

## Table of contents

**Supplementary Figure 1. Examples of data extraction methods of ETL process - MS-SQL User-Defined Function: PT\_BASIC (Patient Basic Information) data manipulation language (DML) structure**

**Supplementary Figure 2. Examples of data extraction methods of ETL process - MS-SQL User-Defined Function: DG\_CNCR (diagnosis\_cancertype) data manipulation language (DML) structure**

**Supplementary Figure 3. Examples of data extraction methods of ETL process - MS-SQL User-Defined Function: EM\_LAB (test\_diagnostic test) data manipulation language (DML) structure**

**Supplementary Figure 4. Examples of data extraction methods of ETL process - MS-SQL User-Defined Function: EM\_IMEX (test\_imaging test) data manipulation language (DML) structure**

**Supplementary Figure 5. Examples of data extraction methods of ETL process - MS-SQL User-Defined Function: OP\_OPNN (surgery\_surgical findings) data manipulation language (DML) structure**

**Supplementary Figure 6. Examples of data extraction methods of ETL process - MS-SQL User-Defined Function: TX\_CHTH (treatment\_chemotherapy) data manipulation language (DML) structure**

**Supplementary Figure 7. Examples of data extraction methods of ETL process - MS-SQL User-Defined Function: TE\_DEAD (progress\_death information) data manipulation language (DML) structure**

**Supplementary Figure 8. Development of a SQL table-valued function to convert free text remote metastasis details to YCDL site codes, enhancing adaptability and simplifying ETL query management.**

**Supplementary Figure 9. Examples of data extraction methods: Python User-Defined Function.**

**Supplementary Figure 10. Overview of Python script-based operations for complicated elements**

**Supplementary Figure 11. Overview of the automated collection, processing, and interpretation of next-generation sequencing data**

**Supplementary Figure 12. Continuous QC steps within distinct closed-loop systems**

**Supplementary Figure 13. Architecture of the developed clinical decision support system (CDSS).**

**Supplementary Table 1. Accuracy and missing data rates of segmented data in YCDL from surgical and molecular pathology for various cancer types**

**Supplementary Movie 1: Demonstration of CDSS longitudinal tumor tracking and visualization.** This video demonstrates the visualization of tumor volume changes over three selected time points during cancer treatment. In the left section, CT images in axial and coronal planes display the segmented total tumor volume, color-coded

in yellow, green, and red to represent different stages of disease burden depending on response to systemic therapy. The segmentation process currently involves manual contouring. The right section illustrates bar graphs showing the total tumor volume (in cc) and the number of metastatic lesions for each time point. This visualization aids in assessing changes in disease burden during treatment.

**Supplementary Movie 2: Timeline and comprehensive overview of a patient's healthcare journey, showing the chronological sequence of events and progression of the patient's treatment.** This video illustrates the timeline of key landmark events for a patient throughout their cancer treatment. At the top, the timeline spans from the first to the most recent visit, with small bars representing each encounter. A slightly shaded area indicates the current time frame being covered. Users can zoom in and out and adjust the time frame. Each encounter highlights the department visited. In the treatment section, the timeline displays the administration of systemic therapies, such as 1st, 2nd, and 3rd line chemotherapy and targeted agents, along with the corresponding cycles (1, 2, 3, ...). The second line of the treatment section shows radiotherapy and surgery. In the progression section, the timeline indicates progression events and specifies the sites where progression occurred. The functional imaging section displays PET-CT, CT, and MRI scans, allowing viewers to access corresponding radiology reports. The gene/lab/pathology section presents the patient's tumor marker (e.g., CEA) test results.

**Supplementary Data 1: Excel sheets containing data feature definitions and formats for all features across all cancer types**

**Supplementary Figure 1. Examples of data extraction methods: MS-SQL query - Extract data from the YCDL table by joining or unioning with the DSC table.**

**PT\_BASIC (Patient Basic Information) data manipulation language (DML) structure example**

| DB name |                | Table name     |                             |                               |       |                                                                                                                                                                             |
|---------|----------------|----------------|-----------------------------|-------------------------------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MAIN    | DSC_cancertype | CNCR_PATINFO   | (Patient Basic Information) | MAIN                          | WHERE | MAIN.SvrId = 10 --Hospital location<br>MAIN.BirthYmd NOT LIKE " --Exclude entries with blank date of birth<br>MAIN.Sex NOT LIKE " --Exclude entries with blank gender       |
| JOIN    | INNER JOIN     | DSC_cancertype | CNCR_TUMOR_RGT              | (Cancer Patient Registration) | RGT   | ON<br>MAIN.SvrId = RGT.SvrId<br>MAIN.AlsUnitNo = RGT.AlsUnitNo                                                                                                              |
|         |                |                |                             |                               | WHERE | RGT.VldYn = 'Y' --Data with valid cancer registration number<br>RGT.DelYn = 'N' --Data not deleted<br>RGT.FsDiagYmd >= '20060101' --Diagnosis date is after January 1, 2006 |
|         | INNER JOIN     | YCDL           | YCDL_CODE_TBL               | (YCDL Code Table)             | CD    | ON<br>RGT.SvrId = CD.SvrId<br>RGT.ICDOCd = CD.CD_VAL<br>CD.CD_TYPE = 1090600 --ICDO Code: Modify code type conditions according to cancer type                              |

**Supplementary Figure 2. Examples of data extraction methods: MS-SQL query - Extract data from the YCDL table by joining or unioning with the DSC table.**

