

EVALUATING PREDICTION ALGORITHM OF MALIGNANT  
VENTRICULAR ARRHYTHMIA FOR EARLIER PREDICTION TIME  
ON HETEROGENOUS DATABASES

WILLIAM MOK WEN LENG



Dr. Nurul Ashikin Binti Abdul Kadir  
School of Electrical Engineering  
Faculty of Engineering  
Universiti Teknologi Malaysia  
81310 UTM Johor Bahru  
Johor  
Email : ashikin.kadir@utm.my

UNIVERSITI TEKNOLOGI MALAYSIA

EVALUATING PREDICTION ALGORITHM OF MALIGNANT  
VENTRICULAR ARRHYTHMIA FOR EARLIER PREDICTION TIME  
ON HETEROGENOUS DATABASES

WILLIAM MOK WEN LENG

A project report submitted in partial fulfilment of the  
requirements for the award of the degree of  
Master of Engineering (Computer and Microelectronic Systems)

School of Electrical Engineering  
Faculty of Engineering  
Universiti Teknologi Malaysia

JULY 2022

## DEDICATION

*To my parents, siblings, supervisor, and my friends.*

## **ACKNOWLEDGEMENT**

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Ts. Dr. Nurul Ashikin Binti Abdul Kadir, for encouragement, guidance, critics, and friendship. I am also very thankful to my co-supervisor Dr Heng Wei Wei for her guidance, advice, and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to Intel Malaysia for funding my Master study. Universiti Teknologi Malaysia (UTM) also deserve special thanks for their assistance in supplying the relevant literatures.

My fellow postgraduate student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family member.

## ABSTRACT

Prediction of malignant ventricular arrhythmia (mVA) is utmost imperative to enable earlier medical intervention and prevent sudden cardiac death (SCD). However, patients with a history of coronary artery disease (CAD) and congestive heart failure (CHF) are at higher risk of SCD. This thesis aimed to develop a reliable mVA prediction algorithm with high performance and an earlier prediction time and evaluate in a more authentic situation mixed with other cardiac diseases which are CAD and CHF. This was done by testing the algorithm on multiple online databases which are Sudden Cardiac Death Holter Database (SDDDB), MIT-BIH Normal Sinus Rhythm Database (NSRDB), Long Term ST Database (LTSTDB) and BIDMC Congestive Heart Failure Database (CHFDB). Heart rate variability (HRV) analysis with support vector machine (SVM) was employed in the prediction algorithm due to its reliability observed in previous works. To investigate the statistical relationship between all databases, 65 features were extracted from first, second, third, and fourth minute HRV signal before mVA onset and before two hours mark of control signals. Experimental results show a significant difference in HRV of mVA signals and other non-mVA signals, including six time-domain features and seven nonlinear features. Six feature combinations from time-segment-specific classification were found to perform best in predicting imminent mVA in situation mixed with CAD and CHF. High accuracy of 97.33% with 89.47% sensitivity and 100% specificity was achieved. For classification of the four distinct databases, four feature combinations of  $pNN50$ ,  $MaxNN$  and  $CVI$  with  $CVNN$ ,  $SD2$ ,  $SD1a$ , or  $SDNNa$  achieved a high accuracy of 98.67% with 100% sensitivity and 98.21% specificity. For exploration of earlier prediction time, the six best-performing feature combinations in predicting imminent mVA with other non-mVA signals were selected for classifier training and testing in leave-one-out cross-validation classification on 120-minutes signal. A balanced performance with reasonably high accuracy of 73.33%, sensitivity of 73.68%, specificity of 73.21% and 91.14 minutes of earliest prediction time was achieved by combination of  $pNN50$ ,  $SD1d$ ,  $SDNNa$  with Gaussian radial basis function (RBF) SVM and moving average of 15 minutes.

## ABSTRAK

Ramalan aritmia ventrikel malignan (mVA) adalah amat penting untuk membolehkan perubatan lebih awal dan mencegah kematian mendadak serangan jantung (SCD). Walau bagaimanapun, pesakit yang mempunyai sejarah penyakit arteri koronari (CAD) dan kegagalan jantung kongestif (CHF) mempunyai risiko yang lebih tinggi untuk mengalami SCD. Tesis ini bertujuan untuk mencipta algoritma ramalan mVA yang berprestasi tinggi dan masa ramalan yang lebih awal serta dinilai dalam situasi yang lebih tulen bercampur dengan penyakit jantung lain iaitu CAD dan CHF. Ini dilakukan dengan menguji algoritma pada beberapa pangkalan data dalam talian iaitu *Sudden Cardiac Death Holter Database* (SDDB), *MIT-BIH Normal Sinus Rhythm Database* (NSRDB), *Long Term ST Database* (LTSTDB) dan *BIDMC Congestive Heart Failure Database* (CHFDB). Analisis kebolehubahan kadar jantung (HRV) dengan mesin vektor sokongan (SVM) telah digunakan dalam algoritma ramalan kerana kebolehpercayaannya diperhatikan dalam kerja-kerja terdahulu. Untuk menyiasat hubungan statistik antara semua pangkalan data, 65 ciri telah diekstrak daripada isyarat HRV minit pertama, kedua, ketiga dan keempat sebelum permulaan mVA dan sebelum tanda dua jam isyarat kawalan. Keputusan eksperimen menunjukkan perbezaan ketara dalam HRV isyarat mVA dan isyarat bukan mVA lain, termasuk enam ciri domain masa dan tujuh ciri tak linear. Enam kombinasi ciri daripada klasifikasi khusus segmen masa didapati berprestasi terbaik dalam meramalkan mVA yang akan berlaku dalam situasi bercampur dengan CAD dan CHF. Ketepatan tinggi 97.33% dengan sensitiviti 89.47% dan spesifisiti 100% telah dicapai. Untuk pengelasan empat pangkalan data yang berbeza, empat kombinasi ciri *pNN50*, *MaxNN* dan *CVI* dengan *CVNN*, *SD2*, *SD1a* atau *SDNNa* mencapai ketepatan tinggi 98.67% dengan sensitiviti 100% dan spesifisiti 98.21%. Untuk penerokaan masa ramalan yang lebih awal, enam kombinasi ciri berprestasi terbaik dalam meramalkan mVA yang akan berlaku telah dipilih untuk klasifikasi pengesahan silang cuti satu keluar pada isyarat 120 minit. Prestasi seimbang dengan ketepatan yang agak tinggi iaitu 73.33%, kepekaan 73.68%, kekhususan 73.21% dan 91.14 minit masa ramalan terawal dicapai dengan gabungan *pNN50*, *SD1d*, *SDNNa* dengan fungsi asas jejari (RBF) Gaussian SVM dan purata bergerak selama 15 minit.

