Automated Cardiac Rhythm Diagnosis for Electrophysiological Studies, an Enhanced Classifier Approach

Julian Bostock

Submitted towards the degree of Doctor of Philosophy

at

City University London

Centre for Health Informatics

Supervisor: Dr Peter Weller

August 2013

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

List of Tables

List of Tables (continued)

List of Illustrations

List of Illustrations (continued)

Acknowledgements

Heartfelt thanks to my wife Christendai for her love, inspiration, support, tolerance and patience and my wonderful daughters for their encouragement.

A thousand thanks go to Dr Peter Weller for his guidance, instruction, encouragement, praise and very subtle suggestions - ever the diplomat!

Thank you to my clinical colleagues for their help and assistance, particularly to Dr Michael Cooklin who provided clinical supervision and acted as the primary domain expert throughout the research as well as providing access to his patients. Special thanks to Prof. Eric Rosenthal for assistance in the catheter laboratory data collection planning and provision of patients and to Prof. Jas Gill, Dr Aldo Rinaldi and Dr Mark O'Neill for their kind support and encouragement and for providing patients to enrol in the clinical study.

I offer humble thanks to Cardiovascular Services for sponsorship of my research and particularly all my colleagues in the Cardiac Physiology department at Guy's and St. Thomas' for their kind words, help and assistance, particularly with my timetable.

To all my valiant circle, I thank you for supporting me throughout my ups (many) and downs (few), to this culmination of my research.

Declaration

I grant powers of discretion to the University librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

INTRODUCTION

Heart function can be impaired by rhythm disturbances (cardiac arrhythmia), illustrated by electrocardiogram (ECG) recordings. Computerised arrhythmia diagnosis is well established for ECG's but less for intracardiac electrophysiological (EP) testing. Accurate diagnosis is pre-requisite for delivering appropriate treatment to patients however existing algorithms misdiagnose a proportion of arrhythmias. Studies suggested artificial intelligence (AI) classifiers are accurate using ECG and intracardiac electrogram features and reviews suggested new features might augment diagnosis. This study aimed to develop an accurate cardiac rhythm diagnostic algorithm for electrophysiological (EP) studies with potential application as a generic rhythm classifier.

METHOD

An ethically approved prospective clinical study collected clinical history, right atrial and right ventricular intracardiac electrograms, beat-to-beat cardiac stroke volume, body motion and body temperature data during EP studies. An iterative system development life-cycle was used, including knowledge management and classifier development sub-processes. Domain expert knowledge and clinical arrhythmia diagnosis were modelled, synthesised as AI classifiers and used to classify cardiac rhythms.

RESULTS

Data collected from 65 patients was pre-processed into instances for classifier inputs. Decision tree, naïve Bayes, neural network, support vector machine and inference engine classifiers developed using Matlab showed good performance and were combined as a production system in a mixture-of-experts multi-classifier system. 18 different rhythms were classified, with the naïve Bayes classifier used to classify 11 rhythms, decision tree 4 rhythms, neural network and support vector machine one each, unclassified instances by the inference engine classifier and final class allocation using decision rule. Production system showed overall correct clasification rate 0.960; error 0.040; mean sensitivity 0.855; mean specificity 0.977; mean κ 0.767; mean positive predictive value 0.792; mean negative predictive value 0.975; mean Pearson's phi 0.787, with $P < 0.004$ (equivalent to $P = 0.05$ for 18 way Bonferroni comparison) supporting no difference with the gold standard. Correct classification, sensitivity, specificity, Cohen's kappa and positive predictive value showed values of 1.0 for inappropriate sinus tachycardia, focal atrial tachycardia and ventricular tachycardia and > 0.9 for sinus node dysfunction and atrio-ventricular nodal/ junctional tachycardias. Temperature, accelerometry and QT interval were assessed as features by a comparison of algorithm performances with each feature removed and found not to affect classification performance. An evaluation showed 10 beat analysis performed better than 5 beat analysis.

CONCLUSIONS

Modelling of the clinical diagnosis process produced an AI based mixture-of-experts multi-classifier system, which accurately diagnosed different 18 cardiac rhythms. The naïve Bayes classifier performed best and classified 11 rhythms. Features for clinical symptoms and predisposing factors, atrial electrogram morphology and changes in stroke volume were found to influence rhythm classification. High performances encourage further development and potential future improvements include: a larger sample dataset; inclusion of His and coronary sinus electrograms; data mining for unknown features with significant influence on diagnosis; binary classification. The aim to classify rhythm using artificial intelligence suitable for use during EP studies was satisfied and the research hypothesis that it outperformed current algorithms was accepted. The system was likely to be able to accept updates but needs conversion as a precursor to use in a live clinical environment.

Abbreviations and Acronyms

1HB	First degree atrio-ventricular block
2HB	Second degree atrio-ventricular block (Mobitz type I)
2HB2	Second degree atrio-ventricular block (Mobitz type II)
α	Type I error
β	Type II error
A	Intracardiac wave equivalent to the P wave component of the ECG
AA	The interval between successive A waves
AAVRT	Antidromic atrio-ventricular reciprocating tachycardia
ADC	Analogue-to-digital converter
AED	Automatic external defibrillator
AI	Artificial intelligence
AF	Atrial fibrillation
AIRS	Artificial immune recognition system
ANN	Artificial neural network
AP	Accessory AV connection
ARB	Artificial recognition ball
AT	Focal atrial tachycardia or atrial tachycardia
AUC	Area under the (ROC) curve
AV	Atrio-ventricular
AV ratio	The ratio between AA and VV intervals
AVNRT	Atrio-ventricular nodal reciprocating tachycardia
B point	onset of rapid upstroke of *dZ/dt* preceding C wave
Bagging	Bootstrap aggregation
BX	The interval between B and X points of *dZ/dt*
c′	Citations per paper per year since publication
C	Citations per paper
C wave	Peak of *dZ/dt* following R wave
CCR	Correct classification rate (Accuracy index)
CHB	Complete atrio-ventricular block
CI	Confidence interval
$Class_{NB}$	Class label (naïve Bayes classifier)
$Class_{DT}$	Class label (decision tree classifier)
$Class_{SVM}$	Class label (support vector machine classifier)
$Class_{NN}$	Class label (neural network classifier)
$Class_{IE}$	Class label (inference engine classifier)
CPU	Central processing unit

Abbreviations and Acronyms (continued)

CT	Computerised tomography
ΔZ	Thoracic impedance change with time (full notation $\Delta Z(t)$)
ΔZ_R	Respiratory component of thoracic impedance change
ΔZ_C	Cardiac component of thoracic impedance change
dis	Disagreement measure
dP/dt	First time derivative of change in pressure
dZ/dt	First time derivative of change in thoracic impedance
DF	Double-fault measure
DSP	Digital signal processing
DT	Decision tree
ECG	Electrocardiogram
EGM	Intracardiac electrogram
EP	Electrophysiological testing
f_{in}	Sample rate
f_{out}	Desired sample rate
F_1	F_1 score, a measure of test accuracy
FFBP	Feed-forward, back-propagation
FFT	Fast Fourier transform
FIR	Finite impulse response
FIS	Fuzzy inference system
FP	False positive
FN	False negative
FCNN	Fuzzy clustering-neural network
FES	Fuzzy expert system
FJT	Focal junctional tachycardia
g_0	Acceleration due to gravity
GB	Gigabyte of digital information
GD	Generalised diversity
GDA	Generalised discriminant analysis
GEE	Generalised estimating equation
GSTT	Guy's and St. Thomas' NHS Foundation Trust
H	Hypothesis
H_0	Null hypothesis
H_1	Research hypothesis
HRS	Heart Rhythm Society
HRT	Heart-rate turbulence

Abbreviations and Acronyms (continued)

HRV	Heart rate variability
I, II, III	Standard (Einthoven) ECG lead I, II or III
ICD	Implantable cardioverter-defibrillator
ICG	Impedance cardiography
ID	Identification number
IE	Inference engine
IEEE	Institute of Electrical and Electronics Engineers
IF	Impact factor
IFVT	Idiopathic fascicular ventricular tachycardia
ILR	Implantable loop recorder
IST	Inappropriate sinus tachycardia
IVC	Inferior vena cava
κ	Cohen's kappa; inter-rater agreement
kB	Kilobyte of digital information
k-NN	*k*-nearest neighbour
L	Up-sampling factor
LDA	Linear discriminant analysis
LR+	Positive likelihood ratio
LR-	Negative likelihood ratio
LVET	Left ventricular ejection time
MAR	Missing at random
MAT	Multifocal atrial tachycardia
MB	Megabyte of digital information
MCAR	Missing completely at random
MNAR	Missing not at random
MCS	Multi-classifier system
ME	Mixture-of-experts
MeSH	Medical subject headings
MRAT	Macro-re-entrant atrial tachycardia
MRI	Magnetic resonance imaging
MV	Minute ventilation
MVI	Minute ventilation index
MWI	Moving window integration
N	Number of samples
NaN	Not a number
NB	Naïve Bayes

Abbreviations and Acronyms (continued)

NHS	National Health Service
NHS REC	National Health Service Research Ethics Committee
NN	Neural network
NPJT	Non-paroxysmal junctional tachycardia
NPSA	National Patient Safety Agency
NSR	Normal sinus rhythm
OAVRT	Orthodromic atrio-ventricular reciprocating tachycardia
OR	Odds ratio
OTVT	Outflow tract ventricular tachycardia
OVA	One-versus-all
OVO	One-versus-one
ϕ	Pearson's phi
P	Probability
P_{crit}	Critical probability value for significance testing (such as 0.05)
P	The P wave component of the ECG
PAC	Premature atrial contraction
PCA	Principal component analysis
PDA	Personal digital assistant
PJRT	Permanent junctional reciprocating tachycardia
POTS	Postural orthostatic tachycardia syndrome
PP	The interval between successive P wave components of the ECG
PPV	Positive predictive value
PR	The interval between successive P wave and R wave components of the ECG
P:R	Arithmetic ratio between P and R waves
PVC	Premature ventricular contraction
PVT	Polymorphic ventricular tachycardia
PWM	Pulse width modulation
Q	Yule's Q
QRS	The QRS complex, a component of the ECG
QRST	The QRS and T wave complex, a component of the ECG
QT_c	The corrected QT interval (using Bazett's formula)
r	Pearson's product-moment correlation coefficient
R	The R wave component of the QRS complex
RA	Right atrium (atrial)
RBF	Radial basis functions
RCT	Randomised clinical trial

Abbreviations and Acronyms (continued)

RFA	Radiofrequency ablation
ROC	Receiver operating characteristic
RR	Respiratory rate
RR	The interval between successive R wave components of the ECG
RSA	Respiratory sinus arrhythmia
RSS	Square root of the sum of squares
RV	Right ventricle (ventricular)
RVPP	RV pulse pressure
S1	Unipolar atrial EGM
S2	Composite atrio-ventricular EGM
S3	Unipolar ventricular EGM
SA	Sinus arrest
SAB	Sino-atrial block
SB	Sinus bradycardia
SCD	Sudden Cardiac Death
SDLC	Systems development life-cycle
SDNN	Standard deviation of mean RR interval between normal heart beats
Se	Sensitivity
SICD	Subcutaneous ICD
SIDS	Sudden infant death syndrome
SNRT	Sinus node re-entry tachycardia
Sp	Specificity
ST segment	A component of the ECG waveform, the ST segment
ST	Physiological sinus tachycardia
SV	Cardiac stroke volume
SVI	Stroke volume index
SVM	Support vector machine
SVT	Supraventricular tachycardia
SVTab	SVT with aberration
t	Time
τ	Statistical power
T	The T wave component of the QRST complex
TFIT	Thoracic flow inversion time
TP	True positive
TN	True negative
TV	Tidal volume

Abbreviations and Acronyms (continued)

μ	Measure of difficulty
U	The U wave component of the QRSTU complex
V	Intracardiac wave corresponding to a QRS complex component of the ECG
VTC	Vector timing and correlation (QRS morphology)
VTCA	Vector timing and correlation in the atrium (P wave morphology)
VF	Ventricular fibrillation
VL	Voltage at the left arm
VLSI	Very-large-scale integration
VMU	Vector magnitude units
VT	Monomorphic ventricular tachycardia or ventricular tachycardia
VV	The interval between successive V waves
WCT	Wilson's Central Terminal
X point	Nadir of *dZ/dt* wave following C wave
Z	Absolute impedance
Z_0	Base thoracic impedance

Chapter 1 Introduction

1.1 Overview

The concepts of heart rhythms and their diagnosis will be introduced with the disciplines of cardiac rhythm diagnosis, using clinical examination, non-invasive and invasive investigations as well as treatment options summarised. Currently available technologies for automated arrhythmia analysis will be reviewed and the problem of misdiagnosis, motivation for the study, will be outlined.

1.2 Background

The heart is a sophisticated muscular pump, designed to efficiently deliver oxygen-rich blood to the tissues and organs of the body. This function can become impaired due to heart disease, including disturbances of the heart rhythm (cardiac arrhythmia). A normal heart rhythm is a regular, coordinated sequence of heart beats with each beat initiated by a wave of electrical activation of heart muscle cells, starting in the atria and travelling to the ventricles, called depolarisation. A similar sequence of mechanical contraction follows electrical activation, providing the heart's pumping action. Arrhythmia is an electrical disturbance, followed by its mechanical consequence, which upsets this coordinated sequence.

Arrhythmia is among the top ten reasons for hospital admission, can be debilitating and may require high cost treatments, emphasising the importance of timely and accurate diagnosis (Stewart *et al*. 2004). There are a number of different arrhythmias; some consisting of a single abnormal heart beat, others are more sustained and some provoke serious symptoms. There are two main types of sustained arrhythmia, slow heart beats or bradycardia and rapid heartbeats or tachycardia, both of which can be further subdivided into more specific rhythms.

Bradycardia and tachycardia can sometimes result in haemodynamic compromise (circulatory impairment) and hypoxia, a reduced oxygen supply to tissues and organs. Depending on the severity of these effects, some arrhythmias are considered life-threatening and require prompt treatment, for example by a pacemaker or a defibrillator. Ventricular fibrillation (VF) is considered the most serious arrhythmia and is almost invariably fatal unless corrected immediately by an electric shock treatment called defibrillation. All cardiac arrhythmias require accurate diagnosis to guide the choice of treatment.

A diagnosis of arrhythmia is made clinically, supported by evidence from tests to improve diagnostic granularity and accuracy. Several tests can be used to aid differential diagnosis, including the electrocardiogram (ECG) and invasive electrophysiological (EP) testing.

Automated computerised rhythm diagnosis is widely used in ECG's and implantable cardioverter defibrillators (ICD's) and less in EP studies. This study will focus on computerised rhythm diagnosis of intracardiac electrograms, as used in EP testing and in ICD's, seeking to maximise diagnostic accuracy to a uniformly high standard, working towards reducing need for domain expert interpretation, offering improved and clinically valuable differential diagnosis.

1.3 Conventional Cardiac Arrhythmia Diagnosis and Treatment

Physicians assess patients for suspected arrhythmia using well-validated clinical diagnostic techniques and a range of tools, with the ECG playing a central role. Patients with arrhythmia may present with symptoms such as: palpitation; syncope (blackouts); light-headedness; chest pain; breathlessness or fluid retention. Symptom severity, quality and characteristics are important. They include regularity or irregularity of palpitations; whether palpitations are very fast and how symptoms start and stop (National Institutes of Health 2011).

A medical history is collected from the patient, with particular attention to factors known to predispose towards arrhythmia (see Appendix A) (Brignole *et al.* 2001 and 2004; Zipes *et al.* 2006) and effort is made to discover circumstances that might have triggered an arrhythmia.

Arrhythmia may be the result of an identifiable cause, such as: structural heart disease; medication reaction; stress; caffeine intake; electrolyte imbalance; electrical abnormality (channelopathies); mechanical or electrical stimulation, such as chest trauma or electrocution. A physical examination is conducted, including: evaluation of heart sounds; assessment of arterial and venous pulses; measurement of resting blood pressure and an ECG recording and sometimes further diagnostic tests, all of which aid the differential diagnosis of cardiac arrhythmia and formulation of a therapeutic strategy (Zimetbaum & Josephson 1998). If an arrhythmia has been detected, it is likely to be represented using an ECG trace (see Fig. 1.1).

Diagnosis of arrhythmia is a process of clinical evaluation that includes non-invasive testing, such as electrocardiogram (ECG), augmented by echocardiogram and in some cases, invasive electrophysiological (EP) testing.

1.3.1 Non-Invasive Arrhythmia Diagnosis

Non-invasive tests involve no intrusion into the body, usually only requiring application of external skin electrodes, a probe or use of a scanner. The ECG is the classic non-invasive diagnostic test for arrhythmia and is used in various forms, such as: a resting 12-lead ECG; a treadmill or bicycle exercise stress electrocardiogram; an ambulatory electrocardiogram; an ECG loop recorder; a signal-averaged electrocardiogram; tilt-table testing and vectorcardiography. Arrhythmias may be assessed for clinical severity or an underlying structural cause by other non-

invasive examinations, such as: ultrasonic examination of the heart (echocardiography); cardiac computerised tomography (CT); nuclear scan and magnetic resonance imaging (MRI).

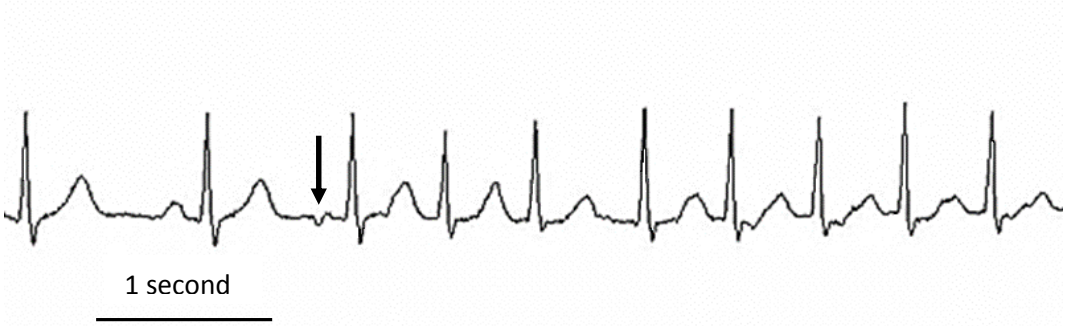

Figure 1.1 An ECG showing onset of an arrhythmia (atrial fibrillation). This ECG tracing, read from left to right, shows two normal heartbeats, followed by an abnormal premature wave (arrowed) then by abnormally rapid, irregular heartbeats. Signal amplitudes are not calibrated in this diagram, as the focus is upon timing changes.

Where non-invasive testing fails to provide a differential arrhythmia diagnosis and it is prognostically important, invasive electrophysiological testing (EP) may be indicated.

1.3.2 Invasive Arrhythmia Diagnosis

Invasive diagnostic techniques involve intrusion into the body, are less well-tolerated, involve discomfort and risks of complication. Typically, in an EP study, four thin multi-polar electrode catheters are passed through veins, using a percutaneous femoral venous approach, under X-Ray guidance, into standard positions within the heart, according to the technique described by Josephson (2002). In a typical EP study, quadripolar electrodes with 5mm inter-electrode spacings are placed in standard positions of: the high right atrium (RA) near to the sino-atrial node or right atrial appendage; a position adjacent to the bundle of His (His) and at the right ventricular apex (RV) with a decapolar electrode with alternating 2mm and 5mm spacings placed within the coronary sinus (CS) and advanced towards the left cardiac border (see Fig. 1.2).

Following positioning of the electrodes, simultaneous continuous recordings are taken of intracardiac electrogram signals from adjacent electrode pairs (see Fig. 1.3). Changes of timing relationships, observed as pattern variations, during pacemaker stimulation and during instances of arrhythmia are noted, to provide insight into the underlying mechanisms of an arrhythmia, leading to a differential diagnosis.

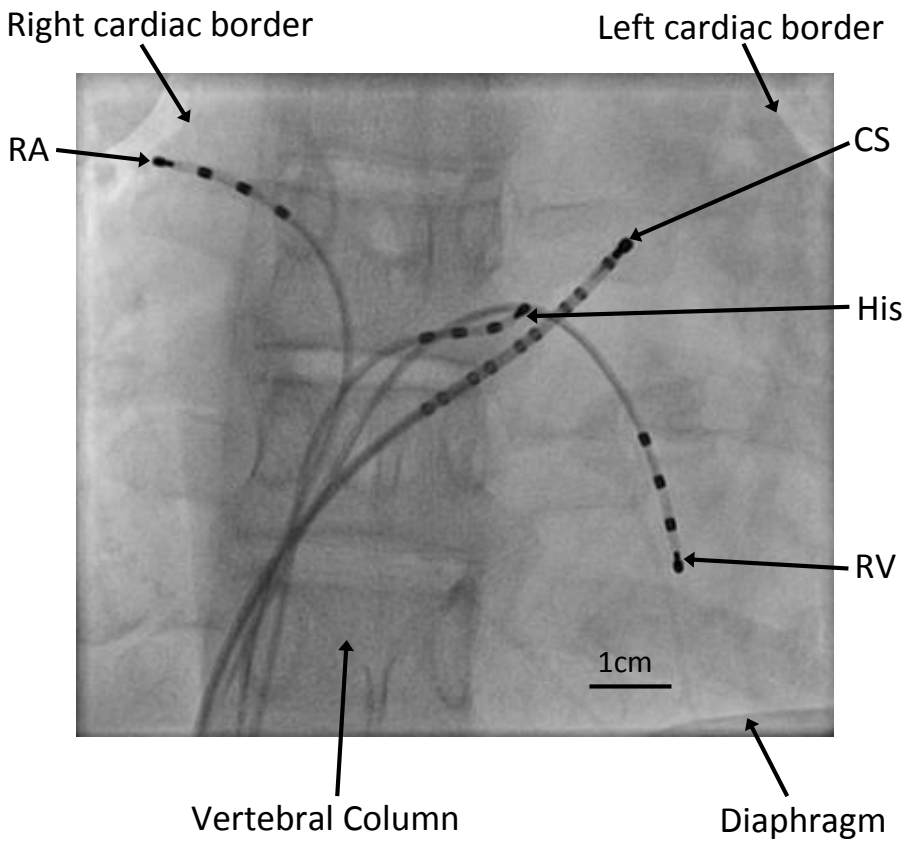

Figure 1.2 X-ray image of electrode catheters in the heart during an EP study. Electrodes in the RA (right atrium), His (bundle of His), RV (right ventricle) and CS (coronary sinus). Patient 11; PA view; Normal cardiac structural anatomy, history of arrhythmia only.

If a suspected arrhythmia is rare, or exhaustive testing during an EP study is inconclusive, an implantable loop recorder (ILR) may be used (Seidl *et al.* 2000). An ILR is the semi-permanent implant of a small ECG recorder placed beneath the skin of the chest wall.

1.3.3 Treatments for Cardiac Arrhythmia

Treatments are tailored, based on the arrhythmia type, symptom frequency and severity, tolerance to drugs and prognoses attached to therapy options. Treatments include non-interventional options such as: lifestyle adjustment; dietary modification and drugs and interventional treatments such as: cardioversion; the implant of electronic devices, such as pacemakers or cardioverter-defibrillators (ICDs); ablation or surgery (Heart Rhythm Society n.d.).

1.3.4 Implantable Pacemakers and Cardioverter Defibrillators (ICDs)

For patients with symptomatic bradycardia, or other special conditions, pacemaker treatment may

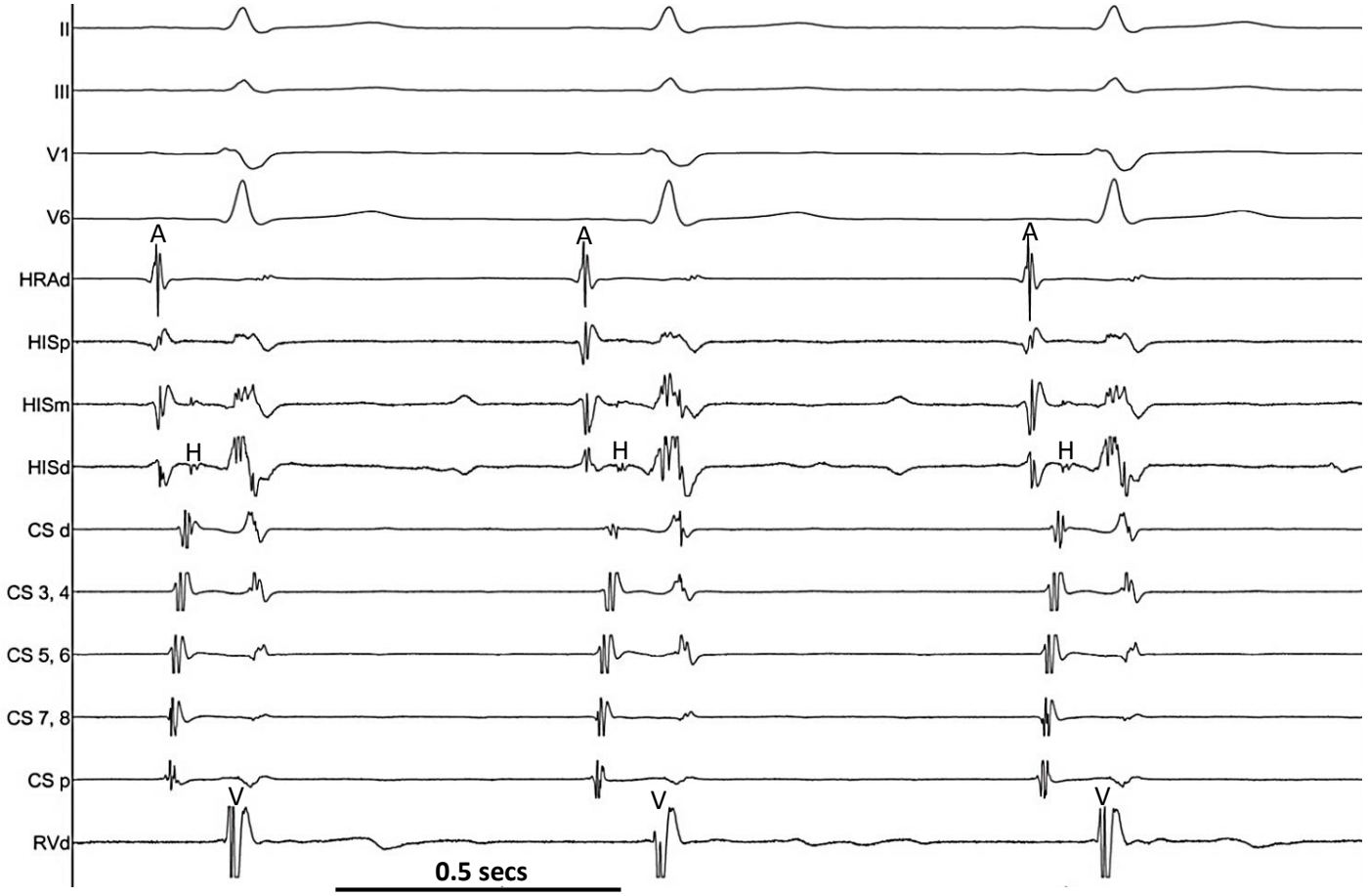

Figure 1.3 Intracardiac electrograms recorded during an EP study. Key (top first): Surface ECG II, III, V1 and V6; HRAd - right atrial electrogram (A); HISp, HISm and HISd - His bundle electrograms (H); CSd to CSp - coronary sinus electrograms; RVd - right ventricular electrogram (V). Archive recording; Normal cardiac anatomy, no non-arrhythmia history; Signal amplitudes uncalibrated as the focus is timing and sequence.

be chosen. Pacemakers have electrodes connected to the heart, usually in the right atrium and right ventricle, in similar positions to those of an EP study. More sophisticated models have an additional electrode within the coronary venous system, allowing connection with the left ventricle

Pacemakers are able to detect cardiac activity and, deliver electrical pulses to stimulate heart beats, automatically maintaining the heart rate and improving cardiac output. Implantable defibrillators incorporate all the capabilities of pacemakers with additional functionality of rapid ventricular stimulation or delivery of high voltage electric shocks, to correct ventricular tachycardia (VT) or ventricular fibrillation (VF) respectively.

1.3.5 Ablation

Cardiac ablation involves the localisation then destruction of very small areas of abnormal heart tissue responsible for a rhythm disturbance. Once an arrhythmia diagnosis is made at an EP study, it is usual to proceed directly to treatment with ablation, if that is indicated. Ablation may also be a stand-alone procedure for certain indications, such as macro-re-entrant atrial tachycardia, paroxysmal or persistent atrial fibrillation. At EP study, arrhythmia localisation is achieved by a process of "mapping" the inner surface of the heart or endocardium, first approximately, using the four standard electrode positions, then more precisely with an adjustable electrode which is moved into different positions. Electrogram timing characteristics at the tip of this adjustable electrode allow the electrophysiologist determine the area within which the abnormality is located. Ablation is then performed by delivery of energy to the localised site of the abnormality, using the same adjustable electrode, typically with radiofrequency electrical energy or less commonly, other energy sources such as cryothermal, microwave or laser may be used.

1.3.6 Arrhythmia Surgery

Occasionally, an arrhythmia can be treated with surgery, most commonly when cardiac surgery is already being performed for another reason, for example a heart valve replacement. The most frequently performed of these treatments is the maze procedure for atrial fibrillation, where a surgeon either ablates with a hand-held probe or places cuts on the left and right atria, provoking scarring which creates electrical barriers to prevent spread of the disorganised atrial depolarisation characteristic of atrial fibrillation (Cox *et al.* 1991).

1.4 Computerised Arrhythmia Diagnosis

The state of current research into computerised cardiac rhythm diagnosis will be summarised, focussing on ECG diagnosis, EP studies, ICD algorithms, bench studies and comparative studies.

1.4.1 Rhythm Classification in ECG Interpretation

The most widely used form of medical computer decision support is computerised analysis of the resting ECG (Tsai *et al*. 2003), where ECG recordings and their automated diagnosis are made available to clinicians who then are able to edit its interpretation, guiding teatment.

1.4.2 Rhythm Classification in Electrophysiological Studies

Rhythm diagnosis made during invasive electrophysiological testing and ablation procedures is non-automated, involving interpretation of signals by a domain expert using a systematic manual process to reach a conclusion. There are no notable studies in the area of computerised diagnosis during EP studies with which to compare algorithm accuracy. Development of automated rhythm diagnosis for electrophysiological studies requires maintenance of high diagnostic accuracy, comparable to that of a domain expert - an experienced cardiac electrophysiology consultant, while reducing the need for the presence of that domain expert.

The extent of computerised analysis during electrophysiological studies is restricted to specialised applications, such as 3D anatomical navigation systems used to guide therapeutic ablation procedures and 3D mapping to precisely locate intracardiac pathways, in both cases where a rhythm diagnosis has previously been made.

1.4.3 Rhythm Classification in ICDs

Implants of ICDs and the availability of automatic external defibrillators (AED's), such as seen in public buildings, are now ubiquitous and their automated rhythm diagnostic capability may be considered among the most advanced available and may serve as a model for comparative purposes. Real-time algorithms, such as those in cardiac implantable devices require rapid, accurate diagnosis with no short-term possibility of alteration. ICD algorithms are considered among the most sophisticated of this type of unsupervised classification algorithm.

The diagnosis of ventricular tachycardia (VT) and ventricular fibrillation (VF) in ICDs depends on sophisticated diagnostic capabilities, enabling not only very slow or very rapid rhythms to be distinguished but also differentiation and classification of rhythms, to deliver the most appropriate therapy (or none) from a range of options. Most ICD's employ binary algorithms, designed to make "treatment / no-treatment" decisions, rather than a formal rhythm diagnosis.

The most diagnostically accurate ICD algorithm was found to have an error of 0.053, sensitivity of 1.000, specificity 0.898 and κ of 0.895, with $P < 0.001$ for differential diagnosis of VT and VF from supraventricular arrhythmia (Gold *et al.* 2012b). Current ICD algorithms have high sensitivities varying from of 98.7% to 100% but much more varied specificities of between 65%

and 95%, varying by manufacturer and model for the detection of life-threatening cardiac arrhythmias (Aliot *et al.* 2004).

ICDs classify atrial and ventricular rhythms by placing them into heart rate "zones", according to what are believed to be appropriate therapeutic options (see subsection 1.3.3). Classifications are often not clinically accurate but, in practice, the delivery of appropriate ATP or shock therapy is generally good. Studies of ICD algorithm performances tend to focus on the differential diagnosis only as it affects treatment, rather than the actual rhythm diagnosis. This differentiation is usually between supraventricular tachycardia, a rapid heart rhythm where no treatment is indicated and ventricular tachycardia, where treatment is indicated (see Appendix E). ICD diagnosis of other rhythms is considered of secondary importance. This is well illustrated in the START study (Gold et al., 2012), where diagnosis of atrial tachyarrhythmia is described only in terms of its effect on the specificity of ventricular arrhythmia diagnosis and thereby its importance "to inhibit therapy". Recent improvements in the accuracy of ICD diagnostic algorithms have centred on digital signal processing, such as filtering rectification and noise rejection of the detected intracardiac electrograms.

1.4.4 Rhythm Diagnosis Algorithms – Bench Studies

The usefulness of classifier algorithms and artificial intelligence (AI) classifiers in the field of rhythm analysis has been demonstrated in small studies (see Chapter 2, subsections 2.8.2 to 2.8.12). AI has so far failed to be widely adopted for this application. Apparent poor physician acceptance of "black box" computer decision making suggests adoption of AI in cardiac implantable devices is against the trend (Kalogeropoulos *et al.* 2003; Erickson & Bartholmai 2002). Factors contributing to this include processing power and time-delays in "greedy" AI processes, such as training neural networks.

1.4.5 Computational Power and Automated Rhythm Diagnosis

Limitations on the computational power of cardiac rhythm classification systems are largely related to size and power requirements.

ECG recorders are usually portable, to allow bedside use in a clinic or hospital environment. There is an increasing availability of lightweight options, including notebook and tablet computer applications which require only a very small interface box with necessary patient cable connections to acquire the ECG. Processing and interpretation is limited only by the constraints of the computing power and power requirements of the device used, which vary considerably.

In ICD's and pacemakers, due to the limitations of size on implanted devices, information gathering is currently at the expense of capability or vice versa. There are competing influences

at work: clinicians wish to include as much technology and capability as possible into an implanted device and maximise the longevity of the device, reducing the need for replacement. These tend to increase circuit complexity, power requirements and increase device size. Patients desire an effective treatment with a cosmetically acceptable result which minimises complications, tending to reduce device size. The balance of these is addressed by manufacturers in collaboration with their customers. Future devices of more compact size and higher processing power may be possible with developments in battery chemistry, reduced power consumption and miniaturisation of circuitry, helping to solve these limitations.

Fewer limitations are applicable to EP studies and to bench studies of rhythm diagnostic algorithms. EP studies are performed using sophisticated PC-based systems which are directly fed with signals from a patient for on-line analysis, currently performed manually. These systems are usually purchased and upgraded periodically, limiting technological capability to that available at the time of purchase and available to the particular manufacturer. Bench studies may be performed on any suitable computer systems available to the researcher, with no size limitation at all, limited only by the availability of suitable technology and the budget of the researcher.

By the nature of these limitations one may envisage a loose process, with bench studies being more aspirational in nature, computational power and power requirements limited only by the budget of the researcher. Following the refinement and improvement processes required to bring a practical solution to market one would anticipate reduced need for computational power and restricted power requirements allowing future inclusion in dedicated systems, such as those used in EP testing, then ultimate inclusion into the capability of implanted devices. One pathway that has been explored is the development of system-on-a-chip, where large scale integration is used to microminiaturise a system into a single chip, allowing integration into implanted devices of the future (Leong & Jabri 1995).

1.4.6 Misdiagnosis in Automatic Rhythm Diagnosis

Studies of the accuracy of ECG interpretive algorithms (Willems *et al.* 1991; Salerno *et al*. 2003a) have found that in some instances expert cardiologists were susceptible to error, compared with computerised diagnosis, as well as the reverse. Willems *et al.* (1991) concluded that computer analysis could be used to support uniform high diagnostic standards. No major studies focused on computerised ECG rhythm diagnosis but the review of Salerno et al. (2003b) suggested evidence to support the position that common ECG interpretation algorithms performed poorly for arrhythmias. Studies by Shah *et al.* (2007) and Poon *et al.* (2005) both examined errors in computerised ECG rhythm diagnosis and found

Until recently the incorrect diagnosis from implantable defibrillator (ICD) algorithms approached

20% (Nunain *et al.* 1995; Rinaldi *et al.* 2004; Sacher *et al.* 2006). In the MADIT II trial, 11.5% of ICD patients received inappropriate shock episodes, accounting for 31.2% of all shocks. In MADIT II, incorrect diagnoses were due to: atrial fibrillation (44%); supraventricular tachycardia (36%) and abnormal sensing (20%). They also found that patients with inappropriate shocks had a greater likelihood of all-cause mortality (P = 0.025) (Daubert *et al.* 2008). The delivery of inappropriate therapies causes discomfort, pain and psychological difficulty to patients as well as unsatisfactory clinical outcome. The MADIT-RIT study found substantial reductions in inappropriate therapy and all-cause mortality were achieved by modifying ICD detection parameters, either increasing the heart rate boundaries for detection or increasing the time to detection of ventricular tachycardia (VT) and VF (Moss et al. 2012). This major improvement in outcome from simple algorithm adjustments has reduced research activity into further development of diagnostic discriminators in ICD's.

Single chamber pacemakers and ICDs have one electrode in the right ventricle and dual chamber pacemakers and ICDs have two electrodes, one in the right atrium, and another in the right ventricle. This enables them both to detect electrical activity from and deliver electrical therapies, such as bradycardia pacing therapy; anti-tachycardia pacing (ATP) and shock therapy to the chambers in which electrodes are placed. ICDs illustrate sophisticated diagnosis using intracardiac electrograms, similar to those collected for interpretation during EP studies, as in the objective of this study.

The relationship between diagnosis and treatment is the key to delivering effective healthcare generally. Misdiagnosis is important where it results in either the delivery of an inappropriate treatment or where a treatment is not delivered when it is required. This evidence has shown that automated cardiac rhythm diagnosis remains imperfect and in many cases lies substantially below that achievable by a human domain expert. Automated cardiac rhythm diagnosis offers the prospect of technology coming to our aid in widening the availability of expert diagnostic capability and of maintaining uniformly high standards of diagnostic accuracy. Atrificial intelligence (AI) technologies offer a prospect of improved diagnostic accuracy and this will be explored as part of this study.

1.5 Towards Accurate Arrhythmia Diagnostic Algorithms

This study hypothesises that new algorithms, offering improved diagnostic accuracy and improved granularity of rhythm classification, could lead to appropriate treatment decisions, supplementing expert interpretation, leading to a reduction in misdiagnosis and with the potential to offer sophisticated accurate diagnostic capability where such expertise is not available. A preliminary study demonstrated the feasibility and desirability of using an artificial neural network (ANN) to develop an improved cardiac rhythm diagnostic algorithm (Bostock 2004).

A challenge of automatic rhythm diagnosis is to synthesise the knowledge and techniques of domain experts, in this case cardiac electrophysiologists. Techniques used in clinical cardiac arrhythmia diagnosis have the potential for inclusion in automatic diagnostic algorithms and this will be attempted in this study. This study will refine previous work, producing improvements to the performance of algorithms in implantable devices, leading to improved clinical outcome and quality of life for patients. Using a structured approach, informatics techniques will be used to synthesise the diagnostic methods of a human cardiac electrophysiologist, particularly focussed on the use of AI classifiers, using data obtained during human electrophysiological testing procedures and evaluate the performances of any production system.

1.6 Research Question

A research question encapsulating the main theme of the research was formulated:

> *"Can cardiac rhythm classification algorithms with greater performance than those current commercially available be developed using artificial intelligence for electrophysiological studies?"*

1.7 Aims and Objectives

It should be noted that given the considerations outlined in Chapter 1, section 1.4, there are no prominent comparable diagnostic systems designed for electrophysiological studies, using intracardiac electrophysiological signals. Performance comparisons are best made against results of bench studies and ICD diagnostic accuracy studies.

The aims of the research were identified based on the research question:

1. *Produce cardiac rhythm classification algorithm with good performance compared with bench studies and those of implantable cardioverter defibrillators, suitable for use during electrophysiological studies and other applications.*
2. *Use of artificial intelligence.*

Critical steps in achievement of these aims were set as objectives and as an outline of the processes in the study. These were: to review the literature; outline a theoretical background; assemble resources and tools and develop a methodology. Following these, data was to be collected and prepared for use in prototype algorithms; divided into training and test sets. Iterative development cycles (prototyping) would be performed using a variety of classifiers and classifier combinations and performances compared with current algorithms.

1.8 Hypotheses

Using the research question, null and research hypotheses were designed to allow satisfaction of the research aims to be statistically tested:

Null hypothesis (H_0)

Prototype cardiac rhythm diagnostic classifiers using AI do not outperform current algorithms.

Research Hypothesis (H_1)

Prototype cardiac rhythm diagnostic classifiers using AI outperform current algorithms.

1.9 Research Plan

A methodology was developed, guided by the literature review and an examination of the theoretical background of classification, with emphasis on approaches already known to be successful. The methodology led to data collection for features as classifier inputs and construction of base classifier units. Classifier development would culminate in a production system with performance characteristics and a level of statistical significance enabling acceptance or rejection of the null hypothesis.

1.10 Summary

Rhythm diagnosis, its methods and current systems for computerised diagnosis were summarised. The need for the research was established from a brief critique of currently available systems and identifying limitations, such as the availability of domain expert diagnosis. The rate of misdiagnosis was identified as the problem for address. This research aimed to synthesise domain expert diagnostic capability, aiming to reduce misdiagnosis.

The research question asked whether new rhythm diagnostic algorithms could be designed using AI, to exceed the performance of existing systems. Based on the research question, aims and objectives, null and research hypothesis were stated and a research plan outlined.

Chapter 2 Literature Review

2.1 Overview

A need for improved accuracy of automatic cardiac rhythm diagnostic algorithms in ICDs was outlined (see Chapter 1, subsection 1.1.2). This review sought to provide an overview of research and standard methods in this field and identify new methodological approaches.

From the research question, aims, objectives and hypotheses (see Chapter 1, sections 1.6, 1.7 and 1.8), inclusion and exclusion criteria were defined, literature sources and search resources identified, search terms defined and a search strategy formulated.

Research which met the inclusion criteria was critically appraised. Search results were analysed to identify prominent authors and principle sources of research. Continuous re-evaluation of the literature was performed during the research to provide supplemental references. Influences on this research from the reviewed literature were summarised.

EndNote X5 (Thomson Reuters 2011) bibliographic software was used for reference management.

2.2 Search Terms

Key words were extracted from the research question, aims and objectives (see Chapter 1, sections 1.6, 1.7 and 1.8):

cardiac rhythm
classification
algorithms
artificial intelligence
electrophysiological studies
compared
bench studies
implantable cardioverter defibrillators

Further examination of the research question identified two domains: cardiac rhythm analysis using AI and implantable defibrillator algorithms.

To generate search terms, the use of natural language or a controlled vocabulary was considered. Medical subject headings (MeSH) terms are suited to medical fields but unsuited to AI

terminology so a decision to use natural language terms was made. Using these key words, search terms, synonyms and subordinate terms were identified:

Term	Synonym	Subordinate term
cardiac rhythm	(none)	
artificial intelligence	(none)	pattern recognition
		neural network
		fuzzy
		support vector machine
		expert system
		decision tree
		genetic algorithm
		evolutionary
		Bayes
Classifier	(none)	
algorithm	process	
comparison	correlation, association	
implantable cardioverter-defibrillator	(none)	

Search terms were combined using Boolean operators, rationalised by limiting combinations to those forming "sentences" within the identified domains. Results were concatenated (see Appendix B) and references stored in EndNote.

2.2.1 Search Scope

Broad subject areas encompassed by the research question were biomedical engineering, computer science and medicine. The search was limited by inclusion and exclusion criteria, selection of source databases, search strategy and the selection of references by relevance.

2.3 Inclusion and Exclusion Criteria

To ensure a sufficiently broad search but limiting results to high quality texts and research most likely to be of relevance, inclusion and exclusion criteria were identified:

2.3.1 Inclusion Criteria:

1. Direct relevance to the research question.
2. Recognised authoritative texts, peer-reviewed research, reviews, conference proceedings (within 5 years) or new technology/ research not found elsewhere; patents; technical papers (grey literature).

2.3.2 Exclusion Criteria:

1. Literature relating to the topic area but not related to the specific focus.
2. Literature proposing new algorithms, using theory alone, without provision of bench test results and standard statistical analysis.

2.4 Literature Sources

For subject areas identified in the scope (see subsection 2.2.1), all relevant high quality literature that met the inclusion criteria (see subsection 2.3.1) were found to be available through online search engines.

The Google search engine was used as a general search tool with Google Scholar as a broadly based general literature search resource. Specialist search resources were selected based on subject areas and likely availability of content relevant to the research question. Scientific bibliographic database resources were evaluated. In the area of general scientific enquiry: Elsevier Sciverse Hub and Web of Science; for biomedical engineering: Institute of Electrical and Electronics Engineers (IEEE) Xplore and IET InspecDirect; for computer science: ACM Digital Library and CiteseerX and for medicine: Embase, and Cinahl.PubMed (Medline portal) and Cochrane Library. During trial searches, Elsevier Science Hub, IET InspecDirect, Embase, Cinahl, ACM Digital Library and CiteseerX were considered to add little to results, either producing duplicate results or lacking sufficient granularity by producing excessive results.

It was noted that, other than for selected cases, non-indexed abstracts from major medical conferences were not available within these resources. To correct for this, recent conference abstracts from Heart Rhythm Society (HRS), the most prominent annual international conference in the field were sought. Three dedicated databases were used: HRS Abstract Archive 2006 - 2010 Abstracts2View™ Archive and Online Itinerary Planners for HRS 2011 and 2012 (Heart Rhythm Society 2010a, 2010b, 2011, 2012).

The sources selected were: Google (restricted to first 10 results by relevance); Google Scholar (restricted to first 10 results by relevance); IEEE Xplore; Science Direct; Web of Science; PubMed; HRS Abstracts and the Cochrane Library. Primary, secondary and grey literature sources were considered. Primary sources were defined as: articles in peer reviewed journals; conference proceedings and textbooks of widely accepted high quality and repute. Secondary sources were defined as: review articles in peer reviewed journals; reference books; Wiki and other Internet sites; biographies.

Grey literature sources were defined as: theses; unpublished conference proceedings; reports. Secondary sources were examined for cited high profile research not previously found using the search strategy.

2.5 Search Strategy

The scope of the search was defined. Search terms identified in section 2.2 were entered as search criteria into the selected resources and the numbers of references returned for each term or combination of terms were assessed.

2.5.1 Granularity of Search

A sufficiently narrow search, retaining the most relevant results, was considered to be ≤ 100 returned results. If the number of returned results was greater than 100, refinement of the search was considered to reduce the results and improve relevance. Selection of a second or third term and use of Boolean operators was used, provided they produced a logical search "sentence". Use of fewer terms and Boolean OR would be used to widen a search if required. Narrowing a search used combinations of search terms using Boolean AND as well as "filters" (such as: topic areas; journal name or keywords) offered within resources. When sufficiently narrow, results were examined for relevance to the study and stored for further examination and refinement.

2.5.2 Relevance

Relevance was considered to be content involving the design, testing and evaluation of methods for cardiac rhythm classification using an AI technique or a commercial ICD. Initial assessment used title and abstract content. References were rated highly, moderately or not directly relevant. Where review articles were found, their reference lists were examined to ensure inclusion of relevant cited literature. Where the same research was represented by a conference paper and subsequent full journal paper the latter was quoted. For highly relevant papers, the content was evaluated and critically reviewed.

2.6 Search Results

An initial literature search was performed in October 2005 and repeated in April 2012. Results of the latter search were analysed. It was noted that approximately 55% of relevant references originated between 2005 and 2012.

Use of a single search term was found to provide insufficient granularity. Use of Boolean AND operators to combine terms in "sentences" of two or three terms achieved sufficient granularity and no further refinement was required (see Appendix B).

Search "sentences" used were:

rhythm
rhythm AND classifier
rhythm AND “artificial intelligence”
rhythm AND “neural network”
rhythm AND “fuzzy”
rhythm AND “support vector machine”
rhythm AND “expert system”
rhythm AND “decision tree”
rhythm AND “Bayes”
rhythm AND “genetic algorithm”
rhythm AND evolutionary
rhythm AND “pattern recognition”
implantable cardioverter-defibrillator
algorithm AND comparison
implantable cardioverter-defibrillator AND algorithm
implantable cardioverter-defibrillator AND algorithm AND comparison

Note that results using the term “cardiac”, and the synonyms “correlation”, “association” and “process”, although tested, were unproductive and are not listed.

Results from both domains (see section 2.2) were pooled. Following elimination of duplicates and an initial evaluation for relevance, returned results were: 4 book chapters, 7 peer-reviewed reviews, 136 peer-reviewed journal papers and 47 peer-reviewed conference papers (within 5 years of 2012). No relevant grey literature or patents were found.

2.6.1 Results by Year of Publication

Publication trends were examined (see Fig. 2.1). Publication in significant numbers commenced in 1989 and plateaued at about 4 publications per annum until 2004. For 2012, mean annual publication had risen to approximately 11 publications per annum.

2.6.2 Results by Journal

Results were analysed by journal and impact factor (IF) for 2010 (BioxBio.com, 2013). The scientific journal with most publications in the field was IEEE Transactions on Biomedical Engineering (21 papers), then Pacing and Clinical Electrophysiology (14 papers). Papers published in journals with very high impact factor: 6 papers in Circulation (IF = 14.816) and 4 in the Journal of the American College of Cardiology (IF = 12.535). The most prominent journal in the field of computer science was Artificial Intelligence in Medicine (IF = 1.568). Surprisingly, a

major journal in the specialised field of electrophysiology, Heart Rhythm (IF = 4.246) was under-represented in results, providing only one result (see Appendix C (a)).

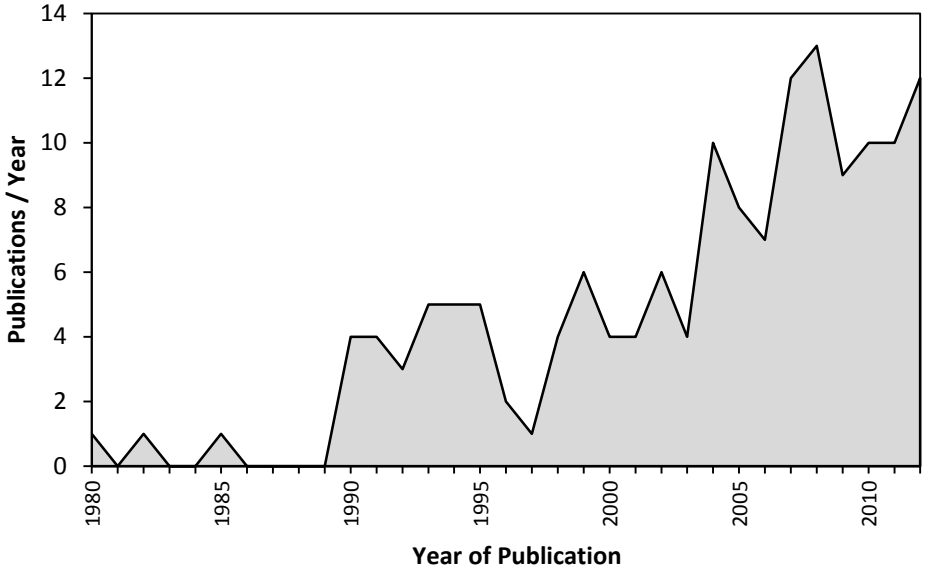

Figure 2.1 Totals relevant publications per year, by year of publication.

2.6.3 Results by Author

Returned results were searched by co-authorship for publication frequency, ranked by number of citations and most recent citation in the field within 10 years (see Appendix C (b)).

The four most prominent authors were all USA based: Dr Michael Gold, Chief of Cardiology at the Medical University of South Carolina and Professor of Medicine at the Medical University of South Carolina; Dr Charles Swerdlow, a cardiac electrophysiologist at Cedars-Sinai Medical Center, Los Angeles and Professor of Medicine at the University of California, Los Angeles; Jeffrey Gillberg MS, Research Director at Medtronic, who has a large portfolio of published articles, reviews and patents related to cardiac rhythm management devices and Dr Kenneth Ellenbogen, Chairman of Cardiology at McGuire Veterans Administration Medical Center in Richmond, Virginia and Professor in Cardiology at Virginia Commonwealth University.

Other prominent authors were: Prof. Peter Macfarlane Institute of Cardiovascular and Medical Sciences, University of Glasgow, renowned for pioneering work on application of computer techniques to ECG interpretation; Associate Prof. Vessela Krasteva, Bulgarian Academy of Science, Sofia, Bulgaria; Yüksel Özbay, Selcuk University, Konya, Turkey; Prof. Stanislaw Osowski, Warsaw University of Technology and Dominic Theuns PhD at Erasmus MC, Rotterdam, the Netherlands.

2.6.4 Conference Papers

Representing recent and unpublished research within 5 years (2007 to 2012), 47 peer reviewed conference papers were found. Major conferences were identified as the IEEE Engineering in Medicine and Biology Society, HRS and Computing in Cardiology conferences. Biomedical engineering and computer science conferences were well represented in the literature, believed to be linked to the custom of publishing full papers, often indexed, whereas medical conferences, with a more usual format of brief published abstract, supported with a poster or oral presentation and commonly non-indexed, were not well represented.

2.6.5 Results by Algorithm

Search results were evaluated for classifier technologies and ranked by both journal and conference papers and total publications since 2007. Neural networks were ranked highest (14 journal and 11 since 2007); followed by fuzzy logic, support vector machines and morphology discrimination (see Appendix D). Uses of hybrid technologies and multi-classification systems were led by neuro-fuzzy classification (4 publications since 2007). Two studies compared multiple technologies (Jovic & Bogunovic 2011; Acharya *et al*. 2004). In a similar ranking for ICD algorithms, morphology and dual chamber detection ranked highest (see Appendix C (c)).

2.7 Selection of References

Having been refined by relevance, search results were further refined using bibliometric scoring. Review articles were examined for supplementary references to ensure broad capture of important sources. Rejected articles were re-examined for re-inclusion based on any importance or relevance "missed" by the inclusion/ exclusion criteria.

2.7.1 Bibliometric Analysis

Relevant results were ranked for highly-citedness using bibliometric analysis. Impact factors were accessed from BioxBio.com (2013) and citation scores from Google Scholar. Three factors were used for ranking: Google Scholar total citations per paper (C); calculated citations per paper per year since publication (c′) (2.1) and impact factor (IF).

$$c' = (C \,/\, (2012 - (year\ of\ publication)) \tag{2.1}$$

References were further refined, satisfying at least one of the following criteria:

1. C ≥ 25 (highly-cited papers).
2. c′ ≥ 8 (recently published highly-cited papers).
3. Published in 2012 in journals with IF > 2 (high profile recently published papers).

2.7.2 Supplementary References

Among review papers, 7 additional references were considered important, although none met the inclusion criteria, they were retained as background (Moss *et al*. 2004; Poole *et al*. 2008; Osswald *et al*. 2000; Langenfeld *et al*. 1998; Ellenbogen *et al.* 1990; Turcott & Pavek 2008; Revishvili 1999). Excluded papers were re-examined for inclusion should they offer significant relevant content. One paper was reinstated in this way (Hintringer *et al*. 2004).

2.7.3 Evidence

In addition to citation data, levels of evidence were evaluated. References were dominated by bench studies using open-access data sources for classifier training, testing and validation. A number of review papers and cohort studies were available. Only two randomised clinical trials (RCT) were available (Gold *et al.* 2012a and 2012b).

2.8 Critical Reviews

The final collection of 47 papers consisted: 4 review papers; 19 papers (1 conference paper) describing single classifier technology; 4 hybrid classifiers; 6 multi-classifier systems; 2 comparing classifiers; 7 papers (1 conference paper) single ICD algorithms and 6 comparative studies of ICD algorithms. Full texts were accessed online and subjected to critical review.

Critical reviews took account of provenance, hypothesis, study design, limitations and impact on this research. All selected references were published by authors and scientific journals of high repute. Papers were grouped as: review papers; single classifiers; hybrid classifiers; multi-classifier systems; comparative studies of classifiers; single ICD algorithms and comparative studies of ICD algorithms. Single classifier papers were further grouped by classifier type: statistical classifiers; syntactic classification; neural network classifiers; fuzzy classifiers; decision tree classifiers; support vector machine classifiers; *k*-nearest neighbour classifiers and other classifier technologies.

2.8.1 Review Papers

Aliot *et al*. (2004) compared arrhythmia detection algorithms in ICDs using a meta-analysis of data published in previous individual studies. An overview of five different manufacturers' algorithms was provided. Clinical results of selected studies were summarised and showed all algorithms having a VT sensitivity of >98% and specificities between 66 and 94%. Hintringer's study (2001) compared four device algorithms using the same arrhythmia test set, showed specificities of 11% for the Boston Scientific, 12% for the Biotronik; 20% for the Medtronic and 28% for the ELA algorithms. This review discussed limitations of many of the studies, advising cautious interpretation due to small study sizes, lack of large randomised studies and that some studies focused on specific tachycardias, making comparisons difficult. They also mention

limitations of imperfect ICD sensing and data storage in capturing and counting arrhythmic events. On statistical analysis of algorithm performance, they noted that specificity illustrated classification performance of supraventricular rhythms and positive predictive values (PPV) measured appropriateness of therapies. The authors criticised lack of use of PPVs in studies, as these are known to have value as accuracy indices (Charbonnier & Tacker, 1994). They recommended use of the generalised estimating equation (GEE) to correct for bias, where a few patients account for many episodes (Zeger *et al*. 1988) and cite Wilkoff *et al*. (2001) in this. They suggested improved uniformity of study design and that new hemodynamic sensors might augment rate detection for therapy decisions. This review of ICD algorithms failed to analyse results in depth, citing differences in statistical technique between studies. The recommendation of PPV as a measure of algorithm accuracy does not solve the issue of a global accuracy measure, as it uses only two cells of the 2 x 2 contingency table, true and false positives, where alternative statistics, such as kappa (κ), use all four cells of the 2 x 2 contingency table, more fully representing accuracy.

An invited review by Kaszala & Ellenbogen (2010) centred on use of sensors in devices. They summarised sensors in current use and included a section on myocardial function relevant to this study. They particularly mentioned the study of Osswald *et al.* (2000) which demonstrated that right ventricular (RV) impedance changes correlated with contractility and that acceleration sensors within the lead achieved a similar effect (Langenfeld *et al*. 1998). Poor specificity of arrhythmia discrimination using hemodynamic sensors (Ellenbogen *et al*. 1990) was briefly discussed and they suggested miniaturised wireless sensors and photoplethysmography (Turcott & Pavek 2008) as potential solutions for this. This paper failed to critically appraise the technologies it summarised, seeking to offer a technological overview and look to the future, rather than assessing the available research.

Andrikopoulos *et al*. (2010) focused their review on device-based monitoring. A summary of arrhythmia diagnostic capabilities of devices, including counters, histograms and electrogram storage was provided. They provided a brief overview of self-diagnostic capabilities related to pacing, sensing and impedance and briefly mentioned automaticity and its associated reduced need for intervention. Heart failure monitoring capabilities with heart rate variability (HRV), intrathoracic impedance, patient activity, intracardiac pressure monitoring, patient alerts and ECG ST segment monitoring to detect ischemic events were summarised. The authors concluded with a suggestion for improvements in the accuracy of diagnostics and for patient alerts as important directions for development. Although this review was likely to guide study design, it offered no assistance in the area of algorithm accuracy and performance.

Jackson *et al.* (2012) emphasised the importance of misclassification of atrial arrhythmias to rates of inappropriate therapy. They quoted the MADIT II trial (Moss *et al.* 2004) and the SCD-HeFT trial (Poole *et al.* 2008) where 34% and 22% of patients respectively received inappropriate shocks. They discussed consistency of detection algorithms, focussing on detection and prevention of inappropriate shock treatments. Arrhythmia detection algorithms were briefly outlined but no suggestions of specific future developments in this area were made, limiting its value for this study.

2.8.2 Statistical Classifiers

Zhang *et al.* (1999) aimed to demonstrate a chaotic complexity measure described by Lempel & Ziv (1976) applied to arrhythmia detection. Classification was non-automated sequential hypothesis testing. They found sensitivity, specificity and accuracy to be incremental with ECG window lengths, >90% at 5 sec and reaching 100% at 7 sec. Statistical analysis and comparisons with previous studies were not made. They failed to discuss methods for automation of their algorithm.

Chiarugi *et al*. (2007) described a method to detect termination of atrial fibrillation (AF) based on two-lead surface ECGs. QRS complexes were eliminated using a QRST cancelling approach based on morphology classification and average cluster template subtraction. Stepwise discriminant analysis was used for feature selection. This paper, within its limitations and specific to the discrimination of AF termination modalities, was well designed and conducted. The study did not directly address rhythm classification beyond AF, limiting its value.

2.8.3 Syntactic Classification

Udupa & Murthy (1980) aimed to develop a syntactic approach to ECG rhythm recognition. The authors succeeded in their aim and produced some initial accuracy measures however they failed to validate these with statistics or comparison with other studies. Syntax analysis was performed using context-free languages to classify the patterns. This short paper described a novel process using a very small development set. There was a simple validation process but no testing with new data. Statistics were lacking and detailed discussion very brief. The paper was useful to this research as a test of concept of syntactic analysis of ECG rhythm.

2.8.4 Neural Network Classifiers

Yang *et al.* (1994) compared an ANN with deterministic logic for ECG classification. Combination methods of the neural-network and deterministic classifications were explored. They reported sensitivity for AF using ANN as 92%, compared with 88.5% using deterministic classification, with no difference in specificity of 92.3%.

Coggins *et al.* (1995) demonstrated classification of QRS morphologies using a purpose-designed very-large-scale integration (VLSI) circuit containing a neural network classifier. The ANN used a 6-node hidden layer and 3 outputs with a "winner-take-all" arbitration. Results showed ≥ 93% accuracy for VT and 77.6% -100% for sinus tachycardia. They offered preliminary results and suggested possible improvements. It was useful to this study as a practical approach to the integration of technology in devices.

Minami *et al.* (1999) used power spectral analysis to generate 5 spectral components as inputs to an ANN classifier implemented on a single chip central processing unit (CPU) and compared classification of their system using surface ECGs against intracardiac electrograms (EGM). For classification into normal, VT or VF rhythms, ECG showed 80% sensitivity and 92% specificity; the EGM showed 68% sensitivity and 86% specificity. This paper offered an example, useful to this research, of effective rhythm classification using an ANN and demonstrated its implementation as a low-power single chip configuration, offering potential for incorporation in implantable devices.

Kara & Okandan (2007) examined specifically the diagnosis of AF as a rhythm diagnosis. Mean power spectral densities were used as features for input to an ANN. On the trained network, classification performance of sensitivity, specificity and positive predictive value, and accuracy were 100%. The authors failed to detail the source of their ECG data. Discussion of study limitations was brief, mentioning only that the method worked off-line and not in real-time. This paper provided an elegant methodological description, though its failings render it of little further value.

Christov & Bortolan (2004) ranked ECG features for premature ventricular contraction (PVC) classification. Using Matlab Neural Network Toolbox, two approaches were taken: the Levenberg–Marquardt algorithm for rapid convergence and the Bayesian framework for good generalisation. For all features sensitivity was 99.7% and specificity 98.5%. Features were ranked using the values of ANN input weights. Additional features improved specificity (up to 98.5% with 26 features). The study concluded that beat classes should have individual optimal feature sets for classification. This demonstrated use of an ANN for feature selection, indicated grouping of features and suggested further research into class-specific feature sets, valuable in this research.

Acharya *et al.* (2008) aimed to show that of ECG power spectral density features, extracted using an autoregressive moving average model improved ANN ECG classification. The authors were able to show that their method held advantage over certain alternate feature extraction alternatives but failed to adequately compare the results with other studies. This paper held interest solely for its use of power spectral density as ECG features providing limited value for this research.

2.8.5 Fuzzy Classifiers

Usher *et al.* (1999) designed a fuzzy inference system for cardiac rhythm classification intended for use in ICDs. The authors concluded the method provided correct rhythm classification and was computationally efficient. Without published results or statistical analysis, these claims were poorly substantiated. This paper was of value to this research in its demonstration of the feasibility of fuzzy systems for this application.

Anuradha *et al.* (2008) looked at four features extracted from ECG intervals used as inputs to a Mamdani fuzzy classifier and assessed classification accuracy. The results showed an accuracy of 93.1%. Having quoted four similar studies (Zong & Jiang 1998; Sugiura *et al.* 1998; Acharya *et al.* 2003; Kannathal *et al.* 2005) they failed to compare their results with these studies to demonstrate any benefit from their approach. Though an interesting study, it was unclear if the use of a fuzzy classifier, the choice of features or wavelet pre-processing was responsible for success. The use of features derived only from RR interval seemed a limitation, so perhaps we can ascribe any benefit to the fuzzy classifier.

2.8.6 Decision Tree Classifiers

Tsipouras *et al.* (2005) designed a knowledge-based rhythm classification system using ECG RR interval features for beat then rhythm classification. Rules based on expert clinical practice were applied sequentially. Detailed results were tabulated and the summary statistics provided were: beat classification 98% accuracy; rhythm classification 94% accuracy. Not unexpectedly, the system performed well on beats pre-selected to match the classification system and less well on un-preselected beats. Described as an expert system and "deterministic automaton", the system *de-facto* comprised a decision tree classification system. There was no attempt at statistical significance measures. The study was useful as an indicator for the utility of expert knowledge in a classification system.

Rodriguez *et al.* (2005) investigated the utility and assessed performance of a personal digital assistant (PDA) to classify ECG beat and rhythm. Weka open source (University of Waikato, New Zealand), AnswerTree (IBM Corp., New York) decision tree and LogitBoost (Friedman *et al.* 2000) boosting applications were used. The authors derived inference rules from cardiology texts and pruned them using reduced-error pruning. A weighted majority vote provided final class. The algorithm was implemented in a PDA. The study was of interest for the extraction of inference rules and rule pruning, rather than algorithm performance.

2.8.7 Support Vector Machine Classifiers

Polat & Gunes (2007) conducted a bench study using an SVM to classify rhythm and PCA as a feature reduction method. Results with all three data allocations were 100% sensitivity and

specific and accuracy. The study suggested PCA feature selection and SVM attained near perfect performance and provided an interesting indicator, within its limitations.

Asl *et al.* (2008) proposed an arrhythmia classification algorithm based on generalised discriminant analysis (GDA) and SVM. Features were reduced using GDA and used as inputs to the SVM. SVM classification using one versus all (OVA) and one versus one (OVO) methods were compared. The SVM with OAA was found to produce marginal performance gain. SVM classifier performance using GDA feature reduction, sensitivity was 95.7%, specificity 99.4% and accuracy 99.1%. The study served to recommend SVM as a classifier, despite design faults.

2.8.8 k-Nearest Neighbour Classifiers

Owis *et al.* (2002), in a bench study, used nonlinear dynamics of ECG signals for rhythm classification. Clustering, Bayes and *k*-Nearest Neighbour (*k*-NN) classification methods were compared. The best overall classifier was *k*-NN (*k*=12) giving a sensitivity of 85% and specificity of 34%. Owis' study, contradicted the results of Zhang (1999), with the resultant conflicting evidence of utility of chaotic features in this research.

Minhas & Arif (2008) demonstrated use of a wavelet transform and a *k*-NN classifier to provide high classification accuracy. The study design, using this classifier to assess the value of features based on a wavelet transform as well as the use of PCA, reduced the impact of this study. The seeming inadvisability of this approach limited its usefulness to this study.

2.8.9 Other Classifier Technologies

Sarkar *et al.* (2008) described a Lorenz plot as a detector for AF and atrial tachycardia (AT) with an irregular ventricular response and a supplementary detector for AT with regular response. The differences in density of the Lorenz plot distributions of RR intervals during various rhythms were illustrated. Validation showed that including the supplementary detector improved sensitivity from 80% to 92% and reduced specificity from 99% to 96%. The authors emphasised the utility of this study for longer term monitoring of atrial tachyarrhythmia, rather than its classification accuracy, limiting its utility for this study.

Bayesian classifiers have not been widely tested in this domain with only one paper found (Brüser *et al.* 2012), looking at diagnosis of AF from ballistocardiograms, which examine body vibrations corresponding to heart beats, tested a naïve Bayes classifier among others and found it not to be a high performer.

2.8.10 Hybrid Classifiers

Wang *et al.* (2001) proposed short-time multifractality for arrhythmia detection. The classifier was a fuzzy Kohonen network (Tsao *et al.).* Data was recorded from patients at ICD implantation. Sensitivity, specificity and accuracy were: for VF (98%, 96% and 97%), VT (95%, 99% and 97%) and AF (98%, 100% and 99%). The authors suggested that a large cardiac rhythm database would be necessary. This paper was of interest for its assessment of fractal dimensions of the ECG and their usefulness as features for rhythm classification.

Linh *et al.* (2003) devised a hybrid neuro-fuzzy classifier for heart rhythm. Misclassification rate was 3.95%, corresponding to an accuracy of 96%. Comparisons were made to four results from the literature but not discussed in detail or the statistical significance of their results.

Polat *et al.* (2006) applied an existing artificial immune recognition system (AIRS) to arrhythmia classification. The algorithm was in four stages: initialization; memory cell identification and artificial recognition ball (ARB) generation; competition for resources and development of a candidate memory cell and memory cell introduction. The study showed an interesting anecdotal use of a relatively new field of AI in rhythm classification.

Exarchos *et al.* (2007) proposed a fuzzy expert system (FES) for ischaemic and arrhythmic beat classification. Construction of decision trees used the C4.5 inductive algorithm with post-pruning. Using the FES, performance was 92.4 to 98.9% sensitivity, 97.6 to 99.9% specificity and 95.8% accuracy and was statistically significant. This was a well-designed study which succeeded in its aims, a possible criticism was the method for extraction of crisp rules, relying on data mining, rather than expert knowledge. The paper convincingly assessed the accuracy of its prototype algorithm and compared its accuracy with competing systems. Several aspects of methodology acted as a model for this research.

2.8.11 Multi-Classifier Systems (MCS)

Leong & Jabri (1992) designed a neural network-decision tree multi-classifier system which used a first-past-the-post arbitration strategy. They used intracardiac electrogram morphology and timing for decision tree classification, augmented by a neural network for 1:1 tachycardias. 99.6% correct classification was achieved however there was no statistical analysis of this result. In this very early description of a multiple classifier system, the authors discussed possible reasons for misclassifications and the limitations of their study. Comparison is made with an ICD algorithm using the same data but no detailed statistical analysis was made of the differences.

De Chazal *et al.* (2004) designed and tested an ECG classification system comparing eight combinations of features and 12 classifier configurations. Their system consisted of linear

discriminant classifiers using Bayesian product integration as a combiner. For supraventricular ectopic beats, sensitivity was 75.9%, positive predictive value 38.5% and false positive rate 4.7%. For ventricular ectopic beats, sensitivity was 77.7%, positive predictive value 81.9% and false positive rate 1.2%. The authors found that feature sets using absolute amplitude and ECG segmentation improved results. They discussed choice of data division methods and stated that dividing data on a beat basis optimistically biased classification results. They also discussed possible reasons for misclassification errors.

De Chazal and Reilly (2006) proposed an adaptive ECG arrhythmia detection where the classifier used expert knowledge about a portion of the recording to improve performance on the rest of the recording. The system processed ECG with a global-classifier, placing beats into one of five classes. A fraction of the beats were subjected to expert validation and corrected. A local-classifier was trained using the validated beats and combined with the global-classifier to form an adapted system to classify future beats. For premature atrial contractions (PAC's) sensitivity was 87.7%, accuracy 95.9% and PPV 47%; for PVC's sensitivity was 94.3%, accuracy 99.4% and PPV 96.2. Global accuracy was 84.5%. A full confusion matrix was provided. The results also showed performance increases with training on the local record. The use of a simple combining strategy served as a good working example of multiple classifier system in practice.

Osowski *et al.* (2004) demonstrated a multiple classifier approach to ECG classification using SVM and a weighted voting combiner. The authors failed to make detailed comparison with other studies or estimate statistical significance. Misclassification rates using a higher order statistics feature extraction was 3.5% and 2.8% using the Hermite feature extraction method and 2.6% with the combined classifier. This study expanded on previously published work by the authors Linh *et al.* (2003).

Ozbay *et al.* (2006) conducted a bench study of ECG classification using a fuzzy clustering-neural network (FCNN) multi-classifier system in a cascade configuration. Inputs were segments of ECG in the RR interval. Outputs of unsupervised self-organizing neurons generated inputs to an ANN classifier. Results were compared between an ANN and the FCNN. Classification accuracy was 98.9% (NN) and 99.9% (FCNN) and mean classification error 0.22. Training time for the FCNN was 60% that of the ANN. They concluded that FCNN generalised better than ANN, with a shorter training time. The study was of interest for its innovation and suggested that this type of classifier may be more computationally efficient for similar accuracy to an ANN.

Ceylan *et al.* (2009) developed a fuzzy c-means clustering-PCA-NN multi-classifier system in a cascade configuration. Feature extraction used PCA, fuzzy c-means clustering and wavelet

transform, then compared performance with an ANN classifier. This was an interesting study of a multi-classifier system in a cascade configuration, let down by poor study design.

2.8.12 Comparative Studies of Classifiers

Jovic & Bogunovic (2011) compared 7 classifiers using RR interval features in a bench study. The best overall accuracy of 99.6% was achieved by random forest. Classification sensitivity, specificity and accuracy for normal versus arrhythmia with decision tree; Bayesian network; ANN; SVM and random forest all approached 100%. Selected results, noted only in passing by the authors, showed very high performance for ANN, SVM and random forest in detection of arrhythmia. This bench study was particularly useful for its comparisons of several classification methods on the same data and its comparisons with other studies.

Acharya *et al.* (2004) examined derivatives of the RR interval, used as inputs to two alternate classifiers, an ANN and a fuzzy classifier and results compared. They found the accuracy of their fuzzy system to be better than the ANN at 84% or better. This was an interesting technical paper however due to its lack of analysis served only as insight to the direction of future research.

2.8.13 Single ICD Algorithms

Kuhlkamp *et al.* (1999) compared arrhythmia discrimination functions of single chamber and dual chamber ICD algorithms (see Chapter 1, subsections 1.1.1, 1.3.1 and 1.3.2). Inappropriate treatment rates were used as the primary measure. They found the rate of inappropriate therapies was significantly higher with the dual chamber detection system. In summary, this study was well constructed and executed within its recognised limitations (single centre, non-randomised) and used standard statistical analysis techniques. The study concluded that, due to flaws in the decision algorithm, the rate of inappropriate therapies in the dual chamber ICD algorithm under evaluation was significantly higher.

Swerdlow *et al.* (2000) aimed to assess the ability of a new ICD with combined atrial and ventricular cardioversion capability to distinguish prolonged episodes of AF that required cardioversion from other AT's. Sensitivity of the ICD algorithm for AT/AF diagnosis was 100% and specificity was 99.99%. This study describes experience with this ICD. The results compared favourably with the quoted comparative study (Wellens *et al*. 1998) but the authors failed to offer detail of statistical significance.

Wilkoff *et al*. (2001) proposed methods for comparative studies and assessed the PR Logic dual-chamber detection algorithm using these methods. Recommendations for statistical corrections of algorithm performance metrics and a brief analysis of the algorithm under question using these techniques were made. The authors noted that absolute sensitivity could not be calculated, as the

ICD stored only data for episodes satisfying the detection criteria and a relative sensitivity was calculated. Absolute specificity of VT/VF detection could not be calculated for the same reasons and an incremental specificity was calculated. Detailed recommendations for statistical corrections of algorithm performance metrics were made with specific guidance on calculation of statistical corrections for sensitivity and specificity made. *P* values for algorithm performance and comparisons with other algorithms were not presented in this study.

Swerdlow *et al.* (2002) investigated a downloadable temporary ICD algorithm enhancement using Haar wavelet transforms of intracardiac electrograms as a morphology feature to aid discrimination of supraventricular tachycardia (SVT) from VT. Using a match threshold of 70%, sensitivity was 100%, GEE corrected positive predictive value was 61% and specificity 78%. The receiver operating characteristic (ROC) curve showed optimal match threshold to be 60–70%. They concluded that 69% of inappropriate detections could have been prevented by optimal programming. The authors failed to define the gold standard. Confidence intervals were quoted but no details of statistical tests used or origin of the quoted *P* value for specificity ($P < 0.01$).

