Primitive Time-Domain Classification of J-Junction Defects Indicative of Congenital Diseases Associated with SCD

Jaymie L. Phillips, Stephen John Rees, Tim Bashford

School of Applied Computing, UWTSD, Swansea, United Kingdom.

Jaymie.phillips@uwtsd.ac.uk; John.Rees@uwtsd.ac.uk; Tim.Bashford@uwtsd.ac.uk

Abstract **- With the advancements in modern medicine, the life expectancy within the United Kingdom has increased. This greater life expectancy has resulted in the requisite management of chronic conditions and as such has increased the burden upon the NHS. This paper proposes a means for early screening of underlying health conditions that could result in early intervention and prevention of underlying cardiac conditions, thus reducing the later cost to the health services. The proposed solution employs a data reduction approach to classification in addition to using fundamental biological concepts to ascertain anomalous operation. The approach has proved relatively successful in this endeavor, though there are still improvements that could be achieved with accuracy.**

Keywords - Signal Processing, ECG, Physionet, Healthcare, Sudden Cardiac Death

I. INTRODUCTION

The advent of modern western healthcare has seen a dramatic increase in life expectancy over the last century. This is particularly highlighted in the last three decades with the complement of blending disciplines to solve new problems. This increase in life expectancy, has consequently seen a rise in the mean age of the population. The greater volume of elderly patients using the National Health Service (NHS) and the ailments associated with this coupled with the needs of other patients has imposed a significant demand upon the NHS. This requirement has led to the excessive strain upon the resources available and has such has required the Inevitable restructuring and allocation of the system [1] The constraints upon available care is seen most predominantly in rural areas, where prospective patients live vast distances from the nearest medical center or hospital. This imposes yet further constraints upon both the healthcare provision and the upon the patients themselves[2]. The distances prevent in many cases any form of rapid response in addition to the incurring of costs to both the healthcare service and the patients in travel expenses and in the limited possibility of early diagnoses.

Such intrinsic problems prompt the requirement of a means for providing immediate care in emergency cases and the possibility of screening to provide information on early warning signs of serious medical complications. It is in satisfying this requirement that this paper proposed a less resource and thus cost intensive means of screening for such cardiac difficulties [3].

II. BACKGROUND

The operation of the human body is driven in no small part by the functionality provided by the one of its internal organs; the heart. The structure of this organ is relatively complex being composed of a number of different specialized tissues, each with a unique purpose for which it is ideally suited. By far, the constituent tissue with the greatest volume is the myocardial tissue. The myocardial tissue is a type of muscle that exhibits a very similar structure to that of smooth muscle tissue in that it is ideally suited to continuous contraction [4]. The heart however, is distinguished by its pattern of contraction which presents as rhythmic. This rhythmic contraction facilitates the propulsion of blood around the body by the timed contraction of different areas of the heart.

The different areas of the heart permit the ability to provide the time-sensitive responses required for correct circulation of the blood. The heart is divided into four different chambers with two residing at the head of the organ progressing on to two further chambers protruding from the mid horizontal point, toward the apex. The two chambers located at the top of the heart are referred to as the atria; these chambers are separated by a wall of cardiac muscle tissue with separate blood vessels attached to facilitate the function of each. The right atrium is fused to a major inlet vein known as the superior vena cava, connecting this chamber to the pulmonary system. The left atria is headed by the aortic artery; the widest blood vessel in the body. The ventricles differ from the atria by many different factors most obviously that of their shape, resembling a far more elongated reservoir compared to that of the atria. The ventricles are separated from the aorta by the presence of two specific valves known as the mitral valve and the tricuspid valve. Such valves allow for a periodic passage between the atria and the ventricles. The two ventricles are separated by a thick cardiac muscle wall termed the septum. Within this septum there are a number of different tissues and fibers instrumental in the operation of the heart; the purkinje tissue and the bundle-of-his being the most prominent. The left ventricle exhibits a far thicker outer wall than that of the right in aid of its servicing of the body with the necessary oxygenated blood [5].

