

# HeLP Challenge

## Breast cancer classification on frozen pathology - engineering advices -

16<sup>th</sup> Jan. 2019  
Young-Gon Kim

University of Ulsan College of Medicine,  
Asan Medical Center



울산대학교  
UNIVERSITY OF ULSAN



ASAN  
Medical Center

# Contents

---

- Data set
- Library
- Measure
- Examples



# Data set

- Class distribution

	Positive class	Negative class
Training	103	54
Validation	19	21
Test	??	??

# Data set

---

- Image
  - Format: ".mrxs"
  - Resolution: 93,952 x 132,352 pixels  
(3개만 58,812 x 125,746 pixels)
  - Scanner: Pannoramic 250 Flash, 3DHISTECH (Hungary)
  - MPP (microns per pixel): 0.221  
(3개만 0.388)
  - Apparent magnification: 20X
  - Image bit depth: 8 bits
  - Color channel: RGBA

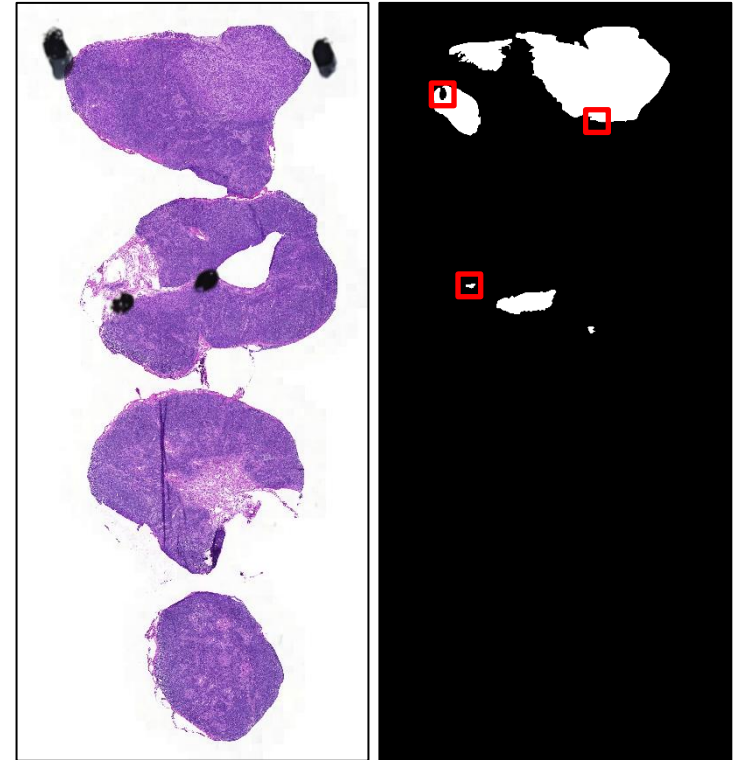
# Data set

- “mrxs” [1]
  - Tile image (jpeg으로 압축)로 구성된 format
    - 동일 폴더에 “example1.mrxs” 파일과, “example1” 폴더가 존재
    - “.mrxs” 파일에는 헤더정보를 포함
    - “example1” 폴더에는 실제 tile image data가 포함
    - File size는 500MB~2GB로 다양함

Data0000.dat	2018-07-24 오전...	GOM 미디어 파일...	6KB
Data0001.dat	2018-07-24 오전...	GOM 미디어 파일...	6KB
Data0002.dat	2018-07-24 오전...	GOM 미디어 파일...	890KB
Data0003.dat	2018-07-24 오전...	GOM 미디어 파일...	560KB
Data0004.dat	2018-07-24 오전...	GOM 미디어 파일...	418,473KB
Data0005.dat	2018-07-24 오전...	GOM 미디어 파일...	433KB
Data0006.dat	2018-07-24 오전...	GOM 미디어 파일...	125,613KB
Data0007.dat	2018-07-24 오전...	GOM 미디어 파일...	36,808KB
Data0008.dat	2018-07-24 오전...	GOM 미디어 파일...	9,857KB
Data0009.dat	2018-07-24 오전...	GOM 미디어 파일...	2,373KB
Data0010.dat	2018-07-24 오전...	GOM 미디어 파일...	613KB
Data0011.dat	2018-07-24 오전...	GOM 미디어 파일...	160KB
Data0012.dat	2018-07-24 오전...	GOM 미디어 파일...	52KB
Data0013.dat	2018-07-24 오전...	GOM 미디어 파일...	15KB
Data0014.dat	2018-07-24 오전...	GOM 미디어 파일...	7KB
Data0015.dat	2018-07-24 오전...	GOM 미디어 파일...	5KB
Data0016.dat	2018-07-24 오전...	GOM 미디어 파일...	39KB
Data0017.dat	2018-07-24 오전...	GOM 미디어 파일...	12KB
Data0018.dat	2018-07-24 오전...	GOM 미디어 파일...	7KB
Data0019.dat	2018-07-24 오전...	GOM 미디어 파일...	230KB
Data0020.dat	2018-07-24 오전...	GOM 미디어 파일...	2KB
Data0021.dat	2018-07-24 오전...	GOM 미디어 파일...	5KB
Data0022.dat	2018-07-24 오전...	GOM 미디어 파일...	2KB
Index.dat	2018-07-24 오전...	GOM 미디어 파일...	513KB
Slidedat.ini	2018-07-24 오전...	구성 설정	14KB

