Improved Arrhythmia Classification Using Select Morphological and Heart Rate Variability ECG Features.

Mark James Dunbar
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, United Kingdom
ec21896@qmul.ac.uk

***Abstract*—Implantable Cardioverter Defibrillators (ICDs) are devices used for the prevention of sudden cardiac death (SCD) through the early detection and correction of dangerous arrhythmia. ICD devices treat dangerous arrhythmia by administering a correctional electrical shock to dangerous arrhythmia events. Effective ICD treatment depends on the accurate discrimination of arrhythmia that require shock therapy from less dangerous forms of arrhythmia and normal heartbeats. This paper presents a machine learning pipeline for the improved classification of arrhythmia heartbeats from ECG signals using a combination of morphological and Heart Rate Variability (HRV) features. Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and Random Forest Classifier (RFC) algorithms were selected to classify heartbeat samples from 44 thirty-minute long ECG recordings using various selections of heartbeat features. Each model was trained and analysed using morphological-only, HRV-only, and combined morphological and HRV feature extraction techniques. 5-fold cross validation was used to reduce over-fitting and bias for model validation. The best performing classifier found was the SVM classifier used with combined HRV and Morphological features, with an F1 score of 0.920, and an accuracy of 97.99%. This model achieved an accuracy increase of approx. 4.5% over the best performing HRV-feature and Morphological-feature only models. These results suggest that the proposed combination of HRV and morphological features lead to a significant improvement in classification accuracy.**

***Index Terms*—Arrhythmia Classification, MIT-BIH, ECG, Heart Rate Variability, ICD**

I. Introduction

Arrhythmia is a general term used to describe an abnormal rhythm of the heart and can arise in many different forms. While most forms of arrhythmia are harmless, some such as Ventricular Tachycardia (VT) can lead to life-threatening events such as Sudden Cardiac Death (SCD), a condition that accounts for approximately 15-20% of all deaths (Podrid, 2022; Srinivasan and Schilling, 2018).

Research on the prevention of arrhythmia related SCD focuses heavily on the early identification and correction of dangerous forms of arrhythmia through the use of Implantable Cardioverter Defibrillator (ICD) devices. ICDs are small battery-operated devices designed specifically to treat life-threatening forms of ventricular arrhythmia. Through constant monitoring, ICDs identify and correct dangerous arrhythmic episodes as they occur, administering therapeutic electrical shocks to restore regular sinus rhythm (Iqbal, Butt, and Jamal, 2022). These devices have proven to be one of the most effective measures of SCD prophylaxis, with clinical studies highlighting a substantial reduction in mortality compared to other existing conventional therapies (Secretariat, 2005), however still face significant setbacks in the form of Inappropriate shock therapy. Inappropriate shock therapy is the event in which an electrical shock is administered falsely by an ICD due to misidentified cardiac events, resulting in an uncomfortable and distressing experience for ICD fitted individuals. The effect of prolonged exposure to inappropriate shock therapy has been linked to an increased risk of patient mortality (Rees et al., 2011), highlighting the importance of correctly identifying cardiac events which require ICD intervention. Inappropriate shocks can be triggered by a wide variety of mechanisms such as periods of increased physical activity, malfunctioning ICD leads, and anxiety (Kulkarni and Link, 2018), however in one study conducted by Cardoso *et al.* (2015), it was found that nearly 80% of all inappropriate ICD shocks are caused by the misclassification of Supra-Ventricular Tachycardia (SVT) as VT.

Machine Learning (ML) is a powerful tool that has been adopted across all stages of patient diagnosis and treatment in cardiac electrophysiology in recent years (Muthalaly and Evans, 2020). ML has proven to be a particularly strong candidate for tackling the challenge of inappropriate shock therapy by applying algorithms to optimise ICD programming (Tzeis et al., 2008). In this paper, several ML techniques will be employed to produce a model pipeline that takes an electrocardiogram signal (ECG) input signal vector, and correctly classifies individually identified heartbeats as normal, or one of several classes of arrhythmia. The classifications produced by the model should identify cases where ICD intervention is appropriate. A new method for arryhthmic heartbeat classification is proposed that uses a combination of heartbeat morphology and Heart Rate Variability (HRV) features to maximise SVT and VT discrimination. HRV is a marker used to quantify changes in a patient's autonomic rhythm, derived from beat to beat intervals of the heart.

Several studies (Parsi, 2021; Asha and Joseph, 2013; Ashtiyani et al., 2018) have identified HRV as a candidate marker for the classification of arrhythmia. Building upon these findings, we will evaluate the effectiveness of these features using a selection of machine learning classifiers.

II. Background

This section provides a detailed overview of the main concepts introduced in this paper, and will also serve as an opportunity to discuss related work and research.

A. QRS Morphological Analysis

Individual heartbeats on an ECG graph can be observed as a series of periodic deflections from the baseline of an ECG. These deflections represent the time evolution of electrical signals across the heart, which are responsible for the heart's mechanical function ("What is an electrocardiogram (ECG)?" 2019). A single sinus-rhythm heartbeat comprises of the P-wave, followed by the QRS complex, which is then also followed by the T-wave, as shown in Fig. 1. The QRS complex is the most prominent feature of a full cardiac cycle, and is the most commonly used feature to identify individual heartbeats.

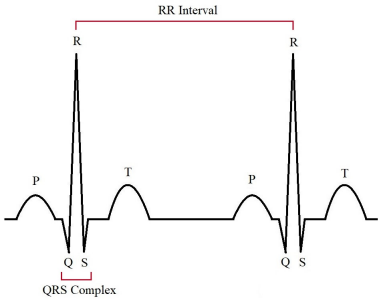

Fig. 1. Example of two consecutive sinus-rhythm heartbeats on an ECG. The QRS complex, RR interval and P and T waves are indicated.