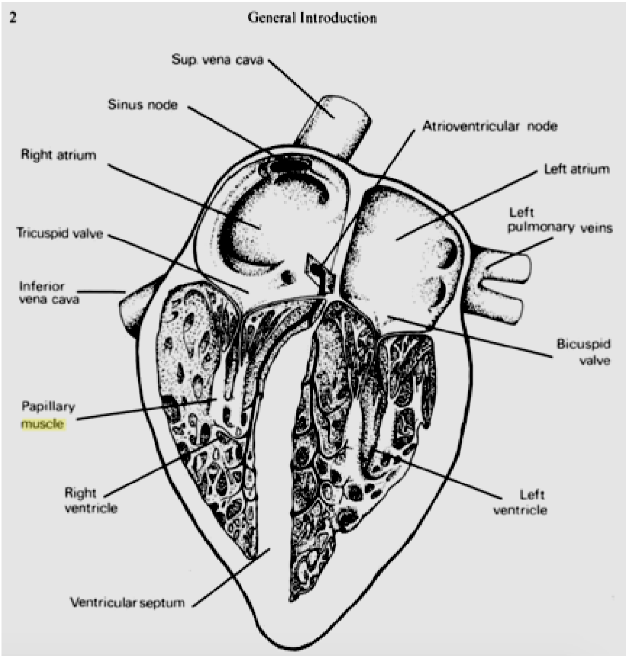

Figure 1. The heart [4]

The structure present in the heart affords the tightly constrained function and timings of the same necessary for life. Each heartbeat is initiated by excitation of the cardiac tissue at the top of the heart surrounding the right atria [6]. This excitation is caused because of the innervation of a small area of nervous tissue known as the Sin-Atrial Node (SAN). From this node, an impulse causes contraction of the atria, forcing the blood through the mitral and tricuspid valves into the ventricles. The excitation created by the SAN travels across the muscle tissue surrounding the atria towards an area at the mid-horizontal point of the heart known as the atrio-ventricular node (AVN). The AVN can be considered to act as a repeater of the signal generated sending the impulse down the septum of the heart through the purkinje tissue and the bundle-of-his fibers toward the apex which is then transferred to the ventricular walls. Upon the signal reaching the ventricular walls, the ventricles contract forcing the blood contained within them out to the body and the lungs. It is this excitation and impulse transfer that can be measured using an electrocardiogram (ECG) [7]. In some cases the typical healthy operation of the heart can be affected by other, more episodes such as ventricular fibrillation (VF) or atrial fibrillation (AF) [8] [9]. Both of these episodes fall into the arrhythmia category but in most cases span shorter periods than other arrhythmias. VF exhibits a variance in the expected pattern of contraction resulting in the both irregular excitation and uncoordinated excitation of the ventricles resulting in the inability to propel blood through the pulmonary artery and the aorta, thus starving the body of oxygen.

Though there are many different conditions that present with variations in the rate of operation, there are those that affect the amplitude of each component of the heartbeat. Sudden Cardiac Death (SCD) has been placed firmly in the spotlight following recent publicity surrounding unexplained cardiac events of notable sports people during play. Upon post-mortem investigation many of these individuals have been found to have suffered from conditions within this category. There have been a number of different cases where SCD has been linked to the individual suffering from underlying chronic conditions such as Brugada Syndrome, Wolf-Parkinson White syndrome and Romano-Ward syndrome [10]. Though Romano-Ward syndrome is typified by the presentation of long-QT both Wolf-Parkinson White syndrome and Brugada syndrome present with displacement indicators; with shorter PR period and presence of "delta wave" and ST-segment elevation respectively [11].

A. Signal Characteristics

The operation of this tissue is predominantly measured utilizing the ECG. The use of this equipment affords depiction of the process and the elucidation of the characteristics of the signals transmitted through the cardiac tissue. Figure 2 provides a graphical representation of a typical healthy adult, highlighting the key components of the cardiac cycle in addition to the notations most commonly attributed to these components.

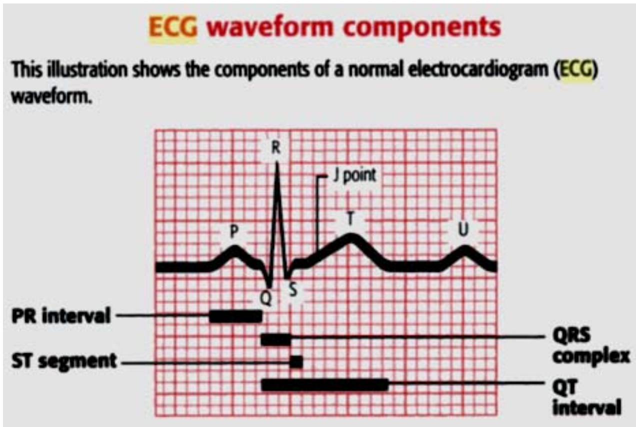

Figure 2. Normal ECG waveform [12]

