

인공지능 기술 육성 및 활성화를 위한 의료영상 공개사업 가이드 라인

Guidelines for Medical image data open
project for developing and advancing
artificial intelligence technology

(앞 표지)

	성명	소 속	직위	위원회 및 직위	가이드라인 버전
가이드라인 제안	황보을	국립암센터	연구원	-	
가이드라인 초안 작성자	이재동	국립암센터	연구원	-	TTAR-01.0001
	박현우	국립암센터	연구원	-	TTAR-01.0001/R1
	안선희	국립암센터	연구원	-	
	백경연	국립암센터	연구원	-	
	이예림	국립암센터	연구원	-	
	신현아	건양대병원	연구원	-	
전문가 또는 위원회	홍길동4	TTA		-	

본 문서에 대한 저작권은 과학기술정보통신부에 있으며, 사전 협의 없이 이 문서의 전체 또는 일부를 상업적 목적으로 복제 또는 배포해서는 안 됩니다.

본 가이드라인 발간 이전에 접수된 지식재산권 파악서 정보는 본 가이드라인의 '부록(지식재산권 파악서 정보)'에 명시하고 있습니다.

본 가이드라인과 관련하여 접수된 파악서 외의 지식재산권이 존재할 수 있습니다.

발행인 : 한국정보통신기술협회 회장

발행처 : 한국정보통신기술협회

13591, 경기도 성남시 분당구 분당로 47

Tel : 031-724-0114, Fax : 031-724-0109

발행일 : 20xx.xx

서 문

1 표준의 목적

본 표준은 공공 및 인간의 지능정보기술 개발 촉진과 지능정보산업 육성을 위한 의료 분야에 인공지능 학습용 데이터(지식베이스) 구축 및 이용환경을 제공하며, 이를 적용하기 위한 가이드라인을 제공하기 위하여 작성되었다.

2 주요 내용 요약

본 표준은 임상적 활용을 위해 모델을 공개하고 활용 방법에 대한 지침을 서술한다. 학습된 모델의 임상적 활용을 위한 모델을 공개하여 수집된 데이터에 관한 기술과 활용 방법에 대한 예시를 제공하고, 모델 강화를 위해 추가 데이터를 가공하기 위한 저작도구를 공개하여 시스템 환경 구성방안을 제시한다.

3 인용 문서와의 비교

- 해당사항 없음.

Preface

1 Purpose of Standard

The standard provides the environment for building and using artificial intelligence learning data (knowledge base) in the medical field for promotion of public and private intelligence information technology development and intelligent information industry development, and to provide guidelines for applying it .

2 Summary of Contents

The standard describes the model for clinical use and provides guidance on how to use it. It provides an example of how to describe and utilize the data collected through model disclosure for the clinical use of the learned model, and We present authoring tool to process additional data to suggest system environment configuration for model enhancement.

3 Relationship to Reference Standards

- None.

목 차

1 적용 범위 1

2 인용 표준 1

3 용어 정의 1

4 약어 2

5 (1 세부) 유방촬영영상 인공지능 학습 데이터 구축 3

 5.1 수집 데이터 상세명세서 3

 5.2 수집 데이터의 정제 및 증강 6

 5.3 어노테이션을 통한 데이터 가공 6

 5.4 데이터 검수 7

 5.5 인공지능 판독 시범 모델 개발 및 데이터 검증 8

6 (2 세부) 안저 영상 인공지능 학습 데이터 구축 11

 6.1 수집 데이터 상세명세서 11

 6.2 수집 데이터의 정제 및 증강 13

 6.3 어노테이션을 통한 데이터 가공 14

 6.4 데이터 검수 15

 6.5 인공지능 판독 시범 모델 개발 및 데이터 검증 17

부속서 A (자유 작성 부속서) 제목 x

부록 I (자유 작성 부록) 제목 x

부록 II-1 지식재산권 협약서 정보 x

 II-2 시험인증 관련 사항 x

 II-3 본 가이드라인의 연계(family) 문서 x

 II-4 참고 문헌 x

 II-5 영문가이드라인 해설서 x

 II-6 가이드라인의 이력 x

인공지능 기술 육성 및 활성화를 위한 의료영상 공개사업 가이드라인

(Guidelines for Medical image data open project for developing and advancing artificial intelligence technology)

1 적용 범위

암은 2011년 이후 국내에서 발생이 감소하는 추세를 보이고 있지만 유독 여성에게 발생하는 유방암만은 지속적으로 뚜렷한 증가하는 추세를 보인다. 유방암의 증가는 전 세계적인 현상으로 세계암통계에 의하면 연령표준 발생률에서 남녀를 모두 합쳐서도 유방암이 가장 많이 발생하는 암종이며, 이를 토대로 국내의 유방암의 증가는 앞으로도 지속될 것으로 판단된다.

건강보험심사평가원 통계자료에 따르면, 2017년 급성 부비동염으로 진료를 받은 인원은 약 411만 명으로, 3년 사이(2015~2017) 10만 명가량(2.9%) 증가한 것으로 나타난다. 비염과 증상이 비슷해서 오진으로 인한 치료제인 항생제 처방 여부가 달라지면, 치료시기를 놓쳐 만성 부비동염으로 진행될 가능성이 높다. 만성 부비동염은 완치가 힘들고 다양한 합병증을 동반하기 때문에, 급성 부비동염 초기일 때의 치료가 매우 중요하다.

본 표준에서 유방암과 부비동염에 대해 의료 지식베이스 구축 및 효용성 검증에 대해 서술하고, 인공지능 기업에서도 활용 할 수 있도록 두 질환에 대하여 의료 인공지능 판독 기술 개발에 공통적으로 활용 가능한 이미지, 영상 등 범용성이 높은 의료 지식베이스 구축·공개하고 이에 대한 의료 지식베이스의 표준 지침을 서술한다.

2 인용 표준

- 해당 사항 없음.

