Notebook

September 8, 2021

1 EDA

1.1 Overview

The data I am using come from the MIT-BIH Atrial Fibrillation Database. It contains annotated ECG data from 23 unique patients each with two simultaneous 10 hour ECG signals. The annotations detailing the rythym at a given time are given in a record's `.atr` files. There are 4 different types of annotations for hearth rythyms: Atrial Fibrilation (AFIB), Atrial Flutter (AFL), AV Junction (J), and "all other rythyms" (N). The dataset mostly captures AFIB and N rythyms while AFL and J only make up around 1% of the data.

```
<pandas.io.formats.style.Styler at 0x7fd4c5b47e20>
```

I am going into this project with little to no background knowledge on how ECGs work and what characteristics each rythym tends to display. Using the data, I'm going to hypothesize the characteristics of each rythym then compare my hypotheses with existing literature.

1.2 Time Domain Analysis

First, I've curated a small random sample that has decent variation (My sampling method struggles with AFL and J due to the low number of records that contain these rythyms). I have plotted these for a surface level comparison.
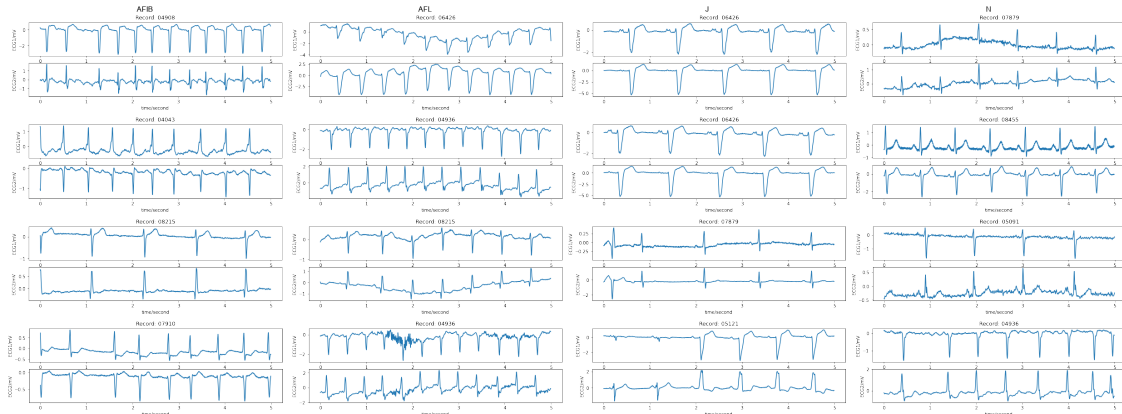
1.2.1 Similarities

- There seems to be a common pattern to each beat:
    1. First comes lead-up activity: a small positive mound just before a beat.
    2. The beat itself (characterized by either a large positive spike, a large negative spike, or small spike quickly followed by a large spike).
    3. Follow-up activity: a medium-sized positive mound immediately after a beat
- Signals tend to return to 0 mV or slightly below 0 mV between beats.

1.2.2 Differences

- Nothing particularly stands out in this view of the data

1.3 Frequency Domain Analysis

Since nothing really stood out between the different rythms, perhaps another view of the data might prove more informative. Using the same sample, I plotted their Discrete Fourier Transforms and noted some observations below.

AFIB

AFL

J

N

- Hard to notice distinct harmonic banding • Noisy †

- Clear harmonic banding • High frequency fundamental

- Sometimes strong harmonic banding • Low frequency fundamental • Slightly noisy †

- Occasionally strong harmonic banding • Records 05091 and 04936 are a little noisy † • Record 08455 has 60 Hz noise(probably a product of the data capture process)