B. Heart Rate Variability

Heart Rate Variability consists of a series of time-frequency measurements that can be used to analyse changes in the heart's rhythm in response to Autonomic Nervous System (ANS) activity (Singh et al., 2018). HRV metrics are derived from the RR intervals of the heart, also known as intra-beat intervals. An RR interval is the time in milliseconds between consecutive R-Peaks on the ECG signal (Fig. 1). Low HRV has been linked to an increased risk of mortality, and is thought to be a contributing factor in the patheogenesis of cardiac arrhythmia. In a study conducted by Fang *et al.* (2019), it was found that compared to patients with a higher HRV, low HRV patients faced a 46% higher risk of all-cause death, and life-threatening cardiac events such as VT, highlighting the significance of the relationship between HRV and overall cardiological function. Further to this, research conducted by Reed *et al.* (2005) found that significant decreases in HRV, particularly in the high frequency (HF), low frequency (LF) and very low frequency (VLF) variability components can precede the sudden onset of VT. In another study investigating the causes of ICD oversensing, Kossaify (2020) identified the analysis of QRS morphology features as the most effective method for SVT discrimination from other heartbeat classifications. Building upon these findings, this paper aims to evaluate the use of a combination of morphology and HRV features to improve VT and SVT heartbeat classification.

HRV consists of both time-domain and frequency-domain metrics. The time-domain metrics of HRV quantify the variability of RR intervals, and frequency-domain. The frequency-domain metrics of HRV are used to estimate the distribution of relative and absolute power among four frequency bands; ultra-low-frequency (ULF), very-low-frequency (VLF), low-frequency (LF) and high-frequency (HF) (Shaffer and Ginsberg, 2017). A minimum of 5 minutes of ECG data is recommended in Fred Shaffer and J. P. Ginsberg's 2017 overview of HRV metrics and norms in order to adequately capture the distribution of VLF, LF and HF frequencies from RR intervals. Based on the lower normal heart rate limit of 60bpm, a sample of 300 RR-intervals will be taken for each heartbeat to ensure at least 5 minutes of previous RR-intervals are captured (ClevelandClinic, 2022).

C. Machine Learning and Arrhythmia Detection

In this section, we will discuss several machine learning classifiers commonly used in heartbeat classification. In this paper, these models will be trained, analysed and compared using three different techniques for feature extraction on the same set of heartbeat data; morphological only features, HRV only features, and finally a fusion of both. Through this analysis, we can identify the performance impact of combining HRV and morphological data across a variety of commonly used ML models, as well as select the best performing model for the final classification pipeline.

1) Support Vector Machine: Support Vector Machines (SVMs) are supervised learning classifiers that use a higher dimensional hyperplane to maximally separate input vectors into two different classes (Shmilovici, 2005). SVM machine learning algorithms are strictly binary classifiers, however they can be adapted to suit multi-class classification problems using one of two methods. Many types of multiclass SVM classifiers have applied to the problem of arrhythmia classification (Kohli, Verma, and Roy, 2010). One of the main advantages SVM classifiers have over other models is that they are particularly robust to overfitting issues, as SVMs identify the maximum margin hyperplane, which maximises the distance between the hyperplanes and the datapoints belonging to each hyperplane. One of the shortcomings of SVM classifiers on the other hand, is that a significant amount of time is required to train them, particularly if a large dataset is used. This problem is particularly exacerbated when intensive model training and testing techniques such as Grid Search Cross Validation are used to train and validate models.

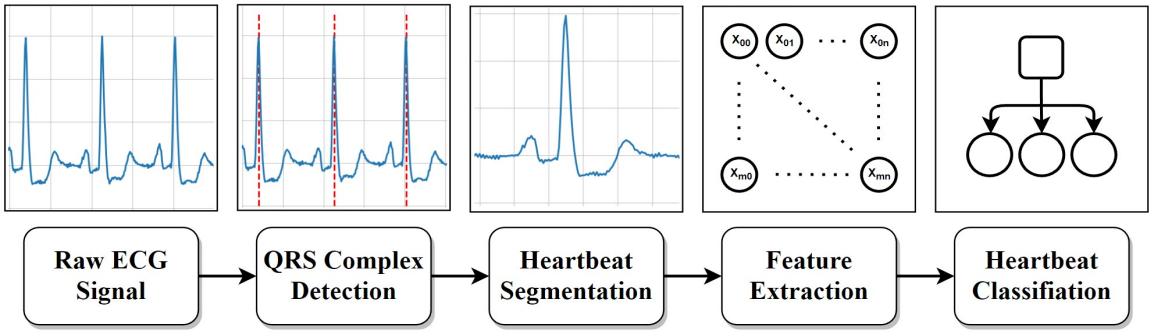

Fig. 2. Diagram of the arrhythmia classification pipeline, from raw ECG data to individual heartbeat classifications.

2) K-Nearest Neighbours: K-Nearest Neighbour (KNN) classifiers are supervised learning classifiers that use the proximity of predicting samples to the nearest k data points to make predictions (Zhang, 2016). KNN works on the assumption that similar data points are grouped closer together. Prediction samples are classified by selecting the k nearest neighbours by distance, using either the *euclidian* or *minkowski* distance. The majority class of these neighbours is then chosen as the predicted class of the prediction sample. The main advantage of KNN classifiers is that they do not require learning, making them much faster to implement than many other models. This means that the only model the KNN classifier requires is the entire dataset. The main drawback associated with using KNN classifiers is determining the optimal value of k. In cases where the nearest neighbours to a prediction sample are equally split between classifications, "tie-breaker" algorithms may have to be implemented to determine the classification of the new datapoint.

3) Random Forest Classifier: Random Forest Classifiers (RFCs) are non-linear ensembles of decision trees (DTs) that operate together to produce a final classification prediction by committee (Schonlau and Zou, 2020). A selection of DTs are fitted using various dataset sub-samples. When classifying a sample, each DT in the RFC votes for their own unique prediction for that sample. The majority of the votes is then used to assign the final prediction of the sample. RFCs are particularly powerful classification algorithms in that they are computationally inexpensive classifiers to use, can process large amounts of data, and are easy to interpret. The main drawback of RFCs, however, is the large number of hyper-parameters that must first be tuned in order to optimize the algorithm.

