

.\Kevin Lu\Code\genSOPData_2023_MIMIC-II directory in one drive under
"BPR01AlgorithmGroup"

Kevin Lu, November 21st 2023

The directory now contains:

1. An edited `mimic_read_matFile.m`(which is updated to process the `matFile` cells one by one)
2. `mimic_save_file.m` (which saves data from the feature extraction to a `mat` file)
3. `mimic_individual_preprocess_saveFeatures_20222023data(matFileFullPath, ECG, Finapres, NIRS)` which takes in a single sample's data and extracts the features

The new pipeline is:

`mimic_test_pipeline => mimic_read_matFile => for i = 1:numberCellsToProcess, call
mimic_individual_preprocess_saveFeatures_20222023data on sample => mimic_save_data.m`

Kevin Lu, November 8th 2023

This directory contains the MATLAB programs for generating SOP Data for 2023 MIMIC-II data,
As well as information on the formatting and other characteristics of the MIMIC-II Data

1) This `readme.txt`, describes the usage and reference information for this functionality

1a) To generate SOP data for MIMIC-II Directory:

1. Download "part_1.mat" from Data(Sample Input and Output folder)
2. Download `mimic_read_matFile.m` and
`mimic_preprocess_saveFeatures_20222023data.m`
3. Download "mimic_test_pipeline.m"
4. Run `mimic_test_pipeline.m` in a MATLAB environment that contains the 20172018 lab
data/MIMIC-I code(I.E. QuanCodes, Qiao, yshi folders)

2) **The main function for reading and processing a single .mat file:**

`mimic_read_matFile.m`

a) It requires a single matlab file path as input

% Syntax: `mimic_read_matFile(matFileFullPath)`

% Inputs: `matFileFullPath` -- input .mat file full path including path and filename

% Because the data is already preprocessed, section start and end removed

b) See `mimic_test_pipeline.m` for testing (just called on "part_1.mat")

3) **The main function for processing a single matlab file is:**

`mimic_preprocess_saveFeatures_20222023data.m`

% Syntax: [`preprocessDataMatFileFullPath`, `featuresMatFileFullPath`, `segDataOutFileDir`]

= `preprocess_saveFeatures_20222023data(matFileFullPath)`

%

% Inputs: `matFileFullPath` -- input .mat file full path including path and filename

%

```
% Outputs: preprocessDataMatFileFullPath -- preprocess data output .mat file full path
%         featuresMatFileFullPath -- features output .mat file full path
%         segDataOutFileDir -- segmented data output directory name
```

3a) Main Edits to preprocess_saveFeatures code:

```
% 1) Kevin Lu October 16, 2023 Initial adaption to MIMIC-II dataset
% a) Changed method header(removed section start and end as the data has
% been preprocessed already and there are no manual data collection errors)
% b) Adapted data channels to MIMIC Data Formatting
% bi) Created appropriate ECG, Finapres, and PPG column vectors from
% MIMIC-II part1.mat file
% c) Changed fsh = 125Hz from 250Hz
% d) Removed STAT Calculations + outputs
%
% 2) Kevin Lu October 23, 2023 Removed some unnecessary preprocessing
% a) Commented out '60Hz_norch_FIR_for_250.mat' ECG filter
% b) Also commented out 60 Hz power line interference noise from Finapres
% signal
% c) Commented out Finapres A/D to mmHg calculation(BP data already in
% mmHg)
```

4) I only call mimic_test_pipeline.m to output SOP data for MIMIC-II

4a) If the other files in the SOP data pipeline are setup(I.E. from 2017 MIMIC-I data), then the code should run using just mimic_read_matFile.m and mimic_preprocess_saveFeatures_20222023data.m

5) Information on MIMIC-II dataset:

In each matrix each row corresponds to one signal channel:

1: PPG signal, FS=125Hz; photoplethysmograph from fingertip

2: ABP signal, FS=125Hz; invasive arterial blood pressure (mmHg)

3: ECG signal, FS=125Hz; electrocardiogram from channel II

%MIMIC-II Data format:

```
% 1x1000 cell array ('p')
% Each cell is another 3x61000 array
% I.E.
% 1. NIRS(PPG) Data: -0.06 -0.07 -0.07...
% 2. Finapres(BP): 67.06, 69.35, 75.36...
% 3. ECG: 1.7595, 1.7185, 1.6843...
```

The MIMIC dataset is split into 10 parts, with a single part containing ~3,578,000 data points. I thought that this data would be enough and did not attempt to combine all 10 parts of the dataset and run the SOP code on it.