**DG\_CNCR (diagnosis\_cancertype) DML structure example**

| DB name |            | Table name      |                 |                                        |      |       |                                                                                                                                                                                                                                                                     |
|---------|------------|-----------------|-----------------|----------------------------------------|------|-------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MAIN    |            | DSC_cancertype  | CNCR_TUMOR_RGT  | (Cancer Patient Registration)          | RGT  | WHERE | RGT.VldYn = 'Y' --Data with valid cancer registration number<br>RGT.DelYn = 'N' --Data not deleted<br>RGT.FsDiagYmd >= '20060101' --Diagnosis date is after January 1, 2006                                                                                         |
| JOIN    | INNER JOIN | YCDL_cancertype | PT_BASIC        | (Patient Basic Information)            | PT   | ON    | RGT.SvrId = PT.SVR_ID<br>RGT.AlsUnitNo = PT.RESCH_REG_NO                                                                                                                                                                                                            |
|         | LEFT JOIN  | DSC_cancertype  | CNCR_PATINFO    | (Patient Basic Information)            | BTH  | ON    | RGT.SvrId = BTH.SvrId<br>RGT.AlsUnitNo = BTH.AlsUnitNo                                                                                                                                                                                                              |
|         | LEFT JOIN  | DSC_cancertype  | CNCR_TUMOR_TRC  | (Cancer Patient Follow-up Information) | TRC  | ON    | RGT.SvrId = TRC.SvrId<br>RGT.AlsUnitNo = TRC.AlsUnitNo<br>RGT.TumNo = TRC.TumNo                                                                                                                                                                                     |
|         | LEFT JOIN  | DSC_cancertype  | CNCR_TUMOR_TRET | (Cancer Patient Treatment)             | TRET | ON    | RGT.SvrId = TRC.SvrId<br>RGT.AlsUnitNo = TRC.AlsUnitNo<br>RGT.TumNo = TRC.TumNo                                                                                                                                                                                     |
|         | LEFT JOIN  | YCDL            | YCDL_CODE_TBL   | (YCDL Code Table)                      | CD   | ON    | RGT.SvrId = CD.SvrId<br>RGT.ICDOCd = CD1.CD_TYPE = 1090600 --ICDO Code: Modify code type conditions according to cancer type<br>RGT.ICDOCdM = CD2.CD_TYPE = 1091700 --ICDO Morphology code<br>output of CASE statement = CD3.CD_TYPE = 1091100 --initial stage code |

**Supplementary Figure 3. Examples of data extraction methods: MS-SQL query - Extract data from the YCDL table by joining or unioning with the DSC table.**

**EM\_LAB (test\_diagnostic test) DML structure example**

| DB name |                | Table name      |                   |                             |     |                                                                                                                                           |
|---------|----------------|-----------------|-------------------|-----------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------|
| MAIN    | DSC_cancertype | CNCR_LAB        | (Diagnostic Test) | MAIN                        | -   |                                                                                                                                           |
| JOIN    | INNER JOIN     | YCDL_cancertype | PT_BASIC          | (Patient Basic Information) | PT  | ON<br>MAIN.SvrId = PT.SVR_ID<br>MAIN.AlsUnitNo = PT.RESCH_REG_NO                                                                          |
|         | INNER JOIN     | DSC_cancertype  | CNCR_DX           | (Encounter/Diagnosis)       | DX  | ON<br>MAIN.SvrId = DX.SvrId<br>MAIN.AlsUnitNo = DX.AlsUnitNo<br>MAIN.AlsChosNo = DX.AlsChosNo                                             |
|         | INNER JOIN     | YCDL            | YCDL_CODE_TBL     | (YCDL Code Table)           | CD1 | ON<br>MAIN.SvrId = CD1.SvrId<br>MAIN.GrOrdCd = CD1.CD_VAL<br>CD1.CD_TYPE = 1110800 --Modify code type conditions according to cancer type |
|         | INNER JOIN     | YCDL            | YCDL_CODE_TBL     | (YCDL Code Table)           | CD2 | ON<br>MAIN.SvrId = CD2.SvrId<br>MAIN.OrdCd = CD2.CD_VAL<br>CD2.CD_TYPE = 1111000 --Modify code type conditions according to cancer type   |

**Supplementary Figure 4. Examples of data extraction methods: MS-SQL query - Extract data from the YCDL table by joining or unioning with the DSC table.**

**EM\_IMEX (test\_imaging test) DML structure example**

| DB name |                | Table name      |                |                             |     |                                                                                                                                         |
|---------|----------------|-----------------|----------------|-----------------------------|-----|-----------------------------------------------------------------------------------------------------------------------------------------|
| MAIN    | DSC_cancertype | CNCR_IMAGE      | (Imaging Test) | MAIN                        | -   |                                                                                                                                         |
| JOIN    | INNER JOIN     | YCDL_cancertype | PT_BASIC       | (Patient Basic Information) | PT  | ON<br>MAIN.SvrId = PT.SVR_ID<br>MAIN.AlsUnitNo = PT.RESCH_REG_NO                                                                        |
|         | INNER JOIN     | DSC_cancertype  | CNCR_DX        | (Encounter/Diagnosis)       | DX  | ON<br>MAIN.SvrId = DX.SvrId<br>MAIN.AlsUnitNo = DX.AlsUnitNo<br>MAIN.AlsChosNo = DX.AlsChosNo                                           |
|         | INNER JOIN     | YCDL            | YCDL_CODE_TBL  | (YCDL Code Table)           | CD1 | ON<br>MAIN.SvrId = CD1.SvrId<br>MAIN.OrdCd = CD1.CD_VAL<br>CD1.CD_TYPE = 1121100 --Modify code type conditions according to cancer type |
|         | LEFT JOIN      | YCDL            | YCDL_CODE_TBL  | (YCDL Code Table)           | CD2 | ON<br>MAIN.SvrId = CD2.SvrId<br>MAIN.EDICd = CD2.CD_VAL<br>CD2.CD_TYPE = 1120700 --EDI code                                             |

**Supplementary Figure 5. Examples of data extraction methods: MS-SQL query - Extract data from the YCDL table by joining or unioning with the DSC table.**

