MSc Project - Reflective Essay

Project Title:	**Improved Arrhythmia Classification Using Select Morphological and Heart Rate Variability ECG Features.**
Student Name:	**Mark James Dunbar**
Student Number:	**210013360**
Supervisor Name:	**Dr Jesús Requena Carrion**
Programme of Study:	**MSc Big Data Science**

The chosen area of research presented focuses on constructing an arrhythmia classification pipeline that utilises a unique set of features to make predictions. The motivation for this study is the treatment of dangerous Ventricular Tachycardia (VT) arrhythmia using Implantable Cardioverter Defibrillator (ICD) devices. ICDs rely on the detection and identification of dangerous forms of arrhythmia such as VT to prevent sudden cardiac death (SCD) in affected patients [1]. Failure to correctly identify arrhythmia can have serious consequences such as inappropriate shock therapy [2], or even SCD if life-threatening arrhythmic heartbeats are not discovered. As such, this research aims to tackle this problem by introducing a new combination of features to produce an effective classification model.

Project Goal and Approach

The end goal of this project is a classification pipeline that takes 30-minute ECG signals as input and classifies each beat in the record as one of three classes: Normal (N), Ventricular Tachycardia (VT) or Supra-Ventricular Tachycardia (SVT). The pipeline functions following 5 basic steps:

1. *Raw ECG input*
2. *QRS Complex Detection*
3. *Heartbeat Segmentation*
4. *Feature Extraction*
5. *Heartbeat Classification*

This research focuses particularly on the *Feature Extraction* and *Classification* stages of the pipeline. Three forms of feature extraction were explored, comparing the use of morphological features [3] only, heart rate variability [4] (HRV) features only, and finally, a novel combination of both. In addition to this, these methods of feature extraction were analysed on a variety of different models to gauge the general impact of these feature sets on classifier performance.

Strengths and Weaknesses

Several strengths and weaknesses have been identified throughout the completion of this project. One strength of this study is that it provides experimental evidence that the use of combined HRV and morphological features provides a new method for the accurate detection and classification of arrhythmia. A novel, yet simple method for the extraction of key morphological QRS features of each heartbeat is proposed in this paper, building on feature extraction techniques first explored by *V. Mondejar-Guerra et. al (2019)* [5]. When used alone on heartbeat data, the proposed morphological feature extraction technique still yielded an average accuracy of ***93.25%*** across the classifier models trained using these features.

This project also provides an introductory exploration into the process of extracting HRV features using Lomb's periodogram [6], as opposed to the more commonly used Fast Fourier Transform (FFT) technique [7], leading on from research first proposed by *G. D. Clifford (2002)* [8]. Using this method, an average accuracy score of ***92.75%*** was achieved across all classifier algorithms tested on HRV data alone. Through conducting further research on this method, this study delivers a deeper understanding of some of the benefits and limitations of using the Lomb method, in particular:

- Lomb's periodogram removes an interpolation step that would otherwise be required to perform an FFT, retaining the resolution of the data.

- Lomb's method is quite complex, taking a significant amount of time to run for each heartbeat sample.

This project also provides a full ECG to heartbeat classification pipeline, which serves to explore the processes that are involved in ICD signal processing, as well as provide context for the data transformations that take place between the raw ECG signal, and the feature dataset used for heartbeat classification. For example; this pipeline may serve as the foundation for the development of a real-time heartbeat classification algorithm suitable for use in ICD devices in future work.

Despite these strengths, the project does hold some weaknesses and issues that should be addressed in future work. As a result of the data used on this project, the experimental results and classification pipeline yielded in this work does not provide an implementable solution for use in ICD arrhythmia classification. ECG signals are very similar, but not identical to the signals obtained by ICDs as they monitor electrical impulses generated by the heart. As ECG data was used, the findings of this work must be viewed as foundational evidence that these techniques can be applied to ICD data in future work, and not as a direct solution to the motivating problem of ICD arrhythmia classification.

Although a pipeline was constructed in this project, this drew some focus away from the exploration and analysis of suitable feature extraction techniques for heartbeat classification. An ideal solution would be to have split these two components into their own research problems; one paper that explores the creation of a fully-fledged classification pipeline for arrhythmic heartbeats, and a second paper that explores the feature extraction techniques used to distinguish arrhythmic heartbeats in more depth.

Limitations and Practical Challenges

One of the main limitations of this study is the lack of open-access ICD data available online. The main motivation for this project was to analyse different methods of heartbeat classification for the improvement of ICD treatment, however the MIT-BIH ECG datasets provided by PhysioNet [9] proved to be the most practical form of data that could be obtained and used. While the results and methods used on this data are directly translatable between ECG and ICD data, it is preferable to use ICD data to explore and identify any problems that may arise through use of such data.

