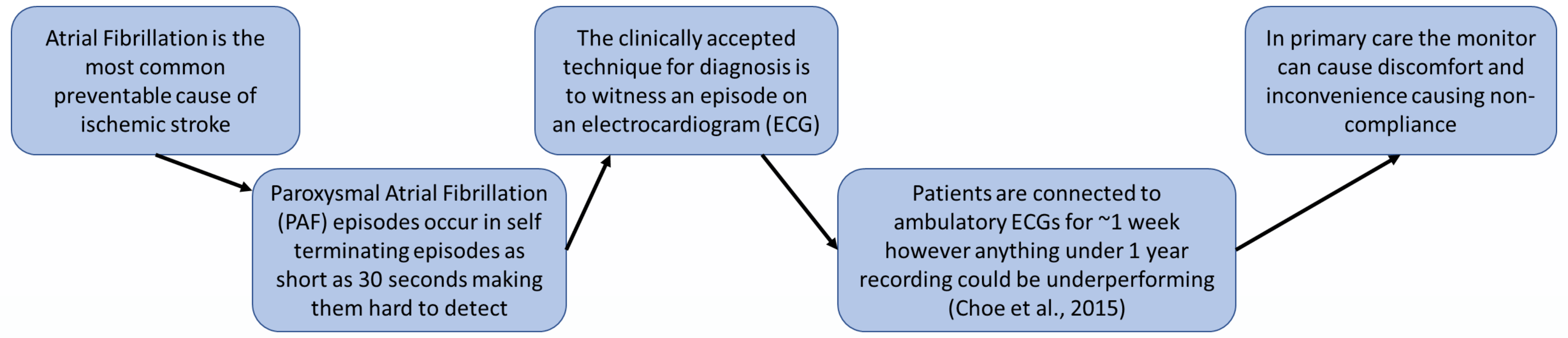
FUNDED BY NIHR | National Institute for Health Research

Utilisation of an open source database for Paroxysmal Atrial Fibrillation detection

Stevie Creasy[1,2], Vadim Alexeenko[1], Jane Lyle[2], Philip Aston[2], Kamalan Jeevaratnam[1]

[1]School of Veterinary Medicine, University of Surrey; [2]Department of Mathematics, University of Surrey

Introduction



Atrial Fibrillation is the most common preventable cause of ischemic stroke

Paroxysmal Atrial Fibrillation (PAF) episodes occur in self terminating episodes as short as 30 seconds making them hard to detect

The clinically accepted technique for diagnosis is to witness an episode on an electrocardiogram (ECG)

Patients are connected to ambulatory ECGs for ~1 week however anything under 1 year recording could be underperforming (Choe et al., 2015)

In primary care the monitor can cause discomfort and inconvenience causing non-compliance

Fig 1: A flow chart showing the aims of this project and describing the most common path from episode to diagnosis, highlighting the issues with the current method.

Aims and Objectives

- To generate a machine learning tool able to classify a normal sinus rhythm ECG as PAF or control with a sufficiently high accuracy.
- To be able to algorithmically select strips of high quality from long ECG signals to improve machine learning accuracy.
- To investigate whether multiple analysis techniques can be combined to give increased accuracy.

Proposed Methods

All techniques used in this project have been previously used to predict cardiac disease with high accuracy.

Symmetric Projector Attractor Reconstruction (SPAR)

- SPAR takes three delayed points within a single heartbeat and plots the positions of these points in 3 dimensions as they traverse the full signal, then by viewing this attractor from a corner it can be projected into 2 dimensions and visualised as a density plot (Aston et al., 2018).