**OP\_OPNN (surgery\_surgical findings) DML structure example**

|            |            | DB name       | Table name        |                             |        |                                                                                                                                                   |                                                                                                                                                                                                                            |
|------------|------------|---------------|-------------------|-----------------------------|--------|---------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MAIN       |            | DSC_LUNG      | CNCR_FRM          | (Document Form)             | FRM    | WHERE                                                                                                                                             | FrmNm LIKE 'Operation note%' OR FrmNm LIKE 'Surgery note%'<br>AttrCd IN ('A0000664','A0000743','A0000728','A0007225') --Surgeon                                                                                            |
|            |            |               |                   |                             |        | Sequence for removing duplicate rows                                                                                                              | ROW_NUMBER() OVER() --Select the most recent surgical record of the same surgery by the same surgeon                                                                                                                       |
| JOIN       | INNER JOIN | YCDL_LUNG     | PT_BASIC          | (Patient Basic Information) | PT     | ON                                                                                                                                                | FRM.SvrId = PT.SVR_ID<br>FRM.AlsUnitNo = PT.RESCH_REG_NO                                                                                                                                                                   |
|            | LEFT JOIN  | DSC_LUNG      | CNCR_FRM          | (Document Form)             | COL_1  | ON                                                                                                                                                | FRM.SvrId = COL_1.SvrId<br>FRM.AlsUnitNo = COL_1.AlsUnitNo<br>FRM.AlsChosNo = COL_1.AlsChosNo<br>FRM.FrmClnKey = COL_1.FrmClnKey<br>COL_1.AttrCd IN (attribute codes where the column value exists) or Value condition     |
|            | LEFT JOIN  | DSC_LUNG      | CNCR_FRM          | (Document Form)             | COL_2  | ON                                                                                                                                                | FRM.SvrId = COL_2.SvrId<br>FRM.AlsUnitNo = COL_2.AlsUnitNo<br>FRM.AlsChosNo = COL_2.AlsChosNo<br>FRM.FrmClnKey = COL_2.FrmClnKey<br>COL_1.AttrCd IN (attribute codes where the column value exists) or Value condition     |
|            | LEFT JOIN  | DSC_LUNG      | CNCR_FRM          | (Document Form)             | COL_3  | ON                                                                                                                                                | FRM.SvrId = COL_3.SvrId<br>FRM.AlsUnitNo = COL_3.AlsUnitNo<br>FRM.AlsChosNo = COL_3.AlsChosNo<br>FRM.FrmClnKey = COL_3.FrmClnKey<br>COL_1.AttrCd IN (attribute codes where the column value exists) or Value condition     |
|            | ...        | ...           | ...               | ...                         | ...    | ...                                                                                                                                               | ...                                                                                                                                                                                                                        |
|            | LEFT JOIN  | DSC_LUNG      | CNCR_FRM          | (Document Form)             | COL_24 | ON                                                                                                                                                | FRM.SvrId = COL_24.SvrId<br>FRM.AlsUnitNo = COL_24.AlsUnitNo<br>FRM.AlsChosNo = COL_24.AlsChosNo<br>FRM.FrmClnKey = COL_24.FrmClnKey<br>COL_1.AttrCd IN (attribute codes where the column value exists) or Value condition |
|            | LEFT JOIN  | DSC_LUNG      | CNCR_FRM          | (Document Form)             | OpCd   | WHERE                                                                                                                                             | AttrNm LIKE 'Operation name%' AND AttrValue <> " " AND AttrDataCd <> " " --Attribute code for the attribute "surgery name"                                                                                                 |
|            |            |               |                   |                             |        | ON                                                                                                                                                | FRM.SvrId = OpCd.SvrId<br>FRM.FrmClnKey = OpCd.FrmClnKey                                                                                                                                                                   |
| INNER JOIN | YCDL       | YCDL_CODE_TBL | (YCDL Code Table) | CD                          | ON     | OpCd.SvrId = CD.SvrId<br>OpCd.AttrDataCd = CD.CD_VAL<br>CD.CD_TYPE = 1180600 --Surgery Code: Modify code type conditions according to cancer type |                                                                                                                                                                                                                            |

**Supplementary Figure 6. Examples of data extraction methods: MS-SQL query - Extract data from the YCDL table by joining or unioning with the DSC table.**