The section of the waveform decorated with the letter 'P' represents by the depolarization of the atria resulting in the contraction of these segments of the heart. This wave artefact is typically very small and in some cases not detected within an ECG reading with a length of approximately 120ms. This is followed by the 'QRS' complex which is the most intrinsically linked and rapid

DOI 10.5013/IJSSST.a.21.02.25

ISSN: 1473-804x online, 1473-8031 print

component collection of the entire cycle. This section occurs as a result of the AVN receiving the impulse transmitted to it by the SAN and the consequent depolarization of the ventricles to facilitate the second half of the cycle. The joining segment between the 'S' and the 'T' components of the representation is known as the 'ST-segment' or in some cases the 'J-Junction' and is instrumental in the diagnoses of cardiac conditions that are not conducive to presenting arrhythmia such as Wolff-Parkinson White and Brugada Syndrome [13]. The 'T' constituent embodies the tail of the waveform a period in which the ventricles are particularly vulnerable because of their electrical instability.

Each of the segments of an ECG recording has a tolerated displacement threshold within which it resides. In cases where a patient is suffering from conditions that affect the normal operation of the heart, these thresholds are exceeded eluding to the presence of issues. The 'QRS-complex' shows a stark contrast to this by the fact that it is not uncommon to observe ranges between 0.5mV for the 'Q' point, an upper limit of 1mV for the 'S' component and 3.0mV for the 'R' amplitude. There is scope for variation between normal and precordial leads here. The 'ST-segment' shows a gradual decline from the 'S' component of ~1.0mV to a lower range of ~0.5mV in the tailing 'T' segment. Some elevations are often seen in precordial leads for this segment, though are confined to an accepted maxima.

The importance of the amplitude of the signal components is matched by the timings of each. The nature of the organ and the requisite synchronization means that the timings of each phase of the cycle of the heart are fundamental to its correct operation. Variance in the timings result in arrhythmias that render the tissue unstable and in extreme cases completely ineffective. The 'P' component including the lead to the 'R' point of the waveform occurs with a window of 120-200ms (milliseconds). The 'QRS-complex' itself is the most rapid construct of this process and exhibits typical periods of between 80 - 120ms. The 'T' component occupies the remaining 310ms of the heartbeat at a rate 0f 75 beats per minute.

III. LITERATURE OF CURRENT TECHNIQUES

Identification of anomalies within cardiac rhythms and displacement through their representation on an ECG is not a new phenomenon and has remained common practice since Einthovens first recording of this data. Though the technology has been available to monitor variations in the operation of the organ, the means by which it is achieved has evolved significantly. Traditionally, identification of deviations would be achieved through the careful and time-consuming consideration by a number of clinical cardiac physicians. Fortunately. With the advent of modern technology the tedious nature of checking every square on a cardiac trace in addition to the possible introduction of human error, digital classification methods have been realized. The intent of such digital classification applications was to reduce the time taken to produce diagnoses for an individual in addition to improving the objective accuracy of said diagnoses. The practice of digital signal classification, particularly in consideration of bio-signals has seen many efforts expended toward the achievement of effective classification. Typically the majority of classification efforts reside in one of two domains; that of time or that of frequency [14]. The approaches toward classification can also be divided into approaches utilizing a feature extraction or feature selection approach.

A. Time Domain

For time-domain analysis and feature extraction from cardiac signals, several pieces of information are calculated from the present waveform including the root mean square of the waveform, the spectral density and the power. These values are then used for the grouping and classification of constituents of the signal in addition to the eventual selection of those elements of interest. This approach permits the least computationally expensive endeavor though can result in difficulties with filtering out noise. The simplest calculation that may be performed upon the waveform of interest is that of the standard deviation. The purpose of this component is in ascertaining the deviation from the midpoint of the waveform and thus the displacement amplitude .

There are a number of approaches that utilize the information present within the time domain in addition to the categorization of distinct differences through the attribution if syntactic values, in a very similar approach to that taken for the final classification of the proposed method [15].

B. Frequency Domain

Though a wealth of information can be gleaned from the use of time-domain feature classification, there are those components that can only be identified within the frequency domain. The most common use of transformation of a waveform into the frequency domain is in the implementation of filters used primarily for the reduction of noise. The most prevalent from of transformation is manifest in the utilization of the works of Fourier. There are many adaptations of this transform used to tackle many problems. Each of these approaches have resulted in varying degrees of progression and success resulting in more efficient and more accurate classification techniques. By far the largest use of such transforms is identifying the disparity between desired and undesired signals.