## TABLE OF CONTENTS

|                  | <b>TITLE</b>                                   | <b>PAGE</b> |
|------------------|------------------------------------------------|-------------|
|                  | <b>DECLARATION</b>                             | <b>i</b>    |
|                  | <b>DEDICATION</b>                              | <b>ii</b>   |
|                  | <b>ACKNOWLEDGEMENT</b>                         | <b>iii</b>  |
|                  | <b>ABSTRACT</b>                                | <b>iv</b>   |
|                  | <b>ABSTRAK</b>                                 | <b>v</b>    |
|                  | <b>TABLE OF CONTENTS</b>                       | <b>vi</b>   |
|                  | <b>LIST OF TABLES</b>                          | <b>ix</b>   |
|                  | <b>LIST OF FIGURES</b>                         | <b>x</b>    |
|                  | <b>LIST OF ABBREVIATIONS</b>                   | <b>xi</b>   |
|                  | <b>LIST OF SYMBOLS</b>                         | <b>xii</b>  |
| <b>CHAPTER 1</b> | <b>INTRODUCTION</b>                            | <b>1</b>    |
| 1.1              | Background                                     | 1           |
| 1.2              | Problem Statement                              | 3           |
| 1.3              | Research Objective                             | 4           |
| 1.4              | Research Scope                                 | 4           |
| 1.5              | Significance of Study                          | 5           |
| 1.6              | Thesis Outline                                 | 5           |
| <b>CHAPTER 2</b> | <b>LITERATURE REVIEW</b>                       | <b>7</b>    |
| 2.1              | Introduction                                   | 7           |
| 2.2              | Sudden Cardiac Death (SCD)                     | 7           |
| 2.2.1            | Malignant Ventricular Arrhythmia (mVA)         | 8           |
| 2.2.1.1          | Ventricular Fibrillation (VF)                  | 8           |
| 2.2.1.2          | Ventricular Tachycardia (VT)                   | 9           |
| 2.2.2            | Coronary Artery Disease (CAD)                  | 10          |
| 2.2.3            | Congestive Heart Failure (CHF)                 | 11          |
| 2.3              | Prediction of Malignant Ventricular Arrhythmia | 11          |

|                  |                                                     |           |
|------------------|-----------------------------------------------------|-----------|
| 2.3.1            | Prediction of mVA: Multi Class Approach             | 13        |
| 2.3.1.1          | Detection of Coronary Artery Disease                | 14        |
| 2.3.1.2          | Detection of Congestive Heart Failure               | 15        |
| 2.3.2            | Prediction of mVA: Earlier Prediction Time          | 17        |
| 2.4              | Prediction of mVA: Research Gap                     | 18        |
| 2.5              | Summary                                             | 21        |
| <b>CHAPTER 3</b> | <b>RESEARCH METHODOLOGY</b>                         | <b>23</b> |
| 3.1              | Introduction                                        | 23        |
| 3.2              | Overall Workflow                                    | 23        |
| 3.3              | Dataset                                             | 24        |
| 3.4              | Data Pre-Processing                                 | 25        |
| 3.5              | Feature Extraction                                  | 26        |
| 3.6              | Statistical Analysis                                | 32        |
| 3.7              | Classification with Support Vector Machine (SVM)    | 32        |
| 3.7.1            | Time-Segment-Specific Hyperparameter Selection      | 33        |
| 3.7.2            | Leave-One-Out Cross-Validation                      | 34        |
| 3.8              | Performance Evaluation                              | 35        |
| 3.9              | Research Tool                                       | 37        |
| 3.10             | Chapter Summary                                     | 37        |
| <b>CHAPTER 4</b> | <b>RESULT &amp; DISCUSSION</b>                      | <b>39</b> |
| 4.1              | Introduction                                        | 39        |
| 4.2              | Features Analysis                                   | 39        |
| 4.3              | Time-segment-specific Performance Analysis          | 47        |
| 4.4              | Leave-one-out Cross-Validation Performance Analysis | 51        |
| 4.5              | Discussion                                          | 56        |
| 4.6              | Chapter Summary                                     | 59        |
| <b>CHAPTER 5</b> | <b>CONCLUSION</b>                                   | <b>61</b> |
| 5.1              | Concluding Remark                                   | 61        |



|     |                                 |                |
|-----|---------------------------------|----------------|
| 5.2 | Recommendations for Future Work | 62             |
|     | <b>REFERENCES</b>               | <b>64</b>      |
|     | <b>Appendices A - B</b>         | <b>69 - 72</b> |

