인공지능 데이터 구축 가이드라인
- 스마트팜 통합데이터 (파프리카) -

담당 역할	기관명
데이터 구축 총괄	원투씨엠㈜
데이터 설계	원투씨엠㈜
데이터수집 및 정제	(사)미래농업포럼
가공(라벨링, 어노테이션)	전북대학교 산학협력단
데이터 검수(자체 검수)	(사)미래농업포럼

구축 가이드라인 작성	원투씨엠㈜	이 종 민
	원투씨엠㈜	이 원 복
가이드라인 버전 (제작일자)	ver 1.0 ('22.03.25)	

목 차

1. 데이터 구축 개요 ····· 3

2. 문제 정의 ····· 4

2.1 임무 정의 ····· 4

2.2 데이터 구축 유의사항 ····· 4

3. 데이터 수집·정제 ····· 6

3.1 원시데이터 선정 ····· 6

3.2 수집·정제 절차 및 방법 ····· 7

3.3 수집·정제 기준 ····· 8

3.4 수집·정제 도구 ····· 8

4. 데이터 가공 ····· 10

4.1 가공 절차 및 방법 ····· 10

4.2 가공 기준 ····· 10

4.3 가공 규격 ····· 13

4.4 가공 도구 ····· 14

5. 검수 ····· 15

5.1 검수 절차 및 방법 ····· 15

5.2 검수 기준 ····· 15

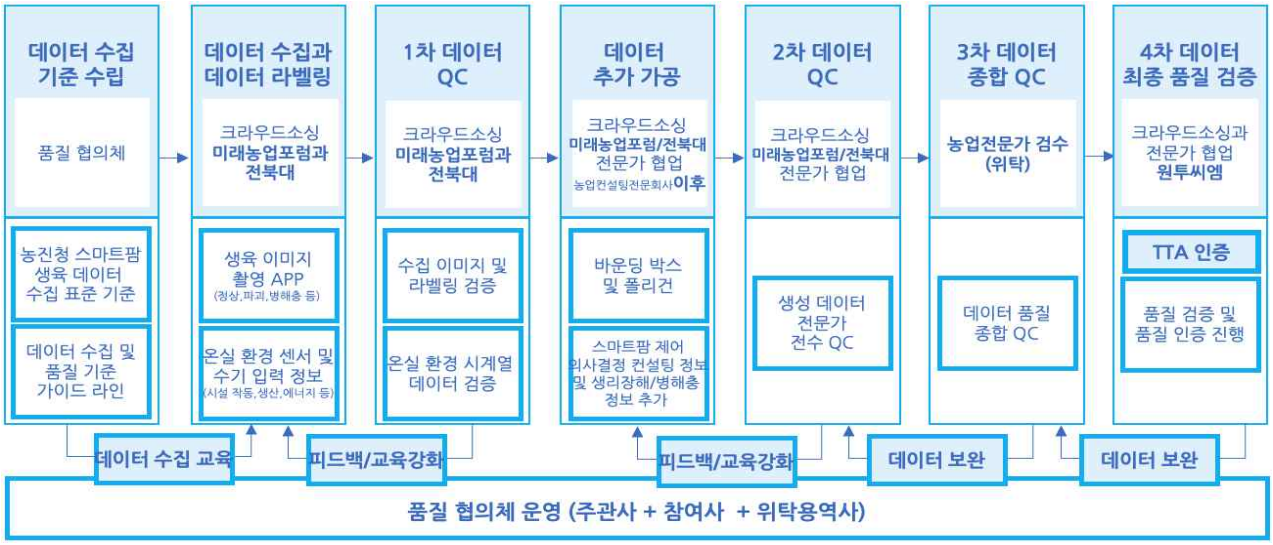
5.3 검수 조직 ····· 17

1. 데이터 구축 개요

▶높은 품질의 확보를 위해 6가지 선행 정비를 통한 데이터 수집 및 가공환경 구축

AI학습을 위한 데이터의 수집 및 가공환경의 구축은 농업관련 전문기관의 스마트팜 작물 생육데이터 취득기준을 참고하여 상세기준을 마련하고 이에 따라 기존 스마트팜 농가 현장에 추가로 필요 센서를 설치하는 것으로 시작되며, 동시에 기보유 스마트팜의 시계열 빅데이터 관리체계를 활용할 수 있는 기반을 마련하도록 구성함. 이렇게 구축된 센서 및 이미지 데이터 수집환경과 더불어 필요 수기데이터의 수집환경도 함께 제공함. 이미지 촬영기능을 기본으로 각종 부가정보, 추가 수기정보 및 바운딩 박스 적용등의 라벨링 기능을 포함하는 작물 재배 현장에서 사용되는 모바일 앱을 제공하며, 수집된 데이터는 어노테이션 플랫폼을 활용, 크라우드 워커들에 의하여 데이터 수집 활동이 진행됨.