3 용어 정의

3.1 디지털 의료 영상 전송 장치 (DICOM)

- 미국 방사선 학회와 전기 공업회가 합동으로 설립한 ACR/NEMA위원회(1996년에 DICOM 위원회로 개칭)가 모체가 되어 의료 화상 전송을 중심으로 정한 규격. 현재는 데이터 보존 규격도 포함되어 있어 표준 규격이 되었다. 1993년에 일단 완성된 이 규격에 의해서 의료 화상 정보를 주고받을 수 있게 되었으며, 1996년에는 디지털 의료 영상 장치(DICOM) 위원회에서 규격을 더욱 강화하였다.

3.2 어노테이션(Annotation)

- 주석처럼 코드에 달아 해당 코드에 특수한 의미를 부여해주는 데이터의 전처리 과정을 뜻하며, 이미지 데이터의 경우 이미지 내 각종 사물을 알아보고 경계선을 구분짓는 것

3.3 비식별화

- 개인정보가 포함된 데이터를 활용할 때, 정보집합물(데이터 셋)에서 개인을 식별할 수 있는 요소를 전부 또는 일부 삭제하거나 대체하는 등의 방법
 - 가명처리(Pseudonymisation): 개인정보 중 주요 식별요소를 다른 값으로 대체
 - 총계처리 혹은 대체(Aggregation, Replacement) : 데이터의 총합 값을 보임으로서 개별 데이터의 값을 보이지 않도록 함
 - 데이터 값(가치) 삭제(Data Reduction): 데이터 공유, 개방 목적에 따라 데이터 셋에 구성된 값 중에 필요없는 값 또는 개인식별에 중요한 값을 삭제
 - 범주화(Data Suppression): 데이터의 값을 범주의 값으로 변환하여 명확한 값을 감춤
 - 데이터 마스킹(Data masking): 공개된 정보 등과 결합하여 개인을 식별하는데 기여할 확률이 높은 주요 개인식별자가 보이지 않도록 처리하여 개인을 식별하지 못하도록 함

3.4 관심영역 (ROI, Region Of Interest)

- 정지 영상의 내용 중 관심을 가지는 특정 영역 혹은 객체(인물, 건물, 동물, 건축물, 차 등의 객체 혹은 산, 바다, 잔디 등의 특정 영역)

3.5 병리 조직 검사

- 얇고 있는 인체 조직을 떼어내어 현미경 관찰을 통해 암 등의 질환명을 결정받는 검사

3.6 부비동

- 비강 (코 안쪽의 주먹만 한 크기의 빈 공간) 주변의 빈 공간으로, 4쌍 (상악동, 전두동, 사골동, 접형동)으로 구성
 - 부비동염에는 mucosal thickness, air-fluid, haziness 등이 존재

4 약어

DICOM	Digital Imaging and Communications in Medicine
WSI	Whole Slide Image
ROI	Region Of Interest

PACS	Picture Archiving Communication System
EMR	Electronic Medical Record
PNS	Paranasal sinus

5. (1 세부) 유방암 조직 이미지 시 데이터 구축



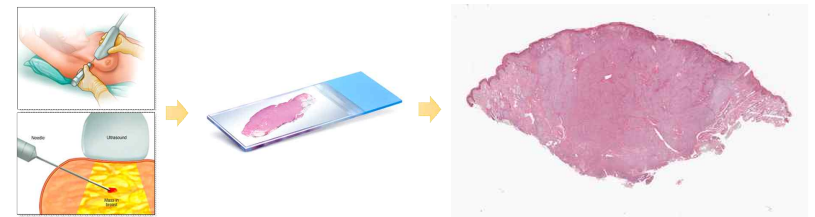
(그림 5-1) 유방암 조직 이미지 데이터 구축 절차

5.1 데이터 수집

- 국립암센터에서 조직검사를 통해 유방암 악성/양성으로 판단된 환자의 유방암 조직 슬라이드를 디지털 스캐너로 스캔하여 Whole Slide Image(WSI)를 수집한다.

5.1.1 WSI 정의

- 환자에게서 얻은 조직을 육안 검경 후 포르말데하이드에 고정, 파라핀 블록 제작, 헤마톡실린/에오신 염색 후 현미경용 유리 슬라이드 완성 후, 슬라이드 스캐너를 이용하여 베투얼 슬라이드를 제작한다. 이를 WSI 라고 칭한다.



(그림 5-2) Whole Slide Image

5.1.2 Whole Slide Image

- 만 19세 이상 성인 여성환자들을 대상으로 악성(Malignant) 3,000명, 양성(Benign)

1,000명데이터 수집

5.1.3 데이터 구조

- 유방촬영 영상의 판독후 악성, 양성, 정상으로 판명된 유방촬영 영상 데이터와 해당 환자의 임상정보를 결합하여 최종 데이터 구조는 표 5-3과 같음

<표 5-3> 최종 구축 데이터 구조

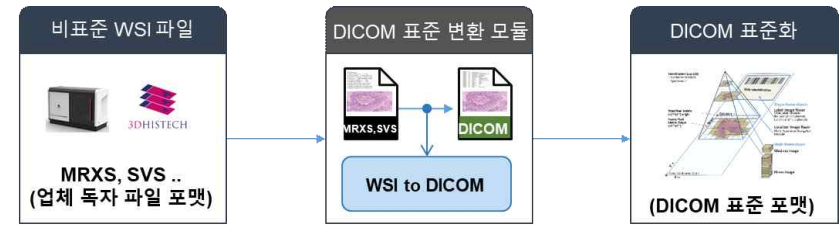
데이터 구조		구분	데이터타입	null	예시	Benign 포함	Malignant 포함
Image Malignant SN003456 volume_20x_1.png volume_20x_2.png ... volume_20x_25.png Benign SN005745 volume_20x_1.png volume_20x_2.png ... volume_20x_25.png Meta data metadata.xlsx metadata.xml	익명화 환자번호	익명화 환자ID	char	N	SN005745	Y	Y
	volumn_n(zoom ratio)_n(index).png 25장 이미지	연령대	char	N	1: 19, 2: 20-29, 3: 30-39, 4: 40-49, 5: 50-59, 6: 60-69, 7: 70-79, 8: 80-89, 9: 90 이상	Y	Y
		검사일시	date	N	1999 01 01.	Y	Y
		검사장비 제조사	char	N	3DHitech	Y	Y
		검사장비 모델명	char	N	P1000	Y	Y
		Class	char	N	Benign / Malignant	Y	Y
		종양 subtype	char	N	Invasive carcinoma	N	Y
		T-stage	char	N	1	N	Y
		N-stage	char	N	2	N	Y

