Department of Computer Science

Submitted in part fulfilment for the degree of BEng

Arrhythmia Detection and Classification Using Evolutionary Algorithms

Sarah Berry

03 May 2023

Supervisor: Christopher Crispin-Bailey

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

TABLE OF EQUATIONS

1 Executive Summary

Arrhythmias are abnormalities in the heart's rhythm and are a prevalent health issue affecting over 2 million people in the UK [1]. They can be detected and classified from an electrocardiogram that records the electrical signal in the patient's heart. Increasing the accuracy of classification algorithms would help healthcare professionals quickly identify issues with minimal intervention and analysis, reducing time spent on diagnostics. This technology would also improve at home and wearable devices where trained personnel may not be available.

While classical neural networks are the most widely used classification method for arrhythmias, this paper investigates whether evolutionary algorithms can offer any potential benefits to this area of research. Specifically, this study uses BioHEL [2], a Bioinformatics-oriented Hierarchical Evolutionary Learning system. The main objective of this paper is to evaluate the effectiveness of BioHEL, which employs an Iterative Rule Learning methodology, in contrast to GAssist, which utilizes the Pittsburgh approach that has been utilised in prior studies.

The data obtained from the MIT-BIH database underwent pre-processing using bandpass filters. The squared double difference method was used to detect R peaks, and the signals were segmented accordingly. Following this, a two-phase procedure was used for arrhythmia detection and classification. Combined with the proposed pre-processing methods, BioHEL achieved an F1-score of 96.6% for arrhythmia detection and 96.2% for arrhythmia classification. BioHEL was determined to be notably superior to GAssist in detecting arrhythmias, with a false negative rate of 3.2%, a reduction of 14%. All aims outlined in the evaluation criteria were met, demonstrating value in the use of evolutionary algorithms in this area of study.

Future research could investigate further configuration file optimisation for different class sizes. Further data cleaning and pre-processing could be done with the help of medical professionals which would increase the accuracy of arrhythmia detection. Improving the generality of the classifiers could also be explored using BioHEL’s Rule Post-Processing Engine.

1.1 Evaluation Criteria

The following criteria are defined to evaluate the merit of using evolutionary algorithms to determine whether future exploration could be beneficial.

- F1 score for detecting arrhythmias ≥ 90%
- Optimise approach to reduce false negative rate ≤15%
- F1 score for classifying arrhythmias ≥ 90%

Maintaining a low false negative rate is crucial in healthcare to avoid any potential undetected issues that may require medical attention. In assessing the performance of the classification algorithms, F1 scores will be used instead of accuracy as it is a more appropriate performance metric for unbalanced classes.

2 Literature Report

Classification of arrhythmias using AI is a widely researched topic, with the main approach being the following steps:

1. Noise reduction
2. R peak detection
3. Segmentation of signal
4. Classification of arrhythmias AI

The first three steps are necessary to produce distinct QRS complexes which are annotated as being normal or with an arrhythmia. These complexes are then fed to the classifier as a single record in the dataset.

2.1 Noise Reduction

Noise is present in ECG signals due to patient movement, electromagnetic interference, equipment quality, and skin preparation. To ensure accurate detection of R peaks in the next step, it is necessary to apply noise reduction. Failure to do so may result in missed or over-detected R peaks, leading to a decrease in the number of samples or incorrect complexes being fed into the AI, ultimately causing inaccuracies and errors.

High-frequency noise can be caused by electromagnetic interference and is in the range 50-60Hz. Whilst this can be physically minimised by shielding leads and lowering the skin-electrode impedance, some noise will still be present. Low-frequency noise, which occurs due to patient movement like breathing and appears as a baseline wander, has a frequency below 0.5Hz. Intermittent signals may also occur in the signal due to equipment faults or incorrect usage.

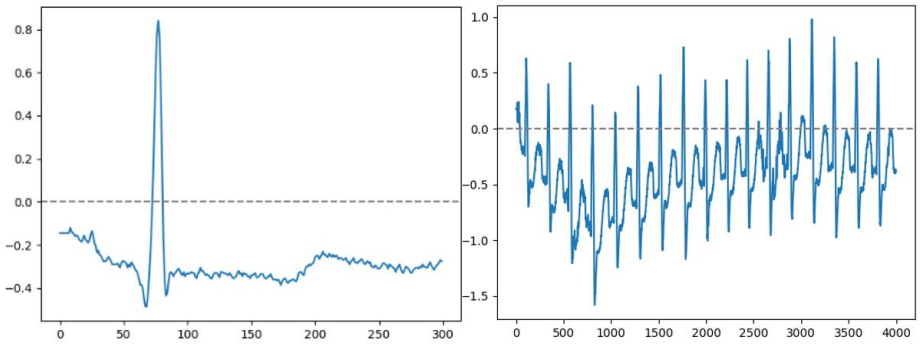

Figure 1: *On left, plot of ECG signal with high-frequency noise. On right, plot of ECG signal with low-frequency noise*

Four noise reduction techniques are detailed below. It can also be noted that one paper focusing on real-time analysis using edge devices uses no pre-processing on the data to reduce computation load [3].

Butterworth filters

A Butterworth bandpass filter, first described by Stephen Butterworth in 1930 [4], is used to allow a specified frequency band of signals to pass through while removing other frequencies. The Butterworth filter has a flat frequency response in the passband, reducing distortion in the original signal.

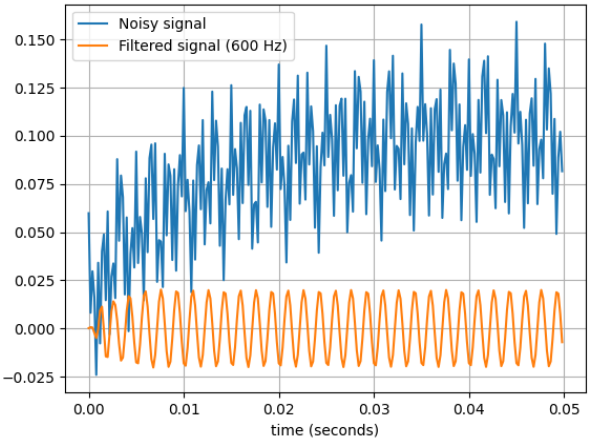

Figure 2: *Example Butterworth bandpass filter as outlined in SciPy cookbook [5]*

Zhang et al [6] utilised a Butterworth filter with a bandpass of 0.5-45 Hz. When combined with their neural network model, it achieved an accuracy of 96% in detecting atrial fibrillation.