## LIST OF TABLES

| <b>TABLE NO.</b> | <b>TITLE</b>                                                                         | <b>PAGE</b> |
|------------------|--------------------------------------------------------------------------------------|-------------|
| Table 2.1        | Previous detection work on CAD                                                       | 15          |
| Table 2.2        | Previous detection work on CHF                                                       | 16          |
| Table 2.3        | Previous prediction work on mVA                                                      | 20          |
| Table 3.1        | List of Databases                                                                    | 25          |
| Table 3.2        | Features extracted from HRV signals                                                  | 27          |
| Table 3.3        | Confusion matrix for two-classes comparison                                          | 35          |
| Table 3.4        | Confusion matrix for three-classes comparison                                        | 35          |
| Table 3.5        | Confusion matrix for four-classes comparison                                         | 36          |
| Table 4.1        | Statistical test result of time-domain features                                      | 40          |
| Table 4.2        | Statistical test result of frequency-domain features                                 | 42          |
| Table 4.3        | Statistical test result of nonlinear features                                        | 43          |
| Table 4.4        | Best-performing features combination for two-classes comparison                      | 48          |
| Table 4.5        | Best-performing features combination for three-classes comparison                    | 49          |
| Table 4.6        | Best-performing features combination for four-classes comparison                     | 50          |
| Table 4.7        | Performance comparison against previous work on time-segment-specific classification | 51          |
| Table 4.8        | Performance comparison against previous work on earliest prediction time             | 57          |

## LIST OF FIGURES

| FIGURE NO. | TITLE                                                                                                                                   | PAGE |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------|------|
| Figure 2.1 | Causes of Sudden Cardiac Death [4]                                                                                                      | 7    |
| Figure 2.2 | Ventricular Fibrillation ECG Waveform                                                                                                   | 8    |
| Figure 2.3 | (a) Monomorphic and (b) Polymorphic Ventricular Tachycardia ECG                                                                         | 9    |
| Figure 2.4 | Coronary artery disease ECG waveform                                                                                                    | 10   |
| Figure 2.5 | Congestive heart failure ECG waveform [21]                                                                                              | 11   |
| Figure 3.1 | Overall workflow of the research                                                                                                        | 24   |
| Figure 3.2 | Signal segments extracted from each ECG records                                                                                         | 26   |
| Figure 3.3 | Flowchart of leave-out-out cross-validation process                                                                                     | 34   |
| Figure 3.4 | Different TP, FP, TN, FN definitions of leave-one-out cross validation and time-segment-specific metrics [12]                           | 36   |
| Figure 4.1 | Overall performance metrics and earliest prediction time for $pNN50$ , $SD2$ , $SD1d$ with Gaussian RBF SVM (C=1000, gamma=1)           | 52   |
| Figure 4.2 | Overall performance metrics and earliest prediction time for $pNN50$ , $SD1d$ , $SDNNa$ with Gaussian RBF SVM (C=1000, gamma=1)         | 53   |
| Figure 4.3 | Overall performance metrics and earliest prediction time for $RMSSD$ , $CVSD$ , $pNN50$ , $SD2$ with Gaussian RBF SVM (C=100, gamma=1)  | 54   |
| Figure 4.4 | Overall performance metrics and earliest prediction time for $CVNN$ , $CVSD$ , $pNN50$ , $SDNNa$ with Gaussian RBF SVM (C=100, gamma=1) | 54   |
| Figure 4.5 | Overall performance metrics and earliest prediction time for $CVSD$ , $pNN50$ , $SD1$ , $SD2$ with Gaussian RBF SVM (C=100, gamma=1)    | 55   |
| Figure 4.6 | Overall performance metrics and earliest prediction time for $SDNN$ , $pNN50$ , $SD2$ , $CVI$ with Cubic SVM                            | 55   |

## LIST OF ABBREVIATIONS

|        |   |                                                  |
|--------|---|--------------------------------------------------|
| AED    | - | Automated External Defibrillator                 |
| ANN    | - | Artificial Neural Network                        |
| ANOVA  | - | Analysis of Variance                             |
| CAD    | - | Coronary Artery Disease                          |
| CHF    | - | Congestive Heart Failure                         |
| CHFDB  | - | BIDMC Congestive Heart Failure Database          |
| DT     | - | Decision Tree                                    |
| ECG    | - | Electrocardiogram                                |
| FN     | - | False Negative                                   |
| FP     | - | False Positive                                   |
| HRV    | - | Heart Rate Variability                           |
| ICD    | - | Implantable Cardioverter Defibrillator           |
| KNN    | - | K Nearest Neighbour                              |
| LI     | - | Line of Identity                                 |
| LTSTDB | - | Long Term ST Database                            |
| MATLAB | - | Matrix Laboratory                                |
| mVA    | - | Malignant Ventricular Arrhythmia                 |
| MVTDB  | - | Spontaneous Ventricular Tachyarrhythmia Database |
| NSR    | - | Normal Sinus Rhythm                              |
| NSRDB  | - | Normal Sinus Rhythm Database                     |
| RBF    | - | Radial Basis Function                            |
| RF     | - | Random Forest                                    |
| SCA    | - | Sudden Cardiac Arrest                            |
| SCD    | - | Sudden Cardiac Death                             |
| SDDB   | - | Sudden Cardiac Death Holter Database             |
| SVM    | - | Support Vector Machine                           |
| TN     | - | True Negative                                    |
| TP     | - | True Positive                                    |
| VA     | - | Ventricular Arrhythmia                           |
| VF     | - | Ventricular Fibrillation                         |
| VFDB   | - | Ventricular Fibrillation Database                |
| VT     | - | Ventricular Tachycardia                          |
| WFDB   | - | Waveform Database                                |