▶품질 검증 및 피드백 단계를 가진 전체 7단계의 공정으로 학습용 데이터를 구축



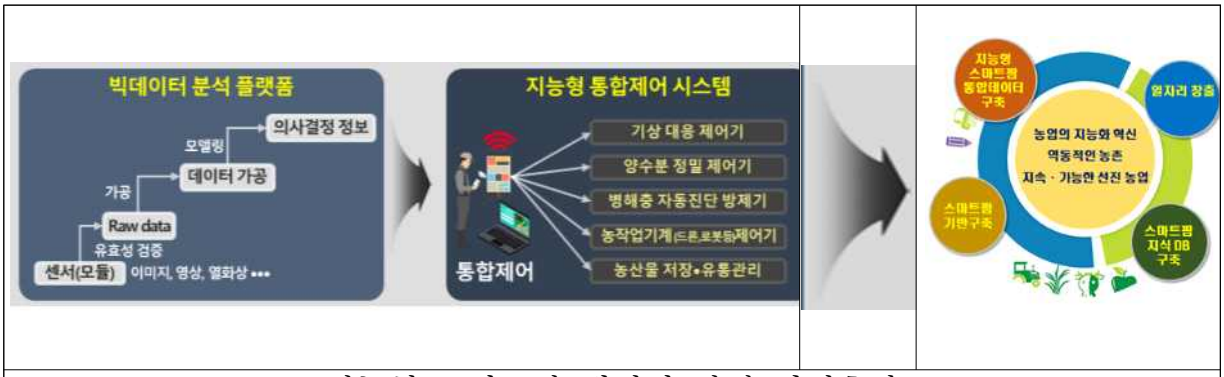
2. 문제 정의

2.1. 임무 정의

● 디지털 농업이 기피 산업에 속한 농업을 미래의 새로운 성장산업으로 육성하는 촉진제 역할을 함으로서, 농가인구의 감소, 고령화가 심화되는 상황 속에서 신규 농업인의 진입장벽은 낮추고, 수익성과 편리성은 강화된 농업의 지능화 혁신을 촉진하여 청년이 돌아오는 지속가능한 농업·농촌을 만들고, 동시에 인공지능 학습용 데이터 수집, 정제・가공, 검증 등에 직접고용과 함께 크라우드소싱 방식을 도입하여 일자리 창출 극대화하는 사업이다.

● 각종 기후변화와 국제 농산물 교역량 증가로 인해 농작물에 발생하는 고위험 문제병해충이 나날이 증가하고 있지만, 전문가가 아니면 정확한 진단이 불가능하여 피해를 줄이기 위한 방제조치 등 신속히 대처하기 어려워 농작물 피해를 체계적으로 줄이기 못하는 상황에서 농업인이 전문기관을 거치지 않고도 현장에서 실시간으로 직접 쉽고 정확하게 진단 및 대처를 할 수 있는 기술을 개발에 4차 산업 혁명의 핵심 기술인 AI를 활용할 수 있도록 병해, 생리장애와 작물보호제 처리 후의 반응에 대한 AI 학습용 데이터 구축을 목적으로 한다.

● 본 사업에 의해 생성된 지능형 스마트팜 통합 데이터 이용하면 작물의 생산성을 좌우하는 환경요인 조절과 재배기술 완전 자동화로 적은 노동력으로도 농산물의 안정적인 생산이 가능한 작물 전주기의 지능형 통합제어 시스템 구축이 가능하고 이에 덧붙여 국내에서 농가의 경제 경쟁력뿐만 아니라 우리나라 스마트농업 분야가 충분한 국가경쟁력을 확보 할 것으로 기대 한다.

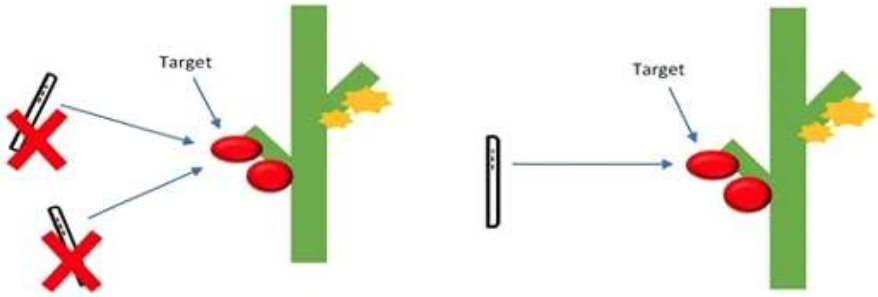


지능형 스마트팜 데이터 사업 기대효과

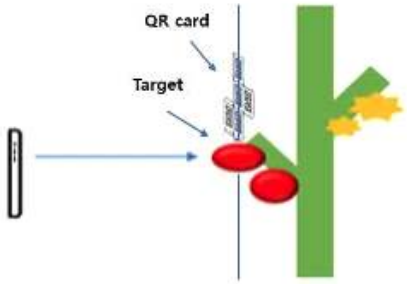
2.2. 데이터 구축 유의사항

반드시 사전 협의된 농가별 촬영 이미지 데이터에 대한 농가 및 촬영자와 저작권 관련 사전 협의를 통해 합법적인 데이터 수집이 이루어 질 수 있도록 함

1) 스마트폰을 가로 또는 세로로 단단히 쥐어 흔들리지 않게 하고 스마트폰과 촬영 대상 사이의 각도가 90도가 되도록 하여 촬영해 주십시오.



2) 촬영해야 할 대상과 인디케이터로 사용하는 QR card는 같은 선, 같은 면에 두고 촬영해 주십시오.



3) 촬영 이미지 하나에는 가급적 QR 코드를 하나만 사용해야 하며 촬영 대상과 QR 코드는 선명하게 촬영될 수 있도록 포커스에 주의해 주십시오.

4) 촬영 이미지에 촬영 대상과 QR 코드를 가리는 다른 물체가 들어가지 않도록 주의해 주십시오.