In a multicentre, open study, Kouakam *et al.* (2004) aimed to evaluate the performance of a new dual-chamber ICD detection algorithm. Sensitivity was 99% and specificity 89%. Inappropriate therapy occurred in 17 cases. They concluded that though performance of the algorithm was satisfactory, diagnosis of 1:1 tachycardias was imperfect.

Kremers *et al.* (2012) demonstrated a use of RV pressure, incorporated in an investigational ICD device, to assess hemodynamic stability during tachycardia and improve diagnostic accuracy. The algorithm used RV pulse pressure (RVPP), RV dP/dt_{max} and electrogram based detection. In this small technology evaluation, 100% sensitivity and specificity was claimed, suggesting use of RV pressure improves arrhythmia discrimination. The study succeeded in testing the utility of the technology, paving the way for evaluation in a randomised clinical trial.

2.8.14 Comparative Studies of ICD Algorithms

Hintringer *et al.* (2001) conducted a study to compare four ICD algorithms of Biotronik, ELA, Guidant and Medtronic, using arrhythmias recorded during invasive electrophysiological studies applied to the inputs of ICDs *in vitro*. The responses were evaluated using interrogation of the ICD. Specificities were compared using McNemar's test (McNemar 1947). Specificities for supraventricular tachyarrhythmia were: Biotronik (12%); Guidant (11 %); ELA (28%) and Medtronic (20%), all very low values compared with the target 100% specificity. The authors identified potential for miscalculations of absolute specificity due to the disproportionate incidence of specific rhythms compared with their actual prevalence. They defended their

approach as a valid comparison, as all four algorithms were tested with the same arrhythmias. There was no comparison with data from other studies and no indication of *P* values.

Gold *et al.* (2002) Morphology-based algorithms had been shown to be effective for arrhythmia discrimination by Langberg *et al.* (1988). The authors proposed a second morphology channel could result in improved discrimination and evaluated a new algorithm based on vector timing and correlation (VTC). The results showed encouraging performance measures. Algorithm performance was assessed by comparison with physician diagnosis. In dual-chamber versus single chamber configurations (see Chapter 1, subsections 1.1.1, 1.3.1 and 1.3.2), sensitivities were 100% and 99% respectively and specificities 97% and 97%. Although no direct comparisons were made with competing morphology algorithms, the authors concluded that their algorithm demonstrated accurate classification of supraventricular and ventricular tachyarrhythmia.

Theuns *et al.* (2004) compared single and dual- chamber algorithms for improved specificity. Correction was made for multiple episodes within a patient using a GEE method. Tachyarrhythmia detection (atrial and ventricular) was not significantly different between both settings ($P = 0.77$). The study was well-designed and used a standard methodology. No comparisons were made with comparable studies.

Hintringer *et al.* (2004) compared specificity and the impact of sample size on specificity of ICD algorithms, using published clinical data and the authors own bench study. Data from multiple studies were pooled and specificities recalculated. Comparison of algorithms used McNemar's test (McNemar 1947) and the relationship between specificity and sample size was using Spearman's correlation coefficient. Specificities for the five ICD algorithms were: Biotronik Phylax AV/ Tachos DR (90%); ELA Defender (89%); Guidant Ventak AV III/ Prizm II DR (89%); Medtronic Gem DR (68%) and St. Jude Photon DR (76%). The correlation between specificity and number of patients reached statistical significance at the 5% level. In a bench test, after correction for prevalence, specificities were: Biotronik Phylax AV/Tachos DR (95%); ELA Defender (99%); Guidant Ventak AV III/ Prizm II DR (94%); Medtronic Gem DR (93%) and St. Jude Photon DR (92%). The interesting relationship between specificity and sample size is unsurprising but previously undocumented in this context. The authors' correction technique, using a weighting based on prevalence, was unproven and there was no explanation of the theory underlying their approach. The comparisons they made between algorithms clearly showed a difference using the same input data but statistical significance was not indicated.

Gold *et al.* (2012a) directly compared two manufacturers ICD algorithms for their ability to prevent inappropriate therapy. PPV for VT were 41.2% for Guidant and 51.3% for Medtronic devices. A standard programming configuration was used and chi-squared tests were used to

compare percentages. Cox proportional hazard models (Cox 1972) were used to obtain hazard ratios. Algorithm PPV's were 41.2% for Guidant and 51.3% for Medtronic devices. Hazard ratios similarly favoured the Medtronic devices (1.34 for inappropriate therapy), $P = 0.003$. Similarly, Kaplan-Meier curve (Kaplan & Meier 1958) for inappropriate therapy-free survival favoured Medtronic ($P = 0.002$). This study was the first well-designed very large prospective RCT of its type. Within its limited scope, it was able to demonstrate differences in performance between two algorithms, supported by high quality statistical analysis.

Gold *et al.* (2012b) assessed the performance of a new algorithm, specifically developed for an entirely subcutaneous ICDs (SICD), compared with established algorithms of dual- and single-chamber ICDs with standard intracardiac electrodes, the START study. The authors found that all systems had sensitivity of between 99 and 100%. Specificity of was 98.0% for the SICD and 64.0-92.0% for single chamber ICDs and 32.7-89.8% for dual chamber ICDs and was significantly better than 2 of the 3 standard ICD systems ($P < 0.001$). This well designed algorithm performance study tested algorithms using the same data resulting in good direct comparability

2.9 Summary of Literature

Using the research question, a search strategy, literature sources resources and search terms were defined. The search included primary, secondary and grey sources and was conducted using Google, Google Scholar, IEEE Xplore, Science Direct, Web of Science, Pubmed, Heart Rhythm Society Abstracts and the Cochrane Library. Granularity was narrowed using Boolean AND combinations of terms and results were limited using inclusion and exclusion criteria.

2.9.1 Reviews

Amongst review papers, Jackson *et al.* (2012) emphasised that up to 34% of patients receive inappropriate therapy, partly due to misclassification of atrial arrhythmias. Aliot *et al.* (2004) analysed ICD algorithm performance research finding studies had poor comparability and there was a need for large scale randomised studies. They recommended standardised statistical analysis and suggested that hemodynamic sensors might augment arrhythmia detection. New sensor technology and diagnostic capabilities in implantable devices was highlighted by Kaszala & Ellenbogen (2010). RV impedance and an acceleration sensor within the lead for contractility monitoring were featured. The poor specificity of haemodynamic sensors (Ellenbogen *et al.* 1990) was discussed and photoplethysmography (Turcott & Pavek 2008) proposed as having potential to solve this. Andrikopoulos *et al.* (2010) focused on device-based monitoring and suggested improved accuracy of diagnostics and patient alerts as an important direction for development.

2.9.2 Single Classifiers

Since 1980, reports AI use for cardiac rhythm classification, served to establish a potential utility and suggest which technologies more suited to the field. In a comparison of studies of individual

classifiers, there were notable differences in study design, in particular: data source, sample size, ECG features and number of classes, with the implication that direct comparison might not be meaningful. Despite this, there was evidence of good classification accuracies, where quoted, of more than 83% (see Table 2.1).

Table 2.1 Comparison of studies measuring performance of different AI classifiers.

Study	Classifier	Classes	Accuracy	Sensitivity	Specificity	PPV
Zhang *et al.* (1999)	Statistical	3	100	100	100	-
Chiarugi *et al.* (2007)	Statistical	5	89-92	-	-	-
Udupa & Murthy (1980)	Syntactic	6	82	-	95	-
Yang *et al.* (1994)	ANN	2	-	92	92.3	-
Coggins *et al.* (1995)	ANN	2	93	-	-	-
Minami *et al.* (1999)	ANN	3	-	80	92	-
Kara & Okandan (2007)	ANN	2	100	100	100	100
Christov & Bortolan (2004)	ANN	2	98.5	99.7	98.5	-
Acharya *et al.* (2008)	ANN	9	83.8	81.7	100	100
Anuradha *et al.* (2008)	Fuzzy	8	93.1	-	-	-
Tsipouras *et al.* (2005)	Decision	6	93	96.9	99.9	83.3
Rodriguez *et al.* (2005)	Decision	3	92.7	-	-	-
Polat & Gunes (2007)	SVM	2	100	100	100	100
Asl *et al.* (2008)	SVM	5	99.1	95.7	99.4	93.5
Owis *et al.* (2002)	*k*-NN	5	-	85	34	-
Minhas & Arif (2008)	*k*-NN	6	99.5	-	-	-

PPV= positive predictive value; Accuracy= correct classification rate

Three studies claimed 100% accuracy (Zhang *et al*. 1999; Kara & Okandan 2007; Polat & Gunes 2007) however these classified only 2 or 3 classes and this "simpler" requirement may have been responsible for over-optimistic results. High values in all performance indices were found using SVM classifiers (Polat & Gunes 2007; Asl *et al.* 2008), suggesting best overall performance. Of poorer performing classifiers with indices of between 80 and 90%, one (Acharya *et al*. 2008) quoted sensitivity of 83.8% and specificity of 100%, reflecting trade-off between sensitivity and specificity, with the accuracy reflecting the lower of the two values. Two studies directly compared classifier performances, using the same data (Acharya *et al*. 2004; Jovic & Bogunovic 2011). Notable among these, Jovic & Bogunovic (2011) compared 7 classifiers, (see Table 2.2).

Jovic & Bogunovic (2011) showed that clustering algorithms performed poorly and that decision tree, Bayesian, ANN, SVM and random forest algorithms, all showing accuracies of > 90%,

appeared more accurate. Random forest, a combination of decision trees in a multi-classifier system, was found to be the best classifier. Despite poor study design, the comparison of classification methods using the same data was useful to act as a guide to study design and suggests a potential advantage of using multi-classifier systems for improving accuracy.

Table 2.2 Comparison of ECG rhythm classifiers. (Data from Jovic & Bogunovic (2011).

Classifier	Accuracy (%)
K-means clustering	49.3
EM clustering	50.8
Decision tree (C4.5)	92.2
Bayes	99.4
ANN	91.4
SVM (linear)	73.5
SVM (squared polynomial)	98.4
RF	99.6

Accuracy= correct classification rate

Table 2.3 Comparison of ANN and fuzzy classifiers. (Data from Acharya *et al*. (2004).

Rhythm Class	Accuracy (%)	
	ANN	Fuzzy
normal	90	92.5
premature ventricular contraction	88	90
complete heart block	81	88
sick sinus syndrome	88.9	90.9
left bundle branch block	85.7	88.9
cardiomyopathy	83.3	86.4
atrial fibrillation	85	88
ventricular fibrillation	81	84
Mean accuracy	**85.3**	**88.5**

Accuracy= correct classification rate

Acharya *et al.* (2004) used RR interval derivatives: spectral entropy; Poincaré plot geometry and largest Lyapunov exponent (presence of chaos) for 8 classes and compared the performances of an ANN with a fuzzy classifier (see Table 2.3).

The fuzzy classifier outperformed the ANN for accuracy in all 8 classes by a mean difference of 3%. The study suggested that fuzzy classifiers had superior performance to ANNs, using features lending themselves to fuzzification, rather than crisp measures.

Given the evidence of these comparative studies, high performing classifiers such as decision trees, Bayesian classifiers, fuzzy classifiers, neural networks and support vector machines were considered for use in this study

2.9.3 Hybrid and Multi-Classifier Systems

Classifier hybrids and multi-classifier systems (MCS) have become increasingly popular as they promise improvement on single classifier performance. Hybrid classifiers have a more sequential configuration whereas MCS commonly rely on a combination strategy for final class allocation. Studies examining the performances of hybrid and MCS algorithms for rhythm classification are summarised in Table 2.4.

Table 2.4 Comparison of accuracies of hybrid (above) and multi-classifier systems (below) (Data from published studies).

Study	Classifier	Classes	Accuracy (%)
Wang *et al.* (2001)	Fuzzy-Kohonen Hybrid	3	97
Linh *et al. (2003)*	Neuro-Fuzzy Hybrid	6	96
Polat *et al.* (2006)	Fuzzy-Artificial Immune Hybrid	unspecified	76
Exarchos *et al.* (2007)	Fuzzy-Expert Hybrid	4	95.8
Leong & Jabri (1992)	ANN-Decision Tree MCS	9	99.6
de Chazal *et al.* (2004)	LDA-LDA MCS	5	84.5
de Chazal & Reilly (2006)	LDA-LDA MCS (Human interaction)	5	95.7
Osowski *et al.* (2004)	SVM-SVM MCS	13	97.4
Ozbay *et al.* (2006)	Fuzzy Clustering-ANN Cascade MCS	10	98.9
Ceylan *et al.* (2009)	FCM-PCA-ANN Cascade MCS	10	99

Accuracy= correct classification rate

In the hybrid classifier studies reviewed, classification was performed for 6 or fewer classes. With the exception of a fuzzy-artificial immune hybrid (Polat *et al*. 2006) accuracies of more than 95.8% were exhibited (Wang *et al*. 2001; Linh *et al.* 2003; Exarchos *et al*. 2007).

For the MCS classifier studies examined, between 5 and 13 rhythm classes were classified. One study (de Chazal *et al*. 2004) showed an MCS which performed with 84.5% accuracy and a second

study by the same group (de Chazal & Reilly 2006) which showed the same system, with an additional phase to correct misclassifications using human expert interaction, improved accuracy to 95.7%. Given the additional stage was non-automatic and consisted of manual reclassification; it was difficult to compare its performance with other studies. With the exception of these two studies, MCS had accuracies of ≥ 97.4%.

2.9.4 ICD Classifiers

In studies of ICD algorithm performance there was wide variation in performance metrics used and important "gaps" in the published data (see Table 2.5).

Table 2.5 Comparative studies of ICD algorithms. (Data from published studies).

Study	Classifier	Classes	Acc. (%)	Se. (%)	Sp. (%)	PPV (%)
Swerdlow *et al.* (2000)	Jewel AF	3	91.6	100	99.9	-
Wilkoff *et al.* (2001)	PR Logic	4	90.3	100	56.1	78.1
Swerdlow *et al.* (2002)	Wavelet	4	-	100	78	61
Kouakam *et al.* (2004)	Ventak AV	4	95	99	89	93.3
Kremers *et al.* (2012)	HemoDx	4	100	100	100	100
Hintringer *et al.* (2001)	Phylax AV	4	-	100	87	-
	Defender IV,	4	-	-	93	-
	Ventak AV III DR	4	-	100	86	-
	Gem DR	4	-	100	83	-
Gold *et al.* (2002)	Dual Chamber VTC	4	99	100	97	99
	SingleChamber VTC	4	98	99	97	99
Theuns *et al.* (2004)	Single Chamber	4	76	100	56	76
	Dual Chamber	4	76	100	60	76
Hintringer *et al.* (2004)	Biotronik Tachos DR	4	-	-	95	-
	ELA Defender IV	4	-	-	99	-
	Guidant Prizm II DR	4	-	-	94	-
	Medtronic Gem DR	4	-	-	93	-
	St Jude Photon DR	4	-	-	92	-
Gold *et al.* (2012a)	Guidant Dual Chamber	4	-	-	-	38.4
	Guidant Single Chamber	4	-	-	-	41.3
	Medtronic Dual	4	-	-	-	51.1
	Medtronic Single	4	-	-	-	52
Gold *et al.* (2012b)	SCICD	4	-	100	98	-
	Dual Chamber ICD	4	-	99	68	-
	Single Chamber ICD	4	-	99	76.7	-

(Key: Acc. = accuracy; Se. = sensitivity; Sp. = specificity)

All but one study emphasised specificity and in several studies, sensitivities were not quoted hence, specificity was used as the main comparator.

Where raw data for confusion matrices was published, reanalysis was performed to provide more complete data. With ICD algorithms, under-detection of arrhythmia has a life-threatening potential and a reasonable expectation is sensitivity near 100%. This was borne out in studies where sensitivities were quoted or calculated (Swerdlow *et al.* 2000; Wilkoff *et al*. 2001; Swerdlow *et al*. 2002; Kouakam *et al.* 2004; Kremers *et al*. 2012; Hintringer *et al.* 2001; Gold *et al*. 2002; Theuns *et al*. 2004).

Many studies were observational and were subject to under-detection error, as arrhythmia detections and performance metrics depended on the algorithm itself detecting the arrhythmia. Between individual studies, specificities for algorithms varied from 56 to 100%. One study claiming 100% performance (Kremers *et al*. 2012) was an experimental algorithm incorporating haemodynamic assessment of arrhythmia augmenting the more usual interval features.

It should be noted that initial studies evaluating the performance of individual algorithms were often manufacturer sponsored and specifics of study design such as manufacturer specific program settings, though scientifically valid, may have shown the algorithm in the best light. Studies which directly compared algorithms using the same data address differences in inter-study design. With a standardised study design, direct comparisons are valid, as errors and biases are common to all. The study by Hintringer *et al.* (2004) provided a comparison of 5 manufacturers' algorithms and showed high specificities of 92-99%, of which the ELA algorithm performed best with a specificity of 99%, which the authors were unwilling to subject to significance testing. Gold *et al.* (2012a) showed statistically significant superiority of Medtronic over Guidant algorithms in preventing inappropriate therapy but did not assess overall algorithm performance.

2.9.5 Feature Selection

Features used as classifier inputs varied between studies. Feature sets for classifiers with accuracies of better than 90% were examined (see Table 2.6).

All the selected studies used more than 5 features, the most common being RR interval (heart rate) and its derivatives, in 7 studies. RR interval derivatives included pattern, regularity, power spectral density and largest Lyapunov exponent. Combined morphology and interval features were used in 2 studies (Leong & Jabri 1992; de Chazal & Reilly 2006) with morphology alone in a further 2 studies (Ozbay *et al*. 2006; Ceylan *et al*. 2009). Wavelet transforms, considered a form of morphological analysis, were used in 2 studies (Minhas & Arif 2008).

Table 2.6 Features used in rhythm classification algorithms. (Data from published studies with accuracies of > 90%).

Study	Features	Feature group 1	Feature group 2
Anuradha *et al.* (2008)	8	RR interval	
Tsipouras *et al.* (2005)	6	RR interval	
Minhas & Arif (2008)	6	RR interval	Wavelet transform
Linh *et al.* (2003)	6	RR interval	
Leong & Jabri (1992)	9	Morphology	Intervals
de Chazal & Reilly (2006)	5	RR, other intervals	morphology
Osowski *et al.* (2004)	13	higher order cumulants	Hermite basis functions
Ozbay *et al.* (2006)	10	Morphology	
Ceylan *et al.* (2009)	10	Morphology	
Jovic & Bogunovic (2011)	11	RR interval	

Use of RR interval features was common to all ICD classifier algorithms. R waves were detected in a number of different approaches, most common being use of wavelet transforms or the Pan-Tompkins method (Pan & Tompkins 1985), combined with threshold detection. Some algorithms added morphology, with or without PP interval and PR relationship and one experimental algorithm added a haemodynamic assessment (Kremers *et al.* 2012).

2.10 Classifier Data Sources

The majority of classifier evaluation studies used surface ECG data from open source databases available from the Physionet PhysioBank (2012) and the UC Irvine Machine Learning Repository (1988). Intracardiac electrograms were used in few of these studies, with the Ann Arbor Electrogram Library at Electrogram.com (n.d.), used by Usher *et al.* (1999). Leong & Jabri (1992) used their own intracardiac electrogram recordings made at invasive electrophysiological (EP) studies.

Observational studies evaluating ICD algorithm performances, of necessity used only intracardiac electrograms available to the device inputs. In their ICD algorithm comparison studies, Hintringer *et al.* (2001 and 2004) used a library of intracardiac electrograms they compiled themselves to compare ICD algorithms using the same data.

2.11 Summary

Review papers suggested a need for large scale randomised studies and recommended standardised statistical analysis. They suggested that hemodynamic sensors might augment

arrhythmia detection and that improved accuracy of diagnostics and patient alerts were important future developments.

Good classification accuracies were found in studies evaluating single classifier technology and (see subsection 2.9.2) suggested that decision trees, fuzzy, neural network and support vector machines classifiers performed well in this domain. There was limited experience with hybrid and multi-classifier systems, but reports suggested performance gain over single classifiers. Hintringer (2004) showed the PARAD+ algorithm in the ELA Defender IV ICD best, with 99% and other manufacturers with 92-95% specificity. Other head-to-head comparisons (Gold *et al.* 2012a and b) showed superior performance of the Medtronic PR Logic over the Guidant Rhythm ID algorithm and of the Cameron SICD algorithm (98% specificity) over conventional ICDs.

The aim of the proposed performance improvement was to raise specificity without cost to sensitivity, for a wide range of rhythms. In the reviewed studies, features used in classification algorithms centred on the ECG and specifically RR intervals and QRS morphology both in classifier technology and ICD algorithm evaluations. Use of additional features appeared lacking other than in early experimentation. These influences aided design and selection of classifiers for inclusion, the development process, methodology and approach to analysis.

Chapter 3 Methodological Approach

3.1 Overview

A standard methodological approach was unclear from the literature review. Two main types of study had emerged, the bench study, evaluating classification algorithms using archived arrhythmia databases or the clinical study, mainly involving implanted devices and their built-in algorithms. Elements of a standard methodology were seen in both, such as commonality of statistical analysis and some indication of available databases and alternative suitable data collection methods. This indicated a need to design a methodology specifically for this study.

The methodological approach was evolutionary, building successive stages according to the findings of preceding stages. It was not always necessary to follow the stages exactly in sequence. The following stages were included:

1. Evaluation of the technologies underpinning pattern recognition and AI classifiers and their suitability for this study (see sections 3.2 to 3.6).
2. Developing a plan to create suitable datasets, decide a partitioning strategy and decide how to deal with missing data (see 3.8 and 3.9)
3. Developing a knowledge-based system, representing the domain and ontology (see section 3.10).
4. Modelling the systems (see subsections 3.10.1, 3.10.2 and 3.10.4).
5. Selecting methods to assess classifier performance (see section 3.12 and subsection 10.2.1).
6. Selecting classifiers and designing a multi-classifier system (see section 3.15).
7. Using a knowledge-base to examine rhythm definitions (see section 3.18).
8. Selecting process models, such as system development, knowledge engineering and classifier development and integrating them into a well-defined process (see section 3.19).
9. Implementing the process model (SDLC) (see Chapter 4).
10. Selection of classifier features (see Chapter 5).
11. Prepare and plan data collection (see Chapter 6).
12. Collect patient data (see Chapter 7).
13. Data preparation and pre-processing (see Chapter 8, sections 8.2 to 8.5).
14. Generation of feature sets and instances (see Chapter 8, section 8.6).
15. Initial data analysis and data quality assessments (see Chapter 9).
16. Design of individual classifiers and classifier tuning (see Chapter 10, section 10.2 to subsection 10.2.7).

17. Iterative development with performance assessments and design modifications (see Chapter 10, subsection 10.2.8 to section 10.9).
18. Select or combine classifiers into a production system (see Chapter 10, section 10.11).
19. Performance testing of the production system (see Chapter 10, subsection 10.11.1).

3.2 Artificial Intelligence in Rhythm Classification

AI has subfields reflecting human intelligence traits, such as: deduction; reasoning; problem solving; knowledge representation; planning and learning. AI is well represented in medicine, particularly the specialist areas of pattern recognition, knowledge representation and ontology, decision support, reasoning under uncertainty, modeling human reasoning and cognitive science, all of which relate to this study.

AI has been used in published examples of cardiac rhythm diagnosis and showed AI algorithm accuracies comparable to commercial systems for cardiac rhythm diagnosis (see Chapter 2, subsection 2.8.12). In commercially available implantable cardiac devices, algorithms are dominated by expert systems using decision trees, often augmented by specific digital signal processing techniques. This research aimed to show the value of AI in the development of new rhythm diagnostic systems with improved accuracy.

3.3 Logical Reasoning in AI

Logical rules can be expressed in natural or formal language and encoding is normally prerequisite for their input into computer systems. In this study, rules in natural language were easily converted to conditional statements for an inductive inference rule-based system. A classifier using an inference engine was included in this study.

3.4 Problem-Solving in AI

Human problem solving is often intuitive, without conscious decision making, so in AI systems, modeling human intelligent behaviour may be advantageous. Heuristic, algorithmic and cognitive techniques can be used to model human problem-solving processes in AI. Heuristic problem solving allows simplification of complex problems, with the limitation that solutions may be of lower quality than more formal methods. Algorithmic or goal-based problem-solving is where a sequence of actions leads to a desirable goal, such as classification of a rhythm.

Algorithmic approaches can help define a process but are less adaptable than human cognitive processes. Simple cognitive models aid visualisation of human cognitive processes and will be used to aid understanding (see subsection 3.10.4 and Fig. 3.3). AI-based medical diagnostic

systems can be considered intelligent agents (Russell *et.al* 2003, p.40). Production algorithm(s) which simulate human cognitive processes were considered here to be cognitive agents.

3.5 Pattern Recognition and Classification

Machine learning is a branch of AI for development of learning algorithms, using data with known properties to predict outcome (Mitchell 1997, p.2). Machine learning fits well with this study, placing cardiac rhythms into classes, based on examples with known diagnosis. Pattern recognition is machine learning with decisions based on recognition of patterns or feature sets, using training data in a learning phase. Patterns are points in *n*-dimensional feature space, with dimension determined by the number of features and similar patterns tending to form clusters. Popular methods include: template matching; statistical classification; syntactic or structural matching and neural networks (Jain *et al*. 2000).

3.5.1 Template Matching

Template matching compares unknown patterns with a stored template. Decisions are typically not statistically or probability based and do not account for variation within a class. It has been used for comparison of cardiac waveforms in previous research (Coggins *et al*. 1995; Swerdlow *et al*. 2002) and will be similarly used in this study.

3.5.2 Statistical Pattern Recognition

Statistical pattern recognition algorithms use a function to divide feature space into decision regions, using decision boundaries, one for each class. The group of techniques includes linear discriminants, principal components analysis, support vector machines and Bayes classifiers amongst others (see Chapter 2, subsections 2.8.2 and 2.8.7).

3.5.3 Syntactic Pattern Recognition

Syntactic or structural matching is pattern recognition where features may be represented by a set of new categorical features allowing the representation of pattern structures with more complex interrelationships between attributes than in statistical classification. ECG waveforms can be represented by a sequence of lines, with normal and abnormal waveforms defined as formal grammars and can be classified by describing them in term of line sequences. In cardiac rhythm diagnosis, this has largely been abandoned in favour of statistical methods (see Chapter 2, subsection 2.8.2).

3.5.4 Unsupervised and Supervised Learning

A form of pattern recognition using unsupervised learning is clustering, with training set data allocated to a class by similarity rather than their known class. Clustering risks creating arbitrary groups of statistically related data points, having no class relationship.

3.5.5 Classification

Classification is pattern recognition using supervised learning. Training sets consist of "patterns" of features, each having membership of a known class, established using human expertise. Different classifiers use different decision functions (see section 3.11). In cardiac rhythm analysis, rhythm examples are reliably classed (diagnosed) by human domain experts, indicating the advisability of a supervised learning approach. Given these influences, cardiac rhythm diagnosis can be considered a classification problem.

3.6 Uncertainty in Classifiers

Probability theory provides a framework dealing with uncertainty in AI applications. Stochastic (non-deterministic) programs use probabilistic methods to solve problems.

With frequency based probability, a hypothesis must be either true or false (0 or 1), conversely Bayesian probability reasons, with premises of uncertain truth, allowing a hypothesis to have a value of between 0 and 1. In this study, pre-test probability for each rhythm diagnosis may be estimated using disease prevalence, where available, post-test probability from a preceding test or an heuristic estimate based on experience. Errors may be introduced into calculations by inaccuracies in the estimation of prior probabilities and by inadequate testing. Bias, error and variance in hypothesis testing is dealt with in more detail in section 3.12.6 and probability in significance testing in sections 3.12.1 to 3.12.4.

The Dempster-Shafer theory of evidence allows the combination of evidence from different sources to produce a degree of belief. Such combination methods have utility in classifier fusion.

3.7 Data Sets for Classifiers

Existing data is unlikely to match the requirements for a specific problem unless the goal is the same. A data set, supporting design and evaluation of a classifier should be appropriate to the task. Existing data sets were considered, to avoid undertaking unnecessary data collection. Most published studies of cardiac rhythm classifier performance use open-source electrocardiogram or intracardiac electrogram data libraries (see Chapter 2, section 2.10). In this study no existing data adequately represented the feature space under investigation so a new data collection exercise was justified.

Data collection should be designed so the feature space provides a good representation of the ontology (see subsection 3.10.2), with all possible pattern combinations included. High dimensional data with a set of features likely to contain all relevant information is advantageous. A balance between training set size and dimensionality needs to be achieved to get the best performance. Dimension reduction during processing by pruning, rather than during data

collection by selection bias helps achieve this. Optimal dimensionality reduction would be a reduced feature set without loss of accuracy.

3.8 Data Partitioning

Classifier input patterns are typically arranged as data arrays, rows representing individual patterns or instances and columns feature variables. One column represents the known class (in the case of supervised learning). To reduce over-training, data re-use is controlled, using partitioning techniques, such as: "hold-out"; *k*-fold cross-validation or bootstrapping. Hold-out works well with large data sets, but rare patterns may not be represented in the training set, reducing performance. *k*-fold cross-validation works well when the data is not large. Bootstrapping works well when the data contains all the available information about the population, allowing the sample to be treated as a population estimate. Bootstrapping is particularly suited to very small data sets. Given that the data set collected in this study was not very small or sufficiently large for hold-out sampling, *k*-fold cross-validation was selected.

3.9 Missing data

AI algorithms which imitate human step-by-step problem-solving are improved with methods for dealing with uncertainty (see section 3.6) and missing data.

Missing data occurs because of dropout or poor data collection methods and research should be designed to minimise missing data (Adèr 2008). Missingness has value on a continuum from missing completely at random (MCAR), through missing at random (MAR) to missing not at random (MNAR) (Graham 2009). In this study, missing data from dropout was considered MAR and from equipment failure MNAR.

For limited missing data, imputation is a typical strategy. Multiple imputations can improve the quality of results but may reduce statistical power (Graham 2009). In this study, for small percentages of data loss, interpolation will be used to impute data.

3.10 A Knowledge-Based System

In AI systems, a human expert is central to knowledge acquisition and in supervised learning so the processing of knowledge is considered important. Models were used to represent the domain and ontology. Evaluation of the domain used a knowledge engineering approach, guiding feature selection.

3.10.1 Domain

The domain was defined as cardiac rhythm diagnosis, a human specialist discipline, with domain experts. To gain deeper understanding of the domain, an analysis was conducted, focused on two

areas relevant to this study: definitions of different cardiac rhythms and the processes involved in cardiac rhythm diagnosis. The domain was illustrated using models. A knowledge acquisition approach was used, with a combination of knowledge transfer and modelling techniques.

Landmark scientific papers, joint working group reports and consensus guidelines were used as source material for knowledge acquisition (Buxton *et al*. 2006; Blomström-Lundqvist *et al*. 2003; Epstein *et al*. 2008; Brugada *et al*. 1991; Saoudi *et al*. 2001; Fuster *et al*. 2011; Bonow *et al*. 2012; Surawicz *et al*. 2009). These formed a body of knowledge from a large number of domain experts, distilled into detailed statements, likely to be superior in scope and accuracy compared with interview results from a single domain expert. Human domain expert activity was limited to accuracy checking and validation tasks.

Major rhythm class descriptions and their diagnostic criteria were extracted into a knowledge base (see Appendix E). The natural language of this knowledge base was translated into progressive structured representations to model cardiac rhythms and their relationships.

A hierarchical taxonomy illustrated disease relationships in a structure recognisable to domain experts (see Fig. 3.1). Cardiac rhythms were represented by six parent classes, eighteen child classes and twenty sub-classes, within which there were 31 different rhythms (see Table 3.1). The rhythms listed were intended to be clinically useful, rather than exhaustive, whilst remaining manageable. Variants of rhythms and rare or difficult to diagnose rhythms were not included. The knowledge base generated (see Appendix E) was intended for re-use during feature selection and classifier development.

3.10.2 Ontology

The hierarchical taxonomy was used as the basis for a relationship diagram used as a simple ontology. The model used was based on Gene Ontology, which places terms as nodes and relations between terms as arcs (see Fig. 3.2).

3.10.3 Knowledge Engineering

Ontology development is the precursor to its use for problem-solving and knowledge engineering provides a systematic method for this. A three stage knowledge engineering approach of knowledge acquisition, knowledge representation and implementation (Edwards 1991) was adopted in this study.

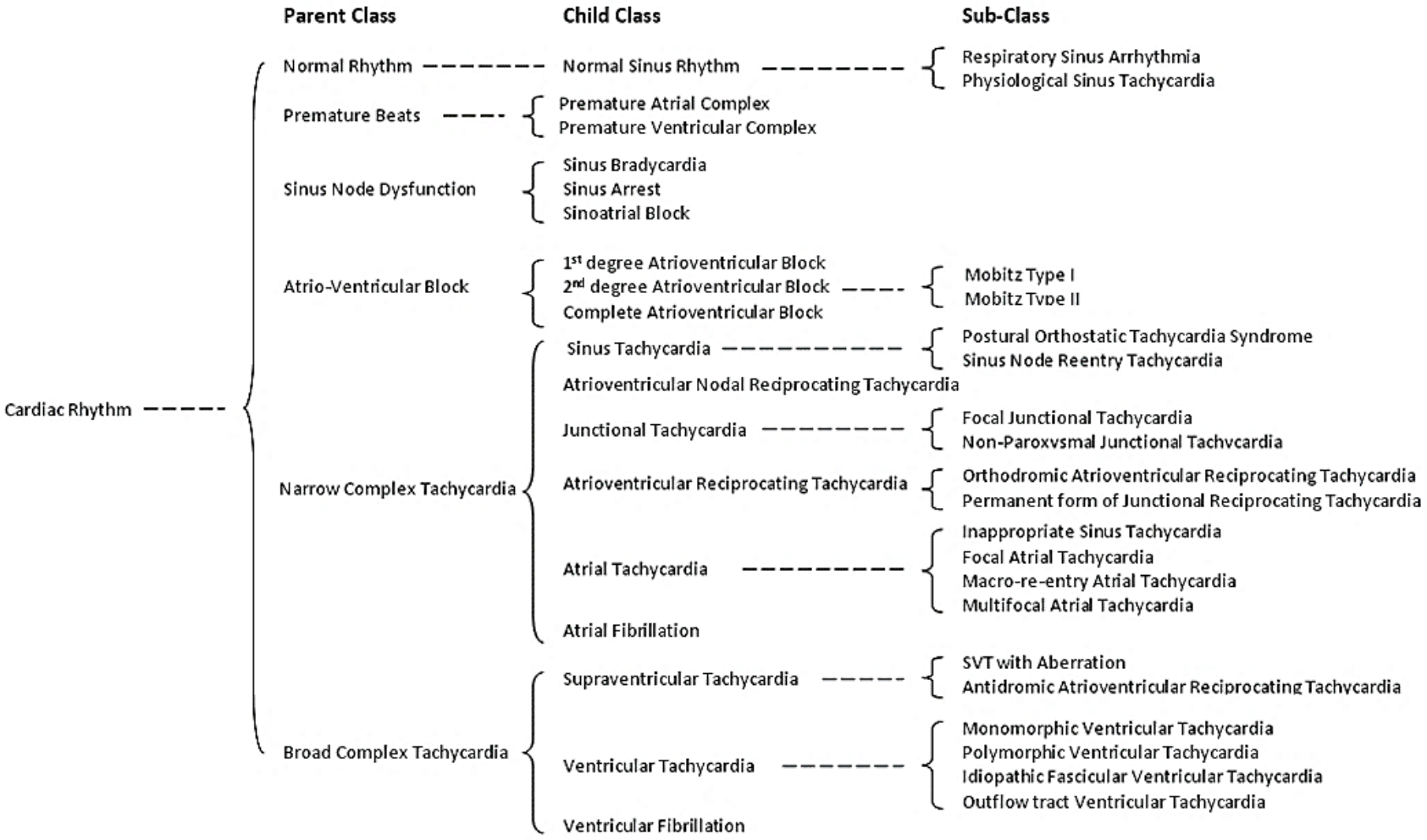

Figure 3.1 A hierarchical taxonomy of cardiac rhythms. Reading from left to right, with increasing granularity for different specific rhythms.

Table 3.1 Listing of rhythm types and their acronyms. (Parent classes in bold type).

Rhythm Description	Acronym
Normal rhythm	
Normal sinus rhythm	NSR
Respiratory sinus arrhythmia	RSA
Physiological sinus tachycardia	ST
Premature beats	
Premature atrial contraction	PAC
Premature ventricular contraction	PVC
Sinus node dysfunction	
Sinus bradycardia	SB
Sinus arrest	SA
Sino-atrial block	SAB
Atrio-ventricular block	
First degree atrio-ventricular block	1HB
Second degree atrio-ventricular block (Mobitz type I)	2HB
Second degree atrio-ventricular block (Mobitz type II)	2HB2
Complete atrio-ventricular block	CHB
Narrow complex tachycardias	
Postural orthostatic tachycardia syndrome	POTS
Sinus node re-entry tachycardia	SNRT
Atrio-ventricular nodal reciprocating tachycardia	AVNRT
Focal junctional tachycardia	FJT
Non-paroxysmal junctional tachycardia	NPJT
Orthodromic atrio-ventricular reciprocating tachycardia	OAVRT
Permanent junctional reciprocating tachycardia	PJRT
Inappropriate sinus tachycardia	IST
Focal atrial tachycardia	AT
Macro-re-entrant atrial tachycardia	MRAT
Multifocal atrial tachycardia	MAT
Atrial fibrillation	AF
Broad complex tachycardias	
SVT with aberration	SVTab
Antidromic atrio-ventricular reciprocating tachycardia	AAVRT
Monomorphic ventricular tachycardia	VT
Polymorphic ventricular tachycardia	PVT
Idiopathic fascicular ventricular tachycardia	IFVT
Outflow tract ventricular tachycardia	OTVT
Ventricular fibrillation	VF

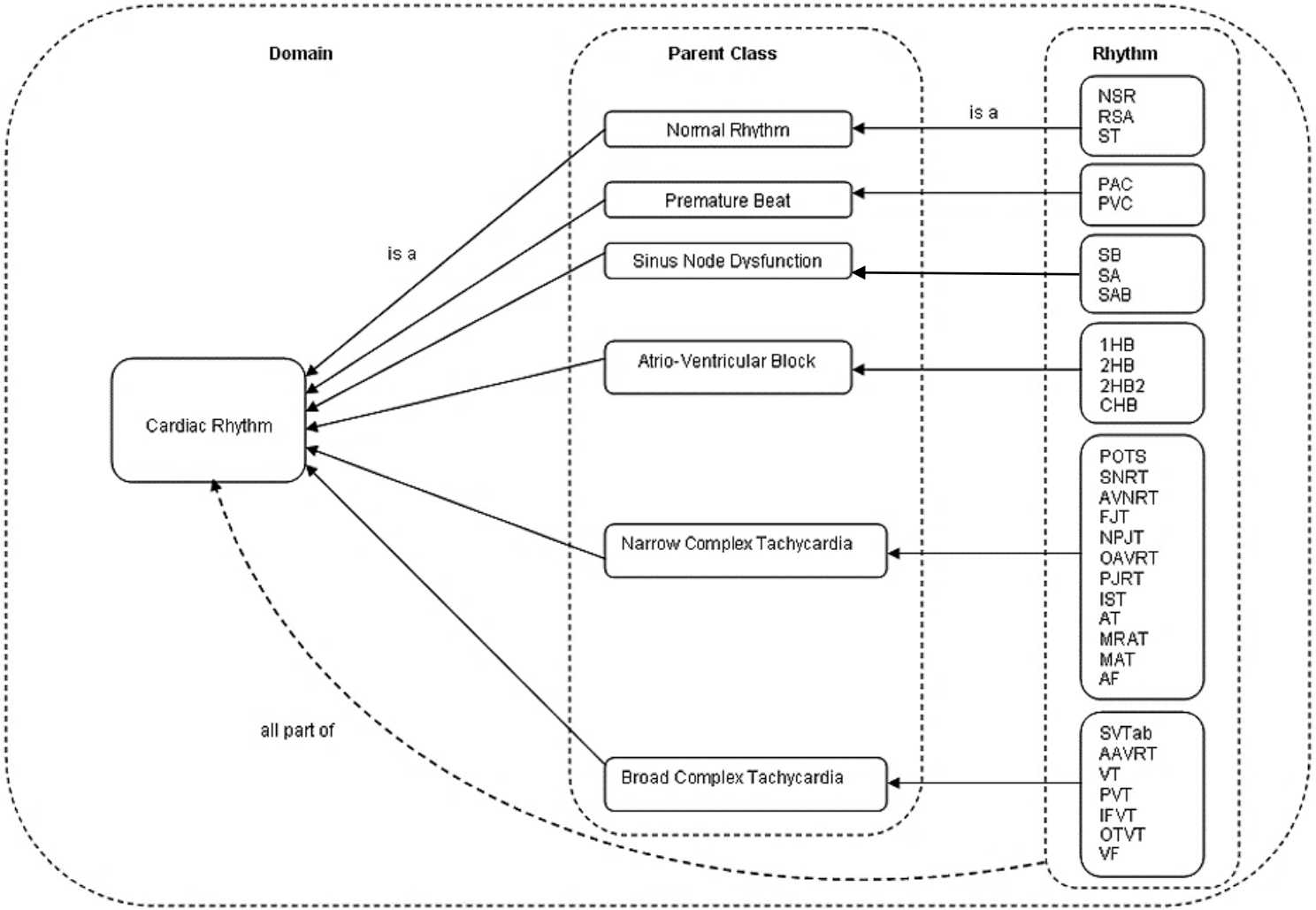

Figure 3.2 Cardiac rhythm ontology diagram. The hierarchical taxonomy of Fig. 3.1 can be shown as nodes and arcs in a directed graph, as a representation of an ontology.

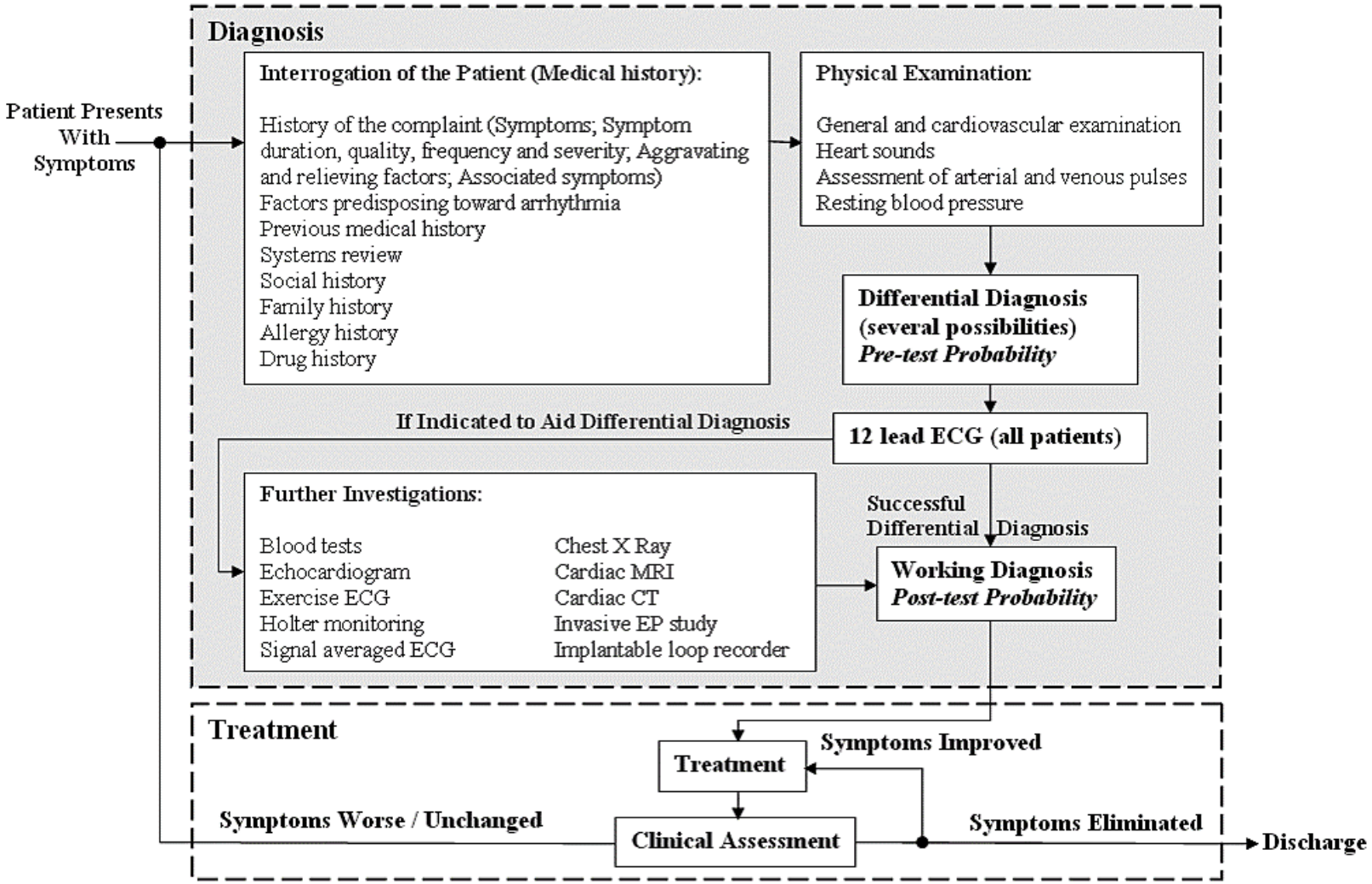

Figure 3.3 A cognitive model of cardiac rhythm diagnosis by a human cardiologist (See text for explanation).

3.10.4 A Cognitive Model of the Diagnostic Process

A synthesis of the human diagnostic process that would aid its understanding was produced. The relationships between rhythms in the domain and ontology models did not represent diagnostic processes well so a specific representation was produced. Knowledge sources (see subsection 3.10.1) were re-used to analyse the descriptions of diagnostic processes and a flow diagram was prepared and subjected to validation by a human domain expert (consultant cardiologist). This process was then condensed into a simple cognitive model (see section 3.4 and Fig. 3.3).

3.11 Types of AI Classifier

Medical diagnosis lends itself well to binary classification, for example disease presence or absence. When selecting classifiers, those having established utility will be considered. High performing classifiers were decision trees, Bayesian, fuzzy, neural networks and support vector machines. Hybrid and multi-classifier systems were also successful in published reports; particularly those including a fuzzy, neural network, expert system or support vector machine classifier unit (see Chapter 2, sections 2.8.10 and 2.8.11).

3.12 Measuring Classifier Performance

The standard method for assessment of classifier performance measures, uses test results presented in a confusion matrix or contingency table.

3.12.1 The Gold Standard Test

In medicine, diagnostic tests are considered classification exercises and performance is measured by comparing classifier outputs to a "gold standard" test, using test data. A gold standard test is the best practical test available that provides an unequivocal diagnosis. Gold standard test results provide class labels to a supervised learning process. In this study this is taken to be human domain expert diagnosis, a specialist cardiac electrophysiologist. This study did not seek to evaluate disagreement between domain experts.

3.12.2 Confusion Matrices and Contingency Tables

Results of classification testing for *n* classes may be expressed as an *n*-way confusion matrix which may then be tested for inter-relationships.

Diagnostic tests commonly use 2-way contingency tables, with test results tabulated against the gold standard test. The tables have four categories: true positive (TP), positive tests result where the condition was present; false positive (FP), positive tests but the condition was absent; false negative (FN), negative tests where the condition was present and true negative (TN), negative tests and the condition was absent. Each individual rhythm diagnosis requires a 2–way contingency table to generate performance indices (see Table 3.2).

Table 3.2 A 2-way contingency table for a diagnostic test.

		Gold Standard Test	
		Condition Present	**Condition Absent**
Test under Evaluation	**Positive Test**	TP	FP
	Negative Test	FN	TN

For each possible class of rhythm diagnosis, represented as "condition present" data, the corresponding "condition absent" data is the summed total of the alternative condition.

3.12.3 Measures of Diagnostic Test Performance

Tests of diagnostic test performance in the medical domain predominantly quote sensitivity, specificity and accuracy. Other useful indices are: false positive and false negative rates; positive and negative predictive values; likelihood ratios; error; Youden's index; diagnostic odds ratio; F_1 score (Jardine & van Rijsbergen 1971); Cohen's kappa (Cohen 1968) and Yule's Q (Yule 1912). In machine learning, Pearson's phi (ϕ) (Pearson 1899), precision and recall are also used. Receiver operating characteristic (ROC) curves illustrate test performance, plotting specificity against sensitivity for varying decision threshold values. The popular area under the curve (AUC) for a ROC is one of the most used measures of classifier performance. AUC was not used in this study as the gold standard decision threshold is considered fixed, providing only one data point, rendering an ROC curve of limited use.

Baldi *et al.* (2000) found that all performance information is contained in the four contingency table values and indices based on less than four values tend to lose information. Performance indices using all four numbers are: sensitivity and specificity (used together); likelihood ratios; error; diagnostic odds ratio; Youden's index (Youden 1950); Cohen's kappa; Yule's *Q* and AUC.

3.12.4 Weak and Strong Learners

A correct classification rate (*CCR*) of > 0.5 represents weak learners well but the transition value to strong learner is ill-defined. ϕ represents correlation well, with 0 meaning correlation no better than random; values just above 0 representing weak learners; values approaching 1 meaning strong learning; +1 for perfect accuracy or −1 for all classifications being incorrect.

3.12.5 Consistency and Generalisability

Classifiers correctly classifying all training examples are considered consistent. Generalisability is where an algorithm produces a plausible output for any input (Bishop 2006, p.2). A rote learner which matches patterns identical with learned examples may be consistent but not generalisable. *CCR* and φ with values of 1 represent perfect consistency. Accuracy tests performed on classifiers will emphasise generalisability, reflecting real world performance.

3.12.6 Error, Bias and Variance

Hypothesis testing is subject to type I error (α), where a true null hypothesis is rejected or type II error (β), where a false null hypothesis fails to be rejected and α is the significance level of a test and β is related to power of the test (1 - β), with power equal to the sensitivity of the test. Type I errors are represented by false positives (FP) and type II by false negatives (FN).

Classifiers with high bias are systematically incorrect when trained with different training sets and high variance if they predict different classes for a particular pattern. Classifier error is related to the sum of bias and variance. Type I and II error, bias and variance will be discussed for the production system.

3.13 Comparing Classifiers

Tests computing *P* value indicate a level of statistical significance and are considered more useful for comparisons, as they are readily performed on 2-way contingency tables. The significance level (α) is the probability below which the null hypothesis will be rejected (usually 5%) with *P* values < α considered statistically significant.

Standard error (σ_X) and confidence interval (*CI*) analysis of performance indices (see subsection 3.12.13) expressed as a proportion were estimated using the normal approximation method of the binomial confidence interval, using a 95% level (CI_{95}) to quantify imprecision (3.1 and 3.2).

$$\sigma_X = \sqrt{\frac{X(1-X)}{n}} \tag{3.1}$$

$$CI_{95} = 1.96\sqrt{\frac{X(1-X)}{n}} \tag{3.2}$$

3.13.1 Bayesian Analysis

Naïve Bayes classification uses algebraic calculations on contingency table data, where final class is that with the highest posterior probability ($P(H|E)$). The odds-ratio form of Bayes' rule applies to classification and contingency table analysis, allowing calculation of posterior probability ($P(H|E)$) from odds, as a result of new data provided by the results of tests. Prior probability

($P(H)$) for the hypothesis (H) is estimated using prevalence of the condition in the gold standard. Prior odds may be calculated from prior probability. Using the odds rule, posterior odds for positive and negative test results are calculated from positive likelihood ration (LR+) and negative likelihood ratio (LR-) indices and then posterior odds are converted to a posterior probability ($P(H|E)$). Posterior probabilities generated in this way for each classifier allow comparisons.

3.14 Classifier Selection

Diagnostic tests commonly use 2 x 2 contingency tables, with test results tabulated against the gold standard test (see subsection 3.12.2 and Table 3.2). Analysis of the table provides understanding of the test. Fisher's exact test is valid for all sample sizes and as complex computation is now possible has largely superseded the chi-squared test, provided row and column totals are fixed.

Best performance is the standard indicator for the individual best classifier (Ruta & Gabrys 2005) (see section 2.9.2). Agreement statistics, such as sensitivity and specificity, are widely used to the describe performance of arrhythmia classifiers, with differences between classifiers evaluated by significance testing. Other performance indices are commonly used, such as the correct classification rate (*CCR*) - a derivation of sensitivity and specificity and Cohen's kappa (κ). κ provides an index of agreement between two "raters" and may be used for to rank classifiers as an alternative to statistical significance testing. κ is widely used to represent overall classifier performance as it incorporates all four components of the contingency table. Where there is low prevalence *CCR* can be an over-optimistic measure of classifier performance and κ is also less reliable. It is recognised that no single performance index adequately represents classifier performance and that use of a wide range of agreement measures is advisable.

3.15 Designing a Multiple Classifier System.

Multiple classifier systems (MCS) aim to improve overall accuracy, combining weak learners to create a strong learner, on the assumption that they are the best possible solutions. Synonyms are: "hybrid systems", "ensemble classifiers", "combinations of classifiers" and "committee machines". In MCS design a number of key factors guide strategy, such as the accuracy and diversity of individual and combinations of classifiers; classifier selection strategies and the choice of combiner. Decision optimisation methods select and optimize the combiner for a fixed ensemble of base classifiers. Coverage optimisation methods create diverse base classifiers assuming a fixed combiner.

3.15.1 Accuracy and Diversity

With a perfect single classifier there are no errors and an MCS is not needed. A classifier that does make errors can be complemented with another classifier, which makes errors on different

objects. Diversity of classifiers in an MCS serves to correct for systematic misclassification of one classifier with correct outputs from another. The relationship between diversity and accuracy in MCS is unclear and intuitive methods for inducing diversity when building MCS appear to work well. Measuring diversity and using the results to guide building systems is less successful. Misclassified instances having different labels for different classifiers are a simple sign of diversity.

Various diversity measures have been proposed, the most widely used examples are: ϕ; Yule's Q statistic; the disagreement measure (*dis*) (Skalak 1996), which measures the ratio of number of observations on which one classifier is correct and the other is incorrect to the total number of observations; the double-fault measure (*DF*) (Giacinto & Roli 2001a), which measures the proportion of cases misclassified by both classifiers; Kohavi–Wolpert variance (Kohavi & Wolpert 1996), which is a bias-variance decomposition of the error of a classifier; inter-rater agreement (κ); the measure of difficulty (μ) , where for a discrete random variable for a random sample from the training set, μ is the variance of this variable over the whole training set, so with decreasing diversity μ increases (Hansen & Salamon 1990) and generalised diversity (*GD*) (Partidge & Krzanowski 1997). Similar to Double-Fault Measure, *GD* is a coefficient representing the ratio of probabilities of one randomly picked classifier failing against both classifiers failing.

Studies advocate producing a pool of classifiers followed by selection of the most diverse and accurate. Giacinto & Roli (2001a) used the double fault measure *DF* (probability of both classifiers being incorrect) and the *Q* statistic, to form a pairwise diversity matrix for a classifier pool and then select classifiers that are least related. Tang *et al.* (2006) demonstrated explicit relationships between diversity measures and margin maximization, showing experimentally they were all ineffective for constructing ensembles with good generalisation.

3.16 Classifier Combiners

Classifier combination strategies may be categorised into 'selection', where specialised classifiers are dedicated to parts of the feature space or 'fusion', where classifiers are combined using a variety of algorithms (Kuncheva 2004). Classical combination strategies include voting, boosting, bagging and stacking. There are many others including mixture-of-experts (ME); cascade and hybrid systems. A combiner may also consist of another layer of classification, with base classifier outputs used as inputs for a classifier used as a combiner.

In a "parallel" combination, identical input patterns are applied to each of *n* classifiers and each outputs a class label. These labels become inputs for a combiner, where a decision rule is applied to allocate the final label. As an example, typical voting decision rules are: simple majority (class with the most votes) (see Fig. 3.4); overall majority (>50%); two-thirds majority or unanimity.

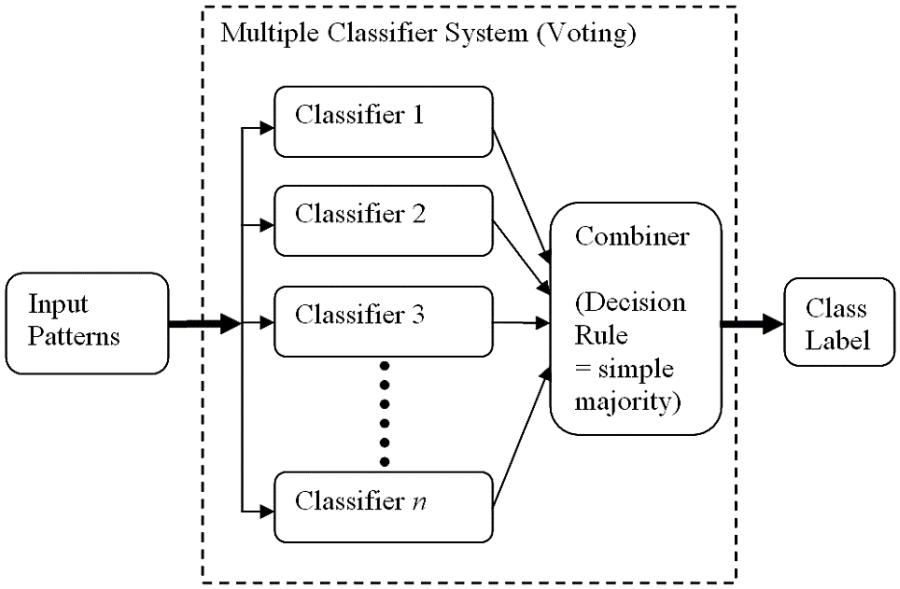

Figure 3.4 Illustration of a simple multi-classifier system (MCS), using a voting combiner. Data for classification is input (left) to each of various classifier units, each independently allocating a class then using a combination strategy (centre) final class label is decided (right).

Weighted voting allocates more competent classifiers a greater weight based on lower classification error, such as misclassifications during training, before application of the decision rule.

Boosting is a combination strategy that iteratively adapts diverse weak classifiers, arranged in a series and is effective for classification of patterns that are difficult to learn. AdaBoost is the most commonly used boosting algorithm (Freund & Schapire 1995). Bagging (bootstrap aggregation) creates multiple data sets by sampling with replacement (Breiman 1996) and the resultant models combined using a combiner. Stacking (Wolpert 1992) consists of layers of classifiers. Classifiers at higher layers learn misclassification errors of classifiers immediately below hence minimising generalisation error. A cascade is a series of classifiers with one classifier active at a time. Only misclassifications are passed to the next classifier in the cascade (Kuncheva 2004). ME uses a separate classifier to control "gating" or "soft" switches to activate specific combinations of "expert" classifiers for each input pattern.

3.16.1 Combining Specialist Classifiers

Specific classifiers may perform well in regions of competence. These can be arbitrary divisions of the feature space, provided estimates of competencies may be made. It is possible for the classifier with the highest accuracy for the whole feature space to be eliminated or a classifier to

be nominated for more than one region. Related features may be combined in groups or subsets and may be separately classified for subsequent combination.

3.17 Current ICD algorithms

Arrhythmia detection in ICD algorithms starts when heart rate enters a tachycardia or bradycardia zone and persists for a number of beats. Diagnosis using heart rate alone is acceptable for some rhythms, such as ventricular fibrillation, but leads to a high incidence of misdiagnosis. ICD algorithms use additional criteria, such as sudden onset; rate stability and sustained high rate and a morphology discriminator to reduce misdiagnosis. Dual chamber ICDs offer atrial signal analysis and atrio-ventricular timing relationships to assist in rhythm determination. ICD manufacturers use this information differently. One algorithm has the form of a decision tree (see Fig. 3.5).

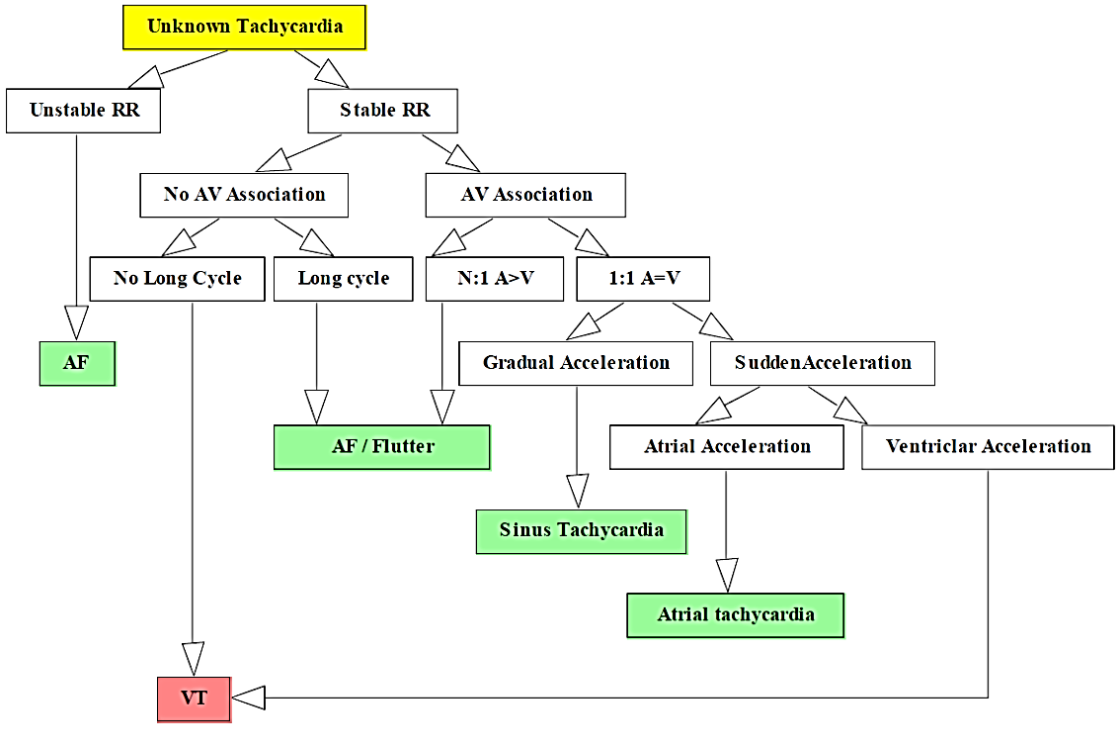

Figure 3.5 A schematic of an ICD decision tree algorithm. (From Aliot *et al.* 2004). A representation of the Sorin PARAD+ ICD algorithm, in the form of a decistion tree. Reading from top to bottom, each node requires satisfaction to proceed down the tree to a diagnosis.

Commercial ICDs rhythm discriminators nominal settings are shown in Table 3.3. Although there is inter-manufacturer variation in the exact method of detecting an arrhythmia, there is common ground. A method common to several algorithms is to assess heart rate on a continuous basis until

the rate increases above a set limit. Two further criteria are then activated: percent "onset", where the heart rate immediately after a detection is compared with immediately before and a percentage change is required to satisfy the criterion; then "stability", where the heart rate after a detection is required to be stable to with a percentage of variation.

Table 3.3 Nominal settings in commercial dual-chamber ICD algorithms.

	Discriminator							
Manufacturer	VT count (beats)	VF count (beats)	Onset (%)	Stability (msec)	Sustained VT	Morphology Threshold (%)	AF Threshold (bpm)	AV
Biotronik	26	8 / 12	20	24/12%	Off	N/A	200	On
Boston	2.5sec	1sec	9	20	3mins	90	170	On
Cameron	18 / 24	18 / 24	N/A	N/A	N/A	50	N/A	N/A
Medtronic	16	18 / 24	81	Off	Off	70	171	On
Sorin	6 / 8	6 / 8	25	63	N/A	N/A	N/A	On
St. Jude	12	12	100ms	80	Off	60	N/A	Off

The Medtronic algorithm (Medtronic, 2010), which is common to all their ICD models will be considered typical. In this algorithm, "onset" compares heart beat cycle lengths, using a rolling average of four beats, between the current average and that of the preceding four intervals, with a programmable percentage defining a sudden onset. Within the pre-set VT heart rate zone, on the third consecutive VT event, "stability" compares the current ventricular interval to each of the previous three and defines the rhythm as unstable if the difference is greater than a programmed value. Satisfaction of these criteria allows a preliminary diagnosis and it is this diagnosis which will be used in this study, for comparative purposes.

3.18 The Basis of Cardiac Rhythm Analysis

The ECG is at the core of analysis of cardiac rhythm. An authoritative text, Wagner (2001, p.44) suggest systematic analysis of the ECG using the following guide:

1. Rate and regularity.
2. P wave morphology.
3. PR interval.
4. QRS-complex morphology.
5. ST-segment morphology.
6. T-wave morphology.

7. U-wave morphology.
8. QT_c interval (the corrected QT interval).
9. Rhythm

Wagner (2001, pp.236-237) also advocates consideration of the ECG features of mechanism and site of origin. Wagner (2001) defines bradyarrhythmia as any rhythm with a rate <60 beats/min, and "tachyarrhythmia" any rhythm with a rate > 100 beats/min. Other arrhythmias do not alter the rate beyond normal limits. Wagner (2001) also suggests that the site of origin of arrhythmias may be determined by rate and atrio-ventricular relationship. It is argued that if the relationship is 1:1, the rhythm originates in either atria or ventricles, if *n*:1 the rhythm originates in the atria or if 1:*n* the rhythm originates in the ventricles. When associated rhythms have differing rates, the rhythm is named for the originating chamber or when rhythms are dissociated, both rhythms should be named. This conventional analysis indicates that rhythm analysis depends on individual beat characteristics, including intervals and sequences. Electro-mechanical features such as electrogram morphologies and the effect of the beat on stroke volume are likely to add to this.

3.18.1 Rhythm Change – Guidelines

Consensus guidelines (Buxton *et al*. 2006; Blomström-Lundqvist *et al*. 2003; Epstein *et al*. 2008; Brugada *et al*. 1991; Saoudi *et al*. 2001; Fuster *et al*. 2011; Bonow *et al*. 2012; Surawicz *et al*. 2009) show that many diagnoses (see Appendix E) are based on the intervals between beats. Other key features are beat morphology of either P or QRS complex in the form of duration, axis, morphology and presence or absence of major waves.

Clues within these consensus guidelines allow estimation of the intervals required to show a rhythm change, its persistence and stability. Criteria are: 10% or 120 msec (RSA); “shorter coupling” interval (PAC, PVC); regular ≤ 2% cycle-to-cycle variation (MRAT); irregularly irregular (AF) and ≥ 3 ventricular beats at > 100bpm (VT). No guideline specifies an interval count for detection and diagnosis of atrially based rhythms.

3.18.2 Number of Beats to Diagnose Rhythm

Given the widespread use of electrogram intervals in this domain (see Chapter 2, subsection 2.9.5), consideration was made of the required sample length to achieve diagnosis, using the electrogram intervals.

Guidelines (see subsection 3.18.1 and Appendix E) suggest that for rhythm change to have occurred, an interval change of 10% or 120ms must occur. Detection requires at least two ventricular intervals (three beats), with R_0 the first beat (onset) of the altered rhythm, R_{-1} and R_{-2} the two preceding beats. To indicate regularity, at least two succeeding intervals should conform

to the regularity criterion of ≤ 2% variation so 5 ventricular beats are required, 2 preceding (R_{-1} to R_{-2}) and 2 succeeding (R_1 to R_2) the first beat (R_0) of each altered rhythm. This can be represented in a conduction diagram (after Langendorf *et al.* 1944) (see Fig. 3.6).

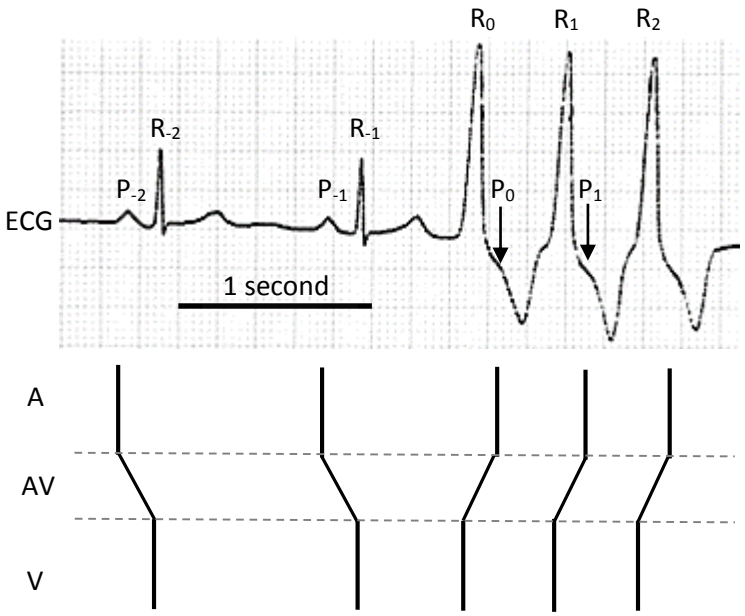

Figure 3.6 An illustration of a 5 beat sequence to diagnose rhythm, indicated by clinical guidelines. Top trace is ECG (read from left to right) showing two normal beats, followed by a ventricular "triplet" arrhythmia. The lower diagram represents atrio-ventricular conduction sequence, with the upper bar (A) atrial depolarisation; the lower bar (V) ventricular depolarisation; the oblique line joining them atrio-ventricular conduction (AV); double lines represent "blocked conduction, where atrio-ventricular conduction is absent in either direction. R_{-2} and R_{-1} represent the two ventricular beats preceding the onset of the arrhythmia, R_0 represents the first ventricular beat of the arrhythmia, R_1 and R_2 the following beats, with P waves labelled in the same way for atrial beats. Note that P_0 and P_1 (arrowed) are not easily visible on the ECG trace.

In a typical ICD algorithm, such as Medtronic PR logic (Medtronic 2010), detailed in section 3.17, at least 5 beats (or 4 inter-beat intervals) within a detection zone, are required to classify an episode as non-sustained VT. Thus, 10 ventricular beats are required to satisfy the requirement for a rolling average of 4 beats before and after arrhythmia onset: 5 preceding (R_{-5} to R_{-1}) the first beat (R_0) of each altered rhythm and 4 succeeding (R_1 to R_4) (see Fig. 3.7).

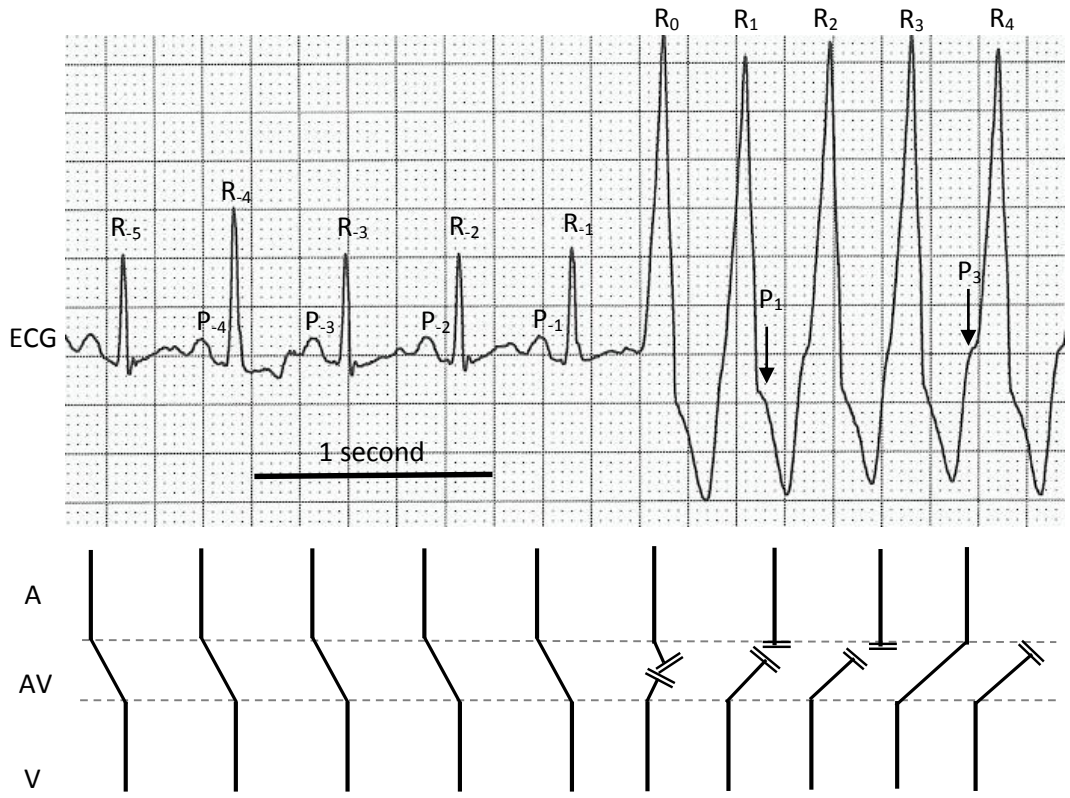

Figure 3.7 An illustration of a 10 beat sequence to diagnose rhythm in an ICD algorithm. Top trace is an ECG (read from left to right) shows five normal beats but at a heart rate of 130 per minute (sinus tachycardia), followed by five beats of ventricular tachycardia. The lower diagram is thr atrio-ventricular conduction. Labelling conventions as for Fig. 3.6. The positions of P_1 and P_3 are arrowed; positions of P_0 and P_2 were not easily visible on the ECG trace and were inferred.

In this study, to satisfy the requirements of the guidelines and those required for a conventional ICD algorithm (see section 3.17), a 10 beat segment will be the basis for the data collection and pre-processing phases of this study, allowing a "like-for-like" performance comparison between a new algorithm and ICD algorithms, using data as similar as possible.

Figure 3.7 illustrates the well-recognised difficulty of demonstrating an important diagnostic feature often associated with ventricular tachycardia, that of ventriculo-atrial dissociation (see Appendix E). The use of intracardiac electrograms is known to facilitate atrial rhythm detection and would enable detection of this feature, so their use is perhaps indicated in this study.

3.19 System development

A project strategy for classifier development was developed, to include necessary steps for design of a classifier, such as feature selection, data collection, pre-processing, classifier testing and the

possibility of iterative development. A method using informatics techniques was preferred. This would include an algorithmic approach, a selection process for technologies and analytical techniques suited to the problem. Consideration of study design and methods of analysis recognisable and familiar to medical practitioners, particularly cardiologists, which might encourage widespread acceptance of the study conclusions was preferred.

A systems engineering approach was successfully used in in a preliminary study (Bostock 2004), using a "single-pass" process model (Jenkins 1969, Flood & Carson 1988). Based on this, a systems development life-cycle model was considered suitable to guide development.

3.19.1 The Incremental Build Model

An incremental build systems development life-cycle (SDLC) model was selected, for its simplicity and practicality. The model is a modification of the waterfall model which allows prototyping with iterative design improvements. There are six stages: user requirements; system specification; system design; increment verification; prototype operation and maintenance, with iterations for major redesign or design modification.

In this study redesigns were limited to encourage rapid prototyping and evaluation, rather than allowing frequent minor modifications. Given that SDLC's are intended for large-scale systems design projects, and that a prototype production system was the objective in this study, the model was used as a general guide.

3.20 Summary

AI and its subfields relevant to this study were briefly explored with concepts of supervised and unsupervised learning and classification introduced. Cardiac rhythm analysis, able to be classed by human domain experts was considered to be a classification problem.

Uncertainty in classifiers was summarised, with the use of probability and hypothesis testing in classification being discussed. Issues in classifier design applicable to this study, such as data dimensionality, collection strategies, set allocation and missing data were reviewed. Cross-validation sampling for data set allocation and imputation to deal with missing data were chosen.

The importance of knowledge in AI, with concepts of domain and ontology, was outlined. An understanding of the cardiac rhythm domain was aided by modelling. A hierarchical taxonomy was constructed as a knowledge representation and refined into an ontology, demonstrating the hierarchical nature of rhythms and their relationships. The diagnostic process was illustrated as a cognitive model, with a view to its use in system design.

The choice of AI classifiers emphasised those with known utility in this field: Bayesian classifiers, fuzzy classifiers, decision trees, neural networks and support vector machines. Classifier performance measurement, the need for a "gold standard" and the use of confusion matrices to analyse results were introduced. Indices derived from these analyses were used as performance measures. Categorisation of classifiers as weak or strong learners and their consistency, generalisabilty, error, bias and variance were outlined. To select classifiers, statistical hypothesis testing using the Fisher's exact test with the Bonferroni correction at a 5% level of significance was used and Cohen's kappa (κ) selected as a simple index of agreement. In choice of classifiers for a multiple classifier system, measurement of accuracy and diversity and the trade-off between them was examined. Classifier combination methods such as voting, boosting, bagging, stacking, cascade of classifiers and ME schemes were overviewed.