† I should be careful when I say "noisy". If you actually look at the signals they derived from, they aren't so noisy per say. However, in the frequency domain, it is hard distinguished harmonic spikes like we see in the other signals.
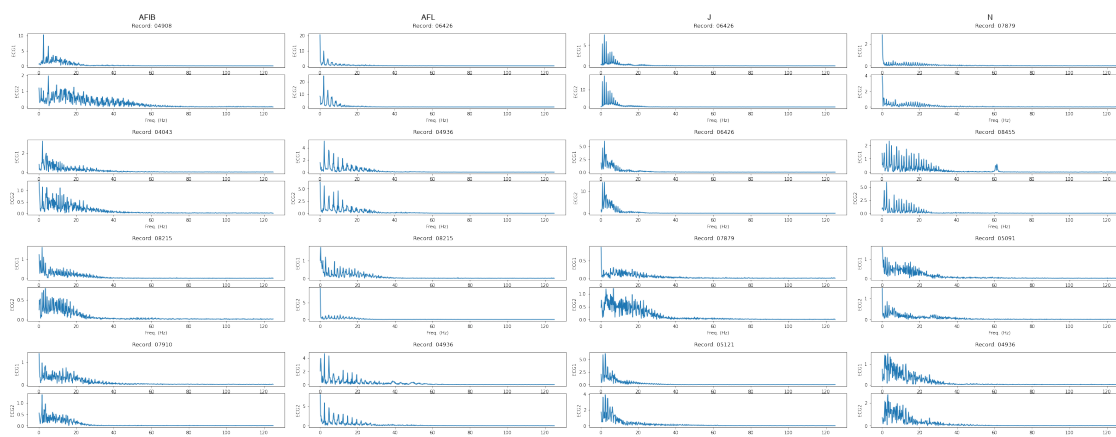
1.3.1 Aggregated FDA

To do a more general comparison, I took 200 samples for each class, applied a Discrete Fourier Transform, and averaged their values to generate the plot below.
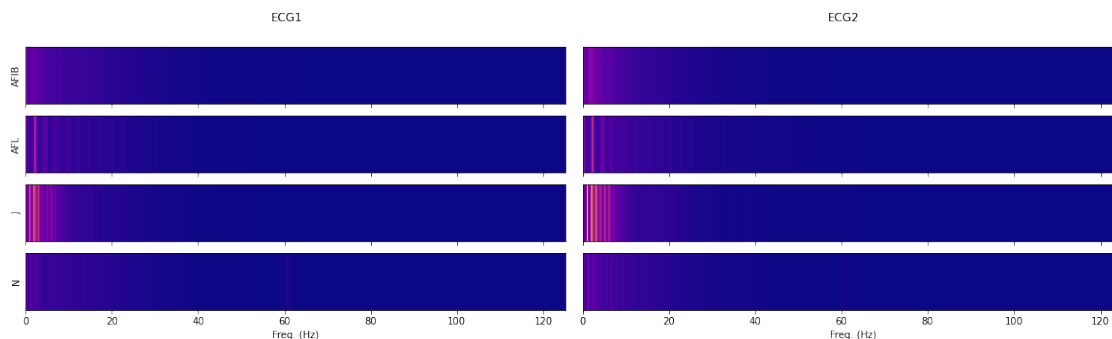
The only real distinction I can make is sharpness in the banding for each rythym. However, I believe this doesn't really inform us much about the characteristics of each class. We can expect banding as the result of the harmonics produced by the beat rythym. The sharpness of this banding is likely a result of the variance in each label. AFL and J will have low variance due to the limited number of samples available for them while AFIB and N have higher variance since they have a larger sample pool. Thus, AFL and J appear "sharper" while AFIB and N are more "fuzzy".

There is a slight contradiction though. AFIB and N do not follow this pattern. N has more sample availability than AFIB, having about 20% more data available by duration and appearing in more unique occasions that last longer than 5 seconds. Still, AFIB appears to have more variance making it hard to distinguish any clear harmonics. Meanwhile, N has some pretty evident harmonic banding.

One explanation could be due to AFIB appearing in 23 records while N only appears in 21. If this is the case, then I theorize that N's appearance in this plot will become more fuzzy as I increase the sample size for each class. I tested this at two other sample sizes. At 1,000, I observed slightly more variance in both N and AFIB, but harmonic banding remained distinguishable in N. At 10,000 there was no noticable difference than that at the 1,000 level. From this experiment, I can't conclusively say that this isn't the case.

Another explanation could be that AFIB rythyms simply have more variation in BPM resulting in this plot appearing less sharp.

One final explanation might be more clear if we reference the previous frequency domain plot. Recall how I noted that the AFIB plots had indistinguishable spikes. This may indicate the AFIB signals are more eratic resulting in their aggregates appearing as they do.