

인공지능 데이터 구축·활용 가이드라인

- 체부암: 폐암 진단을 위한 영상 데이터-

인공지능 데이터 구축	사업 총괄	
	응용 서비스 개발	
	데이터 설계	
	원천데이터 수집 및 정제	 고신대학교복음병원 KOSIN UNIVERSITY GOSPEL HOSPITAL
	데이터 가공	
	데이터 검수	
	교육생 모집	
	AI모델 개발	
가이드라인 작성	에이아이트릭스	변주섭
	고신대학교복음병원	옥철호
가이드라인 버전	V6 2021.05.03	

목 차

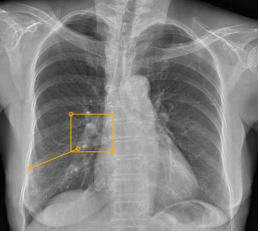
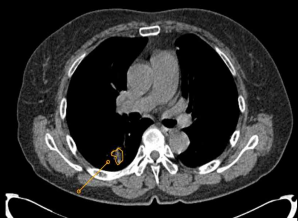
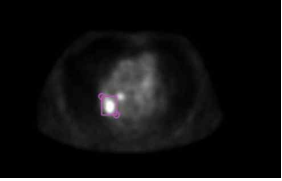
1. 데이터 명세 정보	1
1.1 데이터 정보 요약	2
1.2 데이터 포맷	3
1.3 어노테이션 포맷	4
1.4 데이터 구성	7
1.5 데이터 통계	7
1.6 원시데이터 특성	9
1.7 기타 정보	5
2. 데이터 구축 가이드	10
2.1 데이터 구축 개요	10
2.2 문제정의	10
2.3 획득·정제	11
2.4 어노테이션/라벨링	13
2.5 검수	16
2.6 활용	17

1. 데이터 명세 정보

1.1 데이터 정보 요약

데이터 이름	폐암 AI 학습 데이터	
활용 분야	의료	
데이터 요약	데이터 유형별(X-ray, CT, PET/CT) 양성/악성/정상 구축	
데이터 출처	고신대학교복음병원	
데이터 이력	배포버전	v4
	개정이력	신규
	작성자/ 배포자	옥철호 (고신대학교 복음병원)

1.2 데이터 포맷

영상 모달리티	데이터 포맷	어노테이션 항목	메타데이터
<p>X-ray</p> 	<p>DICOM</p>	<p>단일 결절 부위 Bounding box 체크</p>	<ul style="list-style-type: none"> ○ DICOM 파일 (비식별화 후) <ul style="list-style-type: none"> - age, sex, machine name ○ Json 파일 <ul style="list-style-type: none"> - 환자:json = 1:1 ○ Excel 파일 <ul style="list-style-type: none"> - 양성/악성정보 비식별 환자ID , 나이, 성별, 병리결과, 흡연력 (악성의 경우만 병리결과 수집)
<p>CT</p> 	<p>DICOM</p>	<p>단일 결절 부위 Polygon 체크</p>	<ul style="list-style-type: none"> ○ DICOM 파일 (비식별화 후) <ul style="list-style-type: none"> -age, sex, matrix size, machine name, slice thickness, series description ○ Json 파일 <ul style="list-style-type: none"> - 환자:json = 1:n ○ Excel 파일 <ul style="list-style-type: none"> - 양성/악성정보 비식별 환자ID , 나이, 성별, 병리결과, 흡연력 (악성의 경우만 병리결과 수집)
<p>PET-CT</p> 	<p>DICOM</p>	<p>폐 결절 및 임파선 부위 Bounding box 체크</p>	<ul style="list-style-type: none"> ○ DICOM 파일 (비식별화 후) <ul style="list-style-type: none"> - age, sex, machine name, ○ Json 파일 <ul style="list-style-type: none"> - 환자:json = 1:n ○ Excel 파일 <ul style="list-style-type: none"> - 양성/악성정보 비식별 환자ID , 나이, 성별, 병리결과, 흡연력 (악성의 경우만 병리결과 수집)

1.3 어노테이션 형식

1.3.1 JSON Info 포맷

No	영문명	한글명	필수여부	타입	길이	비고
1	description	슈퍼카테고리	Y	string	50	과제명
2	version	버전	Y	string	5	버전
3	contributor	배포자	Y	string	130	참여기업
4	date_created	데이터 생성일	Y	string	20	yyyy-mm-dd hh-mi-ss