ICD algorithms from major manufacturers and their common and unique features were summarised. Clinical rhythm diagnosis techniques were examined and an outline of data required to achieve a diagnosis made. Both evaluations were to ensure inclusion of key features in any algorithm.

A system development life-cycle approach to design, using an incremental model was selected for algorithm development.

Chapter 4. Implementing the System Development Process

4.1 Overview

The incremental build systems development model, was tailored to incorporate additional processes needed for classifier development: feature selection; data collection; pre-processing; classifier testing and iterative development, as illustrated in the incremental model.

In implementing the first two stages of the incremental model (see Fig. 4.1), a user requirement and system specification were prepared based on the study aims and objectives. System design consisted of the preparatory stages of design followed by successive classifier iterations. During the implementation phase, to aid rapid prototyping and avoid multiple iterations of similar systems, iterations were optimised prior to testing. Increment (iteration) verification used the classifier performance measures introduced in Chapter 3, section 3.12. Prototype operation and maintenance were important only should new information forcing a design change become available prior to satisfaction of the user requirement.

Each increment implies iterative improvement. In this study there was no guarantee of successive improvement, so increments were referred to as iterations.

4.2 User Requirement

In system development, achievements of the goals set in a user requirement indicate when no further iterations are required and the final prototype becomes the production system.

User requirements were elicited from the study objectives (see Chapter 1, section 1.7). These were re-focused on the systems and diagnostic methods which would provide an accurate cardiac rhythm classification algorithm. Elimination of duplication and use of systems development terms condensed these to three goals:

1. Iterations will use AI based classifiers, knowledge management and analytic techniques.

2. Final production system will have a higher performance than existing production

4.3 System Specification

A system specification was developed, containing the information required to describe the system, including function, development, information inputs and outputs.

Modelling is a legitimate means of communicating a system specification and a Hatley-Pirbhai system context diagram was chosen for this (Hatley *et al*. 2000, p.434; Pressman & Ince 2001,

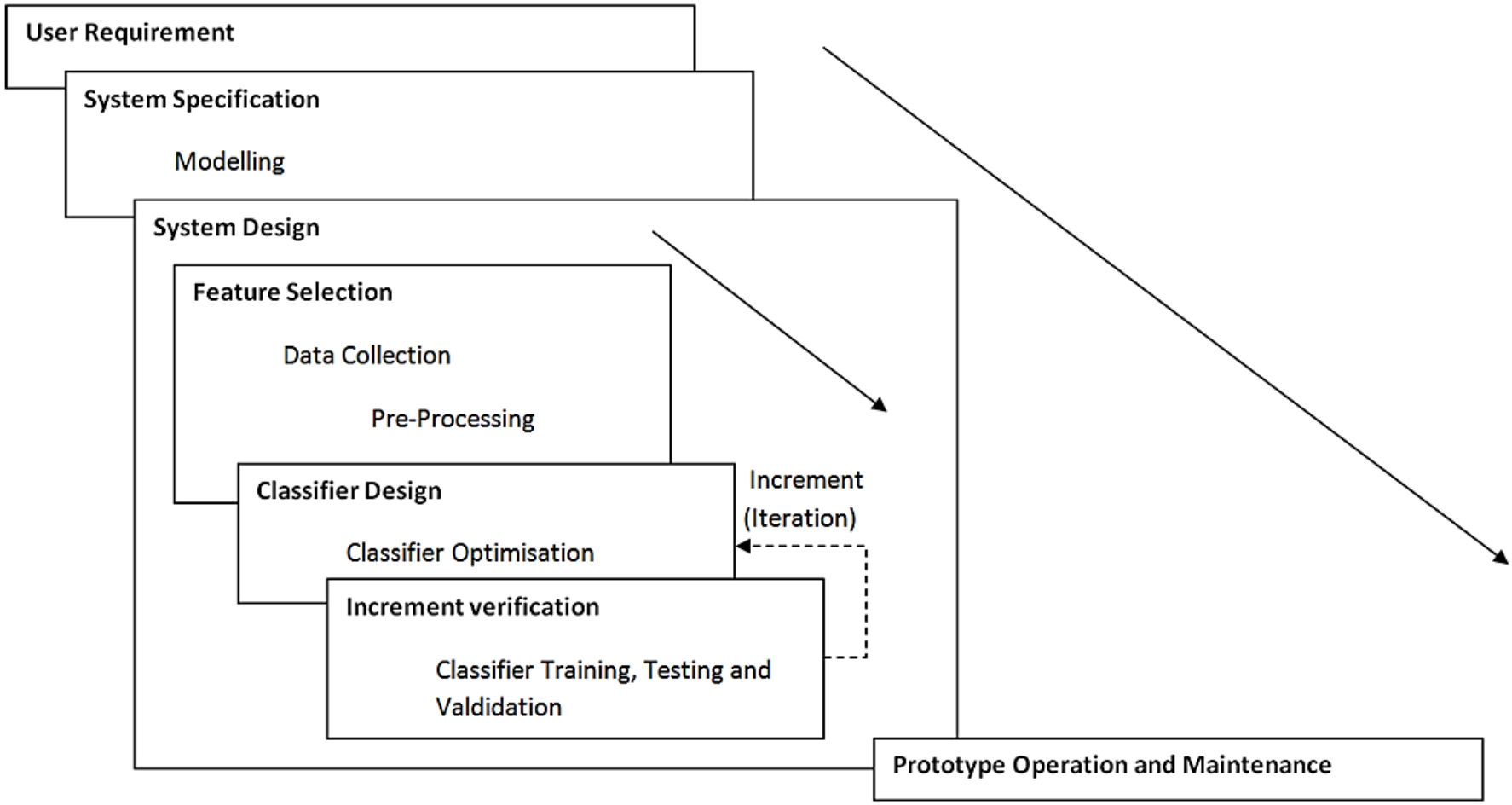

Figure 4.1 Incorporation of classifier design into the incremental build model. Reading from top to bottom (direction arrows), this is an adaptation of the standard waterfall model, showing stages and internal processes. The "System design" stage contains internal sub-stages and an iterative process.

pp.262-263). User requirements were used alongside the cognitive model (see Chapter 3, subsection 3.10.4 and Fig. 3.3) and the processes for classifier design outlined in section 4.1 as the basis for the model (see Fig. 4.2).

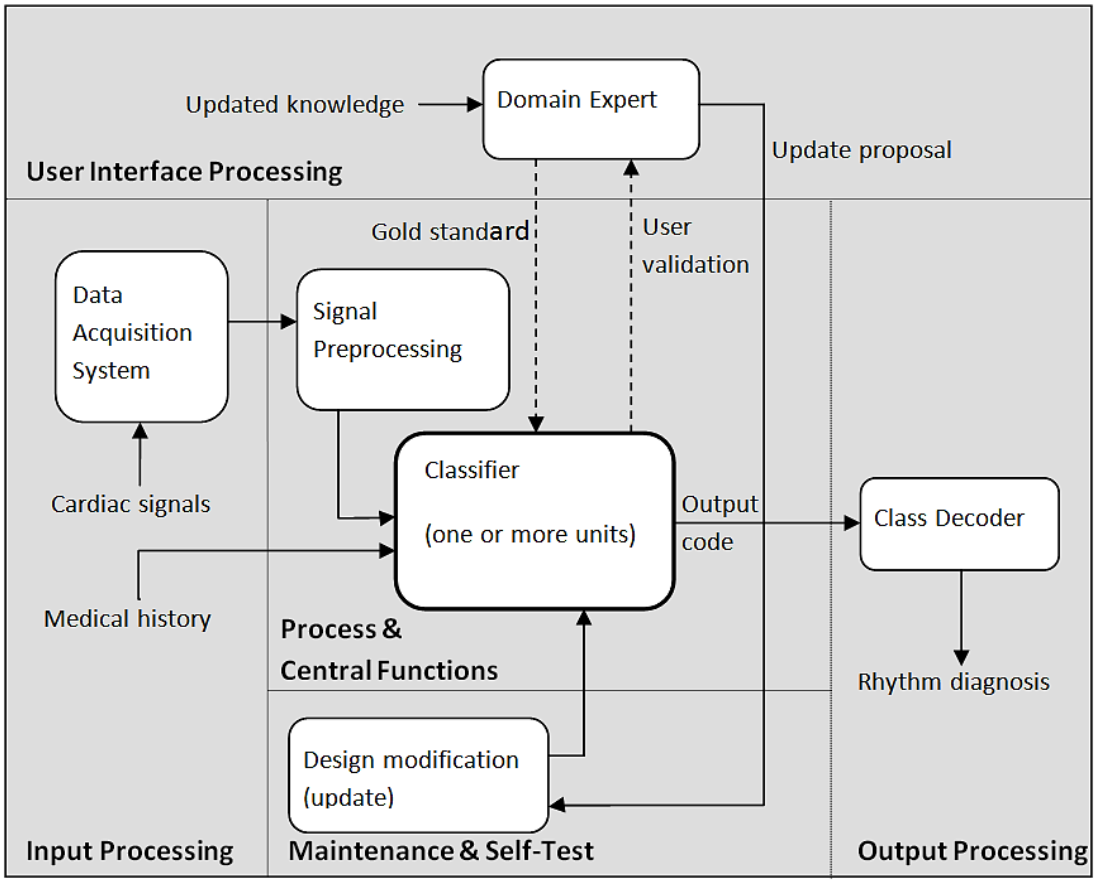

Figure 4.2 Hatley-Pirbhai system context diagram of cardiac rhythm classifier prototype development. The flow runs from left to right: starting with inputs (data) which in this application consists of cardiac signals and clinical history data; followed by central functions of signal pre-processing and classification, which is in turn subject to validation and updates; leading to class label allocation and rhythm diagnosis in the output stage.

This system context diagram was usable as a system specification when accompanied by further explanation and detail.

The system context diagram consisted of major stages as nodes, linked by arcs indicating sequence. Within the design process, the domain expert interacted with the classification system by provision of the gold standard diagnosis, the validation process and recommendation of future updates. System inputs were cardiac signals and medical history which were selected during a

process of feature selection (described in Chapter 5. Material requirements including software and hardware for data acquisition, signal and data pre-processing, classifier prototyping and performance testing are covered in Chapters 6 to 10. Finally, classifier output codes were decoded into the rhythm diagnosis.

4.4 System Design

System design built on the user requirements and system specification (see sections 4.2 and 4.3) and detailed how these were satisfied in the build. In this study, design processes included: feature selection, data collection, data pre-processing and classifier design and build, output class encoding and decoding. Any design modification, due to new knowledge or classifier inadequacy triggered a new iteration.

4.5 Iteration Implementation (Prototype Build)

Evaluation of the literature in Chapter 2 suggested a range of classifiers for implementation which performed well in this domain. Individual classifiers selected as iterations for comparison were summarised in Table 4.1.

Table 4.1 AI Classification technologies.

Classifier Technology
Decision tree
Fuzzy classifier
Naïve Bayes
Neural network
Support vector machine
Inference engine

To satisfy user requirements, classifier performance was to be maximised within the design constraints.

4.6 Iteration Verification (Prototype Testing)

Iterations were verified using various performance measures (See Chapter 3, section 3.12) and classifier performances compared. Later iterations were, selected for use in hybrid or multi-classifier systems. Combination strategies examined in Chapter 3, section 3.16 depended on results: the best performing classifiers overall; the presence of any misclassification and possible use of specialist classifiers in specific areas of the feature or solution spaces.

4.7 Prototype Operation and Maintenance

Comparisons between algorithms suggested requirement for improvement, triggering a design modification or major redesign iteration.

Each stage of the process was documented and evaluated to ensure an efficient, non-repetitive development process. Successive prototypes were based on single classifier units or combinations. Prototype testing included performance measures and expert feedback to highlight problems and suggest modifications. Iterations continued until performance was maximised and the production system was produced.

4.8 Summary

In this chapter, use of the incremental build model for classifier design was outlined. A user requirement was summarised in three goals, for which achievement would signal the end of the development process. These consisted of use of defined AI techniques, resulting in a production system with lower misclassification rate and higher accuracy than commercial ICD algorithms.

A system specification was developed, containing information describing the system. A Hatley-Pirbhai system context diagram was used as a model of system specification, completed with further detail. System design detailed how user requirements and system were satisfied in the build. A range of classifiers were suggested for implementation which performed well in this domain: decision tree, fuzzy, naïve Bayes, neural network and support vector machine.

Iterations were verified using performance measures. Combination strategies were to depend on results, using the best performing classifiers. When user requirements were met the final classifier was to become the production system.

Chapter 5 Feature selection

5.1 Overview

A feature is a measurable characteristic, in this case, of a cardiac rhythm. Features are usually numeric and a set of features may be grouped in a feature vector. The initial stage of classifier development is feature selection.

Classifier development depends on large volumes of data available as training, test and validation sets. Early data collection, without adequate consideration of the data to be included, could be counterproductive, as the features used may not be useful or others inadvertently omitted. To make best use of a data collection opportunity careful selection of features as potential algorithm inputs was considered essential.

Clinical evaluation of patients with suspected arrhythmia includes an assessment of their symptoms; clinical history of predisposing influences for arrhythmia as well as more objective measurements (see Chapter 1, section 1.2). As an aim of the study was to simulate the human diagnostic process, detailed analysis of human diagnostic processes could aid in feature selection.

Cardiac signals were identified as feature groups needed as classifier inputs (see Chapter 2, subsection 2.9.5). More specific definition of these cardiac signals and aspects of clinical history to be collected were required before data collection could proceed.

Three approaches were used for feature identification: firstly, features used in implantable pacemakers and defibrillators for rhythm diagnosis were identified; secondly, the literature was reviewed for other features previously used or feasible for use for rhythm diagnosis; finally, a knowledge elicitation process for feature selection was used and the results from the three approaches were pooled.

5.2 Features used in Implantable Pacemakers and Defibrillators

Heart rate, atrial and ventricular cardiac electrograms and their derivations: intervals, beat counters and morphology were identified as features for cardiac rhythm diagnosis in Chapter 2, subsections 2.9.5. Accelerometers, incorporated in all current devices will be examined for utility as features.

5.2.1 Electrogram Intervals

Intracardiac electrogram waves, recorded from within the heart, correspond to the surface ECG, with characteristics of timing, frequency and vector (amplitude and direction). Different intervals are measured for each cardiac cycle (See Fig. 5.1).

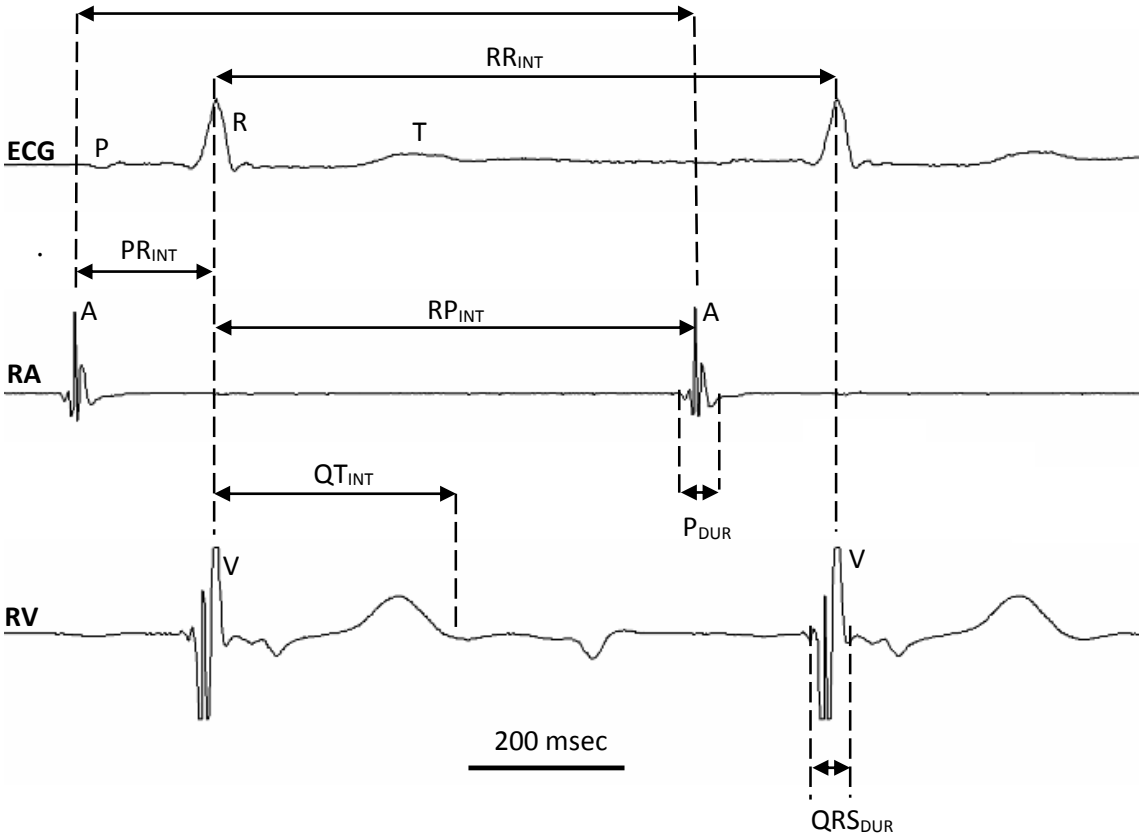

Figure 5.1 Intracardiac electrogram interval measurements. Traces (from the top) are: electrocardiogram (ECG) with P, R and T waves marked; right atrial (RA) bipolar electrogram with A waves marked; right ventricular (RV) bipolar electrogram with V waves marked. Intervals shown are the atrial interval (PP_{INT}), the ventricular interval (RR_{INT}), the atrio-ventricular interval (PR_{INT}), the ventriculo-atrial interval (RP_{INT}), the ventricular electrical recovery time (QT_{INT}), the atrial activation (P) wave duration (P_{DUR}) and ventricular activation (QRS) wave duration (QRS_{DUR}). Note that for this study, P and R waves were measured from their detected peak, rather than the onset. Signal amplitudes are not calibrated in this diagram, as the focus is upon timing intervals.

Important timing intervals used for cardiac rhythm diagnosis are: PP interval, PR interval; RR interval (heart rate); RP interval; QT interval; sequential variability of RR interval (stability and onset); P:R ratio and P/R relationship. P duration and QRS duration relate to wave morphology.

The Hamilton-Tompkins method for the detection of QRS waves from surface ECG's was implemented, adapted for use with intracardiac A and V electrograms. Intracardiac A waves will be detected instead of P waves and intracardiac V waves instead of QRS complexes or R waves

and should be considered synonymous in method descriptions. Peaks and fiducial points were detected, followed by detection of T waves (Hamilton & Tompkins 1986).

Electrogram waveforms recorded from the right atrial and right ventricular electrodes usually contain a single chamber-specific signal. Unlike these, those recorded from the His bundle and coronary sinus electrodes have complex components. The His bundle electrogram, due to the electrode location near the tricuspid valve, contains three major components (A, H and V) and the coronary sinus, which lies in the atro-ventricular groove, contains two (A and V) (see Chapter 1, Figure 1.3). The amenity of His and CS electrograms to computerised analysis is reduced, since more than one peak is usually present, making correct detection of the components complex.

Intracardiac electrograms conform to the single electric dipole model, such that unipolar electrode configurations, with the cathode within the heart and a distant anode, detect a larger signal but are more likely to detect interference, as the dipole is larger (Parker *et al*. 1969). Bipolar configurations have a smaller dipole and with a shorter distance between anode and cathode. Intracardiac electrograms also include "far-field" signals, from adjacent chambers, superimposed on local "near-field" signal. Far-field signals have been found to be less with shorter inter-electrode spacing (Fröhlig *et al.* 1999) and mean that far-field signals were unlikely to be detectable with the 5 mm inter-electrode spacing used in this study. To identify the maximum peaks of the ventricular waveform, and the maximum peaks and fiducial points of the atrial waveform, bipolar electrograms were used, due to low susceptibility to noise. To identify fiducial points of the ventricular waveform, a unipolar far-field electrogram was used.

It is notable that none of the algorithms found during a search of the literature involved analysis of the His and CS electrograms and this aspect of electrogram analysis will not be attempted in this study.

5.2.2 Electrogram Morphology

Morphology detection works by examining right ventricular electrogram morphology for its major features. Atrial electrogram morphology or vector analysis is not currently used in any device. Typically, the intracardiac electrogram is detected and the timing of QRS onset, offset and duration can be measured. Each electrogram has a series of major peaks and troughs, having characteristics of timing, sequence, polarity and amplitude which contribute towards electrogram morphological analysis.

Electrogram morphologies are analysed in a number of ways: comparison of a number of points of the waveform timed to a reference point or by a representation of salient (e.g. wavelet transforms). The St. Jude Medical MD algorithm (see Chapter 2, subsection 2.8.14) represents

electrogram morphology features as triangular waves, allowing a "reconstructed" approximation of the electrogram to be made and used for analysis. During normal rhythm, the right ventricular electrogram peaks and troughs are converted to a series of triangular waves, whose characteristics of sequence, duration and amplitude are stored as a template for comparison. During a suspected arrhythmia, the differences between each triangular wave of the template and the current beat are computed, to provide a percentage match, used to differentiate "normal" from "abnormal" beat morphology.

5.2.3 Accelerometry

Semiconductor accelerometer sensors are the dominant sensor in current pacemakers and defibrillators, converting body motion into data. Accelerometer frequency and amplitude output is converted into a proposed heart rate according to a predefined formula, used to modulate heart rate. Accelerometers have rapid response, are relatively free of interference and can be tuned to patients' individual requirements.

Accelerometer data are not used in arrhythmia discrimination algorithms. If combined, the presence of a tachycardia and low activity levels may be a good indicator of inappropriate tachycardia. The investigational "Smartracking" algorithm for pacemakers, developed to prevent unnecessarily high pacing at heart rates higher than indicated by the accelerometer sensor (Kamalvand *et al.* 1996), demonstrated this in principle, though this was not adopted. Use of multi-axis accelerometers could potentially enable the additional detection of posture and direction of travel, provided appropriate calibration is made.

5.3 Feature Selection by Review of the Literature

Candidate features and sensors not currently used for arrhythmia diagnosis, but with evidence in the literature supporting their use in arrhythmia diagnosis or management were then considered. Published evidence of outcomes research, review articles and case studies were considered. Individual case or cohort studies were considered only where they were the first report, most cited or sole available published evidence.

5.3.1 QRS Duration

QRS wave duration is a very primitive representation of QRS morphology and has been previously used in ICDs with evidence of utility (Klingenheben *et al.* 1998). It was susceptible to variation unrelated to arrhythmia, resulting in poor predictive value (Aliot *et al*. 2004) and has been superseded by morphology assessment.

5.3.2 Heart Rate Variability

Decreased heart rate variability (HRV) was proposed as a predictor of arrhythmia linked to Sudden Infant Death Syndrome (SIDS) (Leistner *et al.* 1980) and in adults following acute myocardial infarction (heart attack) (Kleiger *et al.* 1987). Both groups proposed that this is due to increased sympathetic or decreased vagal tone, predisposing towards ventricular fibrillation. Healthy hearts have a large HRV, with decreased or absent variability suggesting cardiac disease. Measures of heart rate variability are principally time-domain and frequency domain metrics. An example of a time domain metric is standard deviation of the mean of the RR intervals (of the ECG) between all normal heart beats (SDNN) over a period (usually 24 hours). A frequency domain method is the use of fast Fourier transform (FFT) to estimate power spectral density (PSD) analysis of beat-to-beat intervals, over several frequency bands (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology 1996).

On-going research into indices derived from HRV found it to be predictive of ventricular arrhythmias (Malik *et al*. 1989; Osterhues *et al.* 1993; Huikuri *et al.* 1995; Copie *et al.* 1996; Bernardi 1996; Nolan *et al.* 1998). Other methods of analysis have been proposed, including fractal components, chaos theory, entropy and heart rate "turbulence".

5.3.3 Heart Rate Turbulence

Heart-rate turbulence (HRT) is variation of sinus-rhythm cycle length after a single ventricular premature beat. Absence of the heart rate turbulence after ventricular premature beats has been found to be a strong risk stratifier after myocardial infarction (Schmidt *et al.* 1999). Onset and slope are the commonly used measures of HRT. Use of HRT is limited to patients with dominant sinus rhythm and the presence of single ventricular premature beats.

5.3.4 QT interval and T waves

Where the duration of ventricular repolarisation, measured as the QT interval, is abnormally prolonged or shortened, it is known to be associated with sudden arrhythmic death (Phillips & Ichinose 1970; Gaita *et al.* 2003).

QT interval and heart rate are inversely related, with the slope and curvature of the relationship varying between individuals (Batchvarov *et al.* 2002). A correction may be applied to eliminate the influence of heart rate (QT_c). Fixed correction methods (Bazett 1920; Friderica 1920) can be accurate for small heart rate changes (< 5 beats/min) but are inaccurate for larger changes and individualised calculations are indicated (Malik *et al.* 2012 and 2013; Garnett *et al*. 2012). QT interval has also been utilised in pacemaker applications. QT interval was the earliest effective sensor used in rate-controlled pacemakers use to determine pacing rate (Rickards & Norman 1981; Rickards *et al.* 1983; Donaldson & Rickards 1983). The QT sensor remains in clinical use

today, in a combined “blended” sensor with an accelerometer. In relation to this effect, a useful quality of QT interval is that it has an association with blood catecholamine levels (Hedman *et al.* 1990).

A study by Andrássy *et al.* (2007) supports this with its suggestion that QT_c is prolonged with mental but unaltered by physical stress. These studies suggest QT interval has utility in detection of mental stress. In this study changes in the QT interval was considered for use as one among several rhythm discrimination features representing physiological and mental stress. A fixed correction method (Bazett 1920) (5.1) was used to reduce the influence of heart rate on QT interval values and produce an index of stress, rather than to quantify the absolute value of QT_c and in this context, the impact of any error was considered minor.

$$QT_c = \frac{QT}{\sqrt{RR}} \tag{5.1}$$

5.3.5 Peak Endocardial Acceleration and Heart Sounds

The assessment of heart sounds, cardiac auscultation, is a conventional technique used in diagnosis of heart disease. A commercial pacemaker system is available with a specially made electrode, having a peak endocardial acceleration (PEA) sensor able to detect left ventricular contractility and intracardiac heart sounds and has been validated as a guide to device programming (Delnoy *et al.* 2008).

5.3.6 Body Temperature

The relationship between body temperature and metabolic rate is well established (Berggren & Christensen 1950). Body heat is generated by: metabolic processes; exercise; infection; hormonal influences (such as thyroxine, triiodothyronine, adrenaline and noradrenaline); after eating and exposure to high or low environmental temperatures and is regulated by the thermoregulatory mechanism in the hypothalamus area of the brain. A correlation between blood temperature and heart rate in pacemaker technology was demonstrated (Fearnot & Evans 1988) but has fallen out of favour due to the fragility of the electrode-mounted sensors and the slow response time from onset of exertion (Sugiura 1983; Fearnot *et al.* 1984; Alt *et al.* 1988). Its usefulness in arrhythmia detection is unknown, with just one old publication suggesting a relationship (Beaulnes & Day 1957). Axillary temperature is lower than core temperature, however there is good correlation, and it may be inferred that temperature variations are accurately reflected (Fullbrook 1993).

5.3.7 Blood Oxygen Saturation

Venous oxygen saturation was evaluated as a pacemaker rate-response sensor but due to sensor complexity and non-standard electrode failed to gain acceptance (Wirtzfeld *et al.* 1983; Stangl *et*

al. 1988; Lau *et al.* 1995). Ohlssen (2001) revisited this during a trial of an implantable haemodynamic monitor, also finding a problem with sensor failures.

5.3.8 Blood pH

It is known that severe physiological exercise may produce metabolic acidosis (reduced blood pH) (Robergs *et al.* 2004). Camilli *et al.* (1983) demonstrated use of an implantable pH rate-response sensor to detect exercise in animal implants; however this has not been successfully produced for large-scale evaluation. Anderson *et al.* (1968) found that haemodynamic instability, associated with acidosis, following cardiac infarction may also result in arrhythmia. However, there is no published evidence that arrhythmia is a causal factor in acidosis.

5.3.9 Blood Pressure

Blood pressure measurement has been in routine clinical use since the early twentieth century (Korotkoff 1905) and is the standard measure in clinical evaluation of haemodynamic status. Blood pressure is generated by the forceful contraction of the heart and may be reduced (hypotension) if contraction is impaired or raised (hypertension) due to bradyarrhythmia or a chronic clinical condition unrelated to arrhythmia. A sudden drop in blood pressure may make someone black out and is sometimes associated with bradycardia (Yuskis & Griffith 1949).

Ovsyshcher *et al.* (1992) using a pacemaker with a specially made unipolar electrode with pressure sensor, examined the peak of first derivative of the pressure waveform (dP/dt_{max}) as a pacemaker rate-response sensor. Recently, an implantable haemodynamic monitor with a pressure sensor mounted on a right ventricular electrode, also having core temperature and heart rate sensors in the device was used in large-scale clinical trials (Steinhaus 2005) and has been incorporated into a new ICD. A preliminary study suggested that pressure based indices may augment arrhythmia classification algorithms (Kremers *et al.* 2012)

5.3.10 Bio-impedance

Nyboer *et al.* (1940) described a cardiac component associated with thoracic electrical impedance. Thoracic electrical impedance is able to detect changes in cardiac output, respiration and thoracic fluid volume. From this, impedance cardiography (ICG) was developed for physiological monitoring of astronauts in space, by NASA (Kinnen & Kubicek 1963, 1964a and 1964b; Kubicek *et al.* 1966).

Standard pacing electrodes have been used to record intracardiac impedance. Constant-current non-stimulating (very low current and voltage) pulses are injected to a pair of right ventricular intracardiac electrodes and the resulting voltage is detected from the same or from a different pair. The detected signal is modulated by chamber volume and is inversely proportional to right

ventricular volume. Amplitude changes (stroke volume) and timing parameters may be measured, such as the pre-ejection interval (Chirife 1988). Salo *et al.* (1984) made stroke volume measurements using a catheter-based, right ventricular, intracardiac impedance system and Schaldach (1990) suggested the use as a pacemaker sensor (Schaldach 1990; Schaldach & Hutten 1992). Right ventricular impedance changes were found to correlate well with dP/dt_{max} and RV contractility during dobutamine stress testing and it was suggested as a sensor for a closed-loop pacing system (Osswald *et al.* 2000).

A respiration sensor based on impedance was first used in pacemakers in 1984 (Rossi *et al.* 1984). This was refined into a reliable minute ventilation sensor to modulate pacemaker controlled heart rate on exercise (Alt *et al.* 1987; Nappholz *et al.* 1987; Mond 1988; Lau *et al.* 1988). Minute ventilation is the product of tidal volume and respiration rate. Changes in minute ventilation are slightly delayed from onset of physiological exercise and the sensor is best used in combination with an accelerometer, called "sensor blending", which remains in clinical use.

5.4 Feature Selection by Knowledge Engineering

Features important to arrhythmia diagnosis were evaluated in a knowledge engineering process, outlined in a previous study (Bostock *et al.* 2010). Use of a knowledge engineering approach ensures the system includes up-to-date expert knowledge and is updatable.

5.4.1 Domain Expertise

Domain expertise for knowledge acquisition was provided by a series of internationally validated consensus clinical guidelines and highly-cited papers, rather than a single human domain expert. These source documents were considered "expert opinion" evidence, based on meta-analysis of large volumes of study data and sound international expert consensus and were used in this study as the primary source for knowledge elicitation. Source documents used were: Bonow *et al.* 2012; Blomström-Lundqvist *et al.* 2003; Surawicz *et al.* 2009; Buxton *et al.* 2006; Epstein *et al.* 2008; Saoudi *et al.* 2001; Fuster *et al.* 2011 and Brugada *et al.* 1991. Extracted knowledge was then validated by a human domain expert.

5.4.2 Clinical Diagnosis of Arrhythmia

The practice of cardiac rhythm diagnosis not specific to individual rhythms was analysed. Two of the source documents (see Chapter 3, subsection 3.10.1) contained recommendations for the clinical diagnosis of arrhythmia (Blomström-Lundqvist *et al.* 2003; Fuster *et al.* 2011) and were subjected to knowledge elicitation by detailed manual analysis of the text. Key phrases and measured parameters used in cardiac rhythm diagnosis were identified and recorded in as few words as possible. In a similar process, detailed arrhythmia definitions, previously obtained (see

Appendix E) were also subject to detailed manual analysis of the text. Key phrases and measured parameters used were identified.

5.5 Results of Feature Selection

Results of the three feature selection techniques were pooled and listed. Features were merged and functional duplicates eliminated (see Appendix F).

Features were placed into two groups: those likely to be obtained during an initial clinical evaluation (i) and those likely to be obtained during more detailed diagnostic testing (ii).

i. Features likely to be obtained during an initial clinical evaluation, from the clinical history, physical examination or simple investigative testing were: arrhythmia related symptoms; a history of cardiac or pulmonary (lung) disease; predisposing or precipitating factors; physical signs and metabolic imbalances or disease.
ii. Features likely to be obtained during detailed diagnostic testing using an electrocardiogram (ECG) or advanced tests were: the influence of stress; haemodynamic status during arrhythmia; ECG measurements and response to pacing.

5.6 Summary

In summary, important features used in implantable devices are: PP interval, PR interval; RR interval (heart rate); RP interval; sequential variability of RR interval (stability and onset); P:R ratio; P/R relationship; electrogram morphology and accelerometry.

Additional sensors and parameters examined were not all suitable for inclusion in this study. QRS duration has poor predictive value, used in isolation but is incorporated in some of the morphology algorithms. Heart rate variability and turbulence indices have been shown to be predictive of arrhythmia but require long periods of continuous monitoring, rather than providing an “instant” result. QT intervals offer a simple way of detecting catecholamine changes, particularly with mental stress and QT dispersion provides an additional index of arrhythmic risk. T wave alternans was described in as an additional predictor of arrhythmia.

A peak endocardial acceleration sensor, mounted in a pacemaker electrode can detect cardiac function and guide to device programming. Metabolic sensors mounted on intracardiac electrodes, able to detect blood temperature, oxygen saturation of acidity, have been tested. Blood temperature has been demonstrated as a useful pacemaker sensor however the sensors are fragile and have a slow response time. The usefulness of body temperature as a detector of arrhythmia is not established. Blood oxygen saturation was evaluated but failed to gain acceptance due to frequent sensor failures. Use of an implantable blood pH rate-response sensor to detect exercise has not

had large scale human trials and there is no published evidence that arrhythmia is a causally related to acidosis. A blood pressure sensor can be incorporated in an intracardiac electrode and has been successfully used to monitor cardiac function and may have utility in arrhythmia classification. Impedance sensors for respiration and cardiac function assessment are successful and do not require additional sensors.

The knowledge engineering exercise was in two parts, analysis of the general clinical approach to arrhythmia diagnosis and detailed analysis of differential diagnosis of arrhythmia. From the first part, results were dominated by factors related to clinical history and physical examination. From the second, ECG related features, acknowledged to be central to modern cardiac rhythm diagnosis, dominated.

Pooled results from the three feature selection techniques were divided into two groups: those obtained during an initial clinical evaluation and those obtained during more detailed diagnostic testing.

The first group consisted of: arrhythmia related symptoms; a history of cardiac or pulmonary (lung) disease; predisposing or precipitating factors; physical signs and metabolic imbalances or disease. The second of: the influence of stress; haemodynamic status during arrhythmia; ECG measurements and response to pacing. These features will provide the basis for data collection, the next phase of classifier development.

Decisions to include particular sensors were made based on their relevance to the selected features and that there was a corresponding sensor technology. The implementation of features and sensor technologies is discussed further in the following Chapter (see Chapter 6, section 6.6).

Chapter 6 Preparation for Data Collection

6.1 Overview

Features identified in Chapter 5 (see Chapter 5, sections 5.5 and 5.6) identified two groups of features: those likely to be obtained during an initial clinical evaluation and those likely to be obtained during more detailed diagnostic testing. Data in both groups consisted of human clinical data. To collect such data required a clinical study to be designed.

A diagnostic trial was designed as a sub-study. A population was identified and sample size estimation made. The process of ethical approval was described, including its aims, inclusion and exclusion criteria and data security.

Features were examined in detail to establish optimal measurable parameters which would enable their capture. A list of equipment required for data collection was based on the measurable parameters identified. The characteristics of the measured signals were used to form the basis of a minimum technical specification and equipment satisfying the specification was identified and assembled, ready for data collection.

6.2 Guidelines for Diagnostic Trials

In medical research diagnostic trials are clinical studies intended to find better diagnostic tests, and this study involved developing and testing a new classifier, which equated to a diagnostic test.

Many diagnostic trial results are hampered by poor study design and until recently there was a lack of guidance. This has been addressed by the production of internationally recognised guidelines for good practice by the STARD group (STAndards for the Reporting of Diagnostic accuracy studies) (Bossuyt *et al*. 2003), whose recommendations include a 25 point checklist for diagnostic trial design and reporting results (see Appendix G). STARD recommendations do not include sample size estimations, ethical and data security considerations. All were incorporated into study design at the appropriate points.

6.3 Sub-study Population

Previous studies looking at rhythm classification in ICDs (summarised in Chapter 2, subsections 2.8.13 and 2.8.14) used data from different types of collection: retrospective data from arrhythmia libraries and repositories; retrospective data collected from event data stored in ICDs and prospective data collection, using Holter monitoring for standard ECG's or invasive electrophysiological studies for intracardiac electrograms. Unlike this study, the previous studies

reviewed here did not collect continuous temperature, accelerometry or cardiac function data, so failed to provide a methodological model or a suitable source of data.

6.3.1 ICD Implant Patients as Sub-study Population

Patient populations in previous studies of ICD algorithm performances (see Chapter 2, subsections 2.8.13 and 2.8.14) either had existing ICD implants or were studied during ICD implant procedures. This population may not be easily reproducible, so data from a population that deviates from this may be expected to differ.

In the reviewed studies of ICD algorithm performances, sensitivity measures were based on diagnosis of VT and VF diagnosis and specificity measures based on discrimination of SVT. Little attention was paid to other arrhythmias.

Given that this study placed emphasis on the suitability of the final algorithm for use in ICDs, certain data characteristics were required. Firstly, collection of ECG data from skin (surface) electrodes is impractical for an implanted device, secondly, implanted electrode configuration and complexity is limited by compatibility with existing ICD hardware (standard / non-standard) thirdly, durability issues (potential for failure of overly complex or fragile technology) and lastly, current hardware configurations (electrodes and ICDs) undergo rigorous long term testing and are subject to continuous performance review so significant changes in design, though possible, may be inadvisable.

Cardiac electrograms are equivalent parameters to ECG's, recorded directly from or near the heart. To enable standard ICD compatibility, cardiac electrogram data should be collected from standard ICD compatible electrodes placed in the right atrium and right ventricle. The ICD device should be placed in the pectoral region. Left sided implants are preferred, as right-sided implants show higher all-cause mortality rate and defibrillation threshold (Gold *et al.* 2007). Data collection based on this prescription, with prospective data being collected during standard ICD implant procedures, would have seemed advisable. Potential limitations of the approach included:

1. The need to collect data at ICD implant procedures and the risks of infection or complication associated with resultant prolonged procedure times.
2. The need to artificially initiate arrhythmias other than VF during the ICD implant.
3. The low likelihood of many non-life-threatening arrhythmias beyond ICD implant indications.

Given these limitations, an alternative study population was sought for which these limitations were solved: access to a population that had right atrial and right ventricular electrodes placed,

with a low risk of infection or complication, where initiation of arrhythmia is routinely performed and where many different rhythms may be observed.

6.3.2 Electrophysiological Studies Patients as Sub-study Population

A suitable population was patients undergoing EP, which also involved placement of electrodes in the right atrium and right ventricle, among other locations. EP offered a solution to the limitations listed in subsection 6.3.1. Collection of arrhythmic data and artificial initiation of arrhythmias is normal practice during EP and the possibility of encountering very many more types of arrhythmia is greater than during ICD implant procedures. This approach also had some different potential limitations:

1. Electrophysiology electrodes differ in design to implantable ICD electrodes.
2. Right atrial electrode locations are often different to the standard location for a pacemaker or ICD implant (right atrial appendage).

These limitations were addressed as follows:

1. Electrode designs are sufficiently similar in electrode surface area, electrode material and dipole distance to minimise any practical difference. Quadripolar electrodes have proximal electrodes which could be used to simulate shock coils.
2. Right atrial pacing electrodes may often be located in non-standard locations to counter poor electrical characteristics or fixation problems.

6.3.3 Sub-Study Population

Body motion, temperature, respiration and cardiac function data collection are amenable to collection from patients at either ICD implant or EP, as data would be collected over a short period, under clinical laboratory conditions.

Having satisfactorily addressed any limitations, EP was selected as the data collection mode following which detail was added to the methodology, given the type of study involved.

6.4 Sample Size Estimation - Powering the Sub-study

For the required accuracy to be achieved, a minimum sample size should be calculated, by setting a target for power (τ) of the statistical test to be applied.

6.4.1 Effect, α, β and Power

Values for α (type I error or significance level of the test) and β (type II error), which is related to power τ, and a desired effect size (β) are required to calculate τ (6.1).

$$\tau = (1 - \beta) \tag{6.1}$$

Widely accepted convention uses a 4:1 trade-off between β and α error. This trade-off is provided when β is 0.2 and α is 0.05 at a value of 0.8 for τ. Values typically used are τ of 0.8 or 0.9, and α of 0.05 or 0.01. For a power of 0.8, a false null hypothesis will be rejected 80% of the time.

Effect size represents the association between two groups, which for Fisher's exact test is often represented by the phi (ϕ) coefficient or odds ratio (OR). Phi values lie between -1 and +1 and a common interpretation: -1.0 to -0.7 strong negative association; -0.7 to -0.3 weak negative association; -0.3 to +0.3 little or no association; +0.3 to +0.7 weak positive association; +0.7 to +1.0 strong positive association. OR is a measure of effect size with values of 0 to infinity, where 1 represents no relationship; values <1 are not interpretable, so the group expected to have higher odds of an event should be in the first column of the 2 x 2 contingency table and values >1 mean there is a significant difference between groups. OR is a measure of effect size with values of 0 to infinity, where 1 represents no relationship; values <1 are not interpretable, so the group expected to have higher odds of an event should be in the first column of the 2 x 2 contingency table and values >1 mean there is a significant difference between groups.

A simple method to estimate sample size is trial and error, inserting values into the 2 x 2 contingency table until a desired odds ratio (OR) or ϕ is obtained. An alternative is the use of pilot study data or data from previous similar research to estimate a minimum sample size.

6.4.2 Sample Size Estimation from a Pilot Study and Published Studies

This study aims to improve diagnostic accuracy for all types of cardiac rhythm, exceeding the accuracy of commercial ICD algorithms. A convention in accuracy evaluations of commercial ICD algorithms is for a diagnosis of VT and VF to be positive and SVT negative, with accuracy expressed in terms of sensitivity and specificity. Sensitivity is used to represent the correct diagnosis of VT and VF (usually combined) and specificity the correct classification of SVT. This convention was observed in sample size calculations.

Using data from a pilot study (Bostock 2004), sample size estimation was made with pooled data used to test all ICD algorithms. For 800 events, 600 were positive (VT or VF) and 200 negative (SVT), of which 579 were true positives (a); 145 false positives (b); 21 false negatives (c) and 55 true negatives (d). Negative to positive ratio was 200:600 or 0.333. The probability of incorrect diagnosis of SVT (false positive rate, α error) as VT or VF was 0.725 and of correct diagnosis of VT or VF (sensitivity, power) was 0.965. 73 VT or VF events and 25 SVT events are required to

be able to reject the null hypothesis that the two probabilities are equal with a power of 0.8 and α error of 0.05.

Using data taken from the study of Hintringer *et al.* (2001) for ELA PARAD algorithm, there were 86 events, of which 15 were positive (VT or VF) and 71 negative (SVT), 15 true positives (a); 36 false positives (b); 0 false negatives (c) and 35 true negatives (d). Negative to positive ratio was 71:15 or 10.47. The probability of incorrect diagnosis of SVT (false positive rate, α error) as VT or VF was 0.507 and of correct diagnosis of VT or VF (sensitivity, power) was 1 (0.999 used in calculation). 7 VT or VF events and 74 SVT events are required to be able to reject the null hypothesis that the two probabilities are equal with a power of 0.8 and α error of 0.05.

No published study was found against which to compare these proportions in a population undergoing EP study, but Zeldis *et al.* (1980) studied a similar population of 518 patients with palpitations, dyspnoea, discomfort in the chest, dizziness, and syncope, using 24-hour ECG recordings. Of these patients, 40 (7%) had ventricular arrhythmias and 54 (10%) supraventricular tachycardia.

Sample size calculations were performed using PS Power and Sample Size Calculations software, Version 3.0 (Dupont & Plummer 1990), using Fleiss's (1981) generalisation of the method of Casagrande & Pike (1978).

6.4.3 Converting Samples into Patients

The pilot study suggested a sample size of 73 VT or VF events and 25 SVT events whereas data from Hintringer *et al.* (2001) suggested 7 and 74 events respectively were required to provide sufficient power to reject the null hypothesis. Taking the higher value for each suggests 73 VT or VF events and 74 SVT events.

Converting a sample size estimated as units of events into a number of patients is complex. For example, 73 patients examined will not all have VT or VF, so patient numbers will need to be greater than this. Prevalence of VT, VF and SVT in the population of patients undergoing EP is unknown and given that event rates will vary between institutions conducting the procedures, an estimate of the number of patients to be recruited to satisfy power requirements is not possible.

The data from Zeldis *et al*. (1980) indicated a likely prevalence of these arrhythmias in an EP patient population. Using the proportions found in Zeldis *et al*. (1980) study, to find 73 episodes of ventricular arrhythmia would require a study sample size of 73/7 * 100 or 1042 patients and to find 7 episodes would require 100 patients. In the same way, to find 74 or 25 SVT events would require 740 and 250 patents respectively. Given that numbers over 100 were impractical in the

planned single-centre study, during the period allocated to data collection, the lower requirement of 100 patients was accepted.

6.5 Sub-study Ethics Application

An application was made to St. Thomas' Hospital Research Ethics Committee using the NHS REC (National Health Service Research Ethics Committee) process on 9/10/2007. The REC reference was: 07/H0802/119.

The study title was: "Creation of an Intracardiac Electrogram and Physiological Parameter Library for Use in Cardiac Arrhythmia Research". Approval was received on 30/1/2008 (see Appendix H). A protocol amendment was submitted on 8/4/2009 and approval of the amendment was received on 22/05/2009.

6.5.1 Type of Study

This was a single centre, prospective, uncontrolled case-series, involving collection of physiological parameter data at the time of elective interventional cardiac electrophysiological studies (EP), including radiofrequency ablation (RFA).

6.5.2 Sub-study Objective

The objective was to obtain a collection of physiological parameter data corresponding to a wide variety of cardiac rhythms in a large cohort of patients.

6.5.3 Setting

The setting for data collection was the location of performance of EP procedures, the cardiac catheterisation suites at Guy's and St. Thomas' NHS Foundation Trust (GSTT).

6.5.4 Duration of the Sub-study

Data was collected at the time of EP or RFA procedures only and not at any other time.

6.5.5 Recruitment

Given the power calculation (see subsection 6.4.30) of 100 patients, a maximum of 200 patients, set by expedience for the expected eligibility at GSTT, were to be enrolled within the recruitment period of one year. Approved start and end dates of data collection from were 22/5/2009 to 30/4/2010.

6.5.6 Conduct Monitoring

A consultant cardiologist, Dr Michael Cooklin monitored the clinical aspects of this work on GSTT premises. Individual patient studies were conducted under the direct clinical supervision

of the co-investigators who were the consultant physicians caring for the patients undergoing investigation: Dr Michael Cooklin; Dr Mark O'Neill; Dr Jaswinder Gill, Dr Aldo Rinaldi and Dr Eric Rosenthal. Academic Supervision was provided by Dr Peter Weller, Senior Lecturer in Medical Informatics, City University London.

Any adverse event attributable to the study was reported to the monitors who would make an evaluation and recommend continuation or discontinuation of the study, as appropriate.

6.5.7 Potential Risks and Benefits to Participants

There were no known potential adverse effects, risks or hazards. Minor disadvantages were that additional skin electrodes might cause very minor skin irritation (itching or redness). There were no specific benefits to participants other than their contribution to this or other possible future use of the data in research.

6.5.8 Funding

The sub-study was funded by Guy's and St. Thomas' NHS Foundation Trust.

6.5.9 Main Ethical Issues

The main ethical issue identified was effects of the use of additional skin electrodes, temperature probe and accelerometer and the time taken to apply them, prior to the commencement of the study (estimated to be 15 minutes). Patients had this explained to them and had the option to decline consent.

6.5.10 Data Types

Data consisted of a library of anonymised demographic data, clinical history data and physiological signals and measurements collected during cardiac electrophysiological studies and radio-frequency ablation (RFA) procedures. Measurements included those were those normally made during these procedures, including the following parameters: 12-lead ECG, intracardiac electrograms, and measurements not normally made during these procedures: impedance cardiography (non-invasive), body motion accelerometry (non-invasive) and body temperature (non-invasive).

6.5.11 Data Security

Health records were accessible to no-one outside the normal clinical team. Prior to anonymisation, data collection and transfer was on dedicated NHS systems and archived to DVD, held in secure storage, as is customary for a standard procedure.

Data anonymisation was performed, after the data was collected, on GSTT premises by the chief investigator, who had access to the non-anonymised data as a part of his normal duties. Anonymisation consisted of the removal of personal identification data, such as participant name, address, NHS number, hospital number, date of birth (age and sex at data collection will be retained); no other demographic data was collected. Anonymised patient data was identifiable beyond the data collection phase only by a study number. Anonymised data was stored in digital form on DVD, enabling later off−line processing.

6.5.12 Data Retention

The Chief Investigator was named as custodian of the data. The data will be retained indefinitely, with a minimum of 15 years from the date of enrolment.

6.5.13 Data Verification

Cardiac rhythms were verified by two domain experts, at the time of recording, at least one of whom was a cardiac electrophysiology consultant. This provided the "gold standard" reference diagnosis for later use in supervised learning and validation.

6.5.14 Sub-study Inclusion and Exclusion Criteria

A rationale for study population selection was explained in section 6.4. The patient population (see subsection 6.4.3) was patients undergoing invasive diagnostic electrophysiological studies where at least right atrial and right ventricular electrodes would be positioned.

Inclusion and exclusion criteria were:

Inclusion Criteria:

Patients scheduled for non-emergency invasive cardiac electrophysiological study with or without radiofrequency ablation, including children under 16.

Exclusion Criteria:

Patients having prophylactic ablation (such as for atrial fibrillation) without invasive electrophysiological study.

Concurrent participation in any other clinical investigation or trial; prisoners; young offenders; adults in Scotland who are unable to consent for themselves; healthy volunteers; a dependent relationship with the investigator e.g.: those in care homes, medical students; adults with learning disabilities or are unconscious or very severely ill or have a terminal illness or in emergency situations or with mental illness or with dementia; other vulnerable groups.

Justification for the inclusion of vulnerable groups (such as children):

> The study of paediatric arrhythmias was an important aspect of this study and exclusion would limit the value of the data. As the study involved minimal inconvenience or additional risk, this was considered a viable approach.

6.5.15 Participant Recruitment

Recruitment was based on referral for invasive cardiac electrophysiological testing or ablation procedures, based on symptoms with or without tests supporting the diagnosis.

Potential participants were supplied with a written Patient Information Sheet, suited to their age group as defined in NPSA guidelines, containing detail of the investigation, for a minimum of one hour prior to being approached for consent. Where this time was not available, consent was not sought.

6.5.16 Informed consent

All patients were enrolled following an informed consent procedure with the clinician performing the test taking consent. Information and consent procedures were adopted appropriate to each age group, with the full participation of parents, patients and responsible physician. Parental and patient consent was sought for patients under 16. Investigators explained the purpose of the study to the patient and consenting parent or guardian and answered any questions. There were no non-English-speaking patients.

6.5.17 Withdrawal of Consent

Participants were informed of their right to withdraw consent at any time. It was considered unlikely that additional information would become available during the course of the research that might be relevant their continued participation, given the short duration of participation. In that eventuality, the consenting clinician was expected to be fully aware of any new information relevant to the study and was to supply participants with a suitably updated patient information sheet, subject to local research ethics committee approval. If the participant lost capacity to consent during the course of the research, they would be withdrawn from the study with any identifiable data anonymised and retained or disposed of.

6.5.18 Insurance and Liability

All patients were recruited at NHS sites, so the NHS indemnity scheme or professional indemnity applied. Liability for clinical negligence by NHS staff lay with the NHS.

6.6 Converting Features to Measured Parameters

In machine learning, the terms “feature” and its synonyms “variable” and “attribute” refer to a measurable property with a set of features known as an instance or pattern. A “parameter” may be a special type of feature. In clinical medicine, parameter often refers to a measurable factor, or a vital sign, such as temperature, pressure, or the ECG and this definition is adopted for this study.

A measured parameter may have several features derived from it. For example, from the ECG parameter we can derive heart rate, PR, RR, QT intervals and P:R relationship. There were a number of different recordable parameters to best represent those features selected in Chapter 7. Close examination of individual features, given that they would be collected at EP investigations, was made to explore options.

6.6.1 Features from Clinical Evaluation

Following the pattern of human clinical diagnosis, information from the clinical history, physical examination and simple investigative testing was collected during the process of clinical evaluation. In this study, it was considered unnecessary to duplicate this process and this documented data was extracted and encoded without further intervention.

6.6.2 Features from Diagnostic Testing

Features obtained during detailed diagnostic testing included those selected on the basis of ECG diagnosis, supplemented by supporting evidence (see Chapter 5, section 5.5). These were: the influence of ECG measurements; stress; haemodynamic status during arrhythmia and response to pacing.

6.6.3 ECG Features

ECG features could be consolidated to the detection of P, QRS and T waves and measurement of their respective time intervals, morphologies and axes as well as detail of P to R arithmetic relationships and sequences.

In an EP procedure intracardiac electrograms are taken from the right atrium and right ventricle, and the signals are used to represent intracardiac equivalents of the ECG P and QRS waves respectively, as in EP studies (see Chapter 5, subsection 5.2.1 and Fig. 5.1).

Electrograms may be “unipolar”, with a distant, extracardiac electrode having a voltage close to zero as an indifferent (negative) electrode and an intracardiac recording (positive) electrode or they may be bipolar, with both electrodes within the cardiac chamber, separated by a dipole distance, typically 1 cm.

To achieve a unipolar configuration, options for an indifferent electrode include: a composite of surface ECG limb electrodes in Wilson's central terminal (WCT) configuration (Wilson *et al.* 1934); an additional invasive electrode catheter positioned in the inferior vena cava (IVC) or an extracardiac electrode in some other position distant to the heart (pseudo-unipolar).

WCT generates a mean voltage from the right arm, left arm and left leg surface ECG electrodes. Use of WCT for intracardiac electrogram recording often results in unacceptable levels of electrical noise, as intracardiac electrograms require significantly greater amplification compared with surface ECG recordings. Unwanted electrical interference is detected and amplified from each of the three limb electrodes used to generate a WCT, possibly an exacerbation. The use of an additional invasive electrode in the IVC probably provides best results (Stevenson & Soejima 2005) but involves ethical concern, due to the requirement for the placement of an additional intravascular electrode, for a recording that is not clinically indicated. A practical pseudo-unipolar alternative was required. An indifferent electrode suitable for pseudo-unipolar recording can be provided using a skin ECG electrode placed in the left pre-pectoral region. The voltage at this position will be close to that of the left arm (VL) and being distant to the heart will be low compared with the high voltages expected from intracardiac active electrodes in this study. An assumption was made that the contribution of the indifferent voltage (VL) to recorded signal amplitude was small. During an EP study, electrograms (EGM) can be obtained using combinations of pairs of different electrodes (see Fig. 6.1).

6.6.4 Features Related to Stress

A variety of device–based sensors for detecting physiological, emotional and mental stress have been tested (see Chapter 5, subsections 5.2.3 to 5.3.8). The most well-known were considered for use in this study: accelerometry, minute ventilation, temperature and the QT interval of the ECG.

Body motion has a potential to be used as a detector of physical exercise and so represent physical stress. Body motion data is typically acquired in pacemakers (see Chapter 5, subsection) by accelerometry and is measured in units of *g*, multiples of the acceleration due to gravity (g_0). A commonly used index for body motion over time, combining axial data, is the square root of the sum of squares (RSS) of each vector, called vector magnitude units (VMU) (Coleman *et al.* 1997; Steele *et al.* 2000). With RSS, squaring converts the data into positive values and the square root returns an output vector containing VMU as index for body motion as an aggregate of the original data, shown here for *x* and *y* axes only, in a dual-axis accelerometer (6.2).

$$VMU = \sqrt{(x^2 + y^2)} \tag{6.2}$$

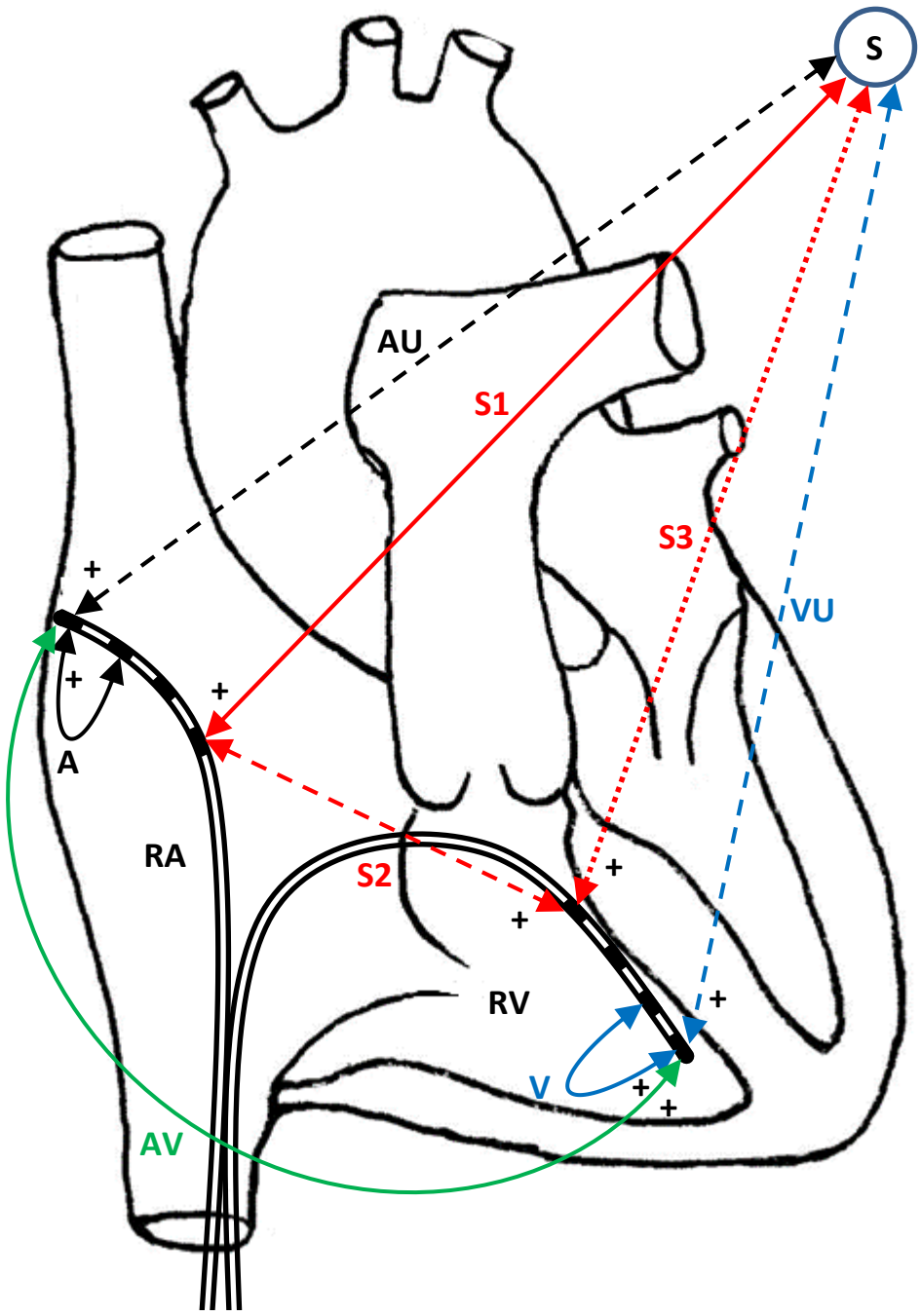

Figure 6.1 Schematic of atrial, ventricular and atrio-ventricular composite electrograms. Quadripolar electrodes are in the lateral RA and RV apex. A skin electrode (S) in the left pre-pectoral region. Key**:** A=Atrial bipolar (black arrows); V=ventricular bipolar (blue arrows); AV=composite AV (green arrows); AU= atrial unipolar (black dashed arrows); VU=ventricular unipolar (blue dashed arrows); S1=far-field atrial unipolar (red arrows); S2=far-field composite A4 to V4 (red dashed arrows); S3=far-field ventricular unipolar (red dotted arrows); S is the skin electrode used as indifferent for pseudo-unipolar electrograms; + indicates a positive (active) recording polarity, otherwise negative.

A respiration sensor based on impedance was first used in a pacemaker in 1984 (Rossi *et al.*1984) and was refined into a reliable minute ventilation sensor to modulate pacemaker controlled heart rate on exercise (Alt *et al*. 1987; Nappholz *et al.* 1987; Mond 1988; Lau *et al.*1988). Changes in minute ventilation, the product of tidal volume and respiration rate, are used in pacemakers to detect increased respirations. The availability of this established sensing technology means that, like body motion, respiration could be used as a sensor to indicate physical stress.

Features related to stress (see Chapter 5, subsections 5.2.3 to 5.3.8) were consolidated to parameters measuring: body motion (by accelerometry); body temperature; respiration by thoracic bio-impedance and QT interval from the ECG.

6.6.5 Haemodynamic Status

Haemodynamic status is most commonly measured using blood pressure. The practicality of measuring blood pressure on a beat-by-beat basis requires invasive blood pressure monitoring, by insertion of a cannula into an artery. Measurement of haemodynamic status by an implantable device has been achieved in a number of ways and these were examined as solutions for data collection.

Recently, an implantable haemodynamic monitor with a pressure sensor mounted on a right ventricular electrode, also having core temperature and heart rate sensors in the device has been used in large-scale clinical trials (Steinhaus 2005) and has been incorporated into a commercial ICD. An alternative is a peak endocardial acceleration, which also involves an electrode-mounted sensor to measure an analogue of right ventricular contractility. Both peak endocardial acceleration and blood pressure would require invasive monitoring, not normally required in experimental procedures, posing an ethical challenge. To be developed for any resultant commercial system, both require an additional sensor, adding to commercial risk.

In contrast, impedance or conductance measurements have been extensively studied in implantable devices and uses standard electrodes, not requiring additional sensors. Stroke volume measurements were made by a catheter-based, right ventricular, intracardiac impedance system (Salo *et al.* 1984). Schaldach (1990) suggested the use of intracardiac impedance measurement as a pacemaker sensor (Schaldach 1990; Schaldach & Hutten 1992). Further study of right ventricular impedance changes were found to correlate well with dP/dt_{max} and right ventricular contractility during dobutamine stress testing and suggested it as a sensor for a closed-loop pacing system (Osswald *et al.* 2000). Standard pacing electrodes have been used to record intracardiac impedance. Two or more electrodes are used, depending on the proprietary system. Constant-current non-stimulating (very low current and voltage) pulses are injected to a pair of right ventricular intracardiac electrodes and the resulting voltage is detected from the same or from a

different pair. The detected signal is modulated by chamber volume and is inversely proportional to right ventricular volume during the cardiac cycle. Amplitude changes (stroke volume) and timing parameters may be measured, such as the pre-ejection interval (Chirife 1988).

ICG is a convenient, validated method to non-invasively use thoracic bio-impedance to monitor both cardiac function and respiration in a simulation of device-based impedance measurement. With an ICG recording, a pair of electrodes is placed at the base of the neck and another pair at the upper abdomen. A high frequency, low amplitude, constant alternating measurement current is passed through the chest, via the outer pair, producing an induced voltage across the inner pair. The measurement current seeks the path of least resistance and blood is the most electrically conductive intervening tissue, so current passes mostly through the thoracic aorta, superior vena cava and inferior vena cava (Summers *et al.* 2003). Underpinning theory of thoracic impedance is that Z_0 is the static base impedance and is indirectly proportional to total thoracic fluid content (Pomerantz *et al.* 1970). Thoracic fluid conductivity is directly proportional to thoracic fluid content (Pomerantz *et al*. 1969 and 1970) and is used as a diagnostic feature for congestion monitoring of heart failure in some ICDs (Yu *et al.* 2005). Total thoracic impedance change with time (ΔZ) can be expressed as the sum of its components (6.3) where the cyclic respiratory impedance component (ΔZ_R) is due to changes in venous and pulmonary blood volume caused by each respiratory cycle and the cardiac impedance change (ΔZ_C) is due to aortic blood volume and velocity changes with each cardiac cycle.

$$\Delta Z = Z_0 + \Delta Z_R + \Delta Z_C \tag{6.3}$$

Use of the first time derivative of ΔZ, dZ/dt reduces the effect of ΔZ_R signal and its peak is proportional to cardiac stroke volume (*SV*) (Kubicek *et al*. 1966; Sramek 1983; Bernstein 1986; Osypka & Schafer 1998; Osypka & Bernstein 1999). A study by Van de Water and colleagues (2003) found dZ/dt correlated well with cardiac output (r = 0.81). Impedance cardiography *SV* formulae include the Sramek-Bernstein equation (6.4) (Sramek 1983; Bernstein 1986) and the Bour equation (6.5 and 6.6) (Charloux *et al.* 2000).

$$SV = \delta . \frac{(0.17H)^3}{4.2} . \frac{dZ/dt_{max}}{Z_0} . LVET \tag{6.4}$$

In the Sramek-Bernstein equation, δ is weight as a fraction of ideal weight, as taken from the Metropolitan Life insurance tables; H is height; dZ/dt_{max} is the systolic peak of dZ/dt; Z_0 is the base impedance; and *LVET* the left ventricular ejection time.

Used in the PhysioFlow system, the Bour equation (6.6) uses a stroke volume index calibration (SVI_{cal}) performed for 30 consecutive beats at rest as a baseline reference (6.5); $Z_{max} - Z_{min}$ is the

Z variation during systole; *TFIT* is the thoracic flow inversion time, from first zero crossing after the Q wave of *dZ/dt* to the nadir after dZ/dt_{max}; $TFIT_{cal}$ is *TFIT* during the calibration period and is weighted by heart rate and arterial pressure difference (systolic minus diastolic); *k* is a constant. Parameters are measured and compared to the baseline.

$$SVI_{cal} = k.\frac{dZ/dt_{max}}{(Z_{max} - Z_{\min)})}.W(TFIT_{cal}) \tag{6.5}$$

$$SV = SVI = BSA.SVI_{cal}.\sqrt[3]{(dZ/dt_{max}/dZ/dt_{maxcal}).\ TFIT_{cal}/TFIT)} \tag{6.6}$$

Impedance measurements provided a validated method to record beat-by-beat haemodynamic status and monitor respiration.

Naidu and co-workers accurately extracted impedance cardiogram B, C and X fiducial points using Matlab with a low computational cost (Naidu *et al.* 2011). Their recommended window size for C wave detection was modified heuristically in this study to the first 40% of each RR interval following the V wave, rather than 20%. Conventionally, the B point is the point of onset for the rapid upstroke of *dZ/dt* preceding the C wave. Naidu, at-odds with convention, used the nadir preceding the C wave to detect B point. An optimal automated algorithm for B detection using this might involve use of the peak second derivative (d^2Z/dt^2) following the R wave (Debski *et al.* 1993; Bour & Kellett 2008). Lozano *et al.* (2007) derived a simple function to accurately estimate RB interval, based on RC interval (6.7).

$$RB_{int} = 1.233RC_{int} - 0.0032{RC_{int}}^2 - 31.59 \tag{6.7}$$

Attempts to derive the second derivative d^2Z/dt^2 in this study resulted in excessive high-frequency noise and were abandoned, leading to use of the Lozano equation.

6.6.6 Pace Termination

Literature on pace termination of arrhythmia overwhelmingly relates to treatment, rather than diagnosis; however overdrive pacing is particularly useful for the treatment of re-entrant arrhythmias and is extensively used during diagnostic invasive electrophysiological studies (EP).The source material used for feature extraction refers to this only once, by way of its ineffectuality in the termination of non-paroxysmal junctional tachycardia (See Appendix E). Pace termination is the subject of recently increased research activity however, in this study, in common with a majority of commercial ICD systems, pace termination will be considered a treatment rather than a diagnostic technique and will be excluded from further evaluation.

6.6.7 Equivalent Measurable Parameters

From the exploration of recording suitable parameters with which to capture the feature data suggested in Chapter 5, section 5.6, data collection was planned using the following measurable parameters: ECG's (intervals, relationship, morphology and vector); body temperature; body motion accelerometry and thoracic bio-impedance.

6.7 Equipment Selection

Based on the parameters outlined (see subsection 6.6.7) any equipment required, its minimum specification, consumables and data collection methodology were considered and outlined.

6.7.1 A Generic Equipment List

Features previously consolidated into recordable parameters were matched to recording solutions: manual notation; intracardiac electrograms; temperature; body motion (accelerometry) and thoracic bio-impedance and using this, a generic equipment list was drawn up.

Demographic and clinical history data would be manually recorded, using a procedure worksheet (see Appendix J); ECG-like features as well as emotional and psychological stress would be recorded using an electrophysiological system; metabolic stress using body temperature; exercise physiologic stress using accelerometry; physiologic stress (respiration) and haemodynamic status using impedance cardiography.

Equipment for data collection was specified, based on capability (to record the required parameters), technical specification and availability.

6.7.2 Technical Specification: Bandwidth, Sampling Rate and Resolution

Digital, rather than analogue, data recording simplifies data export for later off-line post-processing. Sampling converts analogue signals into time-series of digital data, whose accuracy is dependent on sampling rate and signal resolution.

To calculate an appropriate sampling rate, Nyquist–Shannon theorem can be used. The Nyquist rate is the lowest sampling rate (in Hz) which avoids aliasing and is equal to double the bandwidth of the signal. The minimum sampling rate should be greater than the Nyquist rate. Oversampling at a rate above the Nyquist rate does not alter information content but may be advantageous during digital signal processing, for example in noise elimination. Signal resolution is measured in voltage (amplitude) or bit rate, both set by analogue-to-digital converter (ADC) circuitry. An 8-bit ADC output has 256 amplitude steps, 12-bit 4096 steps and 16-bit 65,536 steps. Resolution in volts is the voltage range divided by the number of amplitude steps. Knowledge of signal

bandwidth and the required resolution of a parameter are important when specifying recording equipment.

Intracardiac electrogram bandwidth has been studied (Myers *et al.* 1978; Kleinert *et al.* 1979; Parsonnet *et al*. 1980) and found to be 300Hz, and a Nyquist rate of 600Hz. To specify the electrophysiology system, practical systems typically have 12-bit resolution and 1 kHz or better sampling rates.

Meng *et al.* (2006) performed a power spectral analysis for common movements (walking, running, jumping, skipping) and found they occupy a narrow bandwidth up to 10 Hz with a central frequency 4Hz and frequencies above 10 Hz caused by vibration artefact . This gives a Nyquist rate of 20Hz. Typical commercial accelerometers have 8 to 12-bit resolution.

For temperature measurement, Wurster & McCook (1969) found that the maximum rate of change of skin temperature was 0.008 ^{O}C /sec. With the range of body temperatures compatible with life being 27 to 43 ^{O}C and a resolution of 0.1 ^{O}C required for medical temperature monitoring, the resolution equates to 160 amplitude steps, a bandwidth of 0.08Hz and a Nyquist rate of 0.16Hz. Impedance cardiogram bandwidth is 50Hz (Hurwitz *et al.* 1993), a Nyquist rate of 100Hz. Impedance cardiography systems typically have 12-bit resolution and 200Hz or better sampling rates.

These minimum specifications are summarised in Table 6.1.

Table 6.1 Minimum equipment specifications.

Parameter Description	Bandwidth	Nyquist Rate	Resolution
ECG, Intracardiac Electrograms	0 - 300 Hz	600 Hz	12-bit
Accelerometer	0 - 10 Hz	20 Hz	8-bit
Temperature	0 - 0.08 Hz	0.16 Hz	0.1 ^{O}C
Impedance	0 - 50 Hz	100 Hz	12-bit

6.7.3 Satisfying the Specification

Equipment was selected to meet the specifications in table 6.1.

The Ensite 3000 (St. Jude Medical, St. Paul, MN, USA) electrophysiological system with V7.0 software was selected for use. The system has 12-bit resolution (range of 125mV), a sampling rate of 1200Hz, exceeding the Nyquist rate (see subsection 6.7.2) and 32 channels which may

include ECG's, intracardiac electrograms and up to 8 analogue inputs. It has a digital data export capability, allowing off-line processing.

Body temperature was monitored using a TEMPerNTC USB temperature probe (RDing Technology Ltd, Shenzhen, China) with a probe resolution of 0.06 ^{O}C. No analogue waveform output was available and a Windows XP SP3 desktop computer with a USB interface and the manufacturer's TEMPerNTC v3.2 data logging software was used, set to a sampling rate of 1Hz. The software allowed for probe calibration, to account for any differential between measured and actual temperature.

Body motion was monitored using an Analog Devices ADXL202EB dual-axis accelerometer (Analog Devices Inc., Norwood, MA, USA) with a bandwidth of 500Hz, range of ± 2 *g* and 5 m*g* resolution. No analogue waveform output was available and a Windows XP SP3 desktop computer having an unused serial port (RS232) with X-Analyze v2.02 software (Crossbow Technology Inc., San Jose, CA, USA) was used for data logging, set to a sampling rate of 100Hz.

A PhysioFlow, model PF-05 Lab1 with v1.0.7 software (Manatec Biomedical, Paris, France) impedance cardiography system with a sampling rate of 250 Hz. The PhysioFlow system uses a measurement current of 3.8 mA (peak to peak) at 75 kHz and uses the Bour equations (6.5 and 6.6) to calculate stroke volume.

6.7.4 Additional Equipment

As stated, a desktop computer running Windows XP SP3 was required to run software for the impedance cardiograph, temperature and accelerometry data logging.

6.8 Consumables

Quadripolar electrophysiology catheters for intracardiac electrograms are part of the normal consumables for EP so were not an additional requirement. Additional consumables were: 11 additional standard ECG electrodes per participant for ECG and impedance cardiography. 3 recordable DVD's per participant for data archival.

6.9 Summary

Feature selection identified features consisting of human clinical data. To collect such data required a clinical study and a diagnostic trial was designed as a sub-study.

The STARD guidelines were incorporated into study design. Features were amenable to collection from patients at either ICD implant or EP and the limitations of both were examined.

EP was selected and it was then possible to add detail to the methodology in the knowledge of the type of study involved.

Minimum sample size calculations were performed, using previous research and pilot study data, for α of 0.05 and τ of 0.8. These suggested 73 VT or VF events and 74 SVT events. An estimate of the number of patients corresponding to this number of events was not possible.

A sub-study was designed for data collection. Ethics application was made and suitable approvals obtained. The sub-study was a single centre, prospective, uncontrolled case-series, involving collection of data at EP, aiming to obtain a collection of physiological data corresponding to a wide variety of cardiac rhythms. Data collection was once-only for each participant at the time of EP. Recruitment was to be a maximum of 200 patients over 1 year, at GSTT. Conduct of the study was monitored by a responsible clinician and academic supervisor. There were no major risks or benefits identified. Funding was provided by GSTT. Minor ethical issues of the additional electrodes applied and the time taken to set up equipment with the patient were identified. Data types being collected, data security, anonymisation and data retention arrangements were defined. Data verification was by domain experts, providing the "gold standard". Having identified the population of patients undergoing EP, inclusion and exclusion criteria were defined. Participant recruitment process was outlined. Informed consent and issues of consent withdrawal of consent, participant insurance and researcher liability were considered.

Clinical history, physical examination and simple investigative testing data was collected during clinical evaluation. Features included demographics and clinical history; the influence of ECG measurements; stress; haemodynamic status during arrhythmia and response to pacing. Detailed examination of features identified recordable parameters and matched these to recording solutions and a generic equipment list. These were a procedure worksheet; an electrophysiological system; body temperature; accelerometry; physiologic stress (respiration) and impedance cardiography.