**TX\_CHTH (treatment\_chemotherapy) DML structure example**

| DB name |                | Table name        |                    | Coulmn name                        |                                                               |                                                                                                |                                                                                                                                                                                                                                                                                             |
|---------|----------------|-------------------|--------------------|------------------------------------|---------------------------------------------------------------|------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MAIN    | DSC_cancertype | ① CNCR_REGIMEN    | (chemotherapy)     | CAP                                | SvrlId<br>AlsUnitNo<br>AlsChosNo<br>OrdYmd<br>OrdSeq<br>TblNm | WHERE                                                                                          | StrYmd NOT LIKE '' --Exclude entries with a blank start date<br>--Although there are prescription sequence numbers where ExecYmd is NULL, the condition to exclude blank execution dates is not set<br>INNER JOIN [UFN_YCDL_CODE] CDT1 --Regimen or clinical drug-related prescription code |
|         | DSC_cancertype | ② CNCR_DRUG       | (medication)       | DRUG                               |                                                               | WHERE                                                                                          | DCYn = 'N' --Select non-cancelled drug prescriptions<br>--If the condition to exclude blank execution dates is set, non-institutional prescription details might be omitted. Therefore, this condition is not set. Check the execution status with DrgExecStusCd                            |
|         | DSC_cancertype | ③ TEMP_TUMOR_TRET | (cancer_treatment) | TRET                               |                                                               | WHERE                                                                                          | DelYn = 'N' --Data not deleted<br>IsCtx = 1 --Selected treatment details for chemotherapy                                                                                                                                                                                                   |
| JOIN    | INNER JOIN     | YCDL_cancertype   | PT_BASIC           | (Patient Basic Information)        | PT                                                            | SvrlId<br>AlsUnitNo                                                                            | ON<br>MAIN.SvrlId = PT.SVR_ID<br>MAIN.AlsUnitNo = PT.RESCH_REG_NO                                                                                                                                                                                                                           |
|         | LEFT JOIN      | CTE               | CAP                | (Recursive Query_Chemotherapy)     | C                                                             | ExecYmd<br>OrdCd<br>ClnDeptCd<br>DiagGrpCd<br>RegimenCd<br>Line<br>Cyl<br>MedTret<br>EndYmd    | ON<br>MAIN.SvrlId = C.SvrlId<br>MAIN.AlsUnitNo = C.AlsUnitNo<br>MAIN.AlsChosNo = C.AlsChosNo<br>MAIN.OrdYmd = C.OrdYmd<br>MAIN.OrdSeq = C.OrdSeq                                                                                                                                            |
|         | LEFT JOIN      | CTE               | DRUG               | (Recursive Query_medication)       | D                                                             | ExecYmd<br>OrdCd<br>ClnDeptCd<br>DrgExecStusCd<br>HoldResn<br>CrNo<br>OrdNo<br>ATC<br>ATC_DESC | ON<br>MAIN.SvrlId = D.SvrlId<br>MAIN.AlsUnitNo = D.AlsUnitNo<br>MAIN.AlsChosNo = D.AlsChosNo<br>MAIN.OrdYmd = D.OrdYmd<br>MAIN.OrdSeq = D.OrdSeq                                                                                                                                            |
|         | LEFT JOIN      | CTE               | TRET               | (Recursive Query_cancer treatment) | E                                                             | StrYmd<br>Commt<br>TumNo<br>ICDOCd                                                             | ON<br>MAIN.SvrlId = E.SvrlId<br>MAIN.AlsUnitNo = E.AlsUnitNo<br>MAIN.OrdYmd = E.StrYmd                                                                                                                                                                                                      |
|         | LEFT JOIN      | DSC_cancertype    | CNCR_DX            | (Encounter/Diagnosis)              | DX                                                            | ChosGb<br>ClnDeptCd                                                                            | ON<br>MAIN.SvrlId = DX.SvrlId<br>MAIN.AlsUnitNo = DX.AlsUnitNo<br>MAIN.AlsChosNo = DX.AlsChosNo                                                                                                                                                                                             |
|         | LEFT JOIN      | YCDMLG            | UFN_YCDL_CODE      | (YCDL Code Table Function)         | CDT                                                           | DiagGrpCdNm<br>RegimenCdNm<br>MedTretNm                                                        | ON<br>MAIN.SvrlId = CDT.SvrlId<br>MAIN.DiagGrpCd = CDT2.CD_TYPE = 1210900 --Cancer Group Code<br>MAIN.RegimenCd = CDT3.CD_TYPE = 1211100 --Regimen Code<br>MAIN.MedTret = CDT4.CD_TYPE = 1211500 --Treatment Aim Code                                                                       |



**Supplementary Figure 7. Examples of data extraction methods: MS-SQL query - Extract data from the YCDL table by joining or unioning with the DSC table.**