## LIST OF SYMBOLS

|                 |   |                                                                                                  |
|-----------------|---|--------------------------------------------------------------------------------------------------|
| $\sigma$        | - | Minimal error                                                                                    |
| $\gamma$        | - | Gamma                                                                                            |
| <i>MeanNN</i>   | - | Mean of RR intervals                                                                             |
| <i>SDNN</i>     | - | Standard deviation of RR intervals                                                               |
| <i>RMSSD</i>    | - | Square root of the mean squared differences of successive RR intervals                           |
| <i>CVNN</i>     | - | Coefficient of variation of RR intervals                                                         |
| <i>CVSD</i>     | - | Coefficient of deviation of RR intervals                                                         |
| <i>MedianNN</i> | - | Median of absolute value of successive differences between RR intervals                          |
| <i>MadNN</i>    | - | Median of absolute deviation of the RR intervals                                                 |
| <i>MCVNN</i>    | - | Median coefficient of variation of RR intervals                                                  |
| <i>IQRNN</i>    | - | Interquartile range of the RR intervals                                                          |
| <i>Prc20NN</i>  | - | 20th percentile of the RR intervals                                                              |
| <i>Prc80NN</i>  | - | 80th percentile of the RR intervals                                                              |
| <i>pNN50</i>    | - | proportion of successive RR interval differ more than 50ms                                       |
| <i>pNN20</i>    | - | proportion of successive RR interval differ more than 20ms                                       |
| <i>MinNN</i>    | - | Minimum RR intervals                                                                             |
| <i>MaxNN</i>    | - | Maximum RR intervals                                                                             |
| <i>HTI</i>      | - | Triangular index of RR intervals                                                                 |
| <i>TINN</i>     | - | Triangular interpolation of RR interval histogram                                                |
| <i>LF</i>       | - | Spectral power of low frequencies (0.04-0.15Hz)                                                  |
| <i>HF</i>       | - | Spectral power of high frequencies (0.15-0.4Hz)                                                  |
| <i>VHF</i>      | - | Spectral power of very high frequencies (0.4-0.5Hz)                                              |
| <i>LFHF</i>     | - | Ratio of <i>LF</i> to <i>HF</i> power                                                            |
| <i>LFn</i>      | - | Normalized low frequency                                                                         |
| <i>HFn</i>      | - | Normalized high frequency                                                                        |
| <i>LnHF</i>     | - | Natural logarithm of <i>HF</i>                                                                   |
| <i>SD1</i>      | - | Standard deviation of the distance along the shorter axis of the fitted ellipse of Poincare Plot |
| <i>SD2</i>      | - | Standard deviation of the distance along the longer axis of the fitted ellipse of Poincare Plot  |
| <i>SD1SD2</i>   | - | Ratio of <i>SD1</i> to <i>SD2</i> of Poincare Plot                                               |
| <i>S</i>        | - | Area of ellipse described by <i>SD1</i> and <i>SD2</i> .                                         |
| <i>CSI</i>      | - | Cardiac Sympathetic Index                                                                        |

