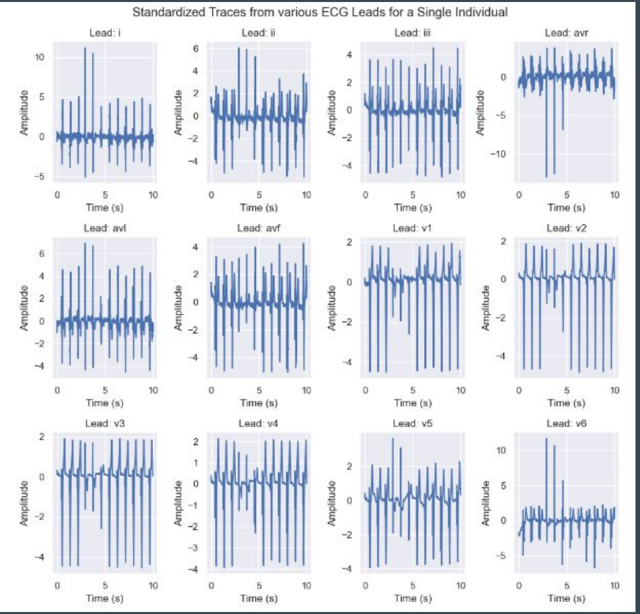
CS109B Final Project: Predicting Cardiac Pathologies from ECG data

Justin Adjasu, George Popoola, Brian Nzuki

Problem Statement

Can a model trained on ECG data accurately predict whether an individual possesses different cardiac conditions? If so, what can we learn from such a model's predictions?

We'll be using coupled ECG & demographic data from Lobachevsky University. An example of the raw ECG data is shown at right.



Preprocessing

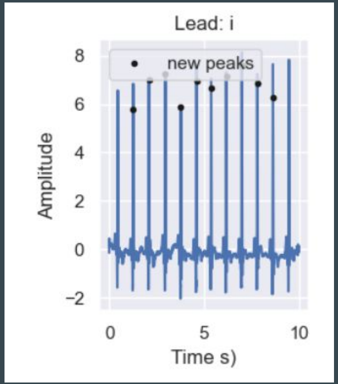
- In our demographic data, each disease was collapsed to a binary variable indicating the presence or absence of the condition, eliminating subtypes
  - This was done to combat the problem of extreme class imbalance among subtypes
  - Note: NaNs were used to encode disease absence in the original data

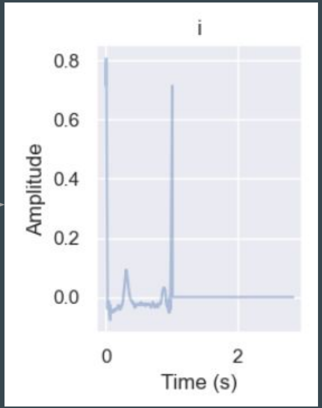
	ID	Sex	Age	Rhythms	Electric axis of the heart	Conduction abnormalities	Extrasystolies
0	1	F\n	51\n	Sinus bradycardia	Electric axis of the heart: left axis deviation	NaN	NaN
1	2	M\n	64\n	Sinus rhythm	Electric axis of the heart: normal	NaN	NaN
2	3	M\n	53\n	Sinus rhythm	Electric axis of the heart: vertical	NaN	NaN
3	4	M\n	56\n	Sinus rhythm	Electric axis of the heart: left axis deviation	Incomplete right bundle branch block	NaN
4	5	M\n	61\n	Sinus rhythm	Electric axis of the heart: horizontal	NaN	NaN

	ID	Sex	Age	Rhythms	Conduction abnormalities	Extrasystolies	Hypertrophies
0	1	1	51	1	0	0	1
1	2	0	64	0	0	0	1
2	3	0	53	0	0	0	1
3	4	0	56	0	1	0	1
4	5	0	61	0	0	0	1

