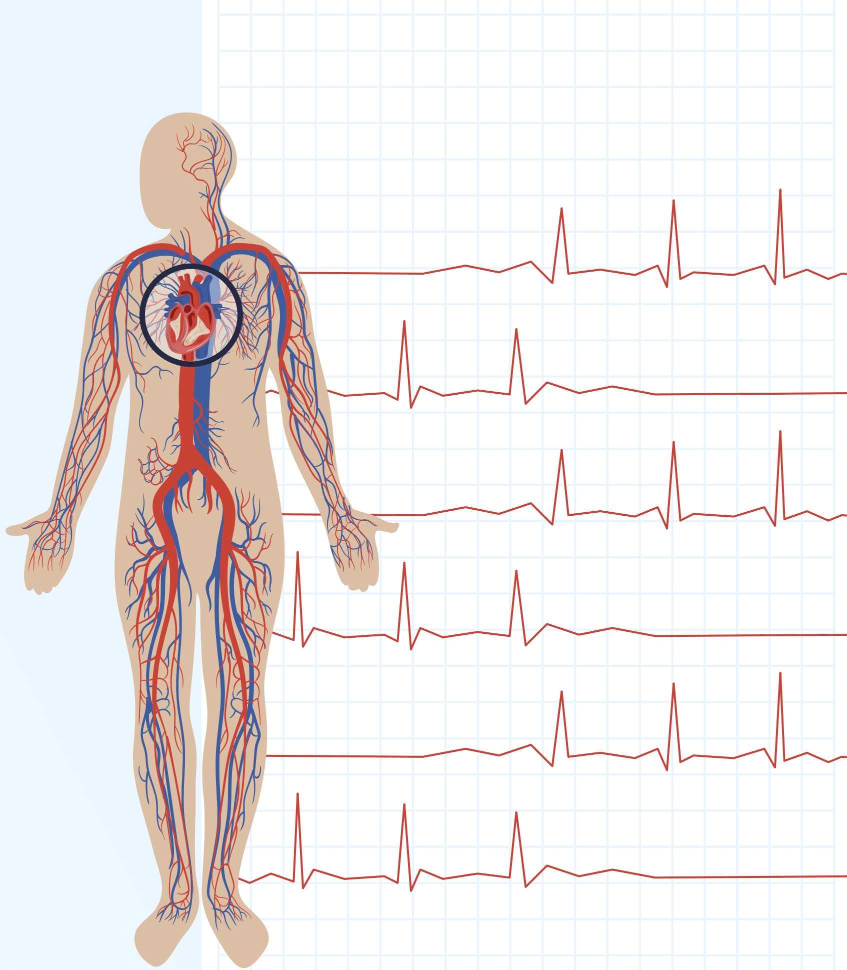
ECG Diagnostic Tool: A Machine Learning Approach



Daniel Kim, Trisha Sanghal, Zachary Fenton

Table of contents

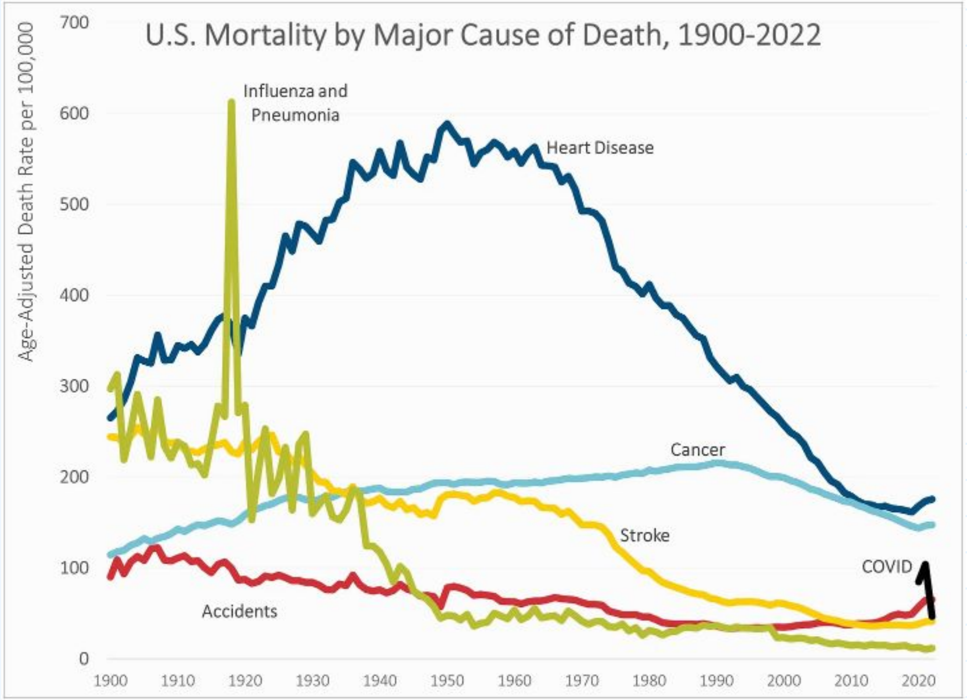
Motivation

Heart disease is the leading cause of death in America.