<작물개체 촬영시 일반 유의 사항(예시)>

3. 수집·정제

3.1. 원시데이터 선정

원시 데이터 선정 규모

항목	데이터 종류	수량(건)	비고
파프리카	이미지 데이터	208,000	스마트팜 6개소이상 (300평이상)
	온실환경 시계열 센서 데이터(5분간격)	225,792	
	생육지표 측정 데이터(파괴/실측)	9,600	
	끈끈이 촬영	96	
	시계열 가공 데이터	20,000	RFP외 추가데이터
	소계	**463,488**	

원시 데이터 분포

▷파프리카 촬영 이미지 : 208,000장
- [1단계] 5월중순~8주간, 1개농가, 100개체 대상 촬영 = 100개체*8주간*20장*1개농가 = 16,000장
- [2단계] 8월중순~16주간, 6개농가, 100개체 대상 촬영 = 100개체*16주간*20장*6개농가 = 192,000장

▷파프리카 센서데이터 (5분간격수집) : 225,792건
- [1단계] 5월중순~8주간, 2개농가, 288건/일 = 8주간*7일*288건*2개농가 = 32,256건
- [2단계] 8월중순~16주간, 6개농가, 288건/일 = 16주간*7일*288건*6개농가 = 193,536건

▷파프리카 생육지표 측정 데이터 (파괴 및 실측) : 9,600건
- 8월중순~16주간, 6개농가, 100개/주 = 16주간*100개*6개농가 = 9,600건

▷파프리카 스마트 온실내 끈끈이 촬영 : 96건
- 8월중순~16주간, 6개농가, 1컷/주 = 16주간*1컷*6개농가 = 96건

3.2. 수집·정제 절차 및 방법

○ 스마트팜 데이터 수집 환경 기준과 이를 충족하는 센서 시설 환경 보완 구축 후 데이터 수집

- 시설 제어 환경을 갖춘 300평 이상의 스마트팜에 외부 및 스마트팜 내부 센서를 보완하여 데이터 수집
- 외부 센서 데이터 종류 : 온도계, 일사량계, 풍향/풍속계 및 감우계 등
- 내부 센서 데이터 종류 : 온도계, 습도계, CO2 측정계 및 데이터 수집 장치 등
- 내부 시설 데이터 종류 : 열화상 카메라, 온도 변화를 분석하기 가진 인공 잎 등
- 수기 입력 데이터 종류 : 생산량, 에너지, 유동팬 작동 등
- 기존 RFP 요구 데이터 외 열화상 카메라를 추가로 설치하여 열화상 데이터를 추가로 수집
- 시계열 수집 데이터를 정제하여 이미지와 연결한 시계열 데이터 분석이 가능한 가공 데이터 추가 생성

○ 데이터의 신뢰성과 품질을 담보하려면 원칙적으로는 한 농가에서 재배하는 작물의 전생애 생육 데이터를 수집하는 것이 합당하나, 사업의 수행기간을 고려하자면 생육데이터 수집기간은 9월~12월(생육 초기~생육 중기 전반기)로 전생애 시계열데이터가 확보되지 않음. 완전한 시계열데이터를 얻기 위해서는 본 사업의 수행기간이 2년(당해년 8월 시작~익년 7월 종료)이거나, 1년이라면 늦어도 1월 말에 사업이 시작되어 2월부터 12월까지 데이터가 수집되어야 하나, 여러 가지 현실적인 문제로 사업수행기간은 제한적이므로 수집될 데이터에도 한계점이 있음. 이를 보완하기 위한 방안으로 작기가 달라서 생애주기가 다른 평지 3개 농가와 고랭지 3개 농가를 선정하여, 평지 농가에서 생육 초기~생육 중기 전반 데이터를 수집하고(9월~12월), 고랭지 농가에서 생육 중기 후반~생육 후기 데이터를 수집하는(7월~11월) 것으로 . 이때 데이터 수집 농가의 시설, 설비, 재배기술 수준이 높은 유사성을 가져야 하며, 재배하는 작물의 품종은 동일해야 한다.

○ 스마트팜용 시계열 빅데이터 관리 체계를 활용한 시계열 데이터 축적 및 가공 데이터 생성

- 시계열 빅데이터 플랫폼을 활용하여 체계적인 시계열 데이터 정제, 수집 진행
- 데이터 전처리 및 일차 가공을 통한 환경 시계열 생육 데이터를 구축

○ 어노테이션 및 크라우드소싱 플랫폼과 연계한 촬영 전용 APP을 제작 배포

- 스마트폰을 통한 이미지 촬영 후 데이터 수집 농장, 환경, 작물 정보 및 시간 정보와 함께 데이터 생성
- 가이드 창에 맞추어 촬영하여 바운딩 박스를 자동 생성하고 스마트폰에서 즉시 바운딩 박스 보정 및 폴리곤 기법을 통해 객체 분류
- 스마트폰 앱에서 생성된 데이터는 실시간 어노테이션 및 크라우드소싱 작업관리 체계로 정보가 적제되어 전체 공정과 연결되어 데이터 관리

○ 50% 농가 대상 스마트팜 제어 컨설팅 지원으로 최적 데이터 생성

- 환경제어 의사결정 컨설팅을 제공을 통하여 스마트팜의 제어를 통한 최적의 생육 환경에서의 데이터 수집을 진행
- 파프리카 농가 3곳에 컨설팅을 지원하여 컨설팅이 진행되지 않는 환경 50%에서의 데이터와 최적의 상황을 컨트롤하는 데이터를 병행 수집함.