III. Methodology

A. Classification Pipeline

The model produced in this paper is designed to take ECG signal data as input, and produce a classification prediction for individual heartbeat signals. Figure 2 shows the pipeline of the proposed model for arrhythmia classification from ECG data. First, ECG signals are fed into the pipeline. Data pre-processing normally takes place in this step, however much of the pre-processing required has already been performed by the dataset provider, as discussed in the *Dataset* section below. Following this, QRS detection is used to identify the index of the R-Peaks of all individual heartbeats within the ECG. Next, individual heartbeat signals are extracted from the ECG signal by selecting several samples before and after the R-Peak. In the feature extraction stage, 301 preceding R-Peaks are used to determine 300 RR-intervals before the heartbeat, and calculate several HRV features. These features are then combined with the morphological heartbeat data extracted in the previous step to produce one final feature vector. Finally, the extracted feature vector is then passed to an ML classifier which determines the heartbeat classification. This paper will mainly focus on the ML classifier model used at the end of the model pipeline, however the significance and functionality of all stages in the proposed pipeline will also be explored and discussed.

B. Data Science Environment

The proposed model will be developed in a Python 3.6 environment, utilising several common data science libraries. The Scikit-Learn Python library (Pedregosa et al., 2011) was used in several roles, from data processing and train/validation dataset splitting, to classification model initialisation, model training, testing and validation. The commonly used Python libraries Pandas (McKinney, 2010) and Numpy (Harris et al., 2020) were used to format, transform, and display matrices data across all areas of the model pipeline. Finally, the AstroPy Python library (Price-Whelan et al., 2018) was used to produce the Lomb periodigram for HRV feature extraction, as discussed in the *Feature Extraction* section below.

C. Dataset

The decision was made to base the machine learning classifier on lead-II ECG data instead of ICD-captured data. This decision will be further discussed in the *Conclusion and Discussions* section of this paper. The ECG dataset consists of 201 thirty-minute long Lead-II ECGs, comprising of 3 ECG databases obtained from PhysioNet's open-source databases (Goldberger et al., 2000; Moody and Mark, 2001). These databases are:

- The MIT-BIH Arrhythmia dataset, consisting of 48 ECG records taken from 47 patients at Boston's Beth Israel Hospital and MIT laboratories.

- The MIT-BIH Supra-Ventricular Arrhythmia dataset containing 78 ECG records from 78 patients, also taken from the Boston's Beth Israel Hospital and MIT laboratories.
- The St Petersburg INCART 12-lead Arrhythmia database, consisting of 75 ECG records taken from 32 patients.

The ECG records in these databases were previously pre-processed by the dataset provider in order to re-sample each ECG signal to the lowest common sample-rate of 128 Hz, and also remove gain (Sharma, 2019). Baseline wander was previously removed from each ECG record by the dataset provider by applying median filtering to each ECG signal (Sharma, 2019). ECG records in the dataset contain two vectors; an ECG signal vector, and an ECG label vector. Both vectors have a standard length of 230,400 samples.

The accompanying label vector provides element-wise annotations for features in the ECG signal vector from one of 19 PhysioNet standard heart-beat annotations (PhysioNet, 2016). Annotations contained in the label vector are one of two classes; non-beat features, and beat-type annotations. Beat-type annotations in the label vector identify the index of heartbeat R-peaks in the ECG signal (see Fig. 1), and the associated beat-type annotation. These label vectors were both annotated and independently reviewed by two or more cardiology experts, with any discrepancies being rectified at the time of data collection (Moody and Mark, 2001).

D. Data Pre-Processing

Additional transformations were applied to the data, producing the final dataset to be used as the input for the classification pipeline. First, several ECG records deemed unsuitable for use were discarded from the dataset:

- Records 102 and 104 were removed from the dataset as they do not contain lead-II signals.
- Records 107 and 217 contained paced heartbeats and were removed.
- Record 207 contained ectopic heartbeats not marked by beat annotations and was removed.

Next, each label vector was remapped to heartbeat classifications recommended by the Association for the Advancement of Medical Instrumentation (AAMI) (2013). During the remapping process, any PhysioNet non-beat annotations were removed. The remaining PhysioNet beat annotations were replaced with the corresponding AAMI label, as shown in TABLE I, reducing the number of unique heartbeat classifications from 15 to 3. The resulting ECG dataset was used to produce an array consisting of each ECG signal of size 196×230400, and the corresponding AAMI label vector of size 196×230400. The ECG data matrix will represent the input of the classification pipeline introduced in the previous section, and will be referred to as the *ECG input dataset* hereafter.

The distribution of heartbeat classes within the dataset after mapping AAMI labels is shown in TABLE II. As a consequence of the large class imbalance between non-ectopic (N) and ectopic heartbeats (VT, SVT), the F1 score is chosen as the primary performance metric for analysing models used on this data. The F1 score also accounts for the consequence of false positive and false negative predictions, whereas the accuracy score metric does not. In the particular case of classifying arrhythmic heartbeats, a false positive prediction can result in serious events such as inappropriate shock therapy, or in the more dangerous case of a false negative; failure to treat a life-threatening arrhythmic episode.

TABLE I
AAMI RECOMMENDED HEARTBEAT CLASSIFICATIONS

AAMI class description	AAMI label	PhysioNet description and beat annotation
Non-ectopic beats	**N**	Normal beats (**N**) Left bundle branch block beat (**L**) Right bundle branch block beat (**R**) Bundle branch block beat (**B**)
Supra-ventricular ectopic beats	**S**	Atrial premature beat (**A**) Aberrated atrial premature beat (**a**) Nodal premature beat (**J**) Supraventricular ectopic beat (**S**) Atrial escape beat (**e**) Nodal escape beat (**j**) Supra-ventricular escape beat (**n**)
Ventricular ectopic beats	**V**	Ventricular flutter wave (**V**) Ventricular escape beats (**r**) Premature ventricular contraction (**E**)

TABLE II
DATASET HEARTBEAT CLASSIFICATION DISTRIBUTION

Heartbeat Classification	Percentage Distribution
N	88.59%
S	7.42%
V	3.99%

E. QRS Complex Detection

QRS complex detection is an essential step required to extract individual heartbeats from a larger ECG signal. There are several forms of existing QRS complex detectors in the field of ECG analysis (Liu et al., 2018), however QRS detection algorithms can be generalised to two main stages:

- **Signal Transformation** - where the ECG signal is transformed such that features of interest are maximised, allowing them to be easily distinguished from the rest of the signal.
- **Decision Ruling** - where a threshold is used to separate the desired features from the rest of the signal.