Electrocardiography (ECG; Pronounced 'Eee-Kay-Gee') is a painless, non-invasive diagnostic tool.

AI support systems for classifying ECGs could provide significant assistance; however, there are 2 major obstacles:

1. The lack of available datasets
2. A well trained model



Campbell, Mary Pat. "The Shape of U.S. Mortality 2: 1900-2022." *The Shape of U.S. Mortality 2: 1900-2022*, STUMP - Meep on public finance, pensions, mortality and more, 15 Nov. 2023, marypatcampbell.substack.com/p/the-shape-of-us-mortality-2-1900.

PTB-XL

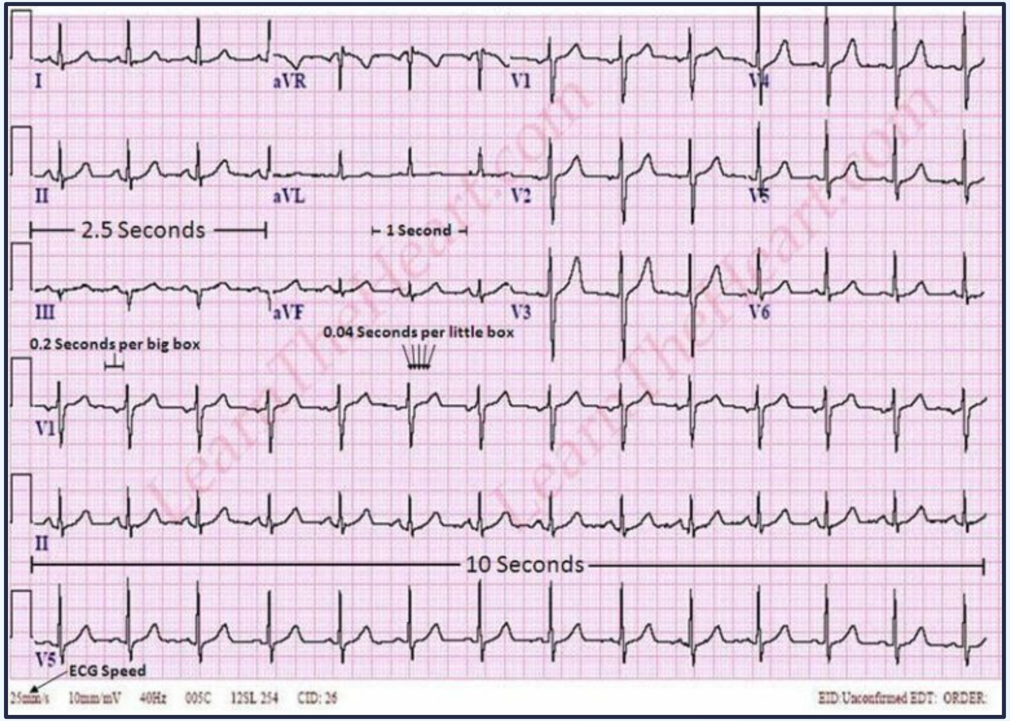
Data

100hz

500hz

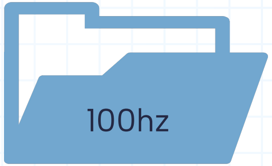
➔ 21,837 records

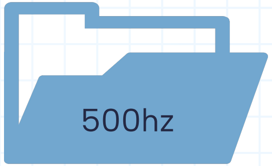
- 18,885 patients
- 12 lead
- 10 seconds in length
- 52% male, 48% female
- Ages 0 to 95



