Problem Statement

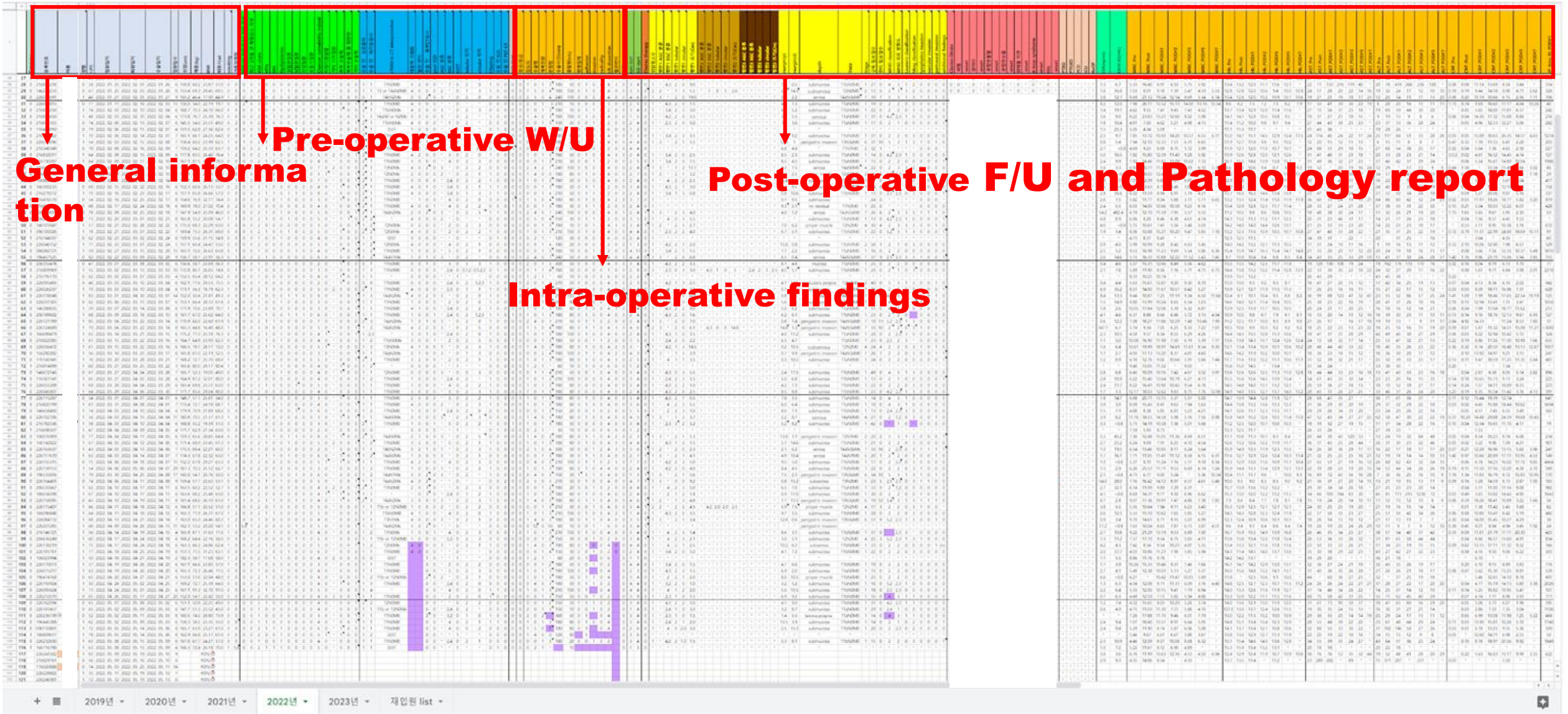
Woohwan Jung (whjung@hanyang.ac.kr)

Fall 2022

HANYANG UNIVERSITY

Data Science Lab

Clinical Data



Problem statement

- Given
  - A prediction date POD #n
    - Among POD#0 (Post-op), POD#1, POD#2, POD#3, POD#5 , POD#7
  - the **available** clinical data of a patient at a specific date after the surgery
  - General information
  - Pre-op, Intra-op, and Post-op reports
  - A pathology report
- Predict the occurrence of the DSL

Available Information

- For the prediction at POD #n, the following information is available
  - General Information
    - E.g.) Age, Height, Weight, BMI, …
  - Pre-op and Intro-op data
  - Blood test results
    - Pre, Post
    - POD#1, …POD#n
  - Pathology reports are available from POD#5

Pre-processing

- Train data
  - Will be provided as it is
- Test data
  - Step 0: Drop the label (DSL) column
  - Step 1: Draw a prediction date
  - Step 2: Masking unavailable information