Equipment for data collection was specified, based on capability, specification and availability. Minimum specifications for the four systems for recording parameters, based on bandwidth, sampling rate and resolution were established. Available equipment that met the technical specification was an Ensite 3000 electrophysiological system, for intracardiac electrogram and waveform recording; a TEMPerNTC USB temperature probe; an Analog Devices ADXL202EB dual-axis accelerometer and a PhysioFlow, model PF-05 Lab1 impedance cardiography system. Availability of this equipment allowed the study to proceed to data collection.

In summary, equipment and consumables required were:

Ensite 3000 (Electrophysiology system)

TEMPerNTC USB temperature probe (Body temperature)

Analog Devices ADXL202EB Dual-axis accelerometer with arm band (Body motion)

PhysioFlow, model PF-05 Lab1 impedance cardiograph

Windows XP SP3 desktop PC (Impedance, Temperature and Accelerometer Data logging)

1 x E7506 Patient Return Electrode (Intracardiac electrograms) (Per Participant)

11 x ECG electrodes (Impedance and ECG) (Per Participant)

3 x Recordable DVD (Data archival and storage) (Per Participant)

PhysioFlow v1.0.7 software (Impedance cardiograph)

TEMPerNTC v3.2 software (Temperature)

X-Analyze v2.02 software (Accelerometer)

Chapter 7 Data Collection

Having specified and identified equipment and consumables (see Chapter 6, section 6.9), the practicalities of data collection are described. Conduct during the study is represented by a typical workflow. Equipment setup, connections, use during procedures and data export are explained. Data collection was conducted at Guy's and St. Thomas' NHS Foundation Trust, London between 22/6/2009 and 27/4/2010.

7.1 Equipment Assembly and Workflow

Equipment was assembled in advance and positioned in the cardiac catheter laboratory to minimise disruption of the working environment and avoid delays. A stainless steel clinical trolley was used for the impedance cardiograph and the Windows PC (see also section 7.5). The remaining large equipment was routinely stored in the laboratory used for the investigations.

A workflow for sub-study patient procedures was devised:

1. Participant enrolled, following informed consent
2. Participant allocated a study identification number (ID), used to identify recordings ensuring anonymisation of data from the time of collection.
3. Windows desktop PC – connections to impedance cardiograph, accelerometer and temperature probe and power up.
4. Electrophysiological system (Ensite) – pin-tip jumpers, all connections (see Table 7.1) and power up. Software loaded and ready for recording.
5. Temperature software started (logged at 1 second intervals).
6. Accelerometer software started (continuous X, Y monitoring) .
7. Impedance cardiograph software started (continuous *dZ/dt* monitoring, SV logged at 1 second intervals).
8. Manual time synchronisation of the 3 systems.
9. Attach sensors and electrodes (15 minutes allowed).
10. Manual notation on worksheet of clinical history and demographic data.
11. Start recording as soon as RA and RV catheters in position.
12. Note arrhythmias induced, medications, proceed to RFA, any other notable event.
13. Archival of data to DVD (one per patient for Ensite data, one for other data

Workflow sequence between stages 5 to 10 was not critical.

Patient's connections appeared similar to those encountered during routine invasive electrophysiological studies (see Fig. 7.1).

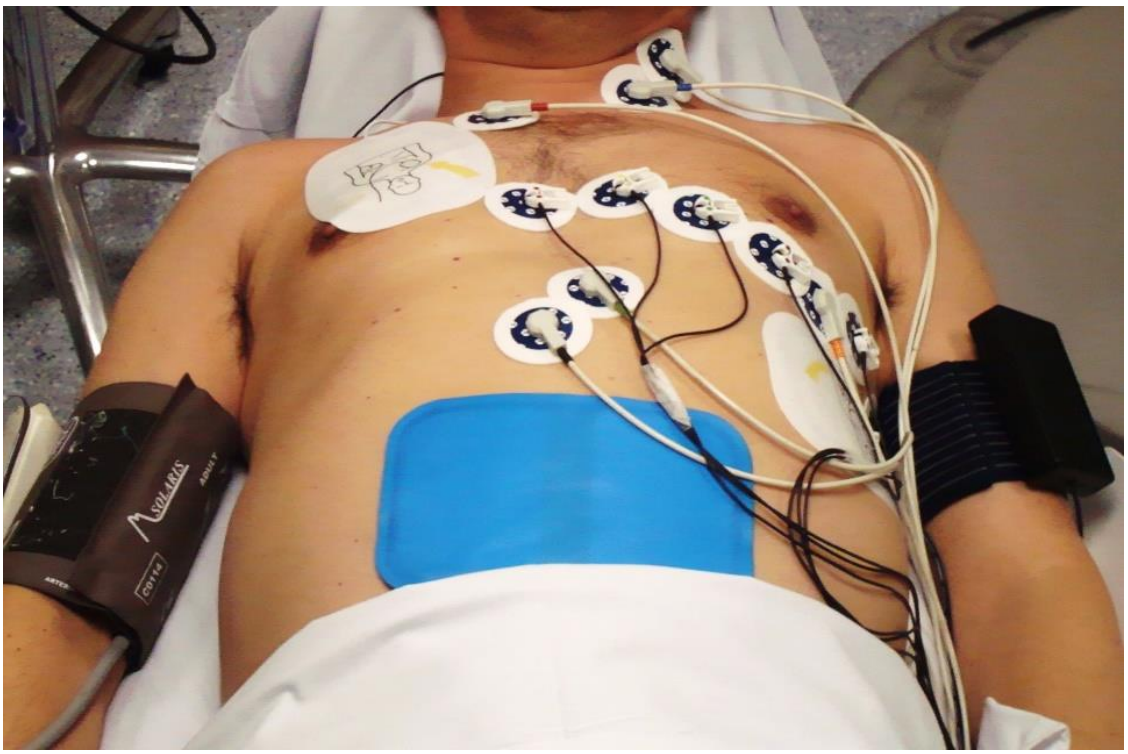

Figure 7.1 Simulation of a patient with systems attached (intracardiac electrodes not shown). The accelerometer is attached to left upper arm, non-invasive blood pressure cuff to the right upper arm. Electrodes attached to the neck and lower sternum are for impedance cardiography. The large white patch electrodes are for standby emergency cardioversion. The remaining ECG electrodes provide a 12-lead ECG to the electrophysiology recording system. The large blue abdominal patch is the reference electrode for the Ensite system. Intracardiac electrode insertions to the femoral veins are not shown. (Subject was a volunteer model).

7.2 Procedure Worksheet

At the start of the study, demographics, clinical history and examination data were manually transcribed from medical records onto a procedure worksheet (see Appendix J). Concurrent notes of events during EP procedures were made, including any adverse events or complications, induced arrhythmias, medications administered and whether a procedure proceeded to RFA.

7.3 Electrophysiological System

An Ensite 3000 (St. Jude Medical, St. Paul, MN, USA) electrophysiology system, with v7.0 software recorded ECG, intracardiac electrograms and Z (ΔZ) waveforms. Four radio-translucent pre-gelled ECG electrodes (4500M, Unomedical, Lejre, Denmark) were connected to the patient's limbs to obtain standard surface ECG waveforms. The Ensite system required a large surface area electrode (E7506 Patient Return Electrode, Covidien, Dublin, Ireland) on the abdomen to act as

a reference electrode. Two 5F (1.67 mm in diameter) Supreme JSN 401443 quadripolar electrophysiology electrode catheters (St. Jude Medical, St. Paul, MN, USA), with 5 mm inter-electrode spacing (see Fig. 7.2), were inserted by the operator under X-Ray control, into the femoral vein, advanced to heart and positioned in the right atrium and right ventricle (see Chapter 1, Fig. 1.2).

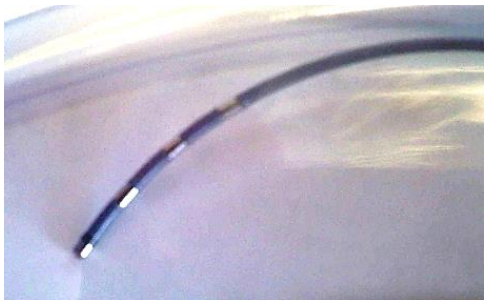

Figure 7.2 Tip appearance of a 5F quadripolar catheter showing the four electrodes.

The quadripolar electrodes were connected to the Ensite system interface module, using pin-tip jumpers connected in parallel with the clinical electrophysiology system. Intracardiac electrograms (see Chapter 6 subsection 6.6.3 and Fig. 6.1) configuration was as shown in Table 7.1 and stored as a template for re-use in all patients in the study.

Table 7.1 Waveforms recorded on the Ensite system.

Channel	Signal Type	Parameter Name	Polarity	
			+	-
1	ECG	Surface I	LA	RA
2	ECG	Surface II	LL	RA
3	ECG	Surface III	LL	LA
4	EGM	Atrial bipolar (A)	A1	A2
5	EGM	Ventricular bipolar (V)	V1	V2
6	EGM	AV Composite (AV)	V1	A1
7	EGM	Ventricular unipolar (VU)	V1	Skin
8	EGM	Coil Composite (S2)	V4	A4
9	EGM	Can to V coil (S3)	V4	Skin
10	Analog 1	PhysioFlow ECG	N/A	N/A
11	Analog 2	PhysioFlow Z curve	N/A	N/A

The Ensite system recorded waveforms continuously during the EP (see Fig. 7.3). Recordings were interrupted during ablation energy application, due to distortion of the signals. Some studies were of long duration, resulting in the generation of large text files of up to 500MB in size, containing up to 4604201 lines of data. Data export was performed using a proprietary application within the Ensite software. Data was exported as a single tab-delimited text file (.TXT) for each period of recording (see Fig. 7.4) and burned to DVDR.

All the waveform data was sampled at 1200 Hz with 12-bit resolution by the Ensite system. Surface ECG signals were band-pass filtered for display between 0.1 – 100Hz and intracardiac electrograms between 4 – 150Hz, both with 50Hz noise filters switched on however data was exported to text file as raw, unprocessed data. ECG and electrogram data had a flat frequency response between 0.1 to 300Hz with data values of 1.0 corresponding to signal amplitude 1mV, with a maximum range of 125mV. A DC analog input from the PhysioFlow impedance cardiograph, provided *Z* (impedance) data with data values of 0.01 corresponding to signal amplitude 1 Ohm.

7.3 Digital Thermometer

A TEMPerNTC USB temperature probe was connected to an unused USB port of the Windows PC (defaults to COM3). The probe was positioned in the patient's left axilla and secured using a thermally insulating self-adhesive polyurethane foam pad (often an unconnected ECG electrode). At the start of each procedure a digital clinical thermometer (ACT 2010, Actherm Medical Corp., HsinChu, Taiwan) was used to measure axillary temperature, placed adjacent to the USB temperature probe. The value was recorded simultaneously with temperature from the probe, to allow for a correction to be applied during post-processing. Temperature was continuously displayed during the procedure.

Temperature data was logged using the TemperNTC software at 1 sample per second and stored in a dedicated folder on the hard disk drive of the PC, as a comma separated volume file (.TXT) (see Fig. 7.5). File sizes were up to 50MB and consisted of up to 1482124 lines of data.

7.4 Accelerometer

An Analog Devices ADXL202EB dual-axis accelerometer was connected to the 9-pin RS232 serial input (COM1) of the Windows PC. The X-Analyze software was run and the accelerometer zeroed so all subsequent motion, during the recording would be compared to the original sensor position. The accelerometer was strapped to the patient's upper arm, without causing constriction. Motion in two axes (X and Y) was recorded throughout the procedure. Data was stored in a dedicated folder on the hard disk drive of the PC, in comma separated volume format, as a text file (.TXT) (see Fig. 7.6).

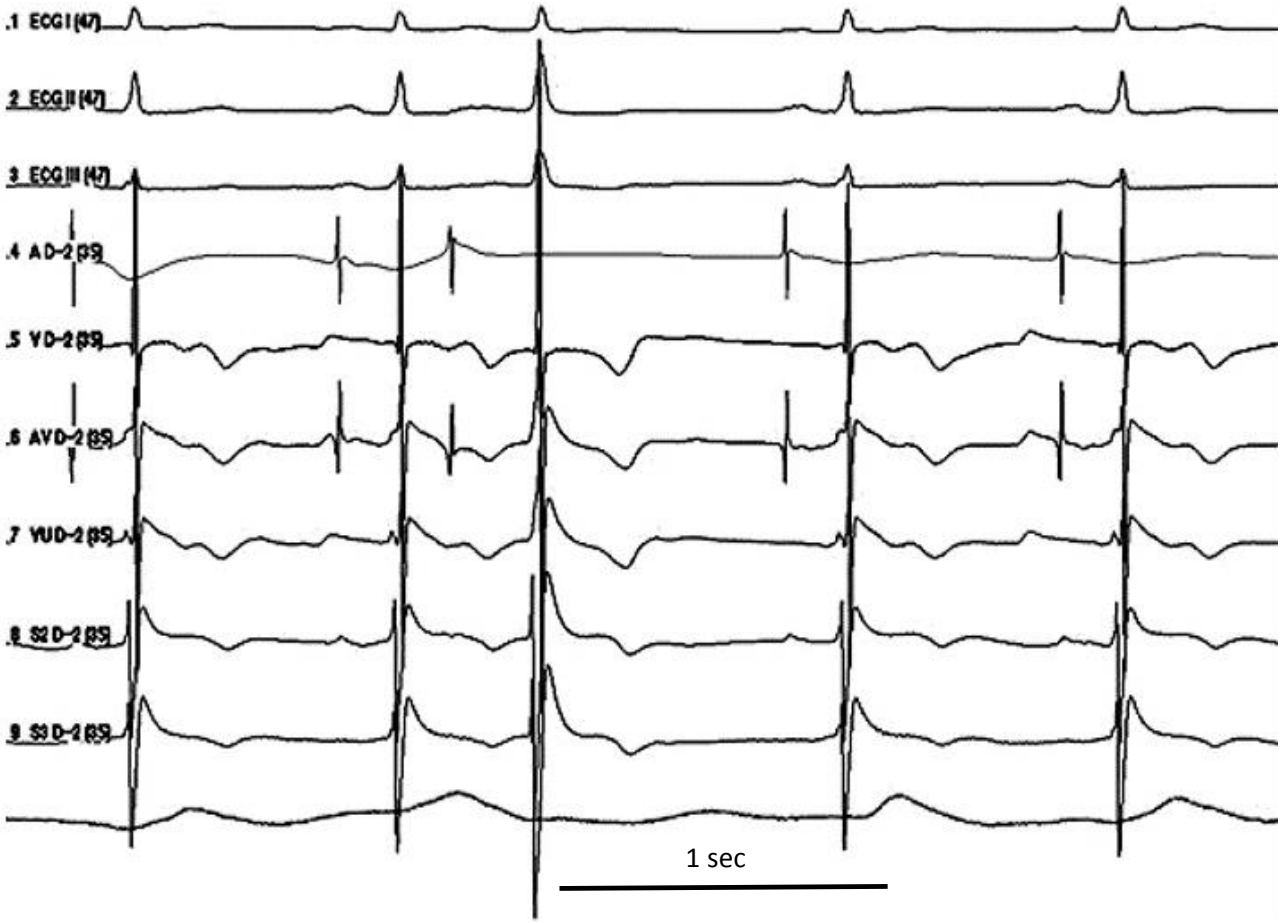

Figure 7.3 An example of waveforms recorded on the Ensite system. Traces are (top down) 1 to 3 - ECG leads I, II, III, 4 - RA bipolar electrogram; 5 - RV bipolar electrogram; 6 -composite atrial and ventricular electrogram; 7 - ventricular pseudo-unipolar electrogram; 8 – far-field composite atrio-ventricular electrogram; 9 – far-field pseudo unipolar ventricular electrogram; lowest (label not shown) - impedance cardiogram Z. Signal gain settings (bracketed) are shown adjacent to channel labels. Gains were arbitrary for wavefrom viewing and not used in signal analysis.

St. Jude Medical Waveform data export; file format revision 4

Exported from study study_esi9080_2009:09:30:09:39:14

Beginning sample number (time): 10:14:40:0411

Ending sample number (time): 10:26:48:1199

Number of waves (columns): 11

Number of samples (rows): 874389

****

* Wave Names and (x y z) coordinates *

****

Wave 0 = ECG I

Wave 1 = ECG II

Wave 2 = ECG III

Wave 3 = A D-2

Wave 4 = V D-2

Wave 5 = AV D-2

Wave 6 = VU D-2

Wave 7 = S2 D-2

Wave 8 = S3 D-2

Wave 9 = ECG

Wave 10 = Z

Begin data

Wave 0	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9	Wave 10
0.063	0.148	0.089	-0.001	0.204	0.213	0.411	-0.328	0.247	1.408	1.254
0.060	0.137	0.081	-0.007	0.195	0.221	0.463	-0.410	0.304	1.408	1.254
0.057	0.126	0.074	-0.010	0.186	0.222	0.486	-0.464	0.335	1.402	1.254
0.055	0.117	0.067	-0.008	0.181	0.216	0.456	-0.441	0.315	1.397	1.249
0.054	0.108	0.059	-0.005	0.179	0.208	0.395	-0.290	0.266	1.392	1.249
0.054	0.100	0.051	0.000	0.177	0.197	0.336	0.011	0.222	1.387	1.254

Figure 7.4 Extract of data export from Ensite system (data from patient KP029).

```
NO , Time , Inner, Outer
1,27/08/2009 09:35:15,waitting,waiting
2,27/08/2009 09:35:15,waitting,waiting
3,27/08/2009 09:35:17,20.91 C,94.83 C
4,27/08/2009 09:35:17,20.91 C,94.83 C
5,27/08/2009 09:35:19,20.91 C,32.38 C
6,27/08/2009 09:35:19,20.91 C,32.38 C
7,27/08/2009 09:35:21,20.85 C,32.36 C
8,27/08/2009 09:35:21,20.85 C,32.36 C
9,27/08/2009 09:35:23,20.98 C,32.15 C
```

Figure 7.5 Extract of temperature data export (data from patient ML015).

```
X-ANALYZE Sensor Data
ADXL202-EB-232A on COM1
File Opened:   03/15/2010     10:03:17
Software Version:      2.02

Initial Settings:
Digital Filtering:     5
X Sens (PWM %/g):      12.500
Y Sens (PWM %/g):      12.500
X Offset (PWM %):      50.000
Y Offset (PWM %):      50.000

Time (s)        X (PWM %)       Y (PWM %)       X (g)   Y (g)
7.221   55.33   59.53   0.426   0.763
7.346   55.36   59.55   0.429   0.764
7.352   55.21   59.56   0.416   0.765
7.392   55.12   59.59   0.409   0.767
```

Figure 7.6 Extract of accelerometer data export (data from patient AC052).

7.5 Impedance Cardiograph

A PhysioFlow, model PF-05 Lab1 impedance cardiograph was operated from a Windows PC, using the PhysioFlow software. The 9-pin RS232 serial output of the PhysioFlow was connected, using a serial to USB converter, to an unused USB port (default COM4) of the PC. The analogue Z output was connected from the 15-pin analogue output (DA 15) of the PhysioFlow, using a proprietary cable, to an unused DC analogue input of the Ensite, by a BNC connection. 6 radio-translucent pre-gelled ECG electrodes (4500M, Unomedical, Lejre, Denmark) were placed on the patient in the manufacturer's recommended positions (see Fig. 7.7). With the Bour equation, used in the PhysioFlow system, there is no absolute measurement of Z_0 so electrode positioning was not critical.

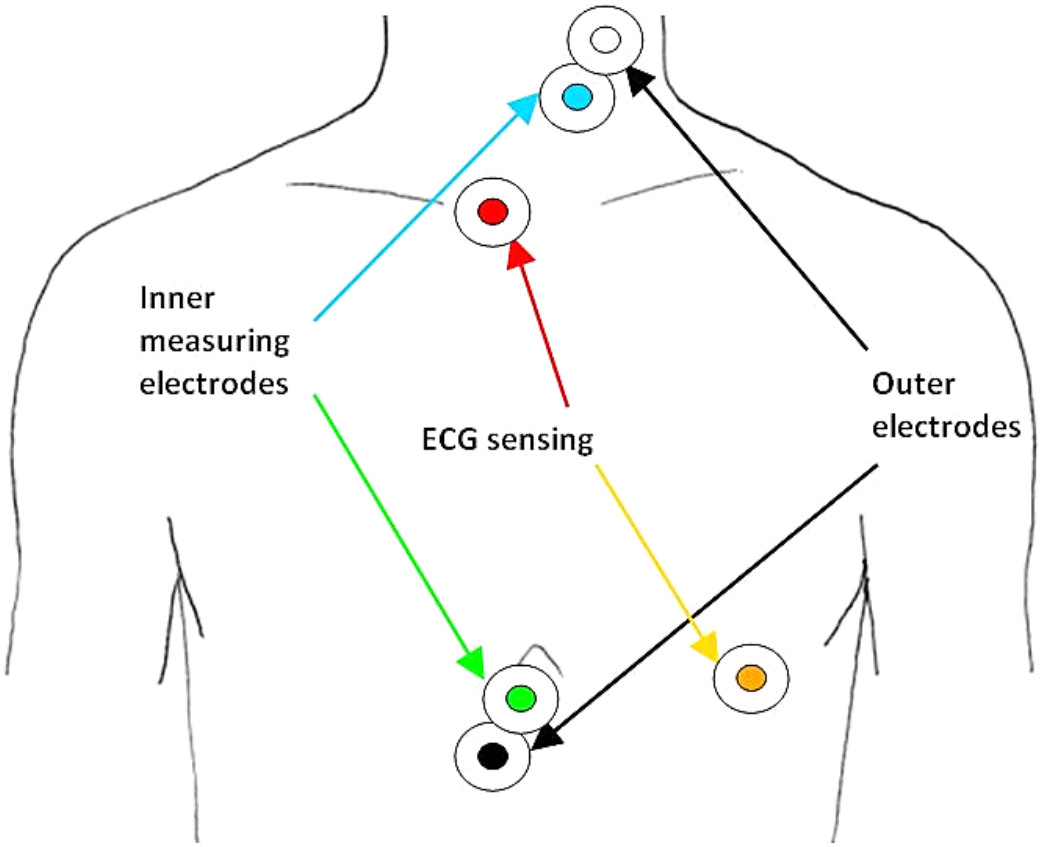

Figure 7.7 Schematic of electrode placements for impedance cardiography. Outer electrodes (White/ Black) apply the AC measurement current; inner electrodes (Blue/ Green) measure the thoracic impedance. Red/ Yellow electrodes provided an ECG reference trace. Exact electrode placements were not critical (manufacturer's recommendation) with approximate positions being based on positions relative to the sternum and neck.

Using the PhysioFlow software, study ID, patient height, weight and starting blood pressure were entered. A 30 second calibration phase was initiated, during which artefact level was monitored and the results displayed.

Recording was started and ICG derived haemodynamic parameters were displayed during the procedure and logged at 1 second intervals. PhysioFlow computed multiple haemodynamic indices which were exported as an Excel file (.XLS) for each patient.

7.6 Clock Synchronisation

The system clock times on Ensite and PC were manually noted simultaneously, to enable data synchronisation to within one second, during post-processing.

7.7 Summary

Equipment was assembled and positioned in the catheter laboratory. A standardised workflow was in place and a procedure worksheet used for documentation during the study.

The Ensite system was used to record ECG, intracardiac electrograms and impedance waveforms (ΔZ). An axillary temperature probe was used to monitor body temperature, a dual-axis accelerometer strapped to the patient's upper arm monitored body motion and impedance cardiography monitored heart function. These parameters were all displayed in real-time, during the procedure.

Data export was performed from the Ensite and the three systems running software on the Windows PC. Data was exported either as text or Excel files and burned to DVDR.

Chapter 8 Data Preparation and Pre-processing

8.1 Overview

Newly collected raw data may be unsuitable for use in classifiers without some form of processing to create usable feature sets. Data preparation takes raw data and produces high quality data as feature sets, containing sufficient information to adequately represent the feature space.

In Chapter 3, sections 3.7 and 3.8 transforming datasets into features were explained and concepts of dimensionality and dimension reduction techniques introduced, in section 3.9 missing data and use of imputation were discussed. These will be enacted within a data preparation regime.

There is no universal recipe for data preparation, other than to transform data into a form suitable as machine learning inputs which will produce the best possible results. Zhang *et al*. (2003) presented a comprehensive review of data preparation processes for pattern recognition, defining them as a staged process of: data collection, cleaning, integration, transformation, reduction, and discretisation (binning). Zhang's structured approach was used and will be described. Waveforms, temperature and accelerometer data were considered continuously monitored parameters and were pre-processed using Zhang's algorithm. Demographic, clinical history, symptomatic and baseline steady state values for haemodynamic parameters were not continuously monitored parameters and detailed pre-processing was not needed prior to their being appended as features. Data quality was assessed following data collection.

Microsoft® Excel® 2010 (Microsoft Corporation, Redmond, WA, USA), Matlab® version 7.8.0, R2009a (The MathWorks Inc., Natick, MA, USA) and a freeware text editor, Win32Pad (Feldman 2007) were used during data preparation.

8.2 Data Cleaning

Data was examined for its completeness, accuracy and consistency by manual examination of a domain expert, at the time of annotation (see also Chapter 9, subsection 9.4.5).

Imported data files were examined for corrupt, inaccurate and missing data, ensuring consistency with other data, including dealing with missing data (see Chapter 3, section 3.9). Unusable data was discarded; any required calibration or time synchronisation factors were applied. There were few missing data within demographics, clinical history and examination data. Where sign or symptom data was not explicitly recorded, it was assumed to be absent. Waveform data was of high reliability, with no corrupted or missing data. Intracardiac electrogram data was considered critical to rhythm classification, so corrupt, inaccurate or missing data was considered lost with further processing infeasible. Waveform data was "topped and tailed" to exclude noisy or

incomplete data at the start of procedures and to include complete ECG cycles, starting and ending the segment during the inter-beat period.

Using Win32Pad, temperature data header row and columns containing sample number, recording date, outer (ambient) temperature and units were not required and were deleted. Data collected prior to attaining a steady state was unreliable and discarded. Likewise, unusable data at the start and end of the file was discarded by "topping and tailing". Where electromagnetic interference interfered with data collection data was also discarded. As temperature was calibrated to skin temperature (see Chapter 7, section 7.3), a calibration factor was calculated. The data was imported to Excel and the calibration factor applied. As the maximum rate of change of skin temperature is 0.008 ^{O}C /sec (see Chapter 6, subsection 6.7.2), and temperature resolution was 0.1 ^{O}C, missing data lasting less than 12 seconds was linearly interpolated without loss of resolution.

Accelerometer data column one was a time stamp and columns four and five contained acceleration on X and Y axes. Columns two and three contained pulse width modulation (PWM) data which was not required and deleted using Win32Pad. Calibration had been performed prior to recording. As the bandwidth of body motion is 10Hz (see Chapter 6, subsection 6.7.2), missing data lasting less than 0.1 seconds was linearly interpolated without loss of resolution.

8.3 Data Integration

The integration target was a concatenation of waveform data with temperature and accelerometer data, using time synchronisation and up-sampling.

8.3.1 Time Synchronisation

As temperature and waveform data were collected on different systems, a time correction was applied to align to waveform data (see Chapter 7, section 7.6). For each waveform segment, times of first and last waveform data points were noted and temperature or accelerometer data outside this range was discarded.

8.3.2 Up-sampling

Up-sampling was required to integrate the temperature and accelerometer data. Temperature and accelerometer sample rates were calculated for each dataset. Data points were considered to be located at sample interval start, so for n samples collected between time t_1 and t_2 there would be n-1 points, enabling sample rate (f_{in}) calculation. Up-sampling then increased f_{in} to the desired sample rate (f_{out}) by an up-sampling factor (L), adding L-1 zero-value points to each sample interval. For each recording, L was calculated for both temperature and accelerometry data, using an Excel worksheet (see Table 8.1).

Table 8.1 Up-sampling Excel worksheet (temperature data, patient ID AC052).

Description		Value	Units
Samples	n	680	
Time of first sample	t_1	09:48:13	hh:mm:ss
Time of last sample	t_2	09:59:41	hh:mm:ss
Sampling period	$t_2 - t_1$	00:11:28	hh:mm:ss
Calculated sample rate	f_{in}	0.9869186047	Hz
Sample interval		1.0132547865	secs
Desired sample rate	f_{out}	1200	Hz
Up-sampling factor	L	1215.905744	
Non-up-sampled interval		1	sample
Up-sampled interval	c	0.0008224322	samples
Expected samples	$f_{out} * (t_2 - t_1) + 1$	825601	samples

Number of samples (n); and f_{out} were entered into the worksheet and values of L and up-sampled interval ($1/L$) computed. Temperature and accelerometry recorded outside the limits of the waveform recording was considered superfluous so data preceding the initial waveform sample (up to 1200 samples) and beyond those required to match the waveform vector size were discarded. This meant the synchronisation of data had a maximum error of 1 second.

Linear interpolation was used to up-sample data. Waveform, temperature and accelerometer data text files were imported into Matlab as variables. An impulse response *x* was created, beginning with the starting sample number and ending in the last sample number to be interpolated (in the above example samples 1 to 680), taken from the Excel worksheet. An up-sampled impulse response *xi* was then created, using the start and end sample numbers together with the up-sampled interval ($1/L$) calculated in the Excel worksheet. An input vector *y* was created from the appropriate column of the imported data matrix, comprising the data between the desired samples numbers as before. One-dimensional data interpolation, Matlab function *interp1q* was used to generate a vector *yi* of the desired length, containing the original and interpolated values. In some recordings, where missing data was significant, a Matlab *padarray* command was used to add values equal to the nearest measured value (no change). Up-sampled data was concatenated with waveforms, using the Matlab *horzcat* function.

8.4 Data Transformation

Recorded waveforms were amenable to digital signal processing (DSP) techniques available using Matlab and data was transformed using filtering and differentiation. In this study, absolute values were considered useful and normalisation of data was not performed.

8.4.1 Clinical History and Examination Data

Data was manually transferred from the procedure worksheet into an Excel worksheet. In the absence of a risk scoring system for arrhythmia, a completely arbitrary encoding system was designed for clinical history and examination data.

Where a drug, condition, sign or symptom was present it was represented in binary form (0 absent or 1 present). Grading systems were used for some parameters (see Appendix K): syncope (0 no symptom , through light headedness to 2, loss of consciousness); symptom character (0 to 8, the sum of 4 characteristics: 0 to 3 for increasing duration of relevant symptomatic episodes; 0 to 3 for increasing; 0 to 1 for sudden onset and offset; and 0 to 1 for termination with vagal manoeuvres; 0 to 3 a count of pre-existing cardiac conditions, including previous cardiac surgery (>3 counted as 3); antiarrhythmic drugs, alcohol intake, nicotine (smoking) caffeine intake were all scored 0 for none to 3 for heavy usage, taken as 20 cigarettes, 3 units alcohol, 3 drugs, 5 cups caffeinated coffee per day).

8.4.2 Interference and Far-Field Electrograms

Reliable identification of waves was limited by unwanted waveform components, such as: power-line interference (50 or 60 Hz); somatic muscular activity; variation in electrode contact and electromagnetic interference. Much of the required signal decomposition was achieved at data collection, as electrode locations were known as part of the study design. Recorded bipolar electrograms contained little far-field information (see Fig. 8.1 a) and b)) (see Chapter 5, subsection 5.2.1).

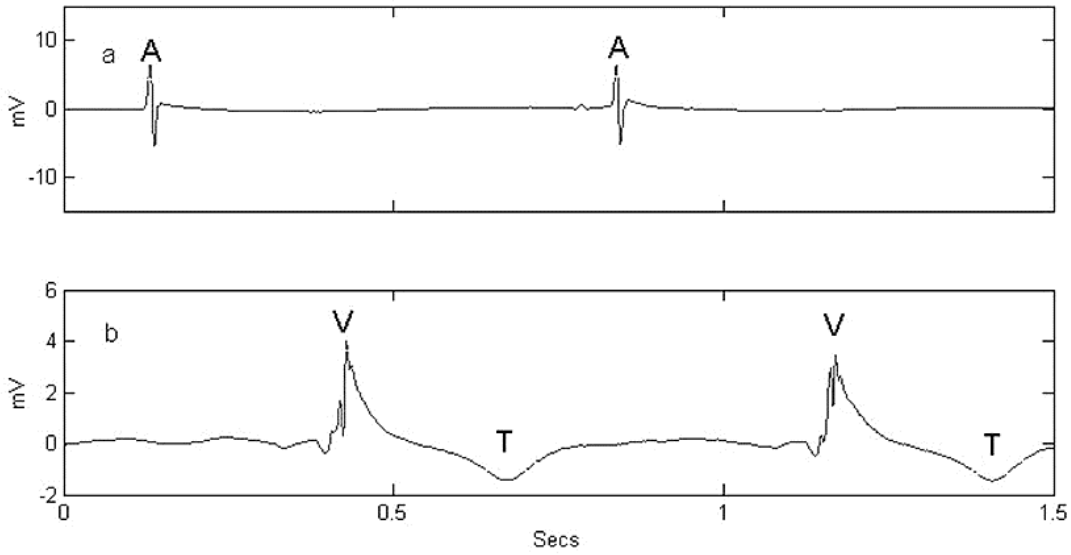

Figure 8.1 Unprocessed bipolar intracardiac electrograms showing component waves. a) Atrial electrogram with A waves labelled; b) Ventricular electrogram with

V and T waves labelled. Note: Electrograms show minimal visible "far-field" interference.

8.4.3 Power Spectral Analysis

As characteristics of recording instruments vary, power spectral analysis by fast Fourier transform (FFT) was performed on waveforms recorded on the Ensite system, using the Matlab *fft* function. The power spectra of atrial and ventricular bipolar electrograms were found to be similar, having major peaks at 10 Hz, 30 Hz, minor peaks at 50 Hz, representing power-line noise, and 70 Hz, and noise above 100 Hz (see Fig. 8.2). ΔZ spectra contained a peak at very low frequencies (1-5 Hz) representing respiration components and a small peak between 5 and 20Hz, representing cardiac components and high levels of noise above 30 Hz.

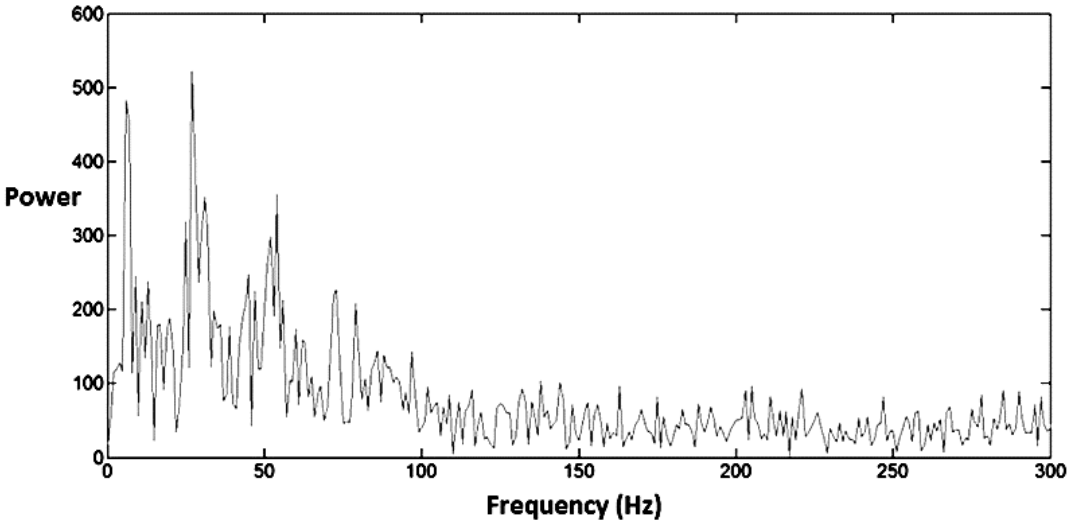

Figure 8.2 Power spectrum of a bipolar atrial electrogram. This shows that the majority of the spectral energy is located below 100Hz.

8.4.4 Filtering and Differentiation

Using the power spectral analyses, filters were selected to reduce unwanted components, while retaining desired signal content. Second order Butterworth filters were selected as having near flat frequency response and reduced computation time. Filters were designed using Matlab Signal Processing Toolbox *sptool* and the Filter Design and Analysis Tool *fdatool*. Filters produced a delay and correction for this was made later in processing.

Bipolar (near-field) electrograms had a 5 to 300 Hz band-pass (*BP_5_300.m*) filter and unipolar (far field) electrograms had 5 to 150 Hz band-pass *(BP_5_150.m*) filter applied, followed by

differentiation, using the Matlab *diff* function, to exclude baseline wander, far-field waves and high frequency noise.

Thoracic electrical bio-impedance decreases for each heartbeat (Patterson 1989), so absolute impedance (Z) requires inversion for conventional positive-going ΔZ waves. Following inversion, cardiac components of ΔZ (see Fig. 8.3 a)) were processed using a 1 to 30 Hz band-pass filter (*BP_1_30.m*) (see Fig. 8.3 b)). Differentiation then generated the derivative *dZ/dt* used for haemodynamic assessment, but also amplified high frequency noise (see Fig. 8.3 c)), requiring a further 12 Hz low-pass filter (*LP12.m*) to be applied (see Fig. 8.3 d)).

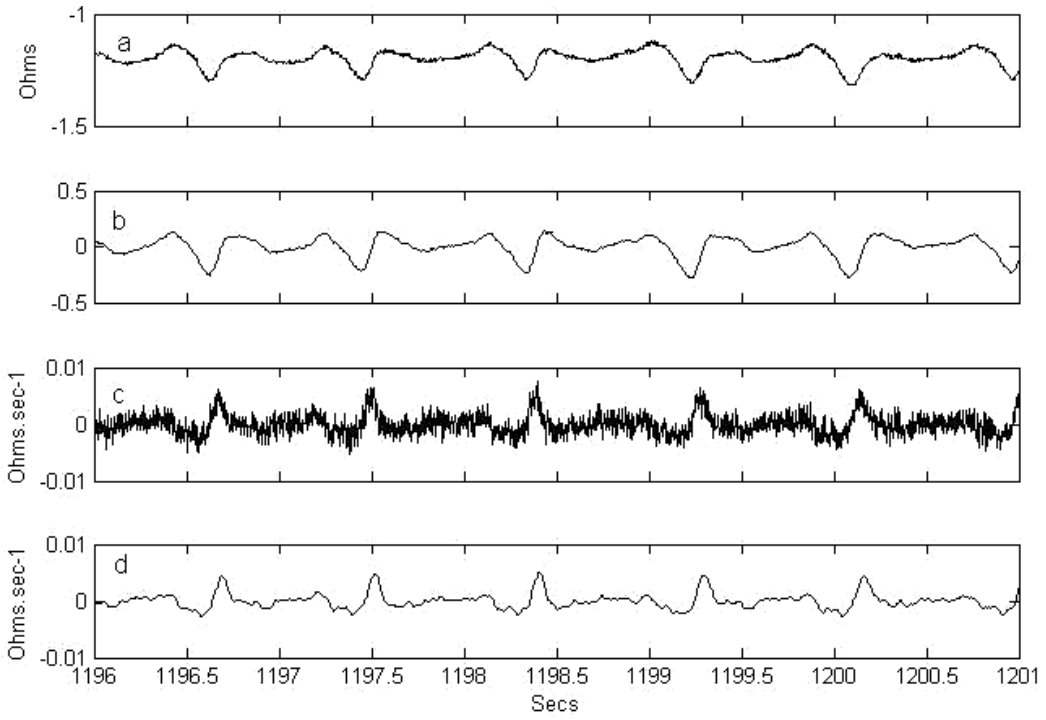

Figure 8.3 Stages in ΔZ cardiac component signal processing. a) unfiltered ΔZ; b) ΔZ with band-pass filtering; c) Differentiated ΔZ (*dZ/dt*); d) *dZ/dt* with secondary low-pass filtering applied.

Temperature and accelerometer data were acceptable in their raw form and did not require filtering.

8.5 Data Reduction

The aim of data reduction is to reduce data without significant information loss, using techniques such as dimensionality reduction (see Chapter 3, section 3.7), feature selection and feature extraction. Given that feature selection was the focus of Chapter 5, dimensionality reduction and feature extraction were the main focus of this section.

8.5.1 Clinical History and Examination Data

As feature selection was driven by known utility in consensus clinical guidelines, it was considered undesirable to reduce dimensionality of clinical history and examination data by grouping or dimensionality reduction.

8.5.2 Intracardiac Electrograms – Peaks and Fiducial Points

Open source Matlab code (Uysal 2011) for the Hamilton-Tompkins method (Hamilton & Tompkins 1986) was modified for this application. The method had the following steps: filtering, to remove noise, baseline wander, far-field V waves and T waves; differentiation, to amplify sharp slopes; rectification and moving window integration, to extract peaks and fiducial points.

Filtering and differentiation of atrial and ventricular bipolar (near-field) electrograms and "S3" unipolar (far-field) electrogram were performed as described in Chapter 6, subsections 6.6.3. Differentiated waveforms were then rectified using the Matlab *abs* function. Rectified waveforms were passed through a non-recursive finite impulse response (FIR) filter (convolution). The moving window integration (MWI) window size *w* was set to 100 msec, using principles suggested by Urrusti & Tompkins (1993) and an impulse response *h* was triggered. MWI was performed by continuous time case convolution integral using the Matlab *conv* function. Convolution increased output vector length by *w*-1 and these additional points were removed. Convolved waves, from a segment of waveform containing normal sinus rhythm near the start of each recording, were analysed manually by domain expert to select optimal sensing thresholds. Sensing thresholds required values higher than that at which over-sensing occurs and low enough for reliable peak sensing. Heuristically, this was optimal at 25% of the maximum value (see Fig. 8.4).

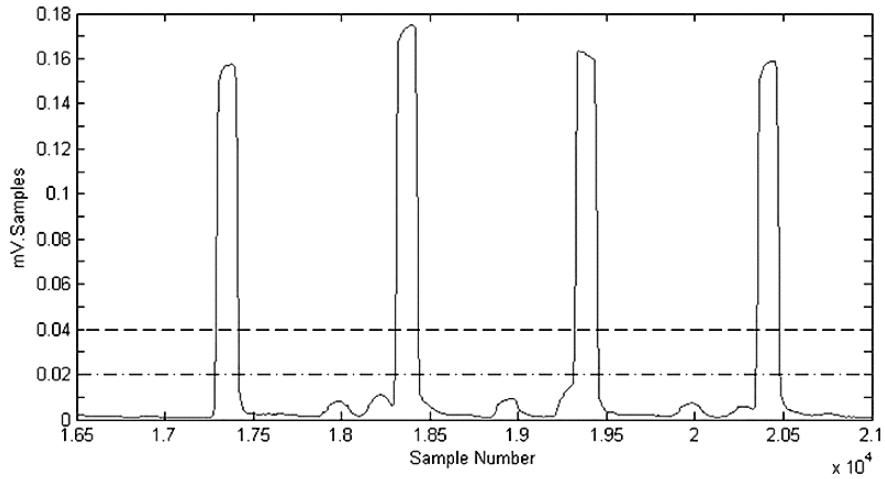

Figure 8.4 Peak detection thresholds. Upper dashed line is at the selected threshold value of 25% of the peak value of 0.16mV.samples, at approximately 0.04mV.samples. (Patient MM030).

In Fig. 8.4, over-sensing occurred below 0.02 mV.samples (and the threshold value was set using the maximum amplitude (0.16 mV.samples), at 25% or 0.04 mV.samples (higher broken line). The sensing algorithm was validated for each recording and adjusted where required.

From the convolved waveform, a logical array of points satisfying the threshold value was generated, with values of 1 within each detection window. The differential of this gave values of +1 at the start and -1 at the end of each window, used later in marking P and R waves for interval measurements. Using *find* functions, left and right sides of each window were located and a correction made for delay during band-pass filtering. A Matlab *for* loop, with the number of detected windows as the index value, identified exact locations and values of peaks within the windows. Fiducial points were then identified, referenced to peak locations. Peak and fiducial points were then verified manually by domain expert inspection (see Fig. 8.5 and 8.6).

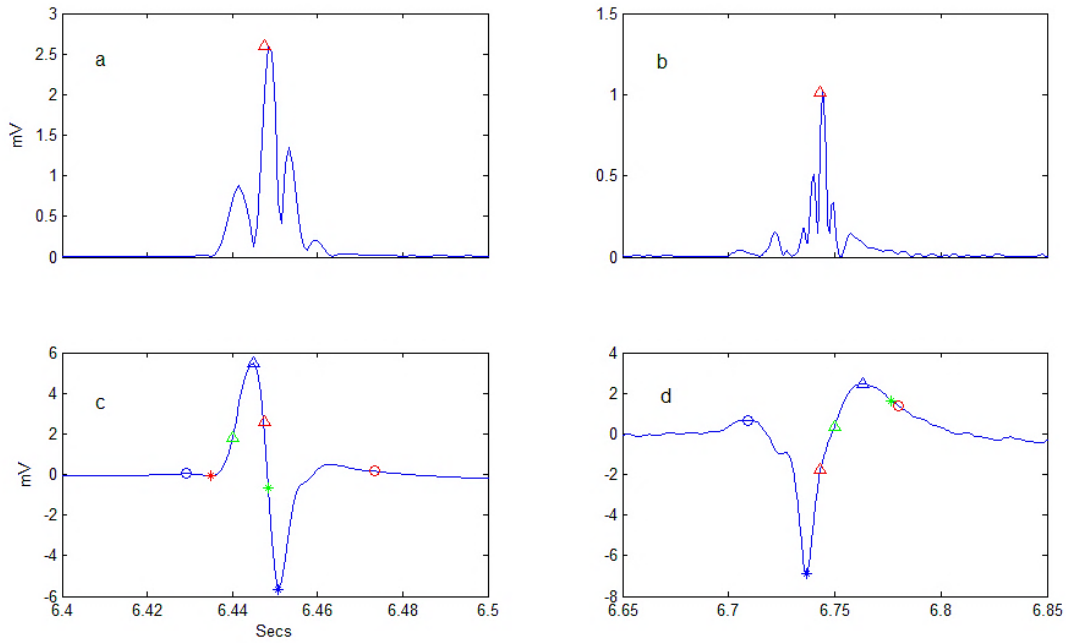

Figure 8.5 Intracardiac electrogram fiducial points. Rectified atrial a) and ventricular b) electrogram peaks (red triangles). Atrial c) and ventricular d) electrograms and fiducial points. Key: Blue o onset; red * first negative wave (Q); green Δ (Q) wave end; red Δ peak; blue Δ maximum positive detection; green * peak end; blue * second negative (S) wave; red o wave end. Note: d) has a QS complex so Q and S fiducial points appear simultaneous. (Patient MM030).

Dimensionality reduction was achieved by reduction of electrogram waveform files into vectors containing the location and value of 8 fiducial points. These contained sufficient information content to accurately reflect the waveforms.

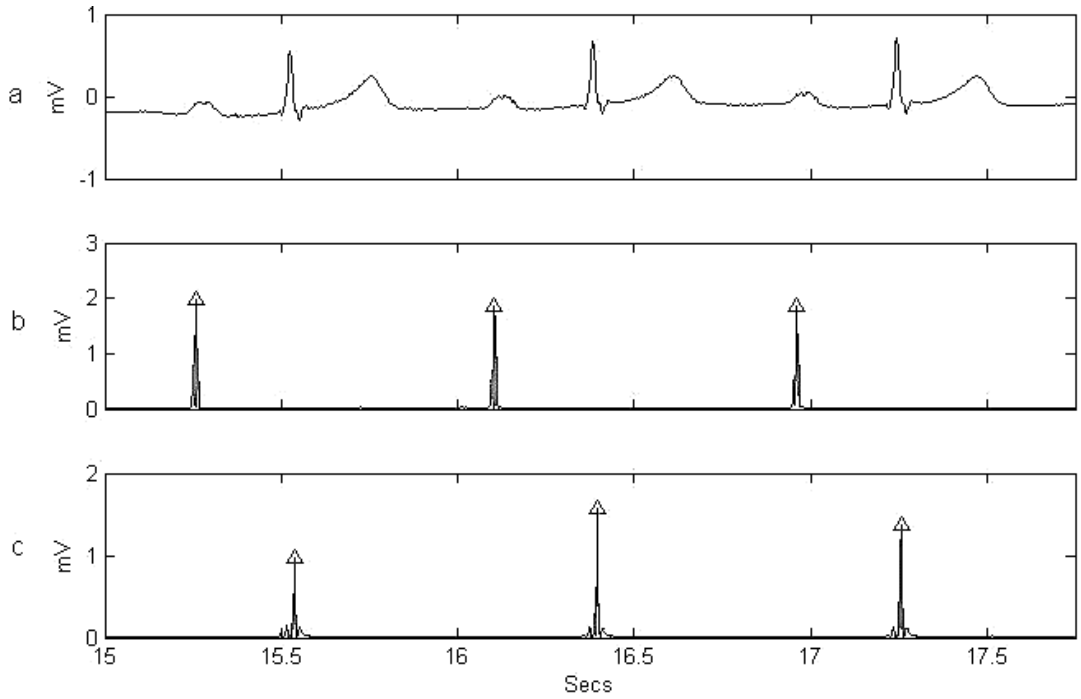

Figure 8.6 Detection of A and V wave peaks. a) ECG trace recorded for reference; b) atrial and c) ventricular rectified electrograms with detected peaks, marked with triangles. (Patient MM030).

8.5.3 T wave Detection

T wave end detection from the bipolar ventricular electrograms used open source MATLAB code *twaveend.m* (Zhang et al, 2006). Previous R wave peak detection, T wave polarity and window size were input. A modification to the code was made to automate T wave polarity input by use of the rectified V waveform, giving positive T waves. Moving window integration was performed, using a window size between the expected maximum V wave width and minimum RR intervals, 240 samples (200 msec) was selected, to limit detection to V wave widths < 200 msec or heart rates < 300 beats/ minute.

Correct detection of T wave ends was confirmed manually by a domain expert (see Fig. 8.7). The only data required from T waves was the location of the T wave end referenced to each R wave, reducing the information to a single short vector.

8.5.4 Impedance Cardiogram Wave Detection and Fiducial points

A Matlab script *ImpedanceCardiac.m* implemented the Naidu method (see Chapter 6, subsection 6.6.5) to detect fiducial points on the *dZ/dt* trace. Stroke volume index, could be extracted from B, C and X points for each beat, reducing the data to point values and locations. Fiducial point locations were confirmed by domain expert inspection. Output from this is illustrated in Fig. 8.8).

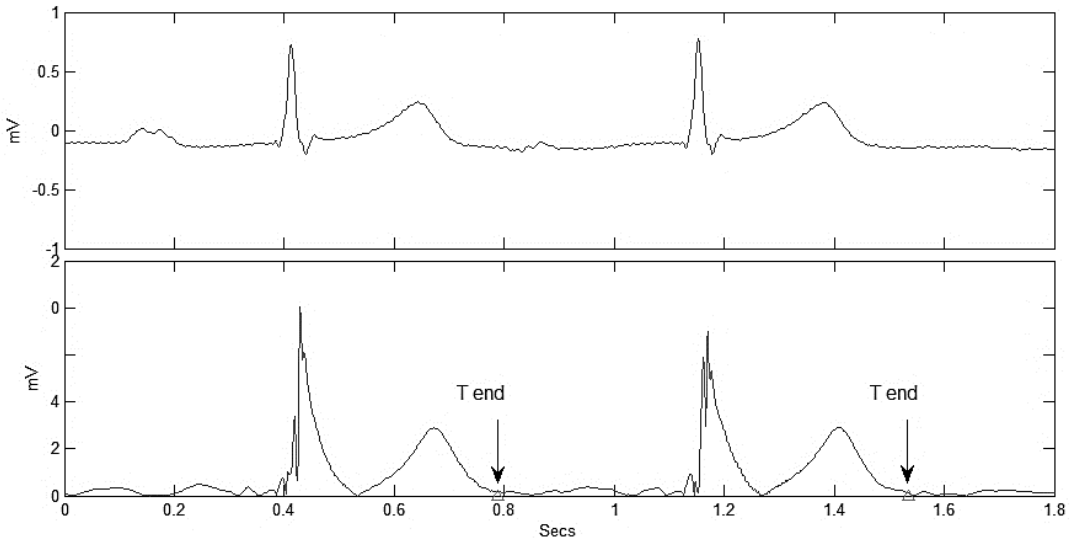

Figure 8.7 T wave end detection. ECG trace (top trace), and rectified ventricular electrogram (bottom trace) showing T wave ends (arrowed). (Patient MM030).

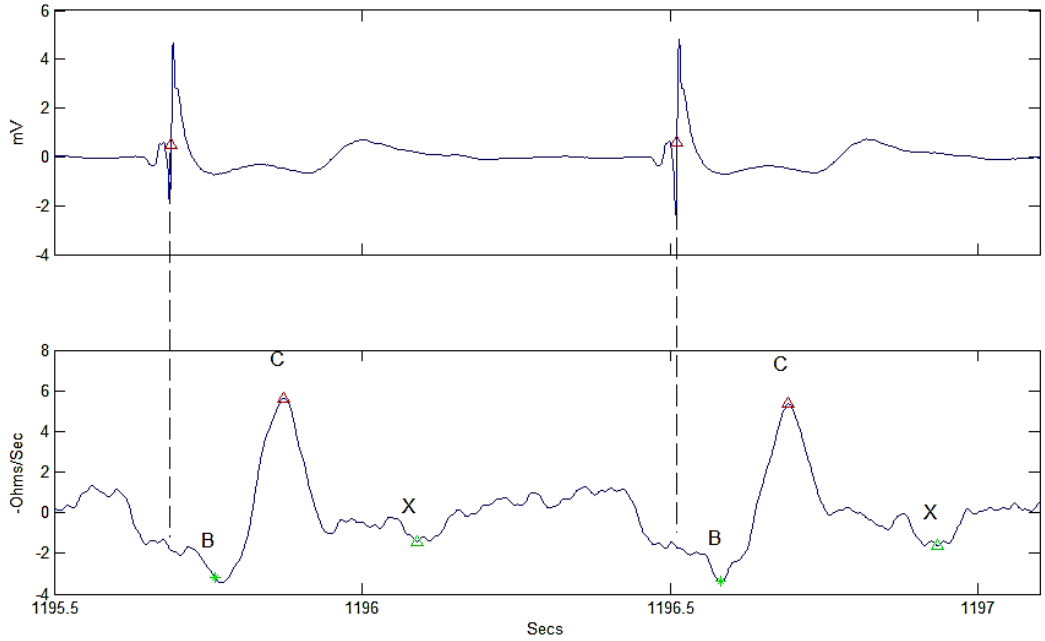

Figure 8.8 Impedance cardiogram fiducial points. Ventricular electrogram (top trace) showing R wave reference points (red trriangles) and impedance cardiogram *dZ/dt* (bottom trace) showing B points (Green stars), C (red triangles) and X (green triangles) waves. (Patient MM030).

8.5.5 Respiration Peak and Trough Detection

A respiration component was extracted from the ΔZ wave using a composite methodology, as there is no well-accepted method. Techniques of filtering (Houtveen *et al.* 2006), convolution (or

averaging) (Korten & Haddad 1989) and peak and trough detection (Wilson *et al.* 1982) were combined into a rational process. Peak detection used open source Matlab code *peakdet.m* (Billauer 2005). Moving averaging was performed using the *tsmovavg* function in Matlab. To allow the detection of a full range of respiratory rates (Rowland & Cunningham 1997), lag was set to 1100 samples (920 msec), for which 65 respirations per minute is the maximum detectable rate.

Within *peakdet.m* function, *delta* (of ΔZ) was set as the minimum variation of ΔZ for maxima and minima detection. The optimal value of *delta* was found heuristically to be 0.012 (equivalent to 1.2 Ohms). A phase shift caused by the filter delay, was less than one quarter of a breath and given that respiratory phase was not of interest in this study, compensation was considered unnecessary (see Fig. 8.9). The location and amplitude of peaks and troughs in each respiration cycle were confirmed by domain expert inspection and reduced the data required to represent respiration into four vectors.

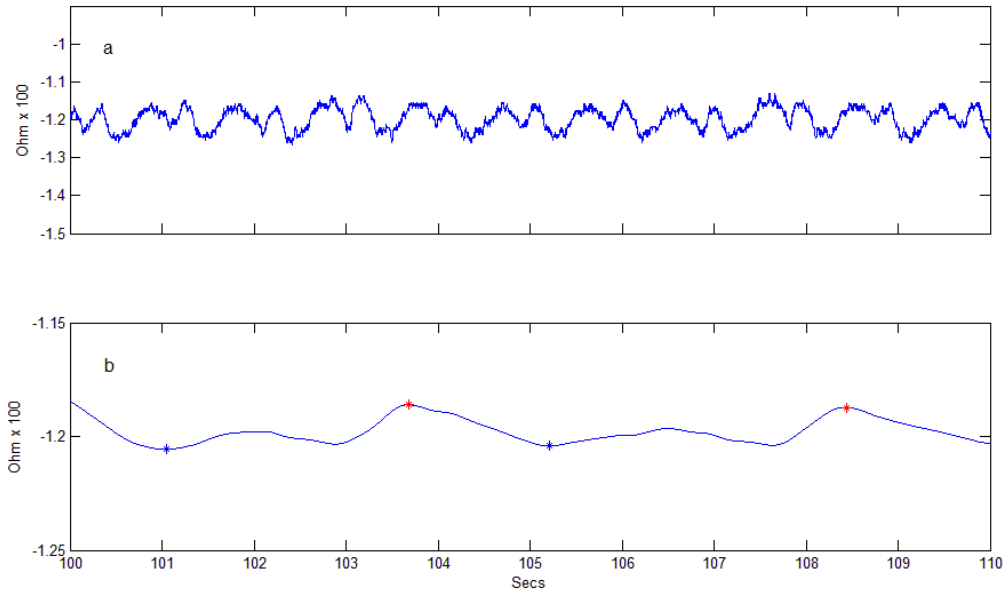

Figure 8.9 Respiration peak detection. (a) Unprocessed ΔZ; (b) Low-pass filtered ΔZ opitimised for respiration. Peak (red dots) and trough (blue dots) detections are shown. (Patient MM030).

8.5.6 Temperature and Body Motion Data Reduction

Moving averaging was performed on accelerometry and temperature data, using the Matlab *tsmovavg* function, eliminating unwanted high frequencies and reducing data, indicated by frequency bandwidths (see Chapter 6, subsection 6.7.2).

8.6 Data Discretisation

Data Discretisation was performed to divide the data into intervals ("bins") or concept hierarchies to supplement data reduction, reduce dimensionality and present a final set of features suitable for use as classifier inputs.

8.6.1 Clinical History and Examination Data

The Excel data containing clinical history and examination data (see subsection 8.4.1) was imported as a matrix to Matlab.

8.6.2 Electrogram Intervals

The Hamilton-Tompkins method for the detection of QRS waves from surface ECG's was implemented, adapted for use with intracardiac electrograms. Intracardiac A waves were detected instead of P waves and intracardiac V waves instead of QRS complexes or R waves and were considered synonymous in method descriptions. Peaks and fiducial points were detected, followed by detection of T waves (Hamilton & Tompkins 1986).

Using A and V wave markers (see subsection 8.5.2), beat-by beat AA interval, VV intervals and heart rate were automatically calculated arithmetically. AV ratio, representing the instantaneous value of the ratio of A and V rates, was calculated directly from the ratio of AA and VV intervals. Atrial rhythms were analysed using A events occurring within the ventricular (VV) intervals of a beat sequence. AV and VA intervals were derived with nested Matlab *for* and *if* loops. Intervals were converted from samples to time intervals using the sampling rate.

An indication from guideline driven feature selection was for a sample of 5 heart beat duration, having 2 beats preceding and 2 succeeding the first beat of an arrhythmia (see Chapter 3, subsection 3.18.2). To maintain comparability with ICD algorithms, a 10 beat sample was also indicated. A decision was to take 10 beat samples of data and to select a subset of that data to satisfy the 5 beat requirement (see also Chapter 10, section 10.5).

8.6.3 Electrogram Morphology

Electrogram morphology was assessed using near-field A wave electrograms and far-field V wave electrograms (S3) (see Chapter 6, subsection 6.6.3), between the detected fiducial points of wave onset and offset. Four methods were used: wave duration (width); axis (Euclidean vector direction); VTC and waveform area. Template-matching pattern recognition was used, with an example of normal rhythm analysed and stored, against which new beats were compared.

A 10 beat sample of mean values of the relevant features sampled during a period of normal sinus rhythm were as used as template values and stored for comparison. All remaining beat values

were compared with the template values and the difference expressed as a percentage or correlation coefficient allocated to each beat.

Wave widths were calculated from the arithmetic difference between wave onset and offset, converted to time intervals.

It was assumed that each electrogram was a vector, having magnitude and direction, and that the direction of vectors, taken from the relative locations of collecting electrodes, were known. It was also assumed that the vectors were arranged as an approximately right-angled triangle with the vector at S3 as its hypotenuse (see Chapter 6, Fig. 6.1) and that the triangle rule for addition of vectors, derived from Euclidean geometry, was applicable. The arctan inverse trigonometric function (8.1) was used to estimate the instantaneous direction of depolarisation θ, for A and V waveforms, from the amplitudes of S_1 and S_2 waveforms (see Chapter 6, Fig. 6.1 and Chapter 7, Table 7.1), given their directions were known.

$$\theta = arctan\left(\frac{S1_{net\ amplitude}}{S2_{net\ amplitude}}\right) \tag{8.1}$$

As absolute values were of less importance than relative change, correction for any inaccuracies of these assumptions was not made and the value of θ was used to provide an index of change of instantaneous vector direction.

Areas under the A and V waves were also estimated. As waveform samples were equally spaced, Newton-Cotes formulae were appropriate for approximating the definite integral with the rectangle mid-point rule being selected, as sampling rate was high and expected error low. The sample value became rectangle height, sampling interval width and the product of height and width was used to output the area. Summed area under the waves was termed "gross" area, irrespective of sign, absolute area, taking account of sign, was termed net area. Rectangular integration used the Matlab *sum* function. Output was rounded to aid subsequent calculation.

VTC was calculated using peak A and V waves as reference points and eight fiducial points were located for each beat. Using the Matlab *corr* function, Pearson's linear correlation coefficient was calculated, between template and new values for eight fiducial points, for each beat and converted to a percentage match.

8.6.4 Derivation of Corrected QT interval

Zhang's *twaveends.m* script was incorporated into a new Matlab script *qtinerval.m* which output corrected QT intervals (QT_c), calculated using Bazett's formula (5.1) (see Chapter 5, subsection 5.3.4). Given the known inaccuracies of Bazett's formula and that the intent was to demonstrate

change, the absolute value was of considered of no importance and the value of QT_c was used to provide an index of change of QT interval.

8.6.5 Haemodynamic Parameter Derivation

A baseline value of stroke volume index (SVI), measured by the PhysioFlow system, the patients height and baseline values of dZ/dt_{max} and BX interval were all inserted into the Sramek-Bernstein equation (6.4) to estimate the base impedance (Z_0) and calibrate further measurements. New values of BX time and dZ/dt_{max} for each beat were then inserted into the equation to calculate beat-by-beat stroke volume index (SVI), using a Matlab script *SramekBernstein.m*.

8.6.6 Respiration Derivatives

Using intervals between peaks, respiratory rate (*RR*) was calculated and from cyclic peak-to-trough depth, an index of tidal volume (*TV*) calculated. *TV* and *RR* product was calculated to produce a single measure of respiration, minute ventilation (*MV*).

8.6.7 Temperature

Temperature was presented as a single feature. In order to discretise, averaged values taken at the time of each beat (R peak) were stored as a feature vector.

8.6.8 Body Motion Data Discretisation

From up-sampled, averaged dual-axis accelerometry data, *xaxis* and *yaxis* variables were created and the Matlab *hypot* function used to calculate the RSS and generate an output *vmu.* Testing and validation of this data were not required since feature extraction was not performed.

8.6.9 Generation of Feature Sets as Classifier Inputs

Time-series feature sets, as with the measurements in this study, typically consist of a matrix, with rows representing time vectors of observations and columns representing the individual features. For supervised learning, an annotation vector, with rows exactly corresponding to the rows of the feature matrix is also required.

Meticulous beat-by-beat visual examination of rhythm for the entirety of each recording was made by a domain expert and then confirmed using concurrent notes made during data collection and with reference to the medical report of the procedure.

Each feature set contained data from 10 beat sequences corresponding to each rhythm occurrence (see Chapter 3, subsection 3.8.2). For each record, a 10 beat sample of normal sinus rhythm was selected for comparative purposes. If the patient was in a persistent arrhythmia, such as atrial fibrillation or flutter this sample was not taken. At each rhythm change within the record, 10 beat

samples were taken, with 5 beats preceding and 4 succeeding the initial beat (R_0) of a rhythm (see subsection 8.6.3). Where rhythms were continuous or R_0 was unclear or not captured, the sixth beat in a sequence was arbitrarily selected as R_0. For this study, a beat was defined as a V wave detection. Rhythm segments were selected for use as feature sets, provided all beats in the relevant 10 beat segment were reliably sensed. All available rhythm examples in each recording, where these criteria were satisfied were collected and each 10 beat segment retained for reference. 5 beat segments (see Chapter 3, subsection 3.8.2) could be derived as a subset of each 10 second sequence and were obtained in a later phase of processing (see Chapter 10, section 10.5).

For each record, the reliable detection of A and V waves, T wave end markers and impedance cardiac and respiration waveforms without evidence of significant over-sensing or under-sensing were confirmed by a domain expert. Detection of A and V wave peaks was made from the concatenated waveforms (see subsection 5.2.1 and Fig. 5.1). Nominal sensitivity values were set to > 20% of mean peak value during 8 seconds of normal sinus rhythm. Respiration detection sensitivity was nominally set to 1.2 Ohms (see subsection 8.5.5). These sensitivities were adjusted where required, to individually optimised values. For each 10 beat segment, 20 features for each beat and 12 clinical history features (see Chapter 6, subsection 6.6.1, subsection 8.4.1 and Table 8.2) were concatenated into vectors of 212 features.

A feature matrix was generated with rows for each rhythm segment. A single column vector was used for rhythm annotation with rows matched with the corresponding feature matrix row.

8.7 Summary

The Zhang algorithm for data preparation was used, where data collection is followed by data cleaning, integration, transformation, reduction and discretisation to produce a suitable set of features for use in classification systems.

Data cleaning was performed using manual editing of text files. There was little missing data demographic and history and Waveform data was found to be of generally high quality with little missing data. From temperature and accelerometer data header rows and columns containing unwanted data or data containing unacceptable levels of interference were deleted. Missing data was replaced wherever practical with interpolated data.

Data integration of waveform data with the remaining data was performed using time synchronisation, up-sampling with linear interpolation and concatenation.

Data transformation was achieved by digital signal processing (DSP) together with differentiation, without normalisation, to improve signal quality of waveforms. Clinical history

Table 8.2 Feature vector component features.

Feature Number	Feature Description	Vector Name
1	AA interval	PPint
2	AV interval	PRint
3	VV interval	RRint
4	VA interval	RPint
5	A/V rates ratio	PRratio
6	A wave axis percent template match	PAXISmatchsamples
7	V wave axis percent template match	QRSAXISmatchsamples
8	A wave duration percent template match	Pwidthmatchsamples
9	V wave duration percent template match	QRSwidthmatchsamples
10	A wave gross area percent template match	Pgrossareamatchsamples
11	A wave net area percent template match	Pnetareamatchsamples
12	V wave gross area percent template match	QRSgrossareamatchsamples
13	V wave net area percent template match	QRSnetareamatchsamples
14	A wave VTC percent template correlation	VTCAsamples
15	V wave VTC percent template correlation	VTCsamples
16	Corrected QT interval	QTcsamples
17	Minute ventilation (relative)	MVsamples
18	Stroke volume index	SVIsamples
19	Body Temperature	TEMP
20	Velocity motion index (VMU)	ACCEL
Features 1 -20 repeated for a total of 10 beats		
201	Tiredness or lethargy symptom	-
201	Chest pain symptom	-
203	Shortness of breath symptom	-
204	Dizziness or blackouts symptom	-
205	Palpitations symptom	-
206	Abnormal heart sounds sign	-
207	Cardiac or pulmonary disease clinical history	-
208	Antiarrhythmic medication Prescribed	-
209	Tobacco consumption	-
210	Alcohol consumption	-
211	Caffeine intake	-
212	Metabolic conditions	-

data was encoded and rationalised. Unwanted signal content in the waveforms was identified and power spectral analysis of recorded waveforms directed choice of filter. Second order Butterworth

filters were used exclusively.To reduce data without significant information loss, data reduction was performed, primarily by feature extraction. Fiducial points, sufficient for subsequent processing into features, were extracted using Matlab scripts for intracardiac electrograms and thoracic bio-impedance, reducing the waveforms into vectors containing their location and value.

In a process of data discretisation, the data points obtained during transformation, final features were extracted, such as electrogram intervals, morphology, respiration and haemodynamic features. Pre-processing computation time, before concatenation, was approximately 20secs of data per second of processing time (desktop PC running Windows 7 with Intel Core 2 Quad 2.8 GHz CPU, 4GB RAM). A time-series matrix of features was produced for each patient, divided into annotated rhythm segments, in preparation for use as classifier inputs. Allowing for some patients who had multiple recorded segments, this produced 75 files with a total size of 3.11GB.

Chapter 9 Data Collection Results

9.1 Overview

Results of the data collection study were summarised in terms of the demographics of patients from whom data was collected, rhythms detected and data quality.

9.2 Demographics of Patients Studied

Between June 2009 and April 2010, 495 patients underwent electrophysiology procedures, with 273 satisfying the inclusion criteria (see Chapter 6, subsection 6.5.14). 67 patients (28%) were approached to participate in the study, 65 of whom consented and participated in the data collection part of the study. There were 4 (6%) failed recordings, from which no data was obtained (patients 3, 4, 5 and 8). Many of the procedures proceeded to therapy for an arrhythmia, using radiofrequency ablation, however this was not specifically documented as it was not relevant to this data collection exercise. No patient studied had significant complication and there were no adverse events.

Of 61 (94%) patients from whom successful recordings were made, 30 (49.2%) were male, of age 42.2 ± 16.3 years with an age range of 16 to 77 years, had a mean body surface area of 1.90 ± 0.25 m^2 and body mass index of 25.8 ± 4.9 kg/ m^2 (mean ± standard deviation). 22 (36%) patients had a known clinical history of cardiac disease; 29 (48%) had significant non-cardiac clinical history; 19 (31%) had a family history of heart disease; 16 (26%) smoked tobacco; 45 (74%) patients admitted alcohol use of whom 1 admitted regular consumption of 3 units or more per day; 8 (13%) had previously diagnosed hypertension; 4 (7%) had diabetes mellitus; 13 (21%) had a history of hypercholesterolemia and 5 (8%) had known significant coronary arterial disease. 55 (90%) patients reported symptoms of palpitations; 10 (16%) reported previous syncope; 19 (31%) dizziness; 6 (9%) reported shortness of breath; 6 (9%) chest pains and 4 (6%) tiredness or lethargy. No patients were asymptomatic; 32 (52%) patients complained of 1 symptom; 20 (32%) patients 2 symptoms and 9 (14%) patients 3 symptoms.

9.3 Rhythms Detected

From the 61 successful recordings, 1109 examples of rhythms (instances) were collected. Within these, 20 (61%) of 33 possible rhythms were represented (see Table 9.1). Each of the 7 rhythm groups were represented: 110 instances of normal rhythms; 783 premature beats; 29 of sinus node dysfunction; 5 of atrio-ventricular block; 108 narrow complex tachycardias; 6 broad complex tachycardias and 78 instances of paced rhythms.

Using data from the study of Zeldis *et al.* (1980) (see Chapter 6, subsection 6.4.3) equivalent proportions were calculated for this study and a comparison made of proportions of arrhythmia incidence found in this study. Proportions were found to be similar (see Table 9.1).

Table 9.1 A comparison of rhythm occurrence in this study with published data. Data from Zeldis *et al.* (1980).

Rhythms	**Zeldis *et al.* (1980)**	**This study**
n (number of patients)	581	61
Premature Ventricular Complexes (PVC)	362 (69%)	48 (78%)
Premature Atrial Complexes (PAC)	226 (44%)	48 (78%)
Sick Sinus Syndrome (SB/SA/SAB)	10 (2%)	2 (3%)
Heart block (SHB2/CHB)	0 (0%)	0 (0%)
Supraventricular Tachycardia (AVNRT/OAVRT)	54 (10%)	18 (28%)
Atrial Flutter (MRAT)	11 (2%)	6 (9%)
Atrial Fibrillation (AF)	28 (5%)	9 (14%)
Ventricular Tachycardia (VT)	40 (7%)	4 (6%)
No arrhythmias	92 (17%)	2 (3%)

A reduced number of patients with no arrhythmias in Zeldis *et al.* (1980) was contrasted with increased percentages of patients with PVC, PAC, SVT, MRAT and AF in this study. These differences may largely be explained by difference between the contrasting nature of non-invasive 24-hour Holter recording and EP studies where electrodes placed in the heart cause mechanical irritability as well arrhythmia provocation during pacing and pharmacological manoeuvres. This comparison serves to support that data in this study reflects the natural class distribution.

Rhythms not represented were: sinus arrest, sino-atrial block, Mobitz type II second degree heart block, complete atrio-ventricular block, postural orthostatic tachycardia syndrome, sinus node re-entry tachycardia, focal junctional tachycardia, non-paroxysmal junctional tachycardia, permanent junctional reciprocating tachycardia, SVT with aberration, antidromic atrio-ventricular reciprocating tachycardia, outflow tract ventricular tachycardia and ventricular fibrillation.

Rhythm classes were condensed based on rhythm group and common diagnostic characteristics. Grouped rhythms were: sinus node dysfunction; second and third degree atrio-ventricular blocks; atrio-ventricular nodal and junctional tachycardias; atrio-ventricular re-entry tachycardias; abnormal tachycardias of the sinus node and ventricular tachycardias. SVT with aberration was absorbed into more specific groups. Grouping of classes resulted in 19 classes and included one

empty class, ventricular fibrillation (see Table 9.2). In this study, grouping resulted in loss of class resolution only between variants of VT.

Table 9.2 Rhythm instances and abbreviations.

Rhythm	Abbr.	Group	Total
Normal sinus rhythm	NSR	% Normal rhythm	52
Respiratory sinus arrhythmia	RSA		49
Physiological sinus tachycardia	ST		9
Premature atrial complex	PAC	% Premature beats	397
Premature ventricular complex	PVC		386
Sinus bradycardia	SB	% Sinus node dysfunction	29
Sinus arrest	SA		0
Sino-atrial block	SAB		0
First degree atrio-ventricular block	FHB	% Atrio-ventricular block	3
Second degree AV block (Mobitz type I/ Wenckebach)	SHB		2
Second degree AV block (Mobitz type II)	SHB2		0
Complete atrio-ventricular block	CHB		0
Postural orthostatic tachycardia syndrome	POTS	% Narrow complex tachycardias	0
Atrio-ventricular nodal reciprocating tachycardia	AVNRT		27
Orthodromic atrio-ventricular reciprocating tachycardia	OAVRT		12
Permanent junctional reciprocating tachycardia	PJRT		0
Focal atrial tachycardia	AT		8
Focal junctional tachycardia	FJT		0
Non-paroxysmal junctional tachycardia	NPJT		0
Sinus node re-entry tachycardia	SNRT		0
Inappropriate sinus tachycardia	IST		3
Macro-re-entrant atrial tachycardia	MRAT		8
Multifocal atrial tachycardia	MAT		17
Atrial fibrillation	AF		23
SVT with aberration	SVTAB	% Broad complex tachycardias	0
Antidromic atrio-ventricular reciprocating tachycardia	AAVRT		0
Monomorphic ventricular tachycardia	VT		2
Polymorphic ventricular tachycardia	PVT		1
Idiopathic fascicular ventricular tachycardia	IFVT		3
Outflow tract ventricular tachycardia	OTVT		0
Ventricular fibrillation	VF		0
Atrial paced rhythm	APACE	% Paced rhythms	48
Ventricular paced rhythm	VPACE		30

Rhythm class frequencies were: 52 instances of normal sinus rhythm, 49 of respiratory sinus arrhythmia, 9 of physiological sinus tachycardia, 397 atrial contractions, 386 premature ventricular contractions, 29 of sinus node dysfunction, 3 of first degree atrio-ventricular block, 2 of 2nd and 3rd degree atrio-ventricular block, 27 atrio-ventricular nodal and junctional tachycardias, 12 atrio-ventricular reciprocating tachycardias, 3 abnormal tachycardias of the sinus node, 8 focal atrial tachycardias, 8 macro-re-entrant atrial tachycardias, 17 multifocal atrial tachycardias, 23 atrial fibrillation, 6 ventricular tachycardias, 48 atrial paced, 30 ventricular paced (see Table 9.2).