Butterworth filter and polynomial function subtraction

Another approach is to apply a low-pass Butterworth filter to reduce high-frequency noise and a baseline subtraction method to reduce baseline wander. One paper [7] used this approach with a cut-off frequency of 100 Hz with an order of 10 for the low pass filter. Then, to reduce baseline wander, a polynomial function of order 8 was fitted to the signal along the isoelectric line (baseline). This is then subtracted from the filtered ECG signal.

No analysis was done in this paper on the effectiveness of the pre-processing however in combination with their R peak detection achieved an impressive sensitivity of 100% when segmenting ECG signals.

Wavelet transforms

Zhang [8] explores the wavelet approach for baseline wander correction and noise reduction. Firstly, a wavelet function and resolution level are selected to capture the baseline. A symlets wavelet with order 10 is suggested. This function is then removed from the original signal to remove the baseline wander. Wavelet shrinkage denoising is then applied, originally proposed by Donoho et al [9]. This is achieved by taking a discrete wavelet transform of the signal, thresholding the coefficients and then taking the inverse wavelet transformation to obtain the smoothed signal. No statistical evaluation was performed on this method; however, a sample signal shows its ability to remove high and low-frequency noise effectively.

Wavelet transforms and median filtering

Cheng [10] proposed a method for noise reduction that also utilises wavelet shrinkage denoising but in combination with median filtering to reduce noise. In this method, large values are removed to obtain the trend of the signal. This is then superimposed with the original signal to eliminate baseline wander. A median filter is then applied with a neighbourhood size of 9. In combination with the neural network proposed in this paper, only using the wavelet transform produced an overall F1 score of 0.864 and an accuracy of 0.885. Just using a median filter transform produced a marginally greater overall F1 score of 0.887 but only an accuracy of 0.884. Combining these two pre-processing methods on the same neural network had the greatest results with an F1 score of 0.887 and an accuracy of 0.888.

2.2 R Peak Detection and Segmentation

To classify beat arrhythmias, ECG signals need to be split into segments with a single discrete QRS complex. To do this, R peaks are detected using de-noised data and then split into windows around the peaks. This prepares the data to be fed into a classifier.

Pan–Tompkins algorithm

A widely known method for signal segmentation is the Pan–Tompkins algorithm [11] which is designed to work in real-time. To detect R peaks, a derivative filter is applied to the signal.

$$H(z) = (18T)(-z^{-2} - 2z^{-1} + 2z + z^{2})$$

Equation 1: *Pan-Tompkins derivative filter*

A 5-point filter with a processing delay of 2 samples and a gain of 0.1 is recommended. The resultant signal is then squared, making all points positive and amplifying peaks. This reduces the number of T peaks being detected as R peaks. A moving-window integration is then applied. This method is highly accurate, only failing to detect 0.675 percent of the beats on the MIT/BIH arrhythmia database [12, 13].

Fast Fourier Transform

A Fast Fourier transform can be used to extract features from the signal by projecting it onto a set of functions which is utilised to find the R peaks [14]. Firstly, FFT is applied to the signal. Then the inverse FFT is then applied to the result. The resultant signal can then be filtered to detect R peaks. This paper does not directly compare using FFT against another segmentation method on the same neural network, however, the combined effectiveness of FFT and the proposed CNN will be discussed in the next section.

Squared Double Difference

As the amplitude differences are proportional to the derivative of the signal, squared double difference can be used to detect R peaks [15]. For each value in the signal, the squared double difference is calculated and stored in a resultant list. This list is sorted into descending order and the peaks above a 3% threshold are classified as possible R peaks.

This method produces an overall detection sensitivity of 99.8% meaning it fails to detect around 0.2% of peaks. This result is a slight improvement on the Pan–Tompkins algorithm although tested on a different database. For the above methods, once the R peak has been detected, a segment can be defined as all values within a given range of a peak.

2.3 Classification

CNNs and DNNs are the most used classifiers for ECG arrhythmias. One proposed DNN called ENCASE [16] reported an F1-score of 84% when tested on 4 categories, normal sinus rhythm, two arrhythmia types and noise. Subsequently, an improved version of AlexNET was proposed [17] which in combination with FFT pre-processing, resulted in an F1-Score of 98%. when tested on 4 classes, normal sinus rhythm and three arrhythmia types.

However, this paper is exploring the use of evolutionary algorithms. There have been studies into the classification of arrhythmias using

GAs, although more limited than classic neural networks. This is where a population of individuals are generated. The fitness of each individual is then evaluated. The individuals are crossed over and mutated to create a new population and evaluation is repeated. This cycle continues until an optimum solution is found.

KELM

The first paper trained a kernel extreme learning machine (KELM), tuning its parameters using a GA [14]. Complex pre-processing was done using wavelet decomposition and moving average filters. It performed well, achieving 99.24% on two classes (Normal vs abnormal) on the PTBDB database [18] [12] and 89.74% on 13 classes on the UCI database [19].

GAssist-ADI

The second paper focuses on the comparison of GA techniques, namely GAssist-ADI, PSO-LDA, GFS-AdaBoost, GANN, SLAVE, XCS, UCS, PSO-ACO and AntMiner [20]. Before classifying, instances of ventricular premature contraction, supraventricular premature contraction, left ventricular hypertrophy, and atrial fibrillation were removed from the dataset. After pre-processing, classification was split into two phases, the first classifies if the signal is normal or abnormal. In phase two, the normal ECG signals are removed, and the arrhythmia types are classified. Six arrhythmia types are split into 3 groups as follows:

1. Tachycardia and Bradycardia
2. Left Bundle Branch Block and Old Inferior Myocardial Infarction
3. Right Bundle Branch Block and Old Anterior Myocardial Infarction

Phase 2 is further split into 3 sub-phases where it will first classify signals in group 1, group 2 then group 3 in a hierarchical fashion. This strategy removes overlapping class features which can confuse classifiers.