A method of QRS detection known as template matching was adopted for the QRS complex detection stage of the classifier pipeline. Template matching is a powerful technique commonly used in computer vision and signal processing. The aim of template matching is to extract a "well-known" feature from a much larger signal by matching a smaller artificial signal that resembles the desired feature (i.e. a QRS

complex) with the main signal (Nakai et al., 2014). In QRS template matching, the cross correlation of the template and the main signal is graphed by incrementally sliding a QRS-like template signal over the larger ECG signal, calculating their cross correlation similarity at each interval. The QRS detection algorithm proposed for use in the classification pipeline can be summarised by the following steps:

1) The ECG signal vector is normalised to ensure that the ECG signal and the QRS filter, f_{qrs}, are scaled correctly.
2) The ECG vector $\vec{x}$ is processed, using the QRS filter, f_{qrs}, to generate the similarity signal vector, $\vec{g}$.
3) A suitable threshold, τ, is chosen to extract the elements of the signal vector where $\vec{g} > \tau$, generating clusters of above-threshold index values for each R-peak.
4) The average of each above-threshold cluster is calculated to obtain the index of the identified R-Peak on the original ECG signal, $\vec{x}$.

The QRS filter used in this stage of the pipeline follows the discrete function:

$$f_{qrs} = sin(x) \tag{1}$$

where $\frac{3\pi}{2} < x < \frac{7\pi}{2}$.

Interpolating the function by increments of $\frac{2\pi}{15}$, generates a QRS template vector of 15 samples in length. A template size of 15 was selected to produce a filter length of 0.12 seconds, based on the upper limit for QRS complex duration, as recommended by the Nottingham University School of Health Sciences (2021). A threshold of $\tau = 0.3$ was chosen after testing the peak detection algorithm on the ECG input dataset for values of τ ranging between $\tau = 0.1$ and $\tau = 1$. The value of τ was chosen as the where the absolute value of the difference between the number of identified R-Peaks and the number of labelled R-Peaks in the dataset is minimised. Threshold values of $\tau < 0.3$ were found to sometimes include other ECG signal features such as P and T waves (see Fig. 1), and values of $\tau > 0.3$ lead to a significant amount of R-peaks being missed by the algorithm. For $\tau = 0.3$ it was found that the algorithm failed to detect a total of 16 R-Peaks within the entire dataset. Fig. 3 displays an example of the normalised ECG signal compared to the calculated similarity score graph, and the resultant R-peak locations.

F. Heartbeat Segmentation

R-Peak indexes obtained in the QRS detection stage were used to reduce each thirty-minute ECG signal into individual heartbeat signals and their corresponding labels. Heartbeat signals were captured using a window of 38 samples ($0.3s$) before each R-peak index, to 64 samples ($0.5s$) after, leaving a fixed-length signal of 102 samples. The 301 preceding R-peaks of each heartbeat identified in the QRS complex detection are also extracted to later calculate short-term HRV metrics for each heartbeat. The resulting output of the heartbeat segmentation stage is two data matrices; a heartbeat signal matrix of size 21142×102, and a preceding R-Peaks matrix of size 21142×301.

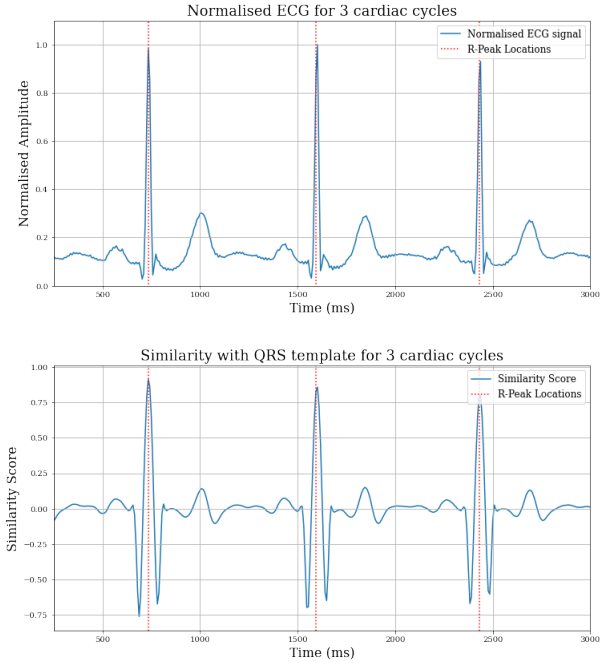

Fig. 3. Normalised ECG segment with the locations of 3 identified R-peaks (above) and the corresponding cross-correlation similarity score (below).

G. Feature Extraction

In the feature extraction step of the classification pipeline, three possible output feature sets will be considered for use. These combinations are; a feature set with only morphological features, a feature set with only the HRV features, and finally, the combined set of both heartbeat signals and HRV metrics. The feature combination used in this step of the classification pipeline will be chosen based on the performance of candidate models using each of the proposed feature set. The processes involved in extracting all of these features will be discussed in this section, however the final pipeline will adopt the best performing feature set after experimentation. Both morphological heartbeat metrics and HRV data from the preceding R-Peaks of each heartbeat are used to generate a feature matrix that will be passed on to the classifier. The extraction method for both categories of features are discussed below:

TABLE III
MORPHOLOGICAL METRICS

Segment	Window index range
pre-beat	[0:15]
p-wave	[15:30]
qrs complex 3	[30:50]
t-wave	[50:90]
post-beat 5	[90:102]

1) Morphological Features: The morphological features extracted from each heartbeat signal consist of five total amplitude metrics from different portions of the signal. This method of feature extraction follows a similar method used by V. Mondéjar-Guerra *et. al* (2019), where several amplitude

measurements were taken from different windowed segments of the beat signal. Five total morphological feature segments were used, each corresponding to a particular feature as shown in TABLE III.