# Data set

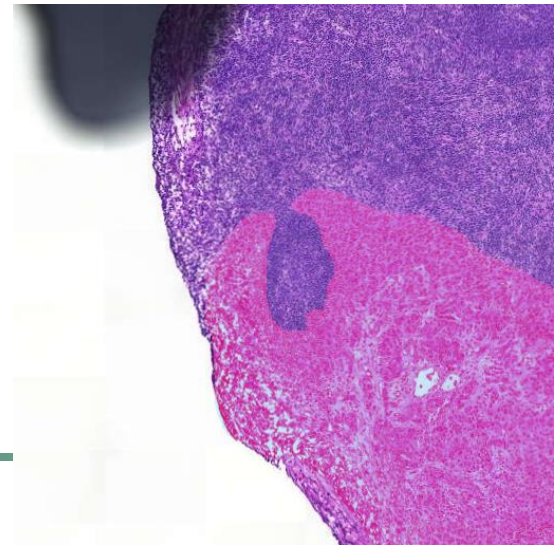
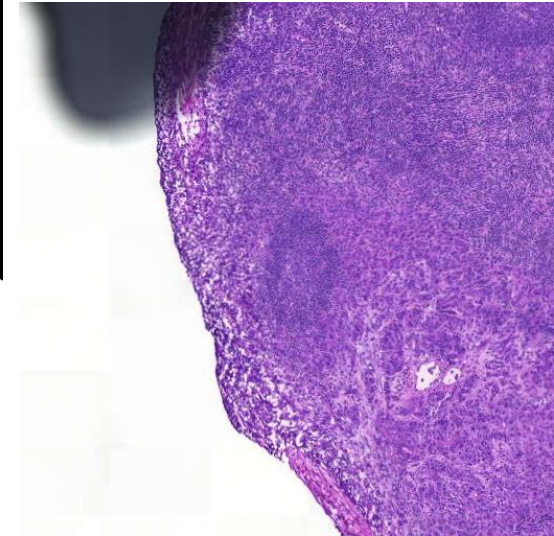
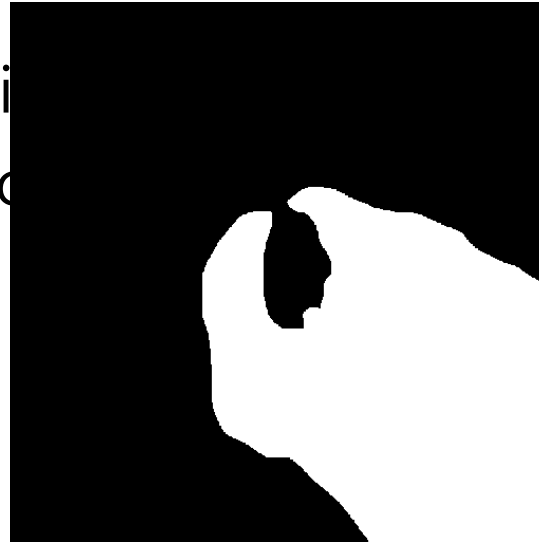
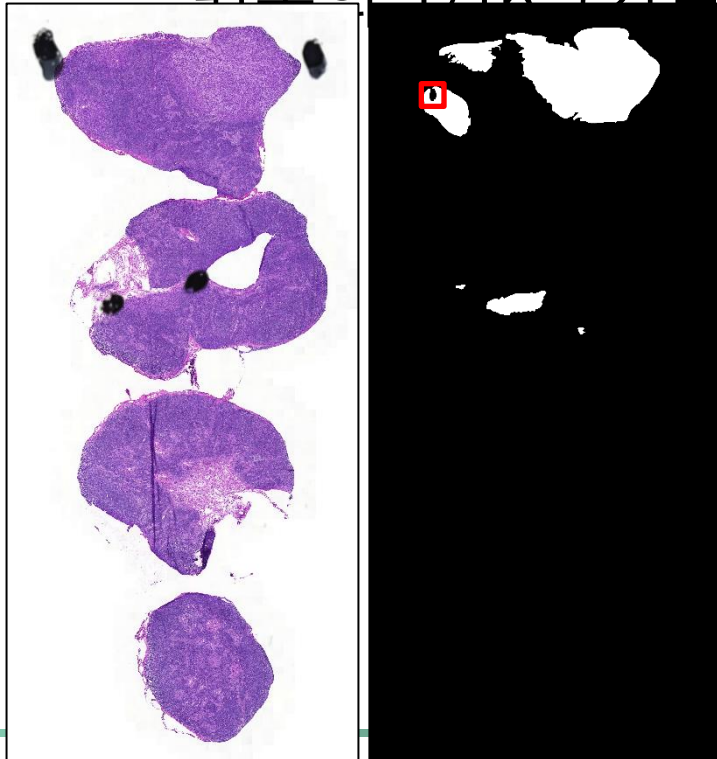
- Mask
  - 흰 영역: tumor region
  - 검정 영역: normal cell or background
  - 원본의 1/16크기로 레이블링
  - 원본데이터와 매칭하려면 min-offset 필요



