Data Appendix

Summary of established data set:

The data used in this project is sourced from a clinical study where echocardiography has been performed on patients with varying degrees of aortic stenosis (1). The dataset includes 40 images of grayscale Doppler signals through the aortic valve in png and jpeg form, each with an overlaid vertical velocity scale bar. To access the dataset, they are stored in a shareable google drive folder. The velocity scale bars on the images have alphabetical and numerical values that consist of:

- 200
- 100
- cm/s
- -100
- -200
- -300
- -400
- "##"bpm

With not all of the scale bars containing this full range, some being shorter, starting at "100" or "cm/s", and ending after "-300." Some may even be in a m/s scale, having values of 1.0, m/s, -1.0, etc.

Questions explored:
1. Is the initial OCR output accurate?
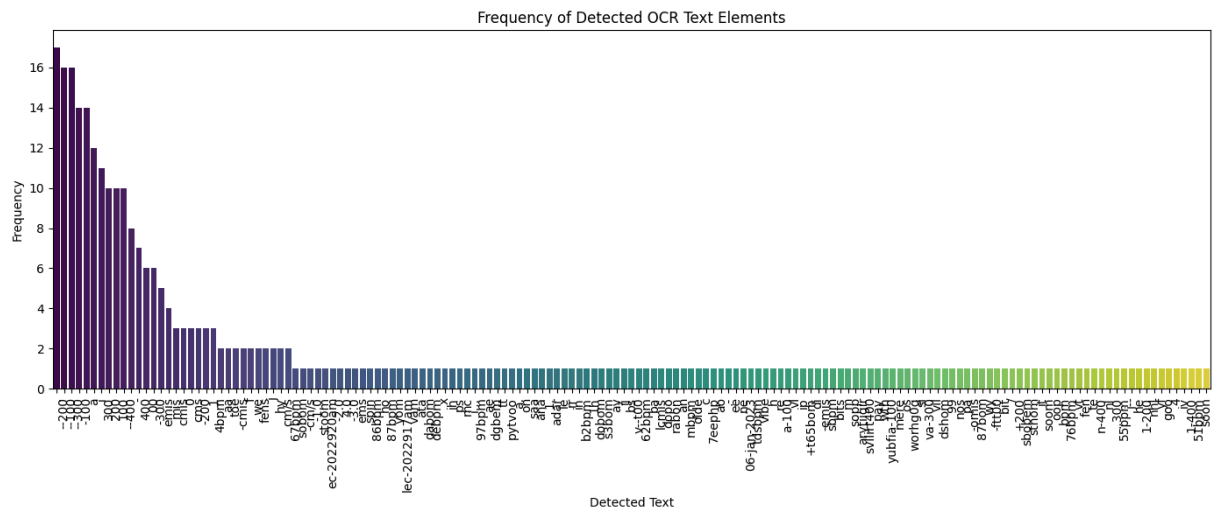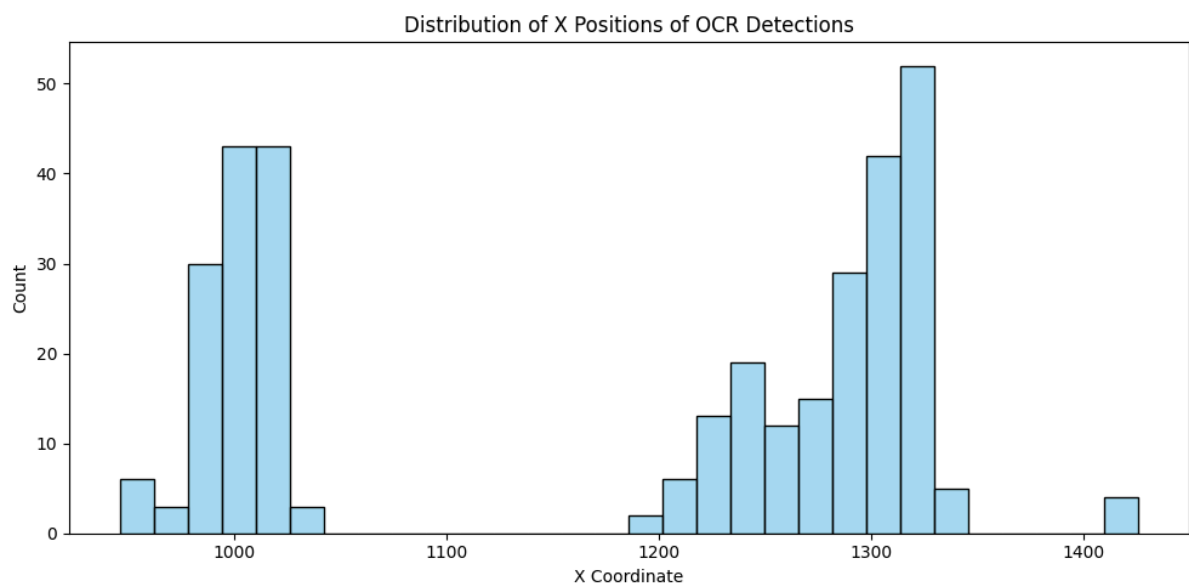2. What is the frequency of text elements?

Figure 1

Figure 2