In the first phase, GAssist-ADI performed the best with a detection rate of 82.7% and a true negative rate of 91.4%. In the second phase, GAssist-ADI also performed the best, achieving 100% accuracy for the 6 arrhythmias. This paper will focus on a comparison between GAssist-ADI and the proposed method.

3 Methodology

3.1 Data Selection

This paper will be using the MIT-BIH Arrhythmia Database to train and evaluate the classification algorithm. It consists of 48 half-hours two-channel ECG recordings: 23 randomly selected from 24-hour recordings and 25 recordings with uncommon arrhythmias. This combination gives a good sample of normal and abnormal ECG signals with an array of arrhythmia types. The signals were recorded with a frequency of 360 samples per second. Each beat is annotated with N for normal sinus rhythm or one of the given arrhythmia types.

3.2 Implementation Details

Pre-processing including filtering, peak detection, segmentation and data file creation will be implemented in Python. The Waveform Database Software Package (WFDB) [21] [12] library will be used to read the data and annotations. To produce the files fed to BioHEL, a bespoke arff file write has been produced. A custom log reader has also been produced to extract the training curve as seen in **Figure 12**.

3.3 Noise Reduction

This study will investigate two noise reduction techniques, the polynomial function subtraction method [7], and the Butterworth bandpass method [4]. These two techniques will be compared because of their high-performance results, such as 96% accuracy for atrial fibrillation detection with the bandpass method and 100% sensitivity for segmenting signals with the subtraction method. The wavelet transform method was discounted as there were no evaluation statistics.

In the original paper on the subtraction method, the ECG signals from the chosen database were only 20 seconds long in contrast to the MIT-BIH database where they are 30 minutes long. Tests will be conducted to assess the applicability of this method on longer signals, as the computation time might rise considerably with a higher order polynomial function.

For the bandpass method, a Butterworth filter will be used to denoise signals. This will be implemented in Python using SciPy. The bandpass used will be 0.5 Hz - 45 Hz as used in the original paper. To specify the input frequencies for the function, they will be divided by the Nyquist frequency (cycles per second) which needs to be calculated.

$$f_n = \frac{1}{2}\Delta t$$

***Equation 2**: Nyquist frequency*

In this case, Δt is the sampling frequency of the records which is 360 samples per second.

3.4 Segmentation

I will be using the Squared Double Difference [15] approach to segment records into discrete waveforms. This method was selected due to its high sensitivity comparable to or exceeding that of the Pan-Tompkins algorithm. The Fast Fourier Transform was discounted due to a lack of performance metrics.

The algorithm will be applied as described in the literature review where the SDD can be calculated by d(j) for j in the range 0 to the length of the signal.

$$d1(i) = e(i+1) - e(i), i = 1,2 \cdots n-1$$

$$d2(j) = d1(j+1) - d1(j), j = 1,2 \cdots n-2$$

$$d(j) = [d2(j)]2$$

***Equation 3**: Squared double difference*

Peaks are detected by filtering the resulting SDD array to the largest 3% of values. This algorithm produces several points close by on the same peak, therefore they need to be filtered so that only one is selected. If a point is ±75ms within another, it will be removed. The QRS region is selected from the original signal, ±75ms around each selected peak.

The signal is then segmented by taking all values 135 points on either side of the peaks. This gives a window with 270 values or 75ms.

3.5 Classification

This paper will be utilising BioHEL [2], a genetics based machine learning system for bioinformatics to classify the segments. This is the successor to GAssist which performed the best out of the classification algorithms tested in KEEL [20]. BioHEL uses a different learning paradigm, namely the Iterative Rule Learning approach, in comparison to GAssist which uses the Pittsburgh approach. This difference makes BioHEL more suitable for large datasets. It uses novel meta-representation AKLR and CUDA based evaluation which decreases training time. With the Iterative Rule Learning approach, a set of rules are produced over generations. Each rule has a condition and corresponding class if the condition is met. Rules are learnt from the most successful individual in the population. The samples that meet the learnt rule condition are then removed and another rule is learnt from a new population. The purpose of selecting this system was to evaluate whether the Iterative Rule Learning approach it employs is superior to those used in previous research papers.

Classification will be in two phases:

Phase 1 – Arrhythmia detection: All instances with normal sinus rhythm labelled in the database will be added to the Phase 1 data with the label “N”. All other instances (arrhythmias) will be added to Phase 1 data with the label “X”. These two classes will then be split into test and train sets and classified.

To choose the arrhythmia classified, a distribution plot was produced. The 7 largest arrhythmias were selected, all having over 1,000 instances.

Beat annotation	Meaning
L	Left bundle branch block beat
R	Right bundle branch block beat
V	Premature ventricular contraction
/	Paced beat
A	Atrial premature beat
f	Fusion of paced and normal beat
F	Fusion of ventricular and normal beat

***Table 1**: In order, for arrhythmias to be classified and their beat annotation symbol.*

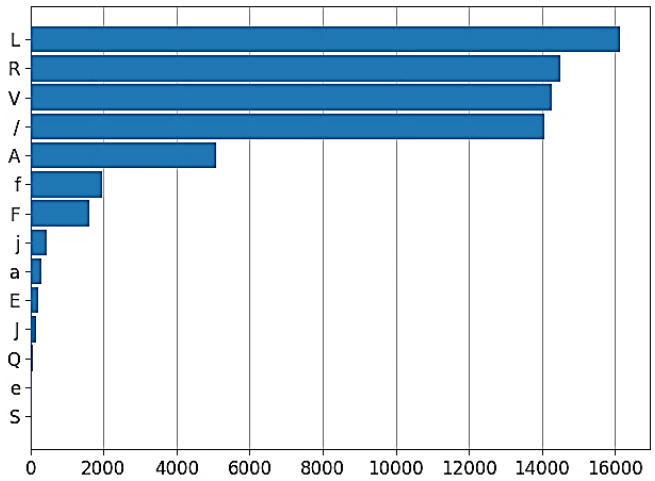

Figure 3: *Arrhythmia instance distribution*

Phase 2 - Arrhythmia classification: Any segments with the “N” beat annotation will be removed as they have been classified in Phase 1. The next largest arrhythmia class will be classified, with all other instances being assigned class “X”. This repeats itself for all major arrhythmia types, removing arrhythmia instances that have already been classified in the previous step.