9.4 Data Quality

Data quality was measured by assessing the completeness, validity, consistency, timeliness and accuracy of the data.

9.4.1 Data Completeness

Of the 61 patient recordings analysed, 22 (36%) patients had incomplete data, including 1 with no impedance data, 4 (7%) no temperature data and 8 (13%) no accelerometer data, 11 (18%) had partial impedance or accelerometer data. All data loss was attributable to technical failures. Incomplete data was entered as NaN (Not a Number) in Matlab, allowing analysis of empty cells, while satisfying matrix requirements. Each of the 1109 feature vectors had 212 values, a total of 235108 data items. Of these, 836 (75.4%) instances were complete with no missing data and 273 (24.6%) instances contained 3480 (1.48%) empty data (NaN) values, resulting in an overall data completeness of 98.52%.

Detailed analysis of the 3480 empty values showed no missing data were due to electrogram data; 350 empty values (0.15% of all data, 3.2% of temperature data) were due to unavailable temperature data; 1710 empty values (0.7% of all data, 15.4% of accelerometry data) were due to unavailable accelerometer data; 1340 empty values (0.57% of all data, 6.0% of impedance data) were due to unavailable impedance data and 80 empty values (0.03% of all data, 0.6% of clinical history data) were due to unavailable clinical history data.

There was partial data where it would be possible to derive a patient-specific mean value, for 491 empty values (14% of all empty cells) from 10 patients, consisting of 170 temperature values (49% of empty temperature values, 1.5% of temperature data) and 321 accelerometry values (19% of empty accelerometry values, 2.9% of accelerometry data). This was not the case for any missing impedance or clinical history data.

In this study, temperature was used as an indicator of metabolic stress and accelerometry to detect physical stress, as body motion. Neither indicator was considered critical to diagnosis but had

potential relevance to the context of a rhythm, such as sinus tachycardia at rest. In the absence of real data, given the low percentages of missing data, use of mean feature values as interpolated values were not expected to introduce significant error.

9.4.2 Data Validity

Data was examined manually during pre-processing for allowed characters and values within the expected range. Matlab generated error messages wherever there were significant data processing errors, allowing for their correction, beyond this, automated validity checks were not considered useful.

9.4.3 Data Consistency

Archived recordings were pre-processed and examined for consistency of the data through the various processes. There was staged backup and replication of data in this study with each phase of pre-processing. Where data copies were made, they were examined to ensure that reconstructed data remained a valid representation of the physiological conditions.

9.4.4 Data Timeliness

Given that the data examined was largely time-series in nature, timeliness was a critical component. Data sequences were rigorously examined during data combination to confirm sampling appearance at the expected interval.

9.4.5 Data Accuracy

Data accuracy, as the degree to which the data correctly reflected the real world of the representation of the physiological conditions was assessed by a domain expert. This was performed manually using visual validation of each instance as a correct and true representation. Annotation of rhythm was appended as part of this process.

An objective assessment of instance accuracy was made by checking that heart rates of tachycardias bradycardias and normal sinus rhythm for adherence to heart rate limits of above 100, below 60 and between 60 and 100 beats per minute, respectively.

9.5 Factors Affecting Data Analysis

Data dimensionality, imbalance and partitioning strategies are known to affect outcome in classifier development and were considered.

9.5.1 Dimensionality

Classifiers are known to perform well with data of low dimensionality. In this study, data had high dimensionality, with 212 features. Dimensionality reduction can be approached in a number of ways.

Collapsing or re-binning, **c**ombining results from two or more columns, with eliminated cell results being combined with other cells. Collapsing entails a loss of information of potential interest and was considered undesirable. Alternatively, a tiny number (such as 0.000000001) may be added to all cells in the table to facilitate convergence however this can give misleading results. Disregarding sparse cells by considering only cells with observed or expected frequencies above a certain value also leads to loss of information and unforeseen results.

An alternative is pruning, where features or aspects of the classifier are removed without significant loss of accuracy. This can be pre-pruning, during classifier development where features are added until maximum accuracy is achieved or post pruning where features are removed until loss of accuracy is detected.

9.5.2 Data Imbalance

This study had a heavily imbalanced dataset, with 783 instances (71%) falling in the two majority classes, consisting of 397 instances of PAC and 386 of PVC. This imbalance made it likely that any successful classifier would over-train for the majority classes and less likely to correctly classify rare classes.

A data partitioning strategy (see Chapter 3, section 3.8), advocated use of a hold-out sampling technique should the dataset be sufficiently large or *k*-fold cross-validation if not. A hold-out partitioning strategy was inappropriate for this dataset, since rare instances could potentially be placed in either training or test set, risking poor performance. In medical diagnosis minority classes are often classes of interest so their accurate classification is of importance, an effective strategy for dealing with this imbalance in the data was sought and techniques which improved performance while avoiding over-training, were considered.

Sampling techniques rebalance data, either oversampling to synthesise data or under-sampling to ignore certain instances. Oversampling was considered undesirable in this domain, where data provenance and integrity are valued and interference discouraged. Alternatives include resampling methods such as cross validation (Bishop 1995, pp.372-375) and bootstrapping. Bootstrap increases variance compared with cross-validation. Kohavi (1995) recommended 10-fold cross validation over bootstrapping for a wide range of data, but Hastie suggested the two methods provide similar results (Hastie *et al*. 2008, pp.253-254). 10-fold cross-validation was

confirmed as the preferred approach to data partitioning. Under-sampling reduces proportions of majority classes but risks training using an unrepresentative sample. In this study, the two majority classes were not considered "high cost", with mis-classifications not critical to classifier performance, as these rhythms were neither debilitating nor life-threatening, further resampling of under-sampled instances was considered unnecessary, though this could be revisited should minority classification prove poor.

The full dataset was resampled using under-sampling, using the Matlab *datasample* function. There were 326 instances for 16 minority classes, a mean sampling frequency of 20. Minority instances were retained while the majority classes were randomly sampled with 20 instances of each, to be in proportion, producing an under-sampled dataset of 366 instances. Both full and under-sampled datasets were used as alternates during classifier development and testing and cross-validation during training to generate classifier models.

9.5.3 Empty Classes

Of 19 possible rhythm classes (see Table 9.2), there was 1 empty class, with no examples recorded. As this study sought to classify rhythms of all types, classification of unlearned rhythms was required. Given the availability of extensive guidelines for cardiac rhythm diagnosis an inference engine would be created to offer an unseen diagnosis. This inference engine would sit in a multi-classifier configuration, offering an alternate diagnosis. Final class label would be decided using a combining scheme.

9.6 Summary

Of 65 patients consented for data collection, 61 data sets were generated, resulting in the production of 1109 feature vectors or instances. Data quality was evaluated, with good data completeness (98.5%), satisfactory accuracy and validity, using visual confirmation by a domain expert for each instance generated. Annotated instances generated by the data collection process were used as classifier inputs.

Chapter 10 System Development and Testing

10.1 Overview

Classifiers selected for implementation were decision trees, fuzzy inference (see subsection 10.2.2), naïve Bayes, neural networks, support vector machines and an inference engine (see Chapter 4, section 4.5). Using features previously described (see Chapter 5, section 5.6) as inputs, classifier performances represented the first iterative stage in the system development life-cycle.

Classifier units were designed and tuned, specifically for the data collected in this study, using a heuristic process. Classifiers were then evaluated for performance indices, taken as an estimate of generalisability. A guideline-based inference engine was designed as a "catch all" for empty classes, unlearned rhythms and unclassified instances. To maintain simplicity of design and encourage adoption, a one-classifier-for-all, rather than one-classifier-per-class design principle was preferred, though this was subject to findings during classifier development.

Comparisons between classifiers were made using classifier error, sensitivity, specificity, κ and P value from a two-tailed Fisher's exact test (see Chapter 3, sections 3.12 and 3.14). For a 5-way comparison with a Bonferroni correction, the critical P value (P_{crit}) was set as 0.010, equivalent to $P < 0.05$.

Initial classifier development using data from an under-sampled data set (see Chapter 9, subsection 9.5.2), as that reduced over-training for the majority PAC and PVC classes. 10-fold cross-validation was used to train classifiers and generate models, other than for the inference engine and the support vector machine, where this was not practical. The classifier generated from the fold with best performance indices was selected and re-tested with all available data.

Performance assessments made during iterations influenced proposals for design modifications. Iterations continued, with modifications, until achievement of target performance.

Classifier combination strategies used the best performing classifiers. Where user requirements were considered met the final classifier model became the production system.

10.2 Classifier Design and Iteration 1

Classifiers were designed using the Matlab Statistics Toolbox for decision tree, naïve Bayes and support vector machine classifiers, Neural Network Toolbox for neural network classifier and a script written as an inference engine classifier.

Class data was multinomial and it was found that neural network classifiers would not easily operate with the chosen class coding. Binarisation of class data was performed, resulting in one column per class, in preparation for use in neural network classifiers with 212 inputs for ANN training. For inputs to support vector machines, Matlab functions did not readily accept empty or NaN values, data was corrected, with empty cells replaced with feature mean values. In the text, data with no imputed values was termed "uncorrected" and data with imputed values for empty or NaN cells was termed "corrected".

Classifier performance was assessed as an estimate of classifier generalisability. Model selection strategy was to avoid models with over-training and while selecting the model with the best performance indices and using that model to derive new performance indices for all instances as an estimate of generalisability.

Classifier models were tuned during development using the options available in Matlab, scored for the number of performance index values best for that model. Where fold models had the same "score", high κ was taken as indicative of acceptable "trade-off" between sensitivity and specificity.

10.2.1 Statistical Testing of Classifier Performances

In this study, each specific rhythm diagnosis was considered to have its own diagnostic test, in a one-versus-all (OVA) classification. In this context, a 2 x 2 contingency table (see subsection 3.12.2 and Table 3.2) was constructed for each rhythm diagnosis, with the gold standard provided by domain expert diagnosis, in this case the domain expert was a cardiac electrophysiologist in a United Kingdom teaching hospital. A TP was counted when a specific rhythm diagnosis was made and confirmed by the domain expert; a TN when the specific rhythm diagnosis was not made and this was confirmed by a domain expert; a FP when a specific rhythm diagnosis was made and the domain expert made an alternative diagnosis; a FN when a specific rhythm was not diagnosed but the domain expert determined that the rhythm was present.

The statistical measures used were chosen in line with the conclusions from Chapter 3, sections 3.13 to 3.15. They were: correct classification rate (*CCR*); error; sensitivity; specificity; Cohen's kappa (κ) and *P* value, calculated using the Fisher exact test. Additional coefficients: prevalence; positive predictive value (*PPV*); negative predictive value (*NPV*); odds ratio (*OR*) and its derivatives relative risk (*RR*) and Yule's *Q*. Coefficients of association, Pearson's phi (ϕ), an equivalent of Pearson's product-moment correlation coefficient *r* when applied to contingency tables, and the F_1 score were used. Type I (α) and II (β) errors were also calculated.

For this study, with an OVA classification, correct classification rate (*CCR*) was defined as the proportion of all rhythms correctly classified, from the total of all rhythm instances *N*, according to the Gold Standard of expert cardiac electrophysiologist diagnosis (10.1).

$$CCR = \frac{TP+TN}{N} \tag{10.1}$$

Error was defined as the proportion of all rhythms incorrectly classified, according to the Gold Standard of expert cardiac electrophysiologist diagnosis (10.2).

$$Error = \frac{FP+FN}{N} = 1 - CCR \tag{10.2}$$

Sensitivity or true positive rate, was defined as the proportion of all rhythms that were correctly classified as having that rhythm diagnosis (10.3).

$$Sensitivity\ (true\ positive\ rate, recall, power)\ = \frac{TP}{TP+FN} \tag{10.3}$$

Specificity or true negative rate, was defined as the proportion of all rhythms with an alternative diagnosis that were correctly classified (10.4).

$$Specificity\ (true\ negative\ rate) = \frac{TN}{TN+FP} \tag{10.4}$$

For *CCR*, sensitivity and specificity, values of 1 represent perfect agreement and values approaching zero poor results. For error, values of 0 represent perfect agreement.

Cohen's kappa (κ) is a measure of agreement, calculated using the marginal (row and column) totals of the 2 x 2 contingency table (10. 5), including all four proportions.

.

$$\kappa = \frac{(P_o - P_e)}{(1-P_e)} \tag{10. 5}$$

where $P_o = \text{agreement observed } = \frac{(TP+TN)}{(TP+FP+FN+TN)}$

and $P_e = \text{agreement expected due to chance} = (r1.c1) + (r2.c2)$

where $r1 = row\ 1\ marginal\ proportion\ (total\ test\ positive) = \frac{TP+FP}{N}$

and $c1 = column\ 1\ marginal\ proportion\ (total\ G.S\ positive) = \frac{TP+FN}{N}$

and $r2 = row\ 2\ marginal\ proportion\ (total\ test\ negative) = \frac{FN+TN}{N}$

and $c2 = column\ 2\ marginal\ proportion\ (total\ G.S\ negative) = \frac{FP+TN}{N}$

κ can have values of between 0, +1 and -1. Values of 0 are found where agreement is no more than expected by chance, 1 for perfect agreement and -1 for perfect disagreement. According to Altman (1991), κ values of less than 0.2 represent poor agreement; between 0.21 and 0.4 fair agreement; between 0.41 and 0.6 moderate agreement; between 0.61 and 0.8 good agreement and greater than 0.81 very good agreement. Paradoxes of the κ statistic have been identified, such as where there is good agreement associated with low values and where the value changes unpredictably with variation of marginal totals. Most observers agree that κ should be supported with other measures

With Fisher's exact test, *P* values are computed directly from contingency tables (10.6).

$$P = \frac{(TP+FP)!(FN+TN)!(TP+FN)!(FP+TN)!}{TP!FP!FN!TN!(TP+FP+FN+TN)!} \quad (10.6)$$

P values from Fisher's exact test tend to be optimistic, so for comparison of multiple classifiers, the Bonferroni correction is most commonly applied. This correction alters the critical *P* value (P_{crit}) from α (usually 0.05), dividing it by the number (n) of comparisons to be made (10.7).

$$P_{crit} = \frac{\alpha}{n} \quad (10.7)$$

The prevalence (10.8) of a rhythm was defined as the proportion of instances of a specific rhythm from the total *N* (10.8), not the prevalence of the rhythm in the population. The ideal prevalence would be where there were equal proportions of all rhythms being analysed.

$$Prevalence\ (\approx\ prior\ probability = P(H)) = \frac{TP+FN}{N} \quad (10.8)$$

A positive predictive value (*PPV*) was defined as the proportion of true positives from all positive results for a rhythm (10.9) Likewise, a negative predictive value (*NPV*) was defined as the proportion of true negatives from all negative results for a rhythm (10.10). Ideal predictive values correspond to absence of false diagnoses, with all positive and negative diagnoses correct, *PPV* and *NPV* having values of 1.

$$PPV\ (precision) = \frac{TP}{TP+FP} \quad (10.9)$$

$$NPV = \frac{TN}{TN+FN} \quad (10.10)$$

The diagnostic odds ratio (OR) measures the effectiveness of classification (10.11) by the ratio of the odds of being positive for a specific rhythm if the rhythm is present, to the odds of the test being positive if the rhythm is absent. Odds ratio may also be calculated using the ratio of positive (*LR*+) to negative (*LR*-) likelihood ratios.

$$OR = \frac{(TP/_{FN})}{(FP/_{TN})} = \frac{LR+}{LR-} \tag{10.11}$$

where $LR+\,(Likelihood\ ratio\ of\ a\ positive) = \frac{Sensitivity}{1-Specificity}$

and $LR\text{-}\,(Likelihood\ ratio\ of\ a\ negative\,) = \frac{1-Sensitivity}{Specificity}$

Derivatives of the odds ratio were relative risk (*RR*) (10.12) and Yules *Q* (10.13).

$$RR = \frac{(TP/(TP+FP)}{FN/(FN+TN\})} \tag{10.12}$$

$$Yule's\ Q = \frac{((TP.TN)-(FP.FN))}{((TP.TN)+(FP.FN))} = \frac{OR-1}{OR+1} \tag{10.13}$$

An alternative measure of association, Pearson's phi coefficient (ϕ) (10.14) has values from −1 to +1, with a similar interpretation to values of κ (see also Chapter 6, subsection 6.4.1).

$$\phi = \frac{((FP.FN)-(TP.TN))}{\sqrt{(TP+FP)(FN+TN)(TP+FN)(FP+TN)}} \tag{10.14}$$

A similar coefficient to ϕ is the F_1 score (10.15), which reaches its best value at 1 and worst score at 0. A limitation of F_1 is that true negatives are unaccounted for, suggesting that other measures may be more useful.

$$F_1 = 2.\frac{PPV.Sensitivity}{PPV+Sensitivity} \tag{10.15}$$

Type I and II errors were estimated using false positive (10.13) and false negative (10.14) rates.

$$Type\ I\ error\ (False\ positive\ rate) = \alpha = \frac{FP}{FP+TN} \tag{10.16}$$

$$Type\ II\ error\ (False\ negative\ rate) = \beta = \frac{FN}{FN+TP} \tag{10.17}$$

Pre-test and post-test probabilities may be derived from calculated values for prevalence, positive predictive value (*PPV*) and negative predictive value (*NPV*). Prevalence estimates pre-test probability (10.8); *PPV* is an estimate of the positive post-test probability and 1-*NPV* is an estimate of the negative post-test probability.

Statistical testing used the Matlab Statistics and Bioinformatics toolboxes, supplemented with open-access Matlab code for the Fisher exact test, "FisherExactTest22" (Li 2010) and Cohen's kappa, "kappa" (Cardillo 2009).

10.2.2 Decision Tree Classifier

A decision tree was created using *ClassificationTree.fit*, configured to default settings of optimised pruning and leaf merging switched on. A column class vector was required for correct training. Data was partitioned for 10-fold cross validation using *cvpartition* and trained using the under-sampled dataset (see Chapter 9, subsection 9.5.2), with and without correction for empty or NaN cells.

Matlab decision tree tuning options were considered. Prior probabilities for each class, trained with the under-sampled set, was optimal with the default 'empirical' setting. Optimal pruned sub-tree sequence setting was 'on' and pruning criterion was the default 'error' setting. Score transformation function was optimised to the 'symmetric' (2x – 1) setting, the split criterion to the Gini (1912) index and surrogate decision splits at branch nodes to 'off'. The tuned classifier configuration was retained for testing.

10.2.3 Fuzzy Inference

Construction of a fuzzy inference system (FIS) depended to some extent on suitability of the application. Use of a FIS to classify cardiac rhythm was previously reported (see Chapter 2, subsection 2.8.5). Of these studies only Usher *et al.* (1999) used intracardiac electrograms in a similar way to this study but failed to describe performance in any detail. Fuzzy systems are known to perform well in hybrid systems (see Chapter 2, subsection 2.8.10 and Table 2.4).

To formulate fuzzy rules, fuzzification converts numeric values to non-numeric concepts such as high, medium and low. Given that in analysis of cardiac rhythm, intervals have critical threshold values, fuzzification appears inappropriate, diminishes the information content of features, reducing resolution, and appears likely to reduce diagnostic performance. For example, for a diagnosis of first degree heart block, a PR interval of > 0.2 seconds is the accepted cut-off value and leads to a crisp rule:

IF *PR Interval* IS > 0.2 secs THEN *first degree heart block*

Fuzzification of this crisp rule would produce a fuzzy rule:

IF *PR interval* IS *long* THEN *first degree heart block*

Although the imprecision implicit during the fuzzification process would result in resolution loss, this could still allow correct diagnosis but, depending on the training set, also risk misclassification of borderline cases. This suggests that interval based features are unsuited to fuzzification.

Other feature subsets considered for fuzzification, were symptoms, such as: rare, occasional or frequent palpitations, dizziness or syncope; haemodynamic, temperature and accelerometer parameters, where accepted cut-off values were unknown or patient-specific. The relative importance of features will be examined later (see subsection 10.3.1). The potential utility of FIS as a specialist classifier or component of a hybrid or multiple classifier system was deferred for possible later use.

10.2.4 Naïve Bayes Classifier

Data was partitioned for 10-fold cross-validation using *cvpartition.* With a *for-end* loop for each fold, a *NaiveBayes.fit* command created a classifier object. The classifier was then tuned using options available in Matlab. Distributions then prior probability estimations were tested for settings which worked best for this data. The kernel smoothing density estimate distribution and empirical prior probability estimation settings were selected, based on performance. The tuned classifier was retained for testing.

10.2.5 Neural Network Classifier

Feed-forward, back-propagation (FFBP) and radial basis functions (RBF) neural networks are known to be suited to classification. RBF networks are complex and perform best in low-dimensional feature space, whereas data in this study had high dimensionality. A single hidden layer model is considered a universal approximator and additional layers are rarely needed (Bishop 2006, pp.230-231).

An FFBP network model with a single hidden layer was chosen. Number of hidden nodes, transfer functions and learning algorithm were tuned heuristically to maximise performance. Network design used *nfittool*, set to defaults of Levenberg-Marquardt back-propagation (*trainlm*) training algorithm with early stopping based on a criterion of minimum error in the validation set, a hyperbolic tangent sigmoid (*tansig*) hidden layer and linear (*purelin*) output transfer functions. The network structure was stored as m code which was heuristically tuned by experimentation.

To train correctly, a column vector for each class was required, so the class vector was converted into a binary matrix using a simple Matlab script.

Tuning was performed using 10-fold cross validation, with each fold generating a network which was assessed for performance (see Table 10.1).

Table 10.1 Illustration of the performance assessment of cross-validation folds.

	Fold									
	1	**2***	***3***	**4**	**5**	**6**	**7**	**8**	**9**	**10**
CCR	0.86	0.88	0.87	0.86	0.87	0.88	0.90*	0.79	0.87	0.79
Error rate	0.14	0.12	0.13	0.14	0.13	0.13	0.10*	0.21	0.13	0.21
Sensitivity	0.67	0.83*	0.75	0.67	0.75	0.80	0.67	0.67	0.75	0.67
Specificity	0.89	0.89	0.89	0.89	0.89	0.89	0.94*	0.81	0.89	0.81
κ	0.49	0.69*	0.59	0.49	0.59	0.65	0.61	0.33	0.59	0.33
***P* (Fisher exact)**	0.073	0.002*	0.021	0.073	0.021	0.006	0.041	0.143	0.021	0.143

* Optimal values; *CCR*=correct classification rate

In the example shown in Table 10.1, both folds 2 and 7 had best performance values for 3 of the 6 indices. κ was used as a tie-breaker and indicated fold 2 as the best performing fold, so the network from fold 2 was stored and used for tuning and further testing.

Hidden nodes were tuned, starting with 1 hidden node and increasing until no further performance improvements were observed. When optimised for under-sampled data, performance indices were optimal with 7 or 8 hidden nodes. As fewer hidden nodes minimise computing resources, 7 hidden nodes was selected. Preconfigured with 7 hidden nodes, performances with the hyperbolic tangent sigmoid, logistic sigmoid, linear, radial basis function, hard and softmax hidden layer transfer functions were then compared. The softmax function was selected, as it performed best over a wide range of indices. Output layer configurations were evaluated with hidden layer settings set at their optimised values and the best performer was a linear transfer function.

Eight back-propagation training algorithms were evaluated. Bayesian regulation did not offer automatic early stopping so epochs were heuristically limited to 10. Bayesian regulation and Levenberg-Marquardt back-propagation algorithms performed well, with the former chosen as it performed best for the under-sampled training set.

Given that the hidden layer transfer function and learning algorithm had changed during tuning, hidden nodes were retuned, indicating optimal performance between 9 and 12 hidden nodes (see Fig. 10.1).

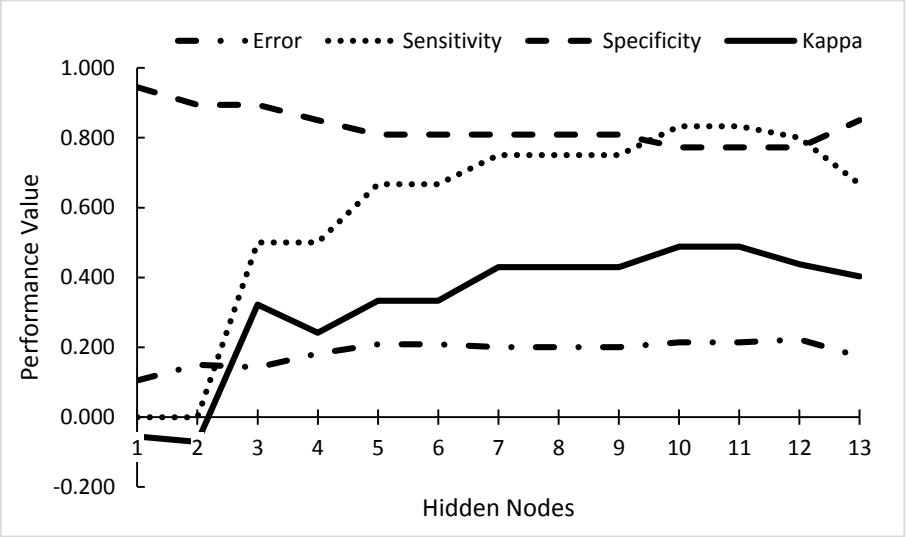

Figure 10.1 Optimisation of hidden nodes by network performance. Increasing numbers of hidden nodes improves performance indices other than error (sensitivity, specificity and kappa) until 10 to 11 hidden nodes, beyind which further increases lead to a reduction in performance for the same indices.

Sensitivity rose and specificity fell as hidden nodes increased, in a trade-off, with error relatively unaffected. A balance between performance indices was reached, with κ maximal (0.488), centred on the selected value of 11 hidden nodes. Performance fell off with more than 12 hidden nodes. During tuning repeat repartitioning for cross-validation produced varied results for unchanged configurations. Where choice between configurations having similar performances was made, consistency of performance was considered, with multiple training and testing cycles.

Final optimised NN configuration for this application was an FFBP using the Bayesian regulation back-propagation learning algorithm; a single layer of 11 hidden nodes with the softmax transfer function and a linear output transfer function. The tuned network was retained for testing.

10.2.6 Support Vector Machine Classifier

Matlab allows support vector machine (SVM) classification into two groups. For multiclass SVM with one binary classifier for each class, open-access Matlab code *multisvm* was used (Neuburger

2012). Support vector machines would train only where partitions contained at least one example for each class and without features containing NaN or empty values.

With one class having two examples, the maximum number of partitions was two, so hold-out partitioning was used with a 10% test set, rather than cross-validation. Repartitioning was performed four times to obtain best performance. Tuning evaluated alternative kernels and showed the quadratic kernel produced best performance. A tuned ensemble of SVM classifiers was generated and retained for further testing.

10.2.7 Guideline-Based Inference Engine Classifier

A knowledge engineering approach was used in design of a decision tree solution for classifying empty classes. Knowledge acquisition of cardiac rhythm definitions from clinical guidelines was performed (see Chapter 3, subsection 3.10.3 and Chapter 5, section 5.4), as an aid to understanding the domain. This knowledge was re-used to design a knowledge-based inference engine.

A simplified approach avoided use of propositional logic and formalisation of inference rules. Rules were converted from natural language into Matlab IF-ELSE-END loops, serving as inference rules and equivalent to IF-AND-THEN conditional statements.

7 additional derived features were required as inputs to decision rules. Additional pre-processing derived these features from the original feature set. Code was checked by a human domain expert. The engine was tested for functionality and retained for testing.

10.2.8 Iteration 1 – Classifier testing

The tuned classifiers were tested, using training sets from the corrected and uncorrected full and under-sampled datasets (see section 10.2.2). With the exception of the support vector machine, 10-fold cross-validation was used, producing 10 trained classifiers, from which the best performing was chosen. The selected classifier was then re-tested with all the data to produce performance indices, as a generalisation estimate (see Table 10.2).

Performance measures for classifiers trained on each of four training sets were assessed. Using the two-tailed Fisher exact test, all tested classifiers trained with all feature sets showed $P < 0.05$ for difference to the gold standard test. The neural network classifier trained with the corrected full dataset performed best overall using all indices, with *CCR* of 0.962; error 0.038; sensitivity 1.000; specificity 0.947; κ 0.906 and $P < 0.001$. Decision tree and neural network classifier technologies performed well when trained with the full data set, with values of > 0.80 for *CCR*,

sensitivity, specificity and κ. The inference engine classifier trained showed best performance for classifiers trained with the undersampled data set.

Table 10.2 Iteration 1 classifier performance measures for different training sets.

		Training Set	
	Performance Index	All instances corrected	Under-sampled corrected
Decision Tree	***CCR***	0.906*	0.536
	Error rate	0.094*	0.464
	Sensitivity	0.904*	0.885
	Specificity	0.988*	0.935
	κ	0.831*	0.520
	***P* (Fisher exact)**	<0.001*	<0.001*
Naïve Bayes	***CCR***	0.832	0.485
	Error rate	0.168	0.515
	Sensitivity	0.365*	0.654
	Specificity	0.995*	0.851
	κ	0.485*	0.221
	***P* (Fisher exact)**	<0.001*	<0.0001*
Neural network	***CCR***	0.962*	0.800
	Error rate	0.038*	0.200
	Sensitivity	1.000*	0.750
	Specificity	0.947*	0.810
	κ	0.906*	0.429
	***P* (Fisher exact)**	<0.001*	0.053
Support vector machine	***CCR***	0.870*	0.203
	Error rate	0.130*	0.797
	Sensitivity	0.942*	0.942*
	Specificity	0.918*	0.814
	κ	0.486	0.273
	***P* (Fisher exact)**	<0.001*	<0.001*
Guideline-based inference engine	***CCR***	0.261	0.571*
	Error rate	0.739*	0.429
	Sensitivity	0.846*	0.846*
	Specificity	0.947*	0.876
	κ	0.551	0.578*
	***P* (Fisher exact)**	<0.001*	<0.001*

*best performance, by training set; *CCR*=correct classification rate

Performance measures for classifiers trained on each of four training sets were assessed. Using the two-tailed Fisher exact test, all tested classifiers trained with all feature sets showed $P < 0.05$ for difference to the gold standard test. The neural network classifier trained with the corrected full dataset performed best overall using all indices, with *CCR* of 0.962; error 0.038; sensitivity 1.000; specificity 0.947; κ 0.906 and $P < 0.001$. Decision tree and neural network classifier technologies performed well when trained with the full data set, with values of > 0.80 for *CCR*, sensitivity, specificity and κ. The inference engine classifier trained showed best performance for classifiers trained with the undersampled data set.

The high performances reached when classifiers were trained with the full data set does not adequately exclude the possibility of over-training for majority PAC and PVC classes. Moderately good overall perfromances when classifiers were trained with the undersampled data sets suggest that over-training affects perfromance but that there is underlying good classification performance. All classifiers were taken forward for further evaluation.

10.3 Design Modifications following Iteration 1

Classifiers produced during classifier development and tested in iteration 1 were taken forward for modifications and subsequent use in a second iteration of the system development life-cycle. Given that performances were good but that maximal performance was not considered to be reached, it was considered that classifier modification rather than redesign was indicated. Design modifications were considered and priority was given to retention of principal features to avoid loss of information content and any indicated additional pre-processing.

10.3.1 Retention of Principal Features

Feature selection was driven by established requirements of cardiac rhythm diagnosis, rather than analysis, such as principal component analysis (see Chapter 7). It remained useful to highlight the relative importance of features and ensure their retention, avoiding a critical loss of information content.

The decision tree classifier resulting from iteration 1 (see subsection 10.2.8), was analysed as a simple and rapid method of feature ranking. For decision trees trained using the full data set, the root node was based on feature 105. Examination of leaves showed that differential diagnosis was possible but clarity was reduced at pruning levels > 14 end nodes there being duplicate leaves for each possible diagnosis (see Fig. 10.2).

In this tree, the root node was based on feature 105 (P:R ratio for beat R_0), level 1 on feature 84 (RP interval for beat R_{-1}), level 2 on feature 184 (RP interval for beat R_4) and level 3 on feature 42 (PR interval for beat R_{-3}).

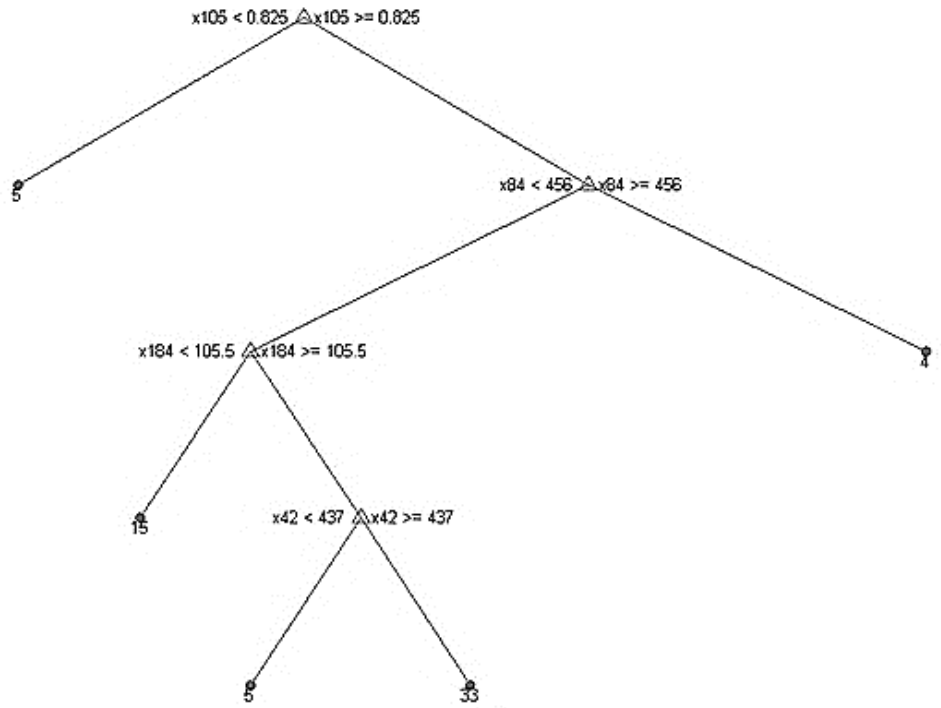

Figure 10.2 Decision tree trained using full data set (pruned to 14 of 18). (Full training set). The root node (x105) is top and is most important, representing P:R ratio for beat R_0, above the nodes of levels 1, 2, 3 and 4 sequentially.

For decision trees trained with the under-sampled data set, reducing the influences of features which lead to diagnosis of the majority PAC and PVC classes, the root node was based on feature 101, PP interval of beat R0. Analysis of trees emphasised the importance of interval-based features in rhythm diagnosis in 73% of nodes in levels 0 to 3; morphology based features 17%; other features were caffeine intake and respiration (MV) (10%) of nodes. The importance of caffeine and respiration features was unclear but noted for later use.

10.3.2 Additional Pre-processing

Consensus guidelines indicated the utility of interval derivatives including onset, inter beat variation, regularity, short coupling intervals and haemodynamic change (see Chapter 3, subsection 3.18.1 and Appendices E and F). Rate, regularity and site of origin are among ECG features considered useful in rhythm analysis. There is also a convention for site of origin based on atrio-ventricular association and relative rates (Wagner 2001, p.44 and pp.236-237).

Of note, ICD algorithms (see Chapter 4, section 4.17) all use rate, rhythm duration, sudden onset, stability, sustained rate duration, beat morphology and atrio-ventricular relationship in differential diagnosis. Sudden haemodynamic change, onset, stability and chamber of origin were all

previously suggested during feature selection (see Chapter 7, section 7.6).The initial feature set did not provide these derivations, suggesting a need for additional pre-processing to improve performance. This suggestion was supported by the need, during inference engine classifier development (see subsection 10.2.7) for additional processing to obtain derived features in appropriate form. The following additional features were produced:

RR3mean4 – mean RR interval of the first 4 beats of a new rhythm (R_0, R_1, R_2, R_3)
PR3mean4 – mean PR interval of the first 4 beats of a new rhythm (R_0, R_1, R_2, R_3)
RP3mean4 – mean RP interval of the first 4 beats of a new rhythm (R_0, R_1, R_2, R_3)
PRratio3mean4 – mean P:R ratio of the first 4 beats of a new rhythm (R_0, R_1, R_2, R_3)
RR0delta - % change between RR interval means at start of a new rhythm (onset)
Stab012 - deviation RR interval from most recent 3 beat mean (stability)
Stress - indicator of stress (QT interval, respiration, motion, haemodynamics)

These indications motivated further feature pre-processing to improve classifier performance. New derived features would include: an indicator of stress, sudden haemodynamic change and sudden rhythm onset, rhythm stability and chamber of origin.

10.3.3 Indicator of Stress

No single index, using a combination of parameters to represent physiological and emotional stress and activity was found in the published literature. A simple index was devised using features previously selected to represent stress: minute ventilation index (MVI); haemodynamic stress in the form of stroke volume index (SVI); body motion accelerometry and corrected QT interval (QT_c).

Threshold values were applied:

Increase in MVI > 10% (MV (V_E) increases 0.8 to 0.89 on low level exercise) (Vai *et al.* 1988)
Increase in SVI > 10% (Stratton *et al*. 1994)
Acceleration of > 0.1g (0.2 to 0.4g during walking) (Kavanagh *et al*. 2004)

Body temperature was not processed further, as temperature change is insensitive to stress, with any temperature increase of 0.5°C delayed up to 5 minutes following onset of exercise (Lim *et al.* 2008). The index detected stress when one or more threshold value was exceeded.

10.3.4 Sudden Haemodynamic Change and Sudden Onset of Rhythms

It is recognised that differential diagnosis of arrhythmias with 1:1 atrio-ventricular conduction poses particular challenges.

Fig. 10.3 illustrates the sudden onset of a rhythm change due to a premature ventricular contraction (PVC) at beat number 150 (arrowed), followed by a reduction in RR interval, P:R ratio and stroke-volume index (SVI). An index was devised to represent the percent change of SVI between the first beat of a new rhythm (R_0) and the preceding beat (R_{-1}) (10.18):

$$SVI\ change = 100.\frac{(SVI_0 - SVI_{-1})}{SVI_{-1}} \tag{10.18}$$

Similarly, interval onset measures sudden change in RR interval at the first beat of a new rhythm (see subsections 4.18.1 and 4.18.2 and Fig. 10.4).

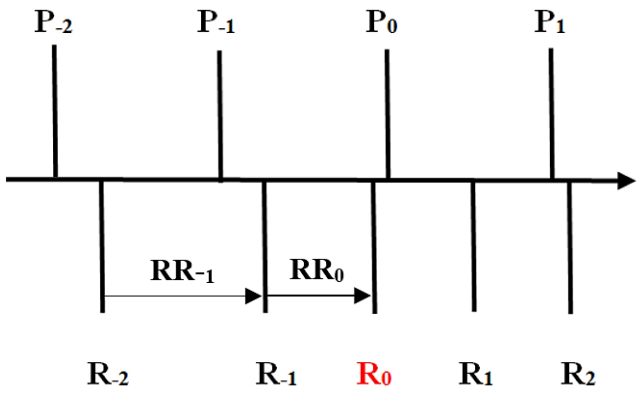

Figure 10.4 Schematic of prematurity using beat-to-beat RR interval (PVC). RR interval RR_0 is much shorter than that of its precesssor, interval RR_{-1}, representing a premature beat. Where this interval is the first of a new rhythm (beats R_0, R_1 and R_2), prematurity as a percentage of the previous RR interval quantifies the sudden onset.

A sudden onset represents prematurity where the interval is shorter or a pause where it is longer. RR interval of the first beat (RR_0) compared with the preceding beat (RR_{-1}) and can be expressed as a percentage change (10.19).

$$Interval\ Onset = 100.\frac{(RR_{-1} - RR_0)}{RR_{-1}} \tag{10.19}$$

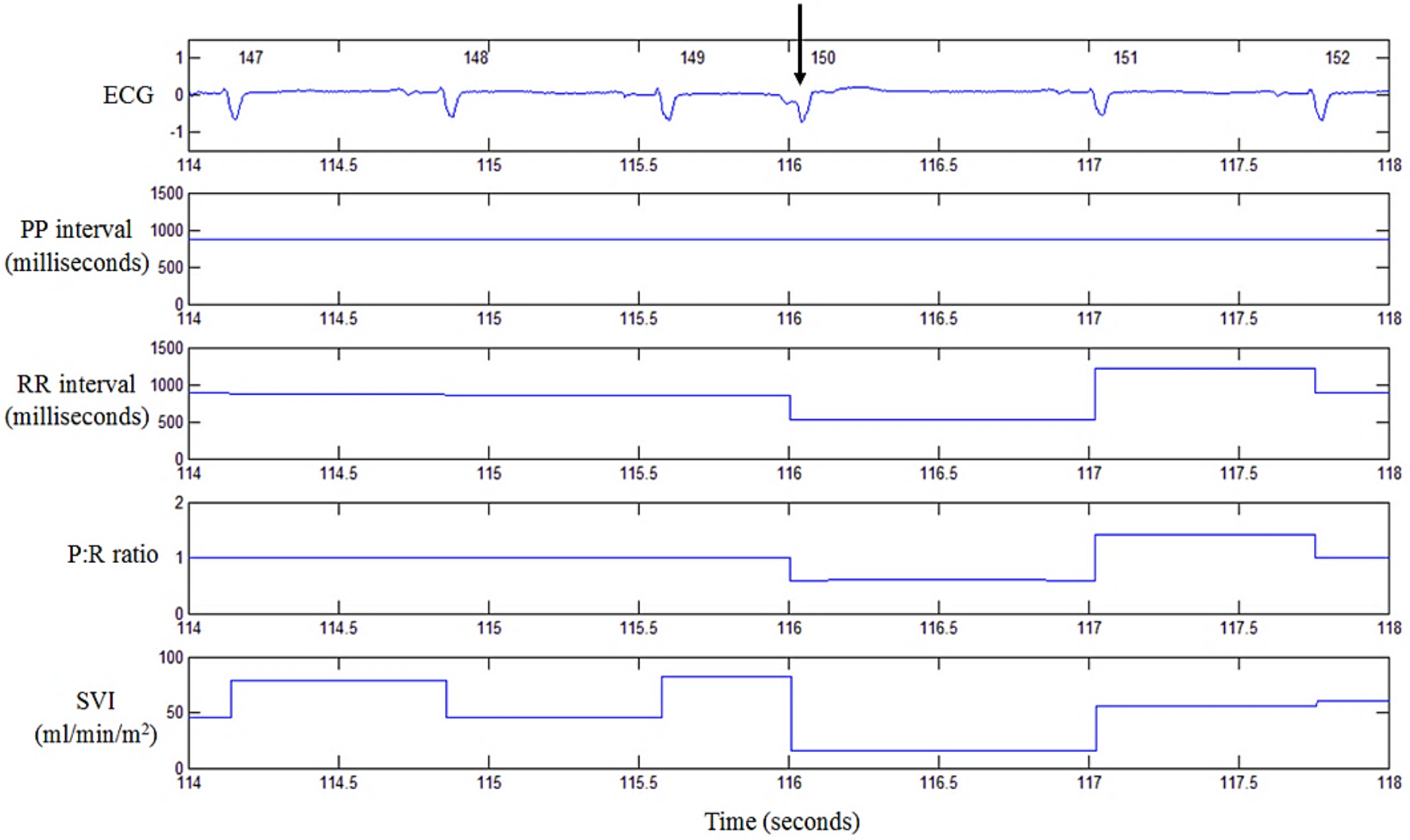

Figure 10.3 Matlab plot of a typical PVC (The PVC is beat 150, arrowed). Showing changes in PP and RR intervals, P:R ratio and SVI).

Positive interval onset values represent a premature beat and negative values a pause. To describe a new ventricular rhythm as having "sudden onset", interval onset should exceed a threshold value, typically > 20% (see Chapter 3, Table 3.3).

10.3.5 Rhythm Stability

Interval stability measures the regularity of a new rhythm once established (see Chapter 4, subsections 4.18.1, 4.18.2 and Fig. 10.5).

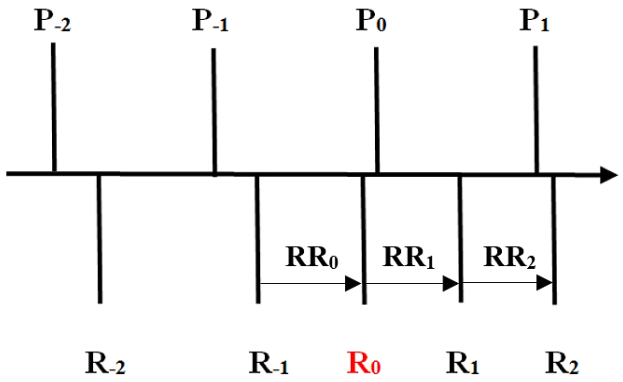

Figure 10.5 Rhythm stability and regularity assessed by RR interval variation. Where RR interval RR_0 is the first of a new rhythm (beats R_0, R_1 and R_2), stability is the maximum variation of RR interval from the mean in the new rhythm.

Stability can be expressed as the maximum variation from the mean of the first three RR intervals of a new rhythm (RR_0, RR_1 and RR_2) (10.20 and 10.21).

$$\overline{RR} = \frac{1}{3}.\sum_{i=0}^{2} RR_i \qquad (10.20)$$

$$Interval\ Stability = \max(|RR_0 - \overline{RR}|, |RR_1 - \overline{RR}|, |RR_2 - \overline{RR}|) \qquad (10.21)$$

Interval stability should exceed a threshold value to enable a new ventricular rhythm to be described as being "stable". A typical stability threshold would be 20msecs (see Chapter 3, Table 3.3).

10.3.6 Chamber of Origin

P:R ratio from beat R_0 was at the root node of the decision tree classifier, indicating its utility in the discrimination of the majority classes, premature atrial complexes (PAC) from premature ventricular complexes (PVC). By analysis of atrial and ventricular intervals leading to a rhythm

change and at the point (onset) of rhythm change, rhythm classification can be indicated by the chamber activating first – the "chamber of origin". The original feature set included interval derivatives describing the relationship between atria and ventricles and their use deciding chamber of origin was explored using simulated rhythm scenarios.

Fig. 10.6 shows a simulation of a PAC with 1:1 atrio-ventricular conduction, where it was clear that the rhythm commenced with P_0 and originated in the atrium.

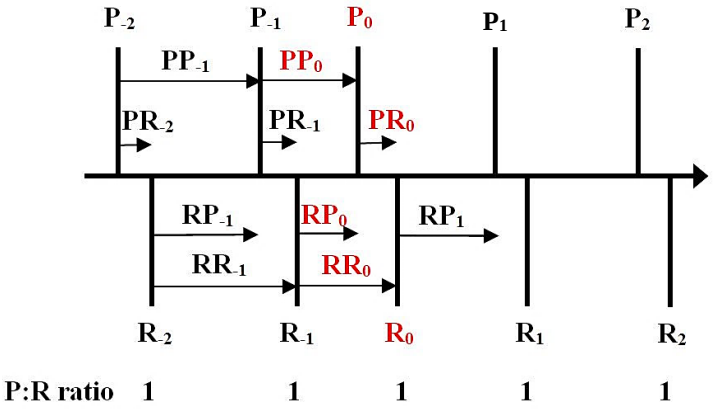

Figure 10.6 PAC with 1:1 conduction (a P:R ratio of 1:1) - chamber of origin is atrial (PAC). Continuous analysis of P:R ratio can indicate the chamber of origin of beats as they occur, with few exceptions (see text).

This is the only case where P:R ratio at R_0 is 1, demonstrating a rule, that if P:R ratio is equal to 1 at the beat of onset (R_0) then chamber of origin is the atrium (10.5). If the P:R ratio is less than 1, then chamber of origin is the ventricle (10.6).

Simulations indicated that the chamber with the faster rate and that P:R ratio may indicate chamber of origin, leading to the following statements:

$$PRratio_0 \geq 1 \Rightarrow Atrial\ Chamber\ of\ Origin \tag{10.22}$$

$$PRratio_0 < 1 \Rightarrow Ventricular\ Chamber\ of\ Origin \tag{10.23}$$

P:R ratio of beat R_0 was in the original feature set and did not require additional pre-processing. It was anticipated that actual values of P:R ratio would not be exact, as heart rate or atrio-

ventrucular conduction time may vary slightly and a margin of error was needed. One scenario not adequately dealt with was the possibility for co-existence of atrial and ventricular arrhythmias. Chamber of origin describes the first impulse at the onset of new rhythm and may be considered inadequate to deal with complex discrimination. Satisfactory differential diagnosis of all possible rhythms requires additional discriminators.

10.3.7 Implementation of Modifications

Threshold values for new features were based on published data (see subsection 10.3.3 and Chapter 3, Table 3.3) and an evaluation of the number of beats required for detection of change (see Chapter 3, subsections 3.18.1). Modifications were implemented using Matlab script and by the concatenation of new features to existing features. Classifier units developed in section 10.2 were re-used.

10.4 Iteration 2

Classifiers were trained using the modified feature set, including four additional features of stress, sudden haemodynamic change and sudden rhythm onset and rhythm stability as iteration 2.

As for iteration 1, performances were again compared for the different classifiers with the full and undersampled data sets (see Table 10.3). The best performing classifier overall was the neural network classifier trained with the corrected full dataset, with *CCR* of 0.926; error 0.074; sensitivity 1.000; specificity 0.900; κ 0.824 and $P < 0.001$, very similar values to those achieved in iteration 1 (*CCR* of 0.962; error 0.038; sensitivity 1.000; specificity 0.947; κ 0.906 and $P < 0.001$). P was < 0.05 for difference to the gold standard test for iteration 2.

Further iterations were to continue, until no further improvement was likely or performance reduction was resultant. Given that further performance improvement was considered unlikely with the inclusion of additional and derived features, a further iteration was proposed to reduce the number of features, on the basis that this dimensionality reduction may have positive effect of performance.

Based on this comparison, iteration 2 was believed to show performance improvement for the decision tree, naïve Bayes, and inference engine classifiers, with the neural network classifier slightly improved performance and the support vector machine classifier showing no overall improvement.

10.5 Optimising a Feature Sub-Set

Increasing the number of features improves classifier performance until a point where peak performance may be exceeded. Given the high dimensionality of the data, with 212 features for

Table 10.3 Iteration 2 classifier performance measures for different training sets.

		Training Set	
	Performance Index	**All instances corrected**	**Under-sampled corrected**
Decision Tree	***CCR***	0.911*	0.595
	Error	0.089*	0.405
	Sensitivity	0.942	0.962*
	Specificity	0.991*	0.873
	κ	0.885*	0.378
	P	<0.001*	<0.001*
Naïve Bayes	***CCR***	0.854*	0.578
	Error	0.146*	0.422
	Sensitivity	0.423	0.827*
	Specificity	0.999*	0.971
	κ	0.574	0.664*
	P	<0.001*	<0.001*
Neural network	***CCR***	0.926*	0.778
	Error	0.074*	0.222
	Sensitivity	1.000*	0.800
	Specificity	0.900*	0.773
	κ	0.824*	0.438
	P	<0.001*	0.030
Support vector machine	***CCR***	0.869*	0.243
	Error rate	0.131*	0.757
	Sensitivity	0.923*	0.904
	Specificity	0.918*	0.819
	κ	0.478*	0.268
	P	<0.001*	<0.001*
Guideline-based inference engine	***CCR***	0.263	0.571*
	Error	0.737	0.429*
	Sensitivity	0.846*	0.846*
	Specificity	0.947*	0.882
	κ	0.551	0.591*
	P	<0.001*	<0.001*

*best performance, by training set

iteration 1 and 216 features for iteration 2, additional features may not improve performance and this is borne out by the disappointing performance in iteration 2.

An optimisation process to reduce dimensionality and potentially improve performance was explored. Given that a stated design approach was to include a cognitive approach (see Chapter 3, subsection 3.10.4) feature selection was based on domain knowledge. The minimum number of beats to diagnose a rhythm, indicated by clinical guidelines (see Chapter 3, subsection 3.18.1), was used as the basis of a cognitive feature selection process.

The minimum beats to diagnose a rhythm change was previously found to require 5 beats (see Chapter 3, subsection 3.18.2): 2 beats preceding a rhythm change and 3 beats following onset of a new rhythm, consisting of beats R_{-2}, R_{-1}, R_0, R_1 and R_2, rather than the 10 beats, used to accommodate beats required by standard ICD algorithms. Selecting features associated with a 5 beat requirement reduced features from 216 to 116.

10.6 Iteration 3

Implementation of modifications used a Matlab script. Classifiers were trained with training sets based on the modified feature set, using a 5 beat analysis. The inference engine script was modified for excluded features. Classifiers were tested with all the data in the same way as iterations 1 and 2 (see Table 10.4).

In iteration 3, the neural network classifier was best performing, when trained with the full data set, with a *CCR* of 0.960; error 0.040; sensitivity 1.000; specificity 0.947; κ 0.896 and $P < 0.001$, compared with results from iteration 2 of *CCR* of 0.926; error 0.074; sensitivity 1.000; specificity 0.900; κ 0.824 and $P < 0.001$, very similar values to those achieved in iteration 1 (*CCR* of 0.962; error 0.038; sensitivity 1.000; specificity 0.947; κ 0.906 and $P < 0.001$). For the neural network classifier, the greatest *CCR* and lowest error was achieved in iteration 1: sensitivity was perfect (1.00) in all 3 iterations; specificity was 0.947 for both iterations 1 and 3 and κ was highest in iteration 1 at 0.906.

10.7 Comparison of Performances between Iterations

Iterations 1, 2 and 3 were evaluated to assess progress towards satisfaction of the user requirements of maximisation of overall performance.

Iteration characteristics were summarised:

- **Iteration 1** - 212 features, 10 beats, initial feature set
- **Iteration 2** - 216 features, 10 beats, features for stress, sudden haemodynamic change, sudden rhythm onset, rhythm stability
- **Iteration 3** - 116 features, 5 beats, features for stress, sudden haemodynamic change, sudden rhythm onset, rhythm stability

Table 10.4 Iteration 3 classifier performance measures for different training sets.

		Training Set	
	Performance Index	**All**	**Under-sampled**
Decision Tree	***CCR***	0.904*	0.630
	Error	0.096*	0.370
	Sensitivity	0.904*	0.788
	Specificity	0.992*	0.957
	κ	0.872*	0.569
	P	<0.001*	<0.001*
Naïve Bayes	***CCR***	0.839*	0.672
	Error	0.161*	0.328
	Sensitivity	0.423	0.673*
	Specificity	0.998*	0.983
	κ	0.566	0.650*
	P	<0.001*	<0.001*
Neural network	***CCR***	0.960*	0.759
	Error	0.040*	0.241
	Sensitivity	1.000*	0.833
	Specificity	0.947*	0.739
	κ	0.896*	0.438
	P	<0.001*	0.018
Support vector machine	***CCR***	0.875*	0.327
	Error rate	0.125*	0.673
	Sensitivity	0.904	0.942*
	Specificity	0.918*	0.813
	κ	0.470*	0.271
	P	<0.001*	<0.001*
Guideline-based inference engine	***CCR***	0.216	0.436*
	Error	0.784	0.564*
	Sensitivity	0.827*	0.096
	Specificity	0.649	0.898*
	κ	0.110*	-0.007
	P	<0.001*	1.000

*best performance, by training set

During the iterative process, all classifiers had been tested and performance measured using four different feature sets, two of which were subsets:

1. **All data** - 1109 instances, uncorrected for missing data
2. **All data corrected** – 1109 instances, corrected for missing data

3. **Under-sampled data** (subset of 1) - 366 instances, uncorrected for missing data
4. **Under-sampled data corrected** (subset of 2) - 366 instances, corrected for missing data

Undersampled data was a subset of all available instances selected at random to balance data, reducing the majority classes to the mean prevalence of the remaining classes (see Chapter 9, subsection 9.5.2).

Classifier maximal performances between iterations 1, 2 and 3 were directly compared for each of the four data sets, by classifier technology (see Table 10.5). Note that support vector machines could not be trained with missing data, precluding input of uncorrected data.

Using the data in Table 10.5, classifiers were scored, by their having the value of each performance index for that classifier technology closest to the optimal value of 1.0 for *CCR*, sensitivity, specificity and κ; 0 for error and values of $P < 0.05$. The maximum score would be six, for six indicators having the most optimal value, with the highest scoring classifier configuration considered best performing. Classifiers with κ of < 0.2 or $P \geq 0.05$ were discarded. κ was arbitrarily given a higher weighting, with the result that tie-breaks favoured the classifier configuration with the highest κ score. For each classifier type, the best performing was selected to take forward for further analysis.

The best performing classifier overall was the neural network classifier, with its highest *CCR* and lowest error in iteration 1; sensitivity was perfect (1.00) in all 3 iterations; specificity was highest at 0.947 for iterations 1 and 3 and κ was highest in iteration 1 at 0.906.

The decision tree classifier performed best in iteration 2 when trained with the corrected full data set, having *CCR* of 0.911; error 0.089; sensitivity 0.942; specificity 0.991; κ 0.885 and $P < 0.001$.

The naïve Bayes classifier performed best in iteration 1 when trained with the uncorrected full data set, having *CCR* of 0.859; error 0.141; sensitivity 0.692; specificity 0.998; κ 0.792 and $P < 0.001$.

The neural network classifier performed best in iteration 1 when trained with the corrected full data set, with a *CCR* of 0.962; error 0.038; sensitivity 1.000; specificity 0.947; κ 0.906 and $P < 0.001$.

The support vector machine classifier performed best in iteration 1 when trained with the corrected full data set, with *CCR* of 0.870; error 0.130; sensitivity 0.942; specificity 0.918; κ 0.486 and $P < 0.001$.

Table 10.5 A comparison of classifier performances by iteration and training set.

	Training Set	All Data			All Data corrected			Under-sampled			Under-sampled corrected		
	Iteration	1	2	3	1	2	3	1	2	3	1	2	3
Decision Tree	***CCR***	0.907	0.886	0.890	0.906	0.911*	0.904	0.523	0.515	0.515	0.536	0.595	0.630
	Error rate	0.093	0.114	0.110	0.094	0.089*	0.096	0.477	0.485	0.485	0.464	0.405	0.370
	Sensitivity	0.923	0.865	0.904	0.904	0.942	0.904	0.865	0.962*	0.808	0.885	0.962*	0.788
	Specificity	0.990	0.992*	0.989	0.988	0.991	0.992*	0.941	0.837	0.957	0.935	0.873	0.957
	κ	0.858	0.850	0.839	0.831	0.885*	0.872	0.537	0.313	0.580	0.520	0.378	0.569
	P	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
Naïve Bayes	***CCR***	0.859	0.843	0.901*	0.832	0.854	0.839	0.476	0.595	0.670	0.485	0.578	0.672
	Error rate	0.141	0.157	0.099*	0.168	0.146	0.161	0.524	0.405	0.330	0.515	0.422	0.328
	Sensitivity	0.692	0.212	0.519	0.365	0.423	0.423	0.808*	0.519	0.500	0.654	0.827	0.673
	Specificity	0.998	0.746	0.753	0.995	0.999*	0.998	0.835	0.720	0.737	0.851	0.971	0.983
	κ	0.792*	-0.014	0.086	0.485	0.574	0.566	0.257	0.069	0.072	0.221	0.664	0.650
	P	<0.001*	0.624	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
Neural network	***CCR***	0.957	0.920	0.885	0.962*	0.926	0.960	0.840	0.760	0.760	0.800	0.778	0.759
	Error rate	0.043	0.080	0.115	0.038*	0.074	0.040	0.160	0.240	0.240	0.200	0.222	0.241
	Sensitivity	1.000*	1.000*	1.000*	1.000*	1.000*	1.000*	0.800	0.667	0.667	0.750	0.800	0.833
	Specificity	0.947*	0.900	0.857	0.947*	0.900	0.947*	0.850	0.773	0.773	0.810	0.773	0.739
	κ	0.862	0.783	0.698	0.906*	0.824	0.896	0.565	0.279	0.279	0.429	0.438	0.438
	P	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	0.0123	0.178	0.180	0.053	0.030	0.018
Support vector machine	***CCR***				0.870	0.869	0.875*				0.203	0.243	0.327
	Error rate				0.130	0.131	0.125*				0.797	0.757	0.673
	Sensitivity				0.942*	0.923	0.904				0.942*	0.904	0.942*
	Specificity				0.918*	0.918*	0.918*				0.814	0.819	0.813
	κ				0.486*	0.478	0.470				0.273	0.268	0.271
	P				<0.001*	<0.001*	<0.001*				<0.001*	<0.001*	<0.001*
Inference engine	***CCR***	0.261	0.263	0.216	0.261	0.263	0.216	0.571*	0.571*	0.436	0.571*	0.571*	0.436
	Error rate	0.739	0.737	0.784	0.739	0.737	0.784	0.429*	0.429*	0.564	0.429*	0.429*	0.564
	Sensitivity	0.846*	0.846*	0.827	0.846*	0.846*	0.827	0.846*	0.846*	0.096	0.846*	0.846*	0.096
	Specificity	0.947*	0.947*	0.649	0.947*	0.947*	0.649	0.876	0.882	0.898	0.876	0.882	0.898
	κ	0.551	0.551	0.110	0.551	0.551	0.110	0.578	0.591*	-0.007	0.578	0.591*	-0.007
	P	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	1.000	<0.001*	<0.001*	1.000

*Best performance for each classifier

The inference engine classifier performed best in iteration 2 when trained with the under-sampled corrected data set, giving a *CCR* 0.571; error 0.429; sensitivity 0.846; specificity 0.882; κ 0.591 and $P < 0.001$.

There were no classifiers showing best overall performance for iteration 3, indicating that performance had declined, satisfying a target criterion for iterative development. The utility of 5 beat rhythm analysis, as implemented in iteration 3, was not held by these results and will be explored further in Chapter 11, subsection 11.2.2.

10.7.1 Selecting Optimal Feature Sub-sets for Classifiers.

The only difference between data sets containing all data and under-sampled data was the reduced number of instances of the majority PAC and PVC classes. This meant that any reduction in performance between classifiers trained with full and under-sampled sets respectively could be attributed to over-training for these majority classes.

Using the data in Table 10.5, classifiers trained using the under-sampled corrected set were scored using the same system as in section 10.7. For each classifier type, the best performing iteration was selected to take forward for further analysis (see Table 10.6).

Table 10.6 Classifier performances by iteration. All classifiers trained with the corrected under-sampled set.

Classifier	Decision Tree	Naïve Bayes	Neural Network	SVM	Inference Engine
Number of features	116	216	116	116	216
Iteration	3	2	3	3	2
CCR	0.630	0.578	0.759*	0.327	0.571
Error	0.370	0.422	0.241*	0.673	0.429
Sensitivity	0.788	0.827	0.833*	0.942	0.846
Specificity	0.957	0.971*	0.739	0.813	0.882
κ	0.569	0.664*	0.438	0.271	0.591
P	<0.001*	<0.001*	0.018	<0.001*	<0.001*

*best performance by classifier technology

For the decision tree classifier, the selected feature set having best performance was the 116 feature set used for iteration 3, iteration 2, showing *CCR* of 0.630; error 0.370; sensitivity 0.788; specificity 0.957; κ of 0.569 and $P < 0.001$.

For the naïve Bayes classifier, the selected feature set having best performance was the 216 feature set used for iteration 2, showing *CCR* of 0.578; error of 0.422; sensitivity of 0.827; specificity 0.971; κ 0.664 and $P < 0.001$.

For the neural network classifier, the selected feature set having best performance was the 116 feature set used for iteration 3, showing *CCR* 0.759; error 0.241; sensitivity 0.833; specificity 0.739; κ 0.438 and *P* of 0.018.

For the support vector machine classifier, the selected feature set having best performance was the 116 feature set used for iteration 3, showing *CCR* of 0.327; of 0.673; sensitivity 0.942; specificity 0.813; κ 0.271 and $P < 0.001$.

For the inference engine classifier, the selected feature set having best performance was the 216 feature set used for iteration 2, showing *CCR* of 0.571; error 0.429; sensitivity 0.846; specificity 0.882; κ of 0.591 and $P < 0.001$. A major change was the reversal in success for iteration 3, using 5 beat analysis, unmasked by minimisation of the influence of over-training for the majority classes by the use of under-sampled data for training.

Feature sets selected were:

Decision Tree Classifier	5 beats (as in iteration 3), 116 features
Naïve Bayes Classifier	10 beats (as in iteration 2), 216 features
Neural Network Classifier	5 beats (as in iteration 3), 116 features
Support Vector Machine Classifier	5 beats (as in iteration 3), 116 features
Inference Engine Classifier	10 beats (as in iteration 2), 216 features

Comparing these classifiers , the naïve Bayes classifier performed best, with *CCR* of 0.578; error 0.422; sensitivity 0.827; specificity 0.971, κ 0.664 and $P < 0.001$. The two-tailed Fisher exact test showed $P < 0.05$ for all the selected classifiers, indicating no significant difference to the gold standard test.

10.8 Performance of Classifiers by Rhythm Type

Classifiers selected in section 10.7.1 were re-trained using corrected under-sampled instances. Each trained classifier was re-tested with all instances and a confusion matrix was generated, using the Matlab confusion *function*. Analysis of the confusion matrices for each classifier was performed and performance indices for individual rhythms tabulated, (see Tables 10.7 to 10.11). Rhythms were allocated to 19 classes, into some of which were pooled (see Chapter 9, section 9.3): normal sinus rhythm; respiratory sinus arrhythmia; physiological sinus tachycardia;

premature atrial contraction; premature ventricular contraction; sinus node dysfunction; first degree atrio-ventricular block; 2nd and 3rd degree atrio-ventricular block; atrio-ventricular nodal junctional tachycardias; atrio-ventricular reciprocating tachycardias; abnormal tachycardias of the sinus node; focal atrial tachycardia; macro-re-entrant atrial tachycardia; multifocal atrial tachycardia; atrial fibrillation; ventricular tachycardias; ventricular fibrillation; atrial paced rhythm and ventricular paced rhythm. Note that ventricular fibrillation was an empty class.

10.8.1 Decision Tree Classifier Performances for Different Rhythms

The decision tree classifier rhythm specific performances are shown in Table 10.7.

Table 10.7 Decision tree classifier performances by rhythm.

Rhythm	CCR	Error	Se	Sp	κ	*P*
Normal sinus rhythm	0.950	0.050	0.788	0.957	0.569	<0.001
Respiratory sinus arrhythmia	0.958	0.042	0.980	0.957	0.651	<0.001
Physiological sinus tachycardia	0.983	0.017	0.556	0.986	0.337	<0.001
Premature atrial contraction	0.802	0.198	0.599	0.914	0.544	<0.001
Premature ventricular contraction	0.805	0.195	0.492	0.972	0.494	<0.001
Sinus node dysfunction	0.982	0.018	0.966	0.982	0.728	<0.001
First degree atrio-ventricular block	0.998	0.002	0.333	1.000	0.499	0.003
2nd and 3rd degree atrio-ventricular block	0.998	0.002	0.000	1.000	0.000	1.000
Atrio-ventricular nodal junct. tachycardias	0.983	0.017	0.852	0.986	0.699	<0.001
Atrio-ventricular recip. tachycardias	0.999	0.001	0.917	1.000	0.956	<0.001
Abnormal tachycardias of the sinus node	0.997	0.003	0.333	0.999	0.399	0.005
Focal atrial tachycardia	0.984	0.016	0.750	0.985	0.394	<0.001
Macro-re-entrant atrial tachycardia	0.994	0.006	0.875	0.995	0.664	<0.001
Multifocal atrial tachycardia	0.946	0.054	0.588	0.951	0.231	<0.001
Atrial fibrillation	0.986	0.014	0.913	0.988	0.625	<0.001
Ventricular tachycardias	0.993	0.007	0.167	0.997	0.197	0.0210
Atrial paced rhythm	0.925	0.075	0.917	0.926	0.482	<0.001
Ventricular paced rhythm	0.978	0.022	0.800	0.983	0.856	<0.001

CCR=correct classification rate; Se=sensitivity; Sp=specificity

The decision tree classifier did not classify 2nd or 3rd degree atrio-ventricular block well from the instances presented as input patterns. The rhythm with which the highest performance parameters was achieved was for atrio-ventricular reciprocating tachycardias, with *CCR* of 0.999; error of 0.001; sensitivity of 0.917; specificity of 1.000; κ of 0.956 and $P < 0.001$. Good performance was achieved for other rhythms, particularly respiratory sinus arrhythmia, with *CCR* of 0.958; error

of 0.042; sensitivity 0.980; specificity 0.957, κ of 0.651 and $P < 0.001$, sinus node dysfunction and atrial fibrillation and ventricular paced rhythm. High specificities of > 0.91 were achieved for all rhythms, with sensitivity being more variable, of between 0 and 0.980 and κ between 0 and 0.856.

10.8.2 Naïve Bayes Classifier Performances for Different Rhythms

The naïve Bayes classifier rhythm specific performances are shown in Table 10.8.

Table 10.8 Naïve Bayes classifier performances by rhythm.

Rhythm	CCR	Error	Se	Sp	κ	*P*
Normal sinus rhythm	0.964	0.036	0.827	0.971	0.664	<0.001
Respiratory sinus arrhythmia	0.990	0.010	0.816	0.998	0.874	<0.001
Physiological sinus tachycardia	0.998	0.002	1.000	0.998	0.899	<0.001
Premature atrial contraction	0.662	0.338	0.164	0.940	0.123	<0.001
Premature ventricular contraction	0.702	0.298	0.718	0.693	0.384	<0.001
Sinus node dysfunction	0.984	0.016	0.966	0.984	0.749	<0.001
First degree atrio-ventricular block	0.999	0.001	0.667	1.000	0.800	<0.001
2nd and 3rd degree atrio-ventricular block	0.998	0.002	0.000	1.000	0.000	1.000
Atrio-ventricular nodal junct. tachycardias	0.998	0.002	1.000	0.998	0.963	<0.001
Atrio-ventricular recip. tachycardias	0.998	0.002	0.833	1.000	0.908	<0.001
Abnormal tachycardias of the sinus node	1.000	0.000	1.000	1.000	1.000	<0.001
Focal atrial tachycardia	1.000	0.000	1.000	1.000	1.000	<0.001
Macro-re-entrant atrial tachycardia	0.999	0.001	1.000	0.999	0.941	<0.001
Multifocal atrial tachycardia	0.940	0.060	1.000	0.939	0.319	<0.001
Atrial fibrillation	0.946	0.054	1.000	0.945	0.415	<0.001
Ventricular tachycardias	1.000	0.000	1.000	1.000	1.000	<0.001
Atrial paced rhythm	0.975	0.025	0.938	0.976	0.750	<0.001
Ventricular paced rhythm	0.995	0.005	0.833	0.999	0.890	<0.001

CCR=correct classification rate; Se=sensitivity; Sp=specificity

The naïve Bayes classifier also failed to successfully classify 2nd or 3rd degree atrio-ventricular block from the instances presented as input patterns. Several rhythm diagnoses had "perfect" performance indices of error 0.000; sensitivity 1.000; specificity 1.000; κ of 1.000 and $P < 0.001$. These were: abnormal tachycardias of the sinus node; focal atrial tachycardia and ventricular tachycardias. The classifier performed very well with several other rhythms: respiratory sinus arrhythmia, physiological sinus tachycardia, sinus node dysfunction, atrio-ventricular nodal and junctional tachycardias, atrial fibrillation and atrial and ventricular paced rhythms.

10.8.3 Neural Network Classifier Performances for Different Rhythms

The neural network classifier rhythm specific performances are shown in Table 10.9.

Table 10.9 Neural network classifier performances by rhythm.

Rhythm	CCR	Error	Se	Sp	*κ*	*P*
Normal sinus rhythm	0.784	0.216	0.942	0.777	0.230	<0.001
Respiratory sinus arrhythmia	0.940	0.060	0.980	0.939	0.566	<0.001
Physiological sinus tachycardia	0.992	0.008	0.000	1.000	0.000	1.000
Premature atrial contraction	0.641	0.359	0.000	0.999	-0.002	1.000
Premature ventricular contraction	0.778	0.222	0.435	0.961	0.448	0.000
Sinus node dysfunction	0.920	0.080	0.931	0.919	0.351	<0.001
First degree atrio-ventricular block	0.997	0.003	0.000	1.000	0.000	1.000
2nd and 3rd degree atrio-ventricular block	0.998	0.002	0.000	1.000	0.000	1.000
Atrio-ventricular nodal junct. tachycardias	0.930	0.070	0.741	0.934	0.313	<0.001
Atrio-ventricular recip. tachycardias	0.989	0.011	0.000	1.000	0.000	1.000
Abnormal tachycardias of the sinus node	0.916	0.084	1.000	0.916	0.056	<0.001
Focal atrial tachycardia	0.993	0.007	0.000	1.000	0.000	1.000
Macro-re-entrant atrial tachycardia	0.993	0.007	0.000	1.000	0.000	1.000
Multifocal atrial tachycardia	0.987	0.013	0.235	0.999	0.359	<0.001
Atrial fibrillation	0.981	0.019	0.087	1.000	0.157	<0.001
Ventricular tachycardias	0.995	0.005	0.000	1.000	0.000	1.000
Atrial paced rhythm	0.895	0.105	0.875	0.896	0.379	<0.001
Ventricular paced rhythm	0.975	0.025	0.933	0.976	0.655	<0.001

CCR=correct classification rate; Se=sensitivity; Sp=specificity

The neural network classifier failed to successfully classify several rhythms: 1st degree, 2nd or 3rd degree atrio-ventricular blocks; physiological sinus tachycardia; atrio-ventricular reciprocating tachycardias; focal atrial tachycardia; macro-re-entrant atrial tachycardia and ventricular tachycardias. The classifier performed well for respiratory sinus arrhythmia, sinus node dysfunction and ventricular paced rhythm and performances for other rhythms were unexceptional.

10.8.4 Support Vector Machine Classifier Performances for Different Rhythms

The support vector machine classifier rhythm specific performances are shown in Table 10.10.

Table 10.10 Support vector machine classifier performances by rhythm.

Rhythm	CCR	Error	Se	Sp	κ	*P*
Normal sinus rhythm	0.833	0.167	0.923	0.829	0.287	<0.001
Respiratory sinus arrhythmia	0.989	0.011	0.980	0.990	0.883	<0.001
Physiological sinus tachycardia	0.996	0.004	0.889	0.997	0.798	<0.001
Premature atrial contraction	0.678	0.322	0.131	0.983	0.140	<0.001
Premature ventricular contraction	0.695	0.305	0.130	0.997	0.159	<0.001
Sinus node dysfunction	0.988	0.012	0.931	0.990	0.800	<0.001
First degree atrio-ventricular block	0.970	0.030	0.000	0.973	-0.005	1.000
2nd and 3rd degree atrio-ventricular block	0.988	0.012	0.000	0.990	-0.003	1.000
Atrio-ventricular nodal junct. tachycardias	0.952	0.048	0.000	0.976	-0.024	1.000
Atrio-ventricular recip. tachycardias	0.460	0.540	0.083	0.464	-0.018	0.002
Abnormal tachycardias of the sinus node	0.997	0.003	0.000	1.000	0.000	1.000
Focal atrial tachycardia	0.993	0.007	0.000	1.000	0.000	1.000
Macro-re-entrant atrial tachycardia	0.993	0.007	0.000	1.000	0.000	1.000
Multifocal atrial tachycardia	0.985	0.015	0.000	1.000	0.000	1.000
Atrial fibrillation	0.979	0.021	0.000	1.000	0.000	1.000
Ventricular tachycardias	0.995	0.005	0.000	1.000	0.000	1.000
Atrial paced rhythm	0.957	0.043	0.000	1.000	0.000	1.000
Ventricular paced rhythm	0.973	0.027	0.000	1.000	0.000	1.000

CCR=correct classification rate; Se=sensitivity; Sp=specificity

The support vector machine was unable to diagnose eight rhythms: abnormal tachycardias of the sinus node; focal atrial tachycardia; macro-re-entrant atrial tachycardia; multifocal atrial tachycardia; atrial fibrillation; ventricular tachycardias; atrial paced rhythm or ventricular paced rhythm. Good performance was exhibited in classification of the following rhythms: respiratory sinus arrhythmia, with error of 0.011, sensitivity 0.980, specificity 0.990, κ of 0.883 and $P < 0.001$; physiological sinus tachycardia, with error of 0.004, sensitivity of 0.889, specificity 0.997, κ of 0.798, $P < 0.001$ and sinus node dysfunction with error of 0.012, sensitivity of 0.931, specificity 0.990, κ of 0.800 and $P < 0.001$.