Test data

Sample data (Before preprocessing)

Sample data (After preprocessing)

Cate	성별	나이	입원일자	퇴원일자	수술일자	입원일수	신장(cm)	체중(kg)	BMI	퇴원시wt	smoking	가족암병력	ASA score	HTN	DM	Dyslipidemia	심혈관질환	뇌혈관질환	신장질환	호흡기질환	Charson comobidity index8	타장기원발암	기타 기저질환	복부 수술력	복부수술력 중 위관련
1	0	52	2022. 01. 02	2022. 01. 07	2022. 01. 03	4	159.3	53.8	21.20	54.1	0	0	1	0	0	0	0	0	0	0	3	0	*	1	0
0	0	53	2022. 01. 03	2022. 01. 10	2022. 01. 04	6	156.0	48.9	20.09	48.3	0	0	1	0	0	0	0	0	0	0	3	0	*	0	0
0	1	58	2022. 01. 04	2022. 01. 15	2022. 01. 05	10	164.9	67.1	24.68	56.0	0	2,4	2	1	1	0	0	0	0	0	4	0	*	0	0
1	0	60	2022. 01. 04	2022. 01. 09	2022. 01. 05	4	152.3	54.0	23.28	54.4	0	0	2	1	0	0	0	0	0	0	4	0	*	0	0
0	1	61	2022. 01. 05	2022. 01. 13	2022. 01. 06	7	161.5	59.0	22.62	53.0	1	0	1	0	0	0	0	0	0	0	4	0	*	0	0
0	0	55	2022. 01. 06	2022. 01. 13	2022. 01. 07	6	158.6	67.1	26.68	68.5	0	2	1	0	0	1	0	0	0	0	3	0	*	1	0
1	1	59	2022. 01. 09	2022. 01. 14	2022. 01. 10	4	163.3	57.6	21.60	57.4	0	0	2	1	0	0	0	0	0	0	3	0	*	0	0
0	1	50	2022. 01. 06	2022. 01. 17	2022. 01. 10	7	179.7	89.3	27.65	85.0	1	0	2	1	1	0	0	0	0	0	4	0	2	0	0
0	1	74	2022. 01. 03	2022. 01. 19	2022. 01. 10	9	162.2	55.5	21.10	51.3	0	2	3	0	0	0	0	0	0	0	5	0	*	1	1
0	1	74	2022. 01. 10	2022. 01. 18	2022. 01. 11	7	165.4	76.5	27.96	72.8	0	0	1	0	0	0	0	0	0	0	5	0	*	0	0

1	환자	Pred	성별	나이	입원일자	수술일자	신장(cm)	체중(kg)	BMI	smoking	가족암병력	ASA score	HTN	DM	Dyslipidemia	심혈관질환	뇌혈관질환	신장질환	호흡기질환	Charson comobidity index8	타장기원발암	기타 기저질환	복부 수술력	복부수술력 중 위관련
2	1	1	0	52	2022. 01. 02	2022. 01. 03	159.3	53.8	21.20	0	0	1	0	0	0	0	0	0	0	3	0	*	1	0
3	2	5	0	53	2022. 01. 03	2022. 01. 04	156.0	48.9	20.09	0	0	1	0	0	0	0	0	0	0	3	0	*	0	0
4	3	1	1	58	2022. 01. 04	2022. 01. 05	164.9	67.1	24.68	0	2,4	2	1	1	0	0	0	0	0	4	0	*	0	0
5	4	0	0	60	2022. 01. 04	2022. 01. 05	152.3	54.0	23.28	0	0	2	1	0	0	0	0	0	0	4	0	*	0	0
6	5	0	1	61	2022. 01. 05	2022. 01. 06	161.5	59.0	22.62	1	0	1	0	0	0	0	0	0	0	4	0	*	0	0
7	6	1	0	55	2022. 01. 06	2022. 01. 07	158.6	67.1	26.68	0	2	1	0	0	1	0	0	0	0	3	0	*	1	0
8	7	2	1	59	2022. 01. 09	2022. 01. 10	163.3	57.6	21.60	0	0	2	1	0	0	0	0	0	0	3	0	*	0	0
9	8	3	1	50	2022. 01. 06	2022. 01. 10	179.7	89.3	27.65	1	0	2	1	1	0	0	0	0	0	4	0	2	0	0
10	9	3	1	74	2022. 01. 03	2022. 01. 10	162.2	55.5	21.10	0	2	3	0	0	0	0	0	0	0	5	0	*	1	1
11	10	0	1	74	2022. 01. 10	2022. 01. 11	165.4	76.5	27.96	0	0	1	0	0	0	0	0	0	0	5	0	*	0	0
12																								

Test data

Sample data (Before preprocessing)