TABLE IV
TIME-DOMAIN HRV METRICS

Metric	Units	Formula
$\overline{RR}$	ms	$\frac{1}{N}\sum_{i=1}^{N} RR_i$
$\overline{HR}$	bpm	$\frac{1}{N}\sum_{i=1}^{N} 60\left(\frac{1000}{RR_i}\right)$
$\overline{HR}_k$	bpm	$60 * \left(\frac{1000}{\overline{RR}}\right)$
HR_{max}	ms	$max\left(60 * \left(\frac{1000}{RR_i}\right)\right)$
HR_{min}	ms	$min\left(60 * \left(\frac{1000}{RR_i}\right)\right)$
$SDNN$	ms	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(RR_i - \overline{RR}\right)^2}$
$SDHR$	ms	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(60 * \left(\frac{1000}{RR_i}\right) - \overline{HR}\right)^2}$
$RMSSD$	ms	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(RR_{i+1} - RR_i\right)^2}$
$NN50$		$count((RR_{i+1} - RR_i) > 50ms)$
$pNN50$	%	$\frac{NN50}{N-1} * 100$

$N = 300$ is the number of RR intervals.

2) HRV Features: HRV features are extracted using the preceding RR intervals of each heartbeat. Before calculating RR-intervals, the preceding R-Peak sample indexes, $\vec{R_s}$ are converted to R-Peak timestamps, $\vec{R_t}$ in milliseconds using the following conversion:

$$\vec{R_t} = \left(\frac{1000}{128}\right)\vec{R_s} \tag{2}$$

Next, the RR intervals are calculated from preceding R-Peaks for each beat:

$$RR_i = R_{i+1} - R_i \tag{3}$$

After the vector of RR intervals, $\vec{RR}$ is calculated, HRV metrics can be extracted through a series of calculations. First, the time-domain HRV features will be calculated from RR data. Each of the equations used to calculate the time-domain HRV metrics are shown in TABLE IV accordingly. There are 10 time-domain HRV metrics in total, which are:

- $\overline{RR}$ - the mean RR interval (ms).
- $\overline{HR}$ - the mean heart rate from each RR interval (bpm).
- $\overline{HR}_k$ - the Kubios style heart rate from the mean RR interval $\overline{RR}$, (bpm).
- HR_{max} - the maximum heart rate from all RR values (bpm).
- HR_{min} - the minimum heart rate value from all RR values (bpm).
- $SDNN$ - the standard deviation of all RR intervals (ms).
- $SDHR$ - the standard deviation of all heart rate values (bpm).
- $RMSSD$ - the root mean square of the difference between successive RR intervals.
- $NN50$ - the number of adjacent RR intervals in the RR data that vary by greater than 50ms.
- $pNN50$ - the percentage of adjacent RR intervals that differ from each other by more than 50ms.

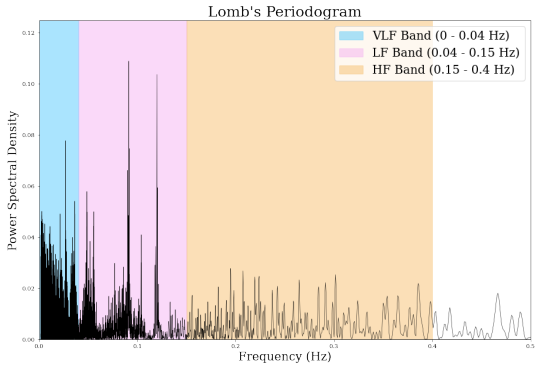

Fig. 4. Lomb's Periodigram for a sample heartbeat's short-term HRV data.

Following this, the frequency-domain HRV features are extracted by applying a periodigram to the RR-intervals. A periodogram is used to identify the dominant frequencies present in a signal by determining the power density distribution (PSD) of frequencies in the signal. As RR intervals are a series of unevenly spaced temporal measurements, a regular Fast Fourier Transform (FFT) cannot be used to produce a periodigram without first interpolating the data to evenly spaced temporal samples, as RR intervals are not evenly spaced time measurements. The Lomb periodigram is a method allows for the extraction of periodic signals from unevenly spaced temporal data, such as RR intervals.

Lomb's method was adopted based on the findings of research conducted by G. D. Clifford (2002), which concluded that significantly better HRV feature estimates were yielded by the Lomb method when compared with FFT. Applying the Lomb periodigram method to RR interval values yields the PSD (ms^2/Hz) across the frequency spectrum. The periodigram PSD is then split into VLF, LF and HF bands, as shown in Fig. 4. These bands are subsequently used to calculate the frequency-domain HRV features described in TABLE V, using the following methods:

- P_{VLF}, P_{LF} and P_{HF}, calculated as the sum of PSD values within each respective frequency band.
- P_{Total}, the sum of the P_{VLF}, P_{LF} and P_{HF} values.
- VLF_{peak}, LF_{peak} and HF_{peak}, the highest frequency values (Hz) in each respective frequency band.
- The LF/HF ratio, calculated using the equation $\frac{P_{LF}}{P_{HF}}$
- LF_{frac}, calculated using the equation $\frac{P_{LF}}{P_{Total}}$

- HF_{frac}, calculated using the equation $\frac{P_{HF}}{P_{Total}}$

Finally, each of the features extracted, x are standardized to a new values, z, by subtracting the mean, and dividing by the standard deviation, as shown in equation 4.

$$z = \frac{x - \mu}{\sigma} \tag{4}$$

This reduces any bias caused by variation in scale between features.