DOI 10.5013/IJSSST.a.21.02.25

ISSN: 1473-804x online, 1473-8031 print

IV. PROPOSED TECHNIQUES

It is the intent of this work to present a non-complex efficient classification method for identifying 'ST-Segment' anomalies using a semi-heuristic thresholding model. To perform the initial isolation of cardiac cycle windows, an adaptation of the Pan-Tompkins algorithm has been used. The classification of ST-Segment anomalies is achieved by considering the biological limits inherent in the tissue and its operation. Using the underlying biological constants, it is possible to construct an ideal model [16]. By observing the biological limitations of the tissues it is possible to identify activity minima and maxima for the rate of contraction in addition to the maximum and minimum voltage amplitudes exhibited by the tissues, It is from such values that the thresholds for the model will be identified. This largely follows the idea behind amplitude envelopes within auditory processing.

To perform the classification of a heartbeat, there are several pieces of information that are typically required. In previous approaches this has manifest in the averaging of the input signal, the RMS of the signal or even the spectral distribution of a signal for time domain analysis. This proposed model considers the use of the amplitude of the incident signals and the timing of those signals only to calculate the possible thresholds for each of the components of the cycle. The calculation for a threshold is achieved by isolating a window of the signal, with a length based upon the minimum possible heart rate, and from this dividing the window into chunks assessing the gradient of the terminal samples of said chunks. The chunk with the highest gradient within this chunk is considered to be the R-peak of the cycle and from this the thresholds for the other components of the cycle, P,ST and J, are derived using a combination of the biological limitations and that of a heuristic analysis of previous windows following the initial assessment.

For some cardiac diseases, not associated with arrhythmia, there is typically a consistent period of normal biological function followed by periods of ST elevations as presented in pericarditis, Brugada and Romano-Ward syndromes. The proposed model allows for calculation of consistent S gradient values based upon previously identified samples coupled with the same approach for the T component. With the retention of such data, the approach is then able to identify a sustained normality within the J transition between the two components. With the reference value established, it is possible for the model to attribute contingencies for normal deviation if observing precordial leads whilst holding a maximum threshold for gradient deviations above which this component is considered to be anomalous and indicative of an underlying issue. The proposed model does not currently hold a contingency for those diseases where no period of normal operation exists.

Most current classification algorithms make use of complex mathematics, most notably that of transforms, to identify trends and patterns within a signal. Though such realizations pose no challenges in a post-processing environment with the time available and the seemingly endless resources, this is in many cases inefficient and difficult to achieve within embedded systems. The nature of embedded systems is such that there are constraints placed upon software implementations as a result of the limited memory, processing and power available. Though there has been some progress towards the increase in resources for some microcontroller offerings in recent years, this has also seen an increase in price for such developments.

It is envisaged that identifying an effective means of classification that can run in real-time on a less complex controller affords the ability to reduce cost in addition to reducing the complexity of performing such screening. In an attempt to further reduce cost it is believed that the use of technology to support the role of the clinician will reduce the necessity for patient or clinician to travel to a medical facility to perform cardiac diagnostic testing and further increase the personal interaction of said clinician with their patient. The increased personal interaction with the patient could result in the identification of further physical or mental health concerns that would not otherwise be highlighted as a result of increased demands upon the time of the clinician.

There is considerable interest in generalized learning algorithms in AI and signal processing, where the arbitrary input signal structure is decomposed or assessed by largely black box techniques. The advent of such generalized algorithms will, no doubt, extend the value of signal analysis to new areas. However, there is a clear issue which is unlikely to be resolved until the algorithmic loop can be closed between the input data (acquisition and preprocessing) and the analysis, such that the analysis outcomes become the driver for modification of the input processing to achieve best aggregate results. At present, this is not available. Accordingly, the use of guided learning, or restructured input based on directed methods, is appropriate given the imperative of a working solution to a problem that is, in fact, potentially fatal.

The approach adopted for this project is, therefore, to optimize the quality of the input signal, to scale both time base and peak signal levels to improve the precision of signal fitting and the probability of correct identification of the ECG trace characteristic features. Fitting the acquired features against a predefined set of models for each potential condition allows a practical, data limited approach to categorization. The essential underpinnings of the modelling of ECG trace characteristics has already been explained in Section III Signal Characteristics. It is worth noting that long-run measured outputs and a wide population of users (test subjects in the research phase) will offer a high utility training set for such future generalized approaches as may emerge for biological signals.