데이터 구조 설명	
o 영상 데이터	
- 양성(benign), 악성(malignant)으로 구분	
- Slide ID는 익명화된 환자 번호 (예 : SN003456)	
- 각 환자당 20x 배율에 따른 512x512 슬라이드, 영상 png 파일 25개	
- 파일명 : volume_20x(배율 정보)_n(index).png	
o 메타 데이터	
- 모든 환자의 메타정보 xls/ xml 파일	
- 익명화 환자ID, 연령대, 검사일시, 검사장비 제조사, 모델명, Class(양/악성), 종양subtype, T,N-stage	

5.2 수집 데이터의 정제

- mrxs, svs와 같은 비표준 파일 포맷인 WSI파일을 표준파일 포맷인 DICOM 으로 변환(그림 5-3)
- 환자 식별이 가능한 정보가 들어있는 Label을 제거하여 비식별화를 진행

- 비식별화 후 DICOM이미지 데이터의 tag정보는 <표 5-2>와 같다.



(그림 5-3) 유방암 조직 이미지 변환 절차

<표 5-3> DICOM 포맷으로 변환된 파일의 Meta Data 구성과 비식별화된 데이터

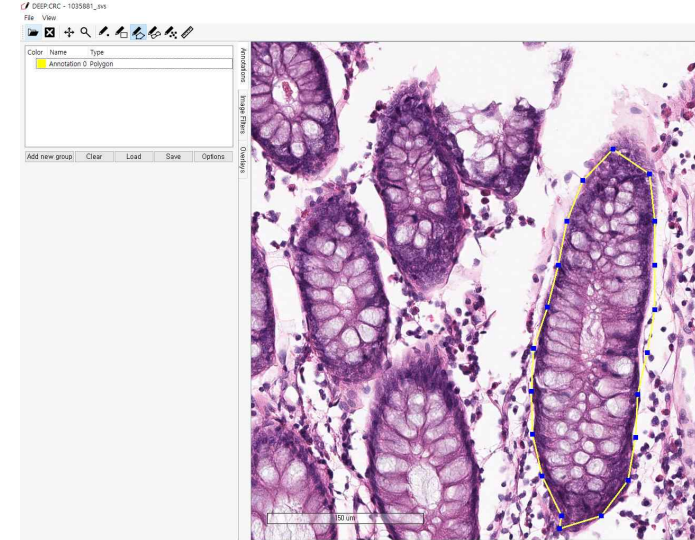
Tag (Group, Element)	VR	TAG Description	Value (예시)
(0008,0008)	CS	ImageType	DERIVED/PRIMARY/LABEL
(0008,0016)	UI	SOPClassUID	1.2.840.10008.5.1.4.1.1.77.1.6
(0008,0018)	UI	SOPInstanceUID	1.2.410.200010.12.1.4.378948042.15904.1579167489.685 (재생성)
(0008,0020)	DA	Study Date	20191126 (재생성 날짜)
(0008,0021)	DA	Series Date	20191126 (재생성 날짜)
(0008,0030)	DT	Study Time	163147 (재생성 시간)
(0008,0030)	DT	Series Time	163147 (재생성 시간)
(0008,0060)	CS	Modality	SM
(0008,0070)	LO	Manufacturer	XXXXXX, Inc.
(0008,1090)	LO	ManufacturerModelName	XXXXX Dimensions
(0010,0010)	PN	Patient's Name	공백 (익명화)
(0010,0020)	LO	PatientID	공백 (익명화)
(0010,0030)	DA	PatientBirthDate	공백 (익명화)
(0010,0040)	CS	Patient's Sex	F(female) or M(male) or O(Other)
(0020,000D)	UI	Study Instance UID	1.2.410.200010.12.1.2.378948042.4388.1563343439.461 (재생성)

(0020,000E)	UI	Series Instance UID	1.2.410.200010.12.1.3.3789 48042.15904.1579167489.6 86 (재생성)
(0020,0013)	IS	Instance Number	1
(0028,0002)	US	SamplesPerPixel	3
(0028,0004)	CS	Photometric Interpretation	YBR_FULL_422
(0028,0006)	US	Planar Configuration	0
(0028,0010)	US	Rows (행)	512
(0028,0011)	US	Columns (열)	512
(0028,0034)	IS	Pixel Aspect Ratio	1.0W1.0
(0028,0100)	US	BitsAllocated	8
(0028,0101)	US	BitsStored	8
(0028,0102)	US	HighBit	7
(0028,0103)	US	PixelRepresentation	0
(0028,2110)	CS	Lossy Image Compression	01
(0028,9110)	SQ	Pixel Measures Sequence	-
>(0028,0030)	DS	PixelSpacing	0.000243W0.000243
(0048,0105)	SQ	Optical Path Sequence Attribute	-
>(0022,0016)	SQ	Illumination Type Code Sequence Attribute	-
>>(0008,0100)	SH	Code Value	111741
>>(0008,0102)	SH	Coding Scheme Designator	DCM
>>(0008,0104)	LO	Code Meaning	Transmission illumination
>(0048,0106)	SH	Optical Path Identifier Attribute	Op_001