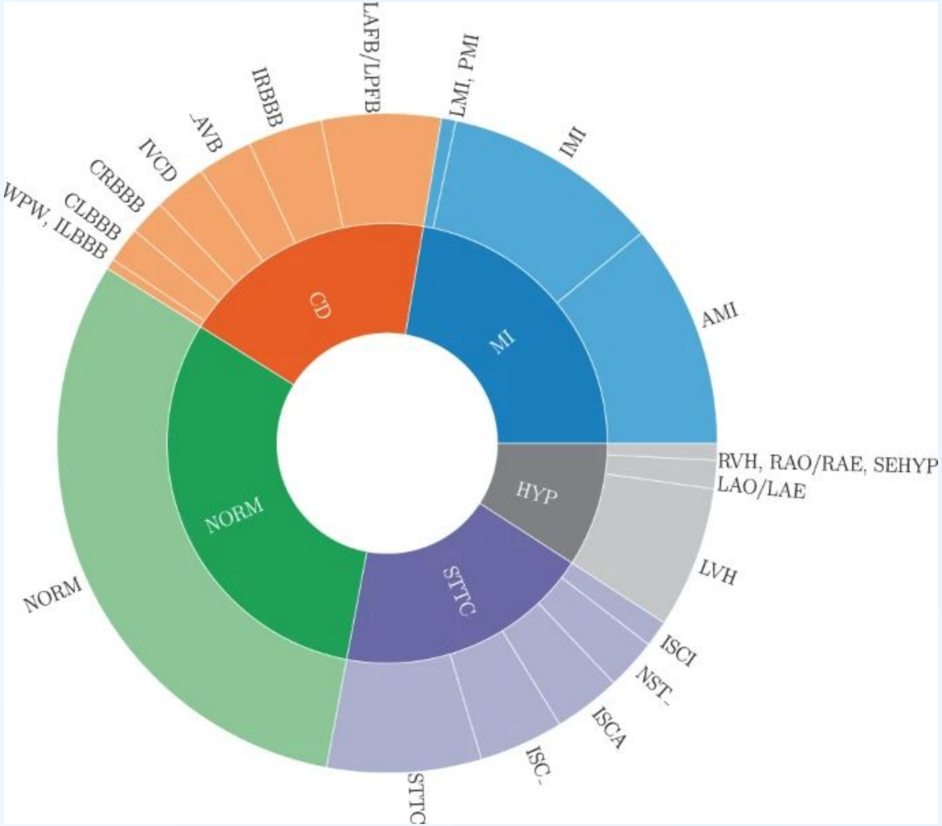
“ECG Boxes to Seconds Calculator - Definition: ECG Values.” *CalCon Calculator - Free Tool for Online Calculations*, 1 Mar. 2023, calconcalculator.com/health/ecg-boxes-to-seconds-calculator/.

Data

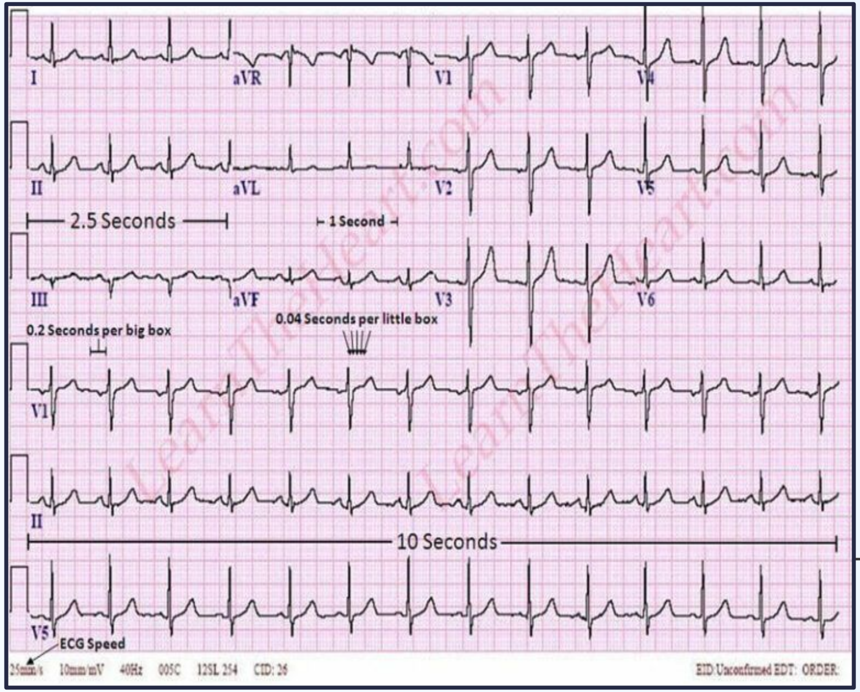




- ➔ 21,837 records
  - 18,885 patients
  - 12 lead
  - 10 seconds in length
  - 52% male, 48% female
  - Ages 0 to 95