An 80/20 train test split will be used as is the industry standard. The train and test sets will then be saved as an arff file using a custom arff file writer in Python. In addition to the test and train data, BioHEL also takes a configuration file with the chosen evolutionary learning system parameters. Paper [22] suggests using a coverage breakpoint of 0.001, an ILAS window of 10-50 and the default class being the majority class for large datasets (10,000 - 500,000 instances).

Whilst the proposed method uses a similar hierarchical classifier strategy as the paper on KEEL [20], instances elimination before Phase 1 will not be done. Removing instances before classification means that the classifiers will have much lower accuracy on real data where these samples cannot be removed.

Class imbalances could cause biases in classifiers, therefore an experiment will be done to see if balancing classes produced better results and importantly reduce the false negative rate.

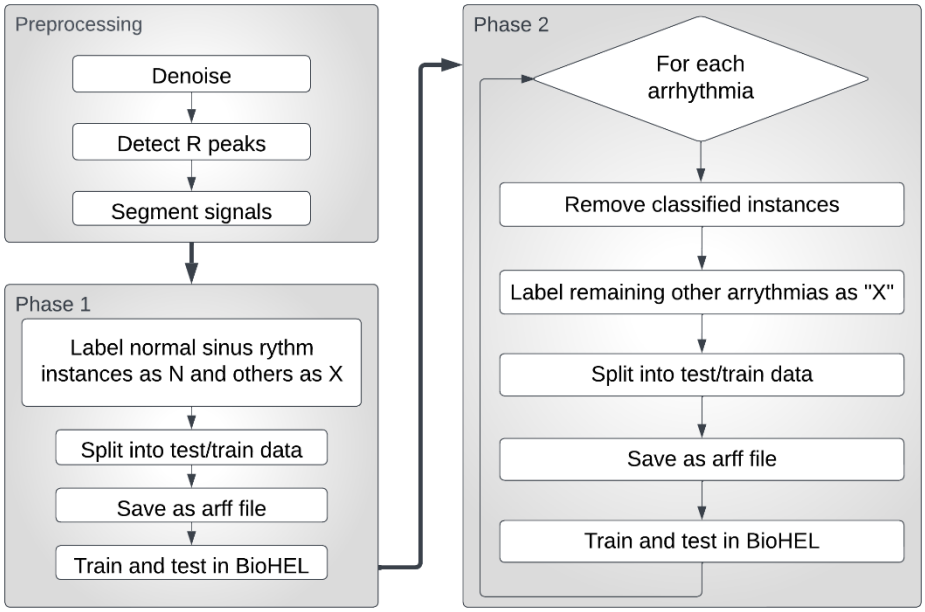

Figure 4: *Methodology workflow*

4 Experimentation

4.1 Noise Reduction

Initial experimentation used the polynomial function subtraction method. The records in the MIT-BIH arrhythmia database are 30 minutes long compared to 20 seconds in the original paper, therefore a higher order is needed. A polynomial function was fitted to the entire signal with an extremely high order of 50. The below figure shows that this was still not enough to be able to determine the baseline wander. Using a higher order is discouraged and can even produce overflow errors. Therefore, this is not a valid technique for this dataset.

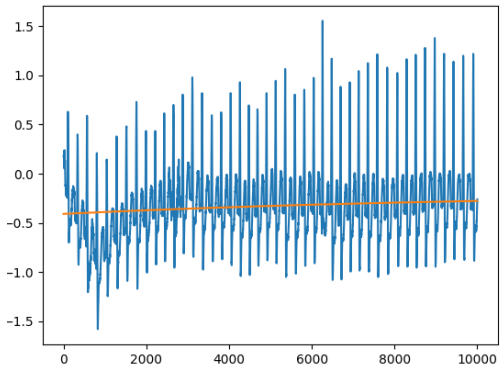

Figure 5: *Polynomial function of a signal shown in orange with an order of 50*

Splitting the signal into 20 second chunks and using an order of 10 produces a good representation of the baseline wander however is very overfitted and could remove information. Furthermore, if the signal is joined up, the boundaries between chunks exhibit sudden jumps as shown in the figure below.

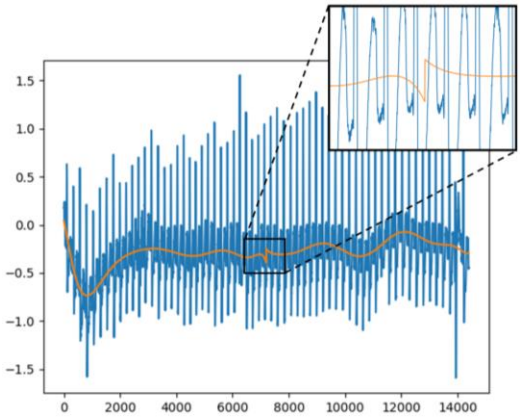

Figure 6: *Polynomial function fitted in orange, exhibiting jump.*

Whilst the records could be split into 20 second subdivisions, when passed through to the next step of segmentation the number of samples detected could be drastically reduced. This is because the complexes at each end of the boundaries could be incomplete, and therefore would be excluded.

With this method ruled out, the Butterworth bandpass was implemented with cut-off frequencies of 0.5Hz and 45Hz.

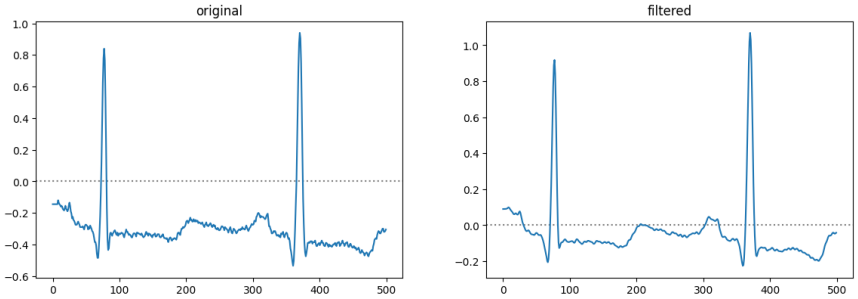

Figure 7: *Left plot showing unprocessed ECG. Right plot showing ECG with bandpass filter applied*

