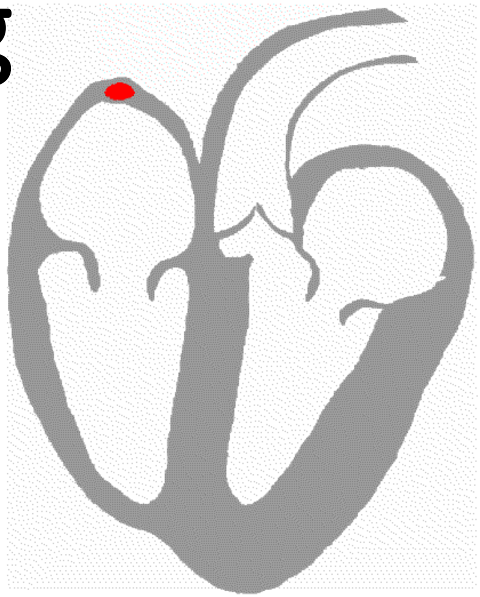
Detecting and Classifying Abnormal Heart Beats from ECG Recordings

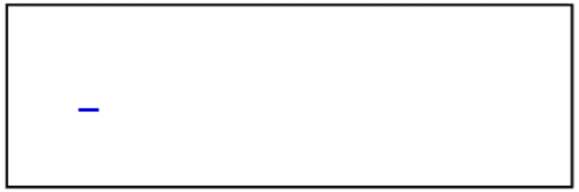
And U. Turken

Galvanize Capstone Presentation

June 9, 2020

Background: Heart-health monitoring





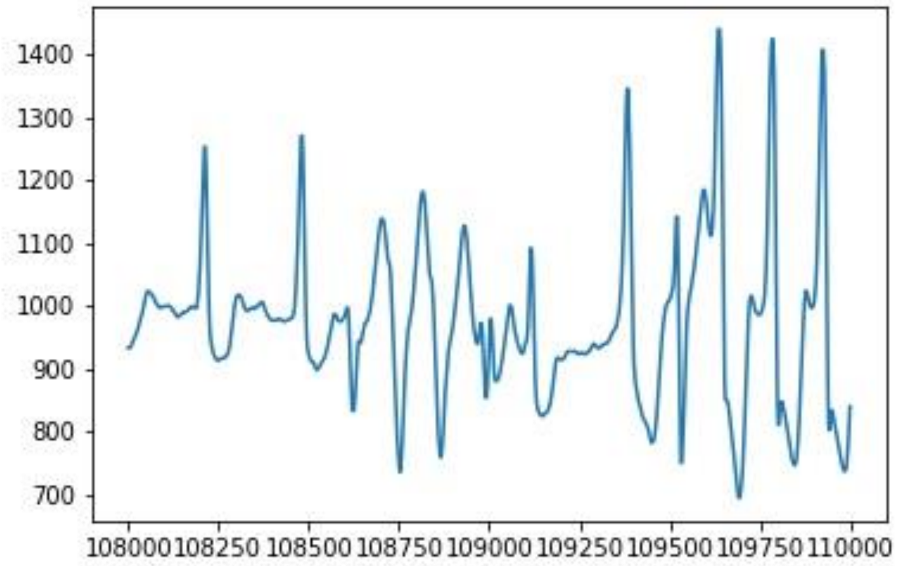
- Abnormal in heart beat rhythms are warning signals
- Heart attacks and stroke can be predicted in advance
- Clinical electrocardiogram (ECG) monitoring
- ECG monitoring with IoT devices (e.g. smart watches)
- ML for automatic ECG abnormality detection
- Train machine learning model with expert-labeled ECG data
- Detect and classify abnormal heart beats using ECG data from same individuals, assess transfer to new individuals

Electrocardiogram (ECG) Data

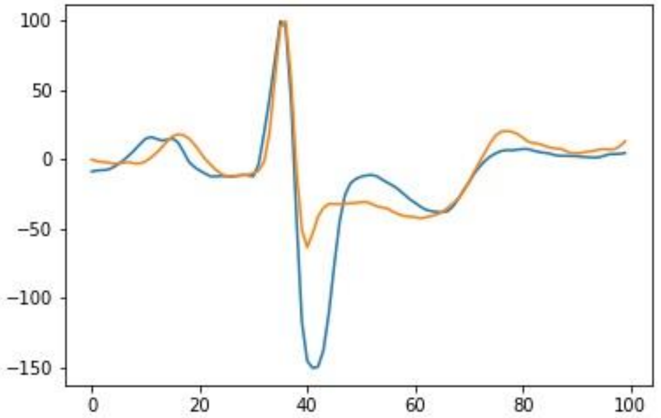
- MIT-BIH Arrhythmia Database from PhysioNet
- 30 minute clinical ECG recordings from 43 individual patients
- ~109,000 expert-labeled heart beats, ~2,500 hear beats / individual
- Variable ECG abnormality across individuals, from 1% <, to > 50% (Overall, 14% of all ECG segments are labeled as abnormal)
- ~50% generally normal, ~50% with significant abnormality
- Type of abnormality recorded on continuous ECG
- Commonly used as a benchmark since 1990s
- Single channel data, recorded at 360 Hz, comparable to smartwatch