Preprocessing

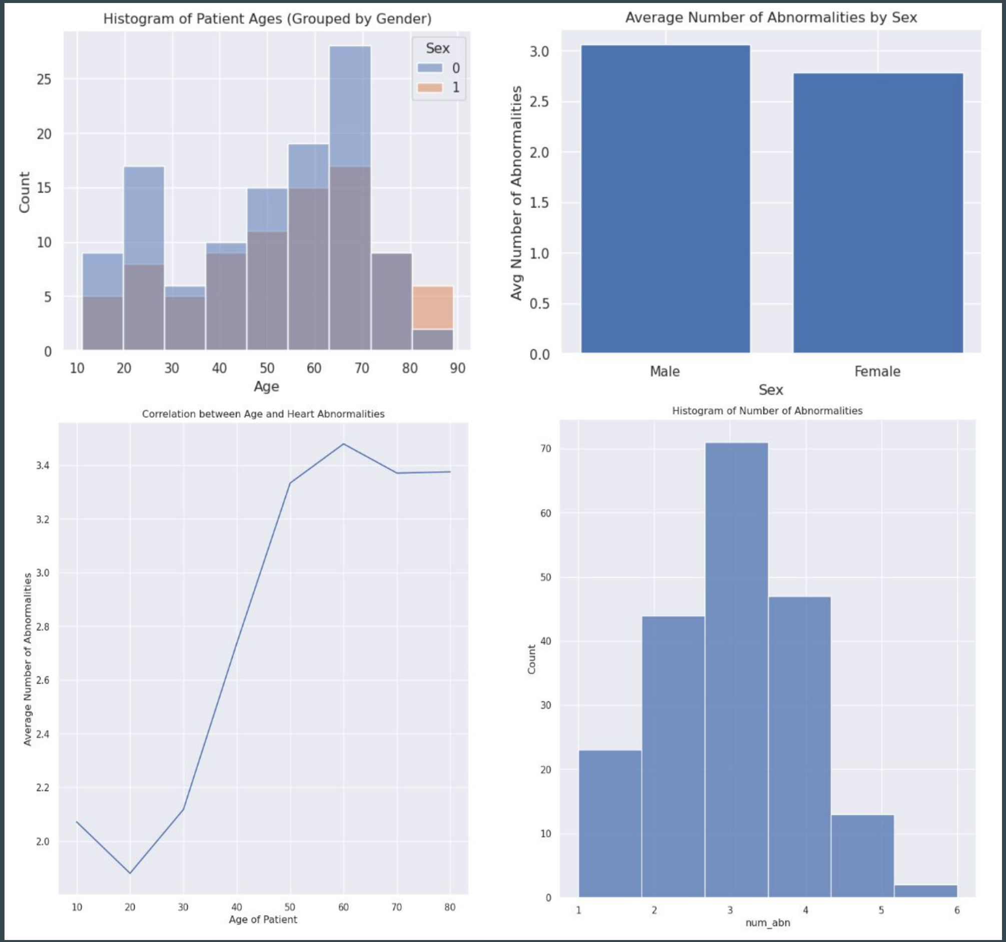
- Each patient's 10s continuous 12-lead ECG recording was broken up into each individual full heartbeat wave observed
  - This allowed us to increase the number and diversity of observations available for training, and is consistent with techniques used in studies doing similar classification work using ECG data
  - As a result, each ECG observation is a 12x1423 matrix