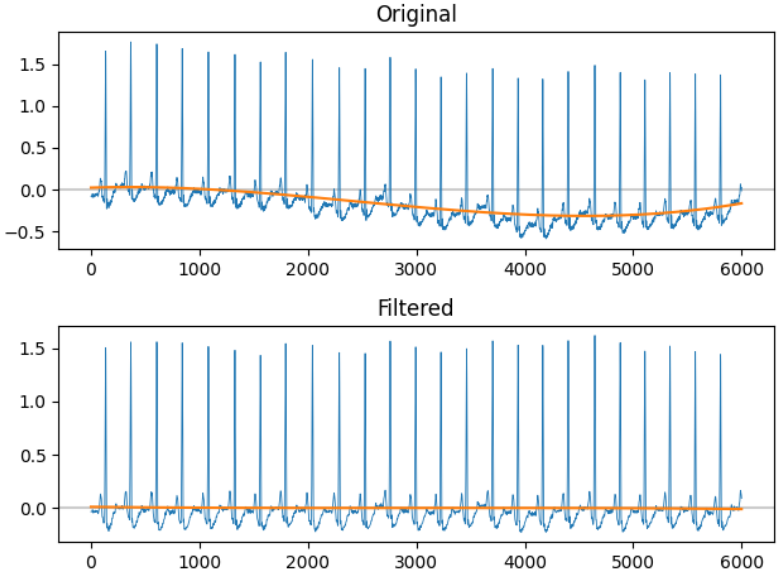

Figure 8: *Top plot shows an unprocessed signal. Bottom plot shows the signal with a bandpass filter applied*

The above figures show good results, with both high and low-frequency noise reduced without distorting the signal.

Due to the many negatives of using the polynomial function subtraction, the Butterworth bandpass is the noise reduction technique selected for this paper.

4.2 Segmentation

To evaluate the effectiveness of using the squared double difference technique, plots were produced for a selection of signals.

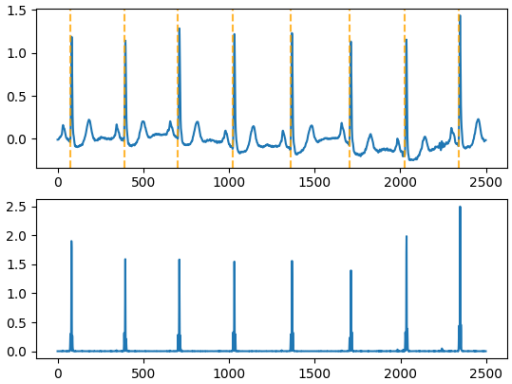

Figure 9: *Bottom plot shows SSD of normal sinus rhythm. Top plot shows selected peaks in orange*

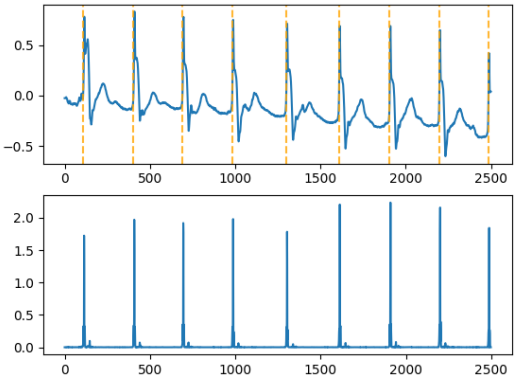

Figure 10: *Bottom plot shows SSD of paced beats. Top plot shows selected peaks in orange*

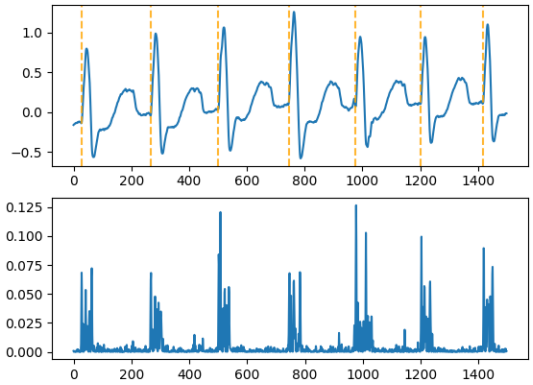

Figure 11: *Bottom plot shows SSD of left bundle branch block beats. Top plot shows selected peaks in orange*

The above figures and further manual tests show that the SDD approach can accurately detect R peaks on normal sinus rhythms and arrhythmia. This will be further assessed in Results and Evaluation. The peaks were then successfully used to create segments 75ms on each side of the R peaks.

4.3 Classification

BioHEL takes a configuration file which sets the parameters for the evolutionary algorithm. The initial configuration file used had the following parameters:

- Initial classifiers 5
- Population size 200
- Windowing ILAS 20
- Coverage breakpoint 0.01

BioHEL was then run with unbalanced train and test data, obtaining an F1-score of 95.6% and a false negative rate of 6.3%.

Variations on this configuration file were tested. The following changes were made accordingly:

- Increase initial classifiers to 20 – This allows for more variation on initialisation to reduce getting stuck on local maximums.
- Decrease coverage breakpoint to 0.001 – This allows the algorithm to converge for a longer time to get better results.
- Increase window ILAS to 50 – Allows the system to scale due to the large training set.

	Starting configuration	Optimised configuration
FN Rate %	6.3	5.3
Accuracy %	97.3	97.5
F1-Score %	95.6	95.9

Table 2: *False negative rate, test accuracy and F1-Score for configuration files*

Whilst the improvements are minimal, there is a reduction in the FN rate which is desired. For equations see Appendix A.

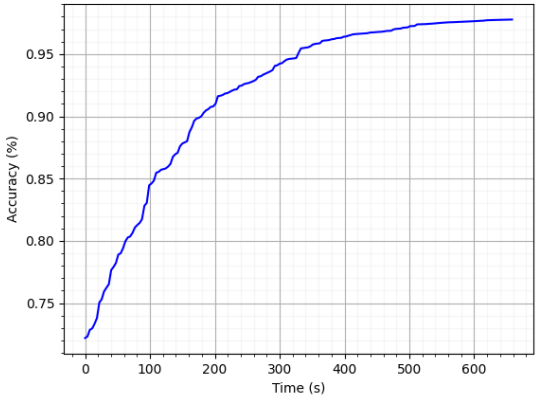

Figure 12: *Training curve for optimised configuration*

The training curve shows BioHEL's ability to converge towards an optimal solution. It is smooth, signalling that there are no trappings in a local optimum.