A further limitation to this research is that the ECG data used does not practically allow for more than ~5 minutes worth of heartbeat preceding data to be collected and analysed. Each ECG record used is 30 minutes long – which, once the 5-minute preceding HRV metrics are calculated for each heartbeat, only ~25 minutes of each record are useable.

This project was conducted without the hands-on supervision and assistance of field experts, resulting in a large amount of time being spend by the researcher (myself) ensuring standard cardiology best practices methodologies were being followed correctly. Though this caveat presented an excellent learning opportunity and challenge for the researcher, this can also be viewed as a limitation due to the time-constraints placed on the project.

Further Work

The next natural step in progressing this research would be to consolidate the findings of this study through further experimentation with ICD data. While the results and methods used on ECG data are translatable between ECG and ICD data in theory, a follow-up study using ICD data would not only provide a classification pipeline that can be directly applied to ICD devices, but also identify any issues that may arise using such data, including the limitations of using the proposed algorithms with operational ICD devices. As ICD devices are battery operated, the computational and energy costs must be considered. More efficient methods and algorithms are favourable, as not to drain the limited resources of the ICD device [10].

This project did not take into account the computational cost of the algorithms used. In particular, problems may arise when applying the Lomb method to heartbeats, as this process is computationally taxing and may place an increased burden on the device's battery life. Furthermore, this method may not be practical or fast enough to be used on real-time heartbeat data.

One additional area of further research would involve exploring the difference in classification performance using 5-minute, 30 minute and even 24 hour HRV feature calculations. HRV metrics are very sensitive to changes in the time window used to calculate them [11], however this study did not capture just how strong an impact the variation of time on such measurements has.

Finally, a large number of end-of-life patients fitted with ICD devices are co-treated with antiarrhythmic drugs [12]. None of the patients included in the arrhythmia dataset used in this project were being treated with such drugs at the time of data collection, highlighting a significant gap in research that must be consolidated in further work. This may involve repeating the research proposed in this paper but with a representative cohort of patients being treated with antiarrhythmics. This would allow for the exploration of the effects of antiarrhythmics on the methods and results detailed in the project paper.

Ethical and Environmental Considerations

Constructing a classifier that could serve to improve the optimisation of ICD functionality was the main motivation behind this project, however ICD devices themselves present several ethical challenges in their usage and treatment [13]. Though the ethical challenges of ICD use are not explored in this paper, I believe it is critical for any researchers contributing to their functionality and improvement to be aware of this before conducting their research.

The data used in this project was sourced from the MIT-BIH ventricular arrhythmia, MIT-BIH supra-ventricular arrhythmia and St Petersburg INCART arrhythmia databases, made available to the public by PhysioNet as open-access data. Physionet is a web-based resource aimed at supporting and stimulating new and current research into clinical and biomedical data. PhysioNet provides a large archive of well-represented physiological data from a wide range of different studies contributed by members from across the research community. PhysioNet data is made publicly available under the OCD Public Domain Dedication and License v1.0 [14]. PhysioNet requests that any data used in research cites the relevant research of paper provided for each PhysioNet dataset.

In terms of the environmental impact of this study, most of this project was carried out in a Google Colaboratory or “Colab” environment, which uses Google hosted runtimes and hardware to execute scripts**.** Google as a company has made significant efforts to drastically offset the impact of their carbon footprint by purchasing large amounts of renewable energy [15].

Personal Development

Personally, this project has had an immense impact on my confidence and mindset as a data scientist. Data Science permeates all areas of modern research, and can be applied to a number of highly specialised fields and problems. Data scientists hold a plethora of tools that can be applied to problems in areas ranging from social networking to business analytics, however, not every expert in these fields has the means to analyse and explore data in a data science capacity. As a result, Data Scientists must adopt a collaborative approach with field experts to produce meaningful, conclusive insights in new territory. Therefore, it is essential for any data scientist to understand the importance of being adaptive, flexible, collaborative, and constantly ready to learn when approaching problems beyond their realm of expertise.

With this in mind, I chose a project beyond my academic understanding to experience the challenges and problems I am presented with when applying my skills in a completely new field. Tackling cardiology concepts from scratch presented a daunting intellectual challenge that I was able to overcome through effective research, and communication with peers within the field. Once a sufficient understanding of the subject matter was obtained, it became clear how to apply the data science expertise I have gained throughout my time as a student at QMUL to reach a solid scientific conclusion.