○ 생리장해/병해충 및 파괴 및 실측 데이터를 목표 데이터 외 추가적으로 수집

- 전체 농가에서 발생하는 병, 해충 (끈끈이 촬영) 이미지 데이터를 전문가가 변별하여 어노테이

션 작업 및 라벨링 작업을 진행함

3.3. 수집·정제 기준

AI학습용 데이터 수집에 있어 수집 데이터의 기준의 수립은 전체 사업 수행 및 최종 결과물의 품질을 결정하는 중요한 시작점이며, 따라서 이와 관련한 면밀한 검토와 기준 마련을 위해 주관기관과 참여기관 및 위탁 용역사를 아우르는 품질 협의체를 구성하여 본 분야에 대한 엄격한 수집기준을 마련하며 본 품질 협의체는 본 사업의 종료시까지 운영함. 수집기준의 마련은 농진청, 농정원 등 농업 관련 전문기관의 스마트팜 작물의 생육데이터 취득기준을 참고하여 상세기준을 마련하고 이를 토대로 데이터를 수집할 농업 작물 재배 현장에 데이터 수집에 필요한 각종 센서 등의 환경을 설치하고 이와 동시에 본격적인 데이터 수집과 라벨링이 진행되는 2단계에서 준용될 데이터 수집 및 품질 기준 가이드 라인을 결정하고 이를 통해 농업 작물 재배 현장에서 데이터 수집과 라벨링을 담당하게 될 기관 및 크라우드 워커들에 대한 데이터 수집 교육을 실시함으로써 앞으로 수집될 데이터 품질 관리에 대한 첫 출발점이 됨

3.4. 수집·정제 도구

(개별 이미지 수집/촬영시 유의 및 고려사항)

재배 작물(파프리카) 촬영시 다음과 같은 사항을 고려하여 촬영 할 수 있도록 모바일 앱을 구성하여 제공하되, 필요시 일반 디지털 카메라를 함께 활용할 수 있음

○ 작물 이미지 촬영 해상도는 Full HD 장축 1920픽셀이상을 확보하며, 16(장축):9(단축)의 가로세로 비율로 촬영함 (1920X1080 혹은 1080X1920의 Full HD급)

○ 장비별, 객체별, 시간별, 종류별, 사람별, 지역별 검토를 위해 촬영이미지의 메타정보인 EXIF정보를 최대한 활용토록 하고 이와 함께 위치정보(GPS)를 활성화 하여 이를 함께 확보함

○ 다양한 상황, 다양한 형태에서의 학습용 데이터를 획득하기 위해 다양한 거리, 각도, 해상도, 조도 등을 반영한 촬영 작업 수행하되, 촬영의 용이성 및 이미지의 정확한 초점상태 확인을 위해 1미터 이내의 거리에서 촬영함(개체 전체 촬영을 위한 원거리 촬영 컷은 제외)

○ 촬영 데이터 각 건당 최소 8장 이상 촬영 (원거리 1장, 정면, 위, 아래 / 중심축 전면 기준 15도, 45도, 90도, 180도 등)하고, 촬영시, 줌인/줌아웃에서 촬영 객체의 화면이 '황금비율의 예시'와 동일하게 해상도와 색상, 화소 등이 최대한 원판과 흡사하게 표현 및 촬영 객체의 상이 틀어지거나 변경되지 않도록 촬영함 (촬영 대상 객체의 경우 전체 촬영 화면 70%이상의 영역을 차지하도록 촬영하여 피사체의 식별이 용이하도록 촬영)

○ 기본적으로 플래시를 이용한 촬영은 금지(촬영시 '플래시 OFF'으로 설정후 촬영 진행)하며, 만약 충분한 조도확보가 어려울 시, 촬영 대상작물의 변경(조도확보가 가능한 위치의 작물 선정) 및 필요시 별도의 조명 설치 검토후 촬영 진행함

○ 병해충 발생상황을 인위적으로 조성하여 촬영하지는 않으나, 본 사업을 위한 현장 작물 개체(파프리카)촬영 기간에 발생한 (1)질병(흰가루병, 잿빛곰팡이병, 잎곰팡이병) (2)해충(온실가루이, 담배가루이)에 대하여 촬영을 수행하고 질병의 경우 촬영부위는 작물, 해충의 경우는 '끈끈이'상의 발생해충을 그 대상으로 함.

○ 병해충 이미지를 촬영시 초점을 확보가능한 최대한 근접한 거리에서 병징이 선명하게 보이도록 촬영하며, 병징(흰가루병, 잿빛곰팡이병, 잎곰팡이병)이 양면으로 발생했을 경우, 각 면의 병징을 별도로 촬영하되 여러개의 중복된 병이 발생한 케이스는 촬영 대상에서 제외함. 또한 병해충이 발생했을 경우, 정상작물을 시작으로 초기, 중기, 말기로 병해충 피해 구분별 영상(정적) 이미지 및 작물

부위별, 감염 정도별, 시기별 다양한 피해 증상에 대한 데이터 확보를 획득함

○ 파프리카 생육 판단을 위한 항목, 스마트팜 내부 환경 데이터 실시간 시계열 데이터 수집 항목, 시설 제어 및 생산량, 에너지 등의 수기 입력 수집 항목, 양수분과 관련한 시계열 가공 데이터, 스크린 운영 관련 시계열 가공데이터, 내부 환경 시계열 가공 데이터, 시설의 환경 설정 데이터에 대한 기준 수립