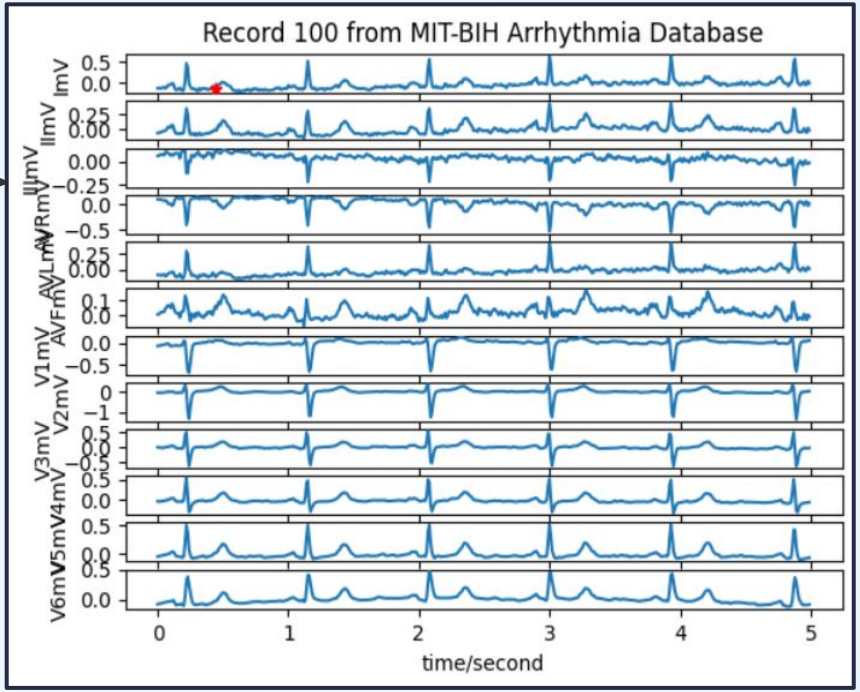


Wagner, Patrick, et al. "PTB-XL, a Large Publicly Available Electrocardiography Dataset." *Nature News*, Nature Publishing Group, 25 May 2020, www.nature.com/articles/s41597-020-0495-6. - Fig 1



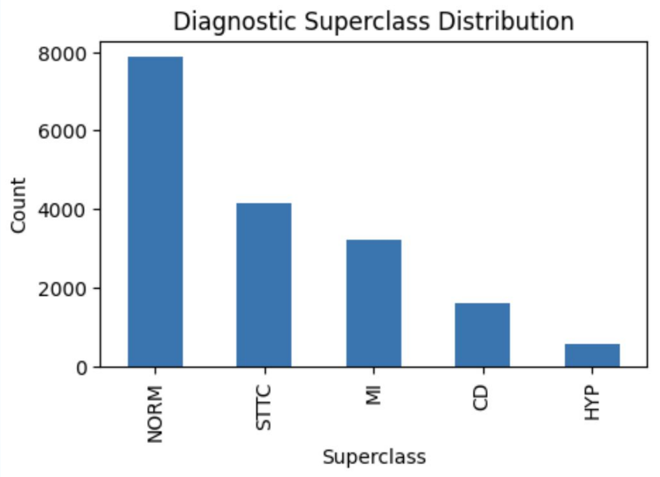
ECG output

WFDB data visualization



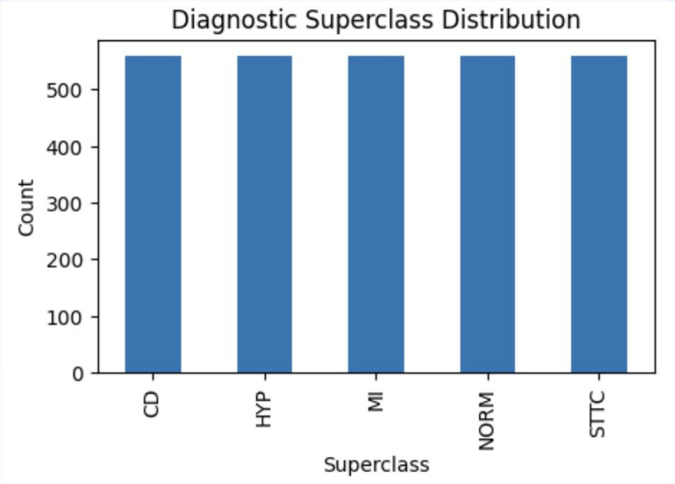
Preprocessing: Correct the Data Imbalance

