**Positive Samples (7259)** = ‘AKI\_features\_60\_cat.txt’ && ‘AKI\_features\_60\_mix.txt’

**Negative Sample (69698)** = ‘NONAKI\_features\_60\_cat.txt’ && ‘NONAKI\_features\_60\_mix.txt’

**Format:**

<demo>|<vitals>|<lab>|<DRG>|<comorbidity>|<imed>|<omed>|<icd>|<AKI\_label>

|  |  |  |
| --- | --- | --- |
| **Feature Group** | **Size** | **Format** |
| Demographics | 3 | age\_race\_gender |
| Vital signs | 5 | bmi\_diastolic\_systolic\_pulse\_temp |
| Lab | 14 | lab1\_lab2\_lab3\_...\_lab14  lab1 = val1,date1;val2,date2;… |
| UHC DRG | 315 | dx1\_dx2\_… |
| Inpatient Medication | 1271 | med1;d1,d2,d3\_med2;d1,d2,d3;… |
| Outpatient Medication |
| ICD9 grouped with CCS | 280 | ccs1;d1,d2,d3\_ccs2;d1,d2,d3;… |
| AKI\_stage | 1 | label,date (label = 0, 1, 2, or 3) |

\*\*\*\*\*NOTE\*\*\*\*\*

If a variable does not exist, then it is ‘0’.

For example, vital = '0\_0\_0\_0\_0' means that all five vital signs were missing/not measured

Otherwise, it is represented as a string of [value,date] pairs separated by ‘;’ for multiple measurements (described above).

For medications and icd codes, a ‘0’ means that there are no recordings in EMR. If it has at least 1 recording, it will be formatted as ‘medIndex1;date1,date2,…\_medIndex2;date1,date2,…\_’

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Categorical Values for the Features:**

|  |  |
| --- | --- |
| **Variables** | **Variable Value** |
| Age at Admission |  |
| 18 – 25 | 1 |
| 26 – 35 | 2 |
| 36 – 45 | 3 |
| 46 – 55 | 4 |
| 56 – 64 | 5 |
| Gender |  |
| Female | 1 |
| Male | 2 |
| Race |  |
| White | 1 |
| African American | 2 |
| Asian | 3 |
| Other | 4 |
| Vitals |  |
| BMI |  |
| Unknown | 0 |
| < 18.5 Underweight | 1 |
| 18.5 – 24.9 Normal | 2 |
| 25.0 – 29.9 Overweight | 3 |
| > 30.0 Obese | 4 |
| Temperature |  |
| Unknown | 0 |
| < 95.0 Hypothermia | 1 |
| 95.0 – 97.7 low body temperature | 2 |
| 97.7 – 99.5 Normal | 3 |
| 99.5 – 104.0 Fever | 4 |
| > 104.0 Hyperpyrexia | 5 |
| Pulse (resting) |  |
| Unknown | 0 |
| < 50 | 1 |
| 50 – 65 Excellent | 2 |
| 66 – 80 Average | 3 |
| 81 – 100 Poor | 4 |
| > 100 | 5 |
| BP (Diastolic / Systolic) |  |
| Unknown | 0 |
| < 120 / < 80 Normal | 1 |
| 120 – 139 / 80 – 89 Prehypertension | 2 |
| 140 – 159 / 90 – 99 Stage 1 hypertension | 3 |
| > 160 / > 100 Stage 2 hypertension | 4 |
| Laboratory Test |  |
| Unknowns | 0 |
| Abnormal values | 2 |
| Albumin: 3.5 – 5.5 | 1 |
| ALT: 7 – 56 | 1 |
| AST: 10 – 40 | 1 |
| Ammonia: 10 – 65 | 1 |
| Calcium: 8.5 – 10.2 | 1 |
| BUN: 7 – 20 | 1 |
| Total Bilirubin: 0.3 – 1.9 | 1 |
| CK-MB: Male < 6.7; Female < 3.8 | 1 |
| CK: Male 52 – 336; Female 38 – 176 | 1 |
| Glucose: 80 – 160 | 1 |
| Lipase: 0 – 400 | 1 |
| Platelets: 150 – 450 | 1 |
| Troponin I: 0 – 0.06 | 1 |
| WBC: 3.5 – 10.5 | 1 |
| Medications (medications\_60.txt) | Feature file variable index starts from 0 |
| Absence | 0 |
| Presence | 1 |
| Admission Diagnosis (adm\_dx\_mod.csv) | Feature file variable index starts from 0 |
| Absence | 0 |
| Presence | 1 |
| ICD9 codes grouped with CCS (ccs\_codes.txt) | Feature file variable index starts from 0 |
| Absence | 0 |
| Presence | 1 |

\*\*\*\*\*NOTE\*\*\*\*\*

In the feature vector file, e.g., AKI\_features\_60\_cat.txt, indices for medications, admission DRG, and icd9 codes starts from 0.

The ‘csv’ files have a header line for column labels, need to ignore. Thus, in case of the csv files, the 2nd row corresponds to the variable index 0 in the feature vector file.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*