TABLE V
FREQUENCY-DOMAIN HRV METRICS

Metric	Units	Description
P_{VLF}	ms^2	HRV power density in VLF range
P_{LF}	ms^2	HRV power density in LF range
P_{HF}	ms^2	HRV power density in HF range
P_{Total}	ms^2	Total HRV power density (0 - 0.4 Hz)
VLF_{peak}	Hz	Peak frequency in VLF band
LF_{peak}	Hz	Peak frequency in LF band
HF_{peak}	Hz	Peak frequency in HF band
LF/HF		LF to HF power ratio
LF_{frac}	%	Fraction of LF in the full 0-0.4 Hz band
HF_{frac}	%	Fraction of HF in the full 0-0.4 Hz band

H. Classifier Training and Evaluation

Three candidate classifiers were chosen for the classification stage of the model pipeline; An SVM Classifier Ensemble, A KNN Classifier, and a RF classifier. Each of these models was trained, tested, and validated in three experiments, with the goal of analysing classification performance on the same set of heartbeats using different combinations of extracted features:

1) Classification performance with morphological features only.
2) Classification performance with HRV features only.
3) Classification performance with combined morphological and HRV features.

The general model training, testing and validation strategy used in this paper can be broken into 4 stages:

1) Feature Dataset Generation: To perform each experiment, 3 model training and validation datasets were extracted from the ECG input dataset. To represent the output from the feature extraction stage of the pipeline, each ECG is split into individual beats signals and their 301 preceding R-Peaks. The label corresponding to each individual beat is also extracted to produce the labels for the training dataset. Finally, the three datasets are generated by performing the relevant feature extraction steps explained in the *Feature Extraction* section. The result is three heartbeat datasets and corresponding labels; a dataset consisting of individual heartbeat morphological features, a feature dataset consisting of individual heartbeat HRV features, and finally a feature dataset consisting of individual heartbeat morphological and HRV features.

2) Train/Validation Dataset Split: The feature datasets for each experiment were split into training and validation feature datasets using a stratified approach, which preserves the distribution of classes between the entire dataset and the training dataset as best as possible, ensuring that the training data is a good representation of the full dataset. After stratification, the percentage distribution of each class in the training data is held to within 2 significant figures of the overall dataset.

TABLE VI
MODEL HYPERPARAMETERS AND DESCRIPTIONS

Model	Hyper-parameters	Hyper-parameter Description
SVM	C	Penalty value for misclassifications.
	degree	Degree of hyperplane polynomial.
	γ	Curvature of the hyperplane.
KNN	N_{nn}	Number of nearest neighbours.
	metric	Measurement method (i.e. Euclidean).
	weights	Weight neighbours' influence.
RFC	*max depth*	Maximum tree depth.
	min samples split	Minimum samples to split a node.
	N_{est}	Number of decision trees to use.

3) Classifier Training and Hyperparameter Tuning: Each model was trained on each of the feature datasets using Grid Search Cross-Validation (GSCV). In the case of the KNN classifier, which does not require training, GSCV allows for the selection of optimal k values for each dataset. GSCV allows an array of hyperparameters to be tested to identify the best model hyperparameter combination. GSCV iterates through each hyperparameter to evaluate, performing k-fold cross validation each time. A fold value of k = 5 is used for all experiments performed in this paper. GSCV splits the data passed to it into k folds of equally sized data sets. Four of the five folds are used to train the model and each unique combination of hyperparameters, and 1 validation fold is used to calculate the F1 score. This step is repeated, each time using a different fold for the validation step. The mean F1 score is then calculated for all 5 iterations. Once GSCV has returned the mean F1 score for every unique combination of model hyperparameters, the hyperparameter combination with the highest F1 score will be returned as the "best" hyperparameter settings for the model. A list of the hyperparameters to be tested for each model with GSCV is shown in TABLE VI.

4) Classifier Validation: For each experiment, each of the "best" tuned models identified in the GSCV stage were validated using the validation dataset. Using k-fold cross validation on the validation dataset, a mean F1 score for k=5 folds was used to return the mean F1 score for each model's performance on the validation dataset.

IV. RESULTS

The performance of each candidate model using the three feature extraction methods discussed are presented below. As discussed in the *Methodology* section, F1 scores will be used as the primary metric for comparing the performances of classifiers.

A. Classifier Performance Using Morphological Features

Trained and tested models were validated using the 20% of the morphological feature dataset as discussed in the methodology section. We see that the KNN classifier performs the best, with an F1 score of 0.673, followed by the RFC with an F1 score of 0.668, and finally, the SVM with an F1 score of 0.644.

TABLE VII
MODEL PERFORMANCE FOR MORPHOLOGICAL FEATURES

Model	Best Hyperparameters	F1 Score	Accuracy
SVM	$C = 10$ degree $= 0$ $\gamma = 0.7$	0.644	92.47%
KNN*	$N_{nn} = 5$ metric $= minkowski$ weights $= distance$	0.673	93.51%
RFC	max depth $= 20$ min samples split $= 2$ $N_{est} = 100$	0.668	93.76%

*Best performing classifier model.

B. Classifier Performance Using HRV Features

Models trained and tested on HRV only data were validated using the 20% validation portion of the data using only HRV features. The SVM classifier produced the highest F1 score of 0.678, followed by KNN with an F1 score of 0.672, and finally, the RFC with an F1 score of 0.644.

TABLE VIII
MODEL PERFORMANCE FOR HRV FEATURES

Model	Best Hyperparameters	F1 Score	Accuracy
SVM*	$C = 5$ degree $= 0$ $\gamma = 0.6$	0.678	93.53%
KNN	$N_{nn} = 5$ metric $= minkowski$ weights $= distance$	0.672	92.55%
RFC	max depth $= 25$ min samples split $= 2$ $N_{est} = 75$	0.644	92.18%

*Best performing classifier model.

C. Classifier Performance Using Morphological and HRV Features

In the final experiment, models were trained and tested using a combination of HRV and Morphological features, and validated using the remaining 20% validation portion of the data. The SVM classifier produced the highest F1 score of 0.920 in this experiment, followed by the RFC with an F1 score of 0.901. The KNN algorithm came last, with an F1 score of 0.882.