|                     |   |                                                                          |
|---------------------|---|--------------------------------------------------------------------------|
| <i>CVI</i>          | - | Cardiac Vagal Index                                                      |
| <i>CSI_Modified</i> | - | Modified <i>CSI</i>                                                      |
| <i>SD1d</i>         | - | Short-term variance of contributions of prolongations of RR intervals    |
| <i>SD1a</i>         | - | Short-term variance of contributions of shortenings of RR intervals      |
| <i>C1d</i>          | - | Contributions of heart rate decelerations to short-term HRV              |
| <i>C1a</i>          | - | Contributions of heart rate accelerations to short-term HRV              |
| <i>SD2d</i>         | - | Long-term variance of contributions of prolongations of RR intervals     |
| <i>SD2a</i>         | - | Long-term variance of contributions of shortenings of RR intervals       |
| <i>C2d</i>          | - | Contributions of heart rate decelerations to long-term HRV               |
| <i>C2a</i>          | - | Contributions of heart rate accelerations to long-term HRV               |
| <i>SDNNd</i>        | - | Total variance of contribution of prolongations of RR intervals          |
| <i>SDNNa</i>        | - | Total variance of contribution of shortenings of RR intervals            |
| <i>Cd</i>           | - | Total contributions of heart rate decelerations to HRV                   |
| <i>Ca</i>           | - | Total contributions of heart rate accelerations to HRV                   |
| <i>PIP</i>          | - | Percentage of inflection points of the RR intervals series               |
| <i>IALS</i>         | - | Inverse of the average length of the acceleration/deceleration segments. |
| <i>PSS</i>          | - | Percentage of short segments                                             |
| <i>PAS</i>          | - | Percentage of RR intervals in alternation segments                       |
| <i>GI</i>           | - | Guzik's Index                                                            |
| <i>SI</i>           | - | Slope Index                                                              |
| <i>AI</i>           | - | Area Index                                                               |
| <i>PI</i>           | - | Porta's Index                                                            |
| <i>DFA_alpha1</i>   | - | Short-term correlation properties of detrended fluctuation analysis      |
| <i>DFA_alpha2</i>   | - | Long-term correlation properties of detrended fluctuation analysis       |
| <i>ApEn</i>         | - | Approximate entropy of RR intervals                                      |
| <i>SampEn</i>       | - | Sample entropy of RR intervals                                           |
| <i>ShanEn</i>       | - | Shannon entropy of RR intervals                                          |
| <i>FuzzyEn</i>      | - | Fuzzy entropy of RR intervals                                            |
| <i>CD</i>           | - | Correlation Dimension                                                    |
| <i>HFD</i>          | - | Higuchi's Fractal Dimension                                              |
| <i>KFD</i>          | - | Katz's Fractal Dimension                                                 |
| <i>LZC</i>          | - | Lempel-Ziv Complexity                                                    |
| <i>Activity</i>     | - | Hjorth Activity                                                          |
| <i>Mobility</i>     | - | Hjorth Mobility                                                          |
| <i>Complexity</i>   | - | Hjorth Complexity                                                        |
| <i>Hurst_rs</i>     | - | Hurst exponent of RR intervals                                           |

## LIST OF APPENDICES

| <b>APPENDIX</b> | <b>TITLE</b>                               | <b>PAGE</b> |
|-----------------|--------------------------------------------|-------------|
| Appendix A      | Classification Python Code                 | 69          |
| Appendix B      | Leave-one-out Cross-validation Python Code | 72          |

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

According to World Health Organization, 17.9 million people died from cardiovascular disease (CVD) in year 2019, accounting for 32% of the global deaths [1]. Sudden cardiac arrest (SCA) accounts for approximately 50% of all deaths attributed to CVD [2]. SCA is an unexpected loss in working of the heart muscles that may cause sudden cardiac death (SCD). In this context, malignant ventricular arrhythmia (mVA), which refers to a group of life-threatening heart arrhythmia comprising of ventricular tachycardia (VT), and ventricular fibrillation (VF) contributed to more than half of the SCD.

SCD can arise with or without a history of heart disease [3]. Patients with a history of coronary artery disease (CAD) are, nevertheless, more likely to develop SCD [4]. The narrowing of coronary arteries, which reduces blood flow to the heart muscles, is known as coronary artery disease (CAD). This is usually caused by cholesterol plaque forming inside the coronary arteries as a result of an unhealthy lifestyle, pollution, smoking habits, or any other unknown factor [5]. If left untreated, CAD can lead to infarction in the coronary arteries, reducing the heart's ability to provide oxygenated blood to the body. Congestive heart failure (CHF) is an incapacity of the heart that can lead to SCA.

An electronic defibrillator or cardiopulmonary resuscitation (CPR) may be used to help restore normal heart function during SCA [4]. A defibrillator is an electrical device that sends electrical stimulation to the heart to restore normal heart function. An implantable cardioverter defibrillator (ICD) may be used to treat patients. In ICD, electrocardiographic indicators are used to predict a future mVA events. Prediction of mVA events is beneficial as it enables earlier attention to the



patient and allow more time for medical intervention. However, ICD is expensive and exposes the individuals to risk of bleeding and device-related infection. Despite advances in signal processing techniques, there is still no optimal approach for forecasting deadly arrhythmic occurrences [4]. Thus, it is utmost important to develop accurate, reliable, and non-invasive risk assessment method for mVA events.

Some invasive methods, such as cardiac angiography and catheterization, are also utilised in clinical practise to diagnose problems in heart function. However, patients often suffer unnecessary pain and discomfort because of the invasive methods. Thus, non-invasive methods should be used over invasive approaches. One of the most extensively utilised non-invasive ways for identifying mVA events is the analysis of electrocardiography-based indicators, also known as electrocardiogram (ECG) [6]. ECG is a graphical representation of the heart's electrical activity, which is obtained through electrodes placed on the skin. During each cardiac cycle, these electrodes detect the tiny electrical changes that occur as a result of cardiac muscle depolarization and repolarization. Several cardiac disorders, including mVA, will induce changes in the ECG pattern, thus it is widely adopted for diagnosing CVD.

Up to three-quarters of patients report warning symptoms such chest pain for a median of one hour before the SCA occurs [7]. Early medical attention after the emergence of warning symptoms can considerably aid in the prevention of SCD. However, patients may not recognise these chronic symptoms and delay the chance of earlier access to medical assistance. Therefore, earlier prediction of mVA using ECG, up to hours ahead of time, is a potentially effective prevention strategy that can reduce mortality due to SCA incidence by enabling patients to obtain appropriate medical intervention and increasing the efficiency of treatment delivery in hospital.

Several research has been carried out to predict mVA events using various machine learning techniques and different feature sets. Researchers have successfully investigated ways to increase the performance metrics and prediction time of their prediction algorithms. Nonetheless, there are several enhancements that could be made to allow for more extensive research. In this paper, earlier prediction of mVA using machine learning on heterogenous databases will be discussed extensively.