1.3.2 JSON Categories 포맷

1.3.2.1 Category 구조

모달리티	supercategory	supercategory_name	grp_id	grp_name	id	name
Chest CT	CCT	Chest CT	1	악성	1	악성
					2	전이성_병변
			2	양성	3	양성
					4	섬유화반흔
					5	폐경화
					6	기흉
					7	무기폐
					8	흉막_삼출
					9	폐섬유화
X-ray	XRY	X-ray	1	악성	1	악성
					2	전이성_병변
			2	양성	3	양성
					4	섬유화반흔
					5	폐경화
					6	기흉
					7	무기폐
					8	흉막_삼출
					9	폐섬유화
PET-CT	PCT	PET-CT	1	악성	1	악성_림프절
					2	폐_병변_악성
			2	양성	3	양성_림프절
					4	폐_병변_양성

1.3.2.2 X-ray

No	영문명	한글명	필수여부	타입	길이	비고
1	supercategory	슈퍼카테고리	Y	string	3	XRY
2	grp_id	그룹 아이디	Y	number		1~2
3	grp_name	그룹 이름	Y	string	2	1 : 악성 2 : 양성
4	id	아이디	Y	number		1~9
5	name	이름	Y	string	10	1 : 악성 2 : 전이성_병변 3 : 양성 4 : 섬유유화반흔 5 : 폐경화 6 : 기흉 7 : 무기폐 8 : 흉막삼출 9 : 폐섬유화

1.3.2.3 Chest CT

No	영문명	한글명	필수여부	타입	길이	비고
1	supercategory	슈퍼카테고리	Y	string	3	CCT
2	grp_id	그룹 아이디	Y	number		1~2
3	grp_name	그룹 이름	Y	string	2	1 : 악성 2 : 양성
4	id	아이디	Y	number		1~9
5	name	이름	Y	string	10	1 : 악성 2 : 전이성_병변 3 : 양성 4 : 섬유유화반흔 5 : 폐경화 6 : 기흉 7 : 무기폐 8 : 흉막삼출 9 : 폐섬유화

1.3.2.4 PET CT

No	영문명	한글명	필수여부	타입	길이	비고
1	supercategory	슈퍼카테고리	Y	string	3	PCT
2	grp_id	그룹 아이디	Y	number		1~2
3	grp_name	그룹 이름	Y	string	2	1 : 악성 2 : 양성
4	id	아이디	Y	number		1~4
5	name	이름	Y	string	10	1 : 악성 림프절성 2 : 폐_병변_악성 3 : 양성 림프절성 4 : 폐_병변_양성

1.3.3 JSON Images 포맷

No	영문명	한글명	필수여부	타입	길이	비고
1	id	이미지 아이디	Y	number		Default=1
2	file_name	이미지 파일명	Y	string	50	영상이미지 파일명
3	patient_id	비식별환자아이디	Y	string	7	비식별 환자 아이디
4	series_id	시리즈 번호	Y	string	7	환자의 영상촬영 순번
5	image_id	이미지 번호	Y	string	7	시리즈의 이미지 순번
6	width	가로길이	Y	number		1~9999999
7	height	세로길이	Y	number		1~9999999
8	date_captured	생성날짜	Y	string	20	yyyy-mm-dd hh-mi-ss
9	modality	모달리티	Y	string	3	CCT: Chest CT PCT: PET CT XRY: X-ray
10	manufacturer	검사장비	Y	string	300	영상촬영 장비 제조사
11	manufacturer ModelName	검사장비 모델	Y	string	300	영상촬영 장비 모델명

1.3.4 JSON Annotations 포맷

No	영문명	한글명	필수여부	타입	길이	비고
1	id	어노테이션 아이디	Y	number		어노테이션 아이디
2	image_id	이미지 아이디	Y	number		시리즈의 이미지 순번
3	category_id	카테고리 아이디	Y	number		1,2,3
5	iscrowd	어노테이션 싱글/멀티 유형	Y	number		0: 싱글, 1: 멀티
6	area	어노테이션 넓이	Y	float		bbox를 활용한 넓이
7	bbox	어노테이션 최대Box 좌표	Y	float		[x1, y1, x2, y2] x, y는 float형식
8	toolname	어노테이션 툴 유형	Y	string	300	어노테이션툴명
9	segmentation	어노테이션 좌표정보	Y	float		[x1, y1, x2, y2, ..., xn, yn] x, y는 float형식
10	width	세그먼트 가로길이	Y	float		픽셀 길이
11	height	세그먼트 세로길이	Y	float		픽셀 길이

1.4 데이터 구성

경로					파일명 (모달리티_비식별ID_시리즈_순번_파일종류)	내용	
/lung_cancer	/x-ray	/benign, malignant, normal	/비식별ID	/시리즈	/dcm	xry_비식별ID_시리즈_순번.dcm	DICOM
		/benign, malignant			/binary	xry_비식별ID_시리즈_순번_카테고리ID_binary.dcm	Binary DICOM
					/json	xry_비식별ID_시리즈_순번.json	정보 : JSON
	/chest-ct	/benign, malignant, normal	/비식별ID	/시리즈	/dcm	cct_비식별ID_시리즈_순번.dcm	DICOM
		/benign, malignant			/binary	cct_비식별ID_시리즈_순번_카테고리ID_binary.dcm	Binary DICOM
					/json	cct_비식별ID_시리즈_순번.json	정보 : JSON
	/pet-ct	/benign, malignant, normal	/비식별ID	/시리즈	/dcm	pct_비식별ID_시리즈_순번.dcm	DICOM
		/malignant			/binary	pct_비식별ID_시리즈_순번_카테고리ID_binary.dcm	Binary DICOM
					/json	pct_비식별ID_시리즈_순번.json	정보 : JSON