10.8.5 Inference Engine Classifier Performances for Different Rhythms

The inference engine classifier rhythm specific performances are shown in Table 10.11.

Table 10.11 Inference engine classifier performances by rhythm.

Rhythm	CCR	Error	Se	Sp	κ	*P*
Normal sinus rhythm	0.901	0.099	0.846	0.904	0.403	<0.001
Respiratory sinus arrhythmia	0.724	0.276	0.592	0.730	0.090	<0.001
Physiological sinus tachycardia	0.972	0.028	0.667	0.975	0.270	<0.001
Premature atrial contraction	0.662	0.338	0.060	0.997	0.073	<0.001
Premature ventricular contraction	0.688	0.312	0.166	0.967	0.162	<0.001
Sinus node dysfunction	0.963	0.037	0.276	0.981	0.262	<0.001
First degree atrio-ventricular block	0.997	0.003	0.000	1.000	0.000	1.000
2nd and 3rd degree atrio-ventricular block	0.986	0.014	1.000	0.986	0.197	<0.001
Atrio-ventricular nodal junct. tachycardias	0.967	0.033	0.444	0.980	0.377	<0.001
Atrio-ventricular recip. tachycardias	0.979	0.021	0.417	0.985	0.293	<0.001
Abnormal tachycardias of the sinus node	0.989	0.011	0.000	0.992	-0.004	1.000
Focal atrial tachycardia	0.973	0.027	0.250	0.978	0.108	0.014
Macro-re-entrant atrial tachycardia	0.970	0.030	0.500	0.974	0.186	<0.001
Multifocal atrial tachycardia	0.976	0.024	0.000	0.991	-0.011	1.000
Atrial fibrillation	0.805	0.195	0.478	0.812	0.057	0.002
Ventricular tachycardias	0.904	0.096	0.500	0.907	0.044	0.014
Atrial paced rhythm	0.957	0.043	0.000	1.000	0.000	1.000
Ventricular paced rhythm	0.973	0.027	0.000	1.000	0.000	1.000

CCR=correct classification rate; Se=sensitivity; Sp=specificity

The inference engine classifier failed to diagnose five rhythms: first degree atrio-ventricular block; abnormal tachycardias of the sinus node; multifocal atrial tachycardia; atrial paced and ventricular paced rhythms. Best performance was achieved for atrio-ventricular nodal junctional tachycardias, with a *CCR* of 0.967; error 0.033; sensitivity 0.444; specificity 0.980; κ of 0.377 and $P < 0.001$. No notably high performances were exhibited.

10.8.6 Optimal Classifier by Rhythm

Subsections 10.8.1 to 10.8.5 indicated that different classifiers perform best for specific rhythms. The performances of each classifier, trained with its optimal feature set, were compared for each rhythm and the best performing classifier for that rhythm was selected, using the scoring system outlined in Chapter 10, section 10.7 (see Table 10.12).

The naïve Bayes classifier, trained with the 216 feature set (iteration 2) performed best for eleven rhythms, the decision tree classifier, trained with the 116 feature set (iteration 3) for four rhythms, and the support vector machine (116 features), neural network (116 features) and inference engine classifiers (216 features) performed best for one rhythm each.

Table 10.12 Optimal classifier for each rhythm by performance.

Classifier	Rhythm	CCR	Error	Sens.	Spec.	κ	P
NB	**Normal sinus rhythm**	0.964	0.036	0.827	0.971	0.664	<0.001
NB	**Respiratory sinus arrhythmia**	0.990	0.010	0.816	0.998	0.874	<0.001
NB	**Physiological sinus tachycardia**	0.998	0.002	1.000	0.998	0.899	<0.001
DT	**Premature atrial contraction**	0.802	0.198	0.599	0.914	0.544	<0.001
DT	**Premature ventricular contraction**	0.805	0.195	0.492	0.972	0.494	<0.001
SVM	**Sinus node dysfunction**	0.988	0.012	0.931	0.990	0.800	<0.001
NB	**First degree atrio-ventricular block**	0.999	0.001	0.667	1.000	0.800	<0.001
IE	**2nd and 3rd degree atrio-ventricular**	0.986	0.014	1.000	0.986	0.197	<0.001
NB	**Atrio-ventricular nodal junct.**	0.998	0.002	1.000	0.998	0.963	<0.001
DT	**Atrio-ventricular recip. tachycardias**	0.999	0.001	0.917	1.000	0.956	<0.001
NB	**Abnormal tachycardias of the sinus node**	1.000	0.000	1.000	1.000	1.000	<0.001
NB	**Focal atrial tachycardia**	1.000	0.000	1.000	1.000	1.000	<0.001
NB	**Macro-re-entrant atrial tachycardia**	0.999	0.001	1.000	0.999	0.941	<0.001
NN	**Multifocal atrial tachycardia**	0.987	0.013	0.235	0.999	0.359	<0.001
DT	**Atrial fibrillation**	0.986	0.014	0.913	0.988	0.625	<0.001
NB	**Ventricular tachycardias**	1.000	0.000	1.000	1.000	1.000	<0.001
NB	**Atrial paced rhythm**	0.975	0.025	0.938	0.976	0.750	<0.001
NB	**Ventricular paced rhythm**	0.995	0.005	0.833	0.999	0.890	<0.001

DT=decision tree; NB=Naïve Bayes; NN=neural network; SVM=support vector machine; IE=inference engine.

All the classification performances for each rhythm had a high *CCR*, with values of between 0.793 for premature ventricular contractions, to 1.000 for abnormal tachycardias of the sinus node, focal atrial tachycardia and ventricular tachycardias. High *CCR* represents high consistency (see Chapter 3, subsection 3.12.5). The bias-variance trade off tends to result in satisfactory performance when there is low classification error, corresponding to high *CCR*. For an 18-way comparison, statistical significance at the 5% level, with the Bonferroni correction $P_{crit} = 0.003$. All the classifiers had $P < 0.001$, for no significant difference to the gold standard, satisfying P_{crit}. Given all five classifiers produced during classifier development were included among these best performing classifiers, all were retained for inclusion in the production system.

10.9 Analysis to Allow Performance Comparisons to an ICD Classifier

Class recoding would have been required to evaluate the production classifier system performances for the differential diagnosis of VT and VF (combined) against supraventricular tachycardia, in a one-versus-one (OVO) classification (see Chapter 2, subsection 2.8.14 and Table 2.5), to make a direct comparison with tachycardia discrimination in ICDs and compare performance indices with the target criteria (see section 10.1).

Given that such a comparison was not a requirement central to satisfaction of the research question, that there was a zero incidence of VF patterns and that "perfect" performance had been achieved for diagnosis of VT by the naïve Bayes classifier (see subsection 10.8.2), the one-versus-all (OVA) analysis offered in section 10.8, comparing classification of VT (and VF) against all other rhythms, was considered an equivalent criterion.

10.10 Classifier Combination

The five trained classifiers (see subsection 10.8.6) were used to build a multi-classifier system (MCS) (see Chapter 3, section 3.16) using a mixture-of experts combination (Kuncheva 2004), which requires a gating classifier to allocate final class (see Chapter 3, Fig. 3.4). A simple voting combiner was not useful for this system, as optimal classification required a rhythm-specific decision, based on the best classifier for each rhythm.

The support vector machine and neural network classifiers were used to classify one rhythm each (see subsection 10.8.6) and were considered specialist classifiers. Both classifiers were rebuilt and re-trained as specialist binary classifiers, using the tuning stages outlined in subsections 10.2.5 and 10.2.6. All the developed classifiers were used in the production system so diversity measures suggested in Chapter 3, section 3.15 were not required to select classifiers.

10.10.1 A Mixture-of-Experts System

A mixture-of-experts system was designed, with gating using a decision rule. The decision rule was to function as an "expert" consultation system, consulting classifiers in reverse sequence of performance, starting with that having greatest performance with the highest number of different cardiac rhythms.

The first to be consulted was the naïve Bayes classifier and if the output ($Class_{NB}$) was any of the eleven rhythms listed for its use (see Table 10.12), those were allocated to the class label. If no class label was allocated, the decision tree was then consulted and if class was allocated ($Class_{DT}$) to any of its four rhythms, then for the remaining unclassified rhythms, the support vector machine ($Class_{SVM}$), neural network ($Class_{NN}$) and inference engine classifiers, ($Class_{IE}$) sequentially.

10.10.2 Unclassified Instances

A strategy was required to deal with failure to classify any instance, where the consultation system had been exhausted. In the classifier design process, the guideline-based inference engine was intended to fulfil such a role. Testing indicated that the inference engine classifier performed well for all rhythms, with a κ of 0.59, making this a reasonable approach.

10.10.3 The Decision Rule

Using the best classifier for each rhythm (see subsection 10.8.6), application of a sequential consultation system (see subsection 10.10.1) and the strategy for unclassified instances (see subsection 10.10.2), a decision rule was derived:

IF $Class_{NB}$ = NSR **OR** RSA **OR** ST **OR** FHB **OR** AVNRT **OR** IST
OR AT **OR** MRAT **OR** VT **OR** APACE **OR** VPACE
THEN $Class_{NB}$ = Class Label
ELSE IF $Class_{DT}$ = PAC **OR** PVC **OR** AVRT **OR** AF
THEN $Class_{DT}$ = Class Label
ELSE IF $Class_{SVM}$ = SND
THEN $Class_{SVM}$ = Class Label
ELSE IF $Class_{NN}$ = MAT
THEN $Class_{NN}$ = Class Label
ELSE $Class_{IE}$ = Class Label

10.11 The Production System

The classifier units, and decision rule gating resulted in a multiple classifier system model (see Fig. 10.7).

10.11.1 Production system performance

A confusion matrix was generated for the production system, tested with all the data (see Table 10.13) and performance indices calculated (see Tables 10.14 and 10.15). Major indices of error, sensitivity, specificity and κ had 95% confidence intervals (CI_{95}) calculated (see Table 10.14).

Performance of the production system classifier was good for all rhythms, with an overall *CCR* of 0.960 and error of 0.040 ± 0.012. Mean values were: sensitivity of 0.855; specificity of 0.977; κ of 0.767; PPV of 0.792; NPV of 0.975; ϕ of 0.787) with all rhythms having *P* values being below the required Bonferroni P_{crit} of 0.004, equivalent to $P = 0.05$ for an 18 way comparison, supporting there being no difference with the gold standard (see Tables 10.14 and 10.15).

Performances were largely preserved from those calculated for the optimal; classifier units, in Table 10.12, with the exception of multifocal atrial tachycardia, with a markedly improved , sensitivity of 0.765, compared with 0.235 achieved with the neural network classifier unit alone. For indices with 95% confidence intervals, CI_{95} values were all within 0.03, meaning imprecision was 3% or better. For all 18 different rhythms, error ranged from 0 to 0.246 and for 15 rhythms, other than for normal sinus rhythm, premature atrial and ventricular contractions, was 0 to 0.014.

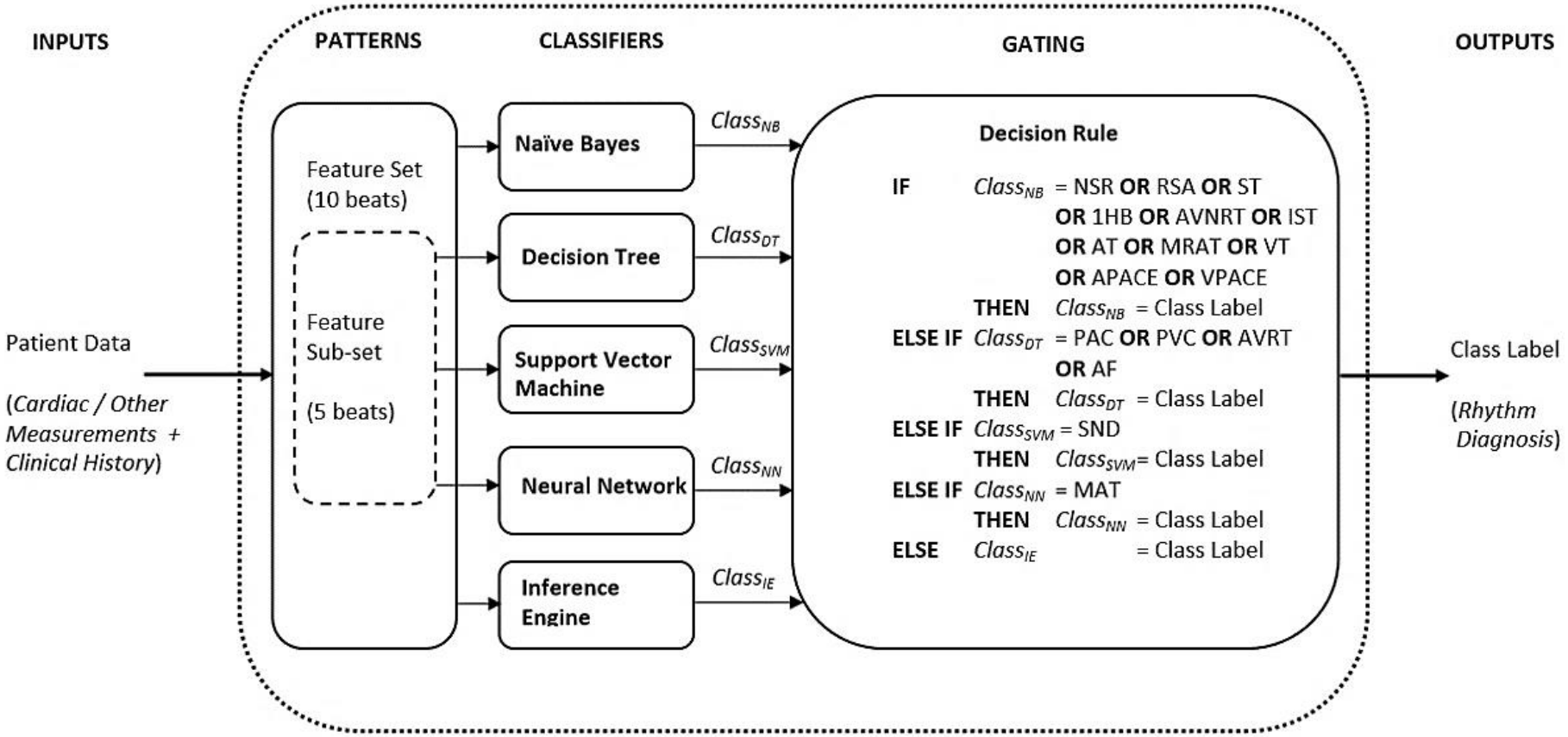

Figure 10.7 Production system flow diagram. Data is input at the left of the system. Data is processed into patterns (instances), containing a set of features, presented as inputs near simultaneously to classifier units. Class allocation of each classifier is passed through a gating decision rule, in sequence: naïve Bayes first, then decision tree, then support vector machine, then neural network and finally an inference engine, to allocate final class.

Table 10.13 Production system output confusion matrix, by rhythm.

		Test Positive (Output Class)																	
		NSR	**RSA**	**ST**	**PAC**	**PVC**	**SND**	**FHB**	**SHB**	**AVNRT**	**AVRT**	**IST**	**AT**	**MRAT**	**MAT**	**AF**	**VT**	**APACE**	**VPACE**	**Total**
Gold Standard Diagnosis (Target Class)	**NSR**	45	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	5	0	**52**
	RSA	6	40	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	**49**
	ST	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**9**
	PAC	113	1	0	237	11	0	0	12	0	0	0	0	0	11	7	0	5	0	**397**
	PVC	135	0	2	56	190	0	0	0	1	0	0	0	0	0	2	0	0	0	**386**
	SND	1	0	0	1	0	27	0	0	0	0	0	0	0	0	0	0	0	0	**29**
	FHB	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	**3**
	SHB	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	**2**
	AVNRT	0	0	0	0	2	0	0	0	25	0	0	0	0	0	0	0	0	0	**27**
	AVRT	2	0	0	0	0	0	0	0	1	9	0	0	0	0	0	0	0	0	**12**
	IST	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	**3**
	AT	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	**8**
	MRAT	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	**8**
	MAT	2	0	0	0	2	0	0	0	0	0	0	0	0	13	0	0	0	0	**17**
	AF	1	0	0	0	1	0	0	0	0	0	0	0	0	1	20	0	0	0	**23**
	VT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	**6**
	APACE	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	44	1	**48**
	VPACE	4	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	24	**30**
	Total	**311**	**42**	**11**	**295**	**209**	**29**	**2**	**14**	**27**	**9**	**3**	**8**	**9**	**25**	**30**	**6**	**54**	**25**	**1109**

Table 10.14 Production system major performance indices, by rhythm.

Rhythm	CCR	Error	±	CI_{95}	Se	±	CI_{95}	Sp	±	CI_{95}	κ	±	CI_{95}	*P*
Normal sinus rhythm	0.754	0.246	±	0.025	0.865	±	0.020	0.748	±	0.026	0.182	±	0.023	<0.001
Respiratory sinus arrhythmia	0.990	0.010	±	0.006	0.816	±	0.023	0.998	±	0.003	0.874	±	0.020	<0.001
Physiological sinus tachycardia	0.998	0.002	±	0.002	1.000	±	0.000	0.998	±	0.003	0.899	±	0.018	<0.001
Premature atrial complex(es)	0.803	0.197	±	0.023	0.597	±	0.029	0.919	±	0.016	0.547	±	0.029	<0.001
Premature ventricular complex(es)	0.806	0.194	±	0.023	0.492	±	0.029	0.974	±	0.009	0.522	±	0.029	<0.001
Sinus node dysfunction	0.996	0.004	±	0.004	0.931	±	0.015	0.998	±	0.003	0.929	±	0.015	<0.001
First degree AV block	0.999	0.001	±	0.002	0.667	±	0.028	1.000	±	0.000	0.800	±	0.024	<0.001
Second/third degree AV block	0.989	0.011	±	0.006	1.000	±	0.000	0.989	±	0.006	0.248	±	0.025	<0.001
AV nodal/ junctional tachycardias	0.996	0.004	±	0.004	0.926	±	0.015	0.998	±	0.003	0.924	±	0.016	<0.001
AV reciprocating tachycardias	0.997	0.003	±	0.003	0.750	±	0.025	1.000	±	0.000	0.856	±	0.021	<0.001
Inappropriate sinus tachycardia	1.000	0.000	±	0.000	1.000	±	0.000	1.000	±	0.000	1.000	±	0.000	<0.001
Focal atrial tachycardia	1.000	0.000	±	0.000	1.000	±	0.000	1.000	±	0.000	1.000	±	0.000	<0.001
Macro-reentrant atrial tachycardia	0.999	0.001	±	0.002	1.000	±	0.000	0.999	±	0.002	0.941	±	0.014	<0.001
Multifocal atrial tachycardia	0.986	0.014	±	0.007	0.765	±	0.025	0.989	±	0.006	0.612	±	0.029	<0.001
Atrial fibrillation	0.988	0.012	±	0.006	0.870	±	0.020	0.991	±	0.006	0.749	±	0.026	<0.001
Ventricular tachycardias	1.000	0.000	±	0.000	1.000	±	0.000	1.000	±	0.000	1.000	±	0.000	<0.001
Atrial paced rhythm	0.987	0.013	±	0.007	0.917	±	0.016	0.991	±	0.006	0.856	±	0.021	<0.001
Ventricular paced rhythm	0.994	0.006	±	0.005	0.800	±	0.024	0.999	±	0.002	0.870	±	0.020	<0.001
ALL (mean)	**0.960**	**0.040**	**±**	**0.012**	**0.855**			**0.977**			**0.767**			**<0.004**

CCR = correct classification rate; Se = sensitivity; Sp = specificity; κ = Cohen's kappa; CI_{95} = 95% confidence interval; *P* values to 3 significant figures

Table 10.15 Additional performance indices by rhythm. Note that prevalence is the rate of occurrence in sample instances not the population.

Rhythm	FPR (α)	FNR (β)	Odds Ratio	Relative Risk	PPV	NPV	F_1 score	Youles Q	Pearsons ϕ	Prevalence
Normal sinus rhythm	0.252	0.135	19.12	16.49	0.145	0.991	0.248	0.901	0.289	0.047
Respiratory sinus arrhythmia	0.002	0.184	2351	112.9	0.952	0.992	0.879	0.999	0.877	0.044
Physiological sinus tachycardia	0.002	0.000	-	-	0.818	1.000	0.900	-	0.904	0.008
Premature atrial complex(es)	0.081	0.403	16.70	4.087	0.803	0.803	0.685	0.887	0.559	0.358
Premature ventricular complex(es)	0.026	0.508	35.92	4.174	0.909	0.782	0.639	0.946	0.568	0.348
Sinus node dysfunction	0.002	0.069	7277	502.8	0.931	0.998	0.931	1.000	0.929	0.026
First degree AV block	0.000	0.333	-	1107	1.000	0.999	0.800	-	0.816	0.003
Second/third degree AV block	0.011	0.000	-	-	0.143	1.000	0.250	-	0.376	0.002
AV nodal/ junctional tachycardias	0.002	0.074	6750	500.9	0.926	0.998	0.926	1.000	0.924	0.024
AV reciprocating tachycardias	0.000	0.250	-	366.7	1.000	0.997	0.857	-	0.865	0.011
Inappropriate sinus tachycardia	0.000	0.000	-	-	1.000	1.000	1.000	-	1.000	0.003
Focal atrial tachycardia	0.000	0.000	-	-	1.000	1.000	1.000	-	1.000	0.007
Macro-reentrant atrial tachycardia	0.001	0.000	717.3	239.8	0.667	0.997	0.755	0.997	0.942	0.007
Multifocal atrial tachycardia	0.011	0.235	292.5	140.9	0.520	0.996	0.619	0.993	0.624	0.015
Atrial fibrillation	0.009	0.130	717.3	239.8	0.667	0.997	0.755	0.997	0.756	0.021
Ventricular tachycardias	0.000	0.000	-	-	1.000	1.000	1.000	-	1.000	0.005
Atrial paced rhythm	0.009	0.083	1156	214.9	0.815	0.996	0.863	0.998	0.858	0.043
Ventricular paced rhythm	0.001	0.200	4312	173.4	0.960	0.994	0.873	1.000	0.873	0.027
ALL (mean)	**0.023**	**0.145**			**0.792**	**0.975**	**0.777**		**0.787**	**0.056**

FPR=false positive rate; FNR=false negative rate; α =type I error; β = type II error; PPV=positive predictive value; NPV=negative predictive valu

Perfect diagnostic performances were exhibited for 3 rhythms, with errors of 0, sensitivities, specificities and κ all having values of 1 and $P < 0.001$, these were: abnormal tachycardias of the sinus node, focal atrial tachycardia and ventricular tachycardias. Poor performance, where κ was < 0.2, was encountered only for normal sinus rhythm, with a *CCR* of 0.754; error of 0.246 ± 0.025; sensitivity of 0.865 ± 0.020; specificity of 0.748 ± 0.026; κ of 0.182 ± 0.023 and $P < 0.001$. Good performance was exhibited by the remaining rhythms.

Prevalence values represented the incidence of the various rhythms in the sample population and had relevance in the interpretation of results. For rhythms with very low prevalence, high performance indices were less reliable, as they were based on very small numbers. This was the case for first degree atrio-ventricular block with prevalence of 0.002; abnormal tachycardias of the sinus node with prevalence of 0.003; ventricular tachycardia with prevalence of 0.005; focal atrial tachycardia with prevalence of 0.007; atrio-ventricular reciprocating tachycardias with prevalence of 0.008 and macro-re-entrant atrial tachycardia with prevalence of 0.008 (see Table 10.18).

Odds ratio (OR) and relative risk values were not computable for some rhythms. High positive and negative predictive values (PPV and NPV) values of 0.5 or greater were found for all rhythms with the exception of normal sinus rhythm, with PPV of 0.145 and second/third degree AV block, with PPV of 0.143. Values of F_1 score and Pearson's ϕ for individual rhythms corresponded well with κ but added little to the interpretation (see Table 10.18).

The confusion matrix (Table 10.13) allowed rhythm diagnoses to be examined in more detail in an attempt to identify and explain misclassifications. No misclassifications were diagnosed for physiological sinus tachycardia, second and third degree heart blocks, inappropriate sinus tachycardia, focal atrial tachycardia, macro-re-entry atrial tachycardia and ventricular tachycardia. Gold standard diagnoses were diagnosed as other rhythms (false negatives) for 6 of 52 (11.5%) diagnoses of normal sinus rhythm; 9 of 49 (8%) of respiratory sinus arrhythmia; 160 of 397 (40%) of premature atrial complexes; 196 of 386 (51%) of premature ventricular complexes; 2 of 29 (7%) of sinus node disease; 1 of 3 (33%) of first degree heart block; 2 of 27 (7%) of atrio-ventricular and junctional tachycardias; 3 of 12 (25%) of atrio-ventricular re-entry tachycardias; 4 of 17 (23%) of multifocal atrial tachycardia; 3 of 23 (13%) of atrial fibrillation; 4 of 48 (8%) of atrial paced rhythm and 6 of 30 (20%) of ventricular paced rhythm.

High rates of misclassification were observed for normal sinus rhythm, with 266 false positives and 7 false negatives, an error 0.246 ± 0.025; premature atrial contractions with 58 false positives and 160 false negatives, an error 0.197 ± 0.023 and premature ventricular contractions with 19 false positives and 196 false negatives, an error of 0.194 ± 0.023. The majority of

misclassifications were included within 113 PAC's (28%) misdiagnosed as normal sinus rhythm; 11 as PVC's; 12 as second or third degree heart block and 11 as multifocal atrial tachycardia; 135 PVC's (34%) misdiagnosed as normal sinus rhythm and 56 (14%) as PAC's. The remaining small number of misdiagnoses did not demonstrate clear patterns.

This analysis showed that PA's and PVC's were frequently misdiagnosed as normal sinus rhythm and PAC's and PVC's respectively into the others class. The key indicators for differentiation of PAC's and PVC's both from sinus rhythm and each other's class were prematurity and QRS morphology. Alterations to the detection thresholds for these features could further improve diagnostic accuracy.

10.12 Summary

Five classifiers were successfully developed and tuned using Matlab - a decision tree, a naïve Bayes classifier, a neural network, a support vector machine and an inference engine. Iterative testing using different data sets showed that different classifiers were optimal for different rhythms. A production classifier system was-designed to make best use of these characteristics. This was a multi-classifier system, in a mixture-of-experts configuration, with decision rule gating.

For the production system, a confusion matrix was generated, and performance indices calculated. Performance was good for all rhythms, with all rhythms having $P < 0.001$ supporting no difference with the gold standard. Performance for multifocal atrial tachycardia, was improved with sensitivity of 0.765, compared with 0.235 achieved with the neural network classifier unit alone. imprecision was 3% or better. Perfect diagnostic performances were exhibited for 3 rhythms, reduced performance, with κ was < 0.2, was for normal sinus rhythm and there was good performance by the remaining rhythms. Low prevalence made performance indices less reliable for 6 rhythms.

Miscalssifications were analysed and high misclassification rates were observed for normal sinus rhythm, premature atrial contractions and premature ventricular contractions. Suggesting that there was potential for alteration of detection thresholds to improve diagnostic accuracy.

Chapter 11 Discussion

11.1 A Holistic Approach

This study adopted a holistic approach to the problem of developing a more accurate cardiac rhythm classifier, re-examining the issue from basic principles. The thinking behind this approach, rather than a piecemeal or focussed approach, such as examining classifier performances without carefully considering feature selection or without considering the classifier development processes was that one could otherwise neglect an import aspect worthy of inclusion.

The problem of misclassification by existing algorithms led to the research question and published research was examined for its contributions. Two main areas of research were identified, involving bench testing of classifier technology with its application in cardiology and clinical research, evaluating algorithms already in use. There was very little research work linking the two fields and this study aimed to bridge the gap, between technological evaluation and real-world application, a field known generically as translational medicine.

This study used heuristics at specific design stages, an algorithmic approach with intelligent agents for rhythm classifier design and a cognitive approach to guide choices in classifier design. The domain of AI was examined and its utility in classification problems outlined, emphasising: Bayesian classifiers; fuzzy classifiers; decision trees; neural networks and support vector machines.

11.1.1 Modelling

Modelling, rarely used in this field, was successfully used to assist in visualisation of the domain and processes, with the ontology being modelled as a hierarchical taxonomy and the clinical diagnostic process in a cognitive model. A systematic approach to the mechanics of system development and testing used an incremental SDLC was used, incorporating sub-structures of classifier development and knowledge engineering, aimed at incorporating domain expert knowledge at appropriate points and the Hatley-Pirbhai system context diagram was used as a system specification.

11.1.2 Data Collection

Data collection was conducted in an ethically approved study, as it involved the collection of human physiological data. The population was chosen from those most easily accessible in a cardiology unit, with high risk of rhythm disturbances. The data collection was successfully planned with equipment specified in detail and was performed on patients while they underwent invasive EP study and was successful in 61 of 65 attempts. The small data loss experienced was

attributed to the complexity of the setup including the number of simultaneously recording systems (Bostock 2010).

11.1.3 Offline Analysis

Offline analysis of collected data was complex and was performed using Matlab scripts. Physiological data and MatLab scripts written specifically for this study are available for download (see Appendix L). This required extensive knowledge of Matlab and was very time-consuming. Data was concatenated into a single file for each patient then digital signal processing and imputation were used to process data into a form suitable for feature extraction. The Zhang algorithm for classifier input data pre-processing was successfully used to guide this process. 10 beat rhythm segments, for each rhythm encountered during a procedure, were extracted as instances and class labelling based on concurrent rhythm annotations made at the time of data collection by the domain expert consultant cardiologist. Generated instances were confirmed by a domain expert for accuracy and validity prior to use as classifier inputs. All these elements were integrated into a continuous process which was found to be a successful and comprehensive approach to classifier development.

11.2 Relative Importance of Features

The suggestion from review papers led to the concept that a re-evaluation of features commonly included in typical comparable algorithms would be useful. Re-examination of features required for a cardiac rhythm diagnostic algorithm started with a baseline of those features widely accepted and used extensively in those studies reviewed in Chapter 2, as having a role, such as heart rate and features of the ECG as well as technologies widely used in cardiac rhythm diagnostic systems, such as RR intervals and morphology. Related technologies previously used or evaluated were also considered, with emphasis on those incorporated in complex implantable devices, ICDs. It was considered expedient to use technologies with known utility, augmenting them with new features rather than a "clean sheet" approach. A knowledge engineering approach then considered the general clinical approach to arrhythmia diagnosis and differential diagnosis of arrhythmia, taken predominantly from clinical guidelines (Bostock 2010). New factors related to clinical history, physical examination and some ECG related features emerged from this process as candidate features.

To test the utility of these new features in a rhythm classifier required new data collection, as similar data did not exist in established databases. The brief examination of principal features using an analysis of decision tree outputs during iterative classifier testing (see Chapter 10, subsection 10.3.1), suggested further analysis to attempt to determine the reasons for its high classification performance related to feature selection.

Relative feature importance was formally assessed using the ReliefF classification algorithm (Kononenko 1994), available in the Matlab statistics toolbox. The 216 feature set was input and compared to the output of the production system, using 10-nearest neighbours per class. The total of all weights was 6.04277. 30 features of high importance were found, with a sum of weights of 1.904365, representing 31% of all feature weights (see Table 11.1 and Fig. 11.1).

Table 11.1 Top 30 features using ReliefF, ranked by weight.

Weight	Beat	Feature description
0.14074	R0	VTCsamples
0.09483		Rx
0.09195	R0	PRint
0.07703		Caffeine
0.07468		SVIonset
0.07332	R0	PAXISmatchsamples
0.06819	R0	RPint
0.06736		Syncope
0.06633	R0	QRSgrossareamatchsamples
0.06522	R0	QRSAXISmatchsamples
0.06316	R1	PAXISmatchsamples
0.06179	R1	PPint
0.06176	R0	VTCAsamples
0.06102	R0	PPint
0.06077		Sm
0.05851	R2	PAXISmatchsamples
0.05841	R1	VTCsamples
0.05633	R0	RRint
0.05481		ETOH
0.05381	R1	RPint
0.05231	R0	QRSwidthmatchsamples
0.05162	R2	VTCsamples
0.05134	R1	RRint
0.05022	R2	PPint
0.05012		Cardiac/Pulm Dis
0.04817	R1	QRSgrossareamatchsamples
0.04811	R1	QRSAXISmatchsamples
0.04800	R-1	QRSAXISmatchsamples
0.04791	R3	PPint
0.04656	R1	QRSnetareamatchsamples
Total	**1.904365**	

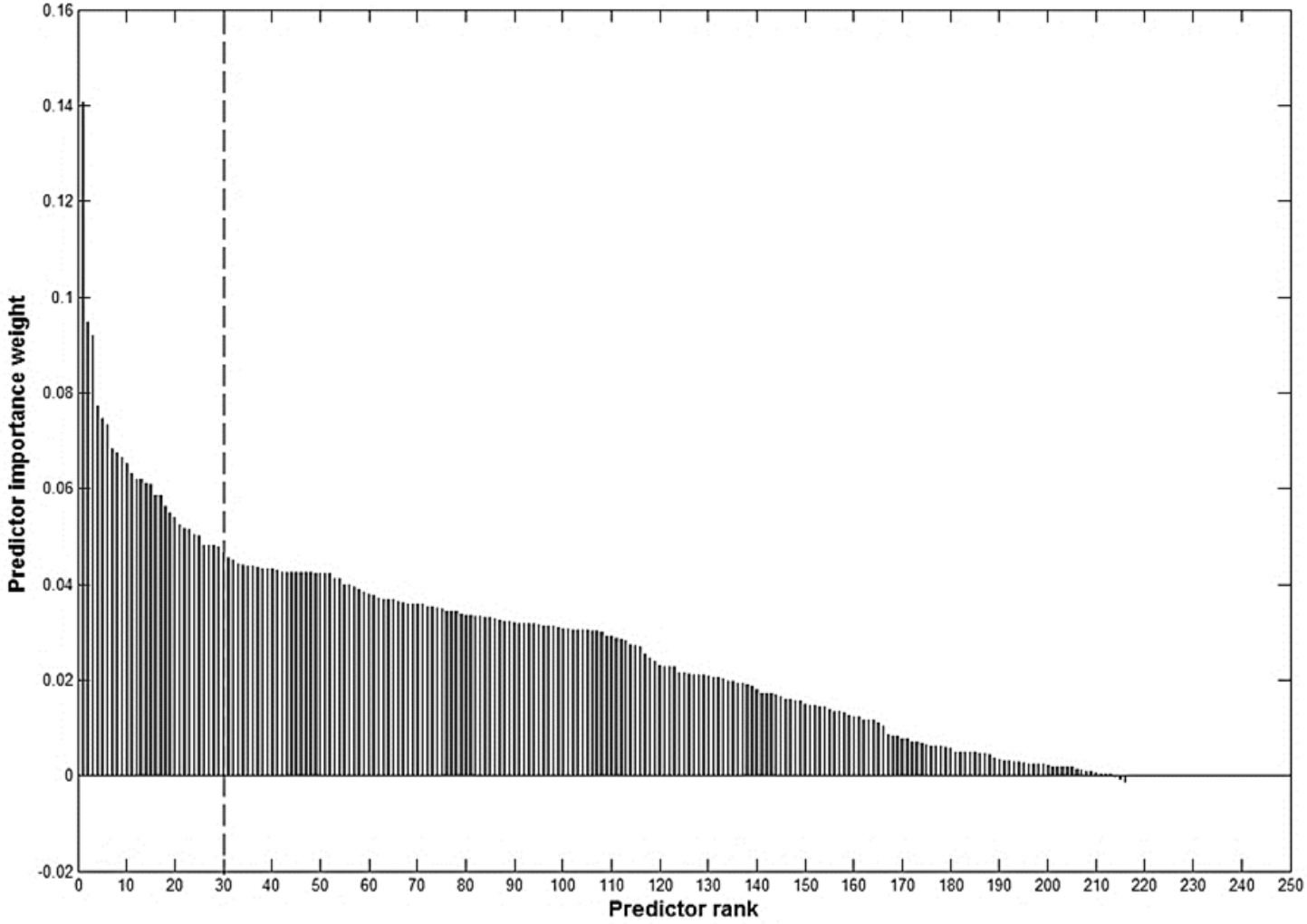

Figure 11.1 Feature importance distribution, from ReliefF analysis. Features are unidentified in this illustration showing that that there are few features with high importance, less than 10 with weighting > 0.05, and many of low importance. Detailed analysis allows key features to be identified (see text).

The feature with highest ranking was VTC for beat R_0, representing QRS morphology of the first beat of anew rhythm. Ten among the top 30 weighted features related to QRS morphology: VTC for beats R_0, R_1 and R_2, the first three beats of a new rhythm; QRS gross area match for beats R_0 and R_1; QRS width match for beat R_0; QRS axis match for beats R_{-1}, R_0 and R_1 and QRS net area match for beat R_1. This placed a very high importance on QRS morphology in rhythm diagnosis.

Nine features related to intervals: PP interval of beats R0, R1, R2 and R3; RR interval of beats R0 and R1; PR interval of beat R0 and the RP interval of beats R0 and R1, also emphasising the importance of interval features.

Five beats: R_{-1}, R_0, R_1, R_2 and R_3, provided features in the top 30 features ranked by weight, supported the belief expressed in Chapter 3, subsection 3.8.2 of the need for only 5 beats to diagnose rhythm, though these were described as beats R_{-2} to R_2.

A new finding of this study was the high importance attributable to symptoms and predisposing factors. 6 of the 12 predisposing factor features collected were ranked in the top 30 features: symptoms of syncope (blackout); a history of cardiac or pulmonary disease; high caffeine intake; high alcohol intake; smoking and prescribed anti-arrhythmic medication. Those features not ranked were: tiredness or lethargy; chest pain; shortness of breath; palpitations; abnormal heart sounds sign; metabolic conditions.

Another new finding was that a change in haemodynamic status, represented by the change in stroke volume index at onset of a new rhythm (SVI onset), ranked fifth among the top 30 features, suggesting a previously unestablished importance in rhythm diagnosis. Additionally, P wave morphology was responsible for 4 features, with P axis match for beats R_0, R_1, R_2 and VTCA for beat R_0.

No features relating to stress, such as QT interval, temperature and accelerometry were placed within the top 30 rank. This was not unexpected, as patients enrolled in this study were immobile for the duration of the data collection exercise, minimising the likely impact of physiological stress.

11.2.1 Influence of Accelerometry, Temperature and QT Interval

Given the absence of accelerometry, temperature and QT interval features within the top 30 ranked according to ReliefF analysis, reanalysis of data was proposed, removing these features, to evaluate their impact on results.

Evaluation began with assessing the influence of accelerometry, temperature and QT interval features individually on performance with each of the 5 classifiers used in the production system. Classifier units were designed using all features, containing data from 10 beats.

For data containing all the features, SVM and NN classifiers were optimal when trained as binary classifiers for sinus node dysfunction and multifocal atrial tachycardia respectively (see Chapter 10, subsection 10.8.6). The specialist classifiers were trained using the modified feature set, excluding accelerometry, temperature and QT interval features, with a 10% hold-out test set. The trained models were stored to be re-used in a modified production system.

Decision trees were generated with optional settings set as for iteration 3, using a 'symmetric' ($2x$-1) transformation function. The tree from the fold with the lowest kappa was selected as best and tested with all instances. Naïve Bayes classifiers were generated with optional settings set as for iteration 3, using 'distribution', 'kernel', 'Prior' and 'empirical'. The classifier from the fold with the lowest kappa was selected as best and tested with all instances. Neural networks were generated using 10-fold cross validation, with optional settings set as for iteration 3, using 11 hidden nodes, a 'softmax' hidden layer transfer function, a 'purelin' output layer transfer function, a 'trainbr' learning algorithm. The network from the fold with the lowest kappa was selected as best and tested with all instances. Support vector machines (SVM) were generated and retested with all the data. The inference engine was tested without training, using all features and the full dataset. Lines of code corresponding to accelerometry and QT deactivated respectively and then together. Note that temperature was not used in the inference structure.

Performances were compared with the corresponding base classifier used in the production system (see Table 11.2). 2-way comparisons were made, using the Bonferroni correction with P_{crit} at 0.025.

The decision tree classifier base unit showed no differences in *CCR*, error, sensitivity, specificity, kappa or *P* value between feature sets where individual features were removed from input data. A performance improvement was observed when all three features (accelerometry, temperature and QT interval) were removed. Positive predictive value (PPV) was improved at 0.978, compared with 0.719; specificity improved to 0.999, compared with 0.983 and kappa improved to 0.903 compared with 0.782, though other indices and *P* values showed no differences.

The neural network classifier demonstrated small performance improvement with individual features removed and a greater performance improvement across all indices with all three features removed: *CCR* improved to 0.963 compared with 0.889; PPV to 0.889 from 0.667; sensitivity was unchanged at 1.0; specificity to 0.947 from 0.857 and kappa to 0.914 from 0.727, though *P*

Table 11.2 Influence of accelerometry, temperature and QT interval on performance. For each classifier, columns (from left to right) represent performance of: the production system base unit (Prod.); the full dataset without accelerometry features (No Acc); without temperature features (No T); without QT interval features (No QT) and without all three (NoATQ).

	Decision Tree					**Naïve Bayes**					**Neural Network**				
	Prod.	No Acc	No T	No QT	NoATQ	Prod.	No Acc	No T	No QT	NoATQ	Prod.	No Acc	No T	No QT	NoATQ
CCR	0.908	0.914	0.908	0.909	0.911	0.853	0.847	0.836	0.882	0.845	0.889	0.920	0.923	0.913	0.963
Error	0.092	0.086	0.092	0.091	0.089	0.147	0.153	0.164	0.118	0.155	0.111	0.080	0.077	0.087	0.037
NL	0.117	0.136	0.136	0.117	0.154	0.615	0.328	0.655	0.462	0.693	0.000	0.000	0.000	0.000	0.000
PL	51.9	83.1	114.3	62.3	894	-	237	183	285	325	7.00	10.0	10.0	10.0	19.0
NPV	0.994	0.993	0.993	0.994	0.992	0.971	0.984	0.969	0.978	0.967	1.000	1.000	1.000	1.000	1.000
PPV	0.719	0.804	0.849	0.754	0.978	1.000	0.921	0.900	0.933	0.941	0.667	0.714	0.750	0.600	0.889
Prevalence	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.222	0.200	0.231	0.130	0.296
Sensitivity	0.885	0.865	0.865	0.885	0.846	0.385	0.673	0.346	0.538	0.308	1.000	1.000	1.000	1.000	1.000
Specificity	0.983	0.990	0.992	0.986	0.999	1.000	0.997	0.998	0.998	0.999	0.857	0.900	0.900	0.900	0.947
kappa	0.782	0.825	0.850	0.804	0.903	0.544	0.769	0.487	0.672	0.451	0.727	0.783	0.806	0.701	0.914
***P* value**	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0003	0.0004	0.0001	0.0057	<0.0001
	Support Vector Machine					**Inference Engine**									
	Prod.	No Acc	No T	No QT	NoATQ	Prod.	No Acc	No T	No QT	NoATQ					
CCR	0.883	0.869	0.966	0.872	0.872	0.193	0.193		0.193	0.193					
Error	0.117	0.131	0.034	0.128	0.128	0.807	0.807		0.807	0.807					
NL	0.063	0.084	0.072	0.105	0.105	0.170	0.170		0.170	0.170					
PL	11.3	11.2	4.65	10.9	11.2	8.77	8.77		8.77	8.77					
NPV	0.997	0.996	0.996	0.995	0.995	0.992	0.992		0.992	0.992					
PPV	0.358	0.356	0.186	0.348	0.356	0.301	0.301		0.301	0.301					
Prevalence	0.047	0.047	0.047	0.047	0.047	0.047	0.047		0.047	0.047					
Sensitivity	0.942	0.923	0.942	0.904	0.904	0.846	0.846		0.846	0.846					
Specificity	0.917	0.918	0.798	0.917	0.920	0.904	0.904		0.904	0.904					
kappa	0.483	0.478	0.253	0.467	0.476	0.403	0.403		0.403	0.403					
***P* value**	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001		<0.0001	<0.0001					

values showed no significant difference.

The naïve Bayes, support vector machine and inference engine classifier base units showed no significant differences in performance between all the feature sets.

Testing showed that no significant differences in performance could be ascribed to the absence of accelerometry, temperature and QT interval data. Equally important, no performance advantage was gained by their removal, reducing the likelihood of any confounding influence.

11.2.2 Comparing the 10 beat and 5 beat Diagnostic Models

The use if iteration 3 in three of the five classifier models used in the mixture-of-experts multi-classifier production system suggested utility for the 5 beat rhythm analysis model. Further testing was performed to determine whether the dimensionality reduction implicit in a 5 beat diagnostic model provided a performance advantage, when compared with a 10 beat model.

The production model was modified to run with datasets containing the minimum possible features, based on the guideline driven feature selection process adhered to in this study and the findings that the influence of accelerometry, temperature and QT interval data were not significant (see subsection 11.2.1). Two datasets were created, based on 10 beats and 5 beats (see Chapter 3, subsection 3.18.2).

Each classifier unit was re-trained with the revised datasets and the corresponding models stored for use in 5 beat and 10 beat multi-classifier systems, having structure identical to the production system (see Chapter 10, Fig. 10.7), except the input sets used either 5 beat or 10 beat data, not both. Analysis used the same procedure as for the production system (see Chapter 10, subsection 10.11.1 and in Figs. 10.16, 10.17 and 10.18).

10 beat analysis (see Tables 11.3 and 11.4) showed small improvements in performance indices over the original production system for some rhythms. 5 beat analysis showed poor overall performance, with high correct classification rates and high specificities with very low sensitivity and kappa values for all rhythms, other than atrial or ventricular paced rhythms (see Tables 11.5 and 11.6).

CCR was clearly higher and error lower for 10 beat analysis, compared with 5 beat analysis for four rhythms: respiratory sinus arrhythmia (RSA) *CCR* of 0.995 compared to 0.636; physiological sinus tachycardia (ST) with *CCR* of 0.995 compared with 0.652; premature atrial complexes (PAC) with *CCR* of 0.960 compared with 0.635 and premature ventricular complexes (PVC) with *CCR* of 0.950 compared with 0.662.

Table 11.3 Major performance indices, for 10 beat analysis - excluding accelerometry, temperature and QT interval features.

Rhythm	CCR	Error	±	CI_{95}	Se	±	CI_{95}	Sp	±	CI_{95}	κ	±	CI_{95}	*P*
Normal sinus rhythm	0.957	0.043	±	0.012	0.885	±	0.019	0.960	±	0.011	0.636	±	0.028	<0.001
Respiratory sinus arrhythmia	0.995	0.005	±	0.004	0.939	±	0.014	0.998	±	0.003	0.946	±	0.013	<0.001
Physiological sinus tachycardia	0.995	0.005	±	0.004	0.778	±	0.024	0.997	±	0.003	0.735	±	0.026	<0.001
Premature atrial complex(es)	0.960	0.040	±	0.011	0.942	±	0.014	0.971	±	0.010	0.914	±	0.017	<0.001
Premature ventricular complex(es)	0.950	0.050	±	0.013	0.930	±	0.015	0.960	±	0.012	0.889	±	0.018	<0.001
Sinus node dysfunction	0.997	0.003	±	0.003	0.897	±	0.018	1.000	±	0.000	0.944	±	0.014	<0.001
First degree AV block	1.000	0.000	±	0.000	1.000	±	0.000	1.000	±	0.000	1.000	±	0.000	<0.001
Second/third degree AV block	1.000	0.000	±	0.000	1.000	±	0.000	1.000	±	0.000	1.000	±	0.000	<0.001
AV nodal/ junctional tachycardias	0.995	0.005	±	0.004	0.852	±	0.021	0.999	±	0.002	0.900	±	0.018	<0.001
AV reciprocating tachycardias	0.993	0.007	±	0.005	1.000	±	0.000	0.993	±	0.005	0.747	±	0.026	<0.001
Inappropriate sinus tachycardia	0.998	0.002	±	0.002	0.333	±	0.028	1.000	±	0.000	0.499	±	0.029	0.003
Focal atrial tachycardia	0.997	0.003	±	0.003	0.625	±	0.028	1.000	±	0.000	0.768	±	0.025	<0.001
Macro-reentrant atrial tachycardia	0.999	0.001	±	0.002	0.875	±	0.019	1.000	±	0.000	0.933	±	0.015	<0.001
Multifocal atrial tachycardia	0.995	0.005	±	0.004	0.765	±	0.025	0.998	±	0.003	0.810	±	0.023	<0.001
Atrial fibrillation	0.993	0.007	±	0.005	0.826	±	0.022	0.996	±	0.004	0.822	±	0.022	<0.001
Ventricular tachycardias	0.995	0.005	±	0.004	0.000	±	0.000	1.000	±	0.000	0.000	±	0.000	1.000
Atrial paced rhythm	0.980	0.020	±	0.008	0.604	±	0.029	0.997	±	0.003	0.715	±	0.027	<0.001
Ventricular paced rhythm	0.993	0.007	±	0.005	0.733	±	0.026	1.000	±	0.000	0.843	±	0.021	<0.001
ALL (mean)	**0.988**	**0.012**	**±**	**0.005**	**0.777**	**±**	**0.017**	**0.993**	**±**	**0.003**	**0.350**	**±**	**0.028**	**<0.001**

CCR = correct classification rate; Se = sensitivity; Sp = specificity; κ = Cohen's kappa; CI_{95} = 95% confidence interval; *P* values to 3 significant figures

Table 11.4 Additional performance indices for 10 beat analysis - excluding accelerometry, temperature and QT interval features. Note that prevalence is the rate of occurrence in sample instances not the population.

Rhythm	FPR (α)	FNR (β)	Odds Ratio	Relative Risk	PPV	NPV	F_1 score	Youles Q	Pearsons ϕ	Prevalence
Normal sinus rhythm	0.040	0.115	185.278	88.951	0.523	0.994	0.657	0.989	0.661	0.047
Respiratory sinus arrhythmia	0.002	0.061	8111.333	338.931	0.958	0.997	0.948	1.000	0.946	0.044
Physiological sinus tachycardia	0.003	0.222	1279.833	384.650	0.700	0.998	0.737	0.998	0.736	0.008
Premature atrial complex(es)	0.029	0.058	535.060	29.393	0.947	0.968	0.944	0.996	0.914	0.358
Premature ventricular complex(es)	0.040	0.070	318.194	24.708	0.925	0.963	0.928	0.994	0.889	0.348
Sinus node dysfunction	0.000	0.103	-	361.000	1.000	0.997	0.945	-	0.946	0.026
First degree AV block	0.000	0.000	-	-	1.000	1.000	1.000	-	1.000	0.003
Second/third degree AV block	0.000	0.000	-	-	1.000	1.000	1.000	-	1.000	0.002
AV nodal/ junctional tachycardias	0.001	0.148	6215.750	259.948	0.958	0.996	0.902	1.000	0.901	0.024
AV reciprocating tachycardias	0.007	0.000	-	-	0.600	1.000	0.750	-	0.772	0.011
Inappropriate sinus tachycardia	0.000	0.667	-	554.000	1.000	0.998	0.500	-	0.577	0.003
Focal atrial tachycardia	0.000	0.375	-	368.000	1.000	0.997	0.769	-	0.789	0.007
Macro-reentrant atrial tachycardia	0.000	0.125	1284.875	224.283	0.826	0.996	0.826	0.998	0.935	0.007
Multifocal atrial tachycardia	0.002	0.235	1771.250	237.033	0.867	0.996	0.813	0.999	0.811	0.015
Atrial fibrillation	0.004	0.174	1284.875	224.283	0.826	0.996	0.826	0.998	0.822	0.021
Ventricular tachycardias	0.000	1.000	-	-	-	0.995	-	-	-	0.005
Atrial paced rhythm	0.003	0.396	538.281	51.370	0.906	0.982	0.725	0.996	0.731	0.043
Ventricular paced rhythm	0.000	0.267	-	135.875	1.000	0.993	0.846	-	0.853	0.027
ALL (mean)	**0.007**	**0.223**	**14.726**	**-**	**-**	**0.993**	**0.958**	**0.873**	**0.350**	**0.056**

FPR- false positive rate; FNR=false negative rate; α = Type I error; β = Type II error; PPV= positive predictive value; NPV=negative predictive value

Table 11.5 Major performance indices for 5 beat analysis - excluding accelerometry, temperature and QT interval features.

Rhythm	CCR	Error	±	CI_{95}	Se	±	CI_{95}	Sp	±	CI_{95}	κ	±	CI_{95}	*P*
Normal sinus rhythm	0.980	0.020	±	0.008	0.308	±	0.027	0.988	±	0.006	0.257	±	0.026	<0.001
Respiratory sinus arrhythmia	0.636	0.364	±	0.028	-	±	-	0.636	±	0.028	0.000	±	0.000	1.000
Physiological sinus tachycardia	0.652	0.348	±	0.028	0.000	±	0.000	0.657	±	0.028	-0.014	±	-	0.057
Premature atrial complex(es)	0.635	0.365	±	0.028	0.003	±	0.003	0.970	±	0.010	-0.036	±	-	<0.001
Premature ventricular complex(es)	0.662	0.338	±	0.028	0.000	±	0.000	0.996	±	0.004	-0.005	±	-	0.555
Sinus node dysfunction	0.978	0.022	±	0.009	0.000	±	0.000	0.999	±	0.002	-0.002	±	-	1.000
First degree AV block	0.971	0.029	±	0.010	0.000	±	0.000	0.974	±	0.009	-0.005	±	-	1.000
Second/third degree AV block	0.991	0.009	±	0.006	-	±	-	0.991	±	0.006	0.000	±	0.000	1.000
AV nodal/ junctional tachycardias	0.973	0.027	±	0.010	0.000	±	0.000	0.997	±	0.003	-0.005	±	-	1.000
AV reciprocating tachycardias	0.986	0.014	±	0.007	0.000	±	0.000	0.995	±	0.004	-0.006	±	-	1.000
Inappropriate sinus tachycardia	0.990	0.010	±	0.006	0.000	±	0.000	0.993	±	0.005	-0.004	±	-	1.000
Focal atrial tachycardia	0.995	0.005	±	0.004	0.000	±	0.000	1.000	±	0.000	0.000	±	0.000	1.000
Macro-reentrant atrial tachycardia	0.979	0.021	±	0.008	0.000	±	0.000	0.986	±	0.007	-0.009	±	-	1.000
Multifocal atrial tachycardia	0.991	0.009	±	0.006	0.000	±	0.000	0.997	±	0.003	-0.004	±	-	1.000
Atrial fibrillation	0.957	0.043	±	0.012	0.000	±	0.000	0.973	±	0.009	-0.021	±	-	1.000
Ventricular tachycardias	0.979*	0.021*	±	0.008	0.000	±	0.000	0.985	±	0.007	-0.008	±	-	1.000
Atrial paced rhythm	0.966	0.034	±	0.011	0.453	±	0.029	0.997	±	0.003	0.588	±	0.029	<0.001
Ventricular paced rhythm	0.978	0.022	±	0.009	0.478	±	0.029	1.000	±	0.000	0.637	±	0.028	<0.001
ALL	**0.928**	**0.072**	**±**	**0.015**	-			0.952			0.076			**<0.001**

CCR = correct classification rate; Se = sensitivity; Sp = specificity; κ = Cohen's kappa; CI_{95} = 95% confidence interval; *P* values to 3 significant figures

Table 11.6 Additional performance indices for 5 beat analysis - excluding accelerometry, temperature and QT interval features. Note that prevalence is the rate of occurrence in sample instances not the population.

Rhythm	FPR (α)	FNR (β)	Odds Ratio	Relative Risk	PPV	NPV	F_1 score	Youles Q	Pearsons ϕ	Prevalence
Normal sinus rhythm	0.008	0.765	37.026	25.941	0.308	0.988	0.267	0.947	0.259	0.015
Respiratory sinus arrhythmia	0.000	1.000	-	-	-	0.636	-	-	-	0.364
Physiological sinus tachycardia	0.011	1.000	0.000	0.000	0.000	0.657	-	-1.000	-0.061	0.341
Premature atrial complex(es)	0.353	0.957	0.083	0.086	0.003	0.970	0.005	-0.846	-0.093	0.021
Premature ventricular complex(es)	0.336	1.000	0.000	0.000	0.000	0.996	-	-1.000	-0.037	0.003
Sinus node dysfunction	0.021	1.000	0.000	0.000	0.000	0.999	-	-1.000	-0.004	0.001
First degree AV block	0.003	1.000	0.000	0.000	0.000	0.974	-	-1.000	-0.009	0.026
Second/third degree AV block	0.000	1.000	-	-	-	0.991	-	-	-	0.009
AV nodal/ junctional tachycardias	0.024	1.000	0.000	0.000	0.000	0.997	-	-1.000	-0.008	0.003
AV reciprocating tachycardias	0.009	1.000	0.000	0.000	0.000	0.995	-	-1.000	-0.006	0.005
Inappropriate sinus tachycardia	0.003	1.000	0.000	0.000	0.000	0.993	-	-1.000	-0.004	0.007
Focal atrial tachycardia	0.005	-	-	-	0.000	1.000	-	-	-	0.000
Macro-reentrant atrial tachycardia	0.007	1.000	0.000	0.000	0.000	0.973	-	-1.000	-0.010	0.014
Multifocal atrial tachycardia	0.006	1.000	0.000	0.000	0.000	0.997	-	-1.000	-0.004	0.003
Atrial fibrillation	0.018	1.000	0.000	0.000	0.000	0.973	-	-1.000	-0.022	0.026
Ventricular tachycardias	0.005	1.000	0.000	0.000	0.000	0.985	-	-1.000	-0.009	0.015
Atrial paced rhythm	0.032	0.094	287.790	157.839	0.453	0.997	0.604	0.993	0.627	0.029
Ventricular paced rhythm	0.022	0.000	-	-	0.478	1.000	0.647	-	0.684	0.020
ALL	**0.047**	**0.364**	**35.474**	**16.911**	**0.538**	**0.968**	**0.583**	**0.945**	**0.546**	**0.079**

FPR- false positive rate; FNR=false negative rate; α = Type I error; β = Type II error; PPV= positive predictive value; NPV=negative predictive value

Sensitivities were low for 5 beat analysis, beng unmeasurable for 14 rhythms, with the exception of normal sinus rhythm (NSR), having sensitivity of 0.885 for 10 beat analysis compared with 0.308 for 5 beat analysis; atrial paced rhythm with a sensitivity of 0.604 for 10 beat and 0.453 for 5 beat analysis and ventricular paced rhythm with a sensitivity of 0.733 for 10 beat and 0.478 for 5 beat analysis. Specificities were high (> 0.6) for all rhythms having both 10 and 5 beat analysis. Kappa (κ) indices were consistently low for 5 beat rhythm analysis, with the exceptions of normal sinus rhythm (NSR) with κ of 0.636 for 10 beat and 0.257 for 5 beat analysis; atrial paced rhythm with κ of 0.715 compared with 0.588 and ventricular paced rhythm with κ of 0.843 and 0.637.

Positive predictive value (PPV) was >0.5 for all rhythms having 10 beat analysis, with the exception of ventricular tachycardias, where no instances were diagnosed and <0.5 for all rhythms having 5 beat analysis.

Pearson's ϕ was >0.57 for all rhythms having 10 beat analysis and consistently low with negative values, for 5 beat analysis, except for normal sinus rhythm (NSR) with ϕ of 0.661 for 10 beat and 0.259 for 5 beat analysis, atrial paced rhythm, with ϕ of 0.731 and 0.627 respectively and ventricular paced rhythm with ϕ of 0.853 and 0.684 respectively. F_1 score and Youles Q values, where calculable, followed the pattern of ϕ scores.

None of the major indices, CCR, error, sensitivity, specificity, κ or P value indicated superior performance, beyond the 95% confidence interval for 5 beat analysis. Likewise none of the additional scores showed improvement in performance for 5 beat analysis. With the exception of normal sinus rhythm (NSR), atrial and ventricular paced rhythms, 10 beat analysis was convincingly superior. These findings did not correspond with those for the production system, where 5 beat based rhythm classification was applied to the diagnosis of six rhythms: PAC; PVC; AVRT; AF; SND and MAT. The explanation for this difference is likely to be the different approaches taken between iterative development (see Chapter 10), where small differences in performance indices by rhythm and different classifier units drove decisions for their inclusion in the production system, compared with this structured test of 10 beat versus 5 beat rhythm analysis.

Though guidelines used as the basis for feature selection (see Chapter 3, subsection 3.8.12) supported the use of a smaller number of heart beats (5 beats) in an approach to rhythm diagnosis, use of 10 beats produced comparable or higher performances, over a wide range of indices, for all rhythms.

In this assessment, performances for 5 beat analysis, other than for normal sinus rhythm and paced rhythms, were found to be poor across all examined performance indices.

Performances for 10 beat analysis for normal sinus rhythm, premature atrial complexes, premature ventricular complexes, and second and third degree heart block were superior to that achieved in the production system, suggesting the potential for use in a future upgrade. Performance indices for ventricular tachycardia were poor and for the remaining rhythms showed little difference (see Chapter 10, Tables 10.17 and 10.18; Chapter 11, Tables 11.3, 11.4, 11.5 and 11.6).

Favourable performance of 10 beat analysis for second and third degree heart block was notable. The individual decision tree, naïve Bayes, neural network and support vector machine classifier units had failed to adequately classify this rhythm during the iterative development phase and the production system successfully used the "catch-all" inference engine classifier but with poor performance.

These results suggest that further optimisation of the production system would be possible, using 10 beat analysis to improve performance for under-performance with normal sinus rhythm and second and third degree AV block.

11.3 Comparisons with Other Studies

Classifier performance of the production system were compared with the studies reviewed in Chapter 2, subsections 2.8.10, 2.8.11 and 2.9.3, which examined single classifier, hybrid and multi-classifier systems and ICD algorithms used for cardiac rhythm diagnosis.

11.3.1 Review Papers

The results from this study support the review of Aliot *et al*. (2004) advising cautious interpretation of small or focused studies however found that use of a wide range of indices better reflects classification performance, rather than their prescribed emphasis on specificity and PPV.

11.3.2 Statistical Classifiers

Zhang *et al.* (1999) examined the performance of their classifier, based on a chaotic complexity measure, for diagnosis of VT and VF from surface ECG recordings, designed for application in AED's, using VT and VF instances form patients undergoing ICD implant and instances of sinus rhythm from the MIT database. They recorded a direct ventricular electrogram though appear not to have used it for analysis. They achieved 100% sensitivity, specificity and accuracy, though there was no comparison between rhythms for individual patients, such as sinus rhythm against an unknown rhythm within the same patient. Our results compare favourably for VT diagnosis only, as no instances of VF were detected.

11.3.3 Syntactic Classification

No studies of syntactic classifiers were available that provided useful statistical measures for comparison.

11.3.4 Neural Network Classifiers

Yang *et al.* (1994) compared deterministic logic with an artificial neural network for differentiating both sinus rhythm with PAC's or PVC's from AF. They found that the artificial neural network achieved sensitivity of AF diagnosis to 92%, with specificity 92.3%, compared with a lower sensitivity of 0.870 (87%) and a higher specificity of 0.991 (99.1%) for atrial fibrillation in our study. The difference emphasises the balance of sensitivity and specificity where there is similar overall performance, one may improve at the cost of the other.

Coggins *et al.* (1995) tested a low power analogue VLSI neural network chip with a 10:6:3 multilayer perceptron, an input bucket brigade device (BBD) and a winner take all output. QRS morphology classifier aimed at application in ICD's. The study showed discrimination for sinus tachycardia with ventricular tachycardia and quote correct classification (*CCR*) of between 77.6% and 100% for ST and 98.3 and 100 for VT, similar results to our study, with *CCR* of 0.998 (99.8%) for physiological sinus tachycardia and 1.000 (100%) for VT.

Minami *et al.* (1999) analysed spectral components of ECG or RV electrogram QRS complexes to discriminate rhythms using a back-propagation neural network. Classification was into three rhythm classes: supraventricular rhythm (including sinus rhythm and supraventricular tachycardias), ventricular rhythm (including VT and PVC) and VF. They provided sensitivity and specificities for all three rhythms using both ECG and electrogram data. Using the RV electrogram, for "supraventricular rhythm" sensitivity was 1.00 and specificity 0.98; for "ventricular rhythm" 0.68 and 1.00 and for VF 0.99 and 0.86 respectively. Given the differences in approach to defining rhythm classes, other than VF, for which we obtained no data, direct comparison was not possible.

Kara & Okandan (2007) developed a LM back propagation neural network to diagnose AF and sinus rhythm from the ECG. Performance for AF was *CCR* 100% with 100% sensitivity, specificity and PPV, compared with *CCR* of 0.988 (98.8%), sensitivity of 0.870 (87%) and specificity of 0.991 (99.1%) and PPV of 0.667 (66.7%) for atrial fibrillation in our study. The simple binary classification in Kara & Okandan's study was less challenging than the multi-class classification in our study, where subtle intermediate diagnoses such a macro-re-entrant atrial tachycardia and multifocal atrial tachycardia were differentiated from atrial fibrillation, perhaps partly explaining the slightly lower performance indices.

Christov & Bortolan (2004) classified PVC's using ECG and vectorcardiogram morphology features and a neural network classifier. The sensitivity of 99.7% and specificity of 98.5% they achieved for PVC classification was superior to our study, with a sensitivity of 0.492 (49.2%) and specificity of 0.974 (97.4%), perhaps due to use of a specialised binary classifier.

Acharya *et al.* (2008) analysed 15 minute segments of ECG using fast-Fourier analysis, input features to a neural network which classified nine conditions, some not related to rhythm, such as congestive heart failure, ischemic/dilated cardiomyopathy and left bundle branch block, and certain rhythms, such as normal sinus rhythm, AF, VF, PVC, complete heart block and sick sinus syndrome. They achieved a sensitivity of 81.72%, specificity of 100% and PPV of 100% for all classes combined. Given their analysis was of long (15 minute) segments of ECG, they did not quote class-specific statistics and their classes included conditions unrelated to cardiac rhythms, so a direct comparison was not performed.

11.3.5 Fuzzy Classifiers

Usher *et al.* (1999) studied classification of AF, VF, SVT and VT using intracardiac electrograms, and a fuzzy inference system. The system had low computational resource requirement but did not offer statistical indices. Anuradha *et al.* (2008) used chaotic features input to a fuzzy system to classify ECG rhythms into 8 classes: left bundle branch block; normal sinus rhythm; PVC; AF; VF; complete heart block; ischemic dilated cardiomyopathy and sick sinus syndrome. Statistical analysis of their results was limited to *CCR* and results for 4 rhythms were amenable to comparison with our study. Of the very few studies examining fuzzy inference for rhythm diagnosis, this study alone showed demonstrable value. Normal sinus rhythm showed a higher *CCR* of 96.77% compared with 0.754 (75.4%) in our study; PVC also showed a higher *CCR* of 93.85% compared with 0.806 (80.6%) in our study. Conversely, AF had a lower *CCR* of 90% compared with 0.988 (98.8%) in our study; complete heart block had a *CCR* of 90% compared with 0.989 (98.9%) in our study and sick sinus syndrome had a *CCR* of 88.9% compared with 0.996 (99.6%) in our study.

11.3.6 Decision Tree Classifiers

Tsipouras *et al.* (2005) designed a knowledge-based decision tree method to classify rhythm into four beat classes using only the RR interval of the ECG: normal, premature ventricular contractions, ventricular flutter/fibrillation and heart block. Results for 3 rhythms were amenable to comparison with our study. Sensitivity was 97.05%, specificity 50.03% and PPV 50.16% for PVC's, compared with similar results of sensitivity 0.492 (49.2%), specificity 0.974 (97.4%) and PPV of 0.909 (90.9%) for our study; ventricular tachycardia had sensitivity 61.33%, specificity 95.45% and PPV of 30.46% compared with higher values of sensitivity of 1.00 (100%), specificity of 1.00 (100%) and PPV of 1.00 (100%) for our study; and for heart block sensitivity

of 100%, specificity 99.96% and PPV of 83.33%, compared with slightly poorer results of sensitivity 1.000 (100%), specificity 0.989 (98.9%) and PPV of 0.143 (14.3%) for our study.

Rodriguez *et al.* (2005) algorithm called MOLEC that classifies ECG beats and rhythms. Tested a variety of different classifiers and found that the decision tree performed best. Optimised using various options and used "info-gain" to rank features. *CCR* was the sole statistical performance index quoted and was amenable to comparison for two rhythms: *CCR* was 97.95%, compared with 1.00 (100%) for VT in our study and *CCR* of 67.35%, compared with 0.754 (75.4%) for normal sinus rhythm in our study, favouring performances from our study.

11.3.7 Support Vector Machine Classifiers

Polat & Güneş (2007) selected ECG features using principal component analysis (PCA) and used a support vector machine to classify rhythm as either normal or arrhythmia. This classification scheme was of insufficient granularity for comparison with our study.

Asl *et al.* (2008) classified rhythms using the RR interval of the ECG, based on generalized discriminant analysis (GDA) feature reduction and a support vector machine classifier. 6 rhythms were classified of which 5 were amenable to comparison with our study: normal sinus rhythm with a *CCR* of 98.94%, higher than the 0.754 (75.4%) in our study; premature ventricular contraction with a *CCR* of 98.96%, higher than 0.806 (80.6%) in our study; atrial fibrillation with a *CCR* of 98.53%, similar to the 0.988 (98.8%) in our study; sick sinus syndrome with a *CCR* of 98.51%, had a similar value to 0.996 (99.6%) in our study and heart block with *CCR* of 100%, was similar to the 0.989 (98.9%) in our study.

11.3.8 k-Nearest Neighbour Classifiers

Owis *et al.* (2002) used nonlinear dynamics to model the chaotic nature of ECG signals. 5 classes of rhythm: normal rhythm and ventricular couplet, VT, ventricular bigeminy, and VF were discriminated comparing several different classifier technologies. They found *k*-NN results generally indicate the highest detection rate among the three classifiers at the price of lowest specificity. Only values for VT were directly amenable to comparison with the results of our study: specificity (all rhythms) was 81.25%, compared to 0.979 (97.9%) in our study and sensitivity for VT was 6.25% and compared very unfavourably with our study, having sensitivity of 1.0 (100%).

Minhas & Arif (2008) used 11 wavelet transform and RR-interval features for ECG classification using a *k*-nearest neighbour classifier, into six beat classes: PVC, PAC; APB), left and right bundle branch block beats, paced beats and normal beats. For normal beats was *CCR* 99.87%,

compared with 0.754 (75.4%) in our study; for PAC's *CCR* was 99.02%, compared with 0.803 (80.3%) in our study and *CCR* of 98.85% for PVC's, compared with 0.806 (80.6%) in our study.

11.3.9 Other Classifier Technologies

Srakar *et al.* (2008) described an AF and AT burden estimator looking at 24 hour ECG samples, which was not directly comparable with this study.

Brüser *et al.* (2012) designed a novel system for detection of AF using cardiac vibration signals recorded by bed mounted sensors, intended as tool for home-healthcare. They ranked seven classifiers from naïve Bayes, linear and quadratic discriminant analysis, support vector machines, random forests, bagged and boosted trees, for which random forests performed best. Classification indices were: for AF sensitivity of 0.938, higher than the 0.870 from our study; specificity of 0.978 was similar to 0.991 from our study; PPV of 0.934 was higher than 0.667 in our study and for normal rhythm sensitivity was 0.897, similar to 0.865 for our study, specificity 0.980, higher than the 0.748 of our study and PPV of 0.792, higher than the 0.145 in our study.

11.3.10 Hybrid classifiers

Wang *et al.* (2001) used ECG short-time multifractality features, to classify VT, VF, and AF rhythms using a fuzzy Kohonen network. For AF, sensitivity was 96.7%, specificity 98.3% and *CCR* 97.8%, compared with lower sensitivity 0.870 (87.0%), similar specificity 0.991 (99.1%) and similar *CCR* of 0.988 (98.8%) and for VT, sensitivity 96.7%, specificity 97.5% and *CCR* 97.2% , compared with sensitivity of 1.00 (100%), specificity of 1.00 (100%) and *CCR* 1.00 (100%).

Linh *et al.* (2003) used Hermite ECG features, with a neuro-fuzzy beat classifier for six beat types: normal beats, PVC, left bundle branch block beats, right bundle branch block beats, PAC, ventricular flutter and ventricular escape beats. They quoted "misclassification" (error), from which *CCR* was able to be deduced. Normal beats showed 1.6% error, a *CCR* of 98.4%, compared with *CCR* of 0.754 (75.4%) in our study; PAC beats showed 9.09% error, a *CCR* of 90.91%, compared with *CCR* of 0.803 (80.3%) in our study and PVC beats showed a 3.6%error, a *CCR* of 96.4%, compared with *CCR* of 0.806 (80.6%) in our study;

Polat *et al.* (2006) classified ECG arrhythmia data using input values obtained from fuzzy weighted pre-processing into 16 classes using an artificial immune recognition classifier system. *CCR* of up to 80.77% was obtained, however detailed class-specific statistics were not provided.

Exarchos *et al.* (2007) studied ischaemic and arrhythmic beat ECG classification using a fuzzy expert system. Rhythm classes were VF, PVC, normal and heart block. Comparable rhythm

classification results were: PVC sensitivity of 92.4%, was higher compared with 0.492 (49.2%) with our study, specificity of 97.6%, compared with 0.782 (78.2%) in our study and *CCR* of 95.8%, compared with 0.806 (80.6%) in our study; for normal beats, sensitivity was 93.6%, compared with 0.865(86.5%) in our study, specificity 97.7%, compared with 0.748 (74.8%) in our study and *CCR* of 95.8%, compared with 0.754 (75.4%); for heart block sensitivity was 98.3 %, compared with 1 (100%), specificity 99.9% compared with 0.989 (98.9%) and *CCR* of 95.8%, compared with 0.989 (98.9%).

11.3.11 Multi-Classifier Systems (MCS)

Leong & Jabri (1992) used atrial and ventricular intracardiac electrogram timing together with a morphology feature in a decision tree and neural network hybrid classifier, with decision rules to allocate final class. Atrial and ventricular intervals were classified using a decision tree and a neural network based morphology classifier only for certain cases, such as ventricular tachycardia with 1:1 conduction. With a view to application in ICD's, four rhythms were classified: NSR, SVT, VT, and VF were included but only an overall *CCR* of 99.6% was quoted, inadequate for useful comparison.

de Chazal *et al.* (2004) classified 2 channels of ECG using morphology and interval features. Each ECG channels had a dedicated linear discriminant classifier to obtain posterior probability estimates and a combiner allocated class to that having the highest posterior probability estimate. They detailed performance statistics for 4 classes, for which two were amenable to comparison with our study: For PAC's, sensitivity was similar at 53.3%, compared with 0.597 (59.7%) for our study; For PVC's, sensitivity was slightly higher at 67.3%, compared with 0.492 (49.2%) for our study. de Chazal & Reilly (2006) refined this classifier. Performance statistics for 2 classes were: for PAC's, *CCR* was higher at 95.9% compared with 0.803 (80.3%) for our study; sensitivity was higher at 87.7%, compared with 0.597 (59.7%) for our study; PPV was lower at 47%, compared with 0.803 (80.37%) for our study and FPR was lower at 3.8%, compared with 0.081 (8.1%) for our study. For PVC's, *CCR* was higher at 99.4% compared with 0.806 (80.6%) for our study; sensitivity was higher at 94.3%, compared with 0.492 (49.2%) for our study; PPV was higher at 96.2%, compared with 0.909 (90.9%) for our study and FPR was lower at 0.3%, compared with 0.026 (2.6%) for our study.

Ceylan *et al.* (2009) examined ECG rhythm classification using a combined Fuzzy Clustering Neural Network Algorithm. Ten rhythms were classified: normal beat, sinus bradycardia, ventricular tachycardia, sinus arrhythmia, atrial premature contraction, paced beat, right bundle branch block, left bundle branch block, atrial fibrillation and atrial flutter. Classification accuracy was 99% for all types, with results not broken down by rhythm type.

Osowski *et al.* (2004) classified ECG rhythms using features from Hermite characterization of and higher order statistics, into 13 heart rhythm classes. They combined two neural classifiers, using a weighted voting scheme into an expert system. Results only quoted "relative" classification errors and were not directly comparable to this study. Ozbay *et al.* (2006) classified ECG rhythms using two neural networks, the first distinguished arrhythmias from normal sinus rhythm and the second to classify rhythm into 10 classes. Accuracy (*CCR*) rates were quoted but not by rhythm.

