Medical Imaging & Intelligent Reality Lab.

www.mi2rl.co

Data Curation

2021.10.14

Jungyu Lee

Convergence Medicine/Radiology,
University of Ulsan College of Medicine
Asan Medical Center

울산대학교 UNIVERSITY OF ULSAN 서울아산병원 Asan Medical Center

00

Contents

MI2RL

01

Data description

20130109	CT	AMC PAC	UNKNOW	CT, Dynan	Pre 3.0 B30f		UNKNOW	15
20130109	CT	AMC PAC	UNKNOW	CT, Dynan	Artery 3.0 B30f		UNKNOW	15
20130109	CT	AMC PAC	UNKNOW	CT, Dynan	75sec delay 3.0 B30f		UNKNOW	15
20130109	CT	AMC PAC	UNKNOW	CT, Dynan	Pre 5.0 MPR		UNKNOW	15
20130109	CT	AMC PAC	UNKNOW	CT, Dynan	Artery 5.0 MPR		UNKNOW	15
20130109	CT	AMC PAC	UNKNOW	CT, Dynan	75sec delay 5.0 MPR cor		UNKNOW	15
20130109	CT	AMC PAC	UNKNOW	CT, Dynan	Processed Images		UNKNOW	15
20130109	CT	AMC PAC	UNKNOW	CT, Dynan	Processed Images		UNKNOW	15

20130723	CT	AMC PAC	UNKNOW	CT, Dynan	PRE 3.0 B30f		UNKNOW	1630
20130723	CT	AMC PAC	UNKNOW	CT, Dynan	PORTAL 3.0 B30f		UNKNOW	1630
20130723	CT	AMC PAC	UNKNOW	CT, Dynan	Artery 3.0 B30f RR		UNKNOW	1630
20130723	CT	AMC PAC	UNKNOW	CT, Dynan	PRE 5.0 SPO		UNKNOW	1630
20130723	CT	AMC PAC	UNKNOW	CT, Dynan	Artery 5.0 SPO		UNKNOW	1630
20130723	CT	AMC PAC	UNKNOW	CT, Dynan	PORTAL 5.0 SPO		UNKNOW	1630

20120226	CT	AMC PAC	UNKNOW	CT, Dynan	ARTERY 3.0 B30f		UNKNOW	4759
20120226	CT	AMC PAC	UNKNOW	CT, Dynan	PORTAL 3.0 B30f		UNKNOW	4759
20120226	CT	AMC PAC	UNKNOW	CT, Dynan	ARTERY 3.0 SPO sag		UNKNOW	4759
20120226	CT	AMC PAC	UNKNOW	CT, Dynan	PORTAL 3.0 SPO		UNKNOW	4759
20120226	CT	AMC PAC	UNKNOW	CT, Dynan	PORTAL 3.0 SPO		UNKNOW	4759

1) Data format

- Liver CT Image : Dicom file
- Folder number(환자 분류) : 34275(1번 ~ 34275번)
- Folder 별(환자 분류 별) 총 CT IMAGE folder 수 : 3
  : 각 폴더별 개수 상이

2) Excel 파일(.csv)로 정리

- 형식 : DCM Meta data / 폴더 경로 / 폴더 당 image 장수(Depth)