**TE\_DEAD (progress\_death information) DML structure example**

| DB name |            | Table name                  |                  |                                        |         |                                                                                                                                                                                                                                                                                                                                                              |
|---------|------------|-----------------------------|------------------|----------------------------------------|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MAIN    |            | DSC_cancertype<br>ALL UNION | ① CNCR_FRM       | (Document Form)                        | FRM     | WHERE FRM.FrmCd IN ('F0000804','F0000817','F0001178','F0002858','F0001112','F0002482','F0002791') --Form codes related to death records<br>FRM.AttrCd IN ('A0003953','A0040412','A0043737','A0050638','A0068984','A0087449','A0016001') --Attribute codes related to date and time of death                                                                  |
|         |            | DSC_cancertype              | ② CNCR_DX        | (Encounter/Diagnosis)                  | DX      | WHERE DX.DscType IN ('41','42') --Discharge type codes: death, death before arrival                                                                                                                                                                                                                                                                          |
|         |            | DSC_cancertype              | ③ CNCR_TUMOR_TRC | (Cancer Patient Follow-up Information) | TRC     | WHERE TRC.DelYn = 'N' --Data not deleted<br>TRC.DeadYmd NOT LIKE '' --Exclude entries with blank date of death                                                                                                                                                                                                                                               |
|         |            | DSC_cancertype              | ④ CNCR_TUMOR_RGT | (Cancer Patient Registration)          | RGT     | WHERE RGT.VldYn = 'Y' --Data with valid cancer registration number<br>RGT.DelYn = 'N' --Data not deleted<br>RGT.DisqualifyYmd IS NULL OR RGT.DisqualifyYmd LIKE '' --Exclude entries with blank disqualification date<br>TRC.DeadYmd IS NULL OR TRC.DeadYmd LIKE '' --Condition to select entries with blank date of death for tumor number disqualification |
| JOIN    | INNER JOIN | YCDL_cancertype             | PT_BASIC         | (Patient Basic Information)            | PT      | ON MAIN.SvrId = PT.SVR_ID<br>MAIN.AlsUnitNo = PT.RESCH_REG_NO                                                                                                                                                                                                                                                                                                |
|         | LEFT JOIN  | DSC_cancertype              | CNCR_TUMOR_RGT   | (Cancer Patient Registration)          | DQLF    | ON MAIN.SvrId = DQLF.SvrId<br>MAIN.AlsUnitNo = DQLF.AlsUnitNo                                                                                                                                                                                                                                                                                                |
|         |            |                             |                  |                                        |         | WHERE RGT.VldYn = 'Y' --Data with valid tumor number<br>RGT.DelYn = 'N' --Data not deleted<br>RGT.DisqualifyYmd IS NULL OR RGT.DisqualifyYmd LIKE '' --Exclude entries with blank disqualification date                                                                                                                                                      |
|         | LEFT JOIN  | DSC_cancertype              | CNCR_TUMOR_TRC   | (Cancer Patient Follow-up Information) | DD_CD   | ON MAIN.SvrId = DD_CD.SvrId<br>MAIN.AlsUnitNo = DD_CD.AlsUnitNo                                                                                                                                                                                                                                                                                              |
|         |            |                             |                  |                                        |         | WHERE RGT.VldYn = 'Y' --Data with valid tumor number<br>RGT.DelYn = 'N' --Data not deleted<br>DD_CD.DeadYmd IS NULL OR DD_CD.DeadYmd LIKE '' --Condition to select death codes for blank date of death                                                                                                                                                       |
|         | LEFT JOIN  | DSC_cancertype              | CNCR_FRM         | (Document Form)                        | DD_CNTE | ON MAIN.SvrId = DD_CNTE.SvrId<br>MAIN.AlsUnitNo = DD_CNTE.AlsUnitNo<br>MAIN.FrmClnKey = DD_CNTE.FrmClnKey                                                                                                                                                                                                                                                    |
|         |            |                             |                  |                                        |         | WHERE DD_CNTE.AttrCd IN ('A0017954','A0017955','A0017956','A0021311') --Attribute codes related to cause of death<br>DD_CNTE.AttrValue NOT LIKE '' --Exclude entries with blank attribute values                                                                                                                                                             |
|         | LEFT JOIN  | DSC_cancertype              | CNCR_DX          | (Encounter/Diagnosis)                  | GB      | ON MAIN.SvrId = GB.SvrId<br>MAIN.AlsUnitNo = GB.AlsUnitNo<br>MAIN.AlsChosNo = GB.AlsChosNo --Research visit number related to death records                                                                                                                                                                                                                  |
|         | LEFT JOIN  | YCDL                        | YCDL_CODE_TBL    | (YCDL Code Table)                      | CD      | ON MAIN.SvrId = CD.SvrId<br>MAIN.DeadCd = CD1.CD_TYPE = 1270400 --Cause of death diagnosis code<br>MAIN.DeadResnCd = CD2.CD_TYPE = 9172600 --Cause of death code: cancer death, non-cancer death, unknown cause<br>MAIN.FrmCd = CD3.CD_TYPE = 9140600 --Form code                                                                                            |

**Supplementary Figure 8. Development of a SQL table-valued function to convert free text remote metastasis details to YCDL site codes, enhancing adaptability and simplifying ETL query management.** When converting the free text of remote metastasis details from source data to YCDL remote metastasis site codes, excessive CASE WHEN conditions were used. However, we developed a SQL table-valued function to extract the data. By managing the CASE WHEN conditions through a string management table, instead of modifying the ETL query whenever new string data occurs, we only need to add the new string to the management table. This approach increases adaptability to new data and simplifies ETL query management.

#### [Previous Method for Extracting Remote Metastasis Site Code Data]

- CASE WHEN in ETL query

```

241      --C15 Esophagus(식도의 악성신생물)
242      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Esophagus%' THEN '15'
243
244      --C16 Stomach(위의 악성신생물)
245      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Stomach%' THEN '16'
246
247      --C17 Small intestine(작은 창자의 악성신생물)
248      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Small intestine%' THEN '17'
249      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%duodenal%' THEN '17'
250      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%duodenum%' THEN '17'
251      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%ileum%' AND [value] COLLATE Korean_Wansung_CI_AS NOT LIKE '%pelvis%' THEN '17'
252      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%small bowel%' THEN '17'
253
254      --C18 Colon(결장의 악성신생물)
255      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Colon%' THEN '18'
256      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%appendix%' THEN '18'
257      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%sigmoid%' THEN '18'
258
259      --C19 Rectosigmoid junction(직장구불결장 이행부의 악성신생물)
260      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Rectosigmoid junction%' THEN '19'
261
262      --C20 Rectum(직장의 악성신생물)
263      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Rectum%' THEN '20'
264
265      --C21 Anus and anal canal(항문 및 항문관의 악성신생물)
266      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Anus and anal canal%' THEN '21'
267
268      --C22 Liver and intrahepatic bile ducts(간 및 간내 쓸개관(담관)의 악성신생물)
269      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Liver and intrahepatic bile ducts%' THEN '22'
270      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%liver%' THEN '22'
271      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%S4%' THEN '22'
272      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%S6%' THEN '22'
273      WHEN TransCd = '03' AND [value] COLLATE Korean_Wansung_CI_AS LIKE '%병원%' THEN '22' --타병원
274
275      --C23 Gallbladder(담낭의 악성신생물)
276      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Gallbladder%' THEN '23'
277      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%gallbladder%' THEN '23'
278      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%GB%' THEN '23'
279
280      --C24 Other and unspecified parts of biliary tract(기타 및 상세불명 담도 부위의 악성신생물)
281      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%Biliary tract%' THEN '24'
282      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%bile duct%' THEN '24'
283      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%biliary duct%' THEN '24'
284      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%CBD%' THEN '24'
285      WHEN [value] COLLATE Korean_Wansung_CI_AS LIKE '%common bile duct%' THEN '24'

```

#### [Method for Extracting Remote Metastasis Site Code Data]

- User-defined function UFN\_CaseMng\_Table in the ETL query

```

45      OUTER APPLY YCDLMNG.dbo.UFN_CaseMng_Table ('YCDL_CLRC', 'TE_MTST', 'TransCd', DSC.TransCd, DSC.TransEtc) ZZ