Then, BioHEL's handling of class imbalances in the context of arrhythmia detection was explored. The normal sinus rhythm class has about 150,000 instances compared to 68,000 in the arrhythmia class. Balancing data can reduce class bias towards the largest class. To do this, the training set was under-sampled, so each class had the same number of samples. The test data is not altered to best represent real data. The table below shows that balancing the data reduces the false negative rate by 1.9%. Although the accuracy and F-1 score had a slight reduction (which could be due to randomness), reducing the

false negative rate for arrhythmia detection is crucial for healthcare. Therefore, this technique will be used.

	Unbalanced	Balanced
FN Rate %	5.3	3.4
Accuracy %	97.5	97.2
F1-Score %	95.9	95.4

Table 3: *False negative rate, test accuracy and F1-Score for balanced classes vs unbalanced classes*

		Predicted	
		N (normal)	X (arrhythmia)
Actual	N	14206	382
	X	221	6338

Table 4: *Final confusion matrix for phase 1 using balancing.*

5 Results and Evaluation

5.1 Noise Reduction and Peak Detection Results

To evaluate the effectiveness of peak detection and noise reduction, 10 seconds of each patient record was tested. This totalled 606 peaks. Among them, 530 peaks were detected without any pre-processing, resulting in a detection rate of 87%. After applying noise reduction techniques, 595 peaks were detected, resulting in a significantly improved detection rate of 98%. The radar plot shows the impact filtering had on the ability to detect R peaks N, R, V, / but most significantly L complexes, increasing the detection rate by 29%. These results imply the proposed noise reduction technique vastly improves the performance of peak detection.

	N	L	R	V	/	A	f	F
No. peaks	420	38	53	38	45	6	4	2
% DR before filtering	88	71	92	87	89	100	100	100
% DR after filtering	100	100	100	95	96	100	100	100

Table 5: *Detection rate of R peaks for different beat types*

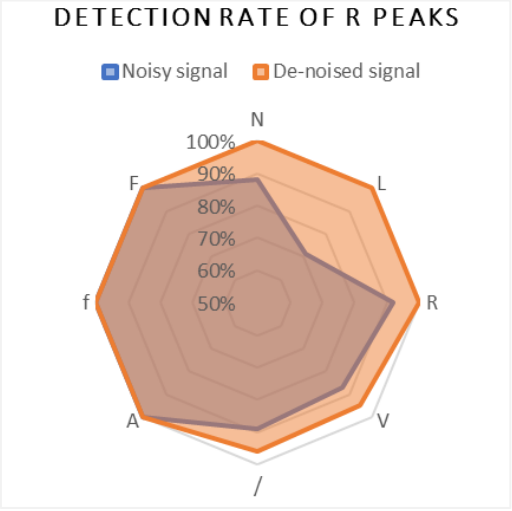

Figure 13: *Radar plot showing the detection rate of R-Peaks for different beat types (Starting at 50%)*

As a result of the high overall detection rate, a large number of high-quality QRS complexes were produced for all arrhythmia types. This will lead to a better F1-Score and reduced FNR as required in the evaluation criteria. Future research could explore increasing the accuracy of detecting V (Premature ventricular contraction) and / (Paced beat).

5.2 Phase 1 Results: Arrhythmia Detection

Classifier	DR %	FNR %	Accuracy %	F1- score %
GAssist-ADI	82.7	17.3	undefined	undefined
BioHEL (proposed)	96.6	3.4	97.1	95.5

Table 6: *Results for Phase 1 detecting arrhythmias*

The Phase 1 results for the proposed method are far superior to GAssist-ADI, reducing the false negative rate by nearly 14%. The reason for this could be attributed to the pre-processing techniques and the fact that BioHEL is specifically engineered to handle large datasets. The input data used had a high number of records (100,000), each with a high number of voltage samples (270). Reducing the false negative rate during arrhythmia detection is critical in healthcare and this shows BioHEL is more suited to this problem.

The classifier was then tested on each patient's record. To do this, a custom bash script was written to automate this as only one test file can be supplied per train cycle. Therefore, the training needed to be repeated for every test with a set seed. The resulting box plot reveals that patients with a higher incidence of arrhythmias achieved a greater median accuracy score. This can be attributed to the model's successful design, which prioritized reducing the false negative rate. Since the model is more prone to wrongly predict arrhythmias rather than normal sinus rhythms, there is a tendency to over-diagnose the latter while minimizing the chances of missing the former.

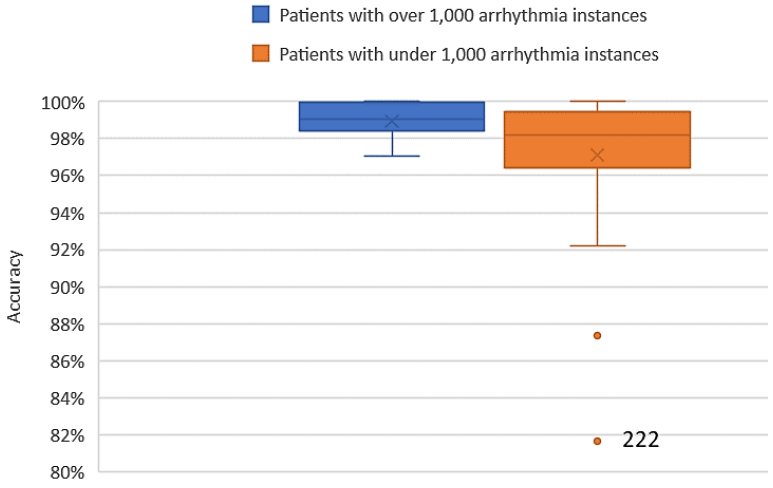

Figure 14: *Box plot showing accuracy of arrythmia detection for patients with a high number of arrhythmias in recording vs those with few*

With patients with under 1,000 arrhythmia instances, there are two outliers. When examining the data, there are some anomalies. For example, with patient 222 (which was classified with the lowest accuracy of 82%) there appears to be extreme levels of noise, perhaps to do with equipment fault or could be incorrectly labelled.