StudyDate	Modality (	Manufactu	Institution	StudyDesc	SeriesDescription (0008,103E)	PatientNar	PatientID (	PatientBirt	PatientSex	PatientAge	BodyPartE	SliceThickr	ImagePosi	ImageOrie	Rows (002	Columns (	PixelSpaci	BitsAllocat	BitsStored	RescaleInt	RescaleSlc	RescaleTyp	path	depth
20140224	CT	AMC PACS	UNKNOWI	CT, Dynamic Liver		UNKNOWI	1	0	M	063Y	Abdomen	5	[-174.900,	[1.000000,	512	512	[0.726562,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00001/21	36
20140224	CT	AMC PACS	UNKNOWI	CT, Dynam	POST	UNKNOWI	1	0	M	063Y	Abdomen	5	[-174.900,	[1.000000,	512	512	[0.726562,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00001/31	36
20140224	CT	AMC PACS	UNKNOWI	CT, Dynam	POST	UNKNOWI	1	0	M	063Y	Abdomen	5	[-174.900,	[1.000000,	512	512	[0.726562,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00001/32	56
20140224	CT	AMC PACS	UNKNOWI	CT, Dynam	POST	UNKNOWI	1	0	M	063Y	Abdomen	5	[-174.900,	[1.000000,	512	512	[0.726562,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00001/33	36
20140724	CT	AMC PACS	UNKNOWI	CT, Dynam	PRE	UNKNOWI	2	0	M	063Y	Abdomen	5	[-166.800,	[1.000000,	512	512	[0.681641,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00002/21	39
20140724	CT	AMC PACS	UNKNOWI	CT, Dynam	POST	UNKNOWI	2	0	M	063Y	Abdomen	5	[-166.800,	[1.000000,	512	512	[0.681641,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00002/31	39
20140724	CT	AMC PACS	UNKNOWI	CT, Dynam	POST	UNKNOWI	2	0	M	063Y	Abdomen	5	[-166.800,	[1.000000,	512	512	[0.681641,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00002/32	61
20140724	CT	AMC PACS	UNKNOWI	CT, Dynam	POST	UNKNOWI	2	0	M	063Y	Abdomen	5	[-166.800,	[1.000000,	512	512	[0.681641,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00002/33	39
20141008	CT	AMC PACS	UNKNOWI	CT, Dynam	Pre 5.0 B30f	UNKNOWI	3	0	M	064Y	Abdomen	5	[-187.6386	[1, 0, 0, 0,	512	512	[0.7226562	16	12	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00003/4	45
20141008	CT	AMC PACS	UNKNOWI	CT, Dynam	Artery 5.0 B30f	UNKNOWI	3	0	M	064Y	Abdomen	5	[-186.6386	[1, 0, 0, 0,	512	512	[0.7226562	16	12	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00003/5	45
20141008	CT	AMC PACS	UNKNOWI	CT, Dynam	Portal 5.0 B30f	UNKNOWI	3	0	M	064Y	Abdomen	5	[-186.1396	[1, 0, 0, 0,	512	512	[0.7207031	16	12	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00003/6	64
20141008	CT	AMC PACS	UNKNOWI	CT, Dynam	Delay 5.0 B30f	UNKNOWI	3	0	M	064Y	Abdomen	5	[-186.6386	[1, 0, 0, 0,	512	512	[0.7226562	16	12	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00003/7	45
20150112	CT	AMC PACS	UNKNOWI	CT, Dynam	PRE	UNKNOWI	4	0	M	064Y	Abdomen	5	[-180.200,	[1.000000,	512	512	[0.712891,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00004/21	43
20150112	CT	AMC PACS	UNKNOWI	CT, Dynam	POST	UNKNOWI	4	0	M	064Y	Abdomen	5	[-180.200,	[1.000000,	512	512	[0.712891,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00004/31	35
20150112	CT	AMC PACS	UNKNOWI	CT, Dynam	POST	UNKNOWI	4	0	M	064Y	Abdomen	5	[-180.200,	[1.000000,	512	512	[0.712891,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00004/32	57
20150112	CT	AMC PACS	UNKNOWI	CT, Dynam	POST	UNKNOWI	4	0	M	064Y	Abdomen	5	[-180.200,	[1.000000,	512	512	[0.712891,	16	16	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00004/33	35
20150721	CT	AMC PACS	UNKNOWI	CT, Dynam	Pre 5.0 B30f	UNKNOWI	5	0	M	064Y	Abdomen	5	[-192.6386	[1, 0, 0, 0,	512	512	[0.7226562	16	12	-1024	1	HU	//192.168.45.125/Liver_LFT/00001-05000/00005/3	45