TABLE IX
MODEL PERFORMANCE FOR MORPHOLOGICAL AND HRV FEATURES

Model	Best Hyperparameters	F1 Score	Accuracy
SVM*	$C = 6$ degree $= 0$ $\gamma = 0.01$	0.920	97.99%
KNN	$N_{nn} = 3$ metric $= minkowski$ weights $= distance$	0.882	97.36%
RFC	max depth $= 30$ min samples split $= 2$ $N_{est} = 125$	0.901	97.39%

*Best performing classifier model.

V. CONCLUSION AND DISCUSSIONS

In this paper, an ML pipeline for ECG arrhythmia heartbeat classification has been successfully developed. Through a new method of feature extraction combining HRV and Morphological features, a significant increase in model F1 score and accuracy has been demonstrated over HRV and Morphological only feature-trained models. Combined HRV and morphological features are shown to yield an approx. increase in F1 score of 0.25, and an increase of 4.5% in accuracy across the best models for each feature set. This result is also shown to hold true across the various classification models used, with SVM, KNN and RFC performances all following the same trend of increasing accuracy and F1 score when combined HRV and morphological features are used.

The main limitation faced by this work is that ICD-captured data was not used to train and analyse the techniques discussed in this paper. After careful consideration, it was concluded that the quality of ECG data provided by PhysioNet's Open Access database was of superior quality and size to any available ICD signal datasets. ECG Lead-II data signals approximate closely the signals recorded by an ICD, meaning that the analysed techniques and results found in this paper should be replicable with ICD data, and should be considered as the most significant area of further work for this paper.

There are many areas in which the overall pipeline can also be improved in further research. The R-peak detection method used had an under-sensing issue, resulting in several heartbeats not being detected. In testing, the proposed QRS detection method missed 16 R-Peaks in total, accounting for $1\mathrm{e}{-10}\%$ of all beats in the dataset. While this is a small portion of the dataset, missed peaks could result in life-threatening events such as arrhythmia being missed, ultimately leading SCD and patient mortality. The QRS complex detection stage is as critical as the classifier model chosen for the final classification stage of the pipeline. QRS complex detection may be improved upon in further iterations by replacing the QRS detection method with a more well-known method such as the Pam-Tomkins QRS detection algorithm (Sathyapriya, Murali, and Manigandan, 2014). In addition to this, it may be possible to use the cross-correlation scores from QRS template matching as features in the classification stage as proposed in reserch conducted by V. Krasteva and I. Jekova (2007).

The improvement of classifier performance with HRV measurement times beyond 5 minutes may also be explored in further research. To explore this, an analysis of the impact of longer-term HRV measurements on classifier accuracy at RR interval monitoring periods of 30 minutes and 24 hours may be conducted.

Acknowledgments

I would like to extend my sincere thanks and appreciation to my supervisor, Dr Jesús Requena Carrion for his guidance and support through all stages of this project.

Additionally, I would like to express my deepest gratitude to Jack Webber and Akash Heer, both of whom offered their incredibly valuable discussions and advice in the field of cardiology.

To my Mum, Dad, Cameron, Olivia, family and friends, thank you for always supporting me and giving me the confidence to achieve my goals.

Finally I would like to thank my partner Alyza Gines (and the Silverpoint crew), who supported me in a multitude of ways throughout this project and degree. In particular, her clinical knowledge and moral support during the tougher periods of my postgraduate studies. I simply could not have done this without you.

References

[1] Philip J. Podrid. *Overview of sudden cardiac arrest and sudden cardiac death*. (Date Accessed 2022-05-05). 2022. URL: https://www.uptodate.com/contents/overview - of - sudden - cardiac - arrest - and - sudden - cardiac-death.

[2] Neil T Srinivasan and Richard J Schilling. "Sudden cardiac death and arrhythmias". In: *Arrhythmia &; Electrophysiology Review* 7 [2018], p. 111. DOI: 10.15420/aer.2018:15:2.

[3] Arshad Muhammad Iqbal, Nausharwan Butt, and Syed F. Jamal. *Automatic Internal Cardiac Defibrillator*. https://www.ncbi.nlm.nih.gov/books/NBK538341/(Date Accessed 2022-07-09). StatPearls Publishing, 2022.

[4] Medical Advisory Secretariat. "Implantable cardioverter defibrillators. Prophylactic use: an evidence-based analysis". In: *Ontario Health Technology Assessment Series 2005;5(14):1-74. Epub 2005 Sep 1*. [2005]. URL: https://pubmed.ncbi.nlm.nih.gov/23074465/.

[5] Johannes van Rees et al. "Inappropriate implantable cardioverter-defibrillator shocks". In: *Journal of the American College of Cardiology* 57 [2011], 556–562. DOI: 10.1016/j.jacc.2010.06.059.

[6] Nitin Kulkarni and Mark S. Link. "Causes and prevention of inappropriate implantable cardioverter-defibrillator shocks". In: *Cardiac Electrophysiology Clinics* 10 [2018], 67–74. DOI: 10.1016/j.ccep.2017.11.006.

[7] Rhanderson N. Cardoso et al. "ICD discrimination of SVT versus VT with 1:1 V-A conduction: A review of the literature". In: *Indian Pacing and Electrophysiology Journal* 15 [2015], 236–244. DOI: 10.1016/j.ipej.2016.02.006.

[8] Rahul G Muthalaly and Robert M Evans. "Applications of machine learning in cardiac electrophysiology". In: *Arrhythmia amp; Electrophysiology Review* 9 [2020], 71–77. DOI: 10.15420/aer.2019.19.

[9] S. Tzeis et al. "Tools and strategies for the reduction of inappropriate implantable cardioverter defibrillator shocks". In: *Europace* 10 [2008], 1256–1265. DOI: 10.1093/europace/eun205.

[10] Ashkan Parsi. "Improved cardiac arrhythmia prediction based on heart rate variability analysis". PhD thesis. 2021.

[11] N. D. Asha and K. Paul Joseph. "Linear and nonlinear analysis of heart rate variability in ventricular tachycardia and supraventricular tachycardia". In: *Journal of Medical Imaging and Health Informatics* 3 [2013], 294–300. DOI: 10.1166/jmihi.2013.1162.