※ benign : 양성 / malignant : 악성 / normal : 정상

1.5 데이터 통계

1.5.1 데이터 구축 규모

유형	목표 건수(건)		
	분류	환자 수	영상 수
X-ray	양성	500명	500장
	악성	3,000명	3,000장
	정상	10,000명	10,000장
흉부 CT	양성	1,000명	150,000장
	악성	2,500명	375,000장
	정상	1,000명	150,000장
PET/CT	양성	500명	125,000장
	악성	3,000명	750,000장
	정상	1,000명	250,000장

1.5.2 기타 활용 통계

유형	연도	저널명	논문명	환자 수
X-ray	2019	Radiology	Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs	4,3292명 (normal :nodule radiograph ratio (34067:9225장) (Malignant nodule 3,892 patient)
	2019	Journal of Thoracic Imaging	Performance of Deep Learning Model in Detecting Operable Lung Cancer With Chest Radiographs	17,211 s (nodule:3500 normal: 13,711)
CT	2020	Academic radiology	Development and Validation of a Modified Three-Dimensional U-Net Deep-Learning Model for Automated Detection of Lung Nodules on Chest CT Images From the Lung Image Database Consortium and Japanese Datasets	Luna 16 data set (888 CT scnas)
PET CT	2018	Lung Cancer	Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks – Initial results	100명(폐암 50명, 정상 50명)

1.6 원시데이터 특성

1.6.1 대상분류

- 고신대학교 복음병원의 의료데이터베이스(EMR, PACS)에서 실제 환자의 x-ray, CT, PET CT 이미지를 획득

1.6.2 제약조건

- 환자의 개인 정보가 포함된 원천 의료데이터는 개인 또는 기관이 직접적인 접근을 할 수 없음
 - 연구계획서를 작성하여 병원의 IRB 조직 승인을 받아야 환자 의료 데이터를 획득 가능하며, 승인된 연구자에 한해서만 의료 데이터에 접근 가능함

1.6.3 속성

- 수집 메타데이터
 - 비식별화 된 환자 ID
 - 검사장비
 - 검사장비의 모델명
 - 양성/악성 정보
 - 조직형 진단(악성데이터의 경우)

1.7 기타정보

1.7.1 포괄성

- 상기의 선행논문에서 활용한 의료 데이터의 수와 비교 했을 때, 이번 과제를 통해 구축하려는 데이터셋의 규모는 충분히 모집단의 특성을 나타낼 수 있다고 판단됨.

1.7.2 독립성

- 환자의 의료정보가 포함되어 있는 의료 데이터(원시 데이터)는 승인된 연구자 이외에는 접근이 불가능함.
- 보건복지부의 보건의료 데이터 활용 가이드라인에 따라, 데이터 활용 및 제 3자의 배포를 위해서는 해당 의료 기관의 데이터 심의기관의 허가를 받아야 함.

1.7.3 유의사항

- 원천 데이터에 대한 환자의 비식별화 작업을 거쳤으나, 민감한 의료 정보가 담긴 데이터이기 때문에 데이터 수정과 배포에 각별히 유의해야 함

1.7.4 관련 연구

- 해당사항 없음.