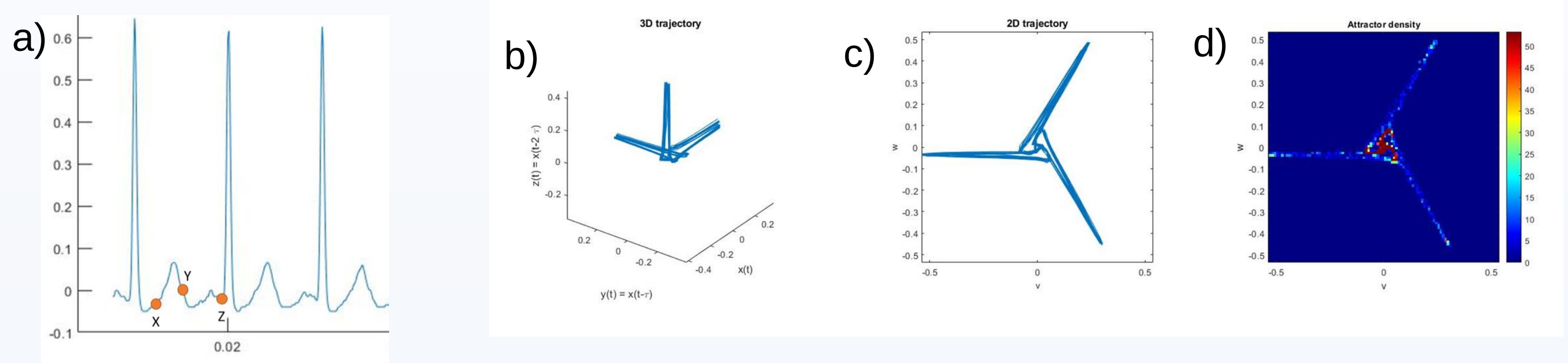


Fig 2: a) Plotting three delay points onto the signal. b) Plotting the delay coordinates through time to produce a 3D plot. c) Viewing the plot through the vector (1,1,1) projects the plot into 2D. d) Converting to a density plot to avoid loss of information.

Complexity

- The signal is first converted to a binary string via a coarse graining technique, threshold crossing assigns a 1 to every point above the threshold and 0 otherwise. Beat detection and feature detection assign 1 at specific features of the signal, such as the R peak, and 0 everywhere else.

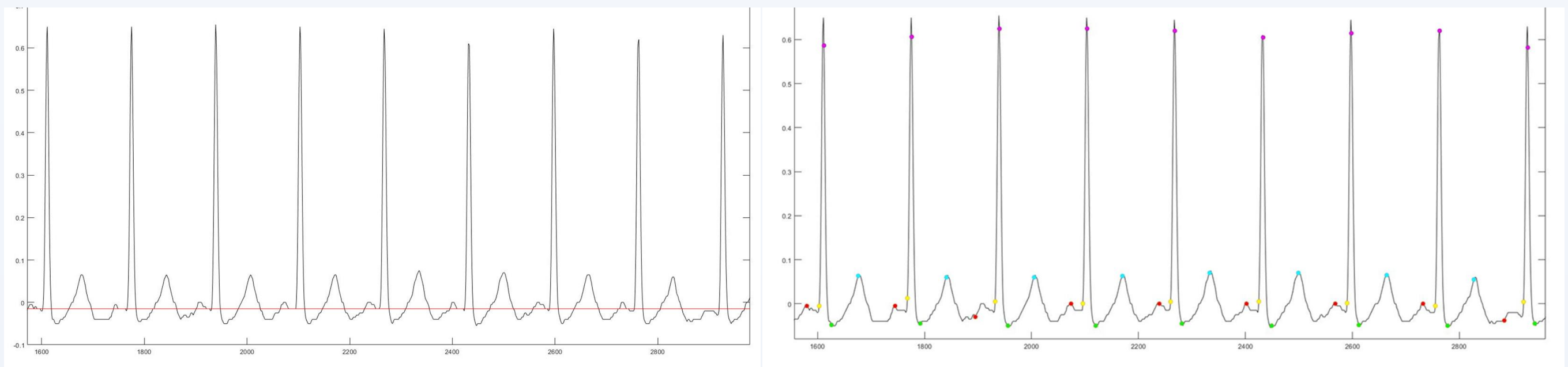


Fig 3: Converting the signal to a binary string. Using the median value of the signal as a threshold, all points above the line become 1 (left). Detection of the Q, R, S and T peaks gives the features for other coarse graining techniques (right).

- Complexity techniques then determine how difficult each binary string is to replicate to determine its complexity. At a fundamental level these can be thought of as pattern detection methods where repeated patterns make the signal less complex (Lempel and Ziv, 1976).