[12] M Ashtiyani et al. "Heart rate variability classification using support vector machine and genetic algorithm". In: *Journal of Biomedical Physics and Engineering* [2018]. DOI: 10.31661/jbpe.v0i0.614.

[13] "What is an electrocardiogram (ECG)?" In: *InformedHealth.org [Internet]*. Institute for Quality and Efficiency in Health Care (IQWiG), 2019.

[14] Nikhil Singh et al. "Heart rate variability: An old metric with new meaning in the era of using mHealth Technologies for Health and exercise training guidance. part One: Physiology and methods". In: *Arrhythmia amp; Electrophysiology Review* 7 [2018], p. 193. DOI: 10.15420/aer.2018.27.2.

[15] Su-Chen Fang, Yu-Lin Wu, and Pei-Shan Tsai. "Heart rate variability and risk of all-cause death and cardiovascular events in patients with cardiovascular disease: A meta-analysis of Cohort studies". In: *Biological Research For Nursing* 22 [2019], 45–56. DOI: 10.1177/1099800419877442.

[16] M.J. Reed, C.E. Robertson, and P.S. Addison. "Heart rate variability measurements and the prediction of ventricular arrhythmias". In: *QJM* 98 [2005], 87–95. DOI: 10.1093/qjmed/hci018.

[17] Antoine Kossaify. "Sensing and Detection Functions in Implantable Cardioverter Defibrillators: The Good, the Bad and the Ugly". In: *Acta Cardiol Sin* 36 [2020], 308–317. DOI: https://doi.org/10.6515\%2FACS.202007_36(4).20191201A.

[18] Fred Shaffer and J. P. Ginsberg. "An overview of heart rate variability metrics and norms". In: *Frontiers in Public Health* 5 [2017]. DOI: 10.3389/fpubh.2017.00258.

[19] ClevelandClinic. *What to know about your heart rate and pulse*. (Date Accessed 2022-05-12). 2022. URL: https://my.clevelandclinic.org/health/diagnostics/17402-pulse--heart-rate.

[20] Armin Shmilovici. "Support Vector Machines". In: *Data Mining and Knowledge Discovery Handbook*. Springer US, 2005, 257–276.

[21] Narendra Kohli, Nishchal K. Verma, and Abhishek Roy. "SVM based methods for arrhythmia classification in ECG". In: *2010 International Conference on Computer and Communication Technology (ICCCT)* [2010], pp. 486–490. DOI: 10.1109/iccct.2010.5640480.

[22] Zhongheng Zhang. “Introduction to machine learning: K-Nearest Neighbors”. In: *Annals of Translational Medicine* 4 [2016], 218–218. DOI: 10.21037/atm.2016.03.37.

[23] Matthias Schonlau and Rosie Yuyan Zou. “The Random Forest Algorithm for Statistical Learning”. In: *The Stata Journal: Promoting communications on statistics and Stata* 20 [2020], 3–29. DOI: 10 . 1177 / 1536867x20909688.

[24] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 [2011], pp. 2825–2830.

[25] Wes McKinney. “Data Structures for Statistical Computing in python”. In: *Proceedings of the Python in Science Conference* [2010], pp. 51–56. DOI: 10.25080/majora-92bf1922-00a.

[26] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585 [2020], 357–362. DOI: 10 . 1038/s41586-020-2649-2.

[27] A. M. Price-Whelan et al. “The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package”. In: *The Astronomical Journal* 156 [2018], p. 123. DOI: 10.3847/1538-3881/aabc4f. URL: https://doi.org/10.3847/1538-3881/aabc4f.

[28] Ary L. Goldberger et al. “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals.” In: *Circulation* 101 [2000], 215–220. DOI: 10.1161/01.cir.101.23.e215.

[29] G.B. Moody and R.G. Mark. “The impact of the MIT-BiH Arrhythmia Database”. In: *IEEE Engineering in Medicine and Biology Magazine* 20 [2001], 45–50. DOI: 10.1109/51.932724.

[30] Nelson Sharma. *ECG Lead 2 Dataset PhysioNet (Open Access)*. (Date Accessed 2022-06-06). 2019. URL: https://www.kaggle.com/.

[31] PhysioNet. *PhysioBank Annotations*. 2016. URL: https://archive.physionet.org/physiobank/annotations.shtml.

[32] Association for the Advancement of Medical Instrumentation (AAMI). “Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms”. In: *AAMI EC57*. 12th ed. Vol. 57. Association for the Advancement of Medical Instrumentation, 2013.

[33] Feifei Liu et al. “Performance analysis of ten common QRS detectors on different ECG application cases”. In: *Journal of Healthcare Engineering* 2018 [2018], 1–8. DOI: 10.1155/2018/9050812.

[34] Yozaburo Nakai et al. “Noise tolerant QRS detection using template matching with short-term autocorrelation”. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* [2014], 34–37. DOI: 10.1109/embc.2014.6943522.

[35] University of Nottingham. *A Beginners Guide to Normal Heart Function, Sinus Rhythm amp; Common Cardiac Arrhythmias*. (Date Accessed 2022-07-16). 2021. URL: https://www.nottingham.ac.uk/nursing/practice/resources/cardiology/function/normal_duration.php.

[36] V. Mondéjar-Guerra et al. “Heartbeat classification fusing temporal and morphological information of ecgs via ensemble of classifiers”. In: *Biomedical Signal Processing and Control* 47 [2019], 41–48. DOI: 10.1016/j.bspc.2018.08.007.

[37] Gari D Clifford. “Signal Processing Methods for Heart Rate Variability”. PhD thesis. 2002.

[38] L. Sathyapriya, L. Murali, and T. Manigandan. “Analysis and detection R-peak detection using modified Pan-Tompkins algorithm”. In: *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies* [2014], pp. 483–487. DOI: 10.1109/icaccct.2014.7019490.

[39] Vessela Krasteva and Irena Jekova. “QRS template matching for recognition of ventricular ectopic beats”. In: *Annals of Biomedical Engineering* 35 [2007], 2065–2076. DOI: 10.1007/s10439-007-9368-9.