# Data set

- Mask

- 흰 영역: tumor region
  - 검정 영역: normal region
- 위보이 1/16크기



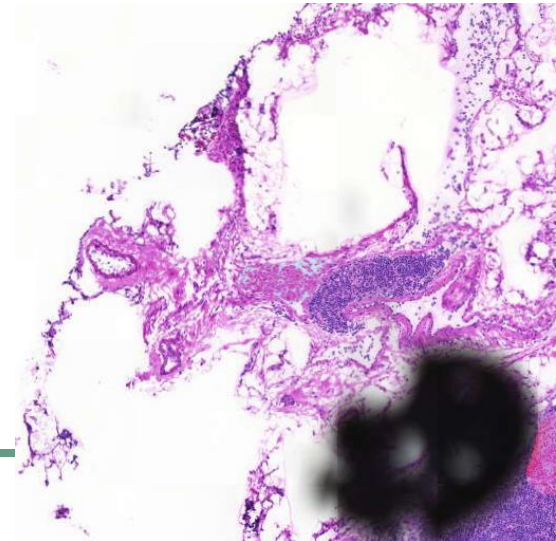
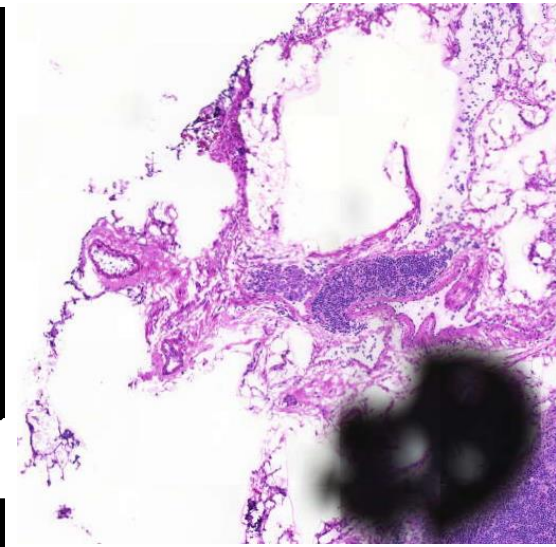
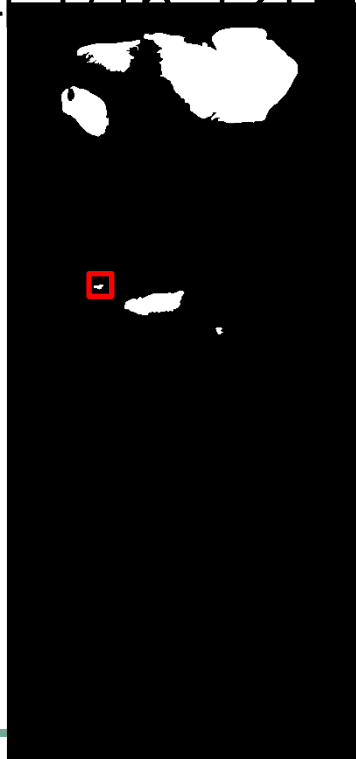
# Data set