2. 데이터 구축 가이드

2.1 데이터 구축 개요

단계	과정	내용		
1단계	기초데이터 획득 및 정제	<ul style="list-style-type: none"> 고신대학교 복음병원의 의료데이터베이스(EMR, PACS)에서 실제 환자 데이터 추출 폐암 3종 이미지 <ul style="list-style-type: none"> X-ray CT PET CT 		
2단계	익명화 및 비식별화	자동 익명화 및 비식별화 SW를 개발하여 공통 규격의 데이터로 가공		
3단계	데이터 어노테이션	X-ray	<ul style="list-style-type: none"> 어노테이션 항목 [1차 작업] <ul style="list-style-type: none"> 병변부위 라인 체크 [2차 작업] <ul style="list-style-type: none"> 병변부위 바운딩 박스 체크 	<ul style="list-style-type: none"> 어노테이션 수행인 [1차 작업] <ul style="list-style-type: none"> 영상의학과 전문의 호흡기내과 전문의 [2차 작업] <ul style="list-style-type: none"> 별도 교육을 받은 인력
		CT	<ul style="list-style-type: none"> 어노테이션 항목 [1차 작업] <ul style="list-style-type: none"> 병변부위 라인 박스 체크 [2차 작업] <ul style="list-style-type: none"> 병변부위 폴리곤 체크 	<ul style="list-style-type: none"> 어노테이션 수행인 [1차 작업] <ul style="list-style-type: none"> 영상의학과 전문의 호흡기내과 전문의 [2차 작업] <ul style="list-style-type: none"> 별도 교육을 받은 인력
		PET CT	<ul style="list-style-type: none"> 어노테이션 항목 <ul style="list-style-type: none"> 병변부위 바운딩 박스 체크 	<ul style="list-style-type: none"> 어노테이션 수행인 <ul style="list-style-type: none"> 핵의학과 전문의 호흡기내과 전문의
4단계	검증	<ul style="list-style-type: none"> 데이터 생성 시 최소 2인의 크로스체크를 통한 품질관리 수행 영상의학과 전문의로 구성된 검증 인력을 구성하여 데이터 검수 외부기관 TTA를 통한 데이터 품질 검증 		
5단계	시모델개발	<ul style="list-style-type: none"> 병변 유/무 진단, 병변 검출 모델 개발 모델 성능 확인 및 제시 		
6단계	응용서비스 개발	<ul style="list-style-type: none"> 이미지를 input하면 병변의 양성/악성 유무와 위치를 표시해주는 프로그램 개발 		

2.2 문제정의

2.2.1 임무 정의

2.2.1.1 추진목적

- 인공지능 기반 흉곽 구조를 이해하는 단순흉부영상 영상 판독 기술 정확도 향상
- 인공지능 기반의 흉부전산단층촬영의 영상판독 기술 정확도 향상 및 폐암 예측도 향상. 특히 간유리음영 소결절에 대한 검사자 간의 판독 결과의 차이와 검사 시간을 줄이기 위한 목적

- PET CT의 소결절에 대한 폐암예측도 향상 최종 조직검사와 비교하여 검사의 정확도 판단

2.2.1.2 활용대상

- 단순흉부영상, 흉부전산단층촬영 영상, PET CT 영상 판독 기술을 종합적으로 판단하여 최종으로 악성의 예측도를 향상하는 프로그램 개발 및 실증.
- 폐암의 진단 및 예후 예측을 위한 인공지능 기반 예측 요인 개발을 위한 실증 사업 솔루션 개발 및 구축, 그리고 사업화 확장.

2.2.2 데이터 구축 유의사항

- 환자의 의료정보가 포함되어 있는 의료 데이터(원시 데이터)는 승인된 연구자 이외에는 접근이 불가능함.
- 보건복지부의 보건의료 데이터 활용 가이드라인에 따라, 데이터 활용 및 제 3자의 배포를 위해서는 해당 의료 기관의 데이터 심의기관의 허가를 받아야 함.

2.3 획득·정제

2.3.1 원시데이터 선정

유형	목표 건수(건)		
	분류	환자 수	영상 수
X-ray	양성	500명	500장
	악성	3,000명	3,000장
	정상	10,000명	10,000장
흉부 CT	양성	1,000명	150,000장
	악성	2,500명	375,000장
	정상	1,000명	150,000장
PET/CT	양성	500명	125,000장
	악성	3,000명	750,000장
	정상	1,000명	250,000장

- 규제 관련 사항

- 원시 데이터 획득을 위해 고신대학교 복음병원의 IRB 승인 획득
- 환자 비식별화 및 익명화 작업을 수행하여 의료 관련 개인정보보호법에 침해되지 않도록 처리

2.3.2 획득·정제 절차

구분	내용	
데이터 획득	<ul style="list-style-type: none"> 고신대학교 복음병원의 의료데이터베이스(EMR, PACS)에서 환자의 폐암 관련 X-ray, CT, PET CT 영상을 획득 	
데이터 정제	공통 사항	<ul style="list-style-type: none"> 익명화 SW를 이용해 환자 ID, 영상 ID를 수집 데이터베이스에 의해 관리되는 기준 ID로 치환하고, 임상 데이터와 함께 가공 기관 및 관련 정보가 함께 관리될 수 있도록 데이터베이스를 구축 Dicom 헤더에 포함된 환자 정보를 제거 영상 내에서 Tagging 혹은 기기별 영상 레이아웃에 의해 포함되는 환자 식별 정보를 제거