5.3 데이터 가공

5.3.1 어노테이션을 통한 데이터 가공

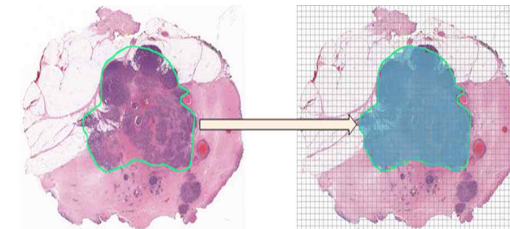
- 어노테이션 기본 UI는 Viewer, Annotation, ImageFilter, Overlays로 구성
- 유방암 조직 이미지 파일을 열고 병변부위 선택하여 어노테이션 수행
- 해당 어노테이션 정보는 .xml 파일로 저장 가능



(그림 5-4) 어노테이션 실행 이미지

5.3.2 타일링(Tiling, crop)을 통한 인공지능 학습데이터 생성

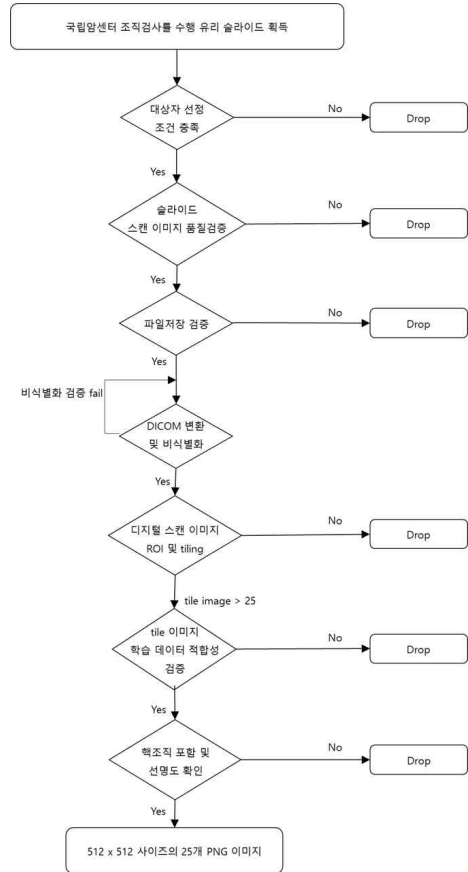
- 유방암 조직에서 어노테이션 실행 후 선택된 병변 이미지를 Tiling
- 타일단위 인공지능 학습 데이터 생성



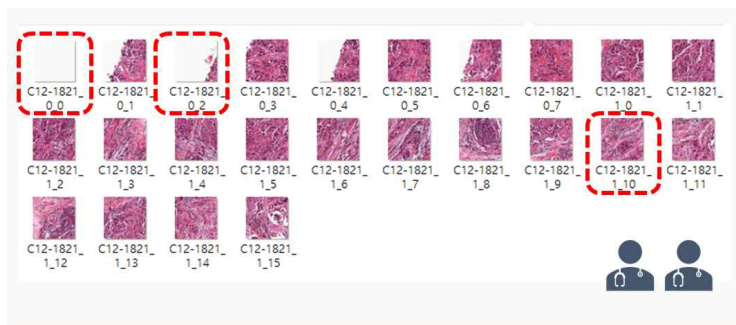
(그림 5-5) 인공지능학습 데이터를 위한 타일 이미지 생성

5.4 데이터 검수

- 의료전문가의 검증 및 학습 부적합 이미지를 제거하기 위해 그림 5-6과 같이 자체 검수 프로세스를 제정하였다.
- 데이터검수 결과 최종적으로 구축되는 데이터 이미지는 그림 5-7과 같다



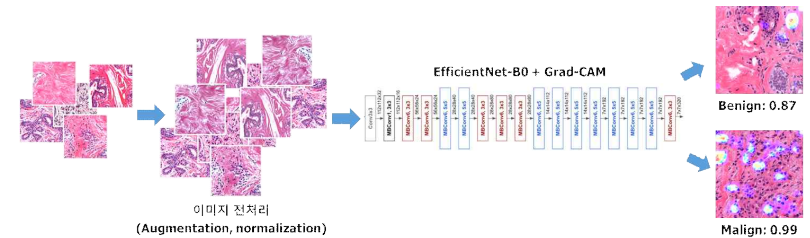
(그림 5-6) 데이터 검수 프로세스



(그림 5-7) 데이터 검수 프로세스 결과

5.5. 인공지능 판독 시범 모델 개발 및 데이터 검증

- 다양한 영상 분석에서 높은 성능을 보여주며, 영상 분석 딥러닝 알고리즘 중 가장 널리 사용되고 있는 CNN (convolutional neural network) 기반의 EfficientNet-B0을 사용하여 인공지능 판독 시범 모델(그림 5-8)을 개발하고 구축된 유방암 조직 이미지의 인공지능 학습데이터로 하용함으로써 데이터의 유효성 검증
- 이미지 전처리로 데이터 증강(random rotation, horizontal flip, vertical flip), Brightness adjustment 수행
- 학습결과 : 512 x 512 사이즈 조직 이미지의 양성, 악성 확률 및 병변 시각화



(그림 5-8) 유방암 조직 이미지 학습모델 구조와 결과 예시

- 유방암 조직 이미지의 총 데이터를 학습(training), 검증(validation), 테스트(test)로 나누어서 유방암 양성과 악성의 판독 학습모델을 개발
- 유방암 판독 학습모델을 검증하기 위하여 n-fold cross validation을 수행하며, 예시로 전체 데이터를 학습용 (80%), 테스트용 (20%)로 나누고, 학습용 데이터에서 n개의 데이터 셋으로 random split
- n-1개 set으로 모델을 학습하고, 1개의 set으로 검증하며 판독 모델의 검증을 위한 data flow는 아래 그림 5-9와 같음
- 인공지능 판독 모델의 성능을 다양한 척도를 기반으로 평가하며, 환자 단위 영상자료 기반의 민감도 (sensitivity), 특이도 (specificity), 정밀도 (precision), ROC curve에서의 AUC를 측정, 측정 결과는 표 5-7과 같음