A. Materials and Resources

In the assessment of the functionality and thus validity of the proposed algorithmic approach, cardiac data of a specific nature was required. The data upon which the testing was founded was that provided by the Physionet data collection. Within this collection, there are a number of public access databases, from which the European ST-T Database and the Sudden Cardiac Death database were selected. The first of these sources contains 90 Two-hour recordings from a holter device and the second 23 twelve-hour recordings.

In the validation of the model proposed, it was decided that the initial manifestation would be constructed for conventional personal computer encompassing moderate resources with regard to the memory and processing power provided; an Intel core I5 running at a core speed of 1.4Ghz and 8GB of Random Access Memory respectively. Further to this the software used to construct the implementation was a desktop version of the g++ compiler including the appropriation of clang. The resources used for the development were selected in accordance to their availability, flexibility and possible future targeting of a dedicated embedded platform; an ARM Cortex-M3 device.

B. Experimental Methodology

The proposed solution was implemented within the C language, adhering strictly to this dialect in ensuring compliance with the language standards and those stipulated by regulatory organizations. In reflection of the purpose for which the solution is intended, further considerations were made with regard to medical device regulations particularly those relating to real time software implementations and software specific to medical devices in respect of IEC 62304 and the quality standard ISO9001. Some of the features of the C language were also avoided in employing the rules for reliable and safe software as stipulated by the MISRA 2012 guidelines. The language extensions afforded by C++ and the associated libraries were deemed unnecessary for the implementation in ensuring reliability and reduction of possible overheads introduced by complex abstraction.

The implementation was centered around a test-harness to facilitate the execution of the algorithmic process. The operation of this solution firstly required incident data as provided by the data sources mentioned. This was achieved by the implementation of data handling functions for the sequential reading of both files and individual blocks of data. In facilitating the output of statistics, data creation functions were also implemented.

Upon the successful read of a block of data from its source, it was fed through the Pan-Tompkins algorithm adaptation to isolate windows representing 'R-R peak' cycle intervals. This was achieved to reduce the data upon which the classification was performed. The following phase to this decimation exercise is that of another data filtration whereby the cycle data is separated to isolate the 'ST-segments' and the remaining components of the cycle windows are discarded with the exception of the end of the 'QRS-complex' retained for relativity. The isolation of these smaller windows has many benefits not least the reduction in complexity and calculations required in classification.

The fourth phase consists of the production of a match-vector component in utilizing the biological thresholds available to identify fluctuations in the ST-segment, in particular large elevations with respect to the retained termination to the 'QRS-complex'. The approach used permits the adjustment of the tolerances in the match-vector thresholds should this be required to adapt to an individual. Should a waveform component reside within the tolerances, this sample is considered to be normal and thus disregard, in the instance of the inverse, this is noted as anomalous. Considerations are made for the differences in normal and precordial leads.

V. RESULTS AND DISCUSSION

In testing the implementation, the first five records for each of the data sets noted were presented to the system. The data handling methods implemented for the solution were responsible for the creation of an output data set detailing the statistics calculated as a result of the operations performed. Contained within this data set is the number of reductions performed, the number of classifications made, and the number of threshold adaptations.

Table I provides an overview of the statistical information obtained from the classification for 5 samples from the European ST Database.

TABLE 1: EUROPEAN ST DATABASE RECORDS

File	P	R-R	ST	Anomalous	Adaptions
E0103	5410	6015	5923	235	411
E0104	5300	5935	5312	176	236
E0105	5345	6001	5255	129	239
E0106	6300	6504	5937	363	287
E0107	6800	6302	5975	249	415

Contained within this table are the isolations made for each of the components of the cardiac cycle using the implementation. File field refers to the name of the file as it appears within the physionet data collection permitting reference to the data used for the testing.

The P field contained within the table is a direct reference to that of the P components identified by the classifier. The R-R field of the table is made in reference of the identification of a full cardiac cycle period with terminal points of the R peak of one cycle to that of another. This approach is detailed in the paper by Pan Tompkins and used to isolate each of the cycles in this approach. The ST field noted is a direct elucidation of the ST section of the cardiac cycle. The adaptations listed present the adaptations made to the identified thresholds throughout the running of the classifier.

Table II provides a representation of the statistical information obtained from the input of 5 samples from the sudden cardiac death database record set into the implemented solution.

TABLE II. SUDDEN CARDIAC DEATH DATABASE RECORDS.