ECG data pre-processing

- Read continuous ECG into Python, visually inspect data quality
- Extract ECG segments for each heart beat
- Individual segments should comparable!
- Extensive pre-processing required
- Apply linear drift correction (detrend)
- Align segments so same peak latency
- Baseline correction
- Minimal temporal smoothing
- Scale ECG peak amplitude to 100
- Remove noisy segments
- Recording artifacts: ~2% of data excluded



Continuous ECG recording



Preprocessed ECG segment = One heart-beat

Modeling Approach

- Assess how well abnormal hear beats can be detected and classified in new unseen ECG data from the same group of individuals (n=28)
- Assess how well abnormal hear beats can be detected and classified in new ECG data from a new group of individuals (n=15)
- Two approaches:
- (1) Use amplitudes for all ECG times points as a feature vector
- (2) Representation learning – apply unsupervised learning to learn a compact representation prior to classification (potentially better transfer to data from new groups of individuals)

Modeling with Scikit-Learn

- Train – Test split (stratified, to preserve individual class frequencies): 75% - 25% of all ECG segments from 28 individuals
- Five-fold cross-validation (CV) for hyper-parameter tuning
- Test out-of-sample generalizability: All ECG segments from further 15
- Multi-class classification: Normal ECG + Four most common abnormal ECG patterns; Precision, recall, F1 to assess performance
- To address class imbalance: “class weight” balancing in Scikit-Learn
- Unsupervised representation learning: “sparse” models for compact components that might capture clinically-relevant aspects of ECG morphology. Pipelined with a classification algorithm, representation tuned based on classifier performance with grid search. Batch learning to manage memory requirements

Modeling - Algorithms

- Simple baseline: Isolation forest for unsupervised anomaly detection (~69% accuracy for labeling abnormal ECG on test data)
- Baseline: K-Nearest Neighbors classifier – tune neighborhood size with CV
- Logistic regression classifier – tune L1 regularization
- Balanced Random Forest – tune number of trees, number of features / tree
- Unsupervised learning: Sparse PCA, Dictionary Learning, ICA – tune number of components, regularization

Results for k-NN, balanced Random Forest and Logistic Regression classifiers (no unsupervised pre-processing)

kNN - Train-test on same group of individuals

	precision	recall	f1-score
N	0.99	1.00	1.00
A	0.98	0.92	0.95
L	0.99	0.99	0.99
R	0.99	1.00	0.99
V	0.98	0.97	0.98
macro avg	0.99	0.97	0.98
weighted avg	0.99	0.99	0.99

kNN - Train on group 1, test on group 2

	precision	recall	f1-score
N	0.85	0.98	0.91
A	0.19	0.23	0.21
L	1.00	0.42	0.59
R	0.33	0.04	0.07
V	0.59	0.70	0.64
macro avg	0.59	0.47	0.48
weighted avg	0.82	0.83	0.80

bRF - Train-test on same group of individuals

	precision	recall	f1-score
N	1.00	0.98	0.99
A	0.81	0.95	0.87
L	0.99	0.99	0.99
R	0.99	0.99	0.99
V	0.88	0.99	0.93
macro avg	0.93	0.98	0.95
weighted avg	0.98	0.98	0.98

bRF - Train on group 1, test on group 2

	precision	recall	f1-score
N	0.89	0.92	0.90
A	0.08	0.44	0.14
L	1.00	0.38	0.55
R	0.52	0.04	0.07
V	0.38	0.86	0.53
macro avg	0.57	0.53	0.44
weighted avg	0.85	0.79	0.78

Log Reg - Train-test on same group of individuals

	precision	recall	f1-score
N	0.95	0.97	0.96
A	0.73	0.66	0.69
L	0.92	0.93	0.93
R	0.86	0.85	0.86
V	0.81	0.64	0.71
macro avg	0.85	0.81	0.83
weighted avg	0.92	0.92	0.92

Log Reg - Train on group 1, test on group 2

	precision	recall	f1-score
N	0.81	0.91	0.86
A	0.07	0.07	0.07
L	0.04	0.00	0.00
R	0.06	0.04	0.05
V	0.17	0.43	0.25
macro avg	0.23	0.29	0.25
weighted avg	0.73	0.77	0.75

k-NN: best performance with n=3; Balanced RF: best performance with 1000 trees; Log Regression: Heavy L1 regularization

N: Normal ECG, A, L, R, V: Four most common ECG abnormality types

Modeling – Results / Conclusion

- Across all models, best performance for prediction abnormal ECG in new data from the same set of individuals
- Using low quality ECG data from a single channel, common machine learning algorithms can successfully be applied to detect abnormal heart activity when past data from same patients are available
- Relatively poor transfer to unseen population (however performance on single ECG segment is a very stringent criterion)
- Contrary to expectation, unsupervised preprocessing gave poorer results (in each cases, ~5-10% lower performance - results not presented further due potential implementation errors)
- Good performance of k-NN and RF suggest that similarity-based approaches might be best suited for this problem. Representation learning remains work in progress

Future work

- Need ML solution for isolating heart beats from continuous ECG
- Deep learning with data from large populations for transferable ECG pattern recognition
- Apply convolutional neural network (CNN) and autoencoder modeling to better characterize ECG morphology
- Apply autoencoder to model variability of ECG features within and across patients
- Characterize ECG abnormality in terms of relative timing and amplitudes of ECG features for better diagnostic interpretability