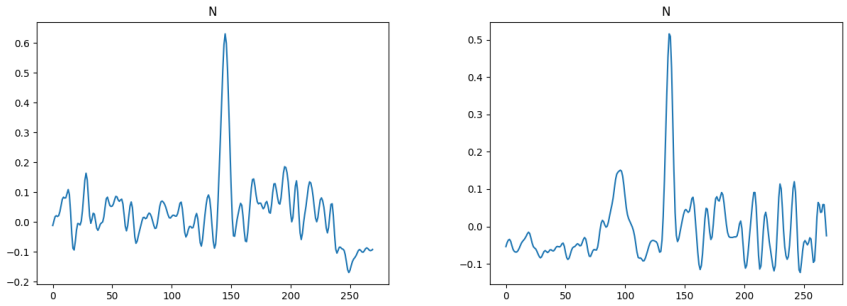

***Figure 15**: Patient 222 samples labelled as normal sinus rhythm instances with anomalies.*

To improve the performance of this classifier, the involvement of medical professionals would be needed to aid in the analysis and validation of the data. Overall, BioHEL was highly successful in detecting arrhythmias and met the evaluation criteria by achieving an F1-Score of less than 90% and a false negative rate of less than 15%.

5.3 Phase 2 Results: Arrhythmia Classification

	L	R	V	/	A	f	F
Accuracy %	98.8	98.8	97.1	98.9	92.5	94.9	94.9
F1-Score %	97.6	98.0	95.6	99.1	92.7	94.6	95.6

***Table 7:** Accuracy results for each arrhythmia classification*

In Phase 2, all seven types of arrhythmias were accurately classified with an F1-Score exceeding 90%. However, the smaller classes demonstrated a reduction in accuracy, which could be attributed to the limited size of their training sets. This results in a lower number of samples for the classifiers to learn features from. Furthermore, the use of the same configuration file for both smaller and larger classes may have also contributed to this disparity. As a next step, it would be beneficial to customize the configuration file for each classifier to improve its performance.

To assess the performance of arrhythmia classifiers on an individual patient level, each patient's record in the dataset was classified using the L classifier after excluding normal sinus rhythms. Overall, the L

classifier showed good performance, although some patients had lower accuracy due to a low number of arrythmias in their record. This means that if one segment was classified incorrectly, there would be a significant impact on the overall accuracy (see labelled datapoint). However, all other patient records had an accuracy above 80%, and patients with more than 500 samples had an accuracy above 95%. Importantly, patients with left bundle branch block had similar classification performance to those without as shown in orange.

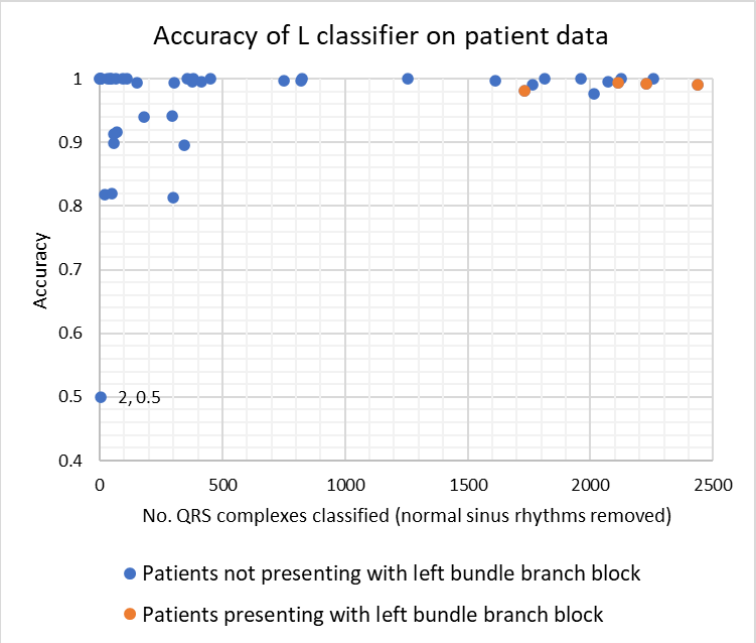

Figure 16: *Accuracy of detecting left bundle branch block on patients against the number of segments classified.*

To find the overall effectiveness of Phase 2, the average F1-score and accuracy below were calculated using a weighted average across the classes. In the second phase of the study, the classification of 7 arrhythmia types achieved an impressive F1 score of 96.2%, as illustrated in the table below. Whilst it did not obtain the 100% reached using GAssist-ADI, no instances were removed pre-classification, meaning these results are not easily comparable. These are very good statistics for the number of arrhythmias classified. BioHEL's ability to deal with large datasets is less significant in Phase 2 because the number of instances is smaller for each classifier.

Classifier	Average F1 score	Average Accuracy
GAssist-ADI	Undefined	100%
BioHEL (proposed)	96.2	97.5

Table 8: *Results for Phase 2 detecting arrhythmias*

5.4 Combined Results

To calculate the overall F1-score and accuracy of the proposed method, the weighted averages are taken over the two phases.

	F1- score %	Accuracy%	No. of classes
Improved AlexNET [17]	98	97	4
ENCASE [16]	84	*	4
KELM [14]	*	89.7	13
GAssist-ADI [20]	**	**	7
BioHEL (proposed)	96.2	97.4	8

* Undefined
** No statistics were provided to evaluate the combination of Phase 1 and 2

Table 9: *Results of the proposed method compared with other methods*

When looking at the combined results of Phases 1 and 2, the proposed method performs well. Whilst it did not attain as high F1-Score as the improved AlexNET model, it can classify 4 additional arrhythmias. It outperformed every other model, notably having a 12% greater accuracy than ENCASE whilst being able to identify 4 more arrhythmias.

6 Conclusion

The results obtained in this paper show that there is value in using an evolutionary classifier for ECG arrhythmia detection and classification. The considerable reduction in the false negative rate in contrast to GAssist indicates that BioHEL is the superior model for arrhythmia detection. However, it is comparable on smaller datasets for arrhythmia classification.

The chosen pre-processing methods such as a Butterworth bandpass filter and peak detection using the squared double difference technique performed well, being able to detect 98% of peaks. This meant that a large number of accurate QRS segments could be used as training and test data, in turn increasing accuracy.

