411 Assignment 2 thoughts:

* Highly unbalanced classes
* In paper, they ignore the Q AAMI class since “it is practically nonexistent”.
* Only MLII provided by all records (so let’s use that?)
* Perform some sort of pre-processing (in the paper they use baseline removal, others use high-frequency noise filtering too).
* In the paper, they use the labels for segmentation, as they focus on classification.
* In the paper, they standardize the data
* In the paper, they add weights to the classes to help deal with imbalance.
* Can’t use accuracy (imbalanced classes). In the paper they use jk index (which we could use to allow for comparison).
* <https://archive.physionet.org/physiobank/annotations.shtml> annotations explained
* <https://i.imgur.com/mFdm9U9.png> standard classes from annotations
* <https://www.scopus.com/inward/record.url?eid=2-s2.0-0034953193&partnerID=10&rel=R3.0.0> more on the dataset

Progress:

* Data processing script
* Do we use all samples, or just ones with corresponding annotations?

test

A more complex issue. The existing system must be setup to ignore ‘brown’ that is touching ‘green’, so that the end of the stem is not thought to be a defect. Unfortunately, this also means the crack behind the stem is ignored.