○ 주관사 중심으로 참여사 및 전문 위탁 용역사로 구성된 품질 협의체를 통하여 기준 수립

○ 데이터 수집 가이드라인과 품질 기준에 대한 가이드라인을 수립

○ 데이터 수집과 어노테이션 , 품질 검증 교육 자료를 생성하여 참여 기관 및 참여 크라우드소싱 인력 대상 교육 진행

4. 데이터 가공

4.1. 라벨링 절차 및 방법

○ 스마트폰 앱에서 촬영 가이드라인을 통한 해당 작물의 전체 및 생육 단계별 제시된 가이드라인에 따른 부위별 직접 촬영을 진행하고 등록된 사용자, 농장, 작물, 시간 정보와 함께 데이터 생성

○ 촬영 시 바운딩 박스를 직접 생성 (촬영 가이드라인 박스내 촬영 후 바운딩 박스를 모바일 폰에서 보정)

○ 가이드라인에 따른 시설 제어 및 에너지 사용량 등에 대한 정보를 수기 입력하여 시간 정보 기준으로 데이터 생성

○ 센서를 통하여 수집되는 데이터는 시계열 데이터 플랫폼에서 정제 및 전처리 과정을 거쳐 시계열 데이터로 적제 생성

본격적인 데이터 수집과 데이터 라벨링에 착수하는 2단계는 미래농업포럼과 전북대 농생명과학대학의 주도하에 1단계에서 마련된 데이터 수집 및 품질 기준 가이드라인을 엄격하게 적용하여 크라우드 소싱으로 구성된 크라우드 워커에 의해 본 수집작업이 진행됨. 생육 이미지 촬영 기능을 기반으로 하는 모바일 앱을 활용, 정상이미지/파괴이미지/병해충 등의 촬영 이미지 및 관련 부가정보를 모바일앱 내 메뉴 선택기능 혹은 텍스트 수기방식으로 입력하여 함께 수집하며 이와 동시에 온실 스마트팜에 설치된 각종 센서장치를 통해 온실 내 시계열 데이터 및 스마트 팜 내부 설비의 구동이력 정보 그리고 출하량,품질등급을 포함한 생산량과 에너지 사용량 정보를 포함한 종합적인 데이터 수집 활동이 전개됨.

추가적으로는 본 사업에 참여하는 총 12개소(파프리카 6개소)의 스마트팜 중 50%에 해당하는 6개소의 수집현장(파프리카 각 6개소의 현장중 각 3개소)에 대해서는 특별한 생육분야 전문가 컨설팅 지원을 추가 진행함으로써 현장에서 수집되는 데이터 품질에 대한 관리를 병행하도록 하여 수집단계에서 보다 높은 수준의 데이터가 확보될 수 있도록 데이터 품질관리에 만전을 기하도록 함

4.2. 라벨링 기준

요구사항ID	품질 요구사항	요구사항 구분		
		구축 공정	구축 데이터	AI학습 모델
RD-001	· 원천데이터는 중복이 없어야 한다.		●	
RD-002	· 원천데이터의 이미지 해상도는 FHD급(1920 × 1080, 약 2 M Pixel) 이상이어야 한다.		●	
RD-003	· 생육 단계별, 명시된 시간별 다양한 분포로 원천데이터가 수집되어야 한다.		●	
RD-004	· 9가지 생육 단계별 원천데이터 구축 수량은 각 20,000장 이상이어야 한다.		●	
RD-005	· 원천데이터 정제 시 가로, 세로 비율(16:9 ~ 4:3 이내)이 급격한 이미지는 없어야 한다.		●	

RD-006	· 라벨링 데이터 포맷은 JSON을 사용해야하며, 바운딩 박스 및 세그멘테이션은 MS COCO 데이터셋의 어노테이션 포맷을 적용해야한다.		●	
RD-007	· 이미지 데이터 포맷은 PNG, JPEG을 사용해야 한다.		●	
RD-008	· 라벨링 데이터의 구문정확성 오류율은 해당 품질 목표를 달성해야 한다.		●	
RD-009	· 라벨링 데이터 구조의 속성 값은 항상 데이터가 존재해야 한다.		●	
RD-010	· 어노테이션 데이터의 바운딩 박스 및 세그멘테이션 정확도는 해당 품질 목표를 달성해야 한다.		●	
RD-011	· 원천데이터 수집은 협약 된 온실 6개소 이상의 재배환경에서 수집되어야 한다.		●	
RD-012	· 일관된 환경에서 원천데이터를 수집되어야 한다.		●	
RP-001	· 정제단계에서 불필요한 데이터 중복, 반복적으로 발생되는 동일 데이터 등을 확인해서 제외해야 한다.	●		
RP-002	· 학습용 데이터는 민간 개방 및 사용에 제약이 없도록 온실 내 원시 데이터 수집 시 개인정보 사용에 따른 동의를 확보해야 한다.	●		
RP-003	· 관련 법령에 따라 개인정보, 초상권 등 민감정보가 포함 된 경우 비식별화 해야한다.	●		
RP-004	· 학습용 데이터 구축을 위한 조직을 구성하여, 역할과 책임 수립 후 관리 및 운영되어야 한다.	●		
RP-005	· 이미지 수집 시 스마트폰 어플리케이션을 활용하여야 한다.	●		
RP-006	· 병해충 발생 데이터는 발생시기, 장소, 환경정보와 연계되어야 한다.	●		
RP-007	· 센서로부터 수집되는 환경 데이터 및 스마트팜 내부 설비 구동 정보는 자동적으로 수집되어야 한다.	●		
RP-008	· 센서로부터 수집되는 환경 데이터는 5분 간격으로 수집되어야 한다.	●		
RP-009	· 수집 이미지의 라벨링 및 어노테이션(바운딩박스, 세그멘테이션 등)이 정상적으로 이루어져야 한다.	●		