Before



Total # of Training Samples:
17,439

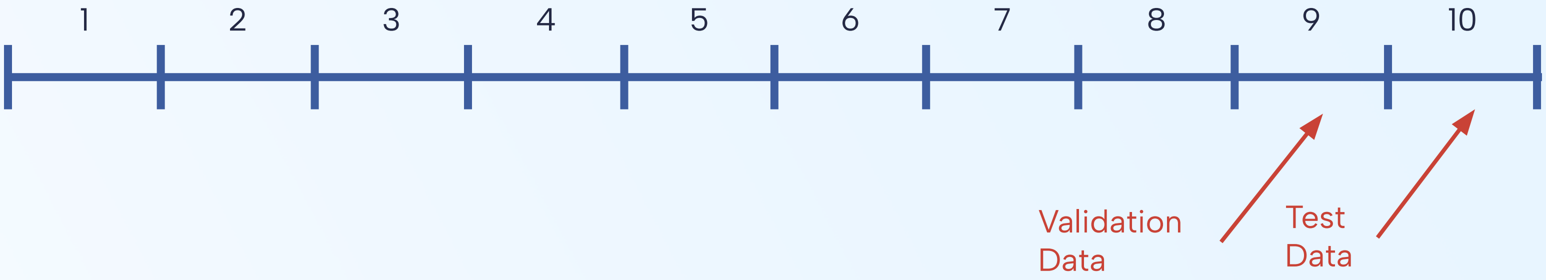
After



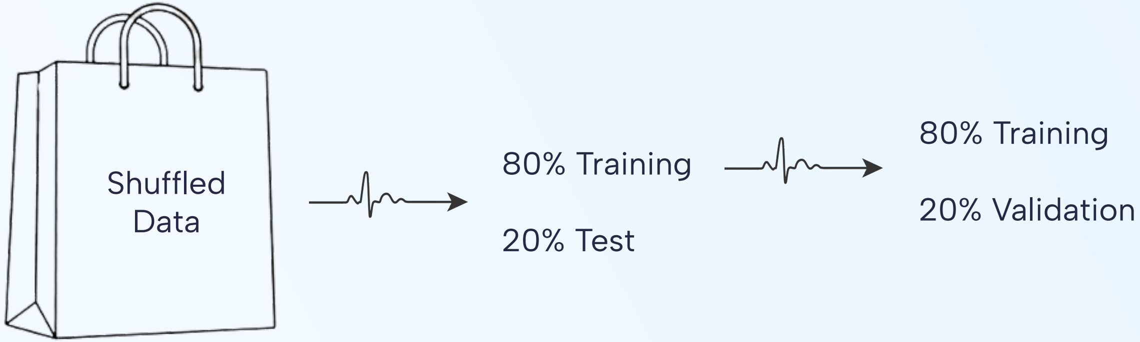
Total # of Training Samples:
2,800

Preprocessing: Randomize the Split

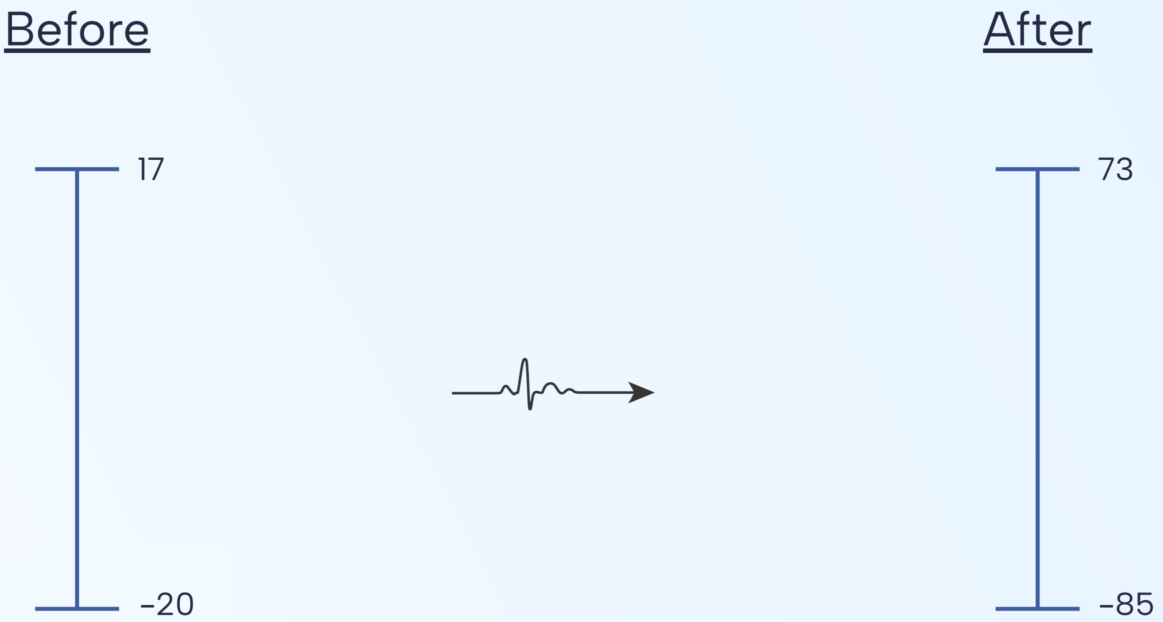
Before



After



Preprocessing: Normalize the Data



Modeling: Baseline Models

Predict at Random

~20% Validation Accuracy

Always Predict Normal

~46% Validation Accuracy

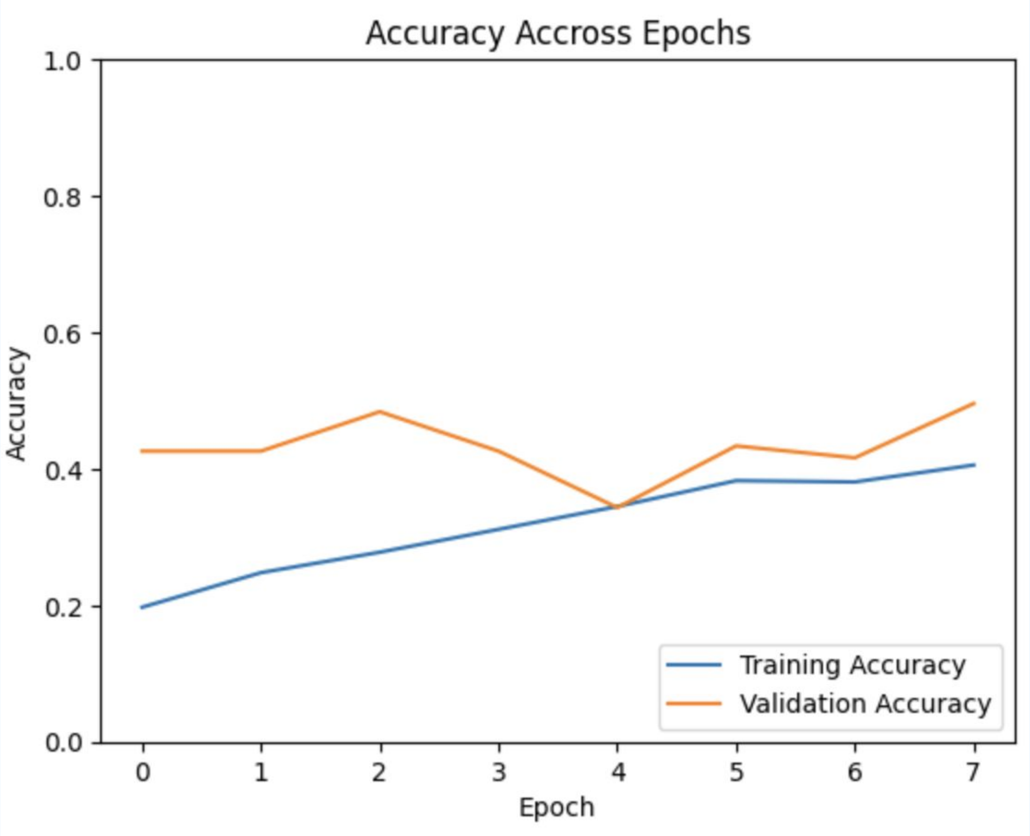
Modeling: Comparing Model Types

Random Forest	SVM	NN
~99% Training Accuracy	100% Training Accuracy	~23% Training Accuracy
~39% Validation Accuracy	~20% Validation Accuracy	~14% Validation Accuracy
Overfitting	Overfitting	

Modeling: What We Landed On

Layer (type)	Output Shape	Param #
conv1d_12 (Conv1D)	(None, 998, 64)	2368
max_pooling1d_12 (MaxPooling1D)	(None, 499, 64)	0
conv1d_13 (Conv1D)	(None, 497, 128)	24704