## 1.2 Problem Statement

There are three research issues discovered from previous research works. Firstly, previous research only shows the possibility of mVA prediction against the normal sinus rhythm (NSR) without taking other CVD into account [8–13]. This may cause false alarms to occur easily as other CVD could be very different from NSR in feature space. Recently, Devi *et al.* have proposed a novel multi-class approach for the prediction of mVA by comparing heart rate variability (HRV) in mVA with NSR and CHF [14]. After that, Rohila *et al.* further improved the study by including CAD database into the classification [15]. However, Rohila *et al.* only have not emphasized on the prediction of mVA by analysing a small duration of the HRV signal.

Secondly, statistical relationship of different cardiac diseases with the development of mVA over time could be investigated by incorporating these arrhythmias in the prediction algorithm. Although Rohila *et al.* [15] explored on the multi-classes approach, but the primary objective of their study was to compare the HRV profile in subjects at risk of mVA. Therefore, feature selection was not applied in their study. However, it is undeniable that there are differences of statistical characteristic in feature space due to onset of mVA events and other existing cardiac disease. Thus, feature selection is utmost significant to allow exploration and investigation on the statistical relationship of different cardiac diseases to the development of mVA events over time.

Thirdly, the performance of the prediction algorithms with earlier prediction time usually come with a lower performance metrics. Most of the previous studies looked at prediction time of up to thirty minutes, with accuracy of greater than 80%. Only few explore the earlier prediction time. Exploration of a longer signal prior to the mVA is advantageous for the finding of an earlier prediction time frame for early medical intervention. On 91-minute ECG readings, Heng [12] found that a combination of heart rate variability features, support vector machine classifier, and firing power approach obtained an overall accuracy of 89.47% and an average earliest forecast time of 77 minutes. This could be further validated and improved.

### **1.3 Research Objective**

This thesis aimed to develop a reliable mVA prediction algorithm with high performance and an earlier prediction time while taking CAD and CHF into account. In short, this research embarked on the following objectives:

1. To improve the prediction algorithm in a more authentic situation mixed with other cardiac diseases such as CAD and CHF on multiple databases.
2. To investigate the statistical relationship of CAD and CHF with the development of mVA.
3. To validate the prediction algorithm using longer signal before mVA events to explore earlier prediction time.

### **1.4 Research Scope**

Based on the research objectives, the scope of this thesis work is listed below:

- ECG signal was the main signal used in this study and was obtained from PhysioBank database [16], which is publicly available to scholars around the world for benchmarking and comparison against previous works. The databases include Sudden Cardiac Death Holter Database (SDDB), MIT-BIH Normal Sinus Rhythm Database (NSRDB), Long Term ST Database (LTSTDB) and BIDMC Congestive Heart Failure Database (CHFDB).
- This study was narrowed to long-term prediction which was prediction of mVAs hours in advance to enable earlier medical intervention.
- HRV analysis with SVM were the main techniques used in this project.
- HRV features were extracted from time-domain and nonlinear analyses.
- Prediction time and performance metrics were used to evaluate the algorithm. The performance metrics included accuracy, sensitivity, and specificity.

## **1.5 Significance of Study**

With advancement of technology, ECG can be obtained easily through mobile devices and wearables. The prediction algorithm developed in this research work may be potentially implemented not only in clinical ECG monitoring system, but also in widely available mobile wearables to keep track of more patients at risk. With the earlier prediction time and more authentic situation mixed with other cardiovascular diseases, patients will be given more accurate warning hours before mVA occurrences. This may significantly improve the survival chances of a patient by enabling earlier diagnosis.

## **1.6 Thesis Outline**

This thesis consists of five chapters, as outlined below.

- Chapter 1 describes the research background, motivations, objectives and scopes of this research.
- Chapter 2 starts with background knowledge on sudden cardiac death and cardiovascular diseases associated with it, including malignant ventricular arrhythmia, coronary artery disease and congestive heart failure. Then, this chapter reviews the previous works on prediction of mVA, CAD and CHF, as well as early prediction time of these diseases.
- Chapter 3 provides an overview of the research methodology used in this research work. The workflows are discussed extensively. This chapter also introduces the research tools.
- Chapter 4 presents the experimental results and discussion. The prediction time and performance of the algorithm are recorded and analysed, and then discussed extensively.
- Chapter 5 concludes this thesis by summarising the research findings and recommendations for future research.