```

- String table referenced within the user-defined function

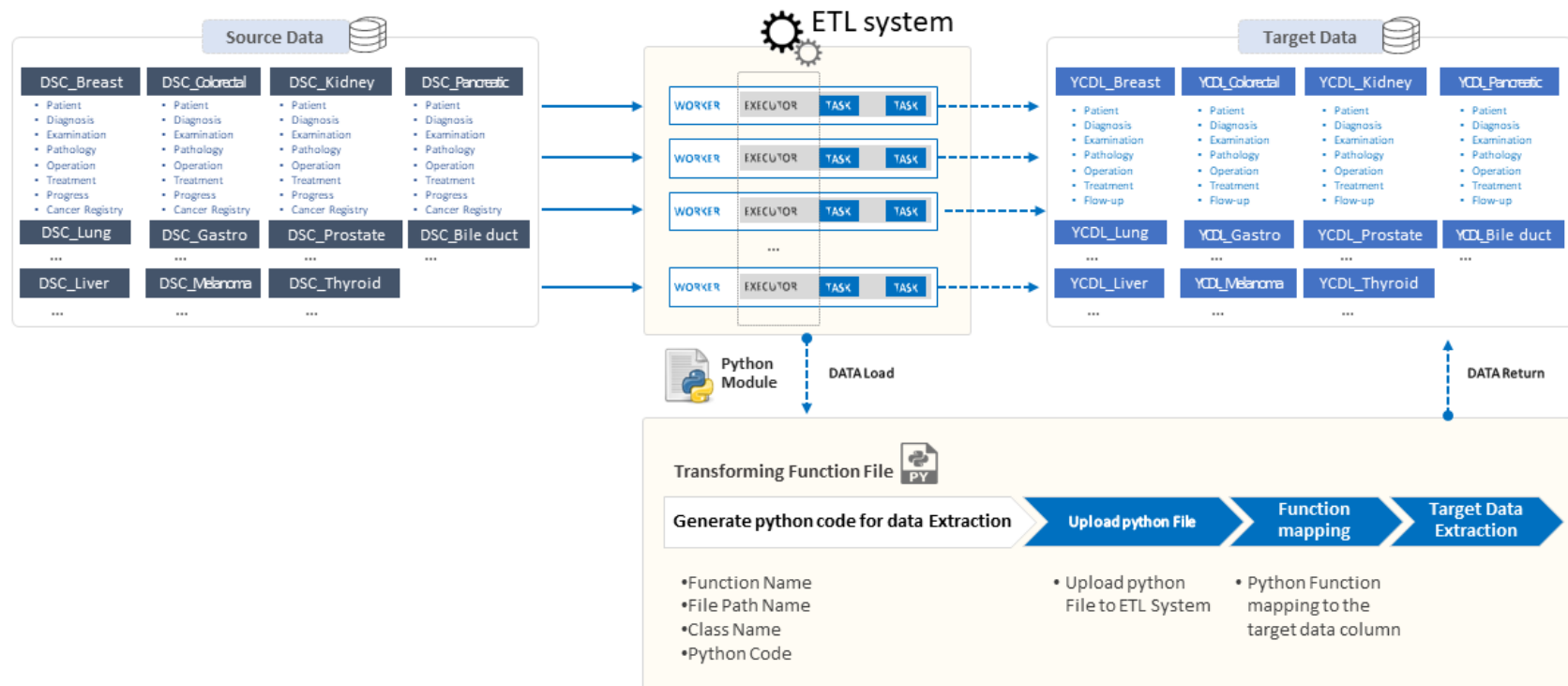
4 SELECT \* FROM DML\_CASE\_MNG (NOLOCK)

15 %

|    | CNCR_GB   | TARGET_TBL | ORV_COL  | ALS_COL | APP_RANK | CASE_STR | THEN_STR | APP_CMD | COLLATION_GB | ELSE_APP_RANK | ELSE_THEN_STR | SUB_CASE          | RMK1                 |
|----|-----------|------------|----------|---------|----------|----------|----------|---------|--------------|---------------|---------------|-------------------|----------------------|
| 1  | YCDL_CNCR | TE_MJTST   | TransEtc | TransCd | 1        | breast   | 42       | IN      |              | NULL          | NULL          |                   | Breast               |
| 2  | YCDL_CNCR | TE_MJTST   | TransEtc | TransCd | 2        | LN       | 63       | LIKE    | CS_AS        | NULL          | NULL          |                   |                      |
| 3  | YCDL_CNCR | TE_MJTST   | TransEtc | TransCd | 3        | Lymph    | 63       | LIKE    |              | NULL          | NULL          |                   | Lymph nodes          |
| 4  | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 4        | 01       | 63       | IN      |              | NULL          | NULL          |                   | 원격탐프절                |
| 5  | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 5        | 02       |          | IN      |              | 99999         | 75            | DML_CASE_MNG.SUB1 | DML_CASE_MNG.SUB1 조회 |
| 6  | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 6        | 03       | 22       | IN      |              | NULL          | NULL          |                   | 간                    |
| 7  | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 7        | 04       |          | IN      |              | NULL          | NULL          | DML_CASE_MNG.SUB1 | DML_CASE_MNG.SUB1 조회 |
| 8  | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 8        | 04       | 31       | IN      |              | NULL          | NULL          |                   | 폐                    |
| 9  | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 9        | 05       |          | IN      |              | NULL          | NULL          | DML_CASE_MNG.SUB1 | DML_CASE_MNG.SUB1 조회 |
| 10 | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 10       | 05       | 57       | IN      |              | NULL          | NULL          |                   |                      |
| 11 | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 11       | 06       | 47       | IN      |              | NULL          | NULL          |                   |                      |
| 12 | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 12       | 07       | 39       | IN      |              | NULL          | NULL          |                   | DML_CASE_MNG.SUB1 조회 |
| 13 | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 13       | 10       | 40       | IN      |              | NULL          | NULL          |                   |                      |
| 14 | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 14       | 12       | 61       | IN      |              | NULL          | NULL          |                   |                      |
| 15 | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 15       | 08       |          | IN      |              | NULL          | 999           | DML_CASE_MNG.SUB1 | DML_CASE_MNG.SUB1 조회 |
| 16 | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 15       | 09       |          | IN      |              | NULL          | 999           | DML_CASE_MNG.SUB1 | DML_CASE_MNG.SUB1 조회 |
| 17 | YCDL_CNCR | TE_MJTST   | TransCd  | TransCd | 15       | 11       |          | IN      |              | NULL          | 999           | DML_CASE_MNG.SUB1 | DML_CASE_MNG.SUB1 조회 |