2.3.3 획득·정제 기준


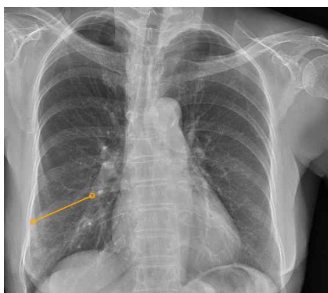
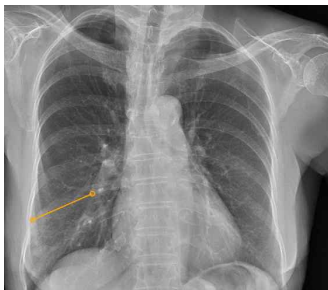
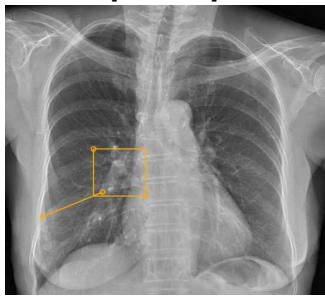
유형	수집기준	배제기준
X-ray	<ul style="list-style-type: none"> Chest PA 영상만을 사용 정상/양성/악성 기준 <ul style="list-style-type: none"> 정상: CT에서 정상으로 확인된 케이스 양성: 조직학적으로 양성으로 확인되거나 CT에서 경과 관찰 중 변화가 없는 경우 악성: 생검 및 수술로 악성으로 확진된 병변으로, CT 판독에서 nodule이 0.4cm이상인 환자의 X-ray 영상 	<ul style="list-style-type: none"> 뚜렷한 인공음영이 있는 경우, Chest AP영상은 제외 CT에서 확인된 결절 중 영상의학과 전문의의 판단에 결절이 보이지 않는 경우는 수집 대상에서 제외함
CT	<ul style="list-style-type: none"> 정상/양성/악성 기준 <ul style="list-style-type: none"> 정상: CT에서 정상으로 확인된 케이스 양성: 조직학적으로 양성이 확인되거나 흉부 CT에서 경과 관찰 중 변화가 없는 경우 악성:nodule이 0.4cm이상인 환자의 X-ray 영상 Slice Thickness: 2~10mm 사이의 영상을 대상으로 함 	<ul style="list-style-type: none"> 움직임에 대한 인공음영이 있는 경우 Slice spacing이 있는 경우는(절편 사이의 간격)영상 수집에서 제외함
PET CT	<ul style="list-style-type: none"> 정상/양성/악성 기준 <ul style="list-style-type: none"> 정상: 폐와 종격동 임파선에 PET negative를 보인 경우 양성: 폐에 결절 가진 환자 중 조직학적으로 양성으로 확인되거나 2년 경과 확인 시 변화가 없는 경우 악성: 생검 또는 수술로 악성이 확인된 경우 	<ul style="list-style-type: none"> DICOM아닌 Image 형태의 영상

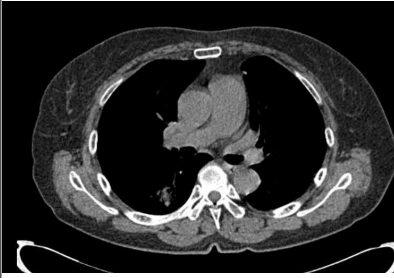
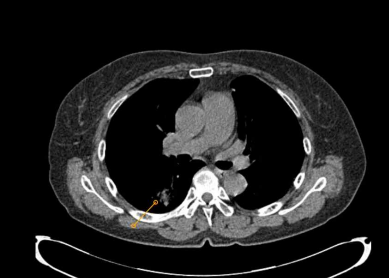
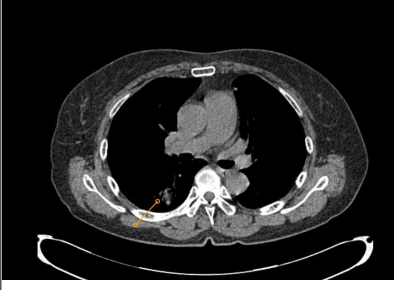
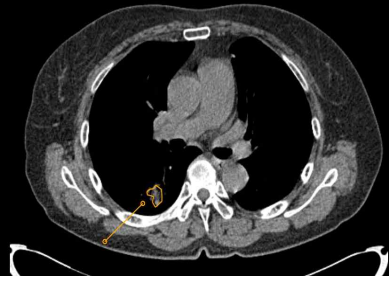
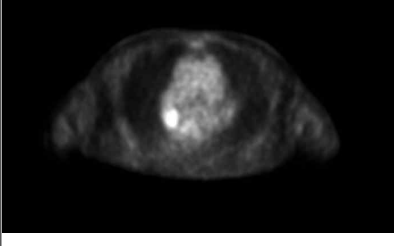
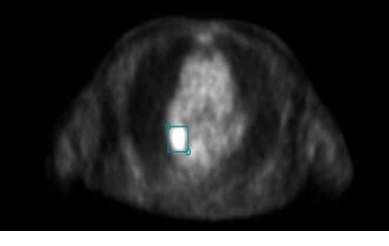
2.4 어노테이션/라벨링