Restitution

- By determining the positions of the R peak, Q peak and T wave end in the signal we can calculate the intervals RR, QT and TQ. These correspond to heart rate, depolarisation and repolarisation time of the heart every beat. Analysis of changes in these intervals allows us to detect cardiac diseases as often a physical change occurs preceding a cardiac event (Huang et al., 2020).

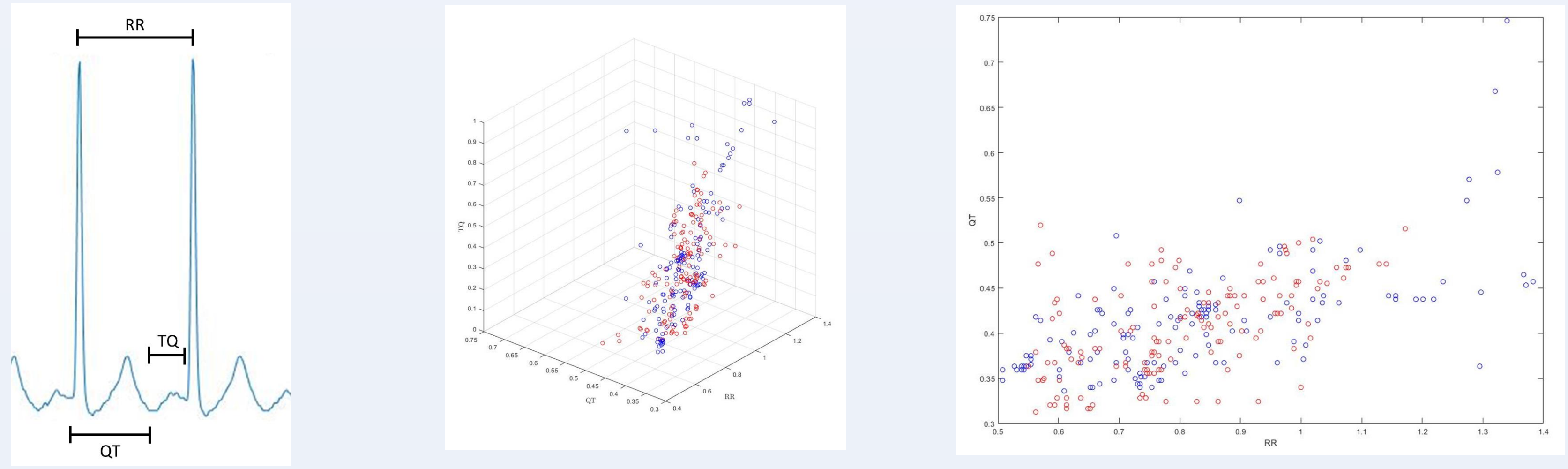
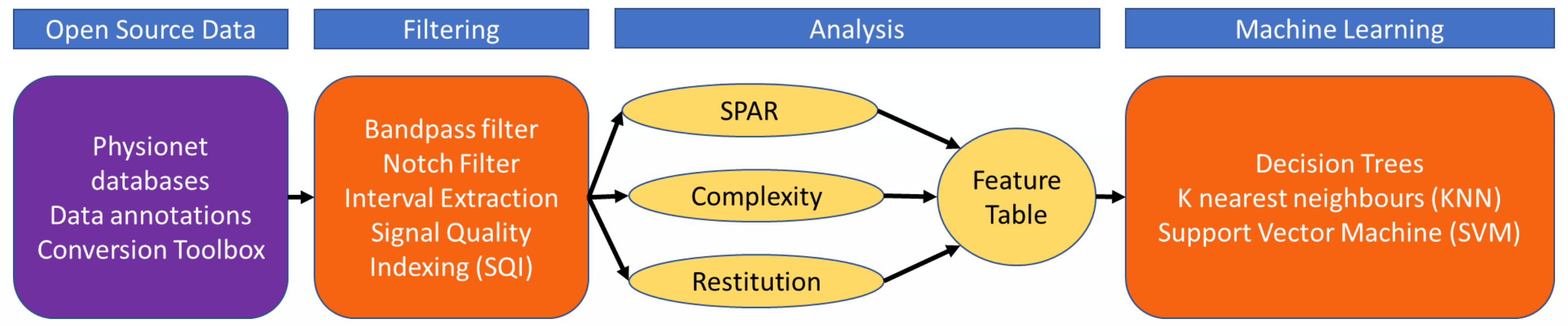


Fig 4: A visual representation of the three intervals on and RCG signal (left). A plot of all three intervals for each subject (centre). A plot showing only the RR intervals against QT intervals (right). In each plots the control subjects are shown in blue and the cases are shown in red.