Train dataset = 80%										Test dataset = 20%
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	
Validation	Train	Train	Train	Train	Train	Train	Train	Train	Train	Test
Train	Validation	Train	Train	Train	Train	Train	Train	Train	Train	Test
Train	Train	Validation	Train	Train	Train	Train	Train	Train	Train	Test
Train	Train	Train	Validation	Train	Train	Train	Train	Train	Train	Test
Train	Train	Train	Train	Validation	Train	Train	Train	Train	Train	Test
Train	Train	Train	Train	Train	Validation	Train	Train	Train	Train	Test
Train	Train	Train	Train	Train	Train	Validation	Train	Train	Train	Test
Train	Train	Train	Train	Train	Train	Train	Validation	Train	Train	Test
Train	Train	Train	Train	Train	Train	Train	Train	Validation	Train	Test
Train	Train	Train	Train	Train	Train	Train	Train	Train	Validation	Test

(그림 5-9) K-Fold Validation 예시

<표 5-7> 구축 데이터 유효성 검증결과

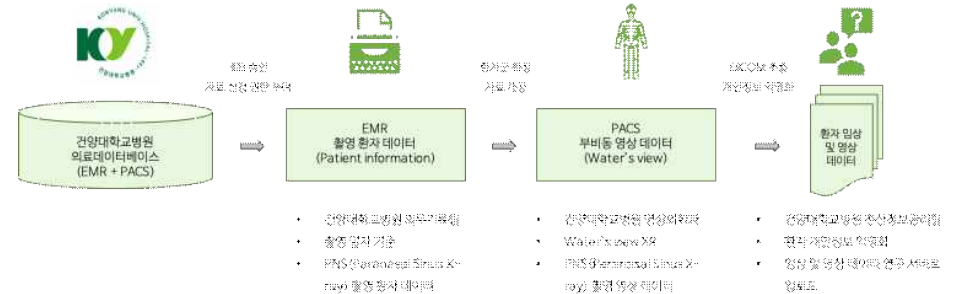
	Sensitivity	Specificity	Precision	AUROC
1	0.991	0.996	0.999	0.999
2	0.993	0.988	0.996	0.999
3	0.991	0.996	0.999	0.999
4	0.995	0.996	0.999	0.999
5	0.998	0.998	0.999	0.999
6	0.994	0.996	0.999	0.999
7	0.991	0.998	0.999	0.999
8	0.993	0.994	0.998	0.999
9	0.993	0.994	0.998	0.999
10	0.991	0.996	0.999	0.999
Avg.	0.993	0.995	0.998	0.999

6. (2 세부) 안저 영상 인공지능 학습 데이터 구축

6.1 수집 데이터 상세명세서

6.1.1 데이터 수집 경로

- 건양대학교 병원에서 운영하고 있는 의료데이터베이스(EMR, PACS)에서 데이터베이스 쿼리를 통하여 부비동영 진단을 위한 부비동 X-Ray 영상을 추출함
 - 각 의료기관 마다 의료 영상을 다루는 장비와 시스템이 상이하여 의료 영상 이미지 마다 서로 다른 resolution으로 생길 수 있는 bias를 방지하기 위해서 각 의료 기관별 고유의 PACS만을 사용
 - 개인정보 보호법 등의 이유로 개인 또는 타 기관에서 각 의료 기관의 의료데이터베이스에 직접적인 접근은 불가능하며, 제한된 인원만이 의료기관과 협의를 하고 IRB 승인절차 필요
- 데이터 수집 절차는 아래 그림 6-1과 같음



(그림 6-1) 데이터 수집 절차

- o IRB 승인 및 자료 신청
 - 건양대학교병원 EMR 접근을 위해 IRB 신청 후, 의무기록팀에 19세 이상 90세 이하 성인 환자 중 부비동 X-Ray 영상 촬영한 환자리스트 요청
- o 부비동 X-Ray 이미지 추출
 - 영상의학과에 환자리스트 기반으로 Waters view 촬영한 환자의 영상 데이터 요청
 - 여러 번 촬영한 환자의 경우 제일 처음 촬영본을 사용 (환자 중복 방지)
 - 전산 정보관리팀에서 폐쇄망 전용 USB를 통해 연구 서버로 업로드
- o 수집 영상 데이터의 비식별화
 - PACS system에서 원천 영상 데이터 추출 시 환자의 개인정보가 담겨있는 tag 정보는 익명화 (환자 번호, 이름, 나이, 생년월일) 하고, private tag 정보는 삭제
 - 난수 생성을 통해 8자리의 일련번호 생성 후 새로운 환자 번호 생성
 - 추출된 이미지 데이터의 파일명이 실제 환자 번호를 포함하기 때문에, 새로운

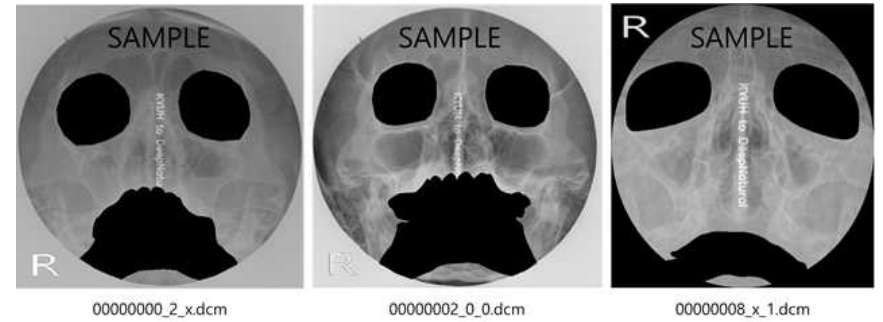
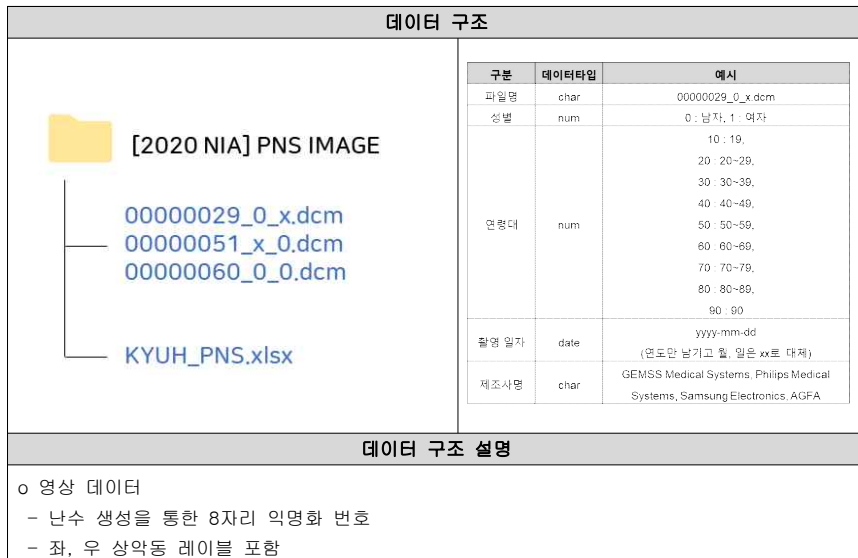
환자 번호로 매핑 후 파일명 변경

