**Readme Instructions**

**Citation to the original paper**

COAM - https://ieeexplore.ieee.org/document/8761846

RETAIN - https://dl.acm.org/doi/pdf/10.5555/3157382.3157490

**Link to the original paper’s repo (if applicable)**

NA

**Dependencies**

1. Python 3.8.12
2. Theano 1.0.5
3. Pickle
4. Numpy 1.21.2
5. Sklearn
6. CUDA / CUDAToolkit 11.31.1

**Data download instruction**

We used MIMIC-III clinical dataset from Physionet website.

**Preprocessing code + command (if applicable)**

python process\_mimic.py ADMISSIONS.csv DIAGNOSES\_ICD.csv PATIENTS.csv mimic3

(In the above command, mimic3 is the name for the files which is appended to the output file extensions that will be created by the process, as shown below)

This command will generate the following cpickled files

mimic3.3digitICD9.seqs, mimic3.3digitICD9.types, mimic3.dates, mimic3.morts, mimic3.pids

mimic3.seqs, mimic3.types

**Training code + command (if applicable)**

python retain.py mimic3.3digitICD9.seqs 942 mimic3.morts C:\Users\sathya\MCS-DS\cs598-DLH\Retain\retain\output1.out --simple\_load --n\_epochs 50 --keep\_prob\_context 0.8 --keep\_prob\_emb 0.5

942 - number of the entire 3-digit ICD9 codes used in the dataset

50 - Number of epochs

0.8 - instruction for the model to reach AOC above 0.8

0.5 - Dropout layer probability

**Evaluation code + command (if applicable)**

python test\_retain.py output1.out.7.npz mimic3.3digitICD9.seqs mimic3.morts mimic3.3digitICD9.types runoutput1.out

output1.out.7.npz - The best model chosen by training the model

runoutput1.out - output file to which logs will be written to

**Pretrained model (if applicable)**

NA

**Table of results (no need to include additional experiments, but main reproducibility result should be included)**

Excel file attached below

****

Table

Description automatically generated