max_pooling1d_13 (MaxPooling1D)	(None, 248, 128)	0
conv1d_14 (Conv1D)	(None, 246, 256)	98560
max_pooling1d_14 (MaxPooling1D)	(None, 123, 256)	0
flatten_4 (Flatten)	(None, 31488)	0
dense_18 (Dense)	(None, 256)	8061184
dropout_10 (Dropout)	(None, 256)	0
dense_19 (Dense)	(None, 128)	32896
dropout_11 (Dropout)	(None, 128)	0
dense_20 (Dense)	(None, 64)	8256
dropout_12 (Dropout)	(None, 64)	0
dense_21 (Dense)	(None, 32)	2080
dropout_13 (Dropout)	(None, 32)	0
dense_22 (Dense)	(None, 16)	528
dropout_14 (Dropout)	(None, 16)	0
dense_23 (Dense)	(None, 12)	204

Total params: 8230780 (31.40 MB)
Trainable params: 8230780 (31.40 MB)
Non-trainable params: 0 (0.00 Byte)

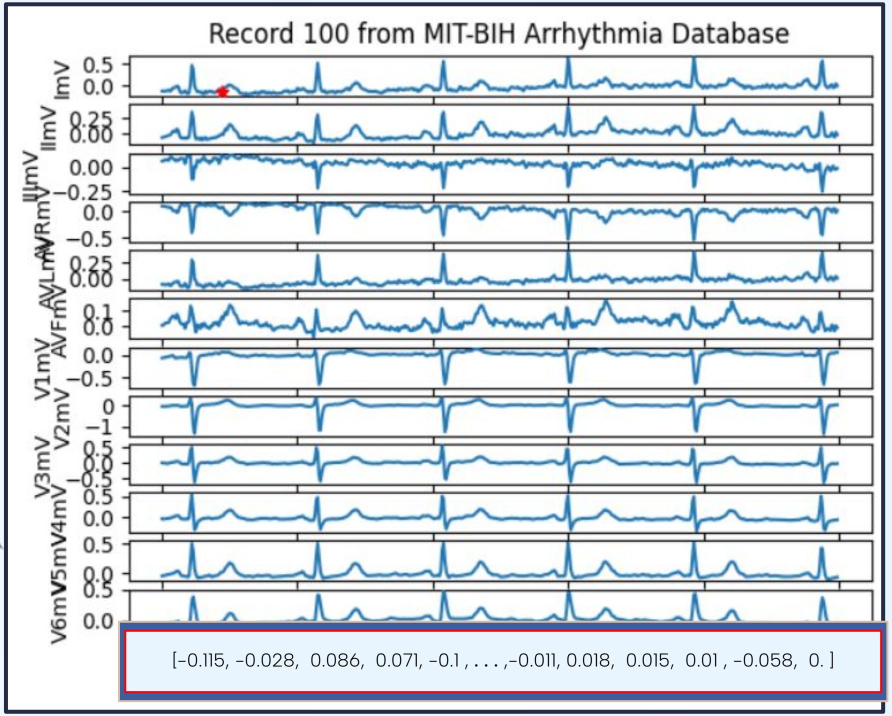


Experiment 1: Meta-data

40% → 43%

Without | With

[-0.115,
-0.028,
0.086,
0.071,
-0.1 ,
. . . ,
-0.011,
0.018,
0.015,
0.01 ,
-0.058,
0.]