11.3.12 Comparative Studies of Classifiers

Jovic & Bogunovic (2011) evaluated the potential usefulness of ECG rhythm classification using features based on heart rate variability (HRV). Although they analysed a number of different classifier units using this, there was no rationale behind the choice of features, other than an exploration of alternative features in this application. They succeeded in classifying rhythm into three broad categories: normal, arrhythmia and supraventricular arrhythmia. HRV is believed to reflect the autonomic nervous system and its relationship with heart rhythm and has been found to be predictive of ventricular arrhythmias, in the context of prevention of sudden cardiac death (SCD) (see Chapter 5, subsection 5.3.2) (Malik *et al*. 1989; Osterhues *et al.* 1993; Huikuri *et al.* 1995; Copie *et al.* 1996; Bernardi 1996; Nolan *et al.* 1998) buts its value in rhythm diagnosis is not established. Acharya *et al.* 2004) succeeded in using HRV derived ECG features to diagnose 5 different rhythms, as well as left bundle branch block and dilated cardiomyopathy, using an artificial neural network and a fuzzy classifier. Neither of these two studies examined the more orthodox features associated with rhythm diagnosis – ECG intervals and QRS morphology. Given the poor comparability of these two studies with the more rational method developed in this work, direct comparisons were considered inappropriate.

Acharya *et al.* (2004) compared an artificial neural network and a fuzzy classifier for 8 classes: left bundle branch block, normal sinus rhythm, PVC, AF, VF, complete heart block, ischaemic/ dilated cardiomyopathy and sick sinus syndrome. For the better performing fuzzy classifier, *CCR* indices for 5 classes were comparable with our study: for normal beats *CCR* was 92.5% compared with 0.754 (75.4%) in our study; for PVC 90.0%, compared with 0.806 (80.6%) in our study; for AF 88.0%, compared with 0.988 (99.8%) in our study; for complete heart block 88.0%, compared with 0.989 (98.9)% for our study and for SSS 90.9% compared to 0.996 (99.6%) in our study.

11.3.13 Single ICD Algorithms

Kuhlkamp *et al.* (1999) compared dual chamber to single chamber ICD algorithms from a single manufacturer. The study focussed on therapy decisions and did not dwell on diagnostic performance, offering no statistical indices suitable for comparison. Swerdlow *et al.* (2000) studied diagnosis of short episodes of AF with prolonged episodes that require cardioversion.

Their detection measures do not correspond well with indices from our study, preventing useful comparison.

Wilkoff *et al.* (2001) examined performance of a dual-chamber ICD detection algorithm. Of indices suitable for comparison with our study, sensitivity for VT was 100.0% and PPV 78.1%, compared with 1.00 (100%) and 1.00 (100%) respectively for our study.

Swerdlow *et al.* (2002) investigated a downloadable algorithm for ICD's to discriminate supraventricular tachycardia from VT, based on morphology differences of ventricular electrograms, using corresponding coefficients of wavelet transforms expressed as a match-percent score. Sensitivity for VT detection was 100%, specificity was 78%, compared with 1.000 (100%) for both indices in our study.

Kouakam *et al.* (2004) assessed a dual-chamber detection ICD algorithm for performance with VT. Using their contingency table values indices had slightly lower values than those in our study, with *CCR* of 0.950 sensitivity 0.983 specificity 0.900 and kappa 0.895, compared with all indices having values of 1.00 in our study.

Kremers *et al.* (2012) used electrogram-based detection with a right ventricular pressure sensor to assess haemodynamic stability during tachycardia in an investigational ICD. VT/VF 100% specificity was 100%, similar to that achieved in our study.

11.3.14 Comparative Studies of ICD Algorithms

Hintringer *et al.* (2001) compared four ICD algorithms using electrograms recorded during EP studies. Four dual chamber ICD algorithms were tested: Phylax AV, Defender IV, Ventak AV II DR, and Gem DR 7271 using 86 arrhythmias recorded from patients undergoing invasive EP studies. Episodes included 7 different rhythm diagnoses: atrial fibrillation, atrial flutter, atrial tachycardia, AV nodal re-entrant tachycardia, AV re-entrant tachycardia, sinus tachycardia, and ventricular tachycardia. Unfortunately, the published analysis quotes supraventricular versus ventricular arrhythmia diagnostic indices in an OVO analysis and did not include raw data allowing re-analysis, precluding rhythm specific comparison other than for VT. The indices for VT for the best performing ICD algorithm, the ELA Defender IV, were sensitivity of 100% and specificity 28%, compared with these having values of 1.00 (100%) in our study.

Gold *et al.* (2002) proposed an electrogram vector timing and correlation (VTC) morphology algorithm for ICD's and assessed performance by comparison with physician diagnosis. Discrimination was between ventricular and supraventricular arrhythmias. Electrograms were collected at ICD placement and tested in a software model. For the dual chamber ICD they found

a sensitivity of 100% and specificity 97% for VT , which compares with sensitivity and specificity of 1.000 (100%) found in our study.

Theuns *et al.* (2004) compared single and dual- chamber ICD algorithms and found no significant difference in tachyarrhythmia detection. Unfortunately, data presented was largely pooled, as the emphasis of their study was on low rates of inappropriate treatment rather than diagnosis and the only comparable result was the sensitivity for VT/VF detection of 100%, the same as in our study.

Hintringer *et al.* (2004) compared specificities of dual chamber ICD algorithms in a bench study which used tachyarrhythmia's recorded during EP studies, processed through ICD devices. For VT specificity was 90% (Biotronik), 89% (ELA Medical), 89% (Guidant), 68% (Medtronic), and 76% (St. Jude Medical), compared with sensitivity and specificity of 1.000 (100%) found in our study.

Gold *et al.* (2012a) directly compared Medtronic and VITALITY 2 ICD algorithms for prevention of inappropriate therapy. PPV's were 41.2% for Guidant and 51.3% for Medtronic ICD's, compared with a PPV of 1.000 (100%) for VT in our study.

Gold *et al.* (2012b) assessed the performance of a subcutaenous ICD algorithm, compared with established ICD algorithms. For VT the sensitivity was 100.0% and specificity 98.0%, similar to the results from our study of 1.00 (100%) and 1.00 (100%) respectively.

11.3.15 Summary of Comparisons with Other Studies

Performance indices for the studies examined in chapter 11, subsections 11.3.1 to 11.3.14 were tabulated by rhythm (see Tables 11.7 to 11.14).

Studies had comparable results for 8 of the 18 rhythms studied: normal sinus rhythm; physiological sinus tachycardia; premature atrial complex(es); Premature ventricular complex(es); Sinus node dysfunction; Second/third degree AV block; Atrial fibrillation and ventricular tachycardias. No studies provided comparative results for 10 rhythms: respiratory sinus arrhythmia; first degree AV block; AV nodal/ junctional tachycardias; AV reciprocating tachycardias; inappropriate sinus tachycardia; focal atrial tachycardia; macro-reentrant atrial tachycardia; multifocal atrial tachycardia; atrial paced rhythm and ventricular paced rhythm.

The production classifier in our study clearly poorly for normal sinus rhythm (see Table 11.7) when compared to published studies evaluated, having inferior performance for all indices with the exception of Rodriguez *et al.* (2005), whose decision tree algorithm had inferior performance to our results. Notably 3 of the studies attempted classification of 5 different rhythms. Lack of

multiple indices in some studies and lack of clearly superior results means that it is difficult to determine the best performing classifier. Of the available data, the best performing was the *k*-nearest neighbour classifier of Minhas & Arif (2008), with a *CCR* of 99.87%.

Table 11.7 A comparison of classifier performances for normal sinus rhythm

Study	n	Classifier	Rhythm	*CCR*	Se	Sp	PPV	*κ*
This study	18	MCS	NSR	75.4	86.5	74.8	14.5	18.2
Anuradha *et al.* (2008)	5	Fuzzy	NSR	96.77				
Asl *et al.* (2008)	5	SVM	**NSR**	98.94				
Rodriguez *et al.* (2005)	2	DT	NSR	67.35				
Minhas & Arif (2008)	3	k-NN	NSR	99.87				
Brüser *et al.* (2012)	2	RF	NSR		89.7	98.0	79.2	
Linh *et al.* (2003)	3	Fuzzy-NN	NSR	98.4				
Exarchos *et al.* (2007)	3	Fuzzy-Expert	NSR	95.8	93.6	97.7		
Acharya *et al.* (2004)	5	Fuzzy	NSR	92.5				

Key: MCS = multi-classifier system; RF = random forest; n = number of rhythms classified; All values %

The sole study we found which had directly comparable results for sinus tachycardia (see Table 11.8) was a neural network of Coggins *et al.* (1995), showed that its *CCR* of 100% closely matched that of our results.

Table 11.8 A comparison of classifier performances for physiological sinus tachycardia

Study	n	Classifier	Rhythm	*CCR*	Se	Sp	PPV	*κ*
This study	18	MCS	ST	99.8	100	99.8	81.8	89.9
Coggins *et al.* (1995)	2	NN	ST	100				

Key: MCS = multi-classifier system; n = number of rhythms classified; All values %

For PAC diagnosis (see Table 11.9), the results of this study compare unfavourably to published results, with the best performing classifier appearing to be the *k*-nearest neighbour classifier of Minhas & Arif (2008), which showed a *CCR* of 99.02%, compared to our result of 80.3%.

Table 11.9 A comparison of classifier performances for premature atrial complexes

Study	n	Classifier	Rhythm	*CCR*	Se	Sp	PPV	*κ*
This study	18	MCS	PAC	80.3	59.7	91.9	80.3	54.7
Minhas & Arif (2008)	3	k-NN	PAC	99.02				
Linh *et al.* (2003)	3	Fuzzy-NN	PAC	90.91				
de Chazal *et al.* (2004)	2	Linear Hybrid	PAC		53.3			
de Chazal & Reilly (2006)	2	Linear Hybrid	PAC	95.9	87.7	47		

Key: MCS = multi-classifier system; n = number of rhythms classified; All values %

Performance indices for PVC diagnosis (see Table 11.10) with the production system fared poorly when compared with published studies with the exception only of the hybrid linear classifier of de Chazal et al. (2004). Best performing classifier for this rhythm was probably the neural network classifier of Christov & Bortolan (2004), with a sensitivity of 99.7% and specificity of 98.5%.

Table 11.10 A comparison of classifier performances for premature ventricular complexes

Study	n	Classifier	Rhythm	*CCR*	Se	Sp	PPV	*κ*
This study	18	MCS	PVC	80.6	49.2	97.4	90.9	52.2
Christov & Bortolan (2004)	1	NN	PVC		99.7	98.5		
Anuradha *et al.* (2008)	5	Fuzzy	PVC	93.85				
Tsipouras *et al.* (2005)	3	DT	PVC		97.05	50.03	50.16	
Asl *et al.* (2008)	5	SVM	PVC	98.96				
Minhas & Arif (2008)	3	k-NN	PVC	98.85				
Linh *et al.* (2003)	3	Fuzzy-NN	PVC	96.4				
Exarchos *et al.* (2007)	3	Fuzzy-Expert	PVC	95.8	92.4	97.6		
de Chazal *et al.* (2004)	2	Linear Hybrid	PVC		67.3			
de Chazal & Reilly (2006)	2	Linear Hybrid	PVC	99.4	94.3	96.2		
Acharya *et al.* (2004)	5	Fuzzy	PVC	90				

Key: MCS = multi-classifier system; n = number of rhythms classified; All values %

For diagnosis of sinus node dysfunction (see Table 11.11), our study out-performed comparable studies for *CCR*, which was the only available index for comparison. Our study showed a *CCR* of 99.6%; a sensitivity of 93.1%; specificity of 99.8%, PPV of 93.1% and kappa of 92.9%.

Table 11.11 A comparison of classifier performances for sinus node dysfunction

Study	n	Classifier	Rhythm	*CCR*	Se	Sp	PPV	κ
This study	18	MCS	SND	99.6	93.1	99.8	93.1	92.9
Anuradha *et al.* (2008)	5	Fuzzy	SND	88.9				
Asl *et al.* (2008)	5	SVM	SND	98.51				
Acharya *et al.* (2004)	5	Fuzzy	SND	90.9				

Key: MCS = multi-classifier system; n = number of rhythms classified; All values %

For second and third degree heart block (see Table 11.12), our classifier had similar performace with the best published algorithm, the support vector machine of Asl *et al.* (2008), which showed a *CCR* of 100%.

Table 11.12 A comparison of classifier performances for second and third degree AV block

Study	n	Classifier	Rhythm	*CCR*	Se	Sp	PPV	κ
This study	18	MCS	2HB	98.9	100	98.9	13.3	24.8
Anuradha *et al.* (2008)	5	Fuzzy	2HB	90				
Tsipouras *et al.* (2005)	3	DT	2HB		100	99.96	83.33	
Asl *et al.* (2008)	5	SVM	2HB	100				
Exarchos *et al.* (2007)	3	Fuzzy-Expert	2HB	95.8	98.3	99.9		
Acharya *et al.* (2004)	5	Fuzzy	2HB	88				

Key: MCS = multi-classifier system; n = number of rhythms classified; All values %

Our study showed fair performance for atrial fibrillation (see Table 11.13) but was slightly out-performed by one study, the binary neural network classifier of Kara & Okandan (2007), which had *CCR*, sensitivity, specificity and PPV of 100%.

Table 11.13 A comparison of classifier performances for atrial fibrillation

Study	n	Classifier	Rhythm	*CCR*	Se	Sp	PPV	κ
This study	18	MCS	AF	98.8	87	99.1	66.7	74.9
Yang *et al.* (1994)	1	NN	AF		92	92.3		
Kara & Okandan (2007)	1	NN	AF	100	100	100	100	
Anuradha *et al.* (2008)	5	Fuzzy	AF	90				
Asl *et al.* (2008)	5	SVM	AF	98.53				
Brüser *et al.* (2012)	2	RF	AF		93.8	97.8	93.4	
Wang *et al.* (2001)	2	Fuzzy-NN	AF	97.8	96.7	98.3		
Acharya *et al.* (2004)	5	Fuzzy	AF	88				

Key: MCS = multi-classifier system; n = number of rhythms classified; All values %

This study performed well for VT diagnosis (see Table 11.14), with perfect performance on all indices, based on a very small number of training examples. Our perfect results were comparable in performance with the statistical classifier of Zhang *et al*. (1999), the neural network classifier of Coggins *et al*. (1995) and ICD algorithms of Kremers et al. (2012) and Theuns *et al*. (2004).

Table 11.14 A comparison of classifier performances for ventricular tachycardias

Study	n	Classifier	Rhythm	*CCR*	Se	Sp	PPV	κ
This study	**18**	MCS	VT	100	100	100	100	100
Zhang *et al.* (1999)	1	Statistical	VT	100	100	100		
Coggins *et al.* (1995)	2	NN	VT	100				
Tsipouras *et al.* (2005)	3	DT	VT		61.33	95.45	30.46	
Rodriguez *et al.* (2005)	2	DT	VT	97.95				
Owis *et al.* (2002)	1	k-NN	VT		6.25	81.25		
Wang *et al.* (2001)	2	Fuzzy-NN	VT	97.2	96.7	97.5		
Wilkoff et al. (2001)	1	ICD	VT		100.0		78.1	
Swerdlow *et al.* (2002)	1	JCD	VT		100	78		
Kouakam *et al.* (2004)	1	ICD	VT	95	98.3	90		89.5
Kremers et al. (2012)	1	ICD	VT			100		
Hintringer *et al.* (2001)	1	ICD	VT		100	28		
Gold *et al.* (2002)	1	ICD	VT		100	97		
Theuns *et al.* (2004)	1	ICD	VT		100			
Hintringer *et al.* (2004)	1	ICD	VT		100	90		
Gold *et al.* (2012a)	1	ICD	VT		100	98.0		

Key: MCS = multi-classifier system; n = number of rhythms classified; All values %

In summary, the production system diagnostic performance of this study was the best performing of all comparable algorithms for 4 of the 8 different rhythms for which comparison data were available: physiological sinus tachycardia, sinus node dysfunction and ventricular tachycardia (see Table 11.15).

Table 11.15 Best performing classifier by rhythm, from comparable studies

Rhythm	Classifier Technology	Study
Normal sinus rhythm	*k*-nearest neighbour classifier	Minhas & Arif (2008)
Physiological sinus tachycardia	Multi-Classifier System	**This study**
Physiological sinus tachycardia	Neural network classifier	Coggins *et al.* (1995)
Premature atrial contraction	*k*-nearest neighbour classifier	Minhas & Arif (2008)
Premature ventricular contraction	Neural network classifier	Christov & Bortolan (2004)
Sinus node dysfunction	Multi-Classifier System	**This study**
Second/ third degree heart block	Multi-Classifier System	**This study**
Second/ third degree heart block	Support vector machine	Asl *et al.* (2008)
Atrial fibrillation	Neural network classifier	Kara & Okandan (2007)
Ventricular tachycardia	Multi-Classifier System	**This study**
Ventricular tachycardia	Statistical classifier	Zhang *et al.* (1999)
Ventricular tachycardia	Neural network classifier	Coggins *et al.* (1995)
Ventricular tachycardia	ICD algorithm	Kremers *et al.* (2012)
Ventricular tachycardia	ICD algorithm	Theuns *et al.* (2004)

Notably, decision trees and naïve Bayes classifiers did not feature highly in published studies and likely is the explanation for their absence from this analysis. Neural network classifiers feature as best or tied for best performer for 4 rhythms, *k*-nearest neighbour for 2 rhythms, a statistical classifier, support vector machine and ICD algorithms for only one rhythm, suggesting that either neural network experimentation is very widespread or that they truly offer high classification performances for a range of rhythms.

The availability of comparable performance indices for only 8 rhythms from published studies suggested that the remaining 10 rhythms were, to our knowledge, classified only in our study. Published studies were found that did classify a number of these rhythms, however results were not of sufficient quality to allow direct comparison. These rhythms were: respiratory sinus arrhythmia; first degree AV block; AV nodal/ junctional tachycardias; AV reciprocating tachycardias; inappropriate sinus tachycardia; focal atrial tachycardia; macro-re-entrant atrial

tachycardia; multifocal atrial tachycardia; atrial paced rhythm and ventricular paced rhythm. Performance indices for these rhythms were taken from our study as "best performances".

Table 11.16 illustrates the best achieved performance indices for the 18 rhythms diagnosed in this study from published studies and from the results of this study. For all the studies, the best achievable *CCR* values are in excess of > 98% and specificity > 99% for all classified rhythms. More variability is apparent for sensitivity with 4 rhythms with low (< 90%) sensitivities achieved for the following rhythms: physiological sinus tachycardia with 81.8%; macro-re-entrant atrial tachycardia with 88.9%; multifocal atrial tachycardia with 52% and atrial paced rhythm with 81.5%, all values from this study. Values of PPV of < 90 were found for 6 rhythms, all provided by this study, and the value of any interpretation is questionable.

Table 11.16 Best published achieved performance for each rhythm, by study.

Rhythm	Study	*CCR*	Se	Sp	PPV
Normal sinus rhythm	Minhas & Arif (2008)	99.87			
Respiratory sinus arrhythmia	**This study**	99.0	95.2	99.2	81.6
Physiological sinus tachycardia	**This study**	99.8	81.8	100	100
	Coggins *et al.* (1995)	100			
Premature atrial complex(es)	Minhas & Arif (2008)	99.02			
Premature ventricular complex(es)	Christov & Bortolan (2004)		99.7	98.5	
Sinus node dysfunction	**This study**	99.6	93.1	99.8	93.1
First degree AV block	**This study**	99.9	100	99.9	66.7
Second/third degree AV block	**This study**	98.9	100	98.9	14.3
	Asl *et al.* (2008)	100			
AV nodal/ junctional tachycardias	**This study**	99.6	92.6	99.8	92.6
AV reciprocating tachycardias	**This study**	99.7	100	99.7	75
Inappropriate sinus tachycardia	**This study**	100	100	100	100
Focal atrial tachycardia	**This study**	100	100	100	100
Macro-reentrant atrial tachycardia	**This study**	99.9	88.9	100	87
Multifocal atrial tachycardia	**This study**	98.6	52	99.6	76.5
Atrial fibrillation	Kara & Okandan (2007)	100	100	100	
Ventricular tachycardias	**This study**	100	100	100	100
	Zhang *et al.* (1999)	100	100	100	
	Coggins *et al.* (1995)	100			
	Kremers *et al.* (2012)			100	
	Theuns *et al.* (2004)		100		
Atrial paced rhythm	**This study**	98.7	81.5	99.6	91.7
Ventricular paced rhythm	**This study**	99.4	96	99.4	80

All values %

It should be noted that many studies published only *CCR* (accuracy) as a performance index and many observers believe *CCR* is inadequate when unsubstantiated to fully reflect classifier performance. Among these was the study by Minhas & Arif (2008), using a *k*-nearest neighbour classifier, the only comparable study which published data for classification of 3 rhythm and features as best classifier for two of them: normal sinus rhythm and premature atrial complexes. Three of the published comparable studies were able to classify 5 rhythms: Acharya *et al.* (2004) and Anuradha *et al.* (2008) both used fuzzy classifiers, but did not feature among the best performing classifiers however the support vector machine of Asl *et al.* (2008) was best performing for second/third degree AV block.

11.4 Considerations for Updates

Given the imperfect classification performances for certain rhythms, consideration was made of an approach to future upgrades.

11.4.1 Re-examination of Performances of Classifier Units

A multi-classifier system is a product of the strengths of its component classifier units. A finding of this study was that the naïve Bayes and decision tree classifier units performed best with the neural network, support vector machine and inference engine classifier units less well

Re-analysis of the results was made, using > 0.5 as an arbitrary minimum level of "good" performance for sensitivity and specificity, corresponding to better than random detection of true positives or true negatives respectively, alongside kappa > 0.75 (Fleiss 1981) and $P < 0.05$, representing rhythms for which each classifier performed well.

For the naïve Bayes classifier, 11 rhythms met these criteria, 10 of which were also those selected by the decision rule for classification by the naïve Bayes classifier (see subsection 10.10.3). The exceptions were normal sinus rhythm, for which kappa was 0.664, failing to meet the above criteria but for which naïve Bayes performed better than all the other classifiers and atrio-ventricular reciprocating tachycardias which met the criteria but were outperformed by the decision tree classifier. The naïve Bayes classifier model selected from cross-validation as best performing was trained on 17 of 18 possible rhythms, the missing rhythm being 2nd and 3rd degree block, explaining its poor performance with this rhythm. This class contained only 2 instances, so of the 10 cross-validation training sets, several could easily contain no examples. Selection of an alternative model could have ensured training with this rhythm but at the cost of overall performance with the remaining rhythms, as suggested by the choice of best performing model that was made.

The naïve Bayes classifier performed well with the largest feature set, containing 216 features. With naïve Bayes, assumption of conditional independence of features is made. A naïve Bayes classifier works well with inputs of high dimensionality and, even if the assumption of conditional independence is unmet, classification performance may be good (Bishop 2006, p.381). Where the underlying distribution is known, the Bayes rule is optimal for classification (Zhang 2004).

During classifier tuning, optimal performance with the options available in Matlab, was with prior probabilities estimated using class prevalence in the training set and using the kernel smoothing density estimate rather than a Gaussian or multinomial distribution. These options were set heuristically with the reasons they worked well being unclear.

The naïve Bayes approach estimates pre-test probability and post-test probabilities, highly relevant to clinical diagnostic process, as illustrated in the cognitive model (see Chapter 3, subsection 3.10.4 and Fig, 3.3). It is postulated that naïve Bayes classification most closely represents the clinical diagnostic process and may account for its superior accuracy.

The decision tree performed well, being trained with the reduced, 116 feature set. 3 rhythms met the above criteria for high performance, (see Fig. 10.10). These were: sinus node dysfunction; atrio-ventricular recip. tachycardias; ventricular paced rhythm. None corresponded to those selected by the decision rule for classification: premature atrial contraction; premature ventricular contractions; atrio-ventricular reciprocating tachycardia and atrial fibrillation. Interestingly, optimal settings were with prior probabilities set according to class prevalence in the training set, as with the naïve Bayes classifier.

Of note, this classifier was trained using the smaller 116 feature set. Reanalysis of the decision tree used in the production system showed branch and leaf structure as in Table 11.17:

Decision nodes corresponded well with clinical diagnostic criteria, with the root node of PP interval of beat R_0 having a threshold value of 988 milliseconds, corresponding to a bradycardia of less than 59 beats per minute. Likewise the RP interval of beat R_{-1}, threshold value of 104 milliseconds, corresponded to the shorter values expected during atrio-ventricular nodal reciprocating tachycardia. Early detection was expected, as instances were rarely spontaneous, having been artificially stimulated and established prior to the nominal first beat R_0.

The stability criterion added in iteration 2 featured twice in this list, emphasising its importance. P wave morphology (Pgrossareasamples) featured highly in diagnosis of atrial pacing and as the node responsible for a split leading to eight leaf nodes (rhythms).

Table 11.17 Production system decision tree nodes.

Code	Beat	Feature	Threshold Value	Branch Level	Leaf node(s)	Split
41	R_0	PPint	988	Root	-	Level 1
24	R_{-1}	RPint	104	1	AVNRT	-
1	R_{-2}	PPint	978	1	RSA	-
65	R_1	PRratio	1.68	2	AF*	Level 3
115	-	Stability	22.5	2	SND, VPACE	-
90	R_2	Pgrossareamatch	4	3	APACE	Levels 4 - 7
84	R_2	RPint	385	4	-	Levels 5 - 7
83	R_2	RRint	600	5	VPACE	-
115	-	Stability	5	5	NSR	-
21	R_{-1}	PPint	812	6	MAT	-
45	R_0	PRratio	0.78	6	PAC*, PVC*	-
108	-	Rx	1.5	7	AT, OAVRT*	-

* Rhythms classified by the decision tree within the production system.

With the support vector machine, performances for 3 rhythms met these criteria (see Table 10.12): respiratory sinus arrhythmia; physiological sinus tachycardia and sinus node dysfunction, of which one, sinus node dysfunction, was the rhythm selected for classification. The support vector machine performed poorly throughout, in contrast to studies supporting its use (Polat & Gunes 2007; Asl *et al.* 2008). Polat & Gunes (2007) used PCA and Asl *et al.* (2008) used GDA for feature selection. Major differences in methodology may be responsible for the performance difference.

For the neural network classifier, none of the rhythms met the criteria (see Tables 10.12 and 10.14). Multifocal atrial tachycardia was selected for classification with the NN classifier, with kappa of 0.359 being the sole poor performance metric, however this improved, when implemented in the production system as a binary classifier, to 0.612 ± 0.029. Networks are known to classify poorly where there are a large number of input features and there is possibility that many of these are redundant, or not required for a diagnosis, making it susceptible to the "curse of dimensionality". Feature reduction to correct for this was discussed in Chapter 10, subsection 10.5 and implemented in iteration 3 but was insufficient to improve performance (see Chapter 10, Table 10.8).

The inference engine was used as a specialist classifier only for 2nd and 3rd degree atrio-ventricular block, despite low error of 0.011 ± 0.006 (value ± CI_{95}) and perfect sensitivity of 1.000 ± 0.000, specificity also good at 0.989 ± 0.006, however κ was very low at 0.248 ± 0.025. The

implication was that all classifiers were poor at detecting 2nd and 3rd degree atrio-ventricular block. The inference engine, which successfully classified this rhythm, used a fixed set of rules and unlike the other classifiers was not subjected to training. The classifier was limited by design to the constraints set by the clinical guidelines on which it was based, so less amenable to modification.

11.4.2 Binary Classification

The iterative development process undertaken in Chapter 10 study was based on a premise that classifier selection and tuning was best performed while evaluating overall performance. A reasonable alternative approach would be to examine classifier performances for each rhythm from the outset, with a binary classification model. The iterative process undertaken in this study led to the inclusion of specialist binary classifiers into our production system, for rhythms where that was indicated to be advantageous. It is possible that further performance improvements could be achieved using specialist binary classification for all rhythms.

11.4.3 Elimination of Unnecessary Features

The Hughes effect suggests that high dimensionality leads to reduced predictive power (Hughes 1968). The high dimensionality of the features in this study resulted from its adherence to inclusion of all the features required to satisfy the diagnostic criteria, as laid out in the guideline documents used to provide cardiac rhythm diagnostic criteria (see Chapter 3, subsection 3.10.1) (Buxton et al. 2006; Blomström-Lundqvist et al. 2003; Epstein et al. 2008; Brugada et al. 1991; Saoudi et al. 2001; Fuster et al. 2011; Bonow et al. 2012; Surawicz et al. 2009) and to allow comparison with ICD algorithms (see Chapter 3, subsection 3.18.2). The implication of the Hughes effect is that there is potential for dimensionality reduction to lead to improved classification performance (see Chapter 3, section 3.7 and Chapter 8, section 8.5).

There was an heuristic approach to dimensionality reduction in this study, with evaluation of alternate feature sets, elimination of features (see Chapter 11, subsection 11.2.1) and the number of beats required to diagnose rhythm (see Chapter 11, subsection 11.2.2). A more algorithmic approach, such as data mining, could lead to elimination of features with little or no influence on diagnosis, for reasons as yet unexplained by current cardiology theory.

11.4.4 Examination of the Number of Beats to Diagnose

This study included a rudimentary comparison of 5 beats analysis and 10 beat analysis (see Chapter 3, subsection 3.18.2 and Chapter 11, subsection 11.2.2). This study did not systematically examine the issue beyond these constraints and it is possible that more detailed examination of the number of beats required to diagnose could provide guidance towards further performance improvements.

11.5 Limitations of the Study

In Chapter 6, subsection 6.4.3 the pilot study (Bostock 2004) and data from Hintringer *et al.* (2001) had suggested 100 patients were required to provide sufficient power to reject the null hypothesis. 61 patients were recruited but nonetheless, significance at the 5% level was achieved, indicating the study was over-powered. Conversely, the low prevalence of many rhythms encountered suggested that a larger sample could have improved the sizes available for training and test sets, enabling fully independent training and testing tests.

With the small dataset of 61 patient recordings and 1109 rhythm instances obtained in the data collection phase of the study, it was necessary to re-use data for training and testing. Cross-validation was used to reduce the risks of over-training but a remaining risk was acknowledged. Testing with new data would help in estimating true generalisability of the classifiers.

The rhythm instances were collected from a single site, un-randomised and unmatched (such as for patient age and sex) study. It is accepted that a multicentre study would provide data from a more diverse population. Given the unknown nature of a patients' arrhythmia before their diagnostic study, any randomisation might be difficult to implement.

Confidence limits were calculated for the major performance measures to check for imprecision. Given the importance of statistical significance using *P* values in studies of this type and given the large number of measures analysed it was elected to limit analysis of this data.

The absence of real-time blood pressure data limited the credibility of haemodynamic assessment, which relied upon a derived parameter – impedance cardiography, perhaps diluting its impact. Likewise, the absence of normal physical activity at the time of data recording, a circumstantial necessity during EP studies, which must be conducted with patients supine and at rest, precluded evaluation the impact of exertional stress indices on rhythm diagnosis.

This study did not evaluate nuances of diagnosis during EP study, such as the value of timed electrical stimulations in differential rhythm diagnosis and interpretation of the multiple waves of the His bundle and coronary sinus electrograms. As this information is considered invaluable to diagnosis during EP studies, it would be desirable to include them in a future development.

11.6 Consideration of Bias

It is possible that patients selected for EP study may include selection bias. There was no treatment allocation relevant to this study, as ablation was considered an extension of the procedure made to avoid unnecessary delay to treatment, rather than allocation to a study treatment group. Referral for EP study was the point at which selection may have occurred, but this was mitigated by the

fact that only patients already referred were approached for inclusion. No systematic selection bias was identified.

Blinding was not possible as investigator and patient were both unaware of the diagnosis prior to the study, as that was the justification for the procedure being performed. There were 4 patients consented for the study for whom no data was collected and may be considered attrition bias, as that uncollected data may have affected results. No reporting bias was expected as all results have been reported in full.

11.7 Consideration of Error

Measures were taken to minimise random error, using careful equipment specification the resolution of measurements, so this was known prior to data collection. Data was continuously monitored during data collection, allowing observed variations or errors to be noted and accommodated. Data quality assessment, used as a form of random error assessment, showed low random error, in the form of acceptable completeness, consistency, timeliness and accuracy analysis (see Chapter 9, section 9.4).

Potential differences of opinion between the domain experts providing the "gold standard" diagnosis was not considered and is a potential source of error in annotation of rhythms.

11.8 Advantages of this Approach to Classification

In this study, we have shown that to work well, data for classifiers should be specific to the task. Re-use of standard databases as sources of data for training and testing, as was the case with the majority of bench studies reviewed, serves to facilitate comparisons between studies but limits any developmental capability to the information contained within that data.

Use of a diverse set of classifier units, enabled rigorous testing of the utility of AI in this domain and sought to extend that utility to maximise performance by selecting the best classifiers for inclusion in a multi-classifier system.

With the use of AI in classification, there is the possibility of updating algorithms, by retraining with new data, should additional or new examples be added to a suitable database. Also, within the limits set by hardware capability, software upgrades to modify classifier structure are possible.

11.9 Computational Cost

Computational resources are used by algorithms in solving computational problems. Efficient algorithms use resources below acceptable levels, such as code execution within a reasonable time on a normal computer. Simple computational resources are computation time, the number of

steps required to solve a problem, memory space, and disk space. The most commonly used measures of computational cost are computation time and memory usage.

Expressions of computational complexity were obtained, using datasets having differing numbers of inputs but with processing code matched. The production system was used, varying the number of instances used in the model. For each dataset, the number of inputs was computed as the sum of the product of number of features and number of instances. Computation time was calculated using MatLab *tic* and *toc* functions and memory usage by the *profile* function (see Table 11.18).

Table 11.18 Computational cost indices.

No. of Instances	No. of Inputs	Computation Time (secs)	Memory Usage (kB)
1109	865020	43.08382	4684
1100	858000	34.53114	1528
1000	780000	31.86179	2096
900	702000	29.52177	1680
800	624000	27.23476	1516
700	546000	24.76662	1324
600	468000	10.25256	1172
500	390000	9.329284	972
400	312000	9.304069	1224
300	234000	8.785004	532

Below 300 instances, the model encountered errors and computation time was not obtained. A simple expression of computational complexity, using big O notation, was possible by analysis of the relationship between the number of inputs and computation time (see Fig. 11.2).

For up to 500 instances, processing time was nearly constant at approximately 9 seconds, a designation of O(1) and the relationship is approximately linear, with processing time related to number of inputs by a factor of 22, suggesting an O(n) designation. Both these big O designations were low, suggesting that the production system might be amenable to implementation using larger datasets without major impact on resources.

Regarding memory usage, the maximum allocation of less than 5MB of memory (see Table 11.8), considerably below the expected norm of 4GB of memory available in a standard PC, supported the view that standard PC computation power would suffice for the processes used here.

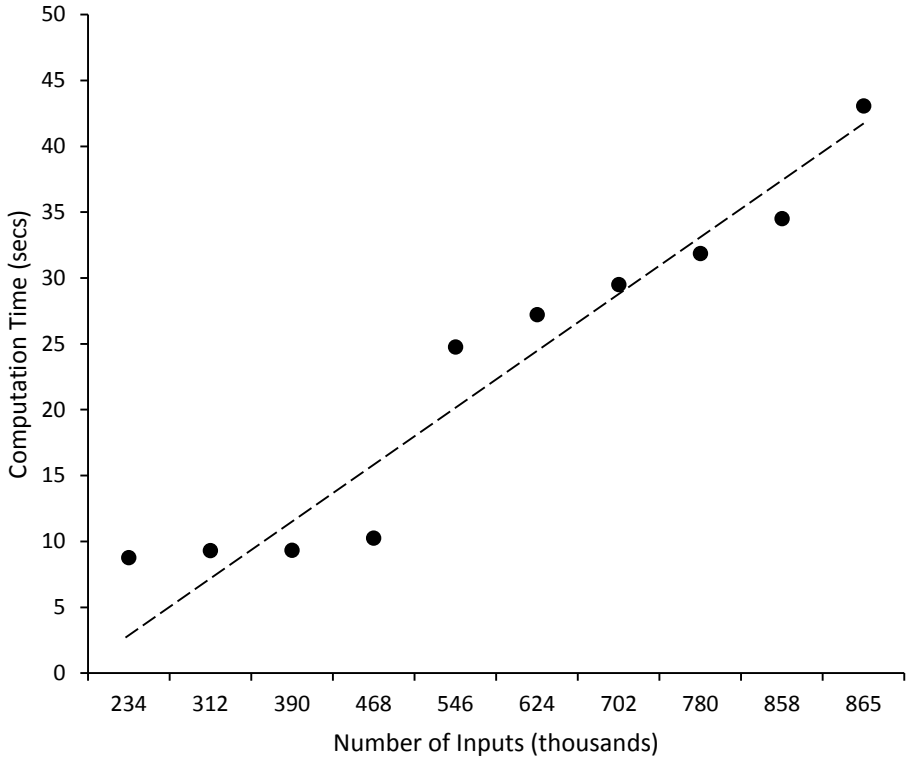

Figure 11.2 Relationship between the number of inputs and computation time. Computation time (complexity) increases with increasing numbers of inputs with an approximately linear relationship (dotted line).

It was noted that these calculations were applied to classifier base units which had already been trained and fixed. The training process itself requires considerably more time and processing power but would only be required during algorithm maintenance, when new training instances became available. Re-training could be performed off-line, following which the fixed classifiers could be updated.

11.10 Summary

The holistic approach to the problem of cardiac rhythm classification taken in this study was emphasised and the appropriate use of heuristic, algorithmic approaches and modelling during classifier development were noted.

Data analysis was performed offline, using MatLab. Data for rhythms encountered during a procedure were concatenated into 10 beat rhythm segments.

Following implementation of the production system analysis of the relative importance of features was performed using ReliefF. Ten of the top 30 ranked features related to QRS morphology, nine

to intervals. 6 to predisposing factors and 4 to P wave morphology features. Stroke volume index onset (SVI onset), ranked fifth, a previously unreported finding.

Re-analysis of data was performed following removal of low ranking features. Accelerometry, temperature and QT interval were selected for this, to evaluate their impact on results. Performance was assessed with each of the 5 classifiers used in the production system and showed no significant differences in performance due to accelerometry, temperature or QT interval data. Equally important, no performance advantage was gained by their removal, reducing the likelihood of any confounding influence.

Testing was performed to compare performance of 10 beat and 5 beat diagnostic models. 10 beat analysis showed small performance improvements in indices over the original production system for some rhythms and 5 beat analysis showed poor overall performance. The lack of correspondence with results from the production system was expected to be due to the different approaches taken between iterative development, compared with the structured test and the different datasets and processing used.

In summary, the production system diagnostic performances achieved in this study were the best performing of all compared algorithms, for 3 of 8 different rhythms for which comparison was made: physiological sinus tachycardia, sinus node dysfunction and ventricular tachycardia.

When published studies were compared with our results, neural network classifiers performed best for 4 rhythms, *k*-nearest neighbour for 2 rhythms, a statistical classifier, support vector machine and ICD algorithms for only one rhythm. Decision trees and naïve Bayes classifiers did not feature in published studies. The production system performance of this study was the best performing for 3 of the 8 different rhythms for which comparison data were available and the remaining 10 rhythms were, to our knowledge, classified only in our study. The results of this study demonstrated that high performance is achievable for a range of different rhythms and was able to classify 18 different rhythms.

This study found that that the naïve Bayes and decision tree classifier units performed best and the classifier units were re-examined in this light. The naïve Bayes classifier performed well with the higher dimensionality 10 beat set for the highest number, 11 of 18 rhythms. It was postulated that the accuracy of the naïve Bayes classification was due to its close representation of the clinical diagnostic process. The decision tree performed best using the 5 beat feature set and analysis of decision nodes showed a good correspondence to clinical diagnostic criteria. The support vector machine performed well for only 3 rhythms and was selected as best performing for only one, in contrast to the views of many exponents of its use in the literature (see Chapter 2, subsection

2.8.7). We suggest this finding was due to major methodological differences with this study. The neural network classifier performed unexpectedly poorly for all rhythms, performing best for only one rhythm. The explanation was proposed to be the known poor performance of networks for a large number of input features, the "curse of dimensionality". Feature reduction was insufficient to improve performance (see Chapter 10, Table 10.8). The inference engine was used as a "catch-all" for otherwise undiagnosed rhythms, of which 2nd and 3rd degree atrio-ventricular block was specified, as all other classifiers were poor for that rhythm.

The study was found to be over-powered, however low prevalence of some rhythms suggested a larger sample could have supplemented training and test sets, avoiding re-use of data. Data was collected form a single site and was un-randomised and unmatched. It was accepted that a multicentre study would provide more diverse data but that randomisation might be difficult to implement. Lack of blood pressure data limited the credibility of assessment, diluting its impact and absence of physical activity during data collection precluded evaluation of its impact on rhythm diagnosis.

A possible selection bias for patients referred for EP study was identified but no systematic selection bias was identified. The absence of a treatment arm was noted and blinding was not possible. Small data losses during data collection were attributed to complexity of the setup and random equipment failures. There was a small potential attrition bias, due to uncollected data and no evidence of reporting bias. Error was minimised by equipment specification and careful monitoring during data collection. Data quality was assessed and showed low random error, acceptable completeness, consistency, timeliness and accuracy.

The development of classifiers specific to the task, using data prospectively collected rather than an existing database, was considered advantageous, a view supported by leading texts on machine learning. Classifier diversity was achieved using a set of different classifier technologies to maximise performance. With a learning capability, classifiers offered the prospect of future upgrade should new data become available.

Computational resource requirement of the production system was estimated and found to be amenable to implementation, without concern over hardware capability or any time constraint.

Chapter 12 Conclusions

The integrated development process presented here succeeded in producing a system, based on a combination of AI technology, in a mixture-of-experts multi-classifier system, which could accurately diagnose 16 of 18 different cardiac rhythms tested. This study was able to classify 18 different rhythms, a greater number than any published study. 4 rhythms were classified with better performance by comparable published studies but for a further 4 rhythms, performance was better than comparable published studies and 10 rhythms were, to our knowledge, classified only in our study.

The algorithm was sufficiently developed for a future stage of testing, in a live clinical environment. The system could be interpreted as having value for use in a variety of clinical applications including invasive EP testing and with potential application in other domains, such as AED's, ICDs and pacemakers.

12.1 Production System Rhythm Diagnostic Algorithm

Clinical diagnostic processes were modelled and detailed domain knowledge used to develop the feature set for the AI classifiers developed in this study. The resultant production system incorporated five classifier units: a naïve Bayes; decision tree, neural network, support vector machine and an inference engine, in a multi-classifier system having a mixture-of–experts configuration. This production system provided high performance indices for all the 18 rhythms assessed. The performance indices of the production system were perfect for three rhythms, including differential diagnosis if VT and VF, though this should be qualified by low prevalence of the rhythms in the study sample.

The naïve Bayes classifier was dominant in the system, responsible for the diagnosis of 11 of 18 rhythms and explanations were offered for this high performance compared to other classifiers.

Several new findings resultant from this study were noted during ReliefF feature analysis (see Chapter 11, section 11.2). The high importance attributable to symptoms and predisposing factors was unexpected. Other new findings were that change in haemodynamic status and P wave morphology were important in rhythm diagnosis. There was conflicting evidence of the efficacy of 5 beat rhythm detection and this warrants further study.

12.2 Contributions

During the course of this study, the following contributions were made:

1. Simulating the cognitive processes of clinical cardiac rhythm diagnosis, emphasising the use of domain knowledge in classifier design.

2. Creation of a library of physiological signals with annotated cardiac rhythms.
3. Use of artificial intelligence-based classifiers to improve automated cardiac rhythm diagnostic accuracy beyond current capability.
4. Naïve Bayes classification as the highest performing classifier in this domain.
5. Multi-classifier systems outperform individual classifier systems in this domain.

This study introduced the use of several features into rhythm classification; all indicated using the knowledge-driven approach adopted from the outset of the study. These were:

1. Symptoms and predisposing factors have a high influence on rhythm diagnosis.
2. Sudden haemodynamic change, detected using impedance-based measurements has value in rhythm classification.
3. P wave (atrial morphology) has value in rhythm classification.

12.3 Achievement of Objectives

The aims produce a rhythm classification algorithm using artificial intelligence, suitable for use during EP studies was considered achieved, though adaptation and further testing of the algorithm will be required.

The study objectives centred on a staged process of classifier development, within the framework of a system development life-cycle. Classifiers were successfully developed and tested using this template.

12.4 Acceptance of the Research Hypothesis

The null hypothesis (H_0) (see Chapter 1, section 1.8) was "*Prototype cardiac rhythm diagnostic classifiers using AI do not outperform current algorithms".* For testing the null hypothesis, a 5% significance level (α) was selected (see Chapter 3, section 3.20). The Fisher exact test was used (see Chapter 3, subsection 3.12.6 and section 3.3) to calculate *P* values (see Table 10.14). For the differences of the production system classification compared with the gold standard, for all rhythms, *P* values were < 0.001. No multiple comparisons with comparable studies were possible, so there was no requirement for a Bonferroni correction. As *P* values were below the level of α, the null hypothesis (H_0) was rejected and the research hypothesis (H_1), that "*Prototype cardiac rhythm diagnostic classifiers using AI outperform current algorithms*" was accepted.

Type I errors (α) were high for five rhythms, normal sinus rhythm at 0.252, premature atrial complexes at 0.081, premature ventricular complexes at 0.026 , second and third degree AV block at 0.011, multifocal atrial tachycardia at 0.011and below 0.010 for all other rhythms.. Type II errors (β) were high for normal sinus rhythm at 0.135, respiratory sinus arrhythmia at 0.184,

premature atrial complexes at 0.403, premature ventricular complexes at 0.508, sinus node dysfunction at 0.069, first degree AV block, at 0.333, AV nodal and junctional tachycardias at 0.074, AV reciprocating tachycardias at 0.250, multifocal atrial tachycardia at 0.235, atrial fibrillation at 0.130, and for atrial paced rhythm at 0.083, ventricular paced rhythm at 0.200 and below 0.010 for the six remaining rhythms.

Despite achievement of the required significance level, the presence of type I error means it remains possible that a true null hypothesis was rejected (see Chapter 10, section). The major effect of the presence of type II error (β), where a false null hypothesis is failed to be rejected was not the case in this study. The Type I errors were represented by false positives (FP) and further improvements in algorithm performance should perhaps focus on reducing these errors.

12.5 Further Research

The low prevalence of certain arrhythmias within the sampled instances (see Chapter 9, section 9.3) suggested the need for a larger sample dataset.

High performances achieved for the wide range of rhythms assessed in this study encourage further development of this line of research into continued improvement of cardiac rhythm diagnostic algorithms. The finding that a naïve Bayes classifier was the best performing was unexpected and requires further support in new studies.

A consequence of the domain knowledge-driven classifier development process adopted in this study was the feature set. Several features, some of which were previously used in standard and investigational implantable cardiac devices, have unknown value, such as: atrial electrogram morphology; electrogram axes; stress indices, haemodynamic sensors and 5 beat detection, and remain poorly used in the context of rhythm classification. The contributions of each to the production classifier were initially unclear from results. ReliefF analysis supported the value and will require further investigation, in dedicated studies.

Many medical devices accept periodic software updates and likewise, existing implantable devices, such as pacemakers and ICDs, are able to accommodate such updates, limited only by hardware capability.

Computational cost was briefly touched upon in this study (see Chapter 1, subsection 1.4.5 and Chapter 11, section 11.9) but the computational requirements for loading such an algorithm into an investigational device was not assessed as part of this study but it is conceivable that investigational software could be developed, incorporating an algorithm of this type, then uploaded into implanted devices for clinical testing.

The benefits of binary classification for all rhythms were unexplored. Further research, using a large data library of rhythm instances would be required to clarify any advantages.

This study relied upon accepted cardiology theory for selection of features and hence the dimensionality of data. Data mining or other algorithmic approaches could be evaluated to include features only with significant influence on diagnosis,

A rudimentary evaluation in this study, which compared classifier performance for rhythm analysis using 5 or10 beats, demonstrated differences, suggesting that a more detailed study of the exact number of beats required to diagnose rhythm with highest classification performance could guide further performance improvements.

The iterative development process undertaken in Chapter 10 study was based on a premise that classifier selection and tuning was best performed while evaluating overall performance. A reasonable alternative approach would be to examine classifier performances for each rhythm from the outset, with a binary classification model. The iterative process undertaken in this study led to the inclusion of specialist binary classifiers into our production system, for rhythms where that was indicated to be advantageous. It is possible that further performance improvements could be achieved using specialist binary classification for all rhythms.

The inclusion of His bundle and coronary sinus electrograms and timed electrical stimulation in differential rhythm diagnosis during EP studies should be evaluated in a future study.

The production classifier accepted a fixed “snapshot” of data from previously recorded physiological data, so conversion for use in a live clinical environment would be required. Real-time, streamed data would need to be continuously input to the algorithm, demanding data buffering capabilities and rapid computation times for timely diagnosis. Software containing this algorithm could be loaded onto an investigational EP analysis system for testing in a clinical environment, during electrophysiological studies. A parallel system, running alongside normal clinical systems should avoid affecting patient outcomes and would be a natural precursor to its adoption in a clinical environment.

Appendix A Factors predisposing toward arrhythmia

Electrical Heart Disorders

Accessory pathways

Wolff–Parkinson–White syndrome

Ebstein's Anomaly of the Tricuspid Valve

Primary electrical disorders (i.e. long QT syndrome, Brugada syndrome)

Survival of cardiac Arrest

Out-of-hospital resuscitation

SCD in the normal heart

Sudden infant death syndrome

Drug-induced torsades de pointes and SCD

Catecholaminergic polymorphic ventricular tachycardia

Previous cardiac Surgery

Valve surgery

Post-surgical Ventricular Septal Defect

Post-surgical Tetralogy of Fallot

Post-surgical Transposition of the Great Vessels

Post-surgical Fontan Repairs

Inflammation near the AV conduction system after surgery in this region

Other cardiac Disease

Chronic coronary heart disease -old myocardial infarction

Congenital heart disease

Myocardial ischemia

Congestive cardiac failure

Heart failure

Aortic stenosis

Mitral valve prolapse

Atrial Septal Defect,

Myocarditis

Endocarditis

Rheumatic heart disease

Cardiomyopathy - Dilated, Hypertrophic, Arrhythmogenic right ventricular

Anomalous origin of coronary arteries

Myocardial bridging

Conditions causing myocardial scarring (sarcoidosis, amyloidosis, tuberculosis)

Pulmonary hypertension

Appendix A Factors predisposing toward arrhythmia (continued)

Neurological Disorders

Carotid sinus hypersensitivity

Enhanced vagal tone (neurogenic syncope)

Neuromuscular diseases (e.g. myotonic dystrophy)

Other Medical Conditions

Anaemia

Malignancies

Chronic Pulmonary disease

Hyperthyroidism

Hyperlipidaemia

Electrolyte disturbance

Hypovolemia

Hypothermia

Pulmonary emboli

Chest trauma

Shock

Pulmonary emboli

Endocrine disorders and diabetes

Pericardial diseases

End stage renal failure

Obesity, dieting and anorexia

Medications

Prescribed compounds (salbutamol, aminophylline, atropine, catecholamines)

Anticancer treatments anthracycline compounds, doxorubicin, Adriamycin, daunorubicin- proarrhythmic side effects

Antiarrhythmics and other cardiac drugs- proarrhythmic side effects

Drug-Drug and Drug-Metabolic Interactions - proarrhythmic side effects

Lifestyle Factors

Physical or mental stress (Physical exertion, Anxiety)

Pyrexia, Fever, infection

Lack of sleep

Premenstrual or menstrual

Trained heart - Athlete

Pregnancy

Use of stimulants (e.g., caffeine, alcohol, nicotine)

Recreational/ illicit drugs (e.g., amphetamines, cocaine, “ecstasy,” cannabis)

Appendix B Search results by search term

	Google	Google Scholar	Pubmed	IEEE Xplore	Science Direct	Web of Science	HRS Abstracts	Cochrane Library
rhythm	0	0	>1000	>1000	>1000	>1000	>1000	0
rhythm AND classifier	8	7	6	49	9	26	0	0
rhythm AND "artificial intelligence"	0	0	20	123	>1000	7	0	0
rhythm AND "neural network"	1	1	15	41	10	43	0	0
rhythm AND "fuzzy"	1	8	9	25	7	16	0	0
rhythm AND "support vector	7	5	6	15	4	12	1	0
rhythm AND "expert system"	5	5	9	6	4	11	2	0
rhythm AND "decision tree"	6	2	3	2	2	4	0	0
rhythm AND "Bayes"	5	3	4	4	1	4	0	0
rhythm AND "genetic algorithm"	0	1	0	1	0	2	2	0
rhythm AND evolutionary	0	0	0	2	1	0	0	0
rhythm AND "pattern recognition"	0	0	35	56	9	16	1	0

Appendix B (continued)

	Google	Google Scholar	Pubmed	IEEE Xplore	Science Direct	Web of Science	HRS Abstracts	Cochrane Library
implantable cardioverter-defibrillator	0	0	>1000	19	>1000	>1000	>1000	1
algorithm AND comparison	0	0	>1000	>1000	>1000	>1000	2	0
implantable cardioverter-defibrillator AND algorithm	9	7	647	12	>1000	491	5	8
implantable cardioverter-defibrillator AND algorithm AND comparison	2	8	23	3	>1000	22	0	1

Appendix C Search results by journal, author and ICD algorithm

a) Results by Journal

Journal	Papers	Impact Factor (2010)
IEEE Transactions on Biomedical Engineering	21	2.154
Pacing and Clinical Electrophysiology	14	1.352
Journal of Cardiovascular Electrophysiology	8	3.703
Circulation	6	14.816
Artificial Intelligence in Medicine	5	1.568
Journal of Electrocardiology	5	1.109
Europace	5	1.839
Physiological Measurement	4	1.567
Journal of the American College of Cardiology	4	12.535
Heart Rhythm	1	4.246

b) Results by lead author

Lead Author Name	Papers	Most Recent	Institution
Gold MR	5	2012	Medical University of South Carolina, Charleston, USA
Swerdlow CD	5	2012	Cedars-Sinai Heart Institute, Los Angeles, USA
Gillberg JM	4	2012	Medtronic Inc., Minneapolis, USA
Ellenbogen K	4	2012	Virginia Comm. Univ. Sch. Med., Richmond, USA
Krasteva V	3	2010	Bulgarian Academy of Sciences, Sofia, Bulgaria
Ozbay Y	3	2009	Selcuk University, Konya, Turkey.
MacFarlane P	3	2007	University of Glasgow, Glasgow, UK
Osowski S	3	2005	Warsaw University of Technology, Warsaw, Poland
Theuns DA	3	2004	Erasmus MC, Rotterdam, Netherlands

c) Results by ICD algorithm

ICD Technology	Papers
Boston Scientific Ventak AV	3
Sorin ELA PARAD+	1
Medtronic PR Logic	1
Boston Scientific Rhythm ID	1
Medtronic Wavelet	1
Medtronic RV Pressure	1

Appendix D Results by AI Technology

Technology	Journal Papers	Journal Papers ≤ 5 Years	Conference Papers ≤ 5 Years	Total ≤ 5 Years
Neural Network	14	4	7	11
Morphology	10	5	0	5
Fuzzy	9	3	3	6
Bayes	5	3	1	4
Expert System	4	4	0	4
Support Vector Machine	4	2	4	6
Wavelet	3	2	3	5
Heart Rate Variability	3	2	1	3
Neuro-Fuzzy Hybrid	2	1	3	4
Decision Tree	2	1	1	2
Hybrid Neural Network	1	1	1	2
Fuzzy-clustering-NN	1	1	0	0
SVM-NN-Perceptron	1	1	0	0
Linear discriminant analysis	1	1	0	0
Phase space matrix	1	1	0	0
Time-Frequency Analysis	2	0	0	0
Syntactic	2	0	0	0
Fuzzy-*k*-NN	1	0	1	0
Right Atrial Pressure	1	0	0	0
Hermite polynomials, Neuro-Fuzzy	1	0	0	0
Chaos	1	0	0	0
Wavelet-NN	1	0	0	0
Fuzzy-Artificial Immune System	1	0	0	0
Statistical	1	0	0	0
Onset	1	0	0	0
Fuzzy-Expert	1	0	0	0
Heart Rate Turbulence	1	0	0	0
Linear Classifier	1	0	0	0
Rule-Based	1	0	0	0
Knowledge-Based Automaton	1	0	0	0

Appendix E Rhythm definitions

Normal Rhythm

Normal Sinus Rhythm (NSR)

"P<120; PR 120-200; QRS<120:QT_c >=440-460" (Bonow *et al.* 2012).

"The P wave on a 12-lead ECG is positive in leads I, II, and aVF and negative in aVR. Its axis in the frontal plane lies between 0 and +90; in the horizontal plane, it is directed anteriorly and slightly leftward and can, therefore, be negative in leads V1 and V2 but positive in leads V3 to V6. The PR interval is normally between 120 ms and 200 ms (220 ms in the elderly). The P waves have a normal contour, but a larger amplitude may develop and the wave may become peaked" (Blomström-Lundqvist *et al*. 2003).

"QRS axis -30 - +90; QRS width ≤ 110ms" (Surawicz *et al*. 2009)

Respiratory Sinus Arrhythmia (RSA)

"Normal P wave morphology/axis. Gradual phasic change in PP interval of more than 10% or 120 ms" (Buxton *et al.* 2006)

Physiological Sinus Tachycardia (ST)

"A cardiac arrhythmia emanating from the sinus node at a rate >100 bpm (cycle length: <600 ms) which demonstrates a gradual onset and termination and is in keeping with the level of physical, emotional, pathological, or pharmacological stress (Buxton *et al*. 2006). Physiological sinus tachycardia is defined as an increase in sinus rate >100 bpm in keeping with the level of physical, emotional, pathological, or pharmacologic stress" (Blomström-Lundqvist *et al*. 2003)

Premature Atrial Complex (PAC)

"A depolarization of the atrium which occurs with a coupling interval shorter than that resulting from the intrinsic heart rhythm" (Buxton *et al.* 2006)

Premature Ventricular Complex (PVC)

"A depolarization of the ventricle which occurs with a coupling interval shorter than that resulting from the intrinsic heart rhythm" (Buxton *et al.* 2006)

Sinus Node Dysfunction

"Sinus node dysfunction manifested as:

Sinus rate inappropriately slow for the conditions.

Sinus arrest, sinoatrial exit block

Prolonged pauses (analogous to sinus node recovery time >1,500 ms or a corrected sinus node recovery time greater 550 ms) following cessation of supraventricular tachyarrhythmias" (Buxton *et al*. 2006)

"SND refers to a broad array of abnormalities in sinus node and atrial impulse formation and propagation." (Epstein *et al*. 2008)

Appendix E Rhythm definitions (continued)

Sinus Bradycardia (SB)

"Sinus bradycardia characterized as sinus rate less than 60 beats per minute (bpm) (cycle length, >1000 msec) with normal P wave axis. Note that P wave morphology may be atypical at slow rates" (Buxton *et al.* 2006)

Sinus Arrest (SA)

"Pause without a P wave, >2.0 s during sinus rhythm; PP interval of pause not a multiple of basic PP interval" (Buxton *et al.* 2006)

Sinoatrial Block (SAB)

"Normal P wave morphology/axis. Pauses with no visible sinus P waves. Constant PR interval

Mobitz I: Progressive decrease in PP interval before pause; PP interval of pause less than twice the preceding PP interval; PP interval following pause greater than twice PP interval preceding pause

Mobitz II: Constant PP interval before and after pause; Pause is an integral multiple (within 100 ms) of normal PP interval" (Buxton *et al.* 2006)

Atrio-ventricular Block

"Classified as first-, second-, or third-degree (complete) block; anatomically, it is defined as supra-, intra-, or infra-His" (Epstein *et al.* 2008)

First degree Atrio-ventricular Block (1HB)

"PR interval >200msec" (Buxton *et al*. 2006)

"Abnormal prolongation of the PR interval (greater than 0.20 seconds)" (Epstein *et al*. 2008)

Second degree Atrio-ventricular Block

"Regular atrial rhythm with intermittent nonconducted P waves" (Buxton *et al*. 2006).

"Advanced second-degree AV block refers to the blocking of 2 or more consecutive P waves with some conducted beats, indicating some preservation of AV conduction. In the setting of AF, a prolonged pause (e.g., greater than 5 seconds) should be considered to be due to advanced second-degree AV block" (Epstein *et al.* 2008)

Second degree Atrio-ventricular Block (Mobitz Type I) (2HB)

"Type I second-degree AV block is characterized by progressive prolongation of the interval between the onset of atrial (P wave) and ventricular (R wave) conduction (PR) before a nonconducted beat and is usually seen in conjunction with QRS. Type I second-degree AV block is characterized by progressive prolongation of the PR interval before a nonconducted beat and a shorter PR interval after the blocked beat" (Buxton *et al*. 2006)

Appendix E Rhythm definitions (continued)

Second degree Atrio-ventricular Block (Mobitz Type I) (2HB)

"Type II second-degree AV block is characterized by fixed PR intervals before and after blocked beats and is usually associated with a wide QRS complex. When AV conduction occurs in a 2:1 pattern, block cannot be classified unequivocally as type I or type II, although the width of the QRS can be suggestive, as just described" (Buxton *et al.* 2006)

Complete Atrio-ventricular Block (CHB)

"Independent atrial and ventricular complexes with atrial rate usually exceeding ventricular rate" (Buxton *et al.* 2006)

"Third-degree AV block (complete heart block) is defined as absence of AV conduction" (Epstein *et al.* 2008)

Narrow Complex Tachycardia (NCT)

"If QRS <120ms, the tachycardia is almost always supraventricular. Haemodynamically stable, QRS <120 is SVT. If no P waves or evidence of atrial activity and RR interval is regular, then AVNRT is most commonly the mechanism. If a P wave is present in the ST segment and separated from the QRS by 70 ms, then AVRT is most likely. In tachycardias with RP longer than PR, the most typical diagnosis is atypical AVNRT, permanent form of junctional reciprocating tachycardia (PJRT) (ie, AVRT via a slowly conducting accessory pathway), or AT." (Blomström-Lundqvist *et al.* 2003)

Postural Orthostatic Tachycardia Syndrome (POTS)

"Orthostatic rise in heart rate of >30 bpm above baseline or >120 bpm within the first 10 min of head-up tilt, accompanied by palpitations, and no significant (<10 mm Hg) fall in systolic blood pressure" (Buxton *et al.* 2006)

"Individuals present with orthostatic intolerance (ie, symptoms on standing , relieved by recumbency) with exaggerated, persistent postural sinus tachycardia (> 30 bpm from baseline or > 120 bpm) within 10 minutes of an upright tilt in the absence of postural hypotension and any demonstrable autonomic neuropathy." (Blomström-Lundqvist *et al.* 2003)

Sinus Node Re-entry Tachycardia (SNRT)

"Arise from re-entrant circuits involving the sinus node's production of paroxysmal, often nonsustained bursts of tachycardia with P waves that are similar, if not identical, to those in sinus rhythm. They are usually triggered and terminated abruptly by an atrial premature beat". (Blomström-Lundqvist *et al.* 2003)

Atrio-ventricular Nodal Reciprocating Tachycardia (AVNRT)

"A regular SVT from re-entry within the AV node and/or perinodal atrial tissue. Subclasses: Slow-fast; Fast-slow; Slow-slow" (Buxton *et al.* 2006).

"Rates of tachycardia often between 140 and 250 per minute. The most common SVT

not usually associated with structural heart disease. involves reciprocation between two pathways, the fast pathway and the slow pathway. Typical AVNRT slow-fast AV-node re-entry. shorter duration (40 ms) P wave during or close to the QRS complex (less than or equal to 70 ms). Atypical AVNRT less common (5-10%) fast-slow AV-node re-entry, producing a long R-P tachycardia or slow-slow AV-node re-entry, P wave is after the QRS (RP interval greater than or equal to 70 ms)" (Blomström-Lundqvist *et al.* 2003)

Focal Junctional Tachycardia (FJT)

"Arises from the atrio-ventricular junction, has a rate >60 bpm (cycle length: <1,000 ms), and may demonstrate dissociation from atrium or ventricle" (Buxton *et al*. 2006).

"Origin from AV node or His bundle. include heart rates of 110 to 250 bpm and a narrow complex or typical BBB conduction pattern. Atrio-ventricular dissociation is often present although one-to-one retrograde conduction may be transiently observed. On occasion, the junctional rhythm is quite erratic, suggesting AF" (Blomström-Lundqvist *et al.* 2003)

Non-Paroxysmal Junctional Tachycardia

"A benign arrhythmia characterized by narrow complex tachycardia 70 to 120 bpm. Mechanism is thought to be enhanced automaticity arising from a high junctional focus or in response to a triggered mechanism. It shows a typical "warm-up" and "cool-down" pattern and cannot be terminated by pacing. Iit may be a marker for a serious underlying condition, such as digitalis toxicity, postcardiac surgery, hypokalemia, or myocardial ischemia. Other associated conditions include chronic obstructive lung disease with hypoxia, and inflammatory myocarditis.There is commonly one-to-one AV association. In some cases, particularly in the setting of digitalis toxicity, anterograde AV-nodal Wenckebach conduction block may be observed" (Blomström-Lundqvist *et al.* 2003)

Orthodromic Atrio-ventricular Reciprocating Tachycardia (OAVRT)

"A re-entrant arrhythmia whose circuit involves the atrium, the AV node, the ventricles, and one or more accessory AV connections. AVRT can be classified as orthodromic AVRT, in which conduction through the AP occurs from the ventricle to the atrium" (Buxton *et al.* 2006).

"Typical accessory pathways are extra nodal pathways that connect the myocardium of the atrium and the ventricle across the AV groove. Delta waves on ECG present in 0.15 to 0.25% of the general population. Pathway conduction may be intermittent, decremental or nondecremental); and capable of anterograde conduction, retrograde conduction, or both. 8% of accessory pathways display decremental anterograde or retrograde conduction. Accessory pathways capable of only retrograde conduction are referred to as "concealed". (Blomström-Lundqvist *et al.* 2003)

Appendix E Rhythm definitions (continued)

Permanent Junctional Reciprocating tachycardia (PJRT)

"Episodes of narrow QRS tachycardia alternating with brief periods of sinus rhythm. During sinus rhythm, the ECG is normal. During tachycardia, negative P waves are typically present in leads II, III and aVF and usually in V4 to V6, with a RP > PR interval. Anterograde conduction is through the AV node, retrograde conduction through an accessory pathway with slow and decremental conduction. Commonly incessant from birth or infancy. Persistence over a long period may lead to tachycardia-induced cardiomyopathy" (Gaita *et al.* 1995).

Inappropriate sinus tachycardia (IST)

"Increase in sinus rate unrelated to, or out of proportion with, the level of physical, emotional, pathological, or pharmacological stress. Can be persistent or intermittent/paroxysmal" (Buxton *et al.* 2006)

"Persistent increase in resting heart rate or sinus rate unrelated to, or out of proportion with, the level of physical, emotional, pathological, or pharmacologic stress" (Blomström-Lundqvist *et al.* 2003)

Focal atrial tachycardia (AT)

"A usually regular cardiac arrhythmia arising from the atrium with a rate >100 bpm (cycle length <600 ms). atrial activation from atrial areas with centrifugal spread, with rates usually between 100 to 250 bpm (rarely at 300 bpm). They may arise from right or left atrial sites" (Buxton *et al.* 2006)

"Regular atrial activation from atrial areas with centrifugal spread. Uusually atrial rates between 100 to 250 bpm and rarely at 300 bpm. Neither the sinus nor the AV node plays a role in the initiation or perpetuation of the tachycardia (Blomström-Lundqvist *et al.* 2003).Regular atrial rhythms at a constant rate ≥100 beats/min originating outside the sinus node region. Atrial tachycardia CL usually is ≥250msec, but it can be as short as ≤200 msec" (Saoudi *et al.* 2001).

Macro-Reentrant Atrial tachycardia (MRAT)

"Arising in the atrium which has a regular rate typically between 250 and 350 bpm (cycle length 240-170 ms) in the absence of antiarrhythmic drugs" (Buxton *et al.* 2006)

"Organized atrial rhythm with a rate typically between 250 and 350 bpm. Includes tachycardias using a variety of re-entry circuits, often occupying large areas of the atrium. The classic type is dependent on the cavotricuspid isthmus (CTI). Circuits that do not use the CTI are less common. Most are related to an atrial scar that creates conduction block and a central obstacle for re-entry. Prior cardiac surgery involving the

atrium, such as repair of congenital heart disease, mitral valve surgery, or the atrial maze procedure, is a common cause "(Blomström-Lundqvist *et al.* 2003).

"Typical atrial flutter, the most common macro-re-entrant atrial tachycardia, usually has a CL between 190 and 250 msec, with ≤2% cycle-to-cycle variation" (Saoudi *et al.* 2001)

Multifocal Atrial Tachycardia (MAT)

"Iirregular tachycardia characterized by three or more P wave morphologies at different rates. Always irregular and frequently confused with AF, but rate is usually not excessively rapid. Most commonly associated with underlying pulmonary disease but may result from metabolic or electrolyte derangements." (Blomström-Lundqvist *et al.* 2003)

Atrial fibrillation (AF)

"A cardiac arrhythmia arising from the atrium with an atrial rate >300 bpm and an irregularly irregular ventricular response in the presence of conduction (Buxton *et al.* 2006). Supraventricular tachyarrhythmia characterized by uncoordinated atrial activation with consequent deterioration of atrial mechanical function. AF is characterized by the replacement of consistent P waves by rapid oscillations or fibrillatory waves that vary in amplitude, shape, and timing, associated with an irregular, frequently rapid ventricular response when atrio-ventricular (AV) conduction is intact. The ventricular response to AF depends on electrophysiological (EP) properties of the AV node and other conducting tissues, the level of vagal and sympathetic tone, the presence or absence of accessory conduction pathways, and the action of drugs. Regular cardiac cycles (R-R intervals) are possible in the presence of AV block or ventricular or AV junctional tachycardia. In patients with implanted pacemakers, diagnosis of AF may require temporary inhibition of the pacemaker to expose atrial fibrillatory activity (4). A rapid, irregular, sustained, wide-QRS-complex tachycardia strongly suggests AF with conduction over an accessory pathway or AF with underlying bundle-branch block. Extremely rapid rates (over 200 beats per minute) suggest the presence of an accessory pathway or ventricular tachycardia" (Fuster *et al.* 2011).

Broad Complex Tachycardia (BCT)

QRS >120 ms

"Differentiate between SVT and ventricular tachycardia (VT). Stable vital signs during tachycardias are not helpful for distinguishing SVT from VT. If SVT cannot be proven, then treat as VT. Dissociation with a ventricular rate faster than the atrial rate generally proves the diagnosis of VT. Fusion complexes represent a merger between conducted supraventricular impulses and ventricular depolarization occurring during AV

dissociation. These complexes are pathognomonic of VT. Retrograde VA block may be present. Demonstration that P waves are not necessary for tachycardia maintenance strongly suggests VT" (Blomström-Lundqvist *et al.* 2003).

"Algorithm for diagnosis of a tachycardia with a widened QRS complex. QRS complex >0.12 second. When an RS complex cannot be identified in any precordial lead, the diagnosis of ventricular tachycardia (VT) is made. If an RS complex is present in one or more precordial leads, the longest RS interval is measured. If the RS interval is longer than 100 msec, the diagnosis of VT is made. If shorter than 100 msec, the next step of the algorithm is considered: whether atrio-ventricular dissociation is present. Ifpresent, the diagnosis of VT is made. If absent, the morphology criteria for VT are analyzed in leads V, and V6. If both leads fulfill the criteria for VT, the diagnosis of VT is made. If not, the diagnosis of supraventricular tachycardia (SVT) with aberrant conduction is made by exclusion of VT" (Brugada *et al.* 1991).

Supraventricular Tachycardia with Aberration

"Typical LBBB or RBBB precordial ECG pattern (Blomström-Lundqvist *et al*. 2003)
The diagnosis of SVT with aberrant conduction is made by exclusion of VT" (Brugada *et al.* 1991).

Antidromic Atrio-ventricular Reciprocating Tachycardia (AAVRT)

"AVRT can be classified as antedromic AVRT, in which conduction through the AP occurs from the atrium to the ventricle (Buxton *et al.* 2006). Antidromic AVRT occurs in only 5 to 10% of patients with WPW syndrome" (Blomström-Lundqvist *et al*. 2003)

Monomorphic Ventricular Tachycardia (VT)

"VT is 3 or more consecutive complexes from the ventricles at >100 bpm (cycle length: <600 ms). Sustained: VT >30 s in duration or requiring termination due to hemodynamic compromise in <30 s. Nonsustained/unsustained: 3 or more beats in duration, terminating spontaneously in <30 s. Narrow complex VT with a QRS duration shorter than 120 ms" (Buxton *et al*. 2006).

"If RS complex is not present in any precordial lead, the diagnosis of VT is made. If an RS complex is present, the longest RS interval in the precordial leads is measured. If the RS interval is longer than 100 msec, the diagnosis of VT is made. If an RS complex is present and the longest RS interval in the precordial leads is 100 msec or less with atrio-ventricular dissociation the diagnosis of VT is made. If an RS complex is present, the longest RS interval in the precordial leads is measured. When the RS interval is 100 msec or less and atrio-ventricular dissociation is not evident, the morphology criteria are analyzed in leads V, and V6. If both leads have morphology compatible with the diagnosis of VT, the diagnosis of VT is made "(Brugada *et al*. 1991)

Appendix E Rhythm definitions (continued)

Polymorphic Ventricular Tachycardia (PVT)

"VT with a changing or multiform QRS morphology at cycle length >180 ms. Catecholaminergic: polymorphic VT associated with syncope and/or cardiac arrest triggered by emotion or exercise in patients whose baseline ECG is normal "(Buxton *et al*. 2006)

Idiopathic Fascicular Ventricular Tachycardia (IFVT)

"A tachycardia that emanates from or requires participation of the distal fascicles of right or left bundle branches" (Buxton *et al.* 2006)

Outflow Tract Ventricular Tachycardia (OTVT)

"Focal VT emanating from the right or left ventricular outflow tract unrelated to structural heart disease" (Buxton *et al.* 2006).

Ventricular Fibrillation (VF)

"Rapid, usually more than 300 bpm (cycle length: ≤180 ms), grossly irregular ventricular rhythm with marked variability in QRS cycle length, morphology, and amplitude" (Buxton *et al.* 2006)

Appendix F Pooled features

Arrhythmia Related Symptoms

fatigue

chest discomfort

dyspnoea

light-headedness

syncope

pre-syncope

palpitations

polyuria

cardiac arrest

duration, frequency, onset of episodes

termination by vagal manoeuvres

vagally mediated AF during sleep/ after large meal

Predisposing/ Precipitating Factors

antiarrhythmic drugs

nicotine

alcohol

caffeine

premenstrual/ menstrual

lack of sleep

Stress

physical stress

mental stress

emotional stress

pathological stress (infection)

pharmacological stress

Cardiac or Pulmonary Disease

pulmonary disease

structural heart disease

atherosclerotic heart diseases

inflammatory myocarditis

post-cardiac surgery

valvular disease

Appendix F Pooled features (continued)

cardiomyopathy
cerebrovascular disease
hypertensive heart disease.
congestive heart failure
congenital heart disease
other conditions (ie, sarcoidosis, tuberculosis)

Metabolic imbalance/ disease
metabolic derangements
electrolyte derangements
digitalis toxicity
thyroid disease
anaemia
hypovolaemia

Physical Signs
irregular pulse
irregular jugular venous pulsations
regular, rapid jugular pulse oscillations
irregular cannon A waves
irregular variation in S1 intensity
S4 absent (heard during sinus rhythm)
cardiogenic shock

Haemodynamic Status
blood pressure
haemodynamic stability
cardiac arrest

ECG
pauses
heart rate
regularity
“warm-up” “cool-down” patterns
Brugada syndrome
no P waves

Appendix F Pooled features (continued)

non-conducted P wave
P rate
P axis
P duration
P wave morphology
negative P wave
PP interval
PR interval
QRS axis
QRS duration
QRS morphology
RR interval stability and onset
P:R ratio
P/R relationship
QRS amplitude
RR interval
QT_c interval
RP interval
F waves
F amplitude
F shape
F timing
fusion complexes
an RS complex
RS interval
LBBB or RBBB pattern
AV conduction pattern
AV association/dissociation
delta wave
3 consecutive complexes
sustained VT >30 s in duration

Response to Pacing
terminated by pacing

Appendix F Pooled features (continued)

Sensors

body motion accelerometry

heart rate variability

heart rate turbulence

QT dispersion

T wave alternans

peak endocardial acceleration blood temperature

blood pressure

impedance respiration

impedance cardiac function

Appendix G STARD checklist

TITLE/ABSTRACT/KEYWORDS

1. Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').