- Mask

- 흰 영역: tumor region

- 검정 영역: normal region

위보이 1/16크기



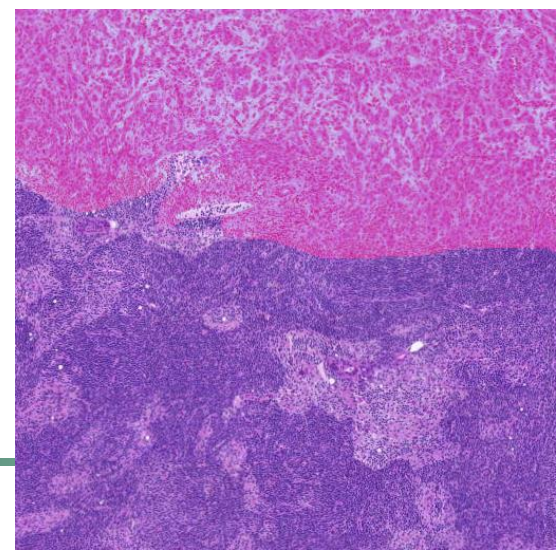
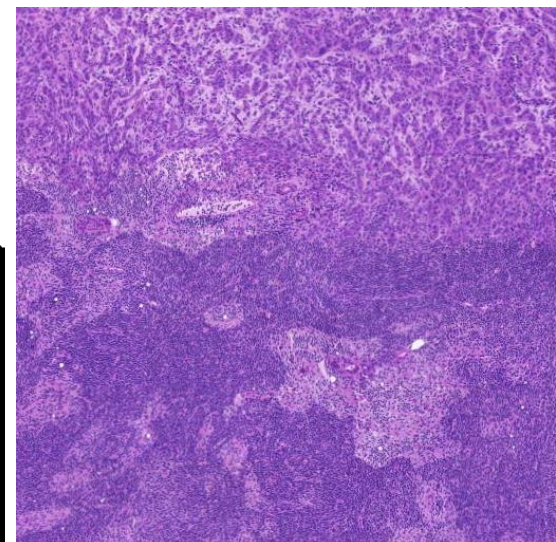
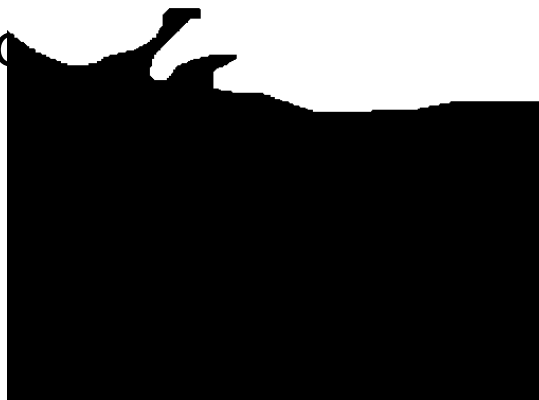
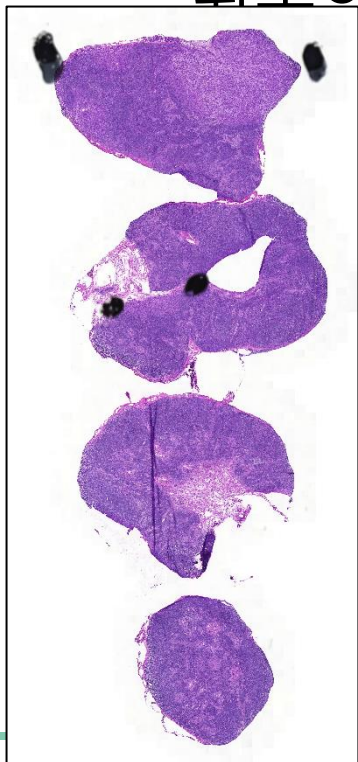