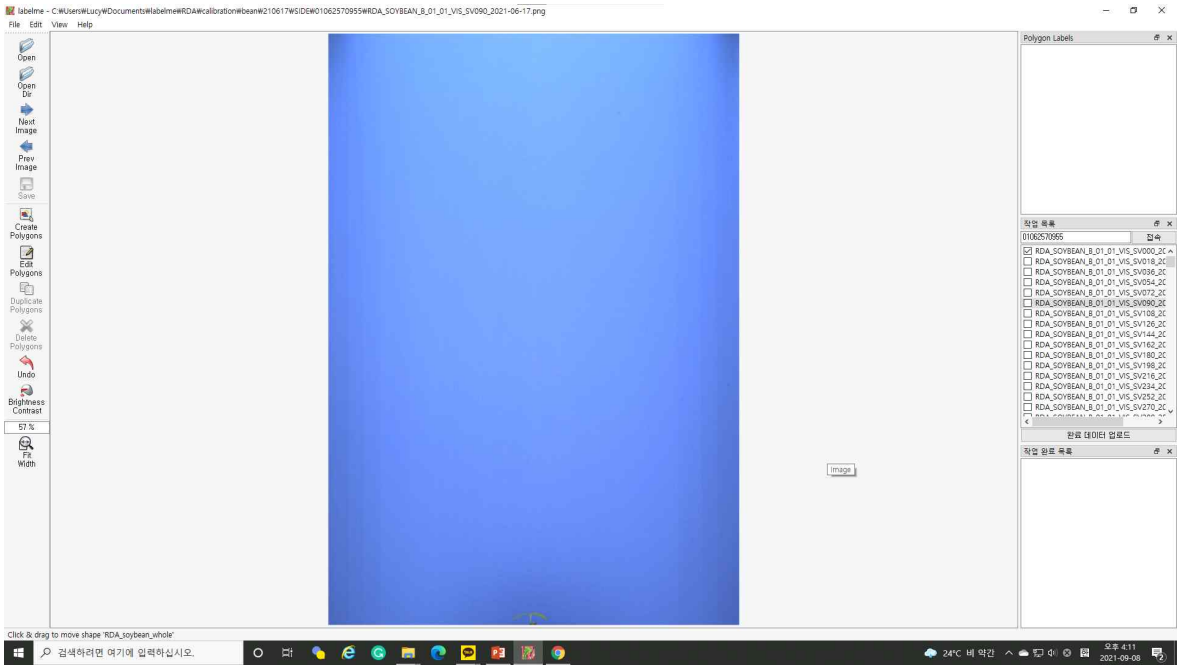
RP-010	· 이미지 데이터와 환경 데이터는 상호 연계가 되어야 한다.	●		
RP-011	· 각 공정 단계별 구축 방법 및 검증 계획에 따라 검증이 실시되어야 한다.	●		
RP-012	· 생육단계, 작물병/병해/생리장애 진전도 구분에 대한 판단을 위해 라벨링 시 작물질병 관련 전문가를 투입해야 한다.	●		
RP-013	· 라벨링/어노테이션에 적절한 도구가 선정되어야 한다.	●		
RP-014	· 라벨링/어노테이션 도구 사용에 따라 산출되는 데이터는 명시된 라벨링/어노테이션 포맷으로 구성되어야 한다	●		
RP-015	· 데이터 수집 및 정제, 라벨링 도구 정의 및 실무 교육을 통하여 크라우드 소싱 인력 작업 데이터 품질을 향상시켜야 한다.	●		
RM-001	· 단일 또는 복수개의 선정 된 인공지능 알고리즘을 적용해서 사업계획서상의 인공지능 학습 성능을 확보해야 한다.			●

4.3. 라벨링 규격

항목코드	촬영항목	어노테이션	라벨	메타정보 (수집앱셀렉트박스옵션명)	비고
a1	생장길이	BB	pap_growth_bb		
a2	최상위개화꽃높이	BB	pap_flower_stem_bb		
a3	줄기두께	BB	pap_stem_dimeter_bb		기본 줄기 두께
			pap_stem_dimeter_tilt_bb		기울어진 줄기 두께
a4	엽장엽폭	BB	pap_leaf_bb		
b	꽃	P	pap_c_flower_poly	미개화	
		P	pap_flower_poly	개화	
		P	pap_p_flower_poly	과실없이 시들은꽃	
c1	열매(착색전)	P	pap_fruit_green_poly	초록색	
c2	열매(착색후)	P	pap_fruit_color_poly	빨강 or 노랑 or 주황	
f	만개꽃	BB	pap_flower_tip_bb		만개꽃잎 끝
		BB	pap_flower_center_bb		만개꽃잎 중앙
		P	pap_flower_poly		만개꽃