File	P	R-R	ST	Anomalous	Adaptions
31	13801	13922	13403	749	1416
32	13912	14104	13111	823	1591
33	13548	14008	14128	622	2008
34	13713	14306	14007	727	1983
35	13811	13698	13951	916	1696

The data contained within this table reflects a similar picture as provided by the previous one for the SCD data set. The fields in this table referred to the filename that was being tested, the P component of the cardiac waveform, the R period of a particular cycle followed by the ST segment and that of the adaptations. The adaptations here denote the adjustments undertaken by the model in accommodating the changes in the recorded signals. IIt is the purpose of this table and that of the one depicted previously to outline the results obtained from initial classification of the components of the waveform using the proposed model.

VI. CONCLUSIONS AND FUTURE WORK

The results obtained from the proposed solution are indicative of the effectiveness of the method. The consideration of the biological characteristics and the variance within, permits an accurate means of identifying those sections of an ECG. The classification of the components of a cardiac cycle was achieved with a high degree of accuracy, particularly that of the R-peak and the ST-Segment of interest.

The identification of anomalous deviations within the ST-segment have proved in some cases to be a little difficult to distinguish. There are several possible reasons for this being the case including the limited pre-filtering of the test data in addition to the possibility of misclassification resulting from similarities within the components of the data.

In addition to the classification, the magnitude of threshold adaptations presented within the statistics proves to be a concern with regard to the overhead introduced in updating this during the classification process. In some cases, the time taken for such operation could pose significant challenges with respect to implementation for an embedded platform. There are a number of improvements that could be made with respect to the proposed model not least the inclusion of a far more complex feedback loop that accounts for stress variations. The nature of the ST-segment displacement issues is such that it could be shadowed by any possible movement artefacts when utilised during physical monitoring scenarios. Further works planned for the implementation include the optimisation of the implementation regarding its resources usage to operate under far more constrained parameters.

REFERENCES

[1] A. a. H. P. a. M. L. a. C. C. J. Bagust, "The projected health care burden of type 2 diabetes in the uk from 2000 to 2060," Diabetic Medicine, vol. 19, no. 1, pp. 1-5, 2002.

[2] J. C. Tomlins, "Is it possible for the NHS to become fully digital ?," in Consumer informatics and digital health, Springer, 2019, pp. 359-374.

[3] G. L. M. J. P. P. N. R. J. F. R. Higgs, "Using Geographic Information Systems to investigate variations in accessibility to 'extended hours' primary healthcare provision," Health \& social care in the community, vol. 27, no. 4, pp. 1074-1084, 2019.

[4] E. a. C. G. a. S. J. Canale, Cardiac Muscle, Springer Berlin Heidelberg, 2012.

[5] A. Katz, Physiology of th eHeart, Wolters Kluwer Health/Lippincott Williams \& Wilkins Health, 2010.

[6] R. Klabunde, Cardiovascular Physiology Concepts, Wolters Kluwer Health/Lippincott Williams \& Wilkins Health, 2011.

[7] J. Levick, An introduction to cardiovascular physiology, Elsevier, Science, 2013.

[8] M. Rajskina, Ventricular fibrillation and sudden coronary death, Springer Us, 2012.

[9] Y. C. A. Yap, Essentials of Atrial Fibrillation, Springer healthcare limited, 2014.

[10] S. C. P. R. F. A. B. A. Nutting, "Antipsychotics for patients with Wolff-parkinson white syndrome and parkinsons," Clinical schizophrenia and related psychoses, no. 1, 2019.

[11] K. S. M. Haugaa, "Cardiac electrical and mechanical alterations-united in the long QT syndrome.," International Journal of Cardiology, vol. 274, no. 1, p. 190, 2019.

[12] L. Wilkins, ECG Interpretation, Lippincott Williams, 2005.

[13] P. a. T. R. a. B. J. a. G. O. a. B. D. a. M. J. a. J. L. a. D. J.-M. a. B. P. a. K. F. a. o. Berthome, "Clinical presentation and follow-up of women affected by Brugada syndrome," Heart Rhythm, vol. 16, no. 1, pp. 260-267, 2019.

[14] J. H. A. H. M. Mantas, Enabling Health Informatics Applications, IOS Press, 2015.

[15] S. a. A. A. B. a. B. M. H. Hamdi, "Syntactic Methods for ECG Signal Diagnosis and QRS Complexes Recognition," in Biomedical Signal Processing, Springer, 2020, pp. 325-355.

[16] J. a. T. W. Pan, "A real-time QRS detection algorithm," IEEE transactions on biomedical engineering, vol. 3, pp. 230-236, 1985.