Experiment 2: Manual Hyperparameter Optimization

43% → 43%

Before → Best Observed

# Conv layers	Conv filters	Dense units	Pool Size
1	[64]	[256, 128, 64, 32, 16]	2
2	[64, 64]	[256, 128, 64, 32, 16]	2
3	[64, 64, 64]	[256, 128, 64, 32, 16]	2
4	[64, 64, 64, 64]	[256, 128, 64, 32, 16]	2
5	[64, 64, 64, 64, 64]	[256, 128, 64, 32, 16]	2
3	[64, 128, 256]	[256, 128, 64, 32, 16]	2
3	[128, 128, 128]	[256, 128, 64, 32, 16]	2
3	[256, 256, 256]	[256, 128, 64, 32, 16]	2
3	[256, 128, 64]	[256, 128, 64, 32, 16]	2
3	[128, 128, 128]	[256, 128, 64, 32, 16]	1
3	[256, 256, 256]	[256, 128, 64, 32, 16]	2
3	[256, 128, 64]	[256, 128, 64, 32, 16]	3
3	[256, 256, 256]	[64]	1
3	[256, 256, 256]	[64, 64]	1
3	[256, 256, 256]	[64, 64, 64]	1
3	[256, 256, 256]	[64, 64, 64, 64]	1
3	[256, 256, 256]	[64, 64, 64, 64, 64]	1

Experiment 3: Keras Tuner Optimization

43%
Before

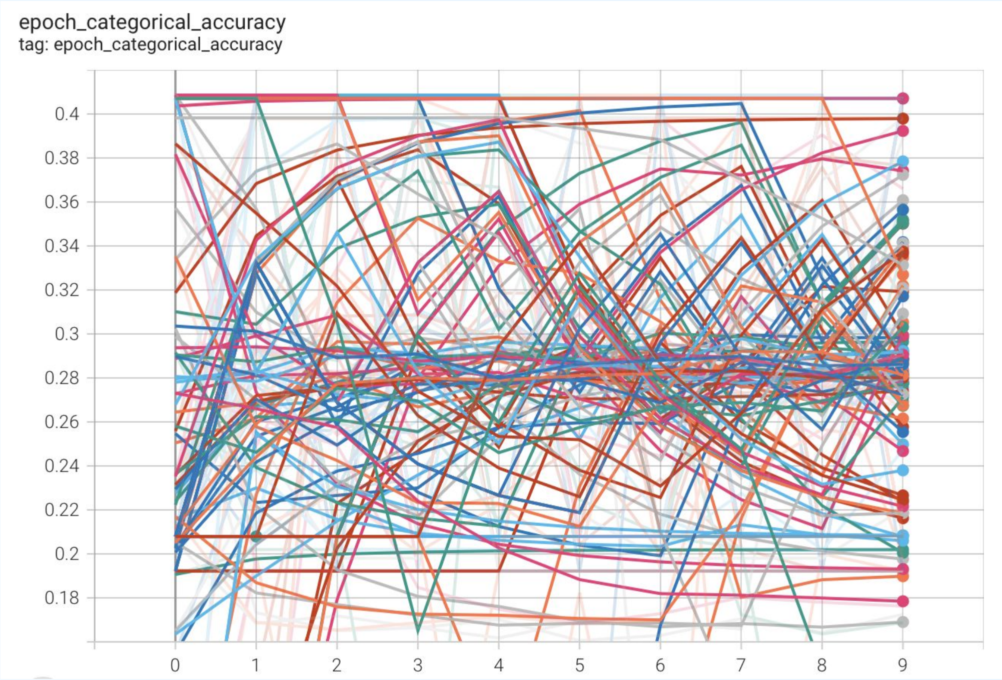
→

47%
Best Model

50%
Test Accuracy

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 1000, 13)]	0
conv1d (Conv1D)	(None, 998, 224)	8960
max_pooling1d (MaxPooling1D)	(None, 499, 224)	0
conv1d_1 (Conv1D)	(None, 497, 224)	150752
max_pooling1d_1 (MaxPooling1D)	(None, 248, 224)	0
conv1d_2 (Conv1D)	(None, 246, 256)	172288
max_pooling1d_2 (MaxPooling1D)	(None, 123, 256)	0
conv1d_3 (Conv1D)	(None, 121, 256)	196864
max_pooling1d_3 (MaxPooling1D)	(None, 60, 256)	0
conv1d_4 (Conv1D)	(None, 58, 256)	196864
max_pooling1d_4 (MaxPooling1D)	(None, 29, 256)	0
flatten (Flatten)	(None, 7424)	0
dropout (Dropout)	(None, 7424)	0
dropout_1 (Dropout)	(None, 7424)	0
dense (Dense)	(None, 32)	237600
dense_1 (Dense)	(None, 32)	1056
dense_2 (Dense)	(None, 32)	1056
dense_3 (Dense)	(None, 32)	1056
dense_4 (Dense)	(None, 32)	1056
dense_5 (Dense)	(None, 32)	1056
dense_6 (Dense)	(None, 12)	396