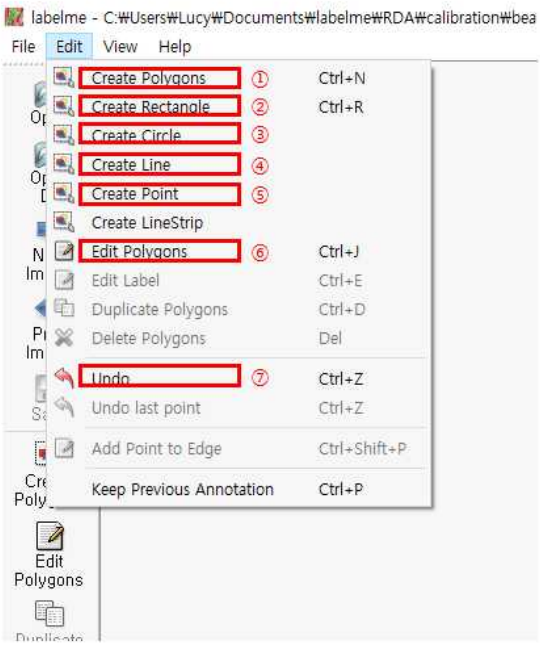
4.4. 라벨링 도구

Labelme 오픈소스 기반의 자체 저작 도구 개발



라벨링 도구 기능 목록

※ labelme tool 기능



1. 폴리곤 형태로 어노테이션 함.
2. 박스 형태로 어노테이션 함.(주로 사용)
3. 원 형태로 어노테이션 함.
4. 선 형태로 어노테이션 함.
5. 점 형태로 어노테이션 함.
6. 어노테이션 수정 가능 버튼(주로 사용)
7. 이전으로 돌아가기 버튼

5. 검수

5.1. 검수 절차 및 방법

(데이터 최종 품질 검증)

엄격한 기준의 자체 점검적 성격의 1차 데이터 QC --> 농업전문가와의 협업을 기반으로 진행되는 2차 전수 데이터 QC --> 객관적 시각의 외부 위탁기관을 통한 전문가 QC인 3차 데이터 종합 QC 이후 마지막 단계인 4차 데이터 최종 품질 검증은 주관기업인 원투씨엠에서 채용한 품질검증 전문 인력의 주도로 진행되며, 이전 단계(3차 데이터 종합QC)를 진행했던 품질관리 전문가들과 협업하여 TTA인증을 진행하게 되며, 필요시 데이터 보완이 필요한 부분에 대하여 재작업 혹은 수정 및 보완이 이루어 질 수 있도록 하여 최종 AI학습용 데이터의 품질을 관리하며 이를 기반으로 본 마지막 최종 품질 검증인 TTA검증을 통해 최종 데이터 인증을 확정함

5.2. 검수 기준

(1차 데이터 QC)

○ 전년도 사업 수행을 진행하였던 전북대와 미래농업포럼을 통하여 크라우드소싱 방식으로 1차 QC 진행

○ 1차 QC는 라벨링 정보의 검증 및 시계열 수지 데이터의 검증을 진행

○ 검증을 위한 시계열 생성 데이터 Viewer를 구성하여 QC에 활용

○ 1차 QC를 위하여 참여자 대상으로 교육을 진행하고 QC의 진행 결과를 전문가가 직접 체크하는 방식으로 1차 QC 진행

○ 1차 QC 진행 결과는 수집 데이터 작업자에게 직접 피드백이 전달되어 교육강화 또는 데이터 보정의 단계를 진행함.

이전 단계에서 수집된 모든 촬영 이미지와 부가정보 및 시계열 데이터는 역시 미래농업 포럼과 전북대 농생명과학대학의 주도하에 크라우드 소싱으로 구성된 크라우드 워커에 의해 1차로 수집된 데이터에 대한 QC가 진행됨. 수집된 이미지 및 수집활동간 각 이미지에 라벨링된 데이터에 대한 검증이 구성원간 교차적으로 진행되어 그 검수 품질을 높이고 온실 스마트팜에서 수집된 온실 환경 시계열 데이터 검증이 이루어짐. 본 QC과정의 결과를 토대로 이전 단계인 2단계 작업결과물에 대한 피드백과 추가로 필요한 교육에 대한 내용을 도출하여 2단계 작업진행에 대한 교육 및 재교육을 강화함으로써 점차로 작물 재배 현장 수집 데이터에 대한 기본 품질 향상에 대한 수준을 높일 수 있도록 함

(데이터 추가 가공)

○ 전년도 사업 수행을 진행하였던 전북대와 미래농업포럼을 통하여 크라우드소싱을 통하여 어노테이션 (폴리건) 작업을 진행하고 전문가 (병해충 부분 : 전북대 및 농가 컨설팅을 통한 제어 의사결정 정보: 이후)에서 데이터 추가 가공

○ 플랫폼을 통하여 시계열 가공 데이터 (일평균, 최고, 최저 , 평균 등)를 추가 가공 생성

이전 3단계를 통과한 수집 데이터들은 어노테이션 플랫폼으로 전송되어 데이터 추가 가공이 실시됨. 미래농업 포럼, 전북대 농생명과학대학 및 농업분야와 관련된 전문가와 협업하며 농업컨설팅전문회사 '이후'가 함께 참여하게 되고 크라우드 소싱으로 구성된 크라우드 워커는 어노테이션 플랫폼에서 어노테이션(폴리건) 작업을 진행함. 바운딩 박스, 폴리곤 작업 그리고 스마트팜 제어 의사 결정 컨설팅 정보 및 생리장해/병해충 정보가 추가되는 본 4단계에서는 전문가 집단과의 협업을 통해 보다 안정적인 데이터의 품질 관리가 이루어지는 환경이 조성되며 외부 위탁기관을 통한 농업전문가 검수가 이루어지는 2차 데이터 QC의 준비단계로서의 의미가 부여됨