# Data set

- Mask

- 흰 영역: tumor regi
- 검정 영역: normal o

위보이 1/16크기

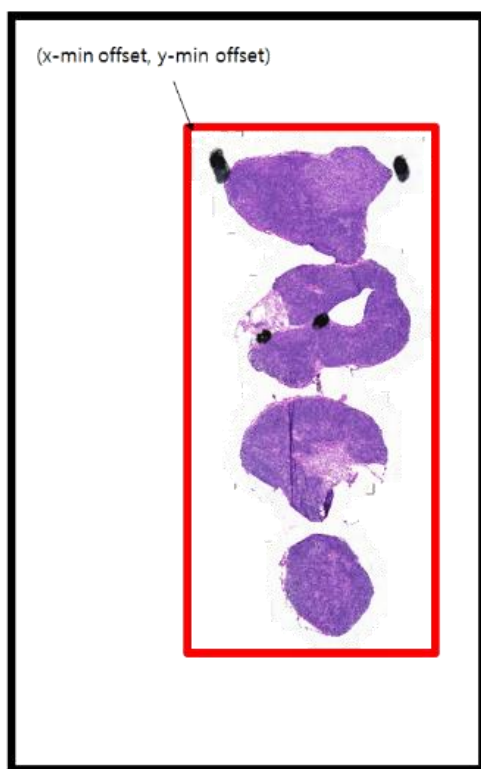


# Data set

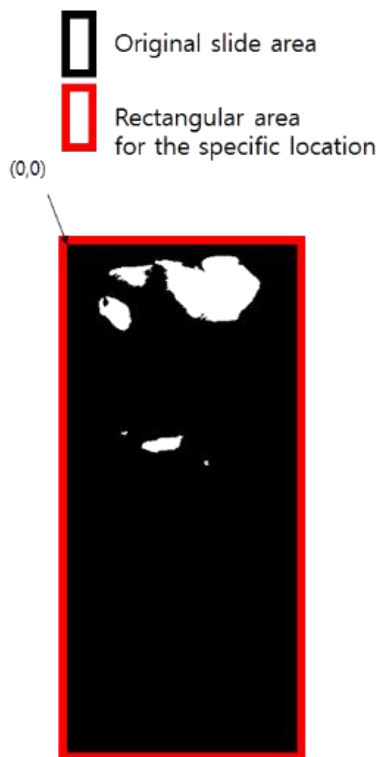
- Dataset\_information.csv

- File name, class, x-min offset, y-min offset 제공

Image views



Slide1 at level 4



Metastasis regions of slide1 at level 4  
(Label 1)

	A	B	C
1	File name	x-min offset	y-min offset
2	Slide002	304	0
3	Slide003	7744	39056
4	Slide004	6160	27856
102	Slide152	480	34656
103	Slide154	30464	71584
104	Slide155	320	41248



# Library

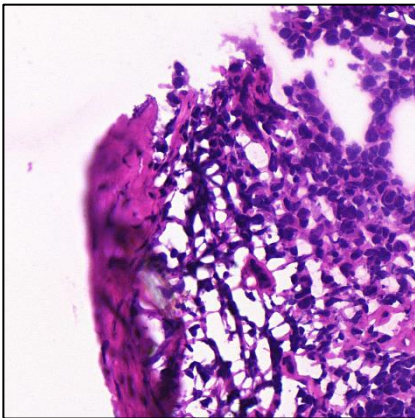
- OpenSlide [2]
  - Tile image로 저장된 format을 처리할 수 있는 library

```
image_region = ex_slide.read_region ([0,0], 4, [ex_slide.level_dimensions[4][0], ex_slide.level_dimensions[4][1]])
```

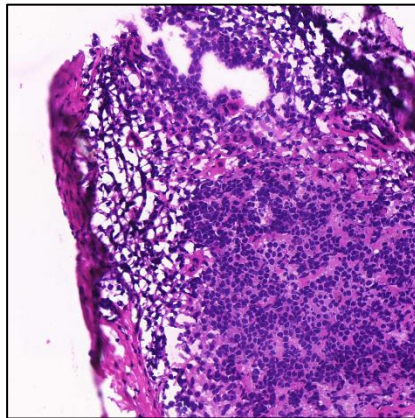
**Read\_region([시작좌표], level, [크기]): level0 기준 시작좌표에서 level에서의 크기로 영역 반환**