02

Curation methods

300914

1) Depth < 10

199176

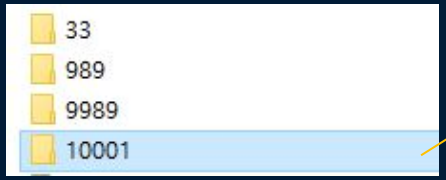
2) in MPR and Coronal

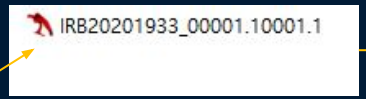
3) Group

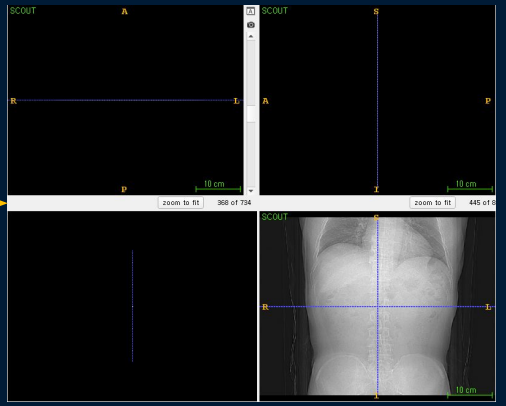
4) in SPO

5) Result

1) **CT Image 수 (Depth) < 10** 인 경우 제외
: Processed Image, SCOUT의 경우에 해당







전체 300914개에서 101738개의 폴더 제외.
199176 개 남음.

02

Curation methods

300914

1) Depth < 10

199176

2) in MPR and Coronal

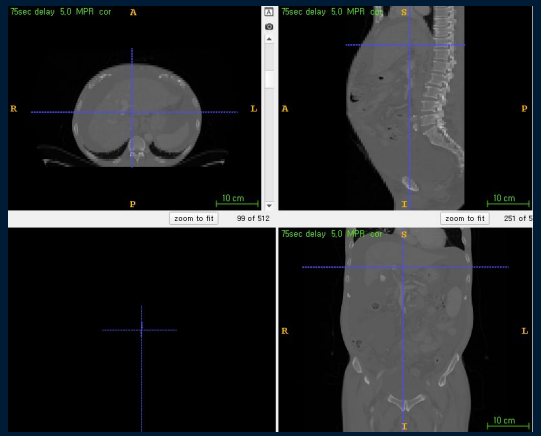
137710

3) Group

4) in SPO

5) Result

2) **Series Description in MPR, Coronal** 의 경우 제외



[15번 환자 9번 폴더]

[1531번 환자 17번 폴더]

전체 199176개에서 61466개의 폴더 제외.
137710 개 남음.

Curation methods

300914

1) Depth < 10

199176

2) in MPR and Coronal

137710

3) Group

137710

4) in SPO

5) Result

3) **Pre, Artery, Portal, Delay Group**

*4 phase 에 해당하지 않는 Folder(46017)는
완전히 다른 Group으로 분류

20150112	CT	AMC PAC	UNKNOW	CT, Dynan	PRE	UNKNOW	4
20150112	CT	AMC PAC	UNKNOW	CT, Dynan	POST	UNKNOW	4
20150112	CT	AMC PAC	UNKNOW	CT, Dynan	POST	UNKNOW	4
20150112	CT	AMC PAC	UNKNOW	CT, Dynan	POST	UNKNOW	4

Pre	Artery	Portal	Delay	Numder of Folder
1	1	1	0	11104
1	1	1	1	10277
1	0	0	0	6290
2	1	1	0	802
1	1	0	1	356
1	0	1	1	317
1	1	0	0	311
2	2	2	0	251
1	2	1	1	136
1	1	2	0	85
0	1	1	1	80
0	1	1	0	62
2	1	2	0	61
1	2	1	0	57
2	2	1	0	53