## REFERENCES

- [1] “Cardiovascular diseases (CVDs).” [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). [Accessed: 12-Jan-2022].
- [2] J. D. Sara, M. F. Eleid, R. Gulati, and D. R. Holmes, “Sudden cardiac death from the perspective of coronary artery disease,” *Mayo Clin. Proc.*, vol. 89, no. 12, pp. 1685–1698, Dec. 2014.
- [3] O. Yousuf, J. Chrispin, G. F. Tomaselli, and R. D. Berger, “Clinical Management and Prevention of Sudden Cardiac Death,” *Circ. Res.*, vol. 116, no. 12, pp. 2020–2040, Jun. 2015.
- [4] N. T. Srinivasan and R. J. Schilling, “Sudden cardiac death and arrhythmias,” *Arrhythmia Electrophysiol. Rev.*, vol. 7, no. 2, pp. 111–117, 2018.
- [5] U. R. Acharya *et al.*, “Linear and nonlinear analysis of normal and CAD-affected heart rate signals,” *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 55–68, Jan. 2014.
- [6] V. Goyal, D. S. Jassal, and N. S. Dhalla, “Pathophysiology and prevention of sudden cardiac death,” <https://doi.org/10.1139/cjpp-2015-0366>, vol. 94, no. 3, pp. 237–244, Sep. 2015.
- [7] T. D. Rea and R. L. Page, “Community approaches to improve resuscitation after out-of-hospital sudden cardiac arrest,” *Circulation*, vol. 121, no. 9, pp. 1134–1140, Mar. 2010.
- [8] A. G. Raka, G. R. Naik, and R. Chai, “Computational Algorithms Underlying the Time-Based Detection of Sudden Cardiac Arrest via Electrocardiographic Markers,” *Appl. Sci. 2017, Vol. 7, Page 954*, vol. 7, no. 9, p. 954, Sep. 2017.
- [9] J. P. Amezcua-Sanchez, M. Valtierra-Rodriguez, H. Adeli, and C. A. Perez-Ramirez, “A Novel Wavelet Transform-Homogeneity Model for Sudden Cardiac Death Prediction Using ECG Signals,” *J. Med. Syst.*, vol. 42, no. 10, pp. 1–15, Oct. 2018.
- [10] D. Lai, Y. Zhang, X. Zhang, Y. Su, and M. B. Bin Heyat, “An Automated Strategy for Early Risk Identification of Sudden Cardiac Death by Using Machine Learning Approach on Measurable Arrhythmic Risk Markers,” *IEEE*

- Access*, vol. 7, pp. 94701–94716, 2019.
- [11] A. Parsi, D. Byrne, M. Glavin, and E. Jones, “Heart rate variability feature selection method for automated prediction of sudden cardiac death,” *Biomed. Signal Process. Control*, vol. 65, no. December, 2021.
- [12] W. W. Heng, “Short-term Prediction of Malignant Ventricular Arrhythmia using Phase Space Reconstruction and Heart Rate Variability from ElectroCardiogram,” Universiti Teknologi Malaysia, 2020.
- [13] H. Fujita *et al.*, “Sudden cardiac death (SCD) prediction based on nonlinear heart rate variability features and SCD index,” *Appl. Soft Comput.*, vol. 43, pp. 510–519, Jun. 2016.
- [14] R. Devi, H. K. Tyagi, and D. Kumar, “A novel multi-class approach for early-stage prediction of sudden cardiac death,” *Biocybern. Biomed. Eng.*, vol. 39, no. 3, pp. 586–598, Jul. 2019.
- [15] A. Rohila and A. Sharma, “Detection of sudden cardiac death by a comparative study of heart rate variability in normal and abnormal heart conditions,” *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 1140–1154, Jul. 2020.
- [16] A. L. Goldberger *et al.*, “PhysioBank, PhysioToolkit, and PhysioNet,” *Circulation*, vol. 101, no. 23, Jun. 2000.
- [17] G. L. Sumner, V. P. Kuriachan, and L. B. Mitchell, “Sudden Cardiac Death,” *Encycl. Cardiovasc. Res. Med.*, pp. 511–520, Jan. 2018.
- [18] S. G. Priori *et al.*, “2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death The Task Force for the Management of Patients with Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death of the European Society of Cardiology (ESC) Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC),” *Eur. Heart J.*, vol. 36, no. 41, pp. 2793–2867, Nov. 2015.
- [19] R. D. Anderson *et al.*, “Catheter Ablation of Ventricular Fibrillation,” *Hear. Lung Circ.*, vol. 28, no. 1, pp. 110–122, Jan. 2019.
- [20] V. Jahmunah *et al.*, “Computer-aided diagnosis of congestive heart failure using ECG signals – A review,” *Phys. Medica*, vol. 62, no. March 2020, pp. 95–104, 2019.
- [21] “Dr. Smith’s ECG Blog: CHF Exacerbation with Old LBBB: Is There New

- Infarction or Not?” [Online]. Available: <http://hqmeded-ecg.blogspot.com/2013/11/chf-exacerbation-with-old-lbbb-is-there.html>. [Accessed: 18-Jul-2022].
- [22] R. K. Tripathy, A. Zamora-Mendez, J. A. de la O Serna, M. R. A. Paternina, J. G. Arrieta, and G. R. Naik, “Detection of life threatening ventricular arrhythmia using digital taylor fourier transform,” *Front. Physiol.*, vol. 9, no. JUN, p. 722, Jun. 2018.
- [23] M. G. Kiuchi *et al.*, “New Approaches in the Management of Sudden Cardiac Death in Patients with Heart Failure—Targeting the Sympathetic Nervous System,” *Int. J. Mol. Sci. 2019, Vol. 20, Page 2430*, vol. 20, no. 10, p. 2430, May 2019.
- [24] F. Sessa *et al.*, “Heart rate variability as predictive factor for sudden cardiac death,” *Aging (Albany. NY)*, vol. 10, no. 2, pp. 166–177, Feb. 2018.
- [25] M. Shi *et al.*, “Early Detection of Sudden Cardiac Death by Using Ensemble Empirical Mode Decomposition-Based Entropy and Classical Linear Features From Heart Rate Variability Signals,” *Front. Physiol.*, vol. 11, p. 118, Feb. 2020.
- [26] C. Fernández Biscay, P. D. Arini, A. I. Rincón Soler, and M. P. Bonomini, “Classification of ischemic and non-ischemic cardiac events in Holter recordings based on the continuous wavelet transform,” *Med. Biol. Eng. Comput.*, vol. 58, no. 5, pp. 1069–1078, May 2020.
- [27] G. Altan, N. Allahverdi, and Y. Kutlu, “Diagnosis of Coronary Artery Disease Using Deep Belief Networks Publication Info,” *Turkey) EJENS*, vol. 2, no. 1, pp. 29–36, 2017.
- [28] A. Davari Dolatabadi, S. E. Z. Khadem, and B. M. Asl, “Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM,” *Comput. Methods Programs Biomed.*, vol. 138, pp. 117–126, Jan. 2017.
- [29] R. R. Sharma, A. Kumar, R. B. Pachori, and U. R. Acharya, “Accurate automated detection of congestive heart failure using eigenvalue decomposition based features extracted from HRV signals,” *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 312–327, Apr. 2019.
- [30] R. K. Tripathy, M. R. A. Paternina, J. G. Arrieta, A. Zamora-Méndez, and G. R. Naik, “Automated detection of congestive heart failure from electrocardiogram signal using Stockwell transform and hybrid classification