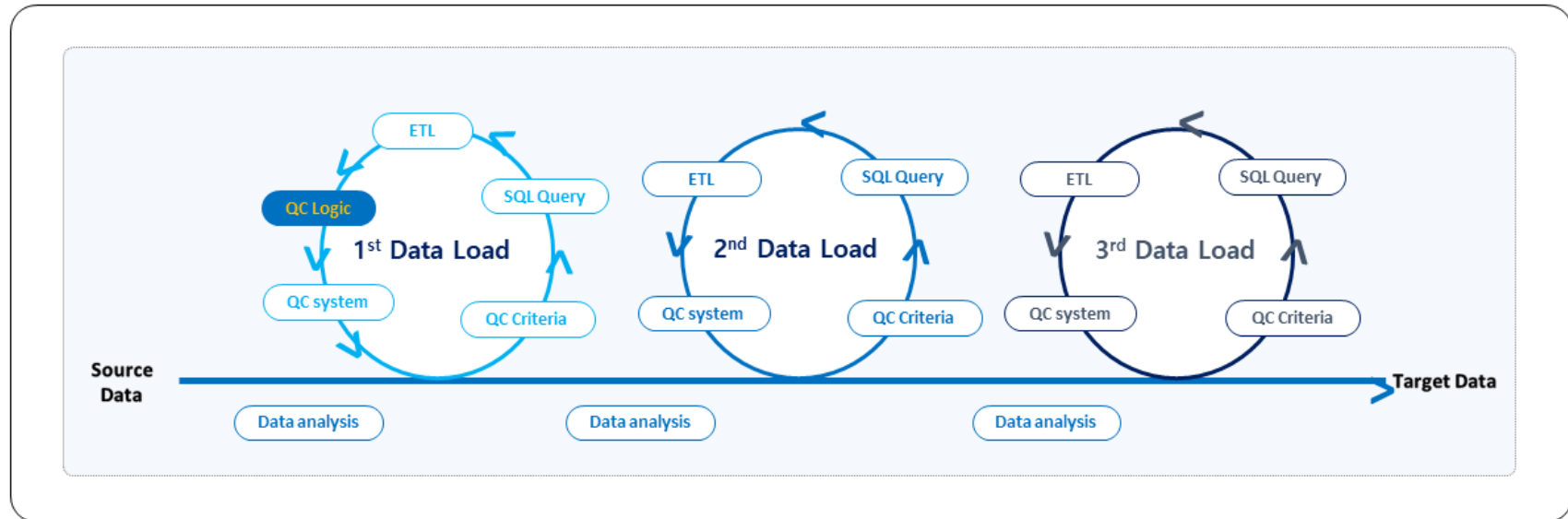


**Supplementary Figure 10. Overview of Python script-based operations for complicated elements.** For certain datasets requiring intricate extraction protocols, bespoke ETL strategies were devised using Python scripts crafted for each specific operation.

