ECG Classification for Arrhythmia Detection

Files

- beats_nsvfq_48p.csv
  - **Students can ignore this file**
  - The complete dataset.
    - Each row contains 250 values of ECG beat (the first 250 columns in the same row)
    - The last entry in each row is the label in ASCII (e.g. 78 in ASCII means N which denotes arrhythmia type).
  - We have already created train/validation/test sets from it.

- Other three files
  - **Students should use these files**
    - train_data.csv for training
    - val_data.csv for hyperparameter tuning
    - test_data.csv for testing/reporting the final results
  - Structure
    - Each row contains 250 values of ECG beat (the first 250 columns in the same row)
    - The last entry in each row is its numeric label (e.g. 0 means N which denotes arrhythmia type)
      - ASCII labels are replaced with numeric labels as given on next slide.

Labels

Arrhythmia Class	ASCII Label (Used in beats_nsvfq_48p.csv)	Numeric Label (Used in all other csv files)
N	78	0
/	47	1
L	76	2
R	82	3
e	101	4
j	106	5
A	65	6
a	97	7
J	74	8
S	83	9
E	69	10
F	70	11
V	86	12
f	102	13
Q	81	14

Labels

- train_data.csv, val_data.csv and test_data.csv
  - Contain numeric labels (0 to 14) and NOT ASCII labels

- Students are free to modify the labels in any way if needed
  - Specify which arrythmia class is assigned to which label.

Data Pre-processing

- The following steps are already done for the provided csv data
  - QRS peak detection
  - Beat segmentation
  - Splitting the data into train/val/test sets
  - Task-1 and Task-2.1 in the PDF are already covered here
    - **Students do not need to do these**

- Students are free to use any additional data pre-processing methods if needed.

Rare Classes

- We are aware that some classes have very less data.
  - Try to achieve as much accuracy,f1 score etc as you can on those classes using existing data.
  - We will take into account that performance on those classes gets hampered by lack of enough data and we will assess your final submission accordingly.

- Workarounds for reporting results in presence of such rare classes
  - Report the confusion matrix
  - Report metrics like accuracy/f1 score/sensitivity etc. for each class separately
  - Report aggregate metrics (e.g. overall f1-score) using weighted averages across classes.

Thank You!

Questions?