Conclusion

In this brief essay, the various contributions, strengths and limitations, ethical and environmental impacts that this project entails have been discussed. Furthermore, this essay also briefly discusses the personal impact that this project has had on me before and after the completion of this work. As a Physics graduate in the years before pursuing my postgraduate studies in Data Science, I have always enjoyed solving complex problems and explaining them in a concise, understandable manner. Data Science spans almost every domain of research in the modern world and has taught me how easily I can develop and expand my skills in a variety of new fields. I have become increasingly confident in my knowledge of machine learning, artificial intelligence, statistics, and programming through this project – putting every one of these areas of expertise to the test. This project is the ultimate culmination of my postgraduate studies at QMUL and has given me the confidence to tackle future problems in my career to come.

References

1. WG;, K. B. A. S. (2022). *Ventricular tachycardia and sudden cardiac death*. Mayo Clinic proceedings. Retrieved August 3, 2022, from https://pubmed.ncbi.nlm.nih.gov/19252119/
2. Li, A. C. (2020, December 10). *The significance of shocks in implantable cardioverter defibrillator recipients*. AER Journal. Retrieved August 3, 2022, from https://www.aerjournal.com/articles/significance-shocks-implantable-cardioverter-defibrillator-recipients
3. Science Direct. (2022). *Waveform morphology*. Waveform Morphology - an overview | ScienceDirect Topics. Retrieved August 3, 2022, from https://www.sciencedirect.com/topics/medicine-and-dentistry/waveform-morphology
4. Cleveland Clinic. (2022). *Heart rate variability (HRV): What it is and how you can track it*. Cleveland Clinic. Retrieved August 4, 2022, from https://my.clevelandclinic.org/health/symptoms/21773-heart-rate-variability-hrv
5. Mondéjar-Guerra, V., Novo, J., Rouco, J., Penedo, M. G., & Ortega, M. (2019). Heartbeat classification fusing temporal and morphological information of ecgs via ensemble of classifiers. *Biomedical Signal Processing and Control*, *47*, 41–48. https://doi.org/10.1016/j.bspc.2018.08.007
6. VanderPlas, J. T. (2018). Understanding the lomb–scargle periodogram. *The Astrophysical Journal Supplement Series*, *236*(1), 16. https://doi.org/10.3847/1538-4365/aab766
7. Maklin, C. (2019, December 29). *Fast fourier transform*. Medium. Retrieved August 5, 2022, from https://towardsdatascience.com/fast-fourier-transform-937926e591cb
8. Gari D. Clifford. *Signal Processing Methods for Heart Rate Variability*. PhD thesis. 2002.
9. PhysioNet. (2020). *MIT-BiH Arrhythmia Database Directory*. MIT-BIH Arrhythmia Database Directory (Introduction). Retrieved August 12, 2022, from https://archive.physionet.org/physiobank/database/html/mitdbdir/intro.htm
10. Boriani, G., Merino, J., Wright, D. J., Gadler, F., Schaer, B., & Landolina, M. (2018). Battery longevity of implantable cardioverter-defibrillators and cardiac resynchronization therapy defibrillators: Technical, clinical and economic aspects. an expert review paper from Ehra. *EP Europace*, *20*(12), 1882–1897. https://doi.org/10.1093/europace/euy066
11. Li, K., Rüdiger, H., & Ziemssen, T. (2019). Spectral analysis of heart rate variability: Time Window Matters. *Frontiers in Neurology*, *10*. https://doi.org/10.3389/fneur.2019.00545
12. Bollmann, A., Husser, D., & Cannom, D. S. (2005). Antiarrhythmic drugs in patients with implantable cardioverter-defibrillators. *American Journal of Cardiovascular Drugs*, *5*(6), 371–378. https://doi.org/10.2165/00129784-200505060-00004
13. STRÖMBERG, A., FLUUR, C., MILLER, J., CHUNG, M., MOSER, D., & THYLÉN, I. (2014). ICD recipients’ understanding of ethical issues, ICD function, and practical consequences of withdrawing the ICD in the end-of-life. *Pacing and Clinical Electrophysiology*, *37*(7), 834–842. https://doi.org/10.1111/pace.12353

14. Open Data Commons. (2022). *Open data commons*. Open Data Commons Public Domain Dedication and License (PDDL) v1.0 - Open Data Commons: legal tools for open data. Retrieved August 12, 2022, from https://opendatacommons.org/licenses/pddl/1-0/
15. Google. (2022). *Commitment to a carbon-free future*. Google Sustainability. Retrieved August 8, 2022, from https://sustainability.google/commitments/carbon/#leading-at-google