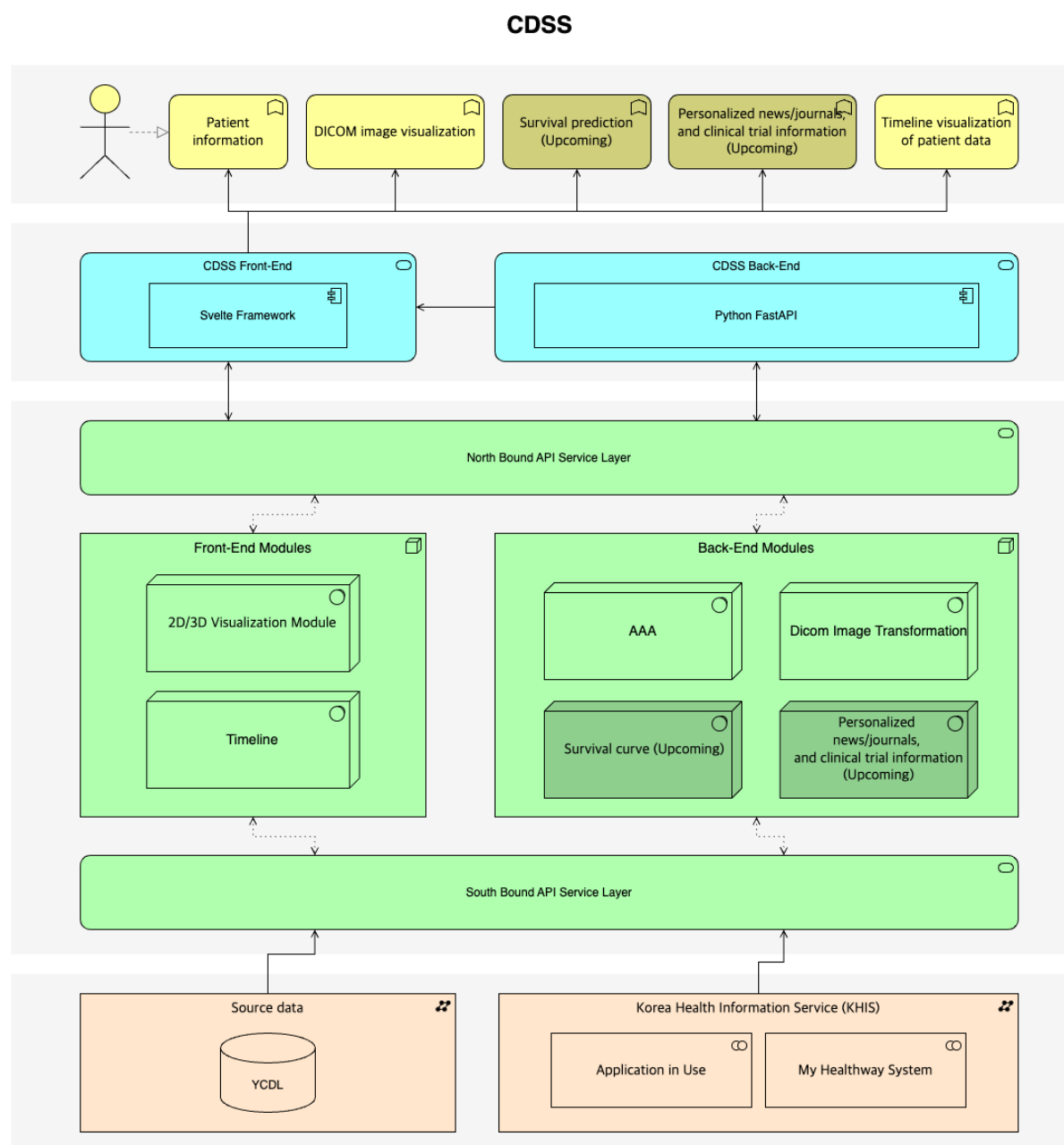




**Supplementary Figure 12. Continuous QC steps within distinct closed-loop systems.** A series of data quality control steps gradually enhanced the accuracy of the cleansed target data with minimal intervention.



**Supplementary Figure 13. Architecture of the developed clinical decision support system (CDSS).** The CDSS User Interface (UI) front-end components include: (1) patient information, (2) DICOM image visualization for PACS-integrated 3D tumor display, and (3) a longitudinal view of the complete patient journey with a timeline of key events and data points. Additional components for survival prediction, personalized news/journals, and clinical trial information are under development. The CDSS Front-End, built using Svelte and React, provides an interactive UI and visualizes data received from the Back-End. Key functionalities include user interaction management and data visualization. The CDSS Back-End, implemented with FastAPI, processes data requests from the Front-End and performs computations. It handles RESTful API requests, DICOM data processing, AI-based survival analysis, web scraping services, and visualizes DICOM data using the VTK library. The Front-End Modules include a 2D/3D Visualization Module and a Timeline Module. The Timeline Module arranges patient data chronologically, supporting zooming and dragging interactions. The Back-End Modules include an AAA Module for Authentication, Authorization, and Accounting using JWT, and a DICOM Image Transformation Module for converting DICOM images into VTK images using the VTK library.





(**Note:** Modules depicted with a darker shade are under development for future integration and are not currently available for use.)

**Supplementary Table 1. Accuracy and missing data rates of segmented data in YCDL from surgical and molecular pathology for various cancer types**

| Type                | Cancer Type          | Number of features | Accuracy (%) | Missing rate (%) |
|---------------------|----------------------|--------------------|--------------|------------------|
| Surgical Pathology  | YCDL_BRST_16_PH_SRGC | 32                 | 91.48        | 4.86             |
|                     | YCDL_CLRC_16_PH_SRGC | 28                 | 89.00        | 9.07             |
|                     | YCDL_LUNG_16_PH_SRGC | 23                 | 98.08        | 0.96             |
|                     | YCDL_GSTR_16_PH_SRGC | 23                 | 92.69        | 6.34             |
|                     | YCDL_LVER_16_PH_SRGC | 32                 | 98.75        | 0.47             |
|                     | YCDL_MLNM_16_PH_SRGC | 26                 | 86.54        | 10.69            |
|                     | YCDL_KDNY_16_PH_SRGC | 26                 | 95.87        | 2.07             |
|                     | YCDL_PRST_16_PH_SRGC | 20                 | 95.40        | 1.70             |
|                     | YCDL_THRD_16_PH_SRGC | 28                 | 91.14        | 1.14             |
|                     | YCDL_PNCT_16_PH_SRGC | 22                 | 92.45        | 7.18             |
| Molecular Pathology | YCDL_BLDT_16_PH_SRGC | 33                 | 89.49        | 0.49             |
|                     | YCDL_BRST_17_PH_IMML | 16                 | 98.37        | 1.38             |
|                     | YCDL_CLRC_17_PH_IMML | 16                 | 100.00       | 0.00             |
|                     | YCDL_LUNG_17_PH_IMML | 10                 | 94.60        | 5.40             |
|                     | YCDL_GSTR_17_PH_IMML | 10                 | 94.60        | 5.40             |
|                     | YCDL_LVER_17_PH_IMML | 11                 | 98.72        | 0.55             |
|                     | YCDL_MLNM_17_PH_IMML | 16                 | 95.38        | 3.75             |
|                     | YCDL_KDNY_17_PH_IMML | 15                 | 98.67        | 0.40             |
|                     | YCDL_PRST_17_PH_IMML | 9                  | 92.00        | 8.00             |
|                     | YCDL_THRD_17_PH_IMML | 11                 | 99.45        | 0.36             |
|                     | YCDL_PNCT_17_PH_IMML | 13                 | 98.92        | 0.00             |
|                     | YCDL_BLDT_17_PH_IMML | 13                 | 99.85        | 0.00             |