2.4.1 어노테이션/라벨링 절차

구분	X-ray	CT	PET CT
1차 수행	- 수행인: 영상의학과, 호흡기내과 전문의 - 수행 내용: 병변부위 라인 체크	- 수행인: 영상의학과, 호흡기내과 전문의 - 수행 내용: 병변부위 라인 체크	- 수행인: 핵의학과, 호흡기내과 전문의 - 수행 내용: 병변부위 바운딩 박스 체크
2차 수행	- 수행인: 별도 교육을 받은 인력 - 수행 내용: 병변부위 바운딩 박스 체크	- 수행인: 별도 교육을 받은 인력 - 수행 내용: 병변부위 폴리곤 체크	해당사항 없음

2.4.2 어노테이션/라벨링 기준

유형	어노테이션 항목		어노테이션 수행인
X-ray	○ 1차 작업 - 병변 부위 라인 체크 (처리 데이터 수가 많아, 1차 작업(프리어노테이션)을 전문의가 수행한 후 교육을 수행한 인력이 2차 작업을 수행함) [수행 전]		○ 전문의 - 영상의학과 - 내과 전문의
			
X-ray	○ 2차 작업 - 병변부위 Bounding box 표시 [수행 전]		○ 별도의 교육을 받은 교육생
			
CT	○ 1차 작업 - 병변부위 라인 체크 [수행 전]		○ 전문의 - 영상의학과 - 내과 전문의
		[수행 후]	

			
	○ 2차 작 - 병변부위 폴리곤 표시		
	[수행 전]	[수행 후]	
			○ 데이터 레이블링 관련 교육을 수료한 의공학, 공학, 의학 등 전공자
	○ 병변부위 바운딩 박스 체크		
	[수행 전]	[수행 후]	
PET CT			○ 전문의 - 핵의학과 전문의 - 내과 전문의

2.4.2.1 어노테이션 조직 구분

- 내부 전문가 그룹: 영상의학과, 내과, 핵의학과 전문의
- 외부 교육생: 데이터 레이블링 관련 교육을 수료한 의공학, 공학, 의학 등 전공자

2.4.2.2 데이터 어노테이션 관련 교육 진행

- 어노테이션을 수행하는 전문의 및 외부 수행자를 대상으로 어노테이션 툴(저작 도구) 사용 방법, 어노테이션 방법 등 교육 진행

교육 구분	내용	수행 기간	비고
기본 교육	<ul style="list-style-type: none"> • 암종의 이해 • 데이터 형태에 대한 이해 • 어노테이션 도구 교육 • 실습 	<ul style="list-style-type: none"> • 암종별 1일 • 반복 수행 	온/오프라인
심화 교육	<ul style="list-style-type: none"> • 암종의 세부 교육 • 데이터 어노테이션 방법 • 데이터 어노테이션 결과 평가 방법 	<ul style="list-style-type: none"> • 암종별 1주 • 반복 수행 	온 /오프라인

	•실습	
--	-----	--

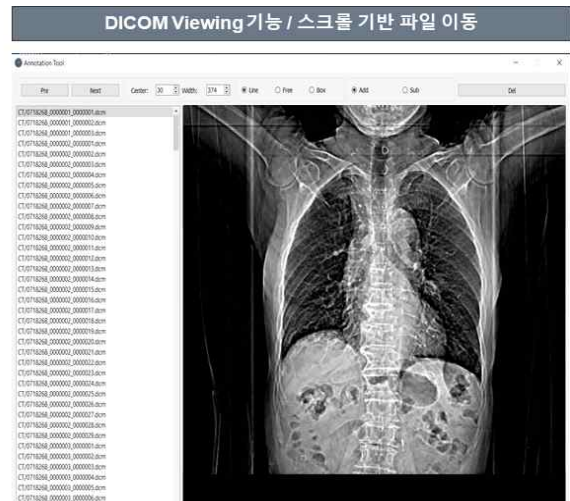
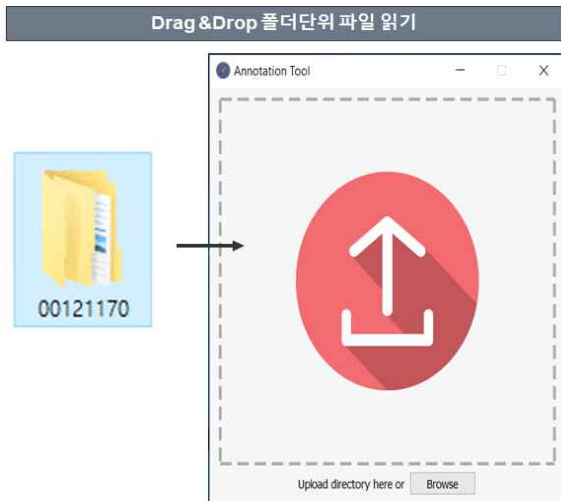
2.4.3 어노테이션/라벨링 도구