Open Research Practice

- Once we have the feature tables from our analysis we can then pass this to machine learning models to perform cross validation and determine the accuracy of our model.



Open Source Data: Physionet databases, Data annotations, Conversion Toolbox

Filtering: Bandpass filter, Notch Filter, Interval Extraction, Signal Quality Indexing (SQI)

Analysis: SPAR, Complexity, Restitution → Feature Table

Machine Learning: Decision Trees, K nearest neighbours (KNN), Support Vector Machine (SVM)

Fig 5: The pipeline from data input to machine learning output. The open source data is converted to a matlab format using a conversion toolbox. The signals are then filtered with bandpass and notch filters. The three techniques previously discussed are fun to generate a feature table which is passed to machine learning to determine the accuracy of the model.

- Machine learning tools are susceptible to overfitting if not enough data is used, this is where physionet.org becomes a valuable resource.
- Physionet provides an extensive archive of physiological signals across a wide range of databases along with a library of software allowing for conversion and processing of these signals. We are able to access large databases containing ECGs from patients with PAF both during episodes and during normal sinus rhythm. These include a range of signals recorded both in a clinical setting but also those from ambulatory ECGs, similar to those that we would expect to be used with the output of this project.
- Whilst control signals are easy to access, having case signals recorded under the same conditions allows us to tune our machine learning parameters to give accuracies upward of 80%.
- Using toolboxes available through physionet allows us to view the annotations generated by cardiologists. These highlight areas of signal noise, ectopic beats and similar anomalies that may hinder our analysis.
- Normally studies would spend long periods of time collecting data for their analysis. However physionet allows us to begin our calculations earlier in the process by providing a large database of data to test our model on.

PhysioNet
The Research Resource for Complex Physiologic Signals

Limitations

Using an open source database introduces certain difficulties when used:

- We are unable to control the methods used for data collection. This can create issues with low sampling rates, noisy signals or imbalance between controls and cases.
- The choice of subjects is out of our control, sometimes patients are recommended for ECG monitoring due to pre-existing conditions that may impact the data. We may also find ourselves classifying other factors rather than case against control.

Preliminary Results

We have generated some preliminary results from the PAF prediction challenge on physionet, these are currently without rigorous cross validation and need a better approach moving forwards.

SPAR

- Our first runs of SPAR machine learning have given an accuracy of around 60%, however we are yet to optimise the parameters used in the method and so we expect this to increase.

Complexity

	Fine Tree	Medium Tree	Coarse Tree	Linear SVM	Quadratic SVM	Cubic SVM	Medium KNN (10)	Coarse KNN (100)	Weighted KNN
LZ76BD	84.0	83.6	84.1	80.5	78.8	74.1	83.8	82.3	84.2
LZ78BD	83.9	83.2	81.6	80.5	65.8	59.3	82.0	82.2	82.4
TiBD	83.7	83.3	82.2	80.5	63.5	76.7	84.8	80.4	84.9
LZ76FD	81.5	81.0	80.9	80.5	74.5	71.1	80.9	80.7	81.6
LZ78FD	83.0	83.3	81.9	80.5	72.0	75.7	84.7	80.2	84.4
TiFD	80.2	82.0	81.3	80.5	63.3	54.3	81.6	81.3	79.4
All BD methods	87.2	85.5	84.1	80.5	84.1	73.0	86.1	83.2	87.0
All FD methods	82.1	83.7	81.7	80.5	80.5	71.9	81.8	80.2	82.4
All methods	85.8	84.5	83.2	80.5	84.4	81.3	86.1	82.4	86.8

Table 1: A table containing the accuracies of machine learning methods when using results from different combinations of coarse graining and complexity methods. The machine learning methods used were decision trees, support vector machine (SVM) and K nearest neighbours (KNN). Only beat detection (BD) and feature detection (FD) were used as coarse graining techniques whilst all complexity techniques were used.