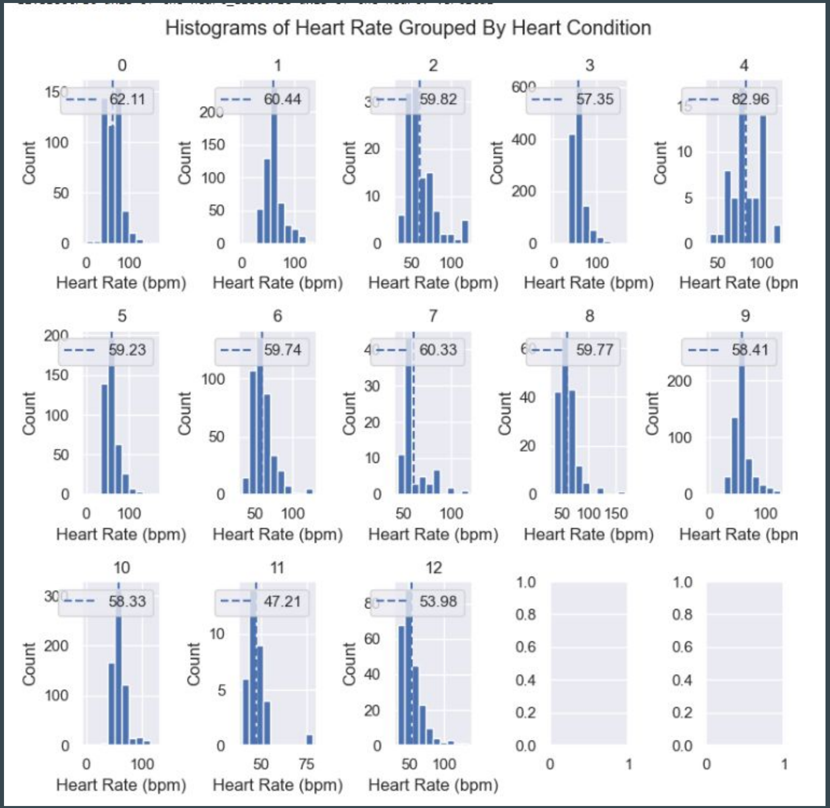
Exploratory Data Analysis

- Avg. number of abnormalities generally increase with age (peaks around 60)
- Most patients report 3 abnormalities on average
- Males report higher numbers of abnormalities on average
- Will evaluate effect of these occurrences on training



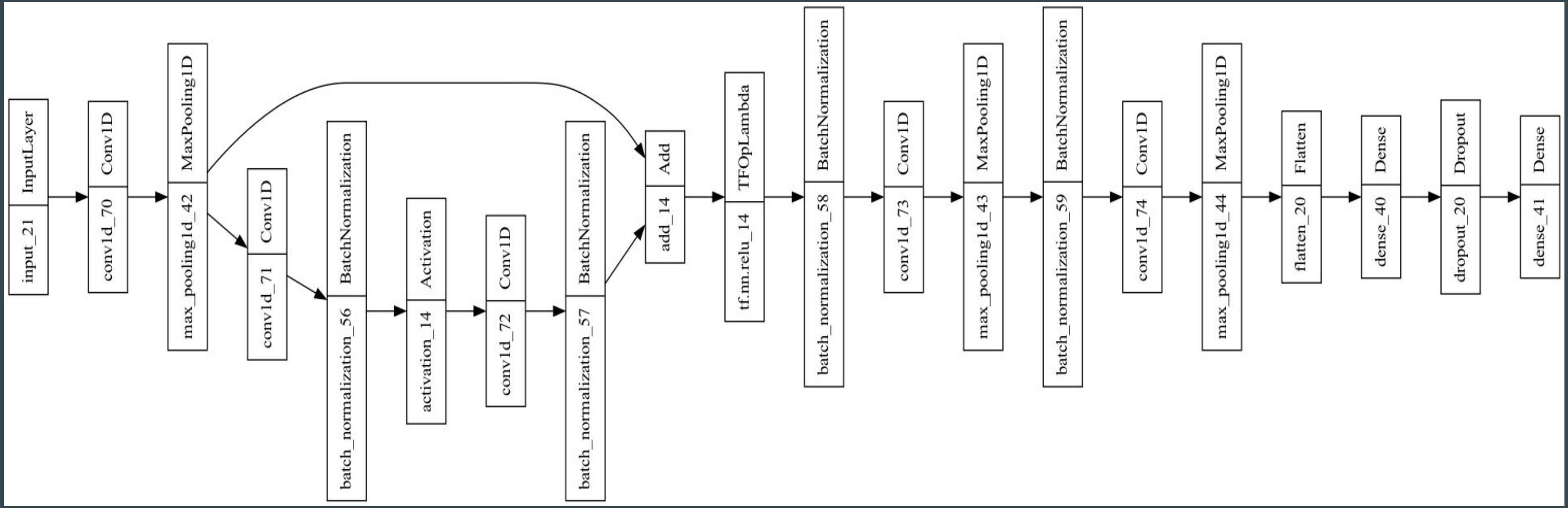
Exploratory Data Analysis

- Also did EDA of key standard metrics available from ECG trace
- Example at right shows heart rate
  - Clear differences in means between classes, indicates this could be one of the many ECG trace features the model picks up on in training



Baseline Model Definition and Training

- Makes use of 1D Convolution layers, MaxPooling, etc., all done across the time dimension of each lead
    - For a more in-depth understanding of full model architecture see Jupyter Notebook