**return type:** `class PIL.Image.Image` [\[source\]](#)

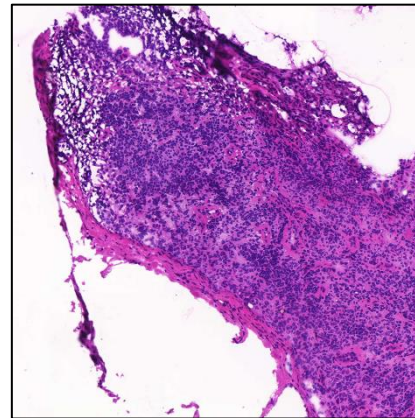
동일 시작좌표로부터 level만 다르게 한 경우의 예시



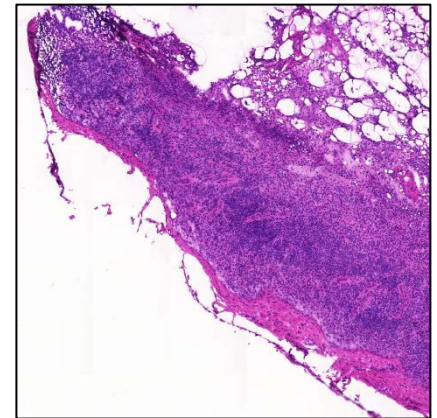
Level 0



Level 1



Level 2



Level 3

[2] <https://openslide.org/formats/>

# Measure

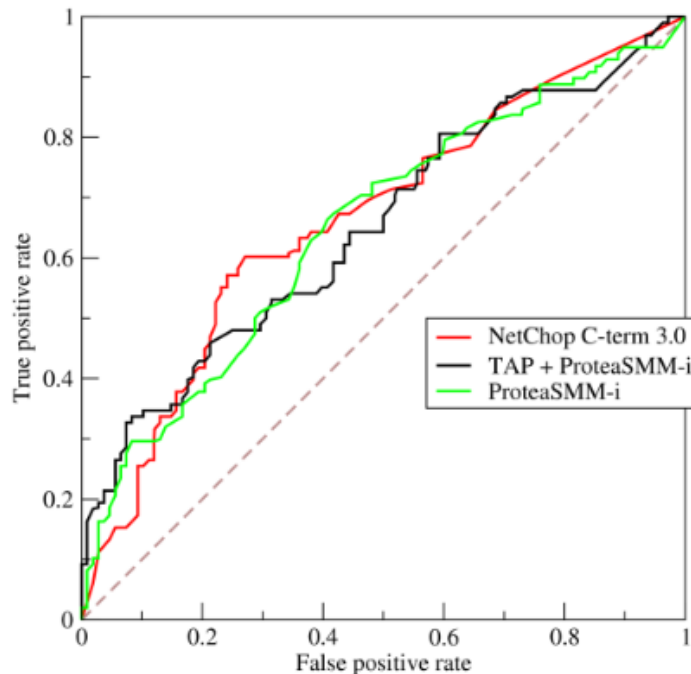
- 제출 형태
  - "Output.csv"
  - 슬라이드 이름, tumor가 있을 확률

	A	B
1	Slide198	0.0003
2	Slide199	0.415464
3	Slide200	0.35496
98	Slide295	0.547655
99	Slide296	0.21654
100	Slide297	0.7892



# Measure

- AUC (Area Under Curve)
  - ROC 커브에서의 밑 면적
    - 가로축: 1 – 특이도 (specificity)
    - 세로축: 민감도 (sensitivity)



# Examples

- Inference time comparisons
  - Pixel by pixel classification
  - Patch size: 448x448 (level 0)
  - Heat map (cutoff: 0.9)
  - GPU: GTX 1080 Ti

Sampling (Level0)	N of inference	Time each slide	Exp. Total	AUC
1600 (Level4*100)	4,756 (58 x 82)	2분	3.5시간	0.91
800 (Level4*50)	18,954 (117 x 162)	8분 (2분*4배)	0.5일	0.95
320 (Level4*20)	118,900 (290 x 410)	50분 (2분*25배)	3.5일	0.96

## Clinical Collaborators @ Asan Medical Center

### Radiology

Joon Beom Seo, SangMin Lee<sup>A,B</sup>, Dong Hyun, Yang, Hyung Jin Won, Ho Sung Kim, Seung, Chai Jung, Ji Eun Park

### Neurology

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

### Cardiology

Jaekwan Song, Jongmin Song, Younghak Kim

### Internal Medicine

Jeongsik Byeon, Kang Mo Kim

### Pathology

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

### Surgery

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

### Anesthesiology

Sung-Hoon Kim, Eun Ho Lee

### Emergency Medicine

Jeong Hyun Lee, Gilsun Hong, Dong-Woo Seo