6.1.2 영상 데이터

- 의료영상데이터의 표준인 DICOM 영상을 기준으로 사용
 - PNS waters' view
- 데이터 판독 기준
 - mucosal thickness: 상악동 내 점막 비후가 2mm인 경우
 - air-fluid: 상악동 내 공기와 액체 음영이 가로로 명확히 보이는 경우
 - haziness: 상악동 전체에 혼탁화 음영이 확인되는 경우 (total opacification)
- 수집 목표: 총 8,000건 (상악동 레이블링)
 - 정상 7,000건, mucosal thickness 650건, air fluid 200건, haziness 150건
- 수집 분량의 적정성
 - 2019년 수행된 유사 연구에서 활용한 데이터 규모 및 비율을 비교해보았을 때, 데이터 수집 분량은 적절한 것으로 판단할 수 있음
 - 분당서울대병원 연구진; normal (81.8%), mucosal thickness (8.2%), air fluid (2.6%), haziness (3.9%); [Kim, Youngjune, et al. "Deep learning in diagnosis of maxillary sinusitis using conventional radiography." Investigative radiology 54.1 (2019): 7-15.]

6.1.3 데이터 구조

<표 6-1> 부비동 X-Ray 영상 데이터 구조도



(그림 6-2) 부비동 X-Ray 영상 데이터 예시

<표 6-2> 부비동 X-Ray 영상 DICOM tag 정보

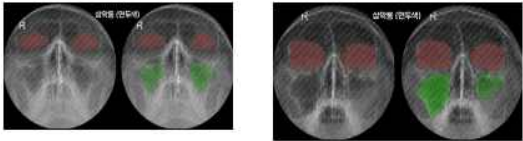
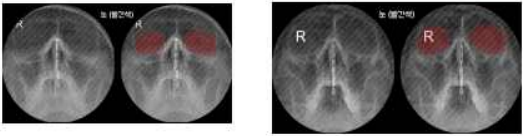
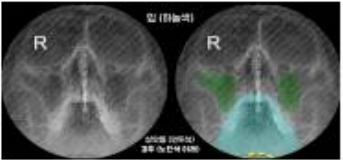
Tag	VR	TAG Description	Value (예시)	Null
(0008,0008)	CS	ImageType	DERIVED PRIMARY	Y
(0008,0016)	UI	SOPClassUID	1.2.840.10008.5.1.4.1.1.1.2	
(0008,0018)	UI	SOPInstanceUID	1.2.840.113681.2887049286.1	
(0008,0060)	CS	Modality	MG	
(0008,0070)	LO	Manufacturer	XXXXXX, Inc.	
(0008,1090)	LO	ManufacturerModelName	XXXXX Dimensions	
(0010,0020)	PN	PatientID	AN_20190813_171118_000000_0_0	
(0010,0030)	DA	PatientBirthDate	19000101	
(0018,1164)	DS	ImagerPixelSpacing	0.070000 0.070000	
(0018,5101)	CS	ViewPosition	CC	
(0020,0062)	CS	ImageLaterality	L	Y
(0028,0002)	US	SamplesPerPixel	1	
(0028,0004)	CS	PhotometricInterpretation	MONOCHROME2	
(0028,0010)	US	Rows	3328	
(0028,0011)	US	Columns	2560	
(0028,0030)	DS	PixelSpacing	0.065238 0.065238	
(0028,0100)	US	BitsAllocated	16	
(0028,0101)	US	BitsStored	12	
(0028,0102)	US	HighBit	11	

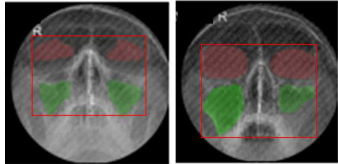
(0028,0103)	US	PixelRepresentation	0	
(0028,1050)	DS	WindowCenter	2047	
(0028,1051)	DS	WindowWidth	4096	
(0028,1052)	DS	RescaleIntercept	0	Y
(0028,1053)	DS	RescaleSlope	1	Y
(0028,1054)	LO	RescaleType	US	Y
(0040,0244)	DA	PerformedProcedureStepStartDate	2014-xx-xx	Y
(0040,0245)	TM	PerformedProcedureStepStartTime	0	Y
(7fe0, 0010)	OW	PixelData	Array of 2794166 elements	

6.2 수집 데이터의 정제

- 영상으로 인한 개인 재식별 방지를 위해, 환자의 눈과 입 부분을 마스킹하는 비식별화 처리를 진행함
- 모델 학습 시 부비동염 여부가 판단되는 상악동 부분을 집중할 수 있도록, 좌/우 상악동이 포함되도록 bounding box 가공을 진행함

<표 6-3> 부비동 X-Ray 영상 정제 조건

전제조건	개인을 식별할 수 없도록 상악동을 가리지 않는 선에서 최대한 눈과 입을 가리도록 작업 
눈 마스킹	눈 뼈 라인에 맞춰 외곽선을 polygon 포인트 설정 
입 마스킹	치아 및 보철을 포함한 입 영역 외곽을 polygon 포인트 설정 

