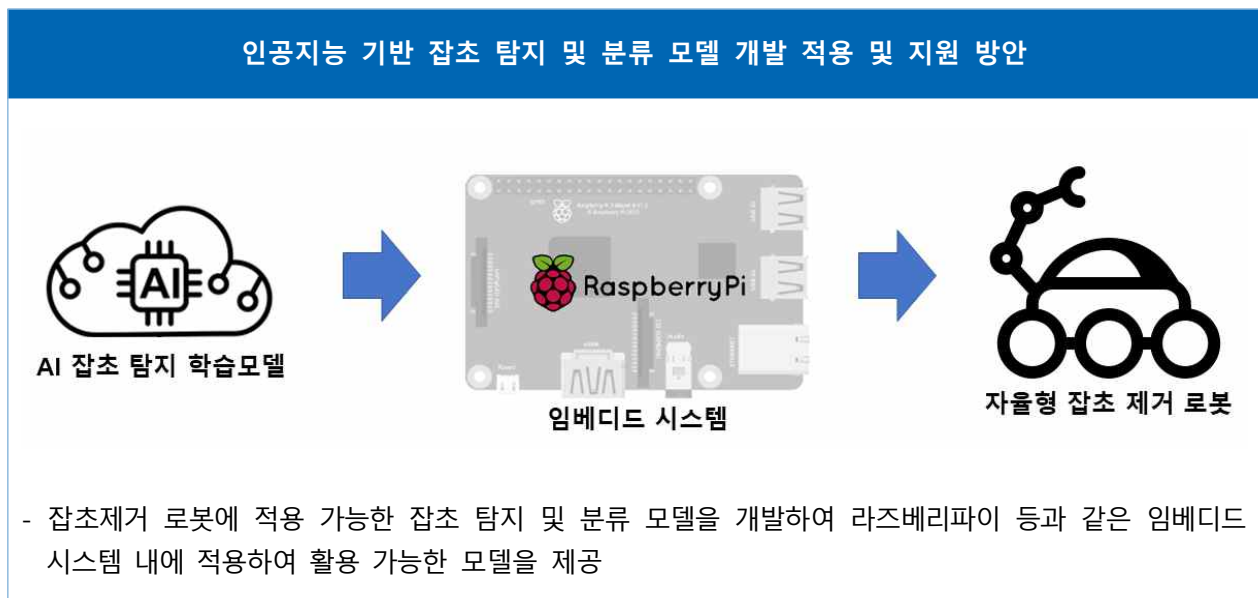
○ 모델 테스트

① (잡초) 인공지능 기반 잡초 탐지 및 분류 모델

- (개발 목표) 농기계 기반 잡초제거 로봇에 활용 가능하도록 밭과수원에 존재하는 잡초를 탐지하고 분류하는 모델을 개발
- (개발 내용) 구축되는 학습데이터를 활용 Mask R-CNN 기반 컬러 이미지에서 잡초 라벨링 데이터를 학습하여 잡초를 탐지하고 Segmentation을 통해서 잡초 분포 영역 및 잡초 크기를 예측할 수 있는 모델 개발
- (적용 방안) 잡초제거 로봇에 적용 가능한 잡초 탐지 및 분류 모델을 개발하여, 라즈베리파이에 탑재하여 바로 활용 가능한 모델을 제공

2.6.1.2 서비스 활용 시나리오

○ 인공지능 모델을 적용한 기술 혁신 지원 방안



2.6.2 데이터 제공

활용정책	세부 내용
배포방안	- AI Hub 사이트를 통하여 다운로드 가능하며, 한국지능정보사회진흥원(NIA)의 정책에 따름 - AI Hub 사이트를 통하여 데이터 공유 시스템을 구축하고 연구자 혹은 기업에서 비교적 자유로운 데이터 학습을 가능
데이터 활용시 필요조건	- 상업적 목적으로는 데이터베이스를 직접 공유하지 않고, 학습모델만 제공함 - 연구 목적 활용의 경우, 기관 및 활용 목적 등에 대한 검토 후 데이터셋을 제공하도록 하며, 상업적 목적으로 활용하는 기업의 경우에는 학습하고자 하는 인공지능 네트워크를 제공 받음