1) 저작도구

- 학습용 데이터 저작도구는 양성 영역 경계 어노테이션, 종양영역 경계 어노테이션, 자동 어노테이션 결과 저장 등을 제공하는 경북대학교 산학협력단에서 개발한 Window/Python 기반 프로그램을 사용함

2) 기능

- 폴더 읽기 / DICOM VIEWING 기능



- Annotation Drawing 편의 및 Output

Annotation 편의 도구

Center: Width:

Line Free Box

Add Sub

Window Level 밝기 조절

- DICOM 이미지의 Window Level 조절을 통해 Detail할 Annotation 가능하도록 함

Drawing 도구

- Line, Free(Polygon), Box 형태의 Drawing 도구를 제공하여 다양한 형태의 Annotation이 가능하도록 함

Drawing 보정

- Add(Annotation 끼리 Merge) or Sub(Annotation 부분 제거) 기능을 사용하여 Annotation 보정을 빠르게 할 수 있도록 함

Annotation Output

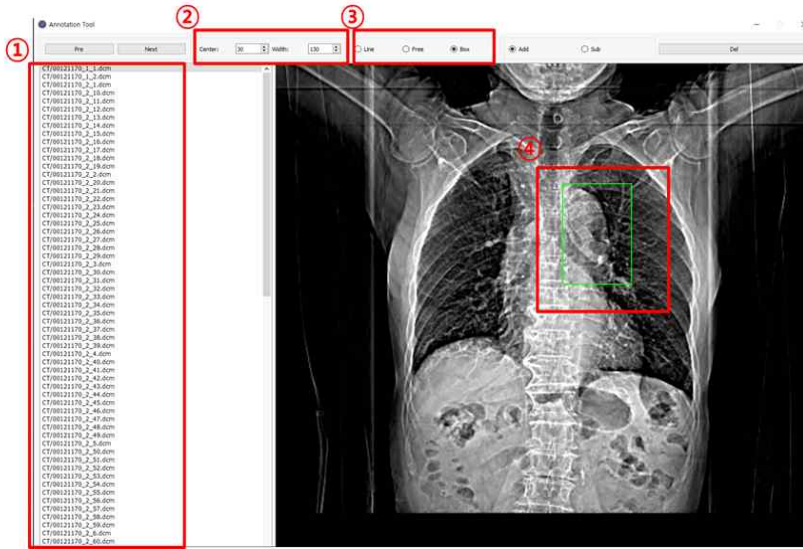
Binary 형태의 DICOM Output

JSON Output

```

[[{"id": 0, "option": {"tag": "isCopyCheck": false, "isSelected": "lesionA1", "annotationData": {"visible": true, "active": false, "color": "#f67081", "handle": [{"x": 471.685177, "y": 104.82496777027027, "highlight": true, "active": true, "line": 469.53692871621628, "y": 113.41695238484871}}, {"x": 113.41695238484871, "y": 113.41695238484871, "highlight": true, "active": true, "line": 534.88517738484848, "y": 143.50083344564564}}, {"x": 143.50083344564564, "y": 143.50083344564564, "highlight": true, "active": true, "line": 534.712043918919, "y": 129.69252833783784}}, {"x": 129.69252833783784, "y": 129.69252833783784, "highlight": true, "active": true, "line": 471.68517738484848, "y": 104.82496777027027, "highlight": true, "lines": [{"x": 469.53692871621628, "y": 113.41695238484848}], "invalidHandlePlacement": false, "label": "0", "canComplete": false, "copySoundingBox": [{"left": 469.53692871621628, "top": 104.82496777027027, "width": 85.1760767567562, "height": 366.86021659459464}], "area": "823.82288340394151", "toolSaveColor": "#f67081", "save": true, "status": "pass"}]]
                    
```


3) Annotation 절차



- ① 어노테이션 할 이미지 선택
(마우스 스크롤로 이동 가능)
- ② 필요 시 이미지 window level 조정
- ③ 어노테이션 종류 선택
(처음에 한번 선택 시 변경 전까지 선택이 유지됨)
- ④ 어노테이션 수행

2.5 검수

2.5.1 검수 절차

(1) X-ray 및 흉부 CT 데이터 어노테이션 및 검수 방법

- 1단계: 전문의가 1차 프리어노테이션 수행
- 2단계: 교육생이 2차 어노테이션 진행
- 3단계: (1차 검수) 파일 생성 여부 확인(어노테이션 된 데이터(Binary DICOM)와 메타데이터(JSON)파일이 올바르게 생성되었는지 파이썬 스크립트로 검수 진행
- 4단계: (2차 검수) 어노테이션 정확성 검토: 영상의학과 전문의에 의해 교차검증 진행