전제조건	눈과 상악동 영역을 포함한 bounding box 가공 작업
상악동 영역	마스킹 된 눈과 상악동 영역을 포함하여 bounding box 설정 

6.3 데이터 가공

- 모바일 저작도구를 활용하여 부비동 X-Ray 영상에 대한 질환명 레이블링 작업을 의미함
- 이비인후과 전문의 2인과 영상의학과 전문의 1인 총 3인의 의료전문가가 판독을 진행함
- 좌/우 상악동을 각각 판독할 수 있도록, 프로젝트 target 여부를 '2'로 설정 후 질환에 대한 레이블링을 지정함 (0 : normal, 1 : mucosal thickness, 2 : air fluid, 3 : haziness)
- 하나의 프로젝트에 3인의 판독자를 할당하여, 영상별 교차 검증을 수행함



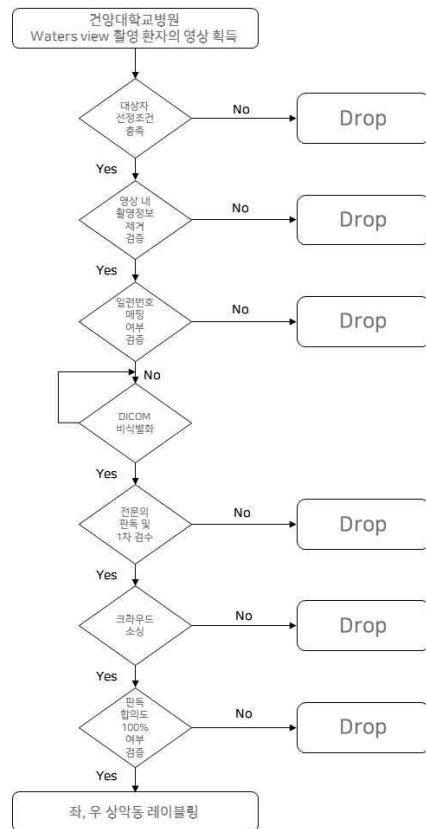
(그림 6-3) 저작도구 화면 예시

<표 6-4> 저작대구 프로젝트 정보 예시

프로젝트명	영상 개수	라벨링 개수	라벨 구성	판독자
PNS_2020	600	1,200	0, 1, 2, 3, 9	Dr. 1
				Dr. 2
				Dr. 3

6.4 데이터 검수

- 자체 검수 프로세스 기준 마련을 통해, 데이터 구축 전 과정에 대한 검수를 수행
- 저작대구를 활용하여 영상에 대한 레이블링 작업과 검수 작업을 동시에 진행
- 3인의 의료 전문가의 100% 합치된 영상만을 최종 데이터셋으로 활용함
- 판독 현황과 판독자별 진행률을 실시간으로 모니터링할 수 있으며, 교차 검증 결과 확인을 통한 데이터셋 구축 현황 파악 가능



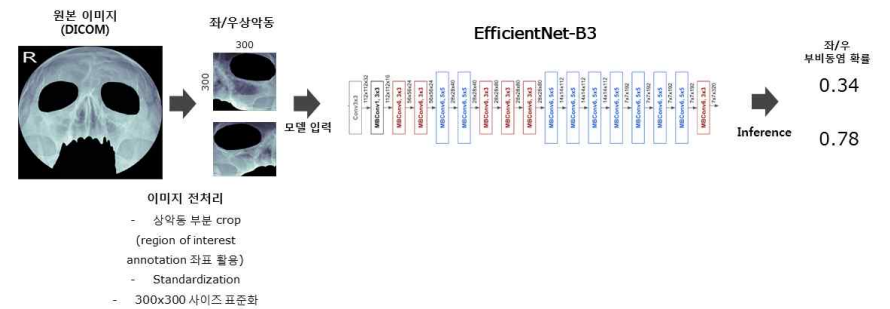
(그림 6-4) 자체 검수 프로세스 순서도

<표 6-5> 교차검증 결과 예시

Project	Image	Target	Consensus	Doc 1	Doc 2	Doc 3
PNS2015	****	0	O	Normal	Normal	Normal
PNS2015	****	1	O	Normal	Normal	Normal
PNS2015	****	0	X	Normal	Normal	Air fluid
PNS2015	****	1	O	Normal	Normal	Normal

6.5. 인공지능 판독 시범 모델 개발 및 데이터 검증

- 목적 : 구축된 인공지능 학습용 부비동 X-Ray 영상 데이터를 활용하여 부비동영 진단 모델 개발
- 영상 전처리:
 - 1) 상악동 영역 bounding box 좌표를 활용하여 crop한 뒤, 좌/우 상악동으로 분리하여 영상의 배경 등 모델 입력에 불필요한 부분을 줄여 계산량 절약
 - 2) Standardization (z-transformation) : 모델 학습이 쉽도록 영상의 intensity를 평균 0, 분산 1이 되도록 스케일링
 - 3) 300*300 사이즈로 변환하여 모든 영상의 사이즈를 통일
- 유효성 평가 : 정상과 부비동영 판독 시범모델의 검증을 위해 10-fold cross validation을 수행
- 판독 척도 : 민감도 (sensitivity), 특이도 (specificity), 정밀도 (Precision), ROC (Receiver Operating Characteristics) curve의 AUC (Area Under the Curve)를 측정



(그림 6-5) 부비동 X-Ray 영상 학습 모델 구조도

<표 6-6> 학습모델 목표 달성 지표

구 분	목표	달성도	목표대비 달성도
sensitivity	≥ 0.85	0.92	108%
specificity	≥ 0.85	0.99	116%
Precision	≥ 0.85	0.93	109%
AUROC	≥ 0.87	0.99	114%

A.1 OOOO (스타일 적용-대항목/소항목)

...