52

“53(0~52)가지 경우 존재”

- **Group 기준**

a) Number of Folder > 50
: 30317(CT Image Folder : 89179)

b) Number of Folder < 50
b-1) Pre or Artery
: 162(CT Image Folder : 710)
b-2) Portal or Delay
: 43(CT Image Folder : 205)
b-3) The rest
: 83(CT Image Folder : 440)

c) 예외 경우 존재

Curation methods

300914

1) Depth < 10

199176

2) in MPR and Coronal

137710

3) Group

137710

4) in SPO

96363

5) Result

- **Group Results & SPO Filtering**

" 1110, 1111, 1011, 1001 등 each 4 phase Group 제외"

Pre	Artery	Portal	Delay
1	1	1	0
1	1	1	1
1	0	0	0

1. df_Pr2_Ar_Po:
   Delay는 존재하지 않고 Pre만 2개 존재하는 경우. (3208 row, 802)
   → 64 row, 16명 남음

2. df_Pr2_Ar2_Po2:
   Delay는 존재하지 않고 Pre, Artery, Portal 이 각 2개씩 존재하는 경우. (1506 row, 251)
   → 모든 환자가 정리되었음

3. df_Pr_Ar2_Po_De:
   각각 1개 존재하면서 Arter만 2개 존재하는 경우.(680 row, 136)
   → 650 row, 130명 남음

4. df_Pr_Ar_Po2:
   Delay는 존재하지 않고 Portal 만 2개 존재하는 경우.(340 row, 85)
   → 128 row, 32명 남음

5. df_Pr2_Ar_Po2 :
   Delay는 존재하지 않고 Pre, Portal이 두개씩 존재하는 경우(305 row, 61)
   → 모든 환자가 정리 되었음

Curation methods

- **Group Results & SPO Filtering**

6. df_Pr_Ar2_Po:
   Delay는 존재하지 않고 Arter 만 2개 존재하는 경우.(228 row, 57)
   → 28 row, 7명 남음

7. df_Pr2_Ar2_Po:
   Delay는 존재하지 않고 Pre와 Arter 가 2개 존재하는 경우. (265 row, 53)
   → 모든 환자가 정리되었음.

8. df_Pr_or_Ar_up:
   Pre 또는 Arter 가 2개 이상 존재하는 경우.(710 row, 162)
   → 538 row, 116명 남음

9. df_Po_or_De_up:
   Portal 또는 Delay가 2개 이상이 있는 경우.(205 row, 43)
   → 126 row, 27명 남음

10. df_any_up :
    Pre, Arter, Portal, Delay 중 어느 것이라도 2개 이상 있는 경우. (40 row, 83)
    → 14 row, 2명의 환자 남음

02

Curation methods

300914

1) Depth < 10

199176

2) in MPR and Coronal

137710

3) Group

137710

4) in SPO

96363

5) Result

- Group Results

Pre	Artery	Portal	Delay	적용 전	SPO 제외이후
2→1	1	1	0	3208	64
2→1	2→1	2→1	0	1506	0
1	2→1	1	1	680	650
1	1	2→1	0	340	128
2→1	1	2→1	0	305	0
1	2→1	1	0	228	28
2→1	2→1	1	0	265	0
>2→1	>2→1	1	1	710	538
1	1	>2→1	>2→1	205	126

Curation methods

300914

1) Depth < 10

199176

2) in MPR and Coronal

137710

3) Group

137710

4) in SPO

96363

5) Result

- **Group Results 예외**

11. df_du :
한개의 row의 Series Description 값이 “ARTERY, PORTAL, DELAY” 인 경우.(69 row, 23)

StudyDate (0008,0020)	Modality (0008,0060)	Manufacturer (0008,0070)	InstitutionName (0008,0080)	StudyDescription (0008,1030)	SeriesDescription (0008,103E)
20200821	CT	AMC PACS	UNKNOWN	CT,Liver dynamic (with enhance) (No oral water)	ARTERY, PORTAL, DELAY
20200821	CT	AMC PACS	UNKNOWN	CT,Liver dynamic (with enhance) (No oral water)	ARTERY, PORTAL, DELAY