(2차 데이터 QC)

○ 최종 취합된 데이터는 미래농업포럼 및 전북대 등 전문가와 크라우드 소싱인력과의 협업를 통하여 라벨링 중심의 생성 데이터 전수 품질 검증 및 피드백 진행

미래농업포럼, 전북대학교 농생명과학대학 등의 농업 전문가들이 크라우드 소싱으로 이루어진 참여자들과 함께 2단계~4단계까지의 과정을 통해 수집되고 생성된 데이터를 전수QC를 함으로써 또 한번 더 객관적이고 엄격한 기준에 맞는 품질관리과정이 진행되며 이전 1차 데이터 QC때와 마찬가지로 지금까지의 작업결과물에 대한 피드백과 추가로 필요한 교육에 대한 내용을 도출하여 적용하여 교육을 추가로 강화 함으로써 작물 재배 현장 수집 데이터에 대한 기본 품질 향상에 대한 수준을 높일 수 있도록 함

(3차 데이터 종합 QC)

○ 외부 위탁기관을 통해 농업 전문가의 객관적이고도 엄격한 데이터 품질 종합 검수

○ 주관기업인 원투씨엠에서 채용한 품질검증 전문인력을 중심으로 최종 품질 검증과 TTA 인증을 준비

이전까지의 과정이 주로 미래농업포럼과 전북대 농생명과학대학의 주도로 수집과 가공이 진행된 데이터에 대해 이루어진 1차 QC가 교차 검증 방식에 기반한 검수적 성격을 지녔다면, 본 6단계를 통한 3차 데이터 QC는 외부 농업 전문 위탁기관을 통해 1단계~5단계까지 수집되고 생성된 모든 데이터에 대해서 보다 객관적이고 엄격한 기준에 의해 진행되는 중요한 품질 검수 및 관리 단계의 성격을 가졌음. 이전 단계까지의 QC가 수집되고 생성된 데이터 그 자체와 연결된 부가정보등과의 관계 그리고 시계열 데이터가 올바르게 수집되었는지에 대한 정확성을 확인하는 데 중점이 있는 QC활동이었다면, 6단계에서 이루어지는 3차 데이터 종합 QC는 이렇게 수집되고 생성된 데이터에 대해 본 사업이 목표한 AI학습에 필요한 데이터인지에 대한 적격 여부를 판단하는 적확성 확인 측면을 포함한 종합적 개념의 QC라고 할 수 있음.

5.3. 검수 조직

○ AI학습용 데이터 실무협의체 구성

◆ 컨소시엄 참여기업의 실무자 협의체를 구성하여 과제 진행상황과 문제점, 개선 방안을 논의하고 과제 진행에 차질이 없도록 일정을 관리하여 과제 진행 방향 조율 및 시너지 강화

<실무협의회 구성 및 역할>

구성	역할
데이터 수집 협의체	• 데이터 수집이 각 컨소시엄 별로 일정에 맞춰 제대로 수집 되는지 확인하고 수량, 품질 등의 이슈에 대해서 공유하고 공동 대응함을 목표 • 부득이한 일정 차질에 대비하여 신속 대응팀의 일정을 공유하고 지원하는 역할이 주된 협의체의 의제로서 프로젝트가 원만히 수행될 수 있도록 지원하고 공유하는 것을 목적으로 함
추출, 정제 가공 등 데이터 생산 가공 협의체	• 수집된 데이터에서 프레임별 이미지를 생산하고 학습, 활용 가능한 모든 일련의 과정을 논의하고 협의하는 협의체 • 툴 사용에 대한 안정성과 편의성을 논의를 통해서 개선점을 도출하고 이를 활용하여 작업을 수행하는 실무자들이 일정에 따라 생산 속도 향상에 기여하는 목적 • 주요일정(초기데이터 제출일, 중간보고, 최종보고 등) 데이터 제출에 차질이 없도록 크라우드 워커 인력을 통한 데이터 확보 계획 수립 • 크라우드 워커 데이터 확보만으로 데이터 제출이 부족할 것을 사전에 파악하고 데이터 확보에 상용인력 투입 계획 수립
AI 학습 알고리즘 협의체	• 학습 데이터를 기반으로 AI 학습 알고리즘을 통해 학습이 제대로 이루어지고 있고 응용 상용화 서비스 적용을 위한 다양한 활용 방안 및 기술적 지원 방안을 논의하는 협의체 • 적용할 알고리즘 리스트를 다양하게 논의 하여 최적의 결과와 성과를 도출하는 것이 목표임
품질 표준화 협의체	• 외부 품질 검증 외에 내부에서 생산된 데이터의 품질을 높이고 라벨, 비식별화 조치 등 활용가능한 안전한 데이터 생산을 위해서 실무 협의체를 구성하여 프로젝트 진행상의 경험을 공유하고 이를 개선하여 최적의 품질 유지 방안을 도출하는 것을 목표로 함