INTRODUCTION

2. State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.

METHODS

Participants

3. Describe the study population: The inclusion and exclusion criteria, setting and locations where the data were collected.
4. Describe participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the (evaluated) index tests or the (golden) reference standard?
5. Describe participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.
6. Describe data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?

Test methods

7. Describe the reference standard and its rationale.
8. Describe technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.
9. Describe definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.
10. Describe the number, training and expertise of the persons executing and reading the index tests and the reference standard.
11. Describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.

Statistical methods

12. Describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).
13. Describe methods for calculating test reproducibility, if done.

Participants

14. Report when study was done, including beginning and ending dates of recruitment.

Appendix G STARD checklist (continued)

RESULTS

15. Report clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, co morbidity, current treatments, recruitment centers).

16. Report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).

Test results

17. Report time interval from the index tests to the reference standard, and any treatment administered between.

18. Report distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.

19. Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.

20. Report any adverse events from performing the index tests or the reference standard. *Estimates*

21. Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).

22. Report how indeterminate results, missing responses and outliers of the index tests were handled.

23. Report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.

24. Report estimates of test reproducibility, if done.

DISCUSSION

25. Discuss the clinical applicability of the study findings.

Appendix H Letter confirming ethics approval

NHS

National Research Ethics Service

St Thomas' Hospital Research Ethics Committee

South London REC Office 3
Ethics Committee Office
Governors' Hall Suite,
Ground Floor South Wing
St Thomas' Hospital
London
SE1 7EH

Telephone: 0207 188 2257
Facsimile: 0207 188 2258

30 January 2008

Mr Julian Bostock
Technical Head of Invasive cardiology
Cardiothoracic Centre
St. Thomas' Hospital
London
SE1 7EH

Dear Mr Bostock

Full title of study: **Creation of an Intracardiac Electrogram and Physiological Parameter Library for Use in Cardiac Arrhythmia Research.**

REC reference number: **07/H0802/119**

Thank you for your letter of 27 January 2008, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Ethical review of research sites

The Committee has designated this study as exempt from site-specific assessment . There is no requirement for [other] Local Research Ethics Committees to be informed or for site-specific assessment to be carried out at each site.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Appendix H Letter confirming ethics approval (continued)

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Application		09 October 2007
Investigator CV	2 - January 2007	09 October 2007
Protocol	1	09 October 2007
Covering Letter		09 October 2007
Letter of invitation to participant		
GP/Consultant Information Sheets		09 October 2007
Participant Information Sheet: Parents/Guardians of Children under 16	1	05 October 2007
Participant Information Sheet: Adults	1	05 October 2007
Participant Information Sheet: Young People 12-16 yrs	2	27 January 2008
Participant Information Sheet: Young People 8-11 yrs	2	27 January 2008
Participant Information Sheet: Young People 4 - 7 yrs	2	27 January 2008
Participant Consent Form: Assent form for Children	1	05 October 2007
Participant Consent Form: Parents/Guardians	1	05 October 2007
Participant Consent Form	1	05 October 2007
Response to Request for Further Information		27 January 2008
CV for Educational Supervisor - Dr Peter Weller		09 October 2007

R&D approval

All researchers and research collaborators who will be participating in the research at NHS sites should apply for R&D approval from the relevant care organisation, if they have not yet done so. R&D approval is required, whether or not the study is exempt from SSA. You should advise researchers and local collaborators accordingly.

Guidance on applying for R&D approval is available from http://www.rdforum.nhs.uk/rdform.htm.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Website > After Review

Here you will find links to the following

a) Providing feedback. You are invited to give your view of the service that you have received from the National Research Ethics Service on the application procedure. If you wish to make your views known please use the feedback form available on the website.

Appendix H Letter confirming ethics approval (continued)

b) Progress Reports. Please refer to the attached Standard conditions of approval by Research Ethics Committees.
c) Safety Reports. Please refer to the attached Standard conditions of approval by Research Ethics Committees.
d) Amendments. Please refer to the attached Standard conditions of approval by Research Ethics Committees.
e) End of Study/Project. Please refer to the attached Standard conditions of approval by Research Ethics Committees.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nationalres.org.uk .

07/H0802/119	Please quote this number on all correspondence

With the Committee's best wishes for the success of this project

Yours sincerely

[redacted]

Dr Adrian Williams
Chair

Email: [redacted]

Enclosures: *Standard approval conditions*

Copy to: R & D Office, Guy's & St Thomas' NHS Foundation Trust

Appendix H Letter confirming ethics approval (continued)

RESEARCH IN HUMAN SUBJECTS OTHER THAN CLINICAL TRIALS OF INVESTIGATIONAL MEDICINAL PRODUCTS

Standard conditions of approval by Research Ethics Committees

1. Further communications with the Research Ethics Committee

1.1 Further communications during the research with the Research Ethics Committee that gave the favourable ethical opinion (hereafter referred to in this document as "the Committee") are the personal responsibility of the Chief Investigator.

2. Commencement of the research

2.1 It is assumed that the research will commence within 12 months of the date of the favourable ethical opinion.

2.2 In the case of research requiring site-specific assessment (SSA) the research may not commence at any site until the Committee has notified the Chief Investigator that the favourable ethical opinion is extended to the site.

2.3 The research may not commence at any NHS site until the local Principal Investigator (PI) or research collaborator has obtained research governance approval from the relevant NHS care organisation.

2.4 Should the research not commence within 12 months, the Chief Investigator should give a written explanation for the delay. It is open to the Committee to allow a further period of 12 months within which the research must commence.

2.5 Should the research not commence within 24 months, the favourable opinion will be suspended and the application would need to be re-submitted for ethical review.

3. Duration of ethical approval

3.1 The favourable opinion for the research generally applies for the duration of the research. If it is proposed to extend the duration of the study as specified in the application form, the Committee should be notified.

4. Progress reports

4.1 Research Ethics Committees are required to keep a favourable opinion under review in the light of progress reports and any developments in the study. The Chief Investigator should submit a progress report to the Committee 12 months after the date on which the favourable opinion was given. Annual progress reports should be submitted thereafter.

4.2 Progress reports should be in the format prescribed by NRES and published on the website (see http://www.nres.npsa.nhs.uk/applicants/review/after/progress.htm#submission)

SOPs version 3.0 dated June 2005: v3.1 minor editing to NRES name change 1 April 07
SL-AC2 Approval conditions (research other than CTIMP)

4.3 The Chief Investigator may be requested to attend a meeting of the Committee or Sub-Committee to discuss the progress of the research.

5. Amendments

5.1 If it is proposed to make a substantial amendment to the research, the Chief Investigator should submit a notice of amendment to the Committee.

5.2 A substantial amendment is any amendment to the terms of the application for ethical review, or to the protocol or other supporting documentation approved by the Committee, that is likely to affect to a significant degree:

(a) the safety or physical or mental integrity of the trial participants

(b) the scientific value of the trial

(c) the conduct or management of the trial.

5.3 Notices of amendment should be in the format prescribed by NRES and published on the website, and should be personally signed by the Chief Investigator.

5.4 A substantial amendment should not be implemented until a favourable ethical opinion has been given by the Committee, unless the changes to the research are urgent safety measures (see section 7). The Committee is required to give an opinion within 35 days of the date of receiving a valid notice of amendment.

5.5 Amendments that are not substantial amendments ("minor amendments") may be made at any time and do not need to be notified to the Committee.

6. Changes to sites *(studies requiring site-specific assessment only)*

6.1 Where it is proposed to include a new site in the research, there is no requirement to submit a notice of amendment form to the Committee. The Site-Specific Information (SSI) form of the application form together with the local Principal Investigator's CV should be submitted to the relevant REC for site-specific assessment (SSA).

6.2 Similarly, where it is proposed to make important changes in the management of a site (in particular, the appointment of a new PI), a notice of amendment form is not required. A revised SSI form for the site (together with the CV for the new PI if applicable) should be submitted to the relevant REC for SSA.

6.3 The relevant REC will notify the Committee whether there is any objection to the new site or Principal Investigator. The Committee will notify the Chief Investigator of its opinion within 35 days of receipt of the valid application for SSA.

6.4 For studies designated by the Committee as exempt from SSA, there is no requirement to notify the Committee of the inclusion of new sites.

7. Urgent safety measures

7.1 The sponsor or the Chief Investigator, or the local Principal Investigator at a trial site, may take appropriate urgent safety measures in order to protect research participants against any immediate hazard to their health or safety.

7.2 The Committee must be notified within three days that such measures have been taken, the reasons why and the plan for further action.

8. Serious Adverse Events

8.1 A Serious Adverse Event (SAE) is an untoward occurrence that:

(a) results in death
(b) is life-threatening
(c) requires hospitalisation or prolongation of existing hospitalisation
(d) results in persistent or significant disability or incapacity
(e) consists of a congenital anomaly or birth defect
(f) is otherwise considered medically significant by the investigator.

8.2 A SAE occurring to a research participant should be reported to the Committee where in the opinion of the Chief Investigator the event was related to administration of any of the research procedures, and was an unexpected occurrence.

8.3 Reports of SAEs should be provided to the Committee within 15 days of the Chief Investigator becoming aware of the event, in the format prescribed by NRES and published on the website.

8.4 The Chief Investigator may be requested to attend a meeting of the Committee or Sub-Committee to discuss any concerns about the health or safety of research subjects.

8.5 Reports should not be sent to other RECs in the case of multi-site studies.

9. Conclusion or early termination of the research

9.1 The Chief Investigator should notify the Committee in writing that the research has ended within 90 days of its conclusion. The conclusion of the research is defined as the final date or event specified in the protocol, not the completion of data analysis or publication of the results.

9.2 If the research is terminated early, the Chief Investigator should notify the Committee within 15 days of the date of termination. An explanation of the reasons for early termination should be given.

9.3 Reports of conclusion or early termination should be submitted in the form prescribed by NRES and published on the website.

10. Final report

10.1 A summary of the final report on the research should be provided to the Committee within 12 months of the conclusion of the study. This should include information on whether the study achieved its objectives, the main findings, and arrangements for publication or dissemination of the research including any feedback to participants.

11. Review of ethical opinion

11.1 The Committee may review its opinion at any time in the light of any relevant information it receives.

11.2 The Chief Investigator may at any time request that the Committee reviews its opinion, or seek advice from the Committee on any ethical issue relating to the research.

12. Breach of approval conditions

12.1 Failure to comply with these conditions may lead to suspension or termination of the favourable ethical opinion by the Committee.

Appendix J Procedure worksheet

REC Ref: 07/H0802/119 **R&D No: RJ109/N04**

Creation of an Intracardiac Electrogram and Physiological Parameter Library

Study ID: _____ **Procedure date: _____**

DOB: _____dd/mm/yy Sex (M/F): _____

Height: _____cm Weight: ___kg Build: _____ slim/average/heavy)

Initial HR: _____bpm Initial BP: _____/___ mmHg Initial Temp:_____^{o}C

Initial Rhythm: _____Temp time: _____

Clinical history:

Known cardiac diagnosis (such as ARVC, WPW, Congenital HD and Valve Disease):___

_____

Other known conditions (e.g.Carotid Sinus Hypersensitivity; Metabolic diseases): _____

_____

Family Hx: ____Smoking: _____Hypertension: _____DM: _____

Hypercholesterol: _____Coronary Disease:_____

Extremes of lifestyle (such as sedentary or athletic): _____

_____

Caffeine_____Alcohol_____Recreational drugs:_____

_____

Current Prescribed Medications:_____ _____

_____

_____ _____

_____ _____

Symptoms:_____

Heart sounds (as noted):_____

_____

Echo Ejection fraction:_____% Date: _____

EchoSummary: _____

Holter Monitor: _____ Date:_____

Thyroid function: _____TSH:_____Date:_____

Notes: _____

_____

ESI time: _____

Sensis time: _____

XP time: _____

Appendix K Feature coding scheme

Arrhythmia Related Symptoms	**Score**
Fatigue	0 or 1
chest discomfort	0 or 1
dyspnoea	0 or 1
polyuria	0 or 1
syncope, light-headedness, pre-syncope	0-2
palpitations	0 or 1
cardiac arrest	0 or 1
vagally mediated AF during sleep/ after large meal	0 or 1
duration, frequency, onset, termination by vagal manoeuvres	0-8
Physical Signs	
irregular pulse	0 or 1
irregular jugular venous pulsations	0 or 1
irregular variation in S1 intensity, S4 absent during arrhythmia	0 or 1
cardiogenic shock	0 or 1
Cardiac or Pulmonary Disease	0-3
Predisposing/ Precipitating Factors	
antiarrhythmic drugs	0-3
nicotine	0-3
alcohol	0-3
caffeine	0-3
premenstrual/ menstrual	1
lack of sleep	1
Metabolic imbalance/ disease	1

Appendix L List of Matlab scripts written for this study.

Link to Matlab Scripts:

Link to Physiological Signals:

Convert_Classes_to_Binary.m

DecisionTree_Train.m

FeatureCheck.m

ImpedanceCardiac.m

ImpedanceResp.m

InferenceEngineRhythms.m

Iteration2_FeatureProcessing.m

Iteration3_FeatureSelection.m

Missing_Data.m

NaiveBayes_Train.m

NN_Train.m

NNRhythms.m

ProductionSystem.m

QTInterval.m

Script_001.m

ScriptFeatures.m

ScriptFeaturesNoResp.m

SramekBernstein.m

SVM_train.m

Undersampling.m

Appendix M. Formulae

Journal Citations per Paper per Year $c' = (C \,/\, (2012 - (year\ of\ publication))$ (2.1)

Standard Error $\sigma_X = \sqrt{\frac{X(1-X)}{n}}$ (3.1)

Binomial Confidence Interval (95%) $CI_{95} = 1.96\sqrt{\frac{X(1-X)}{n}}$ (3.2)

Bazett's Formula $QT_c = \frac{QT}{\sqrt{RR}}$ (5.1)

Power Calculation $\tau = (1 - \beta)$ (6.1)

Vector Magnitude Units $VMU = \sqrt{(x^2 + y^2)}$ (6.2)

Total Thoracic Impedance Change $\Delta Z = Z_0 + \Delta Z_R + \Delta Z_C$ (6.3)

Sramek-Bernstein Equation $SV = \delta . \frac{(0.17H)^3}{4.2} . \frac{dZ/dt_{max}}{Z_0} . LVET$ (6.4)

Bour Calibration Equation $SVI_{cal} = k . \frac{dZ/dt_{max}}{(Z_{max} - Z_{\min)}} . W(TFIT_{cal})$ (6.5)

Bour Equation $SV = SVI = BSA . SVI_{cal} . \sqrt[3]{(dZ/dt_{max}/dZ/dt_{maxcal}) . TFIT_{cal}/TFIT)}$ (6.6)

Lozano Equation $RB_{int} = 1.233RC_{int} - 0.0032{RC_{int}}^2 - 31.59$ (6.7)

Arctan Trigonometric Function $\theta = arctan\left(\frac{S1_{net\ amplitude}}{S2_{net\ amplitude}}\right)$ (8.1)

Correct Classification Rate $CCR = \frac{TP+TN}{N}$ (10.1)

Classification Error $Error = \frac{FP+FN}{N} = 1 - CCR$ (10.2)

Sensitivity $Sensitivity\ (true\ positive\ rate, recall, power) = \frac{TP}{TP+FN}$ (10.3)

Specificity $Specificity\ (true\ negative\ rate) = \frac{TN}{TN+FP}$ (10.4)

Appendix M. Formulae (continued)

Cohen's Kappa	$\kappa = \frac{(P_o - P_e)}{(1 - P_e)}$	(10. 5)
Fisher's Exact Test, *P* value	$P = \frac{(TP+FP)!(FN+TN)!(TP+FN)!(FP+TN)!}{TP!FP!FN!TN!(TP+FP+FN+TN)!}$	(10.6)
Bonferroni Correction, *P* value	$P_{crit} = \frac{\alpha}{n}$	(10.7)
Prevalence	$Prevalence\ (\approx\ prior\ probability = P(H)) = \frac{TP+FN}{N}$	(10.8)
Positive Predictive Value	$PPV\ (precision) = \frac{TP}{TP+FP}$	(10.9)
Negative Predictive Value	$NPV = \frac{TN}{TN+FN}$	(10.10)
Diagnostic Odds Ratio	$OR = \frac{(TP/_{FN})}{(FP/_{TN})} = \frac{LR+}{LR-}$	(10.11)
Relative Risk	$RR = \frac{(TP/(TP+FP)}{FN/(FN+TN\})}$	(10.12)
Yules *Q*	$Yule's\ Q = \frac{((TP.TN)-(FP.FN))}{((TP.TN)+(FP.FN))} = \frac{OR-1}{OR+1}$	(10.13)
Pearson's Phi	$\phi = \frac{((FP.FN)-(TP.TN))}{\sqrt{(TP+FP)(FN+TN)(TP+FN)(FP+TN)}}$	(10.14)
F_1 Score	$F_1 = 2.\frac{PPV.Sensitivity}{PPV+Sensitivity}$	(10.15)
Type I Error	$Type\ I\ error\ (False\ positive\ rate) = \alpha = \frac{FP}{FP+TN}$	(10.16)
Type II Error	$Type\ II\ error\ (False\ negative\ rate) = \beta = \frac{FN}{FN+TP}$	(10.17)
Stroke Volume Index, % Change	$SVI\ change = 100.\frac{(SVI_0 - SVI_{-1})}{SVI_{-1}}$	(10.18)
Interval Onset	$Interval\ Onset = 100.\frac{(RR_{-1} - RR_0)}{RR_{-1}}$	(10.19)

Appendix M. Formulae (continued)

Mean RR interval (3 beats) $$\overline{RR} = \frac{1}{3} . \sum_{i=0}^{2} RR_i \quad (10.20)$$

Interval Stability

$$Interval\ Stability = \max(|RR_0 - \overline{RR}|, |RR_1 - \overline{RR}|, |RR_2 - \overline{RR}|) \quad (10.21)$$

Atrial Chamber of Origin

$$PRratio_0 \geq 1 \Rightarrow Atrial\ Chamber\ of\ Origin \quad (10.22)$$

Ventricular Chamber of Origin

$$PRratio_0 < 1 \Rightarrow Ventricular\ Chamber\ of\ Origin \quad (10.23)$$

References

Acharya, R., Kumar, A., Bhat, P. S., Lim, C. M., Iyengar, S. S., Kannathal, N., & Krishnan, S. M. (2004). Classification of cardiac abnormalities using heart rate signals. [Research Support, Non-U.S. Gov't]. *Med Biol Eng Comput, 42*(3), 288-293.

Acharya, U. R., Sankaranarayanan, M., Nayak, J., Xiang, C., & Tamura, T. (2008). Automatic identification of cardiac health using modeling techniques: A comparative study. *Information Sciences, 178*(23), 4571-4582

Aliot, E., Nitzsche, R., & Ripart, A. (2004). Arrhythmia detection by dual-chamber implantable cardioverter defibrillators. A review of current algorithms. [Review]. *Europace, 6*(4), 273-286.

Alt, E., Theres, H., Heinz, M., Matula, M., Thilo, R., & Blomer, H. (1988). A new rate-modulated pacemaker system optimized by combination of two sensors. *Pacing Clin Electrophysiol, 11*(8), 1119-1129.

Altman, D.G. (1991). Practical statistics for medical research. London: Chapman and Hall.

Anderson, R., Gardner, F. V., Honey, H., Noble, I. M., & Woodgate, D. W. (1968). Relation between metabolic acidosis and cardiac dysrhythmias in acute myocardial infarction. *Br Heart J, 30*(4), 493-496.

Andrassy, G., Szabo, A., Ferencz, G., Trummer, Z., Simon, E., & Tahy, A. (2007). Mental stress may induce QT-interval prolongation and T-wave notching. *Ann Noninvasive Electrocardiol, 12*(3), 251-259.

Andrikopoulos, G., Tzeis, S., Theodorakis, G., & Vardas, P. (2010). Monitoring capabilities of cardiac rhythm management devices. [Research Support, Non-U.S. Gov't Review]. *Europace, 12*(1), 17-23.

Anuradha, B., Reddy, V., & Veera, C. (2008). Cardiac arrhythmia classification using fuzzy classifiers. *Journal of Theoretical and Applied Information Technology, 4*(4), 353-359.

Asl, B. M., Setarehdan, S. K., & Mohebbi, M. (2008). Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artif Intell Med, 44*(1), 51-64.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A., & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review]. *Bioinformatics, 16*(5), 412-424.

Batchvarov. V. N., Ghuran, A., Smetana, P., Hnatkova, K., Harries, M., Dilaveris, P., Camm, A. J., Malik, M. (2002). QT-RR relationship in healthy subjects exhibits substantial intersubject variability and high intrasubject stability. *Am J Physiol Heart Circ Physiol*. Jun;282(6):H2356-63.

Bazett, H. C. (1920). "An analysis of the time-relations of electrocardiograms". *Heart* (7): 353–370

References (continued)

Beaulnes, A., & Day, M. (1957). Effect of temperature on arrhythmia in isolated rabbit artia. *J Physiol, 137*(1), 86-94.

Benditt, D. G., Mianulli, M., Fetter, J., Benson, D. W., Jr., Dunnigan, A., Molina, E., . . . Almquist, A. (1987). Single-chamber cardiac pacing with activity-initiated chronotropic response: evaluation by cardiopulmonary exercise testing. [Clinical Trial Comparative Study Randomized Controlled Trial Research Support, Non-U.S. Gov't]. *Circulation, 75*(1), 184-191.

Berggren, G., & Hohwu Christensen, E. (1950). Heart rate and body temperature as indices of metabolic rate during work. *Arbeitsphysiologie, 14*(3), 255-260.

Bernardi, L., Valle, F., Coco, M., Calciati, A., & Sleight, P. (1996). Physical activity influences heart rate variability and very-low-frequency components in Holter electrocardiograms. [Comparative Study]. *Cardiovasc Res, 32*(2), 234-237.

Bernstein, D. P. (1986). A new stroke volume equation for thoracic electrical bioimpedance: theory and rationale. *Crit Care Med, 14*(10), 904-909.

Billauer, E. (2005). peakdet: Peak detection using MATLAB Retrieved 10/8/2013, from http://billauer.co.il/peakdet.html

BioxBio.com. (2013). SCI Journal Impact Factor. Retrieved 30/7/2013, from http://www.bioxbio.com/if/

Bishop, C. M. (1995). *Neural networks for pattern recognition*: Oxford university press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*: Springer.

Blomstrom-Lundqvist, C., Scheinman, M. M., Aliot, E. M., Alpert, J. S., Calkins, H., Camm, A. J., . . . Trappe, H. J. (2003). ACC/AHA/ESC guidelines for the management of patients with supraventricular arrhythmias--executive summary. a report of the American college of cardiology/American heart association task force on practice guidelines and the European society of cardiology committee for practice guidelines (writing committee to develop guidelines for the management of patients with supraventricular arrhythmias) developed in collaboration with NASPE-Heart Rhythm Society. [Guideline Practice Guideline Review]. *J Am Coll Cardiol, 42*(8), 1493-1531.

Bonow, R. O., Ganiats, T. G., Beam, C. T., Blake, K., Casey, D. E., Jr., Goodlin, S. J., . . . Shahian, D. M. (2012). ACCF/AHA/AMA-PCPI 2011 performance measures for adults with heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures and the American Medical Association-Physician Consortium for Performance Improvement. [Practice Guideline]. *J Am Coll Cardiol, 59*(20), 1812-1832.

References (continued)

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . de Vet, H. C. (2003). Toward complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative. [Consensus Development Conference Research Support, Non-U.S. Gov't Review]. *Am J Clin Pathol, 119*(1), 18-22.

Bostock, J. (2004). *A New Diagnostic Algorithm for Implantable Cardioverter-Defibrillators - A Systems Approach.* M.Sc, City University, London.

Bostock, J., Weller, P., & Cooklin, M. (2010). *Selection and collection of multi parameter physiological data for cardiac rhythm diagnostic algorithm development.* Paper presented at the Journal of Physics: Conference Series.

Bour, J., & Kellett, J. (2008). Impedance cardiography: a rapid and cost-effective screening tool for cardiac disease. [Review]. *Eur J Intern Med, 19*(6), 399-405.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123-140.

Brüser, C., Diesel, J., Zink, M.D., Schauerte, P., Leonhardt, S. (2012). Automatic detection of atrial fibrillation in cardiac vibration signals. *IEEE Trans Inf Technol Biomed,* 2012 Oct 16 [Epub ahead of print].

Brignole, M., Alboni, P., Benditt, D., Bergfeldt, L., Blanc, J., Thomsen, P. B., . . . Kenny, R. (2001). Part 2. Diagnostic tests and treatment: summary of recommendations. *Europace, 3*(4), 261-268.

Brignole, M., Alboni, P., Benditt, D. G., Bergfeldt, L., Blanc, J.-J., Thomsen, P. E. B., . . . Janousek, J. (2004). Guidelines on management (diagnosis and treatment) of syncope–Update 2004 The task force on Syncope, European Society of Cardiology. *Eur Heart J, 25*(22), 2054-2072.

Brugada, P., Brugada, J., Mont, L., Smeets, J., & Andries, E. W. (1991). A new approach to the differential diagnosis of a regular tachycardia with a wide QRS complex. *Circulation, 83*(5), 1649-1659.

Buxton, A. E., Calkins, H., Callans, D. J., DiMarco, J. P., Fisher, J. D., Greene, H. L., . . . Redberg, R. F. (2006). ACC/AHA/HRS 2006 key data elements and definitions for electrophysiological studies and procedures: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (ACC/AHA/HRS Writing Committee to Develop Data Standards on Electrophysiology). [Consensus Development Conference]. *J Am Coll Cardiol, 48*(11), 2360-2396.

Cammilli, L., Alcidi, L., Shapland, E., & Obino, S. (1983). Results, problems and perspectives with the autoregulating pacemaker. *Pacing Clin Electrophysiol, 6*(2 Pt 2), 488-493.

Cardillo, G. (2009). Cohen's kappa. Retrieved 2/6/2013, from http://www.mathworks.co.uk/matlabcentral/fileexchange/15365-cohens-kappa

References (continued)

Casagrande, J. T., & Pike, M. C. (1978). An improved approximate formula for calculating sample sizes for comparing two binomial distributions. [Comparative Study Research Support, U.S. Gov't, P.H.S.]. *Biometrics, 34*(3), 483-486.

Ceylan, R., Özbay, Y., & Karlik, B. (2009). A novel approach for classification of ECG arrhythmias: Type-2 fuzzy clustering neural network. *Expert Systems with Applications*, 36(3), 6721-6726.

Charbonnier, F., & Tacker Jr, W. (1994). Algorithms for arrhythmia analysis in AEDs. *Defibrillation of the Heart, Tacker WA (Editor)*.

Charloux, A., Lonsdorfer-Wolf, E., Richard, R., Lampert, E., Oswald-Mammosser, M., Mettauer, B., . . . Lonsdorfer, J. (2000). A new impedance cardiograph device for the non-invasive evaluation of cardiac output at rest and during exercise: comparison with the "direct" Fick method. [Comparative Study]. *Eur J Appl Physiol, 82*(4), 313-320.

Chiarugi, F., Varanini, M., Cantini, F., Conforti, F., & Vrouchos, G. (2007). Noninvasive ECG as a tool for predicting termination of paroxysmal atrial fibrillation. [Research Support, Non-U.S. Gov't]. *IEEE Trans Biomed Eng, 54*(8), 1399-1406.

Chirife, R. (1988). Physiological principles of a new method for rate responsive pacing using the pre-ejection interval. *Pacing Clin Electrophysiol, 11*(11 Pt 1), 1545-1554.

Christov, I., & Bortolan, G. (2004). Ranking of pattern recognition parameters for premature ventricular contractions classification by neural networks. *Physiol Meas, 25*(5), 1281-1290.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disgreementor partial credit. *Psychological Bulletin, 70*(4), 213-220.

Coggins, R., Jabri, M., Flower, B., & Pickard, S. (1995). A hybrid analog and digital VLSI neural network for intracardiac morphology classification. *Solid-State Circuits, IEEE Journal of, 30*(5), 542-550.

Coleman, K. J., Saelens, B. E., Wiedrich-Smith, M. D., Finn, J. D., & Epstein, L. H. (1997). Relationships between TriTrac-R3D vectors, heart rate, and self-report in obese children. [Research Support, U.S. Gov't, P.H.S.]. *Med Sci Sports Exerc, 29*(11), 1535-1542.

Copie, X., Hnatkova, K., Staunton, A., Fei, L., Camm, A. J., & Malik, M. (1996). Predictive power of increased heart rate versus depressed left ventricular ejection fraction and heart rate variability for risk stratification after myocardial infarction. Results of a two-year follow-up study. [Comparative Study Research Support, Non-U.S. Gov't]. *J Am Coll Cardiol, 27*(2), 270-276.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 187-220.

References (continued)

Cox, J. L., Schuessler, R. B., D'Agostino Jr, H. J., Stone, C. M., Chang, B. C., Cain, M. E., ... & Boineau, J. P. (1991). The surgical treatment of atrial fibrillation. III. Development of a definitive surgical procedure. *The Journal of Thoracic and Cardiovascular Surgery*, *101*(4), 569-583.

Daubert, J. P., Zareba, W., Cannom, D. S., McNitt, S., Rosero, S. Z., Wang, P., . . . Wilber, D. J. (2008). Inappropriate Implantable Cardioverter-Defibrillator Shocks in MADIT II Frequency, Mechanisms, Predictors, and Survival Impact. *J Am Coll Cardiol*, *51*(14), 1357-1365.

de Chazal, P., O'Dwyer, M., & Reilly, R. B. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. [Comparative Study Evaluation Studies Research Support, Non-U.S. Gov't Validation Studies]. *IEEE Trans Biomed Eng*, *51*(7), 1196-1206.

de Chazal, P., & Reilly, R. B. (2006). A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. [Evaluation Studies Research Support, Non-U.S. Gov't]. *IEEE Trans Biomed Eng*, *53*(12 Pt 1), 2535-2543.

Debski, T. T., Zhang, Y., Jennings, J. R., & Kamarck, T. W. (1993). Stability of cardiac impedance measures: aortic opening (B-point) detection and scoring. [Research Support, U.S. Gov't, P.H.S.]. *Biol Psychol*, *36*(1-2), 63-74.

Delnoy, P. P., Marcelli, E., Oudeluttikhuis, H., Nicastia, D., Renesto, F., Cercenelli, L., & Plicchi, G. (2008). Validation of a peak endocardial acceleration-based algorithm to optimize cardiac resynchronization: early clinical results. [Research Support, Non-U.S. Gov't Validation Studies]. *Europace*, *10*(7), 801-808.

Donaldson, R. M., & Rickards, A. F. (1983). Rate responsive pacing using the evoked QT principle. A physiological alternative to atrial synchronous pacemakers. *Pacing Clin Electrophysiol*, *6*(6), 1344-1349.

Dupont, W. D., & Plummer Jr, W. D. (1990). Power and sample size calculations: a review and computer program. *Control Clin Trials*, *11*(2), 116-128.

Edwards, J. S. (1991). *Building knowledge-based systems: Towards a methodology*: Halsted Press.

Electrogram.com. (n.d.). Ann Arbor Electrogram Libraries. Retrieved 2/8/2013, from http://electrogram.com/browser.new/default.asp?volume=Volume%20One

Ellenbogen, K. A., Lu, B., Kapadia, K., Wood, M., & Valenta, H. (1990). Usefulness of right ventricular pulse pressure as a potential sensor for hemodynamically unstable ventricular tachycardia. [Research Support, Non-U.S. Gov't]. *Am J Cardiol*, *65*(16), 1105-1111.

References (continued)

Epstein, A. E., DiMarco, J. P., Ellenbogen, K. A., Estes, N. A., 3rd, Freedman, R. A., Gettes, L. S., . . . Yancy, C. W. (2008). ACC/AHA/HRS 2008 Guidelines for Device-Based Therapy of Cardiac Rhythm Abnormalities: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the ACC/AHA/NASPE 2002 Guideline Update for Implantation of Cardiac Pacemakers and Antiarrhythmia Devices) developed in collaboration with the American Association for Thoracic Surgery and Society of Thoracic Surgeons. [Practice Guideline]. *J Am Coll Cardiol, 51*(21), e1-62.

Erickson, B. J., & Bartholmai, B. (2002). Computer-aided detection and diagnosis at the start of the third millennium. *J Digit Imaging, 15*(2), 59-68.

Exarchos, T. P., Tsipouras, M. G., Exarchos, C. P., Papaloukas, C., Fotiadis, D. I., & Michalis, L. K. (2007). A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree. [Research Support, Non-U.S. Gov't]. *Artif Intell Med, 40*(3), 187-200.

Fearnot, N. E., & Evans, M. L. (1988). Heart rate correlation, response time and effect of previous exercise using an advanced pacing rate algorithm for temperature-based rate modulation. [Research Support, Non-U.S. Gov't]. *Pacing Clin Electrophysiol, 11*(11 Pt 2), 1846-1852.

Fearnot, N. E., Jolgren, D. L., Tacker, W. A., Nelson, J. P., & Geddes, L. A. (1984). Increasing cardiac rate by measurement of right ventricular temperature. [Research Support, Non-U.S. Gov't]. *Pacing Clin Electrophysiol, 7*(6 Pt 2), 1240-1245.

Feldman, G. (2007). Win32Pad - Free, small, fast and powerful text editor. Retrieved 10/8/2013, from http://www.gena01.com/win32pad/

Fleiss, J. L. (1981). Statistical methods for rates and proportions: Wiley (New York).

Flood, R., & Carson, E. (1988). Dealing with Complexity: An Introduction to the Theory and Practice of Systems Science: Plenum, New York.

Freund, Y., & Schapire, R. E. (1995). *A desicion-theoretic generalization of on-line learning and an application to boosting.* Paper presented at the Computational learning theory.

Fridericia, L. S. (1920). "The duration of systole in the electrocardiogram of normal subjects and of patients with heart disease". *Acta Medica Scandinavica* (53): 469–486.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The annals of statistics, 28*(2), 337-407.

Frohlig, G., Helwani, Z., Kusch, O., Berg, M., & Schieffer, H. (1999). Bipolar ventricular far-field signals in the atrium. [Comparative Study]. *Pacing Clin Electrophysiol, 22*(11), 1604-1613.

References (continued)

Fulbrook, P. (1993). Core temperature measurement in adults: a literature review. [Comparative Study Review]. *J Adv Nurs, 18*(9), 1451-1460.

Fuster, V., Ryden, L. E., Cannom, D. S., Crijns, H. J., Curtis, A. B., Ellenbogen, K. A., . . . Wann, L. S. (2011). 2011 ACCF/AHA/HRS focused updates incorporated into the ACC/AHA/ESC 2006 Guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines developed in partnership with the European Society of Cardiology and in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. [Practice Guideline]. *J Am Coll Cardiol, 57*(11), e101-198.

Gaita, F., Haissaguerre, M., Giustetto, C., Fischer, B., Riccardi, R., Richiardi, E., . . . Warin, J. F. (1995). Catheter ablation of permanent junctional reciprocating tachycardia with radiofrequency current. *J Am Coll Cardiol, 25*(3), 648-654.

Gaita, F., Giustetto, C., Bianchi, F., Wolpert, C., Schimpf, R., Riccardi, R., Grossi, S., Richiardi, E., Borggrefe, M. (2003). Short QT Syndrome: a familial cause of sudden death. *Circulation*. 26; 108(8):965-70.

Garnett, C. E., Zhu, H., Malik, M., Fossa, A. A., Zhang, J., Badilini, F., Li, J., Darpö, B., Sager, P., Rodriguez, I. (2012). Methodologies to characterize the QT/corrected QT interval in the presence of drug-induced heart rate changes or other autonomic effects. *Am Heart J*. 163(6):912-30.

Giacinto, G., & Roli, F. (2001). Design of effective neural network ensembles for image classification purposes. *Image and Vision Computing, 19*(9), 699-707.

Gini, C. (1912). Italian: Variabilità e Mutabilità (Variability and Mutability). Cuppini, Bologna

Gold, M. R., Shorofsky, S. R., Thompson, J. A., Kim, J., Schwartz, M., Bocek, J., . . . Lang, D. J. (2002). Advanced rhythm discrimination for implantable cardioverter defibrillators using electrogram vector timing and correlation. [Case Reports Research Support, Non-U.S. Gov't]. *J Cardiovasc Electrophysiol, 13*(11), 1092-1097.

Gold, M. R., Shih, H. T., Herre, J., Breiter, D., Zhang, Y., & Schwartz, M. (2007). Comparison of defibrillation efficacy and survival associated with right versus left pectoral placement for implantable defibrillators. [Comparative Study Randomized Controlled Trial]. *Am J Cardiol, 100*(2), 243-246.

Gold, M. R., Ahmad, S., Browne, K., Berg, K. C., Thackeray, L., & Berger, R. D. (2012a). Prospective comparison of discrimination algorithms to prevent inappropriate ICD therapy: primary results of the Rhythm ID Going Head to Head Trial. [Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't]. *Heart Rhythm, 9*(3), 370-377.

References (continued)

Gold, M. R., Theuns, D. A., Knight, B. P., Sturdivant, J. L., Sanghera, R., Ellenbogen, K. A., . . . Burke, M. C. (2012b). Head-to-head comparison of arrhythmia discrimination performance of subcutaneous and transvenous ICD arrhythmia detection algorithms: the START study. [Comparative Study Multicenter Study]. *J Cardiovasc Electrophysiol, 23*(4), 359-366.

Graham, J. W. (2009). Missing data analysis: making it work in the real world. [Review]. *Annu Rev Psychol, 60*, 549-576.

Hamilton, P. S., & Tompkins, W. J. (1986). Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. [Research Support, Non-U.S. Gov't]. *IEEE Trans Biomed Eng, 33*(12), 1157-1165.

Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 12*(10), 993-1001.

Hastie, T., Tibshirani, R., & Friedman, J. J. H. (2001). *The elements of statistical learning* (Vol. 1): Springer New York.

Hatley, D. J., Hruschka, P., & Pirbhai, I. A. (2000). *Process for system architecture and requirements engineering*: Dorset House Pub.

Heart Rhythm Society. (2010a). Heart Rhythm Society Abstract Archive 2006 - 2010 Abstracts2View™ Archive. Retrieved 30/4/2012, from http://www.abstracts2view.com/hrs/

Heart Rhythm Society. (2010b). Heart Rhythm 2010 Online Itinerary Planner. Retrieved 4/8/2013, from http://www.abstractsonline.com/plan/start.aspx?mkey={11A396D9-0C49-4809-A099-B5E5C705D19C}

Heart Rhythm Society. (2011). Heart Rhythm 2011 Online Itinerary Planner. Retrieved 4/8/2013, from http://www.abstractsonline.com/plan/start.aspx?mkey={29F06DE6-6576-4BFC-8E3B-A7C36155D815}

Heart Rhythm Society. (2012). Heart Rhythm 2012 Online Itinerary Planner Retrieved 4/8/2013, from http://www.abstractsonline.com/plan/start.aspx?mkey={BAEF2DB4-7615-4F2C-851A-E5D7461EBD4E

Heart Rhythm Society. (n.d.). Treatment Retrieved 4/8/2013, from http://www.hrsonline.org/Patient-Resources/Treatment#axzz2aYal4IR5

Hedman, A., Hjemdahl, P., Nordlander, R., & Astrom, H. (1990). Effects of mental and physical stress on central haemodynamics and cardiac sympathetic nerve activity during QT interval-sensing rate-responsive and fixed rate ventricular inhibited pacing. [Research Support, Non-U.S. Gov't]. *Eur Heart J, 11*(10), 903-915.

References (continued)

Hintringer, F., Schwarzacher, S., Eibl, G., & Pachinger, O. (2001). Inappropriate detection of supraventricular arrhythmias by implantable dual chamber defibrillators: a comparison of four different algorithms. [Comparative Study]. *Pacing Clin Electrophysiol, 24*(5), 835-841.

Hintringer, F., Deibl, M., Berger, T., Pachinger, O., & Roithinger, F. X. (2004). Comparison of the specificity of implantable dual chamber defibrillator detection algorithms. [Comparative Study]. *Pacing Clin Electrophysiol, 27*(7), 976-982.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on,* 14(1), 55-63.

Huikuri, H. V. (1997). Heart rate dynamics and vulnerability to ventricular tachyarrhythmias. [Research Support, Non-U.S. Gov't Review]. *Ann Med, 29*(4), 321-325.

Humen, D., Anderson, K., Brumwell, D., Huntley, S., & Klein, G. (1983). A pacemaker which automatically increases its rate with physical activity. *Cardiac pacing* (pp. 259-264): Springer.

Hurwitz, B. E., Shyu, L. Y., Lu, C. C., Reddy, S. P., Schneiderman, N., & Nagel, J. H. (1993). Signal fidelity requirements for deriving impedance cardiographic measures of cardiac function over a broad heart rate range. [Research Support, U.S. Gov't, P.H.S.]. *Biol Psychol, 36*(1-2), 3-21.

Jackson, L. R., 2nd, Daubert, J. P., & Thomas, K. L. (2012). Expanding the benefits of implantable cardioverter-defibrillator therapy: "is less more"? [Review]. *Prog Cardiovasc Dis, 54*(4), 372-378.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22*(1), 4-37.

Jardine, N., & van Rijsbergen, C.J. (1971). The use of hierarchic clusteringin information retrieval. *Information storage and retrieval, 7*(5), 217-240.

Jenkins, G. A. (1972). 4 The systems approach. *Systems behaviour*, 56.

Jenkins, G. M. (1969). The Systems Approach. *Journal of Systems Engineeering, 1*, 3-49.

Josephson, M. E. (2002). *Clinical cardiac electrophysiology: techniques and interpretations.* Lippincott Williams & Wilkins.

Jovic, A., & Bogunovic, N. (2011). Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features. *Artif Intell Med, 51*(3), 175-186.

Kalogeropoulos, D. A., Carson, E. R., & Collinson, P. O. (2003). Towards knowledge-based systems in clinical practice: Development of an integrated clinical information and knowledge management support system. *Comput Methods Programs Biomed, 72*(1), 65-80.

References (continued)

Kamalvand, K., Kotsakis, A., Tan, K., Bucknall, C., & Sulke, N. (1996). Evaluation of a new pacing algorithm to prevent rapid tracking of atrial tachyarrhythmias. *Pacing Clin Electrophysiol, 19*(11 Pt 2), 1714-1718.

Kannathal, N., Puthusserypady, S. K., Lim Choo, M., Rajendra Acharya, U., & Laxminarayan, S. (2005, 17-18 Jan. 2006). *Cardiac State Diagnosis using Adaptive Neuro-Fuzzy Technique.* Paper presented at the Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the.

Kaplan, E.L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J Am. Stat. Ass., 53*(282),457-481.

Kara, S., & Okandan, M. (2007). Atrial fibrillation classification with artificial neural networks. *Pattern Recognition, 40*(11), 2967-2973.

Kaszala, K., & Ellenbogen, K. A. (2010). Device sensing: sensors and algorithms for pacemakers and implantable cardioverter defibrillators. [Research Support, Non-U.S. Gov't Review]. *Circulation, 122*(13), 1328-1340.

Kavanagh, J. J., Barrett, R. S., & Morrison, S. (2004). Upper body accelerations during walking in healthy young and elderly men. [Clinical Trial Comparative Study]. *Gait Posture, 20*(3), 291-298.

Kinnen, E., & Kubicek, W. (1963). Thoracic Cage Impedance Measurements. Cancellation of Respiration Effects on Transthoracic Impedance. Techn Docum Rep Sam-Tdr-63-99. *Tech Doc Rep U S Air Force Syst Command Electron Syst Div, 94*, 1-5.

Kinnen, E., Kubicek, W., Hill, P., & Turton, G. (1964a). Thoracic Cage Impedance Measurements. Tissue Resistivity in Vivo and Transthoracic Impedance at 100 Kc. Techn Docum Rep No. Sam-Tdr-64-5. *AMD TR Rep*, 1-14.

Kinnen, E., Kubicek, W., & Patterson, R. (1964b). Thoracic Cage Impedance Measurements. Impedance Plethysmographic Determination of Cardiac Output (a Comparative Study). Techn Docum Rep No. Sam-Tdr-64-15. *AMD TR Rep*, 1-8.

Kleiger, R. E., Miller, J. P., Bigger, J. T., Jr., & Moss, A. J. (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. [Comparative Study Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *Am J Cardiol, 59*(4), 256-262.

Kleinert, M., Elmqvist, H., & Strandberg, H. (1979). Spectral properties of atrial and ventricular endocardial signals. *Pacing Clin Electrophysiol, 2*(1), 11-19.

Klingenheben, T., Sticherling, C., Skupin, M., & Hohnloser, S. H. (1998). Intracardiac QRS electrogram width--an arrhythmia detection feature for implantable cardioverter defibrillators: exercise induced variation as a base for device programming. [Comparative Study]. *Pacing Clin Electrophysiol, 21*(8), 1609-1617.

References (continued)

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection.* Paper presented at the IJCAI.

Kohavi, R., & Wolpert, D. H. (1996). *Bias plus variance decomposition for zero-one loss functions.* Paper presented at the ICML.

Kononenko, I. (1994). *Estimating Attributes: Analysis and Extensions of RELIEF*. Paper presented at the Machine Learning: ECML-94: European Conference on Machine Learning, Catania, Italy, April 6-8, 1994. Proceedings.

Korotkoff, N. (1905). To the question of methods of determining blood pressure. *Rep Imp Milit Med Acad, 2*(8), 365-367.

Korten, J. B., & Haddad, G. G. (1989). Respiratory waveform pattern recognition using digital techniques. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *Comput Biol Med, 19*(4), 207-217.

Kouakam, C., Kacet, S., Hazard, J. R., Ferraci, A., Mansour, H., Defaye, P., . . . Lambiez, M. (2004). Performance of a dual-chamber implantable defibrillator algorithm for discrimination of ventricular from supraventricular tachycardia. [Evaluation Studies Multicenter Study]. *Europace, 6*(1), 32-42.

Kremers, M., Herre, J., Aaron, M., Krueger, S., Gold , M., Zmijewski, M., & Whitman, T. (2012). *Chronic Evaluation of a New ICD Detection Algorithm Incorporating Hemodynamics in Heart Failure Patients.* Paper presented at the Heart Rhythm 2012: 33rd Annual Scientific Sessions, 9th–12th May 2012., Boston, MA.

Kubicek, W. G., Karnegis, J. N., Patterson, R. P., Witsoe, D. A., & Mattson, R. H. (1966). Development and evaluation of an impedance cardiac output system. [Comparative Study]. *Aerosp Med, 37*(12), 1208-1212.

Kuhlkamp, V., Dornberger, V., Mewis, C., Suchalla, R., Bosch, R. F., & Seipel, L. (1999). Clinical experience with the new detection algorithms for atrial fibrillation of a defibrillator with dual chamber sensing and pacing. [Clinical Trial Comparative Study]. *J Cardiovasc Electrophysiol, 10*(7), 905-915.

Kuncheva, L. I. (2004). Combining Pattern Classifiers. *Methods and Algorithms. Wiley, Chichester*.

Langendorf, R., Katz, L. N., & Simon, A. J. (1944). Reciprocal beating initiated by ventricular premature systoles. *British heart journal*, 6(1), 13.

Langenfeld, H., Krein, A., Kirstein, M., & Binner, L. (1998). Peak endocardial acceleration-based clinical testing of the "BEST" DDDR pacemaker. European PEA Clinical Investigation Group. [Clinical Trial Multicenter Study]. *Pacing Clin Electrophysiol, 21*(11 Pt 2), 2187-2191.

References (continued)

Lau, C. P., Antoniou, A., Ward, D. E., & Camm, A. J. (1988). Initial clinical experience with a minute ventilation sensing rate modulated pacemaker: improvements in exercise capacity and symptomatology. [Clinical Trial Randomized Controlled Trial]. *Pacing Clin Electrophysiol, 11*(11 Pt 2), 1815-1822.

Lau, C. P., Tai, Y. T., Lee, I. S., Erickson, M., & Yerich, C. (1995). Utility of an implantable right ventricular oxygen saturation-sensing pacemaker for ambulatory cardiopulmonary monitoring. *Chest, 107*(4), 1089-1094.

Leistner, H. L., Haddad, G. G., Epstein, R. A., Lai, T. L., Epstein, M. A., & Mellins, R. B. (1980). Heart rate and heart rate variability during sleep in aborted sudden infant death syndrome. [Comparative Study Research Support, U.S. Gov't, P.H.S.]. *J Pediatr, 97*(1), 51-55.

Lempel, A., & Ziv, J. (1976). On the complexity of finite sequences. *Information Theory, IEEE Transactions on, 22*(1), 75-81.

Leong, P. H., & Jabri, M. A. (1992). MATIC--an intracardiac tachycardia classification system. *Pacing Clin Electrophysiol, 15*(9), 1317-1331

Li, G. (2010). Fisher's Exact Test for beginners. Retrieved 2/6/2013, from http://www.mathworks.co.uk/matlabcentral/fileexchange/26499-fishers-exact-test-for-beginners

Lim, C. L., Byrne, C., & Lee, J. K. (2008). Human thermoregulation and measurement of body temperature in exercise and clinical settings. [Review]. *Ann Acad Med Singapore, 37*(4), 347-353.

Lindemans, F. W., Rankin, I. R., Murtaugh, R., & Chevalier, P. A. (1986). Clinical experience with an activity sensing pacemaker. [Clinical Trial]. *Pacing Clin Electrophysiol, 9*(6 Pt 2), 978-986.

Linh, T. H., Osowski, S., & Stodolski, M. (2003). On-line heart beat recognition using Hermite polynomials and neuro-fuzzy network. *Instrumentation and Measurement, IEEE Transactions on, 52*(4), 1224-1231.

Lozano, D. L., Norman, G., Knox, D., Wood, B. L., Miller, B. D., Emery, C. F., & Berntson, G. G. (2007). Where to B in dZ/dt. [Research Support, N.I.H., Extramural]. *Psychophysiology, 44*(1), 113-119.

Malik, M., Farrell, T., Cripps, T., & Camm, A. J. (1989). Heart rate variability in relation to prognosis after myocardial infarction: selection of optimal processing techniques. [Research Support, Non-U.S. Gov't]. *Eur Heart J, 10*(12), 1060-1074.

Malik, M., Acar, B., Gang, Y., Yap, Y. G., Hnatkova, K., & Camm, A. J. (2000). QT dispersion does not represent electrocardiographic interlead heterogeneity of ventricular repolarization. [Research Support, Non-U.S. Gov't]. *J Cardiovasc Electrophysiol, 11*(8), 835-843.

References (continued)

Malik, M., Hnatkova, K., Kowalski, D., Keirns, J. J., van Gelderen, E. M. (2012). Importance of subject-specific QT/RR curvatures in the design of individual heart rate corrections of the QT interval. *J Electrocardiol*. 45(6):571-81.

Malik, M., Hnatkova, K., Kowalski, D., Keirns, J. J., van Gelderen, E. M. (2013). QT/RR curvatures in healthy subjects: sex differences and covariates. *Am J Physiol Heart Circ Physiol*. 305(12):H1798-806.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychmetrika, 12*(2),153-157.

Medtronic Inc. (2010). Protecta XT CRT-D D354TRG - Clinician Manual (EE). Item Number: M930576A001 Retrieved 30/09/2012, from http://manuals.medtronic.com/wcm/groups/mdtcom_sg/@emanuals/@era/@crdm/documents/documents/wcm_prod052457.pdf

Meng, Q., Li, B., & Holstein, H. (2006). Recognition of human periodic movements from unstructured information using a motion-based frequency domain approach. *Image and Vision Computing, 24*(8), 795-809.

Minami, K., Nakajima, H., & Toyoshima, T. (1999). Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. *IEEE Trans Biomed Eng, 46*(2), 179-185.

Minhas, F. U., & Arif, M. (2008). Robust electrocardiogram (ECG) beat classification using discrete wavelet transform. [Evaluation Studies]. *Physiol Meas, 29*(5), 555-570.

Mitchell, T. M. (1997). Machine learning. WCB: McGraw-Hill Boston, MA:.

Mond, H., Strathmore, N., Kertes, P., Hunt, D., & Baker, G. (1988). Rate responsive pacing using a minute ventilation sensor. *Pacing Clin Electrophysiol, 11*(11 Pt 2), 1866-1874.

Moss, A. J., Greenberg, H., Case, R. B., Zareba, W., Hall, W. J., Brown, M. W., . . . Elkin, A. D. (2004). Long-term clinical course of patients after termination of ventricular tachyarrhythmia by an implanted defibrillator. [Evaluation Studies Research Support, Non-U.S. Gov't]. *Circulation, 110*(25), 3760-3765.

Moss, A. J., Schuger, C., Beck, C. A., Brown, M. W., Cannom, D. S., Daubert, J. P., ... & Zareba, W. (2012). Reduction in inappropriate therapy and mortality through ICD programming. *New England Journal of Medicine*, 367(24), 2275-2283.

Myers, G. H., Kresh, Y. M., & Parsonnet, V. (1978). Characteristics of intracardiac electrograms. [Research Support, U.S. Gov't, P.H.S.]. *Pacing Clin Electrophysiol, 1*(1), 90-103.

Naidu, S., Pandey, P. C., & Pandey, V. K. (2011). *Automatic detection of characteristic points in impedance cardiogram.* Paper presented at the Computing in Cardiology, 2011.

References (continued)

Nappholz, T. A., Lubin, M., & Valenta Jr, H. L. (1987). Metabolic-demand pacemaker and method of using the same to determine minute volume: U.S. Patent No. 4,702,253. Washington, DC: U.S. Patent and Trademark Office. National Institutes of Health (NIH). (2011). How Are Palpitations Diagnosed? Retrieved 4/8/2013, from http://www.nhlbi.nih.gov/health/health-topics/topics/hpl/diagnosis.html

Neuburger, C. (2012). Multi Class SVM. Retrieved 2/6/2013, from http://www.mathworks.co.uk/matlabcentral/fileexchange/39352-multi-class-svm

Nolan, J., Batin, P. D., Andrews, R., Lindsay, S. J., Brooksby, P., Mullen, M., . . . Fox, K. A. (1998). Prospective study of heart rate variability and mortality in chronic heart failure: results of the United Kingdom heart failure evaluation and assessment of risk trial (UK-heart). [Clinical Trial Multicenter Study Research Support, Non-U.S. Gov't]. *Circulation, 98*(15), 1510-1516.

Nunain, S., Roelke, M., Trouton, T., Osswald, S., Kim, Y., Sosa-Suarez, G., . . . Torchiana, D. (1995). Limitations and late complications of third-generation automatic cardioverter-defibrillators. *Circulation, 91*(8), 2204-2213.

Nyboer, J., Bagno, S., Barnett, A., & Halsey, R. (1940). Radiocardiograms: electrical impedance changes of the heart in relation to electrocardiograms and heart sounds. *J Clin Invest, 19*, 963.

Ohlsson, A., Kubo, S. H., Steinhaus, D., Connelly, D. T., Adler, S., Bitkover, C., . . . Bennett, T. (2001). Continuous ambulatory monitoring of absolute right ventricular pressure and mixed venous oxygen saturation in patients with heart failure using an implantable haemodynamic monitor: results of a 1 year multicentre feasibility study. [Clinical Trial Comparative Study Multicenter Study Research Support, Non-U.S. Gov't]. *Eur Heart J, 22*(11), 942-954.

Osowski, S., Hoai, L. T., & Markiewicz, T. (2004). Support vector machine-based expert system for reliable heartbeat recognition. [Comparative Study Evaluation Studies Research Support, Non-U.S. Gov't Validation Studies]. *IEEE Trans Biomed Eng, 51*(4), 582-589.

Osswald, S., Cron, T., Gradel, C., Hilti, P., Lippert, M., Strobel, J., . . . Pfisterer, M. (2000). Closed-loop stimulation using intracardiac impedance as a sensor principle: correlation of right ventricular dP/dtmax and intracardiac impedance during dobutamine stress test. *Pacing Clin Electrophysiol, 23*(10 Pt 1), 1502-1508.

Osterhues, H. H., Eggeling, T., Hoher, M., Weismuller, P., Kochs, M., & Hombach, V. (1993). Value of different non-invasive methods for the recognition of arrhythmogenic complications in high-risk patients with sustained ventricular tachycardia during programmed ventricular stimulation. *Eur Heart J, 14 Suppl E*, 40-45.

References (continued)

Osypka, M., & Schafer, E. (1998). *Impedance cardiography: advancement in system design.* Paper presented at the 10th International Conference on Electrical Bioimpedance, Barcelona, Spain, April.

Osypka, M. J., & Bernstein, D. P. (1999). Electrophysiologic principles and theory of stroke volume determination by thoracic electrical bioimpedance. [Review]. *AACN Clin Issues, 10*(3), 385-399.

Ovsyshcher, I., Guetta, V., Bondy, C., & Porath, A. (1992). First derivative of right ventricular pressure, dP/dt, as a sensor for a rate adaptive VVI pacemaker: initial experience. *Pacing Clin Electrophysiol, 15*(2), 211-218.

Owis, M. I., Abou-Zied, A. H., Youssef, A. B., & Kadah, Y. M. (2002). Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification. [Research Support, Non-U.S. Gov't]. *IEEE Trans Biomed Eng, 49*(7), 733-736.

Ozbay, Y., Ceylan, R., & Karlik, B. (2006). A fuzzy clustering neural network architecture for classification of ECG arrhythmias. [Comparative Study]. *Comput Biol Med, 36*(4), 376-388.

Pan, J. & Tompkins, W.J. (1985). A real-time QRS detection algorithm. *IEEE Trans Biomed Eng, 32*(3), 230-236.

Parker, B., Furman, S., & Escher, D. J. (1969). Input signals to pacemakers in a hospital environment. *Ann N Y Acad Sci, 167*(2), 823-834.

Parsonnet, V., Myers, G. H., & Kresh, Y. M. (1980). Characteristics of intracardiac electrograms II: Atrial endocardial electrograms. [Research Support, U.S. Gov't, P.H.S.]. *Pacing Clin Electrophysiol, 3*(4), 406-417.

Partridge, D., & Krzanowski, W. (1997). Software diversity: practical statistics for its measurement and exploitation. *Information and software technology, 39*(10), 707-717.

Patterson, R. P. (1989). Fundamentals of impedance cardiography. *IEEE Eng Med Biol Mag, 8*(1), 35-38.

Pearson, K. (1900).Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quatitaively measurable. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 195*, 1-47.

Phillips, J., & Ichinose, H. (1970). Clinical and pathologic studies in the hereditary syndrome of a long QT interval, syncopal spells and sudden death. *Chest*. 58(3):236-43.

Physionet. (2012). PhysioBank Archive Index-ECG databases. Retrieved 2/8/.2013, from http://www.physionet.org/physiobank/database/#ecg

Polat, K., & Güneş, S. (2007). Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine. *Applied Mathematics and Computation, 186*(1), 898-906.

References (continued)

Polat, K., Şahan, S., & Güneş, S. (2006). A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia. *Expert Systems with Applications, 31*(2), 264-269.

Pomerantz, M., Baumgartner, R., Lauridson, J., & Eiseman, B. (1969). Transthoracic electrical impedance for the early detection of pulmonary edema. *Surgery, 66*(1), 260-268.

Pomerantz, M., Delgado, F., & Eiseman, B. (1970). Clinical evaluation of transthoracic electrical impedance as a guide to intrathoracic fluid volumes. *Ann Surg, 171*(5), 686-694.

Poole, J. E., Johnson, G. W., Hellkamp, A. S., Anderson, J., Callans, D. J., Raitt, M. H., . . . Bardy, G. H. (2008). Prognostic importance of defibrillator shocks in patients with heart failure. [Multicenter Study Randomized Controlled Trial Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *N Engl J Med, 359*(10), 1009-1017.

Poon, K., Okin, P. M., Kligfield, P. (2005). Diagnostic performance of a computerbased ECG rhythm algorithm. *J Electrocardiol.* 38(3):235-238.

Pressman, R. S., & Ince, D. (2001). *Software engineering: a practitioner's approach* (5th ed.): McGraw-Hill New York.

Rajendra Acharya, U., Subbanna Bhat, P., Iyengar, S. S., Rao, A., & Dua, S. (2003). Classification of heart rate data using artificial neural network and fuzzy equivalence relation. *Pattern Recognition, 36*(1), 61-68.

Revishvili, A. (1999). Dual-chamber implantable cardioverter-defibrillator with active discrimination of supraventricular tachycardia. *Prog Biomed Res, 2*, 166-171.

Rickards, A. F., Donaldson, R. M., & Thalen, H. J. (1983). The use of QT interval to determine pacing rate: early clinical experience. *Pacing Clin Electrophysiol, 6*(2 Pt 2), 346-356.

Rickards, A. F., & Norman, J. (1981). Relation between QT interval and heart rate. New design of physiologically adaptive cardiac pacemaker. *Br Heart J, 45*(1), 56-61.

Rinaldi, C., Simon, R., Baszko, A., Bostock, J., Elliot, D., Bucknall, C., & Gill, J. (2004). A 17 year experience of inappropriate shock therapy in patients with implantable cardioverter-defibrillators: are we getting any better? *Heart, 90*(3), 330-331.

Robergs, R. A., Ghiasvand, F., & Parker, D. (2004). Biochemistry of exercise-induced metabolic acidosis. [Biography Historical Article Portraits Review]. *Am J Physiol Regul Integr Comp Physiol, 287*(3), R502-516.

Rodriguez, J., Goni, A., & Illarramendi, A. (2005). Real-time classification of ECGs on a PDA. [Evaluation Studies Research Support, Non-U.S. Gov't Validation Studies]. *IEEE Trans Inf Technol Biomed, 9*(1), 23-34.

Rossi, P., Plicchi, G., Canducci, G., Rognoni, G., & Aina, F. (1984). Respiration as a reliable physiological sensor for controlling cardiac pacing rate. *Br Heart J, 51*(1), 7-14.

References (continued)

Rowland, T. W., & Cunningham, L. N. (1997). Development of ventilatory responses to exercise in normal white children. A longitudinal study. *Chest, 111*(2), 327-332.

Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., & Edwards, D. D. (2003). *Artificial intelligence: a modern approach* (2nd ed.): Prentice Hall, New Jersey.

Ruta, D., & Gabrys, B. (2005). Classifier selection for majority voting. *Information fusion, 6*(1), 63-81.

Sacher, F., Probst, V., Iesaka, Y., Jacon, P., Laborderie, J., Mizon-Gérard, F., . . . Takahashi, Y. (2006). Outcome after implantation of a cardioverter-defibrillator in patients with Brugada syndrome a multicenter study. *Circulation, 114*(22), 2317-2324.

Salerno S.M., Alguire P.C., Waxman H.S. (2003a). Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence. *Ann Intern Med*. 138 :751-760.

Salerno S.M., Alguire P.C., Waxman H.S. (2003b). American College of Physicians. Training and competency evaluation for interpretation of 12-lead electrocardiograms: recommendations from the American College of Physicians. *Ann Intern Med*. 138:747-750.

Salo, R. W., Pederson, B. D., Olive, A. L., Lincoln, W. C., & Wallner, T. G. (1984). Continuous ventricular volume assessment for diagnosis and pacemaker control. *Pacing Clin Electrophysiol, 7*(6 Pt 2), 1267-1272.

Saoudi, N., Cosio, F., Waldo, A., Chen, S. A., Iesaka, Y., Lesh, M., . . . Schoels, W. (2001). A classification of atrial flutter and regular atrial tachycardia according to electrophysiological mechanisms and anatomical bases; a Statement from a Joint Expert Group from The Working Group of Arrhythmias of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. [Guideline Practice Guideline Review]. *Eur Heart J, 22*(14), 1162-1182.

Sarkar, S., Ritscher, D., & Mehra, R. (2008). A detector for a chronic implantable atrial tachyarrhythmia monitor. *IEEE Trans Biomed Eng, 55*(3), 1219-1224.

Schaldach, M. (1990). Automatic adjustment of pacing parameters based on intracardiac impedance measurements. [Review]. *Pacing Clin Electrophysiol, 13*(12 Pt 2), 1702-1710.

Schaldach, M., & Hutten, H. (1992). Intracardiac impedance to determine sympathetic activity in rate responsive pacing. [Clinical Trial Multicenter Study]. *Pacing Clin Electrophysiol, 15*(11 Pt 2), 1778-1786.

References (continued)

Schmidt, G., Malik, M., Barthel, P., Schneider, R., Ulm, K., Rolnitzky, L., . . . Schomig, A. (1999). Heart-rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction. [Research Support, Non-U.S. Gov't]. *Lancet, 353*(9162), 1390-1396.

Seidl, K., Rameken, M., Breunung, S., Senges, J., Jung, W., Andresen, D., . . . Klein, G. (2000). Diagnostic assessment of recurrent unexplained syncope with a new subcutaneously implantable loop recorder. *Europace, 2*(3), 256-262.

Shah, A. P., Rubin, S. A. (2007). Errors in the computerized electrocardiogram interpretation of cardiac rhythm. *J Electrocardiol.* 40(5):385-90.

Skalak, D. B. (1996). The sources of increased accuracy for two proposed boosting algorithms. Paper presented at the Proc. American Association for Artificial Intelligence, AAAI-96, Integrating Multiple Learned Models Workshop.

Sramek, B. B., Rose, D., & Miyamoto, A. (1983). Stroke volume equation with a linear base impedance model and its accuracy, as compared to thermodilution and magnetic flowmeter techniques in humans and animals. Paper presented at the 6th International Conference on Electrical Bioimpedance, Zadar, Yugoslavia.

Stangl, K., Wirtzfeld, A., Heinze, R., & Laule, M. (1988). First clinical experience with an oxygen saturation controlled pacemaker in man. *Pacing and Clinical Electrophysiology, 11*(11), 1882-1887.

Stangl, K., Wirtzfeld, A., Heinze, R., Laule, M., Seitz, K., & Göbl, G. (1988). A new multisensor pacing system using stroke volume, respiratory rate, mixed venous oxygen saturation, and temperature, right atrial pressure, right ventricular pressure, and dP/dt. *Pacing and Clinical Electrophysiology, 11*(6), 712-724.

Steele, B. G., Holt, L., Belza, B., Ferris, S., Lakshminaryan, S., & Buchner, D. M. (2000). Quantitating physical activity in COPD using a triaxial accelerometer. *CHEST Journal, 117*(5), 1359-1367.

Steinhaus, D., Reynolds, D. W., Gadler, F., Kay, G. N., Hess, M. F., & Bennett, T. (2005). Implant experience with an implantable hemodynamic monitor for the management of symptomatic heart failure. [Multicenter Study Research Support, Non-U.S. Gov't]. *Pacing Clin Electrophysiol, 28*(8), 747-753.

Stevenson, W. G., & Soejima, K. (2005). Recording techniques for clinical electrophysiology. *Journal of cardiovascular electrophysiology,* 16(9), 1017-1022.

Stewart, S., Murphy, N. F., Walker, A., McGuire, A., & McMurray, J. J. (2004). Cost of an emerging epidemic: an economic analysis of atrial fibrillation in the UK. [Research Support, Non-U.S. Gov't]. *Heart, 90*(3), 286-292.

References (continued)

Stratton, J. R., Levy, W. C., Cerqueira, M. D., Schwartz, R. S., & Abrass, I. B. (1994). Cardiovascular responses to exercise. Effects of aging and exercise training in healthy men. [Comparative Study Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S.]. *Circulation, 89*(4), 1648-1655.

Sugiura, T., Hirata, H., Harada, Y., & Kazui, T. (1998, 29 Oct-1 Nov 1998). Automatic discrimination of arrhythmia waveforms using fuzzy logic. Paper presented at the Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE.

Sugiura, T., Nakamura, Y., Mizushina, S., Hasegawa, T., Yoshimura, K., & Harada, Y. (1983). A temperature-sensitive cardiac pacemaker. *J Med Eng Technol, 7*(1), 21-23.

Summers, R. L., Shoemaker, W. C., Peacock, W. F., Ander, D. S., & Coleman, T. G. (2003). Bench to bedside: electrophysiologic and clinical principles of noninvasive hemodynamic monitoring using impedance cardiography. [Review]. *Acad Emerg Med, 10*(6), 669-680.

Surawicz, B., Childers, R., Deal, B. J., Gettes, L. S., Bailey, J. J., Gorgels, A., . . . Wellens, H. (2009). AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part III: intraventricular conduction disturbances: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society. Endorsed by the International Society for Computerized Electrocardiology. [Practice Guideline Review]. *J Am Coll Cardiol, 53*(11), 976-981.

Swerdlow, C. D., Brown, M. L., Lurie, K., Zhang, J., Wood, N. M., Olson, W. H., & Gillberg, J. M. (2002). Discrimination of ventricular tachycardia from supraventricular tachycardia by a downloaded wavelet-transform morphology algorithm: a paradigm for development of implantable cardioverter defibrillator detection algorithms. [Research Support, Non-U.S. Gov't]. *J Cardiovasc Electrophysiol, 13*(5), 432-441.

Swerdlow, C. D., Schsls, W., Dijkman, B., Jung, W., Sheth, N. V., Olson, W. H., & Gunderson, B. D. (2000). Detection of atrial fibrillation and flutter by a dual-chamber implantable cardioverter-defibrillator. For the Worldwide Jewel AF Investigators. *Circulation, 101*(8), 878-885.

Tang, E. K., Suganthan, P. N., & Yao, X. (2006). An analysis of diversity measures. *Machine Learning, 65*(1), 247-271.

References (continued)

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. . [Guideline Practice Guideline]. *Circulation, 93*(5), 1043-1065.

Theuns, D. A., Klootwijk, A. P. J., Goedhart, D. M., & Jordaens, L. J. (2004). Prevention of inappropriate therapy in implantable cardioverter-defibrillators Results of a prospective, randomized study of tachyarrhythmia detection algorithms. *Journal of the American College of Cardiology*, 44(12), 2362-2367.

Tsipouras, M. G., Fotiadis, D. I., & Sideris, D. (2005). An arrhythmia classification system based on the RR-interval signal. [Comparative Study]. *Artif Intell Med, 33*(3), 237-250.

Turcott, R. G., & Pavek, T. J. (2010). Identification of hemodynamically unstable arrhythmias using subcutaneous photoplethysmography. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *J Cardiovasc Electrophysiol, 21*(4), 448-454.

UC Irvine Machine Learning Repository. (1988). Center for Machine Learning and Intelligent Systems-Heart Disease Data Set. Retrieved 2/8/2013, from http://archive.ics.uci.edu/ml/datasets/Heart+Disease

Udupa, J. K., & Murthy, I. S. (1980). Syntactic approach to ECG rhythm analysis. *IEEE Trans Biomed Eng, 27*(7), 370-375.

Urrusti, J. L., & Tompkins, W. J. (1993). Performance evaluation of an ECG QRS complex detection algorithm. Paper presented at the Engineering in Medicine and Biology Society, 1993. Proceedings of the 15th Annual International Conference of the IEEE.

Usher, J., Campbell, D., Vohra, J., & Cameron, J. (1999). A fuzzy logic-controlled classifier for use in implantable cardioverter defibrillators. [Comparative Study]. *Pacing Clin Electrophysiol, 22*(1 Pt 2), 183-186.

Uysal, F. (2011). QRS Complex Detection and ECG Signal Processing. Retrieved 10/8/2013, from http://matlabz.blogspot.co.uk/2011/04/contents-cancellation-dc-drift-and.html

Vai, F., Bonnet, J. L., Ritter, P., & Pioger, G. (1988). Relationship between heart rate and minute ventilation, tidal volume and respiratory rate during brief and low level exercise. [Clinical Trial Randomized Controlled Trial]. *Pacing Clin Electrophysiol, 11*(11 Pt 2), 1860-1865.

Van De Water, J. M., Miller, T. W., Vogel, R. L., Mount, B. E., & Dalton, M. L. (2003). Impedance cardiography: the next vital sign technology? [Comparative Study]. *Chest, 123*(6), 2028-2033.

Wagner, G. (2001). Marriot́s practical electrocardiography 10 Ed: Wolters Kluwer Health.

Wang, Y., Zhu, Y. S., Thakor, N. V., & Xu, Y. H. (2001). A short-time multifractal approach for arrhythmia detection based on fuzzy neural network. [Comparative Study Research Support, Non-U.S. Gov't]. *IEEE Trans Biomed Eng, 48*(9), 989-995.

References (continued)

Wellens, H. J., Lau, C. P., Luderitz, B., Akhtar, M., Waldo, A. L., Camm, A. J., . . . Ayers, G. (1998). Atrioverter: an implantable device for the treatment of atrial fibrillation. [Clinical Trial Controlled Clinical Trial Multicenter Study]. *Circulation, 98*(16), 1651-1656.

Wilkoff, B. L., Kuhlkamp, V., Volosin, K., Ellenbogen, K., Waldecker, B., Kacet, S., . . . DeSouza, C. M. (2001). Critical analysis of dual-chamber implantable cardioverter-defibrillator arrhythmia detection : results and technical considerations. [Clinical Trial Evaluation Studies Multicenter Study Research Support, Non-U.S. Gov't]. *Circulation, 103*(3), 381-386.

Willems J.L., Abreu-Lima C., Arnaud P. (1991). The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med*. 325 1767-1773.

Wilson, A. J., Franks, C. I., & Freeston, I. L. (1982). Algorithms for the detection of breaths from respiratory waveform recordings of infants. *Med Biol Eng Comput, 20*(3), 286-292.

Wilson, F. N., Johnston, F. D., Macleod, A. G., & Barker, P. S. (1934). Electrocardiograms that represent the potential variations of a single electrode. *American Heart Journal*, 9(4), 447-458.

Wirtzfeld, A., Heinze, R., Liess, H. D., Stangl, K., & Alt, E. (1983). An active optical sensor for monitoring mixed venous oxygen-saturation for an implantable rate-regulating pacing system. *Pacing Clin Electrophysiol, 6*(2 Pt 2), 494-497.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks, 5*(2), 241-259.

Wurster, R. D., & McCook, R. D. (1969). Influence of rate of change in skin temperature on sweating. *Journal of applied physiology*, 27(2), 237-240.

Yang, T. F., Devine, B., & Macfarlane, P. W. (1994). Artificial neural networks for the diagnosis of atrial fibrillation. *Med Biol Eng Comput, 32*(6), 615-619.

Yu, C. M., Wang, L., Chau, E., Chan, R. H., Kong, S. L., Tang, M. O., . . . Lau, C. P. (2005). Intrathoracic impedance monitoring in patients with heart failure: correlation with fluid status and feasibility of early warning preceding hospitalization. [Research Support, Non-U.S. Gov't]. *Circulation, 112*(6), 841-848.

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer 3*(1), 32-35.

Yule, G.U. (1912). On the methods of measuring association between two attributes. *J Royal Stat. Soc., 75*(6), 581-642.

Yuskis, A. S., & Griffith, G. C. (1949). Orthostatic hypotension and orthostatic tachycardia. *Postgrad Med, 6*(6), 475-479, illust.

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. [Research Support, U.S. Gov't, P.H.S.]. *Biometrics, 44*(4), 1049-1060.

References (continued)

Zeldis, S. M., Levine, B. J., Michelson, E. L., & Morganroth, J. (1980). Cardiovascular complaints. Correlation with cardiac arrhythmias on 24-hour electrocardiographic monitoring. *Chest, 78*(3), 456-461.

Zhang, H. (2004). The optimality of naive Bayes. *In Proceedings of the FLAIRS Conference*. 1(2), 3-9.

Zhang, Q., Manriquez, A. I., Medigue, C., Papelier, Y., & Sorine, M. (2006). An algorithm for robust and efficient location of T-wave ends in electrocardiograms. [Evaluation Studies Validation Studies]. *IEEE Trans Biomed Eng, 53*(12 Pt 1), 2544-2552.

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence, 17*(5-6), 375-381.

Zhang, X. S., Zhu, Y. S., Thakor, N. V., & Wang, Z. Z. (1999). Detecting ventricular tachycardia and fibrillation by complexity measure. [Comparative Study Research Support, Non-U.S. Gov't]. *IEEE Trans Biomed Eng, 46*(5), 548-555.

Zimetbaum, P., & Josephson, M. E. (1998). Evaluation of patients with palpitations. [Review]. *N Engl J Med, 338*(19), 1369-1373.

Zipes, D. P., Camm, A. J., Borggrefe, M., Buxton, A. E., Chaitman, B., Fromer, M., ... & Zamorano, J. L. (2006). ACC/AHA/ESC 2006 Guidelines for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death. A Report of the American College of Cardiology/American Heart Association Task Force and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Develop Guidelines for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death). *Journal of the American College of Cardiology*, 48(5), e247-e346.

Zong, W., & Jiang, D. (1998). Automated ECG rhythm analysis using fuzzy reasoning. Paper presented at the Computers in Cardiology 1998.