This paper aimed to achieve three specific objectives in detecting and classifying arrhythmias: attaining an F1 score of at least 90% for detecting arrhythmias, optimizing the approach to reduce the false negative rate below 15%, and achieving an F1 score of at least 90% for classifying arrhythmias. All these criteria were met and exceeded (95.5%, 3.4% and 96.2% respectively). Most impressively, combined with pre-processing methods, BioHEL’s ability to detect arrhythmias with a false negative rate of 3.4% far outstrips GAssist-ADI which had a false negative rate of 17.3%.

Potential avenues for future research include exploring the optimization of configuration files for different class sizes. The expertise of medical professionals could be used to detect and remove corrupt data (as may be the case in patient 222). Furthermore, the potential for improving the generality of the classifiers could be explored by utilizing BioHEL's Rule Post-Processing Engine, which was not investigated in this study.

The obtained results were found to be at least as good as, if not better than, those of classical neural network models that are commonly used in this area of study. This indicates that employing an evolutionary algorithm, particularly one that utilizes an Iterative Rule Learning approach, is a promising technique for arrhythmia detection and classification, which can be further explored in future research.

Appendix A

$$Accuracy = \frac{TP + TN}{N + P}$$

Equation 3: *Accuracy*

$$Detection\ rate = \frac{TP}{P}$$

Equation 4: *Detection Rate*

$$False\ Negative\ Rate = \frac{FN}{FN + TP}$$

Equation 5: *False negative rate*

$$F1 - Score = \frac{TP}{TP + 0.5(FP + FN)}$$

Equation 2: *F1-Score*

Bibliography

[1] “NHS Arrythmia,” 14 Sept 2021. [Online]. Available: https://www.nhs.uk/conditions/arrhythmia/. [Accessed April 2023].

[2] J. Bacardit et al , “Improving the scalability of rule-based evolutionary learning,” *Memetic Computing,* vol. 1, no. 1, pp. 55-67, 2009.

[3] M. Farag, “A Tiny Matched Filter-Based CNN for Inter-Patient ECG Classification and Arrhythmia Detection at the Edge,” *Sensors (Basel, Switzerland),* vol. 23, no. 3, p. 1365, 2023.

[4] S. Butterworth, “On the Theory of Filter Amplifiers,” *Experimental Wireless and the Wireless Engineer.,* vol. 7, pp. 536-541, 1930.

[5] WarrenWeckesser and KotMorderca, “SciPy Cookbook,” April 2020. [Online]. Available: https://scipy-cookbook.readthedocs.io/items/ButterworthBandpass.html. [Accessed Jan 2023].

[6] Xiangyu Zhang et al, “Over-fitting suppression training strategies for deep learning-based atrial fibrillation detection,” *Medical & Biological Engineering & Computing,* vol. 59, pp. 165-173, 2021.

[7] J. Aspuru et al, “Segmentation of the ECG Signal by Means of a Linear Regression Algorithm,” *Sensors (Basel),* vol. 19, no. 4, p. 557, 2019.

[8] D. Zhang, “Wavelet Approach for ECG Baseline Wander Correction and Noise Reduction,” in *IEEE Engineering in Medicine and Biology*, Shanghai, 2005.

[9] D. L. Donoho et al, “Wavelet Shrinkage: Asymptopia?,” *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 57, no. 2, pp. 301-337, 1995.

[10] J. Cheng, “ECG signal classification based on deep CNN and BiLSTM,” *BMC Medical Informatics and Decision Making,* vol. 21, 2021.

[11] J. Pan and W. Tompkins, “A Real-Time QRS Detection Algorithm,” *IEEE Transaction on Biomedical Enginerring,* Vols. BME-32, no. 3, pp. 230-236, 1985.

[12] A. Goldberger et al, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,” *Circulation Electronic,* vol. 101, no. 23, pp. 215-220, 2000.

[13] GB Moody and RG Mark, “The impact of the MIT-BIH Arrhythmia Database”, *IEEE Eng in Med and Biol,* vol. 20, no. 3, pp. 45-50, 2001.

[14] S. Dalal and V. Vishwakarma, “Classification of ECG signals using multi-cumulants based evolutionary hybrid classifier,” *Sci Rep,* vol. 11, no. 1, p. 15092, 2021.

[15] D. Sadhukhan and M. Mitra, “R-Peak Detection Algorithm for Ecg using Double Difference And RR Interval Processing,” *Procedia Technology,* vol. 4, no. 1, pp. 873-877, 2012.

[16] S. Hong et al, “ENCASE: An ENsemble ClASsifiEr for ECG classification using expert features and deep neural networks,” in *Computers in Cardiology (CinC)*, 2017.

[17] A. Kumar and C. Arvind, “Classification of ECG signal using FFT based improved Alexnet classifier,” *PloS one,* vol. 17, no. 9, 2022.

[18] R. Bousseljot, D. Kreiseler et al, “Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet,” *Biomedical Enginerring,* vol. 40, no. 1, p. 317, 1995.

[19] D. Dua and C. Graff, “UCI Machine Learning Repository,” University of California, Irvine, School of Information and Computer Sciences, 2017. [Online]. Available: http://archive.ics.uci.edu/ml. [Accessed April 2023].

[20] K. Waseem, A. Javed et al , “Using evolutionary algorithms for ECG Arrhythmia detection and classification,” in *Seventh International Conference on Natural Computation*, Shanghai, 2011.

[21] C. Xie, L. McCullum et al, “Waveform Database Software Package (WFDB) for Python,” PhysioNet, 2023. [Online]. Available: https://doi.org/10.13026/9njx-6322..

[22] M. Franco, N. Krasnogor et al, “GAssist vs. BioHEL: critical assessment of two paradigms of genetics-based machine learning.,” *Soft Computing,* vol. 17, no. 1, pp. 953-981, 2013.

[23] NT. Bui and G. Byun, “The Comparison Features of ECG Signal with Different Sampling Frequencies and Filter Methods for Real-Time Measurement,” *Symmetry,* vol. 13, no. 8, p. 1461, 2021.

[33] F. Buendía-Fuentes and M. A. Arnau-Vives, “High-Bandpass Filters in Electrocardiography: Source of Error in the Interpretation of the ST Segment,” *International Scholarly Research Notices,* vol. 2012, 2012.