12. df_post:
pre만 해당하는 환자의 post(8658 row, 3370)

20150112	CT	AMC PAC	UNKNOW	CT, Dynan	PRE	UNKNOW	4
20150112	CT	AMC PAC	UNKNOW	CT, Dynan	POST	UNKNOW	4
20150112	CT	AMC PAC	UNKNOW	CT, Dynan	POST	UNKNOW	4
20150112	CT	AMC PAC	UNKNOW	CT, Dynan	POST	UNKNOW	4

Question

300914

1) Depth < 10

199176

2) in MPR and Coronal

137710

3) Group

137710

4) in SPO

96363

5) Result

1) df_any_up : 하나라도 2개 이상인 항목이 있는 경우
   - 12442
   a) Pre vs Pre **70 kev** folder number : (21, 41)
   b) Artery **140 kvp** vs Artery **70 kev (△)** (101, 61)
   c) Portal **depth 57(△)** vs Portal **depth 35** (141, 181)

   - 21401
   a) 모두 동일한데 중복인 경우 (4, 10005)

Question

300914

1) Depth < 10

199176

2) in MPR and Coronal

137710

3) Group

137710

4) in SPO

96363

5) Result

2) df_Pr_Ar2_Po2 : Delay가 없고, Artery, Portal이 2개인 경우
   - 33435

   a) Portal vs Portal **re** (6, 7)

3) df_du : 한개의 row의 Series Description 값이 “ARTERY, PORTAL, DELAY” 인 경우
   - 4038
   - 7956 등.

4) df_post : post가 여러 폴더 존재 depth에서만 차이
   - 162

이후 추가 작업

1) 중복 된 파일 kev, depth로 제거 : kev의 경우 100~140이 정상적인데 70으로 되어있는 잘못된 ct 모두 제거
2) post가 3개 존재하는데 이전 폴더에 series description이 비어있는 경우는 해당 폴더가 pre인 것을 확인 4 phase로 저장
3) post가 중복으로 되어있는 파일 : pre가 있고 3개가 있으면 artery, portal, delay로 설정. 2개, 1개, 4개 등의 경우 Not used로 분리
4) post가 중복으로 있는데 pre도 없는 경우 불완전하게 시간순으로 구분하여 interminate에 따로 분리.
5) post가 4개인 1명의 환자 각각 4 Phase로 추가

Curation results

- **Final results**

1) 4 Phase (전체 기준 : 93778 row, 29969(29445+523+1))

Pre	Artery	Portal	Delay
29822	26012	25661	14379

2) lab 기준 4546

300914

1) Depth < 10

199176

2) in MPR and Coronal

137710

3) Group

137710

4) in SPO

96363

5) 이후 추가 작업

95874

Collaborators

Radiology

Joon Beom Seo, SangMin Lee[A,B], Dong Hyun, Yang, Hyung Jin Won, Ho Sung Kim, Seung Chai Jung, Ji Eun Park, So Jung Lee, Jeong Hyun Lee, Gilsun Hong

Pathology

Hyunjeong Go, Gyuheon Choi, Gyungyub Gong, Dong Eun Song

Cardiology

Jaekwan Song, Jongmin Song, Young-Hak Kim

Anesthesiology

Sung-Hoon Kim, Eun Ho Lee

Neurology

Dong-Wha Kang, Chongsik Lee, Jaehong Lee, Sangbeom Jun, Misun Kwon, Beomjun Kim

Surgery

Beom Seok Ko, JongHun Jeong, Songchuk Kim, Tae-Yon Sung

Internal Medicine

Jeongsik Byeon, Kang Mo Kim

Emergency Medicine

Dong-Woo Seo