- scheme,” *Comput. Methods Programs Biomed.*, vol. 173, pp. 53–65, May 2019.
- [31] A. A. Bhurane, M. Sharma, R. San-Tan, and U. R. Acharya, “An efficient detection of congestive heart failure using frequency localized filter banks for the diagnosis with ECG signals,” *Cogn. Syst. Res.*, vol. 55, pp. 82–94, Jun. 2019.
- [32] by Scott David Greenwald BSE and S. David Greenwald, “The development and analysis of a ventricular fibrillation detector,” 1986.
- [33] F. Jager *et al.*, “Long-term ST database: A reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia,” *Med. Biol. Eng. Comput.*, vol. 41, no. 2, pp. 172–182, 2003.
- [34] D. S. Baim *et al.*, “Survival of patients with severe congestive heart failure treated with oral milrinone,” *J. Am. Coll. Cardiol.*, vol. 7, no. 3, pp. 661–670, 1986.
- [35] B. A. Bleyer and I. M. Salla, “INTERPOLATION BY CUBIC SPLINES.”
- [36] G. Cappiello *et al.*, “A statistical index for early diagnosis of ventricular arrhythmia from the trend analysis of ECG phase-portraits,” *Physiol. Meas.*, vol. 36, no. 1, p. 107, Dec. 2014.
- [37] P. S. Hamilton and W. J. Tompkins, “Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database,” *IEEE Trans. Biomed. Eng.*, vol. BME-33, no. 12, pp. 1157–1165, 1986.
- [38] E. Ebrahimzadeh *et al.*, “An optimal strategy for prediction of sudden cardiac death through a pioneering feature-selection approach from HRV signal,” *Comput. Methods Programs Biomed.*, vol. 169, pp. 19–36, Feb. 2019.
- [39] T. Pham, Z. J. Lau, S. H. A. Chen, and D. Makowski, “Heart rate variability in psychology: A review of hrv indices and an analysis tutorial,” *Sensors*, vol. 21, no. 12, pp. 1–20, 2021.
- [40] M. D. Costa, R. B. Davis, and A. L. Goldberger, “Heart rate fragmentation: A new approach to the analysis of cardiac interbeat interval dynamics,” *Front. Physiol.*, vol. 8, no. MAY, p. 255, May 2017.
- [41] C. Yan, P. Li, L. Ji, L. Yao, C. Karmakar, and C. Liu, “Area asymmetry of heart rate variability signal,” *Biomed. Eng. Online*, vol. 16, no. 1, pp. 1–15, 2017.



- [42] M. Sabeti, S. Katebi, and R. Boostani, "Entropy and complexity measures for EEG signal classification of schizophrenic and control participants," *Artif. Intell. Med.*, vol. 47, no. 3, pp. 263–274, Nov. 2009.
- [43] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656.
- [44] A. Ishikawa and H. Mieno, "The fuzzy entropy concept and its application," *Fuzzy Sets Syst.*, vol. 2, no. 2, pp. 113–123, Apr. 1979.
- [45] M. Y. Boon, B. I. Henry, C. M. Suttle, and S. J. Dain, "The correlation dimension: A useful objective measure of the transient visual evoked potential?," *J. Vis.*, vol. 8, no. 1, pp. 6–6, Jan. 2008.
- [46] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Phys. D Nonlinear Phenom.*, vol. 31, no. 2, pp. 277–283, Jun. 1988.
- [47] M. J. Katz, "Fractals and the analysis of waveforms," *Comput. Biol. Med.*, vol. 18, no. 3, pp. 145–156, Jan. 1988.
- [48] A. Lempel and J. Ziv, "On the Complexity of Finite Sequences," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 75–81, 1976.
- [49] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, Sep. 1970.
- [50] B. Yu, D. Z. Pan, T. Matsunawa, and X. Zeng, "Machine learning and pattern matching in physical design," *20th Asia South Pacific Des. Autom. Conf. ASP-DAC 2015*, pp. 286–293, Mar. 2015.
- [51] P. Gomes, H. Silva, and P. Margaritoff, "pyHRV - Open-Source Python Toolbox for Heart Rate Variability," *Proc. Int'l Conf. Electr. Electron. Comput. Eng.*, pp. 822–828, 2018.