		성별	나이	입원일자	퇴원일자	수술일자	입원일수	신장(cm)	체중(kg)	BMI	퇴원시wt	smoking	가족암병력	WBC_Pre	WBC_Post	WBC_POD#1	WBC_POD#2	WBC_POD#3	WBC_POD#5	WBC_POD#7
1	1	0	52	2022. 01. 02	2022. 01. 07	2022. 01. 03	4	159.3	53.8	21.20	54.1	0	0	5.27	4.15	7.09	*	*	*	*
2	0	0	53	2022. 01. 03	2022. 01. 10	2022. 01. 04	6	156.0	48.9	20.09	48.3	0	0	7.74	12.40	8.39	4.81	3.62	4.53	4.80
3	0	1	58	2022. 01. 04	2022. 01. 15	2022. 01. 05	10	164.9	67.1	24.68	56.0	0	2,4	7.66	9.13	7.68	7.16	7.01	5.75	7.38
4	1	0	60	2022. 01. 04	2022. 01. 09	2022. 01. 05	4	152.3	54.0	23.28	54.4	0	0	13.55	7.80	12.10	*	5.97	4.23	6.56
5	0	1	61	2022. 01. 05	2022. 01. 13	2022. 01. 06	7	161.5	59.0	22.62	53.0	1	0	8.92	13.70	10.75	11.64	6.79	6.00	7.20
6	0	0	55	2022. 01. 06	2022. 01. 13	2022. 01. 07	6	158.6	67.1	26.68	68.5	0	2	4.59	7.55	7.93	7.01	4.35	4.31	*
7	1	1	59	2022. 01. 09	2022. 01. 14	2022. 01. 10	4	163.3	57.6	21.60	57.4	0	0	6.08	9.48	14.30	7.25	*	*	*
8	0	1	50	2022. 01. 06	2022. 01. 17	2022. 01. 10	7	179.7	89.3	27.65	85.0	1	0	5.91	14.65	12.15	9.27	5.92	4.16	6.99
9	0	1	74	2022. 01. 03	2022. 01. 19	2022. 01. 10	9	162.2	55.5	21.10	51.3	0	2	6.13	16.49	17.72	13.31	9.07	6.06	8.37
10	0	1	74	2022. 01. 10	2022. 01. 18	2022. 01. 11	7	165.4	76.5	27.96	72.8	0	0	4.98	6.82	7.01	7.82	6.48	6.35	6.17

Sample data (After preprocessing)

1	환자	Pred	성별	나이	입원일자	수술일자	신장(cm)	체중(kg)	BMI	smoking	가족암병력	WBC_Pre	WBC_Post	WBC_POD#1	WBC_POD#2	WBC_POD#3	WBC_POD#5	WBC_POD#7
2	1	1	0	52	2022. 01. 02	2022. 01. 03	159.3	53.8	21.20	0	0	5.27	4.15	7.09	*	*	*	*
3	2	5	0	53	2022. 01. 03	2022. 01. 04	156.0	48.9	20.09	0	0	7.74	12.40	8.39	4.81	3.62	##	*
4	3	1	1	58	2022. 01. 04	2022. 01. 05	164.9	67.1	24.68	0	2,4	7.66	9.13	7.68	*	*	*	*
5	4	0	0	60	2022. 01. 04	2022. 01. 05	152.3	54.0	23.28	0	0	13.55	7.80	*	*	*	*	*
6	5	0	1	61	2022. 01. 05	2022. 01. 06	161.5	59.0	22.62	1	0	8.92	13.70	*	*	*	*	*
7	6	1	0	55	2022. 01. 06	2022. 01. 07	158.6	67.1	26.68	0	2	4.59	7.55	7.93	*	*	*	*
8	7	2	1	59	2022. 01. 09	2022. 01. 10	163.3	57.6	21.60	0	0	6.08	9.48	14.30	7.25	*	*	*
9	8	3	1	50	2022. 01. 06	2022. 01. 10	179.7	89.3	27.65	1	0	5.91	14.65	12.15	9.27	5.92	*	*
10	9	3	1	74	2022. 01. 03	2022. 01. 10	162.2	55.5	21.10	0	2	6.13	16.49	17.72	13.31	9.07	*	*
11	10	0	1	74	2022. 01. 10	2022. 01. 11	165.4	76.5	27.96	0	0	4.98	6.82	*	*	*	*	*

Evaluation criteria

$$Recall@K = \frac{\#\ of\ complications\ among\ the\ top - K\ patients}{\#\ of\ complications\ in\ the\ test\ set}$$

We'll use a combination of Recall@Ks

Example) $Recall@10 + Recall@50 + Recall@100$

You need to submit a sorted list of the patients in the test set