Total params: 969,004
Trainable params: 969,004
Non-trainable params: 0



Conclusions

Model

Metric	Value	Conclusion
Test Accuracy	50%	✗
Validation Accuracy	47%	✗
F1	0.59	~
Precision	0.5	~
Recall	0.7	~

Implications

- Without training: medical students have a 42% accuracy while residents have a 55.8% accuracy.[5]
- This implies that this initial model may perform better than your average medical student if you want to get your EKG looked at (not recommended)
- Some potential negative societal impacts include inequity due to biased datasets for certain classes (sex, race, etc.)

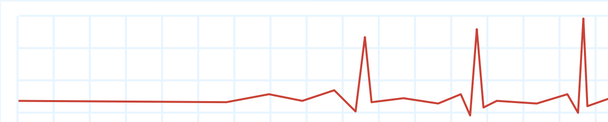
Next Steps

- Continue to optimize data pre-processing
- Re-encode using all relevant metadata
- Scope different model types (RNN, MLP, Hybrid)
- Continue to optimize model hyperparameters (try Adamax loss function)

Contributions

	Daniel	**Trisha**	**Zach**
Code	Hyperparameter optimization	Data preprocessing, baseline modeling	Data loading, Model development
Slides/Presentation	• Experiments • Conclusion	• Preprocessing • Modeling	• Motivation • Data
Research	• Tensorflow hyperparameter optimization automation • ECG studies	• Supervised vs Unsupervised models	• Initial data set • ECG research

- All members contributed to NeurIPS verification with equal effort.



References

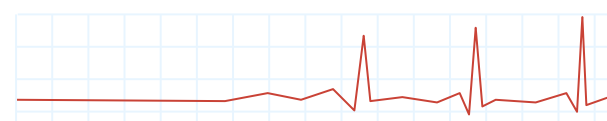
1. *Wagner, Patrick, et al. "PTB-XL, a large publicly available electrocardiography dataset." Scientific Data, vol. 7, no. 1, 25 May 2020, https://doi.org/10.1038/s41597-020-0495-6.*
2. *Wagner, P., Strodthoff, N., Bousseljot, R., Samek, W., & Schaeffter, T. (2022). PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3). PhysioNet. https://doi.org/10.13026/kfzx-aw45.*
3. Goldberger, A., et al. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220." (2000).
4. Wagner, Patrick, et al. "PTB-XL, a Large Publicly Available Electrocardiography Dataset." *Nature News*, Nature Publishing Group, 25 May 2020, www.nature.com/articles/s41597-020-0495-6.
5. Cook DA, Oh SY, Pusic MV. Accuracy of Physicians' Electrocardiogram Interpretations: A Systematic Review and Meta-analysis. *JAMA Intern Med*. 2020;180(11):1461-1471. doi:10.1001/jamainternmed.2020.3989

Data License:

https://physionet.org/content/ptb-xl/view-license/1.0.3/

GitHub:

https://github.com/zfenton/UCBMIDS_W207_finalProject



NeurIPS Checklist

❏ 1. For all authors...
- ✓ (a) Do the main claims made in the abstract and introduction accurately reflect the papers contributions and scope? **- N/A**
- ✓ (b) Have you read the ethics review guidelines and ensured that your paper conforms to them? **- YES**
- ✓ (c) Did you discuss any potential negative societal impacts of your work? **- YES**
- ✓ (d) Did you describe the limitations of your work? **- YES**

❏ 2. If you are including Theoretical Results…
- ✓ (a) Did you state the full set of assumptions of all theoretical results? **- N/A**
- ✓ (b) Did you include complete proofs of all theoretical results? **- N/A**

❏ 3. If you ran experiments…
- ✓ (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **- YES**
- ✓ (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **- YES**
- ✓ (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? - **YES** - in github
- ✓ (d) Did you include the amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **- YES**

❏ 4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets…
- ✓ (a) If your work uses existing assets, did you cite the creators? **- YES**
- ✓ (b) Did you mention the license of the assets? **- YES**
- ✓ (c) Did you include any new assets either in the supplemental material or as a URL? **- YES**
- ✓ (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? **- YES**
- ✓ (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **- YES**

✓ 5. If you used crowdsourcing or conducted research with human subjects…
- ✓ (a) Did you include the full text of instructions given to participants and screenshots, if applicable? **- N/A**
- ✓ (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? **- N/A**
- ✓ (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **- N/A**