Restitution

- We can see that using higher quality strips yields better results for restitution.

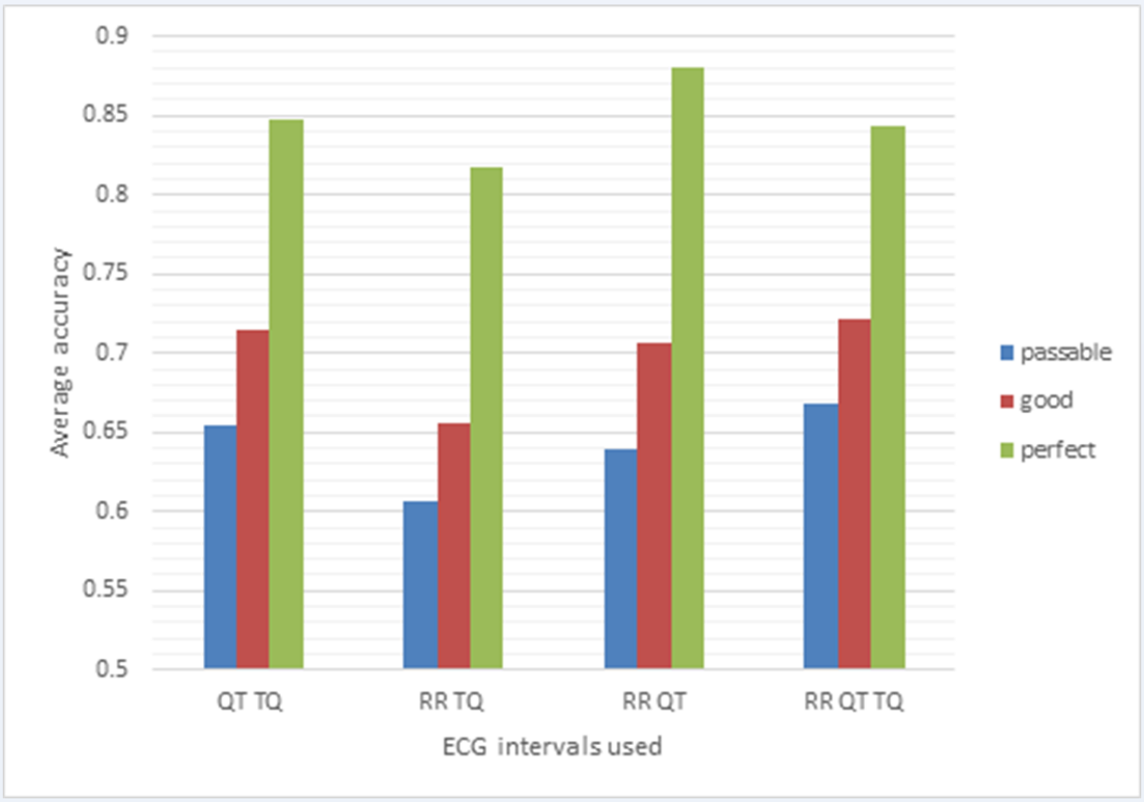


Fig 6: Accuracy plots for each combination of restitution intervals when using different levels of strip quality.

References and Acknowledgements

- Choe, William C. et al. 2015. “A Comparison of Atrial Fibrillation Monitoring Strategies After Cryptogenic Stroke (from the Cryptogenic Stroke and Underlying AF Trial).” *American Journal of Cardiology*
- Aston, P.J. *et al.* (2018) “Beyond HRV: Attractor reconstruction using the entire cardiovascular waveform data for novel feature extraction,” *Physiological Measurement*
- Lempel, A. and Ziv, J. (1976) “On the Complexity of Finite Sequences,” *IEEE Transactions on Information Theory*
- Huang, Y.H. *et al.* (2020) “ECG Restitution Analysis and Machine Learning to Detect Paroxysmal Atrial Fibrillation: Insight from the Equine Athlete as a Model for Human Athletes,”
- Pilia, N., Nagel, C., Lenis, G., Becker, S., Dössel, O., Loewe, A. (2020). ECGdeli - An Open Source ECG Delineation Toolbox for MATLAB

This work was funded by an NIHR grant awarded to Dr Kamalan Jeevaratnam.