(2) PET CT 데이터 검수 방법

- 핵의학과,내과 전문의가 병변 부위 bounding box 표시

2.5.2 검수 조직도

종류	1차 검수	2차 검수
X-ray	조범상 (영상의학과 전문의)	김현숙 (영상의학과 전문의)
Chest CT	정미희 (영상의학과 전문의)	권진환 (영상의학과 전문의)
	백창규 (영상의학과 전문의)	
	조은미 (영상의학과 전문의)	
	신영경 (영상의학과 전문의)	
PET CT	이순성 (핵의학과 전문의)	이석모 (핵의학과 전문의)
	최수정 (핵의학과 전문의)	
	심혜경 (핵의학과 전문의)	

2.5.3 검수 기준

2.5.3.1 데이터 획득 단계(원본추출) :

- PACS에서 영상추출을 위한 환자 대상자선정 및 추출 리스트 검토 2인 이상 확인 (암종별 정보, 환자정보 누락은 없는지, 정확한 검사 종류,날짜,시간 등)
- PACS에서 익명화하여 추출한 영상 파일을 원내 DATA서버로 옮겨 데이터 적절성 교차확인

- (DICOM head 정보, 영상의 누락은 없는지, 매핑정보(new ID))
- EMR ID 와 NEW ID의 매핑정보 재확인 후 클라우드 서버에 업로드

2.5.3.2 2차 검수

- 교육생이 어노테이션을 시행하는 경우 90%의 정확도를 기준으로 90% 이상 일치 시 통과, 정확도 90% 미만 시 annotation 재수행
- 2차 검수 시 2명의 영상의학과 의사가 교차 검증 (의견이 다를 경우 최종 검수자가 병리 결과 및 EMR, PACS 데이터를 확인하여 최종 결정함)

2.6 활용

2.6.1 활용 모델

2.6.1.1 모델 학습

유형	모델 개발 계획	학습 분배량
X-ray	<ul style="list-style-type: none"> • 인공지능 모델 개발 목표 <ul style="list-style-type: none"> - 병변 유무 판단 - 병변 위치 표기 - 병변 양성/악성 진단 • AI 베이스 라인 모델 <ul style="list-style-type: none"> - 수집 데이터에서 성능 향상을 위한 모델 개선, 신규 모델 적용을 수행하며, 목표 성능을 상회하는 모델에 대한 소스코드 등을 공개함 	* 학습:시험:검증 비율은 환자수를 기준으로 8:1:1로 구성되며, 이미지 장수는 차이가 발생할 수 있음.
CT	<ul style="list-style-type: none"> •인공지능 모델 개발 목표 <ul style="list-style-type: none"> - 병변 유무 판단 - 병변 위치 표기 - 병변 양성/악성 진단 •AI 베이스 라인 모델 <ul style="list-style-type: none"> - 딥러닝 알고리즘을 적용한 이미지 분석을 통해 각종 암 병변 여부를 판단하는 딥러닝 모델들로 DeepLab, VGG, InceptionNet과 같은 모델을 사용 - 수집 데이터에서 성능 향상을 위한 모델 개선, 신규 모델 적용을 수행하며, 목표 성능을 상회하는 모델에 대한 소스코드 등을 공개함 	* 학습:시험:검증 비율은 환자수를 기준으로 8:1:1로 구성되며, 이미지 장수는 차이가 발생할 수 있음.
PET CT	<ul style="list-style-type: none"> • 인공지능 모델 개발 목표 <ul style="list-style-type: none"> - 딥러닝 알고리즘을 적용한 이미지 분석을 통해 각종 암 병변 여부를 판단하는 딥러닝 모델들로 DeepLab, VGG, InceptionNet과 같은 모델을 사용 - 수집 데이터에서 성능 향상을 위한 모델 개선, 신규 모델 적용을 수행하며, 목표 성능을 상회하는 모델에 대한 소스코드 등을 공개함 	* 학습:시험:검증 비율은 환자수를 기준으로 8:1:1로 구성되며, 이미지 장수는 차이가 발생할 수 있음.

2.6.1.2 서비스 활용 시나리오

과제명	데이터 유형	응용서비스
폐암 AI 학습용 이미지 데이터 구축	X-ray	<ul style="list-style-type: none"> ○ 병변 검출 및 탐지 ○ 병변의 종류(양성/악성)구분
	흉부 CT	<ul style="list-style-type: none"> ○ 병변 검출 및 탐지 ○ 병변의 종류(양성/악성) 구분
	PET CT	<ul style="list-style-type: none"> ○ 병변 검출 및 탐지 ○ 폐 결절과 임파선의 양성/악성 감별 ○ 악성 폐결절과 악성 임파선 검출

2.6.2 데이터 제공

2.6.2.1 K-ICT 빅데이터 센터를 통한 헬스케어 데이터 사용 신청 및 활용

- 사이트 내 “데이터 안전영역 예약”에서 데이터 사용 신청