부 록 I (스타일 적용-본문상단제목)

(본 부록은 가이드라인을 보충하기 위한 내용으로 가이드라인의 일부는 아님)

제목 (굵림, 15포인트, 굵게, 가운데정렬)

1.1 0000 (스타일 적용-대항목/소항목)

000 0000 ……

부 록 II-1 (스타일 적용-본문상단제목)

(본 부록은 가이드라인을 보충하기 위한 내용으로 가이드라인의 일부는 아님)

지식재산권 요약서 정보 (굵림, 15포인트, 굵게, 가운데정렬)

(※ 본 가이드라인 발간 이전에 접수된 지식재산권 요약서가 있는 경우에 작성하며, 해당 사항이 없는 경우, 각 항목별로 '해당 사항 없음'으로 기재하고, 본 양식을 삭제하지 않음)

II-1.1 지식재산권 요약서(1) (스타일 적용-대항목/소항목)

- 발명의 명칭
- 권리자의 성명
- 등록(출원) 번호
- 등록(출원) 연월일
- 실시조건
- 요약서 접수일

II-1.2 지식재산권 요약서(2) (스타일 적용-대항목/소항목)

- 발명의 명칭
- 권리자의 성명
- 등록(출원) 번호
- 등록(출원) 연월일
- 실시조건
- 요약서 접수일

부 록 II-2 (스타일 적용-본문상단제목)

(본 부록은 가이드라인을 보충하기 위한 내용으로 가이드라인의 일부는 아님)

시험인증 관련 사항 (글꼴, 15포인트, 굵게, 가운데정렬)

(※ 시험인증 관련 사항이 있는 경우에 작성하며, 해당 사항이 없는 경우, 각 항목별로 '해당 사항 없음'으로 기재하고, 본 양식을 삭제하지 않음)

II-2.1 시험인증 대상 여부 (스타일 적용-대항목/소항목)

(※ 본 가이드라인이 법정강제(의무)인증, 법정임의인증, 민간임의인증 등의 대상이 되는 지를 구체적으로 기술함)

II-2.2 시험표준 제정 현황

(※ 본 가이드라인의 내용에 시험방법이 포함되어 있는 경우, 해당 항목 기재)

(※ 본 가이드라인과 관련하여 시험방법 표준이 별도로 제정되어 있는 경우 해당 시험표준번호와 표준명 기재)

부 록 II-3 (스타일 적용-본문상단제목)

(본 부록은 가이드라인을 보충하기 위한 내용으로 가이드라인의 일부는 아님)

본 가이드라인의 연계(family) 표준 (글꼴, 15포인트, 굵게, 가운데정렬)

(※ 본 가이드라인을 이용함에 있어서 같이 이용 또는 참고해야하는 표준 또는 문서가 있는 경우 작성하며, 해당 사항이 없는 경우, '해당 사항 없음'으로 기재하고, 본 양식을 삭제하지 않음)

II-3.1 ... (스타일 적용-대항목/소항목)

.....

(※ 표준(또는 기술보고서) 열거, 관계 서술, 표준(또는 문서)별 역할, 표준(또는 문서)별 구조도 등 자유 양식으로 작성)

부 록 II-4 (스타일 적용-본문상단제목)

(본 부록은 가이드라인을 보충하기 위한 내용으로 가이드라인의 일부는 아님)

참고 문헌 (굵림, 15포인트, 굵게, 가운데정렬)

(*가이드라인을 기술하기 위해 참고한 관련 자료에 대해 작성하며, 해당 사항이 없는 경우, '해당 사항 없음'으로 기재하고, 본 양식을 삭제하지 않음)

출처 및 근거 등의 정보를 나타내기 위한 표준 또는 문서와 수치, 계산식, 그림 등을 인용한 표준 또는 문서에 대해 기재한다.

- [1]
- [2]

예)

[3] ITU-T X.509, "Information technology - Open Syms Interconnection - The Directory : Authentication Framework", 2000.

[4] 저자, 제목, 출판사, 발행연도
.....

※ 상기 기재된 참고 문헌의 발간일이 기재된 경우, 해당 표준(문서)의 해당 버전에 대해서만 유효하며, 연도를 표시하지 않은 경우에는 해당 표준(권고)의 최신 버전을 따름

※ 단순한 용어정의, 계산식, 그림 등을 참조하는 경우는 본 항목에 해당 표준(권고) 등의 정보를 기재함

부 록 II-5 (스타일 적용-본문상단제목)

(본 부록은 가이드라인을 보충하기 위한 내용으로 가이드라인의 일부는 아님)

영문가이드라인 해설서 (굵림, 15포인트, 굵게, 가운데정렬)

(*영문가이드라인에 한해 목차 대항목별 주요 내용 요약 작성하며, 영문가이드라인이 아닌 경우, '해당 사항 없음'으로 기재하고, 본 양식을 삭제하지 않음)

II-5.1 (본문 내용 서술 항목)000000
.....

II-5.2 (본문 내용 서술 항목)0000000000
.....

부속서 A 000000
.....

부 록 II-6 (스타일 적용-본문상단제록)

(본 부록은 가이드라인을 보충하기 위한 내용으로 가이드라인의 일부는 아님)

가이드라인의 이력 (글꼴: 15포인트, 굵게, 가운데정렬)

판수	발생일	가이드라인번호	내용	담당 조직
제1판	2004.01.01	제정 x.xx-xx.xxxx	-	xx분야 작업반 (AAA)
오류정정	2014.06.01	오류정정 x.xx-xx.xxxx-Cor1	(*오류정정 내용 작성)	xx분야 작업반 (AAA)
오류정정	2014.10.01	오류정정 x.xx-xx.xxxx-Cor2	(*오류정정 내용 작성)	xx분야 작업반 (AAA)
제2판	2015.12.01	개정 x.xx-xx.xxxx/R1	(*1판 대비 개정 내용 작성)	xx분야 작업반 (AAA)

부 록 III-1 (스타일 적용-본문상단제록)

(본 부록은 가이드라인을 보충하기 위한 내용으로 가이드라인의 일부는 아님)

가이드라인의 예시 및 추가 설명 (글꼴: 15포인트, 굵게, 가운데정렬)