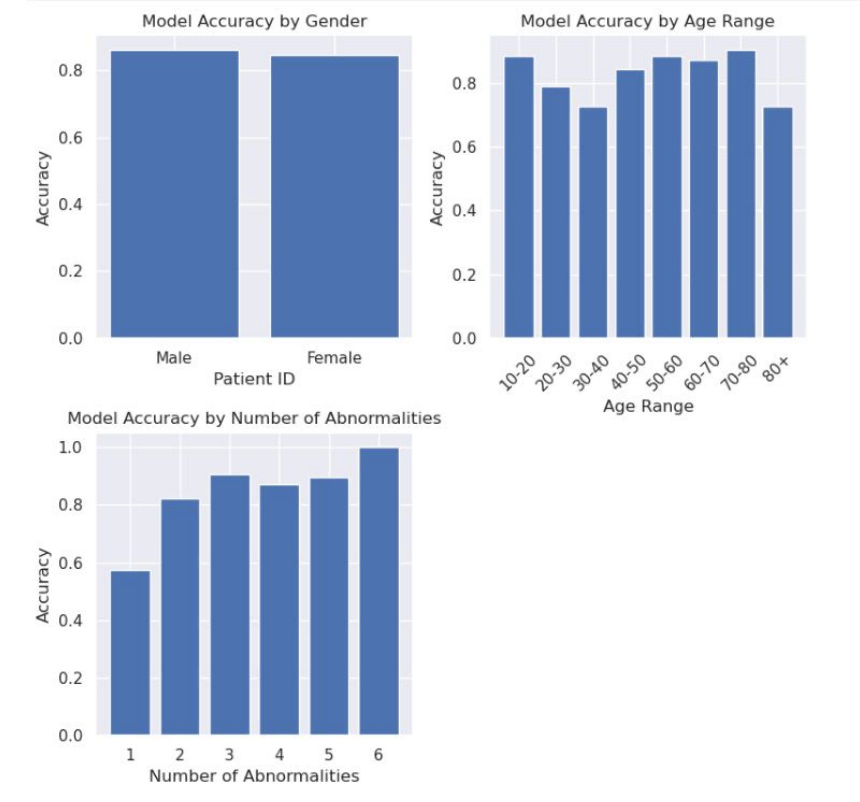
Baseline Model Definition and Training

Response Variable	Training Accuracy	Validation Accuracy	Test Accuracy
Ischemia	99.94%	92.29%	93.64%
Hypertrophy	100%	93.28%	88.05%
Conduction Abnormalities	100%	91.05%	88.63%
Non-specific Repolarization Abnormalities	100%	95.26%	85.61%

Baseline Model Evaluation

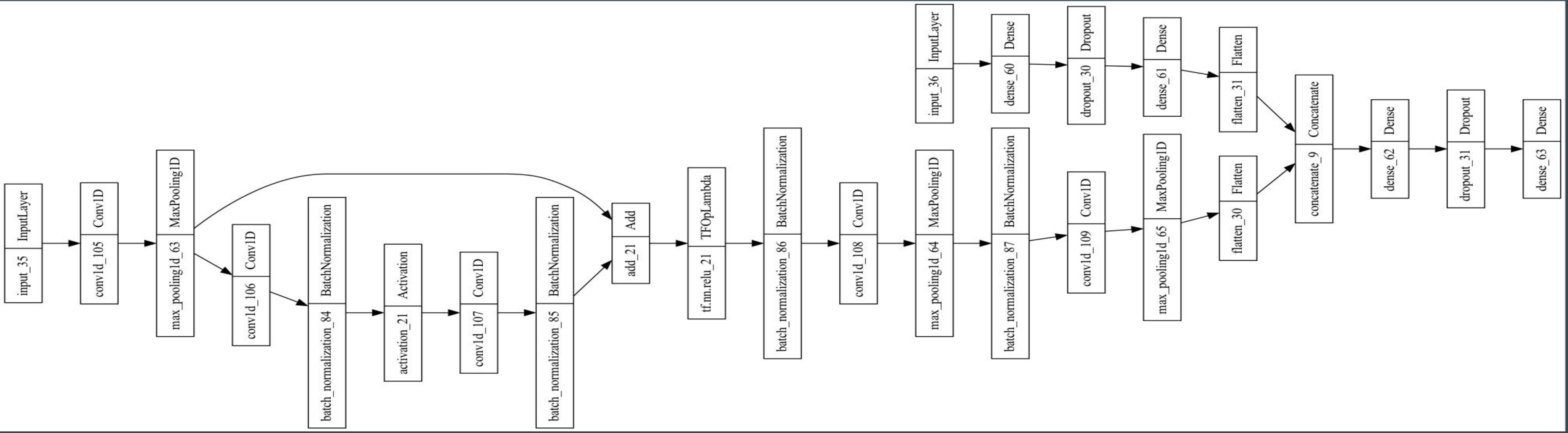
Model Commonalities:

- Reached peak validation accuracy around ~30 epochs
- Performed best with learning rate of .0001
- Some noticeable accuracy biases across different demographic groups
  - Ex. for Non-specific repolarization abnormalities at right
  - Varied between different models, see notebook for details



Final Model Definition and Training

- To address demographic accuracy imbalance, final model is updated to include a FFNN branch using the demographic data available, which is then concatenated to the flattened output of the convolutional layers
  - Goal to "contextualize" ECG information with demographic data



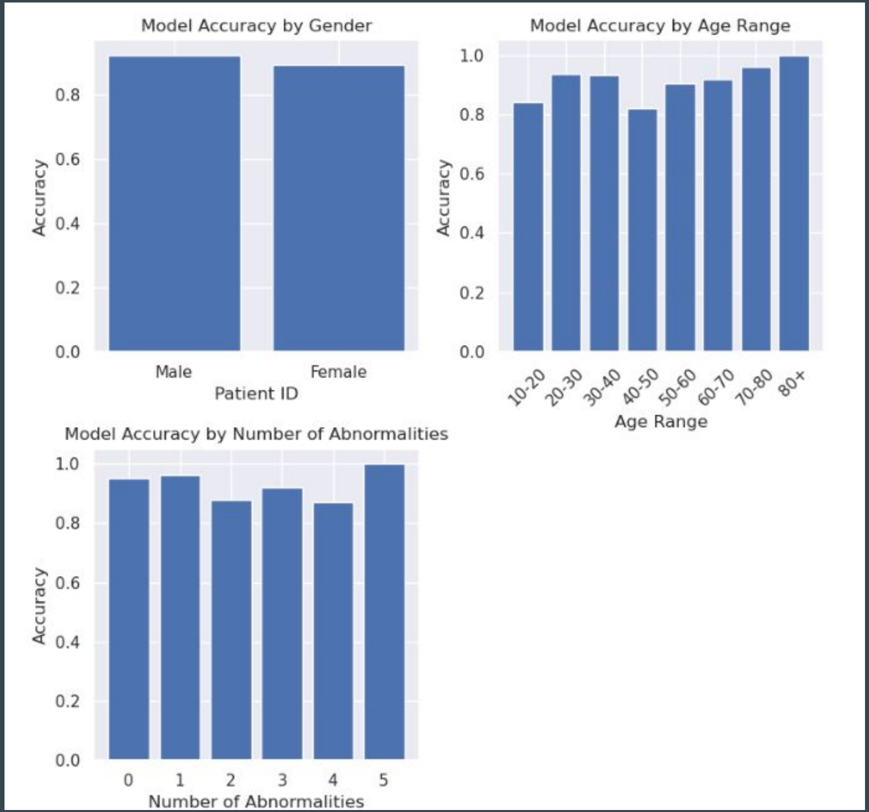
Final Model Definition and Training

- Improvements in all Test Accuracies aside from Ischemia

Response Variable	Training Accuracy	Validation Accuracy	Test Accuracy	Baseline Test Accuracy
Ischemia	99.94%	94.46%	93.45%	93.64%
Hypertrophy	100%	94.28%	90.84%	88.05%
Conduction Abnormalities	100%	93.68%	90.32%	88.63%
Non-specific Repolarization Abnormalities	100%	94.79%	91.10%	85.61%

Final Model Evaluation

- Demographic biases appeared to decrease in magnitude, but did not disappear completely
- Even w/ biases we observe higher absolute accuracies for each sub-population



Conclusion & Interpretation

- Our final model achieves test accuracies of above 90% across all different heart conditions studied, which in a real world setting would be viable as a supplement to existing diagnostics by medical professionals
- Our model's sensitivities are in agreeance with existing medical knowledge on what parts of heart function (and thus ECG waves) best signal the presence of the heart conditions we considered.
    - See evaluation sections of trained models in Jupyter

Future Work

- More data
  - Our dataset includes near 2000 different individual ECG waves, but they come from only ~200 individual patients; having a more diverse patient set would allow our model results to be more easily generalized
- GANs
  - Being able to generate characteristic ECG waves for the differing heart conditions could be useful as a way to visually prototype what the waves of afflicted individuals would look like
    - Could be useful in bolstering identification skills of physicians
    - Could be compared to what would be expected based solely on physical and anatomical knowledge
- Transfer Learning
  - It could be of benefit to investigate how a more sophisticated model architecture could perform when tuned towards our specific task, such as the ones made available by a Nature paper from 2021: https://www.nature.com/articles/s41598-021-84374-8